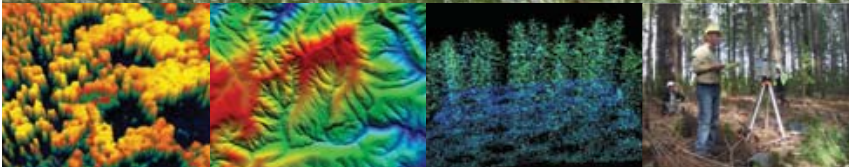


SilviLaser 2011

11th International Conference on
LiDAR Applications for Assessing
Forest Ecosystems



16 - 20 OCTOBER 2011
UNIVERSITY OF TASMANIA

CONFERENCE HANDBOOK

WELCOME

We wish to welcome you to the 2011 edition of SilviLaser. Over the past nine years the annual SilviLaser conferences have developed into a well-known meeting place for a rapidly expanding community of researchers, data providers and practitioners interested in the application of LiDAR technology towards the understanding and management of forest ecosystems. Many of you have travelled a long way to be here.

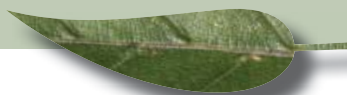
Invited speakers, researchers, authors, Organising and Scientific Committee members have worked hard to develop a strong program. The geographic, thematic and technological diversity of the talks, papers and posters demonstrates the continued vitality of the research, development and technology deployment taking place across the world. We invite you to listen to the ideas, discoveries and implementation success stories that will be presented to you over the next few days.

We also invite you to catch up with friends and colleagues, make new acquaintances, consolidate or start new collaborations. We hope the beautiful University campus and the city of Hobart will contribute to making this conference a rewarding one for you.

We thank conference and field excursion Hosts, conference Partners, Sponsors and Supporters for the assistance that made this conference possible.

Have a good time!

Organising Committee



Conference Organising Committee

Jan Rombouts – ForestrySA (Conference Convenor)

John Armston – Department of Environment and Resource Management, Queensland

Mark Brown – University of Melbourne

Nicholas Coops – University of British Columbia, Canada

Darius Culvenor –CSIRO Land and Water, Victoria

Nicholas Goodwin – Department of Environment and Resource Management, Queensland

Andrew Haywood – Department of Sustainability and Environment, Victoria (Head of the Scientific Committee)

David Mannes – Forestry Tasmania

Jon Osborn – University of Tasmania

Russell Turner – Forests New South Wales

Conference Hosts



Conference Partners



Field Trip Sponsor



Name Badge Sponsor



Conference Supporters



Conference Secretariat

Conference Design Pty Ltd
228 Liverpool Street
Hobart Tasmania 7000 Australia
Email: info@cdesign.com.au
Web: www.cdesign.com.au
Ph: 03 6231 2999
Fax: 03 6231 1522
International
Ph: +61 3 6231 2999
Fax: +61 3 6231 1522



SilviLaser 2011

Scientific Committee

- Andrew Haywood** – Department of Sustainability and Environment, Victoria
- Yasumasa Hirata** – Forestry and Forest Products Research Institute, Japan
- Erik Næsset** – Norwegian University of Life Sciences, Norway
- Matti Maltamo** – University of Eastern Finland
- Ross Nelson** – NASA, USA
- Ross Hill** – Bournemouth University, UK
- Sorin Popescu** – Texas A&M, USA
- Mike Wulder** – Canadian Forest Service, Canada
- Barbara Koch** – Freiburg University, Germany
- Benoit St-Onge** – Quebec University, Canada
- Chris Hopkinson** – Applied Geomatics Research Group, Canada
- Nicholas Coops** – University of British Columbia, Canada
- Darius Culvenor** – Commonwealth Scientific and Industrial Research Organisation, Australia
- Andrew Robinson** – University of Melbourne, Australia
- Nicholas Goodwin** – Queensland Department of Environment and Resource Management, Australia
- Ian Ferguson** – University of Melbourne, Australia
- Kim Lowell** – University of Melbourne, Australia
- Mathius Disney** – University College London, England
- Wolfgang Wagner** – Vienna, University of Technology, Austria
- Svein Solberg** – Norwegian Forest and Landscape Institute, Norway
- Hans-Erik Andersen** – US Forest Service
- Håkan Olsson** – Swedish University of Agriculture
- Johan Holmgren** – Swedish University of Agriculture
- Paul Treitz** – Queens University, Canada
- Felix Morsdorf** – University of Zurich, Switzerland
- Markus Hollaus** – University of Technology, Austria
- David Evans** – Mississippi State University, USA
- Terje Gobakken** – Norwegian University of Life Sciences
- Ralph Dubayah** – University of Maryland, USA
- Sanna Kaasalainen** – Finnish Geodetic Institute
- Cris Brack** – Waiariki Institute of Technology, New Zealand

Invited Speakers



PROFESSOR MATTI MALTAMO

University of Eastern Finland, Finland School of Forest Sciences

Presentation title: *Airborne laser scanning based stand level management inventory in Finland*



DR IAIN WOODHOUSE

School of GeoSciences, University of Edinburgh

Presentation title: *Looking forward to LiDAR's colourful future*



DR MIKE WULDER

Research Scientist, Forest Inventory and Analysis

Presentation title: *LiDAR-plots: A new wide-area data collection opportunity*



DR DAVID JUPP

CSIRO Australia

Presentation title: *Ground-based and airborne LiDAR - a natural combination*



CHRISTIAN WITTE

Manager Remote Sensing Centre, Environment and Resource Sciences, Queensland Department of Environment and Resource Management

Presentation title: *LiDAR in Australia - Progressing from digital elevation models to environmental monitoring*

SilviLaser 2011

Program

Sunday 16 October 2011	
1700-1900	Icebreaker Reception Bruni Room, Royal Yacht Club
Monday 17 October 2011	
0800	Registration Uni Centre Foyer
	Opening Session Stanley Burbury Theatre
0900-1000	Chair: Jan Rombouts, Conference Convenor Opening Address: Gordon Duff, CEO of CRC Forestry Welcome to UTAS: Professor Peter Rathjen, Vice-Chancellor Address by Forestry Tasmania representative
1000-1030	Keynote Speaker: Professor Matti Maltamo <i>Airborne laser scanning based stand level management inventory system of Finland</i> Matti Maltamo* , Petteri Packalén, Eveliina Kallio, Jyrki Kangas, Janne Uuttera, Juho Heikkilä
1030-1100	Morning Tea
	Session 1: Deploying LiDAR applications in organisations and business Chair: Jan van Aardt
1100-1115	<i>A national review of airborne LiDAR application in Australian forest agencies</i> Russell Turner, Nicholas Goodwin, Jeremy Friend, David Mannes, Jan Rombouts, Andrew Haywood
1115-1130	<i>Airborne LiDAR based forest inventory in Bangladesh for REDD plus MRV: scope and potentiality</i> Parvez Rana* , Hanna Holm, Tuomo Kauranne
1130-1145	<i>Stand level inventory of eucalypt plantations using small footprint LiDAR in Tasmania, Australia</i> Robert Musk
1145-1200	<i>Building a case for LiDAR-derived structure stratification for Australian softwood plantations</i> Russell Turner* , Amrit Kathuria, Christine Stone
1200-1300	Lunch
1300-1330	Keynote Speaker: David Jupp <i>Ground based and airborne LiDAR – a natural combination</i> Chair: Nicholas Coops
	Session 2: Wood resource assessment and value recovery Chair: Nicholas Coops
1330-1345	<i>Harvesting productivity analysis using LiDAR</i> Muhammad Alam* , Martin Strandgard, Mark Brown, Julian Fox
1345-1400	<i>Scaling plot to stand-level LiDAR to province in a hierarchical approach to map forest biomass in Nova Scotia</i> Chris Hopkinson, David Colville* , Danik Bourdeau, Suzanne Monette, Robert Maher
1400-1415	<i>Estimating stand volume from nonparametric distribution of airborne LiDAR data</i> Doo-Ahn Kwak* , Taejin Park, Jong Yeol Lee, Woo-Kyun Lee
1415-1430	<i>A method for linking TLS- and ALS-derived trees</i> Andreas Fritz* , Holger Weinacker, Barbara Koch
1430-1445	<i>Reducing extrapolation bias of area-based k-nearest neighbour predictions by using individual tree crown approaches in areas with high density airborne laser scanning data</i> Johannes Breidenbach* , Erik Næsset, Terje Gobakken
1445-1500	<i>Stem detection and measuring DBH using terrestrial laser scanning</i> Martin Van Leeuwen* Nicholas Coops, Glenn Newnham, Thomas Hilker, Darius Culvenor, Michael Wulder
1500-1530	Afternoon Tea



	Session 3: Wood resource assessment and value recovery	
	Chair: Gero Becker	
1530-1545	<i>Tree biomass estimation using ALS features</i> Minna Rätty* , Ville Kankare, Xiaowei Yu, Markus Holopainen, Mikko Vastaranta, Tuula Kantola, Juha Hyyppä, Risto Viitala	
1545-1600	<i>Stand level species classification and biomass estimation using LiDAR height, intensity, and ratio parameters</i> Taejin Park* , Doo-Ahn Kwak, Woo-Kyun Lee, Jong-Yeol Lee	
1600-1615	<i>Effect of scan coverage on stem diameter measurement using terrestrial LiDAR</i> Akira Kato* , L. Monika Moskal, Tatsuaki Kobayashi	
1615-1630	<i>Stem curve measurement using terrestrial laser scanning</i> Xinlian Liang* , Juha Hyyppä, Ville Kankare, Markus Holopainen	
1630-1645	<i>Estimating single-tree branch biomass of Norway spruce by airborne laser scanning</i> Marius Hauglin* , Janka Dibdiakova, Terje Gobakken, Erik Næsset	
1645-1700	<i>Airborne laser scanning-based stem volume imputation in a managed, boreal forest area: a comparison of estimation units</i> Jari Vauhkonen* , Petteri Packalén, Juho Pitkänen	
	Free Evening	
Tuesday 18 October 2011		
	Registration	Uni Centre Foyer
0820-0830	Housekeeping	Stanley Burbury Theatre
0830-0900	Keynote Speaker: Christian Witte <i>LiDAR in Australia – Progressing from digital elevation models to environmental monitoring.</i> Chair: Darius Culvenor	
	Session 4: Bio-diversity forest health & environmental applications	
	Chair: Darius Culvenor	
0900-0915	<i>Applying terrestrial LiDAR to derive gap fraction distribution time series during bud break</i> Kim Calders* , Jan Verbesselt, Ham Bartholomeus, Martin Herold	
0915-0930	<i>Foliage profiles from ground based waveform and discrete point LiDAR</i> Jenny Lovell* , David Jupp, Eva van Gorsel, Jose Jimenez-Berni, Chris Hopkinson, Laura Chasmer	
0930-0945	<i>Generating an automated approach to optimize effective leaf area index by Canadian boreal forest species using airborne LiDAR</i> Heather Morrison* , Chris Hopkinson, Laura Chasmer, Natascha Kljun	
0945-1000	<i>Change detection of mountain vegetation using multi-temporal ALS point clouds</i> Mattias Nyström * , Johan Holmgren, Håkan Olsson	
1000-1015	<i>Stability of LiDAR-derived raster canopy attributes with changing pulse repetition frequency</i> Allyson Fox* , Chris Hopkinson, Laura Chasmer, Ashley Wile	
1015-1030	<i>Characterizing peat swamp forest environments with airborne LiDAR data in Central Kalimantan (Indonesia)</i> Hans-Dieter Viktor Boehm* , Veraldo Liesenberg, Juergen Frank, Suwido Limin	
1030-1100	Morning Tea	
	Session 5: Bio-diversity forest health & environmental applications	
	Chair: Yasumasa Hirata	
1100-1115	<i>Comparison of the spatial pattern of trees obtained by ALS based forest inventory techniques</i> Petteri Packalén* , Jari Vauhkonen, Eveliina Kallio, Jussi Peuhkurinen, Juho Pitkänen, Inka Pippuri, Matti Maltamo	

SilviLaser 2011

1115-1130	<i>Fusion of airborne LiDAR and WorldView-2 MS data for classification of depth to permafrost within Canada's sub-Arctic</i> Laura Chasmer* , Chris Hopkinson, Heather Morrison, Richard Petrone, William Quinton
1130-1145	<i>Using high density ALS data in plot level estimation of the defoliation by the common pine sawfly</i> Tuula Kantola* , Paivi Lyytikäinen-Saarenmaa, Mikko Vastaranta, Ville Kankare, Xiaowei Yu, Markus Holopainen, Mervi Talvitie, Svein Solberg, Paula Puolakka, Juha Hyyppä
1145-1200	<i>Assessing spatial variation for tree and non-tree objects in a forest-tundra ecotone in airborne laser scanning data</i> Nadja Thieme* , Ole Martin Bollandsås, Terje Gobakken, Erik Næsset
1200-1215	<i>Exploring horizontal area-based metrics to discriminate the spatial pattern of trees using ALS</i> Inka Pippuri* , Eveliina Kallio, Matti Maltamo, Petteri Packalén, Heli Peltola
1215-1230	<i>Comparison of discrete return and waveform airborne LiDAR derived estimates of fractional cover in an Australian savanna</i> John Armston* , Mattias Disney, Philip Lewis, Peter Scarth, Peter Bunting, Richard Lucas, Stuart Phinn, Nicholas Goodwin
1230-1330	Lunch
1330-1400	Keynote Speaker: Mike Wulder <i>LiDAR-plots: A new wide-area data collection opportunity</i>
	Session 6: Large area applications Chair: Andrew Haywood
1400-1415	<i>Airborne LiDAR sampling of the Canadian boreal forest: Planning, execution & initial processing</i> Chris Hopkinson, Laura Chasmer* , Michael Wulder, Nicholas Coops, Trevor Milne, Allyson Fox, Christopher Bater
1415-1430	<i>Assessing the accuracy of GLAS topography estimation by using airborne Light Detection And Ranging (LiDAR) measurements</i> Han Meng* , Bernard Devereux, Gabriel Amable
1430-1445	<i>Characteristics of satellite LiDAR waveform in tropical rain forests from the comparison with canopy condition derived from high resolution satellite data</i> Yasumasa Hirata
1445-1500	<i>Model development for the estimation of aboveground biomass using a LiDAR-based sample of Canada's boreal forest</i> Christopher Bater* , Michael Wulder, Nicholas Coops, Chris Hopkinson, Samuel Coggins, Erik Arsenaault, André Beaudoin, Luc Guidon, R Hall, Philippe Villemaire
1500-1515	<i>Biases in an airborne profiling survey of Hedmark County, Norway</i> Ross Nelson, Erik Næsset, Liviu Ene, Göran Ståhl, Timothy Gregoire
1515-1545	Afternoon Tea
	Session 7: Poster presentations Chair: Johannes Breidenbach
1545-1550	<i>Early assessment of industrial needs: harvesting and allocation decisions supported by ALS and TLS</i> Gero Becker* , Thomas Smaltschinski, Martin Opferkuch, Holger Weinacker
1550-1555	<i>Remotely sensed crown structure as an indicator of wood quality: A comparison of metrics from aerial and terrestrial laser scanning</i> Thomas Adams* , David Pont, Jonathan Harrington
1555-1600	<i>Developing LiDAR interpretation software for wood resource inventory in Forests NSW</i> Russell Turner* , A Farjad, J Trinder, S Lim
1600-1605	<i>Towards automated and operational forest inventories with T-LiDAR</i> Othmani Ahlem* , Piboule Alexandre, Krebs Michael, Stolz Christophe, Lew-yan-voon Lew
1605-1610	<i>3-D modelling of forest structure for parameterization of radiative transfer models</i> Martin Van Leeuwen* , Nicholas Coops, Glenn Newnham, Thomas Hilker, Darius Culvenor, Michael Wulder



1610-1615	<i>Evaluation of nonlinear equations for predicting diameter from tree height for Pinus radiata (D. Don) in an airborne laser scanning-based plantation inventory</i> Huiquan Bi* , Julian Fox, Yun Li, Yuancai Lei, Yong Pang
1615-1620	<i>Revisiting the status of space-borne LiDAR missions for assessing structural and biophysical forest parameters in the context of sustainable management of Earth resources</i> Sylvie Durrieu* , Ross Nelson
1620-1625	<i>Vegetation classification in the Swedish sub-arctic using a combination of optical satellite images and airborne laser scanner data</i> Mattias Nyström* , Karin Nordkvist, Heather Reese, Johan Holmgren, Håkan Olsson
1625-1630	<i>LiDAR data and cooperative research at Panther Creek, Oregon</i> James Flewelling* , George McFadden
1630-1635	<i>LiDAR estimation of quadratic mean canopy height and stem density in native sclerophyll forests</i> Yadav Prasad Kandel* , Julian Fox, Stefan Arndt, Stephen Livesley
1635-1640	<i>Using a flux footprint model and airborne LiDAR to characterize vegetation structure and topography frequently sampled by Eddy Covariance: Implications for MODIS product validation</i> Laura Chasmer* , N Kljun, Chris Hopkinson, S Brown, T Milne, K Giroux, A Barr, K Devito, I Creed, Richard Petrone
1640-1645	<i>Satellite vs. airborne LiDAR estimates of aboveground biomass and forest structure metrics at footprint scale</i> Sorin Popescua , Kaiguang Zhaoa, Amy Neuenschwanderb, Chinsu Linc
1645-1650	<i>The significance of managed and natural vegetation on house survival during wildfires</i> Anders Siggins* Glenn Newnham, Raphaele
1730	Coaches depart UTAS for MONA
1800-1900	Tour of Award-winning Museum of Old and New Art
1900-2300	Conference Dinner Moorilla (coaches will return delegates to their accommodation venues)
Wednesday 19 October 2011	
0800	Registration
0845-0900	Housekeeping
0900-0930	Keynote Speaker: Iain Woodhouse <i>Looking forward to LiDAR's colourful future</i> Chair: Jenny Lovell
Session 8: Emerging Technologies Chair: Jenny Lovell	
0930-0945	<i>Error assessment and mitigation for hyper-temporal UAV-borne LiDAR surveys of forest inventory</i> Luke Wallace* , Arko Lucieer, Darren Turner, Christopher Watson
0945-1000	<i>A new photon counting LiDAR system for vegetation analysis</i> Jaqueline Rosette* , Christopher Field, Ross Nelson, Phil DeCola, Bruce Cook
1000-1015	<i>Sorted Pulse Data (SPD) format: A new file structure for storing and processing LiDAR data</i> Peter Bunting* , John Armston, Daniel Clewley, Richard Lucas
1015-1030	<i>Tree detection, delineation, and measurement from LiDAR point clouds using RANSAC</i> Peter Tittmann* , Sohail Shafii*, Bruce Hartsough, Bernd Hamann
1030-1045	<i>Mobile terrestrial laser scanning in urban tree inventory</i> Mikko Vastaranta , Tuula Kantola*, Markus Holopainen, Ville Kankare, Harri Kaartinen, Antero Kukko, Matti Vaaja, Juha Hyypä, Hanna Hyypä
1045-1115	Morning Tea

SilviLaser 2011

	Session 9: New Methods and Algorithms Chair: Petteri Packalén
1115-1130	<i>Another dimension from LiDAR - Obtaining foliage density from full waveform data</i> Thomas Adams* , Peter Beets, Christopher Parrish
1130-1145	<i>The Sorted Pulse Data Software Library (SPDLib): Open source tools for processing LiDAR data</i> Peter Bunting* , John Armston, Daniel Clewley, Richard Lucas
1145-1200	<i>Comparison of point cloud data reduction methods in single-scan TLS for finding tree stems in forest</i> Paula Litkey, Puttonen Eetu, Liang Xinlian*
1200-1215	<i>Automated log counting: Proof of concept algorithm</i> Hamish Marshall
1215-1230	<i>Optimal LiDAR gridding parameterization for effective leaf area estimation in the boreal forest Yukon Territory, Canada</i> Heather Morrison* , Chris Hopkinson, Michael Wulder
1230-1330	Lunch
	Session 10: Fire and Water Chair: Jacqueline Rosette
1330-1345	<i>Developing a regional canopy fuels assessment strategy using multi-scale LiDAR</i> Birgit Peterson* , Kurtis Nelson
1345-1400	<i>LiDAR-based estimation of forest floor fuel loads using a novel distributional approach</i> Jan van Aardt* , Mary Arthur, Gretchen Sovkoplak, Tyson Lee Swetnam
1400-1415	<i>Using airborne survey to map stream form and riparian vegetation characteristics across Victoria</i> Nathan Quadros* , Rick Frisina, Paul Wilson
1415-1430	<i>Full waveform LiDAR for assessment of river health</i> David Moore* , Alys Wall, Thomas Hollaus
1430-1500	Afternoon Tea
1500-1600	Panel Discussion Moderator: Nicholas Coops
1600-1630	Wrap Up Session Sorin Popescu
1630-1645	SilviLaser 2012 Handover "First Return"
1645	Conference Close
	Free Evening

Exhibition

Allocation of Booths

01	ESRI		05	KEOPSYS – FIBRE LASER	
02	FUGRO SPATIAL SOLUTIONS		06	AAM	
03	TERRANEAN		07	PHOTOMAPPING	
04	RIEGL				

Posters

LiDAR assisted wood resource

- 1 *Early assessment of industrial needs: harvesting and allocation decisions supported by ALS and TLS*
Gero Becker*, Thomas Smaltschinski, Martin Opferkuch, Holger Weinacker
- 2 *Remotely sensed crown structure as an indicator of wood quality: A comparison of metrics from aerial and terrestrial laser scanning*
Thomas Adams*, David Pont, Jonathan Harrington
- 3 *Developing LiDAR interpretation software for wood resource inventory in Forests NSW*
Russell Turner*, A Farjad, J Trinder, S Lim

Deploying LiDAR applications

- 4 *Gearing toward the potential of LiDAR application in Malaysian forestry*
Mohd Hasmadi Ismail
- 5 *Towards automated and operational forest inventories with T-LiDAR*
Othmani Ahlem*, Piboule Alexandre, Krebs Michael, Stolz Christophe, Lew-yan-voon Lew
- 6 *Laser scanning by echo signal digitization and waveform processing*
Martin Pfennigbauer*, Andreas Ullrich*

New methods and algorithms

- 7 *Crown coverage calculation based on ALS data*
Lothar Eysn*, Markus Hollaus, Klemens Schadauer, Andreas Roncat
- 8 *3-D modelling of forest structure for parameterization of radiative transfer models*
Martin Van Leeuwen*, Nicholas Coops, Glenn Newnham, Thomas Hilker, Darius Culvenor, Michael Wulder
- 9 *Evaluation of nonlinear equations for predicting diameter from tree height for Pinus radiata (D. Don) in an airborne laser scanning-based plantation inventory*
Huiquan Bi*, Julian Fox, Yun Li, Yuancai Lei, Yong Pang

Large area mapping and assessment

- 10 *Revisiting the status of space-borne LiDAR missions for assessing structural and biophysical forest parameters in the context of sustainable management of Earth resources*
Sylvie Durrieu*, Ross Nelson
- 11 *Vegetation classification in the Swedish sub-arctic using a combination of optical satellite images and airborne laser scanner data*
Mattias Nyström*, Karin Nordkvist, Heather Reese, Johan Holmgren, Håkan Olsson
- 12 *LiDAR data and cooperative research at Panther Creek, Oregon*
James Flewelling*, George McFadden
- 13 *Satellite vs. airborne LiDAR estimates of aboveground biomass and forest structure metrics at footprint scale*
Sorin Popescua, Kaiguang Zhaoa, Amy Neuenschwanderb, Chinsu Linc

Biodiversity, forest health & environmental applications

- 14 *LiDAR estimation of quadratic mean canopy height and stem density in native sclerophyll forests*
Yadav Prasad Kandel*, Julian Fox, Stefan Arndt, Stephen Livesley
- 15 *Modelling light conditions in forests using airborne laser scanning data*
Werner Mücke, Markus Hollaus
- 16 *Using a flux footprint model and airborne LiDAR to characterize vegetation structure and topography frequently sampled by Eddy Covariance: Implications for MODIS product validation*
Laura Chasmer*, N Kljun, Chris Hopkinson, S Brown, T Milne, K Giroux, A Barr, K Devito, I Creed, Richard Petrone
- 17 *Estimation of Leaf Area Index based on airborne laser scanning and imaging spectroscopy*
Pyare Pueschel, Henning Buddenbaum, Joachim Hill
- 18 *Satellite vs. airborne LiDAR estimates of aboveground biomass and forest structure metrics at footprint scale*
Sorin Popescua, Kaiguang Zhaoa, Amy Neuenschwanderb, Chinsu Linc

Fire Management

- 19 *The significance of managed and natural vegetation on house survival during wildfires*
Anders Siggins* Glenn Newnham, Raphaele

Airborne laser scanning based stand level management inventory in Finland

Maltamo, M.,¹ Packalén, P.,¹ Kallio, E.,¹ Kangas, J.,² Uuttera, J.,³ and Heikkilä, J.⁴

¹ University of Eastern Finland, School of Forest Sciences, matti.maltamo@uef.fi,
petteri.packalen@uef.fi, eveliina.kallio@uef.fi

² Metsähallitus, jyrki.kangas@metsa.fi

³ UPM Kymmene Oyj, janne.uuttera@upm-kymmene.com

³ Forestry Development Centre TAPIO, juho.heikkila@tapio.fi

Abstract

In Finland, a new ALS based stand level management inventory was developed during last few years. The system is based on area based approach of ALS data. Additionally, the spectral and texture features of the aerial images are utilized in order to improve the separation of the tree species. The species-specific stand attributes are simultaneously estimated with a nearest neighbour imputation. The new airborne laser scanning based stand level management inventory system has been successful. During just a few years almost all actors of practical forestry have modified their inventory and planning systems to be compatible with the new inventory procedure which will cover almost 3 000 000 hectares in 2011. This paper describes the background, development and practical application of this inventory system.

Keywords: area based approach, nearest neighbor imputation, operational inventory, species-specific stand attributes

1. Background

In Finland two main forest inventories are sampling based National Forest Inventory (NFI) for large-scale and inventory by compartments for stand level management. When NFI data are combined with remote sensing data and other auxiliary information in multi-source NFI, reliable estimates for small areas can also be obtained (Tomppo 2006). However, the areas considered in multi-source NFI are still considerably larger than one stand and also the information needs of stand level management inventory are different. From the perspective of practical forestry the accuracy requirement for the stand level inventory is about 15–30% RMSE in stand volume (Uuttera *et al.* 2002).

Traditionally, the information for stand level management has been collected with a stand-wise field inventory method, that is, inventory by compartments, in which species-specific forest characteristics are estimated using subjective angle count sampling and partly visual assessment (e.g. Koivuniemi and Korhonen, 2006). This method includes stand delineation from aerial photographs, field visits to each stand, and calculation of stand attributes of interest, mainly volume by tree species and timber assortments. The stand characteristics assessed in the field include age, basal area, mean diameter, and height. These data are used in forest planning, for example, to determine the need for silvicultural operations. In practical forestry the data acquisition costs are about 10 euro per hectare and the costs of the whole forest planning process are over 20 euro. However, in private forestry this process is highly subsidized by the state, which pays about 70% of the cost. Annually this method has been applied to over 1 000 000 million hectares of private forests and additionally also to considerable areas of state and forest company forests.

Inventory by compartments has been applied since the 1950s, which means that practically the whole country has been inventoried several times. During the last decades the development of this inventory system was related to field measurements and calculation routines (e.g. Kilkki and Päivinen 1986; Kangas *et al.* 2004). During the last 15 years there has, however, been strong emphasis on modernizing this method completely. The main reasons for this development are the high costs, subjectivity, and inaccuracy of the basic method. In the 1990s there was already a lot of research concerning the development of this method towards remote sensing applications (e.g. Päivinen *et al.* 1993; Varjo 2002). However the accuracy demands and usability for operational purposes were not fulfilled by different optical imageries. Usually, the RMSE of the stand level RMSE of total volume exceeded 30% (Hyypä *et al.* 2000; Uuttera *et al.* 2006). It was also difficult to separate tree species and the heterogeneity between images was an issue.

The situation concerning the usability of remote sensing data changed when airborne laser scanning (ALS) data became available. For example in Norway the first studies already showed the accuracy and indicated the usability of these data in the operational stage (Næsset 1997; 2004) The method developed in Norway is based on the variables calculated from the height value distribution of the low pulse density ALS data over a certain area and is, therefore, called an area based approach (or canopy height distribution approach) (see Næsset 2004). In Finland the application of ALS data began with the single tree detection approach, where individual trees were recognized from the canopy height model constructed from high pulse density ALS data (Hyypä and Inkinen 1999). The accuracy of this approach was also already promising in the first studies (Hyypä and Inkinen 1999; Maltamo *et al.* 2004) but high data acquisition costs, a lack of algorithms to detect tree species, and considerable threat of bias of tree and stand attribute estimates restricted the development of this method towards the operational stage.

2. Airborne laser scanning based stand level management inventory system of Finland

2.1 System description

Correspondingly, as in other Nordic countries (Næsset 2002; Holmgren 2004) and later in many other countries (Hudak *et al.* 2006; Jensen *et al.* 2006; Hollaus *et al.* 2007; Rombauts *et al.* 2008; Latifi *et al.* 2010), the area based approach was also tested experimentally in Finland starting in 2004 (Suvanto *et al.* 2005; Maltamo *et al.* 2006). The accuracy obtained for stand total volume was superior compared to earlier studies based on either field measurements (Haara and Korhonen 2004) or other remote sensing data (see e.g. Uuttera *et al.* 2006). However, the requirement for estimated species-specific stand attributes was not fulfilled by this approach either.

To overcome the tree species problem, Packalén and Maltamo (2006; 2007; 2008) combined ALS data with aerial images. As in the other applications, here also the independent variables of the system are those calculated from the height and the density distributions of the low-resolution (pulse density $< 1 \text{ pulse}\cdot\text{m}^{-2}$) ALS data. Additionally, the spectral statistics and the texture metrics of the aerial images are utilized in order to improve the separation of the tree species. The fusion of ALS data and aerial images was further developed in Packalén *et al.* (2009) in order to improve the accuracy of species-specific predictions. The dependent variables of the system are the most essential stand sum and mean attributes, namely volume, basal area, number of stems, mean diameter, and mean height estimated separately for Scots pine, Norway spruce, and tree species group deciduous species.

Since the modelling phase is multivariate the stand attributes are simultaneously estimated with a nearest neighbour (NN) imputation (see Packalén and Maltamo 2007). However, other modelling alternatives such as the Bayesian approach are also possible (Junttila *et al.* 2008). The obtained results have indicated highly accurate results for the stand totals and also for main tree species whereas the accuracy is worse for minor tree species. The chosen non-parametric approach also allows the estimation of species-specific diameter distributions which are compatible with stand attribute estimates (Packalén and Maltamo 2008). These diameter distribution estimates are based on tree diameter measurements of reference plots and, thus, describe the local variability. Alternatively, it is also possible to predict parameters of some theoretical diameter distribution models by using ALS data or to utilize predicted stand attributes and existing parameter models of theoretical diameter distribution models.

Accurate and georeferenced reference plot measurements are a keystone in the new stand level management inventory method (see e.g. Gobakken and Næsset 2009). Georeferencing enables the extraction of the aerial data from exactly the same point as where the field measurements were carried out. For now the number of field reference plots is about 500 plots within the area of each inventory campaign. The ALS based inventory concerns young, maturing, and mature forests but basically seedling stands have been out of the scope so far. The reference plot measurements should represent the existing variation within the inventory area. For the NN method, the limitation is that it is impossible to use it for extrapolation purposes: if the data are not representative, the lowest values of any distribution will be overestimated, while the highest estimates will be underestimated. The multivariate modelling task makes this even more complex since some of the attributes estimated are rare. To capture the true variation within the forest area, the placement of the reference plot measurements should be considered carefully. Usually, the field plot data are not a probability sample in Finland.

When applying the constructed model the stand attributes are estimated by means of plot level reference measurements and aerial data metrics in a wall-to-wall manner. The independent variables are calculated for the field reference plots and for the cells of a 16 m × 16 m grid, which is laid over the inventory area. The cells of the grid are used as estimation units, for which the forest characteristics are imputed from the reference plots by means of aerial data. The use of a grid corresponds with the Norwegian application (Næsset 2004) and more or less similar systematic approaches have been proposed for remote sensing applications in general for stand level inventories (e.g. Poso 1994). Stand level estimates are aggregated from the grid cells that fall inside the boundaries of each stand. Instead of using grids, micro-stands may be applied as well (e.g. van Aardt *et al.* 2006). The reasoning behind the use of micro-stands is that they divide the area into homogenous parts where the prediction of stand attributes may be more reliable. ALS data also provide excellent possibilities for segmentation.

2.2 Practical arrangements

Since scientific studies have shown that the species-specific estimation accuracy obtained with the ALS based inventory is comparable with the traditional field inventory method (Haara and Korhonen 2004; Maltamo *et al.* 2009; Packalén and Maltamo 2007), the actors in the Finnish forest sector were willing to adopt the new, accurate, and less fieldwork-intensive inventory procedure in the hope of reduced costs. In the case of private forestry it is assumed that a 60% cost saving would occur when compared to field inventory. The scientific basis for the new system was developed in the projects “The Use of Airborne Laser Scanning in the Estimation of Accurate Forest Resources” and “The Use of Airborne Laser Scanning and Aerial Photographs in the Inventory of Timber Sortiments by Tree Species” funded by TEKES, the National Technology Agency of Finland (Maltamo 2007; Maltamo and Kallio 2011). The practical forestry organizations were also involved in these projects.

The change from the old system to the new one was surprisingly fast. In Finland there is no similar tradition of using remote sensing data in stand level inventory as in other Nordic countries. In Finland the use of remote sensing data was earlier restricted to visual stand delineation. Usually there is also a conservative attitude towards changes but in this case the reason for the rapid movement might be twofold. Firstly, there was already a long history of finding new remote sensing based solutions, but suitable data were not found. Secondly, there was a close co-operation between researchers and actors of practical forestry when this system was developed.

In practical applications the ALS and aerial data acquisition, processing of the raw ALS data and aerial images into the independent variables, stand attribute modelling, and calculation of inventory results are done by service providers but the process differs between organizations. In the following the practices of three different organizations are presented.

The forest company UPM Kymmene Oyj started its pilot projects in 2004 and the first fully operational project was in 2008. Currently about half of its 900 000 hectares forest area in Finland has been inventoried by using an ALS application. In UPM Kymmene Oyj micro-stands are applied since their usability is better when planning silvicultural operations. Typically, service providers also collect field reference data as part of the campaign. The situation is almost similar in the state forests managed by Metsähallitus. ALS based forest inventory is already in the operational stage in forest planning systems. So far, about 1 200 000 hectares have been inventoried. There are also other purposes for the data, such as planning the cutting of the marked stands and the need for ditch cleaning.

In the case of private forestry the plan is that the total area to be inventoried with the new method in Finland is about 1.5 million hectares per year, which means that all the private forests will be inventoried during less than one decade. In the case of privately owned forests, the estimated stand level forest characteristics are often further processed to holding-specific forest plans and treatment schedules. The results of the inventory are also used for guidance of forest owners (so-called *Metsaan.fi*). Until now, the regional Forestry Centres, of which there are 13 altogether, have been the major institution that offers forest planning services to the private forest owners. After some pilot studies conducted during the last five years the first practical forest inventories in privately owned forests with the new inventory method were done in 2010 when local Forestry Centres organized bidding in order to conduct inventory of certain areas.

The difference between other organizations and private forestry is that these centres measure the reference plot sample in the privately owned forests with the new inventory method. The reference measurements are carried out independently within the area of each regional Forestry Centre and the intention is that each centre would measure about 500–800 circular sample plots from the mature forests in their inventory area. A set of sample plots is also measured from the sapling and seedling stands. The forest planning experts who previously assessed stand level forest characteristics by means of angle count sampling now measure accurate information on georeferenced sample plots. They also concentrate more on planning, guidance, and service for forest owners. Besides the reference plot measurements, fieldwork is still needed for checking the treatment proposals for seedlings that cannot be estimated accurately enough with the new method. There is also a need for some forest classification information (site class, soil class, etc.) on age and biodiversity issues where information must be taken from old inventory data or must be field checked.

The sample plot placement carried out by the Forestry Centres attempts to mimic the NFI with varying cluster and shorter plot distances, stratification, and subjectively allocated additional measurements. The stratification is performed on the basis of the old inventory data on the forest characteristics and the locations of different stands within the inventory area. The strata

are forest site type, dominating tree species, basal area, and mean diameter. This design aims to obtain a good non-probability sample of the forests of the inventory area that includes the true variations and also the extremes of all the variable distributions.

2.3. Specific features

Besides the basic work of the development of the new inventory system many other issues have been tested experimentally as well. Since the NFI also provides a systematic network of field plots, the question of whether this data would also be utilized in stand level management inventory was studied (Maltamo *et al.* 2009). Concerning NFI plots it must be remembered that the plots are angle count plots, which means that 100% coverage with remote sensing data is impossible. According to the results by Maltamo *et al.* (2009) the effect of using angle count samples on accuracy is, however, minor but the current georeferencing of the plots is not adequate and also the sampling design is not optimal for the stand level inventory purposes.

The National Land Survey of Finland is using ALS data for national DTM production (Ahokas *et al.* 2008). The technical requirements for the ALS data in national terrain modelling are almost identical to the requirements of forestry applications but in terrain modelling the primary aim is to acquire leaf-off data. However, using the same (leaf-off) data in national terrain modelling and forestry applications would mean significant cost savings in both campaigns. The usability of leaf-off ALS data in stand level forest inventory was examined by Villikka (2010). The overall conclusion was that leaf-off ALS data are suitable for area based forest inventory in which deciduous and coniferous trees need to be separated. However, the narrow time window when leaf-off ALS data can be collected may restrict the applicability. Correspondingly, in DTM production the laser scanner used may change during one campaign, but this is a critical point for forestry applications (Næsset 2009). In general the results of Villikka (2010) were better with leaf-off than leaf-on data. In addition, leaf-off ALS data per se had the ability to discriminate between deciduous and coniferous trees, which may decrease the inventory costs if the acquisition of aerial images is therefore avoided entirely and if there is a possibility of joint ALS data acquisition between forestry and land survey organizations. Nowadays leaf-off data are already widely applied in private forestry data acquisition in Finland. Altogether about 1 000 000 hectares will be inventoried by leaf-off data in 2011.

One specific feature is also young sapling stands, which usually cover about 25% of the inventory area. Considerable savings of the costs of field checks which are currently a bottleneck in the system could be avoided if at least part of these stands could be covered by ALS inventory. The information needs of these stands differ. Instead of species-specific stand attributes it is more important to know the timing of the next silvicultural treatment. The developed ALS based inventory as such will not provide such information although it can be expanded to cover seedlings as well. There have been some efforts to predict characteristics of young stands or alternatively the need for treatments (Næsset and Bjerknes 2001; Korpela *et al.* 2008; Närhi *et al.* 2008). Research concerning the inclusion of seedling stands as a part of ALS based stand level inventory is currently going on in Finland.

One important issue is the bioenergy content of forests. Like volume, biomass can also be predicted by using ALS. If the ALS based characterization of a stand is at tree level, biomass components (stem, branches, stump) can also be calculated (Kotamaa *et al.* 2010). It is also possible to find stands with the possibility of bioenergy cutting or other silvicultural thinning and to characterize the amount of removed biomass or logging residues in thinnings (Kotamaa *et al.* 2010; Pyörälä 2010; Räsänen 2010; Vastaranta *et al.* 2011).

3. The future

The new airborne laser scanning based stand level management inventory system has been successful in Finland. During just a few years almost all actors of practical forestry have modified their inventory and planning systems to work with ALS data and the inventory will cover almost 3 000 000 hectares in 2011. Nowadays the area based approach is applied but due to the use of diameter distribution models the characterization of tree stock can be transformed to tree level. Although this inventory system is still quite new there is already a lot of research work related to the development of this system or completely new systems.

In relation to the current system, the most serious drawback is still the separation of tree species. Although the main tree species can usually be described, the error in the case of minor tree species can still be very high. If the estimated main tree species of a stand is wrong this is a severe error for a holding-specific forest plan. Development related to this issue might be related to improved algorithms or other data sources such as hyperspectral data, whose operational use is, however, still questionable. Other big issues are the characterization of multilayered and seedling stands as well as information on biodiversity aspects and site classes.

The operational application of ALS data also still has many bottlenecks which are not directly related to the ALS technique. For example, co-operation between different organizations (e.g. ALS raw data processing and field measurements), rapid changes in established practices in organizations, restrictions of information systems, different local conditions in different parts of the country, and weather conditions for optical image acquisition cause difficulties in operational use. For example, the linkage of estimated stand attributes or diameter distributions and forest planning systems might need some improvement.

If the whole inventory approach is changed in the near future one question might be whether single tree detection would be applicable for operational purposes. During the last five years the prices of ALS data have gone down and there has been development of the algorithms to predict tree species and stand attributes (e.g. Korpela *et al.* 2010; Vauhkonen *et al.* 2010). As a result of this development the differences in accuracy between these two approaches have more or less vanished (e.g. Packalén *et al.* 2008; Vastaranta *et al.* 2009; Peuhkurinen *et al.* 2011). One interesting alternative is the so-called semi-individual tree crown approach (Breidenbach *et al.* 2010), where segments including zero, one, or several trees are imputed with non-parametric methods, thus avoiding bias in the resulting stand level estimates.

If the possible change from an area based approach to single tree detection is discussed there are different views. Changing the inventory system from one remote sensing system to another might be easier now since the first remote sensing based inventory system is now in operational use. On the other hand, if no considerable improvement in the description of tree stock or in the accuracy were to be achieved, the change might be needless from an operational point of a view. There might also be some new difficulties related to information systems. Additionally, although single tree detection can, in principle, characterize each tree it is common for certain trees or tree groups not to be observed. This might happen especially in clustered stands where this kind of information is crucial from the silvicultural point of view, and if this information is not obtained the benefit of single tree detection is lost.

The development of inventory systems is, of course, not only related to the choice between area based and single tree detection approaches. There are different types of ALS data, such as full waveform data, and even different platforms. Their usability in different types of inventory varies and in the case of stand level management inventory these issues are still an open question. It should also be remembered that to be applicable in practice their operational use should cover very large areas annually.

References

- Ahokas, E., Kaartinen, H., and Hyypä, J., 2008. On the quality checking of the airborne laser scanning-based nationwide elevation model in Finland. 21st ISPRS Congress Beijing 2008. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(B1/I), 267–270.
- Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., and Solberg, S., 2010. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment*, 114 (4), 911–924.
- Gobakken, T. and Næsset, E., 2009. Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research*, 39 (5), 1036–1052.
- Haara, A. and Korhonen, K.T., 2004. Kuvioittaisen arvioinnin luotettavuus. *Metsätieteen aikakauskirja*, 4/2004, 489–508. (In Finnish.)
- Hollaus M., Wagner W., Maier B., and Schadauer K., 2007. Airborne laser scanning of forest stem volume in a mountainous environment. *Sensors*, 7, 1559–1577.
- Holmgren, J., 2004. Prediction of tree height, basal area and stem volume using airborne laser scanning. *Scandinavian Journal of Forest Research*, 19, 543–553.
- Hudak, A.T., Crookston, N.L., Evans, J.S., Falkowski, M.J., Smith, A.M.S., Gessler, P.E., and Morgan, P., 2006. Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. *Canadian Journal of Remote Sensing*, 32, 126–138.
- Hyypä, J. and Inkinen, M., 1999. Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland*, 16, 27–42.
- Hyypä, J., Hyypä, H., Inkinen, M., Engdahl, M., Linko, S., and Zhu, Y.H., 2000. Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. *Forest Ecology and Management*, 128, 109–120.
- Jensen, J.L.R., Humes, K.S., Conner, T., Williams, C.J., and DeGroot, J., 2006. Estimation of biophysical characteristics for highly variable mixed-conifer stands using small-footprint lidar. *Canadian Journal of Forest Research*, 36, 1129–1138.
- Junttila, V., Maltamo, M., and Kauranne, T., 2008. Sparse Bayesian estimation of forest stand characteristics from ALS. *Forest Science*, 54, 543–552.
- Kangas A., Heikkinen E., and Maltamo M., 2004 Accuracy of partially visually assessed stand characteristics – A case study of Finnish forest inventory by compartments. *Canadian Journal of Forest Research*, 34, 916–930.
- Kilkki, P. and Päivinen, R., 1986. Weibull function in the estimation of the basal area DBH-distribution. *Silva Fennica*, 20, 149–156.
- Koivuniemi, J. and Korhonen, K.T., 2006. Inventory by compartments. In: A. Kangas and M. Maltamo (Eds.). *Forest Inventory. Methodology and Applications. Managing Forest Ecosystems*. Vol. 10. Springer, Dordrecht: 271–278.

- Korpela I., Tuomola T., Tokola T., and Dahlin B., 2008. Appraisal of seedling-stand vegetation with airborne imagery and discrete-return LiDAR – an exploratory analysis. *Silva Fennica*, 42 (5), 753–772.
- Korpela, I., Ørka, H.-O., Maltamo, M., Tokola, T., and Hyypä, J., 2010. Tree species classification in airborne LiDAR data: influence of stand and tree factors, intensity normalization and sensor type. *Silva Fennica*, 44, 319–339
- Kotamaa, E., Tokola, T., Maltamo, M., Packalén, P., Kurttila, M., and Mäkinen, A., 2010. Integration of remote sensing-based bioenergy inventory data and optimal bucking for stand-level decision making. *European Journal of Forest Research*, 129, 875–886.
- Latifi, H., Nothdurft, A., and Koch, B., 2010. Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. *Forestry*, 83 (4), 395–407.
- Maltamo, M., 2007. *Laserkeilauksen käyttö metsävarojen tarkassa inventoinnissa (METSÄLASER)*. Loppuraportti. 15 p. (In Finnish.)
- Maltamo, M. and Kallio, E., 2011. *Laserkeilauksen ja ilmakuvien käyttö metsävarojen puu- ja puutavaralajeittaisessa inventoinnissa (METSÄLASER2)*. Loppuraportti. 21 p. (In Finnish.)
- Maltamo, M., Eerikäinen, K., Pitkänen J., Hyypä, J., and Vehmas, M., 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote Sensing of Environment*, 90, 319–330.
- Maltamo, M., Eerikäinen, K., Packalén, P., and Hyypä, J., 2006. Estimation of stem volume using laser scanning based canopy height metrics. *Forestry*, 79, 217–229.
- Maltamo, M., Packalén, P., Suvanto, A., Korhonen, K.T., Mehtätalo, L., and Hyvönen, P., 2009. Combining ALS and NFI training data for forest management planning – a case study in Kuortane, Western Finland. *European Journal of Forest Research*, 128, 305–317.
- Næsset, E., 1997. Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sensing of Environment*, 51, 246–253.
- Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80, 88–99.
- Næsset, E., 2004. Practical large-scale forest stand inventory using a small airborne scanning laser. *Scandinavian Journal of Forest Research*, 19, 164–179.
- Næsset, E., 2009. Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, 113, 148–159
- Næsset, E. and Bjerknes, K-O., 2001. Estimating tree heights and number of stems – in young forest stands using airborne laser scanner data. *Remote Sensing of Environment*, 78, 328–340.
- Närhi, M. Maltamo, M, Packalén, P., Peltola, H., and Soimasuo, J., 2008. Kuusen taimikoiden inventointi ja taimikonhoidon kiireellisyyden määrittäminen laserkeilauksen ja metsäsuunnitelmätietojen avulla. *Metsätieteen aikakauskirja*, 1/2008, 5–15. (In Finnish.)

- Packalén, P. and Maltamo, M., 2006. Predicting the volume by tree species using airborne laser scanning and aerial photographs. *Forest Science*, 52, 611–622.
- Packalén, P. and Maltamo, M., 2007. The k-MSN method in the prediction of species specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment*, 109, 328–341.
- Packalén, P. and Maltamo, M., 2008. The estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. *Canadian Journal of Forest Research*, 38, 1750–1760.
- Packalén, P., Pitkänen, J., and Maltamo, M., 2008. Comparison of individual tree detection and canopy height distribution approaches: a case study in Finland. In: R. Hill, J. Rosette, and J. Suárez (Eds.), *Proceedings of SilviLaser 2008, 8th International Conference on LiDAR Applications in Forest Assessment and Inventory*, 17–19 September 2008. Heriot-Watt University, Edinburgh, UK: 22–29.
- Packalén, P., Suvanto, A., and Maltamo, M., 2009. A two stage method to estimate species-specific growing stock by combining ALS data and aerial photographs of known orientation parameters. *Photogrammetric Engineering & Remote Sensing*, 75, 1451–1460.
- Päivinen, R., Pussinen, R., and Tomppo, E., 1993. Assessment of boreal forest stands using field assessment and remote sensing. In: *Operalization of Remote Sensing. Proceedings of Earsel 1993 Conference*, 19–23 April 1993. ITC Enshedene, the Netherlands. 8 p.
- Peuhkurinen, J., Mehtätalo, L., and Maltamo, M., 2011. Comparing ALS based individual tree and area based methods in forest inventory. *Canadian Journal of Forest Research*, 41, 583–598.
- Poso, S., 1994. Metsätalouden suunnittelu uusiin puihin. Voidaanko silmävaraisesta kuvioittaisesta arvioinnista luopua? *Metsätieteen aikakauskirja*, 1/1994.
- Pyörälä, P., 2010. *Bioenergiakohteiden tunnistaminen laserkeilauksella*. University of Eastern Finland, Faculty of Science and Forestry, candidate's thesis. 20 p. (In Finnish.)
- Räsänen, I., 2010. *Puuston ensiharvennustarpeen ja tilajärjestyksen määrittäminen laserkeilauksella*. University of Eastern Finland, Faculty of Science and Forestry, master's thesis. 58 p. (In Finnish.)
- Rombouts, J., Ferguson, I.S., and Leech, J.W., 2008. Variability of LiDAR volume prediction models for productivity assessment of radiata pine plantations in South Australia. In: R. Hill, J. Rosette, and J. Suárez (Eds.), *Proceedings of SilviLaser 2008, 8th International Conference on LiDAR Applications in Forest Assessment and Inventory*, 17–19 September 2008. Heriot-Watt University, Edinburgh, UK: 39–49.
- Suvanto, A., Maltamo, M., Packalén, P., and Kangas, J., 2005. Kuviokohtaisten puustotunnusten ennustaminen laserkeilauksella. *Metsätieteen aikakauskirja*, 4/2005, 413–428. (In Finnish.)
- Tomppo, E., 2006. The Finnish multi-source National Forest Inventory – small area estimation and map production. In: A. Kangas and M. Maltamo (Eds.). *Forest Inventory. Methodology and Applications. Managing Forest Ecosystems*. Vol. 10, Springer, Dordrecht: 195–224.
- Utterä, J., Hiltunen, J., Rissanen, P., Anttila, P., and Hyvönen, P., 2002. Uudet kuvioittaisen arvioinnin menetelmät – arvio soveltuvuudesta yksityismaiden metsäsuunnitteluun. *Metsätieteen aikakauskirja*, 3/2002, 523–531. (In Finnish.)

- Uuttera, J., Anttila, P., Suvanto, A. and Maltamo, M., 2006. Yksityismetsien metsävaratiedon keruuseen soveltuvilla kaukokartoitusmenetelmillä estimoitujen puustotunnusten luotettavuus. *Metsätieteen aikakauskirja*, 4/2006, 507–519. (In Finnish.)
- van Aardt, J.A.N., Wynne, R.H., and Oberwald, R.G., 2006. Forest volume and biomass estimation using small-footprint lidar-distributional parameters on a per-segment basis. *Forest Science*, 52, 636–649.
- Varjo, J., 2002. Metsäsuunnittelun tietohuollon järjestäminen tulevaisuudessa. *Metsätieteen aikakauskirja*, 3/2002, 537–540. (In Finnish.)
- Vastaranta, M., Holopainen, M., Haapanen, R., Yu, X., Melkas, T., Hyypä, J., and Hyypä, H., 2009. Comparison between an area-based and individual tree detection method for low-pulse density ALS-based forest inventory. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(3/W8), 147–151.
- Vastaranta, M., Holopainen, M., Yu, X., Hyypä, J., Hyypä, H., and Viitala, R., 2011. Predicting stand-thinning maturity from airborne laser scanning data. *Scandinavian Journal of Forest Research*, 26, 187–196.
- Vauhkonen, J., Korpela, I., Maltamo, M. and Tokola, T., 2010. Nearest neighbor imputation of tree attributes from airborne laser scanning data. *Remote Sensing of Environment*, 114, 1263–1276.
- Villikka, M., 2010. *The Suitability of Leaf-off Airborne Laser Scanner Data for Species Group Level Forest Inventory*. University of Eastern Finland, Faculty of Science and Forestry, master's thesis in Forest Planning and Economics. 44 p.

A national overview of airborne lidar application in Australian forest agencies

R.Turner¹, N.Goodwin², J. Friend³, D.Mannes⁴, J.Rombouts⁵ and A.Haywood⁶

¹ Forest Science Centre, Department of Primary Industries New South Wales, Sydney,
NSW

Russell.Turner@industry.nsw.gov.au

² Remote Sensing Centre, Department of Environment and Resource Management,
Ecosciences Precinct, Dutton Park. QLD

Nicholas.Goodwin@derm.qld.gov.au

³ Planning, Environment and Silviculture, Forest Products Commission, Bunbury, WA

jeremy.friend@fpc.wa.gov.au

⁴ Resource Information, Forestry Tasmania, Hobart, TAS

david.mannes@forestrytas.com.au

⁵ Resource Planning, Forestry South Australia, Mount Gambier, SA

Rombouts.Jan@forestrysa.com.au

⁶ Resource Planning, Department of Sustainability and Environment, VIC

Andrew.Haywood@dse.vic.gov.au

Abstract:

This paper provides a narrative of airborne lidar application across Australian forest agencies. It includes a brief history of early lidar research and operational trials, as well as current programs and future directions on a state by state basis. This review demonstrates a diverse range of lidar applications and increasing adoption of lidar technology within state agencies across Australia.

Keywords: Airborne lidar, remote sensing, national review, forestry

1. Introduction

It is now ten years since the first lidar trials were conducted in Australian forests and, assisted by the growing accessibility of lidar datasets and the development of new processing procedures and software tools, there has been a dramatic escalation in lidar use in forest agencies. Today most forest agencies have experienced a paradigm shift from explorative research to large scale operational programs. Lidar technology is having a significant impact on Australian forest management, and continues to revolutionise wood inventory programs and harvest planning processes.

The lidar forestry community in Australia is relatively small but active; and with representatives in every state. Several national lidar forestry forums have been held across Australia to share ideas and experience. The first was held in Brisbane (Queensland) in 2002, the second was in Hobart (Tasmania) in 2007, and most recently another workshop was again held in Hobart in 2010. A consistent issue raised at these workshops is the importance of disseminating

information on significant lidar research and potential operational applications to the general forestry community within Australia. The aim of this report is to provide a useful summary of past and present lidar work in each state forest agency. Of course, a comprehensive review of every research trial and operational program is well beyond the scope of this paper, however, a general overview will give a sense of the wealth of information and experience that has emerged over the past decade, and will serve to guide future research directions.

2. State overview

This overview concentrates on forest agencies within the six states of Australia (i.e. Queensland (QLD), New South Wales (NSW), Victoria (VIC), Tasmania (TAS), South Australia (SA) and Western Australia (WA)). As far as the authors are aware, airborne lidar has not yet been utilised for forestry purposes in the two territories; the Australian Capital Territory (ACT) and the Northern Territory (NT). This paper focuses on public commercial forests and plantations, but it should also be noted that lidar use in private plantations and public national parks and reserves has also increased significantly. A basic overview of lidar application is presented on a state by state basis.

2.1 Queensland

The remote sensing centre within the Queensland Department of Environment and Resource Management (DERM) has used lidar over the last decade for quantifying a range of biophysical attributes and to support the implementation of DERM's vegetation management policy and programs. The Injune Landscape Collaboration Project (ILCP) (Lucas *et al.* 2010a) has been an important lidar research site for Queensland since 2001. Injune is located in the Brigalow Belt of central Queensland and the vegetation consists mainly of open poplar box (*Eucalyptus populnea*) woodland with patches of denser white cypress pine (*Callitris glaucophylla*) regeneration. Lidar was used to evaluate its utilisation in estimating biomass and a set of forest structural attributes (Tickle *et al.* 2001, and Lucas *et al.* 2006) and results showed that stand-based lidar derived biomass models were highly correlated with field data ($R^2 = 0.92$, SE = 12 Mg/ha). This site was reflown with lidar in 2009 and is now the focus of further research into detecting forest structural change over time.

DERM has been involved in several research studies to better understand the relationship between field data, terrestrial laser scanning, sensor configuration, multi-temporal lidar, and forest structure. For example, a series of field monitoring plots throughout Queensland have had several repeat airborne lidar acquisitions between 2000 and 2009 (Lucas *et al.*, 2010b) to better understand the impacts of drought and land management practices on forest structure and species composition. In 2005 DERM, in collaboration with the University of Newcastle, investigated the use of lidar intensity and crown transparency to distinguish between forest species in white cypress woodland (Moffiet *et al.* 2005). This study showed that vegetation types could be discriminated using the proportion of singular returns and a "porosity" index based on the proportion of lidar points penetrating the canopy.

Forestry Plantations Queensland (FPQ), now owned by Hancock Queensland Plantations Pty Ltd, has incorporated lidar into their operational program. Over a three year period (2005 – 2007) FPQ captured lidar data across their entire softwood plantation estate (188,000 ha) (<http://www.fpq.net.au/>). This information was used mostly for digital terrain model (DTM) derived products (e.g. slope, hillshade, and contours) to assist with classifying harvesting terrain classes in pre-harvest planning (e.g. above and below 24° slope). Lidar was also used to characterise forest structure including stratification of stands based on height (equivalent to site index) to assist pre-harvest inventory.

The largest airborne lidar acquisition to date has been the Protecting Our Coastal Communities (POCC) project which covers an area of around 60,000 km² along the Queensland coast (see Figure 1). Commencing in 2009, this project, jointly funded by the state and local government, involved multiple providers/sensors for different regions. Data was captured with an average sampling density of 2 returns per sq.m. Although primarily intended for planning flood risk and urban development along the coastline, it also provided an excellent baseline for monitoring coastal forests and an opportunity for research such as the calibration/validation of spaceborne sensors, input for forest monitoring applications, as well as the extraction of many spatial products.

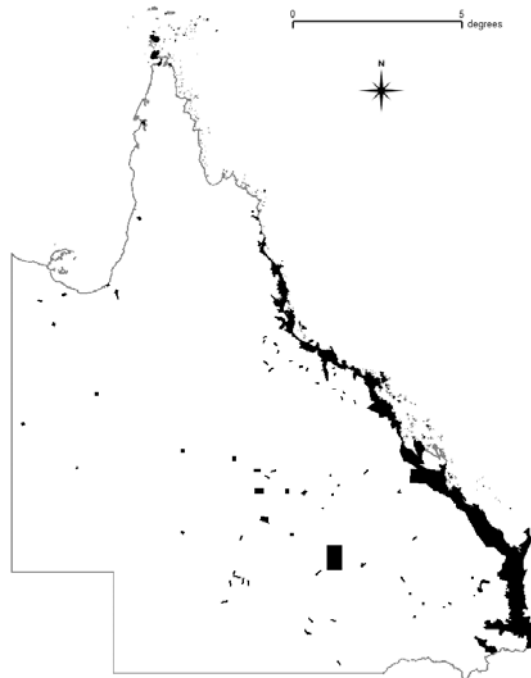


Figure 1: Current lidar coverage (in black) within Queensland. Note: this is only the lidar data held by DERM with the total area exceeding 68000 km².

Lidar data has since been utilised in over 50 sites strategically located to capture the variability in forest structure and composition. In addition to developing new relationships between field data and lidar, this data has been used operationally to calibrate and validate statewide methods and products (Armston *et al.* 2009). For example, lidar derived layers have been used to validate foliage projective cover (FPC) and woody extent which were derived from Landsat (25m) and MODIS (250m) products and to explore the relationships between canopy variables (Armston *et al.*, 2008; Gill *et al.* 2009; Scarth *et al.*, 2008 and Witte *et al.* 2000). More recently, the southeast Queensland region of the coastal capture was used to produce a calibrated FPC layer using 20 field sites located across the 8500 km² area and results showed that field measurements of FPC were highly correlated with first return lidar data ($R^2=0.92$, RMSE=5%). In addition to providing a baseline of FPC for southeast Queensland, it may help to improve the calibration of FPC using satellite imagery in forested areas with steep topography.

2.1 New South Wales

In NSW, around 2.2 million hectares of state-owned commercial forests are managed by Forests New South Wales (FNSW). With 230,000 ha of softwood plantation, FNSW is also the largest softwood plantation owner in Australia. The first lidar trial began in 2001 when 1,000 ha of

eucalypt forest were flown on the Central Coast. The focus of this trial was above-ground biomass assessment and an automated canopy segmentation process was developed (Turner 2006). In the same year, the first large scale trial occurred as a spin-off from a much larger (1.8 million hectare) catchment study by the Murray Darling Basin Commission (Liu et al. 2003). This project provided data across 90,000 ha of river red gum forests (*Eucalyptus camaldulensis*) along the southern border. The data was used to remap road and drainage networks, identify and estimate thinning resources, and plan harvesting events. In addition, a small (450ha) wood inventory trial showed it was possible to predict maximum height ($R^2 = 0.9$), mean dominant height ($R^2 = 0.76$), basal area ($R^2 = 0.72$) and gross volume ($R^2 = 0.79$) (Turner & Webster 2005, and Turner 2007).

In 2004 a study funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) acquired lidar data across 1,813 ha of native forest near Coffs Harbour on the North Coast. The project investigated the influence of scanning at different altitudes (1000, 2000, and 3000 m), footprint sizes (0.2, 0.4, and 0.6 m), scan angles (10° and 15° angle off nadir) and point sampling densities (0.18 to 1.9 m) on forest structure assessment (Goodwin *et al.* 2006). By 2006, another regrowth forest site, covering an area of 12,800 ha, had been flown on the Central Coast. The Jilliby Catchment Area (JCA) was a multiagency collaborative research project and data was used for two studies running simultaneously. The first focused on the application of airborne lidar for spatially mapping forest fuel characteristics (Roff *et al.* 2006 and Turner 2007), while the other study explored the potential for forest health monitoring by mapping Bell Miner Associated Dieback (BMAD), (Haywood & Stone, 2011).

Another 6,000 ha trial in coastal eucalypt forests was completed in 2008. A number of compartments were selected to investigate the benefits of lidar data in harvest planning and field supervision (Turner 2007 & 2008). It was estimated that lidar data reduced the planning effort by 2 to 3 person days per compartment plan and provided a 10% time saving in field supervision during harvesting, with less walking required to locate exclusion zones and merchantable trees. Automated drainage maps from lidar-derived DTMs were also tested. Precision surveys along two creek lines indicated an excellent correlation between survey and DTM elevation ($R^2 = 0.99$) with a mean elevation error of 0.6 m, while automated drainage networks had a centreline mean error of 1.65 m.

A second large scale study was implemented in early 2008 when 240,000 ha of native forest were flown in North Central NSW near the town of Baradine. The Pilliga Remote Sensing (PILRES) project covered cypress pine/eucalypt woodland on predominantly flat terrain. A wood resource inventory utilised airborne lidar (to provide height and stocking) and multispectral digital photography (to define stands of commercial forest types). Lidar-derived canopy height models (CHMs) were used for a strategic thinning program and the high resolution DTMs assisted the update of road and drainage networks. The project was expanded by another 129,000 ha during 2009 across 93 state forests scattered throughout western NSW. This was also the first time that simultaneous lidar and digital photography (colour infra-red) was acquired, making it easier to combine attributes at crown level from both sensors (i.e. data fusion).

The first case study in a NSW softwood plantation was initiated in July 2008. The Plantation Airborne Resource Inventory Appraisal (PARIA) project was undertaken in a 5,000 ha *Pinus radiata* plantation in south-central NSW. The study evaluated airborne lidar and digital multispectral aerial photography for wood resource inventory, structure stratification and forest health monitoring (Stone *et al.* 2008, Stone *et al.* 2010 and Turner *et al.* 2011). A remote sensing guide for softwood plantation managers (Turner and Stone 2010) was also produced as well as a new in-house ArcGIS lidar toolbox.

In late 2009, FNSW and the University of New South Wales (UNSW) obtained an Australian Research Council (ARC) grant to investigate full-waveform lidar applications and to develop a commercial software package for on-screen interpretation. In 2010, full-waveform lidar was acquired at three sampling densities (i.e. 2, 5 and 10 pulses per sq.m) over a 140 ha pine plantation west of Sydney. The data is being used to develop a new waveform processing technique to minimise lidar negative tree height bias (Park *et al.* 2011) and to model 3D structure with both airborne and terrestrial laser scanner (TLS) data (Park *et al.* 2010). Development of the lidar interpretation software is well underway and a working prototype should be ready for field testing by December 2011.

FNSW's largest operational program is currently in progress throughout the northern tablelands and mid-north coast of NSW. The project is capturing almost 296,000 ha of state forest, including around 17,500 ha of pine plantation. The data will be used for general forest management, forest inventory, stand structure assessment and drainage mapping.

Increasingly, lidar data is becoming more accessible from other sources. Local shire councils are acquiring lidar for flood mitigation and urban planning, and data covering around 198,000 ha of state forest has been made available. In 2008, the NSW State Government also purchased a Leica ALS50-II laser scanner that is operated by the Land and Property Management Authority (LPMA) based in Bathurst. LPMA has an ongoing lidar acquisition program for their state mapping needs and where data is captured over state forests it is also provided to FNSW. Since 2001, the accumulative total of all airborne lidar coverage in NSW state forests has reached 943,000 ha, which represents 43% of the total native forest estate. In addition, around 22,000 hectares of softwood plantation have been captured covering almost 10% of the total plantation area. Figure 2 illustrates the widespread coverage of this data across NSW.



Figure 2. NSW map showing the location of lidar coverage in state forests (in black).

2.3 Victoria

The Victorian Department of Sustainability and Environment (DSE) is responsible for the sustainable management of 7.8 million hectares of public native forests (DSE 2008). DSE has been investigating the use of lidar as a land management tool since 2001 (Choma *et al.*, 2005).

Data capture started predominantly for DEM construction (Choma *et al.*, 2005) but has expanded to include forest structural inventory (Haywood and Sutton 2009), forest stand delineation (Haywood and Stone, 2009), biomass estimation (Kandel *et al.*, 2009), fuel load mapping (Haywood *et al.*, 2010). DSE have recently developed a method to assess riparian vegetation health using lidar in conjunction with aerial imagery (Johannsen *et al.*, 2010). This has included the recent acquisition of 26,000 km of lidar acquisition across Victoria's stream network (see Figure 3).

All of the lidar data used in DSE's projects has been acquired through the Victorian Coordinated Imagery and Elevation Program's (<http://www.land.vic.gov.au/Land/>). The Program's mission is to develop an efficient and effective service and cost sharing model for the acquisition of spatial imagery and elevation products for the State of Victoria and the Program Purchase Partners.

The Program coordinates the purchase of aerial imagery and elevation products across Victoria for a range of government and non-government organisations. It is designed to facilitate uses, imagery needs, reduce costs, avoid duplication and to contract manage projects. It streamlines the acquisition, storage and access to aerial images and elevation products for end users.

Instead of having to purchase images for a specific purpose, Purchase Partners can share images and use them again for a range of different purposes, from urban growth area planning, property searches and asset management, through to tracking native vegetation. This has enabled the Victorian land management community to leverage off a coordinated investment.

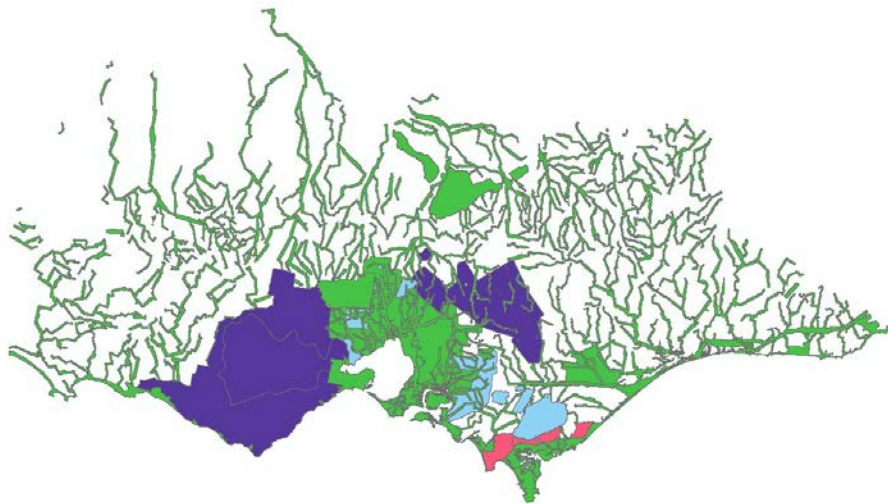


Figure 3: Map showing extent of lidar coverage in Victoria.

2.4 Tasmania

Forestry Tasmania (FT) has been investigating the use of lidar as a forest management tool since 2004 when several small areas were captured to evaluate a variety of applications, including operational planning and forest inventory. Although data capture for this exercise was designed primarily with a research focus, the early investigations showed promise for operational use but at the time the application of the technology was considered too expensive (Bennett 2005).

A fresh look at the results from the 2004 trial concluded that it may be possible to make a financial case for lidar if savings could be realised from across a wider range of applications.

To investigate this further, a large scale operational trial was established in north-east Tasmania in 2007 with the acquisition of 32,000 ha. The study covered a diverse range of forest types and carefully monitored and evaluated the benefits of lidar on forest management and operational planning. To compare lidar with traditional approaches, many coupes were “double planned” (i.e. with and without the availability of lidar data) by different forest planners. This provided a direct estimate of savings that could be realised across the forest management spectrum. Results from the operational trial were sufficiently encouraging for FT to commence the operationalisation of lidar in 2009 (Mannes & Stone 2009). Lidar capture has since been progressively rolled out and to date over 700,000 ha of data has now been acquired mainly over State Forest and surrounding lands including private forests (see Figure 4). Lidar derived data is being used in Tasmania across the spectrum of forest management activities, ranging from engineering, forest harvest management, and resource inventory, to plantation management and general mapping. Where lidar data exists, it is fully integrated into day to day forest management activities.



Figure 4. Tasmania map showing the location of lidar coverage (in grey).

Research into lidar continues to have a strong operational focus. Most research activity is in the forest inventory sphere, and involves investigation of better metrics and methodologies for the prediction and projection of timber variables (timber volume, specifically eucalyptus volume, stem size distribution, basal area and stocking levels) from lidar. Active investigation of image analysis techniques to assist with mapping and forest characterisation is ongoing and is beginning to yield positive results. Research is also underway to develop better ways to display lidar derived data to increase its utility for field foresters. Ongoing monitoring of the operational effectiveness of lidar continues, to determine whether operational results are being achieved in ‘real’ forest management.

2.5 South Australia

The first forestry related airborne lidar campaign was undertaken by the South Australian Forest Corporation (ForestrySA) in 2002. The objective of the trial was to capture data over radiata pine plantations covering a range of age classes to gain an understanding of their potential for applications such as terrain mapping, forest inventory and site quality assessment. The CSIRO utilised some of these early data (with funding from the now Forest and Wood Products Australia) and results from this initial research were reported in Lovell *et al.* (2003) and Lovell *et al.* (2005).

The work by CSIRO demonstrated that lidar data effectively showed spatial variation in stand height and density, suggesting potential application for site quality (SQ) assessment. This led to further work by ForestrySA (Rombouts, 2006) and triggered two more research oriented lidar campaigns in 2006 and 2007 to demonstrate and refine lidar-based SQ assessment across a range of sites. In 2007, ForestrySA also participated in a regional project involving several government departments to build a Digital Elevation Model (DEM) for the lower south-east of the state, and mainly for hydrological modelling purposes. As a result ForestrySA acquired low density data over its entire south-eastern pine plantation estate (105,000 ha). Figure 5 provides a map of the geographic extent of lidar coverage in South Australia. These data were also successfully utilised to demonstrate that SQ could be accurately assessed with relatively sparse lidar sampling density acquired at high altitude (Rombouts *et al.*, 2010).

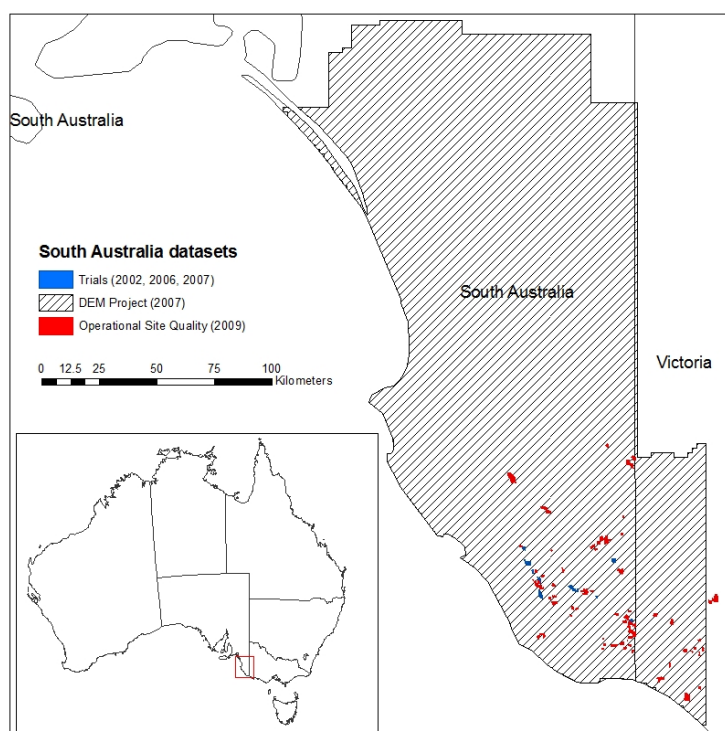


Figure 5: Map showing lidar extent within South Australia.

In late 2007, ForestrySA made the decision to take lidar-based SQ assessment into operational use and conventional SQ assessment was discontinued. The first operational SQ lidar campaign took place in December 2008 and the first SQ maps were uploaded in corporate databases in July 2009. The outcome of the project, in particular the efficiency of the field sampling design, was evaluated and possible improvements identified (Rombouts *et al.*, 2010). A second operational survey is scheduled for the summer of 2012. Having successfully implemented lidar-based SQ assessment, ForestrySA is now turning its attention to product oriented pre-harvest inventory.

In addition to work with airborne lidar systems, ForestrySA has also investigated Terrestrial Laser Scanners (TLS) from 2003 to 2006 as an industry partner with CSIRO validating the Echidna full-waveform laser scanning system (Lovell *et al.* 2003; Culvenor *et al.* 2006). Commencing in 2008, ForestrySA was also involved with the Collaborative Research Centre for Forestry in testing commercial TLS for pre-harvest inventory purposes (Murphy *et al.*, 2010). A follow-up project in collaboration with Treemetrics from Ireland is currently underway. The objective is to capture detailed information on tree stems and defects, enabling optimisation of cutting patterns for each tree and stand, within the constraints imposed by prevailing market

conditions. Another objective is to explore the integration of ground and airborne lidar data to take advantage of the complementary qualities of both sensor types.

2.6 Western Australia

Aerial lidar for forestry applications is still in its infancy within Western Australia. Although a number of agencies within WA are utilising lidar for DTM purposes, there had been limited application within the forestry sector. The first forestry related application was completed in 2007 by the Department of Environment and Conservation (DEC). This was a trial conducted in four native forest sites, totalling 1,288 ha, but no further lidar trials or projects have been undertaken by DEC.

In 2010, the Forest Products Commission (FPC) began a major resource assessment of its 40,000 ha radiata pine plantation estate. Due to limitations in staffing resources and strict project deadlines, a purely traditional ground-based assessment program was not considered feasible, whereas lidar technology offered the potential to capture broad scale plantation resource information in a more timely and efficient manner.

FPC began its first aerial lidar program in February, 2011 acquiring data across 13,500 ha of pine plantation within a region known as the Blackwood Valley, located 50km south-east of Bunbury (see Figure 6). This region was considered suitable for a trial as it contains a significant area of steep terrain, has limited access for ground-based assessment teams and it is relatively spatially compact compared to the majority of the FPC estate.



Figure 6. Map showing the extent of lidar coverage (black) in WA softwood plantations.

The primary focus of the project is to use lidar data, in conjunction with multispectral imagery, to establish estimates of current plantation characteristics such as stocking, height and stand volume. Once these derivatives have been attained, they will be used to assist in long-term wood supply modelling applications. Data will also be used to derive additional value added products, such as DTM's and slope maps. This information will then be passed to planning and operational staff where it can be used to assist making on-ground operations more efficient. The project will also undertake a cost/benefit analysis of the process to compare it with conventional ground-based inventory methods. Moreover, the economics and practicality of collecting lidar data across less spatially compact plantation resource areas will be investigated. If significant benefit can be demonstrated then it is likely that airborne lidar technology will become an integral part of the FPC inventory assessment program.

3. Summary

Airborne lidar forestry applications in Australia have rapidly advanced over the past decade. Today every state has at least trialled lidar over their native forests or plantations, and most forest agencies have shifted from researching potential lidar applications to a larger and more ambitious operational phase. The state summaries have illustrated the wide range of forestry applications that are either under investigation or already operational. Applications include wood resource inventory, harvest planning and supervision, site quality assessment, road and drainage mapping, forest classification, strategic thinning evaluation, forest fuel assessment, canopy health monitoring and biomass estimation. This report clearly demonstrates the high level of state agency support and commitment to developing lidar forestry applications. It follows that lidar technology will continue to create inroads into mainstream operational and planning activities and make a significant contribution to forest management in Australia.

Acknowledgements

The authors wish to thank the following people for their kind assistance in collating information for this national overview.

Christine Stone (Dept. Primary Industries, NSW)
Tony Brown, Paul McBain and Morgan Roche (Forests NSW)
Glenn Jones (Land and Property Management Authority, NSW)
John Armston (Dept. of Environment and Resource Management, Qld)
Paul Rampant (Dept. of Environment & Conservation, WA)
Colin Reugebrink and Michelle McAndrew (formerly Forestry Plantations Queensland)
Andrew Yates (Parks and Conservation, ACT)

References:

- Armston, J.D., Denham, R.J., Danaher, T.J., Scarth, P.F., and Moffiet, T., (2009). Prediction and validation of foliage projective cover from Landsat-5TM and Landsat-7TM ETM+ imagery for Queensland, Australia, *Journal of Applied Remote Sensing*, vol. 3, no.1, 033540-28.
- Armston, J. et al. (2008). A first look at integrating ALOS PALSAR and Landsat data for assessing woody vegetation in Queensland, *Proceedings of the 14th Australasian Remote Sensing and Photogrammetry Conference*, Darwin, Australia, October 2008. (Eds Edwards, E., and Bartolo, R.)
- Bennett (2005) – Prospects for Lidar at Forestry Tasmania , Forestry Tasmania internal report.
- Choma, A., Ratcliff, C. and Frisina, R. (2005). Evaluation of Remote Sensing Technologies for High-resolution Terrain Mapping. *Proceedings of SSC 2005 Spatial Intelligence, Innovation and Praxis: The national biennial Conference of the Spatial Sciences Institute*, September, 2005. Melbourne: Spatial Sciences Institute. ISBN 0-9581366-2-9
- Culvenor, D., Newnham, G., Rombouts, J., Jupp, D. and Lovell, J., (2006). Pre-harvest Forest Inventory from Ground-based Laser Scanning, *USDA Forest Service 8th Annual Forest Inventory and Analysis Symposium, Monterey, October 2006*.
- Department of Sustainability and Environment (2008) Victoria's State of the Forests Report.

Department of Sustainability and Environment, Melbourne, Australia.
www.dse.vic.gov.au/sfm.

- Gill, T. K., Phinn, S.P., Armston, J.D., and Pailthorpe, B.A., (2009). Estimating tree-cover change in Australia: challenges of using the MODIS vegetation index product, *International Journal of Remote Sensing*, vol. 30, no.6, pp. 1547-1565.
- Goodwin, N., Coops, N. & Culvenor, D.S. (2006). Assessment of forest structure with airborne Lidar and the effects of platform altitude, *Remote Sensing of Environment*. 103:140-152.
- Haywood, A., Mellor A., and Siggins, A. (2010). Fuel Hazard Mapping of the Victorian Central Highlands Using Lidar Data. *5th Australian Remote Sensing & Photogrammetry Conference*, 13th – 17th September, 2010. Alice Springs Northern Territory, Australia.
- Haywood, A. and Stone, C. (2009). Object-based analysis of forest stand delineation on high spatial resolution imagery using open source software. In: *Forestry: a climate of change*, Thistlethwaite, R., Lamb, D. and Haines, R. (eds). pp. 142 -150. *Proc. IFA Conf.*, Caloundra, Queensland, Australia, 6 – 10. September 2009.
- Haywood, A. and Stone, C. (2011). Mapping eucalypt forest susceptible to dieback associated with bell miners (*Manorina melanophys*) using laser scanning, SPOT 5 and ancillary topographical data. *Ecological Modelling*, 222, 1174–1184.
- Haywood, A and Sutton, M. (2009). Estimating forest characteristics in young Victorian ash regrowth forests using field plots and airborne laser scanning data. In: *Forestry: a climate of change*, Thistlethwaite, R., Lamb, D. and Haines, R. (eds). pp. 131–141. *Proc. IFA Conf.*, Caloundra, Queensland, Australia, 6 – 10 September 2009.
- Johannsen, K., Grove, J., Hoffman, C. Kollar,S., and Phinn, A. (2010). Object-based image analysis of bank condition using airborne lidar and high spatial resolution image data. In *Victoria, Australia.*” *5th Australian Remote Sensing & Photogrammetry Conference*, 13th – 17th September, 2010. Alice Springs Northern Territory, Australia.
- Kandel, Y.P., Fox, J. C., Culvenor, D., Arndt, S. K., and Livesely, S. J. (2009). Estimating aboveground biomass of native sclerophyll forest using airborne LiDAR. *IUFRO Division 4.01 Conference: “Meeting multiple demands for forest information: New technologies in forest data gathering”*, 17 – 20 August 2009, Mount Gambier- South Australia
- Liu, G., Choma, A., Clark, S. and Tierney, G. (2003). Extracting fine resolution hydrographical features from located airborne laser scanning point clouds for hydraulic modelling. *Proceedings of the Spatial Sciences Conference*, Canberra, Australia.
- Lovell J.L., Jupp D.L.B., Culvenor D.S. and Coops N.C.C. (2003). Using airborne and ground based ranging Lidar to measure canopy structure in Australian forests. *Canadian Journal of Remote Sensing* 29:607-622.
- Lovell J.L., Jupp D.L.B., Newnham G.J., Coops N.C. and Culvenor D.S. (2005). Simulation study for finding optimal lidar acquisition parameters for forest height retrieval, *Forest Ecology and Management*, 214(1-3): 398-412.
- Lucas, R., Bunting, P., Armston, J., Lee, A., and Campbell, G. (2010a), The Injune Landscape

Collaborative Project: An update on research activities. *In: Proceedings of the Australasian Remote Sensing and Photogrammetry Conference*, Alice Springs, Australia, 13-17 September 2010.

- Lucas, R., Lee, A., Armston, J., Carreiras, J.M.B., Viergever, K.M., Bunting, P., Clewley, D., Moghaddam, M., Siqueira, P. and Woodhouse I. (2010b). Quantifying Carbon in Wooded Savannas: The Role of Active Sensors in Measurements of Structure and Biomass. *In: Ecosystem Function in Savannas: Measurement and Modelling at Landscape to Global Scales*, Eds. M.J. Hill and N.P. Hanan, Taylor and Francis.
- Lucas, R.M., Cronin, N., Lee, A., Moghaddam, M., Witte, C. and Tickle, P., (2006). Empirical relationships between AIRSAR backscatter and Lidar-derived forest biomass, Queensland, Australia. *Remote Sensing of the Environment*, vol. 100, pp. 407 – 425.
- Mannes. D., and Stone, M. (2009). Operational Trial to Operational Reality. *In: Proceedings of ForestTECH 2009 conference*, Albury, NSW, Australia. pp 94-107.
- Moffiet, T., Mengersen, K., Witte, C., King, R., and Denham, R. (2005) Airborne laser scanning: Exploratory data analysis indicates potential variables for classification of individual trees or forest stands according to species. *ISPRS Journal of Photogrammetry & Remote Sensing*. Vol 59(5): 289-309.
- Murphy, G. E., Acuna, M. A. & Dumbrell, I. (2011). Tree value and log product yield determination in Radiata pine plantations in Australia: Comparisons of terrestrial laser scanning with a forest inventory system and manual measurements. *Canadian Journal Forest Research*, 40:2223-2233
- Park, H., Lim, S., Trinder, J. and Turner, R. (2010). 3D surface reconstruction of terrestrial laser scanner data for forestry. *In: Proceedings of IGARSS*
- Park, H., Turner, R., Lim, S., Trinder, J. and Moore, D. (2011). Analysis of pine tree height estimation using full waveform lidar. *In: Proceedings of the 34th International Symposium for Remote Sensing of the Environment (ISRSE)*, Sydney, Australia, 10-15 April 2011.
- Roff, A., Taylor, G., Turner, R., Day, M., Mitchell, A. and Merton, R. (2006). Hyperspectral and lidar remote sensing of forest fuel loads in Jilliby State Conservation Area. *In: Proceedings of the 13th ARSPC conference*.
- Rombouts, J. H. (2006). Exploring the potential of airborne Lidar for site quality assessment of radiata pine plantations in South Australia: initial results. *Paper presented to Research Working Group 2, Forest Measurement and Information Systems, Biennial meeting*, 21-24 November 2006, Woodend, Australia.
- Rombouts, J. H., Ferguson, I. S. and Leech, J. W. (2010). Campaign and Site effects in Lidar prediction models for Site Quality assessment of radiata pine plantations in South Australia. *International Journal of Remote Sensing*, 31, 1155-1173.
- Rombouts, J. H., Ferguson, I. S., Leech, J. W. and Culvenor, D. S. (2010). An evaluation of the field sampling design of the first operational Lidar based site quality survey of radiata pine plantations in South Australia. *Conference Proceedings Silvilaser 2010, 14-17 September 2010, Freiburg Germany*.
- Scarth, P., Armston, J., and Danaher, T. (2008). On the Relationship between Crown Cover,

Foliage Cover and Leaf Area Index. *In: Proceedings of the 14th Australasian Remote Sensing and Photogrammetry Conference*, Darwin, Australia, October 2008. (Eds Edwards, E., and Bartolo, R.)

- Stone, C., Turner, R., Kathuria, A., Carney, C., Worsley, P., Penman, T., Bi, H., Fox, J. & Watt, D. (2010). Adoption of new airborne technologies for improving efficiencies and accuracy of estimating standing volume and yield modelling in *Pinus radiata* plantations (PNC058-0809). *Final Report for the Forest & Wood Products Australia Project PNC0-0809*. Available on the FWPA website www.fwpa.com.au
- Stone, C., Turner, R., and Verbesselt, J. (2008) Integrating plant health surveillance and wood resource inventory systems using remote sensing. *Australian Forestry*, Vol. 71, No.3, pp 245-253.
- Tickle, P.K., Lee, A., Witte, C., Lucas, R.M., Jones, K., Austin, J., Denham, R. & Good, N.M. (2001). The Operational Use of Airborne Scanning Lidar and Large Scale Photography within a Strategic Forest Inventory and Monitoring Framework. *Geoscience and Remote Sensing Symposium*, 2001, IGARSS '01, IEEE 2001 International Vol 3: 1000-1003.
- Turner, R. (2006). An airborne lidar canopy segmentation approach for estimating above-ground biomass in coastal eucalypt forests. *PhD Thesis*. School of Biological, Earth and Environmental Sciences, University of New South Wales, NSW. July 2006., 373 pp.
- Turner, R (2007). An overview of Airborne lidar applications in New South Wales state forests. *Proceedings of the ANZIF Conference*, June 2007, Coffs Harbour, NSW, Australia.
- Turner, R. (2008). Applications for remote sensing tools in forestry operations. *Proceedings of the Australian ForestTECH Conference* – April 2008, Albury, NSW, Australia.
- Turner, R. & Stone, C. (2010). Guide to acquisition and processing of remote sensing data for softwood plantations. *Document prepared for Forest & Wood Products Australia and a deliverable associated with FWPA Project PN058-0809*. Available on the FWPA website www.fwpa.com.au
- Turner, R., Stone, C., Kathuria, A. and Penman, T. (2011). Towards an operational lidar resource inventory process in Australian softwood plantations. *Proceedings of the 34th International Symposium for Remote Sensing of the Environment (ISRSE)*, Sydney, Australia, 10-15 April 2011.
- Turner, R. & Webster, M. (2005). Estimating wood volumes in Australian River Red Gum (*Eucalyptus camaldulensis*) forest by combining airborne Lidar and a digital mapping camera. *Proceedings of Spatial Science Conference*. September 2005. Melbourne: SSI. ISBN 0-9581366-2-9
- Witte, C., Denham, R., Turton, D., Jonas, D., Tickle, P. and Norman, P. (2000). Airborne laser scanning: A tool for monitoring and assessing the forests and woodlands of Australia. *In: Proceedings, 10th Australasian Remote Sensing Conference*, 21-23 August 2000, Adelaide, Australia. Paper No. 166, pp 348-362.

Airborne LiDAR based forest inventory in Bangladesh for REDD plus MRV: scope and potentiality

Md. Parvez Rana^{1,2*}, Timo Tokola¹, Hanna Holm³, and Tuomo Kauranne³

¹School of Forest Sciences, University of Eastern Finland, P.O.Box - 111, FI – 80101, Joensuu, Finland. Email: parvezra@student.uef.fi

²Department of Forest Resource Management, Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden, Email: parvez_200207@yahoo.com

³Arbonaut Ltd. Kauppakatu 21, 80100 Joensuu, Finland (firstname.lastname@arbonaut.com)

Abstract

Nowadays, the accurate measurements of carbon stock for carbon trading in REDD plus (Reducing Emissions from Deforestation and Forest Degradation in Developing Countries) countries are going highly demanding. IPCC (Intergovernmental Panel on Climate Change) Tier 3 level accuracy for estimation of emissions from deforestation and forest degradation requires detailed national inventory of key carbon stocks, repeated measurements and modeling. Present study has been carried out to know scope and potentiality of the airborne LiDAR based forest inventory in Bangladesh for REDD plus MRV (monitoring, reporting and verification). Here we supposed a hybrid method where the integration of airborne LiDAR data with satellite imagery and ground truth data based forest inventory in Bangladesh. As the forest of Bangladesh is highly dynamic and inaccessible due to hilly and mountainous area, this method will give an accountable and transparent report of carbon stock. We also highlighted the limitation of this approach in a developing country like Bangladesh due to poor economic and technical condition. Till now there is no record of application of airborne LiDAR system for forest inventory in Bangladesh. Finally, we recommended that the Forest Department of Bangladesh with financial and technical help from international organization can do a pilot project in Sundarban Mangrove Forest.

Keywords: LiDAR, REDD, remote sensing, Bangladesh, Forest inventory

1. Principle of airborne laser scanning (ALS) system

Airborne laser scanning, LiDAR (Light Detection and Ranging) is an optical remote sensing technology which measures the properties of scattered light radiating from of a distant target. An airplane or helicopter-mounted sensor sends laser pulses towards ground and records the elapsed time between beam launch and return signal registration. A typical LiDAR system consists of three main components: a Global Positioning System (GPS) to provide position information, an Inertial Navigation System (INS) for attitude determination and a laser scanner to provide the range from the laser-beam firing point to its footprint (Bang et al. 2008). Some of the LiDAR pulses are reflected from tree canopy, trunks, branches, leaves or lower vegetation, but they also penetrate through the canopy layer reaching the ground, thereby profiling a three-dimensional point cloud image of the forest (Gautam and Kandel 2010). In addition, by varying the wavelength of the light transmitted, pulse frequency and duration, and other factors, LiDAR can be used in a variety of applications to detect numerous substances.

There are two main categories of airborne LiDAR systems: small-footprint discrete-return LiDAR and large-footprint, waveform-recording LiDAR. Small-footprint discrete-return LiDAR devices measure either one (single-return systems) or a small number (multiple-return systems) of heights by identifying, in the return signal, major peaks that represent discrete objects in the path of the laser illumination. Large-footprint waveform-recording devices record the time-varying intensity of the returned energy from each laser pulse, providing a record of the height distribution of the surfaces illuminated by the laser pulse (Li 2009). Slender than 1 meter is usually categorized as small footprint and suitable for estimation of forest attributes at stand or single tree level, whereas large footprint reaching dozens of meters are normally designed for topographical survey of planet.

Airborne LiDAR based forest inventory is two types such as Individual Tree Detection (ITD) and Area Based method. The area based method is considered the more cost-efficient approach due to its lower pulse density ($<1/m^2$), although it needs large amounts of expensive fieldwork, compared with the individual tree detection (pulse density $>5/m^2$) to perform accurately. Area based method also called as distribution-based method that uses the canopy height or vertical distribution of laser echoes for estimating area-based forest inventory parameters (e.g. mean height, stem number, basal area, and volume) by statistical means (Naesset 2004).

2. LiDAR technology for REDD plus Monitoring, Reporting and Verification

REDD, Reducing Emissions from Deforestation and Forest Degradation in Developing Countries, refers to “an effort to create a financial value for the carbon stored in forests, offering incentives for developing countries to reduce emissions from forested lands and invest in low-carbon paths to sustainable development” (UNREDD 2010). The reason for including REDD in the carbon market is to allow developing countries to earn money simply by conserving their forests. The participating developing countries will receive a dual benefit from REDD. They can reduce their own country-specific emissions on the one hand, as well as earn money on the other by selling credits thus achieved to rich countries, helping the latter meet their own emission-reduction targets. Within the mechanism of REDD, rich countries are allowed to pay for protecting tropical forests as a cost-effective alternative to cutting their own GHG. Developed nations can buy credits from developing nations that are backed by reliably measured carbon stored in tree growth, thereby offsetting the higher cost of achieving corresponding reductions by cutting their own emissions.

The significance of estimation of accurate carbon stock in the tropical forest has been growing radically. However, due to topographical complexity with lack of up-to-date data has been a challenging task to get a cost-effective, efficient and equitable approach to meet the carbon monitoring, reporting and verification. In addition, the conventional remote sensing technique with incorrect field measurement has made it difficult to gain accurate estimations of the biomass of tropical forests. Therefore, a hybrid approach, wherein airborne laser scanning is used with satellite image and sample field measurement, is a viable solution that can increase accuracy as well as reduce cost. Such an approach is feasible because wall-to-wall LiDAR meets even higher accuracy standards that are needed in operational forest management. When followed by adaptive statistical estimation, LiDAR is able to “teach” automatic satellite data interpretation to achieve almost the same level of accuracy, even when the area covered by LiDAR is just ten percent of the total.

LiDAR, an emerging remote sensing technology, can improve the assessment of forests and make determinations of the amount of carbon stocks in a particular forest much more reliable. In addition, LiDAR technology is especially well-suited for use in the tropics, as it is less sensitive to weather conditions and sun angles than satellite imaging technology. The IPCC has proposed different levels of methodology for estimating greenhouse gas (GHG) emissions. These options are specified at different Tiers, which relate to methodological complexity. Moving from Tier 1 to Tier 3, the estimation of eligible forest carbon content requires steadily increasing accuracy, whereas the feasibility of estimation methods increases in the opposite direction. At Tier 3, higher accuracy—if it can be verified—is rewarded by a much higher compensation level per ton of carbon, due to better reliability of the assessed amount of carbon captured in forests. As well, Tier 3 estimates can be achieved and verified by adopting a LiDAR-based biomass and carbon assessment method. LiDAR-based forest inventory has become the method of choice in operational forest management.

In August 2010, Bangladesh joined in UN-REDD (The UN-REDD Programme is the United Nations Collaborative initiative on Reducing Emissions from Deforestation and forest Degradation in developing countries) as a partner country. Bangladesh Forest Department has been furnished a carbon stock inventory project for Sundarban Mangrove Forest and Eight Protected Areas. In addition, 'Sundarbans Forest Carbon Inventory-2009' conducted by Forest Department (FD) of Bangladesh under assistance from USDA Forest Service, USAID and other collaborators. The 16th UNFCCC Conference of Parties (2009) in Mexico emphasizes on the integration of indigenous

knowledge (IK) and collaboration of indigenous people to the role of conservation, sustainable management of forest and enhancement of forest carbon stock. For the successful implementation of REDD plus mechanism relies on the accurate estimation of forest carbon stock that requires that there is an effective monitoring, reporting and verification system in operation. According to the Intergovernmental Panel on Climate Change (IPCC), moving from Tier 1 to Tier 3 requires higher accuracy and Tier 3 will be rewarded by a much higher compensation level per ton of carbon, due to better reliability of the assessed amount of carbon captured in forests.

3. An overview of Bangladesh Forest

Bangladesh occupies a unique geographic location (20°34'N – 26°38'N latitude to 88°01'E – 92°41'E longitude) – spanning a relatively short stretch of land between the mighty Himalayan mountain chain and open ocean. According to Forest Department and some other sources (Rana *et al.* 2009) forest cover is nearly about 2.53 million ha representing approximately 17.5% of the country's total surface area (Figure 1, Table 1). Officially, Bangladesh Forest Department manages 1.53 million hectares of forest land of the country. Besides, 0.73 million ha of unclassified state forests (USF) are under the jurisdiction of district administration (Roy 2005). The annual deforestation rate in Bangladesh is 3.3% which is highest among the south-east Asian countries (Poffenberger 2000). Contribution of the forestry sector to Bangladesh GDP is 3.3% at current prices and about 2% of the country's labor forces are employed in this sector (Siddiqi 2001).

Table 1 Forest types and areas in Bangladesh

Forest type		Location	Area (million ha)	Remarks
Hill forest	Managed reserved forest (evergreen to semi-evergreen)	Eastern part of the country (Chittagong, Chittagong Hill Tracts and Sylhet)	0.67	Highly degraded and managed by the Forest Department.
	Unclassed state forest (USF)	Chittagong Hill Tracts	0.73	Under the control of district administration and denuded mainly due to faulty management and shifting cultivation. Mainly scrub forest.
Plain land forest	Tropical moist deciduous Forest	Central and north-western region (Dhaka, Mymensingh, Tangail etc.)	0.12	Mainly <i>Sal</i> forest but now converting to exotic short rotation plantations. Managed by the Forest Department.
Mangrove	Sundarbans	Southwest (Khulna, Satkhira)	0.57	World's largest continuous mangrove forest and including 0.17 million ha of water.
	Coastal forest	Along the shoreline of twelve districts	0.10	Mangrove plantations along the shoreline of 12 districts. Managed by Forest Department.
Village forest		Homestead Forests all over the Country	0.27	Diversified productive system. Fulfill majority of country's domestic timber, fuelwood and bamboo requirements.
Plantation in tea and rubber gardens		Chittagong Hill Tracts and	0.07	Plantations of various short rotation species (mainly

	Sylhet		exotics).
Total forest		2.53	17.49 % of country's total landmass

Source: Mukul *et al.* (2008)

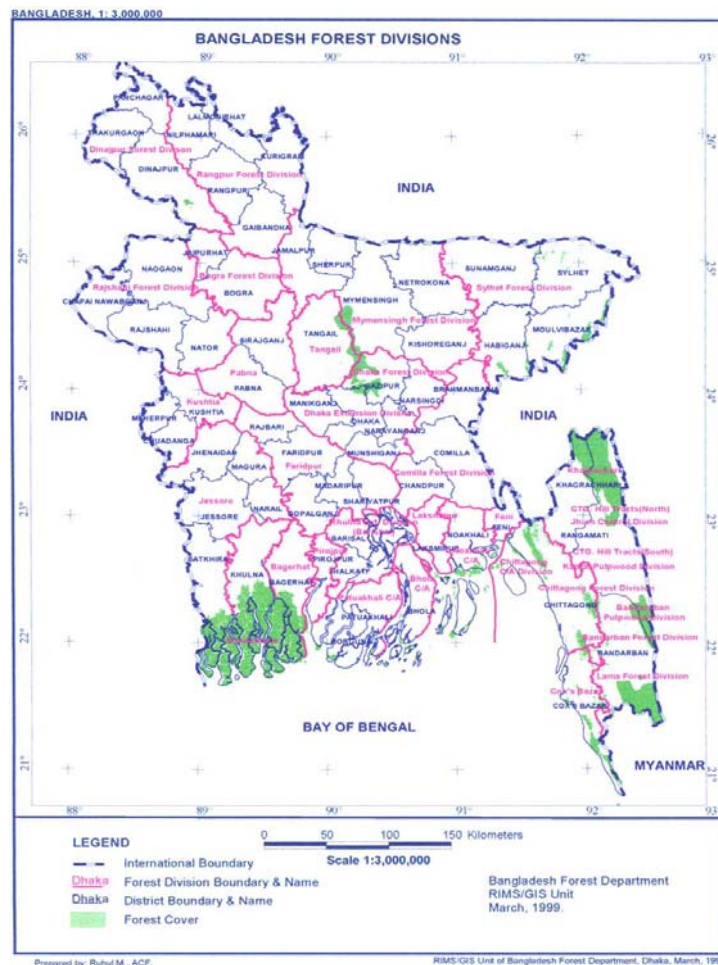


Figure 1: Map showing the distribution of forest in Bangladesh (Bangladesh Forest Department 2008)

5. Sampling design followed for forest inventory in Bangladesh

The last National Forest Inventory (NFI) in Bangladesh has been carried out in 2005-2007 by following systemic sampling method with technical suggestion from FAO Forest Resources Development Service (FOMR). “Tract” was regarded as a sampling unit for ground truth data measurement. 299 tracts or samples site was the result for field measurement after following systematically distribution of sample sites with an interval of 15 minutes latitude and 10 minutes longitude where the sample coordinates represent the South-West corner of the tracts (NFTRA 2007). Each tract represents a square of 1km*1km (1km²) contains 4 plots (the dimension of 20m*20m or 0.5ha). The plot orientation in each tract was such as Plot # 1 North-ward, Plot # 2 East-ward, Plot # 3 South-ward and Plot # 4 West-ward. If the plots are in the land use class of “Forest”, circular subplots was established to collect data on tree regeneration. As tree diversity is very high with respect to wide variety of size, age and species composition, it was recommended that Concentric Circular Sample Plots (CCSPS) are used to tallying trees (NFTRA 2007; van Laar and Akcal 1997). A circular plot that was regarded as subplot with radii of 3.99 m (50m²) that was placed with their center at 5m, 125m and 245m from the plot starting point along the plot central axis (see Figure 2). In addition, the CCSP is easy to locate in a rugged and fragile topography like in Bangladesh and it has less edge effect as compared to square/rectangular plot. All the living trees above the diameter of 5cm was tallied and measured within the radius of the circular plot.

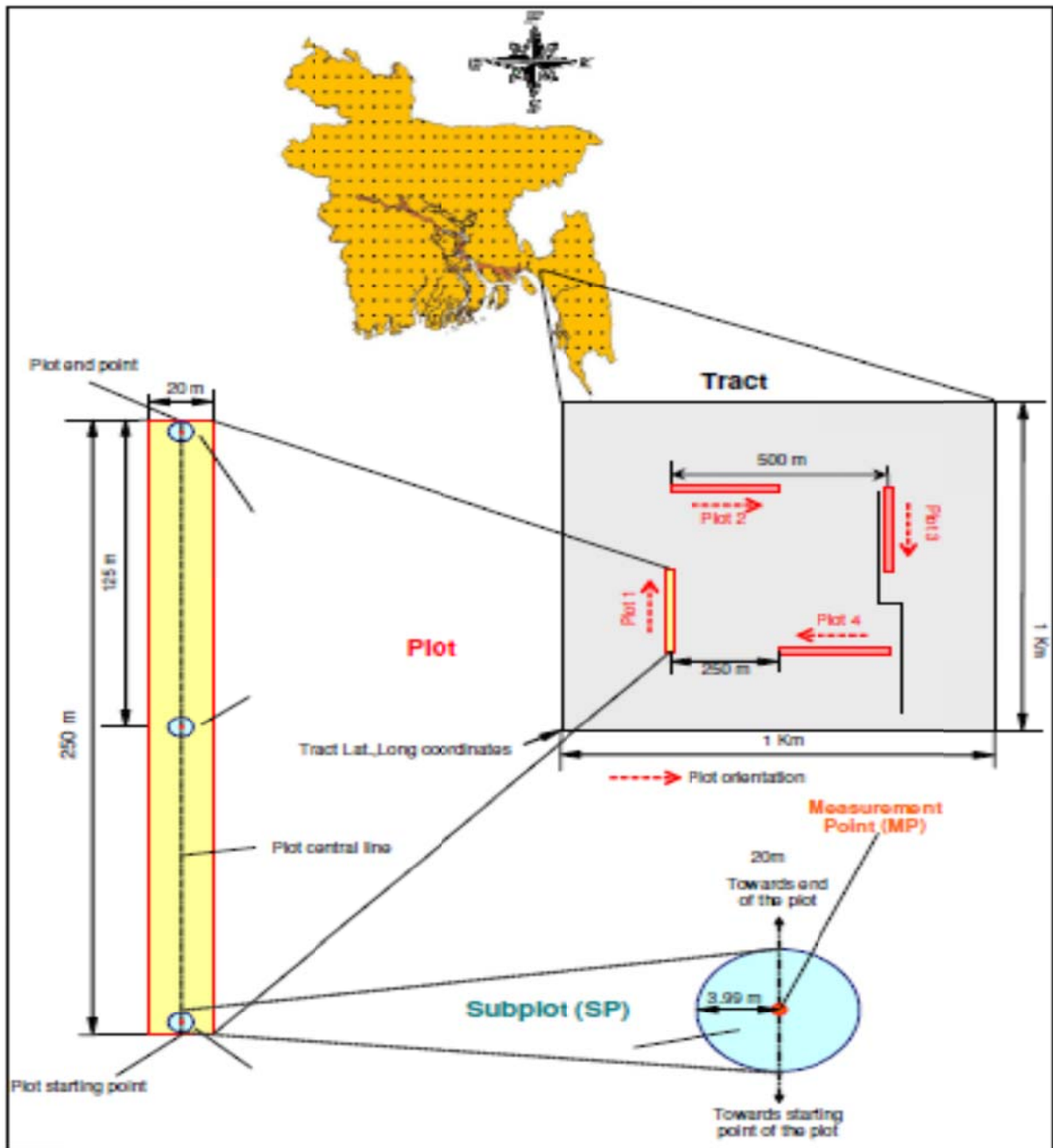


Figure 2: Tract, plot and sub-plot design (NFTRA 2007)

4. Sundarban Mangrove Forest, a priority area for LiDAR based forest inventory

The Sundarbans is the single largest continuous mangrove forest of the world with an area of about 6,01,700 hectare which is 4.07% of total land mass of the country and 40% of total forest land. Sundarban harbours 334 species of trees, shrubs and epiphytes and 269 species of wild animals. The forest inventory of 1998 exhibits that there are 12.26 million cubic meter timber is available from the species of Sundri (*Heritiera fomes*), Gewa (*Excoecaria agallocha*), Keora (*Sonneratia apetala*), Baen (*Avecennia officinalis*), Dhundul (*Xylocarpus granatum*), Passur (*Xylocarpus mekongensis*) etc with 15cm and above diameter (NFTRA 2007). Sundri is the most important tree species in the Sundarban which is distributed over 73% of the reserve. Extent of Sundri is followed by Gewa (*Excoecaria agallocha*), Baen (*Avecinnia offcecinialis*), Passur (*Xylocarpur mekongensis*), Keora (*Sonneratia apetala*) etc. There are some other non-wood forest products like Golpata (*Nypa fruticans*), honey,

wax, fish, crab etc which are also of high value. Considering the importance of preserving biodiversity of the Sundarbans, the UNESCO had on December 6, 1997, declared the forest as the 798th 'World Heritage Site'. It is also one of the two RAMSAR sites of the country.

However, mangrove forest is facing severe problem including over exploitation of forest resources, conversion of forest stand to shrimp pond. In addition, nowadays, salinity intrusion due to declining fresh water flows is also going a big issue. Furthermore, tree mortality especially top dying of *Sundri* tree makes an uncertain future of mangrove forest. Poor forest management strategy with weak law enforcement reducing the productivity of mangrove forest.

A good forest management plan is necessary for Sundarban mangrove forest which is ecologically suitable, economically feasible, and socially acceptable that achieves the core objectives of sustainable forest management. Till now there is no record of application of airborne LiDAR system for forest inventory in Bangladesh. Therefore, an integrated approach like as LiDAR mapping in this forest is recommended. Area based LiDAR inventory is recommended in Sundarban mangrove forest for accurate carbon estimation for REDD plus monitoring, reporting and verification. The strength of area based method lies in their robustness, relative simplicity, lower computational requirements and proven ability to generate unbiased estimates of dynamic and inaccessible Sundarban mangrove forest attributes including, carbon stock, basal area, volume, mean diameter and mean height.

6. Potentiality and advantage of LiDAR application for forest inventory in Bangladesh

Forest inventory in tropical countries like Bangladesh is quite difficult and challenging. Due to heterogeneous characteristics of forest, it takes long time and more costly for measurement of forest resources. In addition, the access to the forest area is also difficult due to steep terrain and hilly area. Remote sensing technique is a important way to measure the forest resources without entering the forest. However, conventional remote sensing technique is not efficient for measurement of forest resources in hill forest of Bangladesh. Airborne laser scanning is an active remote sensing technique that permits observation of the vertical structure of forest. LiDAR data has a great advantage compared to conventional optical remote sensing data/imagery, which suffers from saturation and only shows the topmost layer of the vegetation, while laser pulses penetrate through even a dense canopy.

LiDAR is able to meet the demanding accuracy requirements of operations planning better than any previous technology, also in large-scale forest inventory. As species diversity is very high in Bangladesh, the accurate measurements of carbon stock in each type of forests are very necessary. Therefore, the greatest advantage and potentiality of LiDAR is that it is highly capable of monitoring three-dimensional forest structure. Forest carbon content can be measured very accurately from LiDAR pulse with describing estimated tree height, above-ground biomass, timber volume and crown parameters. LiDAR can also measure forest biomass from individual trees to vegetation on a local, regional and national level.

LiDAR technology will be well-suited for forest inventory in Bangladesh as it is less sensitive to weather conditions and sun angles than satellite imaging technology. In addition, the integration of Airborne Laser Scanning (LiDAR), satellite image and field measurement with advanced statistical models on sample plots, it becomes most powerful tool for sustainable forest resource management. Finally it is already proved that the LiDAR based carbon stock measurement is more reliable in tropical, temporal and other regions (Figure 3) (Næsset 2009)

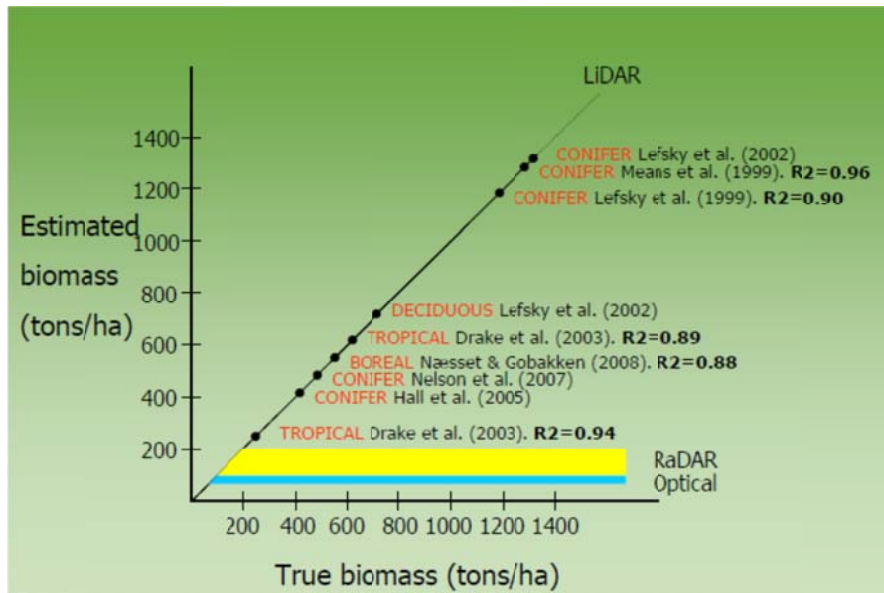


Figure 3: Airborne LiDAR data for forest biomass estimation in different forest type

Some of advantages of LiDAR technology are given below

1. LiDAR based forest inventory reduces the costs for large areas
2. LiDAR gives very high resolution digital elevation information for large forest areas than conventional Digital Elevation Models (DEMs)
3. LiDAR pulse can penetrate very dense forest vegetation
4. LiDAR data can collected during day or night under clean weather condition and less suffer from saturation problem
5. Nowadays, integration of LiDAR data with aerial images and field measurement are the best way of forest resource management.
6. The interpreted LiDAR data layers are easy to integrate with other form of data in GIS analysis.
7. LiDAR data helps pinpoint locations where filed data will be useful

7. Limitation of LiDAR application for forest inventory in Bangladesh

1. High costly technique on a per-acre basis than aerial photography considering Bangladesh perspective
2. Require skilled manpower and software for processing and analyzing LiDAR data sets
3. Although the growing use of LiDAR technology for forest resource management or forest inventory, all the parameters cannot measured by using LiDAR data and the models are not always understood well enough to generalize final findings from local studies.
4. In the tropical forest like forest of Bangladesh where density of vegetation is so high that dense undergrowth may be confused with bare ground.
5. From the previous studies, it was found that vegetation height calculated from LiDAR data is less than the height obtained through aerial images and field measurement.
6. Poor economic condition of Bangladesh
7. Logistic problem
8. Political and bureaucratic problem

8. Concluding remarks

Above all, the LiDAR technology with satellite imagery and field measurement based forest inventory for accurate carbon stock measurement and robust benefit sharing with collaboration of local and

indigenous people has significant prospect in Bangladesh. The Forest Department and Government of Bangladesh with financial or technical help from international organization can make a prosperous and effective management of forest resources. For Tier 3 level REDD plus monitoring, reporting and verification of forest resources, this hybrid method (LiDAR with satellite imagery and ground truth data) will be very effective in Bangladesh. This integrated system will be able to produce consistent reports and a source of verification for any geographic area in Bangladesh.

9. Reference

- Bang, K. I., Habib, A.F., Kusevic, K., Mrstik, P., 2008. Integration of terrestrial and airborne lidar data for system calibration. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B1. Beijing 2008 Commission I, WG I/2. Bangladesh Forest Department, 2008. Banabhaban, Agargaon, Dhaka-1207, Bangladesh.
- Gautam, B.R. and Kandel, P.N., 2010. Working paper on LiDAR mapping in Nepal. Submitted to the Ministry of Forests and Soil Conservation, Federal Democratic Republic of Nepal.
- Li, Y., 2009. Towards Small-footprint Airborne LiDAR-assisted Large Scale Operational Forest Inventory. A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, University of Washington, USA.
- Mukul, S.A., Uddin, M.B., Uddin, M.S. and Khan, M.A.S.A., 2008. Protected areas of Bangladesh: current status and efficacy for biodiversity conservation. *Proc. Pakistan Acad. Sci.*, 45(2), 133-140.
- Naesset, E., 2004 Practical large-scale forest stand inventory using a small footprint airborne scanning laser. *Scand. J. For. Res.*, 19, 164-179.
- Næsset, E., 2009. Some challenges in forest monitoring. Proposing a “research platform” for development, training, and validation of spaceborne technologies in Brazil. A Presentation. Brazilian Norwegian Workshop on Forest and Marine Monitoring, Ilhabela, SP, Brazil, March 9-12, 2009. 24 slides.
- NFTRA, 2007. National Forest and Tree Resources Assessment 2005-2007, Bangladesh. Bangladesh Forest Department, Banabhaban, Agargaon, Dhaka-1207, Bangladesh.
- Poffenberger, M., 2000. Communities and forest management in South Asia. IUCN, DFID and Asia Forest Network, Indonesia. 35-46pp.
- Rana, M.P., Uddin, M.S., Chowdhury, M.S.I., Sohel, M.S.I., Akhter, S., Koike, M., 2009. Current status and potentiality of forest resources in a proposed biodiversity conservation area of Bangladesh, *Journal of Forest Science*, 25(3), 167-175.
- Roy, M.K., 2005. Nishorgo Support Project: Designing a Co-management Model for the Protected Areas of Bangladesh. In: Nishorgo, Protected Area Management Program of Bangladesh. Dhaka, Bangladesh. 1-5 pp.
- Siddiqi, N.A., 2001. Mangrove forestry in Bangladesh. Institute of Forestry and Environmental Sciences (IFES), University of Chittagong, Chittagong. 201 pp.
- UNREDD, 2010. <http://www.un-redd.org/> last accessed 4th August 2011
- van Laar, A., Akca, A., 1997. Forest Mensuration, Cuvillier Verlag, Gottingen 418 p.

Stand level inventory of eucalypt plantations using small footprint LiDAR in Tasmania, Australia

Dr Robert Musk

Forestry Tasmania rob.musk@forestrytas.com.au

Abstract

Models derived using Breiman's Random Forests algorithm have been identified in past studies as having greater predictive accuracies than those derived using nearest neighbour imputation approaches. This is attributed to the algorithms ability to model complex interactions among predictor variables and its resistance to overfitting. These two properties are of particular value in modelling LiDAR-derived variables where strong colinearity is a common feature. In this study, the random forest algorithm is applied to a large inventory dataset to generate mapped estimates of forest stand structure. The ability of the algorithm to identify an optimal set of candidate variables is assessed by means of an iterative model fitting procedure. The study area comprises a eucalypt hardwood plantation estate in northern Tasmania, Australia. Model pseudo R^2 values were 74.6% for basal area, 96.0% for mean dominant height, 64.2% for stocking and 83.9% for merchantable stand volume respectively.

1. Introduction

Forestry Tasmania is the agency responsible for the management of Tasmanian state forests. This estate is managed for a variety of outcomes, including conservation and timber production. Embedded within the estate is a plantation resource of *Eucalyptus nitens* and *E. globulus* covering 42 000 Hectares that is predominantly managed for a high value sawlog yield. Most of this plantation comprises first rotation plantings on sites formerly carrying native forest. The spatial variability in stand condition and woody weed load is very high, as is the demand for precise wood inventory. Until recently, field-based inventory techniques have been the sole source of the necessary planning information. Consequently, management options have been constrained by the poor spatial resolution of information that field-based inventory can provide. In the past 18 months the agency has commenced acquisition of small footprint discrete return LiDAR data over the entire estate. Amongst several other objectives, the organisation seeks to improve the precision of its wood inventory using this dataset. The approach taken to realise this objective follows that of numerous other workers (e.g. Evans et al., 2006; Lim et al., 2003; Magnussen & Boudewyn, 1998; Næsset, 1997a; Næsset, 1997b; Nelson et al., 1988; Reutebuch et al., 2005; Ritchie et al., 1993) and involves the derivation of raster-based LiDAR variables, such as height quantiles and proportion of returns from within, or below, specific height classes, which are then used to develop models to predict and map stand-level metrics of forest structure.

An important strand of research within the field of LiDAR-based forest inventory has focussed upon identifying optimal statistical approaches to model development and application. Early research utilised traditional parametric approaches, such as stepwise least squares regression, as statistical inference tools (e.g. hypothesis testing, confidence limits etc.) were central to fulfilling the objective of establishing the validity of LiDAR as an inventory data source. More recently, research has focussed on identifying modelling approaches that can be used to extract the maximum information content from the LiDAR data. The Random Forests (RF) algorithm (Breiman 2001) has been identified as the most accurate in a number of forest remote sensing applications when compared against nearest neighbour imputation approaches (Falkowski et al.,

2009; Hudak et al., 2008; Latifi et al., 2010; Lawrence et al., 2006), principally because of its ability to model complex interactions among predictor variables and its resistance to overfitting (Cutler et al., 2007). These two properties are of particular value in modelling LiDAR-derived variables where strong colinearity is a common feature. The major drawbacks of the approach are twofold. First, as a nonparametric classifier, an RF model cannot generate predictions outside the range of the data used to build it and so relies on comprehensive sampling of the target population. Second, RF cannot be used for statistical inference. The algorithm generates a ‘forest’ of bootstrap sampled Classification and Regression Trees (CART). The importance of predictor variables can be assessed solely using measures of change in model mean square error and pseudo- R^2 with variable removal (Cutler et al., 2007). In this study, the random forest algorithm is applied to a large inventory dataset to generate mapped estimates of forest stand structure. The ability of the algorithm to identify an optimal set of candidate variables is assessed by means of an iterative model fitting procedure.

2. Methods

2.1 Study area

The study area comprises Forestry Tasmania’s eucalypt hardwood plantation estate covering approximately 16 800 Ha in the north east of Tasmania centred about $43^{\circ} 20' S$ $147^{\circ} 45' E$. Figure 1 presents a map of Tasmania showing the extent of the State Forest managed by Forestry Tasmania and the plantation resource in the north east of the state. These plantations are managed for a variety of harvest products, but the principle aim is to produce high value, pruned solid wood. To this end, most stands are managed over a 22-25 year rotation with three pruning lifts on selected trees and a single thinning operation between 5 and 12 years of age. Management decisions regarding the scheduling, and within-compartment extent of pruning, thinning and harvesting operations rely on access to precise stand information.

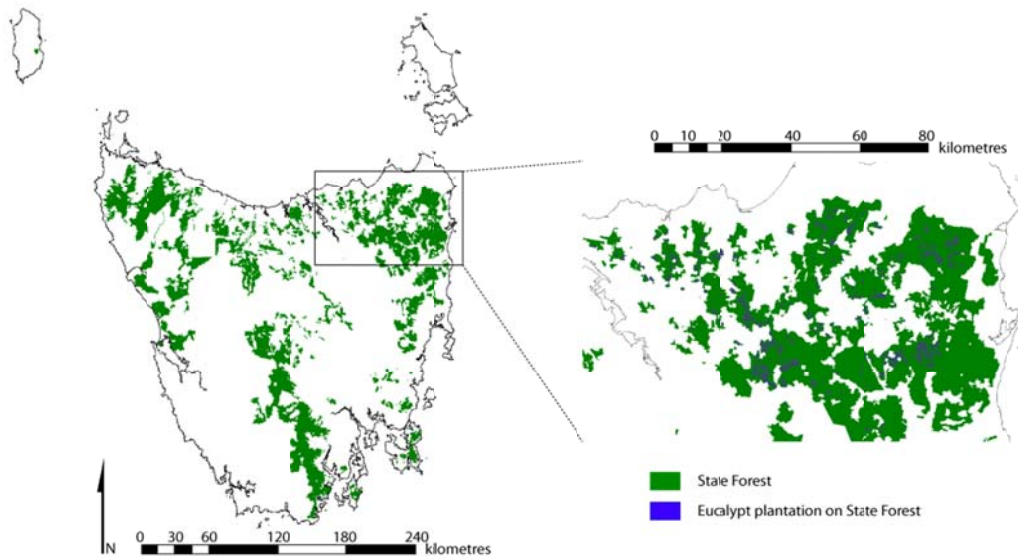


Figure 1: Tasmanian State Forest, with the north east eucalypt plantation estate highlighted. The north east region is centred about $43^{\circ} 20' S$ $147^{\circ} 45' E$.

2.2 Field data

Field plots ($N = 157$) were established using a stratified random sampling schema. Compartment age formed the basis of the strata and all field plot measurements were undertaken during the summer of 2009-2010. Field plots were circular with a 12.3 m radius (450 square metres). Field plot centres were located using a Trimble Pro-XR GPS unit. At least 200 GPS

points were recorded at each plot. These were differentially corrected in a post processing step that ensured sub-metre planimetric (X, Y) standard errors for all plots. On each field plot, the magnetic bearing and horizontal distance to all trees with stem overbark diameter at 1.3 metres above ground (DBH) over 5cm was recorded. The tree species, total height and DBH of all trees over 5cm was recorded. Potential stem product at harvest was also assessed and recorded using Forestry Tasmania inventory standards. All tree DBHs were measured using diameter tapes. All tree total heights and distances were measured using a Vertex ultrasonic hypsometer. Ocular estimates of average canopy height and percentage canopy cover for all understory species with estimated canopy cover over 10% were also recorded. Three photos were taken from the plot centre providing a qualitative record of field plot conditions. These were taken facing north, south and vertically.

The following response variables were calculated using the commercial tree species data:

1. *BA*: Basal area
2. *MDH*: Mean dominant height (*the average height of the 50 tallest trees per hectare*)
3. *N*: Stocking
4. *V*: Merchantable stand volume (*calculated using unpublished tree-level stem volume models developed by Forestry Tasmania*).

2.3 LiDAR data and processing

AMM acquired the LiDAR data during the summer of 2009-2010 using an Optech Gemini discrete-return scanner. The maximum scan angle was set at 25° from nadir and the minimum point density was 200 points per 10 square meters. Up to four returns were recorded per pulse. The data were classified as ground or non-ground using proprietary algorithms and delivered in LAS 1.2 format. Each LAS file comprised a 600×600 metre tile with a 50 metre overlap to tiles adjacent. Vegetation height was calculated after interpolating a 1 metre resolution digital terrain model from the ground-classified returns using a multi-level B-spline (Lee et al. 1997).

LiDAR returns with a 5 metre buffer that were spatially coincident to the field plots were extracted from the delivered data. Rasters of interpolated LiDAR returns and 3D graphs for each plot were produced. A particularly striking feature of the data was the high degree of discrimination between the eucalypt and non-eucalypt returns based on the return intensity. Eucalypt canopy elements typically displayed return intensities less than 18, while non-eucalypt canopy elements typically displayed return intensities greater than 15. The most common non-eucalypt elements were species of the *Acacia* genus.

A range of candidate predictor variables were extracted from the LiDAR data by means of two separate processes. In the first process, the returns were clipped to the extents of the field plot boundaries and numerous grid-based candidate predictor variables were then extracted. These comprised percentiles and moments of both vegetation height and intensity, proportions of returns within vegetation height and intensity strata defined in both absolute and relative terms and two measures of scan angle incidence: the average scan angle and the scan angle range (the latter, a metric that identifies whether the points arise from one or more flight paths). In the second process, a 0.5 metre resolution canopy height model (CHM) was interpolated from the vegetation height data for each field plot. These CHMs were Gaussian filtered with a 3×3 cell kernel. A simple min-max filter was then applied to identify high points in the data. The CHMs were then clipped to the extents of the field plot boundaries and a numerous grid-based candidate predictor variables were then extracted. These comprised total cover (the proportion of pixels with vegetation height > 2 metres), total volume (the sum vegetation height), the number of height maxima identified using the min-max filter, and the moments of the CHM grid cells.

A complete list of the candidate predictor variables derived through these two processing steps appears in Table 1. It includes the age and the species of the commercial crop as derived from the operations database. Several other potential predictors were not used in this study. In particular, the timing and intensity of any pruning and/or thinning operations was ignored since the spatial accuracy of operational boundaries is not sufficient for high resolution mapping purposes. In total, 75 candidate predictor variables were available for modelling.

Table 1: Candidate predictor variables derived from the LiDAR data

Variable(s)	Number of variables	Variable description
<i>hp10, hp20, ... , hp100</i>	10	Height percentiles (10%, 20%, ... , 100%)
<i>rhp10, rhp20, ... , rhp90</i>	9	Height density (proportion of returns with height greater than: 10%, 20%, ..., 90% of maximum height)
<i>hs10, hs20, ... , hs50</i>	5	Height strata (hp90 – hp80, hp90 – hp70, ... , hp90 – hp40)
<i>rhs10, rhs20, ... , rhs50</i>	5	Canopy density strata (rhp90 – rhp80, rhp90 – rhp70, ... , rhp90 – rhp40)
<i>hvar, hskew, hkurt</i>	3	Height moments (variance, skewness, kurtosis)
<i>ip10, ip20, ... , ip100</i>	10	Intensity percentiles (10%, 20%, ... , 100%)
<i>rip10, rip20, ... , rip90</i>	9	Intensity density (proportion of returns with intensity higher than: 10%, 20%, ..., 90% of maximum intensity)
<i>is10, is20, ... , is50</i>	5	Intensity strata (ip90 – ip80, ip90 – ip70, ... , ip90 – ip40)
<i>ris10, ris20, ... , ris50</i>	5	Intensity density strata (rip90 – rip80, rip90 – rip70, ... , rip90 – rip40)
<i>ivar, iskew, ikurt</i>	3	Intensity moments (variance, skewness, kurtosis)
<i>hsurfcov, hsurfvol, hsurftree, hsurfvar, hskew, hkurt</i>	6	Height surface metrics (total cover, total volume, number of height maxima, variance, skewness, kurtosis)
<i>meanscan, rangescan</i>	3	Scan angle (mean, range)
<i>eucAge, eucSPP</i>	2	Commercial crop age and species

2.4 Modelling approach

While the random forests algorithm is generally considered to be robust to overfitting and capable of generating highly accurate predictions in the presence of spurious predictor variables (Cutler, et al., 2007) it was thought necessary to assess this. Of particular interest was the predictive accuracy of the RF model in the presence of a large number of candidate predictor variables. In order to investigate this, a two-stage iterative approach was taken.

In the first stage, RF models were constructed using the all potential predictor variables. The least important predictor variable was then identified as that which caused the least decrease in model pseudo- R^2 upon removal. This predictor variable was then removed from the list of candidate predictor variables and the process repeated until no variables were available for model fitting. As the RF algorithm relies upon bootstrap sampling of the data, the result of this iterative model construction is conflated with the bootstrap sampling error. In order to identify the relative importance of candidate predictor variable removal and the bootstrap sampling error, this iterative process was run 25 times. Upon completion, the results were inspected and the optimal number of candidate predictor variables identified for each response variable.

In the second stage, a similar iterative process was then run 50 times. In each iteration, the process was stopped when the model contained the number of candidate predictor variables identified as optimal in the first stage. At the completion of each iteration the resulting identified predictor variables were recorded. Final model building was undertaken using the list of candidate predictor variables identified in this second stage.

In all cases, RF models were constructed from 500 CARTs. The number of predictor variables randomly sampled as candidates at each split in each CART was set to the total available candidate predictor variables divided by 3. All bootstrap sampling was done with replacement. The minimum size of terminal nodes in each CART was set to 5.

All predictor variables identified in the final RF models were derived over the study area at a 4 meter cell resolution. The chosen random forests models were then used to generate predictions for the response variables across the plantation compartments within the study area.

All LiDAR processing and modelling work was undertaken using open-source software. The principle software tool used was the R language and environment (R Development Core Team 2011) with the additional packages: `sp` (Pebesma and Bivand, 2005; Bivand, Pebesma and Gomez-Rubio, 2008), `MBA` (Finley and Banerjee, 2010), `data.table` (Dowle and Short, 2011), `moments` (Komsta and Novomestky, 2007), `rgdal` (Keitt et al., 2010), `foreach` (Revolution Analytics, 2011) and `Rmpi` (Hao Yu, 2010). The high computing load involved in generating large area rasters was accommodated using 30 Xeon cores in a virtual machine cluster that was managed using openMPI (Gabriel et al., 2004) on an openSUSE 11.0 Linux-based operating system. The processing time using this cluster averaged 0.43 seconds per hectare.

3. Results

3.1 Field data

Scatterplots of the four response variables (BA, MDH, N, V) and two candidate predictor variables (eucAge and eucSPP) appear in Figure 2. Strong relationships are apparent between merchantable stand volume, basal area and mean dominant height. Weaker relationships appear between age and stocking, and other variables. Stocking varies widely and is quite high in some plots, reflecting the presence of double leader stems in many areas. Basal area, stocking and merchantable stand volume all show skew that is typical in these stands. Most of the plots comprise plantings of *Eucalyptus nitens* (eucSPP = 1). There are no substantial response variables differences between crop species.

3.2 Random Forest modelling

Figure 3 present the results of the stage one iterative process in which candidate predictor variables identified as the least important in the RF model built in the first step were removed from the candidate list in the second step. Each coloured line represents the results of a single run with iterative removal of candidate predictor variables. The scatter between the lines about

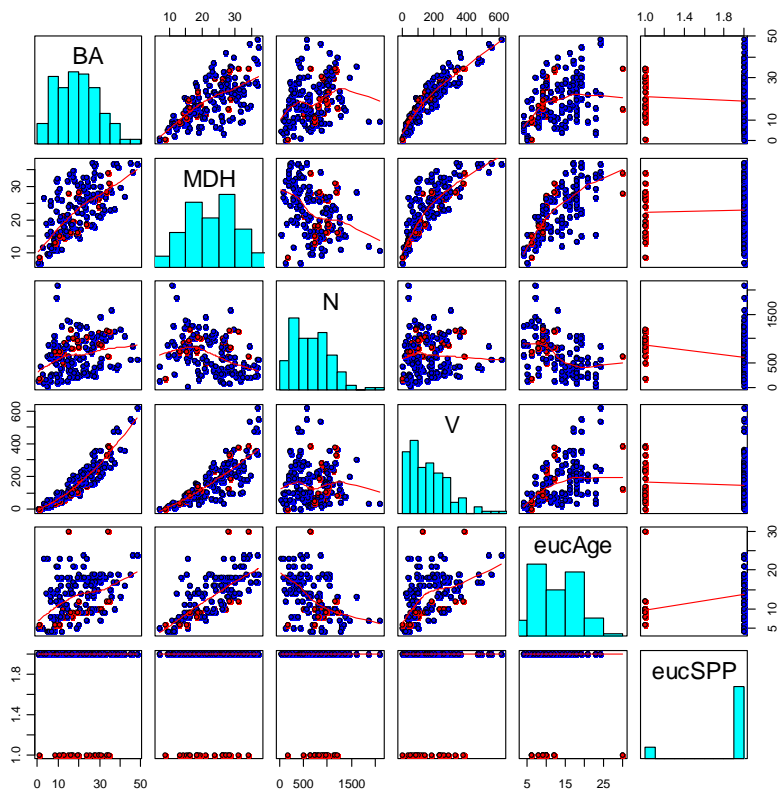


Figure 2: Scatterplots matrices of field plot data: the four response variables (BA, MDH, N, V) and two candidate predictor variables (eucAge and eucSPP). *Eucalyptus nitens* is coded eucSPP = 1. *E. globulus* is coded eucSPP = 2.

the model pseudo R^2 values reflects bootstrap error. In each case, model pseudo R^2 increases slightly with iterative removal, with the smallest increases observed for the mean dominant height model and the largest increases for the stocking model. Evident from these plots is the degree to which model accuracy is impacted by the presence of spurious predictor variables. In particular, the stocking model performs poorly when all candidates are available for fitting and shows a marked improvement with candidate removal. The relationship does not appear to be solely a function of model pseudo R^2 as the merchantable stand volume model shows a larger improvement than does the basal area model, yet also has a larger model pseudo R^2 across the number of available candidate predictor variables. The bootstrap sampling error is a substantial component of the signal in all cases.

Viewing Figure 3, it is evident that approximately 15 predictor variables are required to construct RF models for these forest metrics with maximum predictive accuracy. In stage two of the modelling process, the iterative procedure was rerun 50 times until only 15 predictor variables were left in each model. The list of predictor variables by the number of times each appeared in the final models appears in Figure 4. From the original list of 75 candidate predictor variables, 50 remain. Of these, 39 appear in 10 or more of the 50 models constructed for one or more of the response variables. Of note is the strong degree of separation in predictor variables. Of the 23 variables used to predict stocking, 9 predict nothing else. Of the 21 variables used to predict basal area, 2 predict nothing else. Of the 27 variables used to predict mean dominant height, 12 predict nothing else. Only those 23 variables used to predict merchantable stand volume are shared with other models, reflecting the fact that merchantable stand volume largely integrates the other stand metrics. Also of note is the large number of return intensity derived

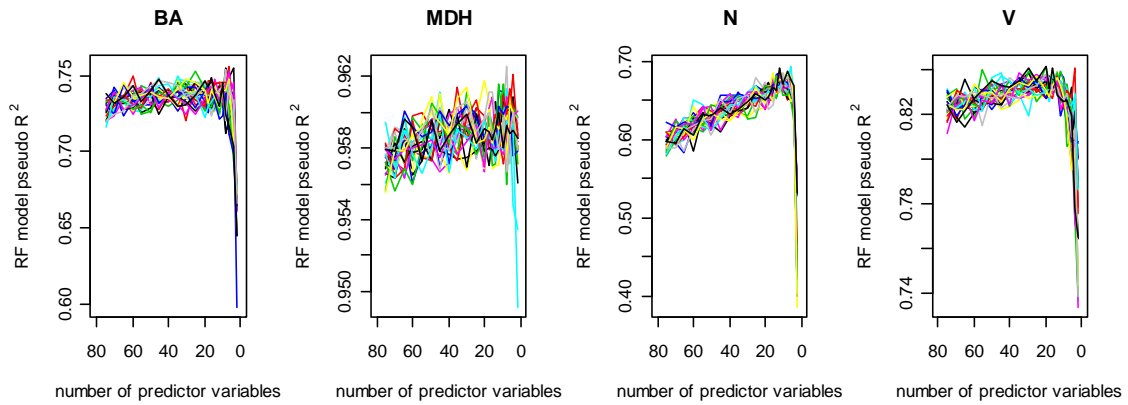


Figure 3: Stage one results. RF model pseudo R² versus the number of candidate predictor variable available for RF modelling. Each coloured line represents the results of a single run with iterative removal of candidate predictor variables.

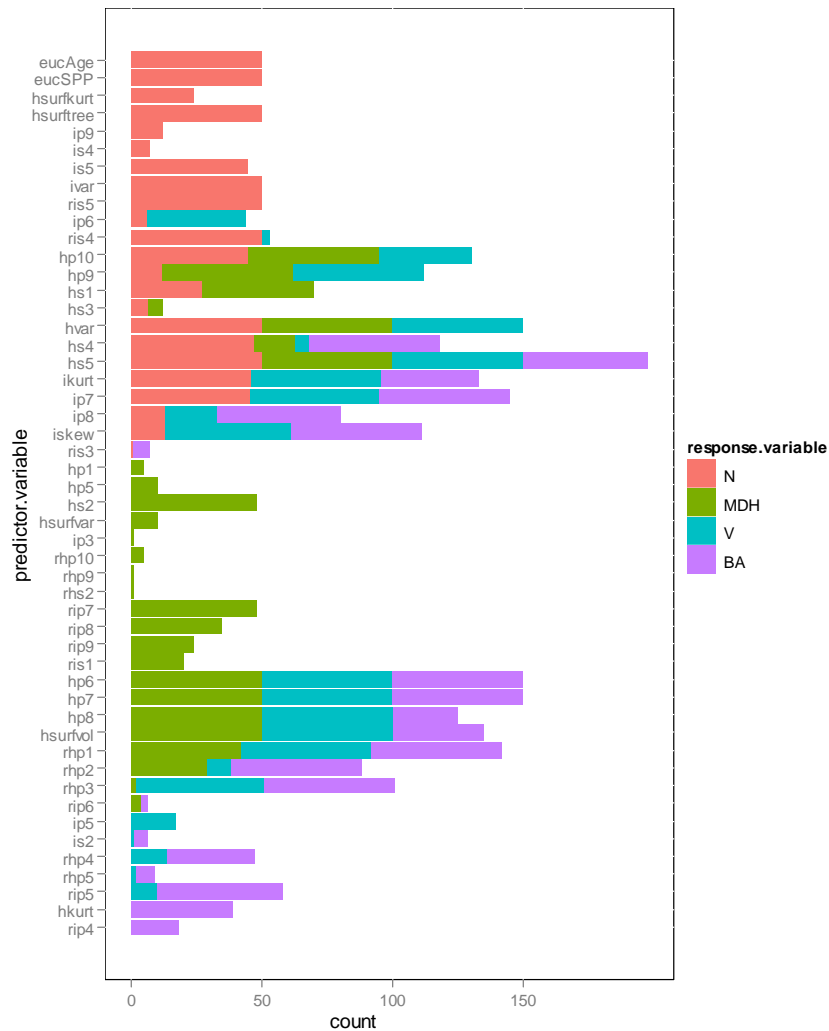


Figure 4: Stage two results. Candidate predictor variables versus the number of iterations in which they are identified in the list of final 15 predictor variables for each of the four response variables

metrics amongst the identified predictor variables indicative of the importance of return intensity in discriminating between eucalypt and non-eucalypt species.

In the final step, the predictor variables identified as candidates in the two stage iterative process were used to construct RF models to be applied across the estate. Figure 5 presents scatterplots of observed versus predicted response variables. Model performance is generally good, with pseudo R^2 values of 74.6% for basal area, 96.0% for mean dominant height, 64.2% for stocking and 83.9% for merchantable stand volume respectively; and root mean square error values of $5.109\text{m}^2\text{ha}^{-1}$ for basal area, 1.490m for mean dominant height, 215.6 stems ha^{-1} for stocking and $48.83\text{m}^3\text{ha}^{-1}$ for merchantable stand volume. Both basal area and merchantable stand volume display some positive bias in the lowest quartile and one or more outliers are present for all models. The very low error in mean dominant height prediction provides some comfort that field data acquisition standards are well implemented as this value approaches the measurement error rate expected in this forest type.

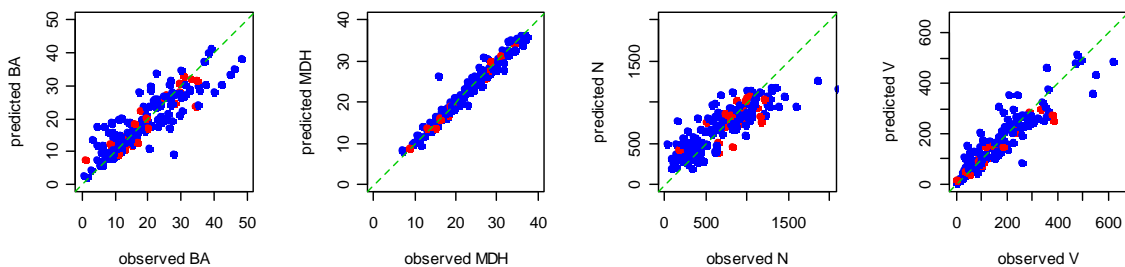


Figure 5 Observed response variables versus those predicted using the RF models derived using the final list of candidate predictor variables. *Eucalyptus nitens* is blue. *E. globulus* is red.

4. Discussion

The random forests algorithm was used to build predictive models that displayed accuracies broadly similar to those published elsewhere (e.g. Lefsky et al., 1999; Magnussen and Boudewyn, 1998; Maltamo et al., 2004; Næsset, E., 2002; Næsset and Økland, 2002). The iterative procedure used to eliminate spurious candidate predictor variables improved the prediction accuracy of the stocking and merchantable stand volume models, but not the basal area or mean dominant height models. Some researchers have considered LiDAR-based estimates of tree and stand height be more accurate than field-based measurements (Næsset and Økland, 2002; Coops et al., 2007). The mean dominant height model presented in this study displayed a predictive accuracy similar to the measurement precision expected of Forestry Tasmania field crews, so these results do not contradict that view. The capacity of LiDAR to accurately measure forest vertical structure leads to models of forest attributes such as basal area that display predictive accuracies somewhat greater than one might expect given the characteristic allometric relationships between stand metrics observed in these forests. The degree to which separate models utilised separate sets of LiDAR-derived metrics is likely to explain this discrepancy. An example of this separation was seen in the return intensity metrics in particular, where the distinctive intensity signature of the eucalypt crop species was exploited to predict stocking.

Future work with this dataset will focus on two areas. Firstly, an appropriate nearest neighbour imputation approach will be identified that can be used to assign field inventory derived stem quality data to raster surfaces and more accurately represent the stand level metric covariance structure. While recognising this will sacrifice predictive accuracy of individual stand level metrics, it will allow the LiDAR data to be used in Forestry Tasmania's growth and yield system to project stand conditions through time and inform strategic planning decisions.

Second, alternative approaches to single tree identification will be investigated to improve the predictive performance of the stocking model in older, more valuable stands.

5. Conclusions

LiDAR data was evaluated for its efficacy in mapping stand-level inventory metrics in Tasmanian eucalypt plantations. The study found model performances were broadly in accordance with other LiDAR studies, with model pseudo R^2 values of 74.6% for basal area, 96.0% for mean dominant height, 64.2% for stocking and 83.9% for merchantable stand volume respectively. An iterative procedure was used to eliminate spurious candidate predictor variables. This process improved the prediction accuracy of the stocking and merchantable stand volume models, but not the basal area or mean dominant height models. LiDAR return intensity was an important predictor variable in all models, reflecting the differences observed in return intensity between eucalypt, and non-eucalypt species.

References

- Bivand, R.S., Pebesma, E.J. And V. Gomez-Rubio (2008) *Applied spatial data analysis with R*. Springer, NY. <http://www.asdar-book.org/>
- Breiman, L. 2001. *Random forests*. Machine Learning 45:15–32.
- Coops, N.C., Hilker, T., Wulder, M.A., St-Onge, B., Newnham, G., Siggins, A., Trofymow, J.A., 2007. Estimating canopy structure of Douglas-fir forest stands from discrete return LiDAR. *Trees* 21, 295–310.
- Cutler, R. D., Edwards, T. C., Jr., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., et al. (2007). Random Forests for classification in ecology. *Ecology*, 88, 2783–2792.
- Dowle, M. and T. Short. (2011). *data.table: Extension of data.frame for fast subset, fast merge and fast grouping*. R package version 1.5.3. <http://CRAN.R-project.org/package=data.table>
- Evans, D. L., S. D. Roberts and R. C. Parker (2006). LiDAR – A new tool for forest measurements? *The Forestry Chronicle* 82(2): 211–218.
- Falkowski, M.J., Evans, J.S., Martinuzzi, S., Gessler, P.E. and A.J. Hudak Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA. *Remote Sensing of Environment* 113 (2009) 946–956
- Finley, A.O. and S. Banerjee (2010). *MBA: Multilevel B-spline Approximation*. R package version 0.0-7. <http://CRAN.R-project.org/package=MBA>
- Gabriel, E., Fagg, G.E., Bosilca, G., Angskun, T., Dongarra, J.J., Squyres, J.M., Sahay, V., Kambadur, P., Barrett, B., Lumsdaine, A., Castain, R.H., Daniel, D.J., Graham, R.L., and T.S. Woodall (2004) Open MPI: Goals, Concept, and Design of a Next Generation MPI Implementation. In *Proceedings, 11th European PVM/MPI Users' Group Meeting*, Budapest, Hungary, September 2004.
- Hudak, A. T., Crookston, N. L., Evans, J. S., Hall, D. E., & Falkowski, M. J. (2008). Nearest neighbor imputation modeling of species-level, plot-scale structural attributes from lidar data. *Remote Sensing of Environment*, 112, 2232–2245.
- Keitt, T.H., Bivand, R.S., Pebesma, E.J. and B. Rowlingson (2010). *rgdal: Bindings for the Geospatial Data Abstraction Library*. R package version 0.6-28. <http://CRAN.R-project.org/package=rgdal>
- Komsta, L. and F. Novomestky (2007). *moments: Moments, cumulants, skewness, kurtosis and related tests*. R package version 0.11. <http://www.r-project.org>, <http://www.komsta.net/>

- Latifi, H., Nothdurft, A. And B. Koch (2010) Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. *Forestry* 83(4): 395-407
- Lawrence, R. L., Wood, S. D., & Sheley, R. L. (2006). *Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest)*. *Remote Sensing of Environment*, 100, 356–362.
- Lee, S., Wolberg, G. and S. Y. Shin (1997) Scattered data interpolation with multilevel B-splines *IEEE Transactions on Visualization and Computer Graphics*, 3(3):229-244
- Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A., Harding, D., (1999). Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment* 70, 339–361.
- Lim, K., P. Treitz, K. Baldwin, I. Morrison and J. Green (2003). Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing* 29(5): 658-678.
- Magnussen, S. and P. Boudewyn (1998). Derivations of stand heights from airborne laser scanner data with canopy based quantile estimators. *Canadian Journal of Forest Research* 28(7): 1016-1031.
- Maltamo, M., Eerikäinen, K., Pitkänen, J., Hyypä, J., Vehmas, M., 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote Sensing of Environment*. 90, 319–330.
- Næsset, E. (1997a). Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sensing of Environment* 61(2): 246-253.
- Næsset, E. (1997b). Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing* 52(2): 49-56.
- Næsset, E., (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 90, 88–99.
- Næsset, E. and T. Økland, (2002). Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment* 79, 105–115.
- Nelson, R., R. Swift and W. Krabill (1988). Using airborne lasers to estimate forest canopy and stand characteristics. *Journal of Forestry* 86(10): 31-38.
- Pebesma, E.J. and R.S. Bivand (2005) Classes and methods for spatial data in R. *R News* 5 (2), <http://cran.r-project.org/doc/Rnews/>.
- R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org/>.
- Reutebuch, S. E., H. Andersen and R. J. McGaughey (2005). Light Detection and Ranging (LIDAR): An Emerging Tool for Multiple Resource Inventory. *Journal of Forestry* 103(6): 286 - 292.
- Revolution Analytics (2011). *foreach: Foreach looping construct for R*. *R package* version 1.3.2. <http://CRAN.R-project.org/package=foreach>
- Ritchie, J. C., D. L. Evans, D. Jacobs, J. H. Everitt and M. A. Weltz (1993). Measuring canopy structure with an airborne laser altimeter. *Transactions of the ASAE (American Society of Agricultural Engineers)* 36(4): 1235-1238.

Building a case for lidar-derived structure stratification for Australian softwood plantations

R.Turner, A.Kathuria and C.Stone

Forest Science Centre, Department of Primary Industries, Sydney, NSW

Russell.Turner@industry.nsw.gov.au

Amrit.Kathuria@industry.nsw.gov.au

Christine.Stone@industry.nsw.gov.au

Abstract:

Conventional resource inventories in Australian softwood plantations usually utilise Geographic Information System (GIS) thematic layers to stratify the resource prior to field sampling. Airborne lidar can offer a viable alternative for stratification but there are no standardised methods for plantation managers. This paper explores issues with current thematic stratification approaches and argues that relatively basic metrics extracted from a lidar-derived canopy height model (CHM) are suitable for constructing better stratification options. The case is supported with findings from an airborne lidar inventory undertaken in a pine plantation in New South Wales (NSW), Australia. Lidar stand level metrics including mean height, mean above mean height, mean dominant height, predicted stocking, canopy cover percentage, occupied volume and height variance were tested as surrogates for plantation structure. The study demonstrated that lidar metrics can predict stand attributes such as age class ($R^2 = 0.91$, RMSE 1.9), thinning treatment (89% accuracy), mean height, ($R^2 = 0.95$, RMSE 4%), stocking ($R^2 = 0.82$, RMSE 26%), basal area ($R^2 = 0.67$, RMSE 19%) and total stand volume ($R^2 = 0.8$, RMSE 19%) across a range of stand structures. Since the metrics tested were highly correlated with survey data it is argued that they could provide a valid basis for a developing a new structure stratification approach to improve sampling design in future plantation resource inventories.

Keywords: Airborne lidar, softwood plantation, stratification and wood resource inventory.

1. Introduction

Australian plantation managers are moving toward the adoption of airborne lidar technology for operational resource assessment. One area of interest to plantation managers is how to use lidar metrics to improve plantation stratification and field sample design. Ground survey costs often represent the greatest expenditure in wood inventory programs and lidar has the potential to significantly improve the cost-efficiency of resource inventories (Maltamo *et al.* 2011).

This paper presents a subset of results from a two year project called the Plantation Airborne Resource Inventory Appraisal (PARIA) sponsored by the Forest & Wood Products Australia (FWPA) and Forest New South Wales (FNSW) (www.fwpa.com.au). The PARIA project explored the use of airborne lidar and multispectral aerial photography for a range of applications (Stone *et al.* 2011; Turner *et al.* 2011) and produced a guide for plantation managers who may not be familiar with remote sensing technology (Turner and Stone 2010).

1.1 Conventional thematic stratification

The main purpose of resource inventory surveys is to collect plantation statistics that are representative of the population. A common traditional approach involves subdividing a plantation into relatively homogenous groups (strata units) and then allocating plots in a

systematic grid to achieve a suitable plot to area ratio (e.g. 1 plot per 4 hectares). Circular plots (0.05 to 0.1 hectares in area) are then field located, measured and later processed to derive a range of plot level attributes expressed on a per hectare basis. Plot estimates are multiplied by the area they represent and then results are aggregated for each stratum and plantation estate.

A successful stratification should ideally meet two important criteria. Firstly the structural variation within each stratum should be as small as possible (i.e. very uniform). Secondly, there should be considerable difference between strata with minimal overlap in structure parameters. In FNSW, resource units are the smallest unit of management and are typically spatially defined with existing GIS thematic layers such as road, compartment, net stocked area, age class and silvicultural treatment boundaries. However, in reality, there are a number of issues regarding the use of thematic layers as a means of stratifying a plantation. For example:

- Subtle variation in structure can still occur within a resource unit, even within the same age and thinning status, hence they can be a coarse predictor of actual plantation variation.
- GIS thematic line-work or categorical labels can sometimes be outdated and consequently any error in these layers also affects the stratification results.
- Age class is unambiguous, but thinning treatment classes are not always consistent between sites and can mean different things in different regions and at different times. This makes it difficult to assume a standardised structure for a given silvicultural event.
- Silvicultural treatment classes are not static over time. For example, two stands may be classed as first thinning (T1), but one may be a recent event while another may have occurred several years earlier. Although both are classed as T1, the older thinning will have fewer canopy gaps and hence a different structure.
- Other factors that affect stand structure are often unaccounted for in GIS thematic layers (e.g. localised storm damage, losses due to drought, pests and diseases, genetic variation in seed sources etc), and this weakens the assumption of stand uniformity.

1.1 Stratification with airborne lidar

In contrast, airborne lidar can provide rapid wall-to-wall continuous variables that are more sensitive to subtle changes in plantation structural variation. Various studies have investigated lidar stratification options. Antonarakis *et al.* (2008) developed a landcover classification approach using lidar metrics including elevation and intensity values. Hawbaker *et al.* (2009) found lidar stratification superior to conventional systematic sampling in mixed forest in Wisconsin, USA. While Maltamo *et al.* (2011) found lidar stratification provided the most accurate results for estimating stand volume and stocking in mixed conifer forests in Norway.

Some studies have utilised an object-based image analysis (OBIA) approach to delineate stands with similar characteristics. Pascual *et al.* (2008) used a two stage approach for classification in *Pinus sylestris* stands in Spain. They initially grouped similar stands using Definiens eCognition® (www.ecognition.com) with a lidar CHM, and then used a k-means algorithm to classify them into structure types. Stone *et al.* (2011) successfully applied eCognition rule-sets based on a lidar CHM to classify landcover and thinning status in pine plantations in Australia. Other studies have used Size-Constrained Region Merging (SCRM) software (Castilla *et al.* 2008; Chen and Hay 2011) to achieve similar results. Although object-based approaches can be effective in segmenting stand polygons, these units can still contain considerable internal variation which has implications for field sampling. Furthermore, Castilla *et al.* (2008) note that OBIA methods require the setting of parameters which are non-intuitive to the user.

Lidar technology offers considerable flexibility when quantifying forest structure. For example, various studies have investigated a wide range of lidar structure attributes including height percentiles (Andersen *et al.* 2005; Hilker *et al.* 2010; Maltamo *et al.* 2011), canopy cover

percentiles (Jochem *et al.* 2011; Vastaranta *et al.* 2011), Weibull distribution (Gobakken & Naesset 2005), height skewness (Antonarakis *et al.* 2008; Latifi *et al.* 2010) and canopy volume as three-dimensional grid cells or voxels (Lefsky *et al.* 1999; Coops *et al.* 2007). However, many lidar metrics can be very complex and computationally problematic when processing large areas of plantation. Moreover, they can require specialist skills and software. For this study, the focus was on utilising lidar metrics that can be readily calculated using commercial software available to most Australian plantation managers. ArcGIS, for example, is used extensively and most foresters are familiar with its operation. In addition, airborne lidar datasets are predominantly supplied as CHMs in raster format which are suitable for analysis in ArcGIS. The lidar metrics presented later in this paper are currently being incorporated into a new ArcGIS lidar toolbox under development at the Forest Science Centre (Department of Primary Industries NSW) specifically designed for plantation managers.

1.3 Objectives

The objectives of this study were to:

1. quantify structural variation in a traditional thematic-based stratification approach used by FNSW, and
2. demonstrate that basic lidar metrics can predict both thematic attributes and survey stand attributes.

An evaluation of results will help to postulate a strategy for developing a lidar stratification approach in softwood plantations.

2. Study area

The study area was situated within Green Hills State Forest (SF), located near the town of Batlow in the Southern Tablelands of NSW, Australia (see Stone *et al.* 2011 for details). The 5,000 ha *Pinus Radiata* plantation has undulating hilly topography and a mean elevation of 750m. The site contains a full representation of age classes, silvicultural treatments and terrain steepness categories.

3. Methodology

3.1 Lidar acquisition and processing

Airborne lidar data was acquired in July 2008 using a Lite Mapper LMS-Q560 ALS system (Riegl, Austria) mounted in a fixed-wing aircraft. The mission specifications included a mean footprint diameter of 60 cm, maximum scan angle of 15° (off vertical), and a mean point density of 2 pulses/m² (based on the non-overlap portion of the swath). Lidar data was provided in LAS format and processed to generate a Digital Terrain Model (DTM) at both 1.0m and 0.5m pixel resolution using a standard linear triangulation surface modelling technique in Environment for Visualizing Images (ENVI) software (Research Systems Incorporated, USA). In addition, a Vegetation Elevation Model (VEM) was generated from all laser points and the DTM was then subtracted from the VEM to derive CHMs at 1.0 m and 0.5 m pixel resolution.

3.2 Sampling design and field surveys

To cover the structural variation within the plantation the site was initially stratified by three thematic factors (i.e. age class, thinning treatment and ground slope categories) into 16 strata classes (Stone *et al.* 2011). Each stratum was randomly allocated four circular plots (except for one) bringing the total to 63 field plots. With known differences in stocking between strata

classes, variable sized plots (with radii ranging from 7 m – 20 m) were utilised to achieve at least 15 trees per plot.

Each individual tree and plot centre was accurately surveyed using a laser theodolite (Leica T1100 total station) and a Differential Global Positioning System (DGPS). Trees were allocated a unique ID number and then measured for stem diameter and tree height (measured twice using a Vertex hypsometer). A localised equation based on stem diameter and height values was then applied to derive gross volume estimates. Individual tree data was then aggregated to generate plot level estimates (stand attributes – SA) including:

- mean height (MH),
- stocking (STK),
- basal area (BA) and
- total standing volume (TSV).

In addition to the 63 research plots, 100 conventional forest inventory plots were measured independently within the same study area at around the same time. Twenty five plots were established in four strata: an unthinned (UT) 1998 age class (AC), a UT 1983 AC, a second thinning (T2) 1983 AC and a T2 1979 AC. Plots were located in the field using a Scout Pak GPS (Juniper Systems) with post-processing differential correction. Plots were measured in accordance with standard FNSW inventory procedures and plot level estimates were derived for stocking, basal area, mean tree height and volume.

3.3 Extraction of lidar metrics

As a surrogate for structural variation, several focal-based stand level metrics were extracted from the 1m CHM using a 0.1 ha circular search windows with a series of Interactive Data Language (IDL) scripts. Within the focal window the following lidar metrics were calculated:

- Mean height (MH): the mean height of all pixels in the focal search window,
- Mean above mean (MAM): the mean height of pixels above MH,
- Mean Dominant Height (MDH): the mean height of pixels found using a 5x5 pixel local maxima search,
- Predicted Stocking (PSTK): The count of pixels found using a 3x3 pixel local maxima search, expressed as stems per hectare,
- Canopy cover percentage (CC): the percentage of pixels above 3m in height,
- Occupied volume (OV): the sum of all pixel heights (i.e. the space occupied from the upper canopy to the ground), and
- Height variance (Var): a measure of how the pixel height values are distributed around the mean height.

The focal statistics above are all derived from a 1m CHM which means they are less sensitive to subtle variations in the original sampling density (e.g. 1 to 5 pulses per m²), or scan angle variations, compared to point statistics. All calculated raster surfaces were generated at 1m pixel resolution. A pixel resolution of 0.5m was also considered but later rejected as it was visibly affected by lidar scan overlap issues. In addition, the processing time for a 1m CHM is considerable shorter. Although processing scripts used in this study were written in IDL they are currently being scripted for ArcGIS through a series of multistage processing steps.

To extract the lidar metrics the survey centre points were initially buffered up to a radius of 35m. Because the field survey plots had variable radii a separate raster indicating the distance of each plot pixel from the centre point was also generated so statistics could be calculated for any nominated plot radius. Plot pixels were converted to a Region of Interest (ROI) and

subsequently used to export the original CHM pixel values and each lidar metric.

3.4 Statistical analysis

The extracted pixel data was imported into the open-source R-statistical package v.2.11.1 (R-Development Core Team 2010) for statistical analysis. Data analysis was conducted in three stages. Firstly, the structural variation within traditional thematic-based strata was quantified to set a performance benchmark. Secondly, the capacity to predict thematic layers with lidar metrics was explored. Lastly, the analysis focused on predicting survey stand attributes with lidar metrics.

3.4.1 Analysis of structural variation within thematic strata

It is not possible to field survey every combination of structural parameters within a pine plantation and consequently strategic inventories typically divide the estate into a smaller number of broad strata classes. For this analysis thematic strata were defined as six different combinations of age and thinning class (table 1).

Table 1: Sample thematic stratification

Stratum	Treatment	Age class	Code
1	Unthinned - UT	Small 10-20	SUT
2	“	Medium 20-30	MUT
3	1st thinning – T1	Small 10-20	ST1
4	“	Medium 20-30	MT1
5	2nd thinning – T2	Medium 20-30	MT2
6	“	Tall > 30	TT2

Mean and confidence interval for the stand attribute mean values were calculated for each strata class. Box-and-whisker plots, created by John W. Tukey (Cleveland, 1993), were used to visually compare the location and spread values of the stand attributes.

3.4.2 Predicting thematic strata with lidar metrics

The conventional stratification approach in FNSW is based on age class and thinning treatment. Because age class is a continuous variable and thinning treatment is a categorical variable, the investigation into the capacity of lidar to predict these attributes was undertaken in two ways.

Age class was modelled using Ordinary Least Square (OLS) regression with the lidar metrics as explanatory variables. Some of these metrics were highly correlated and to avoid multi-collinearity, variables that had a Spearman's correlation coefficient more than 0.7 were not included together in the model (Chatterjee & Hadi, 2006). The final predictor variables were selected manually by comparing alternative models constructed from the initial set of variables. The comparisons were based on the significance of the regression coefficients and adjusted R-square values, which takes into account the number of variables and the model residuals. Model assumptions were checked by looking at the residual plots (Draper & Smith, 1998).

Thinning treatment was predicted using classification and regression tree (CART) analysis with lidar metrics used as explanatory variables for the decision tree algorithm (Breiman et al., 1984). In comparison to most other multivariate statistical techniques (Biermann et al., 2011) CART handles both categorical and metric data without data transformation. The explanatory variable showing the highest statistical association to the response variable is chosen for the split. Initial CART analysis resulted in a very large tree, which was pruned back to an optimal size by

minimising the cross validation error. The Kappa statistic (Cohen 1968) was used as an evaluation of model reliability. The analysis was done using rpart (Therneau et al., 2010).

3.4.2 Predicting stand attributes with lidar metrics

Predictive models for stand attributes (MH, STK, BA and TSV) were developed using the lidar metrics (MH, MAM, MDH, PSTK, CC, OV and Var). Multi-collinearity was avoided by using only one variable from a group of highly correlated variables. Log transformation was used to stabilise the variance in the BA and TSV. Model selection was based on adjusted R-square criterion with residuals diagnostics done to check for the model assumptions. The estimation accuracies of the stand attributes were compared using the root mean square error (RMSE). RMSE and bias were calculated in relative terms (RMSE% and bias%), i.e. the RMSE and bias values for the stand attribute y divided by their observed mean values.

4. Results and discussion

4.1 Structural variation within thematic strata

When using a thematic stratification approach, an assumption is made that the structural variation within a stratum is much less than the variation across all strata. However, when the four stand attributes of interest were plotted as box plots (see Figure 1) results showed that there are common overlaps of the SA values between strata. In fact, the variability within stratum groups is almost as much as the variability between strata groups. For example, mean height (MH) values for medium size trees in the stratum 2 (MUT) share an overlapping range in values with five other strata. Other variables show similar variability.

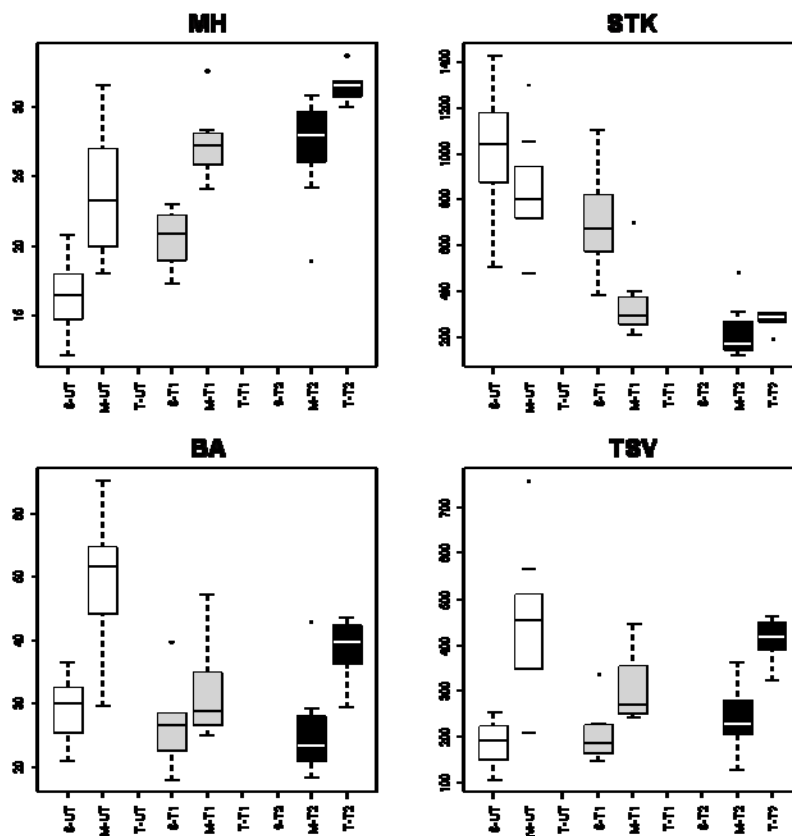


Figure 1: Box plot showing SA variation and overlap for each stratum. The box represents the first and the third quartiles while the whiskers show minimum and maximum values.

Mean stand attribute values, along with their lower and upper confidence limits, are presented in Table 2. Results again indicate that there is considerable variation within stratum and the values in the different classes are overlapping.

Table 2: Mean value of stand attributes by strata class. Upper and lower confidence limits are in parenthesis.

Strat.	MH	STK	BA	TSV
SUT	16.2 (14.7,17.8)	1043 (880,1206)	29.1 (25.9,32.3)	185 (156,214)
MUT	23.6 (21.4, 25.8)	884 (705,942)	49.8 (44.6,54.9)	442 (367,517)
ST1	20.5 (26.7,32)	701 (520,882)	26.7 (21.3,32)	204 (154,255)
MT1	27.3 (25.2,29.5)	342 (212,472)	31.6 (25.4,37.8)	305 (240,370)
MT2	27.3 (25.2,29.4)	212 (145,279)	25.2 (21,29.4)	238 (200,276)
TT2	31.4 (30.5, 32.4)	274 (241,308)	38.8 (34.9,42.8)	413 (374,451)

Results reveal that there is sizeable variation within strata classes and they are not as internally uniform as we would wish. In this example, thematic stratification can be considered a coarse predictor of variation in the key stand attributes of interest (MH, STK, BA and TSV) and consequently there is an increased chance that any stratified plot sampling will not cover the full range in structural variation within the plantation.

However, as a general rule, FNSW usually selects only one age class to sample at any given operational inventory event, although the age class may sometimes be divided into one or two thinning treatment substrata. But even in monoculture single aged stands with one thinning treatment the variation in stand attributes can still be highly variable. To demonstrate this, data from four operational inventories within the study area were analysed (see Figure 2).

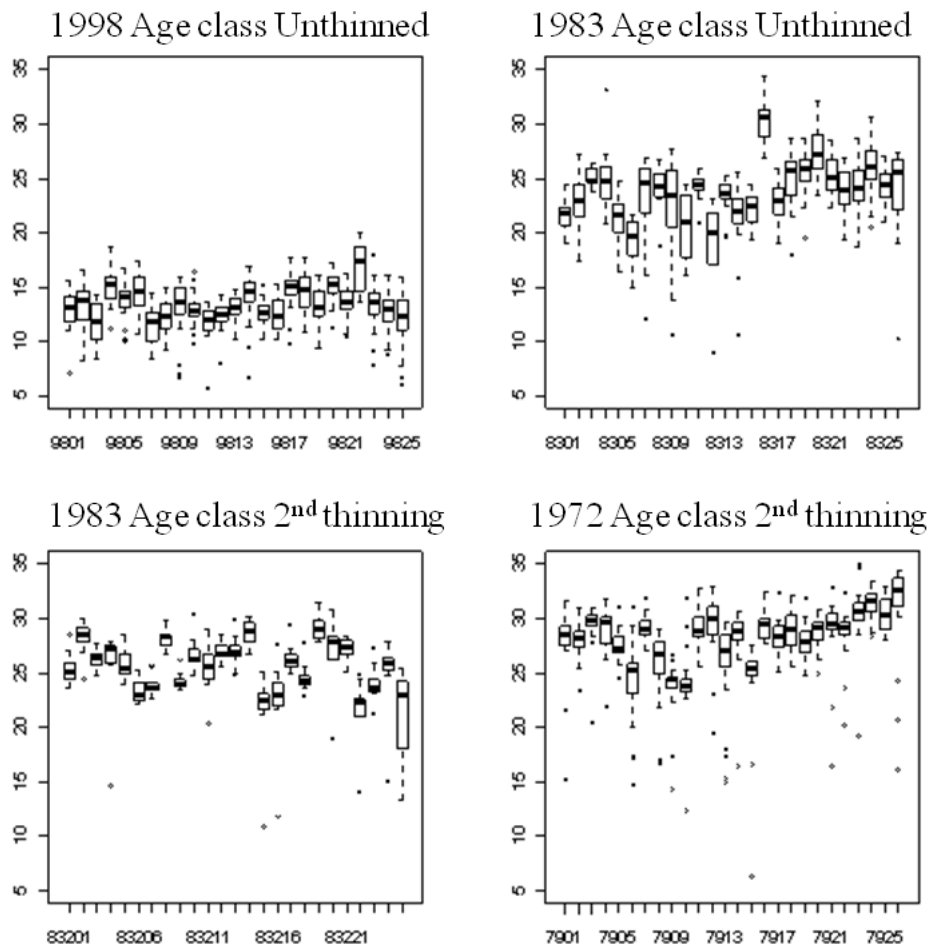


Figure 2 is the box plot of tree heights. Separate plot is done for each age/thinning treatment combination. The plot shows that the plots within an age/thinning treatment class are highly variable.

4.2 Predicting thematic strata with lidar metrics

Plantation managers have traditionally used thematic stratification (related to their management units) for routine wood resource inventory and are most comfortable with this concept. To instil more confidence in an alternative lidar stratification approach, there is merit in demonstrating a link between conventional thematic strata and lidar metrics.

4.2.1 Predicting age class

Normally it would not be necessary to predict the age class of a stand as spatial records should already contain this data. However, in cases where spatial datasets may be out-of-date or non-existent this could potentially be a useful attribute to predict. Furthermore, if lidar can predict stand age it would strengthen the case for the robustness of lidar-derived stratification. Age class was modelled using OLS regression with the lidar independent variables (MH, MAM, MDH, PSTK, CC, OV and Var). Results indicated that two variables (MAM and CC) were significant for modelling age class. Figure 3 shows a graph of the multiple regression

relationship. The fitted model had an R^2 of 0.91 and an RMSE value of 1.9 years.

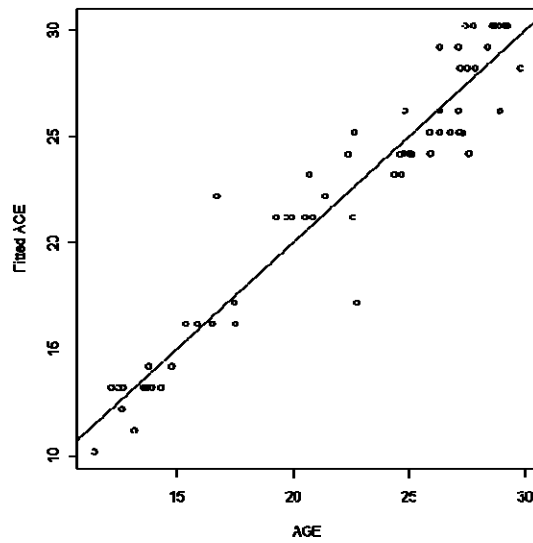


Figure 3: Bivariate plot of the observed age versus the fitted values of the age using the multiple regression model with MAM, CC and IND (a dummy variable to fit segmented regression to data with MAM values of 18 or less and another to MAM values over 18) .

4.2.2 Predicting thinning treatment

As a categorical variable, thinning treatment was predicted using CART analysis with the lidar metrics MH, MAM, MDH, PSTK, CC, OV and Var as input variables. The tree was fully grown and then pruned based on the cost complexity value of 0.02. After pruning, the remaining lidar metrics were CC and MAM. Figure 4 presents the final pruned decision tree.

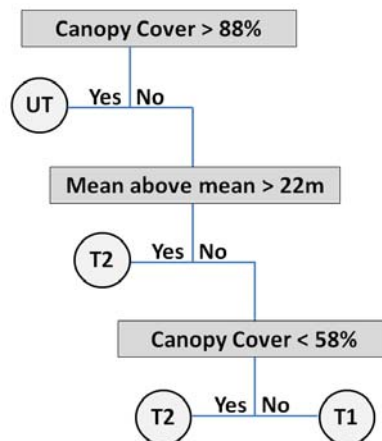


Figure 4: Decision tree classification model for predicting thinning treatment (UT-unthinned, T1-1st thinning and T2-2nd thinning) based on lidar metrics (CC and MAM).

The decision tree had a classification accuracy of 89% (i.e. only 7 plots were misclassified), and a kappa coefficient 0.86. CC proved to be a robust metric for detecting unthinned stands, while T1 and T2 stands were separated using a combination of CC and MAM. The CART analysis results confirm that basic lidar metrics derived from a CHM can be used to predict thinning status, which is an important variable in many field sampling designs.

4.3 Predicting stand attributes with lidar metrics

OLS regression models using lidar metrics, fitted to MH, STK, MH and TSV are presented in Table 3.

Table 3: Results of regression analysis.

Depend. Variable	Model	Adj R ²	RMSE (%)	Bias (%)
MH	$3.715 + (0.826 \times \text{MAM}) + (0.047 \times \text{Var})$	0.95	1.298 (4.3)	0 (0)
STK	$-485.6 + (1.967 \times \text{PSTK})$	0.82	154 (25.6)	0 (0)
Log(BA)	$2.64 - (0.0026 \times \text{Var}) + (0.00027 \times \text{OV})$	0.67	6.6 (19.1)	0.58 (1.7)
Log(TSV)	$4.47 + (0.000469 \times \text{OV}) + (0.00067 \times \text{PSTK}) - (0.012 \times \text{CC})$	0.8	58.8 (19.4)	4.2 (1.4)

Results confirm that the prime stand attributes of interest (i.e. MH, STK, BA and TSV) can be predicted with simple lidar metrics extracted only from a 1m resolution CHM.

5. Conclusion

This study has shown that conventional GIS-based thematic stratification based on age class and thinning status can result in strata with a high degree of internal structural variation and considerable overlap with other strata classes. This affirms that even-aged softwood plantations can still exhibit a high degree of spatial variation within a single age/thinning class. As an alternative, lidar offers a high resolution, rapid and objective tool for stratification that is independent of the inherent issues with conventional thematic layers.

The basic lidar metrics tested in this study (i.e. mean height, mean above mean height, mean dominant height, predicted stocking, canopy cover percentage, occupied volume and height variance) were shown to be useful in predicting stand attributes such as age class, thinning treatment, stocking, mean height and total stand volume, which are all important variables in most plantation inventory programs. These relatively simple lidar metrics can be extracted from a CHM (which is a standard lidar product) using commercial software already available to most plantation managers (e.g ArcGIS) and are also suitable for existing inventory software systems. Moreover, such an approach would be independent of existing GIS thematic layers and therefore have wider application where up-to-date spatial records are not available. Such basic metrics could be used to develop new sampling approaches to improve ground survey efficiency. Future research from the Forest Science Centre will explore a range of probability (design) and model based approaches (McRoberts 2010) to select the most efficient sampling design to support ground-based inventories

Acknowledgements

This study was funded by Forest & Wood Products Australia, Forests NSW and the Department of Primary Industries NSW. The authors wish to thank the following people who provided their assistance: Duncan Watt, Bob Cooper and Rachel Bidder (FNSW); Cath Carney, Peter Worsley, Matt Nagel and Huiquan Bi (DPI NSW).

References:

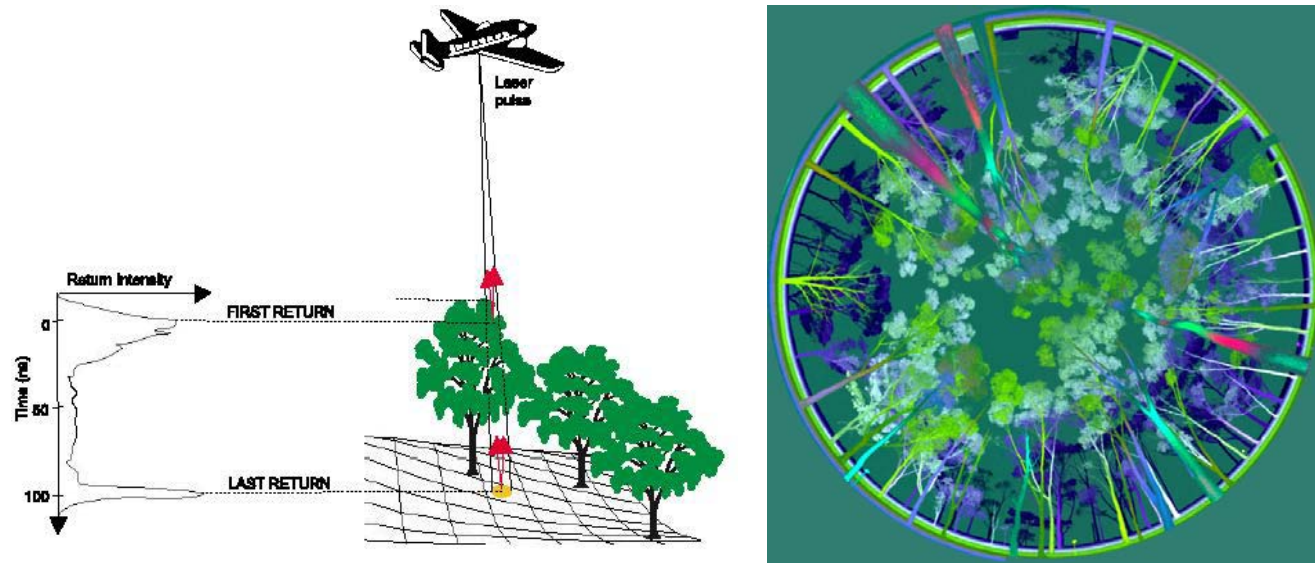
- Andersen, H., McGaughey, R., and Reutebuch, S. (2005) Estimating forest canopy fuel parameters using LIDAR data. *Remote Sensing of Environment* 94 (2005) 441–449.
- Antonarakis, A., Richards, K. and Brasington, J. (2008) Object-based land cover classification using airborne lidar. *Remote Sensing of Environment*, 112: 2988-2998.
- Austin PC (2008) R and S-PLUS produced different classification trees for predicting patient mortality. *J Clin Epidemiol* 61:1222–1226
- Biermann P, M. Lewis, B. Ostendorf and J. Tanner (2011), A review of methods for analysing spatial and temporal patterns in coastal water quality, *Ecol. Indicators* 11, pp. 103–114.
- Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984) *Classification and regression trees*, Wadsworth.
- Castilla, G., Hay, G. And Ruiz-Gallardo, J. (2008) Size-constrained region merging (SCRM). An Automated delineation tool for assisted photointerpretation. *Photogrammetric Engineering and Remote Sensing*, Vol. 74, N0.4, pp 409-419
- Chatterjee S. and Hadi, A.S. (2006), *Regression Analysis by Example: Wiley Series in Probability and Statistics*. John Wiley & Sons Inc. Publication,
- Chen, G. and Hay, G. (2011) An airborne lidar sampling strategy to model forest canopy height from Quickbird imagery and GEOBIA. *Remote Sensing of Environment* 115: 1532-1542.
- Cleveland, W 1993, *Visualizing Data*, Hobart Press, Summit, New Jersey.
- Cohen J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 70 (213-20).
- Coops NC, Hilker T, Wulder MA, St-Onge B, Newnham G, Siggins A, and Trofymow JA (2007) Estimating canopy structure of douglas fir forest stands from discrete-return LiDAR. *Trees Struct. Funct.* Vol. 21:295–310.
- Cuesta,J., Chazette, P., Allouis, T., Flamant, P., Durrieu, S., Sanak, J., Genau, P., Guyon, D., Loustau, D., and Flamant, C. (2010) Observing the Forest Canopy with a New Ultra-Violet Compact Airborne Lidar. *Sensors* 2010, 10, 7386-7403.
- Draper N.R., H. Smith (1998) *Applied regression analysis, Volume 1*, John Wiley & Sons Inc. Publication.
- Gobakken, T. & Naesset, E. (2005). Weibull and percentile models for lidar-based estimation of basal area distribution. *Scandinavian Journal of Forest Research*, Vol 20: 490-502.
- Hawbaker, T., Keuler, N., Lesak, A., Gobakken, T., Contrucci, K. And Radeloff, V. (2009) Improved estimates of forest vegetation structure and biomass with a lidar-optimised sampling design. *Journal of Geophysical Research*, Vol.114, G00E04, doi:10.1029/2008JG000870.

- Hilker, T., van Leeuwen, M., Coops, N., Wulder, M., Newnham, G, Jupp, D. and Culvenor, D.(2010) Comparing canopy metrics derived from terrestrial and airborne laser scanning in a Douglas-fir dominated forest stand. *Trees* Vol. 24:819–832.
- Hilker, T., Wulder, M. and Coops, N. (2008) Update of forest inventory data with lidar and high spatial resolution satellite imagery. *Can. J. Remote Sensing*, Vol.34, No.1, pp 5-12.
- Jaskierniak, D., Lane, P., Robinson, A. and Lucieer, A. (2011). Extracting LiDAR indices to characterise multilayered forest structure using mixture distribution functions. *Remote Sensing of Environment*, 115 (2011) 573–585.
- Jensen, J., Humes, K., Vierling, L. And Hudak, A. (2008). Discrete return lidar based prediction of leaf area index in two conifer forests. *Remote Sensing of Environment*, Vol. 112, pp 1016-1031.
- Jochem, A., Hollaus, M., Rutzinger, M. and Hofle, B. (2011). Estimation of Aboveground Biomass in Alpine Forests: A Semi-Empirical Approach Considering Canopy Transparency Derived from Airborne LiDAR Data. *Sensors* Vol.11, pp 278-295; doi:10.3390/s110100278
- Hopkinson, C. And Chasmer, L. (2009). Testing lidar models of fractional cover across multiple forest ecozones. *Remote Sensing of Environment*, Vol. 133, pp 275-288.
- Latifi, H., Nothdurft, A and Koch, B (2010). Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. *Forestry*, Vol. 83, No. 4, 2010. doi:10.1093/forestry/cpq022.
- Lefsky, M., Harding, D., Cohen, W. and Parker, G. (1999). Surface Lidar Remote Sensing of Basal Area and Biomass in Deciduous Forests of Eastern Maryland, USA. *Remote Sens. Environ.* Vol. 67, 83-98.
- Maltamo, M., Bollandsås, O., Næsset, E., Gobakken, T and Packalén, P. (2011) Different plot selection strategies for field training data in ALS-assisted forest inventory. *Forestry*, Vol. 84, No. 1, 2011. doi:10.1093/forestry/cpq039.
- McRoberts, R.E. (2010) Probability and model based approaches to inference for proportion forest using satellite imagery as auxillary data. *Remote Sensing of Environment*. 114:1017-1025.
- Pascual, C., Garcia-Abril, A., Garcia-Montero, L., Martin-Fernandez, S., and Cohen, W. (2008) Object-based semi-automatic approach for forest structure characterization using lidar data in heterogeneous *Pinus sylvestris* stands. *Forest Ecology and Management*. 255(ii):3677-3685. doi: 10.1016/j.foreco.2008.02.055
- R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Stone, C., Turner, R., Kathuria, A., Carney, C., Worsley, P., Penman, T., Bi H-Q, Fox, J. & Watt, D. (2011). Adoption of new airborne technologies for improving efficiencies and accuracy of estimating standing volume and yield modelling in *Pinus radiata* plantations (PNC058-0809). Final Report for the Forest & Wood Products Australia Project

PNC0-0809. Available on the FWPA website www.fwpa.com.au

- Straub, C., Weinacker, H. and Koch, B. (2008). A fully automated procedure for delineation and classification of forest and non-forest vegetation based on full waveform laser scanner data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37 (B8), 1013–1019
- Therneau T.M, and B. Atkinson. (2010) R port by Brian Ripley. <atkinson@mayo.edu>. rpart; Recursive Partitioning. R-package version 3.1-48.
<http://CRAN.R-project.org/package=rpart>
- Turner, R., Stone, C., Kathuria, A. and Penman, T. (2011) Towards an operational lidar resource inventory process in Australian softwood plantations. *Proceedings of the 34th International Symposium for Remote Sensing of the Environment*, Sydney, Australia, 10-15 April 2011.
- Turner, R. & Stone, C. (2010) Guide to acquisition and processing of remote sensing data for softwood plantations. Document prepared for Forest & Wood Products Australia and a deliverable associated with FWPA Project PN058-0809. Available on the FWPA website www.fwpa.com.au
- Vastaranta, M., Holopainen, M., Yu, X., Hyypä, J., Hyypä, H. And Viitaka, R. (2011). Predicting stand-thinning maturity from airborne laser scanning data. *Scandinavian Journal of Forest Research*, Vol.26: pp 187-196.
- Yu, X., Hyypä, J., Kaartinen, H, and Maltamo, M. (2004). Automatic detection of harvested trees and determination of forest growth using airborne laser scanning. *Remote sensing of Environment*, Vol. 90, pp 451-462.

Ground based and airborne lidar – a natural combination



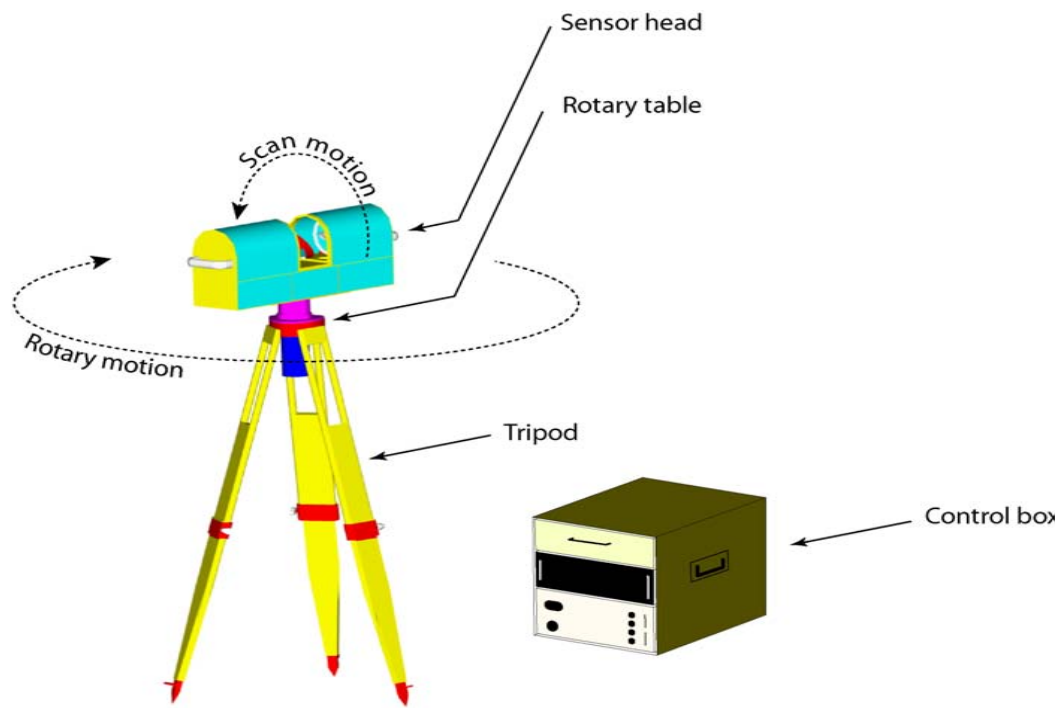
David L B Jupp
Ret. CSIRO Scientist

*SilviLaser 2001, 11th International Conference in LiDAR,
University of Tasmania, Hobart 16-20 October 2011*

Airborne and Ground Based Lidar for forest mapping

- Airborne Lidar for forest mapping (even dual wavelength) has been around for more than 30 years
- Ground based activity maybe half that but is making up ground very fast
- They use similar technology so it is attractive to try and combine them in vegetation mapping
- How well is that being done? Perhaps not as well as it could be.

(eg) A Ground-Based Lidar (Echidna[®])



- Echidna[®] is a ground-based lidar apparatus plus methods designed and patented by CSIRO specifically for forest and vegetation assessment
- The Echidna[®] and its prototype — the Echidna[®] Validation Instrument (or “EVI”) has the following characteristics:
 - Digitizes the full waveform
 - Has variable beam divergence
 - Uses no-gaps hemispherical scanning and beyond Wavelength—1064 nm
 - Linear response and calibration

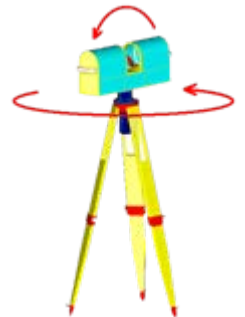
EVI (The Echidna[®] Validation Instrument)



A real Echidna—
in the forest



Scan motion



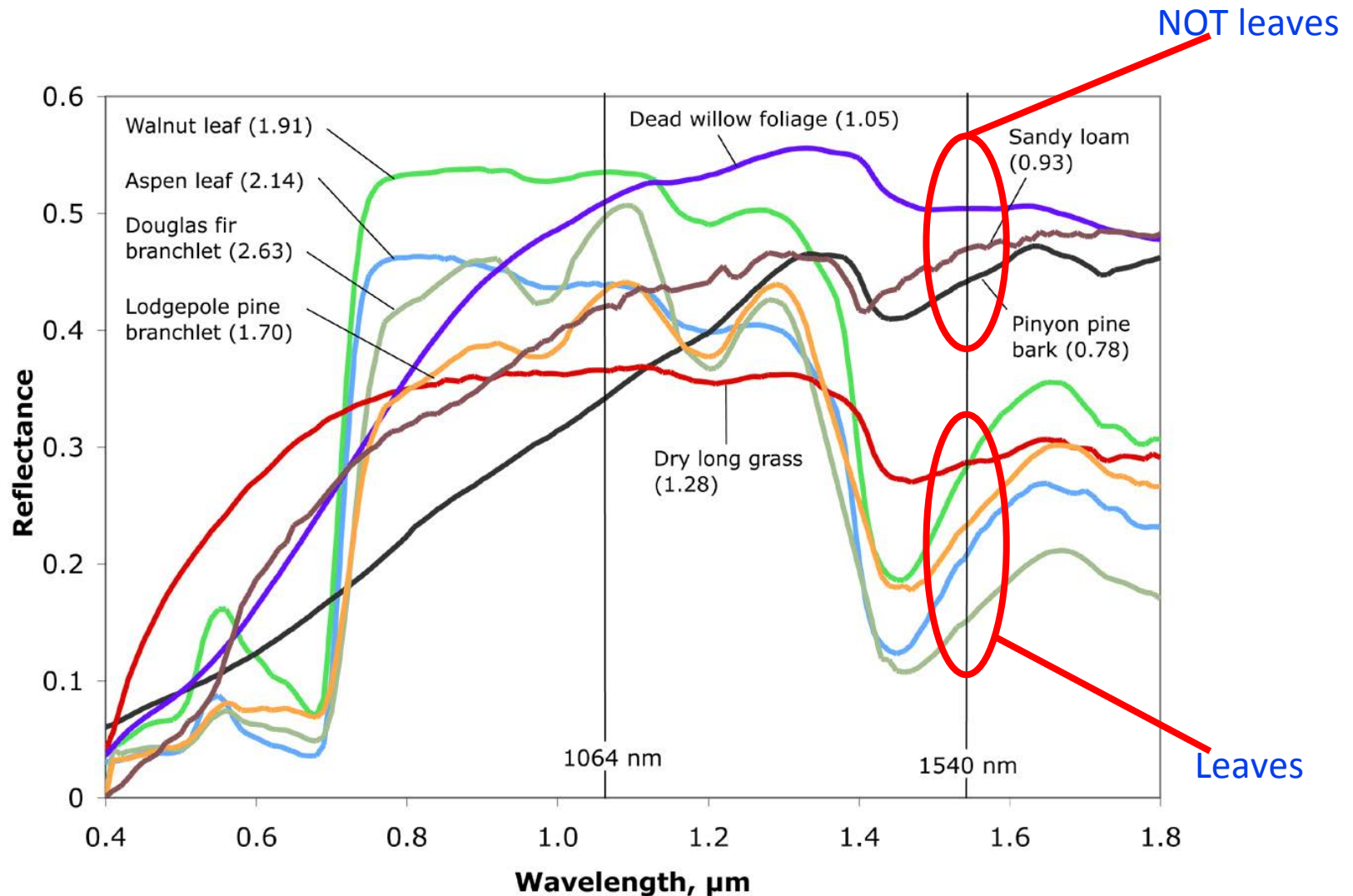
Present Status

- At this time, the EVI is not working ☹️ (But)
- Two new instruments are now being built. One is to work from USA and one from Australia
- It is the Dual-Wavelength Echidna LIDAR (aka DWEL)
- Construction is at Boston University managed by Alan Strahler and Supriya Chakrabarti
- Funding is by US NSF Grant DBI-0923389 in USA and CSIRO and TERN in Australia
- The EVI will soon be mended 😊 to obtain “hand over” calibration and even more valuable data

Dual-Wavelength Echidna LIDAR (DWEL)

- Critical Design Requirements:
 - Meeting the primary Echidna design goals
 - Two wavelengths (1064 nm and 1540 nm) + green marker, simultaneous superimposed pulses
 - Return pulse full waveform digitized as accurately as EVI with aim to be better
 - 1, 2, or 4 mrad scan resolutions, with normal operation at 2 mrad
 - Advanced calibration options
 - Battery operated and ~30 minute scan time
 - Wireless communication with computer
 - Portable (low mass and volume)

Dual Wavelength Identifies Leaves



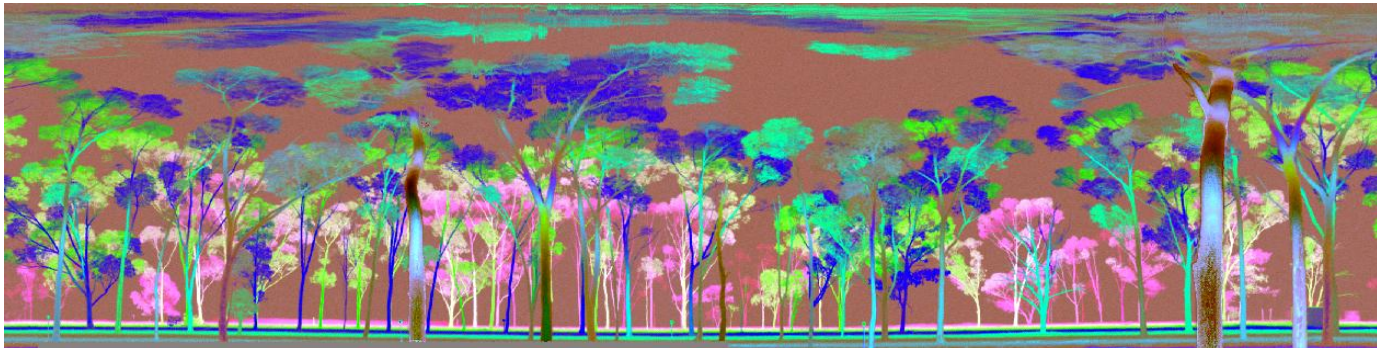
NIR and SWIR are *better* than NIR and Red – not only for eye safe but also for green versus dry and clearer decisions

What an EVI scan (or an artist) “sees” in the forest

Which artist uses Lidar?



Williams F. (1971). Springbrook, Queensland III



Echidna, E. (2005). Larundal, NSW, MNF

Basic reflectance and P_{gap} models

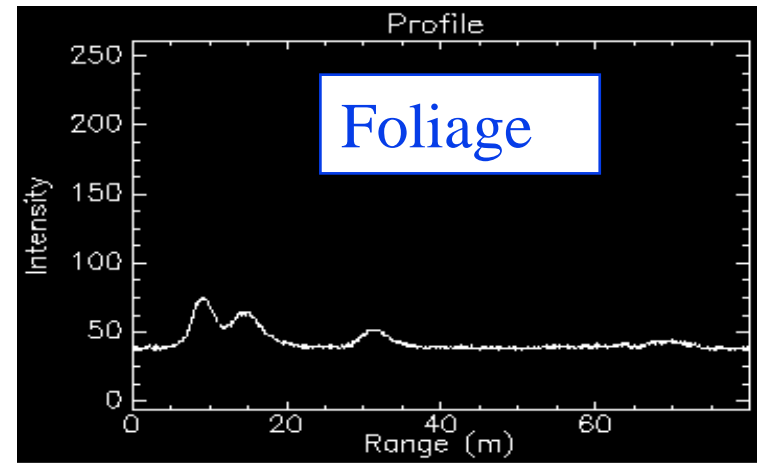
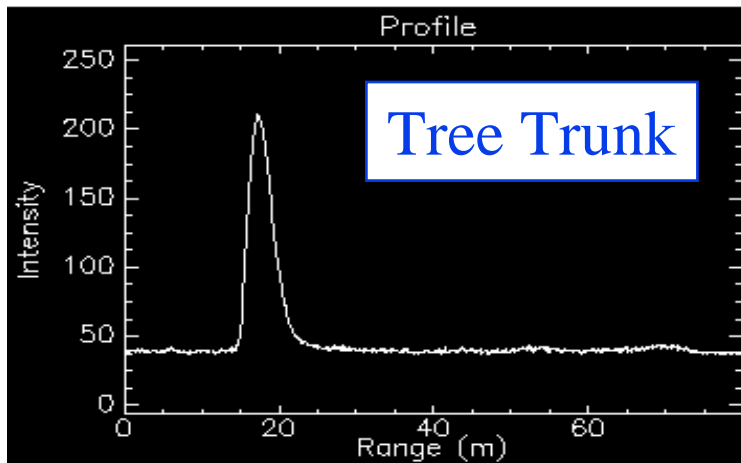
- The basic model of (calibrated) apparent reflectance
- (Jupp et al., 2001, 2005; Strahler et al., 2008) :

$$\rho_{app} = \frac{\Phi_R - \Phi_B}{\Phi_T R_A^2 K(R) \tau_a^2} R^2$$

- The basic model for P_{gap} (used for both cube and point cloud modelling)
- (Ni-Meister et al., 2001; Jupp et al., 2001, 2005) :

$$\rho_{app}(r) = -\frac{\partial P_{gap}(r)}{\partial r} \rho_a$$
$$P_{gap}(r) = 1 - \frac{1}{\rho_a} \int_0^r \rho_{app}(r') dr'$$

Hard (single) & Soft (multiple) Returns



- Pulse characteristics

- Length: 2.4 m (FWHM), strong sharp peak for identification
- Beam divergence: 5 mrad scanned on 4 mrad centres (standard operation, 2-10 mrad available)
- Digitized every 7.5 cm to 140 m range

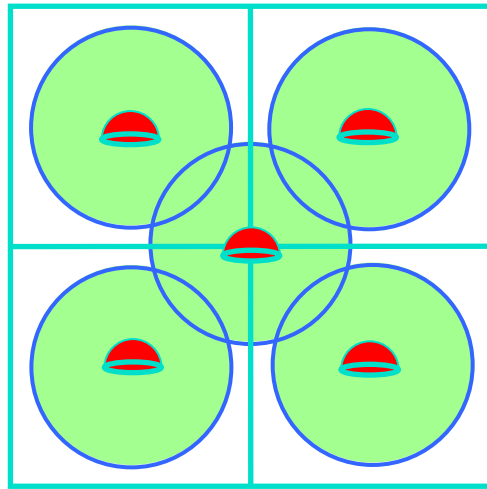
Options for ground based mapping

- Single station scan of multiple angles (objective to scan more sites and spend least time at each and accept higher variance at each site to sample spatial variation of the forest)
- Multiple scans at a site with different positions (may or may not scan same volume). Much lower variances at a site and better - if it can be afforded
- Tomographic scanning aims to sound the same volume from different directions. It produces the best site results but costs the maximum time and effort per site

Examples using BU Experiment Designs

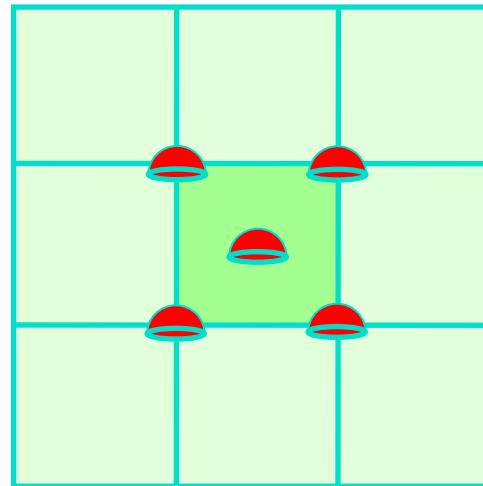
- 1 ha site (100 m x 100 m)
- 5–9 EVI scans (about 60 m radius) centered within the hectare
- Validation area: All stems mapped and DBH measured; heights and crown shapes sampled
- Partial validation area: some height and DBH measurements made
- LAI-2000 + Hemispherical photos for LAI comparisons

2007 (NE)



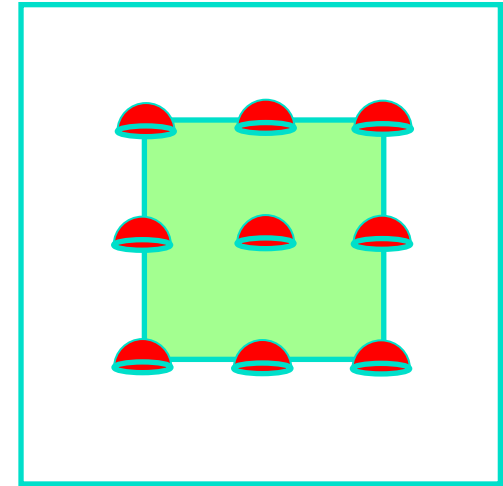
100 m

2008 (Sierras)



100 m

2009 (NE)



100 m



EVI scan point



Validation area



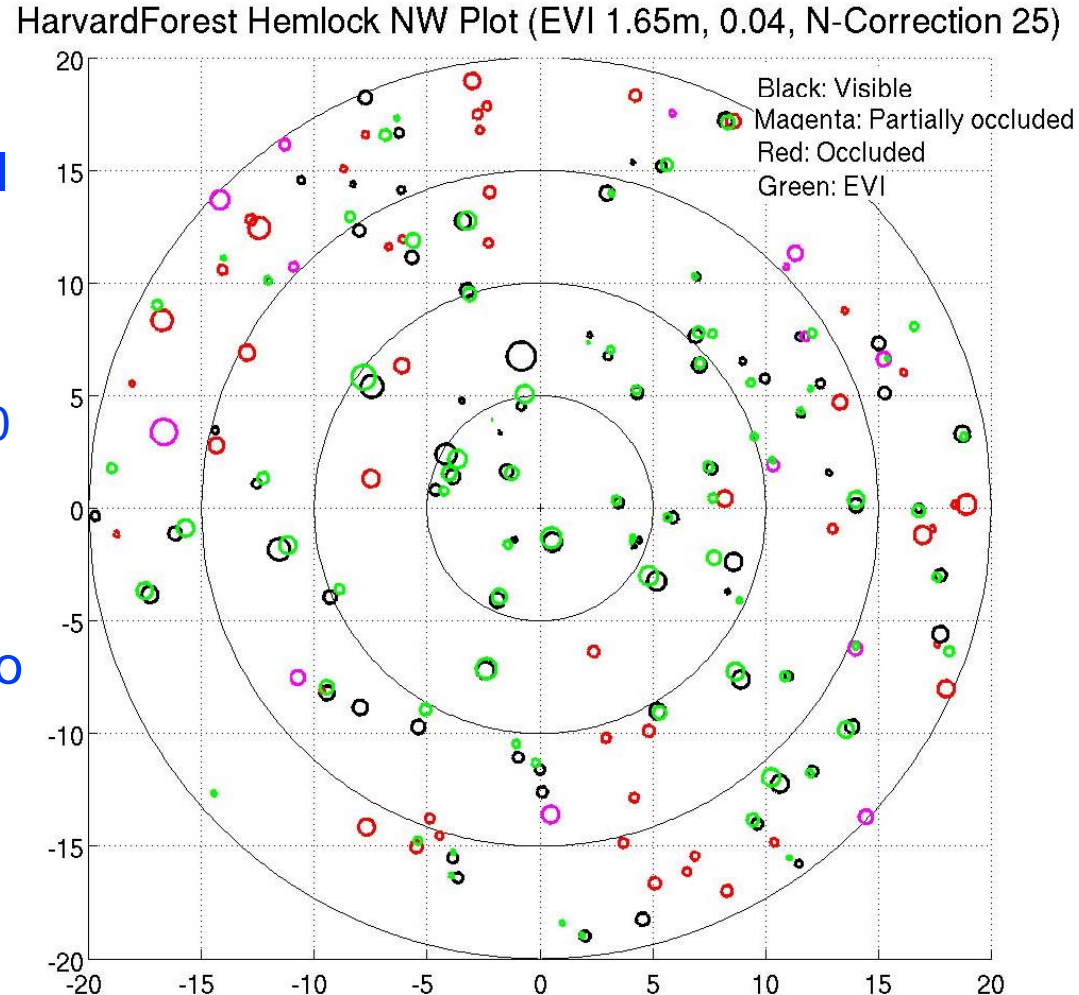
Partial validation area

EVI Single Scan Processing—“Find Trunks” (Yao Tian)

- Field measurements at each scan point
 - Distance and azimuth to all tree stems, DBH
 - Out to 10 m: All trees >3 cm DBH
 - Out to 20 m: All trees >10 cm DBH
 - Noted occlusion of each stem (not obscured, partly obscured, fully obscured) to compare with EVI scan

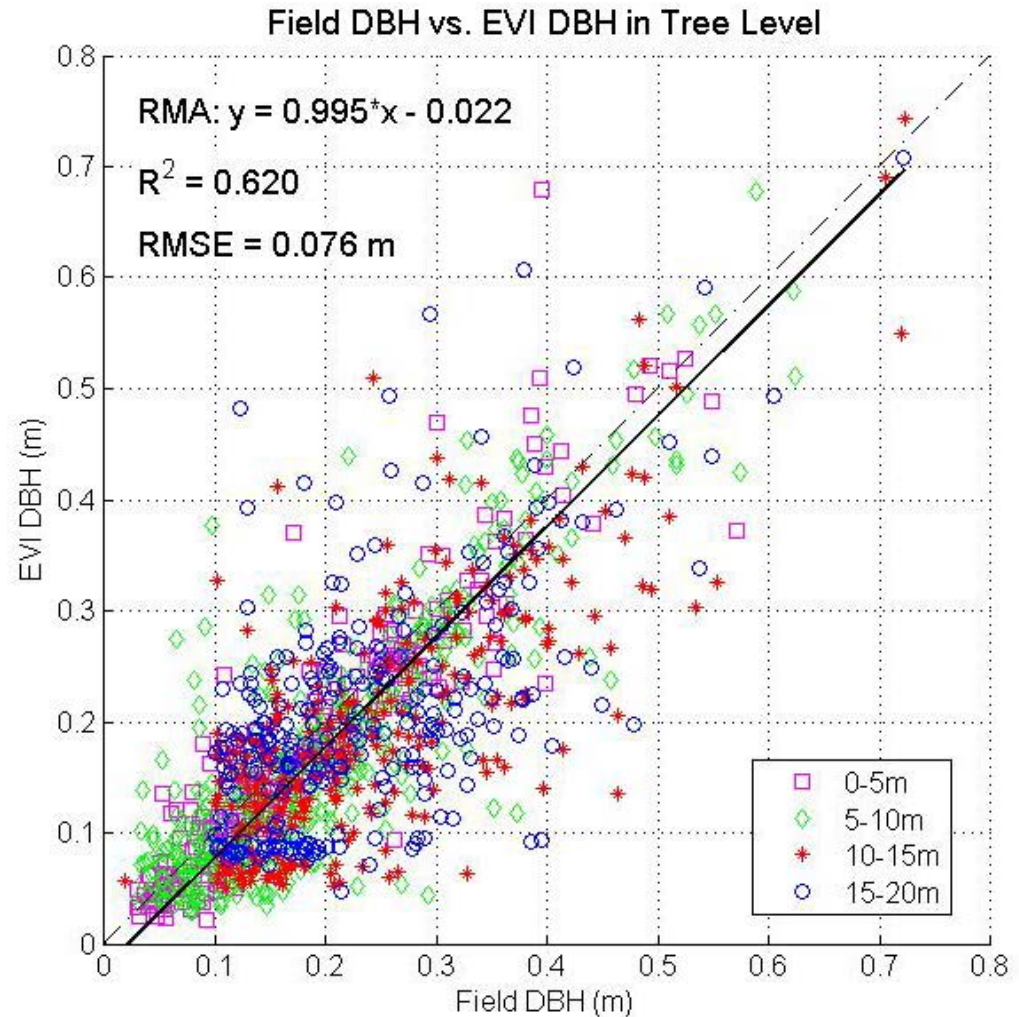


Trees found by EVI
“Visible” trees
“Partly occluded” trees
“Occluded” trees



Diameter Retrievals, New England, 2007 (Yao Tian)

- Slope is not different from 1 with power of 0.99 so overall bias is small
- $R^2=0.936$, site level (6 plots, 5 scans per plot)
- Variance is high due to both spatial variance and angle resolution at farther ranges
- Variance is single scan variance and reduced by obtaining more sites



Methods: Stem Count Density Estimation

- The complete model (Jupp D., 2008) for the number of trees “apparently” within radius R for full circle and all trees is:

$$N(\lambda, D_E; R) = \frac{2\pi\lambda R^2}{(\lambda D_E R)^2} \left[1 - e^{-\lambda D_E R} (1 + \lambda D_E R) \right]$$

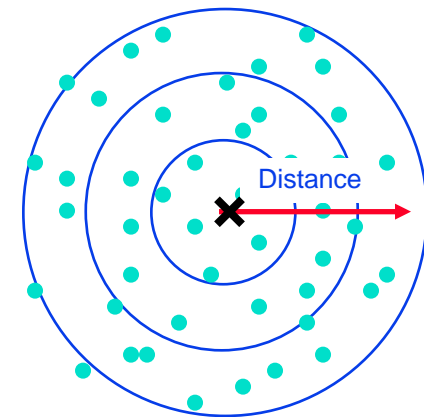
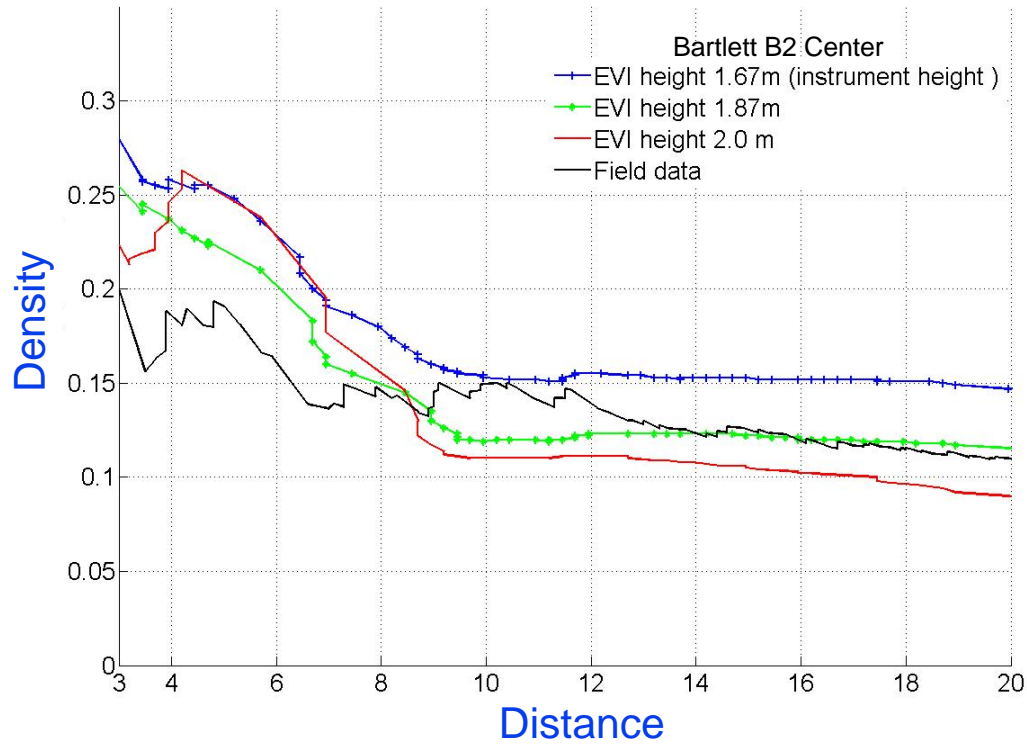
where $N(\lambda, D_E; R)$ is the number of trees “apparently” within radius R , given a true tree count density of λ and an effective tree diameter D_E .

- D_E is the effective occluding diameter, $D_E = \bar{D} + \delta D_E, \delta D_E \geq 0$, where \bar{D} is the mean of tree diameter. D_E depends on the stem density and distribution in the sample plot. As a default, we may use $D_E \approx \bar{D}(1 + C_D^2)^{1/2}$, where C_D^2 is the squared coefficient of variation for tree diameters.

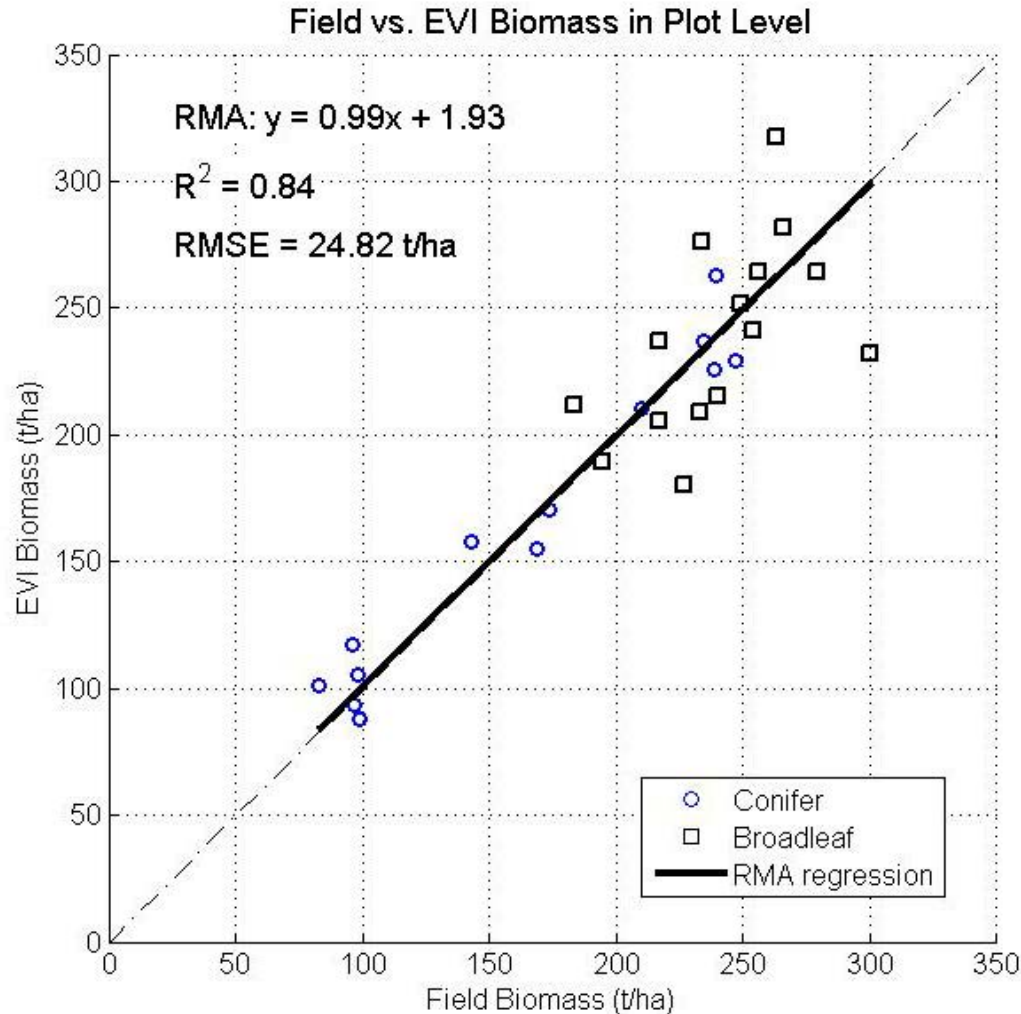
The overall attenuation (λD_E) in this model can be estimated from the horizontal attenuation in the Pgap output. **EVI “relaskop” can also be used with account for occlusion to obtain basal area.**

Stem Count Density (stems/area) (Yao Tian)

Stem count density with increasing plot radius



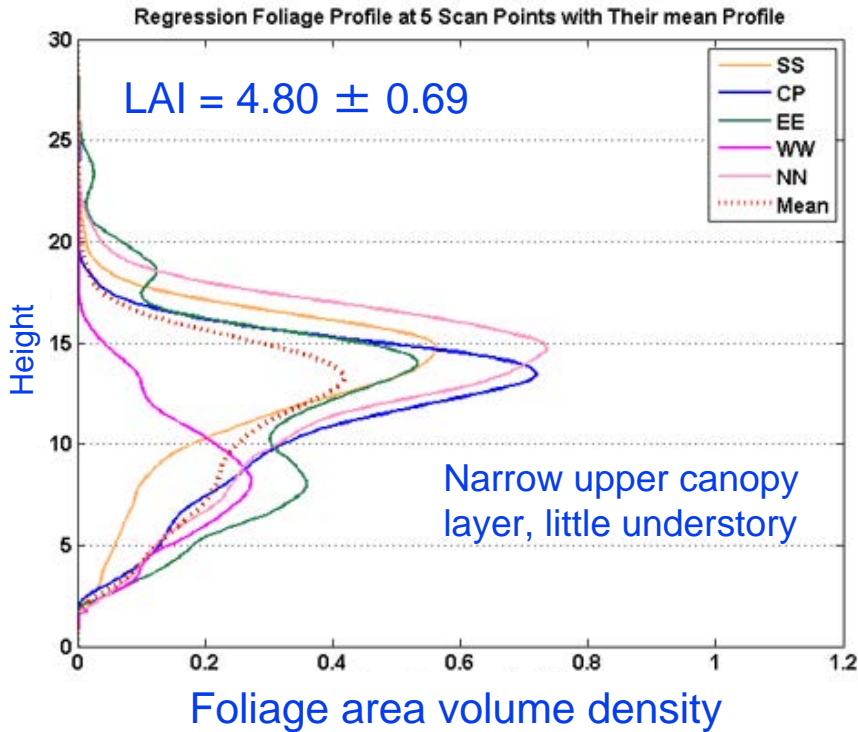
Biomass Estimation, New England, 2007 (Yao Tian)



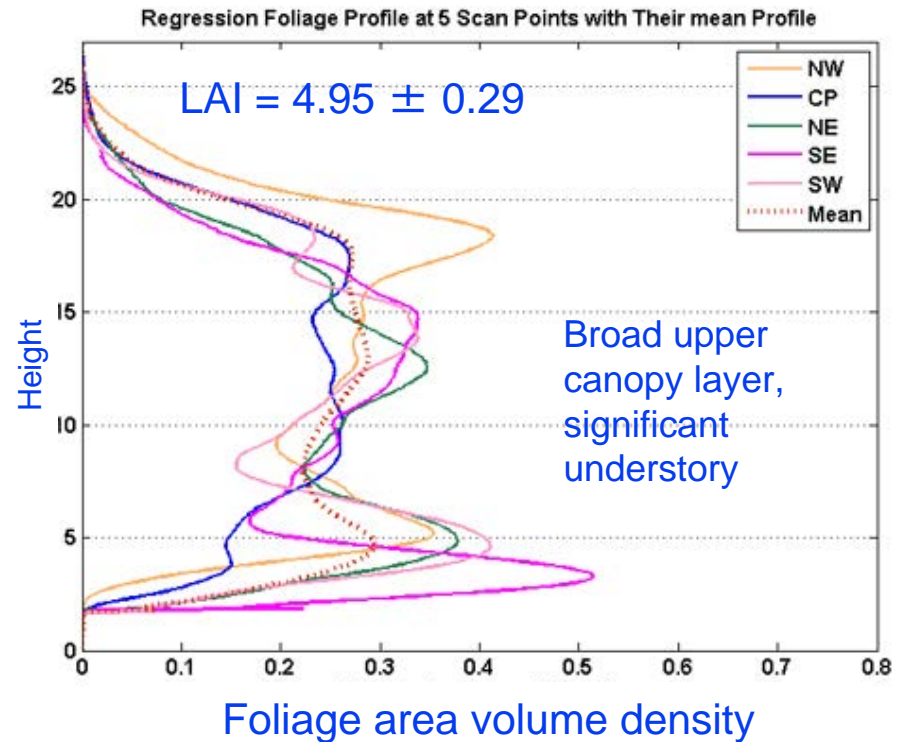
- With EVI-derived mean diameter and stem count density, we can estimate biomass using allometric equations
- Since EVI can't identify species, we used a pooled allometric equation for the leading one or two dominant tree species in the plot
- $R^2 = 0.840$ at the plot level (30 scans)
- $R^2 = 0.975$ at the site level (6 plots, 5 scans per plot)

EVI Foliage Profiles, New England, 2007 (Zhao Feng)

Red spruce stand
(Howland Tower)

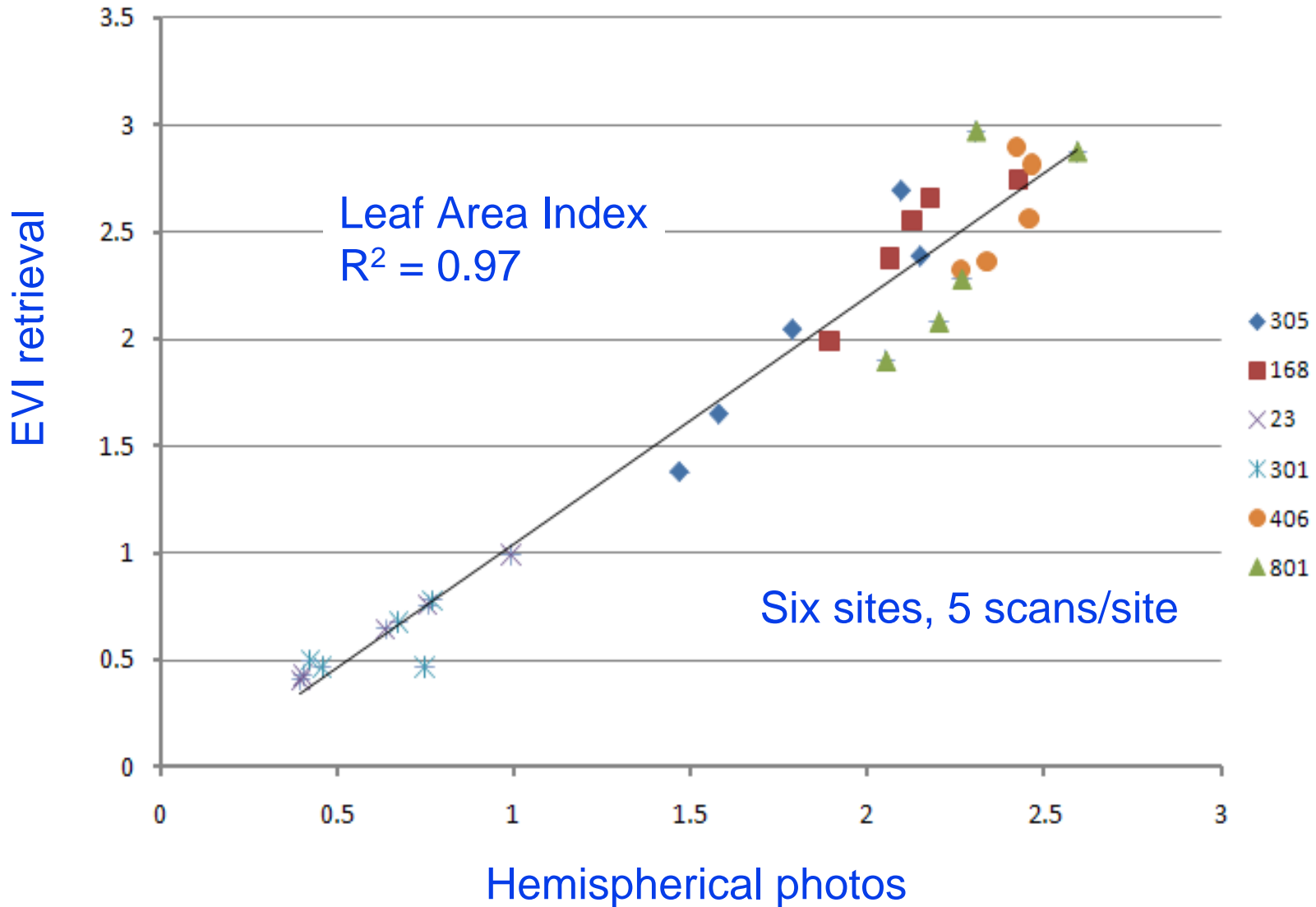


Mature hardwoods
(Bartlett B2)



Regression method retrieval (uses $5^\circ - 60^\circ$ zenith rings)

Leaf Area Index Comparisons, Sierra Nevada, 2008 (Zhao Feng)



Merged Data processing

- Tomographic sounding with merged data is better done in (x,y,z) than polar coordinates as there is a single origin and you need Euclidean transformations to co-register and analyse
- The data can be reduced to a “point cloud” by filtering and a discrete model for the data
- This is a form of data compression and most of the previous results can also be obtained by using these data rather than a polar “cube” of data – but for a single scan image processing has advantages.

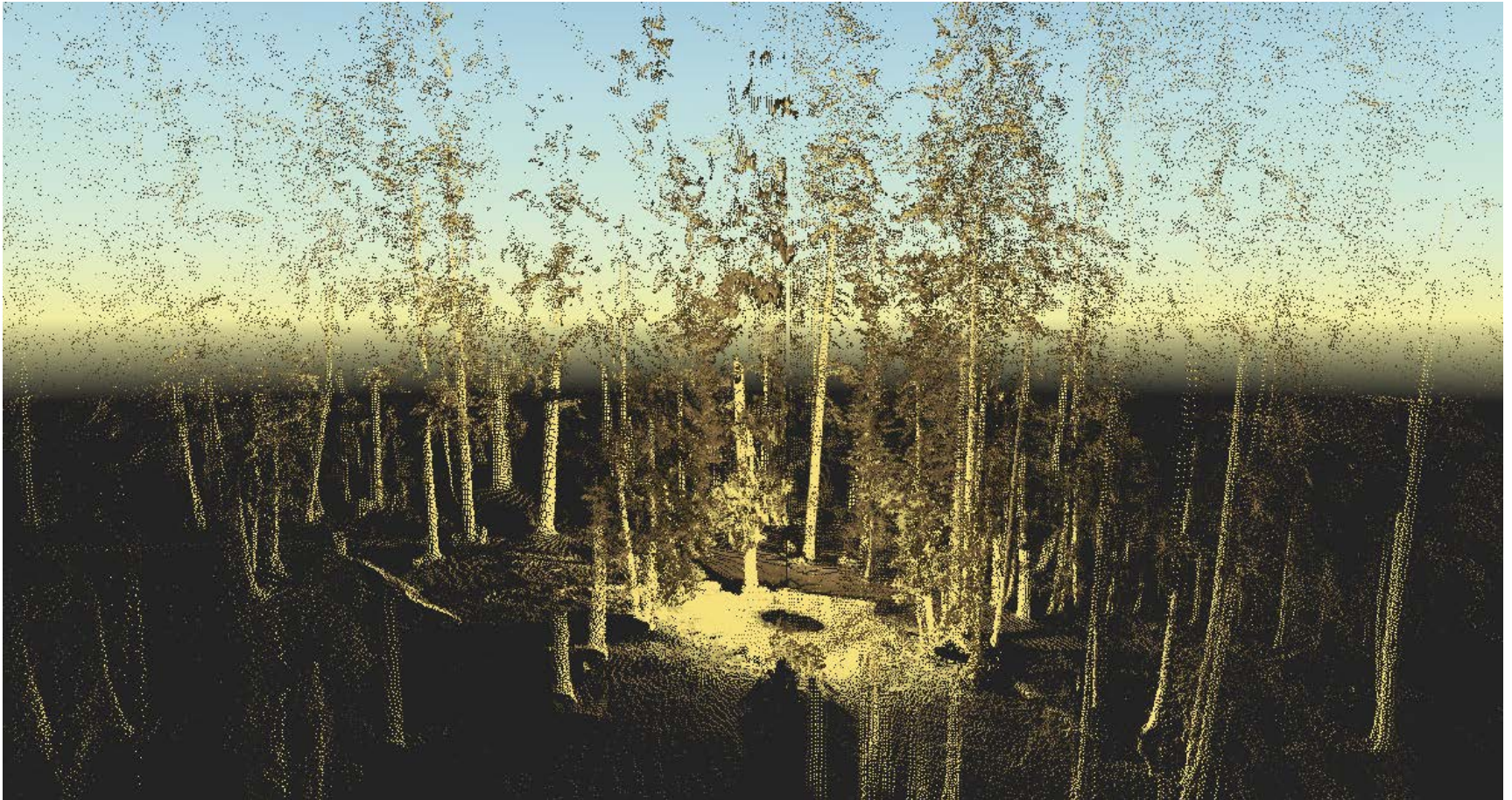
Point Cloud Model

$P_{gap}(r)$ is a step function for separated hard hits

$$\begin{aligned}\rho_{app}(r) &= -\frac{\partial P_{gap}(r)}{\partial r} \rho_a \\ &= \sum_{j=1}^N \rho_j \varphi(r - r_j)\end{aligned}$$

- Model obtained by filtering with the pulse function separating signal and noise and finding peaks
- Multiple scattering and close hits can be modelled by using an increased FWHM where needed
- The sets (ρ_j, r_j) form a point cloud with (θ, ϕ, r) converted to (x, y, z) (plus FWHM in advanced models)
- Original geometry and relationships MUST be recorded

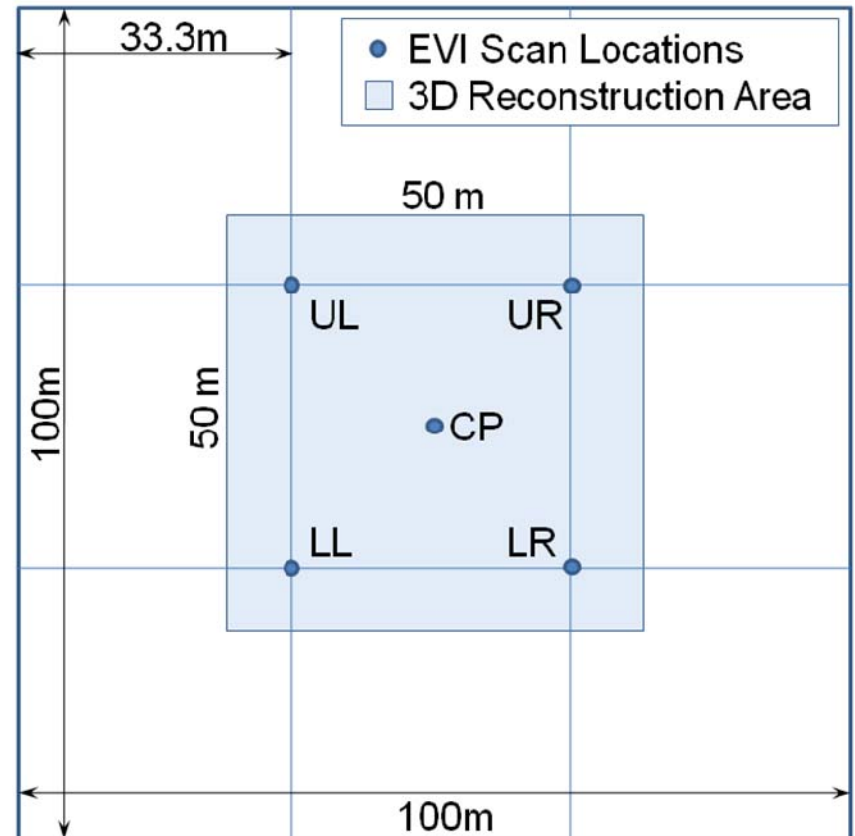
Point Cloud for a Single Scan (Yang Xiaoyuan)



- Site 305 (center scan), a high-elevation red fir stand
- Shade of color shows apparent reflectance at the point

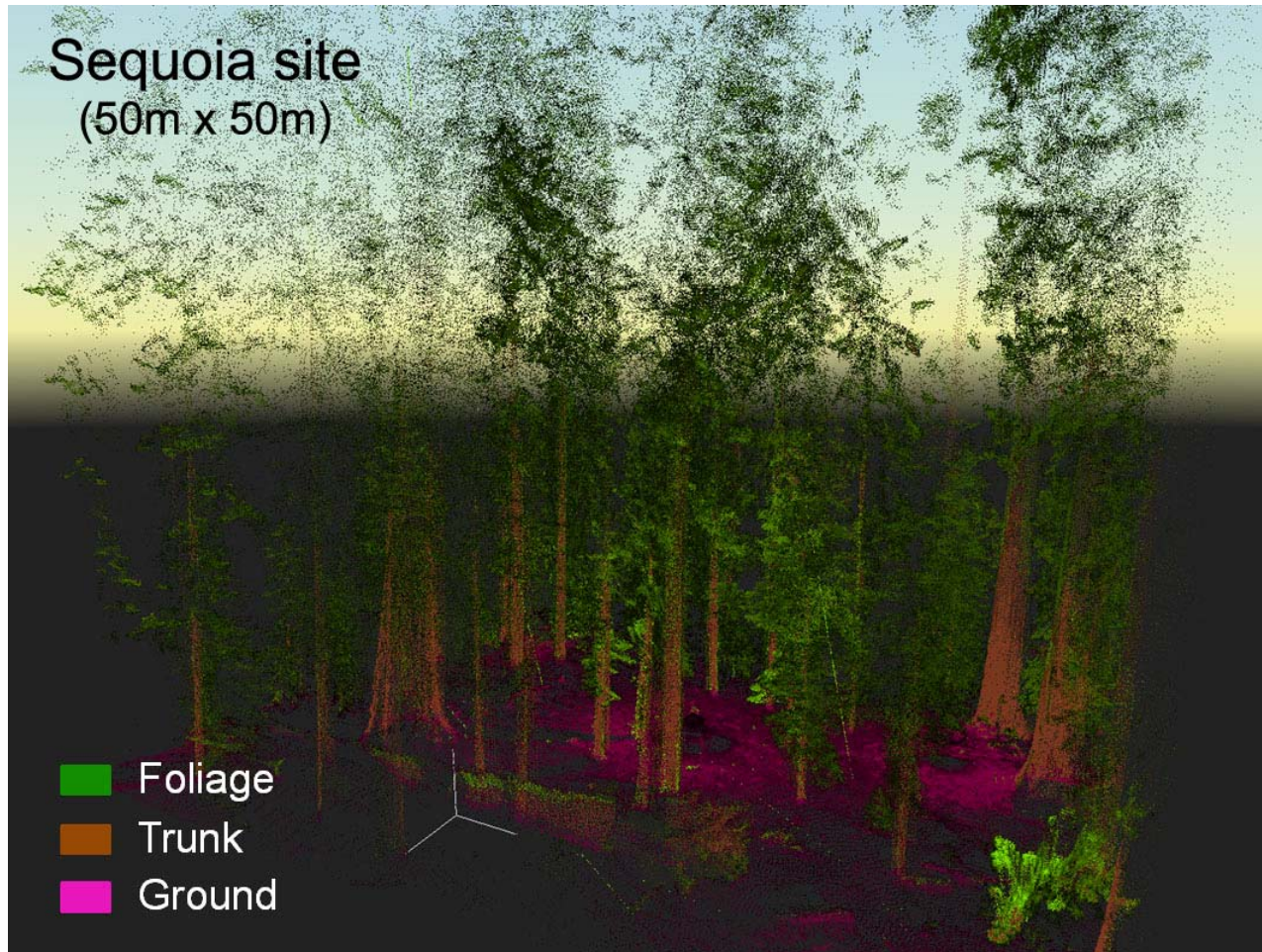
Point Cloud Registration (Yang Xiaoyuan)

- Combine multiple EVI scans
 - Use peg-to-peg position information from field records
- 3D matching procedure
 - X, Y, Z translation
 - X, Y, Z rotation angles



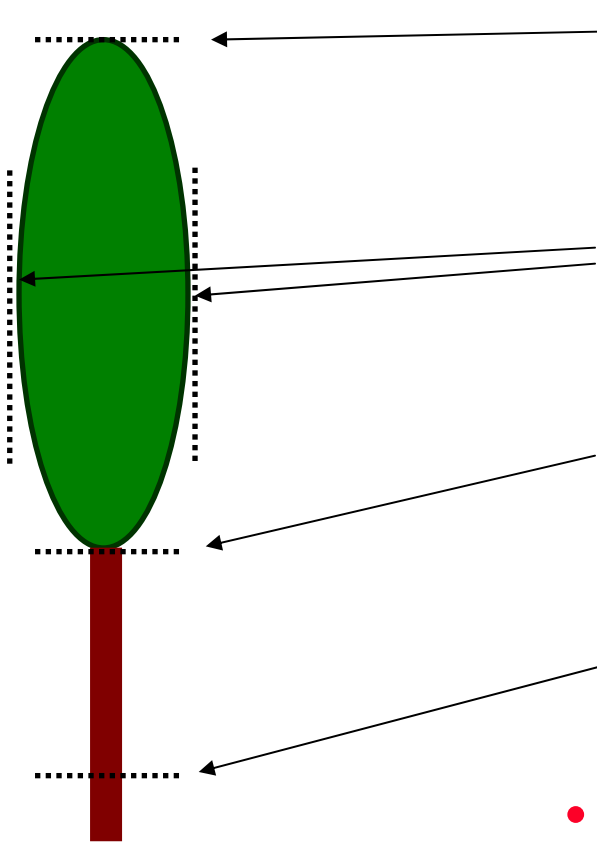
EVI scan design for 2008 Sierra Nevada field sites

Point Cloud Merge & Classification Result (Yang Xiaoyuan)



- Merged point cloud classified into leaf, trunk, and ground returns
- Three trees (2 right, 1 left) are Giant Sequoia with DBH of 3-4 m.

Canopy Parameter Extraction from “Measure Trees”



- Canopy parameters taken from reconstructions
 - Canopy height
 - Tree height measured from the ground to the top of the tree
 - Crown size
 - Measured both along and across slope and then averaged
 - Height to Crown
 - Height measured from ground to the bottom of the tree crown
 - DBH (Diameter at Breast Height)
 - Measured at 1.4 m from ground
- Retrievals compared to field records

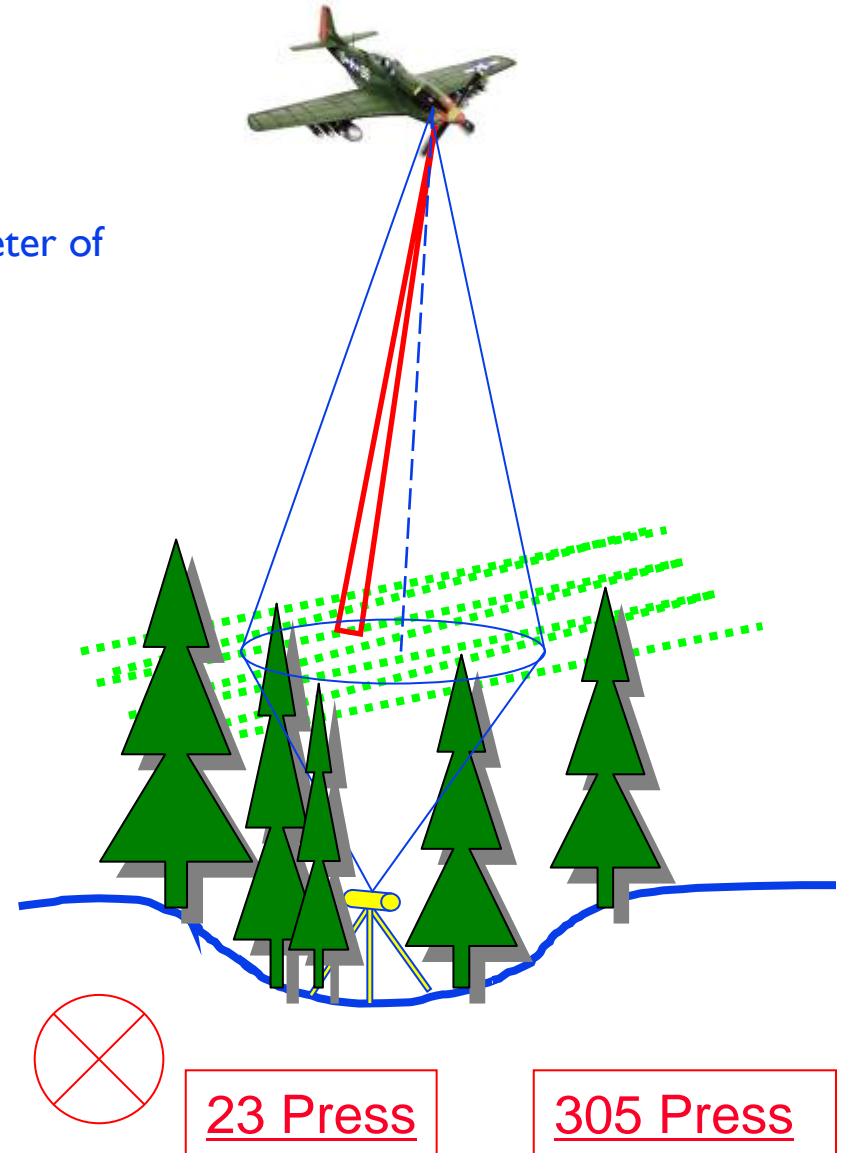
LVIS vs. EVI system (Zhao Feng)

LVIS System Specifications

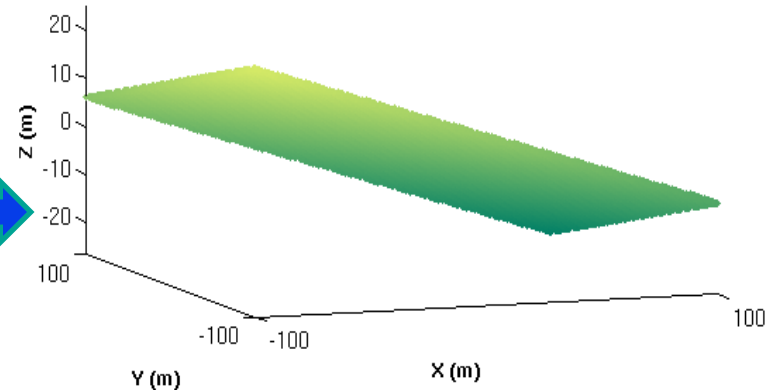
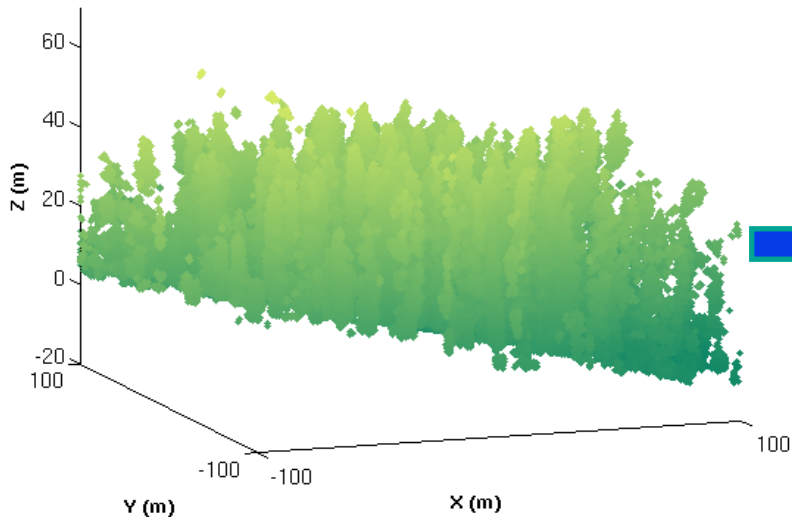
- Wavelength: 1064 nm
- Maximum Altitude: >10km ; ~ 6km for 305 site
- Detector FOV: 8 mrad.
- Maximum Off nadir angle: 12 mrad, nominal diameter of 20meter
- Near nadir view
- Slope is dependent on forest site condition

EVI System Specifications

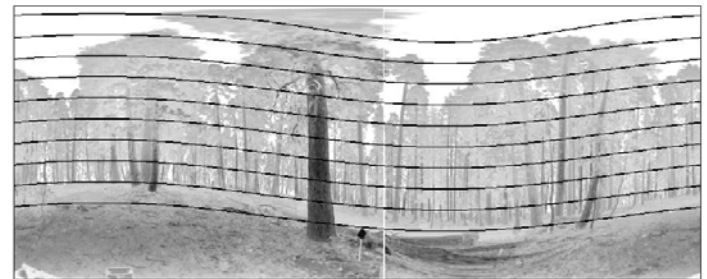
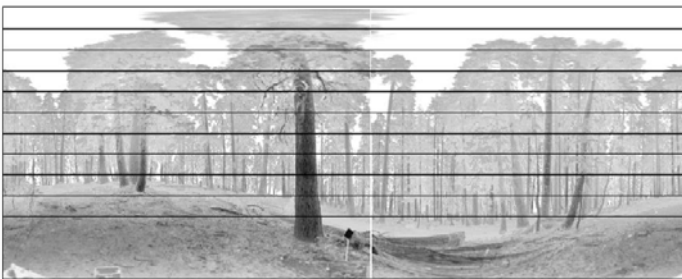
- Wavelength: 1064 nm
- Altitude: 1.7 m above ground
- Detector FOV: 5 mrad with an initial expanded diameter of 23 cm
- Footprint size vary with range
- Full hemispherical view
- Slope can be well characterized



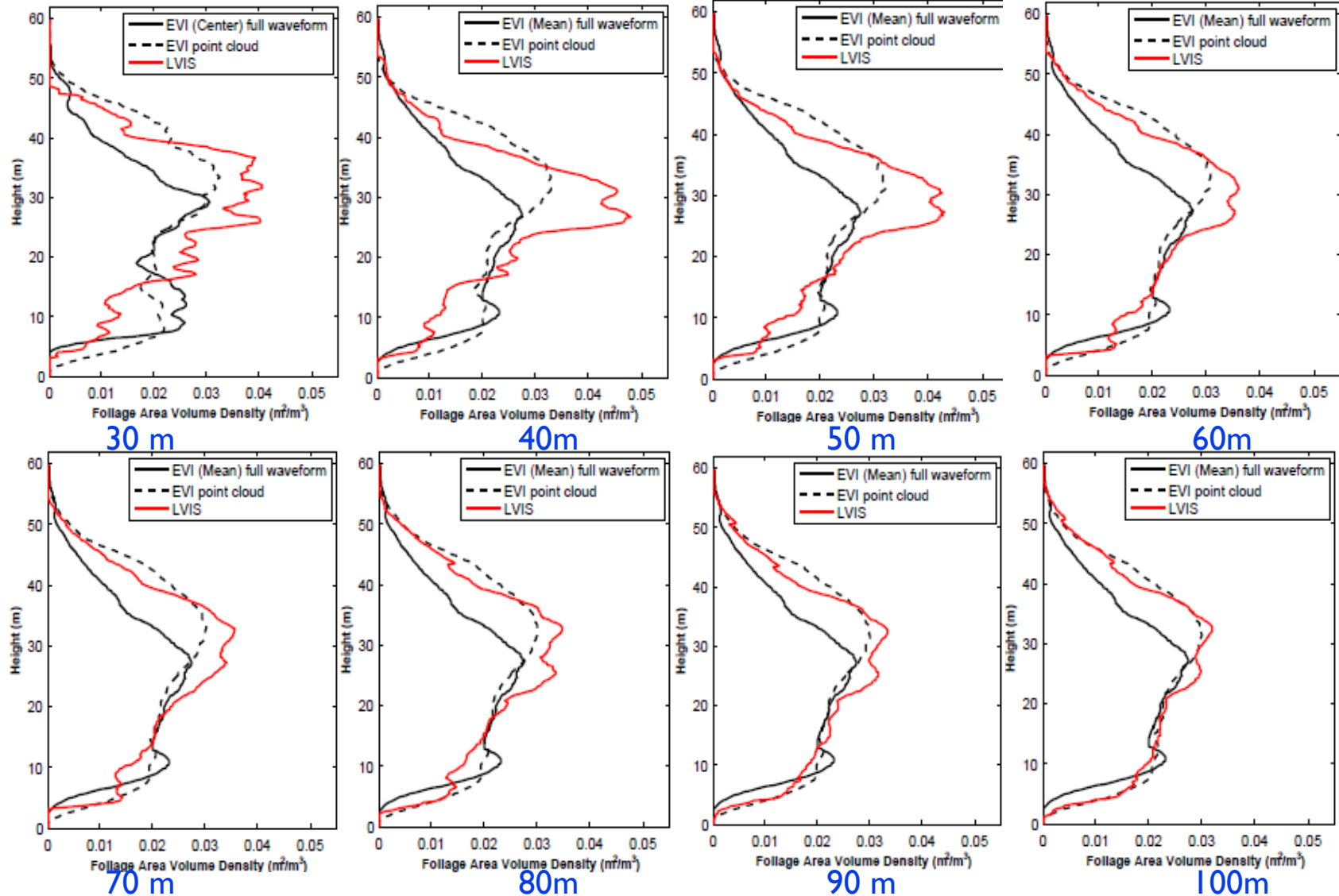
Slope Effect on EVI Full Waveform (Zhao Feng)



$$\cos \gamma = \cos \theta \cdot \cos \beta + \sin \theta \cdot \sin \beta \cdot \cos(\varphi - \varphi')$$



Foliage Profiles (w Slope Removal) (Yang & Zhao)



Note: Local topography corrected

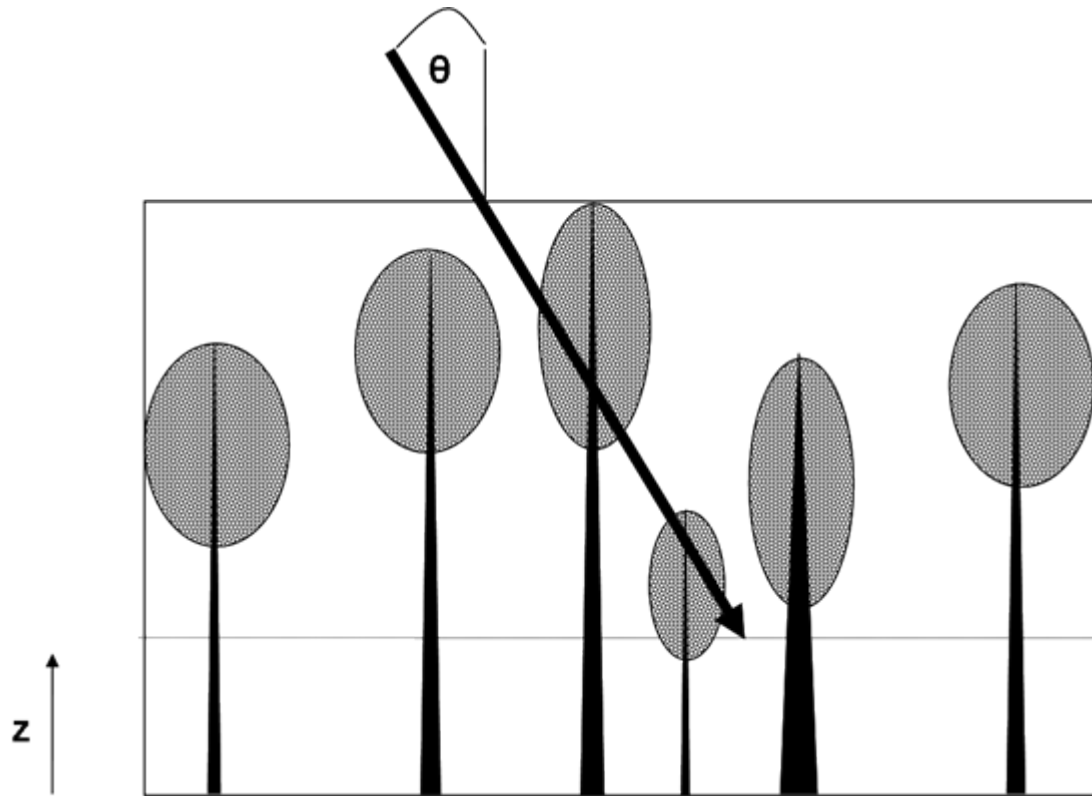
How to fully integrate data from ground and air?

- When the calibration is right, EVI profiles with LAI and height outputs agree in magnitude with ground data and in form with LVIS (LVIS needs scaling to LAI) in the overstorey.
- LVIS “sees” less understorey in most cases.
- To bring airborne data like LVIS and ground based like EVI together, you could use models – but it may also need the airborne data to be calibrated.

Steps to modelling forest and Lidar data

- Work reported here originates with Li Xiaowen (Li and Strahler, 1988) and his work on Pgap for GO model canopies;
- Results from Nilson (1999), GORT (Ni-Meister et al. 2001; 2010) and variants using GO modelling (Newnham 2005, Haverd et al., 2011; Lovell et al., 2011) are consistent;
- Two or more layers used, variable heights, trees with trunks and spheroidal crowns filled with leaves. Models developed adjusted to fit the EVI observed Pgap profile;
- Most work to date has explored forward models based on direct EVI “measure tree” data and parameter sensitivity.

The Geometric Optics (GO) Canopy Model



Pgap model schematic. Allometric relations and GB data can guide Relations between DBH, crown size, height etc. All parameters Can have random distributions.

GO Model for Canopy (Haverd et al, 2011)

For tree (and trunk) density λ the projections of canopy elements
On plane at z m at zenith angle θ is approx by a Boolean Model:

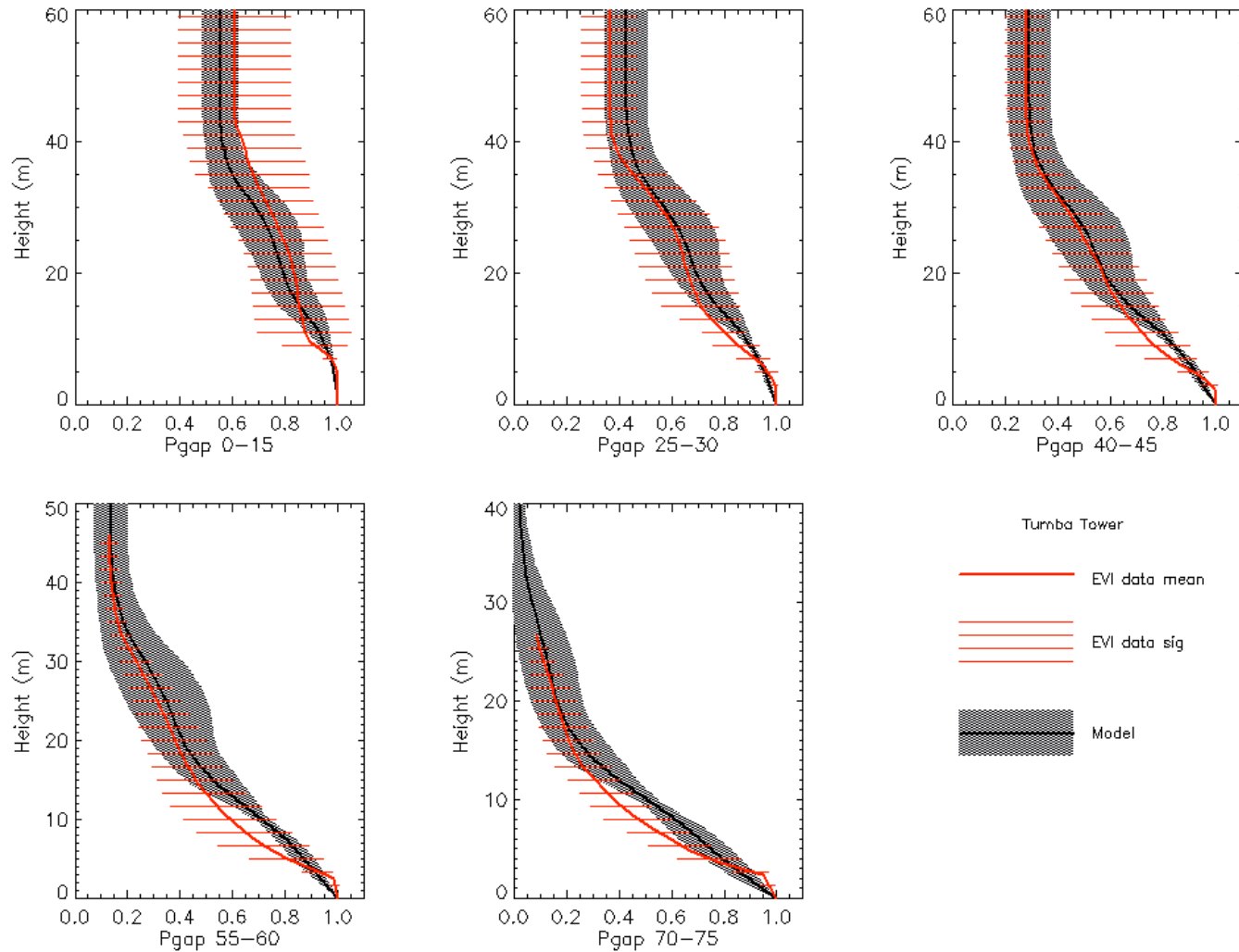
$$P_{gap}(\theta, z) = e^{-\lambda \overline{A_c(\theta, z)(1 - P_{wc}(\theta, z))}} e^{-\lambda \overline{A_t(\theta, z)}}$$

$A_c(\theta, z)$ projected area of a crown element

$P_{wc}(\theta, z)$ gap probability through the crown element

$A_t(\theta, z)$ projected area of a trunk element

Gap Probability Profiles - Mixed Eucalypt (Jenny Lovell, 2011)



How can models “calibrate” airborne?

- If airborne are calibrated it may also be possible to “invert” model aspects that are allowed to vary from those identified at EVI sites using allometry and assuming some basic properties;
- More simply, the relationship between common direct LVIS measures such as RH50 etc may be calibrated against models based on EVI sites and assumed variations incorporating allometry etc. In this case, airborne does not need to be calibrated in the physical measurement sense;
- But it still needs to be done – if it can be done. But if it can be done for LVIS – satellite is possible too

Final Summary

- Field work with GB Lidar has the same trade-offs as all forest measurement;
- Number and extent of sites versus effort at each and cost versus value of information;
- Maybe a few detailed sites with measure trees and many more fast scan (advanced cruising) sites is the best design?
- GB data can be direct and comprehensive, and modelling is possible;
- The potential to “calibrate” airborne and also space borne exists and is waiting to be taken up.

People involved in EVI/DWEL 2000-2011

BU Team:

Alan Strahler
Curtis Woodcock
Crystal Schaaf
Tian Yao
Feng Zhao
Xiaoyuan Yang
Zhuosen Wang
Wenge Ni-Meister
Miguel Román
Mitchell Schull
Yanmin Shuai
Shihyan Lee
Qingsong Zhang

CSIRO Team:

Darius Culvenor
Jenny Lovell
Glenn Newnham
David Jupp (ret.)

DWEL Team:

Supriya Chakrabati
Timothy Cook
Robert Marshall
Ewan Douglas
Jason Martel

Field Team:



Harvesting productivity analysis using LiDAR

Alam, M. M.¹, Strandgard, M.², Brown, M.³ & Fox, J. C.⁴

Department of Forest and Ecosystem Science
The University of Melbourne

¹Email: mmalam@student.unimelb.edu.au

²Email: mnstra@unimelb.edu.au

³Email: mwbrown@unimelb.edu.au

⁴Email: jcfox@unimelb.edu.au

Abstract

Mechanised harvesting operations are common in Australia because of their increased productivity and efficiency, improved worker safety and reduced cost of operations. Most research has found that the productivity and efficiency of a mechanised harvesting system is affected by a number of factors including forest stand characteristics (tree size or piece size, stand density, undergrowth), terrain variables (slope, rocks, woody debris), operators' skill and machinery limitations. The purpose of the study was to use remote sensing technology to quantify these forest stand and terrain factors (particularly slope) and hence derive relationships to predict harvester productivity from remote sensing data.

A case study was conducted in mature radiata pine (*Pinus radiata*) plantation at Mount Burr Reserve Forest, South Australia (37.61° S, 140.44° E). LiDAR (Light Detection And Ranging) flown in 2007 was used to identify and quantify stand and terrain factors (particularly tree size). A time and motion study conducted during final harvest was used to estimate the impact of each factor (tree size and slope) on harvester productivity. Tree size estimates derived from the LiDAR data were grown to the point of harvest using empirical growth models. The point of harvest tree size estimates were ground-truthed against harvester measurements of the same trees. Empirical models were then developed to enable the LiDAR-derived estimates of tree size to be used to estimate productivity of harvesting equipment. The robustness of these relationships will be tested by applying the model to areas not used in the development process.

Key words: Harvesting system, Remote sensing, LiDAR, Productivity, Harvester, Radiata pine

1. Introduction

Productivity (m³ per Productive Machine Hours (PMH)) and efficiency of a harvesting machine and/or system is affected by the characteristics of the forestry machinery, stand condition (i.e. tree shape, tree size, crown size, tree volume and the type and density of trees), extraction rate, (i.e. the ratio between basal area harvested and basal area before harvest), terrain conditions (i.e. slope, ground roughness, ground strength, water course, roads etc.), and the skill of operator (Brunberg *et al.* 1989; Lageson 1997; Visser *et al.* 2009). Many studies have demonstrated 'tree size' to be the most influential factor in harvesting and extraction productivity, with productivity increasing and costs decreasing with increasing tree size (Kellogg and Bettinger 1994; Nakagawa *et al.* 2007; Visser *et al.* 2009). However, the rate of increase in harvesting productivity is less for larger trees and, beyond a "sweet spot", further increases in tree size reduce productivity as the extra time to cut and process the stem outweighs the volume gain (Visser *et al.* 2009). This study will evaluate only 'tree size' (tree volume) as an influential factor affecting harvester productivity for the study area.

Time study is a common practice to measure work or productivity of a system. For machines and equipment, time consumption is calculated for every work element which in turn is used to

estimate its productivity. Work elements of time study in a harvesting system include moving of the machine and/or its boom, felling, delimiting, crosscutting, and bunching (Nakagawa *et al.* 2007).

Early time study methods involved using stopwatches and paper to measure and record time for each machine activity (Howard 1989). This method is tedious, expensive and error prone (Olsen and Kellogg 1983). Another method uses a video camera to capture the harvesting operation activities and record time simultaneously (Wang *et al.* 2003). This method requires field measurements, so extra time and resources are also required. However, such traditional methods are being replaced by computer-based time study methods. In this method, the computer program records time with the built-in clock and collects volume information at the same time. Other studies developed an automated time study of felling and skidding (McDonald 1999; McDonald and Rummer 2000) which recorded machine movement with a GPS and provided gross time study data but not detailed elemental times.

This study used computer-based time study to collect tree volume information including tree Diameter at Breast Height (DBH) and processed length with their GPS (Global Positioning System) locations. Time elements were recorded using Timer Pro Professional software 2010 instead of using computer generated tree-wise time elements in order to achieve time elements for each work element including moving / positioning of the machine, tree felling, processing of each log (cut-to-length), stacking / bunching, travel time, delay time etc.

Time study-based regression equations are generated to express equipment productivity (Gardner 1982). In this study, regression models will be developed which in turn will be used to develop models that will predict the productivity of a harvesting system based on LiDAR (Light Detection And Ranging)-derived estimates of tree size.

LiDAR has been used to detect individual trees, predict tree heights and volume (e.g. Brandtberg 1999; Hyyppa and Inkinen 1999; Lim *et al.* 2003a; Naesset 2004; Woods *et al.* 2008) and diameter distributions (e.g. Gobakkenn and Næsset 2005; Thomas *et al.* 2008). It can also accurately and cost-effectively derive DTM (Digital Terrain Model) or DEM (Digital Elevation Model) compared with conventional photogrammetry especially where the ground surface is not visible in dense forest or in low relief areas such as wetlands or flat plains (Baltsavias 1999; Lefsky *et al.* 2002; Younan *et al.* 2002; Patenaude *et al.* 2004).

Construction of LiDAR-derived Canopy Height Models (CHMs) with distinguishable tree crowns is the prerequisite in order to apply any method for individual tree detection (Holmgren and Persson 2004). The canopy height models (CHMs) are obtained by subtracting the DTMs from corresponding DSMs (Digital Surface Models).

Estimation of tree-level productivity requires the measurements of a large number of trees which is time consuming, laborious, tedious and expensive. However, LiDAR is widely used to derive tree characteristics at the individual or stand level.

The aim of this paper is to (i) Examine the effect of tree size on productivity of a harvester (Valmet 475), for a plantation forest in clear felling operation using a Cut-to-Length (CTL) harvesting method, (ii) Undertake statistical analysis of time elements and productivity of the harvester, (iii) Develop predictive models for the harvester (iv) Evaluate the ability of LiDAR to estimate tree volumes and (v) Develop a productivity model based on LiDAR-derived tree volume.

2. Material and methods

2.1 Study area

A 35-year old radiata pine plantation with a density of 230 trees ha⁻¹ was selected at Mount Burr Reserve Forest (37.61° S, 140.44° E), Brennans locality, South Australia. Trees were planted at 4m X 2m spacing. However, it had been thinned three times prior to the final harvest. Heights and DBHs of the trees ranged from 20 to 35m and 26.4 to 53.5cm respectively. Forest soil comprised of aeolian sands. The site was almost flat with slopes less than 11° (approximately).

2.2 LiDAR Data

LiDAR data was collected for the study area using an ALTM (Airborne Laser Terrain Mapping) 3100 LiDAR system and an inertia measurement unit (IMU) on 20th July, 2007. The scanner system transmitted laser pulses at 1064 nm (near-infrared) and received multiple returns of each pulse. First return density was 2.6m⁻² and returns per pulse were 1 to 4. First and last return pulses were acquired to characterize forest tree structure and terrain surface respectively. The flying altitude was 1100 meter Above Sea Level (mASL), pulse repetition rate was 33000 s⁻¹, maximum scanning angle was 12.5°, beam divergence was 0.20 (mradians) and footprint diameter was 22cm. The horizontal and vertical accuracy of data were 0.55m and 0.20m, respectively.

Processed LiDAR data was supplied by Forestry South Australia in LAS file format with average point spacing of 0.52 points m⁻². Ground and Non-ground returns were classified by the data provider.

Ground and non-ground LiDAR points were used to construct a digital elevation model (DEM) and a digital surface model (DSM) with a 2m cell size for the study area. A slope class map was derived from the DEM. Slope classes were within a range of approximately 0-11 degree where the trees were felled and processed. There were some small areas with more than 11° slope that would not have affected the operation of the harvester. Since LiDAR points provide GPS (Global Positioning System) locations, DSM was used to match field measured tree locations (GPS). The DEM was subtracted from DSM in order to acquire tree heights.

2.3 Time study data

2.3.1 Harvester selection

A harvester (Valmet 475 with a Rosin 997 harvesting head) fitted with DASA Control Systems computer was used to carry out the harvesting operation. The harvester head was properly calibrated at the beginning of the operation in order to accurately record tree measurements. The harvester is designed to perform harvesting operation up to 20° slope and to efficiently handle trees with DBH up to 80cm.

2.3.2 Harvesting operation recording

The harvesting operation carried out by an experienced operator was recorded by a video camera from approximately 20m distance under normal and sunny weather condition on 03 February, 2011 between 11:05:40 am to 12:55:30pm. One hundred and one trees were felled and processed in the operation. Data on time elements and tree characteristics were collected for all trees.

2.3.3 Time elements extraction

The video was played with MS Windows media player and time in centi-second for each work element was recorded with PDA (Personal Digital Assistant, also known as hand-held computer)-based Timer Pro Professional software. Moving / Positioning of the machine, Tree felling, Processing of each log, Stacking / Bunching, Travel Time, Brushing and Clearing were considered as work elements.

Description of time elements:

Moving: Begins when the harvester starts to move and when it stops moving to perform some activity.

Positioning: This is the time between the boom starting to swing toward a tree and machine head is clamped on the tree.

Tree Felling: This is the time between when the felling starts and the tree touches the ground.

Tree Processing: This is the time between the harvester head starts to run and the last processed log is dropped onto the ground.

Stacking / Bunching: This is the time between when the harvester grabs a log and drops it onto the pile.

Brushing/ Clearing: This is the time taken to process unmerchantable trees and clear undergrowth.

Travel time: This is the time to travel to and from where the harvesting took.

Delay time (Mechanical): This is the time that occurs due to mechanical failure.

Delay time (Personal): This is the time that occurs due to operator's personal activity.

2.3.4 Tree characteristics extraction

The DASA onboard computer system, fitted to the harvester was used to measure tree characteristics including tree DBH, log length and volume of processed logs and GPS tree location. Merchantable tree length (m) and volume (m³) for each tree were estimated by summing lengths and volumes of all logs of the relevant trees respectively. Unmerchantable top end of each tree was ocularly measured (average length, 1.75m) and added to the merchantable tree length of the same tree in order to estimate whole tree height (m) of each tree.

2.3.5 Productivity model development

Productivity was calculated individually for each tree using the following formula-

$$\text{Productivity (m}^3\text{/PMH0)} = \text{tree volume (m}^3\text{)} / \text{cycle time (PHM0)} * 60$$

Where tree volume was extracted from the DASA computer system and cycle time is the time spent by the harvester to completely process an individual tree.

Linear regression [$Y = a + b(X)$] analysis was performed to predict productivity of the harvester where X is the independent variable, tree volumes (m³); Y is the dependent variable, productivity (m³/PMH0) and a & b are the coefficients. Statistical software Minitab16 was used to derive the productivity model. In the analysis natural log (LN) of volume was used to fit the data.

2.3.6 Volume model development

The same software package and linear regression equation [$Y = a + b(X)$] was used to predict individual tree volume with the exception that X is the independent variable, tree heights (m); Y is the dependent variable, tree volume (m³) and a & b are the coefficients.

3. Results and Discussion

3.1. Productivity prediction

No relationship was found between tree volume and harvesting time elements other than felling and processing. Therefore these time elements (moving / positioning of the machine, stacking / bunching of logs, travel and brushing and clearing) were averaged separately and added to measured felling, and processing time for all trees to estimate pro rata harvesting cycle times (Nurminen *et al.* 2006). Productive time was defined as machine operating hours excluding delay time (PMH0).

Summary statistics of tree characteristics and cycle time are presented in Table 1 and Table 2 respectively:

Table 1: Tree characteristics of the study area

Attribute	Mean	Std. Dev.	Min.	Max.
Tree height(m)	27.96	2.85	20.71	35.02
DBH(cm)	38.31	5.29	26.40	53.50
Tree vol.(m3)	1.73	0.50	0.78	3.13

Table 2: Pro rata cycle time statistics

Pro Rata Cycle Time (min.)	
Mean	0.941
Standard Deviation	0.187
Minimum	0.672
Maximum	1.740

Figure 1 represents the productivity model for tree volume (m³) vs productivity (m³/PMH0):

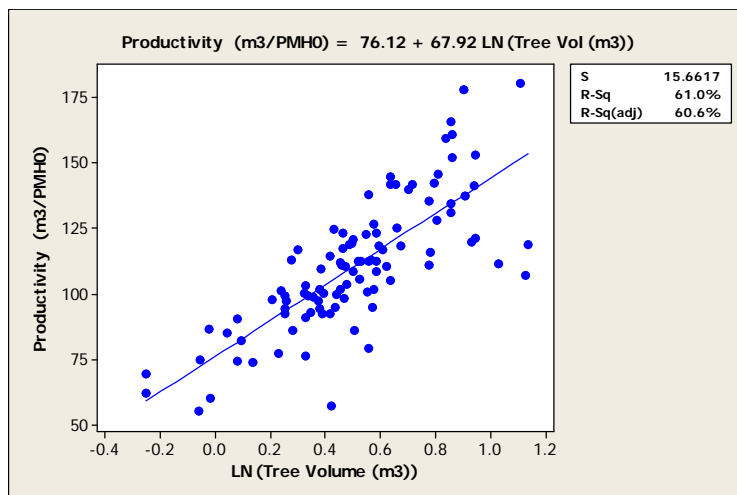


Figure 1: Productivity (m³/PMH0) (pro rata time) against tree volume (m³)

Linear regression based productivity model describes a good fit to the data ($R^2= 61.0\%$) (Figure 1) and RMSE (Root Mean Squared Error) of 15.46. The general trend of the model shows that productivity increases with the increase of tree volume. This result is consistent with other studies (e.g. Jirousek *et al.* 2007; Nakagawa *et al.* 2007).

The study was confined to tree volumes from 0.78m³ to 3.13m³. Therefore, the model may not be suitable to predict productivity beyond this volume range (e.g. greater than 3.5m³). In addition, this model is based on a single data set. To achieve predictability for larger trees (volume) further study is required which would be based on a wider range of tree volumes and more study sites.

3.2. Volume prediction

The model predicting volume from field measured height was developed to predict volume from LiDAR-derived height, because LiDAR can directly derive only height. Figure 2 represents the model predicting volume (m^3) from field measured height:

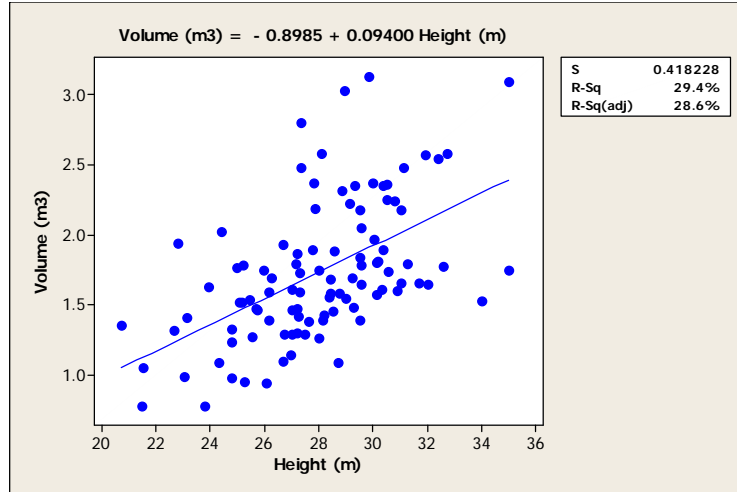


Figure 2: Volume (m^3) prediction from height (m)

The model shows that volume increases with increasing height. However, the model describes only 29.4 % of the variability in the data ($R^2 = 29.4\%$) of the data. Thus the model indicates that tree height alone may not be a good predictor of volume. However, Holmgren *et al.*(2003) estimated stem volume with somewhat lower accuracy from LiDAR-derived tree height and stem number as predicting variables.

This model is limited to a tree height range of 20 to 36m (approximately) and volume range 0.78 to $3.15m^3$. This model may not be suitable for tree sizes out of this range. Because, the rate of increase in machine productivity decreases with increasing tree size with Visser *et al.* (2009) finding that beyond a “sweet spot” further increases in tree size reduce productivity as the extra time to cut and process the stem outweighs the volume gain.

4. Data validation

This model [Volume (m^3) = -0.8985 + 0.09400 Height (m)] would be used to predict volume (m^3) from LiDAR-derived height. In order to check the accuracy of LiDAR-derived height, LiDAR-height class was compared with field height class. LiDAR-derived tree heights were measured from the LiDAR points surrounding field measured trees. Figure 3 represents the distribution of height from field measurements and LiDAR:

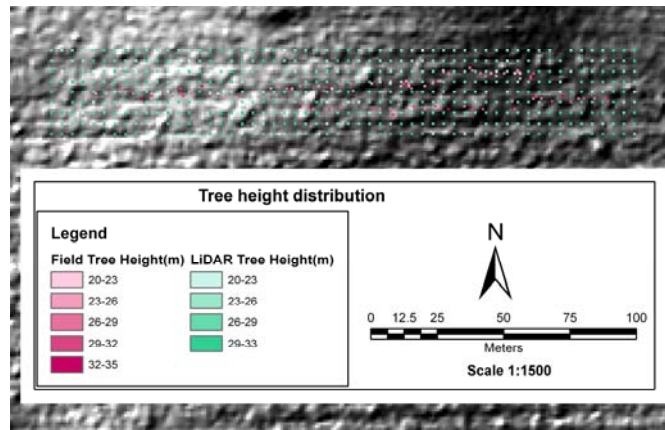


Figure 3: Field and LiDAR-tree distribution (raster format)

Lidar was taken in 2007 and this height has been adjusted at the time of harvesting operation (2011) using the yield table produced for the study area (Lewis *et al.* 1976). It was found that the predominant height increased by 1-1.5m.

The result shows that the mean LiDAR height range (approximately 20-34m, including adjustment) is similar to field measured height (20-35m). Underestimation of LiDAR-derived mean heights is consistent with other studies as LiDAR pulses rarely hit the tip of the tree (e.g. Magnussen and Boudewyn 1998; Heurich 2008).

LiDAR-derived tree height or crown height may not be a reliable predictor for volume estimation (Figure 2). LiDAR-derived tree height and crown diameter can be used to predict individual tree stem diameter and then tree height and stem diameter can be used to calculate stem volume (e.g. Persson *et al.* 2002). LiDAR-derived crown width (Hyypya and Inkinen 1999; Hyypya *et al.* 2001) may enhance its ability to estimate tree volume, because crown width or crown height is highly correlated to DBH (Jakobsons 1970; Sprinz and Burkhart 1987; Gill *et al.* 2000; Peper *et al.* 2001). DBH is a function of tree height and tree height can be derived from LiDAR, thus volume of gross-merchantable timber can indirectly be modelled from LiDAR tree heights (Lim *et al.* 2003b).

5. Conclusion

The methods for estimating productivity based on field measured volume data showed greater predictability. Therefore, LiDAR-derived tree volume may be used to estimate productivity. Volume prediction from height alone shows poor predictability. Therefore it is suggested that DBH and / or crown width functions combined with height should be established for more accurate tree volume prediction and hence better estimation of harvester productivity from LiDAR data.

Acknowledgements

The authors would like to thank Cooperative Research Centre for Forestry (CRC for Forestry) for providing fund for the project and other supports. We also would like to thank Forestry South Australia for organizing harvest operation and providing LiDAR data.

References

1. Baltsavias, E.P., 1999. A comparison between photogrammetry and laser scanning. *Photogrammetry and Remote Sensing*, 54, 83-94.

2. Brandtberg, T., 1999. Automatic individual tree-based analysis of high spatial resolution remotely sensed data. Doctoral Thesis, Swedish University of Agricultural Sciences, Silvestria 118, Uppsala, Sweden.
3. Brunberg, T., Thelin, A. and Westerling, S., 1989. Basic data for productivity standards for single-grip harvesters in thinning operations. Report No 3, The Forest Operations Institute of Sweden: 21.
4. Gardner, R.B., 1982. Estimating production rates and operating costs of timber harvesting equipment in the northern Rockies. General technical report INT: 118, United States Department. of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station Odgen, Utah: 26.
5. Gill, S.J., Biging, G.S. and Murphy, E.C., 2000. Modeling conifer tree crown radius and estimating canopy cover. *Forest Ecology and Management*, 126, 405-416.
6. Gobakkenn, T. and Næsset, E., 2005. Weibull and percentile models for lidar-based estimation of basal area distribution. *Scandinavian Journal of Forest Research*, 20, 490-502.
7. Heurich, M., 2008. Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the Bavarian Forest National Park. *Forest Ecology and Management*, 255, 2416– 2433.
8. Holmgren, J., Nilsson, M. and Olsson, H., 2003. Estimation of tree height and stem volume on plots-using airborne laser scanning. *Forest Science*, 49, 419-428.
9. Holmgren, J. and Persson, A., 2004. Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment*, 90, 415-423.
10. Howard, A.F., 1989. A sequential approach to sampling design for time studies of cable yarding operations. *Canadian Journal of Forest Research*, 19, 973-980.
11. Hyypä, J. and Inkinen, M., 1999. Detecting and estimating attributes for single trees using laser scanner. *Photogrammetric Journal of Finland*, 16, 27–42.
12. Hyypä, J., Kelle, O., Lehtikainen, M. and Inkinen, M., 2001. A segmentation-based method to retrieve stem volume estimates from 3-dimensional tree height models produced by laser scanner. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 969– 975.
13. Jakobsons, A., 1970. The correlation between the diameter of the tree crown and other tree factors - mainly the breast-height diameter. Analysis based on sample trees from the National Forest Survey. Report 14, Department of Forest Survey, Royal College of Forestry Stockholm, Sweden: 75.
14. Jirousek, R., Klvac, R. and Skoupy, A., 2007. Productivity and Costs of the mechanised cut-to-length wood harvesting system in clear-felling operations. *Journal of Forest Science*, 53, 476-482.
15. Kellogg, L.D. and Bettinger, P., 1994. Thinning productivity and cost for mechanized cut-to-length system in the Northwest pacific coast region of the USA. *International Journal of Forest Engineering*, 5, 43-54.

16. Lageson, H., 1997. Effects of thinning type on the harvester productivity and on the residual stand. *Journal of Forest Engineering*, 8, 7-14.
17. Lefsky, M.A., Cohen, W.B., Parker, G.G. and Harding, D.J., 2002. Lidar remote sensing for ecosystem studies. *Bioscience*, 52, 19-30.
18. Lewis, N.B., Keeves, A. and Leech, J.W., 1976. Yield regulation in South Australiaa Pinus Radiata plantations. Woods and Forest Department, South Australia, Bulletin No.23, A.B. JAMES.
19. Lim, K., Treitz, P., Baldwin, K., Morrison, I. and Green, J., 2003a. Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing*, 29, 658-678.
20. Lim, K., Treitz, P., Wulder, M., St-Onge, B. and Flood, M., 2003b. LiDAR remote sensing of forest structure. *Progress in Physical Geography*, 27, 88-106.
21. Magnussen, S. and Boudewyn, P., 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research*, 28, 1016–1031.
22. McDonald, T., 1999. Time study of harvesting equipment using GPS-derived positional data, Forestry engineering for tomorrow, GIS technical papers, Edinburgh University, Edinburgh, Scotland.
23. McDonald, T. and Rummer, B., 2000. Automatic time study of feller-buncher, The 23rd annual meeting of council on forest engineering, COFE, Corvallis, OR.
24. Naesset, E., 2004. Accuracy of forest inventory using airborne laser scanning: Evaluating the first Nordic full-scale operational project. *Scandinavian Journal of Forest Research*, 19, 554-557.
25. Nakagawa, M., Hamatsu, J., Saitou, T. and Ishida, H., 2007. Effects of tree size on productivity and time required for work elements in selective thinning by a harvester. *International Journal of Forest Engineering*, 18, 43-48.
26. Nurminen, T., Korpunen, H. and Uusitalo, J., 2006. Time consumption analysis of the mechanized cut-to-length harvesting system. *Silva Fennica*, 40, 335–363.
27. Olsen, E.D. and Kellogg, L.D., 1983. Comparison of time-study techniques for evaluating logging production. *Transactions of the American Society of Agricultural Engineers*, 26, 1665-1668, 1672.
28. Patenaude, G., Hill, R.A., Milne, R., Gaveau, D.L., Briggs, B.B. and Dawson, T.P., 2004. Quantifying forest above ground carbon content using LiDAR remote sensing. *Remote Sensing of Environment*, 93, 368–380.
29. Peper, P.J., McPherson, E.G. and Mori, S.M., 2001. Equations for predicting diameter, height, crown width and leaf area of San Joaquin Valley street trees. *Journal of Arboriculture*, 27, 306-317.

30. Persson, A., Holmgren, J. and Soderman, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing*, 68, 925-932.
31. Sprinz, P.T. and Burkhart, H.E., 1987. Relationships between tree crown, stem, and stand characteristics in unthinned loblolly-pine plantations. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 17, 534-538.
32. Thomas, V., Oliver, R.D., Lim, K. and Woods, M., 2008. LiDAR and Weibull modeling of diameter and basal area. *Forestry Chronicle*, 84, 866-875.
33. Visser, R., Spinelli, R., Saathof, J. and Fairbrother, S. (2009). Finding the 'Sweet-Spot' of Mechanised Felling Machines. USA: 32nd Annual Meeting of the Council on Forest Engineering (COFE 09). Kings Beach, CA: 10.
34. Wang, J.X., McNeel, J. and Baumgras, J., 2003. A computer-based time study system for timber harvesting operations. *Forest Products Journal*, 53, 47-53.
35. Woods, M., Lim, K. and Treitz, P., 2008. Predicting forest stand variables from LiDAR data in the Great Lakes - St. Lawrence forest of Ontario. *Forestry Chronicle*, 84, 827-839.
36. Younan, N.H., Lee, H.S. and King, R.L., 2002. DTM Error Minimization via Adaptive Smoothing. *IEEE Transactions on Geoscience and Remote Sensing*, 40, 3611-3613.

Scaling plot to stand-level lidar to province in a hierarchical approach to map forest biomass in Nova Scotia

Chris Hopkinson, David Colville, Danik Bourdeau, Suzanne Monette & Robert Maher

Applied Geomatics Research Group,
Centre of Geographic Sciences,
Lawrencetown, Nova Scotia,
B0S 1M0, Canada
Chris.hopkinson@nscc.ca

Abstract

This paper presents a study that used lidar transect, plot and wide area polygon sample data collected across Nova Scotia, Canada from 2005 to 2010 to calibrate and extrapolate above ground forest biomass from permanent sample plots (PSPs) to forest stand polygons to the entire Province. The whole tree dry biomass estimate for the total forest resource inventory (FRI) database in Nova Scotia is $\sim 373 \times 10^6$ tonnes $\pm 39\%$. Where lidar coverage exists, biomass is modelled at the 25 m grid cell resolution, which is a great improvement over the previous ecoregion level estimates, allowing for more effective operational stand management. Given the large spatio-temporal domain of the data sources, one of the major challenges faced in this study was temporal latency between coincident field, lidar and GIS data inputs, which was a significant contributor to the overall level of uncertainty in the result.

1. Introduction

The amount and range of biomass stored within a forested stand is an indicator of its status and ecosystem functioning (Brown, 2002). In Atlantic Canada, revenues from sawlog and pulp wood forestry products, critically important to the rural economy, have been in a steady decline in recent years (APEC, 2008). At the same time, public energy utilities have been rising to the dual challenge of meeting growing energy demands while attempting to reduce greenhouse gas emissions. The Province of Nova Scotia, for example, has committed to 25% renewable energy supply by 2015 and 40% by 2020, and biomass is seen as a potential viable source of long-term carbon-neutral alternative energy to supplement more traditional sources (NSDE, 2010). For all of these reasons, the ability to map forest biomass in Nova Scotia at a scale appropriate for land management has economic, ecological, environmental value.

2. Data sources

The Province of Nova Scotia is a little over 50,000 km² and of this area >80% is forested. The forests of Nova Scotia are catalogued and monitored by the NS Department of Natural Resources (DNR). There are two publicly available and spatially explicit datasets that describe these resources and have been used as the basis for modelling in this project: a PSP database and a FRI GIS database. The PSP database details the attributes of all trees with a stem diameter at breast height (DBH) > 9.1 cm within 3250 plots of 11.3 m radius covering the whole Province. The PSPs are randomly established throughout the forests of Nova Scotia and cover a 400 m² circular area. About half of the 3250 plots were established from 1965 to 1970, while the rest were established between 1998 and 2002. About 650 plots are revisited every year to ensure a five year rotation for each plot. Within the plot, living and dead trees are numbered and several attributes are recorded for each tree including height, DBH, species, signs of disease, cause of death etc.

The FRI database describing the total forest coverage within Nova Scotia contains approximately 1.1 million stand polygons that are delineated from aerial photographs and intended to describe regions of stand similarity within contiguous parcels of land. The FRI database is primarily updated from air photos collected on a ten year rotation. Using paired photographs, interpreters can see in three dimensions and then delineate homogeneous stands of trees to interpret crown closure, stand height, species and land capability. Satellite imagery is used in between aerial photo years to update the FRI for noticeable change, such as clear-cutting. Both PSPs and FRI stands are being continuously updated on a revolving basis as opposed to updating the whole Province at one time.

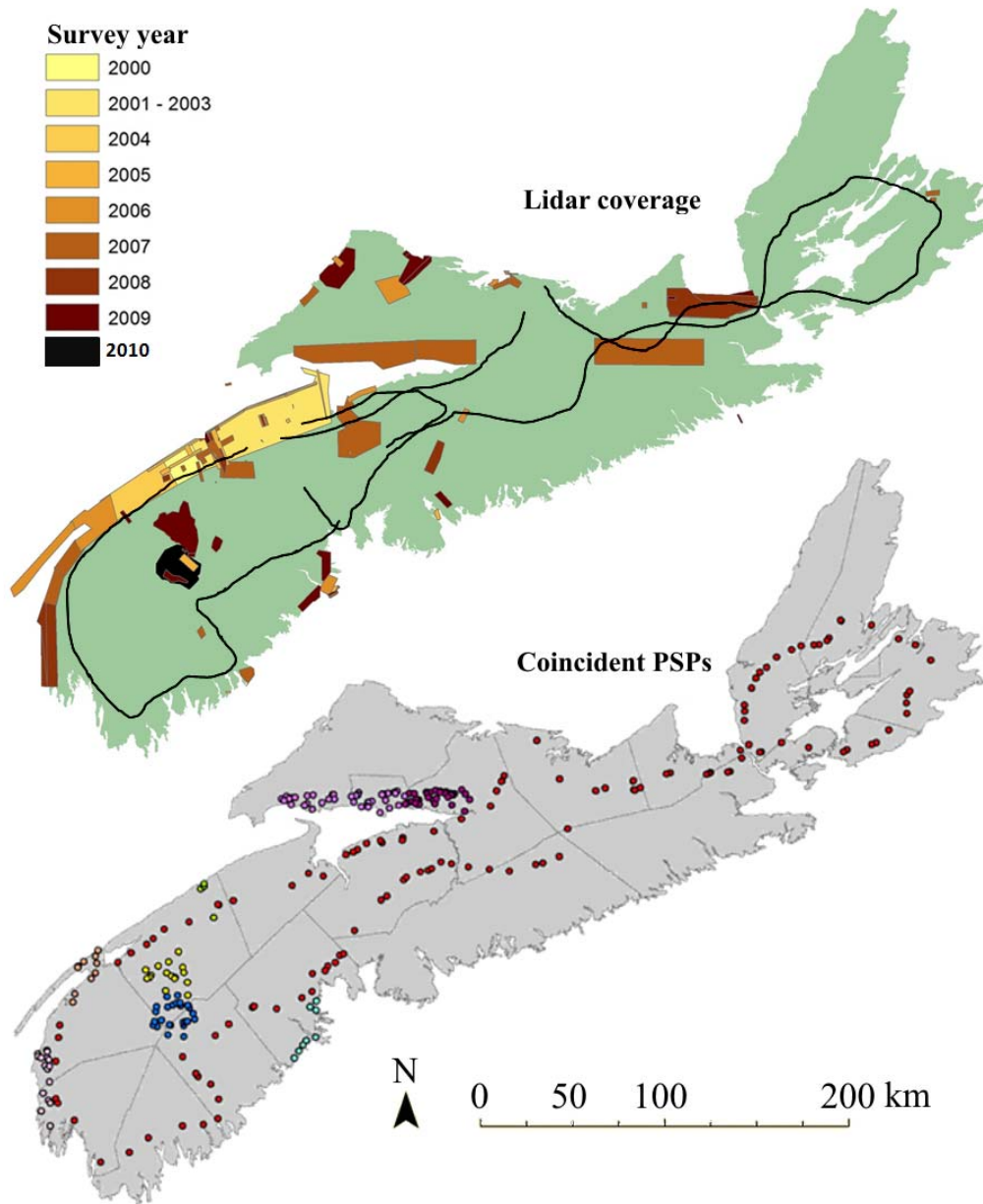


Figure 1: Top - AGRG Lidar survey polygons and sampling transects from 2000 to 2010 within Nova Scotia. Bottom – spatially coincident permanent sample plots that are within two years of lidar survey.

While the DNR GIS databases contain extensive stand-level inventory data covering the Province, the conversion of these data into meaningful estimates of available and sustainable

biomass energy requires calibration (Townsend, 2008). At the local scale, PSP data collected in the field allow reasonably accurate calculation of biomass over small areas (Lambert *et al.* 2005), which have been used to provide coarse estimates of biomass for Nova Scotia down to the ecoregion scale (Townsend, 2008). However, to derive more spatially explicit estimates of biomass at the typical management unit or stand-scale is challenged by the heterogeneity displayed by Acadian mixed wood forests.

To scale between PSP and FRI data layers to develop a spatial model of biomass representing both spatial domains requires a data source that can sample the canopy structure within a plot and allow for effective aggregation at the stand scale. Airborne lidar data have been shown time and again to be ideally suited to the task of plot- and stand-level canopy structure and biomass modelling (Lim *et al.* 2003). Lidar data have been collected across Nova Scotia by the Applied Geomatics Research Group since 2000 (Figure 1). Several polygons and sample transects covering $\sim 10,000 \text{ km}^2$ or $\sim 20\%$ of the area of Nova Scotia have been mapped using an Airborne Laser terrain mapper (ALTM) 3100C (Optech Inc. Toronto, Canada). Only $\sim 50\%$ of the data were suitable for this study, as a threshold of $\sim 1 \text{ point/m}^2$ was applied to ensure a high density of data for subsequent model generation. Of the 3250 provincial PSPs, 281 were spatially coincident with lidar cover, and of these 99 were culled after applying a 2 year temporal buffer. The data sources and associated modelling approaches are described in Table 1.

Table 1: Data sources and attributes used, domains of spatial representation and notes on how the data were used in the Provincial biomass modeling approach

Raw data [number units/ total db units]	Spatial model scale [unit area] (Total db area)	Data attributes	Modeling approach	Modeling purpose
PSP attribute db [258 / 3250] 8%	Plot [400 m ²] (1.3 km ²)	Height, DBH, species, stem count	Based on Lambert et al. (2005). Species divided into hardwood / softwood	Generate biomass 'ground truth' at plot scale
Lidar point cloud [1000km ² /50,000km ²] >2%	Lidar survey coverage [$\sim 1 \text{ m}$ point sampling to 500 km ² polygon] ($\sim 10,000 \text{ km}^2$)	Height percentiles & vertical distribution ratios	Linear, quadratic and JGLS regression models with < 2 variables to predict bole / whole tree biomass	Calibrate FRI stand data by extrapolating PSP-based lidar model
GIS stand polygon [2639 / 1.1 million] 2.4%	FRI stand [$\sim 0.01 \text{ km}^2$ to 10 km ²] ($\sim 42,000 \text{ km}^2$)	Mean canopy height & closure	Linear, quadratic and JGLS regression models of bole / whole tree biomass	Simulate stand level biomass and aggregate up to Province

3. Summary of methods

After initial quality control of the coincident lidar and PSP data, there were 182 PSPs between the years 2005 and 2010 available to train and test a lidar-based model of biomass. PSP data were used to derive ground truth estimates of bole and whole tree dry biomass through the application of a robust individual tree biomass model that was constructed from plot-level sample data collected across Canada (Lambert *et al.* 2005). The lidar biomass data were then used to train an FRI-based model using attributes from 1873 stand polygons which could be applied to the entire Province. Lidar data metrics collected over 13 different survey missions using the same ALTM 3100C sensor were extracted for each of the PSPs. Summary statistics extracted using FUSION (McGaughey, 2010) describing the vertical within-plot lidar frequency distributions and point cloud ratios were used to describe canopy height and cover attributes.

These lidar ‘metrics’ were then correlated with the associated PSP biomass estimates to construct predictive models of biomass. The lidar-based maps of biomass were segmented into corresponding FRI polygons, which were then scaled up to the Province using a new model based on the FRI stand attributes of canopy height and closure.

4. Results & discussion

4.1 Plot-level biomass

Statistical descriptions of each PSP lidar point cloud dataset generated in the FUSION software (McGaughey, 2010) were tested for inter-correlation and suitability for use in multivariate biomass modelling (Table 2). As expected, all height-based frequency distribution metrics demonstrated high inter-correlation, as did most ratio-based metrics. While other derivatives of the frequency distribution are possible, it was decided to keep the PSP biomass modelling approach simple to allow for maximum transferability across diverse lidar datasets. Consequently, models tests were limited to two variables; one height-based and one ratio-based metric, as these demonstrated the least inter-correlation. Furthermore, height metrics are an index of canopy height (e.g. Naesset, 1997) while ratio metrics are an index of canopy cover (e.g. Hopkinson and Chasmer, 2008). These two attributes are logical indices of the two physical dimensions (height and width) that are fundamental to volume, and therefore, biomass calculations.

Table 2: Correlation matrix of selected PSP lidar point cloud frequency distribution attributes extracted from FUSION. Shaded cells denote correlations between height and ratio metrics. **Bold** values illustrate weakest inter-correlation and therefore suitability for multivariate modeling

	ElevMean	ElevStdDev	ElevP70	ElevP75	ElevP90	ElevP95	ElevP99	Perc1stReturns>Mean	PercAllReturns>Mean	AllReturns>Mean/ Total1stReturns*100	Perc1stReturns>1.50	PercAllReturns>1.50	AllReturns>1.50/ Total1stReturns *100
ElevMean	1												
ElevStdDev	0.81	1											
ElevP70	0.97	0.89	1										
ElevP75	0.97	0.91	0.97	1									
ElevP90	0.93	0.95	0.95	0.97	1								
ElevP95	0.91	0.95	0.93	0.95	0.99	1							
ElevP99	0.86	0.93	0.87	0.90	0.96	0.98	1						
Perc1stReturns>Mean	0.76	0.58	0.78	0.75	0.62	0.58	0.51	1					
PercAllReturns>Mean	0.72	0.45	0.71	0.67	0.53	0.49	0.41	0.93	1				
AllReturns>Mean/Total1stReturns*100	0.80	0.72	0.85	0.83	0.74	0.70	0.64	0.97	0.86	1			
Perc1stReturns>1.50	0.71	0.46	0.69	0.67	0.58	0.56	0.49	0.83	0.84	0.79	1		
PercAllReturns>1.50	0.70	0.34	0.63	0.61	0.53	0.51	0.45	0.64	0.73	0.59	0.91	1	
AllReturns>1.50/Total1stReturns*100	0.86	0.66	0.85	0.84	0.78	0.76	0.70	0.82	0.75	0.86	0.91	0.85	1

Using PSP data and the biomass model of Lambert *et al.* (2005), several lidar biomass models were trained and tested for whole tree and bole. For the lidar model, species information was ignored but the ratio of softwood to hardwood stems was considered in PSP model training. Root mean square error (RMSE) for the best polynomial regression whole tree and bole lidar model remained approximately the same in both test datasets at ~ 26%. The explanation of variance in the test data was greater for bole biomass estimates at 75%, than whole tree biomass at 63%. All models were significant at the 99% level of confidence. These results indicated that

the lidar models were robust enough to provide stand-level summaries.

4.2 Stand-level biomass

A challenge that became immediately apparent during the process of relating lidar biomass estimates to FRI stand-level attributes, was the temporal latency between the two datasets (FRI data dating back to 1990s in extreme cases, while lidar data ranging from 2005 to 2010). This latency was most evident when comparing stand-level FRI mean tree height with the mean maximum of the lidar data aggregated into 25 m grid cells (Figure 2). Using canopy height as an indicator, quality control procedures were put in place to systematically remove the most obvious outliers (due to growth and clearcuts) using objective height and date criteria. However, even after this quality control process, the latency between lidar and FRI still has the potential to propagate uncertainty into the model. The nature of this error is such that any stand growth, decay, thinning or clear cut occurring following the last FRI stand update and preceding the associated lidar acquisition will lead to divergence between the lidar and FRI attributes. As long as the forests are in a state of dynamic equilibrium (i.e. the Provincial forest resource as a whole is neither expanding nor contracting), then these stand-level biases will not necessarily lead to a systematic bias in the overall population statistics. In practical terms, this means that we expect the model to display a high level of variance at the stand scale but when aggregating biomass estimates to larger and larger spatial domains, there should be a level of compensation between over- and under-estimates.

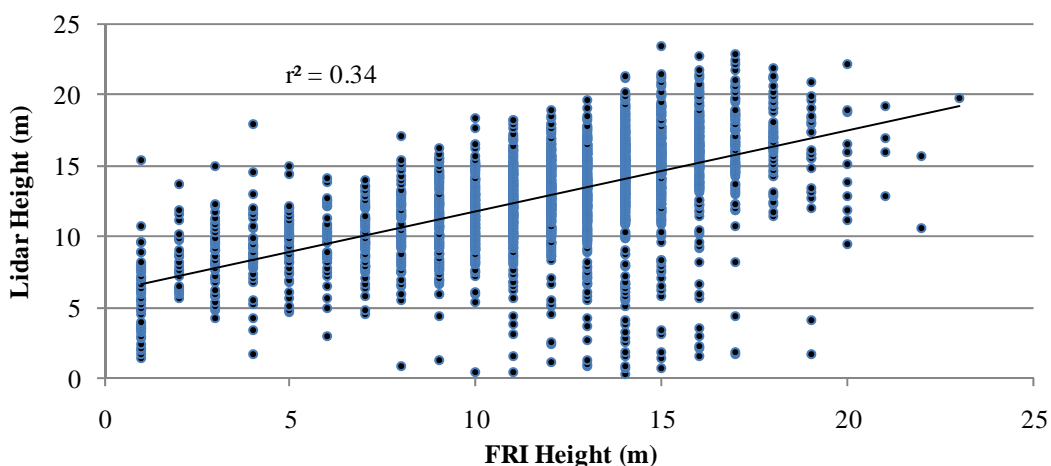


Figure 2: Mean maximum grid level lidar canopy height vs. FRI mean canopy height for stands completely covered by lidar data (n = 2639).

Given the accepted level of uncertainty in the models at this stage of the analysis, no attempt was made to develop highly sophisticated or complex regression models between stand-level lidar biomass and FRI stand attributes. Crown closure and mean canopy height were chosen as the FRI attributes to be used in the stand biomass models as they most closely resembled the lidar metrics used in the previous modelling step. Similar to the PSP results, the RMSE in stand biomass approximated 27% both for whole tree and bole. However, the explanation of variance dropped to 41% and 43%, respectively, most likely a large function of the temporal latency issues described above.

4.3 Nova Scotia's biomass

Using the hierarchical model development approach summarised above, we derived six estimates of total provincial biomass; three for bole (stem wood) and three for whole tree. The

three modelling approaches did not differ in terms of the data used at each stage of model development, rather the differences are simply in terms of the algorithm construction; ranging from simple single variable linear regression to dual variable quadratic and a further dual variable model that mimicked the structure of the model proposed by Lambert *et al.* (2005). Given the Lambert *et al.* (2005) model was used to derive the ‘ground truth’ plot-level dry biomass estimates from which the rest of the lidar and stand-level predictions are based, the model results are to be considered more reliable if expressed as dry biomass. The range of values for total bole biomass within the Province ranged from 253 – 260 x 10⁶ dry tonnes, while whole tree biomass ranged from 365 – 373 x 10⁶ dry tonnes (Figure 3). Bole biomass is the number to refer to if only stems are to be used as fuel wood and tree tops are to be left in situ for nutrient recycling. For the sake of comparison, the Nova Scotia Department of Natural Resources (DNR) estimate of total living merchantable tree stem biomass in the Province is 309 x 10⁶ dry tonnes (Townsend, 2008). The Canadian Forest Service (CFS) has also developed a largely satellite image-based estimate of total above ground biomass at the 1km pixel resolution (Hall *et al.* 2010), and when this is aggregated to the Province scale a value of 362 x 10⁶ dry tonnes results.

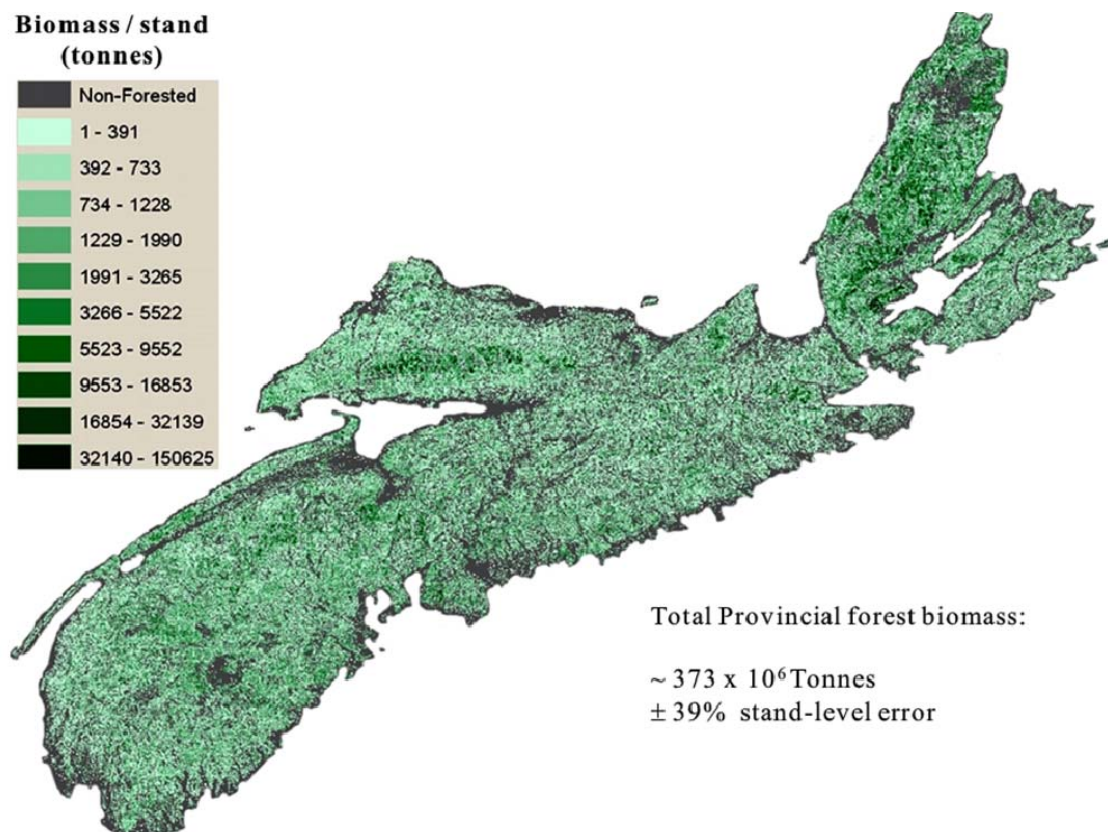


Figure 3: Map of predicted forest stand-level biomass across Nova Scotia

Both the DNR and the CFS estimates of total above ground biomass for the Province of Nova Scotia are lower (by 17% and 3%, respectively) than that generated using the lidar scaling approach described here. The DNR biomass value refers to living whole tree dry biomass but it should be noted that this only considers stems with DBH > 9.1 cm. The whole tree biomass estimate generated in this study includes standing dead stems and given lidar cannot differentiate stems based on DBH it is likely the estimate is further inflated relative to the DNR estimate by inclusion of small and immature tree stems. The primary difference between this and the DNR and CFS approaches is that the approach described here allows calculation of

biomass at the stand scale and is useful for operational planning and decision making. Based on a conversion ratio of 140% for dry to green biomass for typical Acadian mixed wood species (e.g. Shelton and Shapiro, 1976), the estimates above provide values of around 357×10^6 green tonnes for total bole biomass and 514×10^6 green tonnes for whole tree biomass.

Given the approach described uses three modelling steps, each building on the previous, uncertainty will propagate throughout. The RMSE values observed in the model results at each stage, demonstrated errors in the 15% to 30% range. Propagating the RMSE at the PSP, lidar and FRI modelling steps in quadrature compounds to an overall stand- and Province-level error of ~ 39%. This assumes that all errors are random and there is no significant bias.

5. Conclusions

For the time period from 2005 and 2010, 182 PSPs were used to train and test a lidar forest biomass model. This model was then used to train a FRI model from 1869 coincident stands, which was aggregated across all 1.1 million stands in Nova Scotia to arrive at a total above ground forest biomass estimate for the province. This biomass estimate can be expressed several ways but the whole tree dry biomass estimate is $\sim 373 \times 10^6$ tonnes $\pm 39\%$. Where lidar data are available in the Province (about 20% of the land surface area) a spatially explicit estimate of biomass can be generated at the 25 m grid cell resolution. In other areas, biomass can be estimated at the stand-level. The spatial resolution of these estimates constitutes an improvement over previous biomass estimates that were available at the ecoregion (DNR; Townsend, 2008) or 1 km pixel (CFS; Hall et al, 2010) resolutions. These results, therefore, can be used to aid in either stand- or within stand-level forest management practices and in informing forest biomass energy policy in Nova Scotia.

Modeling biomass over such a large area is not without challenges. Greatest of these is obtaining useable model calibration and validation data. In this study, DNR PSP and FRI data were all that were available at the scale required. Both data sources were limited in terms of temporal compatibility with the lidar data that were used to scale between the two. Up to two years of latency in the PSP data is less than ideal given forests grow, die and are managed. However, this was less problematic than the > 10 years of latency for some of the FRI stands. The time discrepancy will introduce larger errors for younger stands and for those that have been clear cut. While objective criteria were used to mitigate such occurrences it is impossible to remove all such instances without manual selection and verification of each stand. Such an approach is not practical at this scale so a substantial amount of model uncertainty remains. However, it is assumed that temporal discrepancies will cause both over- and under-estimation of stand-level biomass, such that there will be a level of compensation.

Acknowledgements

Many thanks to Allyson Fox and Laura Chasmer for their dedication to flying long airborne survey missions and assisting with data processing. Thanks are extended to Dr. Ron Hall of the Canadian Forest Service for providing a 1 km resolution map of total biomass for Nova Scotia. Nova Scotia Power Inc. is acknowledged for funding this project and the Canada Foundation of Innovation for funding AGRG's lidar laboratory. Nova Scotia Department of Natural Resources provided the raw PSP and FRI GIS data layers, while Cory Isenor, Collin Gillis and Heather Morrison assisted with field data collection. Dr Bob McGaughey of the US Department of Agriculture is gratefully acknowledged for his support with FUSION.

References

- Atlantic Provinces Economic Council. 2008. *Building competitiveness in Atlantic Canada's forest industries: A strategy for future prosperity*. Report published by APEC, Halifax, Nova Scotia. 75pp.
- Brown, S. 2002. Measuring carbon in forests: current status and future challenges. *Environ. Pollut.* Vol. 116: 363-372
- Hall, R.J. Skakun, R.S. Beaudoin, A. Wulder, M.A. Arsenault, E.J. Bernier, P.Y. Guindon, L. Luther, J.E. and Gillis. M.D. 2010. Approaches for forest biomass estimation and mapping in Canada. DVD pp. 1988-1991 in *Proc. 2010 IEEE International Geoscience and Remote Sensing Symposium*, Honolulu, Hawaii, USA. July 25-30, 2010. DVD
- Hopkinson, C. and Chasmer, L.E., 2009. Testing lidar models of fractional cover across multiple forest ecozones. *Remote Sensing of Environment*. Vol. 113: 275-288.
- Lambert, M.-C.; Ung, C.-H.; Raulier, F. 2005. Canadian national tree aboveground biomass equations. *Can. J. For. Res.* Vol. 35: 1996-2018.
- Lim K, Treitz, P., Wulder, M.A., St-Onge, B., Flood, M. 2003. LiDAR remote sensing of forest structure. *Progress Phys Geogr.* Vol. 27: 88-106
- McGaughey, R.J. 2010. Fusion/LDV: Software for LiDAR Data Analysis and Visualization. USDA Forest Service. FUSION Manual. March 2010 – FUSION Version 2.80.
- Næsset, E. 1997. Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*. Vol. 52: 49-56.
- Nova Scotia Department of Energy, 2010. *Renewable electricity plan; a path to good jobs, stable prices, and a cleaner environment*. Report published by the Govt of Nova Scotia, Halifax, Nova Scotia. 32pp.
- Shelton, J., Shapiro, A. B. 1976. *The Woodburner's Encyclopedia*. Crossroads Press, Box 33, Waitsfield, VT 05673, U.S.A. 155 pp.
- Townsend, P. 2008. *Forest Biomass of Living Merchantable Trees in Nova Scotia*. Report FOR 2008-9. Nova Scotia Department of Natural Resources. Govt. Nova Scotia publication.

Estimating Stand Volume from Nonparametric Distribution of Airborne LiDAR Data

Doo-Ahn Kwak¹, Taejin Park² & Jong Yeol Lee², Woo-Kyun Lee^{2*}

¹ Environmental GIS/RS Centre, Korea University, Seoul 136-713, South Korea, tulip96@korea.ac.kr

² Division of Environmental Science and Ecological Engineering, Korea University, Seoul 136-713, South Korea, *leewk@korea.ac.kr

Abstract

This study was performed to estimate stand-level volume using the characteristics of vertical and horizontal distribution of airborne Light Detection And Ranging (LiDAR) data. It is found that the height distributional parameters, such as percentile, of LiDAR data reflected on- and in-canopy in a stand have the relationship with stand volume in previous research. However, we assumed that the nonparametric height distribution form of canopy LiDAR returns would be obviously related with the stand volume directly. Nonparametric height distribution was presented to be a continuous line according to the frequency of LiDAR returns by the height. Thereafter, the sum of each height of all canopy returns, which means the area below the continuous line, was compared to stand volume using National Forest Inventory (NFI) data. In addition, for verifying the volume of test stands, the similarity which is the overlapping ratio between the height distribution curves of sample and test stand was calculated. The relationship between the height sum and stand volume was relatively high to be $R^2=0.83$. Based on such relationship, the maximum similarity of each test stand was computed as compared sample stands. As a result, mean similarity and root mean square error (RMSE) of estimated stand volumes were 82% and $34.96\text{m}^3/\text{ha}$ respectively. However, supplementary indices, for non-overlapping part in similar distribution of canopy returns of sample and test stand, are needed to reduce such errors.

Keywords: Nonparametric Distribution, Canopy LiDAR Returns, National Forest Inventory, Similarity, Stand Volume

1. Introduction

The stand volume is also an essential estimator for predicting forest biomass because the volume information can be converted to biomass using the Biomass Conversion and Expansion Factors developed by the Korea Forest Research Institute (KFRI) (Son *et al.*, 2008). Therefore, an accurate forest inventory is critical to carbon sequestration monitoring as well as forest growth stock management (van Aardt *et al.*, 2008). However, accurate and extensive work based on conventional field measurement is labour-intensive, time-consuming and expensive, even if little biased measurements on individual trees and plots are possible (Kwak *et al.*, 2007).

As a solution to these problems, an optical sensor-based remote sensing technique has been used and evaluated over approximately three decades (Tucker *et al.*, 2001). The application of remote sensing to forest measurement using satellite imageries and aerial photographs makes it possible to collect forest growth information for most location including inaccessible areas and to process data immediately (Zhou *et al.*, 2001). However, the optical remote sensing technique suffers the limitations of difficulties in extracting three-dimensional profiles such as the stand height, canopy base height and the crown width, and in the acquisition of an image without the

negative influences of cloud and shade. Recently, airborne Light Detection And Ranging (LiDAR) approaches have attracted attention as a new technology in the field of remote sensing, by which the stand volume and biomass can be estimated using three dimensional (x , y and z) mass points (Lefsky *et al.*, 1999; Popescu and Wynne 2004; Kwak *et al.*, 2010). Several investigations have been performed to prove the utility of LiDAR by estimating the canopy height (Lefsky *et al.*, 1999; Magnussen *et al.*, 1999; Drake and Weishampel 2000; Peterson 2000) and vertical structure (Harding *et al.*, 2001; Parker *et al.*, 2001). As a result, LiDAR remote sensing has been recognized as a reliable provider of important, tree-specific characteristics for estimating the stand volume and biomass (Drake *et al.*, 2002).

Conventional methods require information on the tree height and diameter at breast height (DBH) for estimating the stand volume as these are calculated using allometric functions, such as modelled from field measurements (Kim *et al.*, 2000). However, it is difficult to directly estimate the DBH using an airborne remote sensing technique. To avoid the secondary modelling for DBH deduction, Chen *et al.* (2007) and Kwak *et al.* (2010) suggested a new method to calculate stem volume in single tree level using only the crown geometric volume (CGV) which is the computational crown volume above the crown base height extracted from LiDAR data. Furthermore, the stand volume and biomass have been estimated using height distributional approaches (e.g., median, mode, kurtosis, skewness and percentiles) of a large-footprint LiDAR system (Nelson *et al.*, 1988; Harding *et al.*, 2001; Parker *et al.*, 2001; Drake *et al.*, 2002). Moreover, small-footprint LiDAR has also been used to estimate the stand volume and biomass since the middle of the 1980s (Maclean and Krabill 1986; Nelson *et al.*, 1988). Even if the small-footprint LiDAR data are not waveforms, as can be acquired by a large-footprint LiDAR system, their distributions resemble a waveform because all LiDAR returns are accumulated per sampling unit, resulting in height-frequency distributions according to high vertical stratum (van Aardt *et al.*, 2006). Therefore, the distribution of small-footprint LiDAR data enables characterization of the canopy vertical structure, which is useful in estimating the stand volume and biomass from fine-scale to stand-levels, using height distributional parameters such as percentile, median, mode, kurtosis and skewness (Magnussen and Boudewyn 1998; Means *et al.*, 2000; Næsset 2002).

On the other hand, there were a few researches on the estimation of forest structure using the distribution of LiDAR returns based on probability density function such as normal, weibull and bimodal distribution (Gobakken *et al.*, 2005; Coops *et al.*, 2007; Jaskierniak *et al.*, 2010). However, the estimation based on parametric distribution has the drawback by which the original height value and distribution of LiDAR returns are ignored because the values are generalized during the process. In addition, the distribution of LiDAR returns in uneven forest or stand which has over 2 height clusters may be not normal because they have over 2 stratifications and peaks.

Therefore, in this study, the nonparametric height distribution of LiDAR returns was used to stand volume with Korean National Forest Inventory (NFI) data. As compared with height distribution functions of sample and test stands, the similarity between all stand was analysed. Thereby, the volume of a test stand with the maximum similarity was calculated with the basis of the most similar sample stand.

2. Study area

The study areas were located in Yangpyeong City of Kyeonggi Province (the upper left 127°18'12.7295"E, 37°40'17.6375"N and lower right 127°50'56.4683"E, 37°22'1.8228"N) central South Korea (Figure 1). The area is approximately 878 km² and situated from 160 to 1,157m above sea level. The forest area in study site was dominated by steep hills, and composed to *Pinus koraiensis* (Korean pine), *Larix kaempferi* (Japanese larch), *Pinus densiflora* (Japanese red pine), *Pinus rigida* (Pitch pine) and *Quercus* spp. (Oaks).

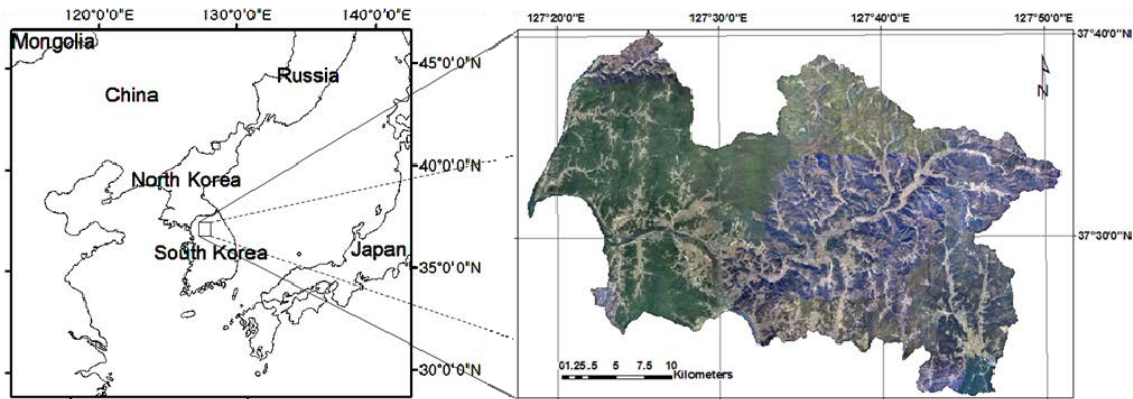


Figure 1. Location of study area presented by mosaic aerial photograph taken from April to May, 2009

3. Acquisition of NFI and LiDAR data

3.1 Selection of sample sites from NFI data

In South Korea, 5th National Forest Inventory (NFI) has been conducted on only forest area (As of 2011, approximately 6.4 million ha) from 2006. The scheme of survey was the systematic sampling with interval of 4 km x 4 km. In addition, there are 4 circular sample plots in one intersection of grid line by 4 km x 4 km. The size per one sample plot was 0.08 ha with 15.96 m radius (Figure 2). In all sites for NFI, the tree species, age, height and diameter at breast height (DBH) of individual tree, and coordinate, elevation, slope and aspect of plot have been measured.

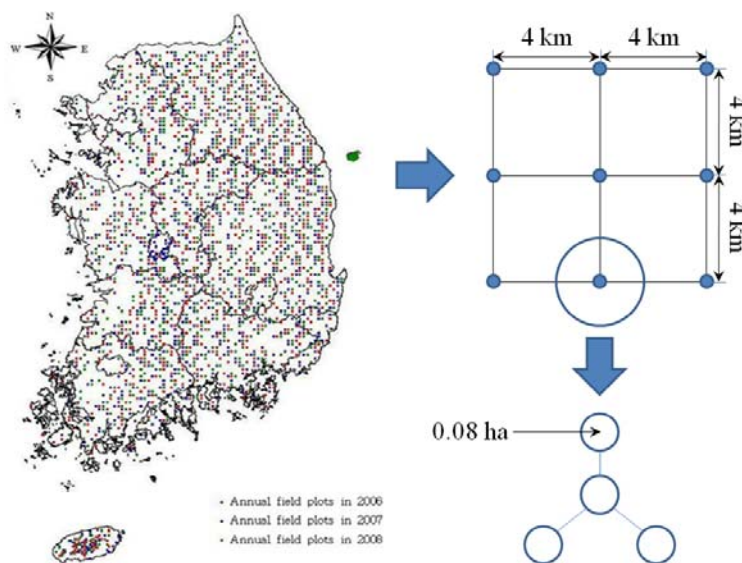


Figure 2. Location distribution and scheme of NFI in South Korea

Surveyed circular 114 plots (0.08 ha, respectively) of Korean pines from NFI data were used as 33 sample for modelling and 81 test plots for verification in Yangpyeong city. The plot (stand) volume was calculated as summing individual tree volumes within each plot. Individual tree volume was also calculated with allometric equations developed by KFRI as shown in Table 1. The coordinate was measured at the center of plot using the GPS Pathfinder ProXR, manufactured by Trimble. To correct for positional errors in the plots surveyed with a single GPS receiver, the acquired GPS data were processed by differential correction using

information of a GPS continuous station near the study area, which acquires precise positional and error correction information every 30 seconds.

Table 1. Allometric equations by class of DBH of Korean Pine (where SV , D and H is the stem volume, DBH and tree height)

Tree species	Class of DBH (cm)	Allometric equation
Korean Pine	2~10 cm	$SV = 0.00006730D^{1.8523}H^{0.9128}$
	12~20 cm	$SV = 0.00006730D^{1.9594}H^{0.9632}$
	20 cm ~	$SV = 0.00006730D^{1.7566}H^{0.9050}$

3.2 Acquisition of airborne LiDAR data

An Optech ALTM 3070 (a discrete LiDAR system) was used to acquire the LiDAR data. The flight was performed from 11th April to 28th May, 2009. The study area was measured at an altitude of 1,000m, with a sampling density of approximately 3~5 points per square meter, with radiometric resolution, scan frequency and scan width of 12bits, 70Hz and $\pm 20^\circ$, respectively. For accurate analysis, we used LiDAR returns within $\pm 10^\circ$ scan angle as overlapped area were eliminated by MicroStation and TerraSolid Program. LiDAR data should be normalized for each return to retain the real height information related to the return hierarchy (van Aardt *et al.* 2006). Therefore, the extracted ground and vegetation returns were normalized to be real height values from the ground, and not height values above sea level. To normalize all returns, a Digital Terrain Model (DTM) was generated using ground returns, with the heights of all returns then subtracted from the DTM heights. Finally, we used the canopy LiDAR returns except understory layer of which height was almost lower than 2.5 m.

4. Method

4.1 Nonparametric distribution of LiDAR data

Previous studies were based on height distributional parameters such as median, mode, percentile, kurtosis and skewness (Nelson *et al.*, 1988; Harding *et al.*, 2001; Parker *et al.*, 2001; Drake *et al.*, 2002). However, such methodologies require the selection of suitable variables and regression analysis throughout all procedure. Moreover, it is dangerous to estimate biophysical parameters, such as tree height, DBH, canopy base height and crown diameter, with the assumption which the distribution of LiDAR data follows the normality in a stand. That's why there are bi- or more-understory by which the stand has over two peaks in the height distribution (Std.) of LiDAR returns (Figure 3). In the case of parametric approach, only mean (μ) and standard deviation (δ) of height of LiDAR returns represent the characteristics of the stand simply. However, in actual forest stand, another statistics with mean and Std. were needed to present the various characteristics. In particular, the forests of South Korea are composed to various tree height and DBH because most forests have been not managed well by owners. Therefore, the assumption of normality of LiDAR returns is not suitable for the analysis of forest biophysical parameters.

Thus we assumed that 1) the distribution of LiDAR returns (particularly canopy LiDAR returns in this study) is nonparametric to describe the stand structure without the loss of its characteristics, and 2) there is the strong relationship between the area of distributional function (sum of height of all canopy returns) and volume within a plot. Furthermore, the sum of height within a plots was calculated under the assumption that the number of LiDAR returns reflected in each plot is same even if they have a few different numbers. Based on such relationship, stand volume was calculated with the similarity between the LiDAR height distribution functions of sample and test sites.

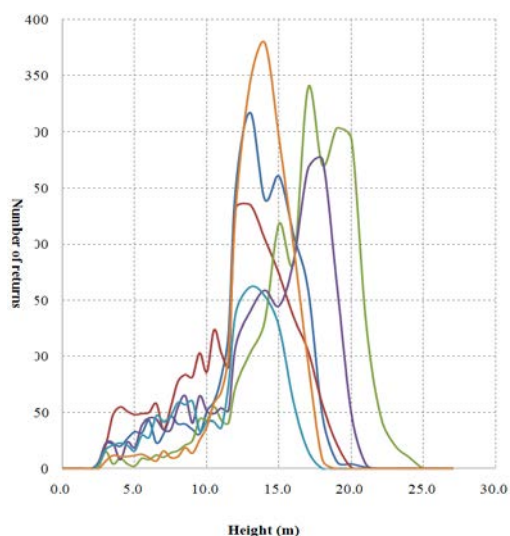


Figure 3. Nonparametric distributions of canopy LiDAR returns

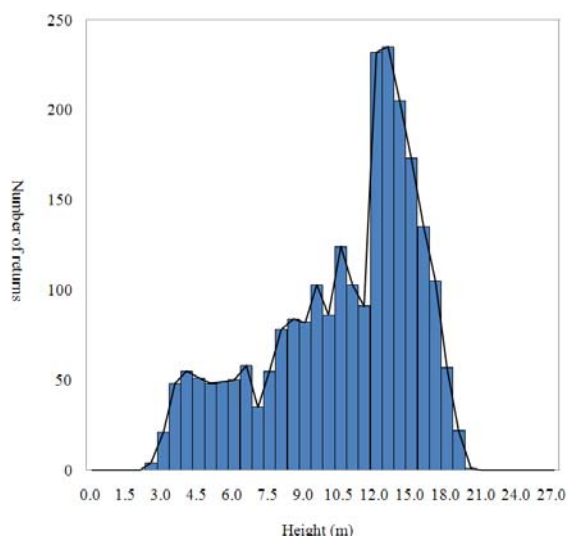


Figure 4. The sum of height of canopy LiDAR returns is equal to area of distributional function

4.2 Relationship between LiDAR height and stand volume

For estimating stand volume using the similarity between the distributional functions of sample and test site, the relationship of area of distributional function and stand volume should be demonstrated. The area of distributional function is equal to the sum of all canopy returns in a stand (Figure 4). In previous studies, it was clarified that percentile data and descriptive statistics were related with stand volume and basal area (Nelson *et al.*, 1988; Harding *et al.*, 2001; Parker *et al.*, 2001; Drake *et al.*, 2002). In this study, it is important that the sum of height of canopy LiDAR returns is correlated with stand volume because the ratio of overlapping area between distributional functions of sample and test sites would be used to estimate stand volume of target test sites directly. However, the estimated regression function was not used to stand volume with the sum of height of canopy LiDAR returns.

4.3 Estimation of stand volume using similarity

The similarity between the distributional function of sample and test sites was calculated with overlapped area of two functions as equation 1.

$$\int_{\alpha}^{\beta} (f(x) \cap g(x)) \quad (1)$$

where α and β are the smallest and highest height, respectively, of overlapped area between distributional functions of sample and test sites. $f(x)$ and $g(x)$ are distributional functions of sample and test site respectively. The overlapping between two functions can be depicted as Figure 6.

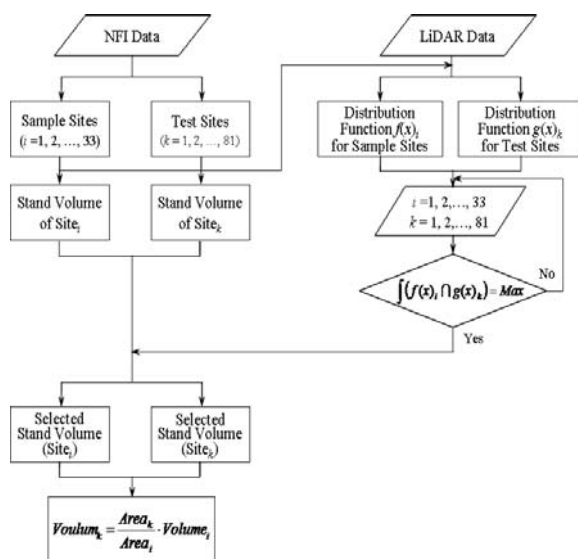


Figure 5. Entire process for estimation of stand volume using the similarity

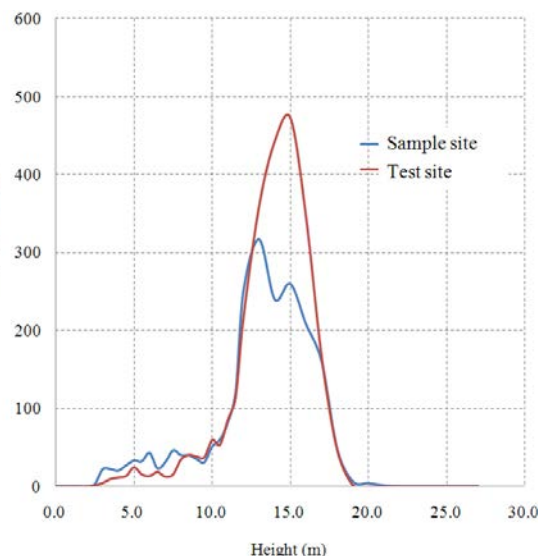


Figure 6. Similarity of distribution function between sample and test site. The similarity is estimated using overlapped area each other.

For estimating the similarity, firstly, we determined 33 and 81 sample and test sites composed of Korean pines. Stand volumes of them were computed using NFI data including tree height and DBH. In addition, canopy LiDAR returns were also extracted from relevant sites, and then the height distributional functions of canopy LiDAR returns for sample and test sites were prepared to estimate overlapped area. The maximum similarity of each test site were computed as compared to the overlapped area with the all distributional functions of sample sites. Thereby, respective 81 test sites were linked with prepared 31 sample sites to have maximum similarity. Finally, a stand volume of test site were calculated with the area ratio of sample and test site and stand volume of sample site which had maximum similarity with a test site (Equation 2).

$$Volume_k = \frac{Area_k}{Area_i} \cdot Volume_i \tag{2}$$

where $Volume_k$ is a stand volume of selected test site, $Volume_i$ is a stand volume of sample site to have maximum similarity with selected test site, $Area_i$ and $Area_k$ are the area of selected sample and test site.

5. Result and discussion

5.1 Correlation of LiDAR height and stand volume

The precedence for clarifying the relationship between the sum of height of canopy LiDAR returns and stand volume were performed as Figure 7. As a result, the coefficient of determination was approximately 0.83 to be relatively high, and there was the linear relationship between them. This is attributed to the fact that the sum of height of canopy LiDAR returns is related with individual tree height and crown diameter in a stand. High value of the sum means that most of LiDAR pulses were almost reflected on high layer of a stand and. Also, several reflection on high layer implies that most of individual crowns were distributed highly and widely on high layer. The stem or stand volume can be computed with tree height and DBH. Crown diameter is also related strongly with DBH. Therefore, the reflection on high layer in a stand represents high stand volume, otherwise low stand volume.

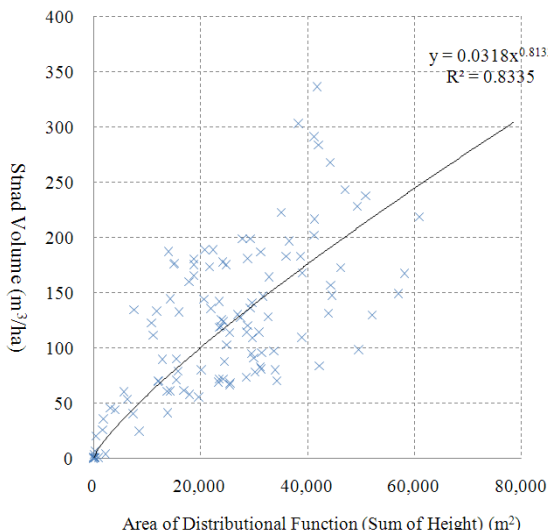


Figure 7. Relationship between sum of height of canopy LiDAR returns and stand volume

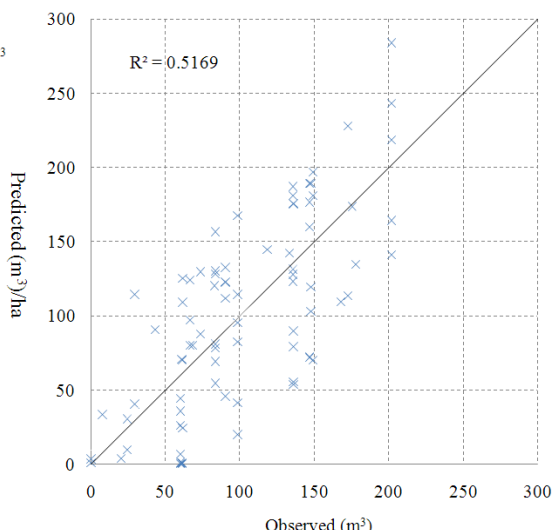


Figure 8. Verification of observed and predicted stand volume using similarity

Based on such relationship, the stand volume could be computed as compared with the similarity of sample and test sites. However, we didn't use the regression function estimated from the relationship between the sum of height and stand volume.

5.2 Analysis of similarity

The distributional similarity of canopy LiDAR returns was estimated with the degree of overlapped area of sample and test sites. Mean, standard deviation, Minimum and Maximum of similarity were analyzed to be 0.82 (82%), 0.10 (10%), 0.33 (33%) and 0.93 (93%), respectively. However, approximately 90% of similarities were distributed around 0.9 when the frequency of similarity depicted on graph (Figure 9). Therefore, the distributions of test sites were almost agreed with those of test sites. When estimating stand volume with the distribution of similarity, we didn't use a few test sites with small similarity considered as outliers.

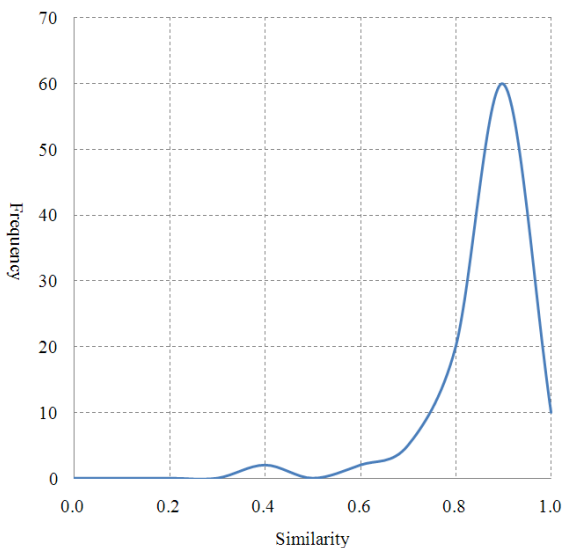


Figure 9. Distribution of similarities between sample and test sites

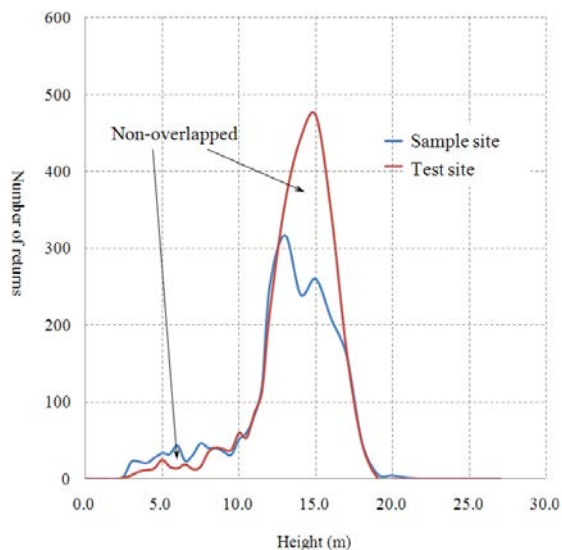


Figure 10. Error by non-overlapped area in distributional function

5.3 Validation of stand volume

After the similarity was analyzed, the stand volume with the maximum similarity between sample and test sites were computed by equation 2. The estimated stand volumes were compared with field measurement from NFI data. As a result, the coefficient of determination (R^2) and root mean square error (RMSE) were estimated to be 0.52 and 34.96 m³/ha respectively (Figure 8) although the similarities were almost distributed around 0.9. This is attributed to the fact that non-overlapped part causes the error as seen in Figure 10. Figure 10 shows the difference of the number of returns in middle height layer between two sites although the distribution pattern of canopy returns are similar each other. This implies that two stands have similar vertical structure but the density of leaves and branches is different. The leaves supplies nutriment by photosynthesis and are related with stand growth directly. Therefore, it is fair that more non-overlapped part leads to more erroneous estimation when stand volume is estimated with only the similarity. To overcome such problems, the addition of explainable indices like leaves density or the development of coefficients for complementing non-overlapped part are required in the future study.

6. Conclusion

We estimated the stand volume using the similarity of nonparametric distribution of LiDAR data in forest area of South Korea whereas the height distributional parameters, such as percentile, of LiDAR data reflected on- and in-canopy in a stand have the relationship with stand volume in previous research. As a result, the similarity of sample and test sites was relatively high, but the accuracy of estimated stand volume was not high to be $R^2=0.51$ and $RMSE=34.96$ m³/ha. That is attributed to the non-overlapped area in distributional functions between sample and test sites. Even if the pattern of height distribution of canopy returns is similar, the leaves density can influence the number of LiDAR pulses reflection and cause non-overlapped part in distributional functions. Therefore, it is needed to derive leaves density indices or complementary coefficients in the future research.

Acknowledgements

This study was carried out with the support of “Forest Science and Technology Projects (Project No. S120911L010130)” provided by Korea Forest Service.

References

- Chen, Q., Gong, P., Baldocchi, D. and Tian, Y.Q., 2007. Estimating basal area and stem volume for individual trees from LiDAR data. *Photogrammetry Engineering and Remote Sensing*, 73, 1355–365.
- Coops, N.C., Hilker, T., Wulder, M.A., St-Onge, B., Newnham, G., Siggins, A. and Trofymow, J.A., 2007. Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. *Trees*, 21, 295–310.
- Drake, J.B. and Weishampel, J.F., 2000. Multifractal analysis of canopy height measures in a longleaf pine savanna. *Forest Ecology and Management*, 128, 121–127.
- Drake, J.B., Dubayah, R.O., Clark, D.B., Knox, R.G., Blair, J.B., Hofton, M.A., Chazdon, R.L., Weishampel, J.F. and Prince, S., 2002. Estimation of tropical forest structural characteristics using large-footprint LiDAR. *Remote Sensing of Environment*, 79, 305–319.

- Gobakken, T. and Næsset, E., 2005, Weibull and percentile models for lidar-based estimation of basal area distribution. *Scandinavian Journal of Forest Research*, 20, 490–502.
- Harding, D.J., Lefsky, M.A., Parker, G.G. and Blair, J.B., 2001. Laser altimeter canopy height profiles: Methods and validation for deciduous, broadleaf forests. *Remote Sensing of Environment*, 76, 283-297.
- Jaskierniak, D., Lane, P.N.J., Robinson, A. and Lucieer, A. 2011. Extracting LiDAR indices to characterise multilayered forest structure using mixture distribution functions. *Remote Sensing of Environment*, 115, 573–585.
- Kim, J.M., Park, K.S., Baek, E.S., Song, Y.K., Ahn, C.Y., Lee, K.H., Lee, M.H., Lee, S.Y., Lee, S.H., Jeong, Y.K., Jeong, J.H., Joo, R.W., Choi, K., Choi, M.S., Song, J.H., Kim, J.W., Kim, J.Y., Park, M.S., Song, T.Y., Kim, J.H., Yang, S.I., Jang, W.H. and Jang, C.S., 2000. *Forest & Forestry Technique*. Korea Forest Service. Daejeon, South Korea.
- Kwak, D.A., Lee, W.K., Lee, J.H., Biging, G.S. and Gong, P., 2007. Detection of individual trees and estimation of tree height using LiDAR data. *Journal of Forest Research*, 12, 425–34.
- Kwak, D.A., Lee, W.K., Cho, H.K., Lee, S.H., Son, Y., Kafatos, M. and Kim, S.R., 2010. Estimating stem volume and biomass of *Pinus koraiensis* using LiDAR data. *Journal of Plant Research*, 123, 421-432.
- Lefsky M.A, Cohen W.B, Acker S.A, Parker G.G, Spies T.A and Harding D.J., 1999. LiDAR remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment*, 70, 339–361.
- Maclean, G.A. and Krabill, W.B., 1986. Gross-merchantable timber volume estimation using airborne lidar system. *Canadian Journal of Remote Sensing*, 12, 7–18.
- Magnussen, S. and Boudewyn, P., 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research*, 28, 1016–1031.
- Magnussen, S., Eggermont, P. and Lariccia, V. N., 1999. Recovering tree heights from airborne laser scanner data. *Forest Science*, 45, 407–422.
- Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L. and Hendrix, C.J., 2000. Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetry Engineering and Remote Sensing*, 66, 1367–1371.
- Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80, 88–99.
- Nelson, R.F., Krabill, W. and Tonelli, J., 1988. Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment*, 24, 247-267.
- Parker, G.G., Lefsky, M.A. and Harding, D.J., 2001. Light transmittance in forest canopies determined using airborne laser altimetry and in-canopy quantum measurements. *Remote Sensing of Environment*, 76, 298–309.

- Peterson, C.J., 2000. Catastrophic wind damage to North American forests and the potential impact of climate change. *The Science of The Total Environment*, 262, 287–311.
- Popescu, S.C. and Wynne, R.H., 2004. Seeing the trees in the forest: using LiDAR and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogrammetry Engineering and Remote Sensing*, 70, 589–604.
- Son, Y.M., Lee, K.H., Park, Y.K., Kim, R.H. and Kwon, S.D., 2008. Management plan for absorption and emission of green house gas in part of forest, Seoul of South Korea: Korea Forest Research Institute.
- Tucker, C.J., Slayback, D.A., Pinzon, J.E., Los, S.O., Myneni, R.B. and Taylor, M.G., 2001. Higher northern latitude normalized difference vegetation index and growing season trends from 1982 to 1999. *International Journal of Biometeorology*, 45, 184–190.
- van Aardt, J A.N., Wynne, R.H. and Oderwald, R.G., 2006. Forest volume and biomass estimation using small-footprint lidar distributional parameters on a per-segment basis. *Forest Science*, 52, 636–649.
- van Aardt, J.A.N., Wynne, R.H. and Scrivani, J.A., 2008. Lidar-based mapping of forest volume and biomass by taxonomic group using structurally homogeneous segments. *Photogrammetric Engineering & Remote Sensing*, 74,1033–1044.
- Zhou, L.M., Tucker, C.J., Kaufmann, R.K., Slayback, D.A, Shabanov, N.V. and Myneni, R.B., 2001. Variations in northern vegetation activity inferred from satellite data of vegetation index during 1981 to 1999. *Journal of Geophysical Research*, 106, 20069–20083.

A method for linking TLS- and ALS-derived trees

Andreas Fritz¹, Holger Weinacker¹ & Barbara Koch¹

¹University of Freiburg, Department of Remote Sensing and Landscape Information Systems, andreas.fritz@felis.uni-freiburg.de

Abstract

Within the past decade progress towards automatic recognition of individual trees and their parameters was made in both TLS and ALS- data based algorithms. In this paper we present an approach to combine single trees derived from ALS and TLS-data in order to gain a higher level of information. Therefore, two data sets are used: 1. a set of 3D-stemfiles generated by the algorithm described in Bienert *et al.* 2007 and 2. a set of detected single trees for the corresponding area of the data set 1 based on the algorithm described in Gupta *et al.* 2010. The 3D-stemfiles include position, information regarding sweep and diameter in 10[cm] height intervals. The ALS-tree description covers the position, maximum crown diameter and length as well as tree top height. This information is used for a hierarchic approach of linking ALS and TLS-derived trees based on three different initial matching algorithms. The estimated position error is taken into account to generate an initial list of matching candidates. The 2D-distance based initial linking method linked 41% of the TLS-trees. It was found that 3D-estimation of the tree top based on sweep information of the TLS-trees led only to minimal more imputations than the 2D-approach. A possible reason is seen in the linear models chosen, which do not reflect the tree shape invariably. Future work focuses on the integration of species information and the quantification of false linkage, which could not be evaluated within this study.

Keywords: terrestrial laser scanning, airborne laser scanning, tree linking, 3D-line fitting

1. Introduction

In recent years, object recognition based on Lidar generated point clouds has been an important topic in both science and industry. Airborne laser scanner (ALS) data in different point density levels is available throughout major countries and updated in useful time intervals. In the field of forestry applications, research related to biomass estimation and single tree detection are among the most regarded publications. High potential has been demonstrated for the retrieval of relevant information regarding economic and ecological parameters of forests from ALS data. Næsset 2002 finds a good estimation of dominant height, mean height, mean diameter, basal area and volume of stands in Scandinavia. In Straub *et al.* 2009, methods for single tree based total stem volume predictions in different forest conditions in southern Germany are compared. In the latest publication of Gupta *et al.* 2010, singletree parameters predicted in the same stands as the previous mentioned study include tree top height and position, crown length and maximum crown diameter. A comparison of different algorithms regarding their performance under different forest conditions can be found in Vauhkonen *et al.* 2010.

Particularly singletree delineation and the related description of quality parameters are of high importance for the diverse European wood industry. Precise economically relevant description of standing trees is among the top applications of terrestrial laser scanning in forests. Liang *et al.* 2008 detect tree trunks in a distance up to 60[m] with a single view scan. Bienert *et al.* 2007 present a commercialized system to detect trees in one scan mode. For each detected tree, a profile fitting in a height interval of 10[cm] was conducted, thereby quantifying diameter, volume and shape of the entire tree.

Combining the features extracted from both ALS and TLS data promises a precise and

comprehensive description of the ecosystem forest and its economic value. Few approaches are known to the author dealing with registration of ALS and TLS generated objects in wooden environments. Olofsson *et al.* 2008 present a method of linking field data to ALS derived trees. The linking is based on dbh, height and distances between ALS- and field-surveyed trees. Ground truth tree data generated in StandforD stemfiles during felling with harvesters is used in Holmgren *et al.* 2010. A 10[m]-radius sample plot size is used and the linking is based on ALS generated tree attributes as crown polygon and height.

In this paper, we present an approach of combining the TLS and ALS datasets on object (tree) level. We use the results of proven algorithms developed at the University Freiburg, Department of Remote Sensing and Landscape Information Systems (Gupta *et al.* 2010) and University Dresden, licensed by TreeMetrics Ltd. (Bienert *et al.* 2007).

2. Material

2.1 Study area

The study area is located in south-west Germany (49.0N 8.4E) in the municipal forest of Karlsruhe. The extent of the test site is 2.1x4.4[km²] and predominantly includes pine, beech and spruce. Within the area, 42 permanent inventory plots were selected to cover typical forest conditions for this site. The plots include scots pine (*Pinus sylvestris*), sessile oak (*Quercus petraea*), red oak (*Quercus rubra*) and European beech (*Fagus sylvatica*).

2.2 TLS data

TLS data acquisition was conducted on the centre point of all plots by TreeMetrics Ltd. in November 2010. The scanning device was a Faro Imager 5006i. Data was processed with the algorithm described in Bienert *et al.* 2007. In a pre-processing step, all returns not within a radius of 15[m] of the plot centre were discarded. The settings were made for high precision and accuracy, which leads to a high omission error. Manual inspection did not reveal any false tree detection, but a high number of undetected stems can be observed. The log length of the detected trees is underestimated due to the strict parameter settings. The centre coordinates of the plots were measured by a surveying company and have a location error of < 5[cm]. The coordinates for each detected stem were calculated with distance and azimuth to the centre. The processing result is a list for each plot containing the detected trees in form of coordinates, diameter and volume in a height interval of 10[cm]. In 42 plots, 831 trees were detected (min=1, max=21, mean=8.9).

2.3 ALS data

In November 2009, ALS data was acquired with a Riegl LMS-Q560 Scanner. The point density is ca. 24[pt/m²]. The method used for detecting and delineating single trees is described in Gupta *et al.* 2010. In addition to the top tree position, two more parameters are calculated for each detected tree: maximum crown diameter and crown length. Within a 20[m] radius of the TLS-scanner position, 1277 trees were detected.

3. Methods

3.1 Data preparation

To reduce false imputation, data cleaning was performed on the TLS-tree set. Here, all small neighbouring trees that (a) are within 1.5[m] and (b) have a dbh difference > 10[cm] were removed. This is to dismiss trees that, although they are detected by TLS, most likely are not

detectable by ALS (i.e. because they are in the understory), as being linking candidates.



Figure 1 view through a TLS-scan with an understory tree covered by a dominant tree. After cleaning the smaller tree has been removed.

Figure 1 shows an example of two trees with a distance to each other of 0.7[m] and dbh-difference of 24.5[cm]. The smaller tree on the left was discarded according to the cleaning condition.

3.2 Initial linking

Firstly, the 2D Cartesian distance for each TLS and its neighbouring ALS tree is calculated when the condition

$$Hd_{TA} < |e| \quad (1)$$

is fulfilled. Where Hd_{TA} refers to the horizontal Cartesian distance between TLS and ALS-trees and e to the combined position error of both datasets.

$$e = \sqrt{e_{ALS}^2 + e_{TLS}^2} \quad (2)$$

e_{ALS} and e_{TLS} are the errors from the two detection algorithms. Each TLS-tree is then associated to potential ALS-trees. A n:m relation is possible, which potentially associates multiple TLS trees to multiple ALS-trees.

3.3 Advanced initial linking

On the test site, a distinct leaning can be observed, particularly pine trees. This implicates that the crown is often not perpendicular to the tree position at dbh. Horizontal distances up to 7[m] between tree top and dbh position were observed in a reference data set. Therefore, we considered additional information provided by the TLS-trees. Those trees have a certain length

which is not related to the actual tree height but more to branchiness, occlusion and parameter settings of the detection algorithm; however, the sweep of the stem is reflected very well. This sweep is used to estimate the potential position of the tree top by fitting a straight line through each 3D-coordinate tuple of the tree. For getting meaningful estimations, this approach was applied only to stems which are longer than 5[m]. The line fitting algorithm is explained in detail in Drixler 1993. The algorithm is based on quadrics and utilizes the concept of eigenvalues. Fitting the 3D-line follows four steps (Drixler 1993 p.47ff.):

1. The coordinates of the TLS-derived tree are reduced with their centre of gravity and the result is stored in matrix A^*
2. Three eigenvalues and eigenvectors of the matrix product of $A^{*T}A^*$ are computed
3. The corresponding eigenvector of the maximum eigenvalue represents the direction vector of the fitted line.
4. The geometric gravity point of the TLS tree points is used as the origin point.
5. The variance-factor σ is calculated with the two smaller eigenvalues as $\check{\sigma} = \sqrt{\frac{\lambda_1 + \lambda_2}{N-3}}$

This approach returns a 3D-straight line defined by 6 parameters (x_{op} , y_{op} , z_{op} , x_{vect} , y_{vect} , z_{vect}). The resolution of the x - y -changes along the stem is in the range of millimetres. This means the points along the stem may reflect small buckles and curvature at the stem bottom what can substantially influence the line fitting. To overcome this uncertainty, an iterative approach was developed that begins at the stem top and moves downwards. Initially, the topmost 10 points are used for an initial line fit. For each iteration, one point is added to the point set used by the fitting algorithm. The change of the variance-factor σ per step is observed and when σ becomes greater than 0.009, the process ceases. The threshold of 0.009 was found empirically.

The fitted line is then elongated to the height of the highest detected tree in the ALS-data. The 3D minimum distance is measured between the fitted line and the ALS trees and if distance values fulfil the criteria described in formula (1), the pair is put on the list of potential matching candidates.

A third list is generated with the help of 3D-boxes. The estimated tree-top and the geometric gravity point, both derived from the line fit approach specify a 3D-box. The parameters maximum crown diameter and tree top height define another 3D-box. When the two shapes intersect the trees are put on the list of possible matching candidates.

3.3 Final linking procedure

The list of potential matches contains all pairs of TLS- and ALS-trees in which, for example, one TLS tree is linked to several ALS-trees and vice versa. Figure 2 outlines the different steps of the linking procedure. The algorithm begins with a random TLS-tree and analyses its neighbourhood situation. It checks how many neighbouring ALS- and TLS-trees are present. The associated neighbours are again checked for their neighbours and so on. This process stops if there are no more partners on the list. The network is stored in a temporary list, which is the starting point for further analyses.

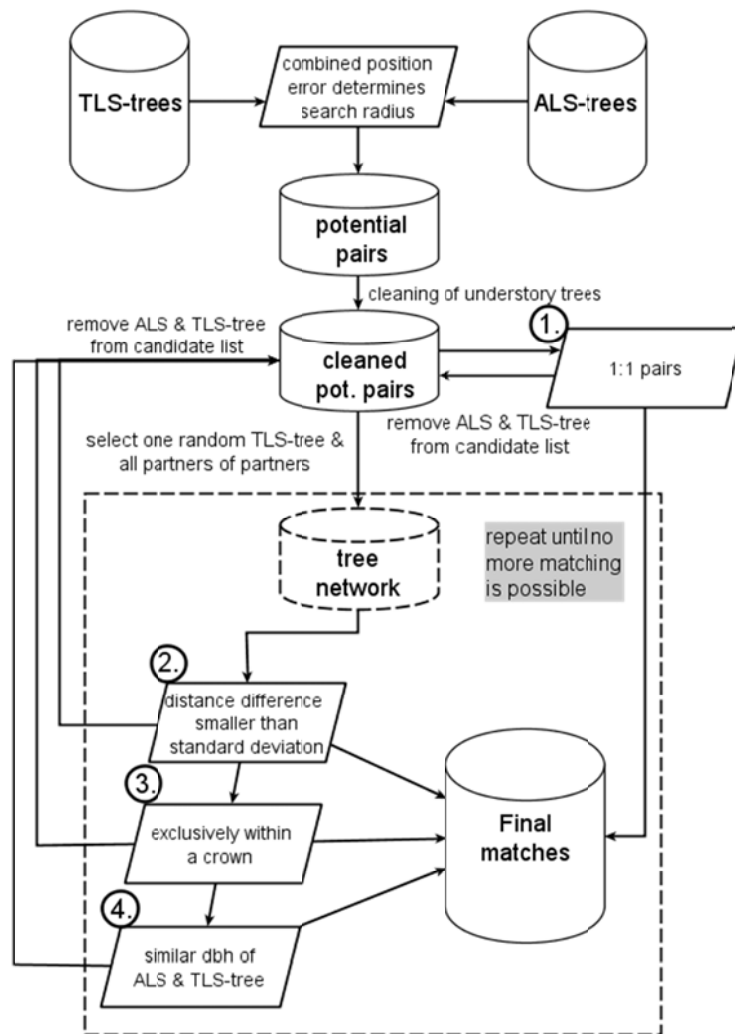


Figure 2 Schematic flow of the steps for linking. Numbers 1-4 indicate the order of decision hierarchy. The actions within the dashed box are repeated until no more TLS-trees are assignable.

3.3.1 Analysing tree networks

In case of a real 1:1 association, the algorithm stops and declares the ALS-TLS-tree pair as successfully linked. A real 1:1 situation exists when the ALS- and TLS-tree do not have further potential partners within the distance of e .

For situations where the initially randomly picked TLS-tree has a network of partners, the horizontal distance, the dbh, the maximum crown diameter of the neighbouring trees are of relevance. The linking is realised according to a decision tree (Figure 2). The algorithm starts with another 1:1 check, since through the following steps new 1:1 situations can arise and goes on with distance comparison.

The distance comparison is defined as those trees will be linked with the tree that is the shortest distance away and if the difference to the second shortest is less than the standard deviation of all distances in the initial list.

When no significant distance difference can be observed, the maximum crown diameter is taken into account. If a 1:1 situation is achieved by checking whether a TLS-tree is exclusively within a buffer defined by the maximum crown-diameter, the ALS- and TLS-tree pair is declared as

successfully linked. If there is still no match possible, a general estimation of the dbh of the ALS tree is made by a logarithmic regression equation based on site specific parameters and literature review. Species information is not available and, therefore, not taken into account. The procedure compares each pair concerning the dbh estimated and measured. The standard deviation of all TLS measured dbh values is taken as the threshold for imputation. When the difference between the measured and estimated dbh is greater than the standard deviation, the pair is discarded from the list. In the case that only one pair meets the condition, it is matched. If more than one combination is possible, then no match is performed.

4. Results

The distance threshold e was calculated with the following values:

1. Since e_{ALS} was not specified in Gupta *et al.* 2010, the author checked the position accuracy manually and concluded that the horizontal distance between the reference tree top and found tree top was up to $\pm 3.2[m]$.
2. In Bienert *et al.* 2007, the position error is not mentioned, but analyses regarding the error of the dbh estimation were performed. Seeing as the diameter is through the stem centre associated with the tree position those result were used as input for e_{TLS} . Bienert *et al.* 2007 found a maximum error for the tree diameter of $\pm 19.6[cm]$ for Sitka spruce.

A test series with $e = 1.5 \dots 7[m]$ applied to the 2D-distance based candidate list showed a strong correlation ($R^2 = 0.98$) between e and the number of matched pairs.

4.1 2D-distance based candidate list

The 2D-distance based initial linking leads to an initial list of 880 possible combinations. After the cleaning of understory trees, 858 combinations remain. Clearing the 1:1 situations leads to matches of 185 TLS-trees to one ALS tree and 673 combinations need further investigations. 6 matches were made through the second 1:1-check. The query regarding the distance difference matched 106 TLS-trees. The maximum crown diameter was responsible for 34 matches and the dbh comparison led to 24 imputations. In total 355 (41%) out of 831 TLS-trees were linked to an ALS-tree.

4.2 3D-line fit based candidate list

After clearing small trees likely covered by a big tree, the initial list of potential pairs within of 3.2[m] minimum 3D-distance of the fitted line to an ALS-tree-top contains 619 combinations (previously 597). In total, 284 (48%) TLS-trees were matched with an ALS-tree. The initial 1:1 situation was dominating the imputation with 179 cases. The distance comparison led to 56 matches and the crown diameter was responsible for 25 links. Analysing the (estimated) dbh of the two tree sets linked 21 TLS-trees. Only 3 pairs were matched through a recheck for 1:1 situations.

4.3 3D-box intersection based candidate list

The 3D-box intersection returns approximately double the amount of potential pairs (1224) compared to the 3D-line-fit based method. After clearing understory trees, 1202 remain. Compared to the other approaches, the number of initially matched pairs is relatively low (57) whereas matching through distance difference (153) dominates. Imputation through crown inclusion and dbh-comparison was responsible for 45 respective 22 pairs. In total 294 (24%) TLS-trees were linked to an ALS-tree.



Figure 3 View through TLS-point cloud (green) with a depth of 1.5[m]. The blue points represent the ALS-trees found by Gupta *et al.* 2010

4.4. Matching reliability

The three different results were analysed in respect of differences in linked tree pairs. 181 pairs were matched with all three initial linking methods. With both the 2D-distance based and the 3D-line fitting based candidate list, 227 pairs were matched. 3D-boxes and 2D-distance based initial list generation led to 216 pairs which appear after the algorithm in both final imputations lists. When comparing the two 3D-approaches, 221 pairs appear in both candidate lists. In total, 454 different TLS-trees could be matched when running all three initial list generators separately and combining the results.

5. Discussion & Conclusion

The results show that depending on the initial linking method approximately one third of the TLS-trees could be linked to an ALS-tree. One has to consider the fact that in our case detection algorithms for terrestrial laser scanners include trees with a dbh of 7[cm] and higher. This implies a substantial number of detected trees that do not take part in the crown layer and are thus barely detectable in airborne laser scanner data. We exclude potentially fully covered TLS-trees that are close to another TLS-tree to reduce that effect, however this does not attack the problem of isolated standing trees that are not dominant enough to appear in the crown layer visible to ALS-scanners. To solve this issue, a more reliable dbh-estimation is essential. In general, the fact that the species information is missing thus far is seen as a major obstacle for a better linking. With known species in either TLS or ALS-tree data, the dbh-estimation could be of a higher accuracy. Recent work described in Heinzl and Koch 2011 show promising results for tree species detection in mixed stands. It is planned to include these findings in the future.

The three approaches of making an initial candidate list do not necessarily produce the same pairs. Only 181 matches were identical in all three approaches. A possible reason for that is the consideration of the third dimension. It was observed that, even with a low variance factor, the fitted line did not always reflect the growth direction of the stem reliably. In these cases, the projected tree top position is very likely incorrect and therefore a wrong ALS-tree gets linked. This problem does affect the 3D-boxes approach as well. The fitting approach returns an

over-parameterized definition of a line since a straight line in 3D is defined by four independent parameters $(\alpha, \beta, \gamma, d)$. The method can be questioned when taking this into consideration. However, we believe that the negative effect of the fitting algorithm itself is not causal for wrong tree top estimation. The high resolution of 2-dimensional direction changes is seen as the major reason for the failing of the fitting. Since the underlying model is a straight line, this leads to an insufficient result in cases where the tree has a distinct sweep along the complete stem. For future work, better indicators need to be found to decide whether or not a line fitting is appropriate or not.

A major problem is to quantify false linkage, since there is no information that can prove correct linking. A solution to that problematic could be to use the TLS-point cloud to assess the accuracy of the tree top position and maximum crown diameter. First visual inspection showed that with good registration, useful accuracy estimation with respect to linking is possible. A cut through the point TLS-point cloud with the detected ALS-trees on top is shown in Figure 3.

The strong linear correlation between the numbers of matched pairs and e is not surprising since the number of potential pairs increases dramatically. The quality of the result is certainly problematic because more unrealistic matches are made when the maximum distance is increased.

In future research the effect of the linking on biomass and forest parameters estimation is investigated.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme *FP7/2007-2013* under *grant agreement* n° 245136. I would particularly like to thank Garret Mullooly for scanning and processing the TLS-data.

References

- Bienert, A.; Scheller, S.; Keane, E.; Mohan, F. and Nugent, C., 2007. Tree detection and diameter estimations by analysis of forest terrestrial laserscanner point clouds. *Proceedings of ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*, Espoo, Finland, 50–55
- Drixler, E., 1993. Analyse der Form und Lage von Objekten im Raum. *Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften*, Reihe C, Heft Nr. 409
- Gupta, S., Koch, B. and Weinacker, H. 2010. Tree species detection using full waveform lidar data in a complex forest. *ISPRS TC VII Symposium – 100 Years ISPRS, 2010, XXXVIII*, 249-254
- Heinzel, J. and Koch, B., 2011. Exploring full-waveform LiDAR parameters for tree species classification. *International Journal of Applied Earth Observation and Geoinformation*, 13, 152 - 160
- Holmgren, J., Barth, A., Larsson, H. and Olsson, H., 2010. Prediction of stem attributes by combining airborne laser scanning and measurements from harvesting machinery In: *Silvilaser 2010, the 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems*. September 14-17, 2010, Freiburg, Germany.

- Liang, X., Litkey, P.; Hyypä, J., Kukko, A., Kaartinen, H. and Holopainen, M., 2008. Plot-level trunk detection and reconstruction using one-scan-mode terrestrial laser scanning data *International Workshop on Earth Observation and Remote Sensing Applications*, June 30 2008-July 2 2008
- Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80, 88 - 99
- Olofsson, K., Lindberg, E. and Holmgren, J., 2008. A method for linking field-surveyed and aerial-detected single trees using cross correlation of position images and the optimization of weighted tree list graphs. In Hill, R., Rosette, J. and Suárez, J. (Eds). *SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment and inventory*. 95-104.
- Straub, C., Weinacker, H. and Koch, B. 2010. A comparison of different methods for forest resource estimation using information from airborne laser scanning and CIR orthophotos *European Journal of Forest Research*, 129, 1069-1080
- Vauhkonen, J., Ene, L., Gupta, S., Heinzl, J., Holmgren, J., Pitkänen, J., Solberg, S., Wang, Y., Weinacker, H., Hauglin, K.M., Lien, V., Packalén, P., Gobakken, T., Koch, B., Næsset, E., Tokola, T. and Maltamo, M., 2010. Comparative testing of single-tree detection algorithms. In: *Silvilaser 2010, the 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems*. September 14-17, 2010, Freiburg, Germany.

Reducing extrapolation bias of area-based k-nearest neighbor predictions by using individual tree crown approaches in areas with high density airborne laser scanning data

Johannes Breidenbach^{1,2}, Erik Næsset² & Terje Gobakken²

¹Norwegian Forest and Landscape Institute, job@skogoglandskap.no

²Norwegian University of Life Sciences

K-nearest neighbor (kNN) approaches are popular statistical methods for predicting forest attributes in airborne laser scanning (ALS) based inventories. Their main upsides are the simplicity to predict multivariate response variables and their freedom of distributional assumptions on the conditional response. One of their largest draw-backs is that predictions outside the range of the reference data inherently result in an under- or overestimation. This property of kNN approaches is known as extrapolation bias and aggravates with an increasing number of neighbors (k) used for the prediction.

This study presents one possibility to reduce extrapolation biases of predictions based on the area-based approach (ABA) by using individual tree crown (ITC) approaches within those specific areas of a low density ALS acquisition where the point density might be sufficiently high for using ITC methods. In the proposed strategy, additional (or artificial) reference plots augmented field measured plots. Artificial plots were created by applying ITC segmentation to a canopy height model derived from high density ALS data. The response variable biomass per hectare was predicted for every segment following a semi-ITC approach. The segment predictions were aggregated on the artificial plot level. The artificial plots were then treated in the same way as the original reference data to make predictions in areas with low density ALS data based on the ABA. It was hereby assumed that the predicted plot level response on the artificial plots is equivalent with the observed plot level response on the original reference data.

The data consisted of 110 reference plots with a smaller data range than the 201 independent validation plots. Considerable extrapolation bias was visible if only the reference plots were used for the prediction. Almost no extrapolation bias was found if the prediction was based on reference plots augmented by artificial plots. The root mean squared error (RMSE) of the biomass predictions based on the reference plots was 39.1%. The RMSE reduced to 29.8% if the reference plots were augmented by artificial plots.

Keywords: Nonparametric regression, kNN, MSN, accuracy, precision, lidar, ALS

1. Introduction

K-nearest neighbor (kNN) approaches are popular statistical methods for predicting forest attributes in airborne laser scanning (ALS) based inventories (e.g., Breidenbach *et al.* 2010b; Hudak *et al.* 2009; Packalén and Maltamo 2006). Their main upsides are the simplicity to predict multivariate response variables and their freedom of distributional assumptions on the conditional response. One of their largest draw-backs is that predictions outside the range of the reference data inherently result in an under- or overestimation (McRoberts 2009). This property of kNN approaches is known as extrapolation bias (Magnussen *et al.* 2010) and aggravates with an increasing number of neighbors (k) used for the prediction. Magnussen *et al.* (2010) proposed a general, model assisted method to dampen extrapolation biases.

This study presents one possibility to reduce extrapolation biases of predictions based on the area-based approach (ABA) (Næsset 2002) by using individual tree crown (ITC) approaches within those specific areas of a low density ALS acquisition where the point density might be

sufficiently high for using ITC methods. In the proposed strategy, additional (or artificial) reference plots augmented field measured plots. Artificial plots were created by applying ITC segmentation to a canopy height model derived from high density ALS data. The response variable biomass per hectare was predicted for every segment following a semi-ITC approach (Breidenbach et al. 2010a). The segment predictions were aggregated on the artificial plot level which resulted in the “observation” of the response at the artificial plot. The artificial plots were then treated in the same way as the original reference data to make predictions in areas with low density ALS data based on the ABA.

Even if low density ALS data are acquired, areas of high density ALS data exist where additional flight lines are flown perpendicular to the main flying direction for the purpose of calibrating the stripes against each other. Lower flying altitudes may even be considered for these additional flight lines in order to increase point density (Næsset *et al.* 2006). This makes the proposed approach easily applicable under operational settings. Since the ALS campaign and the field work were carried out independently, an additional ALS campaign was necessary to acquire the high resolution ALS data for some flight strips in this study.

2. Material and methods

2.1 Study area and field data

The study area was located in the municipality of Aurskog-Høland, in southeast Norway (Figure 1). The forest in the area is dominated by Scots pine (*Pinus sylvestris*) and Norway spruce (*Picea abies*). The topography is smooth with heights above sea level between 120 and 390 m.

The field data consisted of 110 reference plots and 201 independent validation plots. On these field plots, diameter at breast height (dbh) and species were recorded for all trees. Tree height was measured for a subsample of trees on every plot to predict the heights of the unmeasured trees. Single tree biomass was derived using species-specific biomass functions based on dbh and height (Marklund 1988).

The reference plots were sectors (i.e., quarters and halves) of 40 large sample plots with an area of 500 and 1000 m², respectively. Following a purposive sampling design, the 40 large sample plots were located under strips of high density ALS data (Figure 1, lower right hand side). The field work in which also tree coordinates were recorded was carried out in 2008. See Breidenbach *et al.* (2010a) for more details of the reference plot setup. The original number of sectors was 152, each with an area of 250 m². However, in order to better illustrate the strength of using artificial plots, only sectors with a measured biomass between 50 and 170 Mg ha⁻¹ were selected. Thus these selected sectors constituted a sample of 110 reference plots. The reference plots were used to create the artificial plots and for fitting the statistical model using the ABA.

The circular validation plots had an area of 200 m² and stemmed from an operational forest inventory that took place in 2006. The sample plots were distributed according to a stratified systematic design (Figure 1, upper right hand side). Table 1 summarizes characteristics of the reference and validation plots.

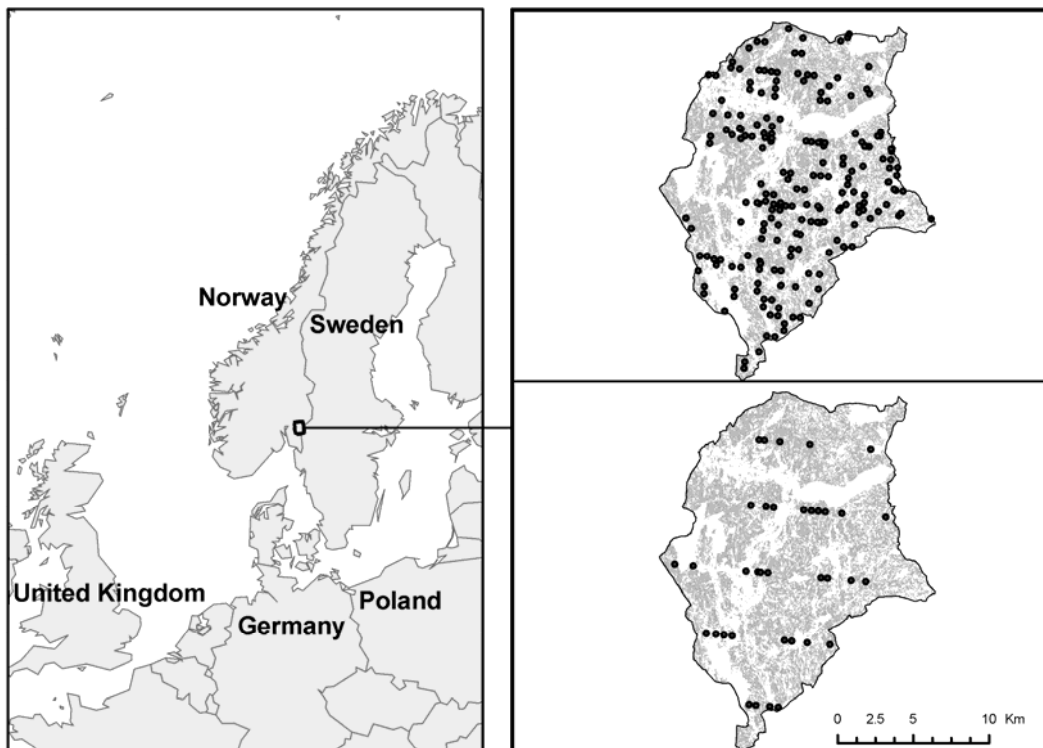


Figure 1: Location of the study area within Norway and positions of the 201 validation plots (upper right hand side) as well as the 40 large sample plots containing the sectors serving as reference plots (lower right hand side).

Table 1: Characteristics of measured biomass on the sample plots (Mg ha^{-1})

Plot type	Minimum	Mean	Maximum	Standard deviation
Reference plots	52.09	92.69	168.30	33.08
Validation plots	18.47	111.30	314.30	62.64

2.2 Remote sensing data

The whole municipality of Aurskog-Høland was covered with low-density ALS data in summer 2005 in order to provide auxiliary information for an operational forest inventory. The Optech ALTM 3100 sensor was operated from a fixed-wing aircraft with a flying height above ground of approximately 1850 m and a half scanning angle of 15 degrees. The pulse repetition and mirror frequencies were 50 kHz and 31 Hz, respectively. Only the first and last returns per pulse were recorded. While the main flight direction was north-south oriented, two flight strips were flown almost perpendicular to the main block in order to calibrate the other strips against each other (Figure 2). The pulse density under these strips was frequently above 1.5 m^{-2} whereas the pulse density in general was around 0.7 m^{-2} .

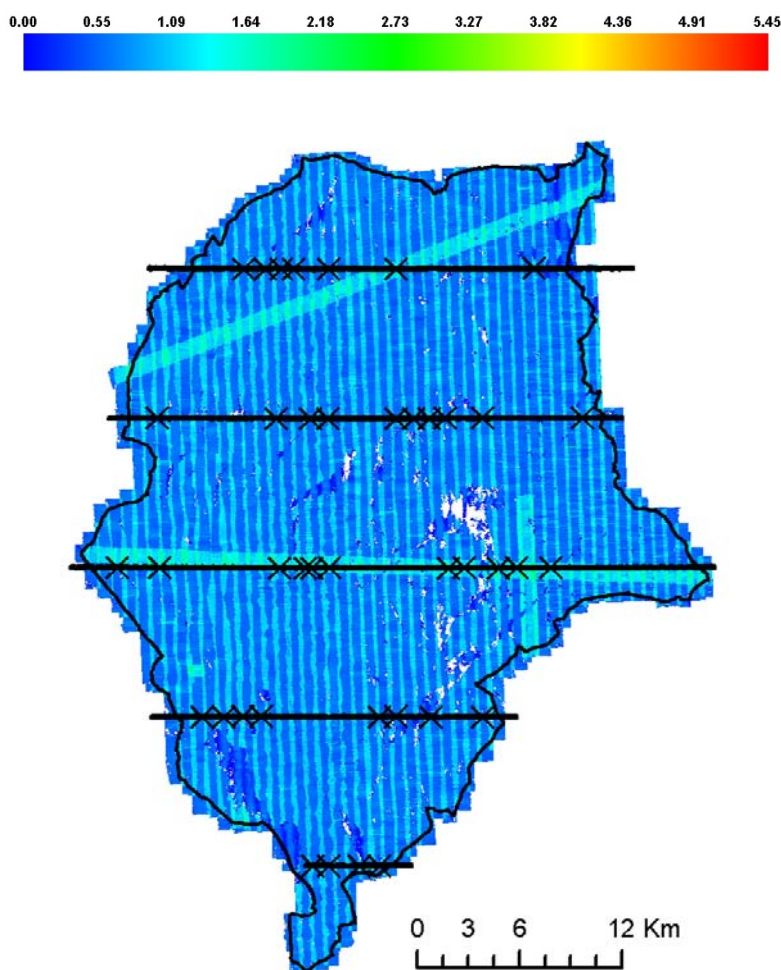


Figure 2: Returns per square meter of low density first return ALS data (background), extent of high density flight strips (horizontal black lines) and locations of large sample plots (crosses).

High density ALS data were acquired along five non-overlapping flight strips covering the 40 large sample plots in summer 2006 (Figure 2). The flight direction was east-west oriented and the acquisition settings resulted in flight strips with a swath of 150 m in south-north direction. Up to four returns per pulse and the intensity of the reflected signal were recorded for approximately seven pulses per square meter. The Optech ALTM 3100 sensor was also used for this data acquisition but was operated from a flying height above ground of approximately 800 m and a half scanning angle of 5 degrees. The pulse repetition and mirror frequencies were 100 kHz and 70 Hz, respectively. Both ALS data sets were processed by the data vendor and delivered in UTM coordinates with ellipsoidal heights.

2.3 Creation of artificial plots and area-based predictions

The proposed approach can be subdivided into two phases. In the first phase, artificial plots are generated in the areas with high density ALS data. The second phase is basically the well-known ABA (Næsset 2002). However, the artificial plots are used in addition to the reference sectors to increase the amount of training data.

In the first phase, the positions of the artificial plots were established such that they were

aligned along the flight lines approximately in the middle of the high resolution ALS strips. An artificial plot center was established every 50 m along the flight lines. It should be noted that the numbers given here were chosen for technical and practical reasons. In theory, artificial plots could be anywhere within the area of high resolution ALS data.

Using an inverse distance weighted algorithm and four neighbors, digital surface models (DSMs) with 0.5 m edge length were created from the high density first return ALS data for areas including a 5 m buffer around the artificial plot center. Tree crowns or small clusters of trees were segmented in the inverted DSMs using a watershed algorithm. Segments were assumed to belong to an artificial plot, if their centroid was within a radius of 8.92 m of the artificial plot center. The mean and coefficient of variation of ALS height and intensity, proportion of first returns, proportion of returns lower than 30% of the maximum height (density3) of the ALS returns within the segments as well as the segment area were derived as predictor variables for the semi-ITC approach.

The same strategy as for the artificial plots was followed to segment tree crowns on the reference plots. The segments were intersected with the tree coordinates in order to link the field measurements with the remote sensing data. Due to omission and commission errors of the segmentation algorithm, there can be no, one, or several trees within a segment.

A nonparametric kNN method based on the normalized Euclidean distance (Crookston and Finley 2008) was used to predict the tree properties of crown segments on the artificial plots. Biomass associated with the closest segment on a reference plot was imputed to the target segments on the artificial plots.

In the second phase, the predicted biomass for the segments was aggregated on the artificial plot level. In the case of the reference plots, the measured tree biomass was aggregated. Both data sets were merged to one large table used as reference in the ABA. Targets in terms of kNN were the 201 independent validation plots. The following seven metrics served as predictor variables in the ABA and were derived from the height distribution of the low resolution ALS data: Maximum, mean, standard deviation, coefficient of variation, interquartile distance, kurtosis, first quartile. The metrics were derived for the artificial, reference and validation plots.

As for generating the artificial plots, a kNN approach was used in the ABA. However, the distance metric was based on canonical correlation. With $k=1$, the method is known as most similar neighbour inference (MSN) (Moeur and Stage 1995). In MSN, the distance between observations, say a target with index t and a reference with index r , is calculated by multiplying the differences in the explanatory variables (X), $d_{tr} = X_t - X_r$, with a weighting matrix. The weighting matrix (W) is derived using canonical correlation analysis that maximizes the correlation between response (Y) and explanatory variables by linear transformation $U_k = \alpha_k Y$ and $V_k = \gamma_k X$; $k = 1, \dots, s$. The variables α_k and γ_k are the canonical coefficients of the response and the explanatory variables respectively. The weighing matrix is then given by $W = \Gamma \Lambda^2 \Gamma'$ where Γ = the matrix of all γ_k and Λ^2 = diagonal matrix of the squared λ_k . This results in the distance function $D_{tr}^2 = d_{tr} W d_{tr}'$.

The goodness of fit of the MSN models was analysed by plotting observed versus predicted values on the validation plots against each other and by root mean squared errors (RMSE) and mean residuals.

3. Results

The range of the predicted values based on the reference plots alone was between 60 and

150 Mg ha⁻¹. Small observed values were overestimated and large observed values were underestimated (Figure 3, left hand side).

Almost no under- and overestimation of small and large observed values was recognized if artificial plots were used to augment the reference plots (Figure 3, right hand side). The use of artificial plots also resulted in considerably smaller root mean squared errors and mean residuals (Table 2).

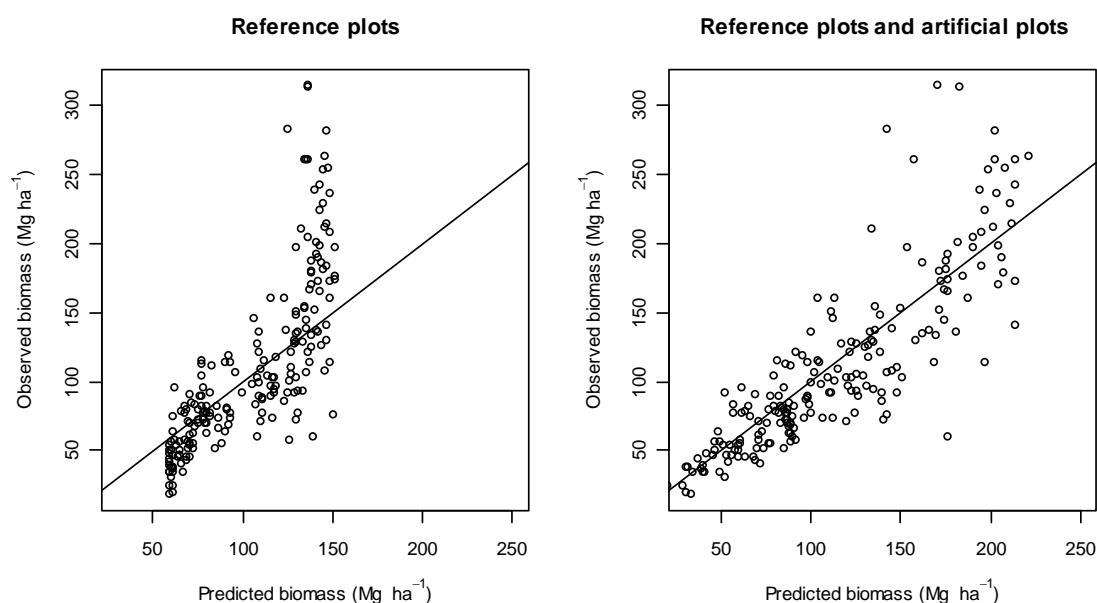


Figure 3: Observed vs. predicted biomass based on the reference plots alone (left) and based on the reference plots augmented by artificial plots (right). The artificial plots were created using the segmented reference plots.

Table 2: Accuracy measures of the predictions

	RMSE (Mg ha ⁻¹)	RMSE (%) ^a	Mean residual (Mg ha ⁻¹)	Mean residual (%) ^a
Reference plots only	43.46	39.06	6.94	6.24
Reference plots and artificial plots	33.14	29.78	-2.60	-2.34

^a Relative to the mean of the observed values on the validation plots.

4. Discussion and conclusion

The range of the predicted values based on the reference plots alone (between 60 and 150 Mg ha⁻¹) was smaller than the range of the observed values in the reference plots (between 50 and 170 Mg ha⁻¹). This was a result of the regression to the mean effect which occurs if the number of neighbors (*k*) is larger than one (McRoberts 2009). The use of artificial plots levelled out the regression to the mean effect and the extrapolation bias (Magnussen *et al.* 2010) that occurred due to the smaller data range of the reference plots compared with validation data. Another less obvious effect was that gaps within the data range were reduced by using the artificial plots which also increased the accuracy of the prediction. The distances in the feature space between a target and the closest neighbors increase with increasing gap size. Gaps are

especially large within areas of the data range that is only sparsely covered with observations (McRoberts 2009).

This study showed that semi-ITC approaches can be useful also if low density ALS data are acquired in order to carry out an ABA forest inventory. It would be an easy task to increase the point density on the flight lines that are flown to calibrate the flight lines against each other in a low density ALS acquisition. A good coordination between planning of the flight and field campaigns is, however, essential. Since field and remote sensing data acquisition were carried out independently in this study, high density ALS data over the reference sectors had to be acquired separately. The necessary ALS data do not result in higher costs if two separate flight campaigns can be avoided. The tree coordinates that are necessary for the semi-ITC approach followed here (Breidenbach *et al.* 2010a) increase the survey costs compared to field measurements sufficient for the ABA. However, the field survey costs for semi-ITC could also be reduced compared to the field data acquisition necessary for the ABA since the measurement of complete sample plots is not required in semi-ITC. In fact, tree coordinates could be measured for single segments or for segments that are organized in polymorph groups. A survey design suitable for semi-ITC could for example be organized in two steps. i) ITC delineation of all areas with high density ALS data. ii) Selection of segments where tree coordinates can easily be measured. III) Measuring of all tree coordinates within the selected segments. The selected segments should cover the whole range of variation in the data. Some kind of randomization scheme during the selection can reduce subjective influence. It is also of importance that the selected segments are not influenced by the fact that they are easy to measure (e.g., by near-by roads). In terms of precision (sometimes denoted as bias), ITC approaches that do not require tree coordinates are not yet in an operational state (Yu *et al.* 2010).

In operational inventories, predictions need to be tree species-specific (e.g., Breidenbach *et al.* 2010b; Packalén and Maltamo 2008). This will be the focus in further studies. It would also be of interest to assess how different segmentation algorithms affect the quality of the artificial plots.

Acknowledgements

We wish to thank Mr Knut Marius Hauglin, Ms Nadja Thieme and Mr Vegard Lien at the Norwegian University of Life Sciences for the fieldwork and several GIS analyses. Thanks also appertain Blom Geomatics, Norway, for collection and processing of the airborne laser scanner data.

Parts of this study have been funded by the Research Council of Norway (project #184652/I10: “New Technologies to Optimize the Wood Information Basis for Forest Industries - Developing an Integrated Resource Information System (IRIS)”), the Forest Trust Fund, and the Nordic Forest Research Co-operation Committee. The work is a contribution to the WoodWisdom-Net ERA-NET project.

References

- Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., & Solberg, S. (2010a). Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment*, 114, 911-924
- Breidenbach, J., Nothdurft, A., & Kändler, G. (2010b). Comparison of nearest neighbour approaches for small area estimation of tree species-specific forest inventory attributes

- in central Europe using airborne laser scanner data. *European Journal of Forest Research*, 129, 833–846
- Crookston, N.L., & Finley, A.O. (2008). yaImpute: An R Package for k-NN Imputation. *Journal of Statistical Software*, 23, 1–16
- Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E., & Falkowski, M.J. (2009). Corrigendum to "Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data"[Remote Sensing of Environment, 112: 2232-2245]. *Remote Sensing of Environment*, 113, 289-290
- Magnussen, S., Tomppo, E., & McRoberts, R. (2010). A model-assisted k-nearest neighbour approach to remove extrapolation bias. *Scandinavian Journal of Forest Research*, 25, 174-184
- Marklund, L. (1988). Biomass functions for pine, spruce and birch in Sweden. *Rapport-Sveriges Lantbruksuniversitet, Institutionen foer Skogstaxering (Sweden)*
- McRoberts, R. (2009). Diagnostic tools for nearest neighbors techniques when used with satellite imagery. *Remote Sensing of Environment*, 113, 489-499
- Moeur, M., & Stage, A.R. (1995). Most similar neighbor: An improved sampling inference procedure for natural resource planning. *Forest Science*, 41, 337-359
- Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80, 88- 99
- Næsset, E., Gobakken, T., Kangas, A., Maltamo, M., Bollandsås, O.M., Aasland, T., & Solberg, S. (2006). Extending and improving methods for operational stand-wise forest inventories utilizing multi-resolution airborne laser scanner data. In Y. Hirata, Y. Awaya, T. Takahashi, T. Sweda & H. Tsuzuki (Eds.), *Silvilaser 2006* (pp. 75-80). Matsuyama, Japan
- Packalén, P., & Maltamo, M. (2006). Predicting the plot volume by tree species using airborne laser scanning and aerial photographs. *Forest Science*, 52, 611-622
- Packalén, P., & Maltamo, M. (2008). Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. *Canadian Journal of Forest Research*, 38, 1750-1760
- Yu, X., Hyypä, J., Holopainen, M., & Vastaranta, M. (2010). Comparison of Area-Based and Individual Tree-Based Methods for Predicting Plot-Level Forest Attributes. *Remote Sensing*, 2, 1481-1495

Stem detection and measuring DBH using terrestrial laser scanning

Martin van Leeuwen^{1*}, Nicholas C. Coops¹, Glenn J. Newnham², Thomas Hilker³,
Darius S. Culvenor², Michael A. Wulder⁴

¹ University of British Columbia, Forest Resources Management, 2424 Main Mall, V6T
1Z4, Vancouver Canada

² Commonwealth Scientific and Industrial Research Organisation (CSIRO), Locked Bad
10, Clayton South VIC 3139, Australia

³ National Aeronautics and Space Administration (NASA), Goddard Space Flight Center,
Biospheric Sciences Branch Code 614.4, Bld 33, #G310, 8800 Greenbelt Road,
Greenbelt, MD 20771, USA

⁴ Pacific Forestry Centre, 506 West Burnside Road, Victoria, British Columbia, Canada

* Corresponding Authors: mvanleeu@interchange.ubc.ca, nicholas.coops@ubc.ca

Abstract

Plot-level forest inventory information is critical for forest management; however, acquisition of forest structural attributes is a time consuming, costly, and often inconsistent task. Recent developments have utilized terrestrial laser scanning (TLS) for rapid acquisition of forest structural detail as point cloud data, suitable for conversion to forest attributes. Processing of these data remains a non-trivial task. In this paper, an efficient and robust method for stem detection from terrestrial laser scanning is presented based on the Medial Axis Transformation. The algorithm effectively eliminates outliers such as branches and uses a small number of parameters. Results show robustness of model parameters and minor errors of commission, while errors of omission are a function of range distance from the scanner. For horizontal distances up to 15 meter from the scanner 85% of trees manually detected from the scans could also be detected automatically. This includes some very small trees (<10 cm) that would not have been included in field inventories. Implications of the current algorithm and related data requirements on acquisition protocols are discussed.

Keywords: terrestrial laser scanning, medial axis transformation, tree detection

1. Introduction

Field inventory information is of paramount importance in formulating management plans and scenarios. The collection of tree-level information through manual measurements provides for detailed and well understood information to support growth and yield programs and also acts as check to inform less detailed forest inventory procedures, such as the development of polygon based inventory information based upon interpretation of air photos. The information on stem attributes can be used to infer a range of structural parameters such as leaf area index (LAI), above and below ground biomass, and volume. Despite the value of plot-level inventory information, the time and cost involved in field acquisitions is currently a limitation to information availability (Baccini et al. 2007) and novel techniques are needed that allow the rapid and inexpensive *in-situ* acquisition of forest structural parameters (Smith et al., 2003).

Ground-based or terrestrial LiDAR is capable of describing forest structure as point cloud data with very high spatial detail. Individual tree stems are typically detected from horizontal slices of point cloud data (Strahler et al. 2008; Maas et al. 2008). Stem attributes such as diameter at breast height (DBH) are then obtained using a least squares circle or a Hough transform (Bienert et al. 2007; Thies and Spiecker 2004). Stem detection can also be achieved

from range-image data. The reduction to a two dimensional space and the availability of image topology improve computation efficiency. Forsman and Halme (2005) segmented range images into smooth surfaces of a size corresponding to the width of tree trunks. Circles were then fitted through the segments and segments were removed if the residuals were large. Liang et al. (2008) detect trunks as clusters in scan projections using cylindrical coordinates of the LiDAR returns.

The potential for range image data, however, has not been fully addressed in the literature and improvements are needed for robust and automatic tree segmentation from point cloud and range image data. One popular and efficient technique used in describing and reconstructing object shapes in image processing is the Medial Axis Transformation (MAT) (Blum, 1967). The medial axis of an object represents a set of points centred within the object's boundaries and provides a description of object shape that can be used for analysis and object classification.

In this paper, we demonstrate that shape description using MAT allows robust detection of tree stems from 2D range image data. The accuracy of the algorithm is assessed against the raw point cloud, and stem locations obtained from field measurements.

2. Method

2.1 Study area

The study area is a plantation forest on Vancouver Island, British Columbia, Canada, approximately 20 km south of Campbell River. The stand studied primarily consists of Douglas-fir [*Pseudotsuga menziesii* var. *menziesii* (Mirb.) Franco], and a minority of western hemlock [*Tsuga heterophylla* (Raf.) Sarg.] and red cedar [*Thuja plicata* Donn. ex D. Don] comprising 17% and 3% respectively. The predominant tree age was 60 years-old and heights varied between 30 and 35 meter.

2.2 Plot establishment

Four plots of 30 by 30 meter were selected for which DBH and stem locations were acquired. Stem locations were measured relative to the plot centres using a vertex (Haglöf, Sweden) range distance measure and compass bearing, while DBH was measured using a diameter tape measure. Terrestrial LiDAR data was acquired using the EchidnaTM Validation Instrument (EVI) (Jupp et al. 2009; Strahler et al. 2008). The EVI instrument emits pulses at 1064 nm at a rate of 2 kHz and records full-waveform return information. Laser pulses are emitted over zenith angles by a rotating mirror that is inclined at 45° relative to optical axis of the laser light source. Azimuthal scanning is achieved by rotating the tripod mounted instrument around its vertical axis (Jupp et al., 2009; Strahler et al., 2008). The resulting scan covers 360 degrees azimuth and 130 degrees zenith. The data were collected in August 2008 using an angular sampling interval of 4 mrad and a beam divergence of 5 mrad. To ensure full plot coverage, five scans per plot were made representing the four corners and centre location.

2.3. Processing

The full waveform LiDAR data sets were processed with the EVI toolkit developed by CSIRO Melbourne, Australia, to obtain a discrete return point cloud representing foliage, trunks and branches. The data was filtered for last returns and then projected on a 2-D plane using a panoramic, or Andrieu projection (Andrieu et al. 1994). The same projection was applied to a suite of EVI output including return intensity, range distance, Cartesian and polar coordinates, and radial distance. Radial distance is the product of range distance and the sine of the zenith angle and represents the horizontal component of range distance.

From the radial distance image, flat objects were identified as pixels for which all 8-connected neighbouring pixels showed deviations less than a tolerance, δ . The resulting binary image (figure 1, top left) can be used in MAT. Several methods for computing MAT exist (e.g. Borgefors et al. 2008). In one method, the medial axis is computed from the Distance Transformation (DT) of a binary image. The DT computes the cumulative shortest distance from feature pixels, having a value of zero, to every non-feature pixel in the original, binary image array (e.g. Borgefors, 2008; Brieu et al. 1995), where distance can be expressed as Euclidean distance or as the sum of horizontal and vertical displacement components, known as Manhattan distance. In this research, Manhattan distance was used. In the DT image, shown as a relief in figure 1 (top right), medial axes appear as top ridges along tree stems and these top ridges were detected based on the change of slope of the DT image pixel values (e.g. Shih and Pu 1995) along the image horizontal axis. Along with cumulative distance of every non-feature pixel, the DT also returns the coordinates of nearest feature pixels and these were used to associate object contours with their medial representation. Figure 1 (bottom left) demonstrates the medial axis representation and shows their associated contour lines (bottom right).

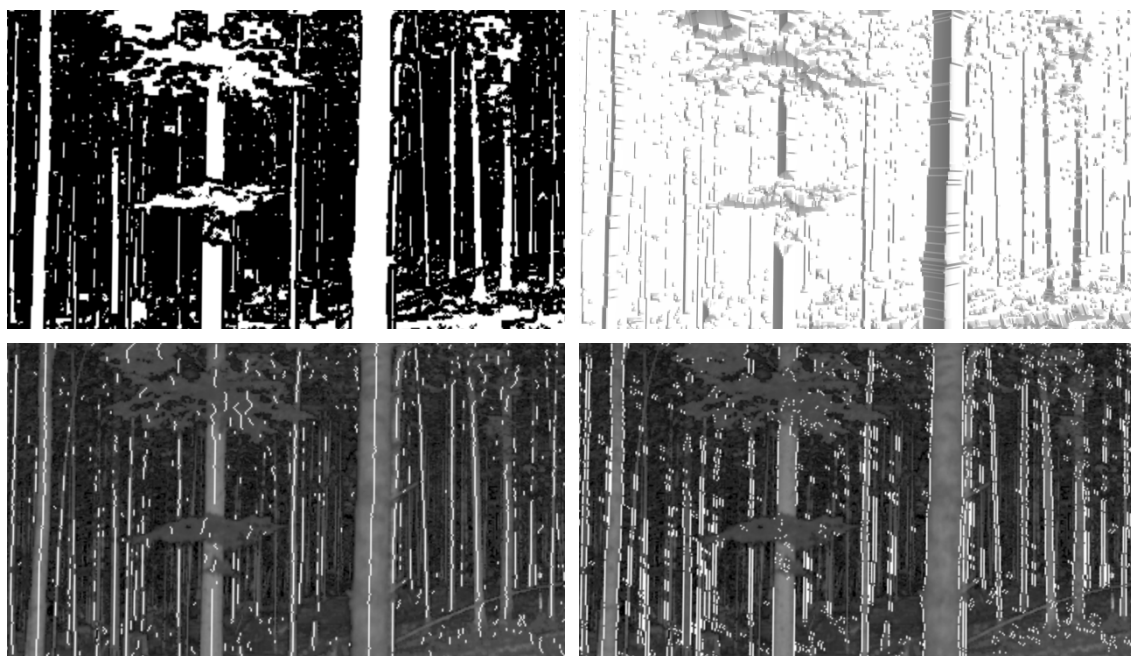


Figure 1: Demonstration of the Medial Axis Transformation. Top left: binary image showing relatively flat surfaces in white. Top right: Distance Transformation computed from the binary image. Bottom left: Medial Axis Transformation derived from Distance Transformation. Bottom right: object contours associated with the medial axes (bottom images are overlaid on return intensity data).

Since the medial axes detected to this stage include all discrete objects in the thresholded radial distance image, filtering is needed to retrieve only the medial axes from tree stems. To do this, objects were classified as stems based on a measure of correlation, ρ , a measure of variation in direction of their medial axes, ζ , and the number of pixels contained in their medial axes, n . For every medial axis, the two associated object contours were examined and assessment was made to what degree these contours represented parallel, planar curves. Hereto, the normalized cross correlation was computed between contours using their pixel y-coordinates. In addition, a measure of local orientation was obtained for every pixel of the medial axes. For a particular medial axis pixel, this was measured as the slope (in image space)

of its two associated contour pixels, and pixels showing a change in slope with that of adjacent pixels beyond the set threshold for ζ , were removed from the medial axes. This measure allows detection and removal of any abrupt changes that indicate incompleteness of object shape or otherwise non-tubular shapes, i.e. objects other than tree stems. Finally, n was used to filter any small objects for which no reliable normalized cross-correlation can be computed.

For each plot, the plot corner scans were co-registered to the centre scan using interactive tools and visual inspection of the point clouds to correct for six degrees of freedom (i.e. translation and Euler angle offsets). The co-registered tree locations were then plotted against field measured tree locations.

3. Result

All scans were processed using threshold parameters $\delta=0.30$ m, $\rho=0.95$, $\zeta=15^\circ$, and $n=24$. The value used for δ was important in separating stems of trees that were standing closely in front of one another, while it was also found insensitive to instrument noise. The correlation coefficient, ρ , was an important parameter in tree detection and measured similarity between the stem contours at the left and right sides of one stem centre. For stems to be detected, a high degree of similarity between their contours is needed. The correlation coefficient was found rather insensitive and values between 0.9 to 0.995 generally produced very similar results. Lower values, however, may result in inclusion of parts of branches and foliage and these errors of commission were deemed of higher cost than omission. Increasing the value for ζ results in an increase of errors of commission, especially in the forest canopy, although this parameter is considered less significant than ρ . In addition, n was an important parameter for removing non-stem related information. Figure 2 shows detected tree stems for one scan. In general, stems were detected with good accuracy for all scans and showed minor errors of commission; however, in rare cases objects at ground level and close to the scanner were falsely classified as tree stems, and some detected trees that appeared close to the scanner showed stem diameters at or below the threshold used for inclusion in field inventories (i.e. 10 cm). The detection rate is a function of range and decreases at greater distance from the scanner. From the trees that could be manually detected in the laser scans within radii of fifteen meters from the scanner's origin, 85% was automatically detected. Figure 3 shows co-registered tree locations from LiDAR scans against the field measured tree locations for one plot. For the co-registered scans an average of 9.25 trees per plot totalling 9.8 % of trees detected in the field inventory were not located in the EVITM derived stem maps. This was a result of both occlusion and the instruments resolution.



Figure 2 Automatic tree detection from range image data. Shown in bright white is the Medial Axis Transformation, filtered using the parameters ρ , ζ , and n .

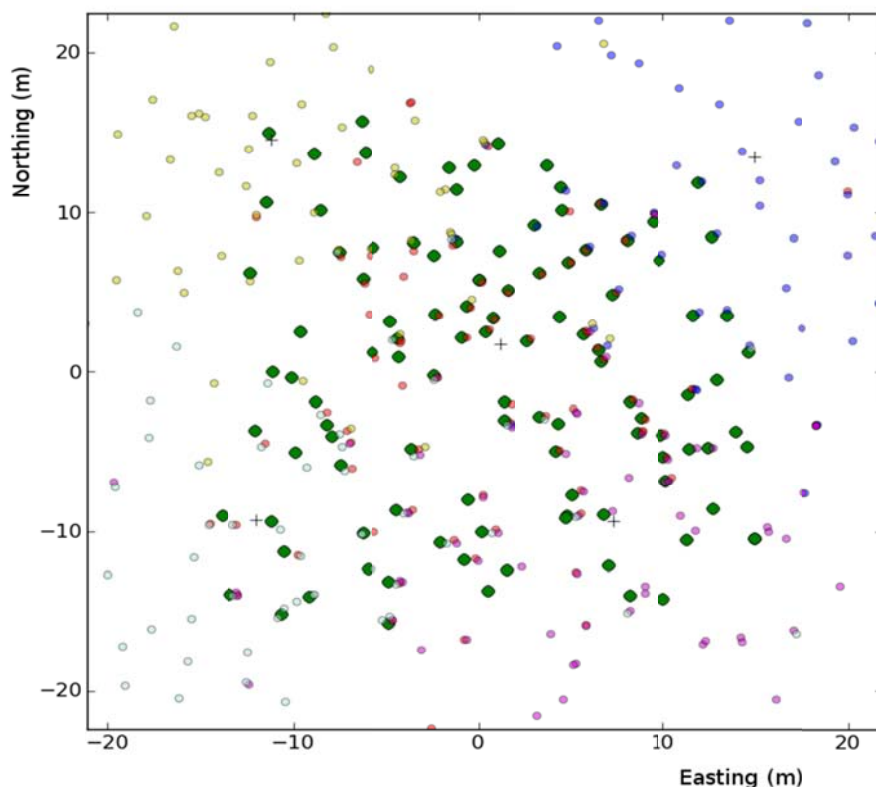


Figure 3: Comparison of field inventory tree locations shown in green and Echidna™ derived tree locations using the presented automatic tree detection algorithm. Individual scans are shown in unique colours (yellow, blue, red, purple, and magenta) and were manually co-registered to a common, local coordinate frame. Individual scan centres are indicated with plus-signs.

4. Discussion

In this paper, a novel method for the detection and reconstruction of tree stems is presented that is both computationally efficient and robust in detecting tree stems from terrestrial LiDAR data. The presented method uses range distance information combined with the Medial Axis Transformation (MAT) and a filtering procedure to detect and segment individual tree stems and stem parts. The computationally efficient algorithm allows processing on relatively low-capacity hardware and immediate processing in the field. The algorithm proved highly robust against errors of commission, while errors of omission are a function of range distance. The computation required on average approximately 3 seconds per scan. A full examination of the sensitivity of each parameter (i.e. δ , ρ , ζ and n) was not conducted. Fallen or heavily leaning tree stems were not detected using the current implementation of the algorithm that only scans along the image horizontal to extract and filter medial axes. However, using either the transpose of the panoramic projections or by implementing a full medial axis transformation, rotation invariance of tree detection can be achieved at the cost of doubling computation time.

Since the detection of trees is achieved in a 2D image space, detection may also be achieved from photographs using thresholds for δ that are based on colour and intensity information and this may open opportunity for low-cost, consumer-grade photo cameras for the automated acquisition of tree count and attribute information. Future research will focus on the automatic reconstruction of branching and canopy structural details to assess effects of canopy structure on physiological functions through integration of airborne and terrestrial LIDAR.

Acknowledgements

We would like to express our thanks to members of the Integrated Remote Sensing Studio, University of British Columbia for their assistance with the field work. Parts of this research are funded by an NSERC Discovery grant to Coops. Additional funding is received from the Canadian Wood Fibre Centre, of the Canadian Forest Service of Natural Resources Canada.

References

- Andrieu, B., Sohbi, Y., Ivanov, N., 1994. A direct method to measure bidirectional gap fraction in vegetation canopies. *Remote Sensing of Environment*, 50, 61-66.
- Baccini, A., Friedl, M.A., Woodcock, C.E., Zhu, Z., 2007. Scaling Field Data to Calibrate and Validate Moderate Spatial Resolution Remote Sensing Models. *Photogrammetric Engineering and Remote Sensing*, 73, 945-954.
- Bienert, A., Scheller, S., Keane, E., Mohan, F., Nugent, C., 2007. Tree detection and diameter estimations by analysis of forest terrestrial laser scanner point clouds. In: *Proceedings of ISPRS Workshop "Laser Scanning 2007 and SilviLaser 2007"*. Espoo, Finland: 50–55.
- Blum, H., 1967. A transformation for extracting new descriptors of form. In: W. Wathen-Dunn (Ed.). *"Models for the Perception of Speech and Visual Form"*. MIT Press: Cambridge, MA, pp. 362-380.
- Borgefors, G., Nystrom, I., Sanniti di Baja, G., 2008. Discrete Skeletons from Distance Transforms in 2D and 3D. *Computational Imaging and Vision*, 37, 155-190.
- Brieu, H., Gil, J., Kirkpatrick, D., Werman, M., 1995. Linear time Euclidean distance transformation algorithms. *IEEE Transactions on pattern analysis and machine intelligence*, 17, 529-533.
- Forsman, P., Halme, A., 2005. 3-D mapping of natural environments with trees by means of mobile perception. *IEEE Transactions on Robotics*, 21, 482-490.
- Jupp, D.L.B., Culvenor, D.S., Lovell, J.L., Newnham, G.J., Strahler, A.H., & Woodcock, C.E., 2009. Estimating forest LAI profiles and structural parameters using a ground-based laser called 'Echidna'. *Tree Physiology*, 29, 171-181.
- Liang, X., Litkey, P., Hyypä, J., Kukko, A., Kaartinen, H., Holopainen, M., 2008. Plot-level trunk detection and reconstruction using one-scan-mode terrestrial laser scanning data. 2008 International workshop on earth observation and remote sensing applications, IEEE, Beijing, 30 June–2 July.
- Maas, H.-G., Bienert, A., Scheller, S., Keane, E., 2008. Automatic forest inventory parameter determination from terrestrial laser scanner data. *International Journal of Remote Sensing*, 29, 1579–1593.
- Shih, F.Y., Pu, C.C., 1995. A Skeletonization algorithm by maxima tracking on Euclidean distance transform. *Pattern Recognition*, 28, 331-341.
- Smith, J.L., Clutter, M., Keefer, B., Zhenkui, M., 2003. The Future of Digital Remote Sensing for Production Forestry Organizations. *Forest Science*, 49, 455-456.
- Strahler, A.H., Jupp, D.L.B., Woodcock, C.E., Schaaf, C.B., Yao, T., Zhao, F., Yang, X., Lovell, J., Culvenor, D., Newnham, G., Ni-Miester, W., Boykin-Morris, W., 2008. Retrieval of forest structural parameters using a ground-based lidar instrument 'Echidna®'. *Canadian Journal of Remote Sensing*, 34, S426-S440.
- Thies M., Spiecker, H., 2004. Evaluation and future prospects of terrestrial laser scanning for standardized forest inventories. In: M. Thies, B. Koch, H. Spiecker, H. Weinacker (Eds.). *Proceedings of ISPRS working group VIII/2 "Laser-Scanners for Forest and Landscape Assessment"*. Freiburg, University of Freiburg: 192–197.

Tree biomass estimation using ALS features

Minna Rätty¹, Ville Kankare¹, Xiaowei Yu², Markus Holopainen¹, Mikko Vastaranta¹,
Tuula Kantola¹, Juha Hyyppä² & Risto Viitala³

¹Department of Forest Sciences/University of Helsinki, Finland,
minna.s.raty@helsinki.fi, first.last@helsinki.fi

²Finnish Geodetic Institute, xiaowei.yu@fgi.fi, juha.hyyppa@fgi.fi

³HAMK University of Applied Sciences, Finland, risto.viitala@hamk.fi

Abstract

Today the estimation of biomass and detection of changes in biomass in large areas is based on coarse remote sensing data and field measurements, which are time consuming, expensive and, above all, in local level inaccurate. The recent development of techniques has offered opportunities to develop new methods, e.g. laser scanning. Airborne laser scanning (ALS) derived features could be used to estimate the total biomass of standing trees. The objective of this study was to make preliminary investigations between accurately measured biomasses in the field and ALS derived features. Study material consisted of 38 sample trees: 19 Scots pines (*Pinus sylvestris*) and Norway spruces (*Picea abies*), which biomasses were accurately measured. ALS derived segments representing the field trees were matched and features for trees were extracted from ALS points within segments. Correlations between biomasses and ALS features were calculated and simple regression models were formulated. The relative residual errors were 21% for Scots pine and 40% for Norway spruce. More empirical tests are needed for ALS based tree biomass estimations.

Keywords: Airborne laser scanning, Biomass, Estimation, Regression

1. Introduction

One of the biggest challenges of the programs aiming to reduce global emissions e.g. Reducing Emissions from Deforestation and Forest Degradation (REDD) is how to measure and monitor forest biomass and its changes effectively and accurately. Remote sensing methods (RS), such as optical and microwave satellite imaging, aerial photography, and laser scanning, are efficient tools for various forest monitoring tasks. The recent knowledge of forest biomass and its changes is based on more or less subjective ground measurements and coarse or medium resolution satellite images. Therefore the accuracy of the biomass estimations, especially in local level (e.g. forest stand), is poor. Though, attempts to improve local estimates have been (Tuominen *et al.* 2010).

Forests are one of the major carbon sinks in the global ecosystem. Because the canopy height, biomass, and carbon pools are functionally related, canopy height, which can be measured accurately by means of ALS, is a critical parameter in terrestrial carbon cycle (Kellndorfer *et al.* 2010). The leaf area index has also been used as a measure of biomass (Koch 2010) and it has been successfully mapped with ALS, using ground calibration (e.g. Solberg *et al.* 2006).

Airborne and terrestrial laser scanning (ALS & TLS) are promising techniques for efficient and accurate biomass detection because of their capability of direct measurement of vegetation structure or tree and stand characteristics (e.g. Holopainen *et al.* 2010). With ALS's ability to directly measure forest structure, including canopy height and crown dimensions, it is increasingly used for forest inventories at different levels. Previous studies have shown that ALS data can be used to estimate a variety of forest inventory attributes including tree, plot and

stand level estimates for tree height (Hyypä and Inkinen 1999; Magnussen *et al.* 1999; Maltamo *et al.* 2004), biomass (Lefsky *et al.* 1999; Bortolot and Wynne 2005; Van Aardt *et al.* 2008), volume (Næsset 1997; Hyypä *et al.* 2001; Wallerman and Holmgren 2007), basal area (Lefsky *et al.* 1999; Means *et al.* 2000; Næsset 2002) and tree species (e.g. Holmgren and Persson 2004; Van Aardt *et al.* 2008; Brandtberg 2009; Vauhkonen *et al.* 2010). ALS is also a promising method for monitoring forest hazards and defoliation, because of its ability to derive vegetation structure properties (Solberg and Næsset 2006; Kantola *et al.* 2010).

Determination of stem and crown biomass requires accurate measurements of stem, bark, branches and needles of an individual tree (Repola 2007; 2009). These measurements are time consuming especially for mature trees. The trees under investigation also need to be felled down for accurate measurements. As far as we know, this kind field data has not yet been used in ALS based biomass estimation. The objective of this study was to make preliminary investigations between accurately measured biomasses in the field and ALS derived features.

2. Material and method

2.1 Study area

The study area is located in the vicinity of Evo, in Southern Finland (61.19°N, 25.11°E). The material was collected from 11 mature and maturing forest stands including 38 sample trees; 19 Scots pines (*Pinus sylvestris* L.) and 19 Norway spruce (*Picea abies* L.) (Table 1).

Table 1: Description of sample trees

	Scots pine			Norway spruce		
	Mean	Std	Range	Mean	Std	Range
dbh, cm	19.6	4.0	15.6	20.9	7.0	26.6
Height, m	18.8	2.4	9	19.2	6.4	20
Age, year	49	5	16	68.6	25.8	88
Crown Ratio	0.57	0.08	0.30	0.25	0.11	0.35
Liv. Branch, kg	46	25	91	106	69	279
Dead Branch, kg	4.2	3.3	13	4.0	4.5	16
Stem Mass, kg	270	122	446	338	281	1034
Total Mass, kg	321	146	541	448	351	1332

2.2 Measurements

2.2.1 Biomass field measurements

First the trees were felled. The total height of the tree, heights of the living and dead crown were recorded. The living crown was divided into four equal length sections and from every section one “typical” branch was selected for dry biomass measurements. Also one dead branch was selected. The felled trees were lopped off and the masses of the branches in these five classes were recorded (one class for the dead branches and four for the living branches).

The stem was cut into logs. The first cut was at the stump height, the second at the middle between the stump and breast heights, the third at breast height (1.3 m), and then starting from 2 m height every meter. The bottom diameter of the logs was measured and the masses were weighed. From every other log a sample disc of 15 cm height were cut for measurements of the moisture content of the bark and stem wood. The bark, stem wood and branches (5 per tree) were dried in an oven at 70 °C temperature for 2-3 days.

2.2.2 Biomass estimation for the sample trees

For every tree the following biomasses were estimated: stem wood, stem bark, and living and dead branches. The branch biomass included both branch wood and bark. The biomasses of the trees were predicted by applying ratio estimation methods. The measured moisture content of the sample discs for both the bark and stem wood separately were applied to the stem mass together with the estimated proportion of bark. Having the sample discs from different heights of each tree, we applied the proportions and ratios measured for the logs that were next to the disc. If some measurement were missing the closest result along the stem was used instead.

The sample branches were used to estimate the branch dry weight from the fresh mass. Ratio estimates for living branch biomass were calculated first by crown sections. The total living branch biomass was the sum of the biomasses of the crown sections. The same method were applied to dead branches.

2.2.3 ALS data and individual tree detection

The ALS data was acquired in July 2009 with a Leica ALS50-II SN058 laser scanner (Leica Geosystem AG, Heerbrugg, Switzerland). The flying altitude was 400 m, at a speed of 80 knots, a half-angle of 15 degrees, a pulse rate of 150 kHz and a footprint diameter of 6 cm. The density of the returned pulses within the plot was approximately 10 hits per m². The ALS data were first classified into ground and non-ground points, using the standard approach of the TerraScan-based method explained in Axelsson (2000). A digital terrain model (DTM) was then developed, using classified ground points, and laser heights above the ground (normalized height or canopy height) were calculated by subtracting the ground elevation from the laser measurements. Canopy heights close to zero were considered as ground returns and those greater than 2 m as vegetation returns (P0, Table 2). The intermediate data between them were considered as returns from ground vegetation or bushes. Only vegetation returns were used for ALS feature extraction. Several features were extracted from the vegetation returns for plot.

A raster canopy height model (CHM) was created from normalised data for all plots inside the coverage of ALS data for individual tree detection (ITD) and crown segmentation. Single tree segmentations were performed on CHM images using a minimum curvature-based region detector. During the segmentation processes, the tree crown shape and location of individual trees were determined. The procedure consisted of the following steps:

1. CHM was smoothed with a Gaussian filter to remove small variations on crown surface. The degree of smooth is determined by the value of standard deviation (Gaussian scale) and kernel size of the filter.
2. Minimum curvatures were calculated. Minimum curvature is one of the principal curvatures. For a surface like CHM, higher value of minimum curvature describes tree top.
3. The smoothed CHM image was then scaled based on the computed minimum curvature resulting in a smoothed yet contrast stretched image.
4. Local maxima were then searched in a given neighbourhood. They were considered as tree tops and used as markers in the following marker controlled watershed transformation for tree crown delineations.

Each segment was considered to present a single tree crown. Laser returns falling within each individual tree segment were extracted and canopy height of these returns were used for deriving tree features. In total, 26 features were generated including basic segment information: crown area, height and volume, and percentiles of canopy height distribution and canopy cover percentiles as proportion of returns below a given percentage of total height (Table 2). The ALS

data segments and trees were matched in order to get the ALS features for the measured trees.

Table 2: Features extracted from ALS data for trees

Feature	Description
X*	Tree location (Easting)
Y*	Tree location (Northing)
Hmax	Heights range
Hmean	Arithmetic Mean of laser heights
HSTD	Standard deviation of heights
CA	Crown area as area of convex hull
CV	Crown volume as convex hull in 3D
P0-P90	Heights percentiles from 0th to 90th (by every 10th)
P100	Heights Maximum
CCP10-CCP90	Percentage of returns below 10% - 90% of total height (by every 10%)
MaxD	Maximum crown diameter when crown was considered a ellipse

A correlation between the biomasses and the ALS features was calculated:

$$\text{Corr}(B, F) = \frac{\sum_{i=1}^n (B_i - \bar{B})(F_i - \bar{F})}{(n-1)s_B s_F}, \quad (1)$$

where B is a biomass (stem wood, stem bark, living branches, dead branches, total), F ALS feature, n is the number of observations and s_B and s_F are the sample deviations for the biomass and ALS feature, respectively. The strength of correlation is classified according to the absolute value of the correlation coefficient either as none (0 – 0.1), weak (0.1 – 0.3), moderate (0.3 – 0.5) or strong (0.5 – 1).

The first regression models (OLS = ordinary least squares) for the total biomasses were searched with statistical software R (R Development Core... 2009). The stepAIC function available at R library MASS can both drop and add variables among the given set of variables by stepping “forward” and “backward”. With the help of the function the variables were selected to the models. The regression models were fitted to both tree species separately.

4. Result

The correlation between the ALS features and biomasses are shown in Table 3. The most important ALS metrics varied between tree species. For the Scots pines the geometric crown features extracted from ALS data and height percentiles (P20-100) in ALS data are the most important, whereas for the Norway spruces only the height percentiles (P0-100) and percentage of returns below 40% of total height show strong correlation.

The visual interpretation of the relationships between ALS features and biomasses of the single trees did not encourage making any transformations to the ALS features before using them as independent variables in regression models. Relationships were more or less linear. Numerous combinations of variables in regression models were tried, and simple models including only two to three independent coefficients in addition to intercept were selected (Table 4). The estimates with these models are plotted in Fig. 1 against the true measured value and also the model residuals against the estimated values. No obvious trends with either tree species could be detected. Norway spruce has larger deviance, but heteroscedasticity could not be seen.

Table 3: Correlations between biomasses and ALS features.

ALS feature	Scots pine					Norway spruce				
	Stem	Stem bark	Liv. Branch	Dead Branch	Total	Stem	Stem bark	Liv. Branch	Dead Branch	Total
Hmax	0.80	0.76	0.52	0.32	0.78	0.14	0.14	0.11	0.11	0.14
Hmean	0.67	0.67	0.37	0.25	0.64	0.20	0.21	0.21	0.22	0.22
HSTD	0.24	0.11	0.03	0.12	0.21	-0.09	-0.08	0.02	-0.02	-0.06
CH	0.55	0.48	0.34	0.16	0.53	0.11	0.13	0.25	0.33	0.16
CA	0.70	0.70	0.75	0.36	0.73	0.16	0.11	0.02	-0.15	0.12
CV	0.60	0.58	0.48	0.30	0.59	0.13	0.08	0.08	-0.07	0.12
P0	-0.20	-0.13	-0.11	0.01	-0.19	0.59	0.56	0.71	0.49	0.66
P10	0.38	0.47	0.28	0.20	0.37	0.72	0.70	0.56	0.36	0.71
P20	0.51	0.53	0.35	0.24	0.50	0.73	0.72	0.54	0.36	0.72
P30	0.56	0.59	0.25	0.22	0.52	0.76	0.72	0.54	0.34	0.73
P40	0.58	0.58	0.24	0.15	0.54	0.75	0.72	0.52	0.32	0.72
P50	0.63	0.61	0.29	0.20	0.59	0.82	0.81	0.56	0.30	0.79
P60	0.65	0.63	0.31	0.23	0.61	0.87	0.85	0.58	0.31	0.82
P70	0.68	0.65	0.33	0.26	0.64	0.89	0.86	0.60	0.31	0.85
P80	0.73	0.67	0.40	0.30	0.69	0.89	0.85	0.59	0.29	0.84
P90	0.78	0.71	0.45	0.33	0.74	0.90	0.85	0.58	0.26	0.84
P100	0.80	0.76	0.52	0.32	0.78	0.91	0.86	0.60	0.24	0.86
CCP10	-0.14	0.26	-0.18	-0.07	-0.13	-0.31	-0.33	-0.36	-0.20	-0.34
CCP20	-0.18	0.23	-0.25	-0.10	-0.18	-0.41	-0.44	-0.44	-0.25	-0.44
CCP30	-0.17	0.23	-0.27	-0.12	-0.17	-0.42	-0.44	-0.44	-0.28	-0.45
CCP40	-0.20	0.18	-0.32	-0.16	-0.21	-0.48	-0.51	-0.49	-0.33	-0.51
CCP50	-0.23	0.15	-0.34	-0.14	-0.23	-0.44	-0.45	-0.45	-0.35	-0.47
CCP60	-0.11	0.18	-0.16	-0.10	-0.11	-0.38	-0.41	-0.44	-0.40	-0.42
CCP70	0.16	0.34	0.21	0.09	0.18	-0.32	-0.36	-0.38	-0.34	-0.36
CCP80	0.34	0.44	0.41	0.28	0.36	-0.04	-0.11	-0.10	-0.23	-0.07
CCP90	0.31	0.35	0.36	0.22	0.33	-0.02	-0.12	-0.02	-0.26	-0.03
MaxD	0.47	0.48	0.50	0.18	0.49	0.01	-0.04	-0.05	-0.12	-0.01

Table 4: Regression models for total biomasses (kg) of single trees.

	Scots pine			Norway spruce		
	Coefficient	Estimate	p-value	Coefficient	Estimate	p-value
(intercept)		-165.5	0.012	(intercept)	-353.5	0.006
Hmean		246.1	0.000	P70	36.9	0.000
P10		-44.6	0.000	CCP20	-369.4	0.042
P50		-185.2	0.000			
SE		33.5			94.8	
R ²		0.82			0.78	

For comparison estimates of total biomass following equations of Repola (2009) were calculated for the trees and plotted against the true value from field measurements (Fig. 2, left). However, the residuals show a systematic shift in estimates larger than 250 kg (Fig. 2, right). In this case, the biases for the Scots pines were 9.1 kg (5.8%) and for the Norway spruce 35.2 kg (15.0%). The RMSEs combining both bias and deviation were 15.0 kg (9.5%) and 54.6 kg (23.2%), respectively. In parentheses were the relative errors.

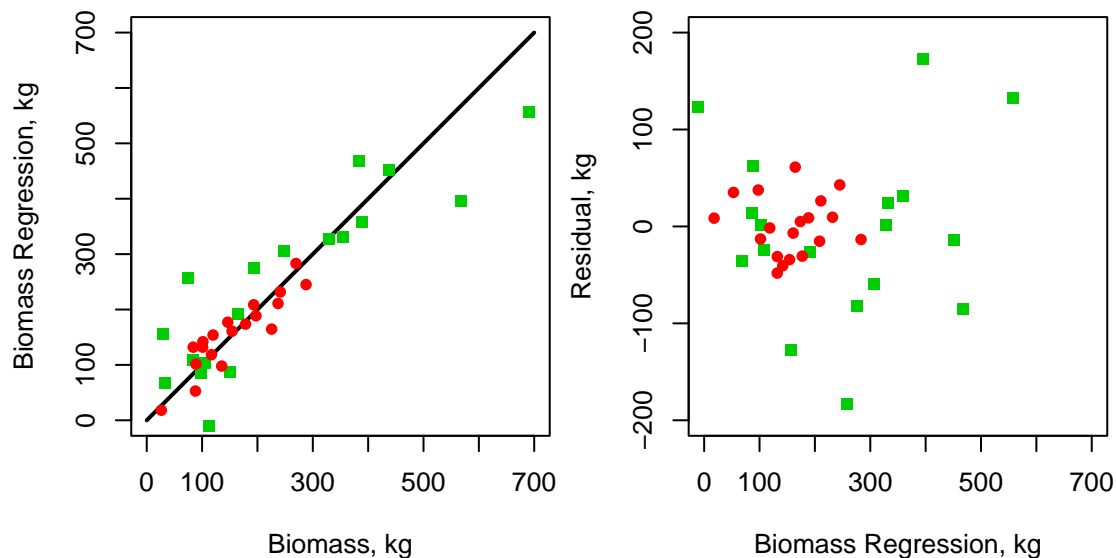


Figure 1: Total biomass of tree (kg) estimated with ALS regression model, vs. field measurements (left) and residuals, vs. estimated total biomass (right), red = Scots pine and green = Norway spruce.

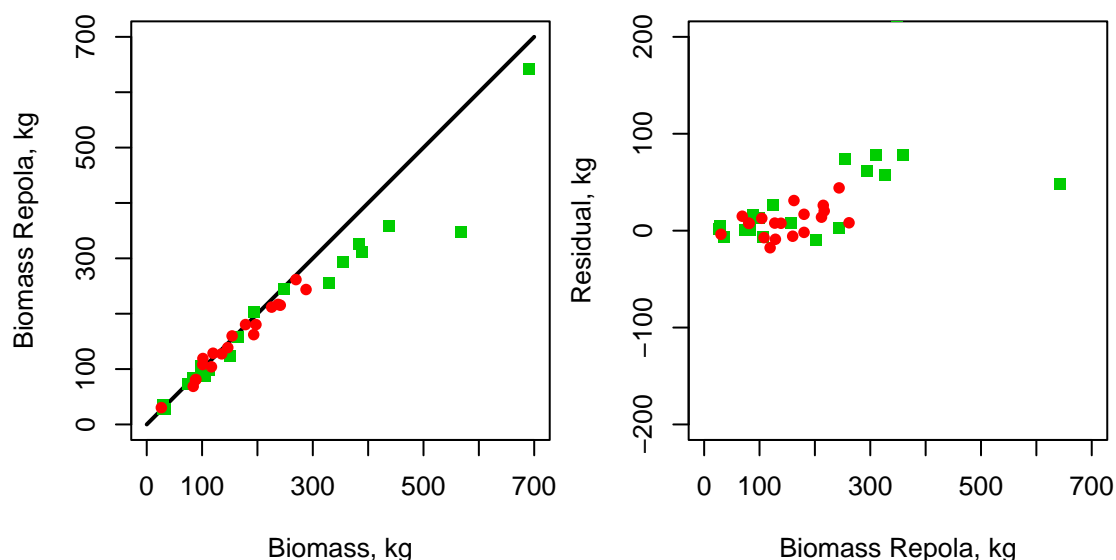


Figure 2: Total biomass of tree (kg) predicted with Repola (2009) vs. field measurements (left) and residuals, vs. predicted total biomass (right), red = Scots pine and green = Norway spruce.

5. Discussion and conclusion

The estimation of biomass for single tree with regression models based on ALS features gave promising results. The residual errors for Scots pine were 34 kg (21%) and for Norway spruce 95 kg (40%) (Table 4). The models of Repola (2009) gave lower RMSEs (9.5% (Scots pine) and 23.2% (Norway spruce)), but similarly the prediction of Scots pine was more accurate than Norway spruce. The coefficients selected for the models are logical. Most of the biomass for the Scots pine is in the stem and therefore the height is important variable (Hmean), but also the density of the crown is included in the model (P50, P10). For the Norway spruce information about the penetration of the laser pulses in the crown (near ground level) and height are important (CCP20 and P70).

This study provided some of the first tests for biomass prediction using ALS data. Results are rather promising, but more field data is needed for developing practical modelling means. Tree biomass estimation accuracy achieved using ALS features could provide alternative or additional information to the current tree biomass models. There are two alternatives to estimate forest biomass for larger areas using tree level information achieved by ALS: 1) using direct ALS feature based models, 2) by deriving dbh, tree species and height of the tree by means of ITD and using biomass models. These variables are possible to estimate with high accuracy using ITD (Maltamo *et al.* 2009, Korpela *et al.* 2010).

Accurate tree level biomass estimation would help globally in the reporting of environmental change-related counter activities. According to our preliminary results, we assume that laser measures offer a good alternative for the traditional field measurements. However, more empirical tests are needed to verify these results.

Acknowledgements

Many thanks to the forestry students at HAMK University of Applied Sciences in Evo who did the hard physical work in field measurements. This study was made possible by financial aid from the Finnish Academy projects “Improving the Forest Supply Chain by means of Advanced Laser Measurements (L-impact)” and “Science and Technology Towards Precision Forestry” (PreciseFor) and the Finnish Cultural Foundation.

References

- Axelsson, P., 2000: DEM Generation from Laser Scanner Data Using Adaptive TIN Models. In *IAPRS*. Vol. 33. part B4/1, pp. 110-117.
- Bortolot, Z. and Wynne, R.H., 2005. Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59, 342-360.
- Brandtberg, T., 2009. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61, 325-340.
- Holmgren, J. and Persson, Å., 2004. Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment*, 90, 415-423.
- Holopainen, M., Vastaranta, M., Rasinmäki, J., Kalliovirta, J., Mäkinen, A., Haapanen, R., Melkas, T., Yu, X., Hyypä, J., 2010. Uncertainty in timber assortment estimates predicted from forest inventory data. *European Journal of Forest Research*, 129:1131-1142.
- Hyypä, J. and Inkinen, M., 1999. Detecting and estimating attributes for single trees using laser scanner. *Photogrammetric Journal of Finland*, 16, 27-42.
- Hyypä, J., Kelle, O., Lehikoinen, M. and Inkinen, M., 2001. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 969-975.
- Kantola, T., Vastaranta, M., Yu, X., Lyytikäinen-Saarenmaa, P., Holopainen, M., Talvitie, M., Kaasalainen, S., Solberg, S. and Hyypä, J., 2010. Classification of Defoliated Trees Using Tree-Level Airborne Laser Scanning Data Combined with Aerial Images. *Remote Sensing* 2(12), 2665-2679.
- Koch, B., 2010. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6), 581-590.
- Kellndorfer, J.M.; Walker, W.S.; LaPoint, E.; Kirsch, K.; Bishop, J. and Fiske, G., 2010. Statistical fusion of lidar, InSAR, and optical remote sensing data for forest stand height

- characterization: A regional-scale method based on LVIS, SRTM, Landsat ETM+, and ancillary data sets. *Journal of Geophysical Research*, 115, G00E08. doi:10.1029/2009JG000997.
- Korpela, I., Ørka, H.O., Maltamo, M., Tokola, T. and Hyypä, J., 2010. Tree species classification using airborne LiDAR—Effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. *Silva Fennica*, 44, 319-339.
- Lefsky, M.A., Harding, D., Cohen, W.B., Parker, G. and Shugart, H.H., 1999. Surface lidar remote sensing of basal area and biomass in deciduous forests of Eastern Maryland, USA. *Remote Sensing of Environment*, 67, 83-98.
- Magnussen, S., Eggermont, P. and LaRiccia, V.N., 1999. Recovering tree heights from airborne laser scanner data. *Forest Science*, 45, 407-422.
- Maltamo, M., Mustonen, K., Hyypä, J., Pitkänen, J. and Yu, X., 2004. The accuracy of estimating individual tree variables with airborne laser scanning in boreal nature reserves. *Canadian Journal of Forest Research*, 34, 1791-1801.
- Maltamo, M., Peuhkurinen, J., Malinen, J., Vauhkonen, J., Packalén, P. and Tokola, T., 2009. Predicting tree attributes and quality characteristics of Scots pine using airborne laser scanning data. *Silva Fennica*, 43(3), 507-521.
- Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L. and Hendrix, C.J., 2000. Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering and Remote Sensing*, 66, 1367-1371.
- Næsset, E., 1997. Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sensing of Environment*, 61, 246-253.
- Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80, 88-99.
- R Development Core Team., 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available online at <http://www.R-project.org>; last accessed June 16, 2011.
- Repola, J., 2009. Biomass equations for Scots pine and Norway spruce in Finland. *Silva Fennica* 43(4), 625-647.
- Repola, J., Ojansuu, R. and Kukkola, M., 2007. Biomass functions for Scots pine, Norway spruce and birch in Finland. *Metlan työraportteja/Working Papers of the Finnish Forest Research Institute* 2007(53). 28 p. Available online at <http://www.metla.fi/julkaisut/workingpapers/2007/mwp053.pdf>, last accessed June 16, 2011.
- Solberg, S. and Næsset, E., 2006. Monitoring forest health by remote sensing. In *Symposium on Forests in a Changing Environment—Results of 20 years ICP Forests Monitoring*, Göttingen, Germany, October 25–28, 2006; Schriftenreihe der Forstlichen Fakultät Göttingen und der Nordwestdeutschen Forstlichen Versuchsanstalt: Göttingen, Germany, 2007a; pp. 99-104.
- Solberg, S., Næsset, E., Hanssen, K.H. and Christiansen, E., 2006. Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sensing of Environment*, 102, 364-376.
- Tuominen, S., Eerikäinen, K., Schibalski, A., Haakana, M. and Lehtonen, A., 2010. Mapping biomass variables with a multi-source forest inventory technique. *Silva Fennica* 44(1), 109-119.
- Van Aardt, J.A.N., Wynne, R.H. and Scrivani, J.A., 2008. Lidar-based mapping of forest volume and biomass by taxonomic group using structurally homogenous segments. *Photogrammetric Engineering and Remote Sensing*, 74, 1033-1044.
- Vauhkonen, J., Korpela, I., Maltamo, M. and Tokola, T., 2010. Imputation of single-tree attributes using airborne laser scanning-based height, intensity and alpha shape metrics. *Remote Sensing of Environment*, 114, 1263-1276.
- Wallerman, J. and Holmgren, J., 2007. Estimating field-plot data of forest stands using airborne laser scanning and SPOT HRG data. *Remote Sensing of Environment*, 110, 501-508.

Stand level species classification and volume estimation using LiDAR height, intensity, and ratio parameters

Taejin Park¹, Doo-Ahn Kwak², Woo-Kyun Lee^{1*} and Jong-Yeol Lee¹

¹Department of Environmental Science and Ecological Engineering, Korea University, Seoul, 136-713, Korea, taejin1392@korea.ac.kr, leewk@korea.ac.kr

²Environmental GIS/RS Centre, Korea University, Seoul 136-713, South Korea, tulip96@korea.ac.kr

Abstract

In this study we use airborne LiDAR to classify tree species and estimate volume at the stand scale using multiple linear discriminant analysis and multiple linear regression analysis. This involved the extraction of 38 independent variables from LiDAR data including height, intensity, and ratio metrics. In stand species classification, the 90 percentile of height ($H_{C,90}$), standard deviation of the intensity ($I_{C,std}$) and vegetation intensity ratio (VIR) were the most suitable variables for explaining each stand species. Hit ratio represented by accuracy in discriminant analysis was 81.7% in stand species classification. Afterward, the regression models were estimated using each variable, with the best model then selected using the corrected Akaike's Information Criterion (AIC_c). $H_{C,90}$, mode of intensity ($I_{C,mode}$) and standard error of mean of intensity ($I_{C,se}$) were applied to optimally explain the stand volume of Japanese Larch (*Larix leptolepis*), with an $R^2 = 0.83$. With the mean of height ($H_{C,mean}$), mode of height ($H_{C,mode}$), standard deviation of intensity ($I_{C,std}$) and range of intensity ($I_{C,range}$) could be used to predict the stand volume of Japanese red pine (*Pinus densiflora*), with an $R^2 = 0.79$. Finally, the 80th height percentile ($H_{C,80}$), $I_{C,mode}$ and the kurtosis of intensity distribution ($I_{C,kurt}$) were applied to predict the stand volume of Oaks (*Quercus* spp.) with an $R^2 = 0.68$.

Keywords: stand species classification, stand volume estimation, LiDAR, Multiple linear discriminant analysis, Multiple linear regression analysis

1. Introduction

Earlier research into mapping forest volume and biomass using airborne LiDAR was based on species specific allometric relationships that were applied to LiDAR derived estimates of individual tree attributes such as maximum tree height, diameter at breast height (DBH) and crown width (Popescu, 2007; Chen *et al.*, 2007). Estimates of DBH typically being developed by the relationship between tree height and crown width (Leckie *et al.*, 2003; Popescu, 2007). However, deriving these individual tree attributes can become complex at the forest stand scale due to variable species composition, density and crown shape (Kim, 2007). To avoid these issues, stand volume has been estimated using height distributional approaches (e.g., median, mode, percentiles) of a large-footprint LiDAR system (Harding *et al.* 2001; Parker *et al.* 2001; Drake *et al.* 2002). The relationships between LiDAR height distributional parameters and stand volume are similar to the allometric relationship between field-measured vertical profiles and the stand volume (Drake *et al.* 2002). In addition, the vertical growth of forest stands has been shown to be correlated with increases in stand biomass levels (Drake *et al.* 2002), and large-footprint LiDAR provides data related with the vertical arrangement of the forest structure from the tree top to the ground (Harding *et al.* 2001). In demonstrating the relationship between LiDAR-derived vertical canopy profiles and biomass, Drake *et al.* (2002) proved the relationship between field-measured vertical profiles, LiDAR-derived vertical canopy profiles and biomass.

To accurately estimate forest volume and biomass over large geographical areas, it is

advantageous to have species information given that allometric equations are typically species specific. Aerial photography and satellite imagery has been widely used to obtain species data by analysing the spectral information. Combining the three dimensional profiles derived from LiDAR system with species information it is then possible to calculate stand volume and biomass. However, the analysis and integration two types of remotely sensed data has the potential to be uneconomical and time-consuming.

In recent years, LiDAR has been shown to discriminate species at the individual tree and stand level (Holmgren and Persson, 2004; Brandtberg, 2007, Kim, 2007). For example, the peak amplitude or intensity of discrete return LiDAR has been indicated to separate tree species (Holmgren and Persson, 2004; Kim, 2007) while the vertical and horizontal profiles could identify crown shape and benefit species discrimination (Kim, 2007). Brandtberg *et al* (2007) also proved that LiDAR intensity data could be used to discriminate coniferous and deciduous forest. Kim (2007) classified 16 tree species using LiDAR data acquired in leaf-off and leaf-on condition.

Therefore, this study attempts to classify tree species at the stand level by applying multiple linear discriminant analysis using information derived from only LiDAR data. Furthermore, classified stand species information was used in multiple linear regression analysis to develop the optimal stem volume model.

2. Materials and methods

2.1 Study area

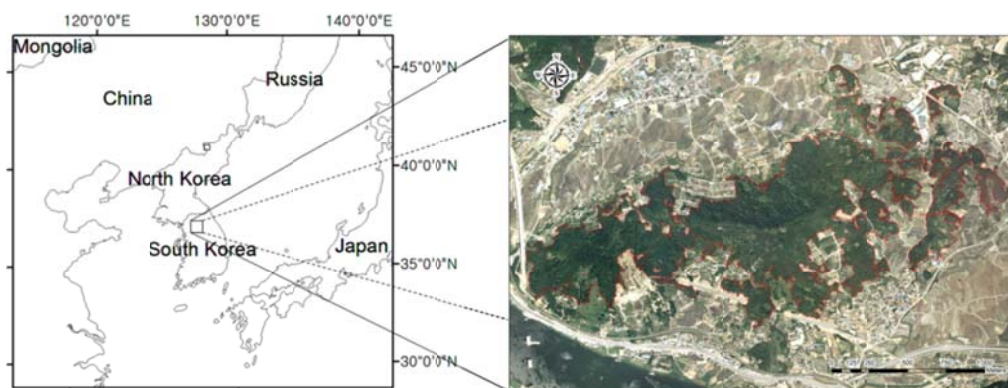


Figure 1. Geographical location of study area with aerial imagery

The study areas were located in the Obin area of Yangpyeong-gun in Gyeonggi-do, (Datum: Geodetic Reference System 1980 (GRS 1980), Coordinates: the upper left 127° 27' 04.08\"E, 37°31'15.80\"N and lower right 127°29'22.89\"E, 37°30'15.45\"N), central South Korea (Figure 1). The forest of the study area occupied approximately 164.1ha and ranged from 21m to 220m above sea level. This area is dominated by gradual hills with the main tree species being Japanese red pine (*Pinus densiflora*), Japanese Larch (*Larix leptolepis*) and oaks (*Quercus* spp.). For this study, 20 sample plots at each tree species were investigated. The stands were selected in such a way as to provide a homogeneous composition of tree species.

2.2 LiDAR and ground data

In this paper, ALTM 3070 (a small footprint LiDAR system) by Optech Inc. was used to acquire the LiDAR data for tree species classification and volume estimation, with the flight performed on the 4th of May 2009. The LiDAR data were acquired from an altitude of 1400m, with a sampling density of 4~9 points per square meter. A field survey was conducted from September to October, 2010. Sample plots consisted of 20 square plots of each stand species,

respectively, each with an area of 0.02 ha (14 meters on a side). Sample plots were used to discriminate stand species and model the stand volume and LiDAR height, intensity and ratio parameters. Due to absence of test plot measurement, this study performed a cross validation to stand species classification. However, estimating stand volume based on optimal regression model didn't evaluate with test plot, therefore further study might measure the test plot and verify the results additionally.

Table 1. Descriptive statistics of the field measurements

Estimator	Statistics	<i>L. leptolepis</i>	<i>P. densiflora</i>	<i>Q. spp.</i>
Stand Age	Max.	50.0	45.0	46.0
	Min.	29.0	35.0	27.5
	Mean.	44.4	40.5	33.4
	Std.	8.2	3.9	5.7
Stand DBH (cm)	Max.	34.2	29.4	22.1
	Min.	12.7	17.8	12.5
	Mean.	25.2	23.2	18.2
	Std.	6.5	3.3	3.3
Stand Height (m)	Max.	22.4	17.4	14.8
	Min.	10.4	9.7	9.8
	Mean.	16.5	13.1	12.2
	Std.	4.6	2.3	1.7
Tree Number	Max.	21.0	55.0	83.0
	Min.	11.0	32.0	13.0
	Mean.	17.7	40.3	39.4
	Std.	3.2	7.1	19.7

3. Method

In this study, the stand level species classification and volume estimation were performed using multiple linear discriminant analysis and multiple linear regression analysis, with LiDAR height, intensity and ratio parameters and field measurements of the stand species and volume as independent and dependent variables, respectively (Figure 2).

3.1 Extraction of LiDAR height, intensity and ratio parameters

The LiDAR height, intensity and ratio parameters can be prepared as representative statistics, such as percentile, mean, maximum, minimum, median, mode, standard deviation, coefficient of variation, standard error of the mean, kurtosis, skewness, the number of canopy and total returns, and range. All parameters were related with the height information of the plots, but especially the kurtosis and skewness were included as explanatory variables due to their relationship to the vertical distribution of LiDAR returns within the plots. In probability theory and statistics, kurtosis is a measure of the "peakedness" of the probability distribution, where a higher kurtosis indicates a greater variance as the result of infrequent extreme deviations, as opposed to frequent deviations. Accordingly, the kurtosis of LiDAR returns reflected within plots might be used to explain how many LiDAR returns are concentrated around the mean height, considering the distribution of simultaneous LiDAR returns. Moreover, an attempt was made to apply skewness as an independent variable for estimating the plot volume. The skewness measures the asymmetry of the probability distribution in probability theory and statistics. Thus, it might be able to predict how many LiDAR returns would be reflected around the ground or canopy within plots via the skewness of LiDAR returns. LiDAR intensity distributional information was also significant in the modeling of tree-specific parameters (Means *et al.* 2000). Additionally, the canopy return ratio, which is the ratio of the number of canopy returns and the total number of LiDAR returns

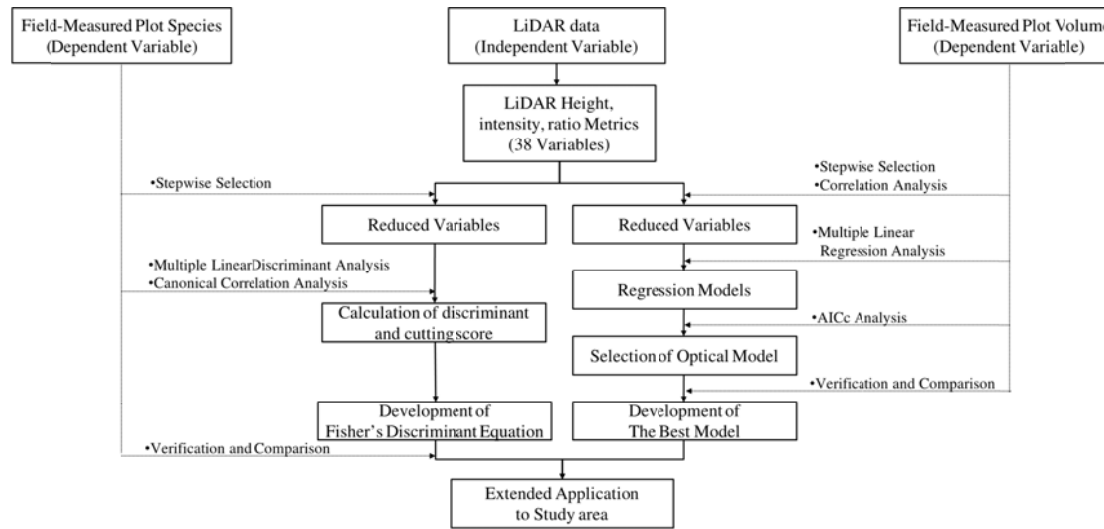


Figure 2. Flowchart of stand species classification and stand biomass estimation

Table 2. Definition of independent variable metrics

Independent variables		
Height metrics based on canopy returns	Intensity metrics based on canopy returns	Ratio metrics based on integrated canopy and ground returns
$H_{C,i}$, $i=10, 20, \dots, 100$ percentile height	$I_{C,mean}$, mean of intensity	Num_T , number of total returns
$H_{C,mean}$, mean of height	$I_{C,max}$, maximum of intensity	Num_C , number of canopy returns
$H_{C,max}$, maximum of height	$I_{C,min}$, minimum of intensity	VRR , vegetation return ratio
$H_{C,min}$, minimum of height	$I_{C,med}$, median of intensity	I_T , sum of total intensity
$H_{C,med}$, median of height	$I_{C,mode}$, mode of intensity	I_C , sum of canopy intensity
$H_{C,mode}$, mode of height	$I_{C,std}$, standard deviation of height	VIR , vegetation intensity ratio
$H_{C,std}$, standard deviation of height	$I_{C,cv}$, coefficient of variation of height	
$H_{C,cv}$, coefficient of variation of height	$I_{C,se}$, standard error of mean of intensity	
$H_{C,se}$, standard error of mean of height	$I_{C,kurt}$, kurtosis of intensity distribution	
$H_{C,kurt}$, kurtosis of height distribution	$I_{C,skew}$, skewness of intensity distribution	
$H_{C,skew}$, skewness of height distribution	$I_{C,range}$, range of intensity	
$H_{C,range}$, range of height		

1
2
3
4
5

1 per plot, was added, because Means *et al.* (2000) and Næsset (2002) demonstrated that the metrics
2 were useful descriptors of the tree volume.

3.2 Multiple linear discriminant analysis for stand species classification

6 To classify three stand species including *P. densiflora*, *L. leptolepis* and *Q. spp.*, this study
7 adopted the multiple linear discriminant analysis with field-surveyed stand species information
8 as a dependent variable. This analysis was fundamentally based on a way that minimizes the
9 inner-class variance as well as maximizes inter-class variance for effective pattern recognition.
10 In this study, 38 independent variables were employed for discriminating the stand species.
11 However, the use of all candidate variables to separate the stand species would be inefficient
12 due to the need for intensive and time-consuming work in collecting and managing all the data.
13 In particular, the discriminant equation is unreliable because the variations in the estimated
14 parameters and species are increased due to multi-collinearity. Therefore, the reduced
15 discriminant equation, with essential explanatory variables, has to be performed using stepwise
16 selection method under 0.05 significant levels.

17 According to Kim (2008), for performing linear discriminant analysis, every independent
18 variable and covariance matrix should be a multivariate normal distribution and identical matrix,
19 respectively. However, independent variables used in this study rejected the normality
20 assumption of linear discriminant analysis through normality test, such as Shapiro-Wilk and
21 Anderson-Darling test. Moreover, it was also dismissed in identical covariance matrix
22 assumption by Box-M test. Nevertheless, Klecka (1980) showed that these rejected assumptions,
23 not a great influence, just decrease the efficiency and accuracy of discriminant analysis slightly.
24 Therefore, multiple linear discriminant analysis makes a progress for classifying stand species.
25 The accuracy of classification result was evaluated by hit ratio represented as explanation of
26 developed discriminant equation. In addition, Fisher's linear discriminant equation was applied
27 to allocate other unknown stand species.

3.3 Multiple linear regression analysis for stand volume estimation

3.3.1 Selection of explanatory variables

32 Multiple linear regression modeling was performed using the field-measured plot volume as a
33 dependent variable. In this study, 38 independent variables were employed for regression
34 modeling of the plot volume, using the LiDAR height, intensity and ratio metrics. However, use
35 of the full model for all candidate variables to estimate the plot volume would be inefficient due
36 to the need for intensive and time-consuming work in collecting and managing all the data. In
37 particular, the regressed model is unreliable because the variations in the estimated parameters
38 and plot volume are increased due to multi-collinearity. Therefore, the reduced model, with
39 essential explanatory variables, has to be regressed using stepwise selection methods under 0.05
40 significant level (Chen *et al.* 2007). Afterward, Variance Inflation Factor (VIF) and correlation
41 coefficient were adopted to reduce the inappropriate variables. Therefore, in the case where the
42 VIFs of stepwise selected independent variables were under 10, correlated independent
43 variables were eliminated for stricter statistics, which had Pearson's correlation coefficients over
44 0.5 (van Aardt *et al.*, 2006).

3.3.2 Selection of regression model

48 Generally, different regression models with various combinations of selected variables can be
49 assessed by their R^2 , adjusted R^2 , Root Mean Square Error (RMSE), Sum of Square Error (SSE),
50 Akaike's Information Criterion (AIC). The coefficient of determination, R^2 , provides a measure
51 of how well future outcomes are likely to be predicted by the model without considering the
52 number of independent variables. However, the fitness assessment between regression models
53 must be carried out using the adjusted R^2 value, which adjusts for the number of explanatory

1 terms in a model. Kvålseth (1985) recommended the following adjusted R^2 equation for
 2 evaluating regression models. However, the adjusted R^2 does not consider the residual variance.
 3 AIC is used to discriminate the suitability and accuracy between different regression models,
 4 which measures the relative distance between estimates and observations (Akaike 1974).
 5 Additionally, AIC considers the residual variance when assessing the fitness of models that are
 6 contrary to the adjusted R^2 . However, in practice, the corrected AIC (AIC_c) must be used,
 7 because secondary-bias may occur when the sample size is small (Burnham and Anderson 2002).
 8 Therefore, AIC_c was used for evaluating and selecting regression models, because the sample
 9 size of 20 observations in our study was relatively small (Equation 2).

$$AIC_c = n \left[\ln \left(\frac{\sum (y - \hat{y})^2}{n} \right) \right] + 2k + \frac{2k(k+1)}{n-k-1} \quad (2)$$

11
 12
 13 When comparing different regression models, the estimated AIC_c values are generally
 14 rescaled with the minimum AIC_c values. This transformation allows the best model to have $\Delta_i =$
 15 0. Models with $\Delta_i \leq 2$ are accepted as providing substantial support, those with $4 \leq \Delta_i \leq 7$ as less
 16 support, and those with $\Delta_i \geq 10$ as having no support (Burnham and Anderson 2002).

18 4. Result and discussion

19 4.1 Stand species classification using multiple linear discriminant analysis

20
 21 Explanatory variables of the LiDAR height, intensity and ratio metrics were extracted via the
 22 stepwise selection method at a significant level of 0.05. The selected independent variables were
 23 90th percentile of height ($H_{C,90}$), standard deviation of the intensity ($I_{C,std}$) and vegetation intensity
 24 ratio (VIR) were the most suitable variables for explaining each stand species (Figure 3). Fisher's
 25 linear discriminant equations of each stand species group were estimated and unknown stand
 26 species could be determined by deriving the higher value through comparison between
 27 calculated value from each equation (Equation 3, 4, 5).

$$29 \text{Larix leptolepis} = 20.271 \cdot H_{C,90} - 0.675 \cdot I_{C,std} + 1010.785 \cdot VIR - 614.361 \quad (3)$$

$$30 \text{Pinus densiflora} = 18.385 \cdot H_{C,90} - 0.585 \cdot I_{C,std} + 1013.633 \cdot VIR - 589.463 \quad (4)$$

$$31 \text{Q. spp.} = 19.310 \cdot H_{C,90} - 0.666 \cdot I_{C,std} + 1038.743 \cdot VIR - 624.776 \quad (5)$$

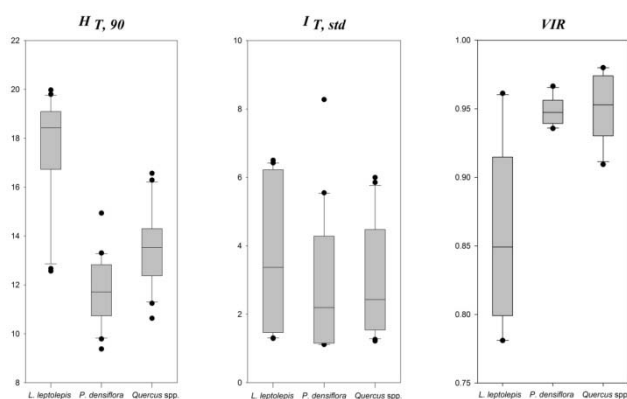


Figure 3. Distributions for different tree species and selected parameters

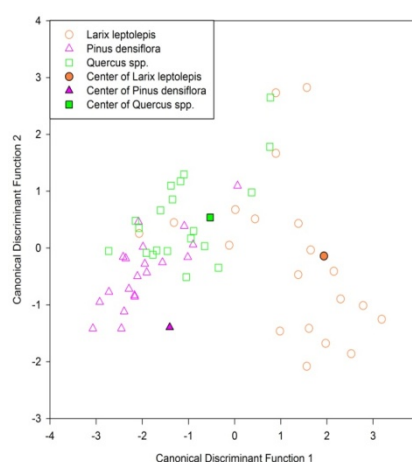


Figure 4. Distribution of discriminant score and centroid by first and second canonical discriminant function

1 Among 60 surveyed sample plots, 49 plots were correctly classified to actual stand species by
 2 cross-validation, in addition, total hit ratio was assessed at 81.7 % (Table 7).

3
 4 Table 7. Stand species classification results by Fisher's discriminant equation
 5

	Species	Predicted species class			Hit ratio(%)
		<i>L. leptolepis</i>	<i>P. densiflora</i>	<i>Q. spp.</i>	
Referenced species class	Frequency	<i>L. leptolepis</i>	17	1	2
		<i>P. densiflora</i>	0	17	3
		<i>Q. spp.</i>	1	4	15
	Probability	<i>L. leptolepis</i>	85.0	5.0	10.0
		<i>P. densiflora</i>	0.0	85.0	15.0
		<i>Q. spp.</i>	5.0	20.0	75.0
					81.7

6 7 4.2 Stand volume estimation using multiple linear regression analysis

8 9 4.2.1 Selection of explanatory variables

10
 11 In case of *L. leptolepis*, five explanatory variables were selected: 90 percentile ($H_{C,90}$), standard
 12 deviation of height ($H_{C,std}$), mode of the intensity ($I_{C,mode}$), standard error of mean of intensity ($I_{C,se}$)
 13 and the sum of intensity (I_T) among the LiDAR height, intensity and ratio metrics, and with low
 14 multi-collinearity due to the VIF of approximately 1 (Table 8). Moreover, explanatory variables were
 15 reselected via a comparison of a correlation analysis between candidate variables. From the result,
 16 $H_{C,std}$ was eliminated due their relatively higher coefficients than 0.5 (Table 9).

17 Likewise, the explanatory variables were selected using the stepwise selection method at a same
 18 significant level to *P. densiflora*. The extracted variables were mean of height ($H_{C,mean}$), mode of
 19 height ($H_{C,mode}$), standard deviation of intensity ($I_{C,std}$) and the range of intensity ($I_{C,range}$). The
 20 multi-collinearity between the selected variables was low compared to their VIFs (below 10), as
 21 shown in Table 8. Using the VIF analysis, the correlations between candidate variables were
 22 investigated via one-to-one comparisons. However, from the results, all four variables were included
 23 for the regression analysis due their relatively low coefficients (below 0.5) (Table 9).

24
 25 Table 8. Results of variable selection to each tree species by variance inflation
 26

Species	Variable	DF	Parameter Estimate	Standard Error	t value	pr> t	Variance Inflation
<i>L. leptolepis</i>	Intercept	1	-11.2321	1.91114	-5.88	<.0001	0.0000
	$H_{C,90}$	1	1.44559	0.28551	5.06	0.0002	4.07915
	$H_{C,std}$	1	-2.24028	0.62636	-3.58	0.0030	4.03682
	$I_{C,mode}$	1	-0.65035	0.19885	-3.27	0.0056	1.00719
	$I_{C,se}$	1	63.19259	20.10955	3.14	0.0072	1.48727
	I_T	1	0.00112	0.000192	5.83	<.0001	1.37517
<i>P. densiflora</i>	Intercept	1	1.19921	2.42319	0.49	0.6279	0.00000
	$H_{C,mean}$	1	0.38271	0.11826	3.24	0.0055	1.48927
	$H_{C,mode}$	1	0.07408	0.04048	1.83	0.0872	1.20480
	$I_{C,std}$	1	2.54207	1.37647	1.85	0.0846	1.52052
	$I_{C,range}$	1	-0.85953	0.24021	-3.58	0.0027	1.86565
<i>Q. spp.</i>	Intercept	1	-0.55562	1.09140	-0.51	0.6181	0.00000
	$H_{C,80}$	1	-0.47026	0.19574	-2.40	0.0297	8.50554
	$H_{C,90}$	1	0.72066	0.18835	3.83	0.0017	8.57314
	$I_{C,mode}$	1	0.06646	0.02167	3.07	0.0078	1.24328
	$I_{C,kurt}$	1	0.84446	0.42291	2.00	0.0643	1.22767

27

Table 9. Results of variable selection to each tree species by correlation coefficient

Species	Variables	$H_{C,90}$	$H_{C,std}$	$I_{C,mode}$	$I_{C,se}$	I_T
<i>L. leptolepis</i>	$H_{C,90}$	1.00000	0.98055	-0.0341	0.18118	-0.15654
	$H_{C,std}$	0.98055	1.00000	-0.03256	0.26466	-0.07958
	$I_{C,mode}$	-0.0341	-0.03256	1.00000	-0.04156	-0.03433
	$I_{C,se}$	0.18118	0.26466	-0.04156	1.00000	-0.14369
	I_T	-0.15654	-0.07958	-0.03433	-0.14369	1.00000
<i>P. densiflora</i>	Variables	$H_{C,mean}$	$H_{C,mode}$	$I_{C,std}$	$I_{C,range}$	
	$H_{C,mean}$	1.00000	0.20974	-0.47106	-0.43012	
	$H_{C,mode}$	0.20974	1.00000	-0.20509	-0.41167	
	$I_{C,std}$	-0.47106	-0.20509	1.00000	0.44358	
<i>Q. spp.</i>	Variables	$H_{C,80}$	$H_{C,90}$	$I_{C,mode}$	$I_{C,kurt}$	
	$H_{C,80}$	1.00000	0.93215	0.15998	-0.07099	
	$H_{C,90}$	0.93215	1.00000	0.25939	0.14843	
	$I_{C,mode}$	0.15998	0.25939	1.00000	0.45776	
	$I_{C,kurt}$	-0.07099	0.14843	0.45776	1.00000	

Explanatory variables of the LiDAR metrics for *Q. spp.* stand were extracted via the same procedure, consequently, the first candidate independent variables shown in Table 8 were the 90 percentile ($H_{C,80}$), 90 percentile ($H_{C,90}$), mode of the intensity ($I_{C,mode}$) and the kurtosis of intensity distribution ($I_{C,kurt}$). The multi-collinearity between the selected variables was weak, since each VIF had a value below 10. Furthermore, as a result of the correlation analysis between the selected variables, $H_{C,80}$ and $H_{C,90}$ were highly correlated, with R^2 values over 0.5 (Table 9). Therefore, $H_{C,90}$ was rejected as candidate independent variables in the final regression analysis because both the probability value from a t -test and the VIF of $H_{C,80}$ were lower than $H_{C,90}$. Eventually, $H_{C,80}$, $I_{C,mode}$ and $I_{C,kurt}$ were adopted for the multiple linear regression analysis for predicting the plot volume.

4.2.2 Selection of regression model

The independent variables in *L. leptolepis* selected using the stepwise selection method and correlation analysis were used in the regression models, which were generated by combining the 4 reduced independent variables. The predictable equation was estimated by adopting the optimal regression model for estimating the plot volume represented by ΔAIC_c below 2 (Table 10). The two optimal regression models found with $\Delta AIC_c \leq 2$ were model no. 1, which estimates the plot volume using $H_{C,90}$, $I_{C,mode}$ and I_C , and model no. 2, where the explanatory variables were $H_{C,90}$, $I_{C,mode}$, $I_{C,se}$ and I_T when AIC_c was employed as the first criterion for selecting the best model. Among these two models, model no. 2 was more significant statistically when other statistics, such as RMSE, SEE, R^2 and Adjusted R^2 , were compared between the two. Therefore, the model showing the best performance when both models are applied to the test plots needs to be determined. Unfortunately, optimal regression model selection procedure was deferred due to absence of test plot measurements.

In *P. densiflora* stand, four explanatory variables were stepwise selected and analyzed for correlation. Therefore, the combinable regression model was estimated, with optimal regression models with ΔAIC_c below 2 chosen (Table 11). However, the best model, and the only one with $\Delta AIC_c \leq 2$, was model no. 1, by which the plot volume could be estimated using $H_{C,mean}$, $H_{C,mode}$, $I_{C,std}$ and $I_{C,range}$, while two models were retained in the case of *L. leptolepis* stand. Models 2 to 7 were eliminated as candidate regression models because their ΔAIC_c was over 2, and they showed a lack of statistical significance with increasing ΔAIC_c .

In the case of *Q. spp.* stand, three explanatory variables were selected and applied to the regression analysis. The results of this regression procedure indicated that one regression model was recommended with $\Delta AIC_c \leq 2$: model no. 1, which estimated the stand volume using $H_{C,80}$, $H_{C,mode}$ and $I_{C,kurt}$. Models 2 to 7 were eliminated as candidate regression models for estimating plot volume of *Q. spp.* stand, because their ΔAIC_c was over 2, and they showed a lack of statistical significance with increasing ΔAIC_c .

1 Table 10. Result of stem volume parameters estimated multiple regression analysis to *Larix leptolepis*

No.	Intercept	$H_{C,90}$	$I_{C,mode}$	$I_{C,se}$	I_T	RMSE	SSE	R ²	Adjusted R ²	AIC _C	ΔAIC _C
1	-4.71040	0.44771	-0.69894	.	0.000753	0.58949	5.5600	0.81366	0.77873	-17.6027	0.0000
2	-5.85002	0.43730	-0.68725	24.2152	0.000782	0.58676	5.1643	0.82693	0.78077	-17.0792	0.5235
3	-5.13540	0.45399	.	.	0.000776	0.68392	7.9517	0.73351	0.70216	-12.4468	5.1559
4	-6.38033	0.44243	.	26.6189	0.000808	0.68341	7.4728	0.74956	0.70260	-11.6893	5.9134
5	-2.79655	0.41578	-0.73511	.	.	0.74849	9.5241	0.68081	0.64326	-8.8381	8.7646
6	-3.44416	0.40871	-0.72885	14.7069	.	0.76551	9.3760	0.68578	0.62686	-7.1515	10.4512
7	-3.18132	0.42134	.	.	.	0.8224	12.1741	0.59200	0.56934	-5.9285	11.6742

2

3 Table 11. Result of stem volume parameters estimated multiple regression analysis to *Pinus densiflora*

No.	Intercept	$H_{C,mean}$	$H_{C,mode}$	$I_{C,std}$	$I_{C,range}$	RMSE	SSE	R ²	Adjusted R ²	AIC _C	ΔAIC _C
1	1.19921	0.38271	0.07408	2.54207	-0.85953	0.59237	5.2634	0.79490	0.74020	-16.6989	0.0000
2	2.64933	0.38166	.	2.59838	-1.01080	0.63437	6.4387	0.74910	0.70205	-14.6681	2.0308
3	3.82277	0.32662	0.07575	.	-0.69251	0.63543	6.4603	0.74826	0.70106	-14.6013	2.0976
4	5.36580	0.32427	.	.	-0.84348	0.67256	7.6898	0.70035	0.66510	-13.1169	3.5820
5	8.03553	.	0.07448	.	-1.00925	0.74067	9.3262	0.63658	0.59383	-9.2581	7.4408
6	9.52304	.	.	.	-1.15547	0.76430	10.5149	0.59026	0.56750	-8.8587	7.8402
7	-1.83127	0.50375	0.12139	.	.	0.76151	9.8583	0.61585	0.57065	-8.1484	8.5505

4

5 Table 12. Result of stem volume parameters estimated multiple regression analysis to *Quercus* spp.

No.	Intercept	$H_{C,80}$	$H_{C,mode}$	$H_{C,kurt}$	RMSE	SSE	R ²	Adjusted R ²	AIC _C	ΔAIC _C
1	-0.71001	0.28685	0.07623	0.31517	0.67066	7.1966	0.67930	0.61917	-12.4426	0.0000
2	-0.78485	0.25922	0.10226	.	0.72128	8.8441	0.60588	0.55951	-10.3196	2.1230
3	-0.46668	0.33777	.	0.52472	0.78235	10.4051	0.53632	0.48177	-7.0687	5.3739
4	2.21758	.	0.11427	.	0.84555	12.8691	0.42652	0.39466	-4.8181	7.6245
5	2.50174	.	0.09646	0.22681	0.83989	11.9922	0.46560	0.40272	-4.2296	8.2130
6	-0.44895	0.31571	.	.	0.95199	16.3130	0.27305	0.23266	-0.0754	12.3672
7	3.58202	.	.	0.48424	0.98272	17.3833	0.22535	0.18231	1.1956	13.6382

4.3 Discussion

The multiple linear discriminant analysis undertaken by this study for stand species identification indicated the most suitable explanatory variables were 90 percentile of height ($H_{C,90}$), standard deviation of the intensity ($I_{C,std}$) and vegetation intensity ratio (VIR). Holmgren and Persson (2004) showed that $H_{C,90}$ was able to be used to be a variable for species identification according to close relation with height of dominant trees in stand, and crown shapes difference by growth differences due to stand species. In addition to $H_{C,90}$, $I_{C,std}$ and VIR were a meaningful variables for discriminating stand species through involving the characteristics of spectral reflectance on each species crown. However, there was not obvious distinction between species in each variable distribution, except $H_{C,90}$ (Figure 3). When considered data acquisition condition, there was an uncertainty whether *L. leptolepis* and *Q. spp.* were leaf-on or –off in 4th of May (Yu, 2011). The leaf unfolding condition could fully influence to $I_{C,std}$ and VIR distribution (Kim, 2007), thus, further study needs a more effective analysis using $I_{C,std}$ and VIR variables at leaf-on conditions.

With the regression model using the LiDAR height metrics, the estimation was performed in each stand species with the $H_{C,90}$, $H_{C,80}$, $H_{C,mean}$, $H_{C,mode}$ and $H_{C,kurt}$ parameter values for predicting the plot volume. It was adjudged that only the canopy LiDAR height metrics-extracted regression model might have an inapplicability problem when estimating the plot volume, even though previous researchers (Chen *et al.* 2007) were able to estimate the plot or stand volume using only the canopy height metrics. However, every variable selected in this study had essential meaning for predicting the stand volume. Therefore, it was not considered possible to predict the stand volumes using only the canopy LiDAR height metrics in our study area.

Intensity data were used to develop the regression models for stand volume estimation. However, the accuracy of the verification decreased steeply when intensity information was included, and their influence was weak when the absolute values of their coefficients were examined. This is because the intensity information of the LiDAR data reflected onto objects could not be normalized according to the object heights above sea level, although the intensity differed according to the medium and height. The intensity data of the LiDAR metrics were demonstrated to be ineffective variables for estimating the plot volume, despite the contrary research of van Aardt *et al.* (2006), because the total corrected intensity could not be estimated according to the increasing elevation.

5. Conclusion

Previous studies classifying tree species at the stand-level used remotely sensed optical imagery, and estimated stand-level volume using DBH and tree height derived from LiDAR data. To scale-up from the individual tree to stand level, the results of this study suggest the use of LiDAR-derived height, intensity and ratio metrics.

The investigation of variables uncorrelated with the metrics will be an important part of our future study for estimating the plot or stand volume. In addition, the consideration of plot density data derived from LiDAR data and acquisition condition will be needed for future accurate volume and species surveys. If these shortcomings can be overcome, LiDAR data will be utilized with increased usefulness in estimating plot- or larger-scale volumes.

Acknowledgements

This study was carried out with the support of “Forest Science and Technology Projects (Project No. S120911L010130)” provided by Korea Forest Service.

References

- 1 Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on*
2 *Automatic Control*, 19: 716-723.
- 3
- 4 Brandtberg, T., 2007. Classifying individual tree species under leaf-off and leaf-on conditions
5 using airborne lidar, *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(5):
6 325-340.
- 7
- 8 Burnham, K.P. and Anderson, D.R. (Eds.), 2002, *Model Selection and Multi model Inference: A*
9 *Practical Information-theoretic Approach, Second edition* (New York: Springer-Verlag
10 Press).
- 11
- 12 Chen, Q., P. Gong, D. Baldocchi, and Y. Q. Tian, 2007. Estimating basal area and stem volume
13 for individual trees from LiDAR data, *Photogrammetric Engineering and Remote Sensing*,
14 73: 1355–1365.
- 15
- 16 Drake, J.B., Dubayah, R.O., Clark, D.B., Knox, R.G., Blair, J.B., Hofton, M.A., Charzdon, R.L.,
17 Weishampel, J.F. and Prince, S., 2002, Estimation of tropical forest structural
18 characteristics using large-footprint LiDAR. *Remote Sensing of Environment*, 79: 305–
19 319
- 20
- 21 Harding, D.J., Lefsky, M.A., Parker, G.G. and Blair, J.B., 2001, Laser altimeter canopy height
22 profiles: Methods and validation for deciduous, broadleaf forests. *Remote Sensing of*
23 *Environment*, 76: 283–297.
- 24
- 25 Holmgren, J. and Å. Persson, 2004. Identifying species of individual trees using airborne laser
26 scanner, *Remote Sensing of Environment*, 90(4): 415-423.
- 27
- 28 Kim, J. H., 2008. Multivariate statistical analysis using SAS. Kyowoosa. Seoul, Korea.
- 29
- 30 Kim, S., 2007. Individual tree species identification using LIDAR- derived crown structures and
31 intensity data, Ph.D. dissertation, University of Washington, Washington.
- 32
- 33 Klecka, W. R., 1980. Discriminant Analysis, Beverly Hills, Sage, London.
- 34
- 35 Kutner, M., Nachtsheim, C and Neter, J. (Eds.), 2004, *Applied Linear Regression Models*, 4th
36 edition (Irwin: McGraw-Hill).
- 37
- 38 Kvalseth, T. O., 1985, Cautionary note about R^2 . *American Statistician*, 39: 279–285.
- 39
- 40 Leckie, D., F. Gougeon, D. Hill, R. Quinn, L. Armstrong, and R. Shreenan, 2003. Combined
41 high-density LiDAR and multispectral imagery for individual tree crown analysis,
42 *Canadian Journal of Remote Sensing*, 29: 633-649.
- 43
- 44 Means, J. E., S. A. Acker, B. J. Fitt, M. Renslow, L. Emerson, and C. J. Hendrix, 2000.
45 Predicting forest stand characteristics with airborne scanning lidar, *Photogrammetry*
46 *Engineering and Remote Sensing*, 66: 1367-1371.
- 47
- 48 Næsset, E., 2002, Predicting forest stand characteristics with airborne scanning laser using a
49 practical two-stage procedure and field data. *Remote Sensing of Environment*, 80: 88–99.
- 50

1 Parker, G.G., Lefsky, M.A. and Harding, D.J., 2001, Light transmittance in forest canopies
2 determined using airborne laser altimetry and in-canopy quantum measurements. *Remote*
3 *Sensing of Environment*, 76: 298–309.
4
5 Popescu, S. C., 2007. Estimating biomass of individual pine trees using airborne lidar, *Biomass.*
6 *Bioenergy*, 31: 646–655.
7
8 van Aardt, J.A.N., Wynne, R.H. and Oderwald, R.G., 2006, Forest volume and biomass
9 estimation using small-footprint LiDAR distributional parameters on a per-segment basis.
10 *Forest Science*, 52: 636–649.
11
12 Yu, N. H., Shin, J. S., Kang, H. G., Shin, H. T., 2011. A Study on the Plants for Phenology of
13 the Mt. Gumwon. *Climate research*. 6(1): 1-87.
14

Effect of scan coverage on stem diameter measurement using terrestrial lidar

Akira Kato¹, L.Monika Moskal² & Tatsuaki Kobayashi³

¹GraduateSchool of Horticulture, Chiba University, Japan akiran@faculty.chiba-u.jp

²Precision Forestry Cooperative, College of the Environment, School of Forest Resources, University of Washington, USA lmMoskal@u.washington.edu

³GraduateSchool of Horticulture, Chiba University, Japan ktatsu@faculty.chiba-u.jp

Abstract

This paper presents a new approach to measure stem diameters based on the data acquired by multiple scanning by terrestrial lidar. Recent terrestrial lidar (Riegl VZ400) has wider coverage and is able to efficiently provide the highest point density data. Stem diameter derived from terrestrial lidar was compared with field measured diameter at breast height (d.b.h) of 42 sample trees. Stem returns of d.b.h were extracted and used to identify the approximated stem centre using principal component analysis. Various scan coverage of stem returns was used in the algorithm developed in this study to assess which is the most appropriate to measure stem diameter. The results show that more than 40% scan coverage of stem returns can produce stem diameter with the error of 5 cm or less using the algorithm. The applied technique can also assess the quality of wood by estimating straightness of stems from the alignment of stem centres at several heights. Furthermore, stem volume which is the most important variable to estimate the amount of carbon can also be measured directly using this technique.

Keywords: Stem diameter, Scan angle, Stem volume, Terrestrial lidar, Principal component analysis

1. Introduction

Above-ground carbon and woody biomass can be predicted using allometric equations with predictor variables such as diameter at breast height (d.b.h) and tree height. Tree parameters derived from remotely sensed data may be used as the inputs in the equations. The effectiveness of airborne Light Detection and Ranging (lidar) to acquire such tree parameters has been demonstrated across a wide range of forest types (Andersen *et al.*, 2006, Hyypä *et al.*, 2001, Kato *et al.*, 2009, and Næsset & Økland., 2002). The applicability of airborne lidar may however be limited by available facilities and cost, which depends on the distance from the airport and the flying time involved in data acquisition, among other things. For this reason airborne data collection is not always feasible in developing countries. Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (REDD) projects focus on the change in the amount of carbon sequestered in forests. Carbon change can be monitored using airborne lidar. But even if airborne lidar is the most promising technology to measure tree parameters (Patenaude *et al.*, 2003), the airborne sensor itself may not be available in the developing countries.

With the recent development of terrestrial lidar a more mobile sensor has become available to acquire high quality data for forest assessment. Even if terrestrial lidar does not offer the area coverage of airborne lidar and is less suitable for the measurement of total tree height, it is comparatively more suitable for the measurement of diameter, which is strongly correlated with biomass and carbon. To monitor the change in the amount of carbon in developing countries, terrestrial lidar is well suited since it can save and display the data in three dimensions, which

can be useful in the event that corner markers of permanent plots are stolen or lost. Moreover, change in the details of objects such as stem growth can be measured directly by multi-temporal data. Efficient data collection by terrestrial lidar provides opportunities for the development of alternative forest inventory techniques.

To ensure the applicability of terrestrial lidar for forest inventory, scanning angle and scanning coverage (tree stem illumination) should be considered to ensure the quality of stem diameter measurements. In past studies, terrestrial lidar has achieved data densities high enough to measure leaf angles (Eitel *et al.*, 2010) and reconstruct the detailed shape of trees (Côté *et al.*, 2009). Terrestrial lidar technologies can detect the details of the objects only if the coverage of scanning is sufficient. Chasmer *et al* (2006) compared terrestrial and airborne lidar systems for forest inventory and found that the tree parameters acquired by terrestrial lidar, especially tree height and crown length, were underestimated. The accuracy of tree parameters can be improved by better placement of the sensor. Hopkinson *et al* (2004) used terrestrial lidar to derive plot level forest metrics. When a clear view of tree top is obtained, underestimation in measurement of tree height is minimized. Stem diameter is typically derived using algorithms to fit a circle or cylinder to the lidar point cloud (for example Tansey *et al.*, 2009). In this study, a new algorithm is introduced to fit the point distribution flexibly to derive stem diameter.

The purpose of this paper is to assess the effect of scan angle and scan coverage on the accuracy of stem diameter measurements. The steps to be taken for the assessment are:

1. to extract the points at the breast height of stem.
2. to estimate the coverage scan angle from the sensor locations.
3. to measure stem diameter and validate it by the field measurement.
4. to assure the quality of data in measuring stem diameter.

2. Method

2.1 Study area

The study area is located in Fujinomiya city, Shizuoka Prefecture, Japan. The coordinates of the study area are 35.255 N and 138.673 E. An experimental scanning in this study was conducted on private land in May 2011. The dominant species is Japanese Cedar (*Cryptomeria japonica*). The terrain is almost flat and there is no understory except for fern (Figure 1).



Figure 1. The picture of study area (left). The horizontal view of terrestrial lidar data with color (right).

2.2 Terrestrial lidar sensor

The terrestrial lidar sensor used in this study was the Riegl VZ400 (Table 1). This is a full-waveform digitizing sensor, which can collect more returns per pulse resulting in high density point clouds suitable for detection of stem shape details.

2.3 Experimental design

The sensor locations were chosen to assess the effect of scan angle and scanning coverage. Five scanning positions were chosen 10 m apart. Forty-two trees were selected as the validation sample for the study. The trees were selected near, away from, and along the sensor locations to obtain a range of scan angles. Stem locations were mapped using a total station and stem d.b.h. were measured using a diameter tape. The sensor location and sampled stem locations are displayed in Figure 2.

2.4 Extraction of laser returns

Based on the mapped stem location, the point clouds associated with each of the 42 stems were extracted and further analysed.

Table 1. Terrestrial lidar sensor setting

Acquisition date	May 16th , 2009
Laser sensor	Riegl VZ-400
Laser wavelength	Near Infrared Red
Max range	150 m (360° x 100°)
Laser point density	125,000 points/second (high speed mode) 42,000 points/second (long distance mode)

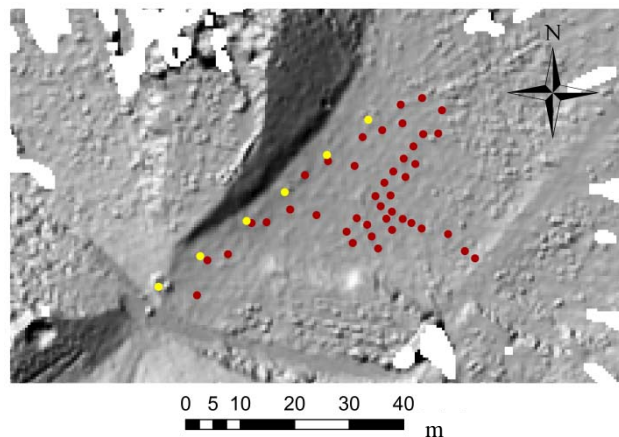


Figure 2. Sensor and stem locations for this study. Yellow points are sensor locations (10 m apart) and brown points are sampled stem locations. The background is the DTM derived from terrestrial lidar data .

2.5 Constructing DTM

Stem diameters were measured at a height of 1.3m above the ground. To define ground level in the lidar data set a Digital Terrain Model (DTM) was constructed based on the smoothed surface of the lowest lying points. First a fifty centimetre resolution grid was generated and the points with minimum height were identified and saved at each grid cell. If there were no data in a grid cell, for example on the obscured side from the sensor location, then the grid cell was assigned a height based on the heights of neighbouring grid cells. In grid cells where laser points were not reflected from the ground but from the understory, the height value was high relative to the surrounding cells. In that case the minimum values of the neighbouring grid cells were assigned.

The procedure was reiterated to smooth the surface.

2.6 Approximating the centre of stem

Points between height 1.25 and 1.35m above the DTM were extracted from the individual stem point clouds. The centre of the stem was approximated in three steps. Firstly, the average coordinate of all extracted points were calculated. Secondly, angles from the average location were obtained and points in some fixed sector were extracted and used to estimate the normal vector, which was determined by eigenvector of principal component analysis. Lastly, the intersected coordinates of these normal vectors were identified and the average of all the intersected coordinates was defined as the approximated centre of stem. The variable angle intervals (from 10 to 180 degrees) were tested to find the optimum extracted angle.

2.7 Scan angle effect on stem diameter

Coverage angle of each stem from the sensor locations was estimated by simulation, or virtual scanning from the actual sensor location to the stem locations. There are two steps for this simulation. In the first step, the coverage angle was estimated without any objects based on two sensor locations. 50 cm grids were generated over the entire study area to estimate the visible side from the two sensors located 50 meters apart and calculate point density. With the both scan angle and point density map, the value of one was assigned to the cells where the point density was more than one and zero was assigned to the others to create a mask, which determined the visible side from two sensors. In this simulation, the maximum distance of lasers was set to be 100 meters from the sensor locations.

2.8 Stem diameter measurement

To measure the stem diameter of terrestrial lidar at breast height, the inner side of the points of returns from stems were used. The outlined points on the surface of stem represented bark and the points at the inner side were only used for the stem measurement. Points at every 5 degrees from the approximated centres were collected and the minimum distance was set to be a sample diameter. All sample diameters were averaged to give the stem diameter of a tree.

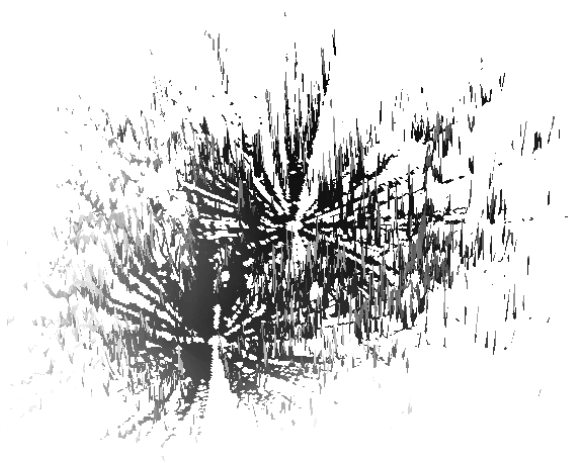


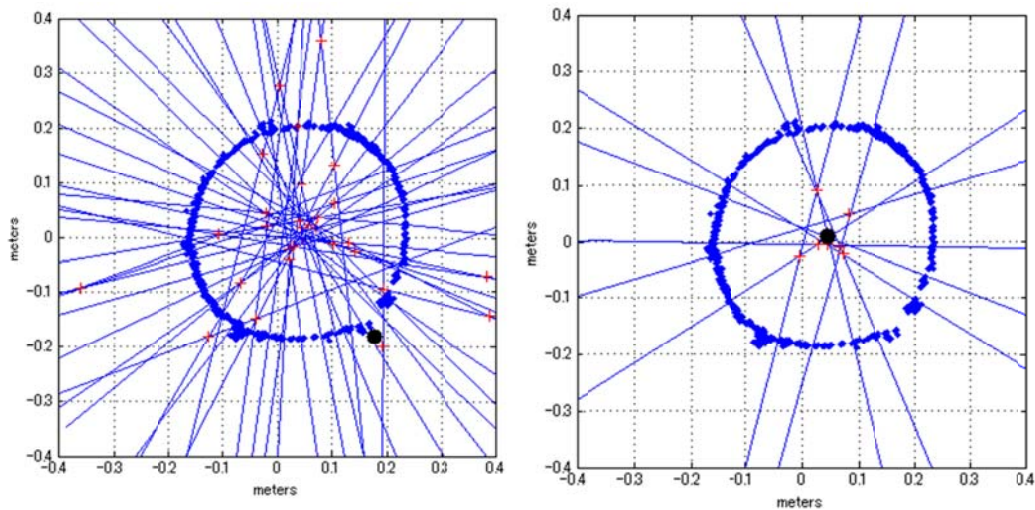


Figure 3. DTM created by two sensor locations 50m apart from each other. An initial surface created only by the minimum height (left) and the final product produced from the iterative process (right).

3. Results and discussion

3.1 Creating DTM

A DTM was generated using the process described above. The final product shown in Figure 3 is smoothed and has no holes. Owing to the high point density generated by the two sensor locations only 50m apart it was possible to generate a continuous DTM from the terrestrial lidar without having to resort to Triangular Irregular Network (TIN).



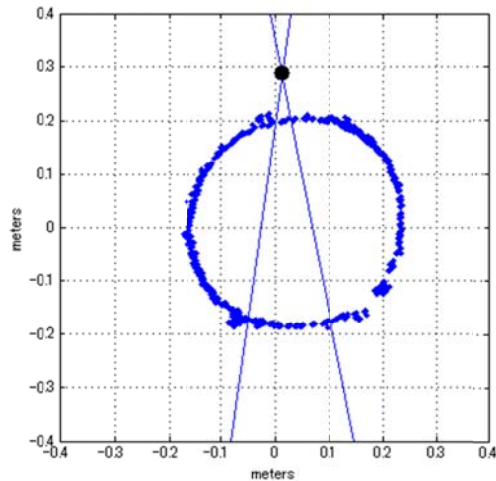


Figure 4. Location of approximated centres in 10 degrees (upper left), 40 degrees (upper right), and 180 degrees (bottom). Blue dots represent the lidar points of 90% coverage sample stem. Blue lines indicate the normal vectors of sampled points with variable degree of the interval. Red crosses indicate the intersected location of normal vectors. A black dot indicates the approximated centre. (0,0) means the average coordinates initially given by the sampled points.

3.2 Approximating the centre of the stem

As an example, points of a stem which had full coverage (360 degrees coverage) were extracted and used for a test to find the centre location. Variable degrees intervals (10, 20, 30, ... up to 180 degrees interval) were applied to determine the normal vector to approximate the centre. The conventional measurement of stem diameter in the field was given by the circumference of stem divided by pi. But the stem measurement from lidar is measured by the average distance between the approximated centre and the inner side of stem points.

Figure 4 shows the normal vector derived from the samples of a fine interval like 10 degrees were influenced by the local variability of points and that the coarse interval like 180 degrees get the approximated centre outside the circle. It is confirmed that the range from 40 to 130 degrees is optimum to get the stem centre.

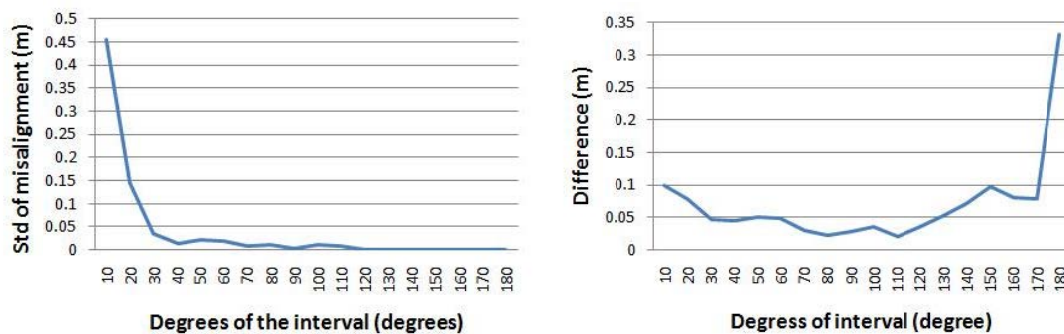


Figure 5. Standard deviation of intersected points (misalignment error) (left) and the difference from the average coordinate temporarily set from the data (right) in variable degrees of the sample interval.

The interval degrees of sampled points to determine the normal vector with the least error was more than 40 degrees (Figure 5) and the misalignment of approximated centres was getting closer from 30 degrees to 130 degrees (Figure 5). This result shows that at least 40 degrees are required to get good approximation of the centre.

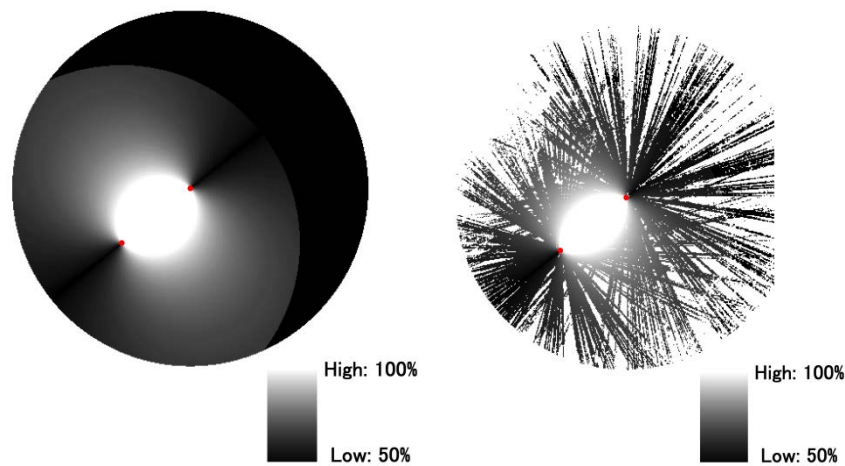


Figure 6. Simulation result to show scan coverage from two sensors without any object (left) and the ideal visible angle is overlaid on the actual scanned map (right). Red points represent the sensor locations.

The white area in Figure 6 corresponds to the zone of full scan coverage (360 degrees coverage, white area in Figure 6). The full scan coverage area is especially located between the sensors.

3.3 Scan angle and scan coverage

The visible side of stems were identified by the simulation based on the sensor locations. The simulation was conducted without any objects (Figure 6) and the range was set from 50% (half obscured) to 100% (360 degree of coverage). The actual scanned image was overlaid on the top of the simulation result to assure the visible side of the angle for the stem location.

3.4 Stem diameter measurement

The relationship between measurement error and scan coverage is shown in Figure 7 (left). The measurement error is the difference between lidar diameter and manually measured diameter. The scan coverage of each of the 42 sampled trees is their location relative to the sensors. Figure 7 (left) shows that when scan coverage exceeds 40 % trees tend to have errors below 5 cm. This indicates that scan coverage of more than 40 % can produce more accurate stem diameter. The correlation between manual and lidar diameter measurements is high ($R^2=0.73$) and the correlation line nearly coincides with the 1:1 line (Figure 7, right). This shows that the lidar stem diameter measurements are unbiased.

3.5 Assurance of scan angle to measure stem diameters

Section 3.2 showed the ideal coverage of scan angle and section 3.3 showed that the coverage angle of more than 40% can produce the stem diameter with the error of 5 cm or less. In Figure 6, the area covered by two sensors has at least 50% coverage in this study site. It was concluded that all trees in this area are qualified to obtain accurate stem diameter from terrestrial lidar sensors set 50 m apart.

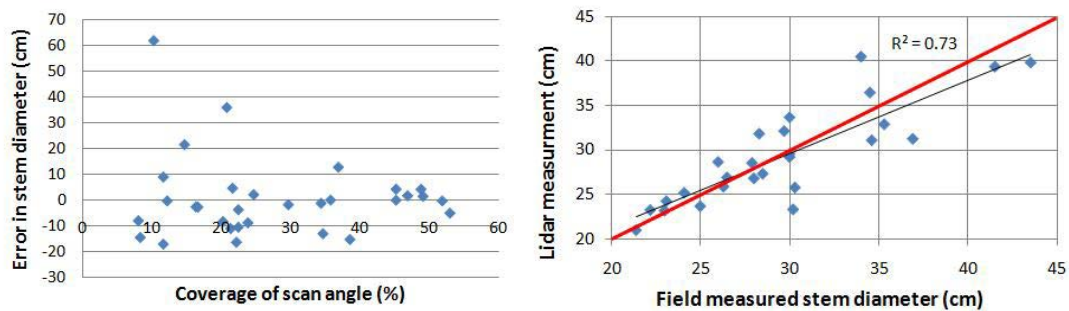


Figure 7. Error of stem diameter in different coverage of scan angle (left). The comparison between lidar and field measurement. Red line represents one to one correlation.

There is not so much understory in this study site but further research is needed to find out the influences from understory to get enough point density to measure stem diameter. In a follow-up study a site with understory should be used to quantify the scan effect for stem measurement.

The technique developed for this study was only used for stem diameter at breast height but the same technique can be applied to any height of stem diameter. For the quality assessment of wood, the straightness of stem can be estimated directly from the alignment of stem centres at several heights using the applied technique. Moreover, stem volume which is the most important variable to estimate the amount of carbon is also directly measured by this technique.

Acknowledgements

This research was supported by the Environment Research and Technology Development Fund (RF-1006) of the Ministry of the Environment, Japan. We thank Fuji Architect Co., Ltd and Riegl Inc., Japan for collecting field data of this research.

References

- Andersen, H. E., Reutenbuch, S. E., & McGaughey, R. J., 2006. A rigorous assessment of tree height measurements obtained using airborne LIDAR and conventional field methods. *Canadian Journal of Remote Sensing*, 32, 355–366.
- Chasmer, C., Hopkinson, C., and Treitz, P., 2006. Investing laser pulse penetration through a conifer canopy by integrating airborne and terrestrial lidar, *Canadian Journal of Remote Sensing*, 32(2), 116-125.
- Côté J-F, Widlowski, J-L., Fournier, R.A., and Verstraete, M.M., 2009. The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial lidar. *Remote Sensing of Environment*, 113, 1067-1081.
- Eitel, J.U.H., Vierling, L.A., and Long, D.S. 2010. Simultaneous measurements of plant structure and chlorophyll content in broadleaf saplings with a terrestrial laser scanner. *Remote Sensing of Environment*, 114, 2229-2237.
- Hopkinson, C., Chasmer, L., Young-Pow, C., and Treitz P., 2004. Assessing forest metrics with a ground-based scanning data, *Canadian Journal of Remote Sensing*, 34 (3), 573-583
- Hyypä, J., Kelle, O., Lehtikoinen, M., & Inkinen, M. 2001 A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 969–975.

- Kato, A., Moskal, L., Schiess, P., Swanson, M., Calhoun, D., & Stuetzle, W. 2009. Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. *Remote Sensing of Environment*, 113, 1148–1162.
- Næsset, E. & Økrand, T., 2002. Estimating tree heights and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment*, 79, 105–115.
- Patenaude G., Milne R., and Dawson T.P., 2003. Synthesis of remote sensing approaches for forest carbon estimation: reporting to the Kyoto Protocol. *Environmental Science & Policy* 8, 161-178.
- Tansey, K., Selmes, N., Anstee, A., Tate, N.J., and Denniss, A. 2009. Estimating tree and stand variables in a Corsican Pine woodland from terrestrial laser scanner data. *International Journal of Remote Sensing*, 19 (10), 5195-5209.

Stem curve measurement using terrestrial laser scanning

X. Liang¹, J. Hyypä¹, V. Kankare² & M. Holopainen²

¹Finnish Geodetic Institute, Department of Remote Sensing and Photogrammetry,
Masala, Finland

(xinlian.liang, juha.hyypa)[@fgi.fi](mailto:)

²University of Helsinki, Department of Forest Resource Management, Finland
(ville.kankare, markus.holopainen)[@helsinki.fi](mailto:)

Abstract

Terrestrial laser scanning (TLS) has been shown to be a promising technology for the accurate forest inventory on the sample plots. The advantages of applying TLS can be improving the accuracy and efficiency of the field measurements. In addition, TLS data have the possibility to provide more tree parameters than what are commonly accepted and employed at the moment. This paper discusses the automatic measurement of the stem curve using TLS. A pine and a spruce were used in the experiment. The stem curve estimated from point cloud was compared to the field measurements. The experiment shows that the estimation of the stem curve from single-scan and merged point clouds are comparable to each other. This result indicates that TLS data has the potential to automatically estimate the stem curve.

Key words: forest inventories, stem curve, terrestrial laser scanning

1. Introduction

The stem curve is an important parameter in many forest applications. Stem curve models can be used to determine saw-wood and pulpwood proportions (Holopainen et al., 2010). The knowledge of the stem curve is often the basis, directly or indirectly, for the forest management, such as harvesting and thinning.

The conventional measuring tools, e.g., callipers, measuring tapes, and hypsometers, are however not suitable for measuring the stem curves of living trees. The exact stem curve can be measured in a destructive way. The tree is cut down and the diameter along the stem is measured, e.g., using callipers. The harvester also records the stem curve. The diameter measurements for the commercial part of the stem at intervals of 10cm are included in the logging machine data.

Alternatively, stem curve models (e.g., Laasasenaho, 1982), were developed to predict stem curves using fixed measurements, usually tree species, the diameter at breast height (DBH) and the total height. If there are several measurements on each stem, splines can be used to interpolate stem dimensions between measured points.

Some specific laser instruments were developed to measure DBH in the field, such as a laser-relascope (Kalliovirta et al. 2005) and a laser camera (Melkas et al. 2008, Vastaranta et al. 2009). The laser-relascope is composed of a laser instrument, a variable-width slot with a fixed length arm, an electronic compass, an electronic inclinometer, and the supporting unit. The DBH is measured by combining the range and angle data. The laser-camera integrates a digital camera and a laser line generator. The stem diameter is measured using the length and relative position of the laser line in the image. These instruments can be used to collect the diameters at different heights; therefore have the potential to measure the stem curve of living trees. However, the measurement becomes less accurate, or impossible, when there are obstacles, e.g., the branches, between the instrument and the stem.

Terrestrial laser scanning (TLS) has been proved to be a promising solution to collect tree parameters (e.g. Thies et al., 2004). The data collected by TLS are high-precision three-dimensional (3D) measurements of the targets in the form of a point cloud. This technique makes automated, non-invasive and expedient field mensuration possible (Hopkinson et al., 2004).

The estimation of the stem curve using TLS has not been studied in detail. A pilot work was reported in (Maas et al., 2008). One Stika spruce in the single-scan data was used in the stem profile determination. The Root Mean Squared (RMS) error of 1.0cm in the best fit part and 4.7cm overall was reported. More studies are needed to understand the estimation of the stem curve of living trees using TLS. As far as we know, this paper is one of the first detailed studies to use single and merged point cloud to estimate the stem curve.

2. Method

2.1 Data acquisition

The study trees were on two sample plots in Evo, Finland (61.19°N, 25.11°E). The plots were scanned in the multi-scan mode in spring 2010. The scans were co-registered using reference balls put on the plot. One Norway spruce and one Scots pine were employed in this study. The merged point cloud of the individual tree was from three nearest scans. The single-scan data was one of the three scans, where the visibility of the stem is good. The registration and the data selection were done using the Leica Cyclone software.

The TLS data were collected using a Leica HDS6100 terrestrial laser scanner (Leica Geosystems AG, Heerbrugg, Switzerland). The scanner uses phase-shift measurements of continuous waves to measure distances and its maximum data acquisition rate is 508,000 points per second.

Fig. 1 and 2 show the trees in point clouds. Fig. 1(a) and (b) are the spruce in the merged point cloud and the single-scan data, respectively. Fig. 2(a) and (b) are those for the pine. Both trees have a lot of branches. They are present at both lower and upper part of the tree as regard to the spruce, and more concentrated at the upper part as regard to the pine.

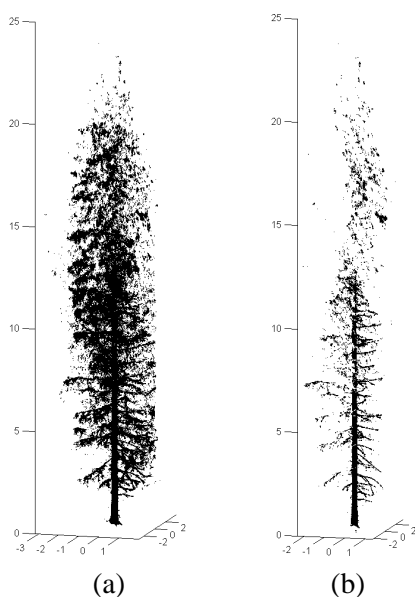


Fig. 1 The spruce in the point cloud

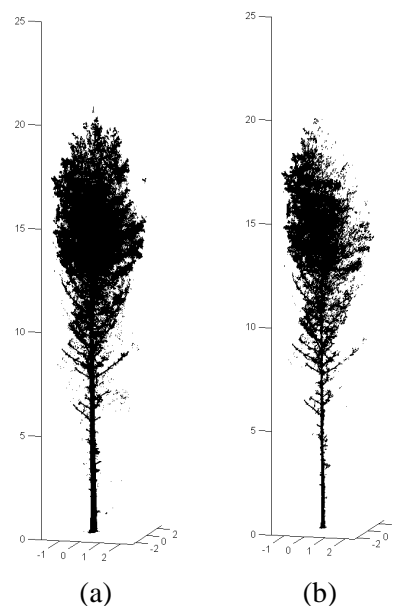


Fig. 2 The pine in the point cloud

The field measurements were collected in summer 2010. The trees were cut down and trimmed. The stem was cut into logs. The first cut was at the stump height, the second at the middle of the stump and the breast height (1.3m), and the third at the breast height. The cuts were done at every meter starting from 2m. The stem diameter was measured using steel callipers at the bottom of the log.

2.2 Method

2.2.1 Data sampling

The distribution of points in the laser scanning data is not even. In general, the further away an object is from the scanning position, the larger is the distance between nearby points. As regard to a single tree in the point cloud, the distance of points on the stem decreases as the height increases. The original data were sampled in the Z direction to reduce the total amount of the data and computing time. Every n of the original points were sampled. The factor n was experimentally determined. It depended on the properties of the data and the object. The factor was selected also in the way that the data from the upper part of the stem are less sampled than those from the lower part. Equation 1 shows the calculation of the factor n . Z is the z-coordinate in meter.

$$n = \begin{cases} 12 - \text{floor}(Z/2) * 3 & Z < 6 \\ 2 & Z \geq 6 \end{cases} \quad (1)$$

2.2.2 Stem modelling

The stem points were identified by points spatial properties and the 3D stem model. The spatial properties were studied in points neighbourhood. A local coordinate system was set up for each point using the eigenvalue decomposition. The axes directions in the system were defined by the eigenvectors. A point was possibly on a stem if it was of low variance along one direction in the local coordinate system and had a close-to-horizontal normal vector in the real world coordinate system. The selected points were grouped. A point was in a group if there was at least one point from the same group within a certain distance. The group with the largest amount of points was selected as the stem.

Stem model was built to the points. The model was a series of 3D right circular cylinders along the stem profile. In each model element, a cylinder was matched to a set of points by means of the robust estimation, where points were weighted to reduce the influence of cross errors, e.g., the branch points. After the first cylinder was built, the parameters of current cylinder were used as the initial estimation for the next one, along the axis and overlapped with the current one. For more details on the robust modelling procedure, readers are referred to (Liang et. al., 2011).

2.2.3 Stem curve estimation

The stem curve was estimated from the 3D stem model. The model elements were selected at the heights where the reference data were measured. The heights were 1.3m and every meter between 2m and 12m. The lowest point in the merged point cloud was used to define the ground level in the scanner coordinate system. The diameters from the corresponding model elements were employed as the diameter estimations.

3. Result

The estimations of the stem curve of the spruce with the merged point cloud and the single-scan are showed in Fig. 3 and 4, respectively. In Fig. 3(a), the solid line is the estimation of the stem curve by means of the TLS data and the automatic method; the dashed line is the reference data recorded in the field. Fig. 3(b) shows the error of the estimation. The error was calculated by subtracting the reference from the estimation at 1.3m and every meter between 2m and 12m. The dashed line shows the range of ± 1 cm. The RMS error is 0.6cm and 0.6cm using the merged

and single-scan data. Fig. 4, 5 and 6 are plotted in the similar way as the Fig. 3.

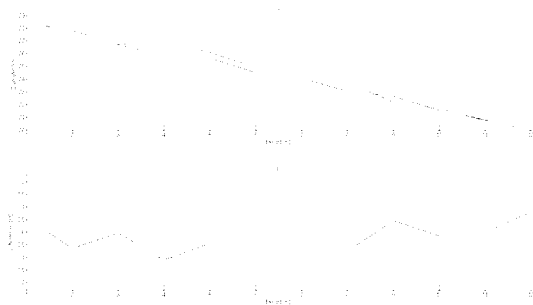


Fig. 3 The stem curve estimation of the spruce using the merged point cloud

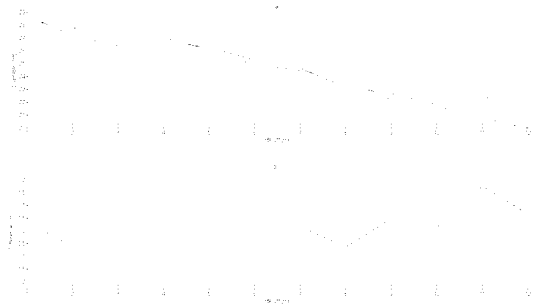


Fig. 4 The stem curve estimation of the spruce using the single-scan point cloud

Fig. 5 shows the estimation of the pine using the merged point cloud. Fig. 6 is that using the single-scan data. The RMS error of the stem curve estimation of the pine is 1.3cm and 1.8cm with the merged and single-scan data, respectively.

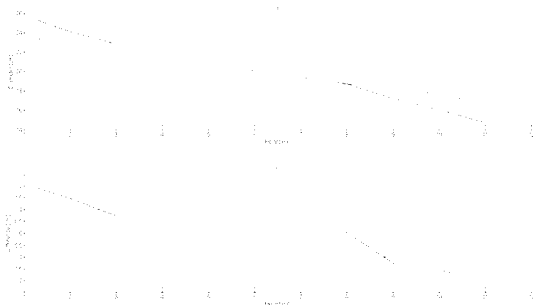


Fig. 5 The stem curve estimation of the pine using the merged point cloud

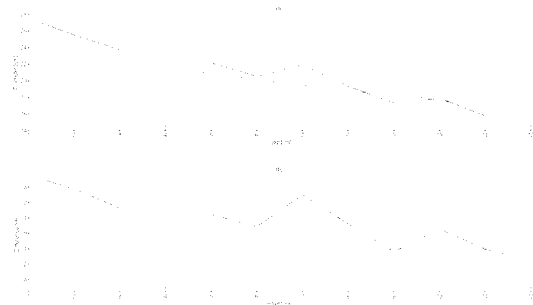


Fig. 6 The stem curve estimation of the pine using the single-scan point cloud

Fig. 7 shows the point distribution of the pine in the XY plane. Points are around 12m above the ground and in the merged point cloud. Fig. 7(a) shows the original point cloud around the stem position. Fig. 7(b) is the detected stem points. More discussion is in section 4.

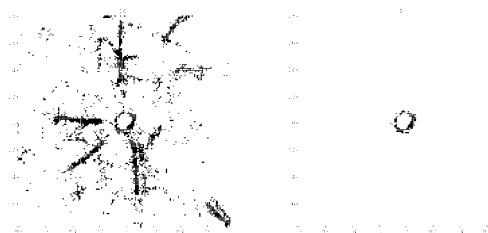


Fig. 7 The point distribution at 12m

4. Discussion

By referring to diameter, it is being implied that stems are circular in cross section. However, the first problem with measuring stem diameter is that tree stems are never exactly circular (West, 2009). The irregularities are generally present at the place where the branches grow or external defects are present, such as knots and bulges.

In TLS data, the stem diameter was automatically estimated by means of 3D modelling. The direction and diameter of the stem are estimated at the same time. The stem diameter was

therefore measured perpendicular to the stem axis, as the field measurement required. The estimation error, however, can be expected because of the irregularities of the cross section and the approximation of the cylinder shape. Given the stem axis, other geometrical primitives can be used to reduce such estimation error. The free-form curve (e.g., Pfeifer and Winterhalder, 2004) has been reported to describe the details of the shape of the stem section in two-dimensional(2D) space. It is possible to improve the estimation of the stem diameter by combining the 3D cylinder and 2D curve fitting. The original points can be first projected to the 3D plane perpendicular to the cylinder, or the stem, axis. The 2D curve is used to fit the points on the plane. Further researches are needed to study the applicability of the combination of the 3D and 2D fitting to reduce the estimation error.

The estimation error can also be introduced by the diameter changes along the stem axis. This is more likely present at the upper part of the stem where the stem diameter decreases rapidly. In such case, the cylinder gives an average value of the diameters along the axis. The diameter of the cylinder can be larger or smaller than the reference, depending on the position the reference data is collected, Fig. 7(b) shows the stem point at the height of 12m. The points were scattered on the plane. It is likely because of the diameter changes along the Z direction.

The accuracy of the stem curve estimation becomes lower in the upper part of the stem. It is influenced by many factors, e.g. the wind and the occlusion effect. The wind, for example, may move clearly the top of the tree around its position. In the multi-scan mode, several scans were made at different positions, usually in turn. It is possible that the stem surface is present at different positions in several scans. In that case, the estimation can be either larger or smaller than the reference. The occlusion effect is typically heavy in the upper part of the stem. Less laser points are reflected from the stem when a lot of branches are present between the scanner and the stem. The influence of the cross error, e.g., crown points, thus becomes more significant when less points are from the stem surface. However, it is worth to note that the influence of the large amount of branches is significant when the occlusion becomes really heavy. The estimation is still accurate for the spruce when a lot of branches are present in the middle and lower part of the tree.

The registration error may also introduce some estimation errors. The registration is done typically using reference balls put on the ground. It can be expected that the registration accuracy becomes lower when the distance to the reference balls growing. Likely, the registration error is larger at the upper part than the low part of the tree. It is not clear yet the distribution of the registration error in the merged data. At the moment, no evidence shows that this error is large.

It is worth to note that the measurement error is also present in the reference data. The measurement of the stem diameter in the field depends on the positions where the measurements are made. For example, it is possible that the difference between major and minor axes is not apparent, and only one measurement is made. It is also possible that the diameter is measured at other heights but recorded as the measurement at the regular height in the field measurement. For example, the stem may thicken because of dead branch or other damage. In that case, the diameter should be measured at a higher position.

The accuracy of the estimation of the spruce is higher than that of the pine. The bark texture may have some influences on that. More experiments are needed for further analyses. The estimation from the single-scan data is less accurate than that got from the merged point cloud. In the single-scan data, the laser point cloud covers 50 percent, at best, of the object surface. It is more difficult and therefore likely less accurate to build the stem model. However, the accuracies from two data sets are close to each other, if the visibility of the stem is fairly good in the single-scan data. This results shows that it is possible to use the single-scan point cloud in

the stem parameter estimation. Further researches are needed to improve the estimation method, to analysis the estimation error, and to understand the effects of merged and single-scan data.

5. Conclusion

The stem curve estimation using the TLS data and automatic method was reported and discussed. The result shows that the TLS data and automatic method can be used to estimate the stem curve on the sample plot. It indicates that the multi-scan TLS data is suitable for the estimation of the stem curve; and the single-scan data can be also used in the estimation of stem curve if the visibility of the stem is rather good.

References

- Holopainen, M., Vastaranta, M., Rasinmäki, J., Kalliovirta, J., Mäkinen, A., Haapanen, R., Melkas, T., Yu, X. and Hyypä, J. 2010. Uncertainty in timber assortment estimates predicted from forest inventory data. *European Journal of Forest Research*, 129(6), 1131-1142.
- Hopkinson, C., Chasmer, L., Young-Pow, C. and Treitz, P., 2004. Assessing forest metrics with a ground-based scanning lidar. *Canada Journal of Remote Sensing*, 34(3), 573–583.
- Kalliovirta, J., Laasasenaho, J. and Kangas, A., 2005. Evaluation of the Laser-relascope. *Forest Ecology and Management*, 204(2-3), 181-194.
- Maas, H.-G., Bienert, A., Scheller, S. and Keane, E., 2008. Automatic forest inventory parameter determination from terrestrial laser scanner data. *International Journal of Remote Sensing*, 29(5), 1579-1593.
- Melkas, T., Vastaranta, M. and Holopainen, M., 2008. Accuracy and efficiency of the laser-camera. In *Proc. SilviLaser 2008, Edinburgh, U.K.*, 315-324.
- Laasasenaho, J., 1982. Seloste: Männyn, kuusen ja koivun runkokäyrä- ja tilavuussyhtälöt (Taper curve and volume functions for pine, spruce and birch). *Communicationes. Institute Forestalis Fenniae*. 108, 74 p. Doctoral thesis.
- Liang, X., Litkey, P., Hyypä, J., Kaartinen, H., Vastaranta, M. and Holopainen, M., 2011. Automatic stem mapping using single-scan terrestrial laser scanning. *IEEE Transactions of Geoscience and Remote Sensing*, In Press.
- Pfeifer, N., and Winterhalder, D., 2004. Modelling of tree cross sections from terrestrial laser-scanning data with free-form curves. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(8/W2), 76-81.
- Thies, M., Pfeifer, N., Winterhalder, D., Gorte, B. G. H., 2004. Three-dimensional reconstruction of stems for assessment of taper, sweep and lean based on laser scanning of standing trees. *Scandinavian Journal of Forest Research*, 6(19), 571 – 581.
- West, P. W., 2009. *Tree and Forest Measurement* (2nd edition). Springer-Verlag, Berlin Heidelberg New York.
- Vastaranta, M., Melkas, T., Holopainen, M., Kaartinen, H., Hyypä, J. and Hyypä, H., 2009. Laser-based field measurements in tree-level forest data acquisition. *Photogrammetric Journal of Finland*, 2009(1), 51-61.

Estimating single-tree branch biomass of Norway spruce by airborne laser scanning

Marius Hauglin¹, Janka Dibdiakova², Terje Gobakken³, Erik Næsset⁴

¹ Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences. email: marius.hauglin@umb.no.

² The Norwegian Forest and Landscape Institute, email: janka.dibdiakova@skogoglandskap.no

³ Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences. email: terje.gobakken@umb.no.

⁴ Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences. email: erik.naesset@umb.no.

Abstract

Dry weight of the branches of 20 trees of Norway spruce was obtained through destructive sampling. Airborne laser scanning data from the same trees were used to calculate crown volume for each tree. The crown volume was derived by using the crown laser echoes with a radial basis function to construct a crown surface. A regression model was fitted to the data, with the crown volume as explanatory variable and the dry weight of the branches as response. The model revealed a strong relationship between the two, with $R^2 = 0.80$. A leave-one-out cross-validation gave a root mean square error of 34%.

Keywords: Airborne laser scanning, biomass, crown volume, bioenergy

1. Introduction

The last ten years has seen an increased interest in the use of biomass for energy purposes. Biomass from forests will most likely be one of several sources of energy that will replace fossil fuels in the future. One obvious example is the utilization of logging residues, biomass that would otherwise have been left in the forest during the logging. The branch biomass constitutes a considerable part of the logging residues. When logging residues become a commercial product from the forest, this resource should be quantified as part of the forest inventory to improve planning of extraction for energy purposes. An increasing part of forest inventories are based on data collected with airborne laser scanning (ALS). While many ALS based operational forest inventories are using the area-based approach as described by e.g. Næsset (2002), also methods targeting single trees have been proposed (Hyypä et al. 2001; Persson et al. 2002; S. Solberg et al. 2006; Wang et al. 2008). The latter methods usually require ALS data with higher resolution, but intend to give information on a single tree level, contrary to the per area information provided by the former. Although not as widely used at the moment, inventory methods targeting individual trees might in the future be more used, depending on the ongoing technological and methodological research and development and future costs for data acquisition. The potential of estimating individual tree characteristics by ALS has been investigated in several studies, including stem volume (Straub & Koch 2011), stem diameter (Popescu 2007), crown base height (Vauhkonen 2010), leaf area index (Roberts et al. 2005) and biomass (Popescu 2007).

When a tree is scanned by an airborne laser scanner, a majority of the laser pulses will echo from the crown, i.e. the branches. This suggests that the ALS data contain useful information on the crown biomass. In fact, much of the information inherent in the ALS data will be directly related to the tree crown and the branches. In a previous study on single-tree biomass estimation from ALS (Popescu 2007) a strong relationship between the ALS data and the branch biomass was reported. The actual branch biomass of each individual tree was however not measured, but obtained through allometric equations with field measured tree height (h) and diameter at breast height (d_{bh}) as explanatory variables. A direct relationship between ALS data and branch biomass were therefore not established.

Single-tree predictor variables derived from the ALS point cloud such as height percentiles and crown diameter have been used by e.g. Popescu et al. (2003) and Straub & Koch (2011). Kato et al. (2009) presents a method for crown surface reconstruction that enables the calculation of crown volume. An intuitive assumption is that this crown volume could be a good predictor variable for estimation of branch biomass.

To our knowledge no previous studies relates ALS data directly to accurate measurements of branch biomass (i.e. obtained with destructive sampling) at a single tree level. The first aim of the present study was therefore to assess the accuracy of ALS based predictive models for single-tree branch biomass of Norway spruce (*Picea abies* (L.) Karst.) using ground measurements of branch biomass. The second aim was to assess the suitability of using an ALS derived crown volume as a predictor variable for branch biomass. This variable was chosen based on the above mentioned assumption, and promising results from a pre-study comparison of some ALS derived variables (not included in the present study).

2. Materials and methods

2.1 Study area

The study area was Aurskog-Høland municipality (59°50'N 11°30'E, 120-390 m above sea level) located in the south-eastern part of Norway. The total area of Aurskog-Høland is 96,000 ha with 67,000 ha productive forest. The forest type is boreal with Norway spruce and Scots pine (*Pinus sylvestris* L.) as the dominant tree species.

2.2 Field data

Field data were collected in June 2009. Two locations were chosen, and from each location 10 trees were selected. The two locations were chosen from potential locations in the intersections of the two east-west strips of already existing ALS data (see section 2.3) and forest roads. In order to avoid edge effects from the forest roads, trees with a distance >10 m to the forest road were preferred. Finally, due to practical reasons, trees with a distance >30 m from the road were not selected.

On all the 20 selected trees the crown projection was measured in the eight cardinal and intercardinal directions. The measurements were carried out with a measuring tape and a compass. The horizontal distance from the stem at breast height to the vertical projection of the branch tip in the given direction was recorded. For all trees d_{bh} was measured with a calliper.

The 20 trees were then felled, and the raw weight of the branches (including needles) of each tree was obtained by weighing the tree before and after the branches were cut off. The weighing was done with a mobile lift mounted on a truck. A Teraoka Seiko OCS-XZL digital scale with load capacity 3000 kg

was used. Samples of entire branches were selected among the living branches of each tree in order to determine the dry weight. In total 11 living branches were taken from each tree, i.e., three branches from the lower part of the crown, three from the middle part and three from the upper part of the crown. In addition two branches were taken from the top of the stem ($d_{bh} < 5$ cm). From all the sampled branches there were taken three sub-samples which were dried and the raw and dry weight of each sub-sample was recorded. For each tree h was measured with a measuring tape after the felling.

The coordinates of each tree were obtained in a two-step procedure: (1) The location of each tree relative to two local reference points was accurately measured with a total station, and (2) the coordinates of the two reference points were obtained by differential Global Navigation Satellite Systems (dGNSS), using dual-frequency receivers observing pseudo-range and carrier phase of the Global Positioning System and the Russian Global Navigation Satellite System. Hasegawa & Yoshimura (2003) found horizontal positional errors in the range of 1 – 30 cm in dGNSS-measurements under conditions comparable to those in the present study.

Characteristics of the 20 trees are summarized in Table 1.

Table 1: Summarized characteristics of the 20 trees in the data material. Field measured d_{bh} , h and dry weight of the branches (BR_{dw}).

	d_{bh} (cm)	h (m)	BR_{dw} (kg)
min	11.2	8.2	8
max	39.8	26.3	156
mean	23.6	19.6	69

2.3 ALS data

ALS data were collected along two strips in east-west direction in the study area. The strips were flown 9 km apart in the north-south direction.

The dataset was collected in June 2006 with an Optech ALTM 3100 sensor on a fixed-wing aircraft. The average flying altitude was 800 meter above ground, the pulse repetition frequency was 100 kHz, the scan frequency 70 Hz and the maximum scan angle was ± 5 degrees from nadir. This gave an average point density on the ground of 7-10 m^{-2} .

Classification of echoes into ground- and vegetation echoes was carried out by the contractor with the TerraScan software (Anon. 2011). The contractor also determined the planimetric coordinates and ellipsoidal height values for all echoes. Echoes classified as ground were used to construct a triangulated irregular network (TIN) terrain model. The height above ground was calculated for all echoes by subtracting the respective TIN heights from the ellipsoidal heights.

First and last recorded echoes were used in the present study.

2.3.1 Single tree segmentation

Several methods for automatic delineation of the ALS point cloud into single tree segments have been proposed. However, automatic segmentation will always omit some trees (omission errors) and include false trees (commission errors) (Vauhkonen et al. 2011). To avoid errors introduced by automatic segmentation, we decided to use the field measured crown projection for selection of echoes

that could be assigned to each individual tree. An eight-sided polygon was formed from the crown projection measurements of each tree, and all echoes within the polygon were assigned to that tree.

2.4 Calculations

2.4.1 Dry weight biomass of the branches

A raw to dry weight ratio was calculated for each branch sub-sample. For each tree a raw to dry weight ratio was calculated as the mean of the ratios obtained from the samples. Finally, the total dry weight biomass of the branches for each tree was calculated as the raw weight of the branches multiplied with the calculated tree-specific raw to dry weight ratio. This is denoted BR_{dw} in the rest of this paper.

2.4.2 ALS derived crown base height¹

The crown base height was estimated from the height of the laser echoes in each tree. A simple procedure was applied, where only the echoes below the median height were considered. These echoes were sorted according to their height above ground, and the echo with the largest vertical distance to the next echo below was set to be the lowermost echo in the crown, and hence the crown base height (Fig. 1). Laser echoes above the crown base height were considered to be crown echoes.

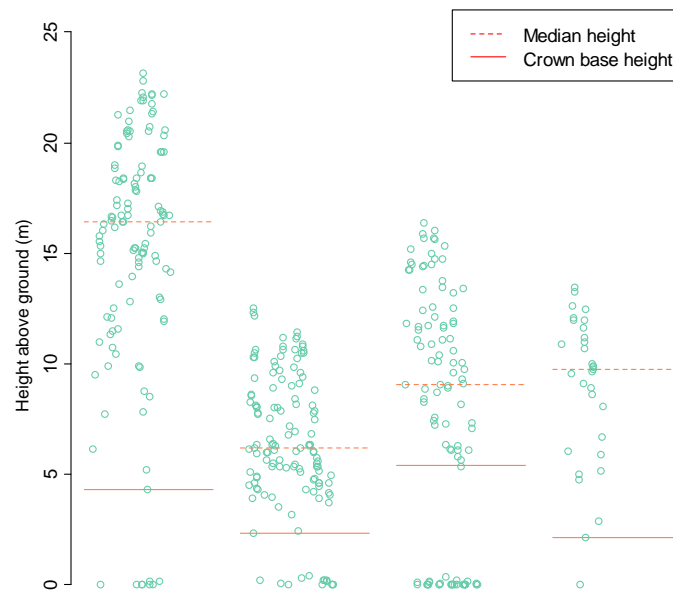


Figure 1: Estimated crown base height for four of the trees in the data material.

¹ Contrary to definitions of crown base height used by e.g. Maltamo et al. (2010) where the base height refers to the height on the tree trunk where the lowermost branches are found, we wanted in this study the crown base height to correspond to the height of the lowermost point on any of the branches in the live crown.

2.4.3 Radial Basis Function derived crown volume

Crown volume was calculated from the crown echoes for each tree by using a radial basis function (RBF) as proposed by Kato et al. (2009). The method was modified to work with less dense ALS data, and the calculations can briefly be summarized as follows:

Points on the surface of the crown to be reconstructed were chosen by calculating a convex hull in the x,y-plane for the echoes in a number of individual height bins. The echoes at the border of the convex hull were marked as surface points, all others as being inside the crown. The topmost and lowermost echoes in the point cloud were always marked as being on the crown surface.

For each surface point two off-surface points were created (see Carr et al. (2001) for details) in a given distance d from the surface point. An RBF can in this case be written as

$$f(x) = \sum_{i=1}^N \lambda_i (|x - x_i|), \quad (1)$$

where x is a point $\in \mathbb{R}^3$, N is the number of all the surface and off-surface points and $|x-x_i|$ is the Euclidian distance. The λ_i is a weight parameter, computed with the distance values d and the surface and off-surface points. When the λ s are determined the RBF in Equation 1 can be evaluated for any given point $x \in \mathbb{R}^3$, and it has the property that $f(x)$ will be zero at the crown surface. A crown surface was approximated by evaluating Equation 1 for values of x in a three-dimensional grid, and constructing a triangulated mesh surface where $f(x) = 0$ (visualized in Fig. 2). The crown volume V_{cr} for each tree was then calculated from the triangulated mesh crown surface by

$$V_{cr} = \sum_{i=1}^M V(t_i), \quad (2)$$

where M is the number of triangles in the mesh, V is a signed volume function and t_i is the tetrahedron formed by the i th triangle and an arbitrary point p . The facing of the i th triangle relative to p , determines the sign of V .

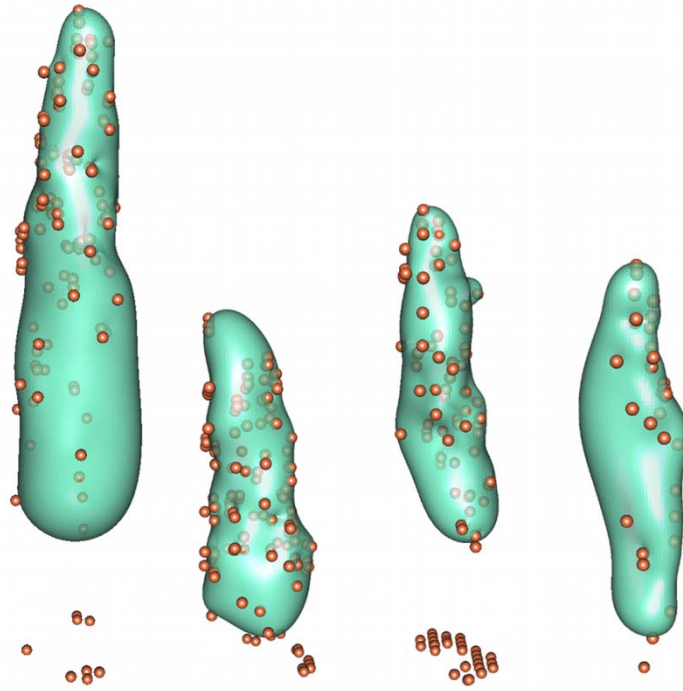


Figure 2: The RBF crown surface for four of the trees in the data material.

2.4.4 Model fitting and validation

A linear regression model with V_{cr} as explanatory variable and BR_{dw} as response variable was fitted to the data. A leave-one-out cross-validation was carried out by leaving out one observation at the time, re-fitting the model with the remaining trees and predicting the biomass for the single tree left out.

2.4.5 Biomass by existing allometric equations

For comparison reasons the biomass of the branches for the 20 trees were also estimated by existing allometric equations with field measured d_{bh} and h as explanatory variables (Marklund 1988).

3. Results

The Pearson's correlation coefficient between the biomass of the branches and the ALS derived crown volume was 0.892, which means that the ALS crown volume explained approximately 80% of the variation in branch biomass in the regression model (Table 2). The leave-one-out cross-validation gave a root mean square error (RMSE) of 34%.

When relating the BR_{dw} to the branch biomass estimated with existing allometric equations with field measured d_{bh} and h as explanatory variables the correlation coefficient was 0.866. The RMSE for these estimates was 31%. Plots of the residuals (Fig. 3) show that the accuracy of the field based estimates from the existing allometric equations is slightly better than the accuracy obtained in the cross-validation of the ALS based regression model.

Table 2: Estimated regression model and associated statistics. RMSE from cross-validation of the model.

Response variable	n	Predictive model	R ²	Observed mean (kg)	RMSE (kg)	RMSE (%)
BR_{dw}	20	$14.765 + 0.562 * V_{cr}$	0.80	69	23	34

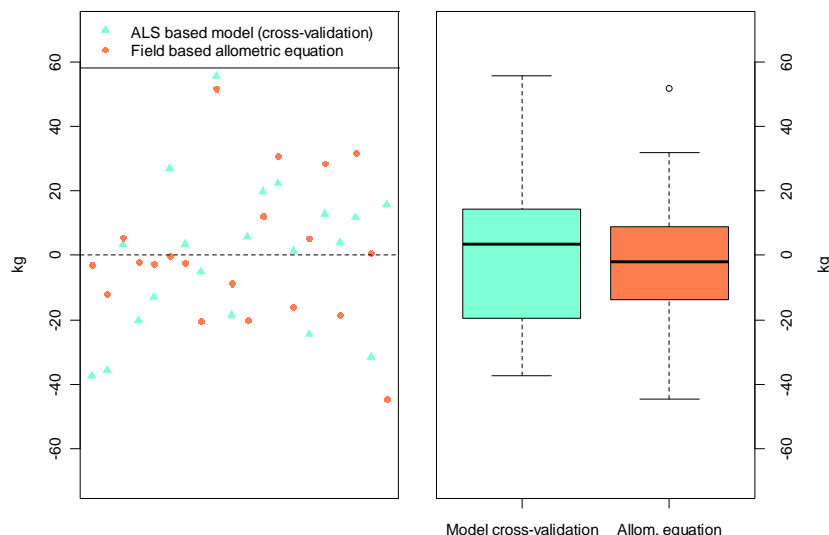


Figure 3: Plots of the residuals from the regression model cross-validation and from the branch biomass estimates based on existing allometric equations. The residuals in the left plot are plotted in increasing order by tree size (d_{bh}).

4. Discussion and conclusions

A strong relationship was found between the ALS data and the biomass of the branches. The accuracy of the ALS based model predictions is almost comparable to the accuracy obtained with allometric equations based on field measurements. This biomass estimate, derived with the existing allometric equation, is in fact equal to a theoretical ‘best case’ biomass estimate following the procedure described by Popescu (2007), assuming that d_{bh} and h can be perfectly derived from the ALS data. In a real situation this will not be the case, and errors introduced when predicting d_{bh} and h from ALS data will directly affect the biomass estimates derived this way. With this in mind the comparison suggests that, in practice, a comparable accuracy can be obtained by predicting the branch biomass directly from the ALS data, without using a model chain with an independent allometric equation. The main challenge with such an approach is however the dependency on field measurements as training data. Destructive sampling is obviously not an option in an operational setting, so other methods to estimate the ground truth values of branch biomass should be explored. Terrestrial laser scanning has emerged as a possible tool for this purpose, but more research is needed on this issue.

Only one ALS derived variable was considered in this study, and including other ALS derived predictor variables will most likely improve the results. The properties of the RBF crown volume could be explored, and also how these are related to ALS point density, scanning angle and echo

return categories. Since the RBF crown volume is highly dependent on the estimated crown base height, more robust methods to estimate the crown base height might improve the results.

In conclusion, the present study revealed a strong relationship between ALS data and accurately measured branch biomass of Norway spruce at a single tree level. Furthermore, the ALS derived variable describing crown volume was shown to be a promising candidate when predicting branch biomass.

Acknowledgements

We wish to thank Blom Geomatics Norway for providing and processing the laser data. We are also grateful to Arne Drømtoorp, Leif Kjöstelsen, Vegard Lien, Tord Storbækken and Espen Martinsen who took part in the field work.

References

- Anon., 2011. TerraScan User's Guide, *Terrasolid Ltd*. Available at: <http://www.terrasolid.fi> [Accessed June 10, 2011].
- Carr, J.C. et al., 2001. Reconstruction and Representation of 3D Objects with Radial Basis Functions. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques, Los Angeles*. ACM Press, pp. 67-76.
- Hasegawa, H. & Yoshimura, T., 2003. Application of dual-frequency GPS receivers for static surveying under tree canopies. *Journal of Forest Research*, 8(2), pp.103-110.
- Hyypä, J. et al., 2001. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*, 39(5), pp.969-975.
- Kato, A. et al., 2009. Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. *Remote Sensing of Environment*, 113(6), pp.1148-1162.
- Maltamo, M. et al., 2010. Comparing different methods for prediction of mean crown height in Norway spruce stands using airborne laser scanner data. *Forestry*, 83(3), pp.257–268.
- Marklund, L.G., 1988. Biomass functions for pine, spruce and birch in Sweden, *Umeå: Swedish University of Agricultural Sciences*.
- Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80(1), pp.88 - 99.
- Persson, A., Holmgren, J & Söderman, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing*, 68(9), pp.925-932.
- Popescu, S.C., Wynne, R.H. & Nelson, R.F., 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, 29(5), pp.564-577.

- Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy*, 31(9), pp.646-655.
- Roberts, S.D. et al., 2005. Estimating individual tree leaf area in loblolly pine plantations using LiDAR-derived measurements of height and crown dimensions. *Forest Ecology and Management*, 213(1-3), pp.54-70.
- Solberg, S., Næsset, E. & Bollandsås, O.M., 2006. Single Tree Segmentation Using Airborne Laser Scanner Data in a Structurally Heterogeneous Spruce Forest. *Photogrammetric Engineering & Remote Sensing*, 72(12), pp.1369-1378.
- Straub, C. & Koch, B., 2011. Estimating Single Tree Stem Volume of *Pinus sylvestris* Using Airborne Laser Scanner and Multispectral Line Scanner Data. *Remote Sensing*, 3(5), pp.929-944.
- Vauhkonen, J., 2010. Estimating crown base height for Scots pine by means of the 3D geometry of airborne laser scanning data. *International Journal of Remote Sensing*, 31(5), pp.1213-1226.
- Vauhkonen, J. et al., 2011. Comparative testing of single-tree detection algorithms under different types of forest. *Submitted for publication*.
- Wang, Y., Weinacker, H. & Koch, B., 2008. A Lidar Point Cloud Based Procedure for Vertical Canopy Structure Analysis And 3D Single Tree Modelling in Forest. *Sensors*, 8(6), pp.3938-3951.

Airborne laser scanning-based stem volume imputation in a managed, boreal forest area: a comparison of estimation units

Jari Vauhkonen¹, Petteri Packalén¹ & Juho Pitkänen²

¹University of Eastern Finland, School of Forest Sciences, P. O. Box 111, 80101 Joensuu, Finland jari.vauhkonen@uef.fi; petteri.packalen@uef.fi

²Finnish Forest Research Institute, Joensuu Research Unit, P. O. Box 68, 80101 Joensuu, Finland, juho.pitkanen@metla.fi

Abstract

In typical airborne laser scanning (ALS)-based inventories, the forest is aggregated from initial estimation units, for which the attributes are produced using variable imputation techniques. The initial units vary in size and shape, being usually either regular grid cells or segments derived from the ALS data. This study compared small grid cells and segments of trees or tree groups as initial estimation units in an ALS-based estimation of species-specific, plot-level volume. The experiments were carried out in a managed, boreal forest area in Eastern Finland, where pine was the dominant species, and spruce and deciduous trees formed the other species groups. The field data consisted of 79 sample plots (400–900 m² in area) and the ALS data had a density of about 12 pulses/m². The estimation was overall very accurate, resulting in best-case root mean squared errors of 13% for the total volume, 23% for pine, 49% for spruce and 90% for the deciduous trees at the plot-level. The total volume was estimated most accurately using a method in which 0 to n trees were imputed per segment. However, the differences between the estimation units were minor. Despite the significant biases in the estimates, the species-specific estimation was most accurate using a single-tree approach, i.e. by considering only the largest trees per segments in the imputation. The species-specific biases were of the same magnitude than the volume not detected by the tree detection algorithm, indicating that the proportion of the detected trees was estimated very accurately.

1. Introduction

Area-based approach (ABA) and individual tree delineation (ITD) are the main strategies to estimate forest attributes from airborne laser scanning (ALS) data (e.g. Hyypä *et al.* 2008). As all trees cannot usually be detected, the ITD is more prone to bias, but comparisons of these approaches have indicated that total stand attributes such as volume and Lorey's height can be estimated with comparable accuracies using both approaches (Packalén *et al.* 2008; Yu *et al.* 2010; Peuhkurinen *et al.* 2011). The comparisons are typically performed using the same data in both approaches, although data requirements differ considerably between the approaches in terms of both ALS and field data.

ABA can be operated using ALS data with a density of <1 pulses/m², but extensive field reference data are required in order to represent the entire phenomenon of interest (e.g. Næsset 2002). The ITD can be performed with a density of 2 pulses/m² (Kaartinen and Hyypä 2008), but typically densities of 5–10 pulses/m² are used. The ITD has a potential to use less field data for model calibration (Hyypä *et al.* 2008), but the trees need to be mapped in the field. Species-specific estimation will cause further differences between the approaches: in the ABA, the species information is predicted with an auxiliary data source such as aerial images (e.g. Packalén and Maltamo 2008), whereas recent Scandinavian studies have shown potential to predict species solely from ALS using the ITD approach (e.g. Korpela *et al.* 2010).

Recent methods have developed towards a combination of ABA and ITD in that segments delineated from ALS data are provided with a summation of field reference attributes rather than treating them as single trees (Breidenbach *et al.* 2010). The utility of the segmentation remains unclear compared to an alternative to use small, regular estimation units, e.g. grid cells in a fashion of a typical ABA (e.g. Næsset 2002) in tandem with high-density ALS data. Furthermore, considering 0 to n trees within a segment may limit the number of similar observations in the reference data and thus requires further comparison to a typical ITD approach, where only the dominant trees are considered and a small amount of reference data may be adequate (Korpela *et al.* 2010; Vauhkonen *et al.* 2010).

The purpose of this study was to test different estimation units in the estimation of species-specific stem volume from high-density ALS data.

2. Methods

2.1 Study area and data

The study area is a typical boreal managed forest area in Eastern Finland (lat. 62°31'N, lon. 30°10'E). The field measurements were carried out on May-June, 2010. Altogether 79 field plots were placed subjectively, attempting to record the species and size variation over the area. Sample plot size was either 20×20, 25×25 or 30×30 meters depending on forest development class. Scots pine (*Pinus sylvestris* L.) was the dominant tree species representing 73% of the volume, Norway spruce (*Picea abies* [L.] Karst.) 16% and deciduous trees 11%.

The ALS data of the area were employed in the mapping of the trees. First, a tree map was produced using the individual tree detection method described in the next section. The tree locations were verified in the field and the undetected trees were positioned using angle and distance observations to the ALS-detected trees. The coordinates for the small trees were then calculated using these observations in a least squares adjustment as explained by Korpela *et al.* (2007). All trees with either DBH \geq 4 cm or height \geq 4 m were mapped and measured for species, DBH and height. Stem volumes were calculated using the species-specific equations presented by Laasasenaho (1982).

ALS data were collected on June 26, 2009 using an Optech ALTM Gemini laser scanning system from approximately 720 m above ground level with a field of view of 26 degrees. Pulse repetition frequency was set to 125 kHz and the instrument was operated in a multi-pulse mode. Each location was covered from two flight lines (side overlap 55%) in order to maximize the probability that trees have ALS hits each side, i.e. that there are no shadowed areas behind trees along the line from the laser scanner to a tree. A nominal sampling density was 11.9 measurements per m². The intensity data were range-corrected by the data deliverer.

One of the plots was left out from the present analysis due to having clearly more larger trees than any other plot. This plot also pointed out as an outlier in terms of ALS-based mean height and intensity. The main characteristics of the field data are presented in Table 1.

Table 1: Stem volume (m³/ha) at the plot level. SD – standard deviation.

	All	Pine	Spruce	Deciduous
Mean	193.7	131.8	40.9	21.0
SD	65.3	79.5	91.7	42.1
Min	79.5	0.0	0.0	0.0
Max	441.7	275.2	389.1	199.9

2.2 Derivation of the estimation units

The core idea of this study was to produce species-specific volume estimates to initial estimation units of different size and shape, which were then aggregated to the plot-level for evaluation. We tested two variants of both grid cells and segments as the estimation units. In the grid approach, we used grids with cell sizes of 5×5 and 10×10, denoted GRID5 and GRID10, respectively. The sizes were selected aiming to a reasonable number of observations and variation within the cells. The grid cells were fit to follow the plot borders, except in GRID10, in which the eastern- and southernmost parts of the 25×25 m plots were left out.

Both segment approaches used the same segments but had differences in terms of field reference data. In SEGM approach, the segments had the sum of field-measured volumes as the reference data in a manner presented by Breidenbach *et al.* (2010). In TREE approach, only the largest tree per segment was used (Vauhkonen *et al.* 2010). The segments were created from ALS-based canopy height model (CHM). The CHM was interpolated to a resolution of 50 cm by taking the maximum ALS point height within a pixel and filling the pixels that had no ALS hits within their area with a median filtering in local windows of 3 by 3 pixels. Hole pixels, with at least seven of the eight neighbours exceeding the height value of the centre pixel by more than five meters, were replaced with the median of the values of the neighbour pixels exceeding that threshold.

In the segmentation, the CHM was first low-pass filtered using Gaussian kernels with the size of the smoothing window increasing as a stepwise function of the heights of the CHM (Pitkänen *et al.* 2004). The segments were created around the local maxima using watershed segmentation with a drainage direction following algorithm. Pixels lower than two meters were masked out from the crown segments and small segments, at most three pixels in size, were combined to one of the neighbour segments based on the smallest average gradient on the segment boundary between two segments. The method is described in more detail by Kaartinen and Hyypä (2008) and Packalén *et al.* (2011), of which the latter also applied it to the data set of this study.

2.3 Volume imputation method

In the imputation, the estimation units were provided with stem volumes from the reference observations considered as nearest neighbours in terms of the extracted ALS-features. The imputation was carried out using the MSN method implemented in *yalImpute* package (Crookston and Finley 2008) of the R statistical computing environment (<http://www.R-project.org/>). MSN with more than one nearest neighbour (*k*-MSN) was additionally tested, with the values of *k* ranging from 2 to 10. The imputations were carried out in a leave-out-one-plot fashion, i.e. the segments belonging to the same plot as the target segment were not available as nearest neighbours. Otherwise the technical details are explained by Breidenbach *et al.* (2010).

Response variables in the imputation were the species-specific sums of the volume within the estimation units in all other approaches except TREE, in which both volume and species were imputed for a segment. As categorical variables were not allowed with the MSN method, the species attributes was coded into dummy variables. A large set of ALS-based independent variables was tested and effects of including different feature groups were also evaluated. The features were calculated either at the level of the estimation unit or at an area level, the latter being either a 250 m² circle in the segment approach or the full reference plot in the grid approach. The features are summarized in Table 2 and their calculation is explained in detail by Vauhkonen *et al.* (2010).

Table 2: Independent variables tested in this study. SD – standard deviation, H – height, Hmax – maximum height, I – intensity, Imax – maximum intensity.

Feature group	Description	Grid	Segment
Hstat	Maximum, mean and SD of H values and proportion of H > 0.5 m	X	X
Hperc	Percentiles 5, 10, 20, ..., 90, 95% of Hmax	X	X
HpercRel	Hperc divided by Hmax	X	X
Hdens	Proportional densities 5, 10, 20, ..., 90, 95% of Hmax	X	X
Istat	Descriptive variables of I values 0-40% down from the treetop	X	X
Iperc	Percentiles 5, 10, 20, ..., 90, 95% of Imax	X	X
Cvol	Volume of the 3D convex hull of the point data above 50, 60, 70 and 80% of Hmax	-	X
Carea	Area of the 2D convex hull of the point data below 10, 20, ..., 100% of Hmax	-	X
Astat	Area-level maximum, mean and SD of H values and proportion of H values > 0.5 m	X	X
Aperc	Area-level percentiles 5, 10, 20, ..., 90, 95% of Hmax	X	X
Adens	Area-level proportional densities 5, 10, 20, ..., 90, 95% of Hmax	X	X
Aint	Area-level mean and SD of I values	X	X

2.4 Evaluation and performance measures

In all cases, the correspondence of the estimates with the reference data was evaluated in terms of RMSE and bias:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

$$bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (2)$$

where n is the number of observations, and y_i and \hat{y}_i are the reference and estimated attributes, respectively. The relative RMSEs were calculated by dividing the absolute RMSE values by the mean of the reference attribute. Paired t-test was used to test if the average difference between the predicted and observed values was zero (“significance of bias” in Breidenbach *et al.* 2010).

3. Results

3.1 Properties of the estimation units

The theoretical potential to produce the attributes with the TREE approach was first evaluated by comparing the field-observed total volume and stem number of the ALS-detected trees to those of all trees (Table 3). The total stem number and volume were underestimated by 46 and 13%, respectively. Pine trees were detected slightly more frequently and spruce trees slightly less frequently compared to the total, whereas the stem number and volume of deciduous trees were underestimated by 76 and 26%.

Table 3: Performance of the ITD algorithm.

	Number of stems		Volume, m ³ /ha	
	Bias	Bias%	Bias	Bias%
All trees	575.7	45.7	26.0	13.4
Pine	229.7	32.4	15.2	11.5
Spruce	112.5	52.8	5.3	13.0
Deciduous	255.4	76.0	5.5	26.1

When intersected with the field data, GRID10 had highest mean volume and lowest coefficient of variation (CV) of the considered estimation units (Table 4). GRID5 had higher mean volume than SEGM, but especially the species-specific CVs were close to each other. The mean volume in the TREE approach was only slightly lower than in SEGM, but the CVs of the approaches differed more from each other.

Table 4: Main properties of the segments.

Method	Mean volume, m ³				Coefficient of variation, %			
	All trees	Pine	Spruce	Deciduous	All trees	Pine	Spruce	Deciduous
GRID5	0.50	0.32	0.12	0.06	84.8	114.4	271.8	330.8
GRID10	2.06	1.31	0.52	0.23	47.5	76.1	220.1	231.5
SEGM	0.31	0.20	0.07	0.03	96.5	119.5	297.4	393.4
TREE	0.27	0.18	0.07	0.03	95.1	122.1	308.3	438.1

3.2 Imputation results with all available variables

The accuracies of stem volume imputation using all available ALS-features are shown in Tables 5 and 6. When evaluated at the level of the initial estimation units, the predictions showed a statistically insignificant bias in nearly all cases, and the obtained RMSEs were smaller than the corresponding CVs (Table 5), the only exception being SEGM with $k=1$. The RMSE values appear to be connected to the size of the estimation units.

The RMSEs usually decreased as a function of the value of k . The increase in the performance was typically higher when increasing k from 1 to a few than e.g. from 5 to 10. Interestingly, the direction of the bias typically changed from overestimation towards underestimation.

When considered at the plot-level (Table 6), the predictions of total volume and the volume of pine trees were both within a margin of 7 percentage points with all methods, while spruce and deciduous trees included more variation between the applied estimation units. The total volume was predicted most accurately using SEGM (RMSE 13%). Despite significant biases, all species-specific volumes were predicted most accurately with the TREE method. Except for spruce, which was estimated accurately using SEGM, all other species were estimated more accurately with grid-based methods. GRID10 showed a better performance than GRID5 in all other cases except the volume of spruce trees. The effect of k was not as obvious when considered at this level, and the accuracy of the TREE estimates tended to decrease with an increasing k , for example.

Table 5: Accuracy of the volume estimates at the level of the initial estimation units.

Method	k	All trees		Pine		Spruce		Deciduous	
		RMSE%	Bias%	RMSE%	Bias%	RMSE%	Bias%	RMSE%	Bias%
GRID5	1	77.6	1.3	111.7	0.9	193.8	1.5	328.2	2.8
	5	63.6	1.0	88.5	0.2	165.1	1.9	267.8	4.0
	10	60.9	1.2	85.1	-0.2	156.3	2.6	262.4	6.1
GRID10	1	33.0	0.3	52.0	0.7	98.8	-2.6	183.3	4.5
	5	28.2	0.8	46.6	-1.0	88.8	-0.1	158.8	13.4
	10	27.7	0.9	44.8	-2.1	87.4	1.5	159.7	16.8*
SEGM	1	61.4	-2.1	105.3	-3.6	183.1	-0.5	429.9	3.1
	5	47.8	-0.9	82.7	-2.7	154.8	2.0	342.2	3.2
	10	46.3	0.0	79.8	-2.0	148.1	1.9	326.8	7.2
TREE	1	56.2	-0.6	94.3	-0.9	154.8	1.7	406.4	-4.1
	5	45.7	1.0	76.5	1.2	137.1	5.0*	331.4	-2.6
	10	44.7	1.4	73.9	1.6	135.3	4.5	328.5	1.1

Table 6: Accuracy of the volume estimates at the plot level.

Method	k	All trees		Pine		Spruce		Deciduous	
		RMSE%	Bias%	RMSE%	Bias%	RMSE%	Bias%	RMSE%	Bias%
GRID5	1	17.1	2.4	28.4	3.5	61.5	2.1	120.8	-4.0
	5	16.7	2.1	26.1	2.6	61.7	2.5	113.7	-2.3
	10	15.4	2.3	24.5	2.2	62.8	3.4	115.9	0.3
GRID10	1	16.8	0.5	26.5	0.4	66.8	-0.5	100.5	2.7
	5	13.8	1.1	24.7	-0.1	66.3	1.5	101.6	8.4
	10	13.7	1.0	24.0	-1.3	60.9	3.2	105.9	11.1
SEGM	1	14.5	-1.5	29.5	-1.9	53.1	-1.1	124.7	-0.1
	5	13.0	-0.2	25.6	-0.7	55.3	1.4	115.2	0.3
	10	12.9	0.6	25.7	-0.1	55.7	1.3	116.5	3.9
TREE	1	19.3	12.8*	23.4	10.5*	48.7	13.9*	90.6	25.3*
	5	19.9	14.0*	24.9	12.7*	54.7	16.6*	90.3	24.0*
	10	19.9	14.2*	24.6	12.9*	55.7	16.0*	97.4	26.3*

3.3 Effects of including different feature groups

Table 7 shows the estimation accuracy when different feature groups were considered. Usually a slightly better performance was obtained when using a subset of all available features. Feature groups *Hdens*, *Iperc* and *Cvol* were not used in any of the estimations.

The performances of the species-specific estimates could be increased by using percentile variables instead of basic height values (*Hstat*), but a more radical increase in the performance was obtained by including intensity variables with these. Using SEGM and TREE, the areas of point clouds at different relative height levels were found to have additional prediction power especially with respect to volumes of spruce. Finally, predictors calculated at an area larger than the initial estimation unit improved the predictions for spruce and deciduous trees in all tested cases.

Table 7: Accuracy of the volume estimates at the plot level using different predictor groups. The values are the mean RMSE% over k values of 1–10.

Feature group	All trees	Pine	Spruce	Deciduous	All trees	Pine	Spruce	Deciduous
	GRID5				GRID10			
Hstat	24.5	60.6	229.2	201.8	18.9	62.4	228.7	207.3
Hperc	20.3	43.7	129.2	192.6	18.0	41.3	142.6	188.5
Previous + Istat	15.9	25.1	60.7	111.3	13.6	24.3	68.7	111.5
Previous+Astat+Aint	15.0	24.6	50.9	101.6	13.8	24.9	62.8	106.2
All	16.4	26.1	61.8	115.9	14.2	24.4	65.0	102.9
	SEGM				TREE			
Hstat	19.5	60.3	140.4	210.7	21.7	63.3	147.4	206.0
Hperc	17.1	38.1	94.3	194.5	19.5	44.4	110.3	197.6
Previous + Carea	14.7	38.2	88.4	159.1	19.3	39.6	90.1	164.7
Previous + Istat	13.8	25.3	63.4	114.4	18.5	23.1	64.9	109.3
Previous + Astat + Adens + Aint	13.9	23.8	51.7	110.1	18.8	23.5	54.5	101.5
All	13.2	25.7	55.0	116.3	19.8	24.7	53.9	92.8

4. Discussion

The predictions produced in this study were overall very accurate compared to previous studies. One should take the different plot size into account, but the total and species-specific accuracies at the plot level were comparable to those reported by Breidenbach *et al.* (2010), yet the estimation here was carried out without the aerial images. Considering tree-level imputation, Vauhkonen *et al.* (2010) reported RMSE of about 30% for stem volume estimates and species recognition accuracy of about 80%, while here the corresponding numbers were 45–56% and about 84%. The difference is likely in the use of different segmentation method between the studies.

The most accurate species-specific predictions were produced by the TREE approach. It is notable that the predictions (Table 6) had biases of the same magnitude than the trees not detected by the ITD algorithm (Table 3) indicating that the properties of the detected trees was estimated very accurately. An explanation may be that single trees are simpler units and thus easier to find more similar reference units in the nearest neighbour search.

The analysis of including different features groups showed that useful predictors vary for the different estimation units. However, some common remarks were made: first, the species-specific estimation with all methods was improved by including ALS intensity metrics. Second, a multi-scale approach, in which the predictors are calculated not only at the level of the estimation unit but also at a larger area was found useful for all tested units.

The results of this study may have implications towards collecting field reference data for ALS-based inventories. More efforts are required to collect an adequate field reference data for the SEGM approach than for the TREE approach, especially if a methodology corresponding to this study (Korpela *et al.* 2007) is used without mapping the smallest trees, i.e. verifying only the ALS-detected trees in the field. The results of the grid based approach should be re-evaluated under operationally more realistic conditions, i.e. using lower resolution area-based data. Finally, the amount of required reference data per method should be tested.

Acknowledgements

This study is a contribution to the 7th Framework Programme of the European Commission project FlexWood – Flexible Wood Supply Chain and additionally supported by the strategic funding of the University of Eastern Finland, granted to project Multi-Scale Geospatial Analysis of Forest Ecosystems.

References

- Breidenbach, J., Næsset, E., Lien, V., Gobakken, T. and Solberg, S., 2010. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment*, 114, 911-924.
- Crookston, N.L. and Finley, A.O., 2008. yaImpute: An R package for kNN imputation. *Journal of Statistical Software*, 23(10), 16 p. Available at: <http://www.jstatsoft.org/v23/i10>.
- Hyypä, J., Hyypä, H., Leckie, D., Gougeon, F., Yu, X. and Maltamo, M., 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, 29, 1339-1366.
- Kaartinen, H. and Hyypä, J., 2008. EuroSDR / ISPRS Commission II Project "Tree Extraction" – Final report. Official publication n:o 53. EuroSDR, Frankfurt am Main, Germany. 60 p. Available at: <http://bono.hostireland.com/~euroedr/publications/53.pdf>.
- Korpela, I., Tuomola, T. and Välimäki, E., 2007. Mapping forest plots: An efficient method combining photogrammetry and field triangulation. *Silva Fennica*, 41(3), 457-469.
- Korpela, I., Ørka, H.O., Maltamo, M., Tokola, T. and Hyypä, J., 2010. Tree species classification using airborne LiDAR – effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. *Silva Fennica*, 44(2), 319-339.
- Laasasenaho, J., 1982. Taper curve and volume function for pine, spruce and birch. Publications of Forest Research Institute in Finland. *Communicationes Instituti Forestalis Fenniae*, 108, 74 p.
- Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80, 88-99.
- Packalén, P. and Maltamo, M., 2008. Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. *Canadian Journal of Forest Research*, 38, 1750-1760.
- Packalén, P., Pitkänen, J. and Maltamo, M., 2008. Comparison of individual tree detection and canopy height distribution approaches: A case study in Finland. In: R.A. Hill, J. Rosette, J. Suárez (Eds.). *Proceedings of SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment and inventory*. Edinburgh, UK: 22-29.
- Packalén, P., Vauhkonen, J., Kallio, E., Peuhkurinen, J., Pitkänen, J., Pippuri, I. and Maltamo, M., 2011. Comparison of the spatial pattern of trees obtained by ALS based forest inventory techniques. Submitted to SilviLaser 2011.
- Peuhkurinen, J., Mehtätalo, L. and Maltamo, M., 2011. Comparing individual tree detection and the area-based statistical approach for the retrieval of forest stand characteristics using airborne laser scanning in Scots pine stands. *Canadian Journal of Forest Research*, 41, 583-598.

- Pitkänen, J., Maltamo, M., Hyypä, J. and Yu, X., 2004. Adaptive methods for individual tree detection on airborne laser based canopy height model. In: M. Theis, B. Koch, H. Spiecker, H. Weinacker (Eds.). *Proceedings of ISPRS working group VIII/2: "Laser-Scanners for Forest and Landscape Assessment"*. Germany, University of Freiburg: 187-191.
- Vauhkonen, J., Korpela, I., Maltamo, M. and Tokola, T., 2010. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment*, 114, 1263-1276.
- Yu, X., Hyypä, J., Holopainen, M. and Vastaranta, M., 2010. Comparison of area-based and individual tree-based methods for predicting plot-level forest attributes. *Remote Sensing*, 2(6), 1481-1495.

Lidar in Australia

DEMs to Environmental Monitoring

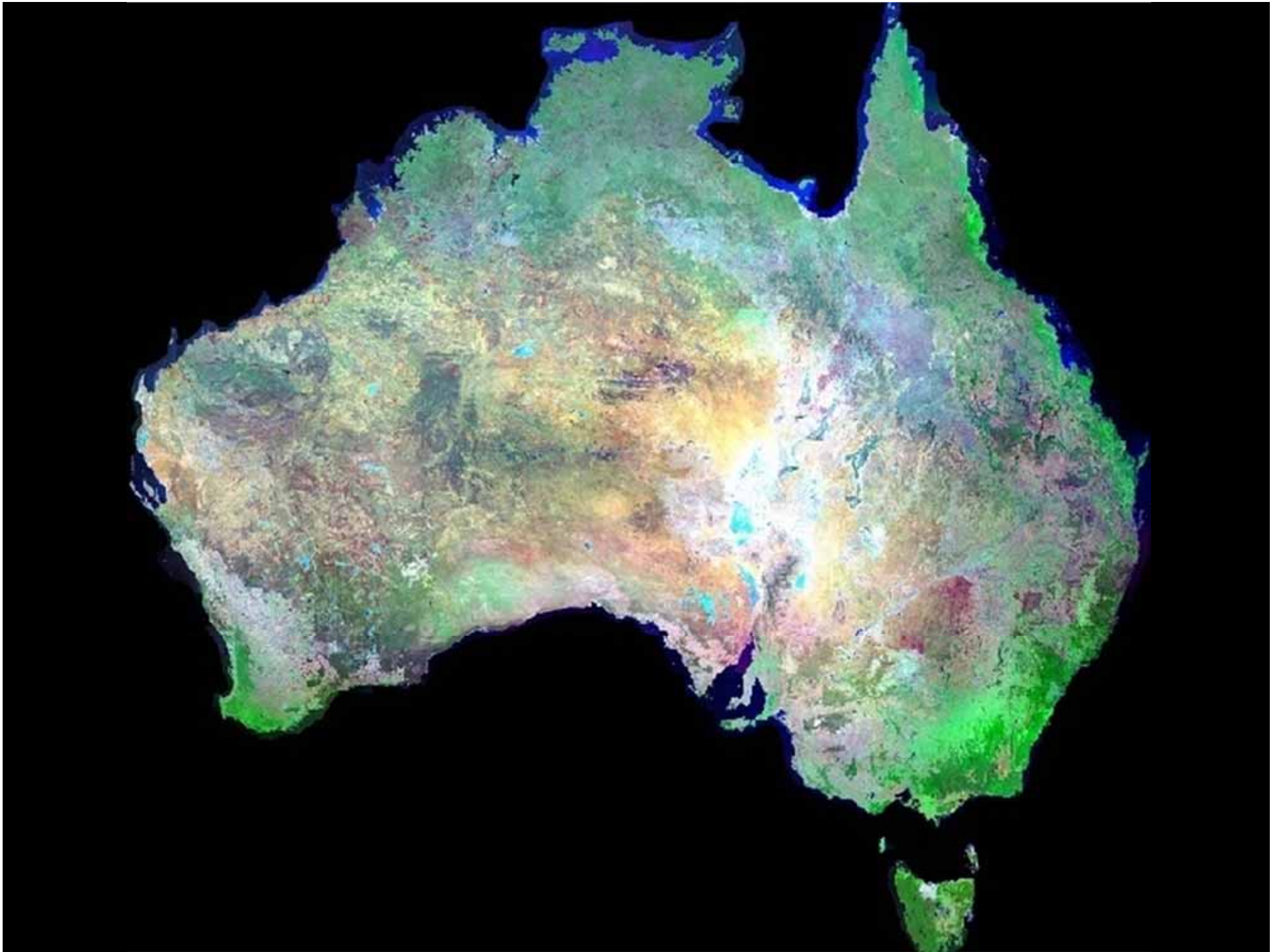
Christian Witte
Manager Remote Sensing Centre
Environment and Resource Sciences



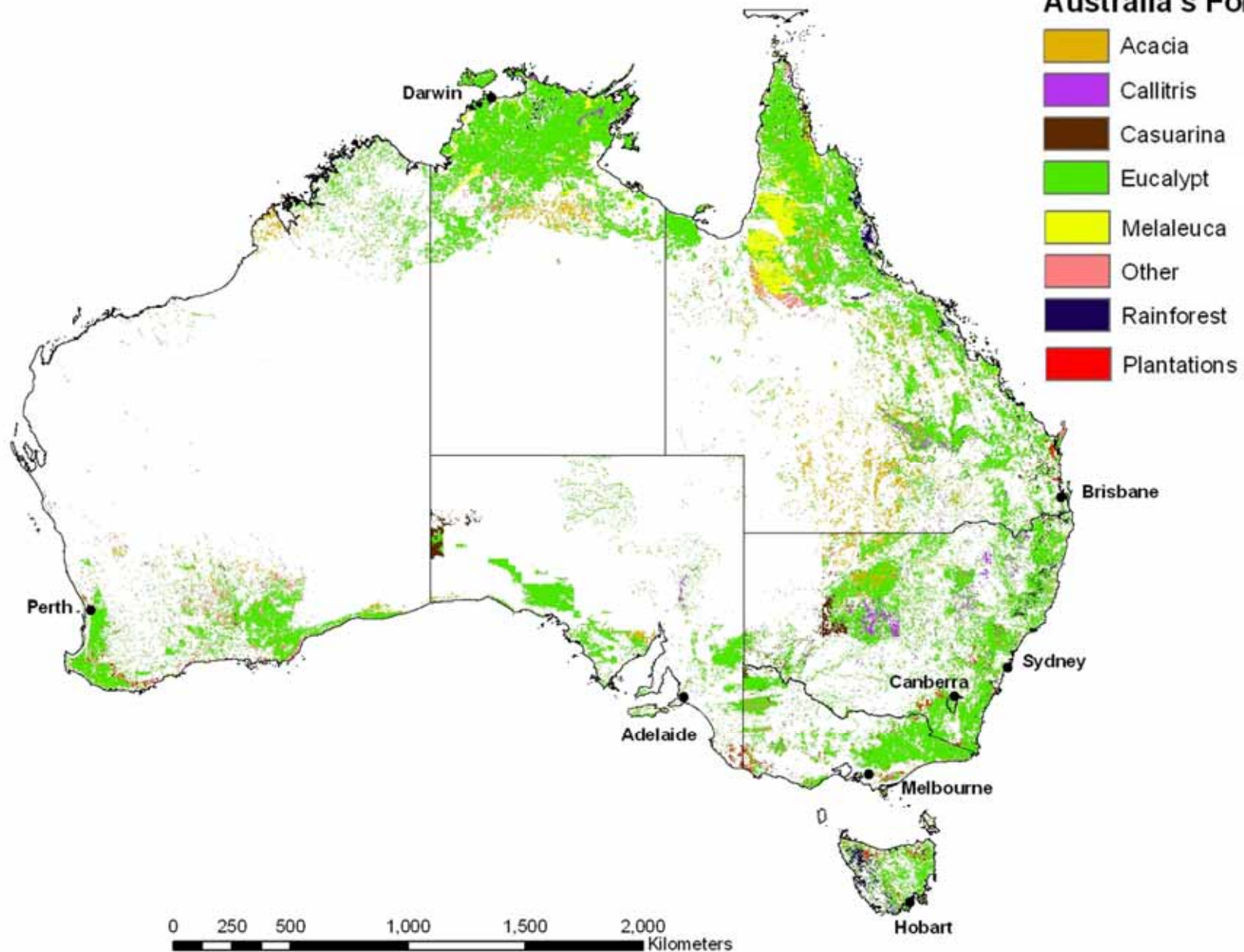


Contributors

- John Armston, Nick Goodwin, Fiona Watson, Peter Todd, Peter Scarth, Dan Tindall, Kevin Chetwynd, Paul Brown (QLD Department of Environment and Resource Management)
- Glenn Jones (NSW Land and Property Management Authority)
- Rick Frisina (Victorian Department of Sustainability and Environment)
- Nathan Qadros (Cooperative Research Centre for Spatial Information)
- Phil Tickle, Nerida Wilson, Hamish Anderson (Geoscience Australia)
- Martin Mutendeudzi and Anthony Hunn (Australian Bureau of Agricultural and Resource Economics and Sciences)
- Darius Culvenor (CSIRO)
- David Jonas, Gail Kelly (AAMHatch)
- David Moore (Terranean)
- Stuart Phinn (University of Queensland)
- Peter Bunting, Richard Lucas (University of Aberystwyth)
- Jacky Croke (Griffith University)
- Matt Coleman, James Bangay (Ergon Energy)

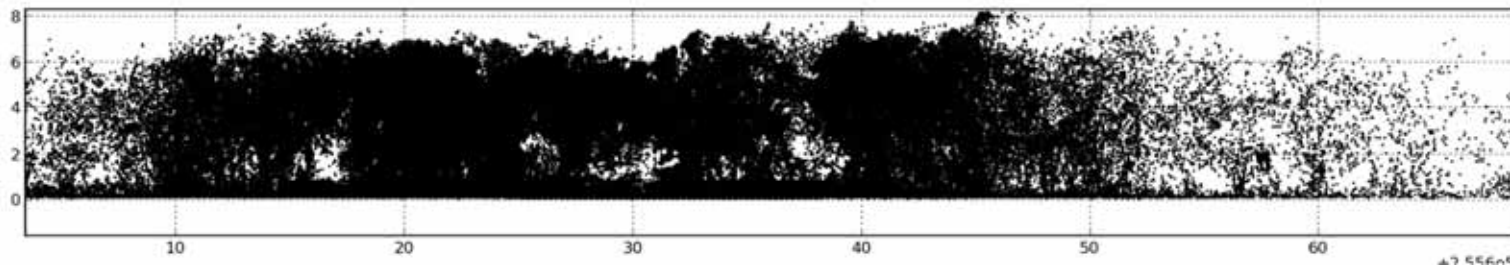


Australia's Forests 2008



Source: State of the Forest Report 2008, ABARES

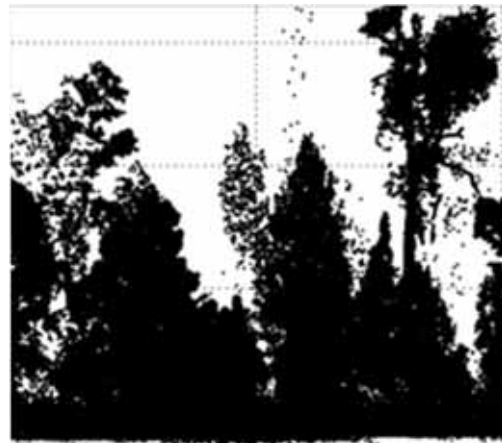
Structure of woody vegetation in Queensland



Mulga scrub



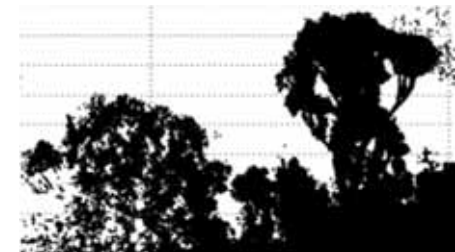
Subtropical
rainforest



Closed savanna
woodland



Open savanna
woodland

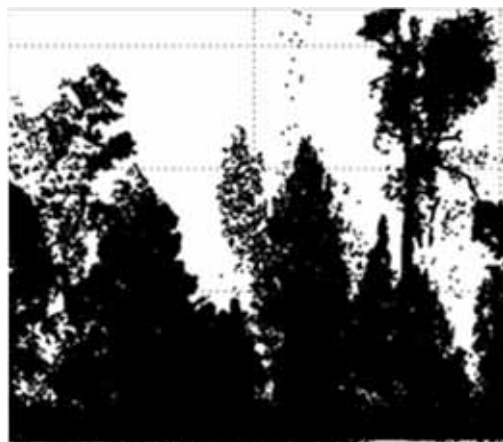


Coastal banksia
woodlands

Structure of woody vegetation in Queensland



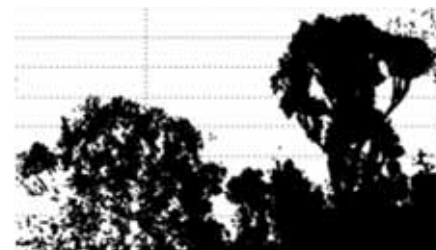
Subtropical rainforest



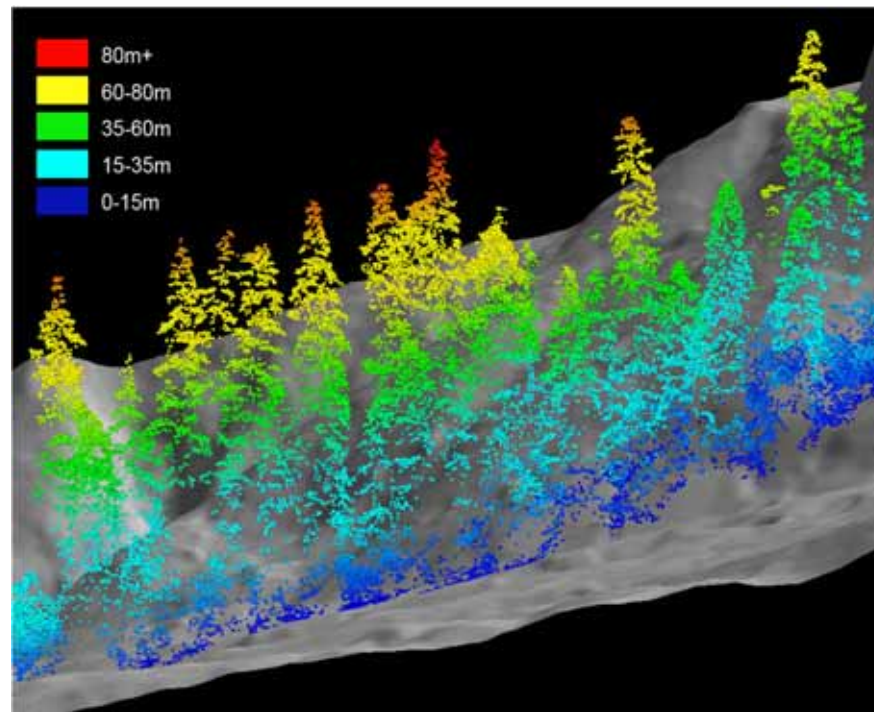
Closed savanna woodland



Open savanna woodland

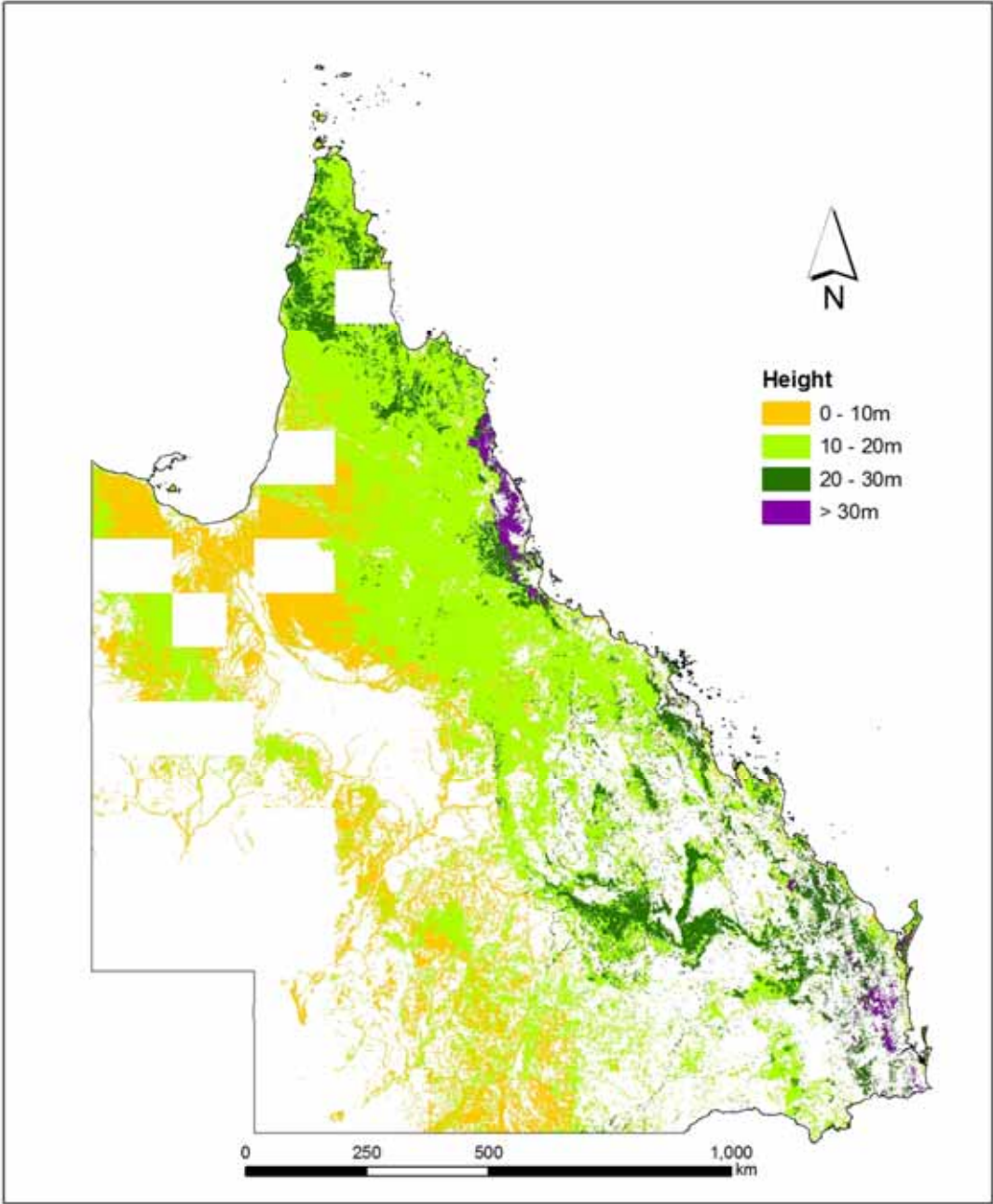


Coastal banksia woodlands

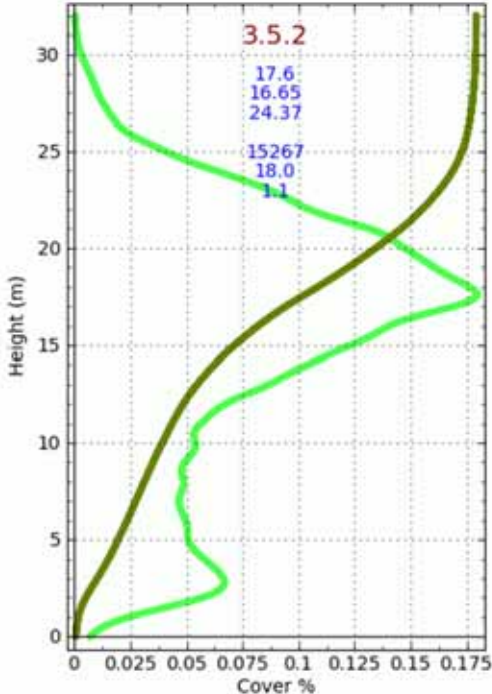


(Image courtesy of Tom Spies and Keith Olsen, OSU College of Forestry)

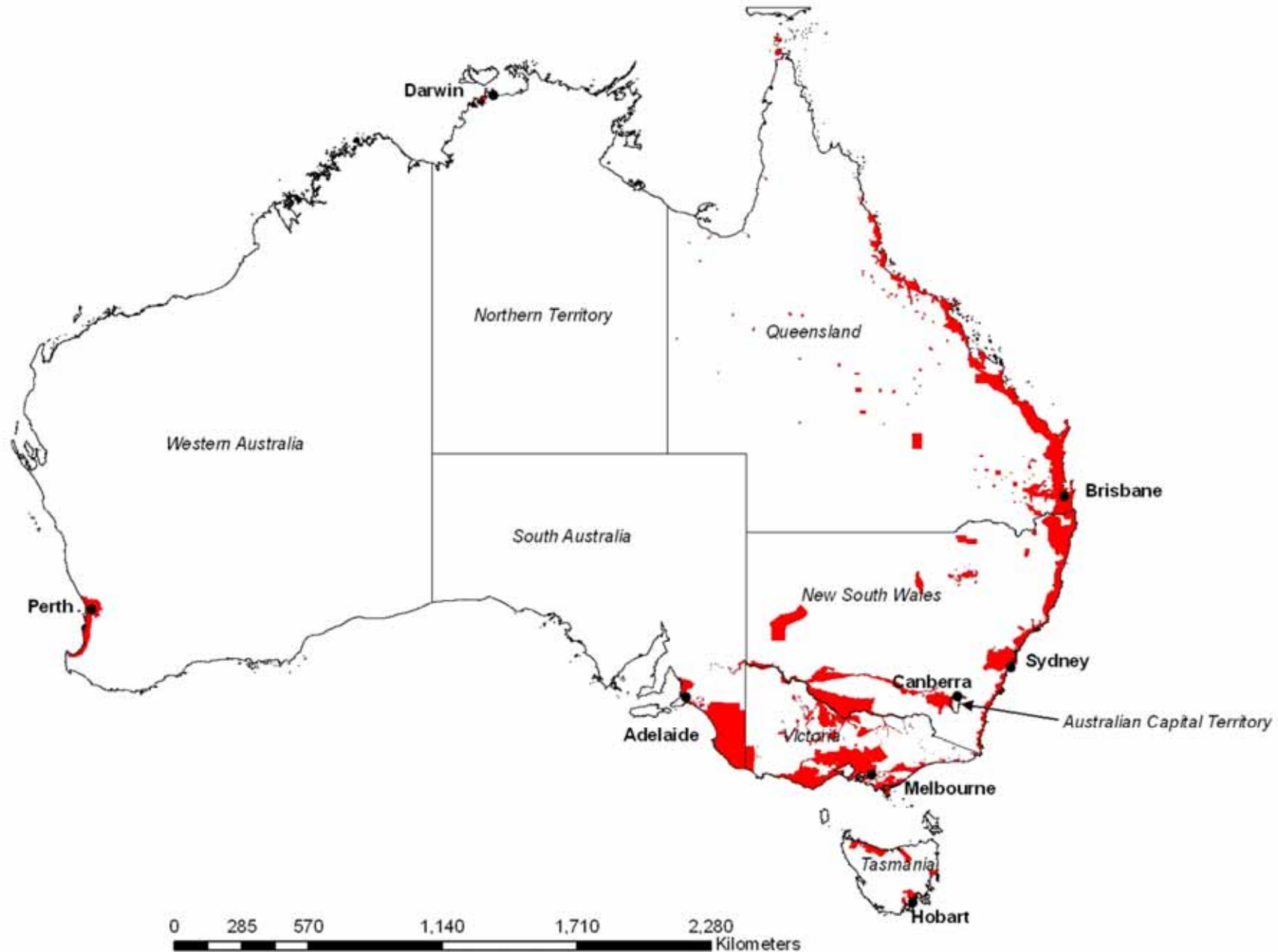
Satellite Lidar Vegetation Height for Queensland



- Aggregation of ICESat Lidar within vegetation mapping polygons



Government funded lidar capture in Australia



Source: Geoscience Australia, Victoria DSE

Current and Likely Future Lidar Projects

Government funded captures

- Capital cities
- Coastal DEMs on-going
- Non-coastal flood risk assessment
- Fuel loads and fire hazard mapping for peri-urban fringes around Melbourne
- Ergon Powerline/Vegetation Monitoring Program

Non-government funded

- Coal-seam gas project planning
- Other mining related captures





Source: Queensland Climate Change Centre of Excellence (QCCCE)

Significant Progress in Lidar Capture in Australia

- Greatly improved licensing and access arrangements
 - Creative Commons Licensing
 - Whole-of-government access
- Lidar Acquisition Specifications and Tender Template
 - Intergovernmental Committee on Surveying and Mapping
- Lead by Geoscience Australia with contributors across the country

Australian Specifications for Lidar Capture - Considerations

- User's won't pay the significant additional funds to achieve higher ground classification accuracy
- Suggested future additions
 - Standard for vegetations points and classifications
 - Standards for full waveform
- Development of a toolkit for assessing data against specifications should be a priority

Room for Improvement

- Commercial-in-confidence issues
- Transparency in processing and classifying point cloud
- Quality assurance
- Efficient data access
- Open source software and file formats
 - Sorted Pulse Data Library or SPDLib

Research and Developmental Priorities

- Systems
 - Multiple lasers
 - Multi-spectral lidar
 - Unmanned aerial vehicles
- Research
 - Efficient integration of technologies e.g. lidar and imagery
 - Further automation in ground classification
 - Storing, processing and analysing full waveform lidar
 - Integration of bathymetry and lidar DEMs

Monitoring with Lidar

Differences in DEMs or vegetations points not related to on-ground changes

- Sensor characteristics
- Capture specifications
- Datums, ellipsoids, survey control
- Classifications
- DEM algorithms

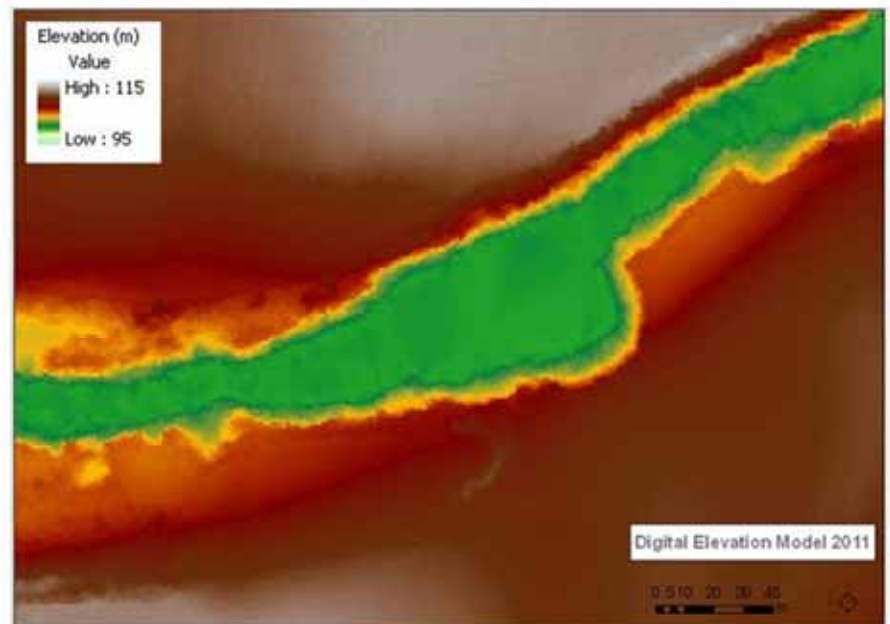
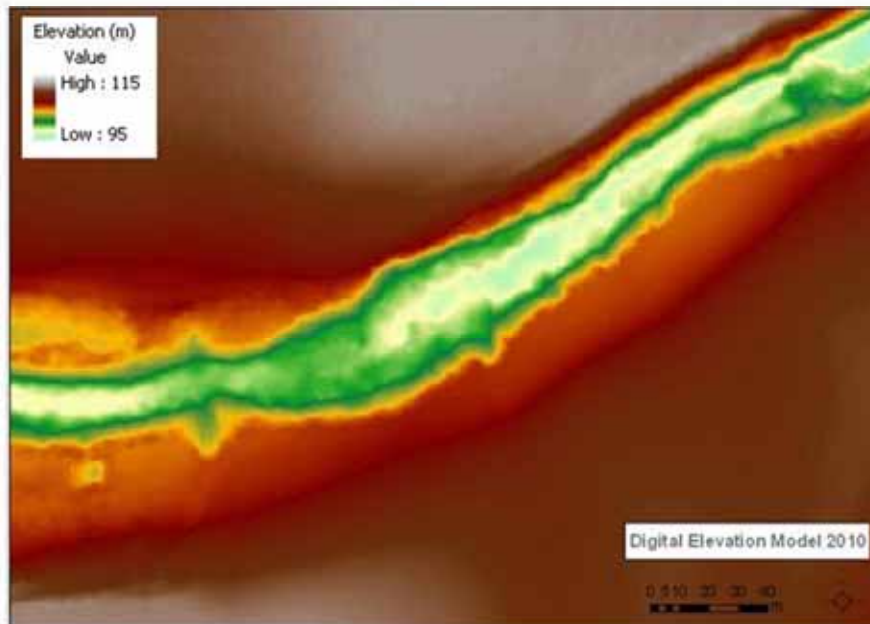
Differences due to on-ground changes

- Of interest
 - Tree growth, tree death, clearing, erosion, deposition
- Not of interest
 - Grass cover, crops

Lockyer Valley Post Flood Geomorphic Assessment



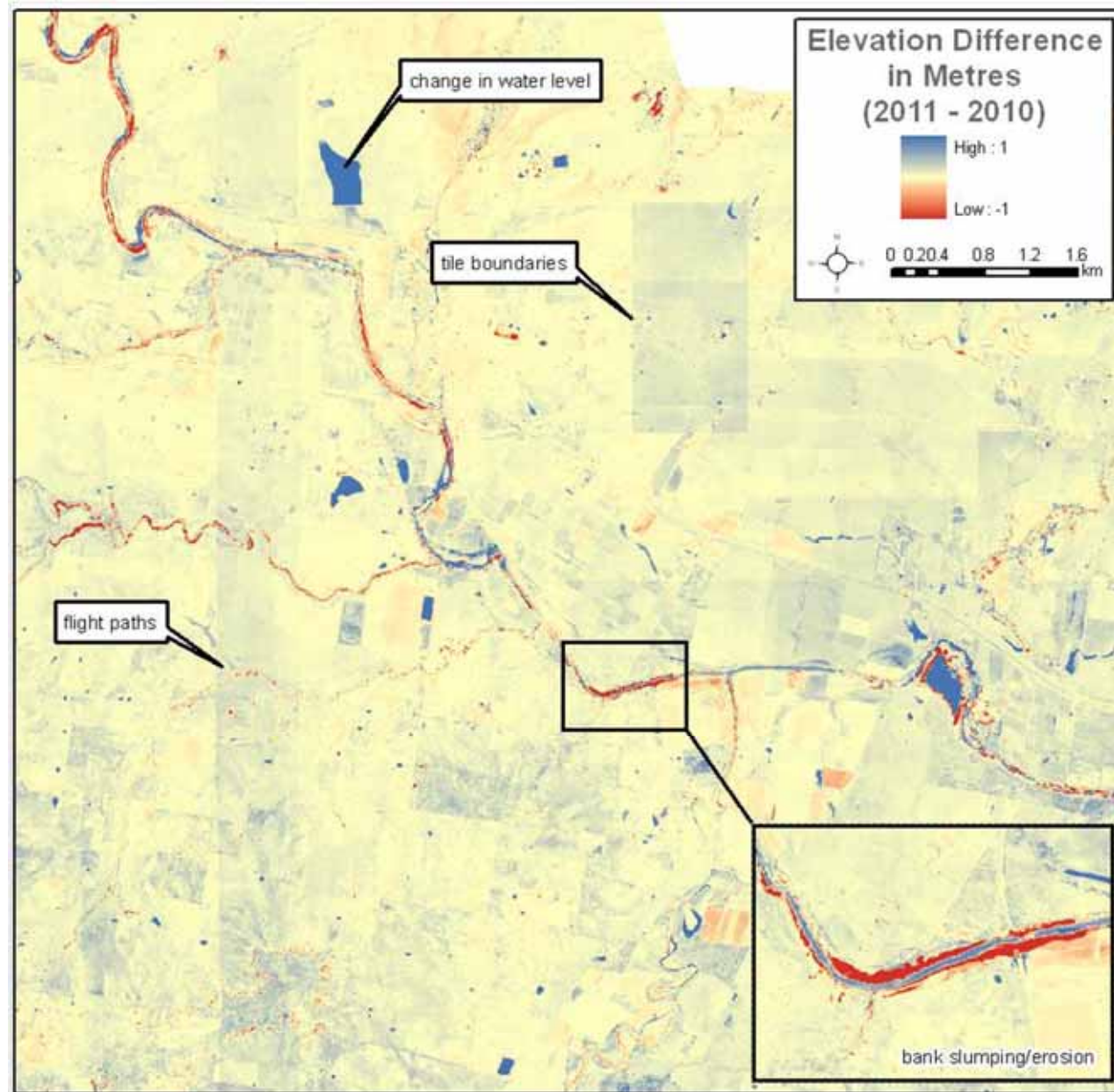
Lockyer Valley Post Flood Geomorphic Assessment



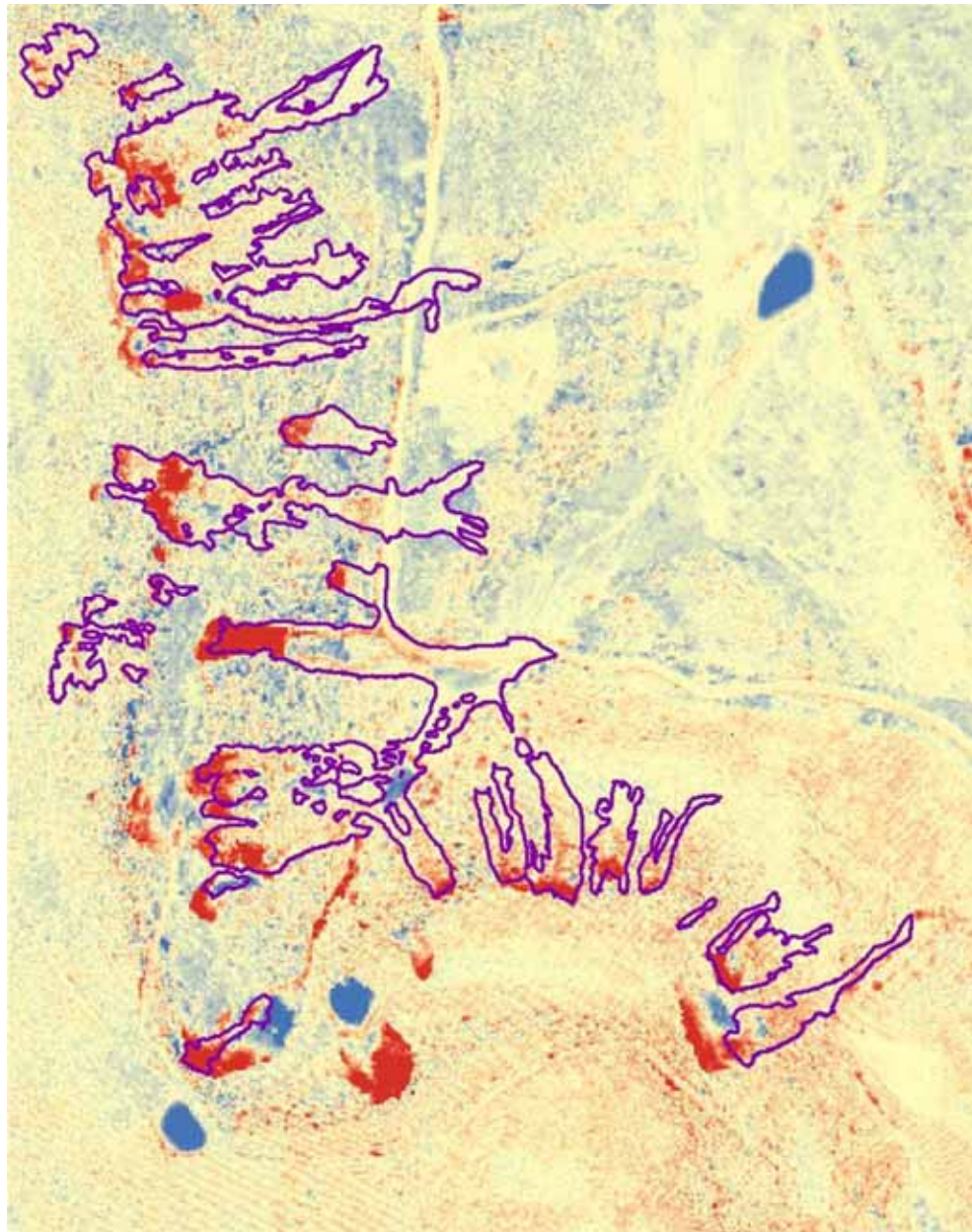
Challenges – Ground cover



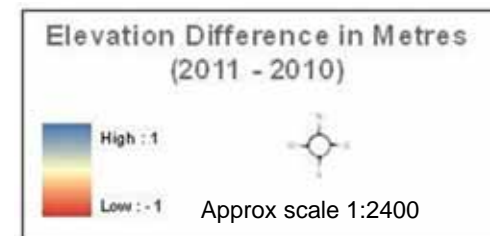
Lockyer Valley Change Analysis Pre- and Post Flood



Lockyer Valley Change Analysis Pre- and Post Flood

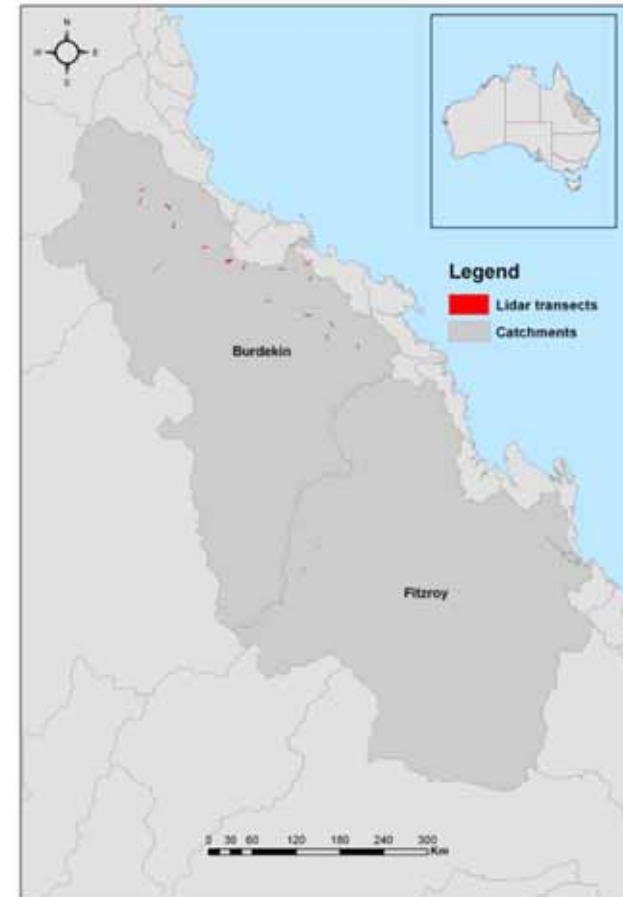


Over 1000 landslips mapped

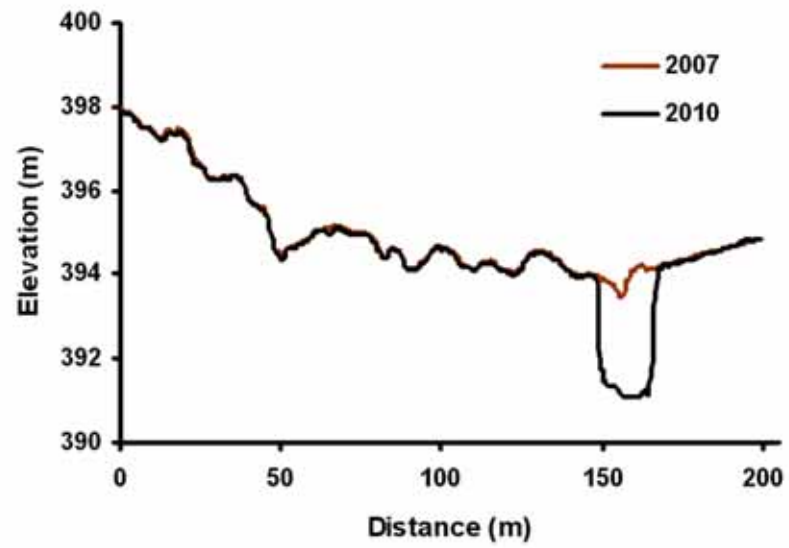
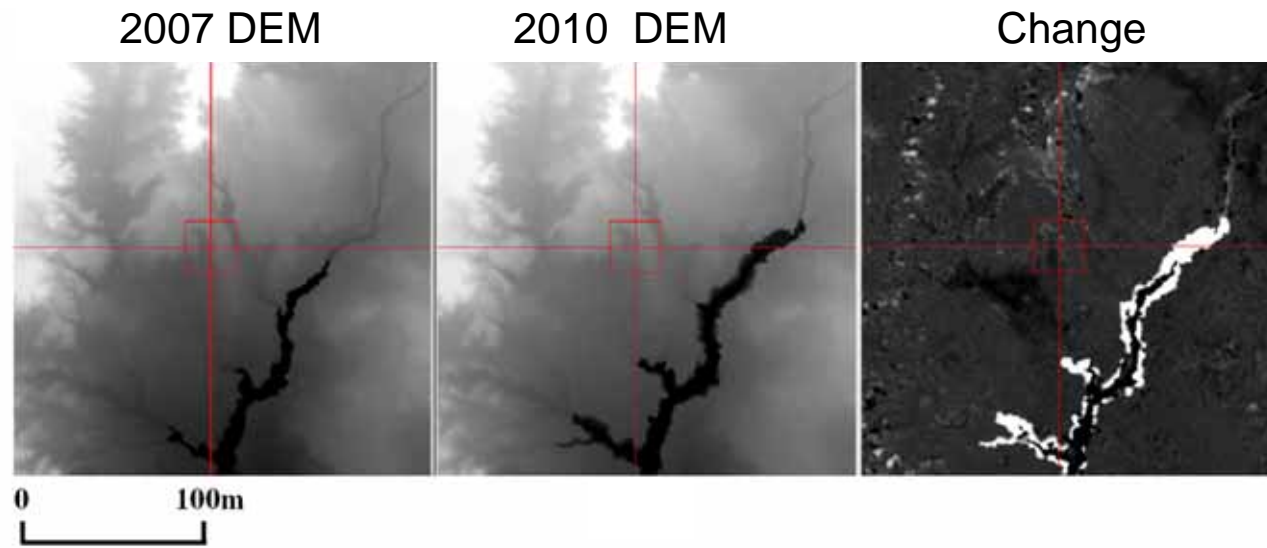


Monitoring Gully Erosion in Great Barrier Reef Catchments

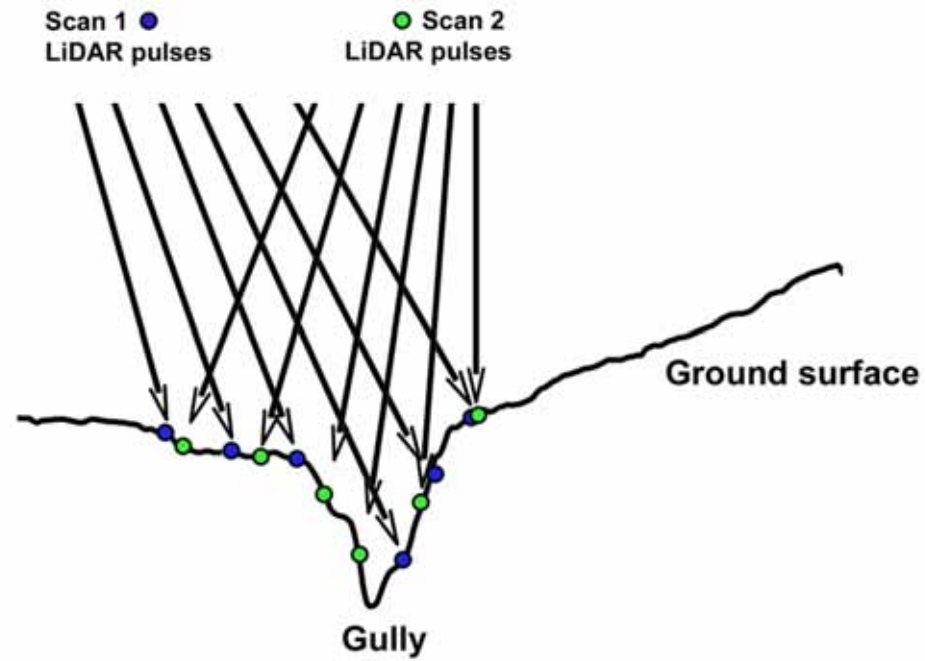
- Burdekin River Catchment:
 - 15 patches of LiDAR (12 to 31 km² in size) acquired in May and June 2010
- Average pulse density of 4.2 m² and an overlap of 50 percent overlap between flight runs.
- Fitzroy River Catchment
 - 5 patches of LiDAR were acquired in 2007 and 2010.



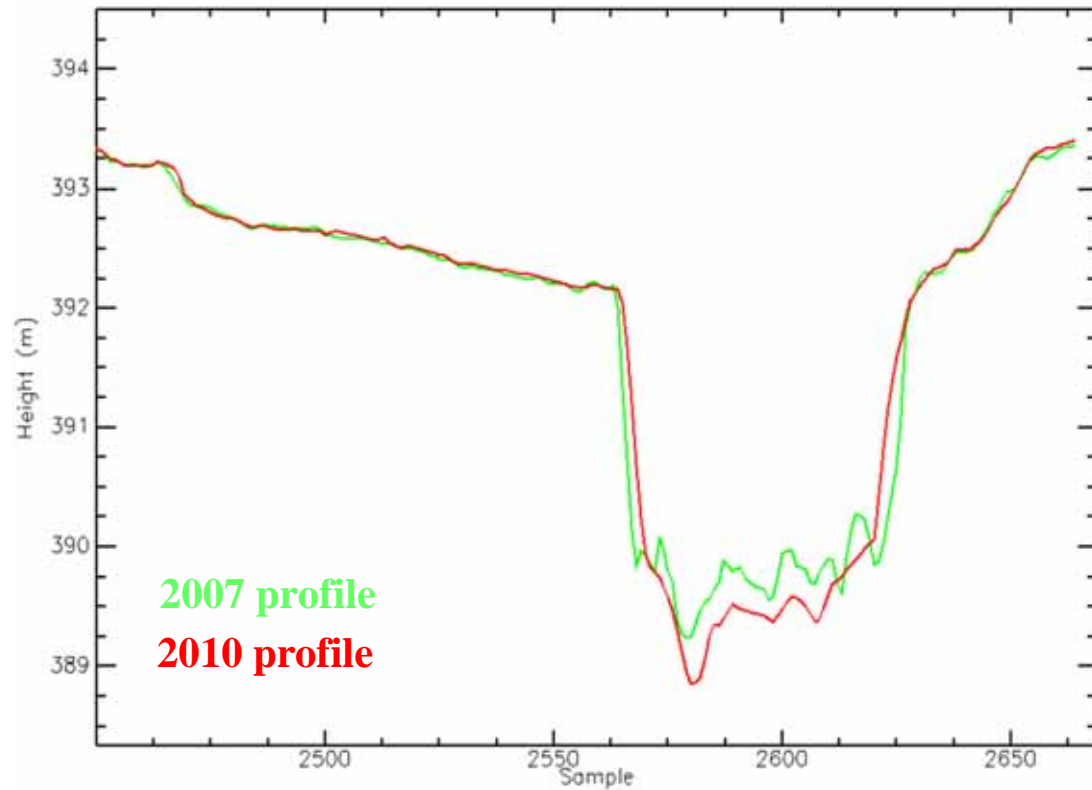
Monitoring Gully Erosion



Appropriate data capture specifications

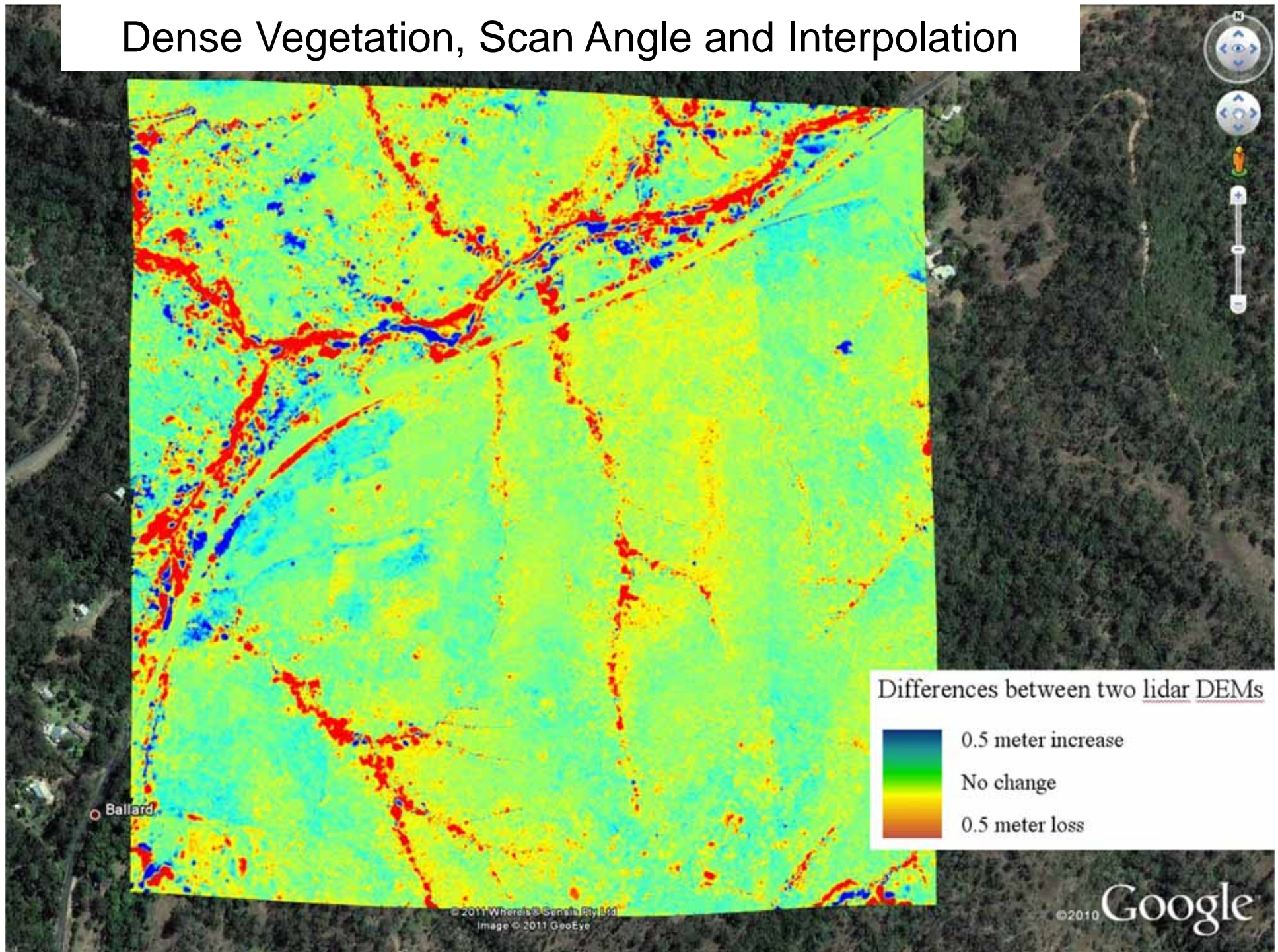


Monitoring Gully Erosion – Does it work?

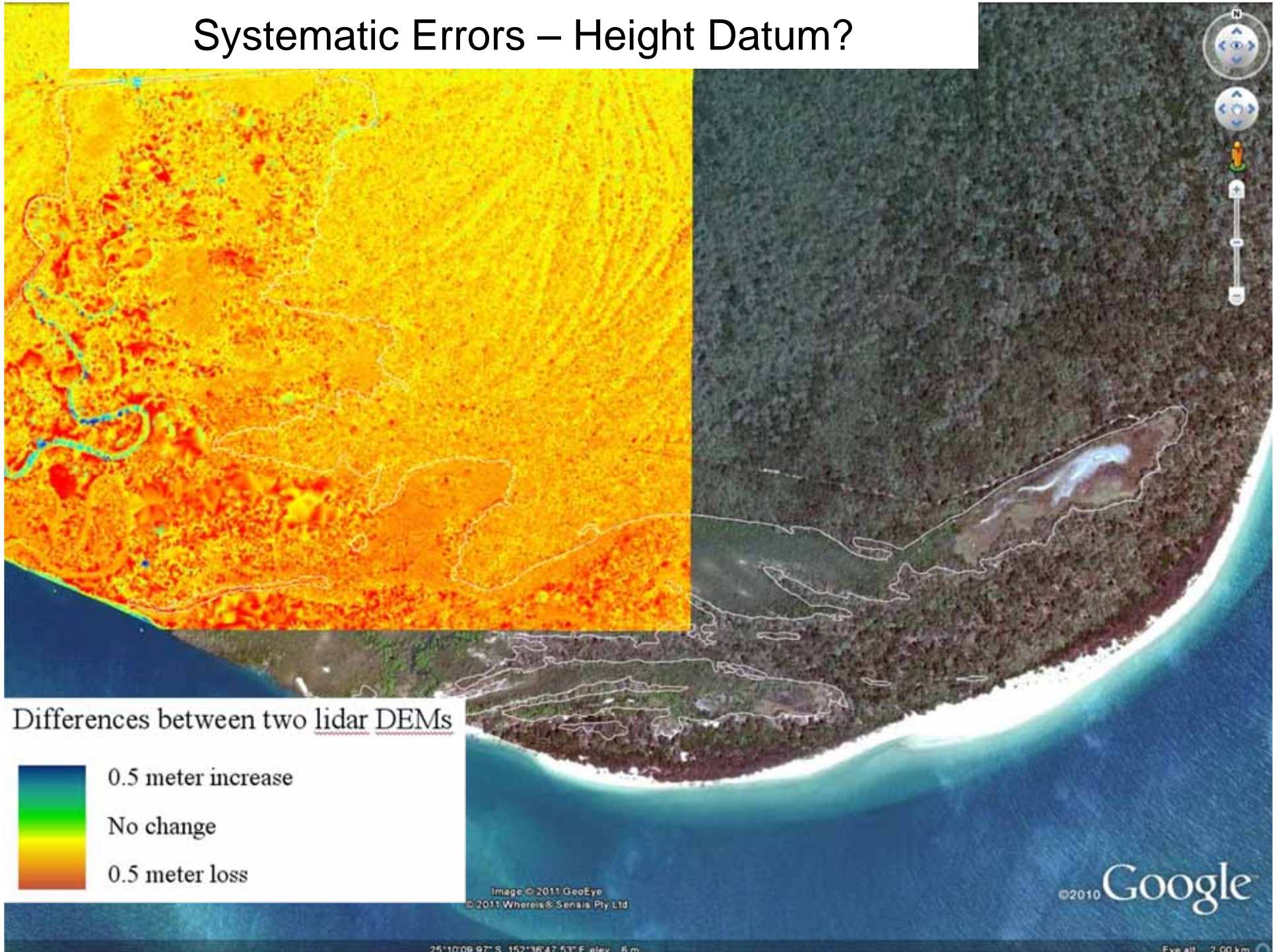


Fitzroy Basin, Qld

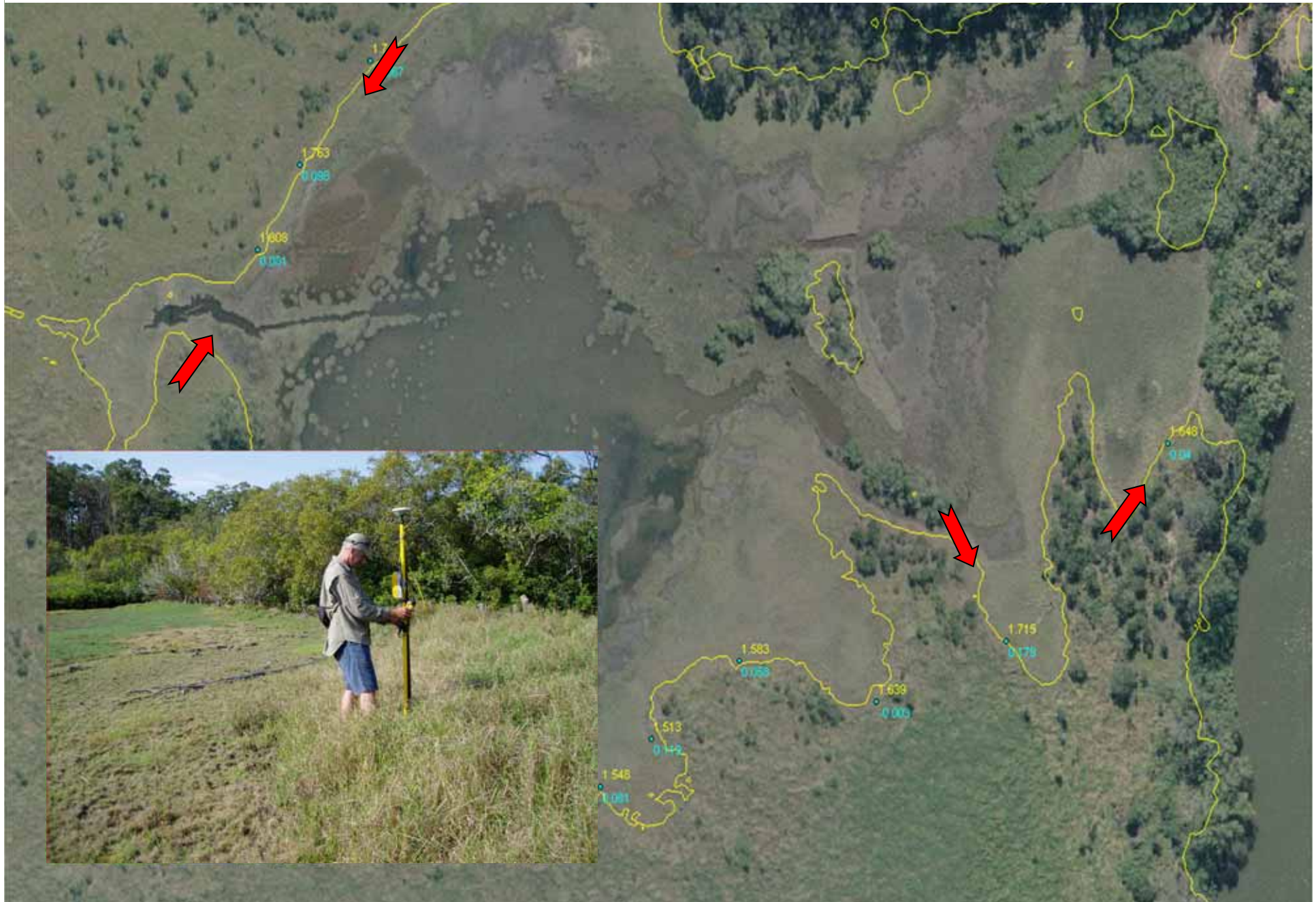
Dense Vegetation, Scan Angle and Interpolation



Systematic Errors – Height Datum?

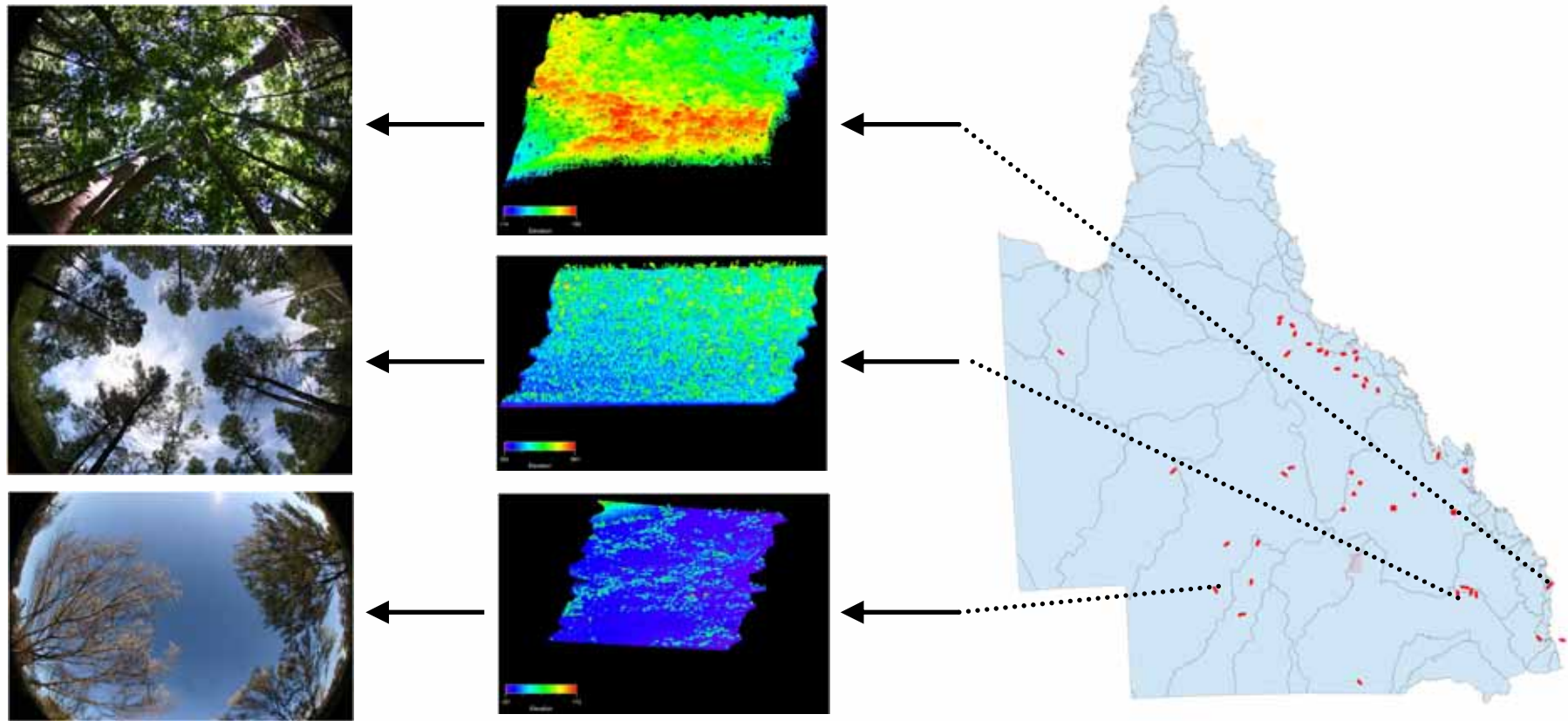


Coastal Example – Highly Accurate Lidar

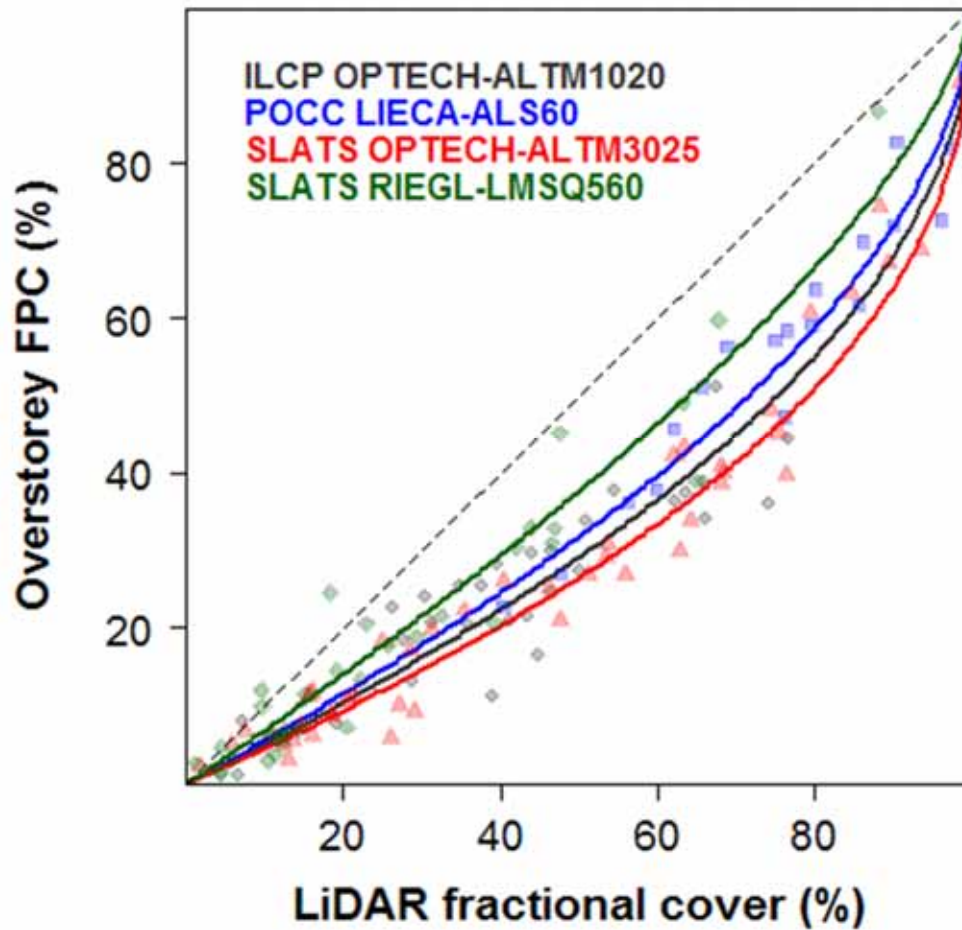


Vegetation Monitoring Sites

- 2-3 field + lidar campaigns between May 2004 and July 2010



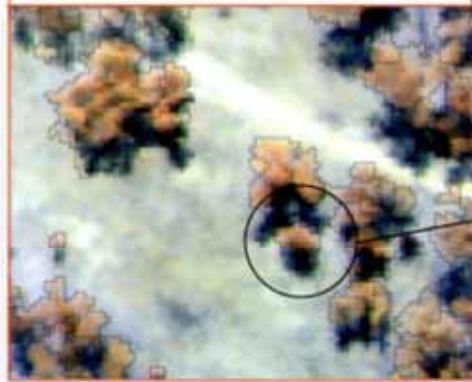
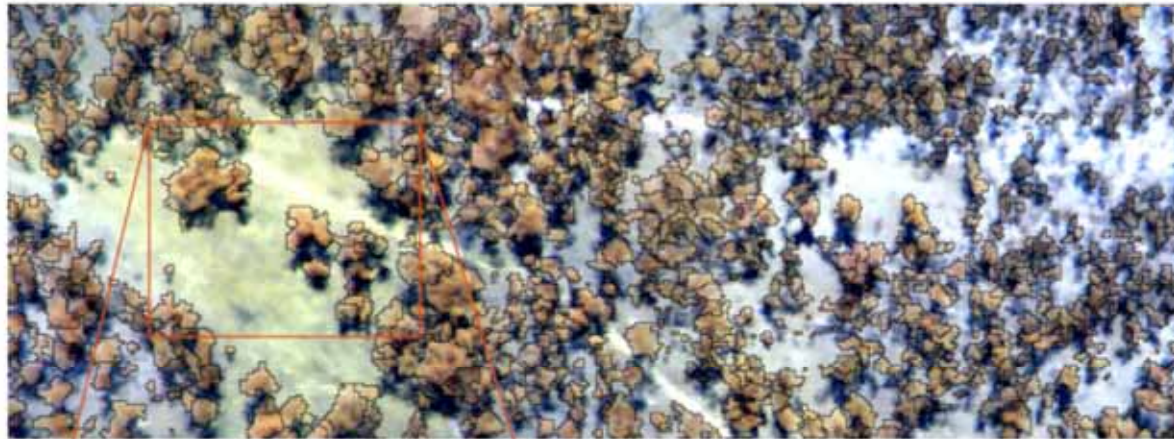
Comparison of discrete-return lidar Foliage Projective Cover (FPC) calibration over large areas



Why are they changing?

- Different sensor properties
- Different acquisition configurations
- Different field calibration datasets
- Variable non-green vegetation cover

Monitoring Trends in Canopy Condition



2000

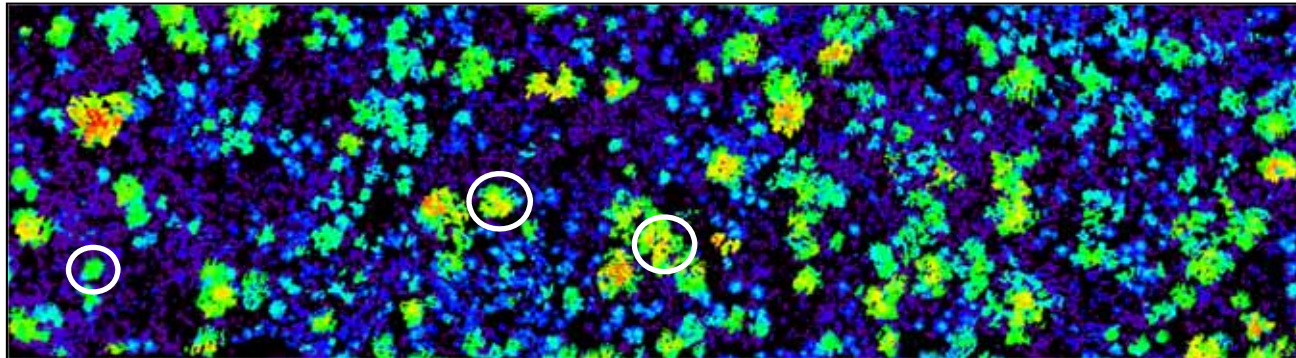


2006

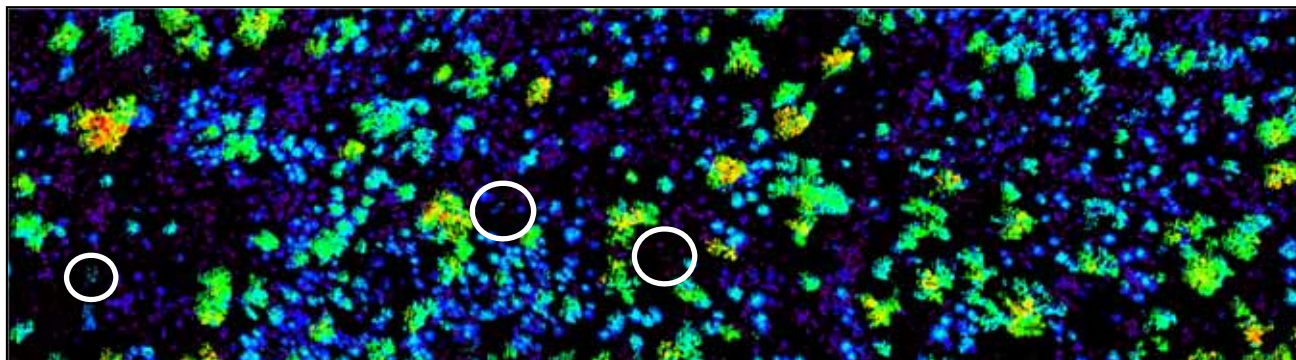
Location: Injune, Qld

Monitoring Trends in Canopy Condition

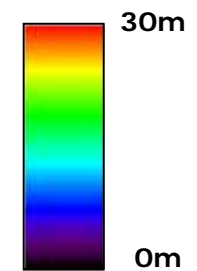
August 2000 – Optech ALTM1020



April 2009 – Riegl LMS-Q560



Height



Location: Injune, Qld

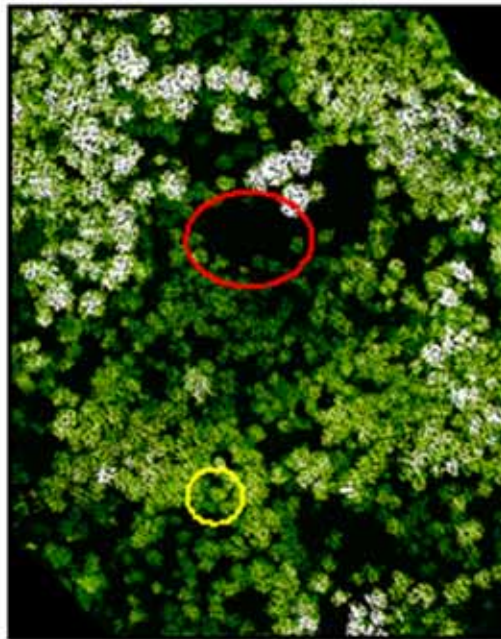
Monitoring Vegetation Change using multi-temporal LiDAR at Brisbane Forest Park, Qld



Maximum Return
Height



June 2005 – Optech 3025



April 2009 – Leica ALS60



Tree loss / dieback

Monitoring Considerations

- Design survey with monitoring in mind
- Use same sensor and specifications if possible
- Obtain raw data, as well as processed
- Consider changes in the environment which will affect DEM or vegetation points and derivatives
 - Water levels
 - Ground cover
 - Fire
- Work out resources required than triple that



RCAMES

OBSERVE • MODEL • SIMULATE



ROAMES Scope

- **Sense** – Gather information through sensors
 - Remote Observation (RO)
- **Model** – Construct working models of objects
 - Automated Modelling (AM)
- **Simulate** – Scenario management in an interactive context
 - Economic Simulation (ES)

everything in our power



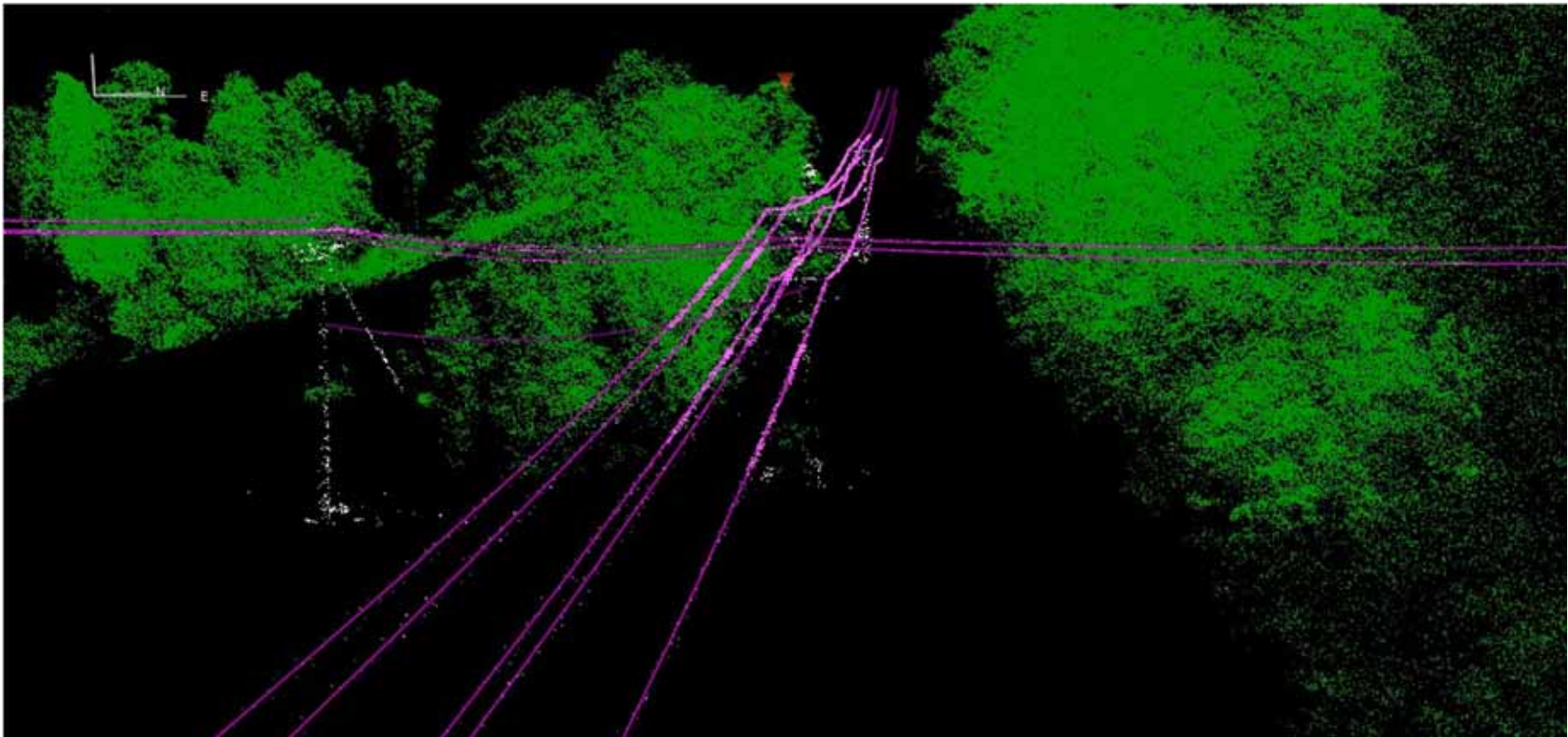
ROAMES

- Custom build scanning system
 - Dual scanners providing ~50 points/sqm
 - 7cm digital photography
- Precise measurement of Ergon Energy's electricity network and surrounding objects on an annual basis
- Up to 1000 reference stations across QLD



ROAMES

- Annual lidar data and imagery for 150,000 linear km across Qld
- Data likely to be available for researchers



Contact: enquiries@roames.com.au

Contributors

- John Armston, Nick Goodwin, Fiona Watson, Peter Todd, Peter Scarth, Dan Tindall, Kevin Chetwynd, Paul Brown (QLD Department of Environment and Resource Management)
- Glenn Jones (NSW Land and Property Management Authority)
- Rick Frisina (Victorian Department of Sustainability and Environment)
- Nathan Qadros (Cooperative Research Centre for Spatial Information)
- Phil Tickle, Nerida Wilson, Hamish Anderson (Geoscience Australia)
- Martin Mutendeudzi and Anthony Hunn (Australian Bureau of Agricultural and Resource Economics and Sciences)
- Darius Culvenor (CSIRO)
- David Jonas, Gail Kelly (AAMHatch)
- David Moore (Terranean)
- Jacky Croke (Griffith University)
- Stuart Phinn (University of Queensland)
- Matt Coleman, James Bangay (Ergon Energy)
- Peter Bunting, Richard Lucas (University of Aberystwyth)

Applying terrestrial LiDAR to derive gap fraction distribution time series during bud break

Kim Calders¹, Jan Verbesselt¹, Harm Bartholomeus¹ & Martin Herold¹

¹Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, P.O. box 47, 6700 AA Wageningen (NL); kim.calders@wur.nl

Abstract

The scientific community is witnessing a significant increase in the availability of different global satellite derived biophysical data sets. However, the use of such data is currently not supported by accurate in-situ biophysical measurement (e.g. canopy structure) in both a research and operational context for the monitoring of forest and land dynamics. Consequently, there is an urgent need for methods to measure in-situ canopy structure accurately and better integrate with improved and innovative remote sensing approaches. This paper explores the use of a ground-based, upward looking LiDAR instrument, combined with a fully automated analysis method to retrieve the gap fraction distribution. Traditional inventory methods for the assessment of forest structure are less objective or based on a 2D approach. We compare the seasonal dynamics of gap fraction distribution from hemispherical photographs and terrestrial LiDAR measurement during bud break.

Preliminary analysis shows that gap fraction distributions derived from terrestrial LiDAR were consistently lower than the values obtained from hemispherical photography. This might indicate that the LiDAR scans at the centre position of the plot are not representing the plot scale variation. However, the LiDAR based methodology is fully automated, requires no operator interference and is more objective, whereas the analysis of hemispherical photographs requires a large number of operator decisions (e.g. thresholding). Further improvements of this LiDAR-based method can still be achieved by (i) a better understanding of scanner settings and data resolution on the derived gap fraction and (ii) integration of target intensity in the analysis. This paper highlighted the high potential and need for a robust method to derive gap fraction distributions to monitor seasonal dynamics in forests.

Keywords: Terrestrial LiDAR, forestry, canopy structure, gap fraction, seasonal dynamics

1. Introduction

Forested areas play an important role in today's society and serve as a source for the production of paper products, lumber and fuel wood. In addition, forests produce freshwater from mountain watersheds, purify the air, offer habitat to wildlife and offer recreational opportunities. To keep these production, ecological and recreational functions balanced, accurate and precise information about forest structure and its biophysical parameters is needed (Warning & Running, 2007). Forest structure closely relates to several biological and physical processes. For example, respiration, transpiration, photosynthesis, carbon and nutrient cycle and rainfall interception heavily depend on the structural arrangement within the forest. Additionally, the terrestrial biosphere acts as a large sink for carbon dioxide (CO₂). In times where global warming is an important issue, accurate and precise measurements of forest variables are needed for the observations of Essential Climate Variables (ECVs) and for reducing emissions from deforestation and forest degradation (REDD). Remote sensing methods are useful tools to obtain this structural and biophysical information and they can be applied over extensive and inaccessible areas.

Hemispherical photography is an indirect optical method, which is commonly used for studying forest canopies. Based on the light attenuation and contrast between elements in the photo (i.e. sky and canopy) it can be used to capture structural information about the canopy architecture. Methodological errors can occur at any stage of image acquisition and analysis and derived structural parameters are therefore subject to several potential error sources (Jonckheere *et al.*, 2004). LiDAR (LIght Detection and Ranging) is an active remote sensing method. In the first instance, no real image is created but only a point cloud, which is obtained in the local coordinate system of the sensor (Pfeifer and Böhm, 2008; Wehr and Lohr, 1999). The basic principle used in laser altimetry is to measure, by some means, the distance (i.e. range) between the LiDAR instrument and the scanned target. Although LiDAR is not dependent on the natural illumination conditions, weather conditions during the data acquisition can be a source of error. The presence of mild to moderate wind will introduce noise in the results due to erratic movements of branches, twigs and foliage (Côté *et al.*, 2009). Danson *et al.* (2007) developed and validated a fully automated and objective method to derive gap fraction based on the xyz-coordinates of the acquired point cloud.

The purpose of this paper is to apply an objective method to assess gap fraction distribution time series during bud break. We will compare gap fraction distributions derived from traditional inventory methods and terrestrial LiDAR measurements at different moment during bud break. The overall canopy gap fraction is defined as the probability of incident radiation being able to pass unhindered through the canopy. Canopy directional gap fraction is the probability that a ray of light will not encounter a canopy element in a given direction. Gap fraction measurements can be used to derive related biophysical parameters such as leaf area index (LAI) (Jupp *et al.*, 2009).

2. Methodology

2.1 Study area

Data has been acquired in Hallerbos, a deciduous forest located in Flanders, Belgium. The region has a temperate maritime (North Atlantic) climate. Previous studies have indicated that these stands were homogenous and constituted of even-aged mature trees (Bequet *et al.*, 2011). Circular inventory plots were established in two mature European beech (*Fagus sylvatica* L.) stands where trees were older than 50 years and no understory was present. Data was collected during four field campaigns in the spring of 2011: April 14 and 15, April 25, May 6 and May 27. Due to extreme weather conditions earlier in the year bud break already initiated around April 7 and no data was collected in leaf-off conditions.

In order to avoid practical problems during the operation of the terrestrial LiDAR, the location of the plot was chosen so that no trees were in close proximity (i.e. closer than 1.5m) of the plot centre. Both plots have a radius of 18 m, establishing an area of approximately 1000 m². Additionally, a 16-points regular sampling pattern was designed for acquiring hemispherical photographs (Figure 1). Directional gap fraction values for the hemispherical photographs were averaged over these 16 photos, which cover the whole plot. This amount of photographs should be sufficient according to Fiala *et al.* (2006), who advised a minimum number of six to ten hemispherical photographs per plot.

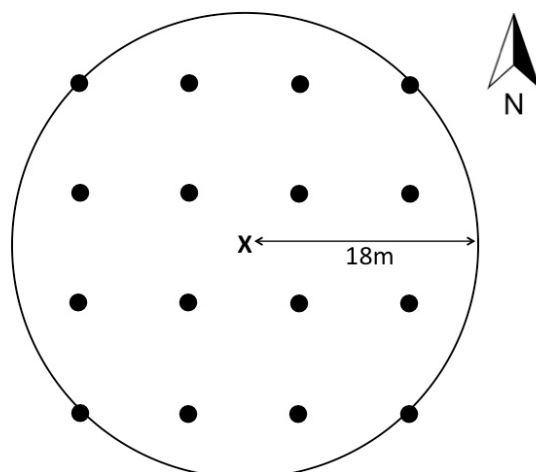


Figure 1: Schematic plot overview. LiDAR measurements were performed at location X using 2 different configurations. Hemispherical photographs were taken at each location on the 16-points regular grid.

2.2 Hemispherical photography

In this study a Nikon Coolpix 8700 (8 megapixels) fitted with an FC-E9 Fisheye lens (180 degrees field of view) was used. The camera was fixed on a pole and the position of the lens was levelled in the horizontal and vertical direction using a water level. The top of the lens was pointed towards the sky and located at 1.3 m above the ground surface. Pictures were taken at the highest resolution and quality with a fixed ISO at 200. Weather conditions during the different field campaigns varied but in general the photographs were taken in overcast sky conditions.

For the processing of the hemispherical photographs, the Hemisfer software (Swiss Federal Institute for Forest, Snow and Landscape Research WSL) was used (Schleppi *et al.*, 2007). Only the blue channel of the image was chosen to be included in the analysis since it gives the maximum contrast between leaf and sky (Zhao *et al.*, 2011) and the lens parameters were set to match those of the Nikon FC-E9. Thresholding is a crucial step in processing the hemispherical photos, which affects all further calculations. Manual thresholding is a subjective task and hard to reproduce consistently. A comparison study by Jonckheere *et al.* (2005) found that the method of Ridler & Calvard (1978) was most optimal and robust for a wide range of light and canopy structure conditions. Therefore this clustering-based automatic threshold method was applied to classify the pixels either as black (canopy) or as white (sky). This method was based on finding the midpoint between two clusters by iteratively estimating the average of two cluster means. The hemisphere was divided in 18 rings of 5 degrees and the directional gap fraction for each of those zenith rings was calculated. In the first step, gap fraction was derived within the defined rings based on the number of black and white pixels. These values were then corrected for non-linearity within the rings and ground slope (Schleppi *et al.*, 2007). Results were calculated for the weighted means of all zenith angles within a ring since there are more pixels along the outer border of a ring than along the inner border.

2.3 LiDAR measurements

The RIEGL VZ-400 3D terrestrial laser scanner was used to acquire terrestrial LiDAR data in the study area. This type of scanner uses a line scanning mechanism that is based upon a fast rotating multi-facet polygonal mirror. This leads to fully linear, unidirectional and parallel scan lines. The instrument operates in the infrared (1550 nm) and has a range up to 600 m. Due to

the on-board online waveform processing, the scanner is able to record multiple target echoes (RIEGL laser measurement systems, 2011).

LiDAR data was acquired at the centre location of the plot (Figure 1). The scanner was mounted on a tripod and tilted 90° . At the scan position two orthogonal scans were made. We consistently orientated the scanner so that the hemisphere was scanned in the East-West-direction by the first scan and the North-South-direction by the second scan. The frame scan angle was set to 180° (i.e. the upper hemisphere) and the line scan angle to 100° . Both frame and line resolutions were 0.04° (Table 1).

Table 1: RIEGL VZ-400 settings for the data acquisition

Frame scan angle	180°
Line scan angle	100°
Frame resolution	0.04°
Line resolution	0.04°
Inclination angle	90°
Measurement time	1 min 33 sec

Directional gap fraction is derived from the method of Danson *et al.* (2007). This method checks whether or not a fired laser pulse hits a scene element on its way. When using this approach, multiple returns do not really have an added value and therefore the data was first filtered in RiSCAN PRO, exporting only the first returns to an ASCII file. The original scan holds approximately 22 to 24 million points, after filtering only 15 to 16 million points were left. The data was then further analysis in MATLAB. The original MATLAB code written by Danson *et al.* (2007) was modified for the use with RIEGL VZ-400 input data. In the first part of the analysis the number of hits in each zenith ring is calculated. This was achieved by transforming the original Cartesian coordinates to spherical coordinates so they could be evaluated against the selected zenith angles.

The RIEGL scanner does not record a return echo if the laser pulse hit no scene element. Therefore the number of fired laser pulses during the scan (i.e. equal to the maximum amount of potential first return echoes) was calculated in the second part of the analysis. This number is dependent on the line and frame scan angles and their corresponding resolution. Based on the number of measured (i.e. real) and modelled (i.e. maximal) hits, directional gap fraction per zenith ring could be calculated.

3. Results

Gap fraction derived from the two orthogonal scan positions showed good agreement at both plots (Figure 2). This behaviour was consistent throughout the different measurement days in spring. Therefore, the LiDAR derived gap fraction values are represented as the average over the two orthogonal scan positions.

The evolution of the directional gap fraction during bud break based on the LiDAR approach is derived for both plots and summarised in Figure 3. The average over the 16 hemispherical photographs was taken to represent the plot average (Figure 4). These values were then compared with the gap fraction estimates from the terrestrial LiDAR scans at different moments in spring (Figure 5).

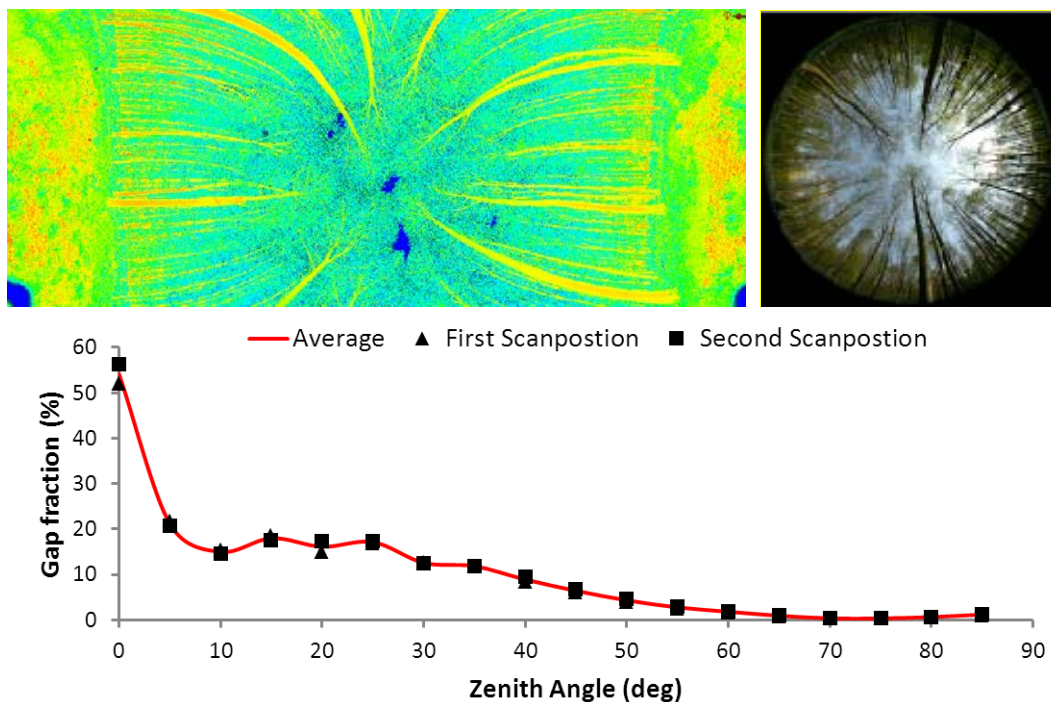


Figure 2: (Top) Intensity of LiDAR returns in a cylindrical projection [left] & Hemispherical photograph of about the centre of the plot (i.e. on one of the inner sampling points of the 16-points regular grid) [right]. (Bottom) Comparison of LiDAR derived gap fraction from two orthogonal scan positions (Plot 2: 15-04-2011).

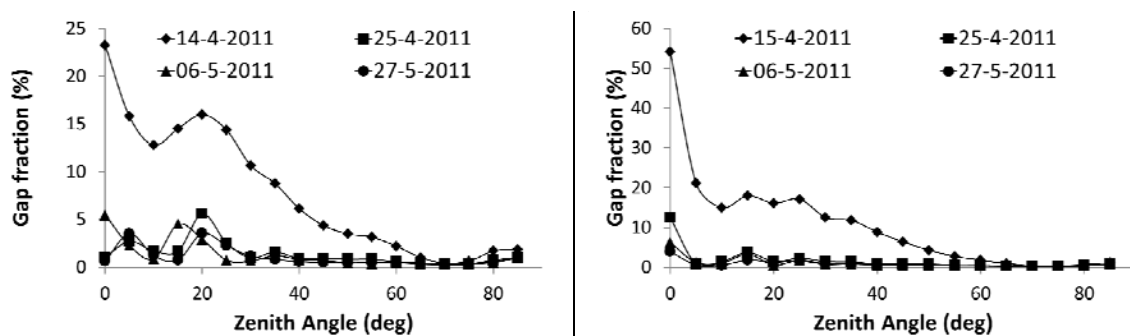


Figure 3: Directional gap fraction during bud break derived from terrestrial LiDAR scans. (Left) Plot 1. (Right) Plot 2.

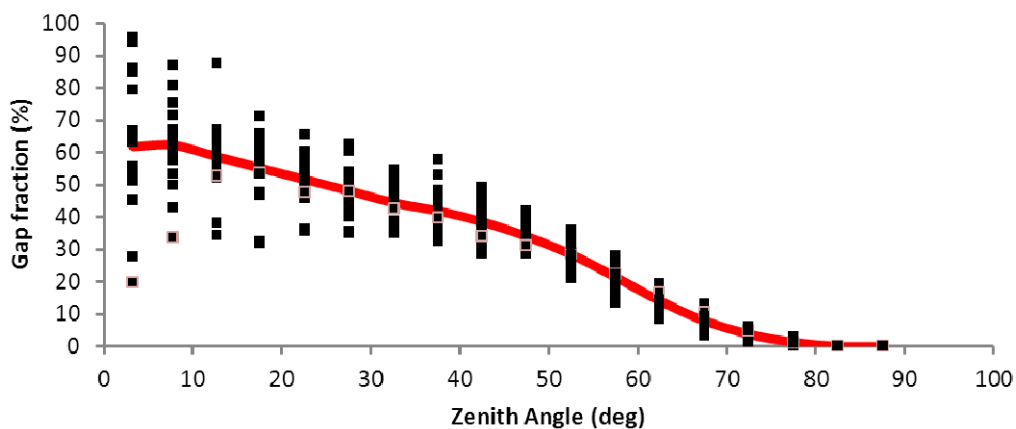


Figure 4: Gap fraction derived from hemispherical photographs. The solid line shows the plot average over the 16 individual locations (Plot 2: 15-04-2011).

Finally, the relative differences between the four points in time were calculated for both methods (Figure 6). Period 1 represent the changes in gap fraction distribution between April 14/15 and April 25, period 2 represents the changes between April 25 and May 6 and the third period represents the changes between May 6 and May 27. Values from the later date are subtracted from the earlier date.

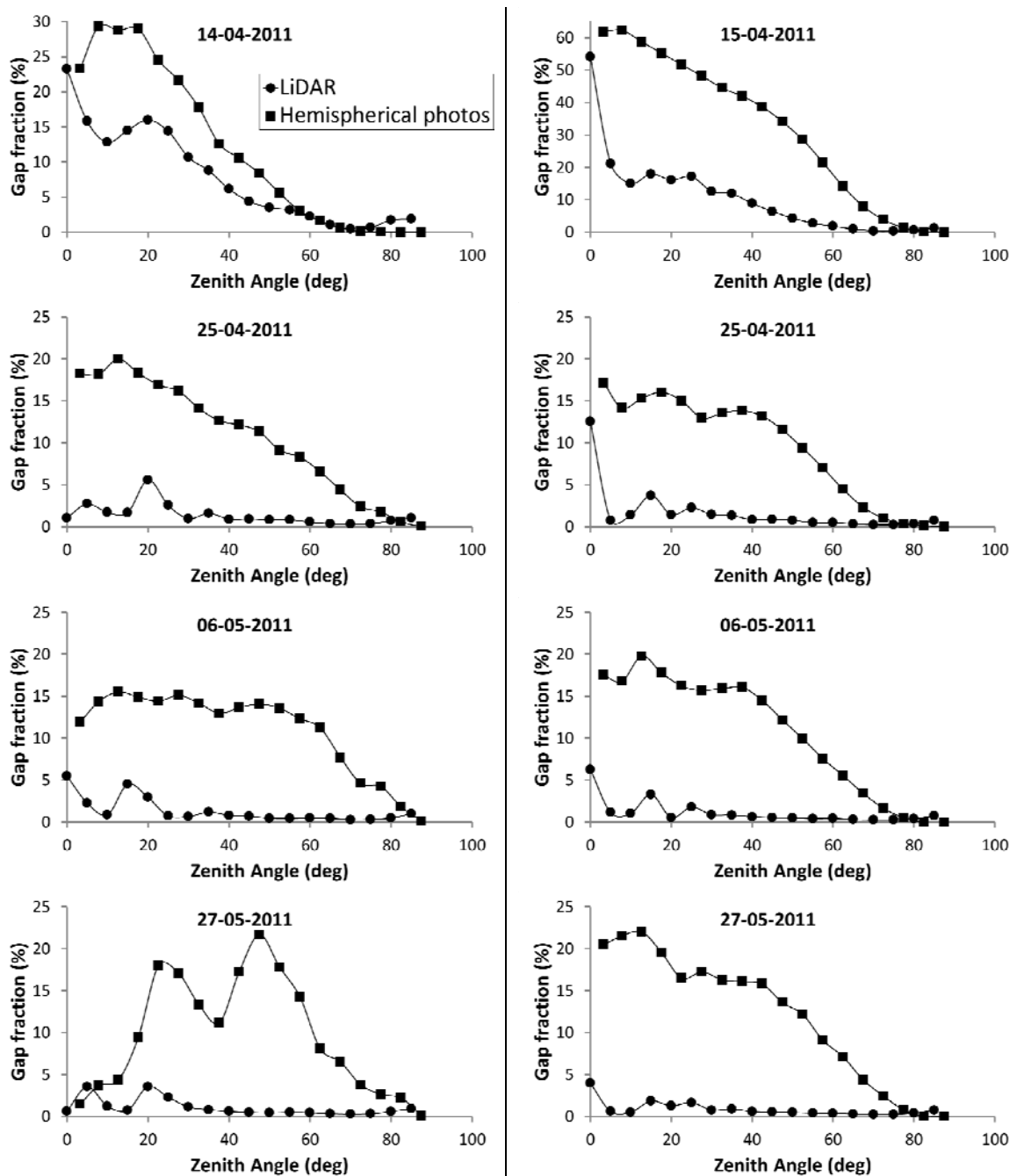


Figure 5: Comparison of gap fraction estimated from terrestrial LiDAR scans and hemispherical photos at different moment in time during spring. (Left) Plot 1. (Right) Plot 2.

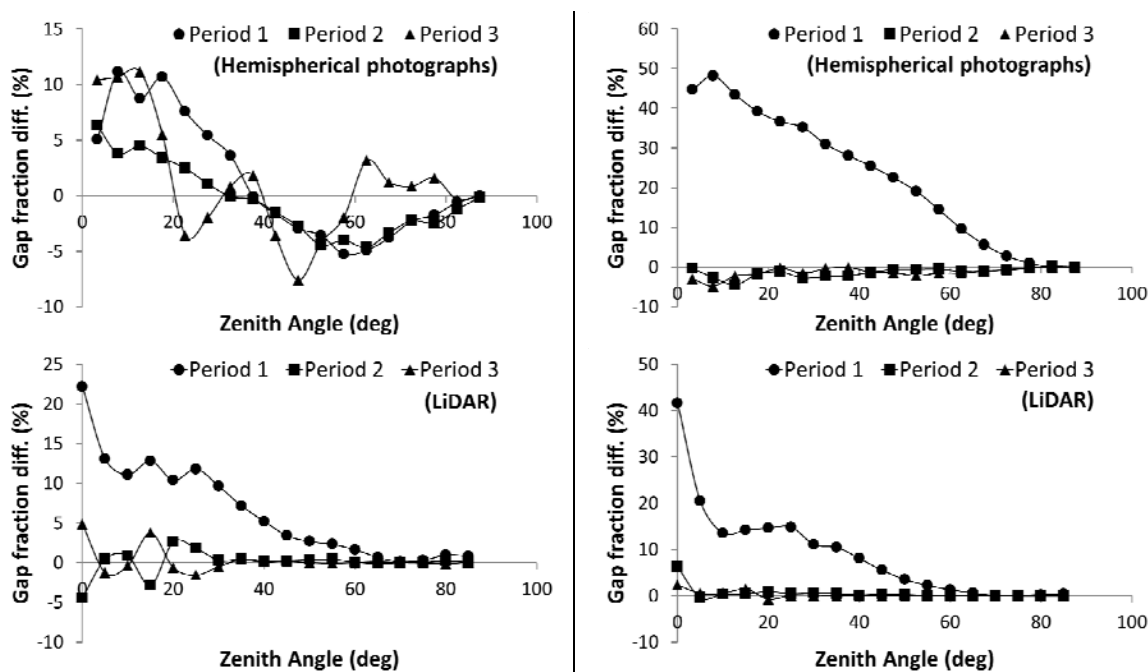


Figure 6: Comparison of the differences in gap fraction between the different time periods for both methods. Period 1: 14/15-04 to 25-04; Period 2: 25-04 to 06-05; Period 3: 06-05 to 27-05. (Left) Plot 1. (Right) Plot 2.

4. Discussion

The directional gap fraction estimates based on terrestrial LiDAR data show similar behaviour for both orthogonal scan positions (Figure 2). These observations are similar to Danson *et al.* (2007) and indicate that the orientation of the scanner in the xy-plane does not affect the obtained gap fraction values.

Overall, there is an underestimation of the LiDAR derived gap fraction compared to the values derived from hemispherical photography (Figure 5). Although the overall trend over the zenith angles is similar, it is more pronounced for the values derived from the hemispherical photos. These results contrast the observations of Danson *et al.* (2007), who found that the calculated gap fraction distributions derived from terrestrial LiDAR only slightly overestimated the values calculated from hemispherical photography. In that study only one hemispherical photograph was taken, located at the exact same position of the scanner, whereas we took the plot average. Figure 4 shows that the directional gap fraction distribution derived from hemispherical photographs can greatly vary within the plot, especially at low zenith angles. It is suggested that LiDAR derived gap fraction shows similar behaviour and a single terrestrial LiDAR scan will therefore not always represent the plot directional gap fraction distribution. Further analysis is needed to test this hypothesis before a firm and strong conclusion can be drawn since other factors might cause this underestimation as well (e.g. scanner settings and data resolution). Therefore, further work is needed to understand the effect of line and frame scan resolution on the derived gap fraction distribution. With the instrument settings used in this work (Table 1) 11245500 laser shots were fired into the scene. This leads to approximately 4 times more data points compared to the hemispherical photographs, which only had 2955806 pixels for analysis. A reduced scan resolution will increase the operating time significantly but will consequently also capture less detailed data (Table 2). As an indication, a scan with both resolutions set to 0.05° will only need 1 minute and with a resolution of 0.1° operation time is reduced to 15 seconds. Additionally, including target intensity values in the LiDAR analysis is expected to improve the derived gap fraction distribution and enable detection of sub-footprint gaps

(Hancock, 2010).

Table 2: Different frame and line scan resolution settings for scan angles 180° x 100°

Frame resolution	Line resolution	# laser shots fired
0.04°	0.04°	11245500
0.05°	0.05°	7198000
0.07°	0.07°	3671388
0.1°	0.1°	1801800

Preliminary time series analysis (Figure 6) illustrates the expected decrease in gap fraction over time caused by canopy closure. Although absolute values differ between both methods, some preliminary conclusions can be drawn from these graphs. The canopy closure in plot 2 happened almost entirely in period 1. Plot 1 shows a more gradual canopy closure over the 3 periods, especially for the gap fraction derived from hemispherical photographs. These values suggest a gradual decrease in gap fraction in the smaller zenith angles and an increase in gap fraction in the larger zenith angles. This temporal behaviour is not supported by the LiDAR results, which show small gap fraction fluctuations in the lower zenith angles and no changes in the larger zenith angles. It is therefore necessary to first conduct the further research suggested in the paragraph above to come to a consistent and reproducible method that can be used to monitor gap fraction distribution changes during bud break.

5. Conclusion

In this work the seasonal dynamics of gap fraction distribution for European beech were derived from terrestrial LiDAR data and hemispherical photographs. Especially in the context of time series analysis, a consistent and reproducible tool is preferred for deriving structural information. Although gap fraction distribution derived from terrestrial laser scanning showed an underestimation compared to the values calculated from hemispherical photographs, the high potential of a robust and automated LiDAR method is highlighted. Since LiDAR is an active remote sensing tool it is also not dependant on special conditions for natural light and does not require user interaction.

Acknowledgements

The original MATLAB code for deriving gap fraction distributions from terrestrial LiDAR was developed and made available by Mark Danson. We would also like to thank him for his useful support.

References

- Bequet R., Kint V., Campioli M., Vansteenkiste D., Muys B. & Ceulemans R., 2011. Influence of stand, site and meteorological variables on the maximum leaf area index of beech, oak and Scots pine. *European Journal of Forest Research*, in press. DOI 10.1007/s10342-011-0500-x
- Côté, J.-F., Widlowski J.-L., Fournier, R. A. & Verstraete, M. M. 2009. The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial lidar. *Remote Sensing of Environment*, 113, 1067-1081.
- Danson, F.M., Hetherington, D., Morsdorf, F., Koetz, B. & Allgöwer, B., 2007. Forest canopy gap fraction from terrestrial laser scanning. *IEEE Geoscience and Remote Sensing Letters*, 4, 157-160.

- Fiala, A. C. S., Garman, S. L., & Gray, A. N., 2006. Comparison of five canopy cover estimation techniques in the western Oregon Cascades. *Forest ecology and management*, 232, 188-197.
- Hancock, S., 2010. Understanding the measurement of forests with waveform lidar. PhD thesis, UCL Geography
- Jonckheere, I., Nackaerts, K., Muys, B. & Coppin, P., 2005. Assessment of automatic gap fraction estimation of forests from digital hemispherical photography. *Agricultural and Forest Meteorology*, 132, 96-114.
- Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M. & Baret, F., 2004. Review of methods for in situ leaf area index determination Part I. Theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology*, 121, 19-35.
- Jupp, D.L.B., Culvenor, D.S., Lovell, J.L., Newnham, G.J., Strahler, A.H. & Woodcock, C.E., 2009. Estimating forest LAI profiles and structural parameters using a ground-based laser called Echidna®. *Tree physiology*, 29, 171-181.
- Pfeifer, N. & Böhm, J., 2008. Early stages of LiDAR data processing. *In Advances in photogrammetry, remote sensing and spatial information sciences: 2008 ISPRS congress book*
- Ridler W. & Calvard S., 1978. Picture thresholding using an iterative selection method. *IEEE Trans. Syst. Man Cyber.*, 8, 260-263.
- RIEGL laser measurement systems, 2011, www.riegl.com, accessed 4th June 2011
- Schleppi P., Conedera M., Sedivy I., Thimonier A., 2007. Correcting non-linearity and slope effects in the estimation of the leaf area index of forests from hemispherical photographs. *Agric. Forest Meteorol.*, 144, 236-242.
- Warning, R. H. & Running, S. W., 2007. Forest ecosystems analysis at multiple scales, 3rd edn, Elsevier, 420p.
- Wehr, A., U. Lohr. 1999. Airborne laser scanning - An introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54, 68-82.
- Zhao, F., Yang, X., Schull, M.A., Román-Colón, M.O., Yao, T., Wang, Z., Zhang, Q., Jupp, D.L.B., Lovell, J.L., Culvenor, D.S., Newnham, G.J., Richardson, A.D., Ni-Meister, W., Schaaf, C.L., Woodcock, C.E. & Strahler, A.H., 2011. Measuring effective leaf area index, foliage profile, and stand height in New England forest stands using a full-waveform ground-based lidar. *Remote Sensing of Environment*, in press. DOI 10.1016/j.rse.2010.08.030

Foliage Profiles from Ground Based Waveform and Discrete Point Lidar

J.L. Lovell¹, D.L.B. Jupp², E. van Gersel², J. Jimenez-Berni², C. Hopkinson³ & L. Chasmer⁴

¹CSIRO Marine and Atmospheric Research, Hobart, Australia, Jenny.Lovell@csiro.au

²CSIRO Marine and Atmospheric Research, Canberra, Australia

³Applied Geomatics Research Group, Lawrencetown, Canada

⁴Wilfrid Laurier University, Waterloo, Canada

Abstract

Terrestrial lidar systems provide a means to characterise the structure of a forest canopy. Their use to measure foliage area volume density depends on the ability to account for sampling effects and intensity calibration of the instrument. This paper presents a theoretical framework for the unbiased calculation of foliage amount using a waveform recording lidar instrument to simulate point cloud data. The method is initially based on the hemispherical scan configuration of the instrument, but is generalised to be applied to point cloud data in a generic coordinate system. The theory is tested with the simulated point cloud data as well as data from a commercial instrument. Foliage profiles from the terrestrial lidar instruments and airborne lidar are compared.

1. Introduction

Leaf area index (LAI) and foliage area volume density (FAVD) are important quantities in the study of the structure and function of canopies e.g. light interception, respiration, transpiration, photosynthesis in multi-layer canopies all depend on these. While a few options exist for ground-based measurement of LAI, profiles of FAVD are more difficult as measurements are needed throughout the height of the canopy. This has been done successfully from the air with both lidar (e.g. Lefsky *et al.* 1999) and radar (Imhoff *et al.* 2000) but these methods often require on-ground calibration. Airborne instruments are also limited in the range of angular sampling and ability to sense structure through the full depth of a dense canopy. This is where ground-based lidar is an attractive option.

Parker *et al.* (2004) used a simple laser rangefinder to sample vertical gap probabilities along a transect and thus calculated vertical foliage profiles. Takeda *et al.* (2005, 2008) also used a simple rangefinder system, but incorporated a 2-axis scan platform to allow angular sampling of the canopy. Their measurements of plant area density within gridded volume elements (voxels) resulted in plant area indices that reflected seasonal variation in the canopy, but significantly overestimated the actual amount of plant material. Hosoi and Omasa (2006) acquired multiple high resolution scans of single trees and used a ray tracing method to calculate contact frequency within voxels. From this they derived profiles of leaf area density which were validated against stratified clipping. Van der Zande *et al.* (2008) extended the Hosoi and Omasa (2006) methodology in a simulation study with virtual forest stands.

Jupp *et al.* (2009) presented a method for estimating LAI profiles using a full-waveform ground-based lidar system. The terrestrial lidar system (TLS) used in the study was the Echidna® validation instrument (EVI) which has the advantage of scanning the full upper hemisphere with no gaps in laser illumination. The geometry of the scan (zeniths and azimuths of the outgoing beams) is also recorded along with the intensity profile of all target reflections.

In TLS that record discrete target locations, there is still the potential to estimate foliage amount and distribution provided sufficient information is recorded to characterise the geometry of the scan pattern and the intensity of the returns. In this paper, we summarise the method of Jupp *et al.* (2009) and develop an equivalent method using discrete point lidar data, based on the polar geometry of the EVI scan. We then develop a voxel-based method for the calculation of the foliage amount per voxel using discrete point lidar data in a general coordinate system. We use a single EVI scan to illustrate the way in which the waveform-based foliage profile method can be modified for use with discrete point data and then use the same data to test the voxel method. Scans of the same site using a discrete-point TLS system are also used to test the voxel method.

2. Theory

Jupp *et al.* (2009) present a method of estimating LAI profiles using a vertically resolved gap probability distribution, P_{gap} .

$$P_{gap}(\theta, z) = e^{-G(\theta)L(z)/\cos\theta} \quad (1)$$

where θ is zenith angle, z is height, $G(\theta)$ is the Ross G-function (Ross, 1981) and L is LAI. Provided with an estimate of $P_{gap}(\theta, z)$, the profile of leaf area can be calculated and thus the foliage area volume density which is the derivative of $L(z)$.

We now explore how P_{gap} can be estimated from TLS data, first following the Jupp *et al.* (2009) method using EVI data, then modifying this for application to discrete point data.

2.1 Hemispheric Waveform Method

The EVI waveform data processing uses a quantity called apparent reflectance. This is the reflectance of a diffuse target filling the beam of the instrument that would return the same intensity as recorded from the actual target. For a waveform recorded at zenith angle, θ , over ranges, r , it has the form

$$\rho_a = \frac{I(\theta, r)R^2}{K(R)\Phi_0} \quad (2)$$

where I is the range-dependent recorded intensity, R is the range to the target, $K(R)$ is a calibration function associated with the geometry of the receiver optics and Φ_0 is the energy of the outgoing pulse. Integrating ρ_a over range provides a step-wise reduction in the power of the outgoing signal brought about by hits on single or multiple targets. This is related to P_{gap} by

$$I_a(\theta, r) = 1 - p(\theta, g)(1 - P_{gap}(\theta, r)) \quad (3)$$

where I_a is the integral of ρ_a , g is the distribution function for facet directions of the targets and p is the mean phase function for the varying facets. Jupp *et al.* (2009) take the phase function as the square of the Ross G-function. In general, the phase function is unknown and if possible should be estimated from the data. Jupp *et al.* (2009) use an initial assumption of $p=1$ and then identify two thresholds in the calculated P_{gap} relating to (i) the maximum P_{gap} value for targets

that fully extinguish the beam (hard target) and (ii) the maximum P_{gap} value for targets that partially extinguish it (soft target), above which all samples are assumed to be true gaps. These are used to scale the P_{gap} in a similar way to the two level separation of gap and vegetation that can be done in hemispherical photograph analysis (Leblanc *et al.*, 2005).

The value of P_{gap} calculated from a single waveform is a realisation of an actual gap, rather than a probability, therefore it is necessary to average the measured values over some spatial region in order to estimate the underlying probability distribution. Using EVI data, it is convenient to average over a ring or sector between zenith angle limits. Jupp *et al.* (2009) describe a method for calculation of a mean foliage profile from zenith-ring averages of P_{gap} . The method uses a ratio of cumulative foliage area ($L(z)$) relative to LAI to provide a profile largely independent of clumping. The range-based zenith-ring averaged P_{gap} data are resampled to common points on the height axis, z and the cumulative LAI profile is defined by

$$\frac{L(z)}{LAI} = \frac{\log P_{gap}(\bar{\theta}, z)}{\log P_{gap}(\bar{\theta}, H)} \quad (4)$$

where H is the height of the canopy and the notation $\bar{\theta}$ indicates that the data are averaged over a range of zenith angles, rather than representing a mean angle. The foliage area volume density profile is then

$$f(z) = LAI \frac{\partial}{\partial z} \left(\frac{\log P_{gap}(\bar{\theta}, z)}{\log P_{gap}(\bar{\theta}, H)} \right). \quad (5)$$

In these equations the value of LAI can also be estimated from the EVI data either from data near the zenith angle $\theta \approx 57.5^\circ$ or from a simple linear canopy model as shown by Jupp *et al.* (2009). The profiles are calculated for a number of zenith rings i.e. different values of $\bar{\theta}$, and then a mean profile is calculated by weighting each profile according to the solid angle subtended by the ring. Jupp *et al.* (2009) note that the measurements based on the EVI data do not separate plant material into leaf and stem. Thus the quantities calculated are in fact plant area index and plant area volume density, however we will maintain the notation of the previous work. This paper is concerned with the comparison of methods applied to two sets of lidar data so the distinction between plant and leaf is unnecessary in this case.

2.2 Hemispheric Discrete Point Method

Waveform EVI data can be converted to (x, y, z) point data by applying a filter based on the known shape of the outgoing laser pulse. There may be single or multiple targets identified from each waveform. The output records from the conversion process include x, y, z coordinates, apparent reflectance, outgoing zenith and azimuth angles, the number of hits from the shot and the hit number of the particular point. There is also a record of shots for which no hits were detected i.e. sky gaps. This point cloud data retains the geometric and sampling advantages of the waveform data but also allows us to demonstrate that it is possible to produce equivalent foliage profile information from discrete data.

We now develop a method to calculate P_{gap} from the point cloud data. It is again useful to accumulate data over zenith rings or sectors. We first define an unscaled P_{gap} term, pg :

$$\begin{aligned}
 pg(\bar{\theta}, r) &= 1 - P_{hit}(\bar{\theta}, r) \\
 &= 1 - \frac{\sum_0^r hit_w}{n_{shot} vol}
 \end{aligned} \tag{6}$$

where hit_w is a weighting value for each hit up to range, r , n_{shot} is the total number of outgoing pulses in the zenith range and vol is the illuminated volume in the zenith range. The weighting of the hits is either the apparent reflectance value of the target points, or a weighting defined by the number of hits from a single shot (e.g. points recorded from a waveform with n hits detected would each be assigned a weight of $1/n$).

The illuminated volume, vol , is a factor that is necessary to take into account the obscuration of some regions by targets at closer range to the instrument. It can be calculated from the point cloud data by identifying the last hit from each shot and calculating the volume contained within these final hits.

The quantity pg needs to be scaled to account for the phase effect. We do this by applying a simple linear scale defined to rescale the quantity to achieve a value of $P_{gap}=1$ for true sky gaps.

$$P_{gap}(\bar{\theta}, r) = 1 - \frac{pg(\bar{\theta}, r)(1 - skyratio)}{(1 - pg(\bar{\theta}, r_{max}))} \tag{7}$$

where $skyratio$ is defined within the same zenith ring or sector as the ratio of the number of shots where no hits were recorded, to the total number of shots and r_{max} is the maximum range covered by the data.

For each of the zenith rings, we resample the P_{gap} to common points on the height axis to give $P_{gap}(\bar{\theta}, z)$ which can be used in equations (4) and (5) in the same way as the waveform-based P_{gap} .

2.3 Voxel Method

In order to develop a more general algorithm that can be applied to point data from different TLS instruments, we now develop the theory from the perspective of voxels. We define foliage area volume density, $f(\theta, \phi, r)$ for a voxel at polar coordinates (θ, ϕ, r) such that

$$\int_0^r f(\theta, \phi, r) dr' = \frac{L(z)}{\cos \theta} \tag{8}$$

where $z = r \cos \theta$. Consider now the illumination of the foliage elements in this voxel by a laser which has passed through the canopy to this point with a gap probability, $P_{gap}(\theta, \phi, r)$. The observed apparent reflectance of the voxel, ρ_a can be expressed as

$$\rho_a = G(\theta) p(g, \theta) P_{gap}(\theta, \phi, r) f(\theta, \phi, r) \rho_t. \tag{9}$$

The total effective area of objects in the voxel is $f(\theta, \phi, r)$ multiplied by the volume of the voxel. It is useful at this point to generalise the expressions by converting to Cartesian coordinates, thus

$$LA_V = \frac{V}{G(\theta) p(g, \theta) P_{gap}(x, y, z)} \rho_a(x, y, z) \quad (10)$$

Where LA_V is the leaf area within the voxel and V is the volume of the voxel. If we divide by the base area of the voxel, we obtain an expression for a leaf area index of the voxel:

$$L_V = \frac{h_z}{G(\theta) p(g, \theta) P_{gap}(x, y, z)} \rho_a(x, y, z) \quad (11)$$

where h_z is the height of the voxel. This can be summed vertically to obtain canopy LAI above the base of the voxel. This provides an algorithm to calculate leaf (or plant) area in voxels given the apparent reflectance returned from that voxel and an estimate of the gap probability at that point in the canopy. This expression is valid only for voxels that are illuminated i.e. not obscured from the instrument due to closer canopy elements. It is therefore best applied in situations where there are multiple scans that provide consistent illumination and overcome obscuration. Where there is incomplete illumination, equation (11) will provide an approximation to the true foliage area volume density sampled according to the illumination of the volume. If the illuminated volume can be characterised, a correction could be made, however it has not been attempted at this stage.

3. Data Collection

Data were collected from 7-10 December 2009 in a tall eucalypt forest near Tumbarumba, Australia (35°36'42"S, 148°06'29"E) at the location of the Tumbarumba FLUXNET site (Leuning *et al.*, 2005). The forest has a 2-layer canopy and significant ground cover of shrubs and grasses. Multiple-return airborne lidar (ALS) were acquired over an area approximately 6 km by 4 km. Terrestrial lidar data were collected with the EVI waveform lidar and an ILRIS discrete return system with settings as shown in Table 1.

Table 1: Terrestrial lidar specifications and settings

	EVI	ILRIS
Wavelength	1064 nm	1550 nm
Pulse repetition frequency	2 kHz	2 kHz
Beam divergence	5 mrad	0.2 mrad
Beam diameter at 50 m	28 cm	1 cm
Field of view	360° azimuth, 130° zenith	50° horizontal and vertical
Number of returns	Waveform digitised at 2 GSs ⁻¹	Single (last)

The EVI dataset is a single hemispherical scan comprising data from the full upper hemisphere covering the field of view with no gaps in laser illumination. The direction (zenith and azimuth) of the outgoing laser pulse are also recorded. Further details of the EVI instrument are given by Jupp *et al.* (2005).

ILRIS is a tripod-mounted eye safe lidar imaging system manufactured by Optech Incorporated, Toronto, Canada. Ranges of over 1 km can be recorded. The settings can be configured either

for speed of data collection or for high data density. Spot spacing is a function of the user configurable horizontal and vertical field of view and distance from the sensor. A nominal spacing at a 50 m range would be on the order of 1 cm. ILRIS scans were undertaken at the same location as the EVI as well as from the four cardinal directions looking in towards the EVI location at a distance of approximately 80-100 m as illustrated in Figure 1. At each of these locations, a ground-parallel and an upwardly inclined canopy-viewing scan were collected, recording the last return for each pulse. The location of each ILRIS position was recorded using a dGPS rover mounted on top of the sensor head. The dGPS positions were differentially corrected to a base station less than 10 km away. However, given the GPS occupations were under canopy and the receiver used was a single frequency (L1 only), only one scan was positioned to the cm level. The remaining locations are accurate to within approximately 0.5 m.

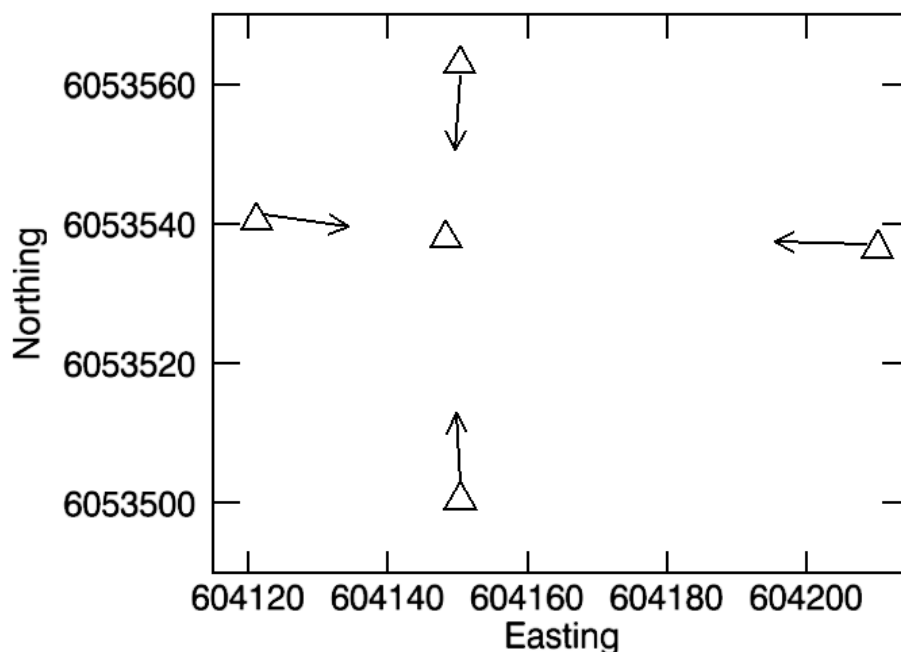


Figure 1: ILRIS scan configuration with 4 scan locations looking into centre of the EVI scan location from north, east, south and west ILRIS scan locations. Each ILRIS scan location has two scans: one parallel with ground and another inclined vertically and aimed into the canopy.

The images in Figure 2 show a comparison of the two terrestrial lidar datasets. The ILRIS data are shown for a square subset ± 40 m from the EVI location as that data covers a larger area and is more dense. The two images are approximately aligned. It is clear that the single viewpoint of the EVI data suffers from occlusion in some areas e.g. there are sectors occluded by the trunks of central trees. These are evident to some extent in the ILRIS data, but are mostly filled by data from different view directions.

4. Demonstration of Results

The EVI waveform data gives us the flexibility to test the theoretical methods using a single dataset, processed and sampled in varying ways. The first and baseline foliage profile that we derived is the waveform-based profile following the method of Jupp *et al.* (2009). This is shown in Figure 3a as the blue line. It shows a 2-layer canopy as expected and sums to an LAI value of 2.4 which is comparable to published values for this site (Leuning *et al.* 2005). Also shown in this figure are a foliage profile derived from the ALS data (red line) for a 40 m by 40 m area and a mean profile derived from 9 areas of the same size in a 3x3 grid centred on the EVI location.

These foliage profiles have been scaled to the expected LAI of the site. Error bars indicate the standard deviation of the FAVD. This demonstrates the spatial variability of the site and indicates uncertainties that may be expected due to errors in positioning the EVI data within the ALS dataset. The FAVD in the understorey is less in the ALS profiles than the EVI profile due to occlusion by the upper canopy.

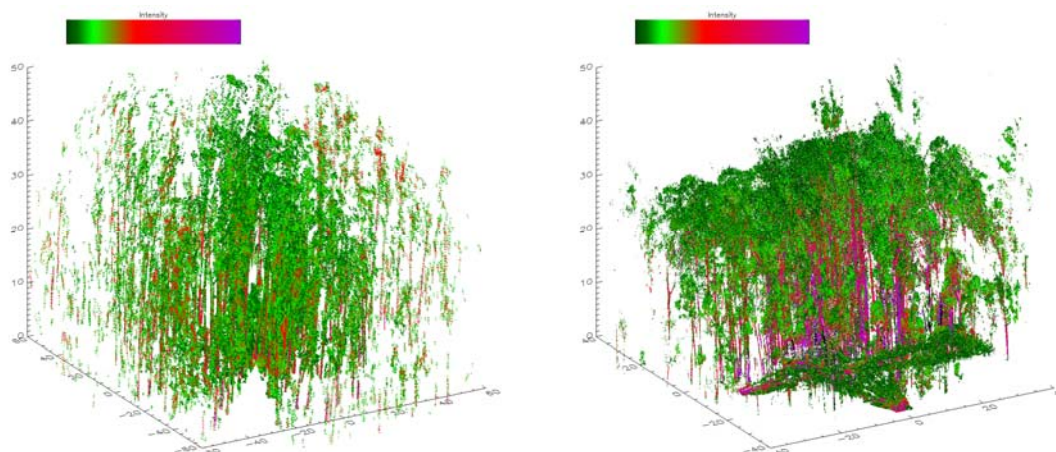


Figure 2: EVI data (left) shown as a point cloud derived from the waveform data and a subset of the ILRIS point cloud (right). The extent of the EVI data is approximately 50 m from the centre of the plot. The ILRIS data has been subset to 80 m x 80 m. The maximum tree height is approximately 40 m. Colours represent intensity of lidar returns.

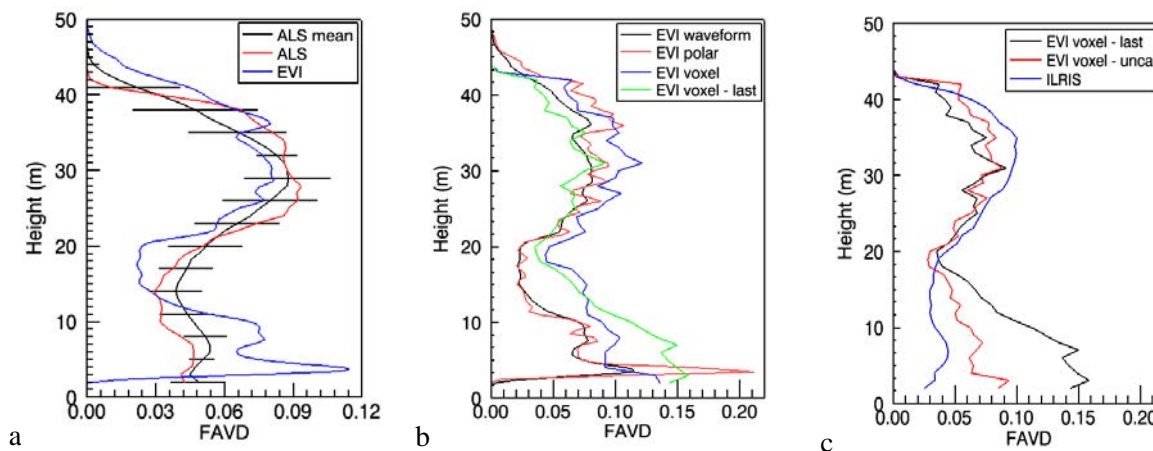


Figure 3: Comparison of foliage profiles: (a) ALS and EVI waveform; (b) EVI various methods and (c) ILRIS foliage profiles compared with the EVI last hit data (right).

Figure 3b shows a comparison of foliage profiles from the EVI data using the waveform, polar point cloud and voxel methods. The EVI waveform profile (black line) is the same as that shown in Figure 3a. The EVI simulated point cloud data were processed using equation (7) to produce gap probabilities for a number of equal zenith angle rings. These are then used in the same way as the waveform-based P_{gap} results to calculate a mean foliage profile over all zenith angles (equation (5)). This process has been done using two kinds of weighting in equation (6). The primary method of weighting (shown in red in Figure 3b) is the apparent reflectance calculated for the point. The alternative, which is of use where calibrated intensity data are not available, is to weight points according to the number of hits recorded in the shot. This produces a similar result and is not shown here. The point cloud result is not as smooth a curve as the

waveform result. This may be due at least in part to the calculation of the illuminated volume which is based on the assumption that all final hits fully extinguish the beam. This is not always true and will introduce some error into the calculation.

The voxel method described in Section 2.1 takes a different approach, calculating the foliage area volume density represented by each hit and assigning this to the appropriate spatial voxel in a grid. Where multiple points fall into a voxel, the average value is used. The theory has been described for multiple-return lidar data, but can also be applied to single returns. This is illustrated by application to the full EVI point cloud as well as a reduced point cloud of just the last returns from each laser shot. These are shown as the blue and green lines respectively in Figure 3b.

The ILRIS point cloud data are last return data and although there is an intensity value recorded, it is not calibrated. We have therefore treated each point with equal weighting and used an approximated value for the apparent reflectance to generate a foliage profile of close to the same magnitude as the EVI data. This is shown in Figure 3c (blue curve) compared with the EVI last hit profile (black) as well as a profile calculated from the EVI last hit data with apparent reflectance approximated to a constant in the same way as the ILRIS data.

The foliage profiles in Figure 3 all show the 2-layer characteristic of the canopy. There is significantly more variation in the foliage amount seen in the understorey layer than in the upper canopy. The voxel-based methods produce larger foliage volumes than the waveform and polar point cloud method. This is probably due to the fact that non-illuminated volumes have been disregarded in this study.

5. Discussion and Conclusion

Simulation of discrete point lidar data from the EVI waveform data has provided a means to test the theory presented here. The polar point cloud method was shown to give very similar results to the waveform method. Thus we are confident that given the right geometric and radiometric information, terrestrial lidar point cloud data can be successfully used to map foliage area volume density within a canopy.

Comparing the results of the multiple scan ILRIS dataset with the equivalent calculations from the EVI data shows some differences that may be associated with the more complete illumination achieved with multiple scans. The EVI profiles from last hits (black and red lines in Figure 3c) show a much larger volume of foliage in the lower canopy than is seen in any of the other profiles. This is not evident in the ILRIS scan and this is probably due to the extra stability provided by the multiple scan locations. The EVI instrument was positioned close to some shrubs and trunks which dominate the lower parts of the foliage profile. The trunks in particular produce high intensity reflections. It is clear that this is a dominant effect since the uncalibrated EVI voxel method, where each point is assigned the same reflectance, produces a profile with less material in the understorey. Deploying the EVI instrument at more than one location and combining the information from multiple scans would help to overcome such bias.

The voxel-based method shows promise and has produced results that are reasonably consistent with the waveform data. However there are aspects of the method that require further investigation. In situations where the scan pattern does not provide complete illumination, the compensation for non-illuminated volume needs to be investigated. If the scan pattern of the lidar is known, then it is possible to map which voxels are illuminated. However, a simple adjustment according to the volume illuminated may not provide the solution because the patterns of illumination are related to the distribution of foliage elements and thus there is inherent bias.

Acknowledgements

We are grateful to Darius Culvenor and Glenn Newnham for assistance with EVI operation and Vanessa Haverd for assistance with fieldwork and reviewing the manuscript.

References

- Hosoi, F. and Omasa, K., 2006. Voxel-based 3-D modelling of individual trees for estimating leaf area density using high-resolution portable scanning lidar *IEEE Transactions on Geoscience and Remote Sensing*, 44 (12), 3610-3618.
- Imhoff, M.L., Johnson, P., Holford, W., Hyer, J., May, L., Lawrence, W. and Harcombe, P., 2000. BioSAR™: an inexpensive airborne VHF multiband SAR system for vegetation biomass measurement. *IEEE Transactions on Geoscience and Remote Sensing*, 38, 1458-1462.
- Jupp, D.L.B., Culvenor, D.S., Lovell, J.L., Newnham, G.J., 2005. Evaluation and validation of canopy laser radar (LIDAR) systems for native and plantation forest inventory. Final Report prepared for the Forest and Wood Products Research and Development Corporation (FWPRDC: PN02.2902) by CSIRO. Available as CSIRO Marine and Atmospheric Research Paper 020 at http://www.cmar.csiro.au/e-print/open/cmar_rp020.pdf.
- Jupp, D.L.B., Culvenor, D.S., Lovell, J.L., Newnham, G.J., Strahler, A.H., Woodcock, C.E., 2009. Estimating forest LAI profiles and structural parameters using a ground based laser called “Echidna®”. *Tree Physiology*, 29, 171-181.
- Leblanc, S.G., Chen, R., Fernandes, R., Deering, D.W. and Conley, A., 2005. Methodology comparison for canopy structure parameters extraction from digital hemispherical photography in boreal forests. *Agricultural and Forest Meteorology*, 129 (3–4), 187-207.
- Lefsky, M.A., Cohen, W.B., Acker, A., Parker, G.G., Spies, T.A. and Harding, D., 1999. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-Fir Western Hemlock forests. *Remote Sensing of Environment*, 70, 339-361.
- Leuning, R., Cleugh, H.A., Zeglin, S.J., Hughes, D., 2005. Carbon and water fluxes over a temperate Eucalyptus forest and a tropical wet/dry savanna in Australia: measurements and comparison with MODIS remote sensing estimates. *Agricultural and Forest Meteorology*, 129 (3–4), 151–173.
- Parker, G.G., Harding, J.H. and Berger, M.L., 2004. A portable LIDAR system for rapid determination of forest canopy structure. *Journal of Applied Ecology*, 41, 755-767.
- Ross, J.K., 1981. *The radiation regime and architecture of plant stands*. Junk Publishers, The Hague, Netherlands.
- Takeda, T., Oguma, H., Yone, T., Yamagata, Y. and Fujinuma, Y., 2005. Comparison of leaf area density measured by laser rangefinder and stratified clipping method. *Phyton – Annales Rel Botanicae*, 45 (4), 505-510.

Takeda, T., Oguma, H., Sano, T., Yone, T. and Fujinuma, Y., 2008. Estimating the plant area density of a Japanese larch plantation using a ground-based laser scanner. *Agricultural and Forest Meteorology*, 148 (3), 428-438.

Van der Zande, D., Jonckhere, I., Stuckens, J., Verstraeten, W.W. and Coppin, P., 2008. Sampling design of ground-based lidar measurements of forest canopy structure and its effect on shadowing. *Canadian Journal of Remote Sensing*, 34 (6), 526-538.

Optimal LiDAR gridding parameterization for effective leaf area estimation in the boreal forest Yukon Territory, Canada

Heather Morrison^{1,2}, Chris Hopkinson^{2,1} & Michael A. Wulder³

¹Acadia University, Wolfville, Nova Scotia, Canada morrison.h@gmail.com

²Applied Geomatics Research Group, NSCC Annapolis Campus, Nova Scotia, Canada

³Pacific Forest Centre, Canadian Forest Service, Victoria, British Columbia, Canada

Abstract

The increased availability of LiDAR-based forestry models raises questions about fundamental procedural steps undertaken before published models begin. The processing stage being investigated in this study is the parameterization of routines used for gridding inputs for forestry models. Grids are a valuable format for modelling as they organize scattered point clouds into manageable pixels for sophisticated processing. Our objective was to examine the effect of grid cell resolution and circular search radii for gridding on resulting leaf area index (LAI) data layers. LAI is employed in models at various scales and was therefore of interest for a range of study objectives. We generated 16 gridded estimates of LAI using unique combinations of cell resolution and search radius for comparison with values measured in the field. Our results determined that cell resolution was not important, allowing for more application flexibility without introducing bias, while search radius was critical for obtaining the most accurate estimates. This type of scale sensitivity analysis is important for any modelled variable that will be applied in a variety of spatial contexts.

Keywords: DHP, LiDAR, leaf area index, Yukon Territory

1. Introduction

Leaf area index (LAI), which is defined as half of the total leaf surface area per unit ground area (Chen *et al.*, 2006) is an important input for biogeochemical and ecosystem-atmosphere models. For most forestry applications estimates of this type are required for large study areas, making direct measurement methods practically impossible. Instead, indirect remote sensing solutions are becoming more available and affordable for modelling metrics such as effective leaf area index (LAI_e) which includes both leafy and non-leafy components of the canopy and can be used to determine true LAI (Chen *et al.*, 1997). Light Detection and Ranging (LiDAR) data provides a three-dimensional representation of the forest canopy that is ideal for modeling LAI_e over large areas. Generating grids is an important time saving step for summarizing irregularly spaced points of LiDAR data and decisions made at this stage will permeate throughout the modelling process. It is important to balance maintaining sufficient variability to represent the natural landscape without compromising computational efficiency or model robustness and reusability. The goal of this paper is to investigate the effect of gridding parameters on estimates of LiDAR-intensity based LAI_e and to highlight optimal settings for a boreal forest landscape.

The LAI_e model being applied in this study is based on a model for estimating gap fraction published by Hopkinson and Chasmer (2007; 2009) that utilizes pulse intensity, range and echo

code. This model is based on a Beer's Law assumption that the canopy can be represented as a turbid medium with randomly distributed foliage, which allows for the estimation of LAIe:

$$LAIe = -\ln(P) / k \quad (1)$$

where P is canopy gap fraction and k is extinction coefficient. A model optimization process was also developed using field measurements of LAIe in Morrison *et al.* (2011).

The few forest-related studies that have used intensity values for forest attribute modeling (e.g. Donoghue *et al.*, 2007) generally do not describe under what conditions the data were gridded before analysis. For this analysis, grids were prepared by assigning to each cell the sum of the intensity values within a specified search radius. Four grid cell resolutions and four search radii were used to generate sixteen uniquely parameterized intensity grids to determine optimal gridding parameters for calculating intensity-based LAIe estimates.

2. Study Area

The study area is located in northern Canada near Watson Lake, Yukon Territory (Fig. 1). The Boreal forest in this region is dominated by a mixture of mature spruce and pine with some early successional birch and aspen in areas that have been recently cut. This study is part of a larger project capturing a broad sample of boreal forest conditions across Canada through the collaboration of the Canadian Forest Service, researchers at the University of British Columbia, and the Applied Geomatics Research Group (Hopkinson *et al.* 2011; Bater *et al.* 2011).

3. Methods

3.1 Ground estimates of LAIe using hemispherical photography

Field data were collected July 30 to August 3, 2010. Thirty plots were established along two East-West transects approximately 850 meters apart. Plots along each transect were at least 100 meters apart and included both treed and clearcut areas (Fig. 1). Plot centers were geolocated using differentially corrected GPS receivers (Leica SR530, Leica Geosystems Inc. Switzerland) with horizontal position quality ranging from 1 cm to 1 m depending on the quality control technique used and the canopy density.

Five digital hemispherical photographs (DHPs) were captured per plot: one at the plot center and four located 20 m from the center in cardinal directions (N, E, S, W) using a compass and measuring tape (Figure 1). Photos were captured using a Nikon CoolPix 8800VR camera fitted with a 180° fisheye lens with the exposure set one 'f stop' lower than normal exposure to improve contrast between foliage and sky. For all images the camera was levelled on a tripod 1.3 m from the ground with the top of each photo oriented towards the north.

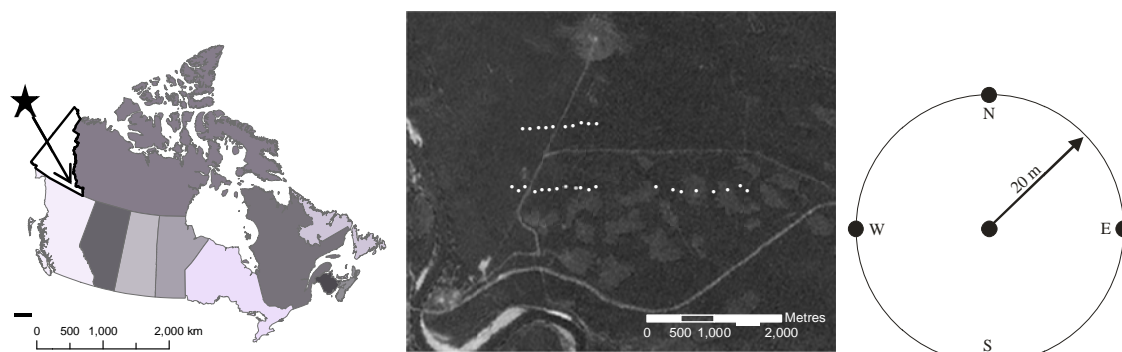


Figure 1: Map showing location of study area in Yukon Territory, Canada (left), relative position of the plot transects (center), and photo capturing layout (right).

Photographs were processed with Gap Light Analyzer (GLA) Version 2.0 developed by Frazer *et al.* (1999). Image analysis in GLA requires a number of steps. After opening the desired photo within the software the image was registered by identifying geographic orientation and circular extent. Next the configuration settings were edited to suit the type of calculations being performed, including site topography information and projection distortion parameters. Given that all the photos for this study were captured using the same camera, settings, and were within the same general area the parameters for these steps were saved to a configuration file and applied to all images being processed. The next step was to classify each image into sky and non-sky pixels. Using the blue channel, a threshold was set for each image to capture the distinction between vegetation and sky. Once an appropriate classification was achieved the final step was to run the calculations and save the output. LAIe values obtained from GLA for this study were estimated by integrating across zenith angles 0° to 75° . A single LAIe estimate for each 20 metre plot was generated for statistical comparisons by averaging the values for the 5 nested photos captured per plot.

3.2 LiDAR data collection and preparation

An airborne LiDAR survey was completed by the Applied Geomatics Research Group on August 3, 2010 using an Optech Inc. ALTM 3100 at a flying height of 1500 m a.g.l., 50 kHz pulse repetition frequency, and 15° scan angle resulting in approximately 1.1 m point spacing. The point cloud was differentially corrected to the same GPS base station that was used for the field plot set up during DHP acquisition. The software package *TerraScan* (Terrasolid, Finland) was then used to classify points into ground, canopy and echo code classes in preparation for intensity-based modelling (Hopkinson and Chasmer, 2009).

3.3 Gridding parameters analysis

Four grid cell resolutions and four search radii were selected for this analysis, generating sixteen gridding parameter variations (Table 1). Cell sizes of 1, 5, 10, and 25 m were used to include the finest resolution given the density of LiDAR returns, up to a size comparable to satellite imaging products. Search radii range was selected to overlap common field mensuration plot areas (100, 200, 400 and 800 m²). Grids were prepared using *Surfer 8* software (Golden Software, USA) by assigning each cell the summed intensity value for all returns captured within the given circular search radius based on the point classification required for model input.

Table 1. Sixteen unique combinations of cell size and search radius for gridding treatment layers

		Search radius (search area)			
		5.6 m (100 m ²)	8.0 m (200 m ²)	11.3 m (400 m ²)	16.0 m (800 m ²)
Cell Size	1 m	1m5.6	1m8.0	1m11.3	1m16.0
	5 m	5m5.6	5m8.0	5m11.3	5m16.0
	10 m	10m5.6	10m8.0	10m11.3	10m16.0
	25 m	25m5.6	25m8.0	25m11.3	25m16.0

3.4 Statistical analyses

LAIe statistics were extracted for thirty, 20 m radius circular plots coinciding with geo-registered plots from field data collection including the minimum, maximum, mean and standard deviation for LAIe in each plot. Twenty meter plots were chosen to minimize the spatial uncertainty of the plot positions by combining five photos nested together. Further research needs to be done to determine the optimal plot size for comparison with DHPs in this context as plot size has been linked to the quality of model estimates by Lovell *et al.* (2003) and Morsdorf *et al.* (2006) when determining forest cover. Statistical analyses were performed using R version 2.11.0 (www.r-project.org). Simple linear regression was performed using the 16 gridding parameter LAIe estimates and DHP LAIe. All best fit lines were forced through the origin as there was no logical reason for DHP LAIe at zero to be different from LiDAR modelled LAIe; i.e. if there is no foliage within a plot, then there should also be no laser points above the ground surface. Two statistical tests were used to evaluate which gridding parameters produced optimal LAIe estimates. Root mean squared error (RMSE) was calculated according to the methodology in Kobayashi and Salam (2000) and prediction sum of squares (PRESS) was calculated according to Myers (1986, p106-111). PRESS is a statistical test where data are fit n times, and in each iteration one data point is removed, estimated based on the linear model, and then the difference between the actual and predicted are summed and squared. This provides a type of model validation without incurring extra cost in collecting large amounts of field data.

4. Results & discussion

The RMSE and PRESS results are presented in Figure 2. Both statistical tests demonstrate that intensity-based LAIe estimates show no significant difference and are stable across all grid cell resolutions for any given search radius. These observations demonstrate that when modeling LiDAR-based LAIe, the choice of grid cell size can be made according to the intended use of the data, without fear of introducing any systematic bias. Moreover, this insensitivity to grid resolution means that LAIe maps generated at multiple spatial resolutions can be directly compared provided appropriate aggregation routines are used.

In both the PRESS and RMSE test, however, predicted LAIe demonstrates systematic variation with the size of search radius used during grid creation, regardless of the final grid cell resolution. This highlights that the choice of search radius is critical and must be underpinned by some logical rationale. Both tests demonstrate that 11.3m produces the lowest overall error and therefore has the most accurate LAIe predictive capability when compared to DHP validation plots.

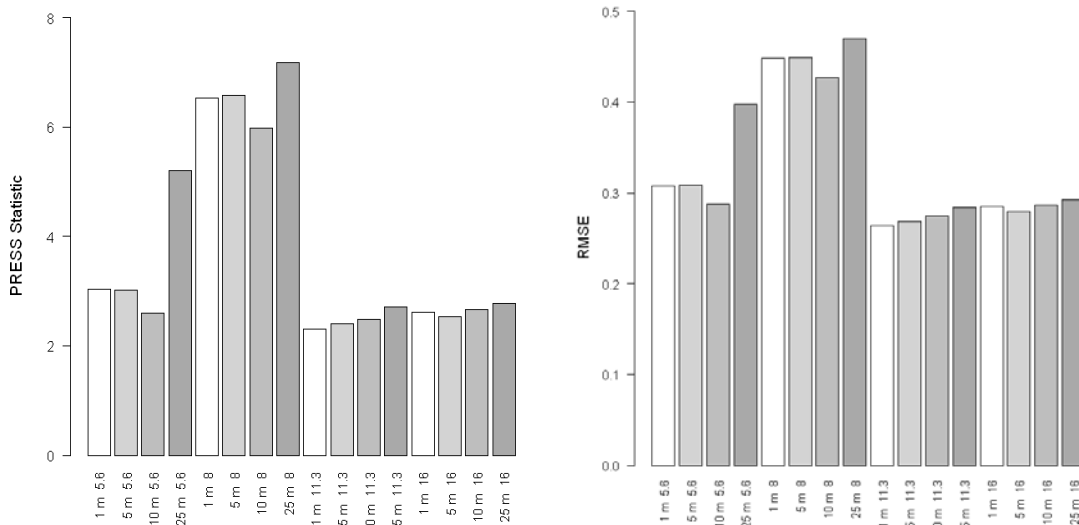


Figure 2. Prediction sum of squares (left) and root mean squared error (right) per treatment (unique cell size and search radius combinations), n=30 for each bar.

Given that the optimal point cloud search radius for recreating plot-level DHP-based LAI is approximately 11.3 m, the DHP angular field of view between the top of the canopy and the photo capture location can be estimated (Figure 3). While mean plot-level canopy varied between ~ 3m and 20m, the average height was approximately 14.3m. With a DHP capture location of 1.3m above the ground, the approximate optimal full field of view directly above the photo point is ~80° (±40°). While the spatial coincidence of DHP LAIe data and LiDAR point cloud attributes are not directly compared here, this does suggest that the domain of DHP LAIe results extends approximately 11.3 m from the photo location. More search radii need to be tested to see if the correlation between point cloud LAIe model results and DHP data continue to diverge as the search radius is enlarged beyond 16 m.

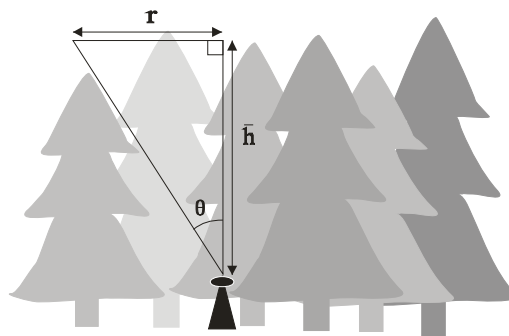


Figure 3. Triangle represents the comparable field of view angle for a hemispheric lens. Given a plot radius (r), and mean canopy height (h) less the height above ground where the camera is positioned.

5. Conclusion

In this study we have demonstrated that LiDAR-based raster LAIe map products are largely insensitive to the final grid cell resolution chosen. This is important because it means high resolution maps can be generated at close to the tree stem level are useful while larger grid cells comparable to wide area satellite coverage data (e.g. Landsat TM) are equally as valid. This is not to say that LAIe will be accurate at, for example, a 1m grid cell resolution, as this cannot be verified from data trained at the DHP plot-level. However, it does suggest that when aggregated to a more 'plot-level' scale, a 1m grid will not introduce any bias. Conversely, the choice of search radius to be applied to the point cloud during grid creation is important, and is likely influenced by the amount of coincident upper canopy surface that can be clearly observed by both the almost planar sampling of LiDAR data and the hemispherically projected DHP image. In this particular boreal forest context, an 11.3 m radius or 80° vertical field of view appeared to provide overall optimal LAIe prediction results and suggests a possible optimal plot size for comparison with DHP imagery. Of some interest is that this radius also corresponds to a typical forest mensuration plot of 400 m².

Acknowledgements

We acknowledge financial support for this work provided by Pacific Forest Centre, Canadian Forest Service with special thanks to Margaret Andrew and Antoiune Lalumiere of the CFS, and Tristan Goulden and Neville Crasto of AGRG for help collecting and geolocating the forest plot data.

References

- Bater, C. Wulder, M.A., Coops, N.C., Hopkinson, C., Coggins, S.B., Arsenault, E., Beaudoin, A., Guindon, L., Hall, R.J., Villemaire, P., and Woods, M. 2011. Model development for the estimation of aboveground biomass using a lidar-based sample of Canada's boreal forest. *Proceedings of the SilviLaser 2011 Conference*, Oct. 16-20, Hobart, Tasmania
- Chen, J.M., Govind, A., Sonnetag, O., Zhang, Y., Barr, A., and Amiro, B., 2006. Leaf area index measurements at Fluxnet-Canada forest sites. *Agricultural and Forest Meteorology*, 140, 257-268.
- Chen, J.M., Rich, P.M., Gower, S.T., Norman, J.M., and Plummer, S., 1997. Leaf area index of boreal forests: Theory, techniques, and measurements. *Journal of Geophysical Research*, 102(D24), 29,429-29,443.
- Donoghue, D.N.M., Watt, P.J., Cox, N.J., and Wilson, J., 2007. Remote sensing of species in conifer plantations using LiDAR height and intensity data. *Remote Sensing of Environment*, 110, 509-522.
- Frazer, G.W., Canham, C.D., Lertzman, K.P., 1999. Gap Light Analyzer (GLA), Version 2.0: Imaging Software to Extract Canopy Structure and Gap Light Transmission Indices from True-color Fisheye Photographs, Users Manual and Program Documentation. Simon Fraser University/Institute of Ecosystem Studies, Burnaby, BC/ Millbrook/NY
- Hopkinson, C., and Chasmer, L., 2007. Modelling canopy gap fraction from LiDAR intensity. *Proceedings of the Silvilaser 2007 Conference*, Sept. 12-14, Espoo, Finland.
- Hopkinson, C., and Chasmer, L., 2009. Testing LiDAR models of fractional cover across multiple forest ecozones. *Remote Sensing of Environment*, 113, 275-288.
- Hopkinson, C., Wulder, M.A., Coops, N.C., Milne, T., Fox, A., and Bater, C.B., Airborne lidar sampling of the Canadian boreal forest: Planning, execution & initial processing. *Proceedings of the SilviLaser 2011 Conference*, Oct. 16-20, Hobart, Tasmania

- Kobayashi, K. and Salam, M.U., 2000. Comparing simulated and measured values using mean squared deviation and its components. *Agronomy Journal*, 92, 345-352.
- Lovell, J.L., Jupp, D.L.B., Culvenor, D.S., and Coops, N.C., 2003. Using airborne and ground-based ranging LiDAR to measure canopy structure in Australian forests. *Canadian Journal of Remote Sensing*, 29(5), 607-622.
- Morrison, H., Hopkinson, C., Chasmer, L., and Kljun, N., 2011. Using a GIS approach to optimize effective leaf area index by Canadian boreal forest species using airborne LiDAR. *Proceedings of the SilviLaser 2011 Conference*, Oct. 16-20, Hobart, Tasmania.
- Morsdorf, F., Kötz, B., Meier, E., Itten, K.I., and Allgöwer, B., 2006. Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment*, 104(1), 50-61.
- Myers, R.H., 1986. *Classical and modern regression with applications*, Duxbury Press, Boston, Massachusetts, p359.

Change detection of mountain vegetation using multi-temporal ALS point clouds

Mattias Nyström¹, Johan Holmgren¹ & Håkan Olsson¹

¹ Section of Forest Remote Sensing, Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Sweden – (mattias.nystrom, johan.holmgren, hakan.olsson)@slu.se

Abstract

Multi-temporal laser scanner data to be used in change detection studies will most likely be acquired with different sensors, flying altitudes, and system parameters. Therefore, calibration is probably needed in order to make laser returns from vegetation comparable between two laser data acquisitions. In this study, two ALS point clouds were acquired with different sensors and flying altitudes. The first data set had 11.5 points m⁻² and was obtained in 2008 with a TopEye MKII scanner and the second with a density of 1.1 points m⁻² was obtained in 2010 with an Optech ALTM Gemini scanner. The test site was located in Abisko in northern Sweden with forest dominated by mountain birch. Six meter radius sample plots were placed in the forest-tundra ecotone and assigned one of the following treatments: (1) reference with no removal of trees, (2) removal of 50% of the total number of stems above 1.5 m, and (3) removal of 100% of the total number of stems above 1.5 m. Histogram matching was used to calibrate the two data sets and sample plots were then classified into the three treatments. The overall classification accuracy was 82% using only the proportion of vegetation returns from the canopy as explanatory variable. Features created from gridded laser data had overall higher classification accuracy than laser features created directly from the point cloud. Histogram matching made the two data sets comparable by reducing the difference between them. These early results show how changes can be detected even with different sensors, flying altitudes, and system parameters.

Keywords: Airborne laser scanning, histogram matching, multi-temporal, LiDAR, change detection

1. Introduction

The effect of ongoing climate change on the sub-arctic and alpine forests has led to interest in monitoring potential changes in the forest-tundra ecotone. In addition to climate change, insect damages, browsing pressure by herbivores as well as anthropogenic impacts will contribute to changes in the sub-arctic forest-tundra ecotone. These changes are difficult to monitor with manual methods because of the complex mosaic pattern of the ecotone.

Airborne laser scanning (ALS) could efficiently be used to map the vegetation in the forest-tundra ecotone (Nyström *et al.* 2010). However, so far there is limited research on using multi-temporal laser data to detect areas with removed trees or estimation of forest growth. One of the few studies is Yu *et al.* (2004) who used ALS data with a resolution of 10 measurements per square meter from two acquisitions with two years in between covering a boreal forest. Individual trees detected from the two datasets were linked and harvested trees could be identified and tree height growth was estimated.

ALS data sets will in the future become available for the same area for a series of acquisitions collected with various resolutions, scanning systems, system parameters, etc. Therefore, research is needed to find methods for efficient calibration of multi-temporal data sets. Calibration methods have earlier been developed for analysis of optical satellite image data and radar data (Coppin *et al.* 2004; Moser *et al.* 2004; Olsson 1993). It is of interest to find out if these calibration methods are suitable also for analysis of ALS data.

One important task is to find features extracted from ALS data that are useful for detection of changes. The features should be efficient for estimation of for example biomass but it should also be possible to calibrate the features between the data sets in areas without changes to reduce differences.

In this study, we used two laser acquisitions from two time points with different scanning systems, system parameters, point densities, and flying altitudes. Between the two laser acquisitions, sample plots were placed out and trees cut down to simulate afforestation. The objectives were to (1) validate the effect of a simple histogram matching algorithm when comparing the two data sets, and to (2) find features from ALS data that are efficient for detection of changes in amount of vegetation in the sub-alpine tree line ecotone using supervised classification.

2. Method

2.1 Study area

The study area is located six km southeast of Abisko in northern Sweden, centered around Lat. N 68°20', Long. E 19°01' (Figure 1) and dominated by mountain birch (*Betula pubescens* ssp. *czerepanovii*), but having also some Junipers (*Juniperus communis*), Rowans (*Sorbus aucuparia*), and Willows (*Salix* spp.) taller than 1.5 m. The birches in the area were of the multi-stem type often with several stems sharing the same root system. Sample plots were placed in the ecotone between birch forest and tundra. This area was characterized by a pattern of forest and mountain heath vegetation and located at an altitude of 480 - 670 m a.s.l.

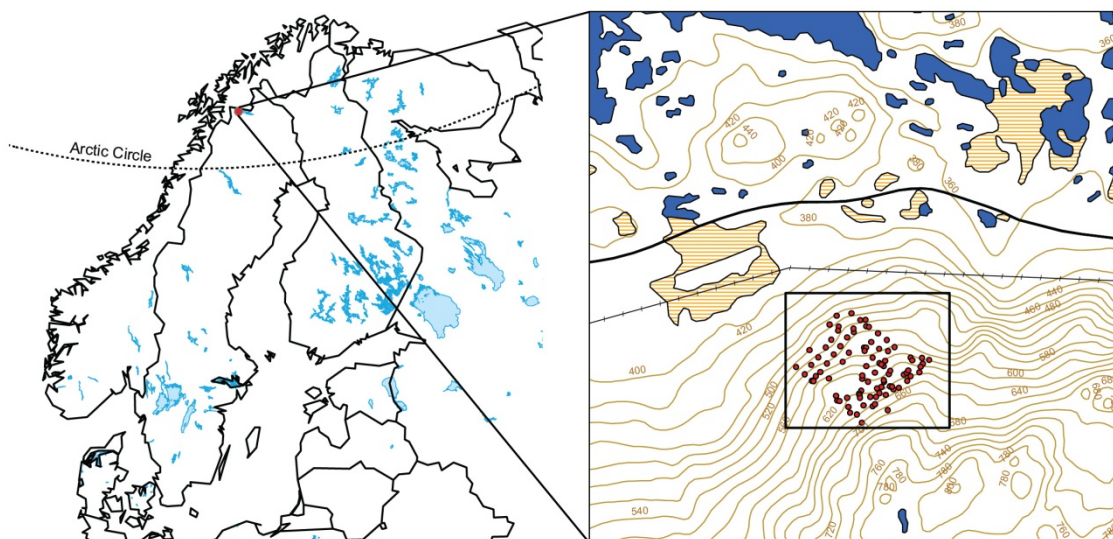


Figure 1: Map of the Nordic countries, showing the study area in northern Sweden. The red dots in the enlarged map are the sample plots and the black polygon is the area used to create cumulative histograms to use in the histogram matching. © Lantmäteriet, I 2010/0345.

2.2 Laser data acquisition

Laser data were acquired under leaf-on conditions with two different laser scanners and at two occasions (Table 1); the first with a TopEye MkII (denoted TopEye) mounted on a helicopter, and the second with an Optech ALTM Gemini (denoted Optech) mounted on a fixed wing aircraft. In both cases, scanning was performed in west-east direction.

Table 1: Summary of the laser scanner properties and flight parameters of the two laser data acquisitions.

Parameter	TopEye MkII	Optech ALTM Gemini
Scanning date	Aug. 1, 2008	Aug. 20, 2010
Flight altitude (above ground)	500 m	1740 m
Footprint	0.5 m	0.5 m
Pulse repetition frequency	50 kHz	70 kHz
Scan frequency	35 Hz	37 Hz
Wave length	1064 nm	1064 nm
Pulse length	4 ns (1.2 m)	6.8 ns (2.0 m)
Scan type	Palmer	Oscillating mirror
Scan width (across flight dir.)	±20°	±20°
Scan width (along flight dir.)	±14°	0°
Point extraction	Up to 2 per pulse	Up to 4 per pulse
Minimum point density ^a	3.0 m ⁻²	0.7 m ⁻²
Average point density ^a	11.5 m ⁻²	1.1 m ⁻²
Maximum point density ^a	34.8 m ⁻²	2.4 m ⁻²

^a based on the 100 sample plots used in this article.

2.3 Field data

The field inventory was performed between the two laser data acquisitions, during four weeks in August 2009 (47 sample plots) and two weeks in June 2010 (43 sample plots), see Table 2 for

details. Sample plots used from 2009 existed from a previous study in the area (Nyström *et al.* 2010). These 47 sample plots had ten meters radius, taller trees and located in areas with older forest than the 43 sample plots from 2010. Only six meters radius of 2009 years sample plots were used in the analysis. All sample plots were revisited in 2010 to assure they were unchanged. In 2010, 43 new sample plots with six meters radius were subjectively positioned to match the following criteria: (1) trees in the range 1.5-2.5 m tall, (2) 10% of the stems can be taller than 2.5 m, but not taller than 3.5 m, and (3) no dominating field (0-0.5 m) and bush layer (0.5-1.5 m) on the sample plot. Figure 2 show one of the sample plots from the field data collection in 2010.



Figure 2: One of the sample plots from the 2010 years field data collection. This sample plot was classified as medium stem density. The yellow/red measuring pole is 3 m tall.

One out of the following treatments was assigned each sample plot:

1. Reference, no treatment.
2. Remove 50% of the total number of stems taller than 1.5 m.
3. Remove 100% of the total number of stems taller than 1.5 m.

Only trees taller than 1.5 m were included in the treatments. Spatial location and height distribution of removed stems on the sample plot were as evenly distributed as possible. The treatment was extended to seven meters radius to minimize problems with horizontal dislocation between field and laser data, but still only six meters was used for the laser data extraction.

Sample plots were divided into three categories depending on total number of stems taller than 1.5 m inside six meters radius:

1. Low stem density: 5-10 stems
2. Medium stem density: 11-50 stems
3. High stem density: 51-100 stems

In this initial analysis, the results are not separated for the three density categories. Table 2 gives a summary of the treatments and densities of the sample plots.

The center position of the sample plot was measured with sub dm accuracy using a Real Time

Kinematic GPS (Trimble R7) with the base station placed six km away at the Abisko research station.

Table 2: Summary of the field data.

Year collected	Treatment (Stem density class: low/medium/high)		
	Reference	50%	100%
2009	47 (16/14/13)	0	0
2010	15 (5/5/5)	13 (5/4/4)	15 (6/5/4)

2.4 Laser data processing

TerraScan (Soinen, 2010) from Terrasolid was used to classify the point cloud into ground and non-ground points (Axelsson, 1999; Axelsson, 2000) with a setting of maximum terrain angle of 88 degrees, iteration angle of 6 degrees, and iteration distance 1.40 m. The statistical software R (R Development Core Team, 2010) and in house developed programs were used to further process the point cloud. No removal of overlapping data was done in either data set.

A DEM with 0.5 m grid size, representing the ground level, was created. Canopy heights (CH) were calculated by subtracting the DEM from the z-value of each laser return. A digital surface model (DSM) was created by assigning each raster cell the maximum z-value for laser returns classified as non-ground. A normalized digital surface model (nDSM) was calculated by subtracting the DEM from the DSM. All data points above 15 m were neglected in the nDSM and the CH to avoid false reflections (the 15 m limit was chosen while no trees taller than 10.5 m were found in the area).

A height threshold of 1.0 m was used on nDSM and CH for measures from the ALS point cloud above ground (denoted laser features). One reason the commonly used 2.0 m height threshold (Nilsson, 1996) was not used here was to obtain reliable laser features for sample plots with trees around 1.5 m tall.

All the following laser features were created from both nDSM and CH. Height percentiles (H_{xx}) in steps of 20 were calculated using canopy heights above the height threshold. In addition, height percentiles were also calculated for H_{95} and H_{99} since these are strong indicators of the height of vegetation. Ten vertical canopy layers (D_x) were estimated in accordance to Næsset and Gobakken (2008) using the height threshold as the lower limit and H_{95} as the upper limit. Standard deviation (H_{sd}) and mean (H_{mean}) of the canopy heights were calculated on values above the height threshold. A sum of squared canopy heights (H_{sum}) was calculated by taking the sum of the squared laser point's height divided by the number of laser points above the height threshold. Vegetation ratio (VR) was calculated by dividing the number of hits above the height threshold with the total number of returns inside the sample plot. Two ways of calculating the vegetation ratio were used: (1) all returns (denoted VR_{all}) and (2) only first returns (denoted VR_{1st}).

2.5 Histogram matching

An area of 1.8 km² (Figure 1) was used to create laser features for 10x10 m pixels. Histogram matching was used to calibrate the two laser data acquisitions to a common distribution. Cumulative histograms were created for each laser feature in the two data sets using 100 bins with minimum and maximum values from each laser feature as bounds. Linear interpolation was used when assigning values in the histogram matching. The histogram created from the TopEye laser features was used as target histogram and the histogram from the Optech laser

features as reference histogram. In total 18450 pixels were inside the bounding box (Figure 1), but several pixels had no laser features because no vegetation was present. Therefore only 10142 pixels were used to create the histograms.

2.6 Multi-temporal comparison

To simulate afforestation, the first laser data collection (2007) was used as “after” and the second laser data collection (2010) as “before”. The similarity of the laser features from two different time points after histogram matching was evaluated by aid of field reference plots that according to a field survey had only normal growth and no unusual changes. The measures used were relative RMSE ($RMSE_r$), and relative bias ($Bias_r$) calculated from the unchanged reference sample plots:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Top_i - Opt_i)^2}{n}} \quad (1)$$

$$RMSE_r = \frac{RMSE}{\bar{y}} \cdot 100 \quad (2)$$

$$Bias = \frac{\sum_{i=1}^n (Top_i - Opt_i)}{n} \quad (3)$$

$$Bias_r = \frac{Bias}{\bar{y}} \cdot 100 \quad (4)$$

where Top_i is the i :th observation of a laser feature from the TopEye data and Opt_i is the i :th observation of the same laser feature from the Optech data, n is the number of observations, and \bar{y} is the mean value of the laser feature, i.e. mean value of Top and Opt . This provided an opportunity to check both the success of the histogram matching, and to obtain an indication of which laser features that provided most similar data for the normal developed vegetation.

2.7 Classification

In order to also check which laser features that best discriminated changed vegetation from unchanged vegetation, sample plots were classified into the three treatments using Linear Discriminant Analysis (LDA) in the statistical software R (R Development Core Team, 2010). The MASS-package (Venables, 2002) was used for the classification. Prior probabilities were set to 1/3 for each of the three classes. Cross-validation was used when calculating the classification accuracy. As explanatory variable a relative difference between the laser features from the two data sets was used and calculated for sample plot i as:

$$\Delta_i = \frac{Top_i - Opt_i}{Top_i + Opt_i} \quad (5)$$

where Top_i is a laser feature from the i :th sample plot in the TopEye data and Opt_i is the same laser feature and the i :th sample plot in the Optech data.

3. Result

Each laser features from the second scanning (Optech) were histogram matched using the first scanning (TopEye) as target. The cumulative histograms created for the vegetation ratio from nDSM (VR_{all}^{nDSM}) is shown in Figure 3. A one-to-one plot of the vegetation ratio before and after histogram matching can be seen in Figure 4.

The similarity of laser features between the two scanings was evaluated by use of RMSE and Bias for each laser feature. Table 3 shows relative RMSE and relative Bias for all features from the 62 reference sample plots before and after histogram matching.

Sample plots were classified into the three classes (reference, 50%, and 100%) using linear discriminant analysis (LDA). Table 3 shows the overall cross-validated classification accuracy for the each laser features as explanatory variable. Table 4 shows the error matrix when using height percentile 60 (H_{60}^{nDSM}) as explanatory variable in the classification and Table 5 when using the vegetation ration (VR_{all}^{nDSM}) as explanatory variable.

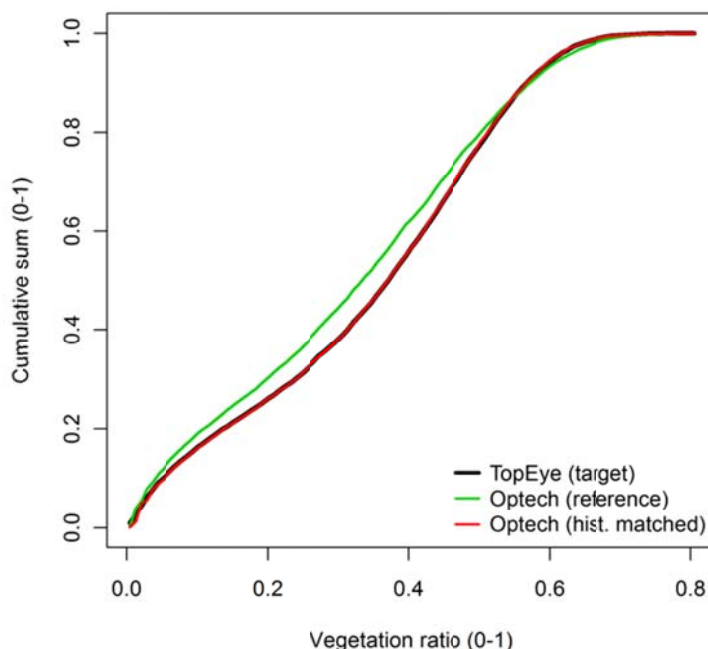


Figure 3: Cumulative histogram of vegetation ratio created using nDSM (VR_{all}^{nDSM}). The red line is the cumulative histogram of Optech after histogram matching using TopEye as target histogram.

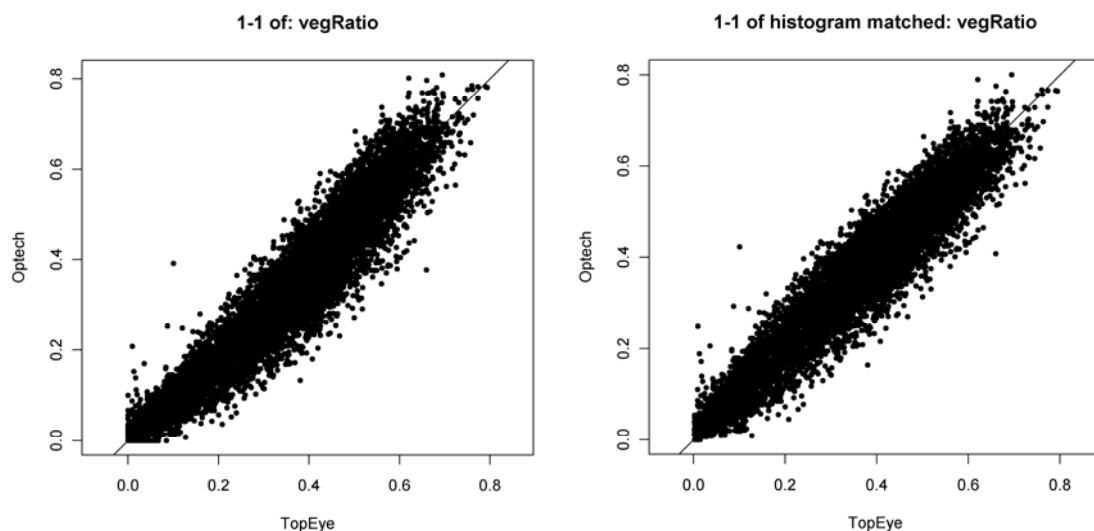


Figure 4: One-to-one plot of vegetation ratio created from nDSM (VR_{all}^{nDSM}). In the left plot is Optech data not calibrated and the right plot is Optech calibrated using histogram matching. The black line is the one-to-one line. When histogram matching is used, the data bend is removed.

Table 3: Relative RMSE and relative bias calculated from the laser features using Equation (2) and (4). In each of the following four categories D_x^{CH} , D_x^{nDSM} , H_{xx}^{CH} , and H_{xx}^{nDSM} , only the feature with lowest relative RMSE after histogram matching is presented. $RMSE_r$ and $Bias_r$ were calculated only using the 62 reference sample plots. The classification accuracy is the overall classification accuracy in the three treatment classes. The table is ordered after $RMSE_r$ -values calculated after histogram matching was performed. The abbreviations of the laser features are explained in section 2.4.

Laser feature	$RMSE_r$ (%)		$Bias_r$ (%)		Classification acc. (%)	
	Hist. match	Not cal.	Hist. match	Not cal.	Hist. match	Not cal.
D_4^{nDSM}	15.6	15.5	-2.5	0.9	80	81
H_{95}^{nDSM}	17.8	19.9	2.7	10.4	79	78
H_{sd}^{nDSM}	18.2	17.9	-1.6	2.7	72	70
H_{100}^{CH}	18.8	21.4	2.1	12.4	78	76
VR_{all}^{nDSM}	19.5	22.9	-3.8	7.6	82	82
H_{sd}^{CH}	20.9	19.3	-3.3	-2.3	68	63
H_{mean}^{nDSM}	24.9	27.0	-5.2	-12.9	58	28
D_2^{CH}	25.2	30.9	-2.9	7.2	79	79
VR_{all}^{CH}	26.8	32.1	-3.7	9.0	79	78
H_{mean}^{CH}	28.0	49.2	-4.7	-31.9	59	30
VR_{1st}^{CH}	29.8	43.1	-5.8	19.7	79	79
H_{sum}^{nDSM}	30.4	30.5	-7.9	-7.6	21	56
H_{sum}^{CH}	35.0	61.4	-8.4	-31.9	13	56
Mean:	23.9	30.1	4.2*	12.0*	-	-

* Mean value calculated using absolute values.

Table 4: Error matrix for supervised classification using height percentile 60 created from nDSM (H_{60}^{nDSM}) as explanatory variable.

		True			User's accuracy
		Ref.	50%	100%	
Predicted	Ref.	59	9	2	84.3%
	50 %	0	0	0	0%
	100%	3	4	13	65.0%
Prod. accuracy		95.2%	0%	86.7%	80.0%

Table 5: Error matrix for supervised classification using vegetation ratio created from nDSM (VR_{all}^{nDSM}) as explanatory variable.

		True			User's accuracy
		Ref.	50%	100%	
Predicted	Ref.	56	4	1	91.8%
	50 %	2	4	0	66.7%
	100%	4	5	14	60.9%
Prod. accuracy		90.3%	30.8%	93.3%	82.2%

4. Discussion

The cumulative histograms (Figure 3) created for the vegetation ratio from nDSM (VR_{all}^{nDSM}) are almost identical when using histogram matching. The one-to-one plot before and after histogram matching (Figure 4) shows that after histogram matching the data is better aligned to the one-one line and the bend in the data is eliminated.

Relative RMSE and relative Bias (Table 3) were calculated to evaluate the similarity between the two data sets. It is clear that features created from nDSM are more similar when comparing the two data sets. Histogram matching reduces the difference for most features. Bias has always a low value when histogram matching is used, but not always lower than not calibrated.

High classification accuracy could be achieved using only one laser feature as explanatory variable. When using only one height percentile, none of the 50% treatment sample plots were correctly classified (Table 4). The reason is that trees taller than 1.5 m still remain on the sample plots and therefore the height percentiles remain almost the same as before removal of the trees. When using vegetation ratio, a difference in density of trees can be detected and therefore it is possible to detect some of the sample plots with 50% treatment, but still the class with largest classification error (Table 5). In future studies it should be considered to use more than one laser feature in the classification, e.g. a measure of heights (e.g. height percentile) in combination with a measure of density (e.g. vegetation ratio).

The laser features created using nDSM values has higher classification accuracy and lower RMSE than features created from point cloud data (CH). Some sample plots were not evenly covered with stems, which can cause large differences in the vegetation ratio feature if the laser data is unevenly distributed. Multi-temporal data therefore requires measures normalized at each time point to avoid problems with uneven distribution of laser points. An example of normalization method is features created from the nDSM. It is of highest importance to further research on normalization methods when analyzing multi-temporal data.

The footprint was about 0.5 m with both laser scanners regardless of the difference in flying altitude. This causes the Optech scanner with low point density to omit vegetation because the ground is not fully illuminated. How this affected the results in this study is hard to say, but when comparing the 3D point clouds from the two acquisitions it is clear that the denser TopEye data clearly show the spatial pattern of the vegetation while the same pattern not is visible in the more sparse Optech data.

The sample plot radius was chosen to be relatively small (6 m) to achieve sample plots with evenly distributed trees and tree heights. Only 43 sample plots were found in the area fulfilling the requirements described in the methods. Therefore 47 sample plots from a previous field inventory were also used as reference plots (not treated). All sample plots from the previous field inventory were revisited in 2010 to check if changes occurred.

If we would like to quantify the changes, histogram matching can be a good method to use and also if we want to interpret a difference image. Table 3 indicate that histogram matching reduce the difference between the two data sets in most of the cases, only a few laser features had higher difference when histogram matching was used. Thus, histogram matching might be a useful method for creating difference images from different laser data acquisitions. A relative calibration, for example histogram matching, will however not necessarily improve an automated classification of changes.

Acknowledgements

This study is part of the research programme Environmental Mapping and Monitoring with Airborne laser and digital images (EMMA), financed by the Swedish Environmental Protection Agency. The laser data sets were acquired through cooperation with the University of Lund and the new national elevation model project at the Swedish National Land Survey. The Abisko Scientific Research Station is acknowledged for supporting the work with field data acquisition.

References

- Axelsson, P.E., 1999. Processing of laser scanner data - algorithms and applications. *Isprs Journal of Photogrammetry and Remote Sensing*, 54(2-3): 138-147.
- Axelsson, P.E., 2000. DEM Generation from Laser Scanner Data Using Adaptive Models. *International Archives of Photogrammetry & Remote Sensing*, 33: 110-117.
- Coppin P., Jonckheere, I., Nackaerts, K., Muys, B. and Lambin, E., 2004. Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, 25(9): 1565-1596.
- Moser, G., Serpico, S.B., De Martino, M. and Coppolino, D., 2004. Automatic partially supervised classification of multitemporal remotely sensed images, *Proceedings of SPIE - The International Society for Optical Engineering 2004*, 126-137.
- Næsset, E. and Gobakken, T., 2008. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment*, 112(6): 3079-3090.
- Nilsson, M., 1996. Estimation of tree weights and stand volume using an airborne lidar system. *Remote Sensing of Environment*, 56(1): 1-7.
- Nyström, M., Holmgren, J. and Olsson, H., 2010. The potential of airborne laser scanning for monitoring the subalpine birch forest ecotone. *Proceeding of Silvilaser 2010*. Freiburg, Germany, 14-17 September 2010.
- Olsson, H., 1993. Regression functions for multitemporal relative calibration of Thematic Mapper data over boreal forests. *Remote Sensing of Environment*, 46, 89-102.

- R Development Core Team (2011). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Soininen, A., 2010. TerraScan User's Guide. TerraSolid Ltd.
- Venables, W. N. and Ripley, B. D., 2002. *Modern Applied Statistics with S. Fourth Edition*. Springer, New York. ISBN 0-387-95457-0.
- Yu, X.W., Hyyppä, J., Kaartinen, H. and Maltamo, M., 2004. Automatic detection of harvested trees and determination of forest growth using airborne laser scanning. *Remote Sensing of Environment*, 90(4): 451-462.

Stability of LiDAR-derived raster canopy attributes with changing pulse repetition frequency

Allyson Fox^{1,2}, Chris Hopkinson^{2,1} Laura Chasmer³ & Ashley Wile²

¹Acadia University, Wolfville, Nova Scotia, Canada

allysonfox@gmail.com

²Applied Geomatics Research Group, Lawrencetown, Nova Scotia, Canada

Chris.hopkinson@nscc.ca

³Cold Regions Research Centre, Wilfrid Laurier University, Ontario, Canada

lechasme@yahoo.ca

1. Introduction

In this study, we compare LiDAR-derived estimates of canopy height, crown closure and fractional cover collected over a three hour period on July 16, 2005 using variable LiDAR survey configurations. Pulse repetition frequency (PRF) was systematically varied over three regenerating and two mature Acadian mixed-wood forest plots in Nova Scotia, Canada. The objective of this study is to determine if differences in PRF influence typical LiDAR-derived raster representations of canopy structure. The three raster representations of canopy structure that are investigated here are: the canopy height model, crown closure, and fractional cover.

Accurate mapping of vegetation structure has important implications for natural resources management and forest harvesting activities (Dubayah & Drake, 2000; Lim *et al.* 2003), assessing the impacts of natural and anthropogenic change on ecosystems (e.g. Weishampel *et al.* 2007), carbon, water, and energy cycling (Lefsky *et al.* 2005; Chasmer *et al.* 2011). In most cases, applications of LiDAR data for monitoring and ecosystem assessment require that: 1) vegetation metrics accurately represent forest attributes so that validation exercises may be limited or no longer required for a range of species types and ages; and 2) temporal datasets can be compared over a period of years to assess ecosystem change. Variations in LiDAR-derived data products due to differences in LiDAR survey configurations, points processing, or rasterisation procedures may vary in magnitude depending on foliage and branching structure of vegetation or vegetation height (e.g. Hopkinson, 2007; Naesset, 2009). When LiDAR data metrics are used within ecosystem or biogeochemical models, slight differences in canopy structural attributes used to parameterize the model could result in compounding errors over time.

Several studies have examined the influence of LiDAR survey configurations on the distribution of laser returns within the canopy (e.g. Holmgren *et al.* 2003; Naesset 2004; Chasmer *et al.* 2006; Hopkinson 2007; Lim *et al.* 2008; Naesset, 2009). In addition to data acquisition settings, the amount of pulse penetration into and through the canopy varies due to the structural characteristics and density of the foliage and ground cover encountered. It has been reported that surveys configured using lower PRFs (typical of older data collections) tend to result in lower laser pulse frequency distributions in the upper quantiles when compared with higher PRF (or more recent) surveys (Hopkinson, 2007; Lim *et al.* 2008). Notwithstanding laser pulse energy plays an important role (e.g. Chasmer *et al.* 2006; Hopkinson, 2007), increasing point density with PRF also increases the probability of sampling tree tops. However, the influence of PRF-induced shifts in the canopy point cloud on derivative raster canopy attributes are not well understood. The objective of this study is to investigate whether or not raster canopy height, crown closure and fractional cover attributes are stable across four different PRF settings over a forested Acadian mixed wood landscape.

2. Method

2.1 Study area

The site is located approximately 5 km south-east of the town of Middleton, within the Annapolis Valley, Nova Scotia, Canada (N 44° 54' 59", W 65° 04' 41") (Figure 1). The area flown is approximately 1 km long by 0.5 km wide, and twenty extraction plot locations equalling approximately 1 hectare in area were defined within this area (Figure 1). The Acadian mixed-wood forest is characteristic of many mixed-wood forests found in Nova Scotia, and comprises of mainly *Acer saccharum* Marsh., *Pinus strobus* L., and *Betula alleghaniensis* Britt.

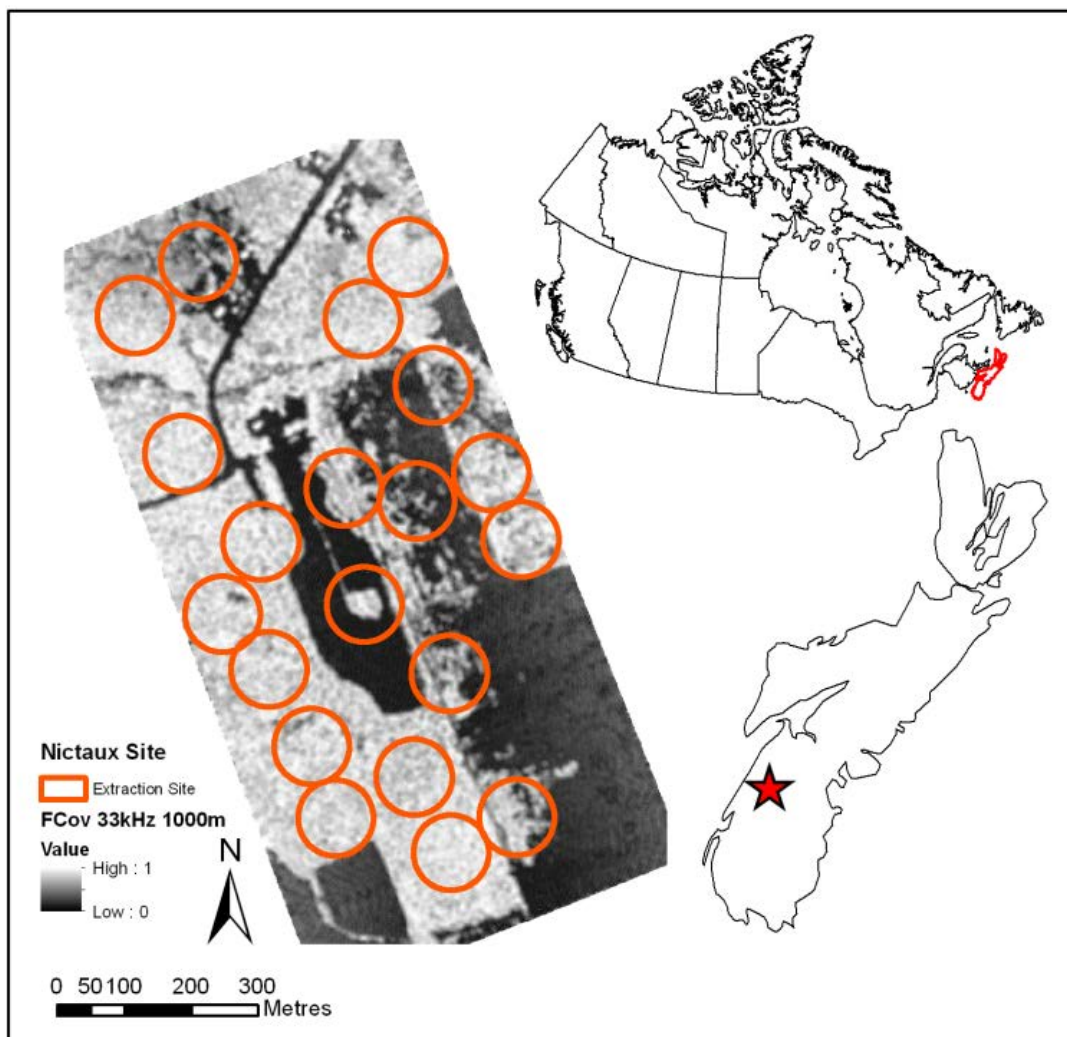


Figure 1: Study area showing fractional cover (33kHz, 1000m altitude, narrow beam) and 20 x 1 hectare LiDAR canopy attribute extraction plot locations.

2.2 Airborne LiDAR data collection and analysis

Airborne LiDAR data were collected during a single flight on July 16, 2005 using an Optech Inc. ALTM 3100, discrete four pulse return system owned and operated by the Applied Geomatics Research Group (AGRG), Nova Scotia. Four LiDAR configurations were flown by varying

PRF (Table 1), and keeping all other data collection parameters equal. All data collections were conducted at 1000 m a.g.l. using a narrow (0.3 mRad) beam divergence (1/e) and a scan angle of ± 20 degrees from nadir.

Table 1. Flight configuration parameters for four data collections.

Configuration	PRF (kHz)	Point Density/m ²
1	33	0.92
2	50	1.30
3	70	1.83
4	100	2.32

Laser returns were classified into ground, below canopy (1.5m threshold) and all hits files within the Terrascan software package (Terrasolid, Finland). Ground returns were used to derive a 1 m resolution digital elevation model (DEM), using an inverse distance weighting approach. A digital surface model (DSM) was created based on a localized maxima algorithm, which uses returns at the maximum height within a specified search radius (in this case a 2.5 m search radius was adopted for all datasets to ensure no data voids). Canopy height surfaces were determined by subtracting the DEM from the DSM to create a canopy height model (CHM) at 1 m resolution for each configuration. Canopy fractional cover was determined as the ratio of the canopy points above 1.5 m to all hits (throughout the canopy to ground). Hopkinson and Chasmer (2009) investigate four LiDAR-based models of canopy fractional cover, and the simple ratio method was adopted in this case as it is widely used and straightforward. Additionally, the CHM was thresholded at 5 m and reclassified into crown (>5 m) and non-crown (<5 m) to develop a binary mask of crown closure. The choice of 5 m was arbitrary and a priori not optimal for all canopy conditions but it was chosen by trial and error as a median canopy height and is used for the sake of illustration. A more in depth analysis is needed to identify an optimal threshold selection based on local canopy conditions but this approach was adequate for the purpose of identifying any systematic PRF dependence.

3. Results and Discussion

Comparisons were performed on plot-level means and maxima of the CHM, fractional cover and crown closure. The 33 kHz data were selected as the baseline datasets, and all observed PRF-dependent differences in the raster canopy attributes were tested for significance using a paired t-Test. In all comparisons the differences were significantly different at the 99% level of confidence (Table 2). Table 2 illustrates the progression of mean height determined by the CHM's, increasing with an increase in PRF. The 33 kHz setting gives the lowest height, and 100 kHz the highest, confirming the anticipated result that to detect higher elements of the canopy, a higher density of pulses is required. Deviations of canopy height per PRF, compared with data collected at 33 kHz are shown in Figure 2.

Figure 3 illustrates canopy height derived from the 33 kHz data and the grid-level height residual between canopy heights derived from 33 kHz and 100 kHz. Differences between the PRFs are emphasized at the edges of crowns, and 99% of the change falls in the range of -3.8 m to +3.8 m. This also illustrates the PRF sampling influence on crown morphology in that the lower sampling density associated with 33 kHz completely fails to sample many smaller individual crown elements in some of the more open areas of the study area.

Table 2 illustrates the percentage of change in mean crown closure, determined above a canopy height threshold of 5 m, from 33 kHz to 50, 70 and 100 kHz, respectively. In general, crown

closure increases slightly with an increase in PRF. Where complete crown closure exists, an increase in PRF deviations in crown closure are often less than 1 percent (Figure 4). However, as canopy openness increases, increases in PRF shows increased variability of crown closure up to 7% (Figure 4).

Table 2. Statistical descriptions of canopy height and fractional cover derived using different PRFs

PRF	CHM mean plot height statistics		Fractional Cover	Crown Closure plot statistics		
	Mean (stdev)	Max		Mean cover % (stdev)	Mean difference from 33 kHz % (stdev)	Maximum mean difference from 33 kHz
33	13.01m (5.93)	22.74m	71 (18)	79.7 (28.2)	-	-
50	13.41m (5.94)	22.79m	82 (11)	81.1 (27.3)	1.4 (1.5)	4.6
70	13.50m (5.96)	22.86m	78 (12)	81.3 (27.2)	1.6 (1.8)	5.6
100	13.63m (5.96)	22.93m	82 (10)	81.7 (26.8)	2.0 (2.2)	6.6

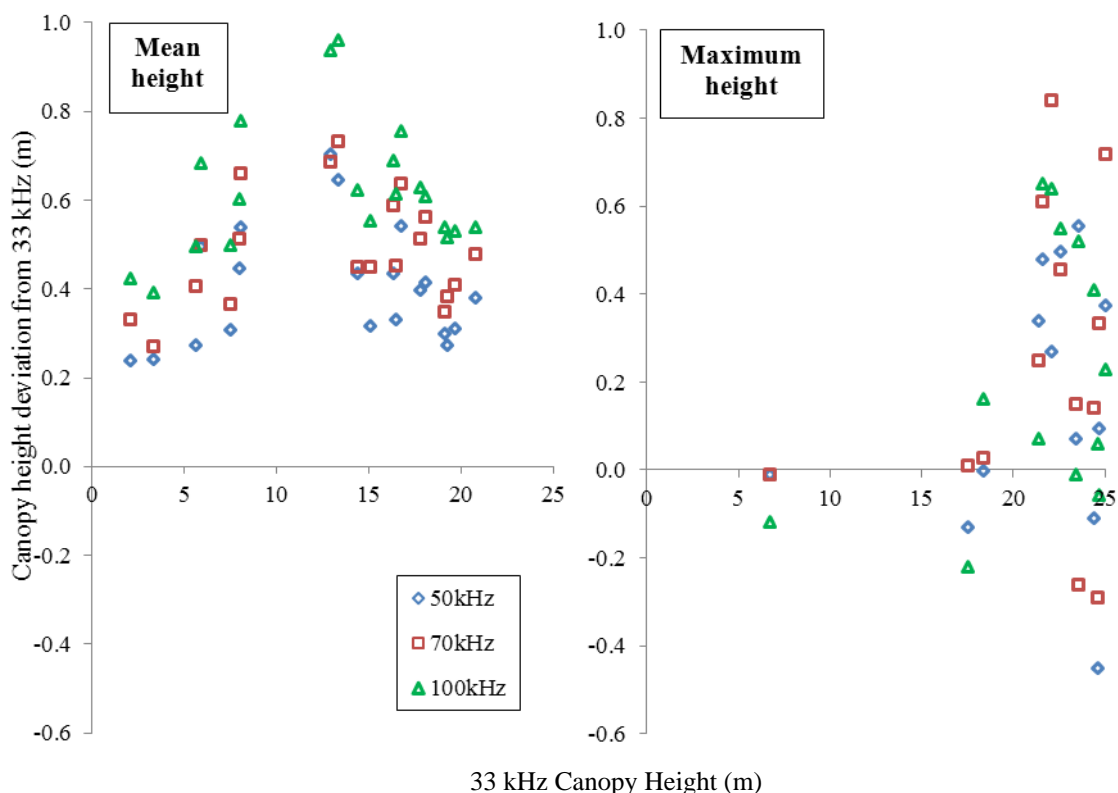


Figure 2. Plot-level mean and maximum canopy height model residuals by PRF (50, 70, 100) from 33 kHz.

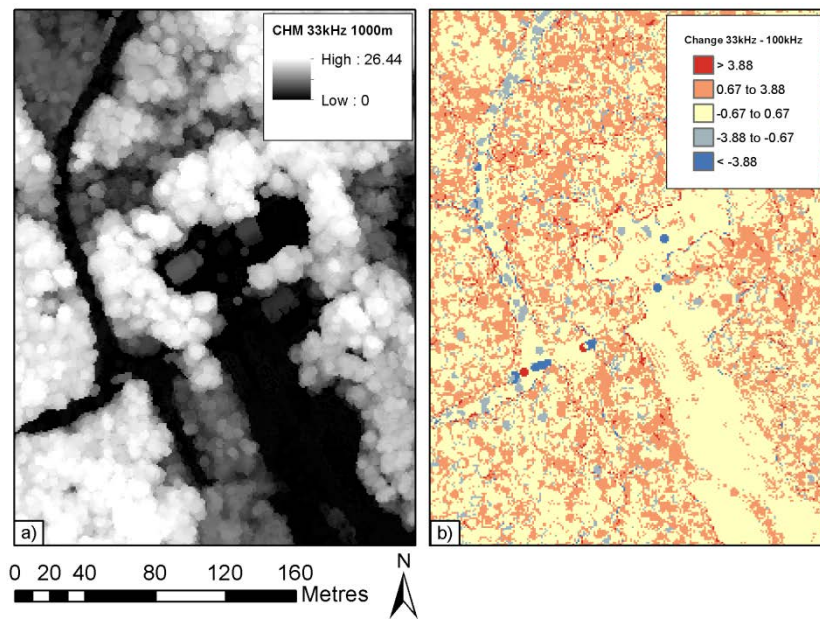


Figure 3: a) Subset of CHM for a mixture of Acadian mixed wood regeneration and mature stands generated from the 33 kHz LiDAR data; b) CHM difference image (100kHz - 33kHz)

Increases in PRF do not systematically cause increases in canopy fractional cover (Table 2). For example at 70 kHz, estimates of fractional cover are lower than that derived from data collected at 50 kHz. However, when compared with 33 kHz, fractional cover derived using higher PRFs are greater and all differences are significant at the 99% level of confidence. The largest difference is at 100 kHz, where canopy fractional cover is 11% greater than at 33 kHz. The deviations of fractional cover per PRF, compared with data collected at 33 kHz are shown in Figure 5. The variation in the type of plots sampled (varying age and openness, amount of understory) and the presence of mid-canopy returns representing canopy cover all influence depth of penetration of pulses into the canopy. Moreover, it has earlier been demonstrated that pulse power plays a critical role in controlling the level of pulse penetration and detection with canopies (Chasmer et al. 2006; Hopkinson, 2007), so it is important to emphasise that canopy representation is not a simple function of sampling point density. The observations here of a variable simple ratio-based fractional cover appear to be indicative of behaviour that is influenced both by pulse power and sampling density. For example, while it is known that increased pulse power increases the chances of ground level returns in continuous canopy cover (Hopkinson, 2007), increased sampling density will increase ground level representation in regions of more open canopy.

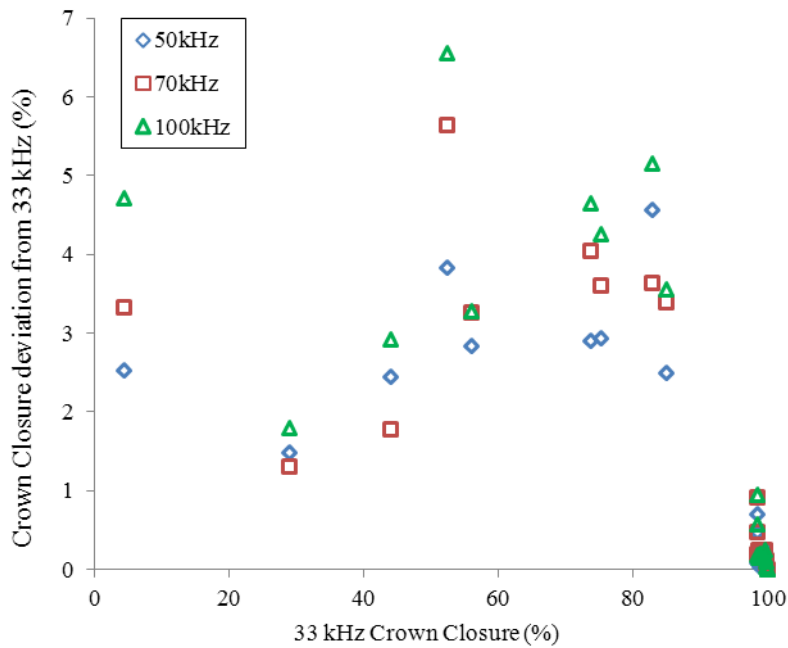


Figure 4. Percent deviation of crown closure from 33 kHz.

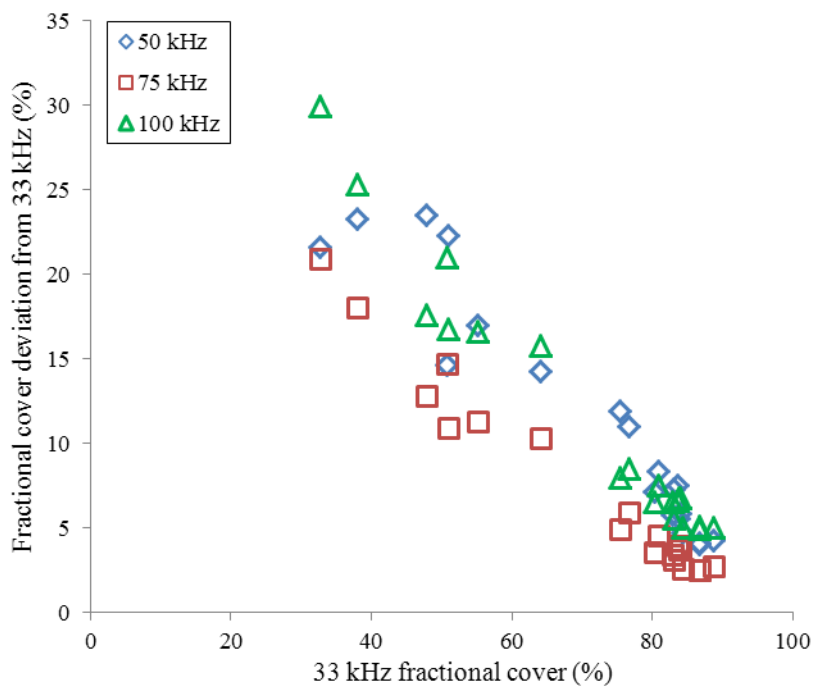


Figure 5. Percent deviation of fractional cover from 33 kHz.

Crown closure can be compared to fractional cover in that crown closure considers the gaps between individual tree crowns, whereas fractional cover is an index of all canopy gaps whether inside or between tree crowns. In theory, then, fractional cover should illustrate a smaller value than crown closure for an equivalent height threshold. In this study, different height thresholds were used (1.5 for fractional cover and 5 m for crown closure) for practical reasons, so the results are not directly comparable. Nonetheless, crown closure does illustrate a slightly higher cover at both 33 kHz (80% as opposed to 71%) and 70 kHz (81% as opposed to 78%). However

differences between crown closure and fractional cover at 50 kHz and 100 kHz are not significant. These observations suggest that fractional cover results are less systematically influenced by changes in PRF (and sampling density) than crown closure derived from thresholded CHMs.

4. Conclusion

The results of this study show that LiDAR-derivative raster canopy attributes are not stable with PRF. Higher canopy elements (such as tree tops) are more frequently sampled at higher PRF due to the increased sampling density, which also causes an upward shift of the CHM. This is important to verify because it is also known that increased PRF coincides with reduced pulse power and weaker detection capability within and below the canopy (Hopkinson, 2007). Average differences in mean canopy height per plot between 33 kHz and 50 kHz, 70 kHz, and 100 kHz are 0.40 m, 0.49 m, and 0.62 m, respectively, and for max plot-level heights are 0.05 m, 0.12 m and 0.19 m, respectively. Differences in the distribution of laser returns through the canopy also affect canopy fractional cover, whereby higher PRFs display some tendency to lead to higher fractional cover estimates by up to 11% on average compared with lower PRFs (e.g. 33 kHz). It is speculated that this increase in the simple ratio fractional cover is more associated with reduced return representation at ground level than it is due to increased detection within the canopy (e.g. Hopkinson, 2007).

Vertical shifts in laser returns throughout the canopy combined with variable sampling coverage of the outer canopy surface caused by varying PRF will result in significant systematic differences in gridded canopy height and CHM thresholded crown closure but equally significant but less systematic differences in canopy fractional cover. Therefore, we conclude that LiDAR derived raster canopy attributes are not stable with PRF and such settings must be considered and accounted for when conducting multi-temporal change detection or site to site comparison studies. Furthermore, these settings should be accounted for (or error margins calculated) if developing and applying LiDAR-based models of vegetation structure, growth or biomass across many different datasets.

Acknowledgements

Many thanks to Heather Morrison, Allison Muise, and Neville Crasto for their support. The AGRG LiDAR laboratory was set up with funds from the Canada Foundation for Innovation.

References

- Chasmer, L., Hopkinson, C., Smith, B. and Treitz, P., 2006. Examining the Influence of Changing Laser Pulse Repetition Frequencies on Conifer Forest Canopy Returns. *Photogrammetric Engineering & Remote Sensing*, 72, 1359-1367.
- Chasmer, L., Kljun, N., Hopkinson, C., Brown, S., Milne, T., Giroux, K., Barr, A., Devito, K., Creed, I., and Petrone, R., 2011, Characterizing vegetation structural and topographic characteristics sampled by eddy covariance within two mature aspen stands using LiDAR and a flux footprint model: Scaling to MODIS, *J. Geophys. Res.*, 116, G02026, doi:10.1029/2010JG001567.
- Dubayah, R.O, and J.B. Drake, 2000, LiDAR remote sensing of forestry, *Journal of Forestry* 98,6:44-46
- Holmgren, J., Nilsson, M., and Olsson, H., 2003. Simulating the effects of LiDAR scanning angle for estimation of mean tree height and canopy closure. *Canadian Journal of Remote Sensing*, 29, 623.
- Hopkinson, C., 2007. The influence of flying altitude, beam divergence, and pulse repetition

- frequency on laser pulse return intensity and canopy frequency distribution. *Canadian Journal of Remote Sensing*, 33, 312-324.
- Hopkinson, C. and Chasmer, L.E., 2009. Testing LiDAR models of fractional cover across multiple forest ecozones. *Remote Sensing of Environment*. 113, 275-288.
- Lefsky, M.A., Turner, D.P., Guzy, M., and Cohen, W.B., 2005, Combining LiDAR estimates of aboveground biomass and Landsat estimates of stand age for spatially extensive validation of modelled forest productivity. *Remote Sensing of Environment*. 95,4: 549-558.
- Lim, K., Hopkinson, C., Treitz, P. 2008. Examining the effects of sampling point densities on laser canopy height and density metrics at the forest plot level. *Forestry Chronicle*. 84, No. 6. pp.
- Lim, K., Treitz, P., Wulder, M., St-Onge, B., and Flood, M., 2003, LiDAR remote sensing of forest structure. *Progress in Physical Geography*. 27,1:88-106.
- Næsset, E., 2004. Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. *Remote Sensing of Environment*, 91, 243-255.
- Næsset, E. 2009. Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*. 113, 148-159.
- Weishampel, J.F., Drake, J.B., Cooper, A., Blair, J.B., and Hofton, M., 2007, Forest canopy recovery from the 1938 hurricane and subsequent salvage damage measured with airborne LiDAR. *Remote Sensing of Environment*. 109,2:142-153.

Characterizing Peat Swamp Forest Environments with Airborne LiDAR Data in Central Kalimantan (Indonesia)

Hans-Dieter Viktor Boehm¹, Veraldo Liesenberg², Juergen Frank¹ & Suwido H. Limin³

¹Kalteng Consultants [www.kalteng.org], Kirchstockacher Weg 2, D-085635, Hoehenkirchen, Germany. E-mail: viktorboehm@t-online.de

²Faculty of Geosciences, Geotechnique and Mining, TU Bergakademie Freiberg, Bernhard-von-Cotta-Str. 2, D-09599, Freiberg, Germany.

³Universitas Palangka Raya and CIMTROP, Jalan Yos Sudarso, Palangka Raya, 73112, Central Kalimantan, Indonesia.

1. Introduction

Natural tropical Peat Swamp Forests (PSF) are important for their rich biodiversity and because they represent important carbon pool (Page *et al.* 2002). However, PSF are decreasing due to conversion into farm land, by excessive draining, the use of shifting cultivation on a large scale, pal oil plantation, illegal logging and forest fire. This increases the interest for understanding in an ecological point of view and mapping such environments as they are recognized as an important source of carbon released in the atmosphere (Sorensen 1993, Page *et al.* 2002, Jaenicke *et al.* 2008, Ballhorn *et al.* 2009, Boehm *et al.* 2010).

According to Hyde *et al.* (2007) airborne Light Detection And Ranging (LiDAR) data is nowadays the best remotely sensor to investigate biophysical parameters (e.g. tree height and canopy diameter which are strongly correlated with above-ground biomass and leaf area index (LAI) (Hajsek *et al.* 2009). In peatland areas the great variety of ecosystems and its ecological rule is still not fully understood. The influence of selective logged areas (e.g. species composition, their structures and canopy properties) on global change issues also remains therefore a big challenge. Consequently, to optimize the biophysical properties characterization, a better understanding of how LiDAR measurements could be useful for ecological studies in such critically endangered forests is still necessary.

In August 2007 we mapped by helicopter different PSF transactions with Riegl LiDAR Technology LMS-Q560 in Central Kalimantan, Indonesia. In this study, our main objectives were: a) to evaluate the peat surface/profile and their associated domes; b) to estimate the dependence of tree canopy height with both peat dome thickness and peat dome slope variations; c) to estimate Above Ground Biomass (AGB) variations based on tree canopy height and physiognomy variations according to optical remotely sensed spectral signatures and field inspection; and d) to demonstrate the applicability of LiDAR technology to detect logging activities in tropical peat swamp forest environments. Additionally, we would like to demonstrate first results of our second LIDAR survey made in August 2011. We found that in an undisturbed tropical PSF area the average tree-height increased from 15.32m to 17.18m by difference of 1.86m which is 12%.

2. Study Area Description

Our study areas consist on one LiDAR transects located inside the Central Kalimantan province, Indonesia (Figure 1). One Landsat image acquired on July 16, 2000 shows the remaining forest in the region in different green colour for 2000. In fact, PSF over the region were impacted with extensive logging activities in the 90th using railways in the swaps for logging roads by the big concessions. The logging leads to severe peat damages with reasonable amount of carbon released to the atmosphere, especially during peat fires in 1994, 1997, 2002, 2006 and 2009.

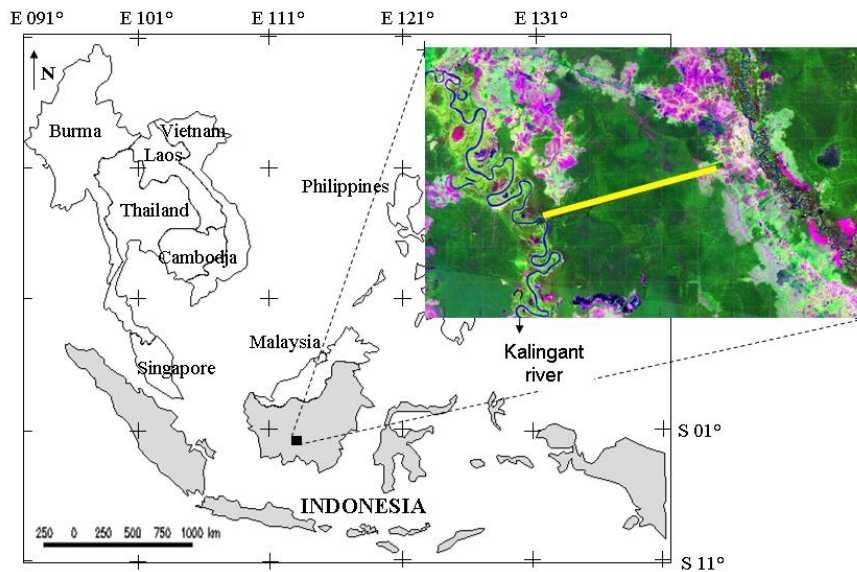


Figure 1: Study area location. (a) Indonesia and (b) subset of Landsat image for Central Kalimantan and the yellow line indicates the LiDAR transect.

The selected LiDAR transect is located between Katingan river and the transmigration village Habaring Hurung (Figure 2). This PSF was selective logged by the concession in the 90th. According to MODerate Imaging Spectroradiometer (MODIS) products, the Leaf Area Index (LAI) of the logged peat swamp forest at the region is in average close to 4, typical for tropical rain forested areas although close to 6 in undisturbed areas (Liesenberg *et al.* 2010). Areas of pasture, small agricultural fields, small villages at the rivers and actual degraded forest resultant from selective logging can be also observed in the surroundings of the selected area.

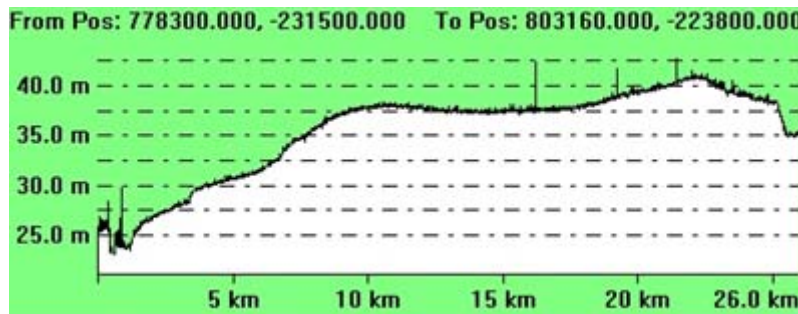


Figure 2: Peat Surface profile of the selected LiDAR transect.

Peat surface profile of the analyzed transect starting left from the river Katingan (23m) reaching to the transmigration village Habaring Hurung, right (Figure 2). A double peat dome is found at 38m and 41m. Between the kilometres 10 and 23 a high plateau is also visible. More south to this a small river can be monitored on the peat area, which is draining the high water content. The transect shows an asymmetrical peat dome on the right with a terrain peak close to 41m; Table 1). The peat surface increases quickly with a maximum of 1.8m/km and an average of 1.0m/km from the Katingan to the first peat dome with 38m.

Table 1: Summary of the LiDAR transect under study

Transect	River Level	Peat Domes	Transect Length	Max. Slope	Av. Tree Height	Max. Tree Height
Katingan – Rungan	23m	38m to 41m	26km	1.8m/km	15.8m	34.5m

2.1 Material and Methods

2.1.1 LiDAR Data Processing

The airborne LiDAR transects were acquired on August 8th, 2007 in a helicopter campaign conducted by Kalteng Consultants and Milan Geoservice GmbH. We collected for the above described tracks approx. 1300ha of PSF with approx. 1.4 laser beams per square meter. The Riegl-airborne laser-scanner LMS-Q560 was mounted under the Bell 206 helicopter. Small footprint LiDAR data was collected for a flight altitude of approx. 500m with a scan angle of 60° with produced a swap-width of approx. 500m (Boehm *et al.* 2007, Boehm and Frank, 2008). We used for this analysis the first and last pulse Laser echoes only, but the full-wave data are available for more detailed analyzes. The Laser scanner had a pulse rate of 66kHz resp. 100kHz with a beam divergence of 0.5mrad or a footprint of 0.25m. The ground backscattering in PSF through the canopy was responsible for 1 to 3% of the 0.5mrad Laser beams. Complementary we explore a second LiDAR dataset acquired on August 5th, 2011 using the same sensor system as 2007 in order to demonstrate change detection over the time.

The DGPS reference station was located at the airport of Palangka Raya which has an elevation of 25.0m above sea level. The position and orientation of the LiDAR system on the helicopter was measured by an Inertial Navigation System (INS) and a GPS located on the tail boom with 256Hz. The Riegl LMS-Q560 airborne Laser scanner system itself allows height measurements with an accuracy of $\pm 0.02\text{m}$. After the correction of the attitude of the helicopter, the elevation accuracy of each Laser beam was $\pm 0.15\text{m}$ with a root mean square error (RSM) of $\pm 0.5\text{m}$ in both x- and y-direction. The processed laser beams were divided into ground surface and over ground classes and converted in order to digital terrain model (DTM) and digital surface model (DSM), respectively, at a spatial resolution of 1m.

2.1.2 Data Analysis

2.1.2.1 Peat Drilling, Peat Surface and Landsat signatures

A total of 31 peat drilling measurements were conducted in order to determine the peat depth of our selected transect. The spatial distribution of the samples keeping an average distance of 600m as well as the peat depth results is shown in Figure 3.

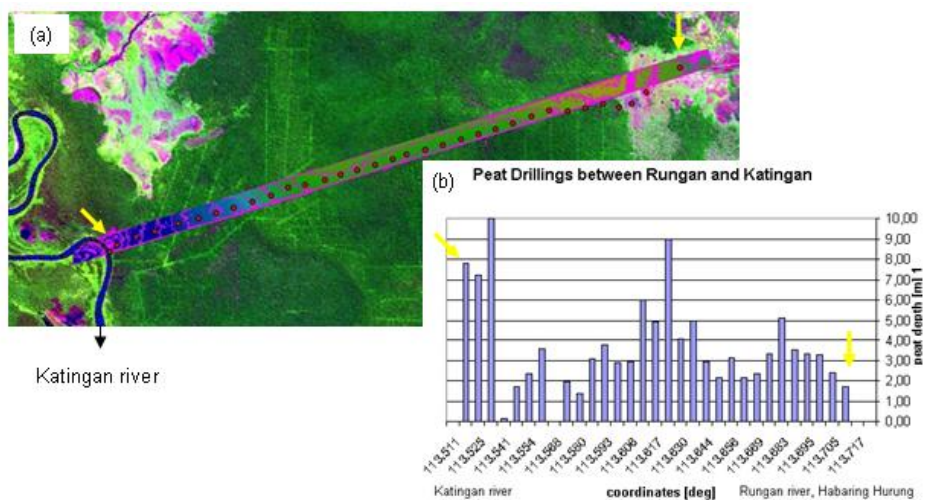


Figure 3: (a) Spatial distribution of the Peat Drilling Samples over the LiDAR transect. Arrows indicate the first and last sample of the peat depth. (b) Results of the Peat drilling measurements showing the Peat depth results up to 10m.

According to this figure, we observe a high variability with the highest values occurring close to the Katingan River and a double dome is also visible compared to peat surface showed in Figure 2. We related the peat drilling measurements with the tree canopy height establishing 1-ha plot over each measurement and the peat dome slope as described better in the next section. A detailed explanation about the procedures in the field may be found in Page *et al.* (2004).

Complementary, we also conducted peat drilling analysis in a second transect whose spatial distribution as well as peat depth is shown in Figure 4. In this figure, 91 peat drillings along the Kalimantan highway from Kasongan to Tangkiling were measured and related with chemical properties (Boehm *et al.*, 2006).

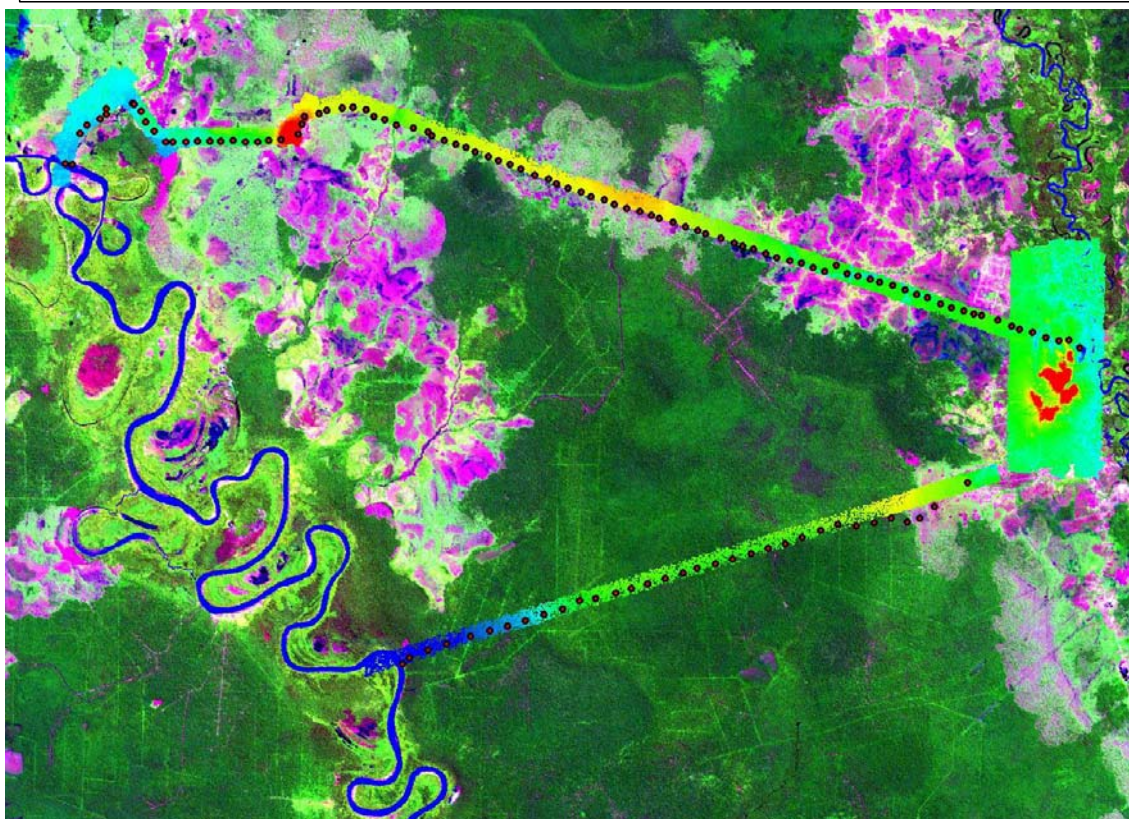
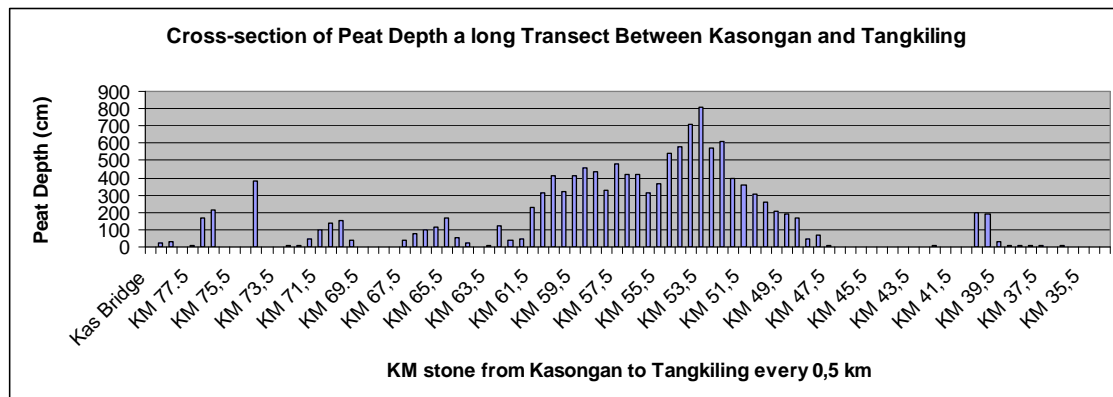


Figure 4: a) Results of the Peat drilling measurements showing the Peat depth results up to 8.1m parallel to the highway Kasongan to Tangkiling with 500m distance. (b) Spatial distribution of the Peat Drilling Samples over the LiDAR transects Kasongan to Tangkiling and Katingan and Habaring Hurung.

Digital Numbers (DN) from Landsat images covering different dates including pre- and after LiDAR acquisition were converted to radiance data and then converted to surface reflectance (hemispherical directional reflectance factor, HDRF) using the Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm. Input data for the radiative transfer code were collected by instruments in a tower maintained by the Asia Flux Network inside the study area as well as MODIS products. We selected different regions based on our LiDAR transect in order to characterize different spectral patterns.

2.1.2.2 Tree Canopy Height vs. Peat Dome Slope Analysis

We calculated the peat dome slope by counting the difference between the DTM values of two sample plots with 1-ha each and converting the altitude difference (e.g. m per km) in an equidistance of 250m. We proceed with the extraction of tree height for the DSM in four sub-plots (e.g. 50mx50m) where we just account for the signal coming from the dominant trees. The sub-plots were used in order to take a sample of each quadrant. The data from each transect was divided into training (70%) and validation (30%) datasets for statistical purposes (Boehm et al. 2010).

The relationship between tree height and peat dome slope employed a linear regression analysis (i.e. $y_j = ax + b$). The slope value for each sample plot was used as the predictor for tree height determination. The linear regression analysis was applied to the test dataset in order to perform the accuracy statistics. The accuracy statistics include the root mean square error (RMSE), Bias and their relative counterparts RMSEr and Biasr as described and explained in Muukkone and Heiskanen (2005). As the peat drilling procedure is very difficult to be determined in the field, we proceed with the analysis twice. The first approach was keeping a distance of the 600m as described in the section 2.1.2.1 and the second one the distance of 250m.

2.1.2.3 Biomass Determination

Above Ground Biomass (AGB) were compiled from sample plots of 1-ha collected in the flown acquisition of the transect keeping a distance of 250m between sample plots. We extract both DTM and DSM values for each measured sample plot. Concerning to the DTM, we only account for the lowest values in order to minimize the inclusion of the return signal coming from tree trunks and branches lying on top of the peat surface. Tree Canopy height was determined by subtracting the DSM and the DTM. The AGB was obtained using an allometric equation proposed by Lefsky *et al.* (1999) (Eq. 1) and Uhl *et al.* (1988) (Eq. 2). Diameter at the Breast Height (DBH) was retrieved using the relationship with tree height ($r^2=0.754$) considering field measurements. Finally, tree height and AGB values were plotted for the selected LiDAR transect and also related with the peat dome attributes.

$$AGB = 1.512 \times h(\text{average})^2 \quad \text{Lefsky et al. (1999)} \quad (1)$$

$$AGB = \exp(-2.17 + \ln 1.02(DBH)^2 + \ln 0.39H) \quad \text{Uhl et al. (1988)} \quad (2)$$

2.1.2.4 Characterization of Logging Activities using LiDAR

Since logging activities are a common practice in tropical rain forest environments, we evaluated the complementary use of both DTM, DSM and Aerial ortho-photographs acquired during the LiDAR survey to detect such practices. We compared it with Landsat images and the procedure was based on both visual interpretation and texture analysis (e.g. GLCM). Since no complementary information was provided by local agencies, such as concession information for forest exploitation, we did not classify whether such practices were legal or illegal.

3. Results and Discussion

3.1 Relationship between Peat Dome Depth and Tree Canopy Height

Linear regressions were carried out to establish the correlation between Peat Depth and both Tree Canopy height and Peat Dome Slope and are shown in Figure 5. A weak relation between the variables was found and explained up to 31% of the variation of the forest attributes (e.g. tree canopy height) and up to 16% the peat depth surface in the selected transect. As observed by Boehm *et al.* (2010) the results may reflect the high degree of forest intervention by previous logging activities in the region. In this transect, both predictors height (e.g. tree canopy height averaged and dominant) were similar.

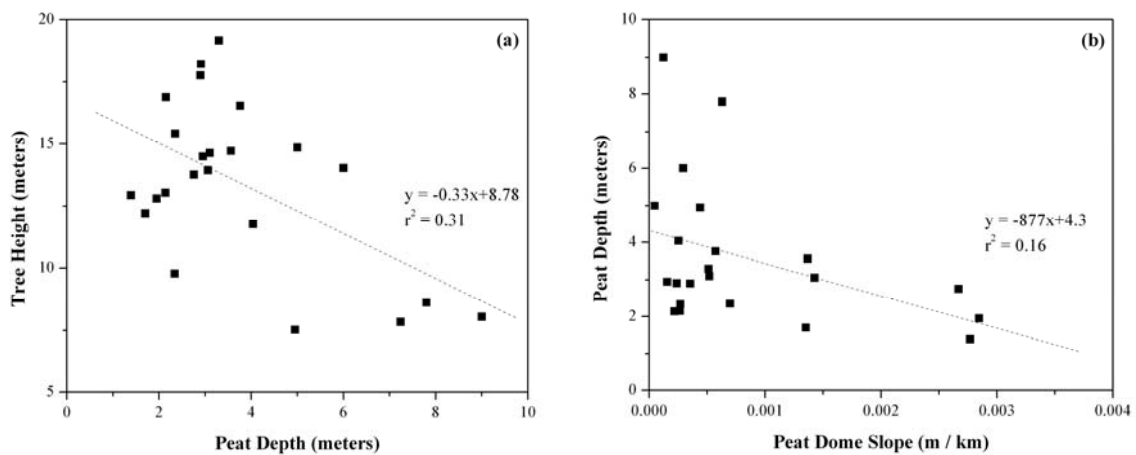


Figure 5: Relationships between (a) Peat Dome Depth and Tree Canopy Height; (b) Peat Dome Depth and Peat Dome Slope.

The relation between Tree Canopy height and Peat dome slope was not affected by the changing the distance between the peat drilling plots (e.g. 600m) and the sample plots (e.g. 250m). Figure 6 shows the relationship between Peat dome slope and both Tree canopy height and AGB. The relationship explains in both situations up to 30% of the variation. At least two non-mutually exclusive hypotheses for positive increase in tree canopy height with peat dome slope can be raised here: (1) soil properties and drainage; and (2) previous disturbance regimes that will be investigated in near future with additional LiDAR data from 2011 as well as field inventory data.

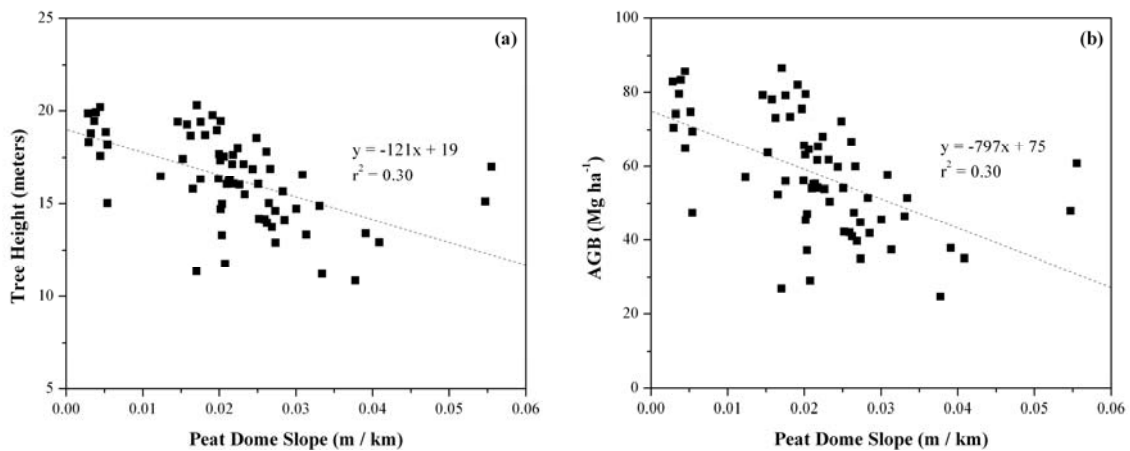


Figure 6: Relationships between (a) Peat Dome Slope and Tree Canopy Height; (b) Peat Dome Slope and Above Ground Biomass.

Landsat signatures showed significant differences in their spectral profiles given samples from

different parts of the peat dome (not shown). The highest differences were observed in the near infrared reflectance which increased from the beginning, middle to the top of the peat dome. This may be an indicative for the occurrence of different physiognomies and canopy architecture changes in the vegetation.

Peat dome slope as a result of the surface topography and peat thickness variations, explains only up to the 30% of the tree canopy height and AGB. Since it does not such forest attributes directly, it may also operate through changes that they may bring about or promote in other characteristics of the peatland, especially related to hydrology, chemistry and organic matter dynamics (balance between peat accumulation and peat degradation) (Jauhiainen *et al.*, 2005, Nishimua *et al.*, 2007). An evaluation taken into account several LiDAR transects sampling different peat characteristics and forest under different disturbance levels should be evaluated for a better assessment of such ecological issues.

3.2 Estimation of Forest Attributes

Several 1-ha sample plots were obtained at the LiDAR transect. We noticed that the average tree height varies from 11 to 20m, whereas the dominant tree canopy height up to 30m (Figure 7a). The lowest tree canopy heights were found close to Katingan river. AGB varies in average from 15 to 200 Mg/ha (Figure 7b). Although a proper field campaign is still necessary to validate the results, an indicative of the LiDAR potential is at least here demonstrated. We intend in near future investigate the potential of the full wave acquisition for the AGB determination instead of using only the first and last LiDAR pulse. A proper LiDAR campaign at ground level with ground LiDAR systems will certainly show promising results, since the accuracy of the AGB is strongly dependent on allometric equations that are normally neither common nor available in tropical regions (Hyde *et al.*, 2007).

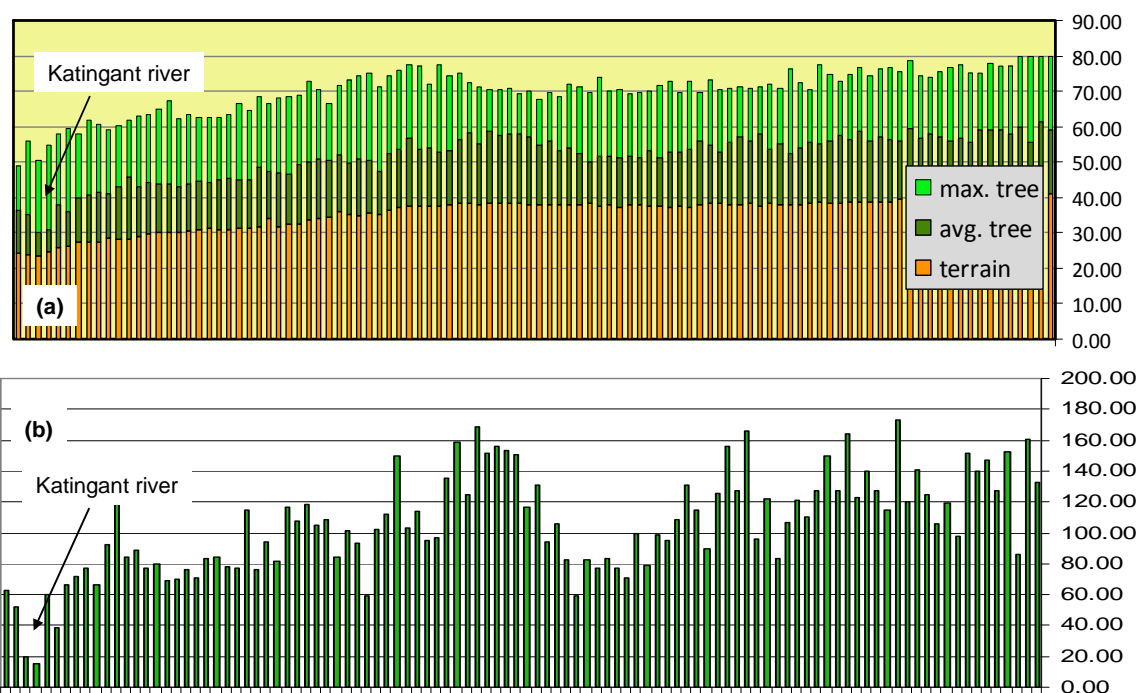


Figure 7: Forest attributes for the 1-ha sample plots from river Katingan to Habaring Hurung considering (a) DTM, average tree height and max. tree height; (b) Average Above Ground Biomass (AGB).

The spatial distribution of the forest attributes extracted from the selected LiDAR transect, such as the Tree Canopy height and AGB are shown in Figure 8. AGB are shown in Figure 9 and 10 employing the two different allometric equations, respectively. The figures show high variability results due to previous selective timber exploitation.

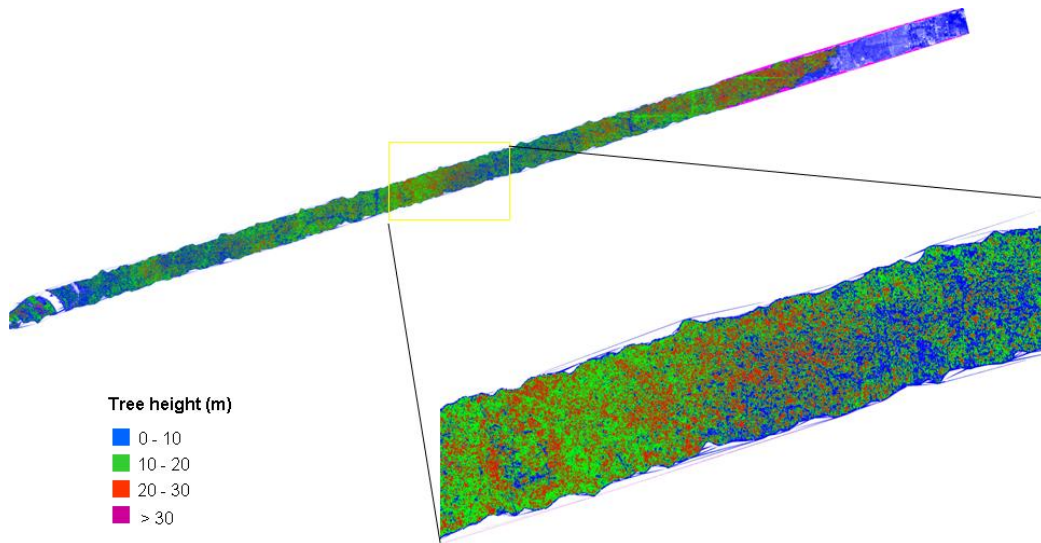


Figure 8: Tree Canopy height variations over the selected LiDAR transect.

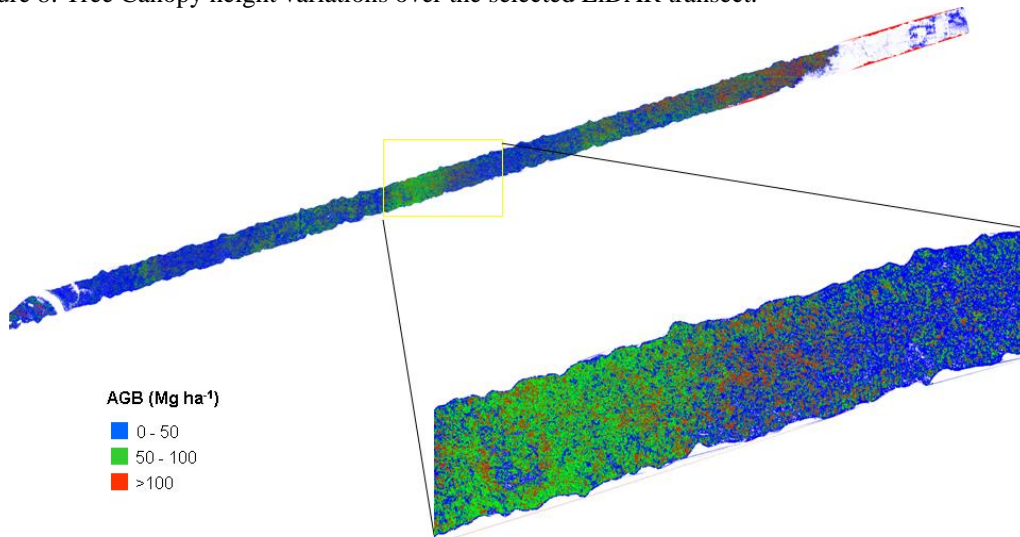


Figure 9: Above Ground Biomass (AGB) variations over the selected LiDAR transect (with Eq. 1).

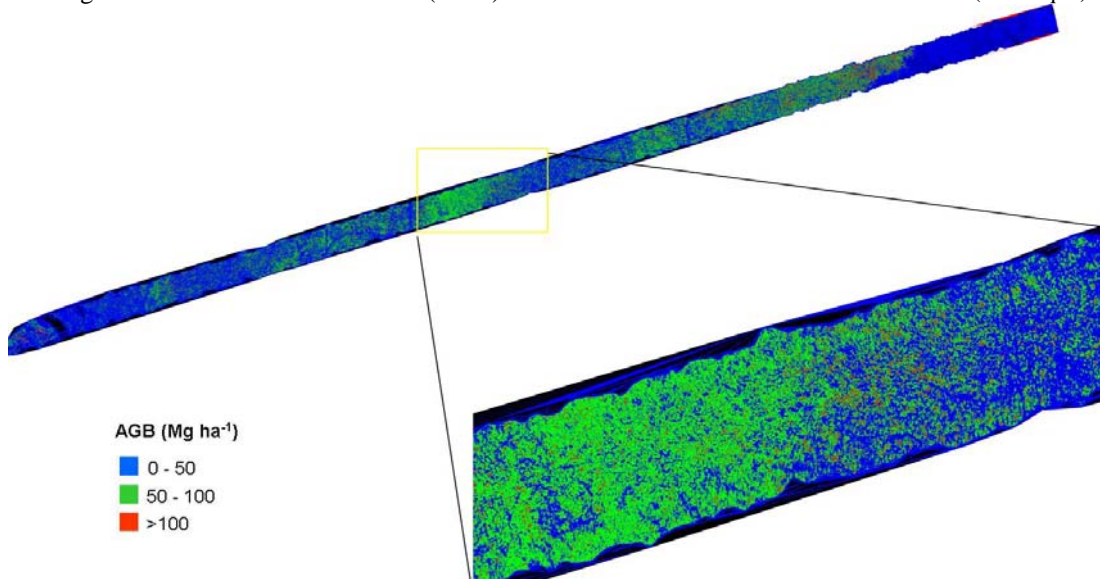


Figure 10: Above Ground Biomass (AGB) variations over the selected LiDAR transect (with Eq. 2).

3.3 Characterization of Logging Effects

Small logging effects on the forest were evident on LiDAR DTM than LiDAR DSM. An assessment of aerial ortho-photographs was also very useful to identify small patches used for wood storage and camping. On the other way, Landsat images were not sufficient to capture such small details besides large and structured channels. The use of texture parameters (e.g. GLCM) and other spectral techniques were also experimented on Landsat images. Figure 11 shows a subset of the LiDAR transect. In this figure, we observe that LiDAR DTM (Figure 11a) capture better the logging tracks than LiDAR DSM (Figure 11b). Such information on regular basis may help local environmental agencies for the monitoring of the endangered forest as well as for REED/MRV practices.

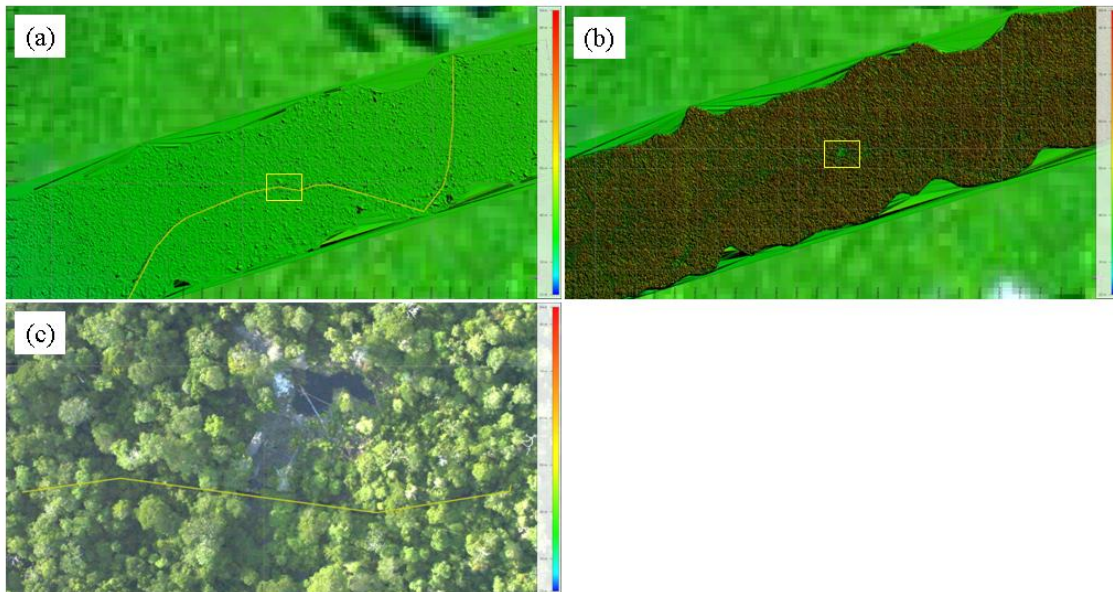


Figure 11: Subset of (a) LiDAR Digital Terrain Model (DTM); (b) LiDAR Digital Surface Model (DSM); and (c) Subset of Aerial Ortho-photograph indicated by yellow square. A yellow line indicates logging channel.

4. Final Remarks

Results showed that LiDAR is an interesting instrument to analyze peat profiles, biomass estimation and relationship between peat dome slope, peat dome depth and tree canopy height. Information on such variables may be useful to assess the dependence of biophysical properties (e.g. tree height, stem diameter and above ground biomass) in distinct peatlands environments. Since peatlands act as a carbon sink, human interventions due to drainage practices for agriculture and selective logging practices of the peat swamp forest may have a stronger impact in the carbon release than relative flat areas that still has to be more investigated with LiDAR technology which we did in a new survey conducted last August 2011.

Nonetheless, further research is being conducted in order to test the dependence of other biophysical parameters to peat dome slope and feature selection techniques for LiDAR data in different vegetation types in Central Kalimantan, Indonesia. In this regard, we will perform a proper analysis with field inventory data conducted from middle July to August 2011. This will allow us to evaluate multi-temporal analysis and to determine the growth or loss of vegetation as well as the peat surface changes.

In our ongoing analysis, we found out that the tree height changes for a selected transect at Sabangau Forest in Central Kalimantan after a period four years was in average up to 2m. This

represents a change of 12% (e.g. from 15.32m to 17.18m, difference of 1.86m). The analysis was performed measuring 20 sample plots of 1ha over a 12km transect for 2007 and 2011. This clearly shows the forest recovery capacity in secondary tropical peat swamp forest under protected conditions (e.g. conservation status).

The presented LiDAR-methodology can be promising in the frame of the REDD (Reducing Emissions from Deforestation and forest Degradation) knowledge of tropical peat swamp forests.

Acknowledgements

The funding for the LiDAR flights and data processing was done by Kalteng Consultants. Additionally we would like to thank the director Mr. Sven Jany, the operators Mr. Jan Giehler, Mr. Matthias Nittel and technician Mr. Detlef Klante from Milan Geoservice GmbH and Mr. Mustafa Syafrudin for the good cooperation during the LiDAR flight campaign. The second author is supported by CNPq/Brazil.

References

- Ballhorn, U., Siegert, F., Mason, M.S., Limin, S., 2009. Derivation of burn scar depths and estimation of carbon emissions with LIDAR in Indonesian peatlands. *Proceedings of the National Academy of Sciences of the USA*, 106, 25-30.
- Boehm, H-D.V., Liesenberg, V., Frank, J., 2010. Relating tree height variations to peat dome slope in Central Kalimantan, Indonesia using small-footprint airborne LiDAR data. In: *Proceedings of the 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems "SilviLaser"*. Freiburg, University of Freiburg.
- Boehm, H-D.V., Frank, J., 2008. Peat Dome Measurements in Tropical Peatlands of Central Kalimantan with a high-resolution Airborne Laser Scanner to achieve Digital Elevation Models. In: *Proceedings of the 13th International Peat Congress, Section 5: Tropical Peatlands*. Tullamore, Ireland.
- Boehm, H.-D.V., Rieley, J.O., Limin, S., Frank, J., Syafrudin M., 2007. Successful Helicopter Flight Trials with Airborne Laser Scanning Technology to measure PSF height and Peat domes in Central Kalimantan. *International Symposium, Workshop and Seminar on Tropical Peatland "Carbon, Climate, Human Interactions, Carbon Pools, Fore, Mitigation, Restoration and Wise Use*. Yogyakarta, Indonesia.
- Boehm, H.-D.V. and Sulistiyanto, Yustinus (2006) Peat depth, minerals below peat, carbon, fires and its characteristics a long transect between Tangkiling and Kasongan, Central Kalimantan. *Presented at International Workshop Tropical Rain Forest and Boreal Forest Disturbance and Their Affects on Global Warming, Palangka Raya, Kalimantan, Indonesia 17-18. Sept. 2006 and published in proceedings*.
- Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, E.D, Fölster, E.H., Higuchi, N., Kira, T., Lescure, J.-P., Nelson, B.W., Ogawa,H., Puig, H., Riera, B., Yamakura, T., 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, 145: 87-99.
- Hajnsek, I., Kugler, F., Seung-Kuk Lee, Papathanassiou, K.P., 2009. Tropical Forest Parameter Estimation by Means of Pol-InSAR: The INDREX-II Campaign. *IEEE Transactions on*

Geoscience and Remote Sensing, 47: 481-493.

- Hyde, P., Nelson, R., Kimes, D., Levine, E., 2007. Exploring LiDAR–RADAR Synergy - Predicting Aboveground Biomass in a Southwestern ponderosa pine forest using LiDAR, SAR, and InSAR. *Remote Sensing of Environment*, 106: 28–38.
- Jaenicke, J. Rieley, J.O., Mott, C., Kimman, P., Siegert, F., 2008. Determination of the amount of carbon stored in Indonesia peatlands. *Geoderma*, 147: 151-158.
- Jauhiainen, J., Takahashi, H., Heikkinen, J.E.P., Martikainen, P.J., Vasander, H., 2005. Carbon fluxes from a tropical peat swamp forest floor. *Global Change Biology*, 11(10):1788-1797.
- Liesenberg, V., Boehm, H.-D.V., Gloaguen, R., 2010. Spectral variability and discrimination assessment in a tropical Peat swamp landscape using CHRIS/PROBA data. *GIScience and Remote Sensing*, 47: 541–565.
- Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A. and Harding, D., 1999. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment*, 70: 339-361.
- Muukkonen, P., Heiskanen, J., 2005. Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data. *Remote Sensing of Environment*, 99: 434-447.
- Nishimua, T.B., Suzuki, E., Kohyama, T., Tsuyuzaki, S., 2007. Mortality and growth of trees in peat-swamp and heath forest in Central Kalimantan after severe drought. *Plant Ecology*, 188(2): 165-177.
- Page, S., Siegert, F., Rieley, J.O., Boehm, H.-D.V., Jaya, A., Limin, S., 2002. The amount of carbon released from peat and forest fires in Indonesia during 1997. *Nature*, 420: 61-65.
- Page, S.E., Wust, R.A.J., Weiss, D., Rieley, J.O., Shotyk, W., Limin, S.H., 2004. A record of Late Pleistocene and Holocene carbon accumulation and climate change from an equatorial peat bog (Kalimantan, Indonesia): Implications for past, present and future carbon dynamics. *Journal of Quaternary Science*, 19(7): 625-635.
- Sorensen, K. W., 1993. Indonesian peat swamp forests and their role as a carbon sink, Chemosphere. *Chemosphere*, 27: 1065-1082.
- Uhl, C., Buschbacher, R., Serrao, E.A., 1988, Abandoned pastures in eastern Amazonia: patterns of plant succession. *Journal of Ecology*, 76: 663-681.

Comparison of the spatial pattern of trees obtained by ALS based forest inventory techniques

Petteri Packalén^{1,3}, Jari Vauhkonen¹, Eveliina Kallio¹, Jussi Peuhkurinen¹, Juho Pitkanen², Inka Pippuri¹ & Matti Maltamo¹

¹Faculty of Science and Forestry, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland, petteri.packalen@uef.fi, jari.vauhkonen@uef.fi, eveliina.kallio@uef.fi, jussi.peuhkurinen@uef.fi, inka.pippuri@uef.fi, matti.maltamo@uef.fi

²Finnish Forest Research Institute, Joensuu Research Unit, P.O. Box 68, FI-80101 Joensuu, Finland, e-mail: juho.pitkanen@metla.fi

³College of Forestry, Oregon State University, 204 Peavy Hall, Corvallis, OR 97331, USA

Abstract

The spatial pattern of trees in a forest can be defined as the locations of the trees in relation to each other. The spatial arrangement of a point (e.g. tree) pattern may be random (Poisson), clustered or regular. In this study the spatial pattern of trees was determined at the plot level by using L function, which is a square root transformation of Ripley's K function. The spatial pattern of tree was summarized into three classes: regular, random and clustered. The study was carried out with 79 sample plots located in a managed forest area in eastern Finland. Tree maps were produced with the individual tree detection (ITD) and semi-individual tree detection (Semi-ITD) and spatial patterns of trees were calculated from the tree coordinates. The spatial pattern of trees was also predicted directly by using patch metrics calculated from the canopy height model as explanatory variables (AREA). The low resolution airborne laser scanning (ALS) data was used in the AREA and the high resolution data in the ITD and Semi-ITD. The Kappa value for the ITD was almost zero, which indicates virtually random classification. The AREA and Semi-ITD methods were clearly more accurate than the ITD. Kappa values for the Semi-ITD and AREA were 0.34 and 0.24, respectively, which nevertheless cannot be considered to be very good. However, determining the spatial pattern of trees by ALS is somewhat unexplored field of study. It should be studied how well the spatial pattern of trees can be determined in different type of forests.

Keywords: individual tree detection, canopy height distribution

1. Introduction

The spatial pattern of trees in a forest can be defined as the locations of the trees in relation to each other. It can be regular, random, clustered or any combination of them (e.g. Pielou 1960, Tomppo 1986). Commonly the spatial pattern is compared to a statistically random pattern, so called Poisson forest (Diggle 1983). Field work for determining the spatial pattern of trees is laborious and expensive. Therefore explicit information about the spatial pattern of trees, such as spatial pattern indices, tree location maps, or competition indices, are quite rarely used in growth and yield models or in any other use case except in scientific research.

Trees within a stand affect development and survival of their neighbours. Therefore it is not *a priori* acceptable to treat trees and their attributes as independent random variates (Penttinen *et al.* 1992). In forestry the spatial pattern of trees is mostly considered implicitly. For example, in thinning one aim is to produce regular spatial pattern of trees on stand-level in order to optimize the use of growing space (Pukkala 1990). In that the case spatial pattern is taken into account

although it's not specifically determined and expressed in numerical form.

Although remote sensing is an obvious tool to determine the spatial patterns of trees in a forested environment there are only very few studies about it. Coops and Culvenor (2000) related scene texture variance to a statistic describing spatial pattern of trees. The study was carried out with simulated high spatial resolution imagery. Their conclusion was that it is possible to estimate the spatial pattern of trees with the employed method if crown size is provided *a priori*. Uuttera *et al.* (1998) estimated the spatial pattern of trees by segmenting single trees from high resolution aerial photographs (1:5000). Their conclusion was that the usability of aerial photographs seems to be limited because clustered spatial patterns were misclassified as regular patterns, and regular patterns as random patterns. The most obvious way to use ALS data in the determination of the spatial pattern of trees is the individual tree detection (ITD). However, Mustonen (2002) seems to be the only study in which the spatial pattern of trees obtained by ITD has been examined. He reported that the detection of clustered spatial pattern of trees is very difficult with the employed method.

The objective is to investigate if the spatial pattern of trees can be determined with different ALS based forest inventory techniques. Some methodological development was done in order to obtain tree positions or direct estimates of the spatial pattern of trees. The spatial pattern of trees may be estimated in addition to conventional tree or plot level attributes used in forest inventory.

2. Method

2.1 Material and study area

The study area is a typical boreal managed forest area in Kiihtelysvaara in eastern Finland. The field measurements were carried out on May-June, 2010. Altogether 79 field plots were placed subjectively, attempting to record the species and size variation over the area. Sample plot size varied between 20x20 and 30x30 meters according to stand development class. Scots pine (*Pinus sylvestris* L.) is the dominant tree species. It represents 73% of the volume, Norway spruce (*Picea abies* [L.] Karst.) 16% of the volume and deciduous trees altogether 11% of the volume. The characteristics of stand attributes at the plot level are presented in Table 1.

Table 1: Mean, Standard Deviation (SD), Minimum and Maximum of stand attributes at the plot level.

	Mean	SD	Min	Max
DBH	15.0	4.0	8.1	28.4
Height	14.4	3.3	8.7	24.1
Stem number	1259	566	467	2875
Volume	197.6	74.6	79.5	502.2

The high resolution ALS data were employed in the mapping of the trees. First, a tree map was produced using the individual tree detection method described in section 2.3. The tree locations were verified in the field and the undetected trees were positioned using angle and distance observations to the ALS-detected trees. The coordinates for the small trees were then calculated using these observations in a least squares adjustment as explained by Korpela *et al.* (2007). All trees with either DBH \geq 4 cm or height \geq 4 m were mapped.

High resolution ALS data were collected on June 26, 2009 using an Optech ALTM Gemini laser scanning system. The test site was measured from an altitude of approximately 600 m above ground level using a field of view of 26 degrees and side overlap of 55%. Pulse repetition frequency was set to 125 kHz. This resulted in a swath width of approximately 320 m. Side

overlap of 55% means that each location is covered from two flight lines. This kind of special configuration was used in order to maximize the probability that trees have ALS hits each side, i.e. that there are no shadowed areas behind trees along the line from the laser scanner to a tree. A nominal sampling density is about 12 measurements per square meter.

Low resolution ALS data were collected on July 18, 2009 using the same Optech ALTM Gemini instrument as high resolution data. The test site was measured from an altitude of approximately 2000 m above ground level using a field of view of 30 degrees and side overlap of 20%. Pulse repetition frequency was set to 50 kHz. This resulted in a swath width of approximately 1050 m and a nominal sampling density of about 0.65 measurements per square meter.

2.2 Determining the spatial pattern of trees

The spatial pattern of tree locations was determined by plots in three classes: regular, random and clustered. The spatial point pattern of a plot was decided by using Ripley's L function, which is a square root transformation of the Ripley's K function. The L-functions were estimated using Lest-function in the R package spatstat (Badeley and Turner 2005). An isotropic edge correction method was used in all estimations. The interpretation of the L-function is based on the comparison of the L-function estimated from the data (in our case the tree locations ($L_{est_{trees}}$)) and L-function of a random xy-locations (L_{random}). If $L_{est_{trees}} < L_{random}$ the spatial point pattern is regular, if $L_{est_{trees}} > L_{random}$ the spatial point pattern is clustered, and if $L_{est_{trees}} = L_{random}$ the spatial point pattern is random. $L_{est_{trees}}$ practically never equals to L_{random} , and there is no general statistical test for interpreting the difference between $L_{est_{trees}}$ and L_{random} . Thus, Monte Carlo method was employed in an interpretation of the functions using the following process: L-function was estimated from random xy-pairs generated for the plot from a uniform distribution ($L_{est_{random}}$). The number of generated xy-pairs was always the same as the number of trees in the plot and the simulation was repeated 1 000 times for each plot. The simulation data was then used to estimate 70% lower and upper confidence levels ($L_{est_{lowerx}}$, $L_{est_{upperx}}$) for the random point pattern for each plot separately. For example, the lower 70% confidence level is the 15th percentile and the upper 70% confidence level is the 85th percentile of the distribution. Then the $L_{est_{trees}}$ were compared to the estimated confidence levels. If $L_{est_{trees}} < L_{est_{lowerx}}$ the plot was classified as regular, if $L_{est_{trees}} > L_{est_{upperx}}$ the plot was classified as clustered, and otherwise the plot was classified as random with the confidence level of 70%. The spatial point pattern was investigated with the search distance (r) of 1–5 meters, and therefore, the final classification was based on determining the areas delineated by $L_{est_{trees}}$, $L_{est_{lowerx}}$ and $L_{est_{upperx}}$ at distances 1-5. The plot was classified to the class that holds the largest area. For instance, if the area delineated by $L_{est_{trees}}$ and $L_{est_{upperx}}$ above $L_{est_{upperx}}$ is larger than the area delineated by $L_{est_{trees}}$, $L_{est_{lowerx}}$ and $L_{est_{upperx}}$, and $L_{est_{trees}}$ and $L_{est_{lowerx}}$ below $L_{est_{lowerx}}$ the plot is classified as clustered. The spatial pattern was regular in 26 plots, random in 38 plots and clustered in 15 plots.

This approach was used to determine the spatial pattern of trees in the field data and in both of the individual tree detection methods (ITD and Semi-ITD).

2.3 Individual tree detection (ITD)

In ITD the high resolution ALS data was used. First a preliminary Canopy Height Model (CHM) was interpolated using a pixel size of 50 cm by taking the maximum ALS point height within a pixel. The pixels that had no ALS hits within their area were marked with a NoData label. The numbers of NoData pixels and pixels considered as hole pixels were then reduced with a median filtering in local windows of 3 by 3 pixels. Each NoData pixel that had at least n height values within its 8-neighborhood was replaced with the median of the height values of the neighbours. Further, a pixel was considered to be a hole pixel, if at least seven of the eight-neighbours exceeded the height value of the center pixel by more than five meters. The hole pixels were also

replaced with the median of the values of the neighbour pixels exceeding the five meters threshold.

Trees were located on the final CHM using watershed segmentation. Before segmentation, the CHM was low-pass filtered with height based filtering (Pitkänen et al. 2004), in which the scale of filtering increases along with the height of the pixel being processed. The aim is that after filtering each tree top would contain one local maximum on the CHM. In practice we selected eight height ranges and used Gaussian filters with different scale parameters for those. Values of the scale parameter σ for smallest and largest trees were selected based on visual comparison by trying a few scale values and then looking at the local maxima of the filtered CHMs on top of the CHM. The first height range was 0–4 m and the selected σ was 0.3. The last height range was over 28.6 m and σ was 1.0. Other six height ranges had equal width of 4.1 m and their σ values were scaled linearly between 0.3 and 1.0.

A height filtered CHM was then used in watershed segmentation, in which watershed regions associated with the local minima in the negative image of the filtered CHM were identified using a drainage direction following algorithm (Gauch 1999, see also Narendra and Goldberg 1980). In this variant of watershed segmentation, each local minimum is first labeled with unique region identifier. For each of the remaining pixels, the eight-neighbours of a pixel are examined to link the pixel to the steepest downhill direction. The links are then followed to a minimum to get a region label for each pixel. The segmentation divides the whole image area to segments. To get boundaries between tree crowns and background, pixels lower than two meters on the height filtered CHM were masked out from the crown segments. Finally small segments, at most three pixels in size, were combined to one of the neighbour segments, being it a tree crown or background, based on the smallest average gradient on the segment boundary between two segments.

In ITD tree locations were obtained from the location of the pixel having highest value within each segment.

2.4 Semi individual tree detection (Semi-ITD)

In the Semi-ITD the idea is that instead of assuming that one segment corresponds to one tree from 0 to n trees are imputed to each segment (Breidenbach et al. 2010). Therefore it does not implicitly produce tree coordinates. In this study the field measured trees were intersected with the segments obtained with the ITD (2.3) and for each tree a relative position (angle & distance) was calculated with respect to the segment's height maximum. In the prediction phase the tree locations were defined using the angle and distance relative to the height maximum of target segment.

The nearest neighbors were searched for using the Random Forest (RF) method implemented in the R package *yaImpute* (Crookston & Finley 2008) with segments' total volume and stem number as response variables. A large number of features from the first-return ALS data were tested, discarding those of low value according to the RF variable importance score. The final combination included 120 ALS features, containing the percentiles and densities calculated at the both segment and plot level, intensity values at 0-40% height down from the treetop, and volume and area of the point clouds below 10, 20, ..., 100% height. The volume and area were derived from the convex hull of the ALS data, calculated in either 2D (area) or 3D (volume); otherwise, the computation of these features is explained in detail by Vauhkonen et al. (2010). The imputation was carried out in a leave-out-one-plot fashion, i.e. the segments belonging to the same plot as target segment were not available as nearest neighbors.

2.5 Area based approach to determine the spatial pattern of trees (AREA)

In AREA the low resolution ALS data was employed. First a CHM was interpolated using a pixel size of one meter by taking the maximum point height within a radius of 1.6 meters from a centre of a pixel. Remaining NoData pixels were filled with the median of the height values in the 8-neighborhood. Filling was repeated recursively until all NoData pixels were filled. Then the CHM was thresholded to represent lower and upper canopy. Lower canopy contains also ground pixels. After preliminary tests thresholding was carried out by plots using an adaptive threshold value:

$$t = h_{\max} - h_{\text{std}}, \quad (1)$$

where h_{\max} is the maximum height in a plot and h_{std} is the standard deviation of first echoes in a plot. The mean threshold value was about 70% of the maximum height in a plot. Metrics calculated from the thresholded image were used as candidate predictors in the classification of the spatial pattern of trees. Metrics are listed in Table 2. These metrics describe the patchiness, the properties of patches and their interconnections. Patch is a spatial concept that describes a continuous area of the same class; here patch is comparable to segment created by thresholding. Patch metrics are common in landscape ecology in which landscapes are analyzed as mosaics of discrete patches (Turner *et al.* 2001). In Table 2 P refers to all patches, LP refers to patches representing lower canopy and UP refers to patches representing upper canopy.

Table 2: Metrics used in the classification of the spatial pattern of trees in AREA.

Metric	Description
$P_{\text{density}}, LP_{\text{density}}, UP_{\text{density}}$	the number of patches per hectare; i.e. patch density
$P_{\text{ave}}, LP_{\text{ave}}, UP_{\text{ave}}$	the average size of the patches
$P_{\text{sd}}, LP_{\text{sd}}, UP_{\text{sd}}$	the standard deviations of the areas of patches
LP_N, UP_N	the average number of pixels in 4-neighborhood belonging to the same patch type

In AREA the spatial pattern of trees was predicted by Linear Discriminant Analysis (LDA) using the field determined spatial pattern as response (Venables and Ripley 2002). Prediction was carried out by Leave-one-out Cross-validation (LOOCV) in order to avoid too positive results. In LOOCV LDA parameters were repeatedly estimated by ignoring the observation for which the prediction is done. The predictor variables to the LDA model were selected by testing all the combinations from 1 to 3 variables. Adding more than 3 predictors did not improve the accuracy any more.

3. Result

Figure 1 presents tree locations obtained by the ITD, Semi-ITD and the ground truth in a clustered and regular plot. Tree locations are overlaid on the top of CHM and ground truth trees are depicted by circles according to the stem diameter. In the regular plot both ITD and Semi-ITD produced somewhat similar map of trees but in the clustered plot predictions differed considerably. The same tendency is observable in most of the plots. The Semi-ITD produced 5188 trees altogether which is considerably higher than 3165 trees obtained by the ITD. However, the true number of trees is 5827 thus even the Semi-ITD underestimated stem number. Visual examination of tree locations produced by the Semi-ITD reveals that in clustered plots predicted tree locations do not correspond very well with the ground truth. However, in terms of the spatial pattern of trees the Semi-ITD produced much more realistic map of trees than the ITD.

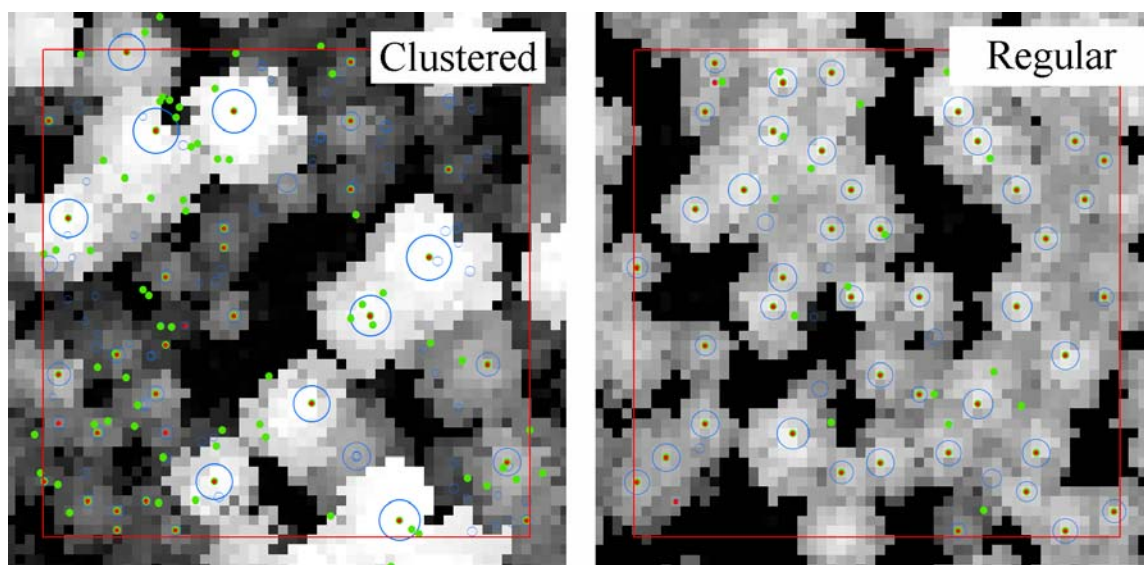


Figure 1: CHM overlaid with tree locations. Blue circles \circ denote ground truth (radius according to stem diameter), red dots \bullet denote ITD trees and green dots \bullet denote Semi-ITD trees. The spatial pattern of trees is clustered in the left plot and regular in the right plot.

The Overall Accuracy (OA) and Cohen's Kappa (κ) for the predictions of the spatial pattern of trees are presented in Table 3. The AREA and Semi-ITD methods were clearly more accurate than the ITD. The κ value of ITD was almost zero, which indicates virtually random classification. In terms of OA the Semi-ITD and AREA were almost equal but in terms of κ Semi-ITD was somewhat better. However, the obtained κ values 0.24 and 0.34 for the AREA and Semi-ITD methods cannot be considered to be very good.

Table 3: The Overall Accuracy (%) and Cohen's Kappa for the predictions of the spatial pattern of trees for the ITD, Semi-ITD and AREA methods.

	Overall Accuracy (%)	Cohen's Kappa
ITD	38.0	0.05
Semi-ITD	57.0	0.34
AREA	55.7	0.24

The error matrices for the predictions of the spatial pattern of trees are presented in Table 4. The ITD did not detect any clustered plot and also most random plots were classified as regular. Thus the ITD tends to coerce random and clustered patterns to regular. The Semi-ITD method detected all the clustered plots correctly but also 9 random plots were misclassified as clustered. Yet it seems to be impossible to discriminate regular and random plot by the Semi-ITD. By the AREA method most random plots were classified correctly whereas most regular and clustered plots were misclassified.

Table 4: The error matrices of the spatial pattern of trees for the ITD, Semi-ITD and AREA methods.

			Predicted			
			Regular	Random	Clustered	Total
ITD	Observed	Regular	26	0	0	26
		Random	34	4	0	38
		Clustered	8	7	0	15
		Total	68	11	0	79
Semi-ITD	Observed	Regular	13	13	0	26
		Random	12	17	9	38
		Clustered	0	0	15	15
		Total	25	30	24	79
AREA	Observed	Regular	10	15	1	26
		Random	7	30	1	38
		Clustered	3	8	4	15
		Total	20	53	6	79

4. Discussion

The spatial pattern of trees can be obtained in ALS based forest inventories several ways. In the ITD tree coordinates are an essential part of the inventory procedure and the spatial pattern of trees can be calculated from the tree coordinates. The Semi-ITD does not implicitly produce tree coordinates. Therefore we extended original Semi-ITD by specifying how tree coordinates can be imputed too. In the area based method tree coordinates are not produced as a part of the inventory procedure but the spatial pattern of trees must be predicted separately at the plot, cell or stand level. Here the spatial pattern was predicted by patch metrics common in landscape ecology. Initial tests indicated that patch metrics are better predictors of the spatial pattern of trees than e.g. traditional height and density metrics used in the area based method.

It is difficult to make a fair comparison of different methods in terms of the spatial pattern of trees because in the AREA method a classification model requiring time consuming field measurements is needed and in the ITD and Semi-ITD methods tree coordinates (i.e. the spatial pattern) are obtained as a part of the inventory procedure. It must also be taken into account that in many use cases the information about the spatial pattern of trees is used jointly with primary attributes (basal area, stem volume, etc.) and that these must be predicted by the same approach. Here the accuracy of primary attributes was ignored entirely although it varies among ITD, Semi-ITD and AREA.

The Semi-ITD provided the most accurate prediction in terms of the spatial pattern of trees. The AREA method was almost as accurate as the Semi-ITD. However, the predictions of the AREA and Semi-ITD methods differed considerably which can be seen in error matrices in Table 4. The Semi-ITD was able to detect clustered patterns quite well although some commission was introduced. The ITD provided the least accurate classification. It tends to force everything toward the regular pattern. This is a logical outcome because an inherent property of the ITD is to detect only those trees which are clearly visible from the above. Results are also consistent with Mustonen (1992) although there are considerable differences in the forest structures of the study areas.

In this study spatial pattern information was analyzed similarly for all the plots of the study area. In the real use case the importance of the information on spatial pattern is different in different stand development stages. In seedling stands the recognition of single trees is usually not possible but (area based) spatial information could still be used, for example, to determine the success of regeneration. In young stands prior to first thinning it is especially important to recognize clustered stands since the timing of the silvicultural operation is considerably different in those stands. In older stands the most important use of spatial pattern information is in spatial growth models. This information can be in the form of tree coordinates in distance dependent growth models, competition indices or just some spatial variable in the model. In multilayered forests, such as clustered plot in Fig.1 the detection of spatial information is, once again, very difficult but on the other hand there is no primary forest management use for such information. All in all, determining the spatial pattern of trees by ALS is somewhat unexplored field of study. Therefore it should be studied how well the spatial pattern of trees can be determined in different type of forests.

References

- Badeley, A. and Turner, R., 2005. Spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software* 12(6). 42 p. Available at: <http://www.jstatsoft.org/v12/i06>.
- Breidenbach, J., Næsset, E., Lien, V., Gobakken, T. and Solberg, S., 2010. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment*, 114, 911-924.
- Coops, N. and Culvedor, D., 2000. Utilizing local variance of simulated high spatial resolution imagery to predict spatial pattern of forest stands. *Remote Sensing of Environment*, 71, 248-260.
- Crookston, N.L. and Finley, A.O., 2008. yaImpute: An R package for kNN imputation. *Journal of Statistical Software* 23(10). 16 p. Available at: <http://www.jstatsoft.org/v23/i10>.
- Diggle, P.J., 1983. *Statistic analysis of spatial point patterns*. Mathematics in biology. Academic press, 148 s.
- Gauch, J.M., 1999. Image segmentation and analysis via multiscale gradient watershed hierarchies. *IEEE Transactions on Image Processing*, 8, 69-79.
- Korpela, I., Tuomola, T. and Välimäki, E., 2007. Mapping forest plots: An efficient method combining photogrammetry and field triangulation. *Silva Fennica*, 41(3), 457-469.
- Mustonen, K., 2002. Puiden tilajärjestyksen ja pituuden määrittäminen laserkeilankuvilta. Metsäsuunnittelun ja -ekonomian pro-gradu. University of Joensuu, 55 p. (In Finnish.)
- Narendra, P.M. and Goldberg, M., 1980. Image segmentation with directed trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2, 185-191.
- Penttinen, A., Stoyan, D. and Henttonen, H.M., 1992. Marked Point Processes in Forest Statistics. *Forest Science*, 38, 806-824.
- Pielou, E.C., 1960. A single mechanism to account for regular, random and aggregated populations. *Journal of Ecology*, 49, 574-584.
- Pitkänen, J., Maltamo, M., Hyypä, J. and Yu, X., 2004. Adaptive methods for individual tree detection on airborne laser based canopy height model. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI, part 8/W2, 187-191.
- Pukkala, T., 1990. A method for incorporating the within-stand variation into forest management planning. *Scandinavian Journal of Forest Research*, 5, 263-275.

- Tomppo, E., 1986. Models and Methods for analysing spatial pattern of trees. *Communicationes Institute Forestalis Fenniae* 138. 65 p.
- Turner, M.G., Gardner, R.H. and O'Neill, R.V., 2001. Landscape ecology in theory and practice. Springer, New York. 401 p.
- Uuttera, J., Haara, A., Tokola, T. and Maltamo, M., 1998. Determination of the spatial distribution of trees from digital aerial photographs. *Forest Ecology and Management*, 110, 257-282.
- Vauhkonen, J., Korpela, I., Maltamo, M. and Tokola, T., 2010. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment*, 114, 1263-1276.
- Venables, W.N., and Ripley, B.D., 2002. Modern Applied Statistics with S. Fourth edition. Springer. 495 p

Fusion of Airborne LiDAR and WorldView-2 MS Data for Classification of Depth to Permafrost within Canada's Sub-Arctic

L. Chasmer¹, C. Hopkinson², H. Morrison², R. Petrone¹, & W. Quinton¹

¹Cold Regions Research Centre, Wilfrid Laurier University, Waterloo ON N2L 3C5
lechasme@yahoo.ca

²Applied Geomatics Research Group, NSCC, Lawrencetown NS B0S 1M0

Abstract

The discontinuous permafrost zone of north-western Canada is characterised by a heterogeneous landscape of tree-covered permafrost plateaus that rise 0.5 m to 2.0 m above the surrounding fens and bogs. The depth to permafrost or “frost table” is influenced to some extent by vegetation canopy cover, which drives complex feedbacks related to permafrost thaw. Spectral remote sensing offers the possibility of large area mapping of canopy and ground surface characteristics that may be used as a proxy for permafrost thaw within remote northern areas. However, this depends on whether or not spectral bands can be used to identify slight variations in vegetation characteristics. The following study compares vegetation and topographic characteristics obtained using airborne Light Detection And Ranging (LiDAR) with high spatial resolution WorldView-2 spectral bands and *in situ* transect measurements of the depth to frost table. The results of this study indicate that the depth to the frost table is related to above ground vegetation cover and tree height, yet relationships are complicated by canopy and understory characteristics, topographic derivatives, and the position of the measured frost-table transect within the fragmented plateau. Comparisons between vegetation structural characteristics and WorldView-2 spectral bands are also examined so that confidence can be applied to depth of frost table estimates from WorldView-2 based on canopy characteristics. Discrete WorldView-2 pixels are related to depth to frost table (bands red, near infrared 1,2) and canopy metrics/topography obtained from airborne LiDAR. Variability is due, in part to absorption of near infrared by shadow fractions observed within WorldView-2 pixels, and spectral reflectance of ground vegetation visible within mixed pixels. High resolution spectral imagery, therefore, provides a link to processes controlling spatial variability of the depth to frost table.

Key Words: permafrost, discontinuous permafrost zone, multi-spectral remote sensing, WorldView-2, airborne LiDAR, vegetation structure.

1. Introduction

The southern boundary of discontinuous permafrost found in north-western Canada is highly sensitive to increases in air temperature via global warming in this region. Several studies have shown unprecedented permafrost thaw and the conversion of permafrost plateaus into bog and fen land cover types using historical aerial photography (e.g. Shur and Jorgenson, 2007; Quinton *et al.* 2010; Chasmer *et al.* 2010, 2011a). The conversion of plateaus into permafrost-free peatlands (ombrotrophic flat bogs and channel fens) strongly affects the hydrological nature of the watershed. Permafrost controls water storage and drainage processes by obstructing and re-directing the movement of water through channel fens and bogs (Quinton *et al.* 2010). The depth to permafrost, herein referred to as the “frost table”, is influenced to some extent by radiation loading (Tarnocai *et al.*, 2004; Chasmer *et al.* 2011a), whereby up to 40% of the variance in frost table depth can be explained by the variation in canopy openness (Hopkinson *et al. in review*). Localised increases in shortwave radiation cause preferential thaw

of the active (or seasonally thawed) layer, which may result in a depression within the permafrost. Water drains towards the depression, resulting in a local increase in soil moisture and thermal conductivity (Hayashi *et al.* 2007).

Spatial variability of the existence of and depth to the plateau frost table has important implications for northern development and resources extraction, hydrology, and carbon and methane cycles. Given that permafrost underlies approximately 50% of the Canadian landmass, and that peatlands, which characterise a large proportion of northern ecosystems, contain approximately 1/3 of the global soil C pool (Gorham, 1991), large area mapping and assessment of proxies of permafrost change (e.g. plateau forest cover) is important. The following study examines relationships between the depth to frost table, vegetation and topographical characteristics measured using airborne Light Detection And Ranging (LiDAR), and spectral bands obtained from high spatial resolution WorldView-2 imagery across a permafrost plateau located within a typical watershed found in the Canadian discontinuous permafrost zone. The results of this study have implications for scaling and assessment of permafrost dynamics over broader areas using lower resolution multi-spectral satellite imagery.

2. Methods

2.1 Study area

The site is located approximately 50 kms south of Fort Simpson, within the Scotty Creek watershed (61° 18'N, 121° 18'W), Northwest Territories, Canada (Figure 1). The watershed is typical of the continental high boreal discontinuous permafrost/peatland region (National Wetlands Working Group, 1997). A supervised classification of IKONOS imagery (2000) (Quinton, *et al.* 2003) showed that approximately 43% of the watershed consisted of permafrost plateaus, ombrotrophic flat bogs (24%), channel fens (21%), lakes (9%), and isolated bogs (4%). Plateau extents are diminishing at an alarming rate (14% aerial reduction since 1970); with 73% of that area converting into bog as opposed to fen land cover types (Quinton *et al.* 2010; Kenward *et al. in review*). Permafrost plateaus support *Picea mariana* Mill. (black spruce) and a variety of shrubs and lichens. The mean annual temperature measured at Fort Simpson is -3°C, while mean annual precipitation is 369 mm (1971 to 2000).

The depth to frost table was measured on August 2nd, 2010 along four ground-survey transects traversing north to south (one transect) and east to west (three transects) across a typical permafrost plateau (Figure 1). Depth to frost table was measured using a graduated steel measurement rod. The rod was pushed into the peat until it encountered frozen ground and could not be pushed in any further. Markings on the rod were used to determine the depth to frost table. Every measurement of frost table depth was geographically located using survey-grade differential GPS measurements (> 10 cm accuracy).

2.2 LiDAR data collection and processing

Airborne LiDAR data were collected by the authors (Applied Geomatics Research Group) for the entire Scotty Creek watershed on August 2nd, 2010 using an Optech Inc. ALTM 3100 four-pulse return system. The survey was parameterised for an averaging flying height of 1500 m a.g.l., with 50 kHz pulse repetition frequency, and a scan angle of ±20°. Returns were classified into ground and non-ground (vegetation) within TerraScan software (Terrasolid, Finland) (Axelsson, 2000). Ground returns were used to create a 2 m resolution digital elevation model (DEM) using an inverse distance weighting approach. A digital surface model (DSM) was created based on the maximum laser pulse return (localised maxima) within a 2 m search radius. Canopy height was determined by subtracting the DSM from the DEM. Finally, gap

fraction was estimated as the ratio of the sum of all laser pulse returns between -1 m and + 0.5 m of the ground surface divided by the sum of all returns within a 1 m search radius.

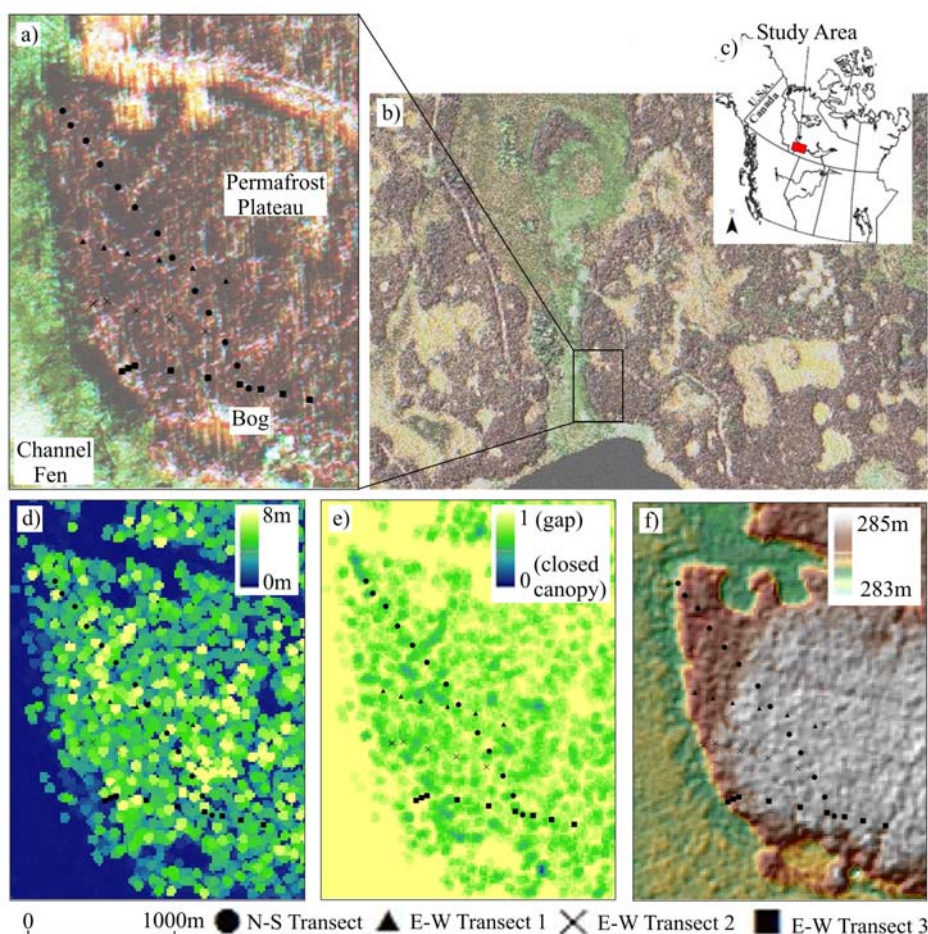


Figure 1: Depth to frost table measurement locations along four transects traversing the western edge of a permafrost plateau displayed as: a) WorldView-2 pan sharpened false colour composite of permafrost plateau, where red channel is WorldView band 6 (red edge), green channel is band 5 (red), and blue channel is band 3 (green); b) sub-area of watershed with LiDAR and WorldView-2 data collections. False colour composite in ‘a’ shown. Green areas = highly productive fen, red areas = treed permafrost plateaus, yellow = less productive bog; c) location of the study site in the Northwest Territories, Canada; LiDAR derived: d) canopy height model; e) gap fraction model; f) digital elevation model.

2.3 WorldView-2 image collection and orthorectification

Space-borne WorldView-2 (DigitalGlobe Corp.) was tasked for imaging of the Scotty Creek watershed between September 17 and October 1, 2010 (resulting in two cloud-free images) via MDA Corp. Canada. WorldView-2 data were obtained as a standard, radiometrically corrected ortho-ready bundle including a 0.65 m panchromatic image (400-900 nm) and eight narrow spectral bands: coastal (400-450 nm), blue (450-510 nm), green (510-580 nm), yellow (585-625 nm), red (630-690 nm), red edge (705-745 nm), near infrared 1 (NIR1, 770-895 nm), and near infrared 2 (NIR2, 860-1040 nm) at 2 m pixel resolution. The look angle was 20° maximum off nadir. Data were processed to UTM coordinates matching that of the LiDAR data (NAD83 CSRS). The watershed was divided into 2 km x 2 km tiles and imported into Geomatica OrthoEngine (PCI Inc. Canada) using the WorldView-2 mosaic and colour-balancing tool. The colour-balanced, mosaiced imagery was then orthorectified using the LiDAR DEM, 2 m

resolution gridded ground surface laser return intensity and 23 tie points located within both datasets at the intersections of trails and seismic lines. The images were orthorectified to better than 1 m accuracy using a second order polynomial transformation in OrthoEngine. Table 2 provides summary statistics of WorldView-2 spectra per band along transects. Pixel digital numbers (DN) were then converted to top-of-atmosphere spectral radiance ($Wm^{-2} sr^{-1} \mu m^{-1}$) based on absolute radiometric calibration factors ($Wm^{-2} sr^{-1} count^{-1}$) provided per band to get a band-integrated radiance ($Wm^{-2} sr^{-1}$). This is then divided by effective bandwidth to get spectral radiance.

2.3 Statistical analysis

LiDAR vegetation (CHM, gap fraction) and elevation (DEM) metrics, and WorldView-2 spectral bands were compared with depth to frost table measurements within 1 m and 2 m radii of the measurement location (pixel/cell averages, maximum, minimum, and standard deviations presented) (Morrison, et al. *this issue of SilviLaser*). Only depth to frost table measurements located on the plateau (as opposed to plateau edges) are described. Processes associated with permafrost thaw at the edge of plateaus can differ from those within plateaus, and therefore, measurements at the edge of plateaus were excluded from this analysis. (A more comprehensive assessment of the LiDAR and frost table depth spatial correlations is presented in Hopkinson *et al. in review*). The numbers of measurements included along transects were: 13 (N-S transect), 6 (E-W transect 1), 6 (E-W transect 2) and 9 (E-W transect 3) for a total of 34 measurements. Comparisons were made using linear and non-linear (where required) regression and tests of significance.

3. Results and Discussion

3.1 General observations along transects

The results of this study provide some promise as to the use of WorldView-2 spectral imagery for detecting i) areas underlain by permafrost; and ii) the relative depth to permafrost as a function of canopy cover. In many cases, airborne LiDAR data may not be needed. This could extend the application of this study to other remote northern environments not surveyed by LiDAR.

Measurements along transects that have tallest trees, fewer within canopy gaps, and slightly upraised elevation are characterised by frost table depths not far below the ground surface (on average) (Table 1, characteristic transects in *italics*). Along one transect, however (E-W Transect 3), close proximity to the southern edge of the plateau, shorter trees, and greater within canopy gaps may have contributed to increased thaw down of the frost table. This is especially evident along parts of the transect that traversed the plateau near an area of plateau collapse and an overland flow channel (Table 1, normal text).

Table 1: Depth to frost table, vegetation and ground surface elevation summary statistics (average and standard deviation (σ)) for transects. Statistics are based on LiDAR metric extractions within 1m of frost table depth measurement locations.

Transect	Average (σ) depth to frost table (m)	Average (σ) canopy height (m)	Average (σ) gap fraction	Average (σ) elevation (m)
<i>N-S Transect</i>	-0.47 (0.08)	4.9 (0.9)	0.75 (0.11)	284.5 (0.32)
<i>E-W Transect 1</i>	-0.52 (0.07)	5.0 (1.3)	0.81 (0.10)	284.6 (0.32)
<i>E-W Transect 2</i>	-0.50 (0.06)	4.5 (0.95)	0.79 (0.10)	284.5 (0.40)
E-W Transect 3	-0.56 (0.09)	3.2 (1.5)	0.87 (0.12)	284.5 (0.30)

Comparisons of average (and σ) band radiance between transects illustrate some variability associated with canopy, understory and ground surface characteristics. Optically bright reindeer lichen (*Cladina mitis*; *Cladina rangiferina*), and senescing short deciduous shrubs (*Oxycoccus microcarpus*; *Ledum groenlandicum*) found along Transects 1 and 3 (E-W) have contributed to increased reflectance in green, red edge, NIR1 and NIR2 bands. Reduced spacing between trees and increased within canopy shadows along transects N-S and E-W 2 may have resulted in reduced reflectance in NIR1 and 2 bands. Are variations in spectral radiance a proxy indicator of frost table depth between transects?

Table 2: WorldView-2 spectral radiance ($\text{Wm}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$) summary statistics within 1 m radius of depth to frost table measurements per transect.

Transect	Ave. (σ) WV2 B1 (coast)	Ave. (σ) WV2 B2 (blue)	Ave. (σ) WV2 B3 (green)	Ave. (σ) WV2 B4 (yellow)	Ave. (σ) WV2 B5 (red)	Ave. (σ) WV2 B6 (red edge)	Ave. (σ) WV2 B7 (NIR1)	Ave. (σ) WV2 B8 (NIR2)
N-S	32.11 (6.31)	27.96 (3.75)	17.74 (1.59)	12.19 (1.66)	9.18 (1.26)	13.88 (1.60)	14.41 (1.27)	10.11 (0.92)
E-W	32.30 (6.35)	28.18 (3.78)	9.17 (0.82)	13.12 (1.79)	9.46 (1.30)	15.65 (1.81)	16.30 (1.44)	12.25 (1.11)
Transect 1	32.09 (6.31)	27.66 (3.71)	17.35 (1.56)	12.29 (1.68)	8.95 (1.23)	13.65 (1.58)	13.50 (1.19)	11.10 (1.01)
Transect 2	32.64 (6.41)	28.62 (3.84)	18.66 (1.67)	13.57 (1.85)	10.41 (1.43)	16.62 (1.92)	16.38 (1.45)	12.61 (1.14)
Transect 3								

3.2 Correspondence between WorldView-2 bands and variability in frost table depth

Variability in frost table depth across all transects did not correspond well with radiance from WorldView-2 spectral bands. At best, increased reflectance in red wavelengths corresponds significantly, but not strongly with increased depth to permafrost ($r^2 = 0.09$, $p < 0.001$, RMSE = 0.07 m for red band). Thus, WorldView-2 spectral radiance is confounded by within and below canopy averaging within pixels and variability when examined across transects. Mixed pixels caused by narrow black spruce canopies and leaning trees (as permafrost thaws), poorly drained and dry upraised peat soils, and groups of shrubs and reindeer lichen alter within pixel spectral reflectance. The location of transects on the plateau surface, and feedbacks between soil moisture (at the edge of plateaus), and fragmentation also confound relationships. Yet, within transects, some interesting relationships between vegetation/ground surface structure, spectral reflectance, and depth to frost table can be found.

3.3 Comparisons between airborne LiDAR metrics and WorldView-2 along transects

Within transects, airborne LiDAR metrics of canopy height, elevation, and to a lesser extent gap fraction correspond significantly with WorldView-2 spectral bands. However, coefficients of variation (r^2) often explain up to only one third of the variability, indicating some complexity remains unexplained:

1. Mean canopy heights within 2 m of individual frost table depth measurements are negatively related to mean reflectance in the red band ($r^2 = 0.34$, $p < 0.001$, RMSE = 1.44 m). Red reflectance decreases with greater pixel fractions containing trees (often positively related to canopy cover ($r^2=0.19$)).
2. Variability of maximum canopy gaps within 1 m radius of frost table depth measurements are positively related to increased variability in NIR2 band ($r^2 = 0.21$, $p < 0.001$, RMSE = 9%) due to reduced within canopy shadowing.
3. Variations in elevation within a 2 m radius of frost table depth measurements are

explained by between pixel standard deviation in the red band, where increased variability often results in increased red reflectance ($r^2 = 0.21$, $p < 0.001$, RMSE = 0.09 m).

Within individual transects, spectral radiance may be used to estimate variability in the depth to frost table, depending on vegetation and ground surface characteristics defined by LiDAR. Taller trees found along N-S Transect are associated with increased reflectance in the green band ($r^2 = 0.52$, $p < 0.01$), increased ground surface shadowing, and reduced depth to frost table (Table 3). Along E-W Transect 1, green reflectance from the ground, as opposed to that from widely spaced trees may be linked to increased tree bole illumination and sensible heat inputs from tree boles into the ground surface (thereby increasing permafrost thaw; Chasmer *et al.* 2011; Hopkinson *et al. in review*). Similar results were also found along E-W Transect 2, which was characterised by some ground surface shadowing and absorption in NIR bands. Spectral reflectance within red edge wavelengths was indicative of senescing black spruce trees, successive deciduous shrub development, and tall grasses often found at waterlogged plateau edges and within areas of plateau slump (Table 3).

Table 3: Strongest relationships between depth to frost table and WorldView-2 bands within transects.

Transect	WorldView band	r^2 (p)	RMSE (m)	Distance from depth to frost table measurement
N-S Transect	Green	0.18 (0.008)	0.06	1 m, minimum radiance
E-W Transect 1	Green	0.94 (<0.001)	0.02	1m, minimum radiance
E-W Transect 2	NIR2	0.81 (<0.001)	0.03	2m, minimum radiance
E-W Transect 3	Red Edge	0.71(<0.001)	0.05	2m, σ

4. Conclusions

This study compares measurements of the depth to frost table with high resolution multispectral WorldView-2 imagery and vegetation structural/topographic metrics derived from airborne LiDAR data. The purpose was to determine if structural/topographic metrics that influence the spatial variability of the depth to frost table can be examined using spectral remote sensing imagery at the end of the growing season. This study has shown that variability in depth of frost table can be estimated using green and NIR bands, however, the strength of the relationships vary with spatial variability of local canopy height, spacing between vegetation, shadows, and proximity to plateau edge (within up to 2 m from the frost table depth measurement). WorldView-2 spectral bands, red and NIR2 can also be used to estimate canopy height, canopy gaps, and elevation.

Acknowledgements

Funding for this project was provided by an NSERC Strategic Grant provided to Dr. Quinton, and a CFI grant provided to the AGRG LiDAR Laboratory for purchase of the airborne LiDAR and infrastructure. Thank-you also to Tristan Goulden, and Allyson Fox for assistance with LiDAR data collection, processing, and field work. Also thanks to the Liidlii Kue First Nation.

References

Axelsson, P., 2000, DEM generation from laser scanner data using adaptive TIN models, *IAPRS* 2000, 33(B4), 110-117.

- Chasmer, L., Hopkinson, C. and Quinton, W., 2010, Quantifying errors in discontinuous permafrost plateau change from optical data, Northwest Territories, Canada: 1947 to 2008. *Canadian Journal of Remote Sensing*, 36(2):S211-S223.
- Chasmer, L., Quinton, W., C. Hopkinson, R. Petrone, and P. Wittington, 2011a, Vegetation canopy and radiation controls on permafrost plateau evolution within the discontinuous permafrost zone, Northwest Territories, Canada, *Permafrost and Periglacial Processes*, DOI: 10.1002/ppp.724.
- Chasmer, L., Kljun, N., Hopkinson, C., Brown, S., Milne, T., Giroux, K., Barr, A., Devito, K., Creed, I., and Petrone, R. 2011b. Characterizing vegetation structural and topographic characteristics sampled by eddy covariance within two mature aspen stands using LiDAR and a flux footprint model: Scaling to MODIS. *Journal of Geophysical Research-Biogeosciences, Special Issue on Advances in Upscaling of Eddy Covariance Measurements of Carbon and Water Fluxes*, 115, doi:10.1029/2010JG001567.
- Gorham, E. 1991, Northern Peatlands: Role in the Carbon-Cycle and Probable Responses to Climatic Warming, *Ecological Applications*, 1(2),182-195.
- Hopkinson, C., Chasmer, L., W. Quinton, Assessing spatial coincidence between forest canopy and discontinuous permafrost features using LiDAR and thermal image data, *Hydrological Processes (in review)*.
- Hayashi, M., Goeller, N., Quinton, W., and Wright, N., 2007, A simple heat-conduction method for simulating the frost-table depth in hydrological models. *Hydrological Processes*, 21,2610-2622.
- Kenward, A., Chasmer, L., Petrone, R., and W. Quinton, Spatial and temporal variability of CO₂ exchanges along a fen-plateau-bog transect within the discontinuous permafrost zone, Northwest Territories: Implications for landscape change, *Ecosystems (in review)*.
- National Wetlands Working Group, 1997. The Canadian Wetland Classification System, 2nd Ed.
- Quinton, W.L., Hayashi, M., and Pietroniro, A., 2003. Connectivity and storage functions of channel fens and flat bogs in northern basins. *Hydrological Processes*. 17,3665-3684.
- Quinton, W., Hayashi, M., and Chasmer, L., 2010. Permafrost thaw in the Canadian sub-arctic: Some implications for water resources. *Hydrological Processes Scientific Briefing*. DOI: 10.1002/hyp.7894.
- Shur, Y.L., and Jorgenson, M.T. 2007. Patterns of permafrost formation and degradation in relation to climate and ecosystems. *Permafrost and Periglacial Processes*. 18,7-19.
- Tarnocai, C., Nixon, F.M. and Kutny, L. 2004. Circumpolar-Active-Layer-Monitoring (CALM) sites in the Mackenzie Valley, northwestern Canada. *Permafrost and Periglacial Processes*. 15,141-153.

Using high density ALS data in plot level estimation of the defoliation by the Common pine sawfly

Tuula Kantola¹, Päivi Lyytikäinen-Saarenmaa¹, Mikko Vastaranta¹, Ville Kankare¹, Xiaowei Yu², Markus Holopainen¹, Mervi Talvitie¹, Svein Solberg³, Paula Puolakka⁴ & Juha Hyyppä²

¹University of Helsinki, Finland, first.last@helsinki.fi

²Finnish Geodetic Institute, first.last@fgi.fi

³Norwegian Forest and Landscape Institute, Norway, first.last@skogoglandskap.no

⁴Finnish Forest Research Institute, Vantaa, Finland, first.last@metla.fi

Abstract

The climate change has been related to the increase of forest insect damages in the boreal zone. The prediction of the changes in the distribution of insect-caused forest damages has become a topical issue. The common pine sawfly (*Diprion pini* L.) is regarded as a significant threat to boreal Scots pine (*Pinus sylvestris* L.) forests. Efficient and accurate methods are needed for monitoring and predicting changes in insect defoliation. In this study, the field work has been carried out in 2009 in Eastern Finland, where *D. pini* has caused considerable damage in managed Scots pine forests. Altogether 95 sampling plots were used in the analysis. A high density ALS data was acquired simultaneously with the field work. The aim of the present study was to test the accuracy of the plot level needle loss predictions determined from the area based and single tree ALS features separately. The Random Forest method (RF) was utilized in the estimation. The best classification accuracy for the test set was 67.4% (area based features). The best plot level accuracy using the tree-wise features was 60.6%, respectively.

Keywords: ALS, Random Forest, defoliation, *Diprion pini*, forest disturbances

1. Introduction

Evergreen coniferous forests dominate the landscape in Finland, covering about 76 % of the land. Forests have been profoundly altered by human activities and most of the massive old-growth forests have been replaced with younger, even-aged managed forests. Climate change and its effects in Finland may be the most serious environmental issue threatening the health of forests. Average annual temperatures have increased more in northern latitudes than the global average temperatures. Ecological balance of forests has been interrupted, causing wide pest damages in managed forests (Moore and Allard 2008). Substantial changes in patterns of forest disturbance have been observed to cover larger areas than ever before (e.g. Lyytikäinen-Saarenmaa and Tomppo 2002). Outbreaks of defoliating insects have increased sharply in the two past decades (Kantola et al. 2010, Karjalainen et al. 2010).

There is an increasing need to map and monitor the area, estimate severity and detect spatial location of the hazard (Lyytikäinen-Saarenmaa et al. 2008, Karjalainen et al. 2010). The remote sensing (RS) methods have related differences in spectral responses to chlorosis, foliage reddening or foliage loss over time, aiming to interpret, classify or correlate to damage caused by pest insects. RS can produce data for large areas of remote, inaccessible forest lands quickly and with a higher cost-efficiency rate than ground surveys (Hall et al. 2007).

Rapid development of airborne laser scanning (ALS) techniques has provided new perspectives to the forest inventory, as well as forest health survey. With the capability of directly measuring

forest structure, including canopy height and crown dimensions, including changes in foliage mass, ALS is increasingly used for forest inventories at different levels (e.g. Maltamo et al. 2004, van Aardt et al. 2008,). ALS is also a promising method for monitoring forest hazards. There are several recent studies for more accurate ALS-based biomass detection (Sohlberg et al. 2006, 2010, Hawbaker et al. 2009). The accuracies for detection of defoliated individual trees have already been proposed (Kantola et al. 2010). Single tree biomass and defoliation level gained high correlation by means of terrestrial laser scanning (see e.g. Hyyppä et al. 2009).

2. Material and methods

2.1. Study area

We carried out the present research in Ilomantsi, eastern corner of Finland (62°53'N, 30°54'E). Dry or dryish forest site types mainly dominate the 34.5 km² wide study area. The dominance of *Pinus sylvestris* is 99.5%. The majority of stands on the area are young to middle-aged stands, having a mean age of 53 years and mean diameter of 14.7 cm. The initial outbreak of the Common pine sawfly (*Diprion pini* L.) was launched in western cost side of Finland in 1997, and was firstly visible in 1999 on the Palokangas area. Since then, the outbreak range in Palokangas has spatially fluctuated between 10 000 to 15 000 ha during the last 12 years. Population density and damage intensity has been fluctuating between years, showing now a chronic nature. The first spots with dead trees appeared in 2008-09, due to severe defoliation during several years.

2.2 Field data

The field measurements were carried out in May and early June 2009 before elongation of current needles, representing the defoliation status of the fall 2008. Total of 113 mature, maturing or seedling felling sampling plots were measured using adaptive cluster sampling (ACS) as an inventory method (see more details from Talvitie et al. 2011). The visual assessment of defoliation intensity was performed simultaneously with tree-wise measurements.

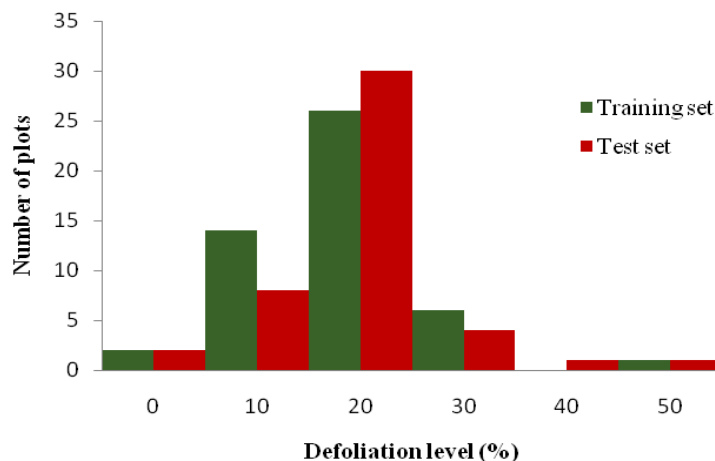


Figure 1. The distribution of the defoliation level of sampling plot in training and test sets.

The sampling plot centres were located with a Trimble Pro XH (Trimble Navigation Ltd., Sunnyvale, CA, U.S.), which can reach up to 30 cm precision. Differential post processing was applied. The individual trees were also located. The defoliation intensity of a single tree was visually assessed from different directions according to Eichhorn (1998). An accuracy of 10 % was used in the assessment of needle loss. The defoliation levels of the plots were calculated as an average of the needle losses of the trees of the two dominant canopy levels.

Altogether 95 sampling plots were used in the analysis. A total of 1377 trees from upper canopy layers were identified. The amount of individual tree detection (ITD) trees per plot varied from 1 to 31. The defoliation levels of the plot varied between 0 and 50 (Figure 1.) and the defoliation levels of the single trees varied between 0 and 100. Most of the trees and plots had 10 to 30% as a level of the needle loss. The data was divided randomly to training and test sets. Training set consisted of 49 plots and 707 trees and the test set of 46 plots and 670 trees.

2.3 Remote sensing material

The ALS data was acquired in October 2008 with a Leica ALS50-II SN058 laser scanner (Leica Geosystem AG, Heerbrugg, Switzerland). The flying altitude was 500 m at a speed of 80 knots, with a field of view of 30 degrees, pulse rate of 150 kHz, scan rate of 52 Hz and size of the laser footprint on the ground of 0.11 m. The density of the returned pulses within the field plots was approximately 20 pulses per m². ALS data were classified into ground or non-ground points using the standard TerraScan approach as explained by Axelsson (2000). A digital terrain model was created using classified ground points. Laser heights above ground (normalized height or canopy height) were calculated by subtracting the ground elevation from corresponding laser measurements. Heights greater than 2 m were considered as vegetation returns, and only these were used for tree feature extraction.

2.4 Plot level feature extraction

In the area based approach the canopy height or vertical distribution of laser returns were used for estimating plot-level defoliation. The laser returns within each plot were extracted from the laser data for the radius of 8 m. Descriptive features were derived individually per plot from the normalized point height for the vegetation points (first and last). The features derived were minimum height (Hmin), maximum height (Hmax), mean height (Hmean) calculated as the arithmetic mean of the laser heights, standard deviation of laser heights (Hstd), coefficient of variation (CV), penetration, percentiles calculated from 10% to 100% of the canopy height distribution at 10% intervals (h10–h100), canopy cover percentiles expressed as proportions of the first returns below a given percentage (10% to 90%) of the total height (p10–p90).

2.5 Individual tree detection and feature extraction

A raster canopy height model (CHM) was created from normalized data for individual tree detection and crown segmentation. Single tree segmentations were performed on the CHM images using a minimum curvature-based region detector (Yu et al. 2010). During the segmentation processes, the tree crown shape and location of individual trees were determined. The procedure consisted of four steps. 1.) The CHM was smoothed with a Gaussian filter to remove small variations on the crown surface. 2.) Minimum curvatures were calculated. 3.) The smoothed CHM image was scaled based on the computed minimum curvature resulting in a smoothed yet contrast-stretched image. 4.) Local maxima were searched for in a given neighbourhood.

Each segment was considered to present a single tree crown. The identification rate was 83 %. Laser returns falling within each individual tree segment were extracted and the canopy heights of these returns were used to derive the tree features. The first returns were used in this study as well as the intensity values of the returns (Table 1). The intensity calibration was achieved by linear regression for modelling the relationship between intensities of two different datasets (Vain et al. 2010) using 3 gravel targets placed in the test area during the campaign.

Table 1. Features extracted from ALS data for individual trees.

Feature	Description
Hmax	Maximum laser height
Hmean	Arithmetic mean of laser heights
Hstd	Standard deviation of heights
CH	Crown height
CA	Crown area as a convex hull
CV	Crown volume as a convex hull in 3D
P10-90	Percentile of canopy height distribution
CCP10-90	Canopy cover percentile as proportion of returns below a given percentage of total height
MaxD	Maximum crown diameter when crown was considered as an ellipse
min	Intensity features
max	"
range	"
std	"
skewness	"
kurtosis	"
COV	"
P10.1-90.1	Percentile of intensity

2.6 Random Forest

The RF algorithm, proposed by Breiman (2001), is a nonparametric estimation approach. The method is composed of a set of regression trees that are constructed from bootstrapped training data. The bootstrapped data consist in general of sets of samples taken randomly with replacement from the original training set. A regression tree is built for each of the bootstrap sets. Random forests are created by averaging over trees. A regression tree is a sequence of rules that split the feature space into partitions having similar values to the response variable. A method based on a classification and regression tree is usually adopted to generate regression trees. At each node of a regression tree, data are split until the leaf nodes contain fewer samples than some preselected value, or the sum of squares of distances to the mean value of the respective group is less than the threshold. RF is described and used for the estimation of tree variables (e.g. in Yu et al. 2010). The R *yaImpute* library (Crookston and Finley 2007-2010) was applied in the RF estimations. Total of 2000 regression trees were fitted in each RF run to gain more consistency.

2.7. Classifications

The defoliation levels of the plots were estimated with RF using two different classifications. The first approach was using two classes (Two classes), having a defoliation level under or 20% or more. The second estimation was made using the 10% accuracy of the needle losses (10% classes). The plot level estimation using the ITD features was made also with the same classifications. The estimation accuracies were tested using the first returns, the last returns and both of them together. The results were studied further in case of the first returns.

3. Results

3.1 Plot level estimation

The best plot level classification accuracy was gained using the first returns with two classes (93.9% for training set and 67.4% for test set) (Table 2.). The best classification accuracy for test set in 10% classification using first returns was 47.8% (79.6% for training set). The

accuracies between different sets of features varied relatively little. The best explanatory features were penetration and different canopy cover percentiles and height features (Figure 2.)

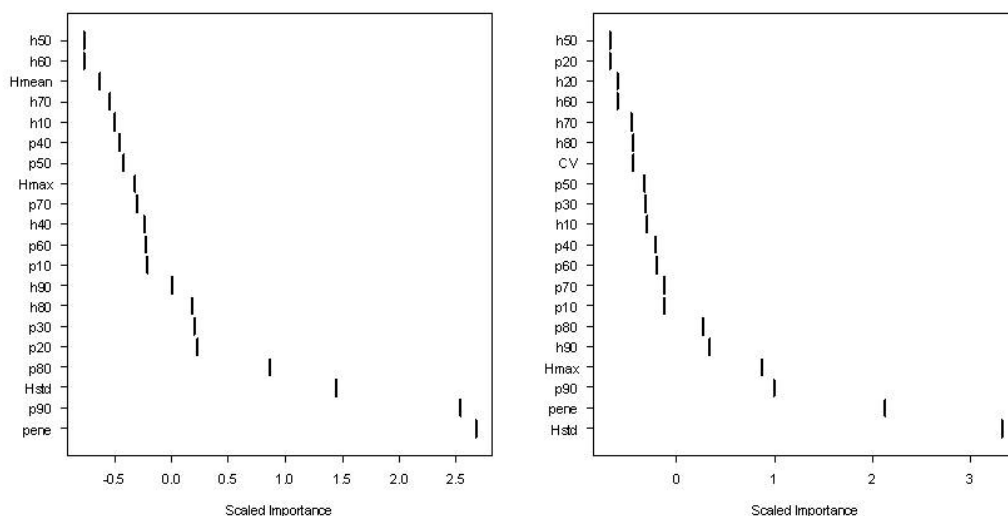


Figure 2. Importance of features in classifying defoliation level of the sampling plots using first returns (two classes on the left and 10% classification on the right side). A high value of a feature indicates an importance to the classification. For feature descriptions, see paragraph 2.4.

Table 2. The plot level classification accuracies (%) for training and test sets with both classification methods using first, last or both returns.

	Two classes			10 % classes		
	First	Last	Both	First	Last	Both
Train set	93.9	91.8	93.9	79.6	85.7	89.7
Test set	67.4	58.7	67.4	47.8	47.8	43.5

3.2 Tree level estimation

In the second approach the plot level estimates were conducted from single tree needle loss estimates (ITD trees). The best classification accuracy for training set was 94.7% (first returns) and for test set 60.6% (both returns) (Table 3.). The tree-wise accuracies were evidently better using first classification. The best tree-wise explaining features using first returns were different intensity features, max diameter, max height and crown features (Figure 3.)

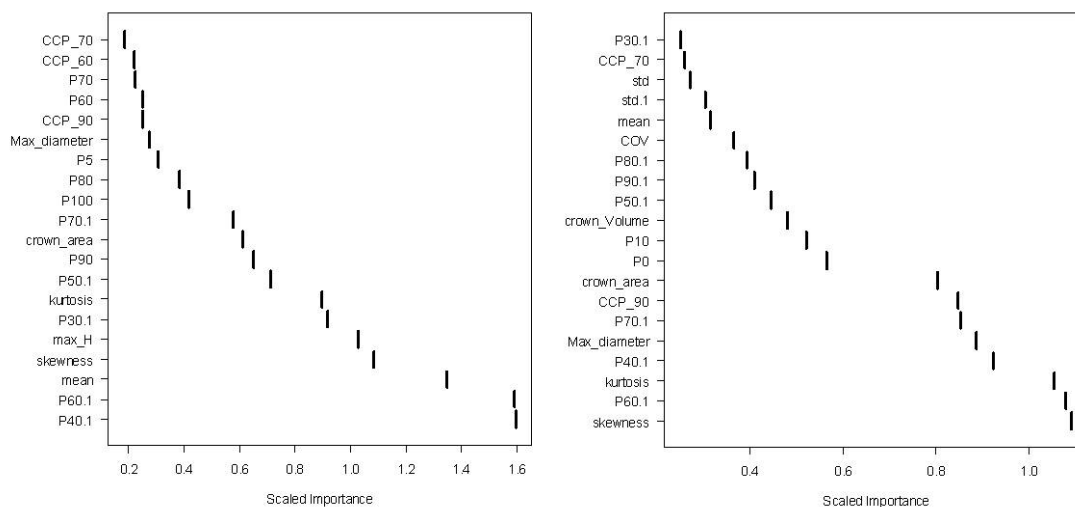


Figure 3. Importance of features in classifying defoliated trees using first returns (two classes on the left and 10% classification on the right side). A high value of a feature indicates an importance to the classification. For feature descriptions, see Table 1.

Table 3. The tree-wise classification accuracies (%) for training and test sets with both classification methods using first, last or both returns.

	Two classes			10 % classes		
	First	Last	Both	First	Last	Both
Train set	94.7	93.9	93.3	87.6	85.9	88.7
Test set	58.1	60.3	60.6	36.1	33.9	37.8

The plot-wise defoliation levels were calculated as an average of the defoliation levels of the ITD trees. The mean defoliation level of the ITD trees varied moderately from the field estimation because all the trees were not identified (Figure 4 and table 3). The best classification accuracy for test set was 60.9 (two classes). Using 10% classes the accuracy was 58.7% for the test set. The trend of minor underestimation of the defoliation level was evident (see Figure 4).

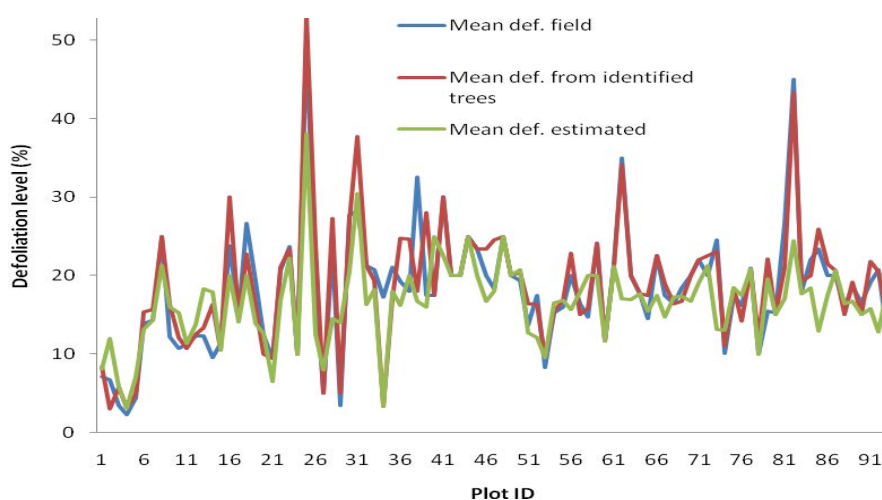


Figure 4. The plot level defoliation levels from the field (blue line), counted as an average from the identified trees (red line) and an average from the estimated defoliation of single trees (green line).

Table 4. Plot level classification accuracies (%) between field estimation, ITD trees and Random Forest estimation based on to different classification using first returns.

		Two classes		10% classes	
		ITD trees	Field	ITD trees	Field
IDT trees	Trainig set	100	77.5	100	77.5
	Test set	100	78.3	100	71.7
Estimated	Trainig set	91.8	73.5	83.7	77.5
	Test set	69.6	60.9	60.9	58.7

4. Discussion

In the present study statistical ALS features were tested for classification of mean defoliation of sampling plots. The RF method was applied for two different approaches. The analyses for both approaches were made with two classifications (Two classes and 10 % classes). To the best of our knowledge this was the first study using ALS features for estimating plot level defoliation. In several studies ALS data has been used in forest characteristics estimation other than

defoliation at stand level (see e.g. Korpela et al. 2010). In the field of forest disturbances the utilization of ALS has been less studied (see e.g. Vehmas et al. 2009 and Solberg 2006). Kantola et al. (2011) tested the ALS data for tree-level defoliation prediction. The classification accuracy was 77% for two defoliation classes (threshold 30% of defoliation).

The results of this study are in some ways comparable with the other studies using RS data in needle loss prediction. Ilvesniemi (2009) used the same Palokangas study area, when investigating the usability of aerial photographs classifying plot level defoliation. The classification accuracies varied between 38% (9 classes) and 87.3% (2 classes). Testing data was not used. Karjalainen et al. (2010) used multitemporal ERS-2 and Envisat satellite images and calculated the SAR backscattering intensities of 400m x 400m grid cells to estimate defoliation. A classification accuracy of 67.8% for test set (two classes) was obtained.

Comparing our results to previous studies with other RS materials ALS does not seem provide much additional information. If defoliation is estimated using are-based approach variation in plot-level basal area hinders the detection of defoliation effect, i.e. basal area dominates pulse penetration. For accurate defoliation detection effect of basal area has to be calibrated. Also the effect of distributions of the ALS features of different fertility classes need also be studied although the variation of different site types was small in the study area. The distribution of the defoliation level of the sampling plots was not even on the study area. The larger proportion of trees of having severe needle loss category could improve the classification accuracy. The more even distribution would also have provided possibility for testing different threshold values for classification.

The results of the present study may give an important finding for detecting and mapping insect damages and improving the inventory of damages by defoliators. ALS based system for monitoring forest health is under interest since operational forestry is adapting ALS inventories in the operational forestry. The classification accuracies might have been presumably lower applying low density ALS data, commonly used at the operational level. However, the pulse densities are assumingly increasing in the future also in practical implementations.

Acknowledgements

This study was made possible by the Maj and Tor Nessling Foundation, Foresters Foundation, Niemi Foundation and the Finnish Academy (project Improving Forest Supply Chain by Means of Advanced Laser Measurements (L-IMPACT)). We also wish to thank the Tornator Ltd.

References

- van Aardt, J.A.N., Wynne, R.H., Scivani, J.A. 2008. Lidar-based mapping of forest volume and biomass by taxonomic group using structurally homogenous segments. *Photogrammetric Engineering & Remote Sensing*. 74(8), 1033–1044.
- Axelsson, P. 2000. DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, 16–23 July 2000, Vol. XXXIII(B4), 110–117.
- Breiman, L., 2001. Random forests. *Mach. Learn.*, 45, 5-32.
- Eichhorn, J., 1998. Manual on Methods and Criteria for Harmonized Sampling, Assessment, Monitoring and Analysis of the Effects of Air Pollution on Forests. Part II. Visual Assessment of Crown Condition and Submanual on Visual Assessment of Crown Condition on Intensive Monitoring Plots. *United Nations Economic Commission for Europe Convention on Long-range Transboundary Air Pollution*, Germany.
- Crookston, N.L. and Finley, A.O., 2007-2010. yaImpute: A R package for efficient nearest neighbor imputation routines, variance estimation, and mapping. *cran.r-project.org*.

- Hall, R.J., Skakun, R.S. & Arsenault, E.J. 2007. Remotely sensed data in the mapping of insect defoliation. In: *Understanding Forest Disturbance and Spatial Pattern. Remote Sensing and GIS Approaches*. (Michael A. Wulder & Steven E. Franklin, eds.). CRC Press, Taylor & Francis Group, Boca Raton. pp. 85–111.
- Hawbaker, T.J., Keuler, N. S., Lesak, A.A., Gobakken, T., Contrucci, K., and Radeloff V.C. 2009. Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. *J. Geophys. Res.* 114, G00E04.
- Hyypä, J., Jaakkola, A., Hyypä, H., Kaartinen, H., Kukko, A., Holopainen, M., Zhu, L., Matikainen, L., Chen, R., Chen, Y., Kaasalainen, S. Krooks, A. Litkey, P., Rönholm, P., Vastaranta, M. & Lyytikäinen-Saarenmaa, P. 2009. Map Updating and Change Detection Using Vehicle-Based Laser Scanning, in *proceedings of JURSE 2009*, 20–22 May 2009.
- Ilvesniemi, S. Numeeriset ilmakuvat ja Landsat TM -satelliittikuvat männyn neulaskadon arvioinnissa. 2009. *Pro Gradu. Metsävarojen käytön laitos*. Helsingin yliopisto. 62 p. (In Finnish).
- Kantola, T., Vastaranta, M., Xiaowei, Y., Lyytikäinen-Saarenmaa, P., Holopainen, M., Talvitie, M., Kaasalainen, S., Solberg, S., Hyypä, J., 2010. Classification of defoliated trees using tree-level airborne laser scanning data combined with aerial images. *Remote Sensing*, 2(12):2665-2679.
- Kantola, T., Vastaranta, M., Yu, X., Lyytikäinen-Saarenmaa, P., Holopainen, M., Talvitie, M., Solberg, S. and Hyypä, J., 2011. Classification accuracy of the needle loss of individual Scots pines from airborne laser point clouds. Submitted.
- Karjalainen, M., Kaasalainen, S., Hyypä, J., Holopainen, M., Lyytikäinen-Saarenmaa, P. Krooks, A., and Jaakkola, A. 2010. SAR Satellite Images and Terrestrial Laser Scanning in Forest Damages Mapping in Finland, in *Proceedings of ESA Living Planet Symposium 2010 ESA Special Publication*. SP-686, 28.6.-2.7.2010, Bergen, Norway.
- Korpela, I., Ørka, H.O., Maltamo, M., Tokola, T. and Hyypä, J., 2010. Tree species classification using airborne LiDAR – effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. *Silva Fenn.*, 44(2), 319-339.
- Lyytikäinen-Saarenmaa, P. and Tomppo E. 2002. Impact of sawfly defoliation on growth of Scots pine *Pinus sylvestris* (Pinaceae) and associated economic losses. *Bulletin of Entomological Research* 92, 137–140.
- Lyytikäinen-Saarenmaa, P., Holopainen, M., Ilvesniemi, S. and Haapanen, R. 2008. Detecting pine sawfly defoliation by means of remote sensing and GIS. *Forstschutz Aktuell*. pp. 14–15.
- Maltamo, M., Mustonen, K., Hyypä, J., Pitkänen, J., Yu, X. 2004. The accuracy of estimating individual tree variables with airborne laser scanning in boreal nature reserve. *Canadian Journal of Forest Research*. 34, 1791–1801.
- Moore, B. and G. Allard. 2008. Climate change impacts on forest health. Forest Health and Biosecurity Working Paper FBS/34E. Rome, FAO.
- Solberg, S., Næsset, E., Hanssen, K.H., and Christiansen, E., 2006. Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sens. Environ.*, 102, 364–376.
- Talvitie, M., Kantola, T., Holopainen, M. & Lyytikäinen-Saarenmaa, P. 2011. Adaptive cluster sampling in inventorying forest damage by the Common pine sawfly (*Diprion pini*). *Journal of Forest Planning*, 16, 141-148.
- Vain, A., Yu, X, Kaasalainen, S and Hyypä, J. 2010. Correcting Airborne Laser Scanning Intensity Data for Automatic Gain Control Effect. *IEEE geoscience and remote sensing letters*. 7(3), 511-514.
- Vehmas, M., Packalén, P. and Maltamo, M., 2009. Assessing deadwood existence in canopy gaps by using ALS data. *Silvilaser 2009 proceedings*, October 14-16, 2009 , Texas, USA
- Yu, X.; Hyypä, J.; Holopainen, M.; Vastaranta, M.; Viitala, R. 2010. Predicting individual tree attributes from airborne laser point clouds based on random forest technique. *ISPRS J. Photogramm. Remote Sens.*, 66, 28-37.

Assessing spatial variation for tree and non-tree objects in a forest-tundra ecotone in airborne laser scanning data

Nadja Thieme, Ole Martin Bollandsås, Terje Gobakken & Erik Næsset

Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, NO-1432 Ås, Norway, nadja.thieme@umb.no, ole.martin.bollandsas@umb.no, terje.gobakken@umb.no, erik.naesset@umb.no

Abstract

Changing climate is expected to have a significant impact on temperature-sensitive ecosystems like the forest-tundra ecotone. In Norway, this ecotone constitutes a large proportion of the total land area and effective monitoring techniques are required. It has been indicated that height and intensity data from airborne laser scanning may hold potentials for monitoring of small trees. In the present study, Voronoi polygons and variograms were employed in order to assess the spatial patterns of trees and non-tree objects located in the forest-tundra ecotone. Patterns both for trees and non-tree objects could be recognised using Voronoi polygons in combination with height and intensity values. Furthermore, variograms and cross-variograms revealed different characteristics for trees and non-tree objects, however, limited to large individual objects located on flat terrain.

Keywords: Forest-tundra ecotone, ALS, intensity, Voronoi, variogram, cross-variogram

1. Introduction

Currently, there is a strong focus on the effects of a changing climate. An increase of mean temperature will affect most ecosystems and the interaction between them (Stenseth et al. 2002; Woodall et al. 2009) and for example the growth in forest ecosystems is expected to be affected. The relatively largest effects of increasing temperatures will happen in forest ecosystems where temperature is the limiting factor, such as in the forest-tundra ecotone. Increased growth (Zheng et al. 2002) and migration of trees into the current alpine zone (Opdam and Wascher 2004) are predicted. It is of great interest to develop methods for monitoring these changes, both for carbon reporting and ecological reasons.

The forest-tundra ecotone constitutes a large proportion of the total land area of Norway and monitoring requires that some remote sensing technique is used. However, it is challenging to use such techniques in these areas because the objects of interest – trees – are both small in size and sparsely distributed. Thus, they are easily confused with other non-tree objects such as rocks and hummocks.

Airborne laser scanning (ALS) data is nowadays frequently used for forest inventory (e.g. Næsset 2007). Studies that utilize ALS data for change estimation have been conducted (e.g. Næsset and Gobakken, 2005). There have also been studies dealing with tree migration into the alpine zone using ALS data (Næsset and Nelson 2007). Most of these studies have only utilised the height information of individual echoes of the ALS point cloud. However, since the ALS data also include information about the intensity of each echo, it might be possible to discriminate between trees and other objects based on this spectral information. Further, the spatial structure of the data may also help discriminating between different types of objects. Spatial correlation or dependency is demonstrated by a variety of biological phenomena (Rossi et al. 1992), and often emerges in patches (Fry and Stephens 2010). Therewith, it is also likely that the spatial variation of ALS echoes classified as vegetation echoes is different around trees

and non-tree objects. For example, if the probability of finding a tree is larger in the close proximity of another tree compared to a non-tree object, this can be exploited to distinguish trees from non-trees. The aim of this study was to test the capability of ALS data to distinguish between trees and non-tree objects in the forest-tundra ecotone using laser height and intensity information and spatial correlation analyses.

2. Materials and Methods

2.1 Study area

The study was carried out using data from a 1,500 km long and approximately 180 m wide longitudinal transect stretching from Tromsø in the northern part of Norway (69°3' N 17°5' E) to Tvedestrand in southern Norway (58°3' N 9°0' E) (see Figure 1). Literately hundreds of mountain forest/alpine elevation gradients were encompassed by the transect. Sample plots were laid out as ground reference in several transitions between mountain forest and the alpine zone, the forest-tundra ecotone. The prevalent tree species were Norway spruce (*Picea abies* (L.) Karst.), Scots pine (*Pinus sylvestris* L.), and mountain birch (*Betula pubescens* ssp *czerepanovii*). Rounded forms with certain occurrences of hummocks, rocks and boulders characterised the terrain surface in most localities along the transect.



Figure 1: Overview of the study area with the selected field sites (black points). The 1,500 km long transect (black line) stretches from to 69°3' N 17°5' E to 58°3' N 9°0' E.

2.2 Field data

The field work was conducted during summer 2008 in order to provide *in situ* tree data within the transect.

The overall field dataset for the entire transect was collected at 36 different field sites allocated along the transect. However, in this first explorative study, only four of these sites were selected for analysis. The selection of these four sites among the 36 was conducted by visual inspection

of digital aerial imagery (Statens kartverk 2009) in order to cover a fairly large range of topographic conditions (see Figure 1).

At the four sites, sample trees were selected according to a modified version of the point-centred quarter sampling method (PCQ) (Cottam and Curtis 1956; Warde and Petranka 1981) with a maximum search limit of 25 m. For each sample tree, tree height, stem diameter at root collar, crown diameter, and tree species were recorded individually. The precise position of each tree was determined using real-time kinetic differential Global Navigation Satellite Systems (dGNSS) employing two Topcon Legacy E+ 20-channel dual-frequency receivers observing pseudo range and carrier phase of both GPS (Global Positioning System) and GLONASS (Global Navigation Satellite System) satellites with an expected precision of 3–4 cm. In total, 70 trees were sampled in the four sites, i.e., 54 mountain birch, 12 Norway spruce, and 4 Scots pine. A summary of tree metrics is given in Table 1.

Table 1: Summary of field measurements of trees.

Tree species	Characteristics	<i>n</i>	Mean	Min	Max
Mountain birch	Height (m)	54	1.14	0.07	4.90
	Diameter at root collar (cm)	53	3.05	0.20	10.60
	Crown area (m ²)	54	0.78	0.002	4.96
Norway spruce	Height (m)	12	1.44	0.07	4.60
	Diameter at root collar (cm)	12	5.60	0.20	16.10
	Crown area (m ²)	12	1.43	0.006	5.29
Scots pine	Height (m)	4	0.31	0.16	0.72
	Diameter at root collar (cm)	4	0.88	0.30	1.70
	Crown area (m ²)	4	0.04	0.008	0.11

2.3 Laser data

Airborne laser scanner data were acquired in two separate acquisitions in July 2006 and July 2007. The first acquisition was conducted in Southern and Central Norway with an Optech ALTM 3100C laser scanning system, whereas the remaining part in Northern Norway was scanned with a Gemini upgraded version of the Optech ALTM 3100C laser scanner, denoted as ALTM Gemini. An 80 km long overlap zone was flown at approximately 65°N with both systems for comparison reasons.

The lasers were carried by Piper PA-31 Navajo aircrafts at an average flying altitude of 800 m a.g.l. and a flight speed of approximately 75 ms⁻¹. Furthermore, the scan frequency was 70 Hz, maximum half angle was 7°, and the average footprint diameter was 20 cm for both acquisitions. Pulse repetition frequency (PRF) however, was 100 kHz for the ALTM 3100C and 125 kHz for the ALTM Gemini, which resulted in mean pulse densities of 6.8 m⁻² and 8.5 m⁻², respectively.

Pre-processing of the laser data was conducted by the contractor (Blom Geomatics, Norway). Planimetric coordinates (*x* and *y*) and ellipsoidal height values were computed for all laser echoes. Laser echoes labelled “last-of-many” and “single”, hereafter denoted as LAST, were used for the derivation of the terrain model. Ground echoes were classified from the planimetric coordinates and the corresponding height data of the LAST return echoes, and a triangulated irregular network (TIN) was derived with an iteration angle of 9° using the TerraScan software (Terrasolid 2010). Laser echoes labelled as “first-of-many” and “single”, hereafter denoted as FIRST, were used in the analyses of the current study. Therefore, FIRST return echoes were projected onto the TIN surface and the corresponding terrain height values were interpolated on these locations. For the spatial analyses, the differences between FIRST return echo heights and the corresponding interpolated terrain height values were computed and stored. Merely FIRST

return laser echoes with positive height values were included in the spatial analyses since these represent the sole indicators for the presence of objects on the terrain surface.

2.4 Assessing spatial variation

For the assessment of the spatial point pattern of the FIRST return laser echoes, Voronoi polygons were used. Voronoi polygons are used in point pattern analysis in numerous disciplines (Boots and Getis 1988). Thereby, a continuous space is divided into regions ensuring that each location in the space is associated with the closest point of the point pattern (Okabe et al. 2000) (see Figure 2). This means “that each Voronoi polygon consists of an area that is closer to a given point than any other point.” (Wulder et al. 2006).

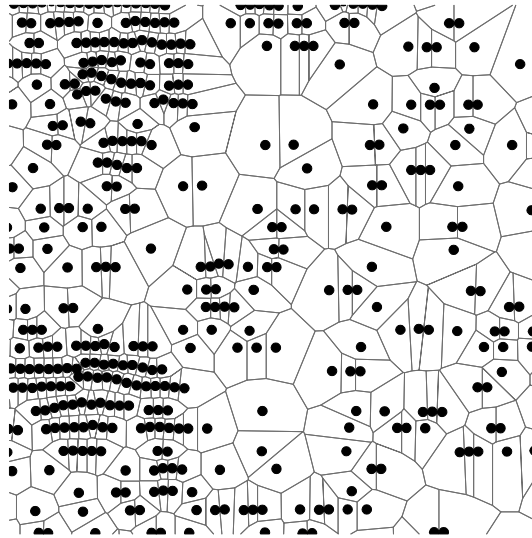


Figure 2: General example of Voronoi polygons.

An adequate area was defined for each field site for the computation of Voronoi polygons in order to avoid edge effects at the sample plot borders. Voronoi polygons were computed for the FIRST return laser echoes and both laser height and intensity values were assigned to the corresponding polygons for spatial pattern analysis of the distinct parameters. Furthermore, the Voronoi polygons were overlaid with the locations of the tree and non-tree objects in order to find patterns characterising these objects. For this purpose, rocks representing non-tree objects were visually identified on digital aerial imagery and localised in a GIS environment prior to analysis.

Furthermore, tree and non-tree objects were defined for assessing the behaviour of the FIRST return laser echoes using geostatistical techniques. Geostatistics in general can involve three interrelated stages of characterising the spatial correlation, interpolating unknown values at each location and random simulation. In this study, the focus was on the first component where the spatial data is characterised by means of semivariogram analysis in order to find differences of the behaviour of spatial correlation between tree and non-tree objects. For this purpose, the spatial dimensions of all field-measured trees were defined by a 6 m buffer around their individual positions in order to provide sufficient point pairs for the geostatistical analysis. Thus, also more distant locations were included for being able to detect the change in the spatial correlation since nearby values are more likely to be similar than values further off (Fry and Stephens 2010). The same procedure was performed for the non-tree objects for the inclusion of an adequate number of point pairs.

A semivariogram, which is usually referred to as a variogram, represents a graphical illustration of the spatial variability of a variable. The spatial correlation of a variable is measured by

calculating the semivariances of multiple pairs of observations as a function of their separation distance (Isaaks and Srivastava 1989) and is referred to as an experimental variogram. The separation distances are usually divided into various distance classes, referred to as lag. The semivariances of a dataset are computed as

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z(x_i) - z(x_i + h)]^2 \quad (1)$$

where $\hat{\gamma}(h)$ is the estimated semivariance estimator for lag h and $n(h)$ the number of data points separated by h (Rossi et al. 1992).

For investigating a spatial co-variability of two variables, the semivariance model is extended to the so called cross-variogram that is computed as

$$\hat{\gamma}_{uv}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [(z_u(x_i) - z_u(x_i + h))(z_v(x_i) - z_v(x_i + h))] \quad (2)$$

where $\hat{\gamma}_{uv}(h)$ is the estimated cross-variance of the two variables for lag h .

Univariate experimental variograms are characterised by an increase in semivariance with distance h when spatial dependence is present. The experimental cross-variogram however can be a decreasing function when the two variables are negatively correlated. Both experimental univariate and cross-variograms may reach the so called sill where the semivariance levels off or increase ad infinitum. The so called nugget effect is characterised by a semivariance value >0 at the origin and represents spatial variability due to measurement errors or distances shorter than the sample spacing.

Prior to the geostatistical analysis, the two laser-derived variables under investigation, positive height values and normalised intensity, were centred and scaled. For each tree and non-tree object, the corresponding laser echoes were extracted using the 6 m buffers and stored separately for further analysis in the statistical computing software R (R Development Core Team 2007). Experimental variograms and cross-variograms were calculated individually using the gstat spatial package (Pebesma 2004). Distance classes were defined as follows since lags closer to 0 are expected to provide more information than lags further off: 0 m, 0.5 m, 1 m, 1.5 m, 2 m, 2.5 m, 3 m, 4 m and 5 m.

Second-order stationarity was assumed, implying constant mean, variance and covariances due to separation only (Webster and Oliver 2001). Furthermore, isotropy was assumed for the spatial distributions of the two variables under investigation.

3. Results and Discussion

The spatial point pattern that was investigated using Voronoi polygons showed promising results concerning polygon sizes in combination with both the laser height values and the normalised intensity. In general, small-sized polygons indicate a high density of FIRST return laser echoes in the neighbourhood (see Figure 2) which further suggests tree or non-tree objects in that area. Large polygons indicate random laser echoes which actually merely are representing noise in the laser data.

Regarding positive laser height values, the spatial pattern characterising trees strongly depended on the terrain surface. For areas with relatively flat terrain and only a few steep slopes, single trees from the field data exceeding certain heights (~ 1.5 m) and crown areas (~ 2 m²) could be recognised in the spatial pattern. Concerning trees with crown areas under 1 m² and heights lower than 1 m, no spatial pattern could be found. For trees between 1 and 1.5 m, a wide crown area is essential for the indication of a potential tree using Voronoi polygons. In quite rugged terrain with some flat areas but also steep slopes, even trees with crown areas around 2.3 m² were hard to identify in the spatial pattern. Furthermore, areas with flat rocky surfaces had low indications for trees by means of the Voronoi polygons. Almost all non-tree objects had a

characteristic spatial pattern independent of the terrain surface. This result was expected since these objects were merely located using aerial imagery by which large sizes – especially in diameter – are required for identification.

For the normalised intensity, distinct differences in the spatial pattern for tree and non-tree objects were found. Even though the intensity values for tree and non-tree objects are similar, trees are characterised by a very heterogeneous pattern of intensity values inside the tree crown, whereas non-tree objects showed a more homogeneous pattern (see Figure 3).

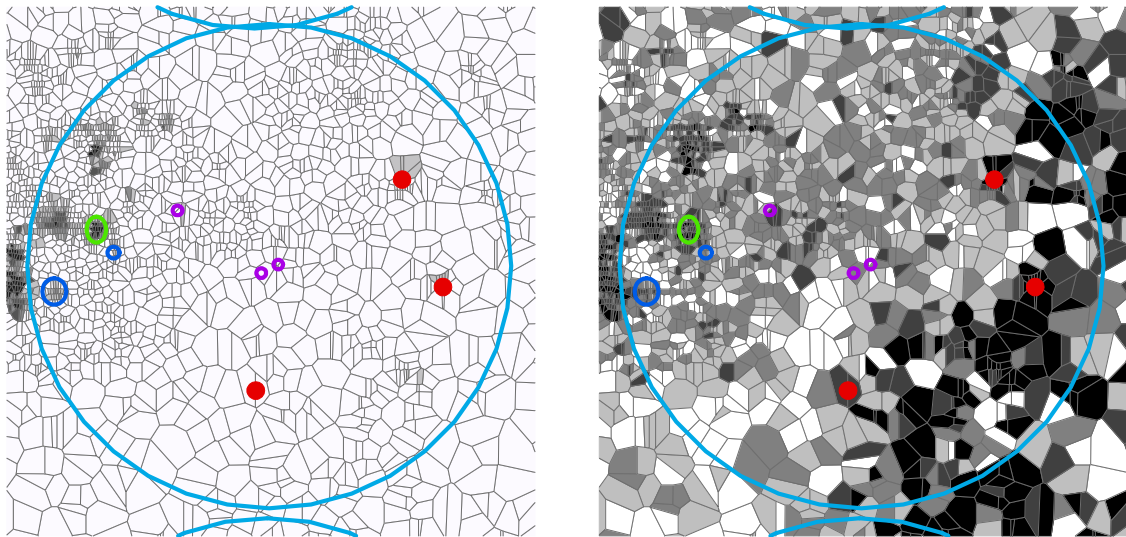


Figure 3: Voronoi polygons with assigned height (left) and intensity (right) values. Red markers label non-tree objects, coloured ellipses trees measured in field.

Concerning the experimental variograms and cross-variograms for tree and non-tree objects, merely variograms with sufficient point pairs were included in the analysis. Some authorities consider 30 to 50 comparisons as sufficient, whereas Webster and Oliver (2001) suggest a minimum of 100 sampling points for isotropic variation. Due to sparse point data, especially for non-tree objects, one tree and eight non-tree objects were discarded. Furthermore, based on the findings using Voronoi polygons, also trees smaller than 1 m were excluded from the geostatistical analysis. This resulted in a total of 31 tree and 15 non-tree objects used in the geostatistical analysis employing experimental variograms and cross-variograms.

Trees are expected to have slightly higher semivariances for both height and intensity values in short distances which furthermore are increasing with increasing distance. This is caused by the more heterogeneous pattern of the height and intensity values inside the tree crown (see Figure 3). Non-tree objects in the forest-tundra ecotone are mostly represented by hummocks, rocks and boulders which are usually solid and compact objects with a more or less homogeneous surface (see Figure 3). Therewith, one would expect low semivariances for short distances and a sudden increase in semivariance for longer distances both related to height and intensity values.

Concerning individual trees located in flat terrain, the experimental variograms for height and intensity and the cross-variogram for both parameters revealed the expected pattern (see Figure 4) with a clear nugget effect followed by a constant increase with increasing distance for the individual variograms. The cross-variogram showed a constant decrease with increasing distance due to the negative correlation between height and intensity. For distinct non-tree objects located in flat terrain, both the individual experimental variograms clearly showed the expected pattern (see Figure 4) with low semivariances within 2 m and an abrupt increase for distances larger than 2 m. The cross-variogram reveals the same characteristics involving the negative correlation between the two parameters.

For smaller trees located in more rugged and varied terrain, both the individual experimental

variograms and cross-variograms got noisier. Furthermore, surrounding trees affect the semivariances of the trees under investigation. Non-tree objects in more rugged terrain in combination with flat rocky surroundings did not show a very clear spatial pattern both concerning the individual variograms for height and intensity and the cross-variograms between these two parameters.

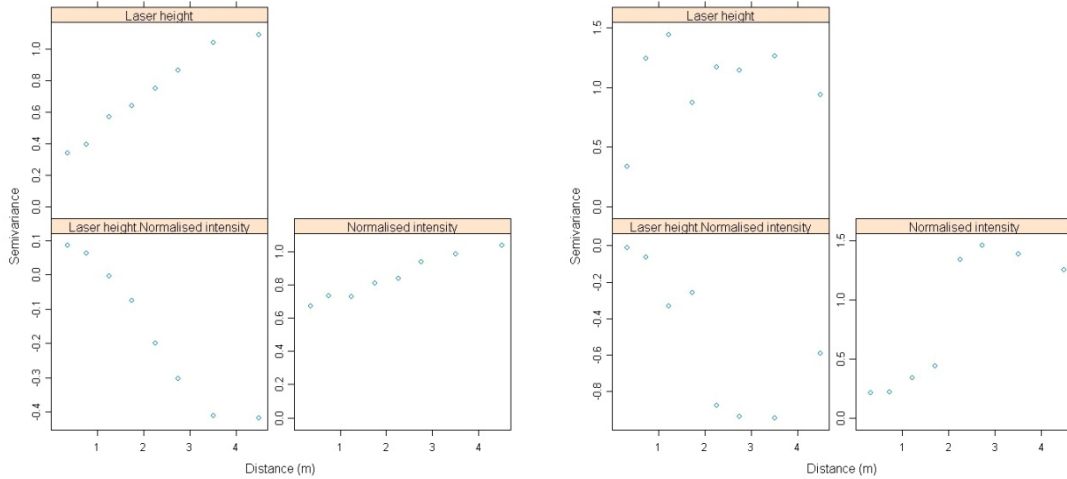


Figure 4: Experimental variograms and cross-variograms for an individual tree (left) and a distinct non-tree object (right).

4. Conclusion

The analysis of Voronoi polygon sizes in combination with positive laser height and normalised intensity values showed promising results in regard to the indication of objects located above the terrain surface. Investigating the spatial pattern of the height and intensity values for small-sized polygons and their neighbourhood, pre-measured trees and non-tree objects could be clearly recognised.

Furthermore, experimental variograms and cross-variograms for tree and non-tree objects revealed characteristics one would expect based on the precedent analysis of the Voronoi polygons. The individual experimental variograms for height and intensity showed a clear nugget effect for trees followed by a constant increase with increasing distance, whereas the cross-variogram revealed a negative correlation between the two parameters with a decreasing function of distance. Distinct non-tree objects revealed low semivariances on short distances followed by a sudden increase for increasing distances in the individual experimental variograms for height and intensity. Furthermore, the cross-variogram followed the same trend with a decreasing function due to the negative correlation of the two parameters.

To conclude, the Voronoi polygon investigation as well as the geostatistical analysis showed promising results for the identification and classification of objects on the terrain surface. However, clear results are limited to flat terrain and large individual objects in order to discriminate between tree and non-tree objects. Further investigation involving for example neighbourhood and terrain properties is needed for being able to better distinguish objects in areas like the forest-tundra ecotone. For this purpose, fitting theoretical variograms using different models in order to obtain nugget, sill and range values might be useful.

Acknowledgements

This research has been funded by the Research Council of Norway (project #184636/S30). We wish to thank Blom Geomatics AS, Norway, for collection and processing of the airborne laser

scanner data. Thanks also appertain to Mr. Vegard Lien at the Norwegian University of Life Sciences, who was responsible for the fieldwork.

References

- Boots, B.N. and Getis, A., 1988. *Point Pattern Analysis*. Scientific Geography Series, Sage Publications, Newbury Park.
- Cottam, G. and Curtis, J.T., 1956. The use of distance measures in phytosociological sampling. *Ecology*, 37, 451-460.
- Fry, D.L. and Stephens, S.L., 2010. Stand-level spatial dependence in an old-growth Jeffrey pine – mixed conifer forest, Sierra San Pedro Mártir, Mexico. *Canadian Journal of Forest Research*, 40, 1803-1814.
- Isaaks, E.H. and Srivastava, R.M., 1989. *An introduction to applied geostatistics*. Oxford University Press, New York.
- Næsset, E., 2007. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. *Scandinavian Journal of Forest Research*, 22, 433-442.
- Næsset, E. and Gobakken, T., 2005. Estimating forest growth using canopy metrics derived from airborne laser scanner data. *Remote Sensing of Environment*, 96, 453-465.
- Næsset, E. and Nelson, R., 2007. Using airborne laser scanning to monitor tree migration in the boreal-alpine transition zone. *Remote Sensing of Environment*, 110, 357-369.
- Okabe, A., Boots, B., Sugihara, K. and Chiu, S.N., 2000. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley, Chichester.
- Opdam, P. and Wascher, D., 2004. Climate change meets habitat fragmentation: linking landscapes and biogeographical scale levels in research and conservation. *Biological Conservation*, 117, 285-297.
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30, 683-691.
- R Development Core Team, 2007. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.
- Rossi, R.E., Mulla, D.J., Journel, A.G. and Franz, E.H., 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs*, 62, 277-314.
- Statens kartverk, 2009. *Norge i bilder*. Available from www.statkart.no [accessed 1 June 2011].
- Stenseth, N.C., Mysterud, A., Ottersen, G., Hurrell, J.W., Chan, K.S. and Lima, M., 2002. Ecological effects of climate fluctuations. *Science*, 297, 1292-1296.
- Terrasolid, 2010. *TerraScan User's Guide*. Terrasolid Ltd., Jyväskylä. Available from www.terrasolid.fi [accessed 6 May 2010].
- Warde, W. and Petranks, J.W., 1981. A correction factor table for missing point-center quarter data. *Ecology*, 62, 491-494.
- Webster, R. and Oliver, M.A., 2001. *Geostatistics for environmental scientists*. Wiley, New York.
- Woodall, C.W., Oswalt, C.M., Westfall, J.A., Perry, C.H., Nelson, M.D. and Finley, A.O., 2009. An indicator of tree migration in forests of the eastern United States. *Forest Ecology and Management*, 257, 1434-1444.
- Wulder, M.A., White, J.C., Dymond, C.C., Nelson, T., Boots, B. and Shore, T.L., 2006. Calculating the risk of mountain pine beetle attack: a comparison of distance- and density-based estimates of beetle pressure. *Journal of Environmental Informatics*, 8, 58-69.
- Zheng, D., Freeman, M., Bergh, J., Røsberg, I. & Nilsen, P., 2002. Production of *Picea abies* in south-east Norway in response to climate change: A case study using process-based model simulation with field validation. *Scandinavian Journal of Forest Research*, 17, 35-46.

Exploring horizontal area-based metrics to discriminate the spatial pattern of trees using ALS

Inka Pippuri¹, Eveliina Kallio¹, Matti Maltamo¹, Petteri Packalén¹ & Heli Peltola¹

¹Faculty of Science and Forestry, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland, inka.pippuri@uef.fi, eveliina.kallio@uef.fi, matti.maltamo@uef.fi, petteri.packalen@uef.fi, heli.peltola@uef.fi

Abstract

Airborne Laser Scanning (ALS) data can be used to accurately determine tree and stand characteristics. We hypothesize here that three-dimensional ALS data can also be used for characterizing the horizontal forest structure like the spatial pattern of trees. This kind of information is of primary interest in forest management. The objectives of this study were (1) to identify ALS point cloud metrics and horizontal texture and landscape metrics, which can be used to determine the spatial pattern of trees and (2) to study how well the clustered spatial pattern of trees can be separated from others.

The field data consisted of 28 microstands, of which 11 were clustered and 17 random or regular. Linear discriminant analysis was used to classify the microstands by means of the metrics calculated from ALS data. The best ALS metrics to determine the spatial pattern of trees were determined by the best overall accuracies (OA) and kappa-values (κ) and based on the significance tests of models and the correlation matrices of metrics.

The classification of the spatial pattern of trees succeeds well based on ALS metrics, with the overall accuracy being 0.89 and kappa-value 0.77. Especially the calculated landscape metrics were found good predictors of the spatial pattern of trees, for example: the number of the ground patches per hectare, the average size of the tree patches and the standard deviations of the size of all patches in the microstand. To conclude, our results were encouraging to detect the spatial pattern of trees based on low density ALS data.

Keywords: Airborne Laser Scanning, spatial pattern of trees, texture metrics, landscape metrics

1. Introduction

In even-aged forestry, one aim of forest management is to produce a regular spatial pattern of trees at stand level in order to maximize the use of growing space (e.g. Pukkala 1990). The spatial pattern of trees in forest can be defined as the locations of the trees in relation to each other. It can be regular, random, clustered, or any combination of them (e.g. Pielou 1960; Tomppo 1986). The scale used in analysis also affects this classification. The determination of the spatial pattern of trees, based on the tree locations in a two-dimensional space, can be estimated statistically. It can be done based on sampling with dedicated sampling designs or by measuring the exact locations of all trees. Commonly the spatial pattern is compared to a statistically random pattern, the so-called Poisson forest (Diggle 1983). Because the field work for measuring the spatial pattern of trees is rather laborious and expensive, the spatial pattern has not been widely utilized in the growth and yield models for simulation of stand development and in forest planning, despite its effects on the pattern of tree growth (e.g. Kilkki et al. 1985; Gavrikov and Stoyan 1995; Pukkala et al. 1998; Coops and Culvenor 2000).

High resolution remote sensing methods can provide the means to get estimates of the spatial pattern of trees without expensive field work. For example, Coops and Culvenor (2000) and Uuttera et al. (1998) have determined the spatial pattern of trees from high spatial resolution imagery and aerial photographs, respectively. Existing ALS studies have concentrated on the vertical structure of forests. Probably the most obvious way to use ALS data in the determination of the spatial pattern of trees would be ITD (individual tree detection), in which the positions of the tallest trees can be located (e.g. Mustonen 2002). However, it is not currently applied in the practical forest inventories, which are operated with the area based ALS.

From the methodological point of view, the landscape and texture metrics are tools for the description of horizontal variation. The landscape metrics are usually used to describe the features, the structure, and the variations of landscape: for example area, fringe, shape, neighbourhood and consistence of landscape (McGarigal and Marks 1995). The landscape metrics calculated from satellite images and aerial photographs have commonly been applied during the last two decades especially for detection of changes in forest coverage, forest fragmentation, and habitat patches (e.g. Hargis et al. 1998; Sachs et al. 1998; Kouki et al. 2001). Similarly, texture metrics (Haralick et al. 1973) have also been widely calculated based on different remote sensing materials (e.g. Shang and Waite 1991; Tuominen and Pekkarinen 2005). It has been noted that certain statistical dependences exist between forest characteristics and the spatial distribution of the grey tones (e.g. Wulder et al. 1998; Packalen and Maltamo 2006).

The objectives of this study were to (1) identify point cloud metrics and horizontal texture and landscape metrics from ALS data, which can be used to determine the spatial pattern of trees and (2) to study how well the clustered spatial pattern of trees can be separated from others.

2. Material and methods

2.1 Field and ALS data and spatial indices

The field data consisted of 28 microstands measured in Janakkala, Southern Finland, in August 2009. The term ‘microstand’ refers to stand that is produced by segmentation using ALS based canopy height model and is smaller than stands in silvicultural sense (here 0.2 and 0.7 hectares). All the microstands were at first thinning phase but not yet thinned. The microstands were dominated by Scots pine (*Pinus sylvestris*) or by Norway spruce (*Picea abies*). T-square sampling was carried out in each microstand, offering efficient and statistically coherent measurements to define the spatial pattern of vegetation in the field (Besag and Cleaves 1973). It is based on point-to-point distance measurements: distances are measured between each sample point and the nearest tree, and between it and its nearest tree in that half of the sampling space. Sample points were located in the crossings of a randomly placed systematic grid at the intervals of about 10 m. For tree measurements a minimum diameter of 5 cm at breast height was used.

The classes of the spatial pattern of trees for the microstands were defined with spatial indices based on the distance measurements (Diggle 1983). Besags and Cleaves’ t_N -index (1973) appeared to be the most suitable for the T-square measurements. Very large or small values of the t_N -index suggest clustered or regular spatial pattern, respectively. When the t_N -index has a value near to 0.5 the spatial pattern is random (Poisson forest). The M-index (Bartlett 1937) can be used as a test statistic for the t_N -index. These statistics were calculated as follows:

$$t_N = n^{-1} \sum_{i=1}^n x_i^2 (x_i^2 + 0.5z_i^2)^{-1}, \quad (1)$$

$$M = 48n (n \log(\bar{a}) - \sum_{i=1}^n \log a_i)(13n+1)^{-1}, \quad (2)$$

where n is the number of the observations, x_i is the distance from a sample point i to the nearest tree, and z_i is the distance from the nearest tree to the nearest tree, $a_i = x_i^2 + 0.5z_i^2$, and $\bar{a} = n^{-1} \sum_{i=1}^n a_i$, $i=1, 2, \dots, n$. Based on calculated spatial indices, only two regular microstands existed in the data. Because we were especially interested in clustered one, microstands were finally classified only into two classes (clustered and random/regular). Thus, the spatial pattern of the trees was random or regular in 17 and clustered in 11 of the measured microstands.

The ALS data were collected in summer 2007 using an Optech ALTM3100 laser scanning system at a flight altitude of 2400 m above ground level, a flight speed of 75 m/s, and a scan angle of 30°. This resulted in a swath width of 1560 m and a nominal sampling density of about 0.62 measurements per m². The measurements of pulses were reclassified to represent first and last echoes. A digital terrain model (DTM) was generated from the ALS data and the ALS heights were converted to an above-ground scale by subtracting the DTM from the orthometric heights.

2.2 ALS metrics

Predictor variables were classified into three groups: 1) point cloud metrics calculated from ALS point cloud and 2) texture and 3) landscape metrics calculated from rasterized canopy height model (CHM). The point cloud metrics considered are: the proportion of laser hits on the vegetation (*veg5*, vegetation limit of 5 m above ground), the proportional canopy densities of 1%, 5%, 10%, 20%, ..., 90%, 95%, 99% (*p1*, ..., *p99*), and the standard deviation (*hstd*) and the average height of the canopy hits (*havg*). The metrics were calculated separately for the first (f) and the last (l) echo data.

First echo data were used to generate CHM for the microstands. CHM was interpolated by taking the maximum height at above-ground scale within a certain radius from the pixel centre. The used radius was 2 m and the pixel size 1 m. The created CHM was classified into two classes to enable the calculation of texture and landscape metrics. The classes were tree pixels (height above 5 m) and ground pixels (height under 5 m). The metrics were calculated for each microstand based on the pixels inside the boundaries of a microstand.

A normalized grey-level co-occurrence matrix was calculated following the principles presented by Haralick et al. (1973). The texture metrics were calculated as an average of all the directions (0, 45, 90, and 135 degrees) with a lag value of 5 pixels. Only one gray tone spatial dependence matrix was created for each plot by direction. The landscape metrics were calculated within a microstand based on ground patches (GP) and tree patches (TP) that are comprised of neighbouring ground and tree pixels in 4-neighbourhood, respectively. The used texture and landscape metrics are presented in Table 1.

Table 1. Description of texture and landscape metrics. GP = ground patches, TP = tree patches and P = all patches.

Texture metric	Description
<i>ASM</i>	angular second moment
<i>cont</i>	contrast
<i>corr</i>	correlation
<i>text_var</i>	variance
<i>idm</i>	inverse difference moment (homogeneity)
<i>savg</i>	sum average
<i>svar</i>	sum variance
<i>sentro</i>	sum entropy
<i>entro</i>	entropy
<i>dvar</i>	difference variance
<i>dent</i>	difference entropy
<i>avg</i>	the normal average of grey-levels
<i>std</i>	the normal variance of grey-levels
Landscape metric	
<i>GP%</i> , <i>TP%</i>	the proportions of the total area of a microstand that are areas of ground patches and tree patches
<i>GP_{density}</i> , <i>TP_{density}</i>	the number of patches per hectare
<i>GP_{ave}</i> , <i>TP_{ave}</i>	the average size of the patches (m ²)
<i>GP_{sd}</i> , <i>TP_{sd}</i>	the standard deviations of the size of patches (m ²)
<i>GP_N</i> , <i>TP_N</i>	the average number of neighbouring pixels belonging to the same class
<i>P_{density}</i> , <i>P_{ave}</i> , <i>P_{sd}</i>	the number, the average size and the standard deviations of the size of all patches

2.3 Classification of microstands and variable selection

In this study we used linear discriminant analysis (e.g Lebart et al. 1984) to classify the microstands (clustered or random/regular) by means of the metrics calculated from ALS data. Wilks lambda (e.g Mardia et al. 1979) was used to test the significance of models. The accuracies of the classifications were calculated by means of overall accuracy (OA) and kappa-value (k) (Rosenfeld and Fitzpatrick-Lins 1986). The goodness of the estimated classification was tested by comparing it with the corresponding classification based on the measured field data.

The maximum number of variables in discriminant functions was fixed at three because of the small number of the microstands. Models were built separately in the three variable groups (point cloud, texture and landscape metrics) by means of testing all the possible combinations of the metrics within each group (including logarithmic, square root, inverse and power two transformations of metrics). The three best metrics were selected from each of the variable groups according to the best overall accuracies and kappa-values of classifications, and based on the significance tests of models and the correlation matrices of the metrics. These metrics were selected into new variable group. Then we built new models with three predictors based on metrics from the mixed groups. Because we were not able to test our models with independent test data, all models were built using leave-one-out cross-validation.

3. Results

3.1 Classification of the spatial pattern of trees

In the most accurate classification model based on landscape metrics the overall accuracy was 0.89 and kappa-value 0.77. Two clustered microstands and one random/regular microstand were classified incorrectly (Table 2). Respectively, in the most accurate classification model based on texture metrics the overall accuracy was 0.75 and kappa-value 0.47 and based on point cloud metrics same values were 0.79 and 0.54. Six clustered microstands and one random/regular microstand were classified incorrectly based on texture metrics and four clustered microstands and two random/regular microstands based on point cloud metrics (Table 2). The most accurate model for the classification of the spatial pattern of trees based on metrics from the mixed group was exactly the same as the most accurate model based on the landscape metrics only (Table 3).

Table 2. The ten most accurate classification models based on three different predictors selected from a group of landscape, texture or point cloud metrics, the overall accuracies and the kappa-values of the classifications and the number of correctly classified microstands. The order of the variables does not indicate the relative importance of the metrics. * = Model statistically significant at the level $p=0.05$.

Metrics	Overall accuracy	Kappa-value	Correct, clustered	Correct, random/regular
Landscape metrics				
$\ln(\text{GP}_{\text{density}}) + 1/(\text{TP}_{\text{ave}}) + \text{P}_{\text{sd}}$	0.89 *	0.77	9/11	16/17
$(\text{TP}_{\text{density}})^2 + 1/(\text{P}_{\text{ave}}) + \text{P}_{\text{sd}}$	0.89 *	0.77	9/11	16/17
$1/(\text{TP}_{\text{ave}}) + 1/(\text{P}_{\text{ave}}) + \text{P}_{\text{sd}}$	0.89 *	0.77	9/11	16/17
$\text{P}_{\text{ave}} + \text{TP}_{\text{N}} + (\text{P}_{\text{sd}})^2$	0.89 *	0.76	8/11	17/17
$(\text{TP}_{\text{density}})^2 + \ln(\text{P}_{\text{density}}) + \text{P}_{\text{sd}}$	0.86 *	0.70	9/11	15/17
$\text{TP}_{\text{N}} + (\text{P}_{\text{density}})^2 + \sqrt{\text{P}_{\text{sd}}}$	0.86 *	0.70	9/11	15/17
$\text{GP}\% + \ln(\text{P}_{\text{ave}}) + (\text{P}_{\text{sd}})^2$	0.86 *	0.69	8/11	16/17
$\text{TP}\% + \ln(\text{GP}_{\text{density}}) + \text{P}_{\text{sd}}$	0.86 *	0.69	8/11	16/17
$\text{GP}_{\text{ave}} + \ln(\text{P}_{\text{ave}}) + (\text{P}_{\text{sd}})^2$	0.86 *	0.69	8/11	16/17
$\ln(\text{TP}\%) + \ln(\text{TP}_{\text{ave}}) + (\text{P}_{\text{sd}})^2$	0.86 *	0.69	8/11	16/17
Texture metrics				
$\ln(\text{ASM}) + \text{idm}^2 + \text{avg}^2$	0.75	0.43	5/11	16/17
$1/(\text{ASM}) + \text{cont}^2 + \ln(\text{savg})$	0.75	0.43	7/11	14/17
$\ln(\text{ASM}) + \sqrt{\text{cont}} + \text{avg}^2$	0.71	0.36	5/11	15/17
$\ln(\text{ASM}) + \ln(\text{dvar}) + \ln(\text{avg})$	0.71	0.34	4/11	16/17
$\text{tex_var} + \text{sentro}^2 + \text{avg}^2$	0.71	0.34	4/11	16/17
$\text{svar}^2 + \text{sentro}^2 + \text{dvar}^2$	0.71	0.34	4/11	16/17
$\text{svar}^2 + \text{sentro}^2 + \sqrt{\text{dent}}$	0.71	0.34	4/11	16/17
$\ln(\text{ASM}) + \text{cont}^2 + \text{svar}$	0.71	0.31	3/11	17/17
$\sqrt{\text{avg}} + \text{std}^2 + \text{sentro}^2$	0.68	0.29	5/11	14/17
$\text{tex_var} + \text{svar}^2 + \text{std}^2$	0.68	0.27	4/11	15/17
Point cloud metrics				
$\sqrt{f_{\text{veg5}}} + \sqrt{f_{\text{p30}}} + l_{\text{p50}}^2$	0.79	0.54	7/11	15/17
$\ln(f_{\text{havg}}) + f_{\text{p20}} + l_{\text{p40}}^2$	0.79	0.54	7/11	15/17
$\sqrt{f_{\text{veg5}}} + f_{\text{p70}}^2 + \ln(l_{\text{veg5}})$	0.75	0.47	7/11	14/17
$\ln(f_{\text{veg5}}) + \sqrt{f_{\text{p20}}} + l_{\text{p40}}^2$	0.75	0.47	7/11	14/17
$\ln(f_{\text{veg5}}) + f_{\text{p60}} + 1/(l_{\text{veg5}})$	0.75	0.47	7/11	14/17
$\sqrt{f_{\text{havg}}} + \ln(f_{\text{p20}}) + \sqrt{l_{\text{p30}}}$	0.75	0.47	7/11	14/17
$1/(f_{\text{havg}}) + f_{\text{p30}} + l_{\text{p50}}^2$	0.75	0.47	7/11	14/17
$f_{\text{hstd}} + f_{\text{p10}} + l_{\text{p30}}^2$	0.75	0.47	7/11	14/17
$f_{\text{hstd}} + \sqrt{f_{\text{p20}}} + l_{\text{p40}}^2$	0.75	0.47	7/11	14/17
$f_{\text{hstd}} + \sqrt{f_{\text{p5}}} + \sqrt{l_{\text{havg}}}$	0.71	0.38	6/11	14/17

Table 3. The ten most accurate classification models based on metrics from mixed group of landscape, texture and point cloud metrics, the overall accuracies and the kappa-values of the classifications and the number of correctly classified microstands. The order of the variables does not indicate the relative importance of the metrics. * = Model statistically significant at the level $p=0.05$.

Metrics	Overall accuracy	Kappa-value	Correct, clustered	Correct, random/regular
$\ln(GP_{density}) + 1/(TP_{ave}) + P_{sd}$	0.89 *	0.77	9/11	16/17
$(P_{sd})^2 + avg + l_{p50}$	0.86 *	0.71	10/11	14/17
$(P_{sd})^2 + \ln(GP_{density}) + 1/(f_{veg5})$	0.86 *	0.69	8/11	16/17
$(GP_{density})^2 + (P_{sd})^2 + idm$	0.82 *	0.63	9/11	14/17
$\ln(GP_{density}) + (P_{sd})^2 + f_{p30}$	0.82 *	0.62	8/11	15/17
$(P_{sd})^2 + \ln(ASM) + \ln(l_{p50})$	0.79 *	0.56	9/11	13/17
$(P_{sd})^2 + f_{veg5} + 1/(l_{p50})$	0.79 *	0.55	8/11	14/17
$P_{sd} + 1/(idm) + \ln(l_{p50})$	0.79 *	0.55	8/11	14/17
$(GP_{density})^2 + (P_{sd})^2 + ASM$	0.79 *	0.55	8/11	14/17
$GP_{density} + (P_{sd})^2 + avg$	0.79 *	0.55	8/11	14/17

3.2 The best combinations of the metrics

The three best selected landscape metrics were $GP_{density}$, TP_{ave} and the P_{sd} (ignoring the transformations). All three metrics gave generally lower values for the microstands which were clustered. The three best selected texture metrics were ASM , idm and avg and all they gave generally slightly lower values for the microstands which were clustered. The three best selected point cloud metrics were f_{veg5} , f_{p30} and l_{p50} . Metric f_{veg5} gave generally slightly lower and f_{p30} and l_{p50} slightly higher values for the microstands which were clustered.

In the classification of spatial pattern of trees the most accurate model based on the selected metrics from three variable classes was the same as most accurate model based on only the landscape metrics (Table 3). The predictor variables were: the number of the ground patches per hectare in logarithmic scale ($\ln(GP_{density})$), inverse of the average size of the tree patches ($1/(TP_{ave})$) and the standard deviations of the size of all patches (P_{sd}). In the ten most accurate models the most common metrics was P_{sd} (ignoring the transformations), which was found as predictor in all models. Also $GP_{density}$ and l_{p50} were quite common (Table 3).

4. Discussion and conclusion

The aim of this study was to identify ALS based metrics which can be used to determine the spatial pattern of trees. In this study the classification of microstands was highly accurate and we found potential horizontal ALS metrics to determine the spatial pattern of trees. However, unfortunately we were not able to test our models with independent test data, and our data were quite limited (only 28 microstands).

Previous studies on the determination of the spatial pattern of trees have used remote sensing imagery and the ITD method with high density ALS. For example, Uuttera et al. (1998) found that if the scale is equal to or less than 1:5000, aerial photographs cannot be used to accurately determine the spatial pattern of trees. Coops and Culvenor (2000) have indicated that it is possible to derive the spatial pattern of trees within a high spatial resolution forest scene provided that the crown size is estimated a priori. Additionally, they pointed out that the total foliage cover of the canopy will affect the ability to predict spatial distribution based on texture variance. Furthermore, in the study by Mustonen (2002) single tree detection with ALS provided correct spatial pattern for the majority of the plots, excluding the clustered plots. Due to differences in the study materials and the methods applied, the results of these previous studies cannot be directly compared to this study.

Especially the landscape metrics had significant role in the identification of the spatial pattern of trees. For example, in the most accurate model all the predictors were landscape metrics: the number of the ground patches per hectare in logarithmic scale ($\ln(GP_{density})$), inverse of the average size of the tree patches ($1/(TP_{ave})$) and the standard deviations of the size of all patches (P_{sd}). Additionally, in each of the ten most accurate models based on metrics from mixed group was as at least one landscape metric.

The more detailed analysis of the best landscape metrics shows that there is lower number of ground patches per hectare ($GP_{density}$), smaller average size of the tree patches (TP_{ave}) and lower standard deviation of the size of all patches (P_{sd}) in clustered microstands compared to random/regular microstands. This suggested that clustered microstands consisted of ground pixels, which are forming not many but large and continuous ground patches and tree pixels, which are forming tree patches (group of trees) surrounded by the ground patches. On the other hand, random/regular microstands consisted of large continuous tree patches and small ground patches. The standard deviation of the size of all patches is higher in random/regular microstands, because in random/regular microstands the tree patches are large and the ground patches small, but in clustered microstands the ground patches are quite large and the tree patches are not very small, because they consist on groups of trees.

Also the results based on the texture metrics and point cloud metrics were quite accurate, but the interpretations of these metrics are not very clear. Three best texture metrics (ASM , idm and avg) gave generally slightly lower values for microstands which were clustered, which means that random/regular microstands are more orderly (ASM is higher, when window is very orderly), homogeneous (idm is higher when the values of grey-level co-occurrence matrix have attended to diagonal) and have higher average of grey-levels. Reason for this could be that random/regular microstands consist of large continuous tree-areas. From point cloud metrics f_{veg5} gave lower values in clustered microstands, which suggests that more ALS hits lay near to the ground than in random/regular microstands. The f_{p30} and l_{p50} gave higher values in clustered microstands, which mean that more ALS hits lay under the 30 % or 50 % of the cumulative heights of vegetation. This might suggest that there is more gap-area in the canopy in clustered microstands and in the random/regular microstands most of the ALS hits lay up to the canopy.

We showed a first attempt to determine the spatial pattern of trees and the need for first thinning based on the use of horizontal ALS metrics (textural and landscape metrics), which in the context of ALS are still quite rare. This is despite the fact that there exist wide possibilities for their use, for example, in studies concerning tree competition, tree species recognition, and site classification. In this study we didn't use the height percentiles of point cloud metrics because there is no theoretical base for that height would affect to the spatial pattern of trees. Our results were promising in regard to the identification of the spatial pattern of trees with low pulse density ALS data. In our study material the microstands consisted of rather young and equally sized trees of same tree layer which may have had effect to the results. However, the approach used should still be studied in more detail in order to generalize the results with larger sample size and especially to find out which are the best ALS metrics for identifying all the three spatial patterns of trees. The information regarding the spatial pattern of trees could be further applied in forest planning including growth simulations and optimization of the timing of forest treatments.

References

- Bartlett, M.S., 1937. Properties of sufficiency and statistical test. *Proceedings the Royal of Society A*, 160, 268-282.
- Besag, J. and Cleaves, J.T., 1973. On the detection of spatial pattern in plant communities. *Bulletin of the International Statistical communities*, 45, 153-158.
- Coops, N. and Culvedor, D., 2000. Utilizing local variance of simulated high spatial resolution imagery to predict spatial pattern of forest stands. *Remote Sensing of Environment*, 71, 248-260.
- Diggle, P.J., 1983. Statistical Analysis of Spatial Point Patterns. Mathematics in Biology. *Academic Press*. London: 148 p.
- Gavrikov, V. and Stoyan, D., 1995. The use of marked point processes in ecological and environmental forest studies. *Environmental and Ecological Statistic*, 2, 331-344.
- Haralick, R.M., Shanmugam, K. and Dinstein, J., 1973. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3, 610-621.
- Hargis, C.D., Bissonette, J.A. and David, J.L., 1998. The behavior of landscape metrics commonly used in the study of habitat fragmentation. *Landscape Ecology*, 13, 167-186.
- Kilkki, P., Pohjola, T. and Pohtila, E., 1985. Puiden ryhmittäisyyden huomioonottaminen harvennussmalleissa. *Silva Fennica*, 19, 137-143. (In Finnish.)
- Kouki, J., Löfman, S., Martikainen, P., Rouvinen, S. and Uotila, A., 2001. Forest fragmentation in Fennoscandia: Linking habitat requirements of wood-associated threatened species to landscape and habitat changes. *Scandinavian Journal of Forest Research*, 3, 27-37.
- Lebart, L., Morineau, A. and Warwick, K.M., 1984. Multivariate Descriptive Statistical Analysis: Corresponding Analysis and Related Techniques for Large Matrices. *John Wiley & Sons*. New York.
- Mardia, K.V., Kent, J.T. and Bibby, J.M., 1979. Multivariate Analysis. London-New York-Toronto-Sydney-San Francisco. *Academic Press 1979*, XV: 521 p
- McGarigal, K. and Marks, B.J., 1995. FRAGSTATS: Spatial pattern analysis program for quantifying landscape structure. *General Technical Report PNW*. Pacific Northwest Research Station: 351 p.
- Mustonen, K., 2002. Puiden tilajärjestyksen ja pituuden määrittäminen laserkeilankuvilta. Metsäsuunnittelun ja -ekonomian pro-gradu. *University of Joensuu*, 55 p. (In Finnish.)
- Packalén, P. and Maltamo, M., 2006. Predicting the plot volume by tree species using airborne laser scanning and aerial photograph. *Remote Sensing of Environment*, 109, 328-341.
- Pielou, E.C., 1960. A single mechanism to account for regular, random and aggregated populations. *Journal of Ecology*, 49, 574-584.
- Pukkala, T., 1990. A method for incorporating the within-stand variation into forest management planning. *Scandinavian Journal of Forest Research*, 5, 263-275.

- Pukkala, T., Miina, J., Kurttila, M. and Kolström, T., 1998. A spatial yield model for optimizing the thinning regime of mixed stands of *Pinus sylvestris* and *Picea Abies*. *Scandinavian Journal of Forest Research*, 5, 263-275.
- Rosenfield, G.H. and Fitzpatrick-Lins, K., 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 52, 223-227.
- Sachs, D.L., Sollins, P. and Cohen, W.B., 1998. Detecting landscape changes in the interior of British Columbia from 1975 to 1992 using satellite imagery. *Canadian Journal of Forest Research*, 28, 23-36.
- Shang, D. and Waite, M.-L., 1991. Utilization of textural and spectral features from Landsat TM imagery for stratifying forest sample plots. *University of Helsinki, Department of Forest Mensuration and Management, Research Notes*, 26.
- Tomppo, E., 1986. Models and methods for analysing spatial pattern of trees. *Communicationes Instituti Forestalis Fenniae*, 138, 65 p.
- Tuominen, S. and Pekkarinen, A., 2005. Performance of different spectral and textural aerial photograph features in multi-source forest inventory. *Remote Sensing of Environment*, 94, 256-268.
- Uutera, J., Haara, A., Tokola, T. and Maltamo, M., 1998. Determination of the spatial distribution of trees from digital aerial photographs. *Forest Ecology and Management*, 110, 257-282.
- Wulder, M.A., LeDrew, E.F., Franklin, S.E. and Lavigne, M.B., 1998. Aerial image texture information in the estimation of northern deciduous and mixed wood forest leaf area index (LAI). *Remote Sensing of Environment*, 64, 64-76.

Comparison of Discrete Return and Waveform Airborne Lidar Derived Estimates of Fractional Cover in an Australian Savanna

J. Armston^{1,2}, M. Disney³, P. Lewis³, P. Scarth^{1,2}
P. Bunting⁴, R. Lucas⁴, S. Phinn¹, N. Goodwin²

¹ Joint Remote Sensing Research Program, Centre for Spatial Environmental Research, School of Geography, Planning and Environmental Management, University of Queensland, j.armston@uq.edu.au

² Remote Sensing Centre, Environment and Resource Sciences, Queensland Department of Environment and Resource Management

³ Department of Geography, University College London, UK

⁴ Institute of Geography and Earth Sciences, University of Aberystwyth, Wales, UK

Abstract

The advance of commercial airborne lidar systems from discrete-return to waveform recording instruments has made repeatable estimates of biophysical variables from these different methods questionable. Using an experimental airborne waveform lidar dataset acquired in an Australian savanna, this study presents a method for the derivation of canopy/ground backscatter coefficients from waveform lidar and a comparison of discrete return and waveform approaches to the estimation of fractional cover. Despite limited validation, the results indicate that waveform estimates of fractional cover can provide consistently higher accuracy than discrete return estimates under varying survey properties. Ongoing work using raw waveform data across larger areas and 3D radiative transfer simulations aims to develop a quantitative understanding of the impact of disparate sensor and survey properties on the detection of change in vegetation structure using commercial lidar instruments.

1. Introduction

The fractional cover of woody vegetation, defined as one minus the gap fraction at a zenith angle of zero degrees, is an important description of plant canopy structure in the biophysical remote sensing and ecological literature. Crown cover projection (CCP), foliage projective cover (FPC) and leaf area index (LAI) are all a function of the spatial and angular distribution of gaps and are common canopy metrics used in mapping, monitoring and modelling applications by natural resource management agencies in Australia (Scarth *et al.*, 2008). Field-based estimates of canopy metrics are difficult to acquire consistently over large areas and there is often insufficient resources to accurately capture the spatial and temporal variability in structure, especially for scaling up to satellite remote sensing (Asner, 2009). Airborne lidar has shown potential to enable scaling between *in situ* and satellite monitoring of vegetation change (Asner, 2009; Lucas *et al.*, 2010) and also provide a viable alternative to traditional field techniques (Armston *et al.*, 2009).

Height and directional gap fraction (one minus the fractional cover) are the only canopy metrics that can be directly retrieved from airborne lidar measurements, with other canopy metrics and above-ground biomass subsequently modelled using different expressions, combinations or spatial variance of these parameters (e.g., Ni Meister *et al.*, 2010). Error is introduced in the estimation of canopy metrics from airborne lidar when empirical methods are applied to

different regions, sensor or survey properties than for which they were developed (Næsset, 2009). As a result, many published empirical relationships between field and lidar estimates of vegetation cover have limited wider application unless recalibrated using field measurements (e.g., Solberg *et al.*, 2010; Armston *et al.*, 2009; Rosette *et al.*, 2009). The rapid advance of commercial lidar sensor technology from single discrete return to waveform recording systems has resulted in often disparate lidar datasets over time and exacerbated this problem. Long-term monitoring programs that employ lidar derived metrics, from often disparate sensor and survey configurations, need to determine if these can replicate the same change observed over space and time using direct field methods. Waveform lidar is required to validate our understanding of the impact of sensor and survey properties, especially for discrete return datasets (Disney *et al.*, 2010; Næsset, 2009).

Methods for estimating fractional cover and related canopy metrics have been developed for discrete return and waveform lidar (Lovell *et al.*, 2003; Ni-Meiseter *et al.*, 2001). However no studies have demonstrated an improvement in fractional cover estimates derived using waveform data in Australian vegetation communities dominated by eucalypt tree species. Discrete return sensors typically only record intercepts, which can be interpreted as a binary measure of signal intensity. Therefore estimates of fractional cover do not account for gaps smaller than the size of the footprint (Liu *et al.*, 2008) and only a limited number of returns above a noise threshold can be recorded. Waveform sensors digitize the entire return signal at a particular temporal sampling interval so do not suffer the same limitations. However estimates of fractional cover are sensitive to canopy/ground reflectivity and therefore the wavelength of the sensor (Ni-Meister *et al.*, 2001). Estimates of fractional cover from both classes of lidar require ground and canopy returns to be separated, which is sensitive to lidar sensor and survey properties and their interaction with canopy structure. An assessment of the relative importance of these differences on estimates of fractional cover in Australian environments is lacking.

The objective of this study was to compare waveform and discrete return estimates of fractional cover for a savanna woodland in northern Queensland, Australia. This paper presents an initial investigation of these objectives through empirical analysis of data currently available from an experimental RIEGL LMS-Q680 waveform lidar survey acquired over an existing monitoring site with repeat field and lidar measurements. These data were used to simulate coincident waveform and discrete return datasets for comparison.

2. Data and Methods

2.1 Study site

The study site is located near Charters Towers in northern Queensland, Australia, and is within the Einasleigh Uplands region at approximately 400 m elevation. This is a region of savanna and woodlands and is subject to livestock grazing. This study utilised three structurally contrasting savanna open woodland field plots (CHAT0101, CHAT0102, CHAT0103; Figure 1) that are part of a larger network of monitoring sites in Queensland for calibration and validation of Landsat-derived woody and herbaceous fractional cover products (e.g., Armston *et al.*, 2009).

The woodlands at CHAT0101 (20.0047°S, 145.6224°E) were dominated by *Eucalyptus drepanophylla* with *Corymbia dallachiana* sub-dominant in the 12–20 m height canopy. *Petalostigma pubescens* and *Maytenus cunninghamii* are also occasionally present in the understorey (3–7 m height). Within CHAT0102 (19.9796°S, 145.6490°E), *Eucalyptus melanophloia* dominated the sparse canopy (8–19 m height), with *Corymbia setosa* and *E. melanophloia* also present. *P. pubescens* dominates a higher density understorey compared to the other two sites. The canopy at the CHAT0103 (20.0230°S, 145.6029°E) site was very sparse

with *Eucalyptus brownii* forming the overstorey and the occasional *Acacia salicina* and *Acacia farnesiana* in the understorey. CHAT0101 and CHAT0102 were located on sand plains with relatively uniform grass cover. CHAT0103 was located on basalt plains with occasional surface basalt boulders and grey to black cracking soils. The grass cover is clumped at CHAT0103 with large areas of bare soil exposed. The terrain at all three sites was flat.



Figure 1: CHAT0101 (left), CHAT0102 (centre) and CHAT0103 (right) near the time of the airborne waveform lidar surveys.

2.2 Field and lidar surveys

The lidar surveys used in this study were acquired on two different dates as shown in Table 1. A RIEGL LMS-Q680 waveform lidar survey was acquired on the 18th June 2010 to quasi-simultaneously capture a range of survey properties (A2–A4). In consultation with the data provider, the A2–A4 survey properties were designed to capture a range of sensor and survey configurations within limits recommended by RIEGL for instrument operation over vegetation. Parallel flight tracks were designed to have 60% overlap at each altitude to ensure a multi-angular airborne dataset over the field sites. Multiple flying heights were designed to capture the changing footprint size and signal to noise level of received waveforms due to the inverse square loss of power per unit area with range. Only the centre flight track at each nominal altitude was used in this study, as directly measured fractional cover validation data was only available at a zenith angle of zero degrees. A RIEGL LMS-Q560 survey was also acquired on the 3rd November 2008 as part of ongoing monitoring at the study site (A1). This survey was acquired with the same centre flight track as the A2–A4 surveys.

A range of coincident measurements were collected at the three 100 m × 100 m field sites. However for the present study, estimates of fractional cover at a nominal zenith of zero were directly measured using three 100 m point intercept (1 m spacing) transects oriented 0°, 60° and 120° from magnetic north. At each 1 m interval along each transect, vertical intercepts were recorded from the overstorey (woody plants greater than or equal to 2 m height) and the understorey (woody or herbaceous plants less than 2 m height) using a GRS densitometer TM. This instrument employed a mirror, two bubble-line levels and a centred cross-hair to project an exact vertical line-of-sight from the sample point in the canopy to the observer. Fractional cover was then calculated as the fraction of observations that were overstorey leaf or wood intercepts.

Table 1: Survey properties for the RIEGL airborne waveform lidar datasets used in this study.

Survey	A1	A2	A3	A4
Acquisition date	2008/11/03	2010/06/18	2010/06/18	2010/06/18
RIEGL sensor	LMS-Q560	LMS-Q680	LMS-Q680	LMS-Q680
Nominal altitude (m)	600	450	900	1200
No. parallel flight lines	1	3	3	3
Swath width (m)	497	520	1039	1386
Pulse rate (kHz)	100	200	200	150
Scan rate (Hz)	60	113	80	63
Pulse density (pulse / m ²)	2.65	3.50	2.10	1.11
Footprint diameter (m)	0.30	0.23	0.45	0.60
Maximum zenith angle (°)	22.5	30.0	30.0	30.0

2.3 Processing lidar waveforms and discrete returns

At the time of this work, the data provider was unable to deliver waveform data for the RIEGL LMS-Q680 surveys. GPS time, Cartesian coordinates (easting, northing, elevation), Gaussian parameters (range, amplitude and width) and pulse parameters (scan zenith, range) produced for each return using the RIEGL RiAnalyze software were available. The Cartesian coordinates were input to a modified version of the progressive morphological filter by Zhang *et al.* (2003) to classify ground and non-ground returns. Discrete return datasets were generated using the Cartesian coordinates of the Gaussian peaks above the noise threshold. Pulses from each acquisition were randomly sampled to a density of 1 pulse / m² to avoid any potential differences in the height distribution of returns.

By assuming the Gaussian model captures the shape of the received waveforms ($I(t)$), as outlined by Wagner *et al.* (2006), the waveforms were reconstructed using the Gaussian parameters. Limited validation of the Gaussian model for the study site was performed using raw waveform data from the A1 survey (Table 1) that was only recently made available without matching RiAnalyze products. Estimates of the Gaussian parameters were derived using non-linear least-squares fitting (Levenberg–Marquardt method) to Equation 1:

$$I(t) = \sum_{i=1}^N A_i e^{-\frac{(t-t_i)^2}{2s_i^2}} \quad (1)$$

where for each return, σ is the noise level, A_i is Gaussian amplitude, t_i is the time (or range) and s_i is the Gaussian standard deviation. Starting parameters were determined from the zero-crossings of the waveform first derivative that were above σ . False returns due to “ringing” in the transmitted pulse were omitted from the starting parameters if their amplitude was less than σ plus the value from an exponential time decay function on the amplitude of earlier local maximums. The σ parameter was set to the same default constant value of 9 used by RiAnalyze (Matthew McCauley, Atlass Pty. Ltd., *pers. comm.*). It is also important to note that several details on the Gaussian decomposition procedure performed by RiAnalyze are proprietary knowledge and therefore exact replication was unlikely.

Examples of the transmitted and received waveforms and corresponding Gaussian model fits are

shown in Figure 2. The gain for the transmitted waveform is unknown and different from the received and therefore cannot be directly used to calibrate waveforms to apparent reflectance. The RIEGL LMS-Q560 and LMS-Q680 instruments record waveform samples in 60 ns sample blocks, with recording of blocks initialised by the signal exceeding a noise threshold. The missing data between 2.5 and 6 m in the received waveform was the result of this “dead time” (RIEGL, *pers. comm.*). For direct comparison of the ground (I_g) and canopy (I_v) components of the raw and Gaussian model received waveforms, their separation was at the first occurrence of the minimum signal between the ground and canopy Gaussian model peaks.

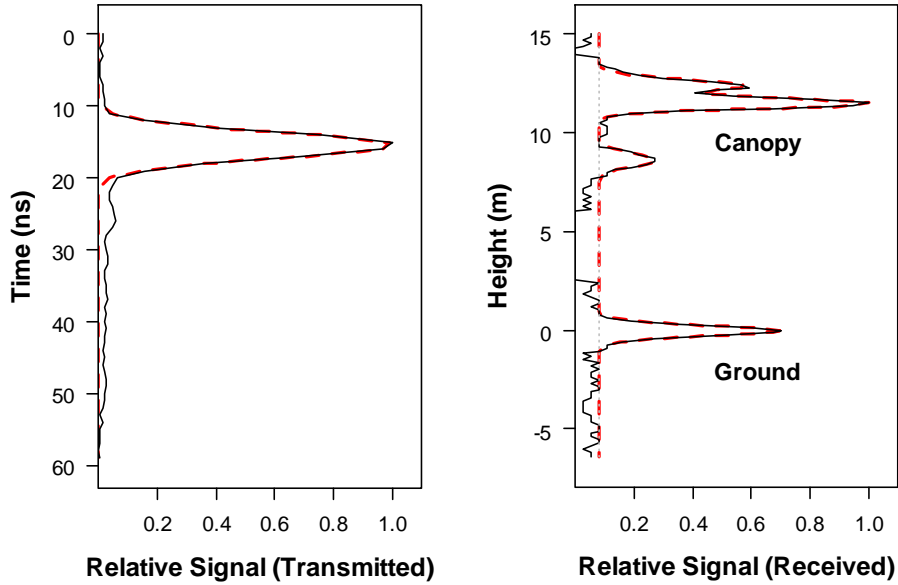


Figure 2: Measured (thin black line) and Gaussian model (dashed red line) transmitted (left) and received (right) waveforms from a single pulse over the CHAT0101 field plot. The waveforms are normalized by the maximum signal and the noise level is shown (dotted grey line). Three canopy returns were derived from the received waveform with peaks at heights of 8.55, 11.44 and 12.36 m above the ground.

2.4 Estimation of fractional cover

Studies that have used discrete return lidar sensors typically estimate fractional cover as the proportion of returns intercepted by the canopy within a data bin (Lovell *et al.*, 2003):

$$P_{cover}(z) = \frac{\sum_{z=i}^{z=\max(z)} C_{v,i}}{N} \quad (2)$$

where C_v is the number of returns from the top of canopy down to height z and N is the total number of canopy and ground returns. We assume here that all pulses are at a zenith of zero degrees. Equation 2 was applied to two categories of discrete returns often used to calculate fractional cover: first returns only (D1) and all returns (D2).

Waveform recording instruments digitise the received waveform and, assuming single scattering only, can be separated into vegetation and ground backscatter components:

$$I = I_g + I_v \quad (3)$$

where I_v is the integrated vegetation backscatter component of the waveform and I_g the integrated ground return. Disney *et al.* (2006) showed greater than 80% of reflectance at the nadir hotspot is single scattered photons and since green foliage typically has low reflectance at 1550 nm, it is reasonable to assume negligible multiple scattering. Following Ni-Meister *et al.* (2001) and assuming the recorded lidar signal is linearly related to the receiver power, fractional cover can then be estimated from uncalibrated waveforms by:

$$P_{cover}(z) = \frac{I_v(z)}{I_v(0) + I_g \frac{v}{g}} \quad (4)$$

where I_v is the integrated vegetation backscatter component of the waveform from the top of the canopy down to height z . Since the returned backscatter is a function of the apparent reflectance, phase function and leaf angle distribution of canopy and ground targets as well as the gap fraction, we need some prior knowledge of the canopy/ground backscatter coefficient ratio v/g to obtain unbiased estimates of fractional cover from waveform data. Previous studies have suggested using a constant (e.g., 2 for 1064 nm; Lefsky *et al.*, 1999), however canopy and ground properties often change between stands. Therefore a method to derive v/g from waveforms is required in order to avoid field calibration, especially in savannas where estimates of fractional cover will be most sensitive to v/g due to sparse canopies and variable backgrounds.

The canopy backscatter component of the waveform is the product of the fractional cover at the zenith of the lidar pulse and the volumetric backscattering coefficient of the canopy (μ_v), and the ground component is the product of the gap fraction and the volumetric backscattering coefficient of the ground (μ_g):

$$I_v = P_{cover}(0) \mu_v \quad (5)$$

$$I_g = [1 - P_{cover}(0)] \mu_g \quad (6)$$

where μ_g and μ_v are a function of the foliage angle distribution and apparent reflectance at the wavelength of the lidar. Combining Equations 5 and 6, the relationship between I_g and I_v is then linear:

$$I_g = \mu_g - \frac{\mu_g}{\mu_v} I_v \quad (7)$$

If constant μ_g and μ_v is assumed to extend to N pulses within a local area, estimation of μ_g and $-\mu_g/\mu_v$ is then a simple linear regression problem ($y = a + bx$). An estimate of μ_g is the intercept (a) and μ_v can be estimated from the slope (b) by $-\mu_g/b$. Fractional cover is then estimated directly using Equation 4. Two waveform estimates of fractional cover are evaluated in this study: v/g derived using the method just described (W1) and with v/g set to a constant of 0.5 for 1550 nm (W2).

3. Results and Discussion

3.1 Accuracy of the Gaussian model

A very close correspondence exists between the measured and Gaussian model estimates of I_g and I_v (RMSE < 6; Figure 3). The error estimates were consistently higher for I_v than I_g , but the difference is negligible. Since the canopy was sparse and the terrain flat, a large percentage of the pulses were single returns from the ground, from which returns that correspond to a Gaussian shape are to be expected. The canopy was composed of clumped scattering elements with variable leaf angle distributions, which resulted in more complex waveforms. Poor fits to a number of individual waveforms with overlapping returns <~60 cm apart (twice the temporal bin spacing of the waveform) have also been observed. However these results provide evidence that Gaussian modelling of RIEGL waveforms in a savanna environment is able to statistically reproduce received waveforms for a large number of pulses.

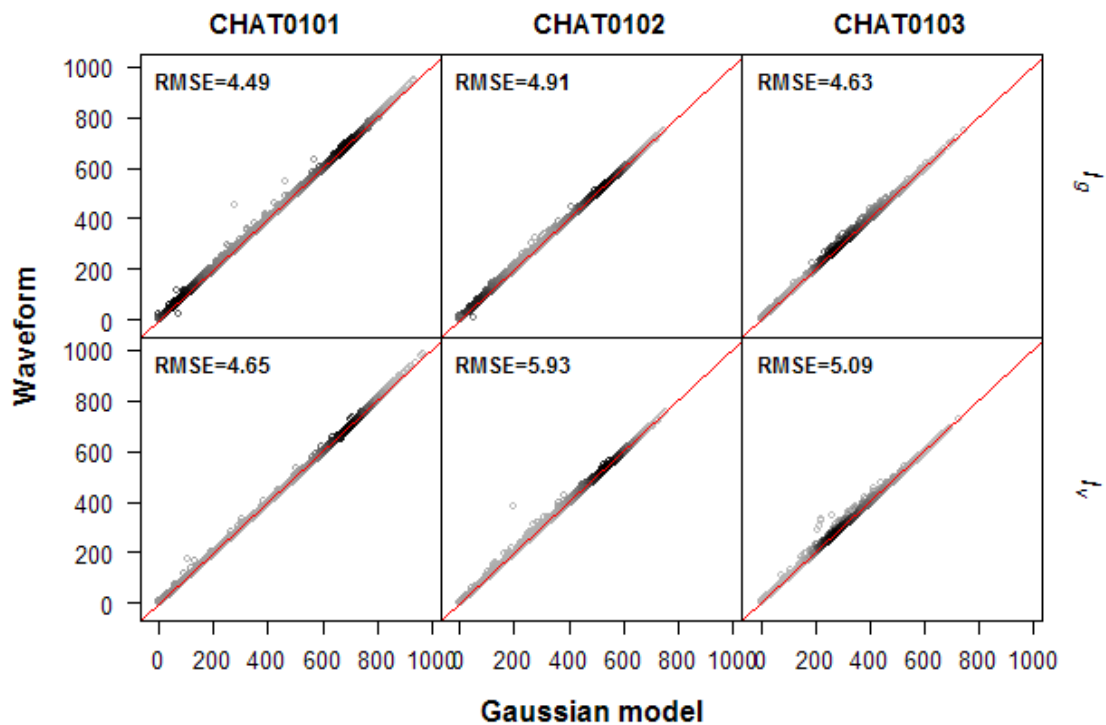


Figure 3: Comparison of waveform and Gaussian model calculated I_g and I_v . Darker regions indicate a higher density of observations. The red line is the 1:1 line.

The advantages of representing the waveform in this way are: raw waveform data are not typically available for most natural resource management agencies; data volumes are greatly reduced; noise is less of an issue when interpreting reconstructed waveforms; and Gaussian distributions have well known properties. The disadvantage is that not all returns necessarily have a Gaussian shape, with this depending on interactions between the shape and duration of the transmitted pulse and vertical canopy structure; and low intensity returns may not be detected (Chauve *et al.*, 2009).

3.2 Relationship between canopy and ground backscatter

The linear relationship between I_g and I_v is shown in Figure 4 for the A2 survey. The small footprint of the RIEGL lidar systems results in high spatial variance in the cross-section but also in the spectral properties of intercepted targets, making robust estimates of v/g difficult. For I_g each received waveform may be backscatter from an individual grass sward or a patch of bare soil between swards. For I_v each received waveform is likely to be composed of highly variable proportions of leaf and woody canopy elements (e.g., dead and foliated branches), which have different spectral properties at 1550 nm for these sites. Jupp and Lovell (2007) also highlight that small footprints result in “speckle” due to high apparent backscatter from individual canopy elements acting as Fresnel reflectors.

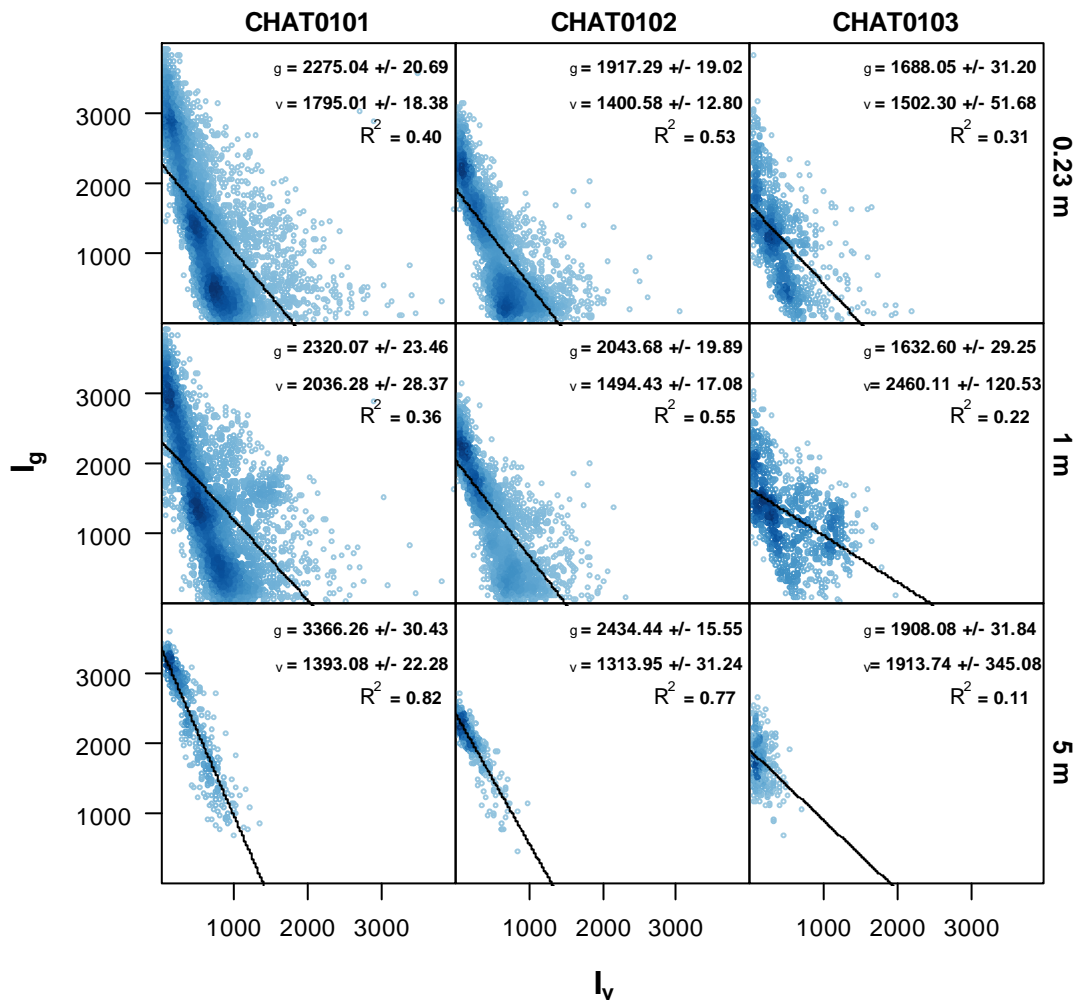


Figure 4: Simple linear regression relationships between I_g and I_v at multiple simulated footprint sizes for the A2 survey at the three field plots. The estimates and uncertainty of v and g derived from this relationship are shown.

To test the effect of footprint size on estimation of v/g , pseudo-waveforms were created at 1 m and 5 m footprint sizes. By integrating all waveforms within a local area and normalizing the signal by the number of pulses to simulate a larger footprint waveform, the distributions of I_g and I_v tend towards unimodal. The smallest deviation from a straight line was at 5 m footprints

for the CHAT0101 and CHAT0102 sites. However the trend did not hold for the CHAT0103 site, as there was little remaining range in I_g and I_v at that footprint size because of the very sparse canopy cover. The spatial heterogeneity in grass cover was high (see CHAT0103 in Figure 1) compared to the uniform grass cover at CHAT0101 and CHAT0102, violating the assumption of constant background. The CHAT0103 site also had dark cracking soils, which resulted in the lowest ρ_g value for all three sites and increased noise. For the next section fractional cover was estimated using ρ_v / ρ_g calculated from the 5 m footprints.

3.3 Effect of survey properties on estimates of fractional cover

Vertical fractional cover profiles and ρ_v / ρ_g derived from each of the lidar survey datasets for each of the three field plots are shown in Figure 5. Fractional cover estimates calculated from discrete returns (D1 and D2) are always higher (differences up to 0.2) than the waveform (W1 and W2) between lidar surveys. There is a trend of decreasing fractional cover with altitude for the discrete return estimates, which is consistent with Morsdorf *et al.* (2008) who used data from a 1560 nm instrument at two altitudes, but in contrast with studies that have reported little or no change in fractional cover using 1064 nm instruments (e.g., Goodwin *et al.*, 2006). Blindness to gaps smaller than the footprint size can lead to overestimation of fractional cover (e.g., Liu *et al.*, 2009), however low intensity returns are more common from canopy elements than the ground which can lead to underestimation of fractional cover if these returns are not separated from noise. The relative importance of these effects depends on wavelength, canopy structure and pulse energy and shape, which are difficult to separate using available measured lidar data.

The calibrated waveforms (W1) appear to be the most consistent for different ranges however ρ_v / ρ_g estimates decrease with increasing range from the A2 to the A4 survey for each site. CHAT0103 is an exception as the fit between I_v and I_g was poor (Figure 4) and the resulting fractional cover too low. The estimation of ρ_v / ρ_g may be compensating for increased noise levels in the waveforms acquired at longer ranges as the number of returns in the canopy detected by the Gaussian decomposition reduces as their amplitude falls below the noise threshold. For example the number of pulses that only have single returns from the canopy at the CHAT0101 site ranges from 38% for the A2 survey (450 m range) to 53% for the A4 survey (1200 m range). This is important as vegetation has low reflectance at 1550 nm so the waveform signal to noise ratio is quite sensitive to pulse energy and range. Another contributing factor could be that the recorded lidar signal is non-linearly related to the receiver power. Other studies that have attempted to calibrate RIEGL sensors have suggested minor differences in loss of received power with range compared to that expected from theory (e.g., Reitberger *et al.*, 2010).

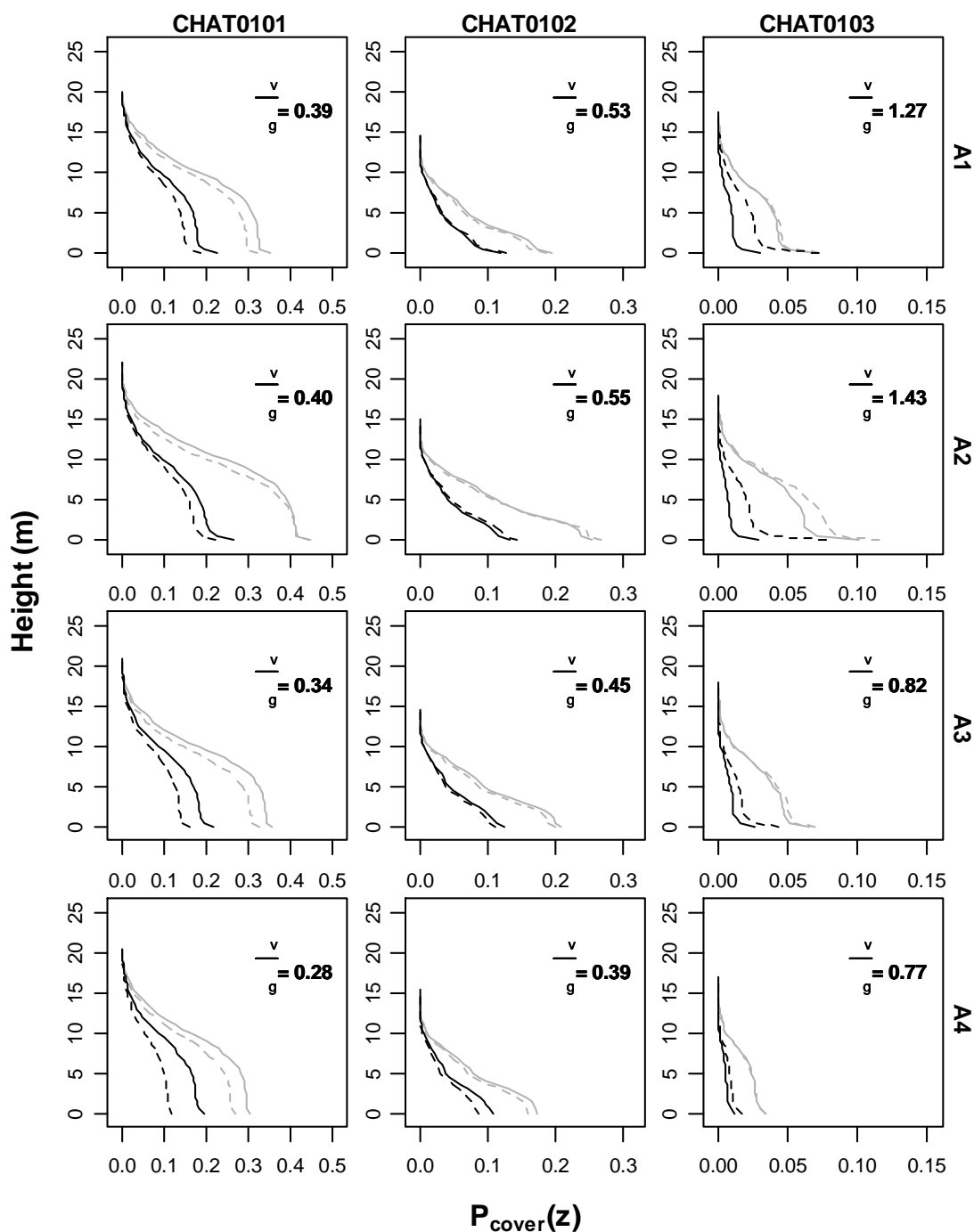


Figure 5: Comparison of discrete return and waveform derived vertical fractional cover profiles for the three field plots at the nominal aircraft altitudes (Table 1). D1 = first returns (grey solid line); D2 = all returns (grey dashed line); W1 = calibrated waveform (solid black line); and W2 = waveform with v/g set to 0.5 (dashed black line). The W1 estimates of v/g are also shown.

The correspondence between field and lidar estimates of fractional cover is shown in Figure 6. The waveform estimates correspond within 5% fractional cover to field measurements compared to within 9% fractional cover for the discrete return estimates. The overestimation exhibited by the discrete return estimates correspond to results shown by other studies using a RIEGL sensor (Miura and Jones, 2010) and discrete return lidar at the same field sites (Armston

et al., 2009). There appears to be less scatter using the estimated v/g rather than the constant of 0.5, however the magnitude of this error falls within that one might expect from binomial sampling error in the field measurements (Armston *et al.*, 2009).

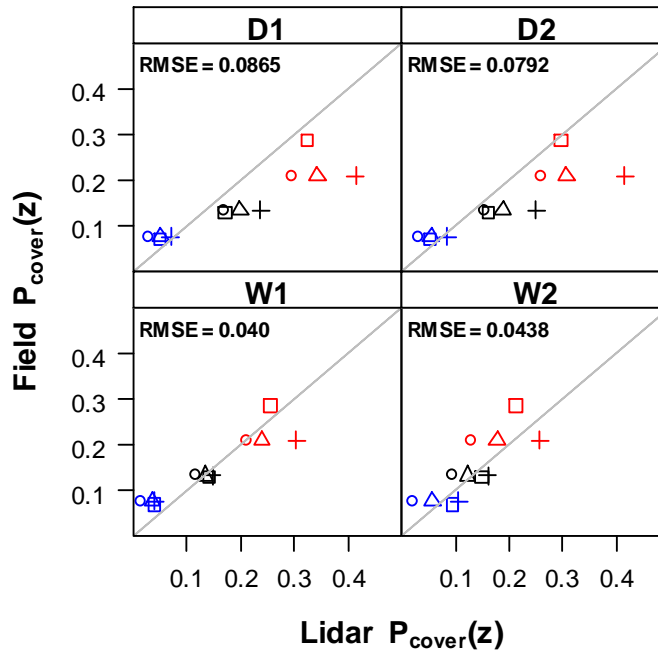


Figure 6: Comparison of fractional cover ($z = 2$ m) estimates derived using discrete return (D1; D2) and waveform (W1; W2) methods. The symbol indicates the source lidar survey (Table 1): square = A1; cross = A2; triangle = A3; and circle = A4. The colour indicates the field site: red = CHAT0101; black = CHAT0102; and blue = CHAT0103. The grey line is the 1:1 line.

4. Conclusions

This study has conducted a preliminary comparison of estimates of fractional cover using field, waveform and discrete return lidar methods. The waveform and discrete return lidar data were simulated using a Gaussian model, which showed a close correspondence with the raw waveforms. Additionally, a method to estimate v/g directly from the waveform data was introduced. Provided the assumptions were satisfied and v/g was estimated with some certainty, waveform methods provided more accurate and consistent estimates of fractional cover than discrete return methods under varying flying heights.

The method to retrieve v/g shows promise, however a quantitative understanding of the limits on retrieval imposed by the interaction between the pulse energy, wavelength, canopy/background heterogeneity and the signal-to-noise of waveforms is required. This will require 3D radiative transfer simulation experiments (e.g., Disney *et al.*, 2010) as measured experimental data are not available to establish the joint sensitivity of v/g and fractional cover estimation to combinations of these sensor and survey properties at a range of fractional cover levels.

Analysis of the raw waveforms is also required as the return detection and Gaussian decomposition procedure used may lead to omission of low backscatter returns. The validation

in this study was confined to a sparsely wooded savanna, however the performance of these methods to estimate fractional cover and derived metrics (FPC, LAI, CCP) are currently being evaluated over approximately fifty monitoring sites from low arid shrubland to coastal rainforest where waveform lidar have been acquired coincident with field and terrestrial lidar surveys.

Improved estimation of the fractional cover of woody vegetation using waveform over discrete return airborne lidar acquired under varying survey properties has important implications for natural resource management agencies employing lidar in large area land cover mapping, monitoring and modelling programs. The results from this study indicate waveform estimates of fractional cover from different flying heights are consistent, enabling comparison of estimates from different surveys over space and time and reducing the need for calibration using field measurements. In future this will lead to a reduction in operational costs as well as uncertainty in reference lidar products used for the calibration and validation of satellite imaging products over large and remote areas in Australia.

Acknowledgements

We would like to thank Robert Denham from the Remote Sensing Centre at the Queensland Department of Environment and Resource Management for assisting with the field work. Michael Schmidt and Christian Witte, also from the Remote Sensing Centre, and two anonymous reviewers are thanked for their comments on the manuscript. We appreciate the efforts of Matthew McCauley from Atlass Pty. Ltd. and Jorg Hacker from Airborne Research Australia to provide raw waveform data.

References

- Armston, J.D., Denham, R.J., Danaher, T.J., Scarth, P.F., and Moffiet, T.N., 2009, Prediction and validation of foliage projective cover from Landsat-5 TM and Landsat-7 ETM+ imagery. *Journal of Applied Remote Sensing*, 3(1), 033540.
- Asner, G.P., 2009. Tropical forest carbon assessment: Integrating satellite and airborne mapping approaches. *Environmental Research Letters*, 4(3), 034009.
- Chauve, A., Vega, C., Durrieu, S., Bretar, F., Allouis, T., Deseilligny, M.P., and Puech, W., 2009. Advanced full-waveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest. *International Journal of Remote Sensing*, 30(19), 5211–5228.
- Disney, M., Kalogirou, V., Lewis, P., Prieto-Blanco, A., Hancock, S., and Pfeifer, M., 2010. Simulating the impact of discrete-return lidar system and survey characteristics over young conifer and broadleaf forests. *Remote Sensing of Environment*, 114(7), 1546–1560.
- Disney, M., Lewis, P., and Saich, P., 2006. 3D modelling of forest canopy structure for remote sensing simulations in the optical and microwave domains. *Remote Sensing of Environment*, 100(1), 114–132.
- Goodwin, N.R., Coops, N.C., and Culvenor, D.S., 2006. Assessment of forest structure with airborne LiDAR and the effects of platform altitude. *Remote Sensing of Environment*, 103: 140–152.
- Jupp, D.L.B., and Lovell, J., 2007. Airborne and ground-based lidar systems for forest measurement: Background and principles. *CSIRO Marine and Atmospheric Research Papers*. CSIRO.

- Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A., and Harding, D., 1999. LiDAR remote sensing of the canopy structure and biophysical properties of douglas–fir western hemlock forests. *Remote Sensing of Environment*, 70: 339–361.
- Liu, J., Melloh, R.A., Woodcock, C.E., Davis, R.E., Painter, T.H., and McKenzie, C., 2008. Modeling the view angle dependence of gap fractions in forest canopies: Implications for mapping fractional snow cover using optical remote sensing. *Journal of Hydrometeorology*, 9(5), 1005–1019.
- Lovell, J.L., Jupp, D.L.B., Culvenor, D.S., and Coops, N.C., 2003. Using airborne and ground–based ranging lidar to measure forest canopy structure in Australian forests. *Canadian Journal of Remote Sensing*, 29(5), 607–622.
- Lucas, R.M., Lee, A.C., Armston, J., Carreiras, J.M.B., Viergever, K., Bunting, P., Clewley, D., Moghaddam, M., Siqueira, P., and Woodhouse, I., 2010. Quantifying Carbon in Wooded Savannas. In M.J. Hill and N.P. Hanan (Eds.), *Measurement and Modelling at Landscape to Global Scales*: Taylor and Francis
- Morsdorf, F., Frey, O., Meier, E., Itten, K.I., and Allgower, B., 2008. Assessment of the influence of flying altitude and scan angle on biophysical vegetation products derived from airborne laser scanning. *International Journal of Remote Sensing*, 29(5), 1387–1406.
- Miura, N. and Jones, S., 2010. Characterizing forest ecological structure using pulse types and heights of airborne laser scanning. *Remote Sensing of Environment*, 114, 1069–1076.
- Næsset, E., 2009. Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, 113(1), 148–159.
- Ni-Meister, W., Jupp, D.L.B., and Dubayah, R., 2001. Modeling LiDAR waveforms in heterogeneous and discrete canopies. *IEEE Transactions on Geoscience and Remote Sensing*, 39(9), 1943–1957.
- Ni-Meister, W., Lee, S., Strahler, A.H., Woodcock, C.E., Schaaf, C., Yao, T., Ranson, J.K., Sun, G., and Blair, J.B., 2010. Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. *Journal of Geophysical Research*, 115, G00E11.
- Reitberger, J., Schnorr, C., Krzystek, P., and Stilla, U., 2009. 3D segmentation of single trees exploiting full waveform LIDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6), 561–574.
- Rosette, J., North, P., Suárez, J. and Armston, J., 2009. Comparison of biophysical parameter retrieval for forestry using airborne and satellite LiDAR. *International Journal of Remote Sensing*, 30(19), 5229–5237.
- Scarth, P., Armston, J., and Danaher, T., 2008. On the relationship between crown cover, foliage projective cover and leaf area index. *Proceedings of the 14th Australasian Remote Sensing and Photogrammetry Conference*, Darwin, NT, Australia.
- Solberg, S., 2010. Mapping gap fraction, LAI and defoliation using various ALS penetration variables. *International Journal of Remote Sensing*, 32(5), 1227–1244.
- Wagner, W., Ullrich, A., Ducic, V., Melzer, T., and Studnicka, N., 2006. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(2), 100–112.
- Zhang, K., Chen, S., Whitman, D., Shyu, M., Yan, J., and Zhang, C., 2003. A progressive morphological filter for removing non-ground measurement of airborne lidar data. *IEEE Transactions on Geoscience and Remote Sensing*, 41(4), 872–882.

Lidar-plots: A new wide-area data collection opportunity

Mike Wulder

Canadian Forest Service

Key collaborators:

Joanne White, Gang Chen, Geordie Hobart – CFS

Nicholas Coops, Christopher Bater, Thomas Hilker – UBC

Christopher Hopkinson, Heather Morrison, Laura Chasmer - NSCC

SilviLaser 2011
Oct 17-20, 2011
Hobart, Australia



Natural Resources
Canada

Ressources naturelles
Canada

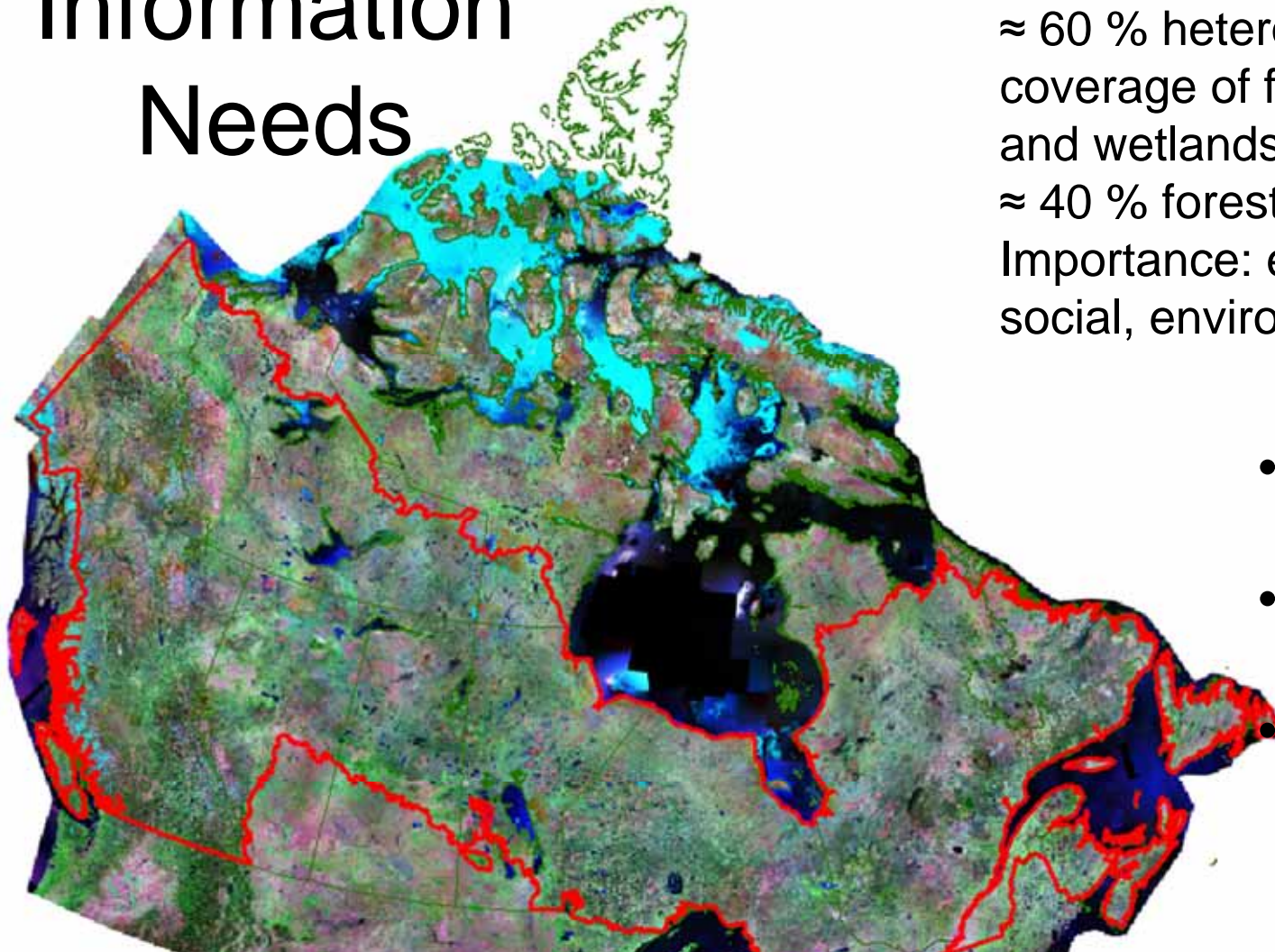
Canada

Abstract

- Forests are an important global resource, playing key environmental and economic roles. Forest monitoring and reporting are typically a national responsibility, with international collaboration and cooperation enabling generation of global statistics. The implementation of quality national monitoring programs is required for the generation of robust national statistics. Programs based upon samples of field plots have proven robust, but are difficult and costly to implement and maintain. Programs with air photo plot based sampling have been developed to mitigate some of the difficulties and limitations with ground plot based programs, especially for large nations. Further, samples of satellite data are becoming increasingly used to produce reliable statistics on forest characteristics and change over large areas. Photo- and satellite-based programs require ground measures to assist in ensuring the quality of attribution undertaken; thereby a conundrum emerges, where ground plots are again desired. To offer a source of detailed data for calibration and validation to large area mapping and monitoring programs we propose the collection and integration of lidar-plots. Light detection and ranging (lidar) has been shown as a data source offering timely and accurate measures of vertical forest structure, relating important information such as canopy height and biomass. Rather than using lidar to produce wall-to-wall coverage we propose the use of transects of scanning lidar to relate the forest conditions present over large areas. Given appropriate sampling statistics can be generated directly from the lidar-plots collected over the transects. In other instances the lidar-plots may be treated as ground plots are typically treated, providing locally relevant information that can be used independently or integrated with other data sources. Issues such as costs and access limitations combine to result in ground plots being endemically in short supply. Lidar-plots are envisioned mitigate this limitation and to produce information to aid and augment monitoring programs and to support science activities.
- In this presentation we further outline the context resulting in the need for detailed forest information over wide areas, we describe the concept for specification, collection, and sharing of lidar-plots; a case study is then presented to illustrate the concept. We conclude with recommendations for alternate implementations, future activities, and improvements to the concept and processes described.
- 30 mins total



Information Needs



9,970,610 km²
≈ 60 % heterogeneous coverage of forests, lakes and wetlands
≈ 40 % forested
Importance: economic, social, environmental

- National
 - Reporting
- Regional
 - Strategic
- Local
 - Operations

**Producing relevant information
across a range of scales**

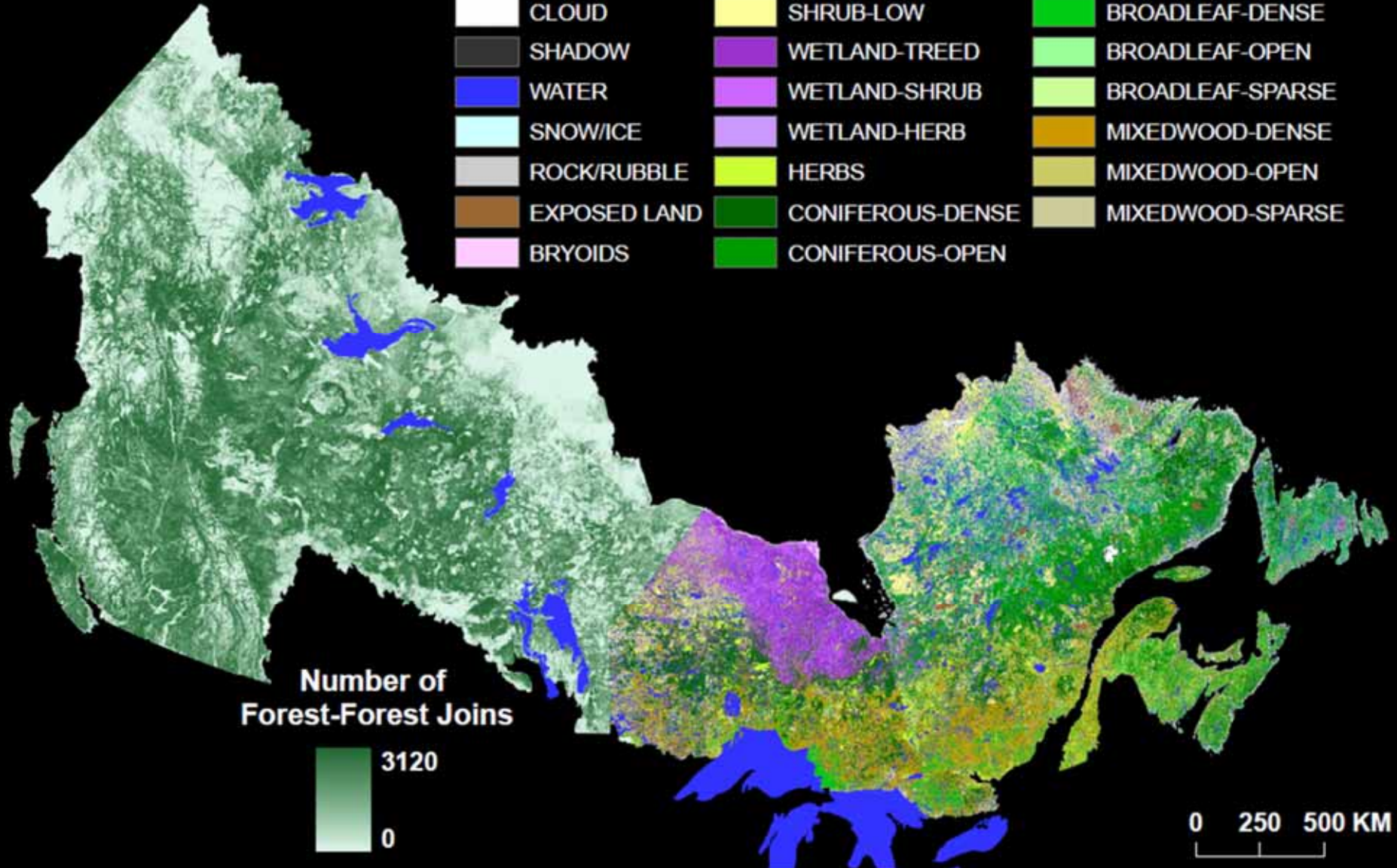


- Information needs:
 - Mapping, monitoring, reporting
 - Inventory type: strategic vs operational
- Intensive vs extensive forestry
- National circumstances

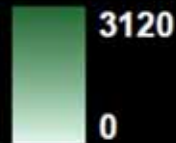
THE FORESTED AREA OF CANADA

EOSD Land Cover Classes

NO DATA	SHRUB-TALL	CONIFEROUS-SPARSE
CLOUD	SHRUB-LOW	BROADLEAF-DENSE
SHADOW	WETLAND-TREED	BROADLEAF-OPEN
WATER	WETLAND-SHRUB	BROADLEAF-SPARSE
SNOW/ICE	WETLAND-HERB	MIXEDWOOD-DENSE
ROCK/RUBBLE	HERBS	MIXEDWOOD-OPEN
EXPOSED LAND	CONIFEROUS-DENSE	MIXEDWOOD-SPARSE
BRYOIDS	CONIFEROUS-OPEN	

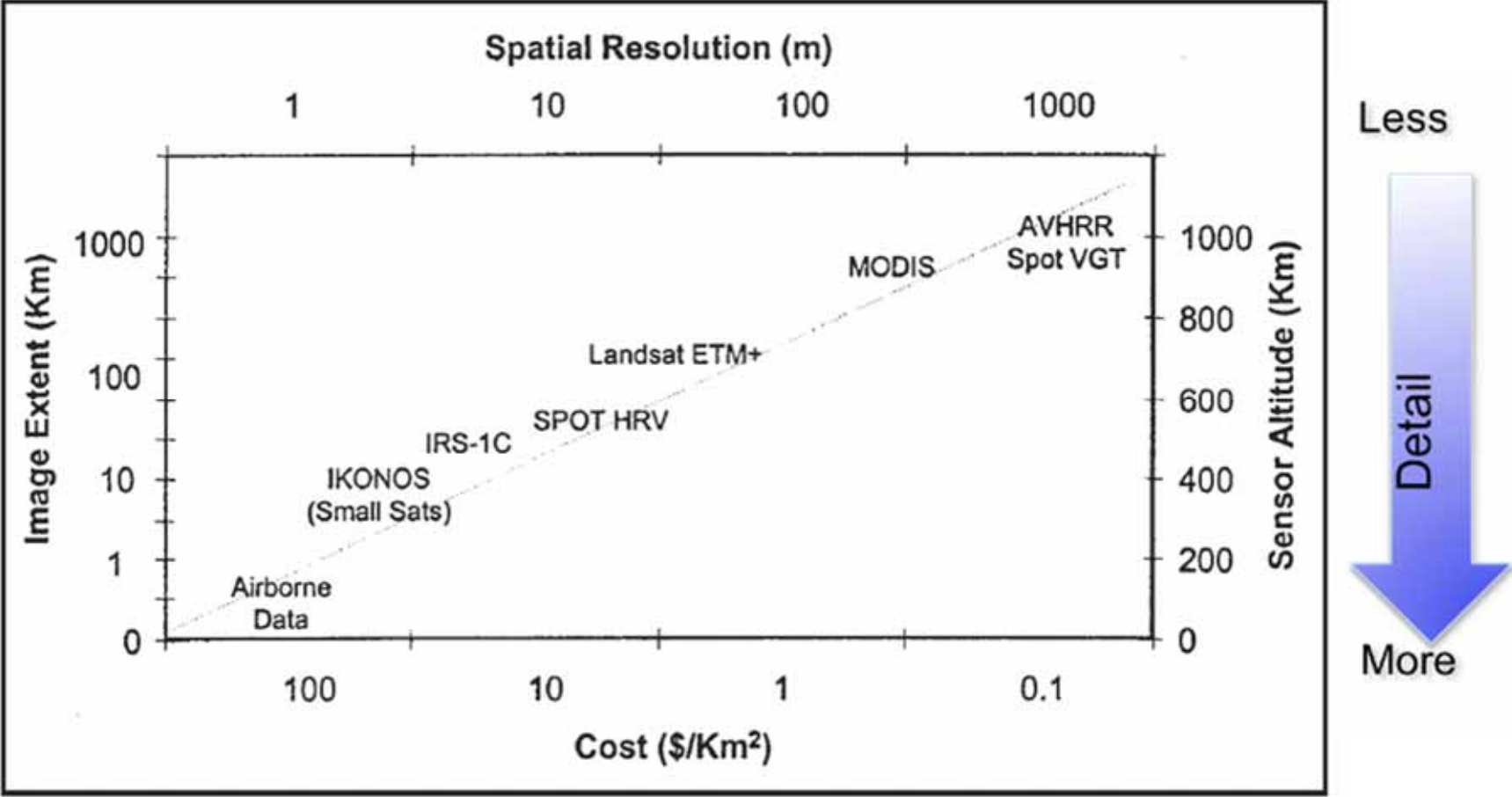


Number of Forest-Forest Joins



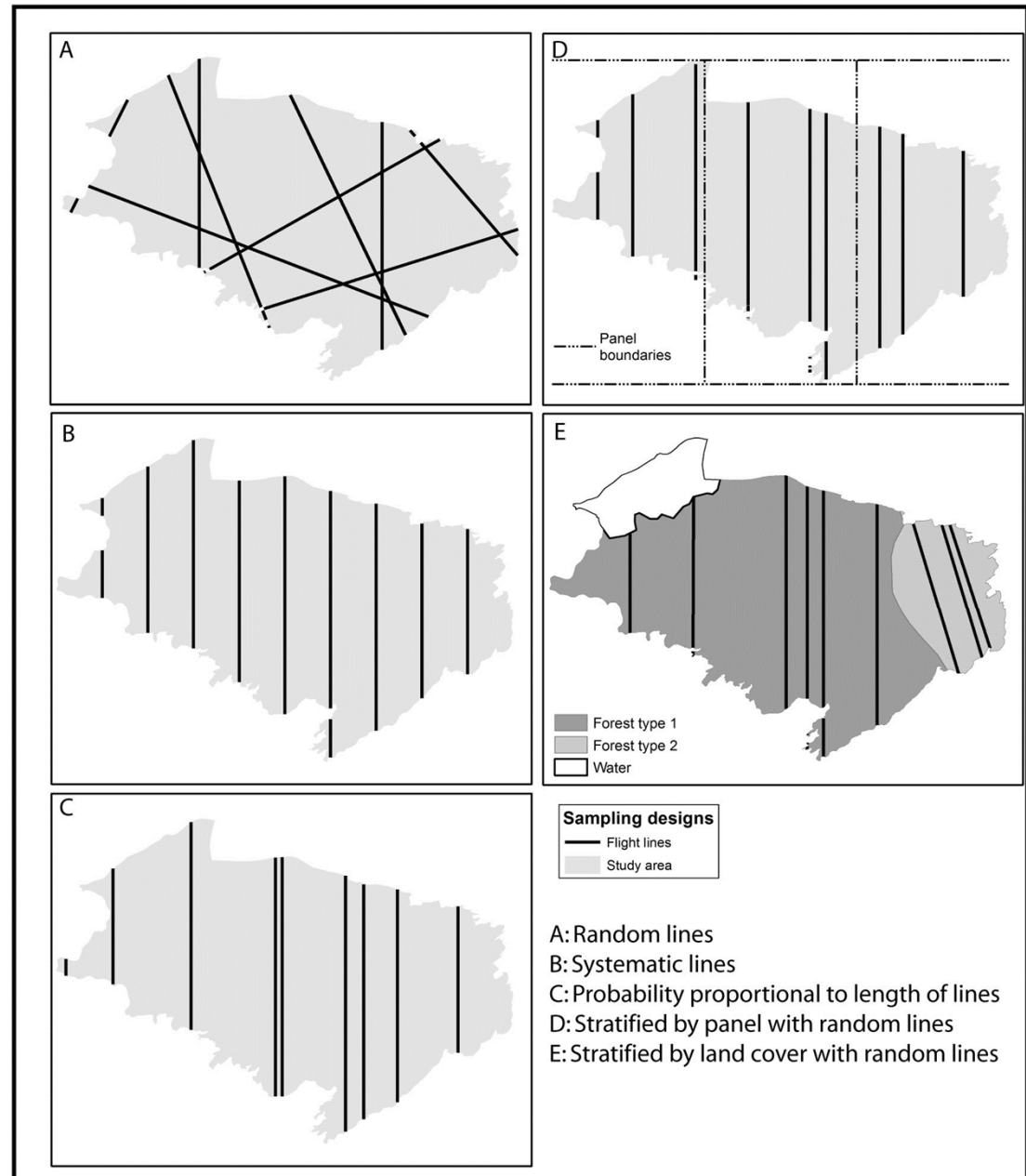
0 250 500 KM

The role of remote sensing in forest inventory and assessment



Relationship Between Spatial Resolution, Extent, and Cost (Source: Franklin, 2002).

Probability Sampling - strategies

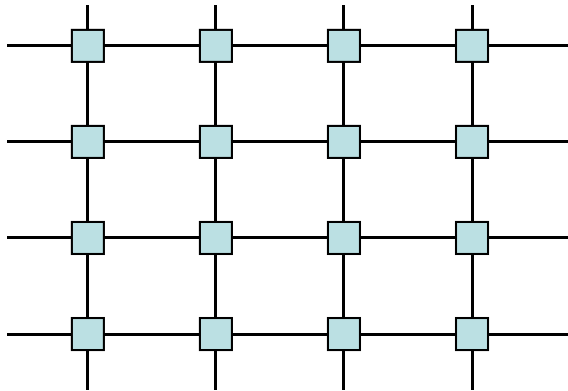


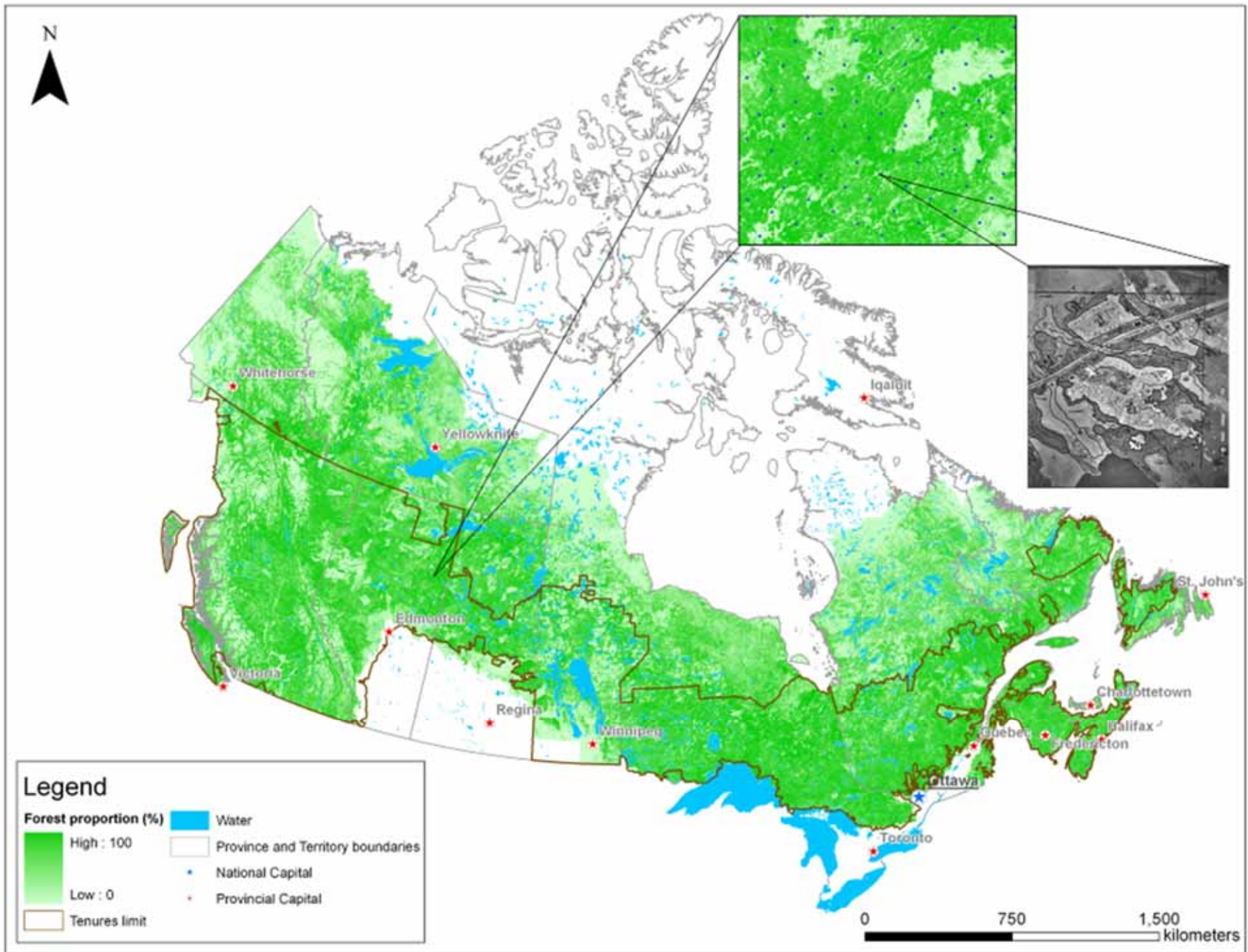
Lidar-plots: The case

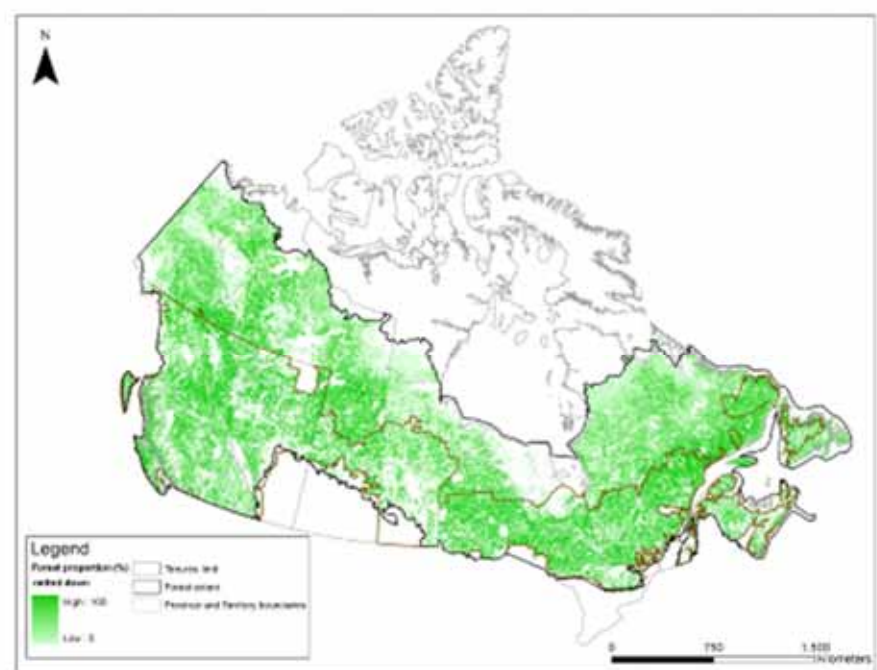
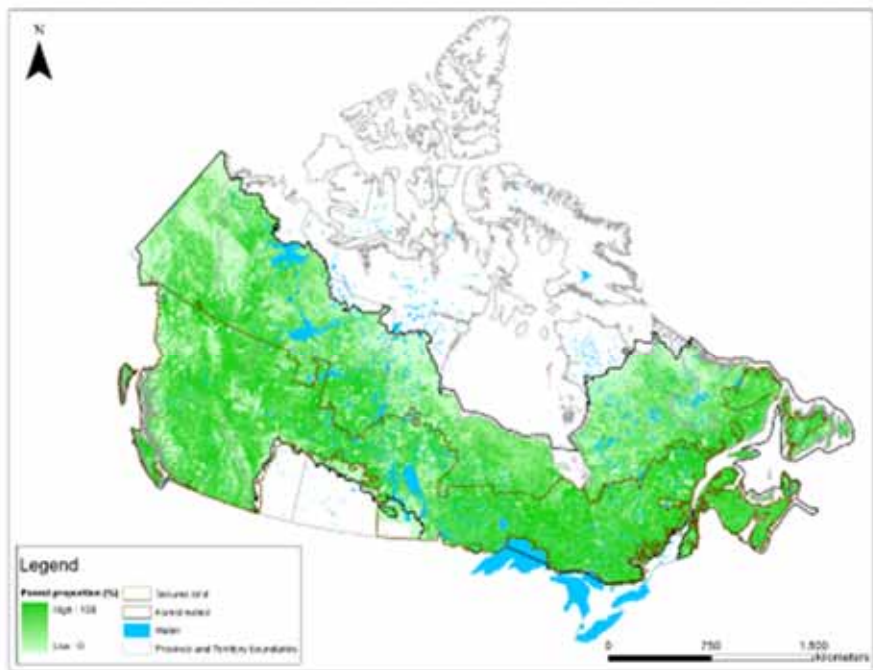
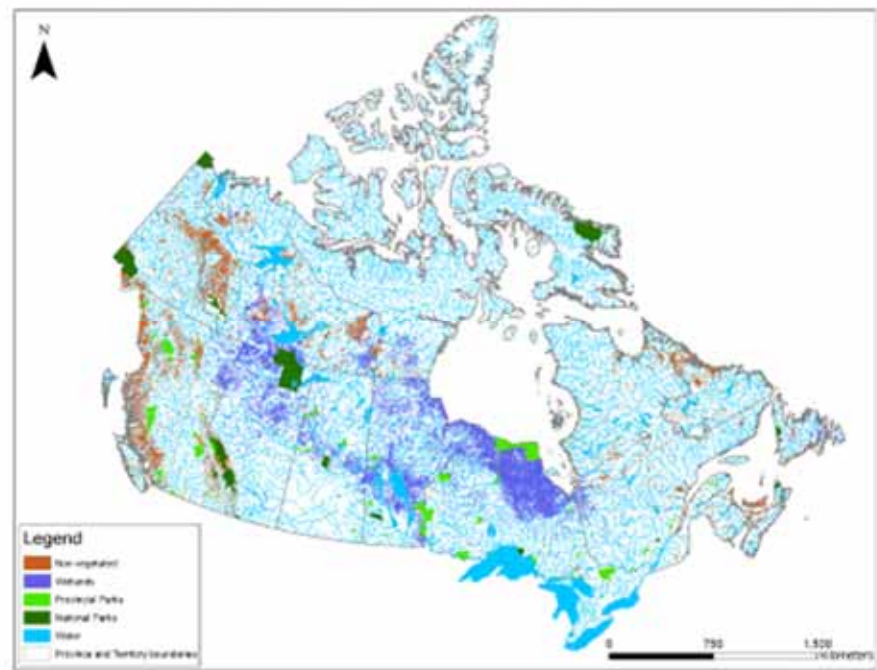
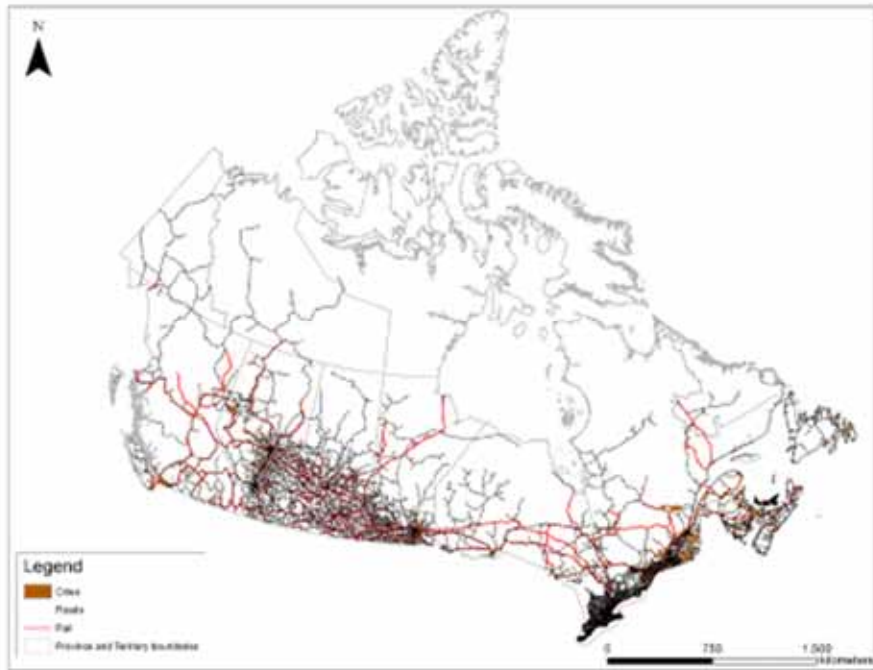
- But, what if you cannot sample sufficiently to enable a implementation of a probability sample?
- What if detailed local information is required to meet monitoring and reporting needs?

National Forest Inventory

- National systematic sample
- Rationale – standardization nationally
- Sample units are 2 x 2 km photo plots on a 20 km grid
- Ground Plots
 - many attributes
 - including DOM and soil C







Lidar

- Inventory
- Monitoring
- Sampling
- Repeatability

Wulder, M.A., C.W. Bater, N.C. Coops, T. Hilker, and J.C. White, 2008. The role of lidar in sustainable forest management. *The Forestry Chronicle*. Vol. 84, No. 6, pp. 807-826.

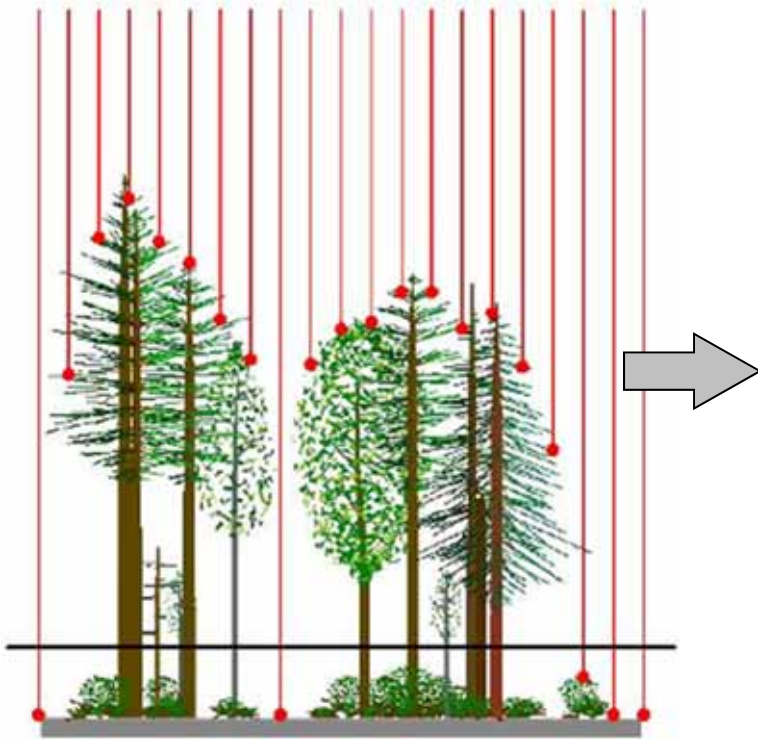
Hilker, T., Wulder, M.A., Coops, N.C (2008). Opportunities for Use of Small-Footprint Airborne Lidar Data and High Spatial Resolution Satellite Imagery to Augment Traditional Forest Inventory, *Canadian Journal of Remote Sensing*, Vol. 34, No. 1, pp. 5-12.

Wulder, M., and D. Seemann, 2003; Forest inventory height update through the fusion of lidar data with segmented Landsat imagery, *Canadian Journal of Remote Sensing*, Vol. 29, No. 5, pp. 536-543.

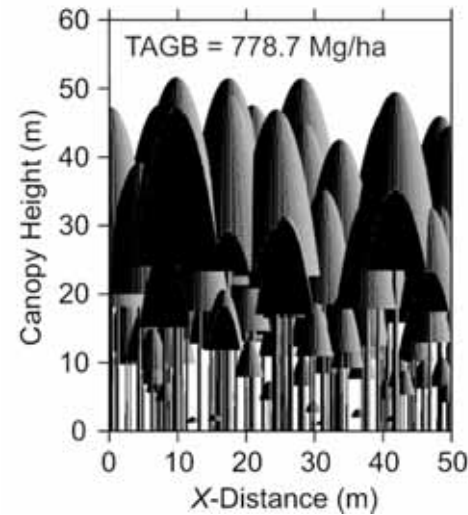
Bater, C.W., Wulder, M.A., Coops, N.C., Nelson, R.F., Hilker, T. and Næsset, E. (2011). Stability of sample-based scanning LiDAR-derived vegetation metrics for forest monitoring. *IEEE Transactions on Geoscience and Remote Sensing*. Vol. 49, No. 2. pp. 2385-2392.

Individual tree measures from the lidar are summarized to produce *critical* plot-level attributes at thousands of plot locations.

Tree-level measures

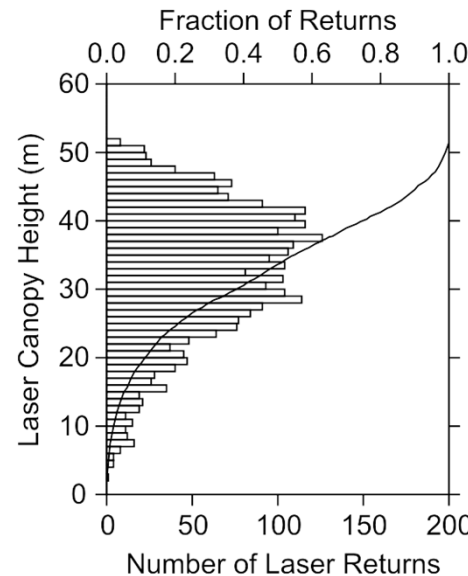


Plot

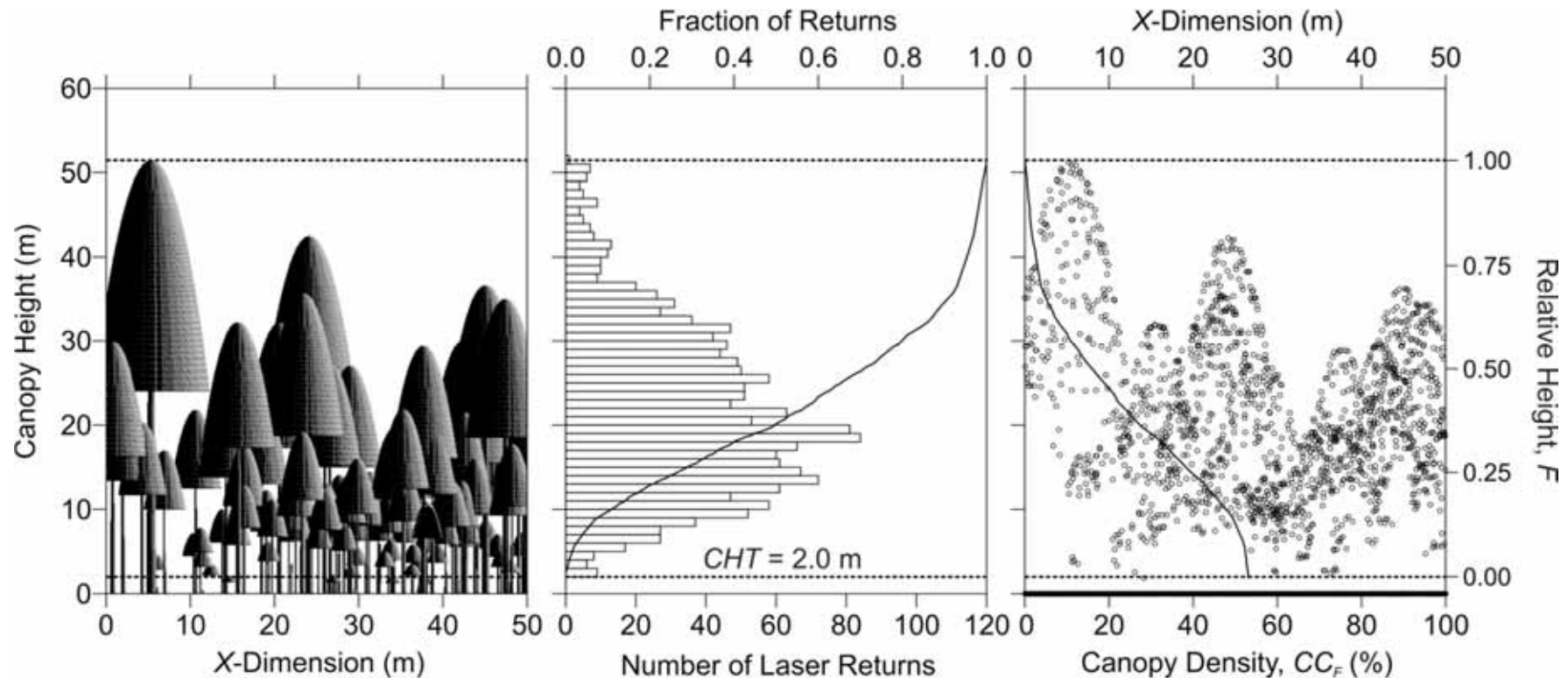


Plot-level attributes

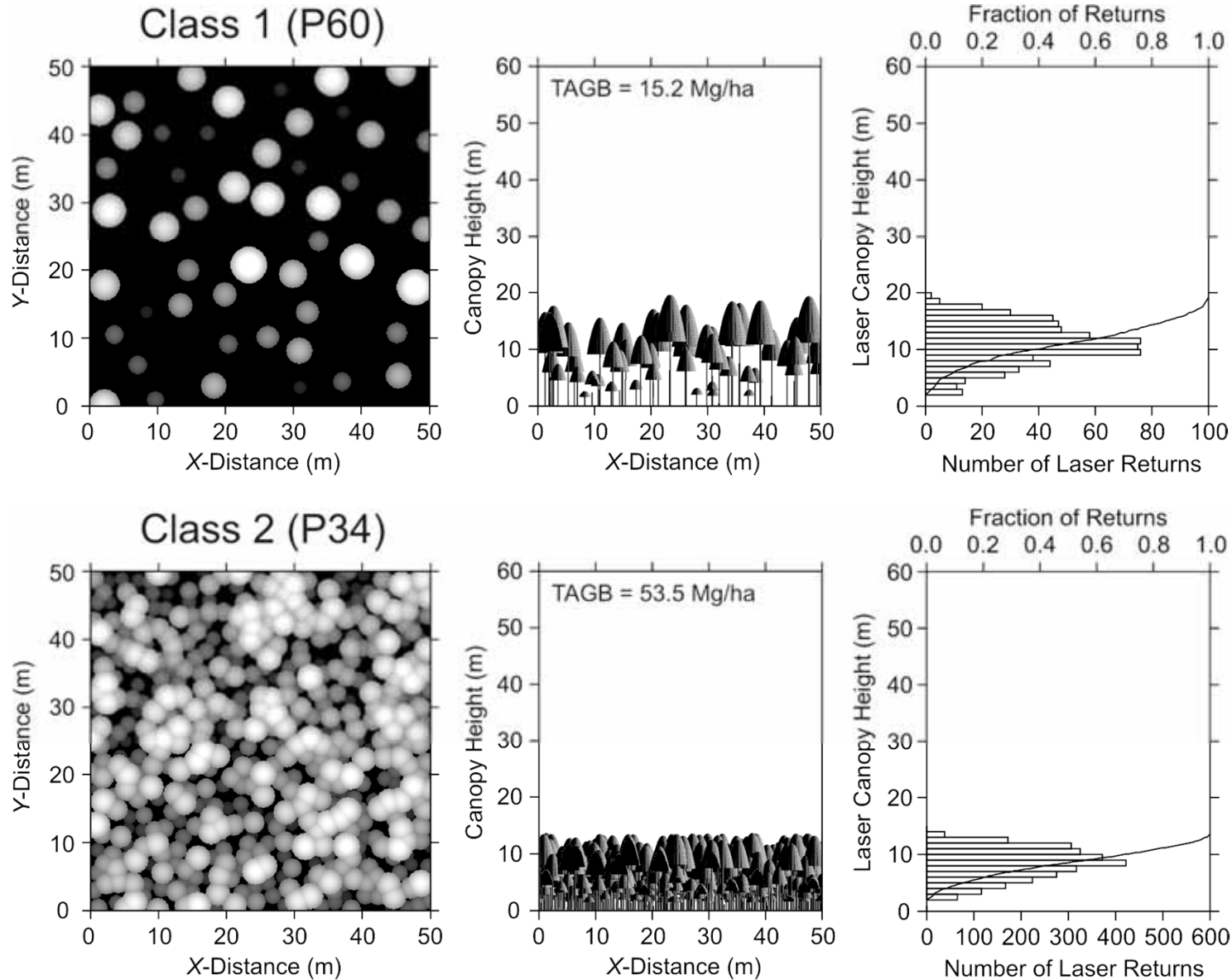
- stand structure
- height (max, mean...)
- crown closure
- volume
- biomass
- gap fraction
- stem density



Spatially generalizing the lidar data

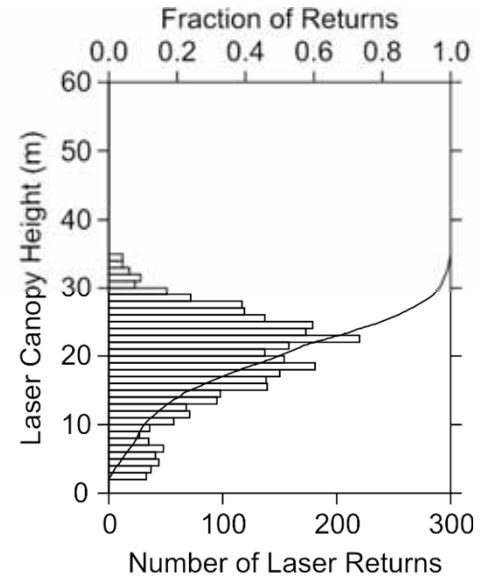
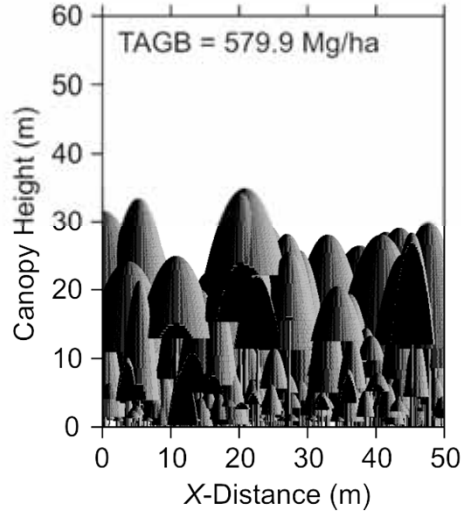
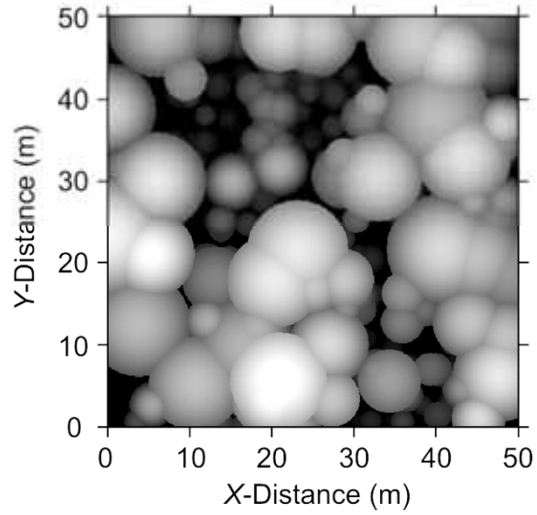


Cross-sectional view of simulated canopy (left), histogram and sample quantiles of the vertical distribution of canopy heights (middle), and canopy density (ratio of heights above threshold F to total number of LiDAR returns in plot) for each relative canopy height F (right).

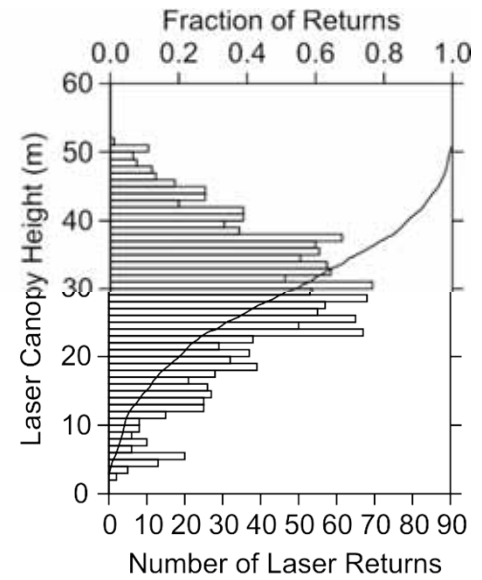
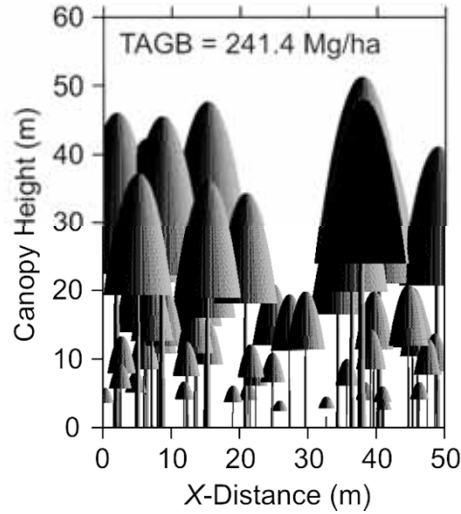
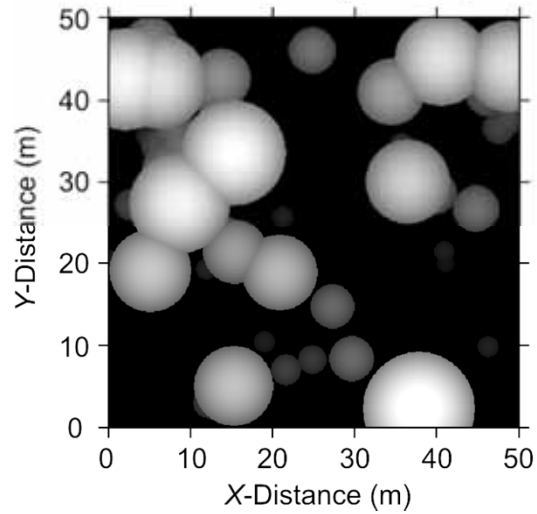


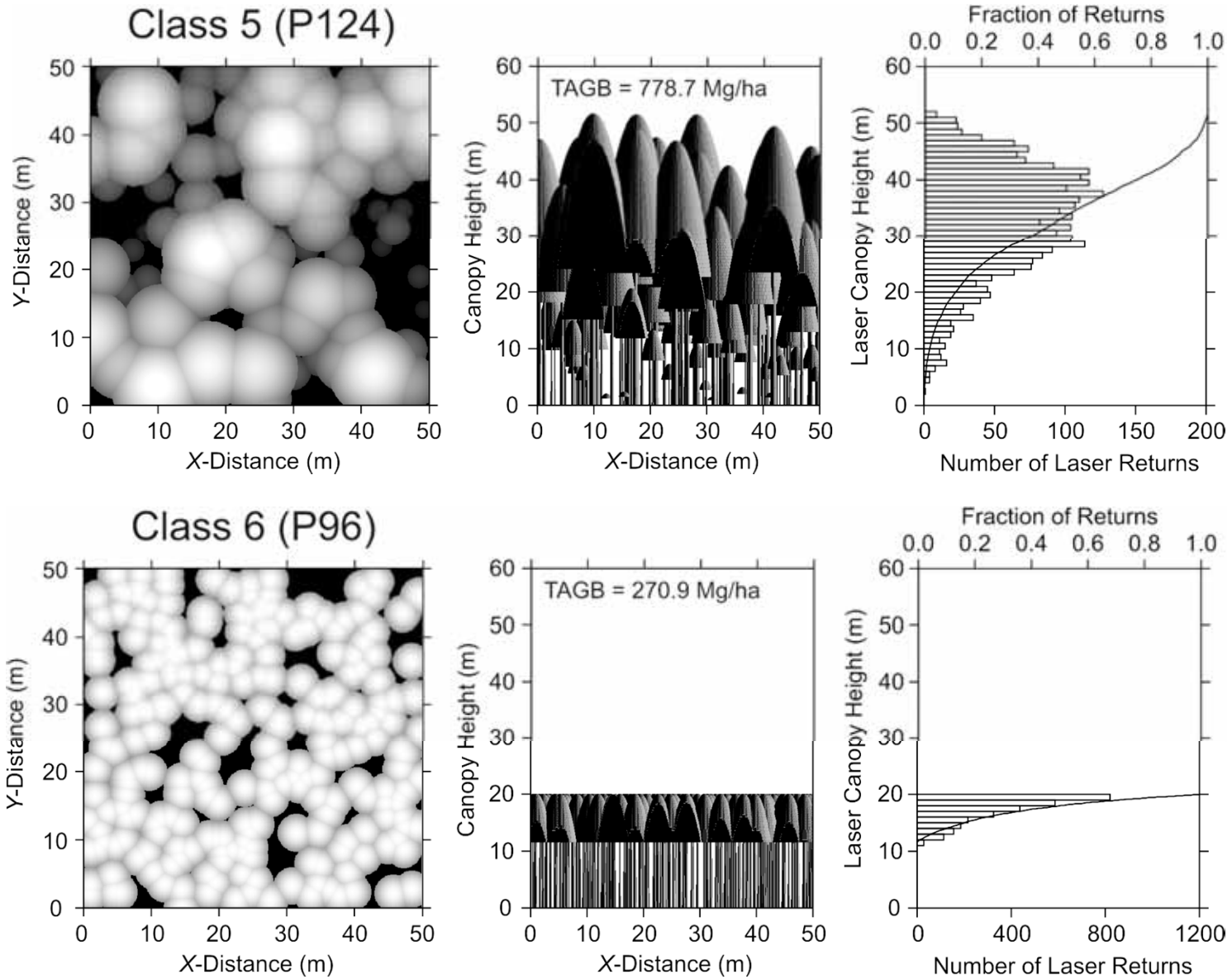
Frazer, G., M. Wulder, and O. Niemann, 2005; Simulation and quantification of the fine-scale spatial pattern and heterogeneity of forest canopy structure: A lacunarity-based method designed for analysis of continuous canopy heights, *Forest Ecology and Management*, Vol. 214, pp. 65-90.

Class 3 (P161)



Class 4 (P129)



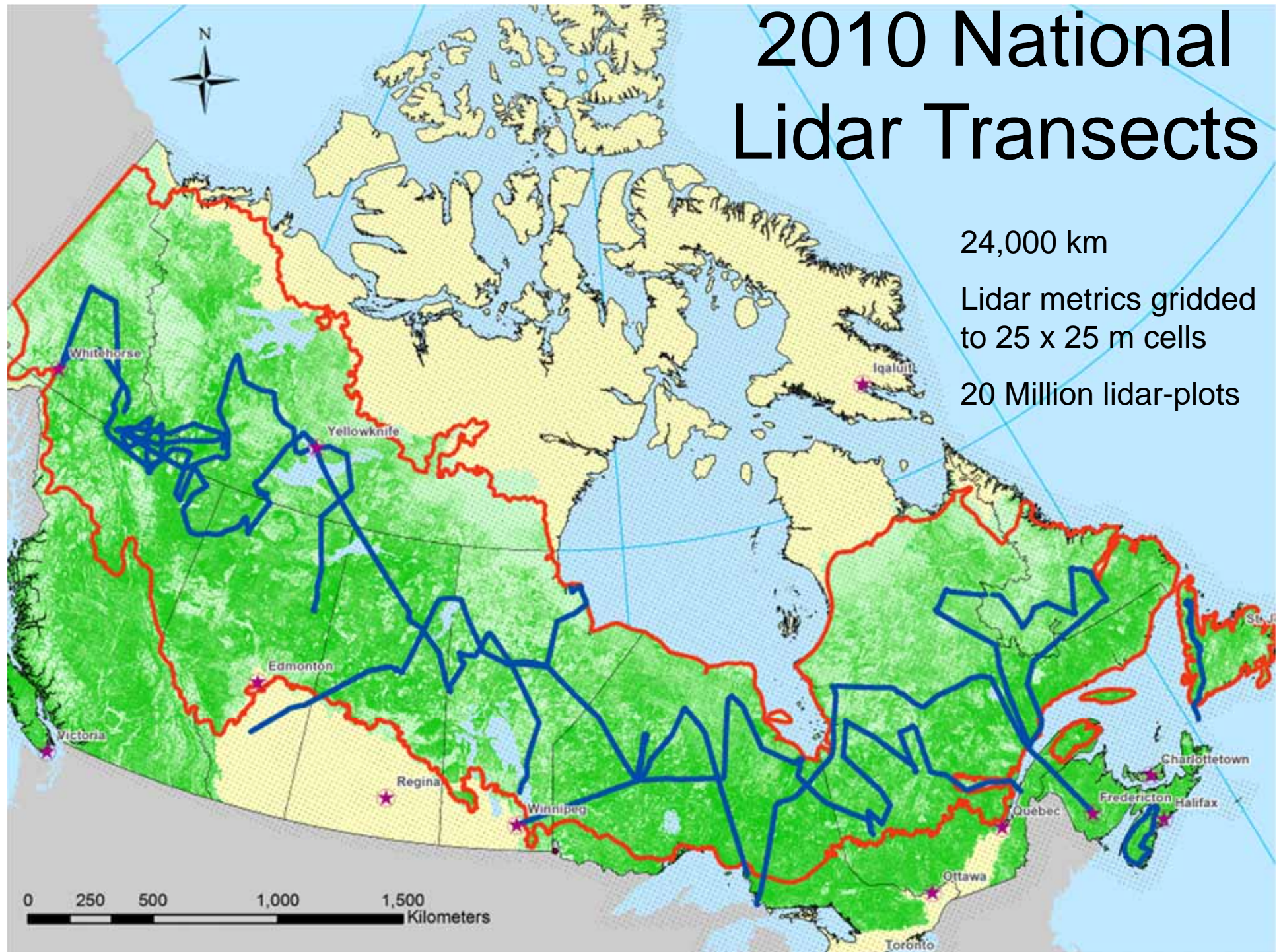


2010 National Lidar Transects

24,000 km

Lidar metrics gridded to 25 x 25 m cells

20 Million lidar-plots

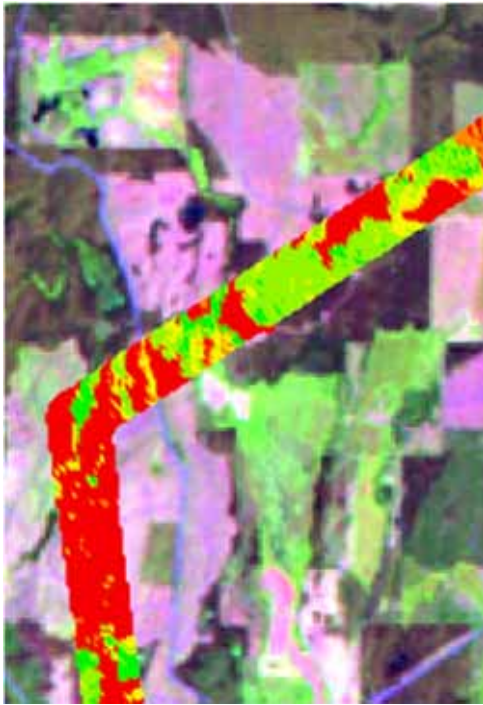


Outcomes – Fast facts

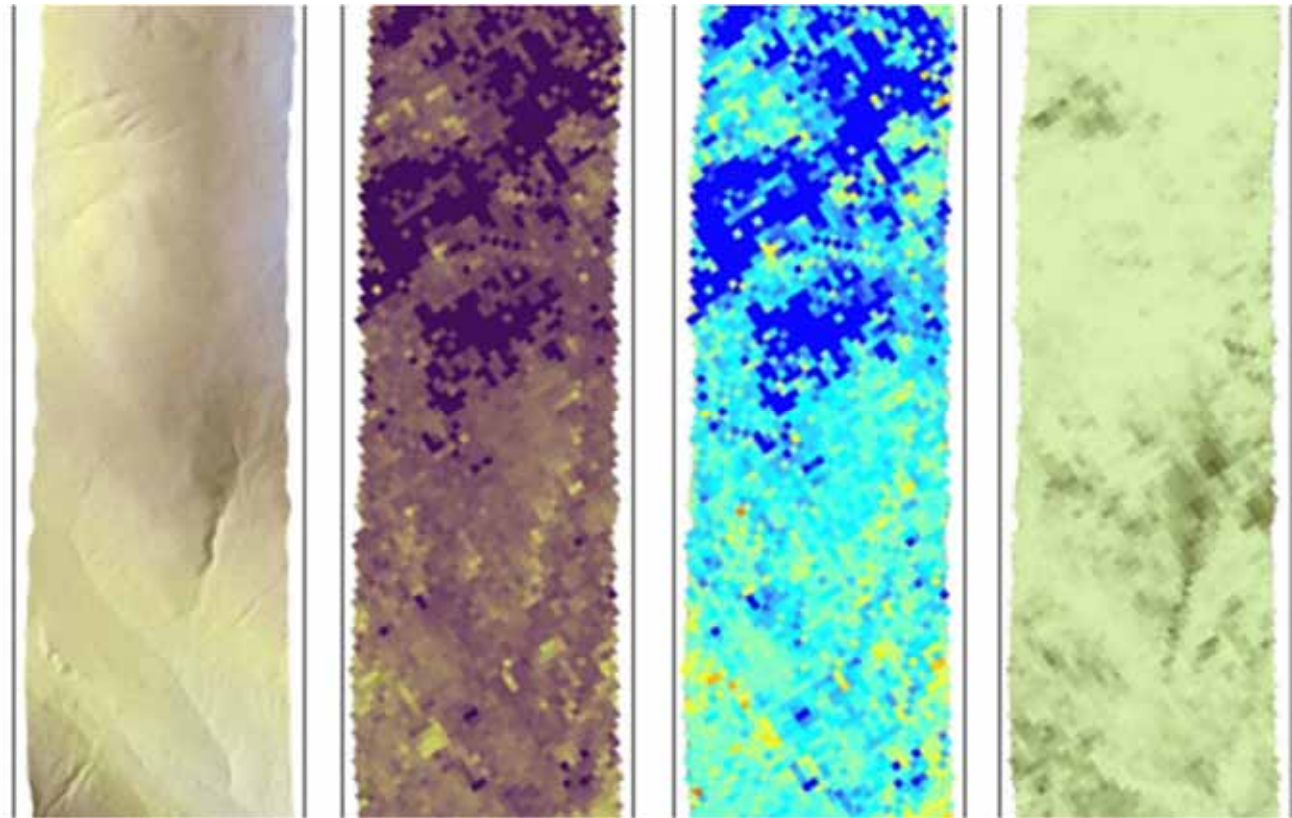
Database size on disk	6.33 GB
Number of plot locations processed	32,103,543
Number of plot locations containing lidar data	20,234,390
Area of plots containing lidar data	1,264,649 ha
Total number of lidar returns	18,542,489,414
Total number of first returns above a 2 m height threshold	4,982,094,975

- Total Length: 24,285.61
- 34 transects, mean length 700 km

Example transect outputs/metrics



Height example,
Landsat backdrop



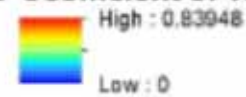
Terrain height



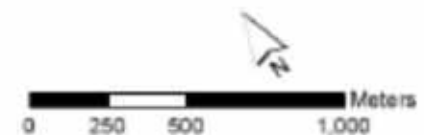
Height 95th percentile



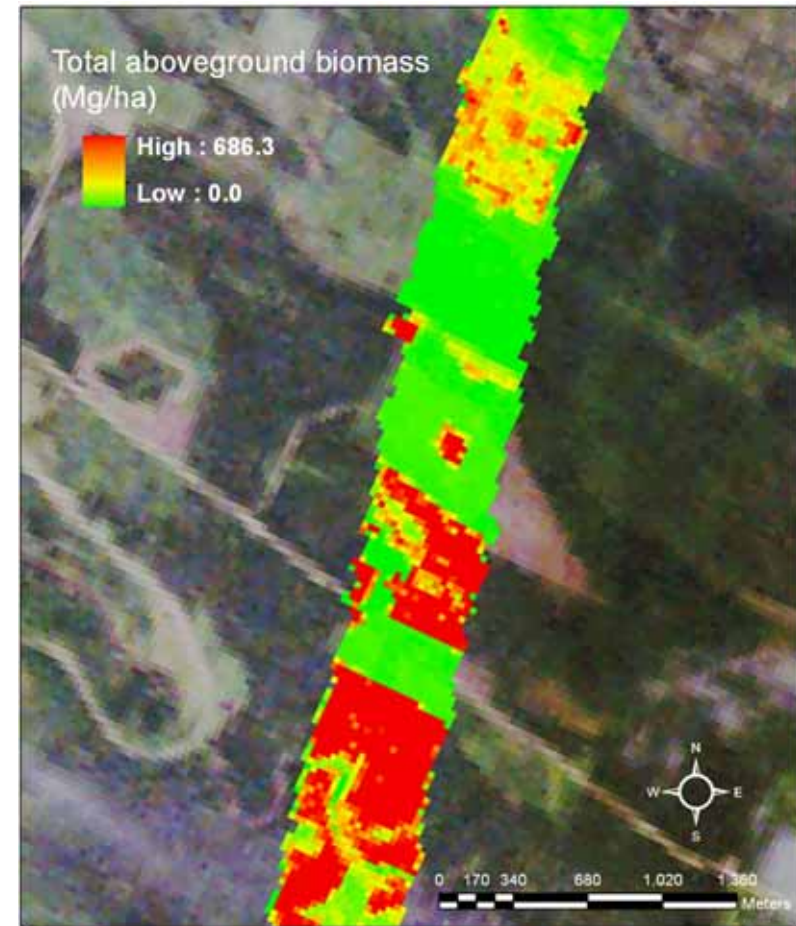
Coefficient of variation



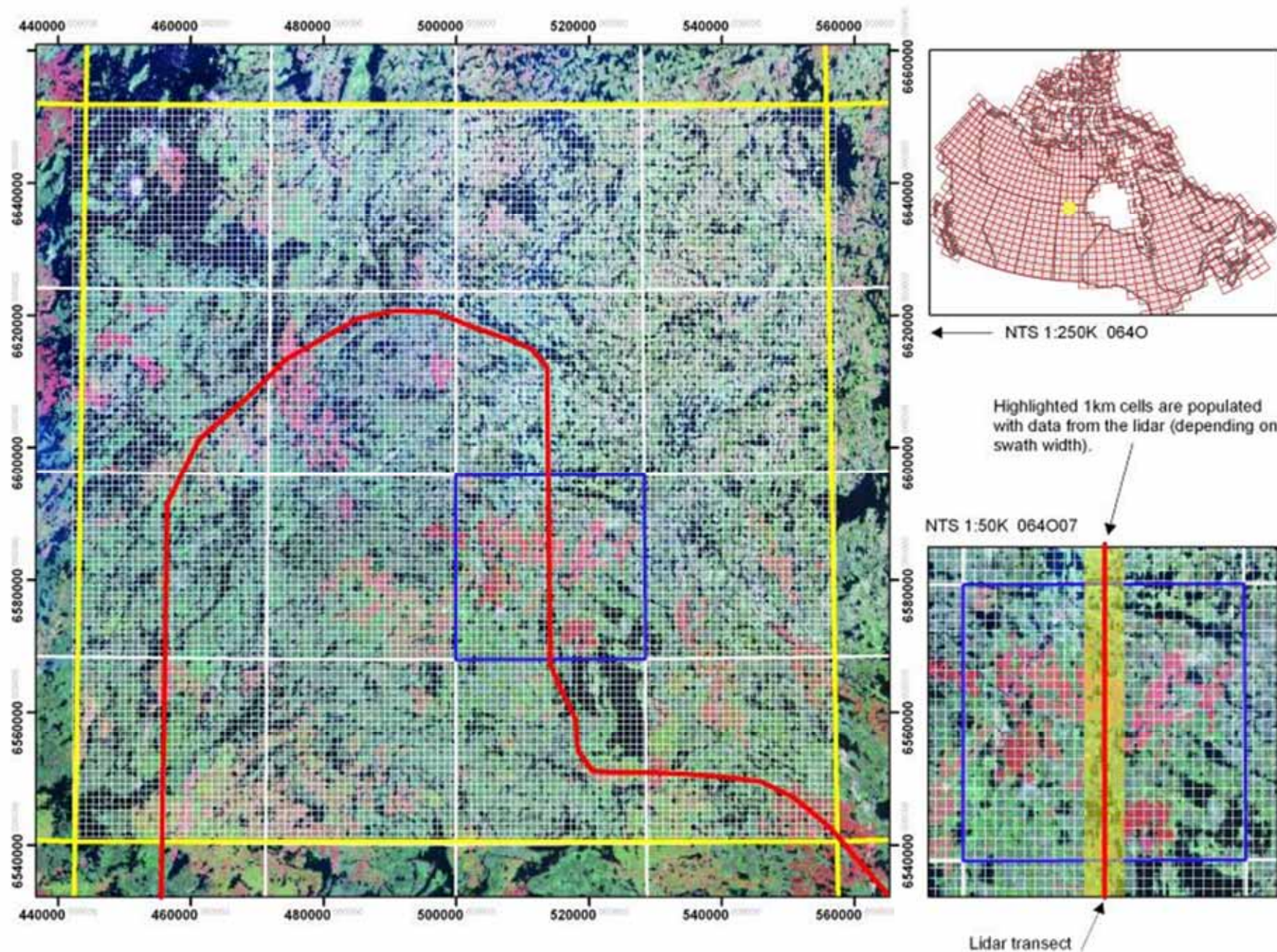
Cover

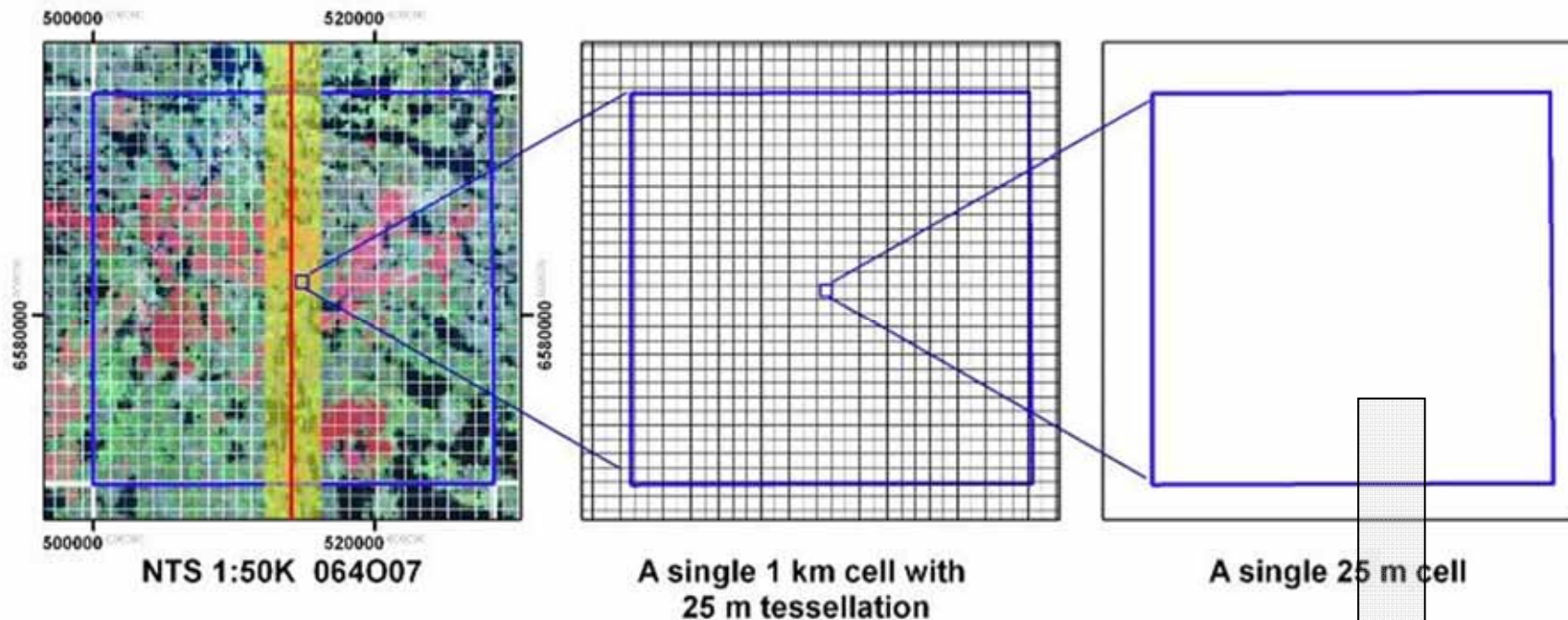


Example product: biomass

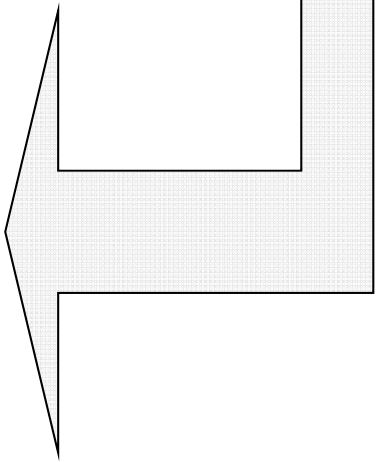


Model development for the estimation of aboveground biomass using a lidar-based sample of Canada's boreal forest; Bater, Coops, et al.





- ATTRIBUTES for each 25 m cell
- | | |
|--------------------------|------------------------|
| Flight date | Stand structure |
| Number of returns | Height |
| Elev minimum | Crown closure |
| Elev maximum | Volume |
| Elev mean | Biomass |
| Elev mode | Volume |
| Elev stddev | Gap fraction |
| Elev variance | Stem density |
| ... | |



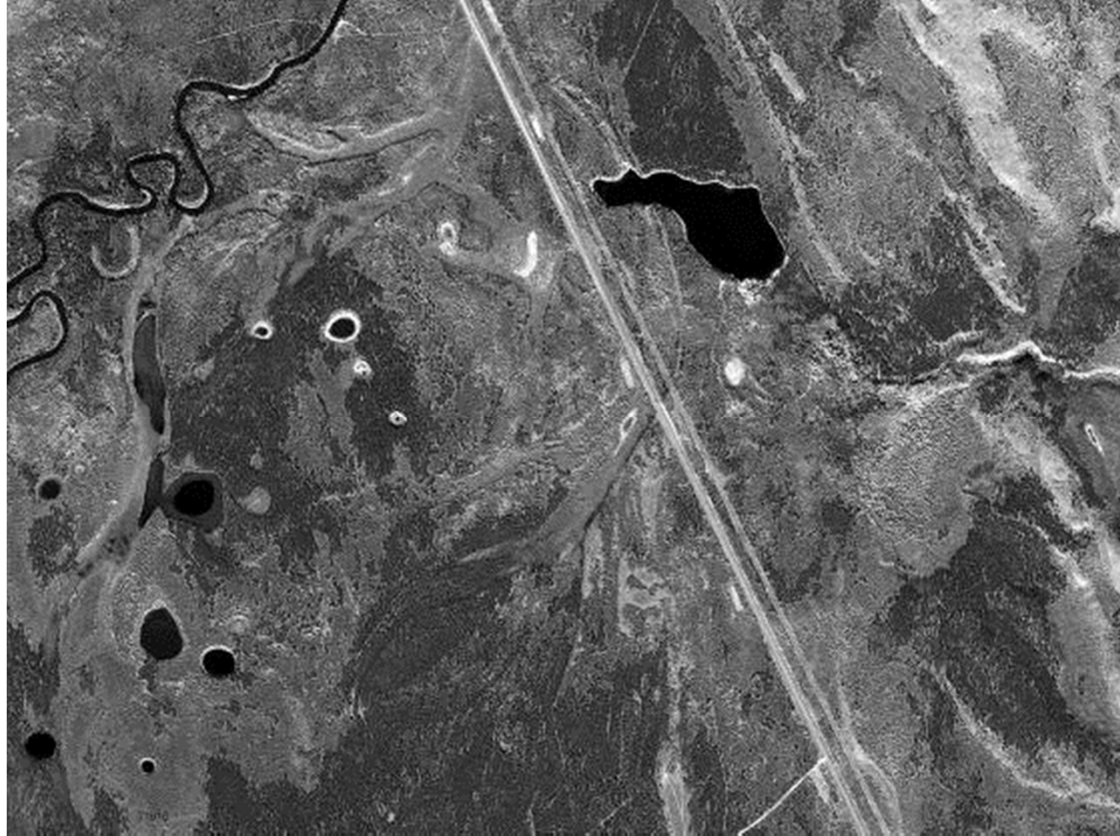
Application example

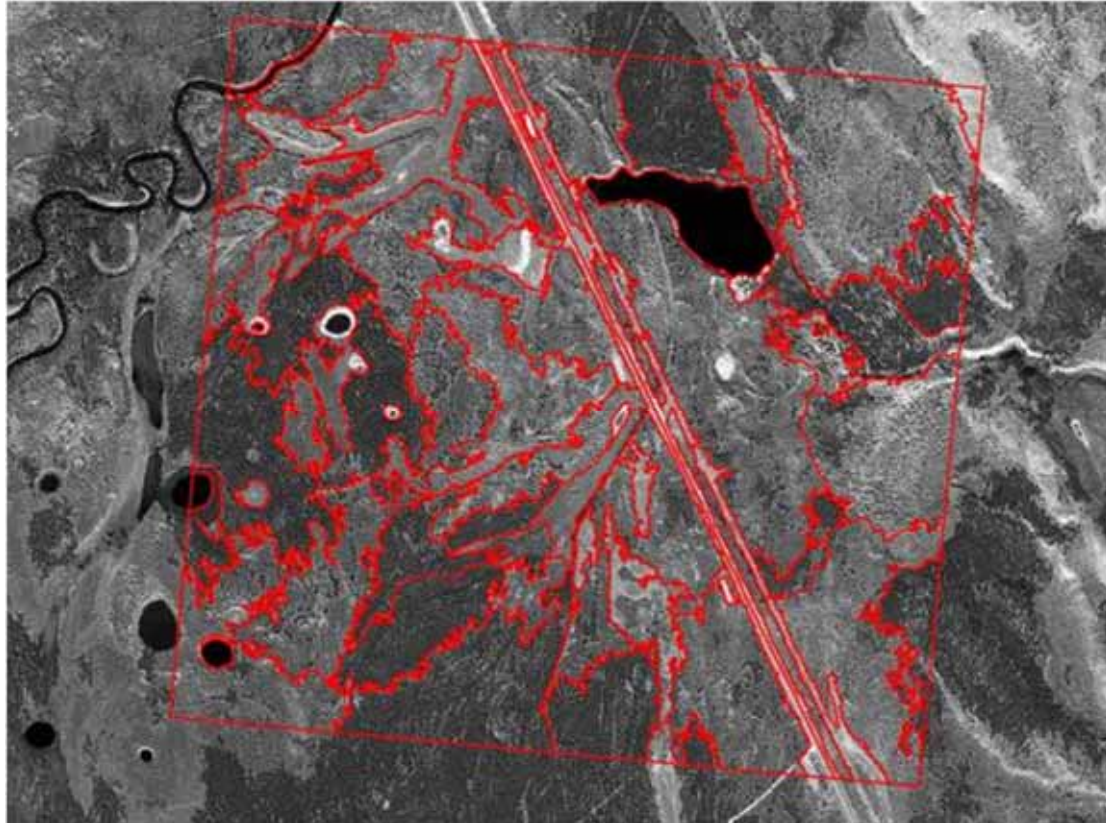
Lidar-plot application Cal / val of Very High Spatial Resolution imagery in support of NFI

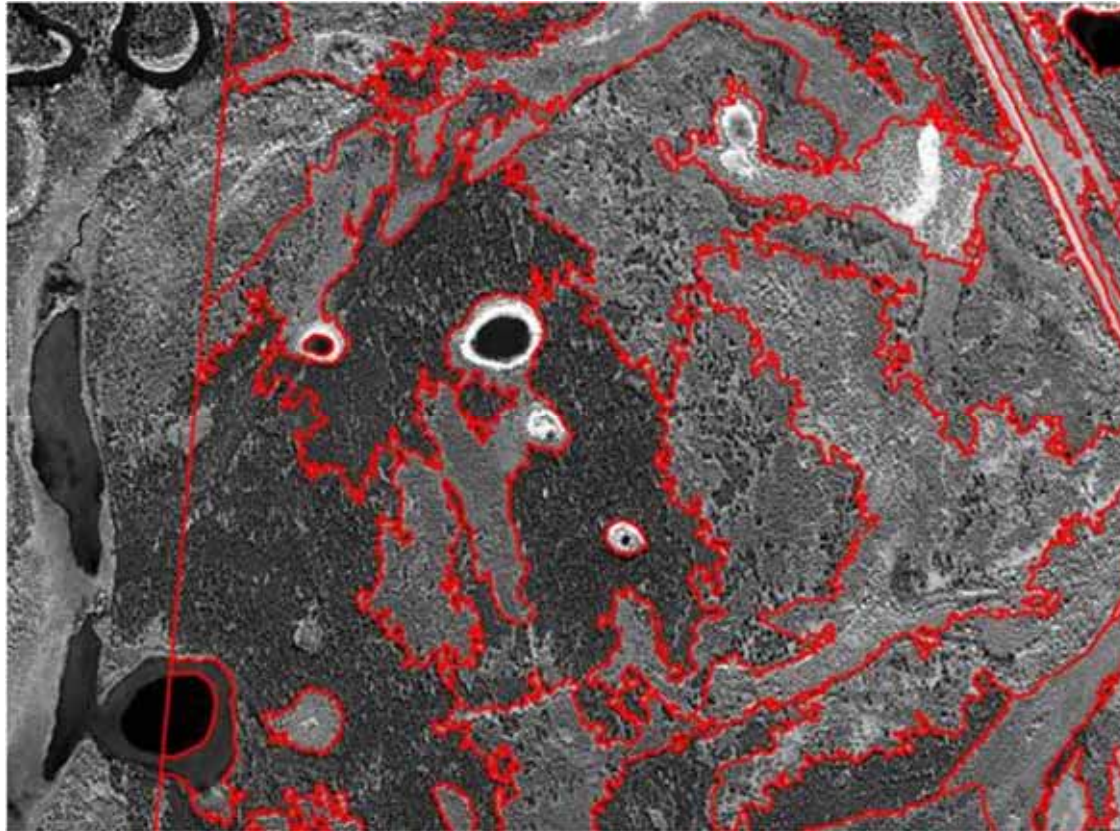
Key refs:

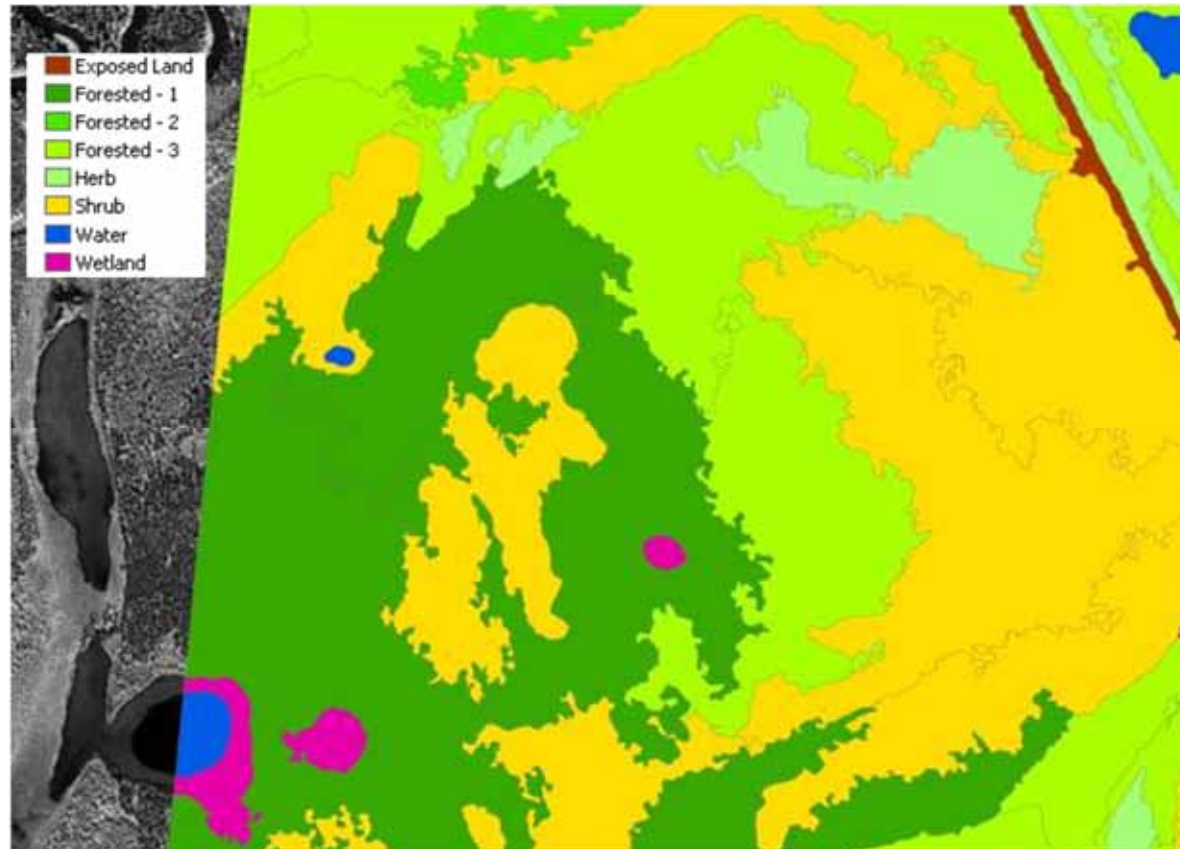
Falkowski, M.J., M.A. Wulder, J.C. White, and M.D. Gillis (2009). Supporting large-area, sample-based forest inventories with very high spatial resolution satellite imagery. *Progress in Physical Geography*. Vol. 33, No. 3, pp. 403–423. (DOI: 10.1177/0309133309342643)

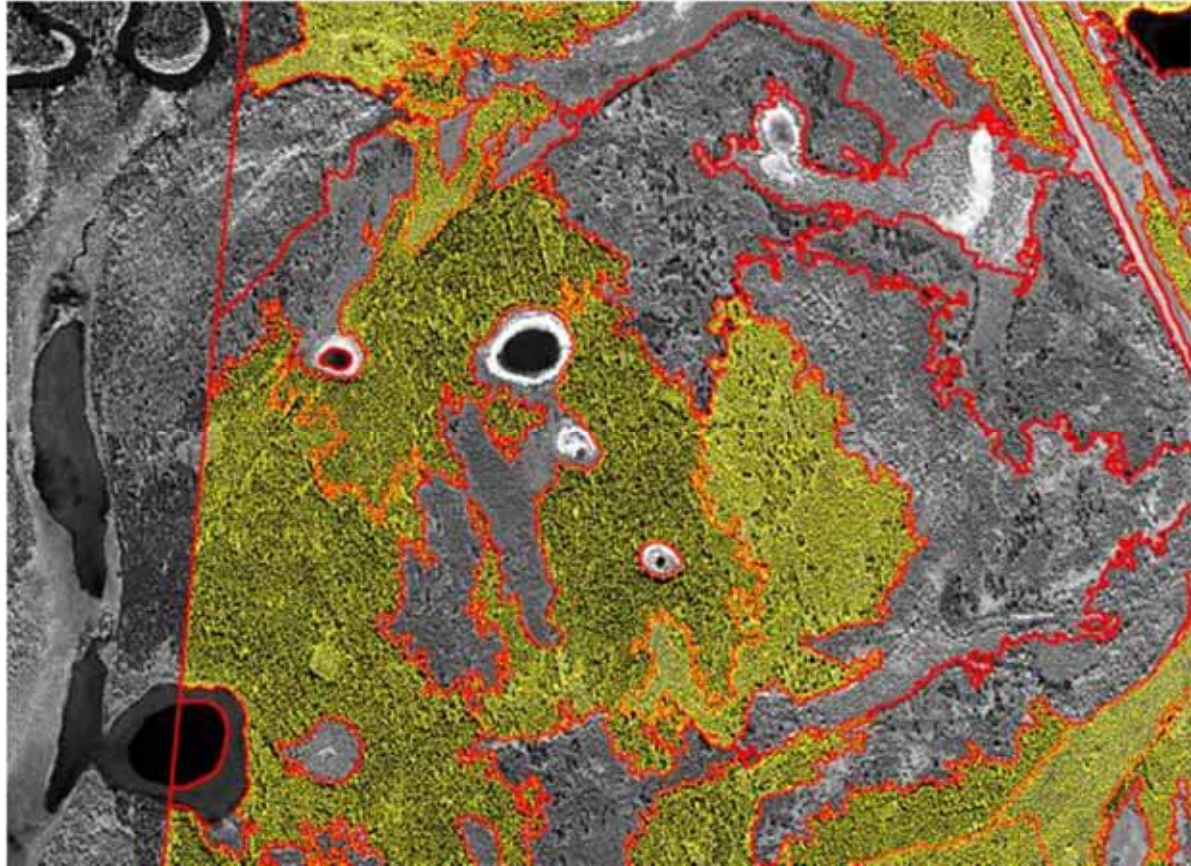
Mora, B., M.A. Wulder, and J.C. White (2010). Segment-constrained regression tree estimation of forest stand height from very high spatial resolution panchromatic imagery over a boreal environment. *Remote Sensing of Environment*. Vol. 114, pp. 2474–2484. (DOI: <http://dx.doi.org/10.1016/j.rse.2010.05.022>)

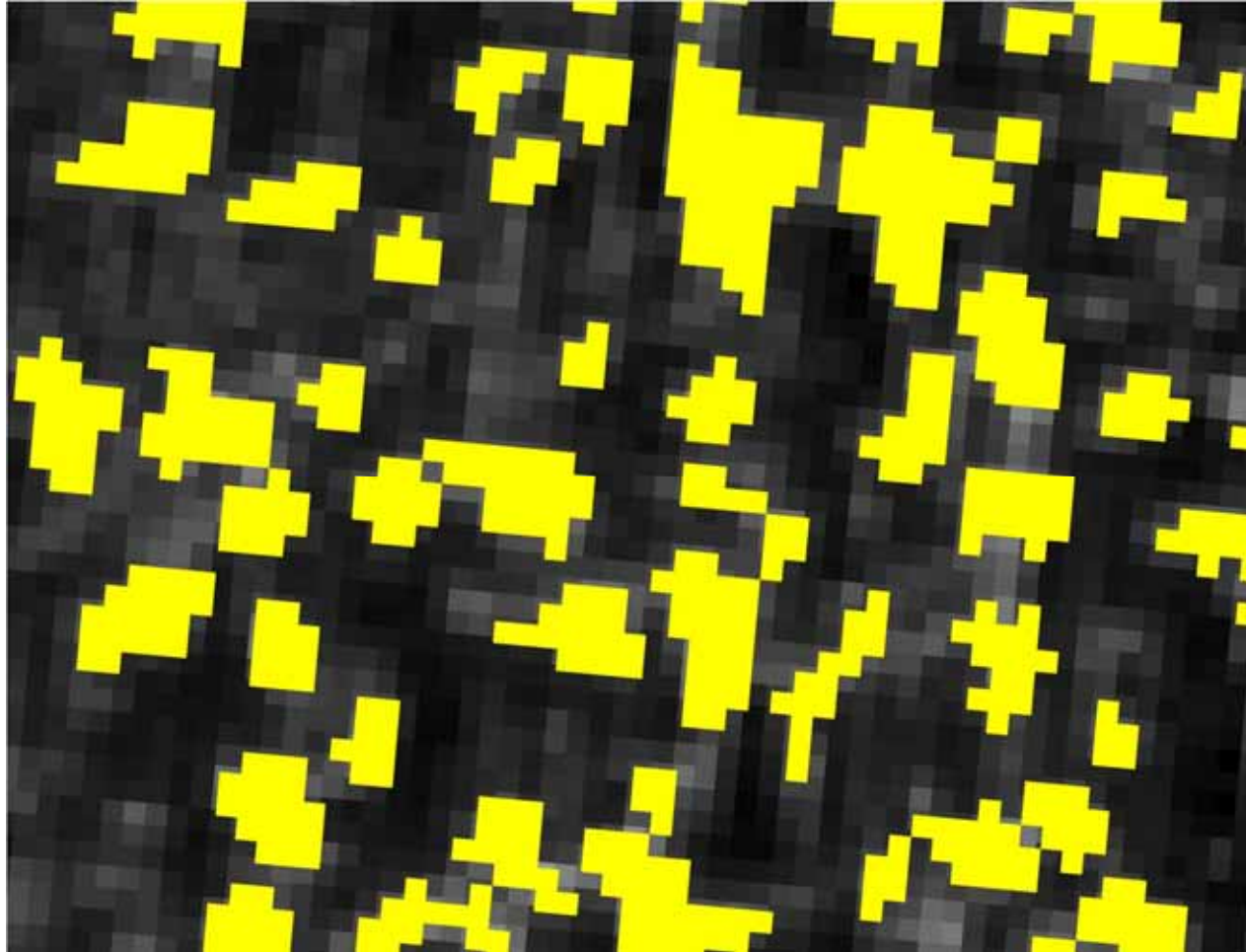












Gougeon, F. (1995). A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian Journal of Remote Sensing*. 21(3):274-284.

Stand and crown metrics

Stand segment metrics

Photo plot ID: 4

Stand ID: 94

Area: 14.85 ha

Stand type: Conifer

Leading species: Lodgepole pine

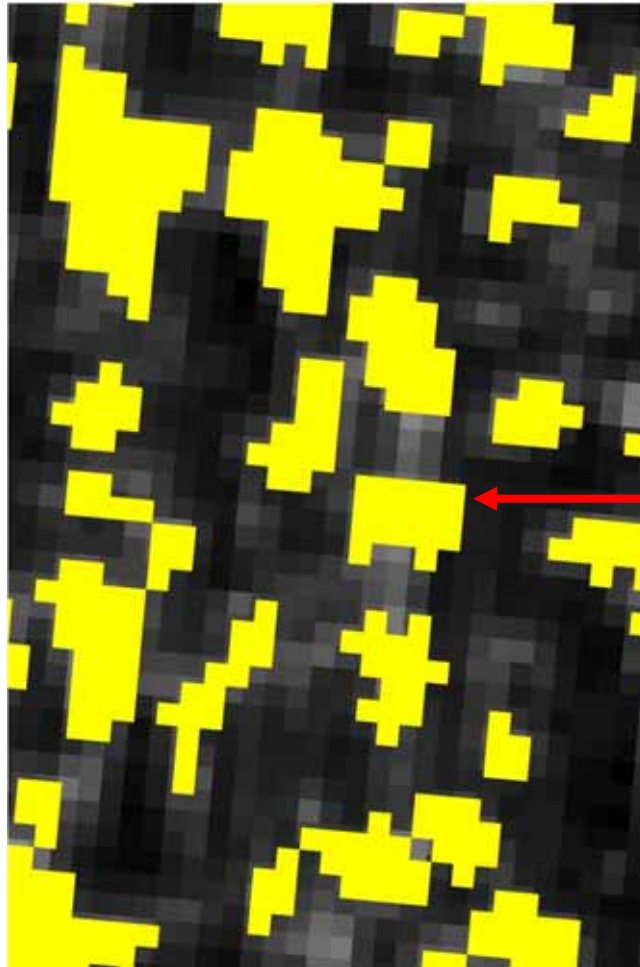
Mean crown area: 17.8 m²

25th percentile: 9 m²

50th percentile: 14 m²

75th percentile: 23 m²

Crown closure: 37%



Crown metrics

Crown ID: 1385

Length: 3.5 m

Area: 6.8 m²

40th
percentile

Outcome

- Forest inventory-like information generation
 - Automated delineation
 - Semi-automated attribution
 - NFI inputs for inaccessible areas
- need for calibration and validation data,
 - Results with lidar-plots replicate ground plot results: RMSE on Ht of <2m

Applications Examples

- Characterization of post-disturbance recovery
- Audit and calibration of growth and yield equations
- Accuracy assessment of national products (Canadian and international)
 - Information independence
- Calibrating / validating modeled national products (based upon optical RS data)

What do we want, where are we going?

- Modeling for detailed attributes
 - Supported by inventory, plots (field, photo)
- Remote sensing:
 - Wall-to-wall
 - Creation of strata
 - Capture of disturbance
- Information needs to drive frequency, spatial detail, categorical detail, of attributes and disturbance
- A combined modeling framework for modeling attributes and using satellite RS for change capture is recommended.
 - Growth can be modeled, depletions mapped and integrated. (on a wide-area tessellation)
 - Higher spatial resolution RS can aid in cal / val
- Multiple outcomes, probabilities assigned

Framework notion:

Integrating multi-scale remote sensing and modeling

- Establish high spatial resolution, fine scale grid
 - Say 25 m cells nationally
- Populate grid through existing spatial data, remote sensing (e.g., Landsat), and modeling
 - GO Modeling
 - Regional expectation from lidar-plots
- Update the grid through satellite change detection
 - Find change, attribute change, update, model
- Confirm / adjust modeled outcomes through sample based, higher spatial resolution remote sensing (VHSR imagery, lidar-plots)
 - cal/val of models with lidar-plot
- Repeat

Thank you

Contact Information:

Mike Wulder

mwulder@nrcan.gc.ca

Publications:

<http://cfs.nrcan.gc.ca/publications/authors/read/11091>



Natural Resources
Canada

Ressources naturelles
Canada



Canada



Airborne lidar sampling of the Canadian boreal forest: Planning, execution & initial processing

Chris Hopkinson¹, Michael A. Wulder², Nicholas C. Coops³,
Trevor Milne¹, Allyson Fox¹, & Christopher W. Bater³

¹Applied Geomatics Research Group, Lawrencetown, Nova Scotia, Canada
Chris.hopkinson@nsc.ca

²Pacific Forestry Centre, Canadian Forest Service, Victoria, British Columbia, Canada
mwulder@pfc.cfs.nrcan.gc.ca

³Department of Forest Resource Management, UBC, Vancouver, British Columbia, Canada
Nicholas.coops@ubc.ca

Abstract

During the summer of 2010, a transcontinental aerial survey mission was performed to acquire 24,000 line km of lidar transects covering >15,000 km² representing all ecozones within Canada's boreal forest. The coverage equates to ~21 million 'lidar plots' at the 25 m grid cell resolution. Each 'plot' contains the position and intensity of 1000 to 2000 laser points, which describe the terrain surface and 3D canopy structure, which will be used to predict forest inventory attributes and to support calibration of wide area satellite-based imagery. Furthermore, in similar fashion to geo-located permanent sample plots, the lidar transect flight path from 2010 can be re-surveyed in the future to facilitate monitoring of forest development and change in a consistent and quantifiable manner. The paper describes the mission planning criteria, survey logistical considerations and customised transect data processing routines.

1. Introduction

Canada's boreal forests cover approximately 3,070,000 km² (Brandt, 2009) and span almost 5000 km from Newfoundland to the Yukon. Monitoring remote and extensive forest resources across such a large area is challenging. While satellite remote sensing is used to meet some information needs, some form of calibration and validation data are required. Further, many scientific questions require plot-level information to relate conditions in a spatially referenced manner. The utility of integrating lidar sampling data with satellite imagery to investigate temporal canopy changes over a study area within Canada's boreal forest has been demonstrated (Wulder *et al.* 2007). Similarly, lidar sampling has also been shown to facilitate the scaling of forest attributes from the plot scale to an entire Province (~50,000 km²) (Hopkinson *et al.* 2011). A systematic collection of lidar transect data representing the entire Canadian boreal forest is envisioned to provide a widely distributed sampling of information that may be used to support calibration / validation activities for monitoring programs, to support research, and to offer experience and insights in support of repeat lidar-based monitoring efforts.

A lidar project requiring coordination at the continental scale of Canada's land mass poses unique challenges. To plan and execute such a mission, the Canadian Forest Service (CFS) initiated partnerships with experienced lidar forestry researchers at the University of British Columbia (UBC) and the Applied Geomatics Research Group (AGRG) in Nova Scotia. This paper will summarize the key elements of the mission planning, data collection, and initial post-processing. Some of the challenges faced and solutions implemented at each stage of the project will be highlighted.

2. Mission Planning

The concept was to adopt the C-CLEAR (Canadian Consortium for Lidar Environmental Applications Research) collaborative research support framework, while using the AGRG airborne lidar equipment and research personnel to facilitate data acquisition and processing. C-CLEAR lidar missions are conducted annually across Canada and occasionally into the Arctic, so this model provided the ideal basis for a boreal-wide transect sampling mission from one side of Canada to the other. The first mission planning task was to identify priority areas and then to map out an approximate route to be taken by the survey aircraft.

The goal of the transect planning was to ensure that a broad sample of boreal conditions was captured, limit the amount of flying over areas with existing management inventories, and to avoid flying too frequently over sparsely tree areas. The first task was to identify and target boreal ecoregions that displayed >50% forest cover to ensure lidar transect sampling was productive; i.e. no point sampling large areas with no forest cover. Given that some forest areas can undergo active forestry operations, these areas are often described within existing forest inventory databases. Therefore, to focus the sampling in ecoregions with minimal existing inventory data, areas displaying more than 75% managed forest cover were excluded. To enable comparisons between lidar transect data and available plot-level image and field data across the country, the National Forest Inventory (NFI) grid node locations (Gillis *et al.* 2005) were used to guide the specific positioning of transect locations. In addition to the national-level sampling, more intensive lidar transect data collection was conducted in the Liard and Hyland Highlands ecoregions in the southern Yukon Territory to facilitate: a) a statistically significant spatial sampling of the forest cover attributes for these ecoregions; and b) some plot-level ground calibration of the lidar transect data for a range of forest attributes (e.g. Morrison *et al.* 2011).

Given AGRG is located in Nova Scotia, on the east coast of Canada, it was decided to initiate data collection here, then traverse across country toward to the Yukon Territory in the north west, with the objective being to capture as much of the target coverage as possible on the outbound leg and fill any large gaps on the return; i.e. fly the full width of the boreal zone two times. The planned survey route was constrained by the operational limitations of the survey aircraft and the presence of suitably equipped airports located to allow adequate sampling of the target areas whilst maintaining progress across the country. Given the strong latitudinal gradients in forest cover in the boreal zone, it was essential that transect sampling had a north-south component to it, thus dictating a 'zig-zag' flight pattern from east to west and back again.

AGRG's survey plane is a twin engine PA-31 Piper Navajo operated by Scotia Flight Centre. Many remote northern airports in Canada have gravel airstrips and only carry jet fuel for helicopters and fixed wing aircraft with turboprop engines. The Navajo cannot use jet fuel and gravel airstrips should be avoided as it has long twin bladed propellers which mobilise dust and debris, and thus could possibly damage the aircraft or the lidar equipment. Furthermore, when fully laden the Navajo requires a runway exceeding ~ 1,100 m in length. While most airports in southern Canada and major northern towns can accommodate these operational considerations, airport options at the northern extent of the boreal forest target zone were limited.

In addition to the spatial sampling and airport location criteria summarized above, further constraints that needed to be factored into the planning were:

1. Based on a flying speed of 150 knots and endurance of 4.5 hours the planned distance between take off and landing should not exceed ~ 1000 km;
2. Due to budgetary limitations, total survey flying time should not exceed 100 hours;
3. Survey routes must avoid restricted airspace.
4. Transects to be collected between late June and end of August

Table 1: Planned survey transects illustrating the anticipated data acquisition, transit and total flight time listed sequentially from east to west across Canada.

Missions	Survey hrs	Transit hrs	Total hrs
Mobilization/Installation	0	4.0 (NB – NS x2)	4.0
Calibration (NS x 2)	10	0	10
1 Schefferville	3.75	2.25 (NS – QB)	6.0 (two flights)
2 Goose bay	2.75	0.5	3.25
3 Sept Isle	2.5	2.5 (QB – NS)	5.0
4 NFLD South	0.75	2.5 (shared)	0.75
5 NFLD North	2.0	2.5 (NFLD – NS)	4.5
6 RDL - Chibougama	4.5	2.5 (NS – QB)	7.0 (two flights)
7 Timmins	4.0	0.5	4.5
8 Moosonee	3.0	0.5	3.5
9 Marathon	3.25	0.5	3.75
10 Pickle	3.75	0.5	4.25
11 Churchill	3.5	0.5	4.0
12 Flin Flon	3.0	0.5	3.5
13 LaRonge	4.0	0.5	4.5
14 Ft McMurray	4.0	0.5	4.5
15 Yellowknife	3.0	0.5	3.5
16 South (100 hr)	2.5	5.0 (transit service)	7.5 (two flights)
17 Ft Nelson	1.0	4.0 (transit service)	5.0
18 Ft Simpson	1.5	0.5	2.0
19 Watson Lake	2.0	0.5	2.5
ELH1 Watson	2.5	0.5	3.0
ELH2 Watson	3.25	0.5	3.75
20 Whitehorse	3.25	0.5	3.75
21 Norman Wells	3.75	0.5	4.25
22 Ft Simpson	3.0	0.5	3.5
23 Hay River	2.75	0.5	3.25
Return Nova Scotia	0	16.5	16.5 (~ 4 - 5 flights)
Mission total	83.25	48.25	131.5

3. Data Collection

Over a period of 67 days from June 14th to August 20th, 2010, the AGRG undertook 34 individual survey flights traversing 13 UTM zones and over 24,000 km of the Canadian Boreal Forest from Newfoundland (56° W, UTM zone 21) in the east to the Yukon (138° W, UTM zone 8) in the west (Figure 2). All provinces and territories were represented apart from Prince Edward Island and Nunavut (where there is minimal to no boreal forest cover) and the longitudinal gradient sampled represents 23% of the Earth's circumference between latitudes 43° N and 65° N. Survey flights ranged from one to five hours in duration, averaging three hours and 700 line kilometres in length. The entire mission took 127 hrs of flying (including transits). Of this, approximately 91 was used for transect data collection and nine for sensor calibration at the start and end of the mission. Three stops totalling ten days were performed en route for scheduled aircraft maintenance and servicing at Fredericton, Calgary, and Yellowknife airports.

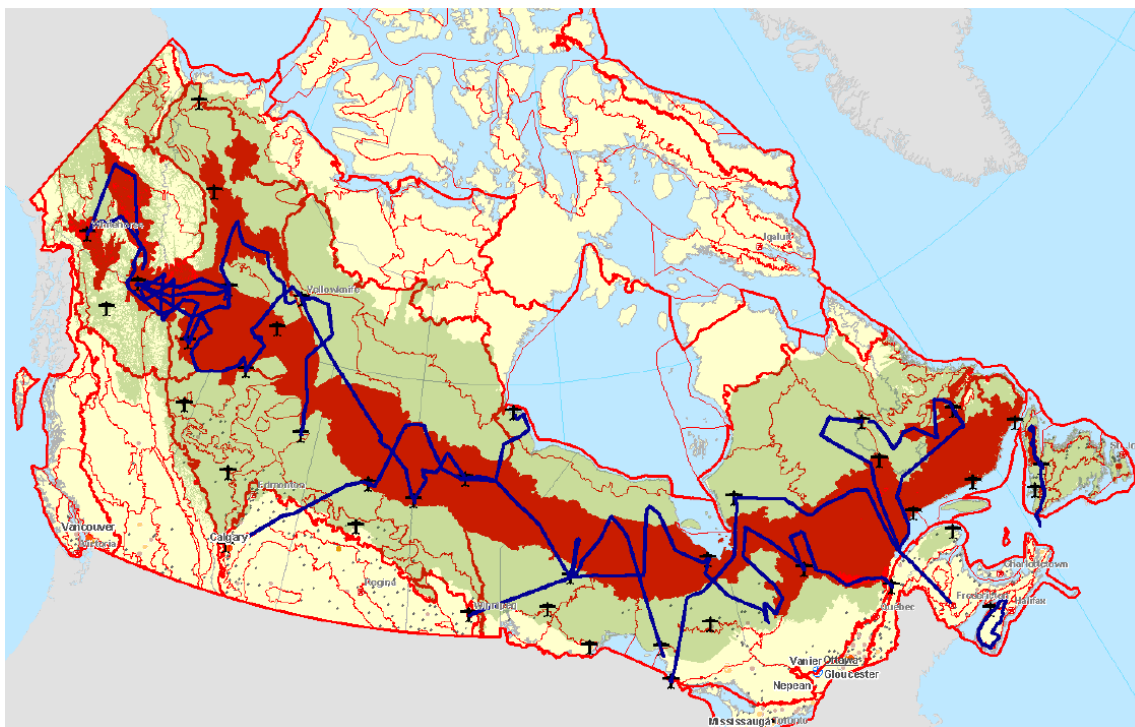


Figure 2: The final lidar sampling transect locations across Canada's Boreal Forest. Area in green represents Canada's boreal forest cover (Brandt, 2009), and the red area illustrates priority ecoregions.

Fire activity in the Boreal Forest during July and August of 2010 was unusually high and this directly impacted approximately one third of the flights by substantially reducing visibility, and forcing diversions away from dense smoke, closed runways and restricted airspace surrounding water bomber activity. A technical problem discovered during calibration and initial test flights was an erratic GPS data gap error due to corroded ground terminals on a radio antenna that passed unfiltered radio signals into the GPS antenna. This intermittent issue resulted in some short data gaps and down time, but did not impact the final data quality. A further logistical challenge encountered concerned the reliability of data contained within the latest Transport Canada Flight Supplement. On three occasions, information concerning fuel and airport service availability was found to be incorrect or out of date. Such minor challenges were expected on a project of this scale but they emphasize the necessity of adaptability and planned contingency.

Due to adverse weather, high relief, excessive fire and smoke conditions, temporary airspace restrictions and airport closures, deviations from the optimal plan were necessary for 24 of the 34 flights. For example, whilst all 34 flights were conducted between altitudes of 450 to 1900 m agl, 11 flights encountered altitudes <900 m agl, and three >1500 m agl (Table 2). Scan angle was kept fixed at 15° for all but four of the flights and PRF kept at 70 kHz for all but seven. Low ceilings forced a scan widening of up to 20°, while high relief dictated a reduction in PRF to 50 kHz. In cases where ceilings or visibility reduced the flying height, data density was minimally impacted and typically increased despite adjusted scan angles. Where relief required a reduction in PRF, data density systematically decreased.

Table 2: Lidar survey transect IDs, timing, flying hours, survey configuration.

Transect	Strips	Survey Flights				Survey configuration		
		JD	Objective/Route	Province	Flying hrs	Alt (m agl)	PRF (kHz)	scan (deg)
		165	Transit + Calibration	NB - NS	2.9			
		166	Calibration	NS	2.4			
Test	1	167	Test transect	NS	3.8	1000-1300	70	15
		169	Transit	NS - QB	2.3			
T01	1	171	Baie Comeau - Goose bay	QB / NFL	3.4	1000-1200	70	15
T02	3	172	Goose bay - Schefferville	QB / NFL	3.2	900-1300	70	15/20
T03	1	172b	Schefferville - Baie Comeau	QB / NFL	3.7	900-1400	70	15
		173	Transit	QB - NB	1.5			
		173b	Test flight	NB	0.9			
		174+	Aircraft service (Fredericton)	NB	0.0			
		186	Transit	NB - NS	0.8			
		187	Transit	NS - NFLD	2.6			
T04	1	188	SW Newfoundland	NFL	1.6	700-1000	70	15
T05	1	192	NW Newfoundland	NFL	1.9	600-1200	70	15/20
		200	Transit	NS - QB	2.0			
T06	2	201	Riviere du loop - Chib	QB	3.3	450-1250	70	15
T07	2	201b	Chibougamau - Val D'Or	QB	4.0	1000-1300	70	15
T08	2	203	Val D'Or - Moosonee	QB / ON	2.1	1000-1200	70	15
T09	2	203b	Moosonee - Pickle lake	ON	4.1	1000-1300	70	15
T10	1	203c	Pickle Lake north loop	ON / MB	1.7	1100-1250	70	15
T11	2	204	Pickle Lake - Winnipeg	MB	2.0	500-600	70	15
T12	1	204b	Winnipeg - Thompson	MB	2.6	700-1150	70	15
T13	3	205	Thompson - La Ronge	MB / SK	3.2	600-1050	70	15
T14	1	205b	La Ronge - Calgary	SK / AB	3.0	1000-1300	70	15
		206+	Aircraft service (Calgary)	AB	0.0			
		210	Transit	AB	2.4			
T15	2	210b	Ft McMurray - Yellowknife	AB / NWT	3.2	900-1250	70	15
T16	3	211	Yellowknife - High Level	NWT / AB	2.5	1150-1300	70	15
T17	2	211b	High Level - Ft Nelson	AB / BC	3.0	750-1000	70	15
T18	6	212	Ft Nelson - Whitehorse	BC / YK	4.4	1200-1500	50	15
T19	1	213	Whitehorse - Watson Lake	YK	3.8	1050-1600	50	15
T20	2	213b	Liard ecozone loop (Watson)	YK	3.2	900-1900	50	15
T21	3	214	Watson Lake - Ft Simpson	YK / NWT	2.0	600-1800	70/50	15/20
T22	3	214b	Ft Simpson south loop	NWT	0.9	1400-1500	50	20
T23	2	214c	Ft Simpson - Watson (plots)	NWT / YK	1.9	900-1900	70/50	15
		215	Aborted	YK	0.2			
T24	4	215b	Watson - Ft Simpson (plots)	YK / NWT	2.5	1200-1400	70/50	15/17
T25	1	215c	Ft Simpson - Yellowknife	NWT	3.6	1200-1300	70	15
		216+	Aircraft service (Yellowknife)	NWT	0.0			
T26	2	218	Yellowknife - Flin Flon	NWT / MB	4.5	1200-1400	70	15
T27	1	218b	Flin Flon - Thompson	MB	2.1	1200-1300	70	15
T28	1	219	Thompson - Churchill	MB	2.0	1200-1250	70	15
T29	1	219b	Churchill - Thompson	MB	2.3	1000-1300	70	15
T30	1	219c	Thompson - Pickle Lake	MB / ON	2.9	1200-1250	70	15
T31	3	220	Pickle Lake - Sioux Ste Marie	ON	4.8	600-1250	70	15
T32	3	223	Sioux Ste Marie - La Grnd Riv.	ON / QB	3.7	1000-1400	70	15
T33	4	223b	La Grnde Riviere - Fredericton	QB / NB	5.1	850-1400	70	15
		230+	Transit + Calibration	NS - NB	6.7			
TOTAL		67			126.7			

4. Transect Data Processing

4.1 Flight light trajectories

Given the need to adapt the sensor and flying configuration to accommodate changing external conditions such as cloud, smoke and terrain relief, the ALTM sensor needed to be stopped and restarted on several occasions in some flights (up to six times in the extreme case). Therefore, during the 34 survey flights, there were actually 69 individual strips of lidar collected, the longest of these being a continuous data stream exceeding four hours in duration. All data were checked in the field shortly after download from the ALTM but final processing took place in the AGRG lab in Nova Scotia.

After download and archival of raw data, the first data processing task was to compute the smoothed best estimated trajectory (sbet) containing both position and orientation data. The ALTM GPS receiver (Trimble BD9500) collected real time GPS signals at 1Hz for the antenna location on top of the aircraft. Meanwhile, multi-axial aircraft accelerations and attitude shifts were recorded at 200Hz at an inertial measurement unit (IMU) located within the sensor head adjacent to the scanner mirror. Trajectory processing uses a Kalman filter to integrate these two data streams to simultaneously estimate and predict the true position and orientation of the aircraft platform. Given the impracticality (high cost and time requirement) of setting up ground base stations at hundreds of locations across Canada, it was originally intended to process all GPS data using Precise Point Positioning (PPP) but after some experimentation it was found that the Canadian active control station (CACS) network and United States continuously operating reference station (CORS) data enabled reasonably accurate differential correction of the airborne GPS trajectory using a 'virtual base station' solution (Boba *et al.* 2008). While the base lines were actually up to several hundred kms in some cases, this capability meant that all points on all trajectories were differentially corrected to accurately known base station locations such that positional errors are anticipated to be better than PPP and likely within 1m throughout

4.2 Points integration

The software used to integrate the raw laser scanner and sbet data was 'Dashmap', a proprietary point processing software package developed by Optech Inc (Toronto, Ontario). Typical user configurable processing settings are the boresight alignment and hardware calibration parameters, factory defaults, range, scanner and altitude masks, atmospheric settings, and intensity normalisation. Output parameters, such as geographic extent, decimation, datum, projection and zone, file formats/paths, etc can also be user defined.

The main Dashmap output definitions for this project were the file format (LAS 1.0), UTM projection (eastings and northings) and the UTM zone for each strip. Because a given transect could cross multiple UTM zones, there were some challenges with zones being incorrectly defined, but these issues were resolved by manually over-riding the default output settings. In total, 69 LAS binary strip files were generated and outputted, containing an average of 300 million points each, ranging from a few million up to around a billion. Both the trajectory outputs and the point cloud data were horizontally and vertically referenced to the International Terrestrial Reference Frame (ITRF), which is equivalent to the WGS84 ellipsoidal datum.

For each emitted laser pulse, there was the possibility of up to four measured returns (first, intermediate, last and single). The echo classification, intensity and scan angle for each pulse are embedded within a LAS file. The LAS binary format is described on the ASPRS web site: http://www.asprs.org/a/society/committees/standards/lidar_exchange_format.html

4.3 Ground Classification

Individual strip file sizes could exceed 30GB and were too large to be handled in most software environments. Therefore a tool was developed to clean the data, classify ground returns, and break the large files down into smaller manageable files of 20 million data points.

‘Lasline’ is a tool developed partially to support the analysis presented here as well as an AGRG project with Nova Scotia Power Inc. to sample and inventory Provincial biomass (Hopkinson *et al.* 2011). Lasline takes raw LAS binary transect files of any size as input then executes the following operations:

1. Data cleaning: isolating high and low laser pulse returns that either float well above the canopy surface or penetrate well below the true ground surface. Such data errors occur due to bird strikes, atmospheric vapour/clouds/aerosols, and/or multi-path of the laser pulse.
2. Ground classification: Ground returns were classified from the transect point cloud using a variant of the algorithm developed by Axelsson (1999) that is also used in Terrascan (Terrasolid, Finland). Prior to implementation, many different parameter sets were tested over various datasets to find a compromise parameter set that produced satisfactory results across a broad range of terrain and land cover scenarios. The Lasline classification routine was found to be faster than the Terrascan routine for an equivalent data volume.
3. Data output: Cleaned and re-classified LAS files were then outputted in 20 million point increments; e.g. for a raw LAS file containing 267million laser points, Lasline would output 13 complete files of 20 million points and one final file of 7 million.

These steps are currently being further expanded to automatically output grid cell-level point cloud metrics in a similar fashion to the USDA tool ‘FUSION’ (McGaughey, 2010). Tools already exist in house that convert point output data to models of forest biometrics, so the intent is to calibrate ‘push button’ tools that automate the workflow from raw LAS binary transect files through to grid-cell level forest attributes ready for input to a GIS.

5. Conclusion

From June to August of 2010, an unprecedented survey of 24,000 km of lidar transects covering >15,000 km² were collected across Canada’s boreal forest. The size of all the LAS files exceeds 500 GB and the coverage equates to ~21 million ‘lidar plots’ at the 25 m grid cell resolution. Each ‘plot’ contains the position and intensity of 1000 to 2000 laser points, which describe the terrain surface and 3D canopy structure, which may then be used for estimating forest inventory attributes (Bater *et al.* 2011). Overall, the completed data collection closely resembled the plan in terms of coverage and timing. The results of this mission represent a rich database describing Canada’s boreal forests during the summer of 2010. These data offer the potential for calibration of wide area satellite-based imagery for spatial upscaling purposes, and will support Canadian government reporting and science programs. Furthermore, in similar fashion to geo-located permanent sample plots, the lidar transect flight path from 2010, or portions thereof, can be re-surveyed in the future to facilitate monitoring of forest development and change in a consistent and quantifiable manner.

Acknowledgements

Many thanks to Heather Morrison, Neville Crasto, Tristan Goulden, Margaret Andrew and Antione Lalumiere for field support. Gary Noell and Jean-Louis Arsenault are gratefully acknowledged for their unending good humour and for taking the Navajo ‘almost’ everywhere

we wanted it! We are also indebted to the brave ‘bird dog’ and water bomber pilots that were kept busy across the Canadian boreal in the summer of 2010; without them, much of the forest canopy data collected would have gone up in smoke! The AGRG lidar laboratory was set up with funds from the Canada Foundation for Innovation. Opportunity funds of the Canadian Forest Service made this lidar-transect project possible.

References

- Axelsson, P. 1999. Processing of laser scanner data—algorithms and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*. Vol. 54 (2-3): 138-147.
- Bater, C. Wulder, M.A., Coops, N.C., Hopkinson, C., Coggins, S.B., Arsenault, E., Beaudoin, A., Guindon, L., Hall, R.J., Villemaire, P., Woods, M. 2011. Model development for the estimation of aboveground biomass using a lidar-based sample of Canada’s boreal forest. *Proceedings of the SilviLaser 2011 Conference*, Oct. 16-20, Hobart, Tasmania
- Boba M., Ussyshkin, V., Slater, M., Sitar, M., Szameitat, W. 2008. Impact of an optimized position and orientation system on the final accuracy of lidar data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XXXVII. Part B3b. Beijing 2008
- Brandt, J. 2009. The extent of the North American boreal zone. *Environmental Reviews*. Vol. 17: 101-161.
- Evans, J.S Hudak, A.T., Faux, R., Smith, A.M.S. 2009. Discrete return lidar in natural resources: recommendations for project planning, data processing and deliverables. *Remote Sensing* Vol. 1: 776-794.
- Gillis, M.D., Omule, A.Y., Brierley, T. 2005. Monitoring Canada’s forests: The National Forest Inventory. *Forestry Chronicle*. Vol. 81 (2): 214-221.
- Hopkinson, C. 2007. The influence of flying altitude and beam divergence on canopy penetration and laser pulse return distribution characteristics. *Canadian Journal of Remote Sensing*. Vol. 33: 312-324.
- Hopkinson, C., Colville, D., Bourdeau, D., Monette, S., Maher, R. 2011. Scaling plot to stand-level lidar to public GIS data in a hierarchical approach to map forest biomass. *Proceedings of the SilviLaser 2011 Conference*, Oct. 16-20, Hobart, Tasmania
- McGaughey, R.J. 2010. Fusion/LDV: Software for LiDAR Data Analysis and Visualization. USDA Forest Service. FUSION Manual. March 2010 – FUSION Version 2.80.
- Morrison, H., Hopkinson, C., Wulder, M.A. 2011. Optimal LiDAR gridding parameterization for effective leaf area estimation in the boreal forest Yukon Territory, Canada. *Proceedings of the SilviLaser 2011 Conference*, Oct. 16-20, Hobart, Tasmania
- Næsset, E. 2009. Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*. Vol. 113 (1): 148-159.
- Wulder, M. A., Han, T., White, J. C., Sweda, T., & Tsuzuki, H. 2007. Integrating profiling LIDAR with Landsat data for regional boreal forest canopy attribute estimation and change characterization. *Remote Sensing of Environment*. Vol. 110(1): 123-137.

Assessing the accuracy of GLAS topography estimation by using airborne Light Detection And Ranging (LiDAR) measurements

Han Meng¹, Bernard Devereux¹, Gabriel Amable¹

¹Geography Department, the University of Cambridge, Cambridge CB2 3EN, UK. hm373@cam.ac.uk

Abstract

Topography estimation is a key factor in forestry studies. The accurate prediction of topography underneath tree canopies will certainly improve the subsequent forest bio-physical characteristics estimation such as tree height, stem volume, biomass/carbon stocks. Thus, the assessment of the accuracy of GLAS topographical estimation is essential before the data can be used for forest bio-physical characteristics prediction. This study proposes the use of airborne LiDAR measurements to assess GLAS ground elevation estimates in a mixed woodland and arable site in south-east England near Thetford, UK, at 52.4°N, 0.81°E, given that airborne LiDAR measurements have already been validated using 'ground-truth' data. GLAS full waveforms are decomposed into up to six Gaussian modes and different indices, such as waveform centroid position (GLA14 position) and GLA01 last peak position, are calculated based on the peak positions of these Gaussian modes. Elevations estimated from these indices are compared with airborne LiDAR elevation estimates for assessment purpose and optimal estimates will be selected based on the results.

Four comparison models are introduced in this study. From these, model 1 (the comparison between GLA14 elevation and non-filtered airborne last return pulses elevation) and model 4 (the comparison between GLA01 last mode elevation and filtered airborne last return pulses elevation) have the best performance with R-squared values of 0.89 and 0.87, respectively, and RMSE values of 3.82 and 4.69, respectively. After removal of outliers for model 4, the R-squared value improves to 0.99 and the RMSE value reduced significantly to 0.66.

A simplified experiment is implemented in this study in order to investigate the impacts on biomass/carbon stock estimates arising from use of different models, with the assumption that there is a uniform average tree height of 20 meters and uniform stem density through the study site.

Key words: GLAS, ICESat, LiDAR, Airborne LiDAR, Topography, Waveform

1 Introduction

The Geoscience Laser Altimeter System (GLAS) was launched in 2003 on-board ICESat by the National Snow and Ice Data Centre (NSIDC). Its primary mission is to measure ice sheet properties. GLAS is a full-waveform LiDAR instrument and has nearly global data coverage. GLAS is also the first satellite-borne Light Detection and Ranging (LiDAR) instrument for continuous global observations of the Earth. Due to its land measurement properties, GLAS has been used in lots of forestry studies.

[Ranson *et al.* \(2004\)](#) explored the possibility of using GLAS waveform data to help with the land cover classification results obtained from MODIS data. [Harding and Carabajal \(2005\)](#) investigated

GLAS measurements of within-footprint topographic relief and vegetation vertical structure. They introduced several common vegetation vertical structure biophysical parameters such as maximum canopy height, crown depth and topographic relief, the upper-most canopy ruggedness, as well as canopy cover, and with procedures for estimation from GLAS waveforms. [Duong et al. \(2006\)](#), [Neuenschwander et al. \(2008\)](#), and [Pang et al. \(2008\)](#) also presented approaches to using GLAS data for retrieval of forest vertical structure parameters such as tree height, etc.

Topography estimation is a key factor in forestry studies. The accurate prediction of terrain height underneath tree canopies will certainly improve the subsequent estimation for forest bio-physical characteristics such as tree height, stem volume and biomass/carbon stock. Thus, the assessment of the accuracy of GLAS topographical estimation is essential before the data are used for subsequent estimation. Exploiting the very high accuracy of airborne LiDAR (15cm RMSE vertically), this study proposes an approach to carry out this assessment by using airborne LiDAR data, provided that the accuracy of airborne LiDAR measurements has already been precisely assessed.

The assumption of this study is that airborne LiDAR measurements are 100% true and can be used to measure the accuracy of GLAS topography estimation and to reveal error budgets in GLAS products when estimating ground elevation. The objectives of the study are to assess the accuracy of GLAS data products in topography estimates by using either non-filtered or filtered airborne LiDAR measurements; to investigate the impacts of using different data products under different models on terrain height estimates and the subsequent forest bio-physical parameters estimates; and to establish the relationships between GLAS and airborne LiDAR measurements. The outcome of this study can assist with a better understanding of GLAS land measurements and also the possibility of using GLAS for terrain surface modelling in biomass/carbon stocks mapping exercises.

2 Study site and Data

The study was implemented at a mixed forest site in Thetford, dominated by oak (*Quercus*) and conifers. It has an area of about 116 ha and is located to the east of Thetford main forest, East Anglia, UK, at 52.4°N, 0.81°E (figure 1). This site mainly consists of pines and broadleaves, with relatively dense under storey vegetation such as hazel (*Corylus colurna*) and hawthorn (*Crataegus laevigata*).



(a) The black dot shows the location of this study site, 52.4°N, 0.81°E. (Source:



(b) A screenshot of the Thetford woodland site from google earth.

ESRI data)

Figure 1: Location and a screenshot of Thetford woodland site

Throughout this area, there are 138 GLAS footprints available for the four-year period from October 2003 to March 2007. ICESat Science Investigator-led Processing System (I-SIPS) at the NASA's Goddard Space Flight Center created 16 standard GLAS data products (GLA01 to GLA16). Land products GLA01 and GLA14 were used in this study.

The GLAS footprints are elliptical rather than circular and their shapes vary for different lasers and operational periods. In practice, it is very difficult to get exactly the same elliptical field site for validation or for science research purposes, so the equivalent circular diameter (64m) has been very frequently used to approximate the footprint shape on ground (Abshire *et al.*, 2005 and Harding and Carabajal 2005). Thus, in this study we use the equivalent 64-meter-diameter circles to represent GLAS footprints.

The trees were leaf-on when the Thetford airborne LiDAR survey took place in June 2009. The ALTM 3033 LiDAR instrument collects 33,000 laser observations per second, with a 40° field of view, a 728m swath at 1,000m flying height and a 20cm laser footprint size at nadir view. The data density is about 1 shot per square meter. As indicated by the system supplier, this LiDAR instrument has an absolute RMS accuracy of better than plus or minus 15cm. About 2GB of data for this area including first-and-last return laser scanning data, intensity data and waveforms were recorded. This study will only use the last return laser scanning data.

3 Methodology

3.1 The pre-processing of GLAS data

The pre-processing of GLAS data mainly dealt with two issues: Saturation effect and coordinate transformation.

Saturation effect is one of the problems that cannot be neglected before GLAS data is used for scientific purposes. GLAS data contains a variety of different variables that have different meanings. These variables can be derived by the variable-deriving software provided by NSIDC, which needs an Interactive Data Language (IDL) environment to run. Amongst these variables, 'I_SatCorrFlg' is the one that indicates if there is a saturation effect on that particular record and how big the effect is. The value of 'I_SatCorrFlg' ranges from 0 through 4. There is no saturation effect if it equals 0; if it is valued less than or equal to 1 or 2, the saturation effect could be corrected by adding the estimate of elevation bias due to saturation reported in i_SatElevCorr; however, if it is 3 or 4, it indicates that the corresponding records are invalid and should be abandoned.

For GLAS land surface data product (GLA14), the 'GLAS Altimetry Product Usage Guidance' provided by NSIDC website has suggested that all the records that have a saturation indicator value larger than 0 should be simply abandoned for scientific research purposes.

Apart from saturation effect, coordinate transformation is another issue that needs to be considered in the data pre-processing stage. In order to compare the performance of GLAS data and that of airborne laser scanning data, the coordinates of these data must be converted into the same system.

GLAS data are initially in the TOPEX/Poseidon reference frame and the airborne laser scanning data used in this study are in OSGB coordinates. It is more convenient and straightforward to process data with units in meters, so the coordinates of GLAS data were converted into OSGB coordinates.

NSIDC provides several coordinate conversion tools that were written using IDL Scripts. The TOPEX to WGS84 conversion tool was first used to convert GLAS TOPEX coordinates into WGS84 ellipsoid. It was stated that the error associated with this conversion is less than 1cm (Duong *et al.*, 2009). Then the converted GLAS data coordinates were transformed into the OSGB reference system using a web-based coordinate conversion tool provided by the UK Ordnance Survey ([ordnance survey website](http://www.ordnancesurvey.co.uk)).

3.2 Elevation derivation from GLAS data

In this study, two GLAS waveform parameters are considered as a possible basis for terrain elevation estimates. These were GLA01 last mode centroid and the centroid of the entire waveform (elevation reported in GLA14). GLA01, the global altimetry data, records the full waveform information. GLAS footprints over vegetated areas usually have a bimodal waveform, shown in figure2.

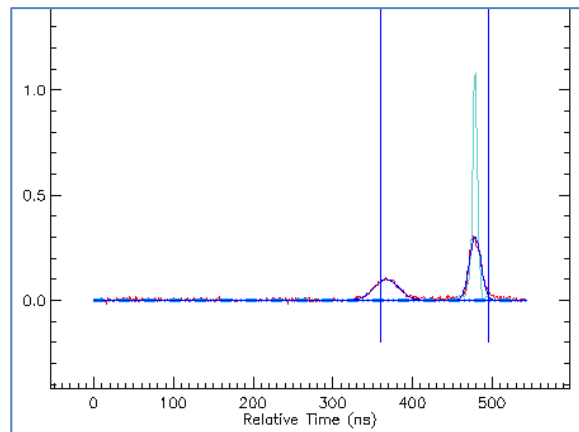


Figure 2: A GLAS bimodal waveform

The narrow and abrupt mode has been considered to be the representative of the underlying ground (Rosette *et al.*, 2008). In figure 2, the light blue narrow and tall mode is the transmitted waveform, same in figure 3. For the convenience of data processing and analysis, GLAS full waveforms can be decomposed into up to six Gaussian modes by using the processing software provided by NSIDC. Amongst these Gaussian modes, the centroid position of the last mode (GLA01 last mode elevation) will be used in this study as one potential indicator of the terrain elevation. Another indicator that will be used in this study is the elevation reported in the second level data product GLA14, which is represented by the centroid position of the complete waveform. Figure3 shows the positions of these two parameters in a GLAS waveform.

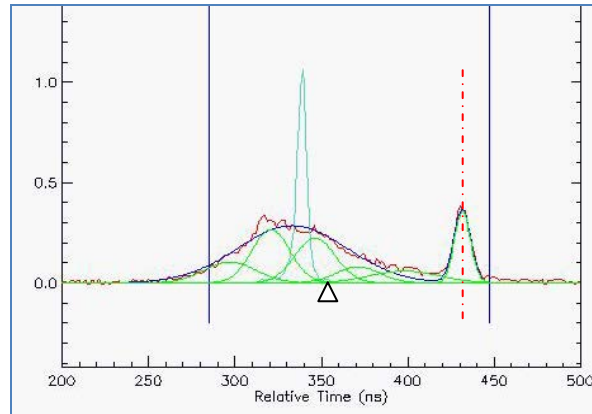


Figure3: The positions of GLA01 last mode (the red dashed vertical line) and GLA14 (small triangle) in a GLAS full waveform. The bimodal waveform (smooth blue curve) was decomposed into six Gaussian modes (green curves).

3.3 The filtering and interpolation of airborne LiDAR data: algorithms

This section will state why airborne LiDAR last return pulses need to be filtered and the filtering algorithm (Popescu *et al.*, 2002) that will be used in this study to filter the data for estimation of the 'ground' Digital Terrain Model (DTM). Previous work in another forest site in the UK - Woodwalton Fen - will be used here to guide selection of the optimized window size for this algorithm. Also, the derivation of airborne LiDAR data within the GLAS footprints will be introduced.

Airborne LiDAR first-and-last return data has been widely used in forestry studies. Literally, the first returns should come from the very top surface of the objects that the laser beam is hitting and the last returns should come from the very bottom, ideally the ground, underneath the tree canopy. However, not all the transmitted laser pulses will be able to penetrate the canopy and reach the ground, especially in densely vegetated areas. Airborne last return pulses therefore, may come from tree branches, under storey vegetation or the terrain surface. If all the last return pulses are simply integrated to generate the terrain model without any filtering process, it is very likely to overestimate the true terrain height. Figure 4 shows the overestimation of airborne last return laser pulses on elevation estimates.

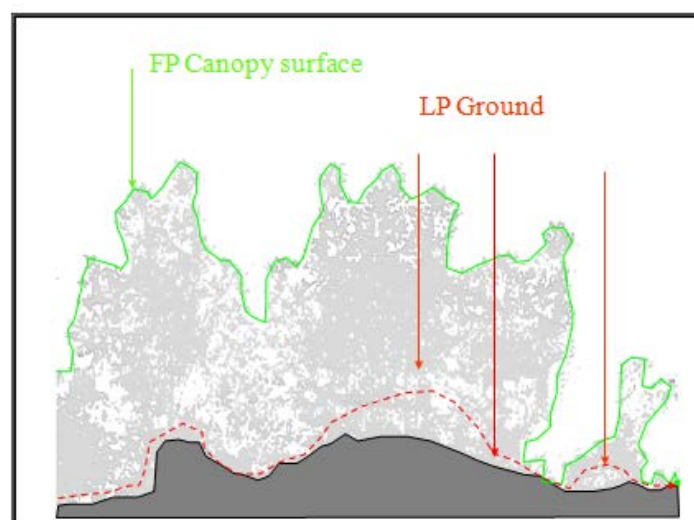


Figure 4: Green arrow represents the First return Pulses (FP) and green curve is the FP Canopy surface; Red arrows represent the Last return Pulses (LP) and the red dashed curve is the terrain surface generated by using LPs (source: [Devereux and Amable, 2009](#))

Accordingly, this study adopted the filtering algorithm proposed by [Popescu *et al.* \(2002\)](#) (refer to the study for details of the algorithm), aiming to filter out as many non-ground returns as possible and to reduce overestimation. Some experiments have been implemented prior to this study ([reference to Meng, H., in prep](#)) in another UK-based study site Woodwalton Fen, trying to find the optimal window size for this algorithm in a boreal forest. Based on these results, a 9m by 9m window was used in this study.

The approximate average diameter of a GLAS footprint is about 64m. In order to compare with GLAS estimates, airborne LiDAR last returns were derived based on this diameter. Subsequent analysis is only based on laser shots that are within the GLAS footprints (figure 5). Both 9m by 9m filtered and non-filtered last returns are used in this study, for the purpose of comparison.

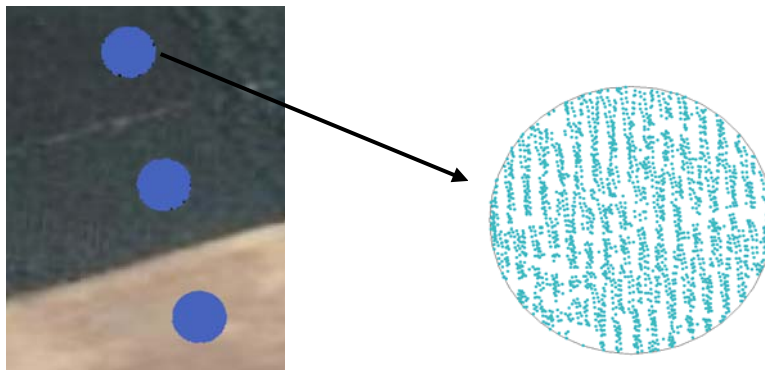


Figure5: Airborne laser shots (blue dots) within GLAS footprints

Figure 5 shows the derivation of airborne LiDAR data within GLAS footprints. The big circles represent GLAS footprints with an average diameter of 64m. The blue dots inside the big circles are the airborne LiDAR laser shots.

3.4 The assessment of GLAS topography estimates accuracy

Topography estimates from the two potential indicator parameters, GLA01 last mode elevation and elevations reported in GLA14 data product, are compared against the average elevation estimates from both non-filtered and filtered airborne LiDAR last returns. For each of the GLAS footprint, there will be two elevation values calculated as the averages of non-filtered and filtered airborne last returns elevation. A linear regression model is used to establish the relationship between these measurements. Figure 6 summarizes the algorithms used for the main methodology described above (sections 3.1-3.4).

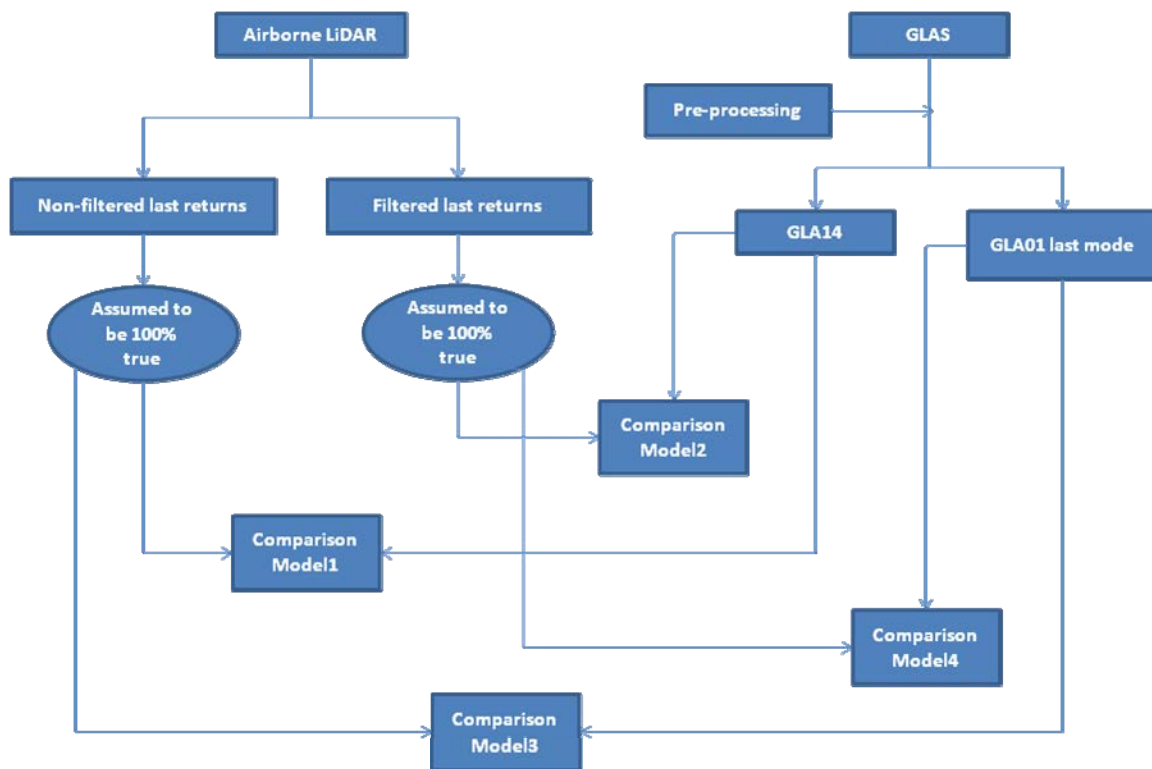


Figure 6: The procedure for the main methodology part

3.5 Stem volume/ biomass estimates - a simplified experiment

This experiment was implemented to demonstrate the impacts of topography estimation from different models on stem volume/biomass estimates. The assumption was that the airborne DTM produced from non-filtered last return pulses or filtered last return pulses is 100% true and there is a uniform average tree height of 20m throughout the entire study site. By comparing GLAS terrain elevation estimates using different data products (GLA01 last mode and GLA14) with airborne LiDAR DTM, four models were generated (table1): Model 1, the comparison between GLA14 elevation and non-filtered airborne LiDAR last returns elevation; Model 2, the comparison between GLA14 elevation and filtered airborne LiDAR last returns elevation; Model 3, the comparison between GLA01 last mode elevation and non-filtered airborne LiDAR last returns elevation; Model 4, the comparison between GLA01 last mode elevation and filtered airborne LiDAR last returns elevation.

Table 1:

The four models from comparisons based on different ‘true grounds’

Assumed ‘true ground’	GLA14	GLA01 last mode elevation
Non-filtered airborne last returns	Model 1	Model 3
9m by 9m Filtered airborne last returns	Model 2	Model 4

For each of these models, there will be terrain height estimates from GLA14 and GLA01 last mode and also average elevation estimates from non-filtered and filtered airborne LiDAR last return pulses for every GLAS footprint. By subtracting GLAS elevation estimates from airborne elevation estimates,

terrain height overestimation/underestimation from that particular GLAS data product can be obtained. Then, the assumed average tree height of 20m will be added to this value to get the estimated tree height by using GLAS topography measurements. The calculations follow equations (1)-(4) for the four models.

$$\text{Model 1 } H = \sum_{i=1}^{127} ((\text{non-filtered ALS}[i] - \text{GLA14}[i]) + 20) \quad (1)$$

$$\text{Model 2 } H = \sum_{i=1}^{127} ((\text{filtered ALS}[i] - \text{GLA14}[i]) + 20) \quad (2)$$

$$\text{Model 3 } H = \sum_{i=1}^{127} ((\text{non-filtered ALS}[i] - \text{GLA01 last mode elevation}[i]) + 20) \quad (3)$$

$$\text{Model 4 } H = \sum_{i=1}^{127} ((\text{filtered ALS}[i] - \text{GLA01 last mode elevation}[i]) + 20) \quad (4)$$

Results will be compared for the four models. As tree height is the key parameter for estimating subsequent forest bio-physical parameters such as stem volume, above-ground biomass and carbon stocks, this experiment gives us a rough idea about how much the difference of estimation could be resulting from the use of the different models.

4 Results

4.1 The pre-processing of GLAS data

There are 138 GLAS footprints available for the Thetford site. During the pre-processing, 11 footprints collected in October 2005 were abandoned due to saturation effects. Thus, only the rest 127 GLAS footprints are analyzed in this study.

After the coordinate transformation process, all the 127 GLAS footprints were converted into OSGB coordinate system. Figure 7 shows the Digital Surface Model of the Thetford site overlaid by GLAS footprints represented by back circles.

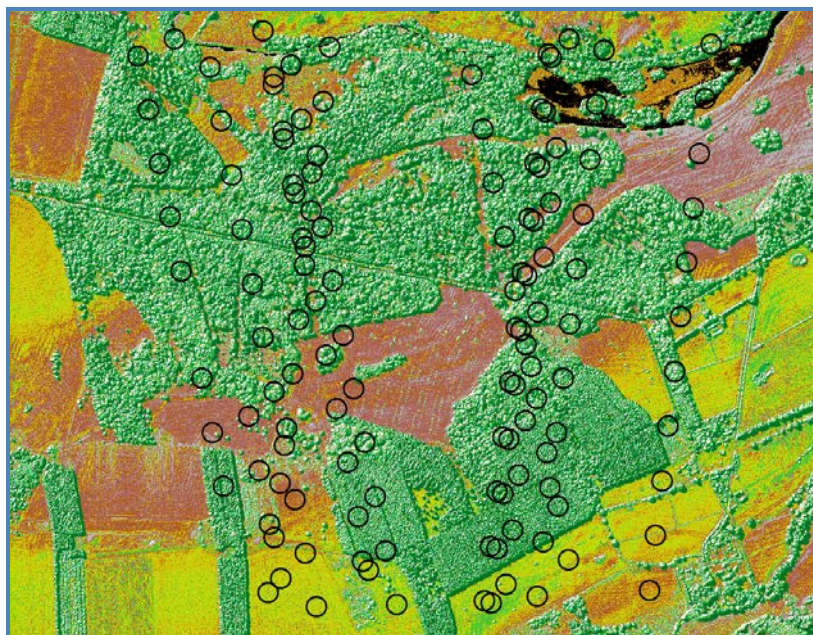


Figure 7: Coordinate conversion results – GLAS footprints overlying the DSM of the Thetford site

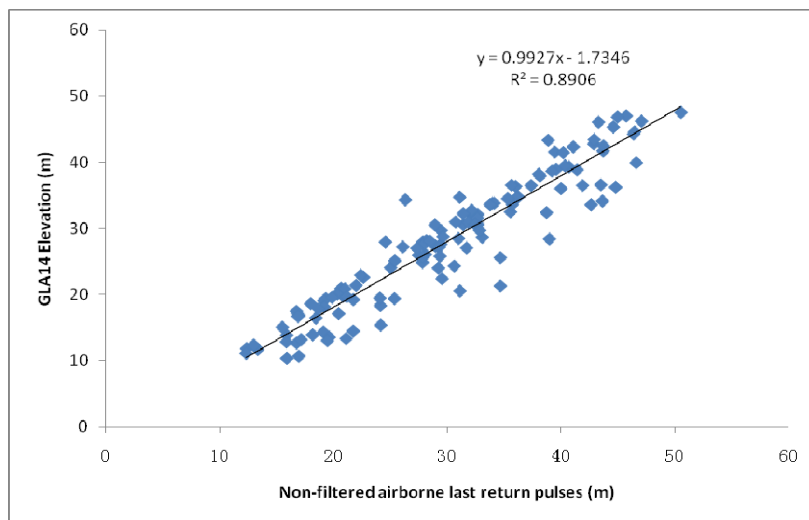
4.2 The filtering and interpolation of airborne LiDAR data: results

The filtering algorithm proposed in [Popescu et al. \(2002\)](#) with a 9m*9m window size was used to filter the airborne LiDAR last return pulses in order to get a less biased terrain surface model. The filtered data has a much lower density of 0.24 points per square meter. Both filtered and non-filtered airborne LiDAR last return pulses that are located within the GLAS footprints were derived for the convenience of analysis. With an average diameter of 64 meters, the area of a GLAS footprint is about 3215 square meters. Thus, there will be roughly 3215 non-filtered airborne laser shots and about 771 filtered airborne laser shots located within each of these GLAS footprints, although the actual laser shots number varies across the whole area.

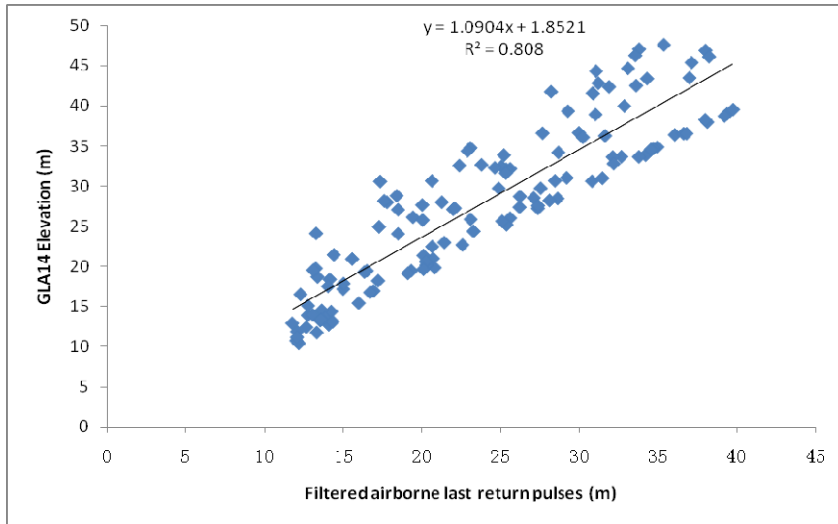
4.3 The assessment of GLAS topography estimation accuracy

We investigated the performance of GLAS measurements for topography elevation estimates by comparing them with airborne LiDAR measurements which have been assumed to be 100% true in this study. The comparisons are shown in figure 8 (a-d).

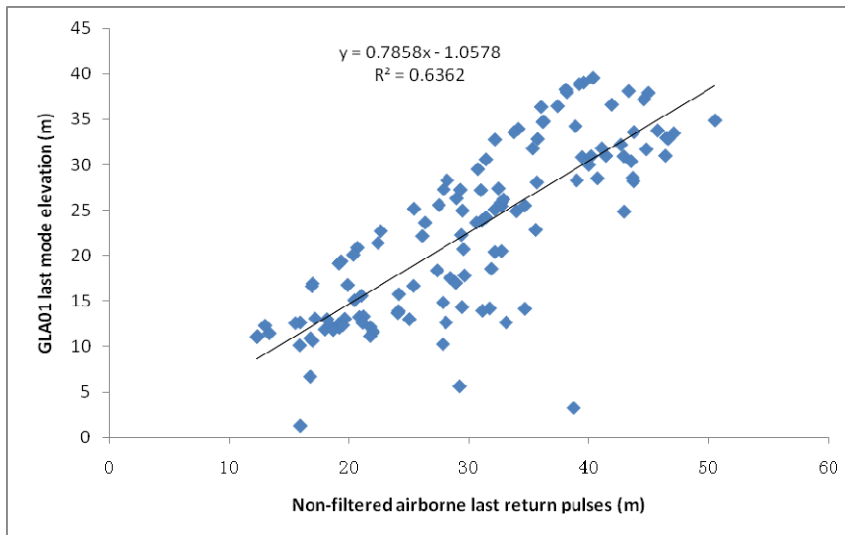
In figure 8 (a) and (c), non-filtered airborne LiDAR last return pulses are assumed to be true and GLA14 and GLA01 last mode elevations are compared with the 'true ground', presented as model 1 and model 3 respectively. Under model 1, an R-squared value of 0.8906 was obtained, which indicates a very strong relationship between non-filtered airborne last return pulses and GLA14 elevation. However, GLA01 last mode elevation (figure 8 (c)) does not seem to match well with this airborne LiDAR measurement, with an R-squared value of 0.6362.



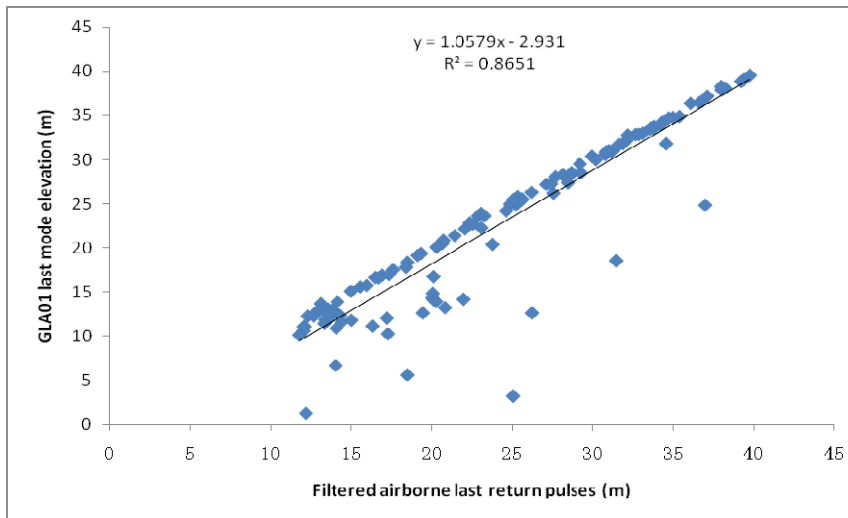
(a)



(b)



(c)



(d)

Figure 8: The regression results from the four models

In the case of figure 8 (b) and (d), 9m by 9m filtered airborne last return pulses are assumed to be the true ground and the performance of GLA14 and GLA01 last mode elevation are again assessed under this assumption. Quite strong relationships are found for these two models, an R-squared value of 0.808 for GLA14 and 0.8651 for GLA01 last mode elevation. Clearly, of all these four models, GLA01 last mode elevation has the strongest relationship with filtered airborne last return pulses.

The pattern shown in figure 8 (d) is very interesting in that the majority of the data form a very straight line with a slope close to 1 and there are several footprints' data clearly away from this line. These data are considered as outliers in this study. A 2m criteria was used to detect these outliers: GLAS footprints with GLA01 last mode elevation 2m larger or less than the corresponding filtered airborne last return pulses average elevation within that footprint were detected as outliers and deleted. This process detected 24 outliers throughout this area. Figure 9 shows the new regression result without these outliers. An R-squared value of 0.9953 was achieved and the slope of the regression line is 0.9726, very close to 1. Table 2 summarizes the elevation estimates for the four models and model 4 with outliers deleted.

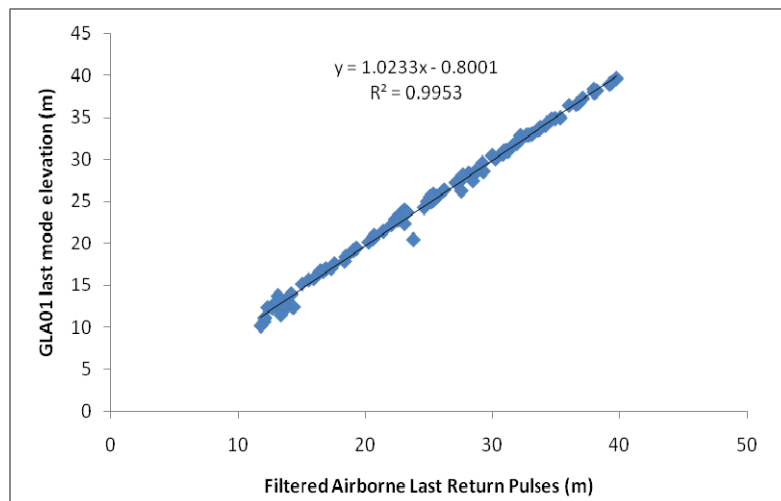


Figure 9: The regression result of Model4 with the detected outliers deleted

Table 2:

The elevation estimates summary for the four models and model 4 without the outliers

	Model 1	Model 2	Model 3	Model 4	Model 4 without outliers
Average difference ¹ (m)	1.95	-4.02	7.84	1.86	0.22
StDev ² of the differences (m)	3.30	4.43	6.55	4.33	0.62
RMSE (m)	3.82	5.97	10.20	4.69	0.66
R-squared	0.89	0.80	0.64	0.87	0.99

¹ Average difference = $(\sum_{i=1}^{127} \text{variances}[i])/127$, negative value means underestimation

² Standard Deviation

4.4 The impacts on stem volume/ biomass estimates

From equations (1) – (4) in section 3.5, tree height was calculated for each of the models: Model 1 H, Model 2 H, Model 3 H, and Model 4 H. For 127 GLAS footprints, the sum ‘true tree height’ is 127 × 20 = 2540m. Difference between the accumulated model height and the accumulated true height were then obtained and the overestimation/underestimation percentage was calculated by using equation (5).

$$\text{Overestimation/underestimation percentage} = \frac{\text{differences}}{\text{accumulated true height}} \times 100\% \quad (5)$$

Table3:

Results: comparison results from this experiment. Model 1 and Model 4 give much better estimates than Model2 and Model3

Model heights	Accumulated Model Height (m)	Accumulated True Height (m)	Differences	Overestimation/underestimation percentage
Model 1 H	2788.14	2540	248.14	9.77%
Model 2 H	2028.93		-511.07	-20.12%
Model 3 H	3535.64		995.64	39.20%
Model 4 H	2776.43		236.43	9.31%

5 Conclusions and discussion

Of the four models tested in this study, model 1 and model 4 gave the strongest relationship between the topography measurements of GLAS data products and airborne LiDAR, with R-squared values of 0.8906 and 0.8651 respectively and RMSE values of 3.82 and 4.69 respectively. The two elevation indicators of GLAS data involved in this study are: GLA14 waveform centroid and GLA01 last mode elevation.

The GLA14 waveform centroid elevation is the measurement of the centroid position of the full GLAS waveform. It averages the object elevations within the footprint such as trees, under storey vegetation and ground. Thus, the surface generated by using GLA14 elevation should be higher than the actual ground and lower than the tree canopy top surface. For non-filtered airborne LiDAR last return pulses, some of them actually hit the ground while the others may just end up hitting tree branches or under storey vegetation before reaching the ground. Therefore, the terrain surface generated from these original last return pulses should also be between the real terrain surface and tree canopy top surface (figure 4). This explains why GLA14 elevation and the average elevation of non-filtered airborne LiDAR last return pulses have such a strong relationship (model 1).

In contrast, GLA01 last mode is from the centroid position of the last Gaussian mode, generated by decomposing the full waveform into up to six Gaussian modes. This represents the lowest elevation within a GLAS footprint. The purpose of the filtering process of airborne LiDAR data is to filter out

the non-terrain points. Thus, compared to the terrain surface generated by using non-filtered airborne LiDAR last return pulses, the one generated by using filtered airborne LiDAR last return pulses is lowered and closer to the real ground. This explains the regression results presented by model 4 which show that GLA01 last mode elevation has a very strong relationship with the average elevation of filtered airborne LiDAR last return pulses.

Although very strong relationships are found in these models, there are some very obvious outliers, especially in model 4 (figure 8 (d)). A criteria of 2 meters was used to detect these outliers. After removing these outliers, the regression result got improved a lot with an R-squared value of 0.9953 and an RMSE value of 0.66. However, we do not really know the reason why 2 meters should be chosen here for outlier detection and this outlier detection method should not be simply replicated to other sites without tests. There are some possible explanations for these outliers. First, GLAS footprints are from different years (from 2003 to 2007) and seasons (February, May and October), which may cause some inconsistency in the data itself. Second, the decomposition process of the full waveform may also have some error. Third, airborne data was collected in 2009, and GLAS data are collected from 2003 through 2007. The time gap between data collection could also be one of the bias causes. Another reason for the outliers might be the lack of GLAS penetration especially in some densely vegetated areas. Terrain slope can be another factor that impacts on the performance of GLAS measurements, which needs further investigation in our future research.

In this simplified experiment, it is also possible that the results may be slightly biased by the difference of tree density through the whole site. Nevertheless, it does demonstrate how much the difference in stem volume/biomass/carbon stocks estimates could be as a result of using different models. When the non-filtered airborne LiDAR last return pulses are assumed to be representing the real terrain surface, the overestimation using GLA14 elevation is 9.77% (model 1) while that using GLA01 last mode elevation is 39.20% (model 3). Similarly, when filtered airborne LiDAR last return pulses are assumed to be representing the real ground (GLA14) the underestimation is 20.12% (model 2). This figure reduces to 9.31% by using GLA01 last mode elevation (model 4). This experiment was carried out for the range of 127 GLAS footprints with a total area of about 40 hectares. The percentage values may vary when taking stem density through the forest into account. Over larger areas, stem volume/biomass/carbon stock estimates could be substantially wrong if inappropriate models are used.

Our study assessed GLAS topography estimates accuracy based on the assumption that airborne LiDAR measurements are 100% true. The assessment results show a very promising potential for the use of GLAS products in large scale biomass/carbon stocks mapping. The impacts of using different GLAS data products on biomass/carbon stocks estimates have also been investigated in this study by using a simplified experiment, based on the assumption that there is a uniform average tree height of 20 meters. The results give us a rough idea of how much the overestimation/underestimation could be by using different models. This addresses the importance of choosing the right model for estimating biomass/carbon stocks by using GLAS data products. Relationships between GLAS data products and airborne LiDAR measurements were also established in this study. It was found that GLA14 has a very strong relationship with non-filtered airborne LiDAR last return pulses and GLA01 last mode has a very strong relationship with filtered airborne LiDAR last return pulses. In future research, the relationship between ground-based measurements and airborne LiDAR measurements

will be established. Together with the relationship between GLAS and airborne LiDAR measurements from this study, it will assist with regional/global biomass/carbon stocks mapping.

Acknowledgements

Many thanks to the US National Snow and Ice Data Center (NSIDC) for providing the GLAS data used in this research and the Unit for Landscape Modelling, the University of Cambridge for collecting the airborne LiDAR data. Dr. James Morison and Mr. Peter Crow working in the Forest Research Agency of the Forestry Commission is acknowledged for providing very helpful advices for this research. Geography Department, the University of Cambridge is also appreciated for supporting part of the travel grant to attend this conference.

6 Bibliography

Abshire, J.B., Sun, X., Riris, H., Sirota, M., McGarry, J., Palm, S., Yi, D., Liiva, P., 2005, Geoscience Laser Altimeter System (GLAS) on the ICESat Mission: On-orbit Measurement Performance, doi: 10.1029/2005GL024028.

Devereux, B. and Amable, G., Airborne LiDAR: Instrumentation, Data Acquisition and Handling, in Heritage G. And Large A., 2009, Laser Scanning for the Environmental Sciences, pp. 49, ISBN: 978-1-4051-5717-9, Wiley.

Duong, H., Pfeifer, N. and Lindenbergh, R., 2006, Analysis of repeated ICESat full waveform data: Methodology and leaf-on/leaf-off comparison, Workshop on 3D Remote Sensing in Forestry, 14th – 15th Feb 2006, Vienna.

Duong, H., Lindenbergh, R., Pfeifer, N. And Vosselman, G., 2009, ICESat Full-Waveform Altimetry Compared to Airborne Laser Scanning Altimetry Over The Netherlands, IEEE Transactions on Geoscience and Remote Sensing, Vol. 47, No. 10.

Harding, D.J., Carabajal, C.C., 2005, ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure, Geophysical Research Letters, Vol. 32, L21S10, doi: 10.1029/2005GL023471.

Meng, H., The use of Light Detection And Ranging (LiDAR) for forest vertical structure studies, PhD thesis in prep, Geography Department, the University of Cambridge.

Neuenschwander, A.L., Urban, T.J., Gutierrez, R., Schutz, B.E., 2008, Characterization of ICESat/GLAS waveforms over terrestrial ecosystems: Implications for vegetation mapping, Journal of Geophysical Research, Vol. 113, G02S03, doi: 10.1029/2007JG000557.

Ordnance survey website: <http://gps.ordnancesurvey.co.uk/convert.asp>.

Pang, Y., Lefsky, M., Sun, G., Miller, M.E., Li, Z., 2008, Temperate forest height estimation performance using ICESat GLAS data from different observation periods, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII. Part B7, Beijing.

Popescu, S.C., Wynne, R.H., Nelson, R.F., 2002, Estimating plot-level tree height with LiDAR: local filtering with a canopy-height based variable window size, *Computers and Electronics in Agriculture*, Vol. 37, pp. 71-95.

Rosette, J.A.B., North, P.R.J. and Suarez, J.C. (2008) 'Vegetation height estimation for a mixed temperate forest using satellite laser altimetry', *international Journal of Remote Sensing*, 29:5, 1475-1493.

Ranson, K.J., et al., 2004, Landcover attributes from ICESat GLAS data in central Siberia, *IEEE*.

Characteristics of satellite LiDAR waveform in tropical rain forests from the comparison with canopy condition derived from high resolution satellite data

Yasumasa Hirata¹, Mui How Phua², Hideki Saito³, Toshiya Matsuura⁴, Wilson Wong², Alexius Korom⁵ & Kanehiro Kitayama⁶

¹Climage Change Office, Forestry and Forest Products Research Institute, Tsukuba, Japan, e-mail: hirat09@affrc.go.jp

²School of International Tropical Forestry, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia, e-mail: pmh@ums.edu.my, w.wilson@ums.edu.my

³Kyusyu Research Center, Forestry and Forest Products Research Institute, Kumamoto, Japan, e-mail: rslsaito@ffpri.affrc.go.jp

⁴Department of Forest Management, Forestry and Forest Products Research Institute, Tsukuba, Japan, e-mail: matsuu50@affrc.go.jp

⁵Universiti Teknologi MARA Sabah Branch, Kota Kinabalu, Sabah, Malaysia

⁶Graduate School of Agriculture, University of Kyoto, Kyoto, Japan

1. Introduction

Tropical forest plays an important role in providing various functions, such as carbon sink, maintenance of biodiversity, but deforestation and forest degradation occur in many developing countries. The importance of a robust transparent national monitoring system was recognized in COP 16 of United Nation Framework Convention on Climate Change and a combination of remote sensing and ground-based forest carbon inventory approaches for estimating, as appropriate, anthropogenic forest-related greenhouse gas emissions by sources and removals by sinks, forest carbon stocks and forest area changes was required. The expectation for the remote sensing technology has risen for monitoring of forest carbon stocks and their changes in developing countries.

While land-cover change can be identified using optical sensors, it has been difficult to identify forest stand structure using conventional remote sensing. LiDAR remote sensing opens a new era for monitoring of three-dimensional forest structure. Particularly, airborne laser scanning has made it possible to reconstruct it widely and spatially. Airborne laser scanning is a remote sensing technique used to acquire three-dimensional data on forest structure, from which a digital surface model (DSM) of the forest canopy and a digital terrain model (DTM) of the ground surface can be obtained. The laser profiler system carried on aircraft on the other hand is a technology that obtains information on canopy height along travelling course. Though increasing information to estimate forest carbon stocks accurately by using airborne LiDAR, various issues such as the cost, the operation, and the regulation remain unsolved to introduce it to monitoring forest carbon stocks in developing countries.

To avoid such issues, use of satellite LiDAR system is another possibility to estimate forest structure. This study aims to investigate characteristics of satellite LiDAR waveform in tropical forests by comparing with canopy structure derived from high resolution satellite data.

2. Method

2.1 Study area

Study area is located in the Tangkulap Forest Reserve, Sabah, Malaysia, which is managed by the Sabah Forestry Department under the Deramakot Forestry District. This forest reserve had been degraded due to intensive and unplanned harvesting in the past. Concern about the current status of the forest reserves has grown considerably in the past decade and therefore, has resulted in some initiatives to reverse the trend and to develop strategies and actions for sustainable forest management (SFM). Within this context, the SFD prepared a 10-year Forest Management Plan (2006-2015) hoping that if properly managed, restored or rehabilitated, the forest reserves have the potential to generate significant environmental and socio-economic benefits (Sabah Forestry Department, 2010).

We established 20 sample plots with 15m radius in the study area and determined the coordinates at the centre of each plot by using a global positioning system (GPS) receiver. DBH and tree height of all standing trees with more than 10 cm DBH were measured. Allometry between tree height and diameter at breast height (DBH) was investigated from field measurement in sample plots (Figure1).

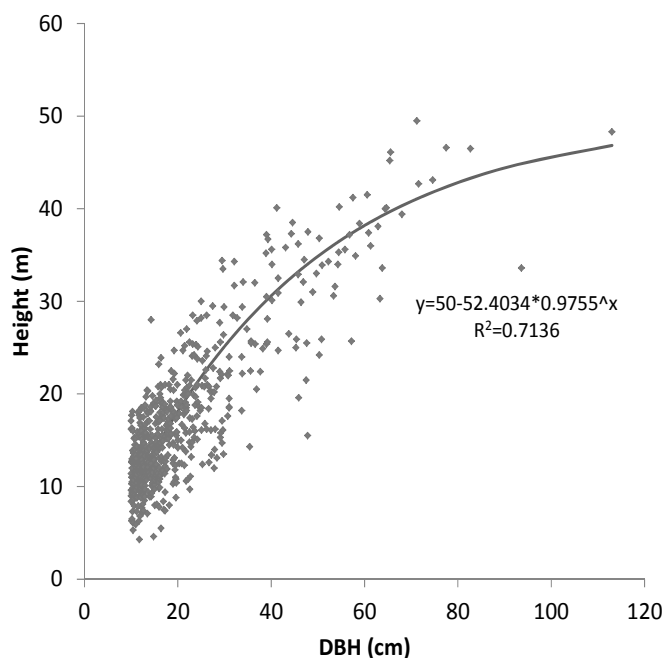


Figure 1. Relationship between DBH and tree height from field data in sample plots.

2.2 Remotely sensed data

2.2.1 Satellite LiDAR data

ICESat GLAS data (release 31), which were acquired on 8 October 2008, were used in this study. ICESat was launched in January 2003 and the mission was terminated in 2009 because of the instrument trouble. ICESat had the orbit of 600km on the ground and it emitted 532 and 1064 nm wavelength laser pulse at 40 Hz simultaneously. The footprint of laser pulse was around 65 m and the interval was about 170 m.

The data for the study area were prepared and provided by the National Snow and Ice Data Center (NSIDC). GLA01data and GLA14 data were used for this study. While GLA14 data file involves the record index, acquisition time latitude, longitude, elevation, range, offsets of signal beginning, signal ending, waveform centroid, and fitted Gaussian peak, original waveform in

each shot of laser pulse was included in the GLA01 data file (Sun, et al., 2008). All waveforms within the study area were plotted to investigate the characteristics for canopy conditions, and they were related to individual footprints 65 m in diameter.

2.2.2 High resolution satellite data

QuickBird panchromatic and multispectral data, which have ground resolutions of 0.61 and 2.44 m, respectively, at nadir were used for this study. The multispectral sensor observes in the following wavelength bands: 450 to 520 nm (blue, band 1), 520 to 600 nm (green, band 2), 630 to 690 nm (red, band 3), and 780 to 900 (near infrared, band 4). The QuickBird data were acquired on 10 October 2007 for the study area.

In the satellite data, treetops showed stronger reflectance than the edges of the crowns. This results from differences in crown geometry and canopy structure. Because the digital number (DN) in the satellite data represents the relative intensity of reflectance, the DN values are larger for the treetops than for the edges of the crowns.

2.3 Data analysis

Reversed QuickBird panchromatic images for each plot by subtracting the DN for a pixel from the maximum DN for the whole image (Wang et al., 2004a) were obtained; then the watershed method (Wang et al., 2004a) to identify individual crowns and extract their areas was used. The watershed method was originally developed to address the influence of terrain on surface water hydrology by modelling the movement of water over a landscape and computing the local flow directions and the gradual accumulation of water moving down slope across the landscape. If we regard the reversed satellite image as analogous to a digital elevation model, the crowns of individual trees resemble depressions in the image. As a result, each crown area can be calculated by using the tools provided by the watershed method. Non-tree areas in the images should be removed during image processing because their presence can lead to overestimation of crown area. To permit this, we determined the threshold between tree crowns and non-tree areas in the satellite data on the basis of the frequency distribution of DN values in mangrove areas and a comparison of the satellite data with data from our field surveys in the study plots. A mask layer for the non-tree areas was then generated by using the threshold value determined by matching the field data to the DN values in the images.

Footprints of laser pulse from ICESat GLAS were identified on the QuickBird image and stand structures in the footprints were estimated from crown information. First, canopy closure within a footprint was calculated using the generated mask. Distribution of individual crown areas within a footprint was investigated for all footprints in the study area. Grade of degradation due to historical selective logging was decided from these two factors, namely, canopy closure and crown size structure. Waveform in each footprint was extracted from ICESat GLAS data and the relationship between stand structure, which was estimated from crown information, and the waveform was investigated. In addition, waveform in oil palm plantation, which was outside of the extent of high resolution satellite data, was also investigated. The condition of the area was identified using Google Earth.

3. Result

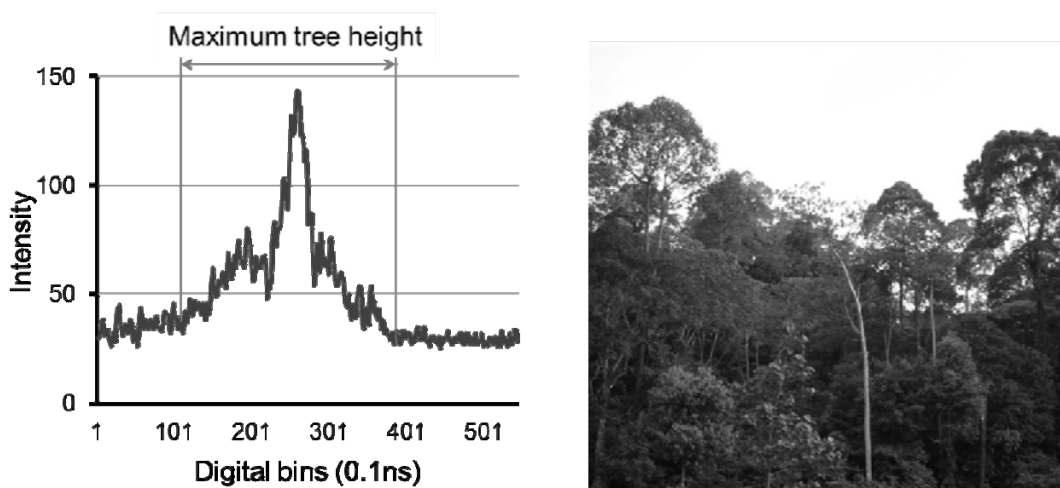
Rates of forest canopy closure within the footprints which were calculated from a mask for distinguishing forest area from non-forest area in high resolution satellite image ranged from 65% to 100 %. It is considered that the difference of rates was caused by the intensity of selective logging. There was difference in distributions of crown areas in the footprints, which were extracted from the panchromatic data. Particularly, regenerated area, where heavy logging

had been conducted in the past, consisted of small crown trees.

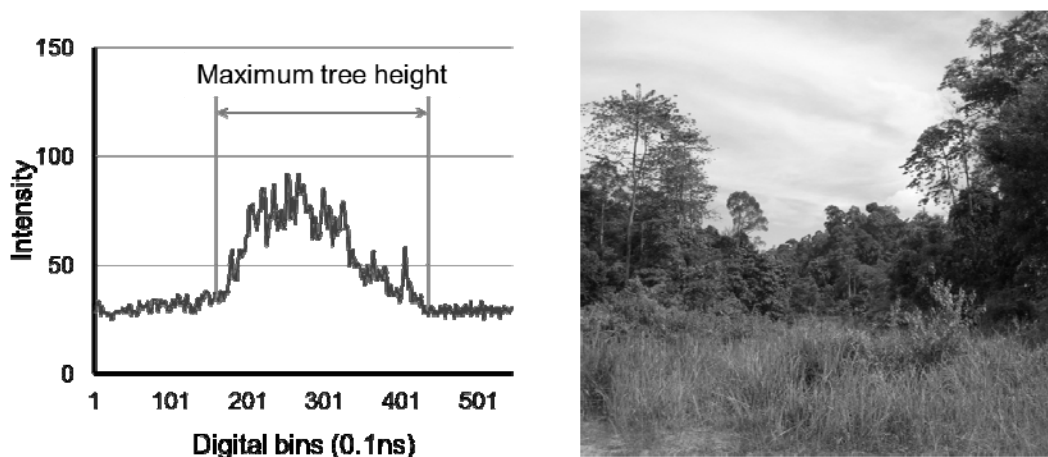
The waveform of a mature forest in the study area, which canopy closure was more than 90% and had relatively large size of crowns, was shown in Figure 2 (a). In a mature forest, maximum tree height derived from LiDAR data was about 43 m. Intensity has a peak around 23m height and it means canopy layer. While there is some difference in height estimation, the waveform described the size composition. In a mature forest, main part of one shot of laser pulse was reflected by crowns of canopy layer and few part of laser pulse reached to lower layers and the ground. In some waveforms for mature forests, it was difficult to recognize the ground level due to weak signal.

The waveform of a degraded forest in the study area, which canopy closure was less than 90% and consisted of various sizes of crowns, was shown in Figure 2 (b). In a degraded forest, Maximum tree height derived from LiDAR data was about 42 m. Intensity was relatively weak through all layers. In a degraded forest due to historical selective logging of valuable trees, some trees which occupied canopy layer have remained. Waveforms in degraded forest were various due to canopy conditions which were recognized from the analysis of high resolution satellite data. It is considered from the result that intensity of a waveform peak of a laser pulse is regarded as an indicator for evaluating grade of forest degradation.

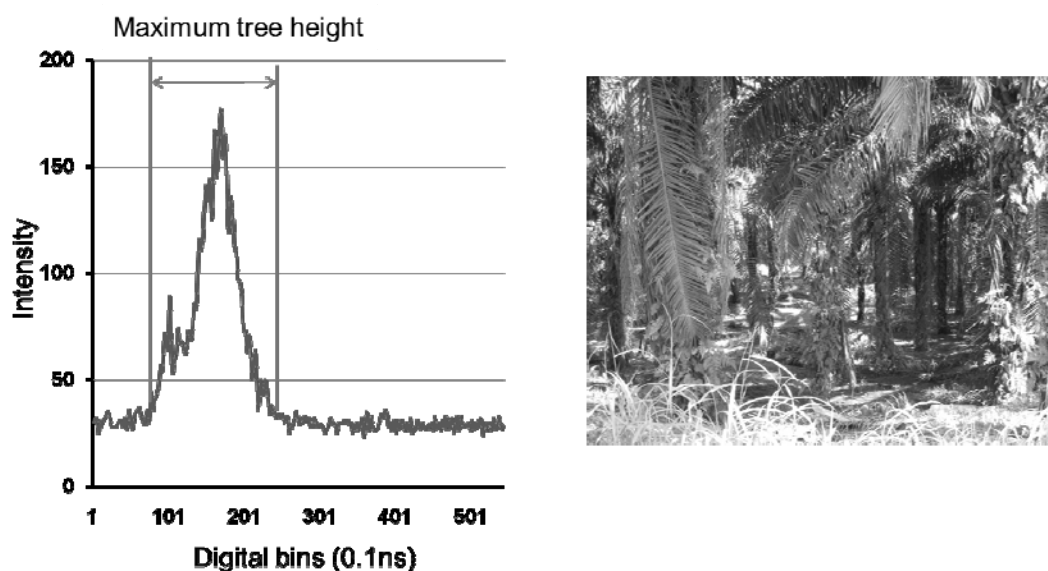
The waveform of a degraded forest in the study area, which canopy closure was less than 90% and consisted of various sizes of crowns, was shown in Figure 2 (c). In oil palm plantation, maximum tree height derived from LiDAR data was about 23 m. Intensity has a peak around 15m height and it means canopy layer. Waveforms in oil plantation were similar and obvious signals from the ground were confirmed in almost of them. In oil palm plantation, trees are planted regularly. There is constant space in a canopy due to this regularity although they have large leaves.



(a) Mature forest



(b) Degraded forest



(c) Oil palm plantation

Figure 2. Waveforms from one shot of laser pulse in different canopy conditions.

4. Discussion

In this study, we investigated characteristics of satellite LiDAR waveform in tropical forests by comparing with canopy structure derived from high resolution satellite data. Stand structure was estimated from waveform of satellite LiDAR data. Length of waveform almost indicated maximum tree height. Peak position of waveform indicated the height of canopy layer. The height and position of peak of waveform indicated the grade of forest degradation. Further studies are required for identify the relationship between waveform of a shot of laser pulse from satellite LiDAR and canopy condition such as canopy closure and distribution of crown area quantitatively.

The application satellite LiDAR data to mapping of ecosystem structure is currently limited by the relatively small fraction of the earth's surface sampled by these sensors (Lefsky et al., 2011). In addition, topographic effect on the data is considerably large in rugged terrain. Some models are required to apply the results concerning forest structure obtained from satellite LiDAR data

to whole forest area spatially. Combination of different types of satellite data might be one possibility to solve this issue.

Acknowledgements

This research was supported by the Environment Research and Technology Development Fund (D-1006) of the Ministry of the Environment, Japan. We thank staffs of Sabah Forestry Department in Deramakot. We thank Dr. Imai for his valuable advice and kind support in the field survey.

References

Lefsky, M.A., Ramond, T., and Weimer, C.S., 2011. Alternate spatial sampling approaches for ecosystem structure inventory using spaceborne lidar. *Remote Sensing of Environment*, 115, 1361-1368.

Sabah Forestry Department, 2010. Tangkulap-Pinangah forest development project. In: Sabah Forestry Department (ed.). *Annual Report 2009*. Sandakan, Sabah Forestry Department: 261-268.

Sun, G., Ranson, K.J., Masek, J., Guo, Z., Pang, Y., Fu, A., and Wang, D., 2008. Estimation of tree height and forest biomass from GLAS data. *Journal of Forest Planning*, 13: 157-164.

Wang, L., Gong, P., and Biging, S., 2004. Individual tree-crown delineation and treetop detection in high-resolution aerial imagery. *Photogrammetric Engineering and Remote Sensing*, 70, 351-357.

Model development for the estimation of aboveground biomass using a lidar-based sample of Canada's boreal forest

Christopher W. Bater^{1,*}, Michael A. Wulder^{2,*}, Nicholas C. Coops^{1,*}, Chris Hopkinson³, Samuel B. Coggins⁴, Eric Arsenault⁵, André Beaudoin⁶, Luc Guindon⁶, R.J. Hall⁵, Philippe Villemaire⁶, & Murray Woods⁷.

¹Integrated Remote Sensing Studio, Department of Forest Resources Management, University of British Columbia.

²Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada.

³Applied Geomatics Research Group, Centre of Geographic Sciences.

⁴Nisga'a Lisims Government.

⁵Northern Forestry Centre, Canadian Forest Service, Natural Resources Canada.

⁶Laurentian Forestry Centre, Canadian Forest Service, Natural Resources Canada.

⁷Southern Science and Information Section, Ontario Ministry of Natural Resources.

*For additional information please contact:

cbater@interchange.ubc.ca; nicholas.coops@ubc.ca; or mwulder@nrca.gc.ca

1. Introduction

The northern forested areas of Canada are largely unmanaged and not subject to inventories with the same level of detail or regularity as southern forested regions. In an effort to augment monitoring and inventory activities, airborne light detection and ranging (lidar) has been employed to obtain plot-level information over a sample of Canada's northern forests. During the summer of 2010, a series of 34 transects were flown over a total length of more than 24,000 km, spanning the width of the Canadian landmass from Nova Scotia to the Yukon, and crossing eight ecozones and 13 UTM zones.

Following data acquisition, a suite of plot-level lidar vegetation metrics were calculated. To develop estimates of forest attributes such as biomass, however, field data were required from the range of conditions found across the region. Co-located field plots and lidar data are required for algorithm development. To that end, datasets were acquired from Quebec, Ontario and the Northwest Territories. The Quebec field plots were coincident with the lidar transects, existing lidar and field data were used for Ontario and Northwest Territories, with careful analysis required due to different lidar acquisition parameters. In this paper we describe the development of regression models for large area estimates of various tree aboveground biomass components using field and lidar datasets of uncommon provenance, with significant differences both in terms of the environments in which they were collected, and the characteristics of the field and lidar surveys. The equations developed are deemed suitable for application and extrapolation across the national series of lidar transects.

2. Methods

2.1 Field data

Field data were obtained from the Northwest Territories, Ontario and Quebec (Figure 1). The datasets are described in detail below. The equations are intended to be relevant over the boreal area of Canada's forest, not the entire forest area depicted.

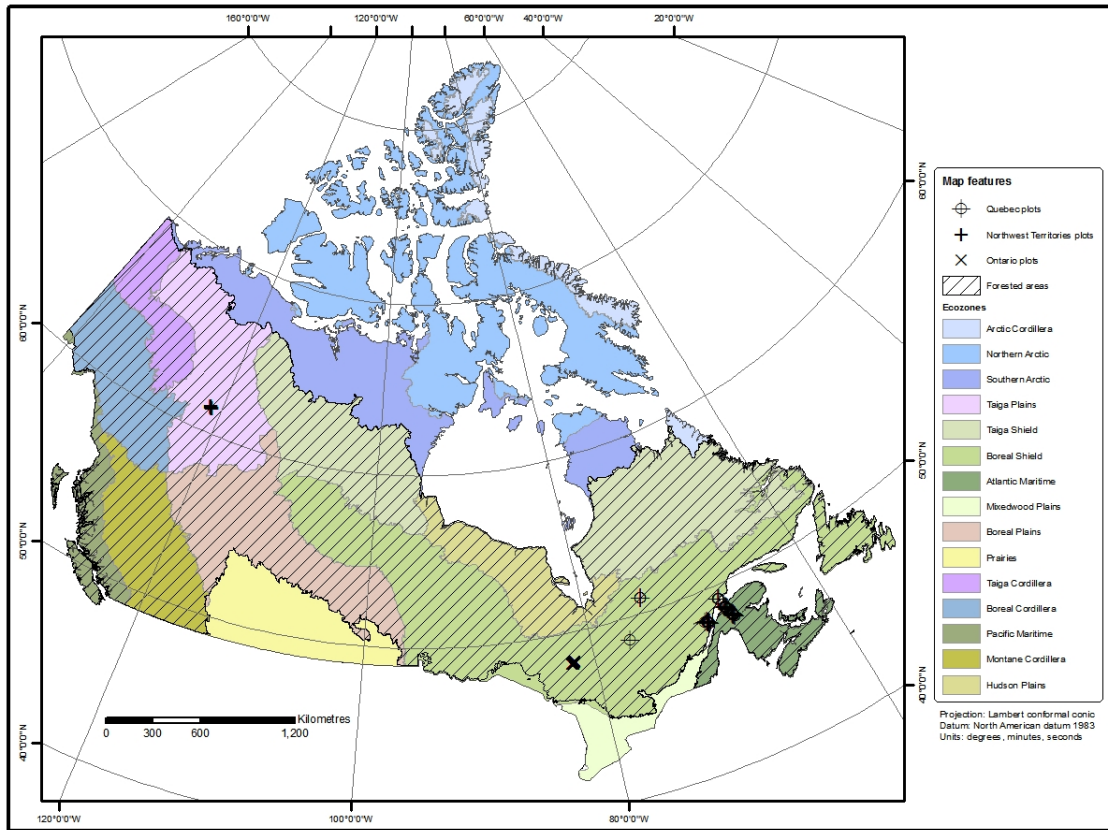


Figure 1. Canadian eozones and locations of field plots by province.

2.1.1 Northwest Territories

Field data were provided by the Northern Forestry Centre, Natural Resources Canada, and consisted of individual tree measurements, including species, height, and diameter at breast height (DBH). Data were collected from square 20 x 20 m (400m²) plots.

2.1.2 Ontario

Field data were provided by the Ontario Ministry of Natural Resources and collaborators. As with those from the Northwest Territories, the data consisted of individual tree measurements, including species, height, and DBH collected from 11.28 m radius (400 m²) circular plots located in the Romeo Malette Forest near the city of Timmins. Plots can be coarsely stratified into mixedwood, intolerant hardwoods, Jack pine, and black spruce.

2.1.3 Quebec

Field data were provided by: Direction des inventaires forestiers Forêt Québec, Ministère des Ressources Naturelles et de la Faune. Tree lists were provided for each plot by species and stems per hectare grouped in 2 cm DBH classes. Tree heights were measured for four to six representative trees per plot. Tree numbers for each DBH class were then calculated based on a 400 m² plot, and hardwood/softwood-specific equations were developed to estimate height for all stems in the plots based on DBH.

2.1.4 Field-based biomass estimates

Biomass components were calculated using national DBH- and height-based “all species” equations originally provided by Lambert et al. (2005) and subsequently updated by Ung et al. (2008). The equations were developed using archival data collected through the Energy for the

Forest Research (ENFOR) program, and provide estimates for wood, bark, stem, branches, foliage, crown, and total biomass. National “all species” equations were selected because no spatially explicit species information is readily available for the majority of the lidar transects. The DBH- and height-based equations take the following form:

$$\begin{aligned}
 y_{wood} &= \beta_{wood1} D^{\beta_{wood2}} H^{\beta_{wood3}} + e_{wood} \\
 y_{bark} &= \beta_{bark1} D^{\beta_{bark2}} H^{\beta_{bark3}} + e_{bark} \\
 y_{stem} &= \hat{y}_{wood} + \hat{y}_{bark} + e_{stem} \\
 y_{foliage} &= \beta_{foliage1} D^{\beta_{foliage2}} H^{\beta_{foliage3}} + e_{foliage} \\
 y_{branches} &= \beta_{branches1} D^{\beta_{branches2}} H^{\beta_{branches3}} + e_{branches} \\
 y_{crown} &= \hat{y}_{foliage} + \hat{y}_{branches} + e_{crown} \\
 y_{total} &= \hat{y}_{wood} + \hat{y}_{bark} + \hat{y}_{foliage} + \hat{y}_{branches} + e_{total}
 \end{aligned}$$

Where y_i is the dry biomass component i of a living tree (kg); i is wood, bark, stem, foliage, branches, crown, or total; \hat{y}_i is the prediction of y_i ; D is the DBH (cm); H is tree height (m); β_{jk} are model parameters with coefficient estimates b_{jk} ; j is wood, bark, foliage or branches; $k = 1, 2$ or 3 ; and e_i are error terms. Stem, crown, and total biomass are then obtained by summing their respective components (Ung et al., 2008). Model parameter estimates are provided in Table 1; the field-based biomass estimates are summarized in Table 2.

Table 1. Model parameter estimates for all-species biomass equations published by Ung et al. (2008)

Parameter	Estimate	Standard error
β_{wood1}	0.0283	0.0004
β_{wood2}	1.8298	0.0075
β_{wood3}	0.9546	0.0101
β_{bark1}	0.012	0.0003
β_{bark2}	1.6378	0.017
β_{bark3}	0.7746	0.0233
$\beta_{branches1}$	0.0338	0.0008
$\beta_{branches2}$	2.6624	0.0182
$\beta_{branches3}$	-0.5743	0.0233
$\beta_{foliage1}$	0.1699	0.0036
$\beta_{foliage2}$	2.3289	0.0184
$\beta_{foliage3}$	-1.1316	0.0235

Table 2. Summary of biomass components estimated using field data and national “all species” equations provided by Lambert et al. (2005) and Ung et al. (2008).

Province	Wood biomass (kg)			
	Mean	Min	Max	Std.Dev.
Northwest Territories	4,667	1,125	10,665	2,218
Quebec	2,646	24	7,312	1,650
Ontario	4,364	495	10,499	1,942
All Groups	4,073	24	10,665	2,070
Province	Bark biomass (kg)			
	Mean	Min	Max	Std.Dev.
Northwest Territories	643	184	1,378	269
Quebec	377	5	1,066	223
Ontario	621	85	1,298	245
All Groups	576	5	1,378	264
Province	Stem biomass (kg)			
	Mean	Min	Max	Std.Dev.
Northwest Territories	5,310	1,308	12,043	2,486
Quebec	3,023	28	8,378	1,870
Ontario	4,985	580	11,797	2,185
All Groups	4,649	28	12,043	2,333
Province	Branch biomass (kg)			
	Mean	Min	Max	Std.Dev.
Northwest Territories	840	250	1,977	363
Quebec	612	9	1,698	371
Ontario	864	96	1,864	344
All Groups	807	9	1,977	366
Province	Foliage biomass (kg)			
	Mean	Min	Max	Std.Dev.
Northwest Territories	188	81	352	59
Quebec	149	5	377	80
Ontario	232	35	664	92
All Groups	206	5	664	90
Province	Crown biomass (kg)			
	Mean	Min	Max	Std.Dev.
Northwest Territories	1,028	331	2,329	418
Quebec	761	13	1,938	445
Ontario	1,095	132	2,260	423
All Groups	1,014	13	2,329	444
Province	Total tree biomass (kg)			
	Mean	Min	Max	Std.Dev.
Northwest Territories	6,338	1,640	14,373	2,887
Quebec	3,784	42	10,261	2,312
Ontario	6,080	712	14,057	2,557
All Groups	5,663	42	14,373	2,739

2.2 Lidar data

2.2.1 Northwest Territories

Lidar data were provided by the Northern Forestry Centre, Natural Resources Canada. Data were acquired by the Applied Geomatics Research Group (AGRG) in August of 2007 using an Optech ALTM 3100. During the survey, the power supply for the laser diode was in a state of rapid but steady degradation, and thus the sampling efficiency of this site was well below optimal. This directly impacted the probabilities associated with ground and canopy returns (due to reduced pulse power) and data density was up to five times below what it should have been. Returns were classed as first, intermediate, last, or single. Because of the power supply issues, many points fell into the single return category and very few were classified into ground and non-ground returns. The AGRG undertook additional processing to create a ground return point dataset that was in turn used to create a 1 m raster of ground elevation. The lidar "all-returns" point data was then overlaid onto this grid and the ground elevations subtracted to generate a canopy point data set.

2.2.2 Ontario

Lidar data were provided by the Ontario Ministry of Natural Resources and collaborators. Data were in standard las format and classified as ground and non-ground returns.

2.2.3 Quebec

Lidar data were extracted from this project's dataset where the transects and Quebec field plots intersected. The desired survey specifications included a flying height of 1,200 magl, a 70 kHz pulse repetition frequency, scan angles of +/-15°, and a nominal pulse density of ~ 2.8 returns/m², with the understanding that flying conditions might necessitate lower or higher flying heights. Scan angle was generally kept fixed at 15°.

2.2.4 Lidar data processing

Lidar data were processed using FUSION software (McGaughey, 2010). A suite of standard plot-level lidar metrics were calculated, including mean first return height; standard deviation, coefficient of variation, and the 95th percentile of first return heights; percentage of first returns above 2m; and percentage of first returns above the first return mean height . Plot sizes were 400 m² and either circular or square depending on their province of origin. Table 3 provides a plot-level summary of return numbers by province.

Table 3. Summary of lidar return numbers within field plots.

Province	Number of plots	First returns above 2m		All returns	
		Mean	Std.Dev	Mean	Std.Dev.
Northwest Territories	40	88	55	189	97
Quebec	41	391	145	745	113
Ontario	120	1086	255	1881	405
All groups	201	746	474	1315	785

2.2 Statistical analysis

Given the large number of possible predictor variables (lidar metrics) which were often strongly intercorrelated, a subset of candidate predictors were selected based on their relatively low intercorrelations and their biological relevance (e.g. Li et al., 2008). Prior to analysis, both predictor variables and forest attributes were transformed to their natural logarithms. Best

subsets multiple linear regression analyses were then performed for each forest attribute. Akaike's Information Criterion (Akaike, 1973) was employed to select the most parsimonious models (Posada and Buckley, 2004). Following the development of regression models, outliers were assessed based on their standardized residuals and leverage values. If a large gap existed between the majority of leverage values and one or a few with very high values, the outliers were removed from the final regression model.

To assess the quality of the final models, the lidar-derived biomass estimates were back-transformed from their natural logarithms to arithmetic units using a bias correction factor (Sprugel, 1983). Bias, mean absolute error (MAE) and root-mean-square error (RMSE) were then calculated from the residuals between predicted and observed values.

3. Results

Results of the multiple linear regression analyses are shown in Tables 4 and 5, and Figure 2. The final regression models explained between 36% and 78% of the variance in the various biomass components. Of the biomass components, foliage biomass was the most poorly predicted, with an adjusted R^2 of 0.36. However, 78% and 76% of the variance in wood and total aboveground biomass were explained, respectively (Table 4).

Table 4. Summary statistics for the multiple linear regression models for the field-based biomass components (dependent variables) and lidar canopy height and cover metrics (predictors).

Dependent variable	n plots	Multiple R	Multiple R^2	Adjusted R^2	df model	df residual	F	p
Foliage biomass	197	0.61	0.37	0.36	2	194	56.08	0.00
Branch biomass	198	0.81	0.65	0.65	2	195	182.28	0.00
Crown biomass	198	0.77	0.60	0.59	2	195	145.63	0.00
Bark biomass	198	0.86	0.73	0.73	2	195	267.69	0.00
Wood biomass	198	0.88	0.78	0.78	3	194	231.02	0.00
Stem biomass	198	0.88	0.78	0.77	3	194	224.43	0.00
Total tree biomass	198	0.87	0.76	0.76	3	194	206.79	0.00

Bias was negligible for all models (Table 5). Total aboveground biomass had a bias of 56.8 kg, or less than 1% of the mean. Foliage biomass, the smallest biomass component, but also the most poorly predicted in terms of its precision, had a MAE of 52 kg, or 25% of the mean value observed in the plots. Conversely, the largest component, wood biomass, had an MAE of 749 kg, or 18% of the mean value observed in the plots, while total tree biomass had an MAE of 1029 kg, or 18% of the mean observed value (Table 5). Foliage biomass had an RMSE of 67 kg or 33%. Wood biomass had an RMSE of 990 kg, or 24% of the mean value observed in the plots, while total aboveground biomass had an RMSE of 1353 kg or 24% (Table 5).

Of the candidate lidar-derived predictor variables, the 95th percentile of first return heights and the percentage of first returns above the first return mean height were employed to estimate the crown biomass components, while mean first return height, the coefficient of variation of the

first return heights, and the percentage of first returns above 2 m were employed to estimate the stem components and total tree biomass.

Table 5. Fit statistics for the multiple linear regression models for the field-based biomass components (dependent variables) and lidar canopy height and cover metrics (predictors). Field plots sizes were 200 m².

Dependent variable	Units	BCF ¹	Bias	MAE ²	RMSE ³
Foliage biomass	kg	1.059	0.87	51.96	67.22
Branch biomass	kg	1.043	4.83	163.38	218.60
Crown biomass	kg	1.045	5.61	212.79	282.12
Bark biomass	kg	1.04	7.00	105.84	136.98
Wood biomass	kg	1.039	51.82	749.33	990.40
Stem biomass	kg	1.039	58.56	854.28	1125.35
Total tree biomass	kg	1.038	56.80	1029.34	1353.21

¹Bias correction factor (Sprugel, 1983)

²Mean absolute error

³Root mean square error

4. Discussion

The objective of this paper was to describe the development of large area estimates of tree biomass components by combining three different field and lidar datasets. The results indicate that, using biologically relevant lidar-derived predictor variables based on height, vertical structural complexity, and cover, over 70% of the variance in stem and total tree biomass can be explained; this despite significant differences in lidar acquisitions, particularly the dissimilarities in return densities (Table 3). In comparison, biomass estimates employing purposefully collected lidar and field data over comparably small areas typically explain between 70% and 90% of the variance in tree biomass (e.g. Popescu, 2007; Li et al., 2008; Næsset and Gobakken, 2008).

It should be noted that the usage of “all-species” national allometric equations to estimate biomass from field-measured height and DBH is not optimal, as prediction errors grow with increased model generalization (Hall and Case, 2008). Nonetheless, without additional information on species within the lidar transects, the equations do provide a means to estimate forest attributes of critical importance.

Acknowledgements

The authors thank D. Pitt (Canadian Wood Fibre Centre), P. Treitz (Queen’s University), and personnel of the Northern Forestry Centre of Natural Resources Canada, the Ontario Ministry of Natural Resources, and the Direction des inventaires forestiers Forêt Québec, Ministère des Ressources Naturelles et de la Faune for providing insights as well as lidar and field data to support this initiative.

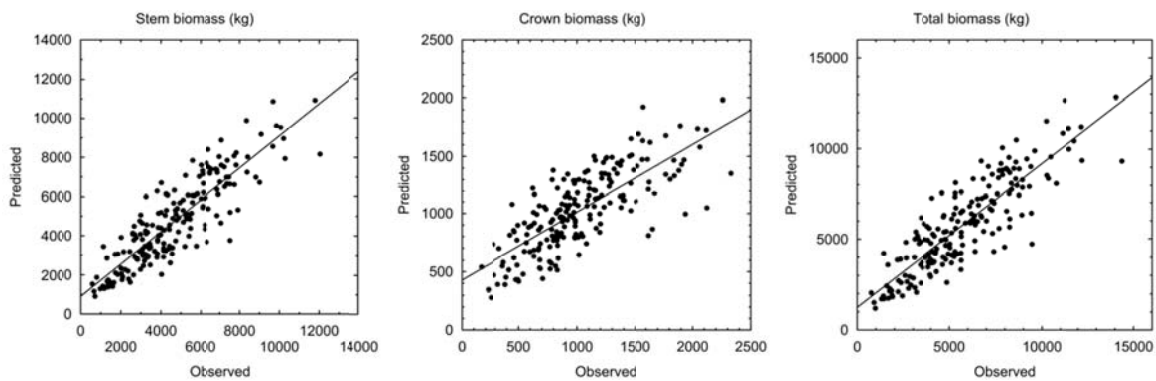


Figure 1. Observed and predicted values of stem, crown and total biomass components estimated using plot-level lidar metrics. Ground plots were 400 m².

References

- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle, in: Second International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary, pp. 267-281.
- Hall, R.J., Case, B.S., 2008. Assessing prediction errors of generalized tree biomass and volume equations for the boreal forest region of west-central Canada. *Canadian Journal of Forest Research* 38, 878–889.
- Lambert, M.C., Ung, C.H., Raulier, F., 2005. Canadian national tree aboveground biomass equations. *Canadian Journal of Forest Research* 35, 1996-2018.
- Li, Y., Andersen, H.E., McGaughey, R., 2008. A comparison of statistical methods for estimating forest biomass from light detection and ranging data. *Western Journal of Applied Forestry* 23, 223-231.
- McGaughey, R.J., 2010. FUSION/LDV: Software for LIDAR data analysis and Visualization (No. Version 2.90). Pacific Northwest Research Station, Forest Service, United States Department of Agriculture.
- Næsset, E., Gobakken, T., 2008. Estimation of above-and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment* 112, 3079–3090.
- Popescu, S., 2007. Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy* 31, 646–655.
- Posada, D., Buckley, T.R., 2004. Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology* 53, 793 -808.
- Sprugel, D.G., 1983. Correcting for bias in log-transformed allometric equations. *Ecology* 64, 209-210.
- Ung, C.H., Bernier, P., Guo, X.J., 2008. Canadian national biomass equations: new parameter estimates that include British Columbia data. *Canadian Journal of Forest Research* 38, 1123-1132.

Biases in an Airborne Profiling Survey of Hedmark County, Norway

Silvilaser 2011, Oct. 16-20, 2011, University of Tasmania, Hobart, Australia

Ross Nelson¹, Erik Næsset², Terje Gobakken², Liviu Ene², Göran Ståhl³, and Timothy Gregoire⁴

1. Code 614.4/NASA-Goddard Space Flight Center, Greenbelt, Maryland, USA, ross.f.nelson@nasa.gov
2. Dept. of Ecology and Natural Resource Management, Norwegian University of the Life Sciences, Ås, N-1432, Norway, erik.naesset@umb.no, terje.gobakken@umb.no, liviu.ene@umb.no
3. Dept. of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå SE-901 83 Sweden, goran.stahl@slu.se
4. School of Forestry and Environmental Studies, 360 Prospect Street, Yale University, New Haven, Connecticut, USA, timothy.gregoire@yale.edu

Abstract

Two biases that affect the accuracy and precision of biomass estimates are explored using a systematic sample of 105 profiling lidar flight lines acquired over Hedmark County, Norway. The first bias deals with inaccuracy that may accrue in mean estimates of biomass for the entire County when the predictive models are generated using a spatially constrained set of ground plots, simulating, for instance, limited access to ground observations. In Hedmark, mean cover class estimates of aboveground dry biomass differed by as much as -12% to +13% when models were developed using coincident ground-lidar observations limited to the north or south ends of the County, respectively. The second type of bias deals with variance inflation that may occur when a systematic lidar sample is treated as a random sample. The model-dependent variance formula used in this study assumes that the lidar flight lines are scattered across the landscape randomly when, in fact, they were flown systematically. Approximate systematic variances are compared with the model-dependent variances. Results at the County level indicate that the model-dependent variance estimates were up to 3.3 times larger than the comparable systematic variance estimates; the amount of inflation varies by cover class. The results indicate that analysts must exercise caution (1) in situations where spatially constrained models are used to estimate biomass across much larger areas of interest, and (2) if an assumption of randomness must be assigned to a systematic collection of lidar flight lines in order to calculate variance estimates.

1. Introduction

Airborne and/or space lasers have been used to sample extensive areas of interest (AOI) to estimate total aboveground dry forest biomass. Most of the studies to date employ model-dependent (i.e., model-based) approaches where biomass equations are developed using spatially coincident field observations and airborne or space lidar measurements. The lidar is then used as a sampling tool to estimate forest biomass across extensive areas (Nelson et al. 2004; Boudreau et al. 2008; Andersen and Winterberger 2009; Helmer et al. 2009; Nelson et al. 2009a,b; Asner et al. 2010; Ståhl et al. 2011; Andersen et al. 2011). The model-dependent approach is attractive in that the analyst can select the location and attributes (e.g., high biomass, low biomass, cover class) of ground plots used to populate his/her predictive biomass models or, if plots already exist, to overfly the existing ground plots. Using a model-dependent approach,

the number of ground plots that need to be collected is driven only by the need to develop robust predictive models, models that encompass (hopefully) the full range of conditions that one expects to find across the entire area of interest.

And therein lies the danger of employing a model-dependent (MD) approach. The accuracy and precision of the MD biomass estimates are based on the assumption that the predictive model is correct, i.e., that the model quantitatively characterizes conditions across the entire study area. Acquiring a statistically rigorous, representative ground sample across an extensive area in order to develop robust models may be economically and/or logistically untenable if significant portions of the AOI are inaccessible, due perhaps to a lack of roads, civil unrest, unexploded ordinance, or topography. But if models are developed using a spatially limited ground sample that does not characterize the entire AOI, then one runs a very real risk of calculating biased estimates of mean biomass per hectare or total biomass.

A second type of bias deals with the effect of treating a systematic sample as a random sample. No unbiased variance estimator for a single systematic sample exists. One solution, commonly employed, is to treat a systematic sample as if it were acquired randomly. Such a solution, however, may lead to inflated variance estimates (Gregoire and Valentine 2008), especially if the AOI exhibits any sort of linear trend in the parameter of interest, in our case, biomass. Gobakken et al. (2011) and Nelson et al. (2011) employed a MD sampling framework developed by Ståhl et al. (2011) to estimate biomass in Hedmark County, Norway. The MD sampling approach integrates airborne laser measurements with ground plots. Their mean and variance estimators assume that (1) the first phase (lidar) sample is randomly collected and (2) the second phase (ground sample) is representative of the entire AOI. Results from the Gobakken and Nelson studies were surprising in that the addition of airborne-laser-based biomass estimates acquired along thousands of kilometers of regularly spaced flight transects did not consistently improve the precision of County estimates. The counterintuitive results suggest that the MD approach may overestimate variances if assumptions concerning randomness are not met. A question then arises. With respect to Hedmark, how large is this inflationary tendency?

The objective of this study is to look at these two types of biases empirically, i.e., (1) bias with respect to mean estimates of biomass when non-representative ground samples are used to formulate the predictive regressions, and (2) bias in variance estimates due to our treatment of a systematic sample of lidar flight lines as a random sample. Preliminary work by L. Ene with a Monte Carlo simulator has shown that the Ståhl et al. (2011) estimators are unbiased if these assumptions are met. It is the objective of this paper to quantitatively describe, for one study area, the magnitude of these biases when the assumptions are not met.

2. Methods

One-hundred-five airborne profiling flight lines were systematically collected over Hedmark County, Norway during the summer of 2006. The east-west flight lines were evenly spaced every three kilometers so that the profiling aircraft would traverse 1483 Norwegian National Forest Inventory (NFI) ground plots measured or due to be measured in 2005, 2006, and 2007. The NFI plots are systematically arrayed across the 27380 km² County on a 3km x 3km N-S, E-W grid. Of the 1483 plots available, 916 were measured by the profiler. In addition, 79 off-grid ground plots were measured by the profiler in agricultural, residential, and urban areas in Hedmark so that a predictive

equation could be formulated for developed areas. These 995 ground plots (= 916+79) measured by the profiler are subsequently referred to as laser-ground plots.

A model-dependent sampling framework (Ståhl et al. 2011) is used to calculate biomass means and standard errors of estimate (SE) for eight cover classes that tessellate the County. These eight classes include four productive forest classes (high, medium, and low productivity forest, and young forest), and four nonproductive forest/nonforest classes (nonproductive forest, land>850m, developed areas, and water). Seven independent equations, one for each cover class except water, are developed. The equations predict ground-measured biomass as a linear function of profiling lidar height and crown density metrics. The nonproductive forest equation is used to predict biomass on any laser segment identified as water and having significant height. These equations, developed County-wide, are also used to predict forest biomass in smaller political units within the County in order to assess the accuracy and precision of biomass estimates as the size of the domain decreased.

2.1. Bias in Mean Estimates of Biomass

Hedmark displays an obvious N-S biomass trend. The southern portion of the County tends to support dense, more highly productive forest; the northern portion is more mountainous and, with increasing elevations above MSL, is less productive. In this portion of the study, we quantify the degree to which biomass estimates change if, for instance, ground plots could only be measured in the southern half of the County but are then used to predict biomass across the entire County. Our reference or “ground truth” data set consists of those cover class estimates generated using all laser-ground plots available across the entire County. Predictive biomass equations are developed for 4 land cover classes – nonproductive forest, low productivity forest, medium productivity forest, young forest.

In order to estimate the size of the bias due to the use of non-representative predictive models, the County is divided into four roughly equal-area sectors, a southern versus northern half and an eastern versus western half. Biomass models were also developed for the four cover classes in each of the four sectors - N, S, E, W - using only those laser-ground plots located within a given sector. These 5 sets of equations, i.e., the reference equations and the 4 sets of sector equations, were applied to all 105 flight lines across the entire County. Only four of the seven cover classes are considered in this portion of the analysis because the other three classes - highly productive forest, land>850m, and developed areas - had less than the 30 laser-ground plots deemed necessary to calculate the biomass model in one or more sectors. These same five sets of equation were applied to Administrative District 1 (AD01) at the south end of the County (5285 km²), Administrative District 4 at the north end of the County (9437 km²), Municipality 1 within AD01 (1036 km²), and Municipality 18 within AD04 (1122 km²) in order to see if bias trends at the County level were maintained when smaller domains were considered.

2.2. Bias in Estimates of Standard Error of Biomass

The bias of the variance estimator was empirically assessed by dividing the 105 systematically-acquired flight lines (fl) into samples so that an empirical estimate of systematic sampling variability could be calculated (Gregoire and Valentine 2008, pg. 56). The 105 fls 3 km apart were divided into 2 sets of 52 fls 6 km apart, 3 sets of 35

fls 9 km apart, etc...., up to 15 sets of 7 fls 45 km apart. The systematic samples considered are reported in Table 1.

Table 1. The 105 systematically-acquired Hedmark flight lines spaced 3 km apart were subdivided into the following systematic samples.

number of systematic samples (n_{sys})	number of fls per sample	dist. between adjacent fls (km)
2	52	6
3	35	9
4	26	12
5	21	15
6	17	18
7	15	21
8	13	24
10	10	30
15	7	45

Empirical, systematic standard errors can be approximated by treating each subset of the 105 flight lines listed in Table 1 as one of n_{sys} mean values instead of treating each subset as one of n_{sys} observations. For instance, given one systematic sample of 105 flight lines, a conservative estimate of the actual systematic variance can be approximated by dividing the existing 105 flight lines into two groups of 52 flight lines. This conservative variance estimate can be substituted for what might be expected if two systematic samples of 105 flight lines had been flown instead of one. This is certainly a rather coarse approximation, but experience has suggested that this estimate, calculated using eqn. 1, is more stable, and smaller, than the inflated MD variance.

$$\hat{var}_{sys}(\hat{b}_j) = \frac{\sum_{r=1}^{n_{sys}} \left(\hat{b}_{jr} - \hat{\bar{b}}_j \right)^2}{n_{sys} - 1} \quad (1) \quad , \quad \text{where} \quad \hat{\bar{b}}_j = \frac{\sum_{r=1}^{n_{sys}} \hat{b}_{jr}}{n_{sys}} \quad (2)$$

where

\hat{b}_{jr} = an estimate of biomass per hectare for the r th systematic sample in the j th cover class, and
 n_{sys} = the number of systematic samples.

The Ståhl et al. (2011) variance estimator (their eqn. 14) has two components, a first term characterizing sampling variability and a second term that incorporates the variation associated with the model coefficients. The systematic variance formula reported directly above (eqn. 1) accounts only for systematic sampling variability; no model error is included. In order to more fairly compare the empirical systematic SEs with MD SEs, the model error term is subtracted from the MD variance estimates prior to comparison. The empirical systematic SEs are compared to the Ståhl et al. SEs with model error removed in order to assess the inflationary effects associated with treating a systematic lidar sample as if it had been acquired randomly.

3. Results

3.1. Bias in Mean Estimates of Biomass

Figure 1 reports mean biomass values estimated using the reference equations (“entire County”) versus four sector equations, i.e., northern half of the County (“north end”), south end, and the east and west sides, for four land cover classes.

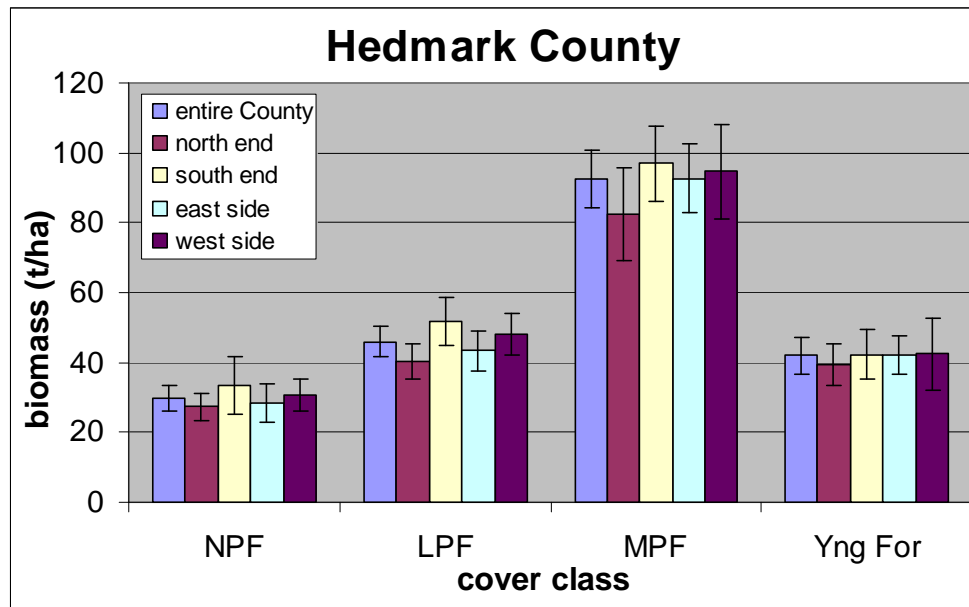


Figure 1. Biomass estimates for four cover classes in Hedmark County – nonproductive forest (NPF), low productivity forest (LPF), medium productivity forest (MPF), and young forest (Yng For). The error bars depict 2 standard errors, i.e., 95% confidence limits on the mean estimates.

Figures similar to Figure 1 were generated for northern and southern administrative districts and for two municipalities located within the administrative districts. The figures (not presented) look very similar to Figure 1, mean differences and cover class trends are consistent as one considers smaller and smaller domains. Considering Hedmark County, estimates based on equations formulated using only north end laser-ground plots underestimated biomass by 5.9% (Yng For) to 12.2% (MPF). South end equations overestimated County means by up to 12.6% (NPF). East-West differences ranged from -5.9% (East equations, LPF) to +4.6% (West equations, LPF). Evidently the most productive Hedmark forests are in the southwest quadrant of the County. Biomass estimates in young forest are essentially unaffected by the provenance of the biomass equations, regardless of the size of the domain considered.

3.2. Bias in Estimates of Standard Error of Biomass

Figure 2 illustrates, for Hedmark County, the relationship between empirical, systematic standard errors (SEs), i.e., the solid lines, and model-dependent estimates of standard error, i.e., the dashed lines, for two cover class groups and one cover class – (1) all 8 cover classes, (2) the four productive forest classes, and (3) low productivity forest. If the model-dependent (MD) estimator (Ståhl et al. 2011, eqn. 14) is unbiased and if there are no biases introduced by ignoring assumptions, then the empirical SEs should, on average, approximately equal the MD estimates of SE across the entire range of sampling intensities considered. Obviously, as Figure 2 depicts, they do not. In Hedmark, considering all cover classes (the blue lines), the MD-SE is, on average,

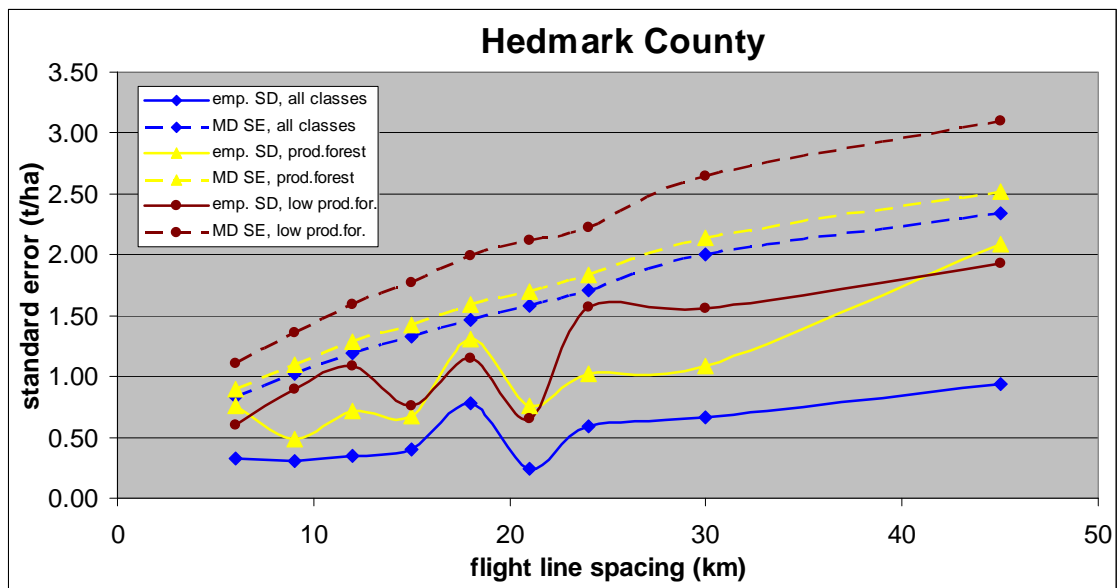


Figure 2. Empirical SDs - solid lines - compared with MD estimates of SE - dashed lines - for three cover classes in Hedmark County: (1) all 8 cover classes - blue, (2) the four productive forest classes - yellow, and (3) low productivity forest - brown.

3.3±1.3 (1 σ) times larger than the empirical SE across all sampling intensities considered. For the productive forest class (yellow), that number is 1.8±0.5 times larger, and for low productivity forest (brown), the multiplier is 1.9±0.6 times larger. This inflationary tendency was also found, though somewhat less pronounced, in AD04 (Figure 3) where multipliers ranged from 1.7±0.7 (productive forest) to 2.3±1.4 (all cover classes). The degree to which the MD variance estimates are inflated is dependent not only on the model used to predict biomass for a given cover class but

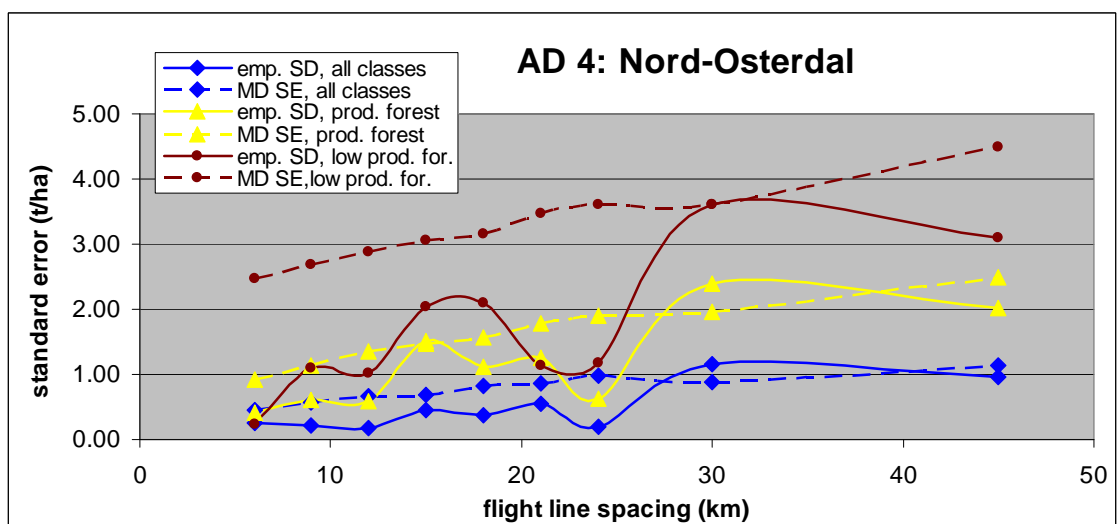


Figure 3. Empirical SDs - solid lines - compared with MD estimates of SE - dashed lines - for three cover classes in Administrative District 4: (1) all 8 cover classes - blue, (2) the four productive forest classes - yellow, and (3) low productivity forest - brown.

also on the domain being considered. Regardless, in this Hedmark study, the tendency toward overestimation of the variance estimate relative to empirical systematic error is unmistakable. This conservative tendency may be due to an inappropriately conceived estimator, i.e., the Ståhl et al. (2011) estimator may inherently overestimate variance, or it may be due to a misapplication of estimator. Which is it?

Monte Carlo simulations runs by L. Ene on a simulated Hedmark forest population have shown that the MD estimator is unbiased if the first phase (lidar) sample is randomly allocated. The inflationary effects noted in Figures 2 and 3 are due to a violation of this assumption, one clearly stated by Ståhl et al. (2011) on page 99. The inflation is most likely exacerbated by the presence of a marked N-S trend in biomass across the County and by the fact that we flew our flight lines E-W. We suspect, though cannot prove, that the MD variance estimates would also have been inflated, though to a lesser degree, had we flown systematic lines N-S since there seems to exist a less remarkable biomass trend E-W (see Figure 1).

4. Discussion

Extrapolation biases can be appreciable and in this study, on an AOI with an obvious biomass trend N-S, differences between means were occasionally significantly different. Situations requiring extrapolation are commonly found in the scientific literature because, basically, one of the strengths of aircraft and satellites is that they can go where field crews cannot. For instance, Nelson et al. (2009a), in Quebec, developed biomass equations using ground plots located only in the southern half of the Province but applied them both north and south. Asner et al. (2010, supporting information, Figure S1(d)) used plots located in the eastern and southern portions of their Peruvian study area and applied them in the central and NW quadrants. Any extensive study of the northern boreal forests or primary tropical forests will face access problems, necessitating extrapolation of optical-, radar-, and/or lidar-ground relationships to the entire AOI. In such situations, analysts, and those who use their results, face a question. Is it better to have potentially biased estimates or no estimates at all? This study is just one example of the magnitude of these extrapolation biases. We do not suggest that biased estimates are without merit but we do recommend (1) that scientists who generate such estimates should clearly state that their numbers may be biased and that the magnitude of the bias is unknown, and (2) that managers or policy makers who make use of these estimates recognize the potential for error.

A comparison of empirical, systematic SEs and MD-SEs quantify the penalty paid for treating a systematic sample of flight lines as a random sample. MD estimator SEs are inflated ~1.5 - 3.5x, depending on the cover class and domain. It is now apparent, based on the results of Ene's Hedmark simulations and based on the empirical results reported in this manuscript, that a significant variance penalty may be inflicted if a systematic sample is treated as a random sample, and this penalty is expected to increase in the presence of a linear trend.

Five years ago, the Hedmark study was undertaken to determine if an airborne laser profiler could be used as a sampling tool to augment a ground-based national forest inventory. We expected that the additional spatial coverage afforded by the laser would increase the precision of the ground sample. The results of that study were surprising in that the SEs of the laser-augmented estimates of biomass were not consistently smaller than the ground-based SEs. It is now apparent that the MD variances were overestimated and that the lidar-based approach was more competitive

than first realized. Given the constraint that, in Hedmark, our flight lines had to traverse ground plots laid out along lines spaced 3 km apart, perhaps the easiest way to approximate systematic sampling error in the context of a flight-line-rich environment is to calculate an empirical systematic variance as per eqn. 1,2. The empirical variance would provide a conservative estimate since our mean estimates would be based on one systematic sample of 105 fls, not, for instance, on two groups of 52 fls or 3 groups of 35 fls. Though this approach would not account for model error, perhaps the model error component of the MD estimator could be added to this systematic sampling term to approximate the variance of biomass estimates generated based on a systematic sample of lidar flight lines. This hypothesis may be investigated with the Hedmark simulator.

References

- Andersen, H.E. and Winterberger, K., 2009. Estimating forest biomass on the western lowlands of the Kenai Peninsula of Alaska using plot- and lidar-based canopy surface structure measurements. In: S. Popescu, R. Nelson, K. Zhao, A. Neuenschwander (Eds.). Proceedings, Silvilaser 2009, College Station, Texas, Texas A&M University: 7 p. (conference CD).
- Andersen, H.E., Strunk, J., and Temesgen H., 2011. Using airborne light detection and ranging as a sampling tool for estimating forest biomass resources in the upper Tanana valley of interior Alaska. *Western J. of Applied Forestry* (in press).
- Asner, G.P., Powell, G.V.N., Mascaro, J., Knapp, D.E., Clark, J.K., Jacobson, J., et al., 2010. High-resolution forest carbon stocks and emissions in the Amazon. *Proc., National Academy of Science*, 107, 16738-16742.
- Boudreau, J., Nelson, R., Margolis, H., Beaudoin, A., Guindon, L., and Kimes, D.S., 2008. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. *Remote Sensing of Environment*, 112, 3876-3890.
- Gobakken, T., Næsset, E., Nelson, R., Bollandsås, O.M., Gregoire, T.G., Ståhl, G., Holm, S., Ørka, H.O., and Astrup, R. 2011. Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. *Remote Sensing of Environment*, submitted.
- Helmer, E.H., Lefsky, M.A., and Roberts, D.A., 2009. Biomass accumulation rates of Amazonian secondary forest and biomass of old-growth forests from Landsat time series and the Geoscience Lidar Altimeter System. *Journal of Applied Remote Sensing*, 3, 033505. [DOI: 10.1117/1.3082116]
- Nelson, R., Valenti, M., Short, A., and Keller, C., 2004. Measuring Biomass and Carbon in Delaware Using an Airborne Profiling LiDAR. *Scandinavian Journal of Forest Research*, 19(6), 500-511. [Erratum. 2005, 20: 283-284.]
- Nelson, R., Ranson, K.J., Sun, G., Kimes, D.S., Kharuk, V., and Montesano, P., 2009a. Estimating Siberian timber volume using MODIS and ICESat/GLAS. *Remote*

Sensing of Environment, 113, 691-701.

Nelson, R., Boudreau, J., Gregoire, T.G., Margolis, H., Næsset, E., Gobakken, T., and Ståhl, G. 2009b., Estimating Quebec provincial forest resources using ICESat/GLAS. *Canadian Journal of Forest Research*, 39, 862-881.

Nelson, R., Gobakken, T., Næsset, E. Gregoire, T.G., Ståhl, G., Holm, S., and Flewelling, J. 2011. Lidar Sampling – Using an Airborne Profiler to Estimate Forest Biomass in Hedmark County, Norway. *Remote Sensing of Environment*, submitted.

Ståhl, G., Holm, S., Gregoire, T.G., Gobakken, T., Næsset, E., and Nelson, R., 2011. Model-based inference for biomass estimation in a Lidar sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research*, 41, 96-107.

SilviLaser 2011, Hobart, Tasmania
Oct. 2011

Early Assessment of Industrial Needs: Harvesting and Allocation Decisions supported by ALS and TLS

Becker, Gero¹, Smaltschinski, Thomas¹, Opferkuch, Martin¹ & Weinacker, Holger²

¹University of Freiburg, Institute of Forest Utilization and Work Science,
fobawi@fobawi.uni-freiburg.de

²University of Freiburg, Department of Remote Sensing and Landscape Information
Systems, holger.weinacker@felis.uni-freiburg.de

1. Introduction

The wood industry has precise requirements towards the qualitative properties of its purchased raw material mainly regarding volume, species, dimension, length, diameter, knottiness, taper, sweep and the absence of defects. Thereof the results of inventories under central European conditions only cover information per species on wood volume and diameter. To plan their own supply with regard to product demands the wood industry needs early information on the prospective wood quantity from the designated harvestable stands with the respective quality properties.

To assess the volume and the distribution of assortments there exist numerous long-standing inventory methods within a certain range of error (SMALTSCHINSKI, 2009). When it comes to quality, objective methods are scarce. Laser technology could provide a technical solution to support quality assessment of sorting trees within a stand prior to cutting. Amongst other partners, the Institute of Forest Utilization and Work Science (FOBAWI) and the Department of Remote Sensing and Landscape Information Systems (FELIS) of the University of Freiburg are working together on this issue in the framework of the European project FlexWood (www.flexwood-eu.org).

Wood quality is determined by external and internal characteristics. Important external features are the stem form, size and number of live and dead branches, the knot-free length of the stem or visibly noticeable stem injuries. Internal properties apply to the annual ring width, density, early and late wood, reaction wood, and internal defects such as fungi or insects.

Here, at first the external quality features are considered and how they can be judged via airborne and terrestrial laser scanning. In addition, already existing recordings of terrestrial inventories and of forest management plans were integrated in this research in order to not exclude the possibilities for improved multiphase sample inventories (de VRIES, 1986). All recordings are in the same projection (Gauss-Krüger strip 3), to allow for simultaneous editing.

2. Material

The test area is located in the SW of Germany north of the city of Karlsruhe. The dataset of the actual forest inventory of Baden-Württemberg was available (KÄNDLER a. BÖSCH, 2002). Within the forest, the plots of the forest inventory form a regular grid of mesh size 100 x 200 m (Fig. 1).

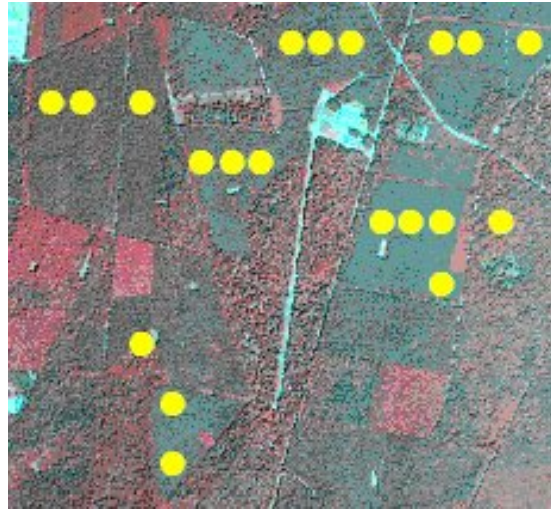


Fig. 1: Research area and plots of the forest inventory (yellow)

The design of the forest inventory provides at each sample plot four concentric circles. The trees are divided in 4 diameter-classes and in the fourth circle only the trees of the biggest diameter class are recorded, in the third only the trees of the third diameter-class and so on (KÄNDLER a. BÖSCH, 2002). From these data age series of the important tree species are deducible including stand parameters like diameter distribution, height curve and volume (SMALTSCHINKI a. BECKER, 2009). For this study in addition to the tree data of the forest inventory all trees of a sample plot were recorded by polar coordinates from the plot centre. The coordinates of the plot centre were remeasured as precise as possible (± 20 cm), by FELIS, to harmonize the georeferenced tree positions with remote sensing data (Fig. 2)

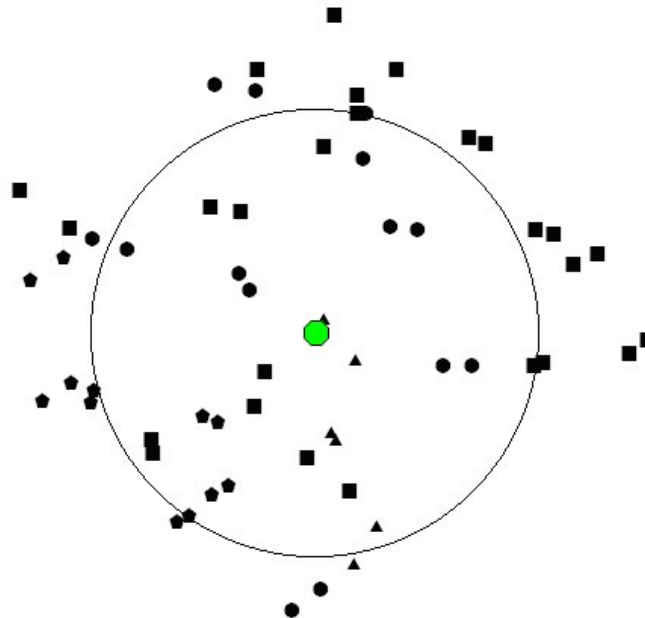


Fig. 2: Concentric circle of a sample plot of the forest inventory with a radius of 12 m and the positions of the plot trees and the plot centre (green) resp. additionally measured trees outside the circle. Rectangle signature *Fagus sylvatica*, circle *Quercus robur*, triangle *Carpinus betulus*, pentagram *Prunus avium*.

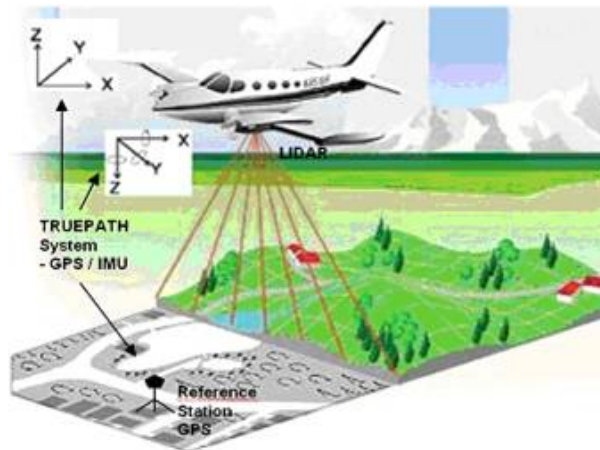


Fig. 3: Airborne Laserscanning

Over the test area the company TopoSys GmbH conducted a flight mission in summer 2007 using the Harrier 56 system. It consists of a full-waveform laser system (ALS) of the company Riegl (Riegl LMS-Q560). The system has a laser recording of the first and the last pulse. The used line scanner records the following four channels: B-G-R-NIR (Fig. 3).



Fig. 4: Trimble terrestrial laser scanner CX 3D

Furthermore, a terrestrial record (TLS) with the laser scanner CX 3D, (Fig. 4) by Trimble has been performed in the test area. The scanner records a hemisphere, so that, beside external quality characteristics on the trunk, the heights of the trees on the plot could be estimated or measured. Per direction 50,000 data points are measured. The angle of progress of the directions can be defined according to the object of focus.

3. Methods

To derive quality relevant data from the dataset of the test area (forest inventory, ALS, TLS) the following analytical steps were conducted.

3.1. 3D-shape crown structure model

At first a dataset of the digital terrain model (DTM) from the test area was created using the “last pulse” of the ALS. The raw data points of the laser scanner can be normalized with this digital terrain model. A vertical structure analysis was used for 3D modeling of the crown (WANG et al, 2005, 2008, KOCH et al., 2004). This analysis is based on a laser point cloud and the previously computed digital elevation model. They can then be assigned to a voxel space. The number of normalized raw data points that fall within a particular voxel represents the value of the voxel. The voxel space was divided into a series of 2D layers to reduce the complexity of the calculations. The actual algorithm for the formation of the crown margins is based on the evaluation of each horizontal projection image.

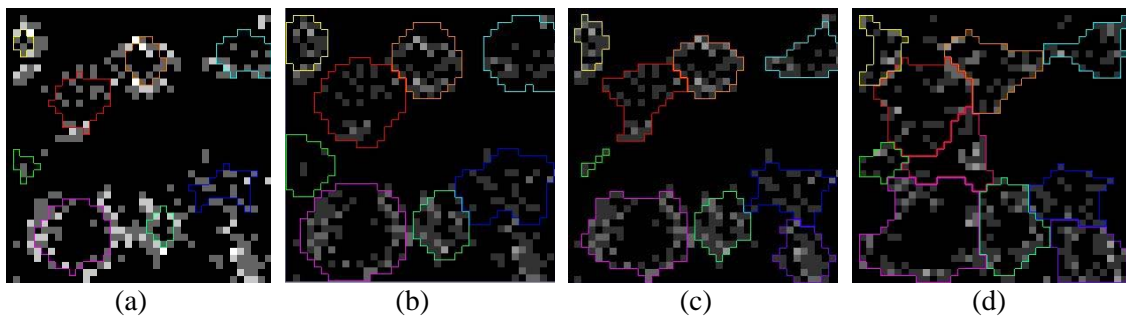


Fig. 5: Hierarchical morphological based crown outline extraction at several height levels. (a) 39m, (b) 38m, (c) 36m and (d) 35m.

The basic principle of the algorithm is the monitoring of crown boundaries in the projection images from top to bottom (WANG et al. 2007, WEINACKER et al. 2004). A hierarchical morphological opening and closing process with a set of predefined structural elements is carried out here (Fig. 5)

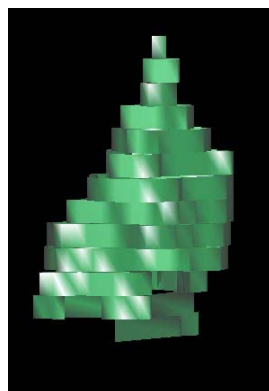


Fig. 6: Prismatic 3D tree structure model

The crown contours from higher levels already evaluated are expanded according to their proximity as “cluster property” in the current projection layer until they coalesce with neighboring regions. The “root node” of a tree is the region which is in the highest layer. A “pre-order” tree traversing process is carried out in order to “visit” all the regions that exist in the layers of different height classes. Because the voxels have a defined height (0.5-1m), 3D prisms can be reconstructed for each of the 2D crown regions in the different crown heights. The combination of these 3D prisms enables the generation of a prismatic 3D tree crown models for each tree (Fig. 6)

3.2. Development of a crown sphere model

For the determination of quality relevant parameters of the upper part of the tree, it is useful to convert the voxels or cells of the tree crowns into continuous crown sphere models (Fig. 7).

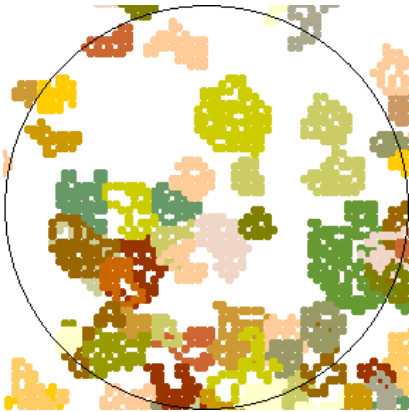


Fig. 7: horizontal projection of 3D points (voxels) of a sample of the inventory. Individual trees have different colors.

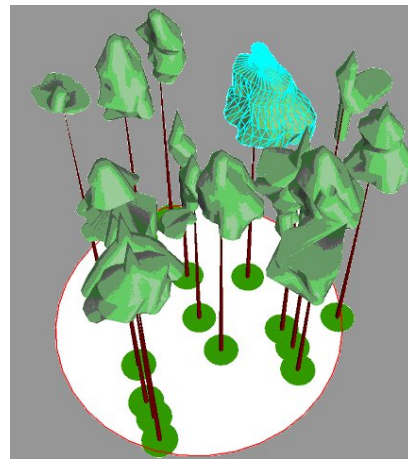


Fig. 8: Crown shapes derived from voxels of an inventory sample.

For that purpose the centers of the voxels were surrounded by a 25 cm buffer strip. The individual polygon rings of the crown were then meshed with each other, so that a closed space was created representing the crown surface (Fig. 8).

Each crown surface is a 3D-Shapefile, which was calculated in ArcGIS 9 as a spatial data set. This allows to link geometries of the crown with attribute data describing these geometries. Quality relevant crown parameters can then be derived and used to predict wood quality features - especially branches, knots - as follows. Fig 8. shows the crown shapes derived from voxels of the forest inventory sample. Fig. 9 shows the crown of the turquoise coloured tree of Fig. 8 in detail. With the quality relevant parameters which can be derived as attribute data from the geometry (triangular mesh) are shown as an example.

The crown length allows to separate the total high of the tree (stem) into the crown parts (which shows green knots as an important quality feature) and the trunk of stem which usually show dry or dead branches or no branches at all.

Well established allometric functions proposed by NAGEL (1997) can be used to derive the diameter of the tree:

$$CrownDiam(d) = (a_0 + a_1d)(1 - \exp[-(d/a_2)^{a_3}]) \quad (1)$$

$$CrownBaseHeight(d, h) = h(1 - \exp[-(b_0 + b_1h/d + b_2d)]) \quad (2)$$

Furthermore there exist well established allometric functions between branch length and branch diameter at the insertion points to the stem (KLÄDTKE, 2005). These functions link the diameter of a branch to its length. According to the European sorting rules for round wood, a 4 cm knot at the log surface is a limit between the classes B (good) and C (low grade). Fig. 10 shows how the branch length can be extracted from the crown sphere model and subsequently indicates the height, where quality decisive knot diameters are exceeded.

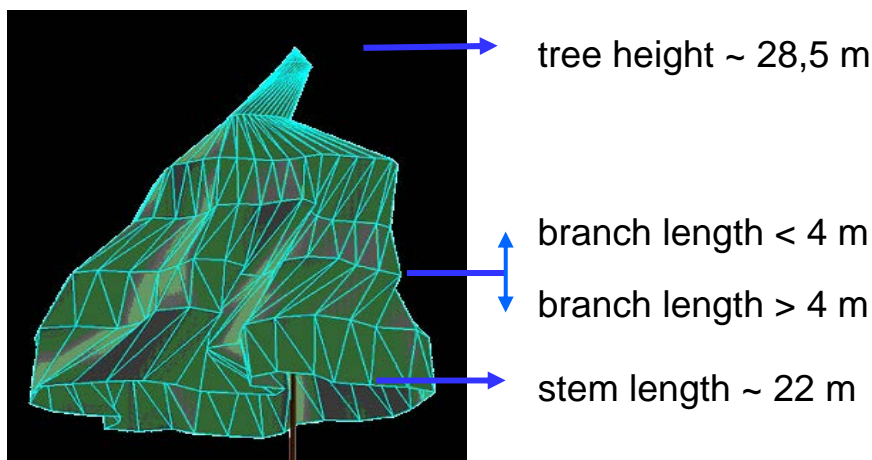


Fig. 9: values derived from the crown shape. Crown length = tree height – stem length

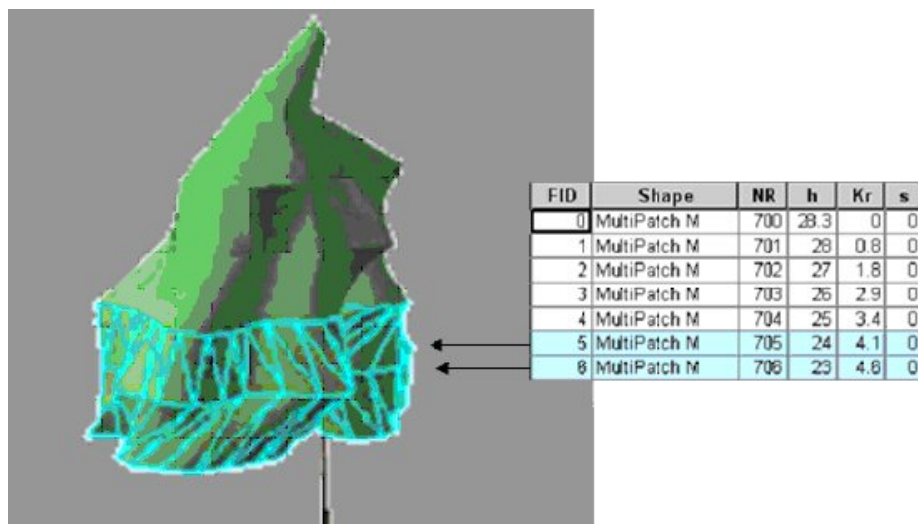


Fig. 10: Geometry and attributes of the spatial data set of the tree crown. The selection of the attribute crown radius (Kr) > 4 m leads to the height levels of 23 and 24 m. Here branches with diameters > 4 cm can be expected.

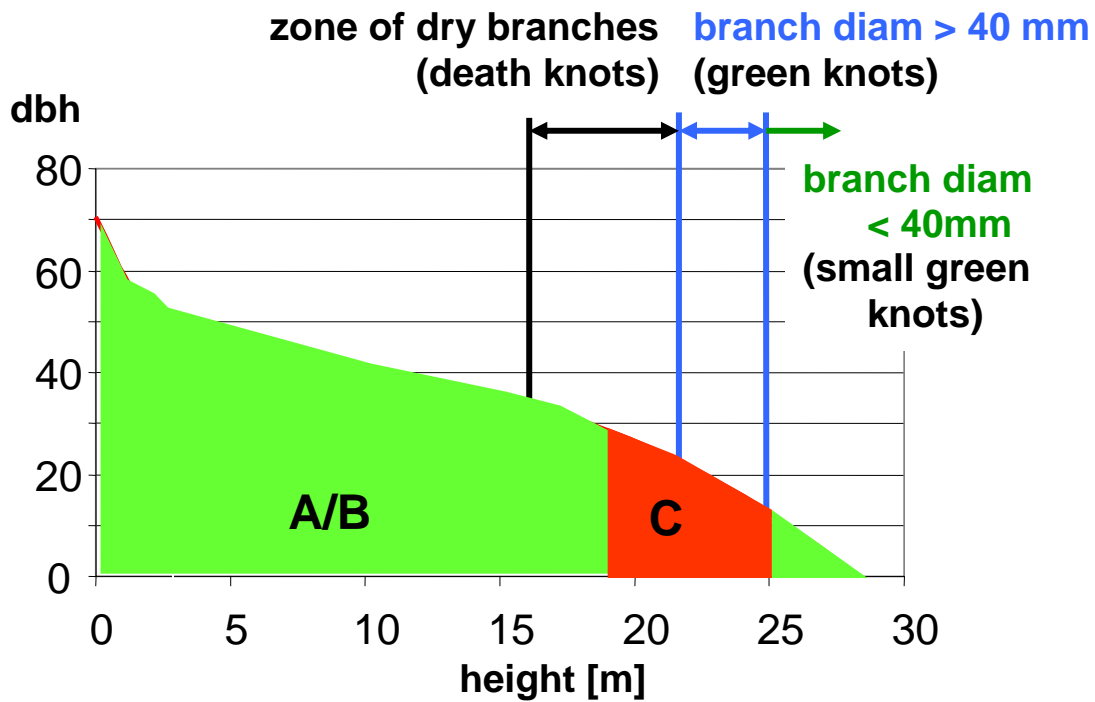


Fig 11: Taper curve of a tree with the quality levels A (very good), B (good) and C (bad)

3.3. Terrestrial laserscanning (TLS)

As shown above ALS-data are only useful to predict quality parameters, namely knot diameter, within the canopy. It is difficult if not impossible to derive quality features from the stem below the crown which normally represents the most valuable part of the whole tree. To get more information about this lower part of the tree, TLS-data may be used (and eventually be combined with crown quality parameters derived from ALS-data). Important stem quality features are length of the utilizable log, shape, diameter distribution along the stem and bark features including bumps, dead knots or bark injuries.

Fig. 12 shows an example of the laser scanning image of the trunk below the crown with quality relevant features like overgrown branches, bumps and bark characteristics.

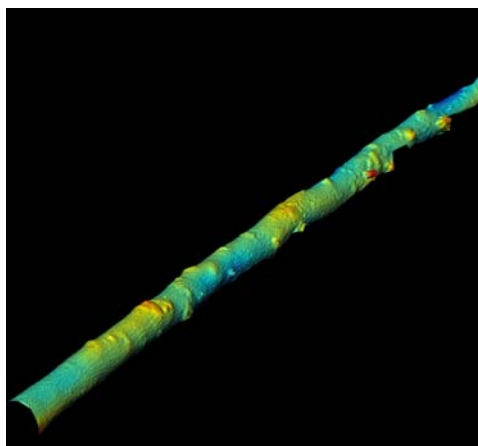


Fig. 12: Terrestrial laser scanning image: analysis capabilities of the trunk surface with overgrown branches and other characteristics

4. Conclusion and outlook

Future objective is to combine existing forest inventory data, with ALS-data combined with multispectral data to identify species, tree positions and crown shape and to derive quality relevant characteristics, namely knot sizes in the upper part of the tree and with TLS-data which can give a more detailed picture of the dimensions and quality parameters of the lower part of the stem.

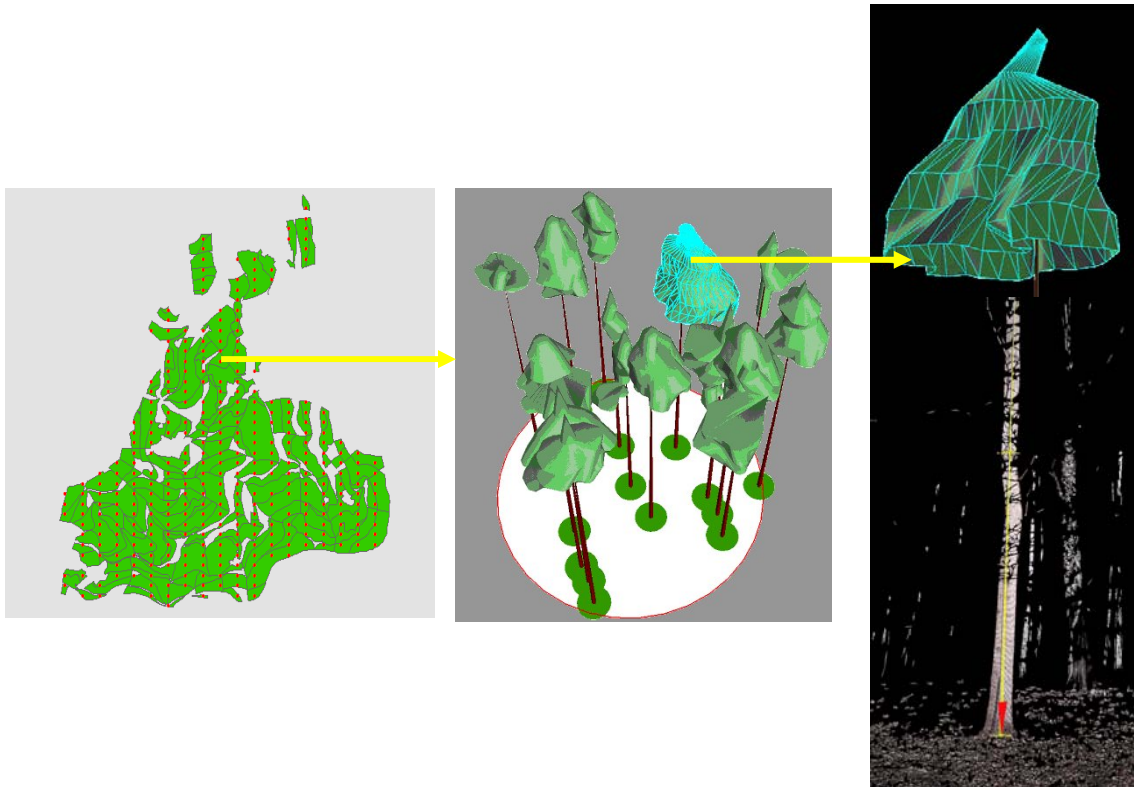


Fig 13: Combination of terrestrial inventory data, airborne and terrestrial laser scanning

Already today, the three inventory approaches allow predicting dimension and quality parameters on a stand level with mean values and distributions. To predict quality features on a single tree level, the challenge is to combine inventory data, ALS-data and TLS-data by exact geo-referencing (Fig. 13).

5. References

- de VRIES, P.G. (1986): Sampling Theory for Forest Inventory. *Springer Berlin, Heidelberg, New York*
- HRADETZKY, J., SCHÖPFER, W. (2001) Das Softwareprodukt Holzernete. *AFZ 56*, S1100-1101
- KÄNDLER, G., BÖSCH, B. (2002) Die Betriebsinventur als Grundlage für Planung, Steuerung und Kontrolle des Forstbetriebes. *Schriftenreihe Freiburger Forstliche Forschung, Bd. 18*. S. 252- 265
- KLÄDTKE, J. (2005) Möglichkeiten und Grenzen waldbaulicher Einflussnahme auf die Holzqualität von Z-Bäumen mittels Z-Baum-Normen. *Forst und Holz*, 60 Jg. Nr. 10: pp 403-407
- KOCH, B., HEYDER, R U., WEINACKER H.(2004) Detection of Individual Tree Crowns in Airborne Lidar Data. *Photogrammetric Engineering and Remote Sensing, Manuskriptnummer 04-046R*.
- NAGEL, J. (1997) BWINPro Program for stand analysis and prognosis. *User's manual Version 3.0*. Nieders. Forstl Versuchsanstalt.
- SMALTSCHINSKI, T. (2009): Precision Forestry und forstliche Wertschöpfungskette. GI-Edition, 29. GIL Jahrestagung März 2009, Rostock, zus. Mit G. Becker. p. 153-156
- SMALTSCHINSKI, T., BECKER, B. (2009): MatchWood - Vom Baum zum Produkt: Wertschöpfung durch Prozessoptimierung im Rahmen naturnaher Waldbewirtschaftung TP 2: Forest Warehouse und Logistik. Inst. Forstbenutzung und Forstl. Arbeitswissenschaft, 45 S.
- WANG Y, WEINACKER H, KOCH.(2007) Development of a Procedure for Vertical Structure Analysis and 3D-Single tree Extraction within Forests based on Lidar Point Cloud Proceeding , ISPRS (Hrsg)
- WANG Y., WEINACKER H., KOCH B.(2005) Automatic non-ground objects extraction based on Multi-Returned Lidar data. *Photogrammetrie, Fernerkundung, Geoinformationssysteme PFG*
- WANG Y., WEINACKER H., KOCH B. (2008) LIDAR Point Cloud Based Fully Automatic 3D Single Tree. *Modelling in Forest and Evaluations of the Procedure; Vol. XXXVII Suppl. Part B8*
- WEINACKER H., KOCH B. & WEINACKER R. (2004) Treevis: A Software System for Simultaneous ED-Real-Time Visualisation of DTM, DSM, Laser Raw Data, Multispectral Data, Simple Tree and Building Models. *Proceedings of the ISPRS working group VIII/2, Freiburg, October 3-6 2004*. ISSN 1682-1750, p. 90-95

Remotely Sensed Crown Structure as an Indicator of Wood Quality

A comparison of metrics from Aerial and Terrestrial Laser Scanning

Thomas Adams

Scion, 49 Sala Street, Rotorua, New Zealand. thomas.adams@scionresearch.com

Abstract

Aerial LiDAR offers a fast and efficient means to estimate wood quantity, but there has been little work to date on wood quality. In this study we investigate the hypothesis that remotely sensed crown structure from Aerial Laser Scanning (ALS) can be used as an indicator of log quality at an individual tree level.

A New Zealand *Pinus radiata* forest was flown with aerial LiDAR at 8 pts per m². Five trees from within the forest were scanned with a terrestrial laser scanner (TLS) to determine external signs of log quality. These measurements were diameter at breast height (DBH), volume, taper, sweep, lean, circularity and average internode distance. In this study we develop a series of metrics from ALS point clouds for each tree to describe the crown structure, which are then correlated against the TLS data. To derive these metrics, novel algorithms were developed for TLS data which extend the level of detail previously obtainable. These algorithms are also detailed in this paper.

As only five trees were studied, the results are proof-of-concept more than outright proofs. The purpose of this paper is to document techniques which will be employed in the future over a much greater sample, proving the preliminary findings presented here. In this small sample we found that crown area from ALS had a moderately strong correlation with DBH and sweep. Crown density from ALS was also moderately correlated to average internode distance. The correlations show that there is at least a moderate connection between crown structure and log properties, and that at higher LiDAR pulse densities and a larger sample size we can expect to describe this connection with greater certainty.

In further studies we also hope to correlate ALS and TLS metrics with internal wood properties, as found from destructive sampling.

Keywords

Aerial LiDAR, Terrestrial LiDAR, ALS, TLS, crown shape, stem shape, wood properties

1. Introduction

There have been numerous studies to quantify the biomass content of forests (Lefsky *et al.*, 2005; Stephens, 2010). For commercial forest managers it is important to know not just the *quantity* but also the *quality* of wood. This is the research question that we address here.

1.1 Wood Quality

Wood quality is the basis on which the value and use of timber is decided, and has been quantified – rigorously or not – for centuries. For an excellent overview of wood quality and its applicability to be remotely sensed the reader is directed to van Leeuwen *et al.*, 2011. In this study we use the term log quality to mean externally detected wood quality indicators on single trees.

Indicators of log quality from Terrestrial Laser Scanning (TLS) and field measurements were compared for with crown metrics from Aerial Laser Scanning (ALS) for five *Pinus radiata* trees

from a commercial plantation in New Zealand. Log quality indicators derived in high detail from field measurement and TLS are extremely time-consuming and costly. The hypothesis tested here is that the canopy structure is related to log quality, and so ALS can offer surrogate indicators across a much larger scale and at a greatly reduced cost per tree.

1.2 LiDAR

Light Detection and Ranging – or LiDAR – has been used in forestry since the early 80s for providing 3D point cloud information on forests (Nelson *et al.*, 1984). Lim *et al.*, (2003) gives a thorough description of the technology, whilst Adams *et al.*, (2011) details the applications for New Zealand commercial forestry. Within commercial forestry LiDAR is predominantly operated from two platforms – aerial and terrestrial. For an explanation of ALS and TLS systems the reader is directed to Lim *et al.*, (2003).

2. Method

2.1 Study site and data collection

Five mature 30 year old trees in Kaingaroa forest, New Zealand were selected as representative in-stand trees. The area was flown with ALS at 8pts per m² in August 2006 by New Zealand Aerial Mapping with an Optech ALTM 3100 EA. The trees were selected to be in a flight line overlap where they would receive double the point density. They were subsequently scanned with TLS in December 2010 with a RIEGL PTM98 Laser Profile Measuring Systemset supplied by Aerial Surveys, set with a horizontal and vertical scan step of 0.036°. Each tree was scanned twice from opposite sides, and three artificial markers were used for scan alignment. After scanning, the trees were measured for diameter at breast height (DBH), felled and the branch locations noted.

2.2 Metric extraction

Table 1 gives a brief summary of the TLS and ALS metrics are given below, followed by an in-depth description.

Table 1 – Summary of metrics derived from TLS and ALS

TLS	
DBH	Diameter at breast height. The mean diameter 1.4m above the ground
Stem volume	The total stem volume from 0.4m above the ground to 25m
Form value	A value indicating taper, see equation 5
Sweep	The maximum deviation of the centreline in any 5m section between 0.4 and 20m up the stem
Lean	The elevation angle between the stem centre points at 0.4m and 20m, and a horizontal plane
Stem circularity	The average circularity (see equation 3) of the stem between 0.4 and 6m
Median internode distance	The median vertical distance between detected branch nodes in the bottom 25m of the tree
ALS	
Crown area	The projected area of the crown viewed from directly above
Crown volume	The net volume of the crown, fitted by convex hull
Crown circularity	The average circularity (see equation 3) of the crown as viewed from above
Crown density	The number of LiDAR canopy returns divided by the crown volume
Net bending moment	The net bending moment on the stem base if every LiDAR return had unit mass
Net water transport distance	The total travel distance required to supply all detected foliage elements with water from the stem base

2.3 TLS data

Initially the two TLS scans for each tree are centred on the mean position of the artificial markers (which were placed around the tree), and clipped to remove points belonging to other trees. The centre points ($\mathbf{P}_1, \mathbf{P}_2$) of the artificial markers are found in both scans and the least-squares solution to

$$\mathbf{R}\mathbf{P}_1 + \mathbf{T} = \mathbf{P}_2 \quad (1)$$

is obtained, where \mathbf{R} is the rotation matrix and \mathbf{T} is the translation vector. This is solved using Horn's quaternion-based method without scaling in Matlab (MathWorks, 2000). In some instances a fourth datum was used – normally a distinctive branch fork high in the tree visible in both scans – to improve the fit in 3D. Once the scans are aligned they are merged. Figure 1 shows a merged point cloud with the two scans in different colours.

2.4 Fitting a geometric model to the stem

The model starts with 0.1m vertical slices of the point cloud, for which the centre is found by a least squares circle fit as in Bienert *et al.*, (2007). If the tree is approximately centred at (a,b) with radius r , we look to minimise the error term ε in

$$(x - a)^2 + (y - b)^2 - r^2 = \varepsilon \quad (2)$$

Where x and y are the coordinates of each return. This minimisation is performed in Matlab using the `fminsearch` function, which is an unconstrained nonlinear optimisation algorithm based on the Nelder-Mead simplex method (Lagarias *et al.*, 1999). The crucial difference here is that we are only looking for an approximate centre, not the diameter of the circle as in Bienert.

This approach is widely used and accepted, but can be thrown off by buttressing, or material close to the stem such as branches or dead needles. To minimise this effect points are removed that are at a distance of $1.25\times$ the previous radius but this is still not perfect, particularly when there is a lot of dead foliage. Figure 2 shows examples of this.

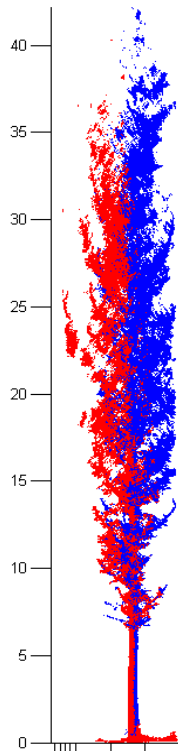


Figure 1 – Aligned point cloud from two scans

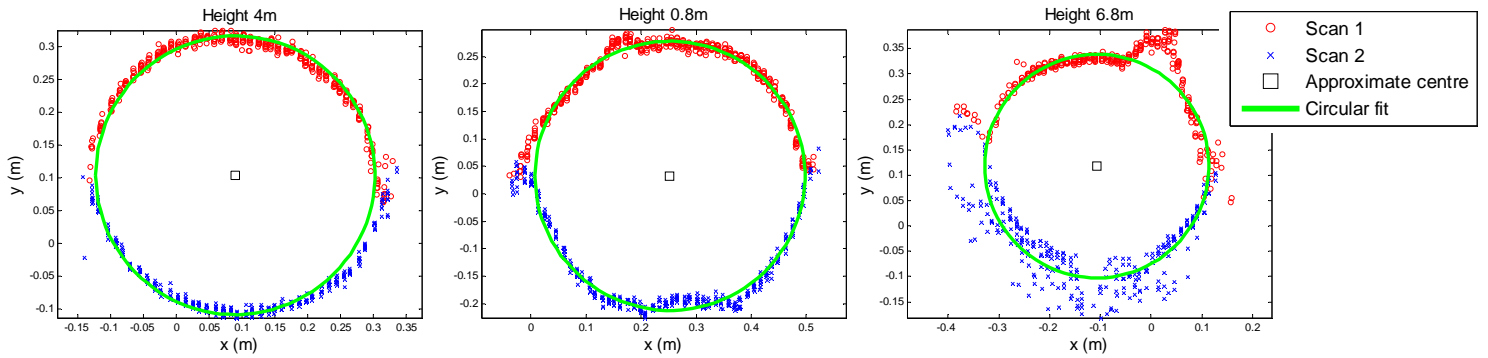


Figure 2 - Circular fitting on a) round section b) buttressed section c) section with dead foliage against stem

In this study we are looking for log quality information, so it is crucial to describe the stem as accurately as possible. To get around the circle-fitting limitations, we describe the stem cross section as a circular harmonic. In polar coordinates (ρ, θ) with the origin on our approximate centre (a, b) , we remove any points more than 0.05m away from our circular fit. The remaining points are then binned according to θ into 100 equally spaced bins from $-\pi$ to π radians. To eliminate the effects of branches and litter, the smallest value of ρ is taken for each bin. The remaining values for $\rho(\theta)$ can be approximated by a Fourier series, and by restricting the number of terms we can eliminate the higher frequency components to produce a smooth curve. Five terms were sufficient to allow for most features. This process is shown in figure 3 for the 4m height section shown in figure 2a. The stem sections shown in figure 2 are shown again in figure 4 with the improved fit (black line), which will be referred to as the harmonic fit. In addition the mean of the harmonic fit gives a slightly improved centre estimate, shown by the blue diamond in figure 4.

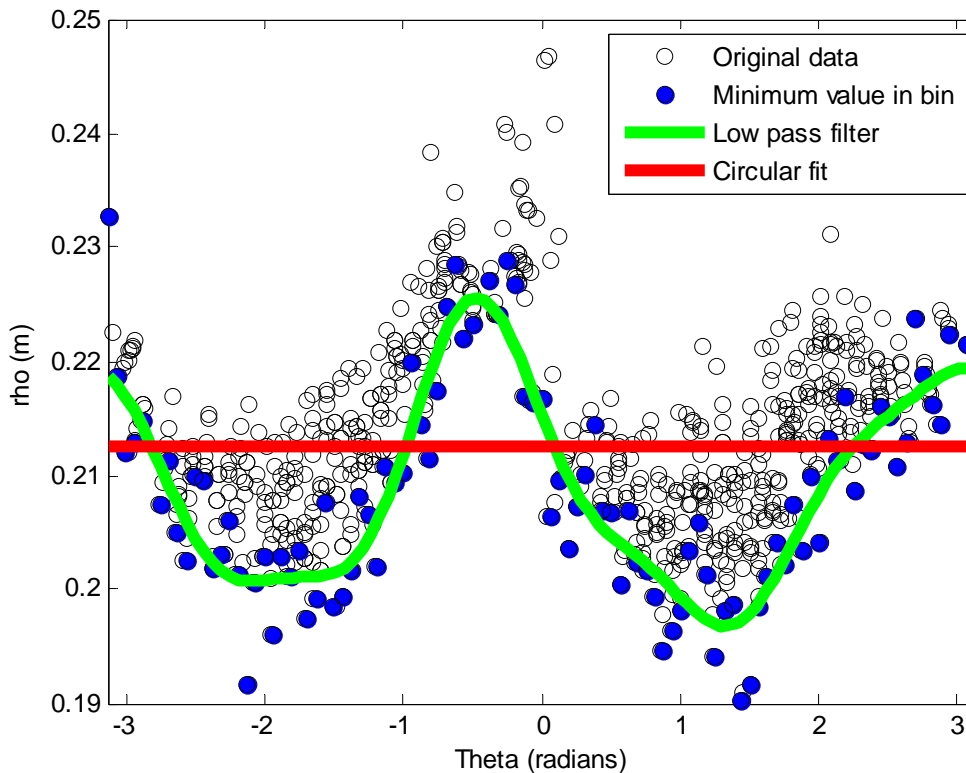


Figure 3 - Fitting a low-pass Fourier series to the stem points in polar coordinates

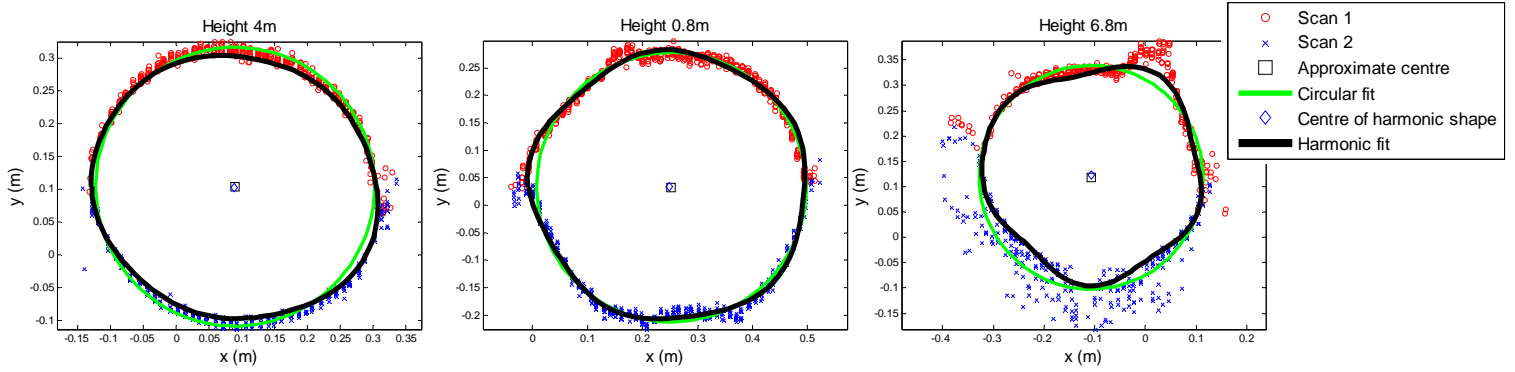


Figure 4 - Stem sections showing circular fit (green) and harmonic fit (black)

The stem tends to become obscured behind branches and other trees higher up. If the algorithm fails to find a good circular fit the vertical window is increased by a further 0.1m. This increase will continue until either a good fit is reached, or the algorithm fails on a 1m section in which case it gives up and moves further up the tree. The failed section can be interpolated later, although in general the data becomes sparse 20-25m up the tree. We have not analysed any of the data above 25m as it is too unreliable. Fortunately wood from the top of the tree is seldom used as timber due to its young age and narrow diameter. Figure 5 shows the series of stem shapes stacked to produce a meshed model of the stem viewed from the side and above.

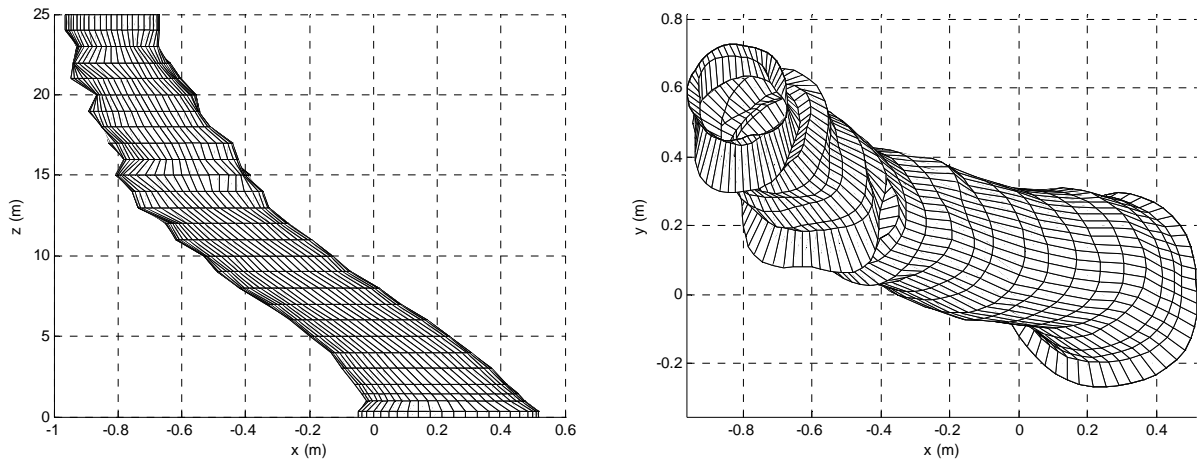


Figure 5 – Mesh construction of stem from TLS scanning.

2.5 Describing the stem

From the above analysis, it is easy to extract a centre point (x,y,z) for each slice, a circumference d , and a mean radius r .

We can also define stem circularity as S where

$$S = \frac{1}{4\pi} \frac{d^2}{A} \quad (3)$$

Where d is the circumference, and A is the area. Note that S is a unitless variable, independent of size. For a circle $S=1$, and will increase as the ratio of circumference to area increases.

In order to remove noise in r , the equivalent of a moving average filter is run across the series. Instead of taking the average value for the window, the 20th percentile was taken, as the values

have a much greater tendency to over-estimate than underestimate. After this filter a 6th-order polynomial is fitted to the calculated metrics x, y, r and S as in Bienert *et al.*, (2007). Figure 6 shows the curve fitting for these metrics.

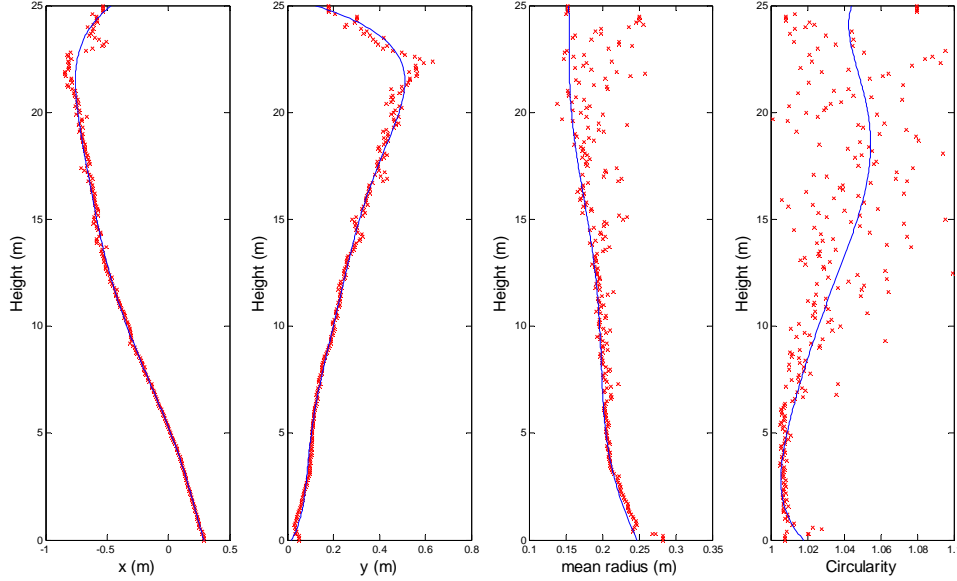


Figure 6 – Polynomial smoothing for stem metrics x, y, r_{mean} and S

We now define single values to describe our tree – DBH, volume, form value, sweep, lean and circularity. DBH is defined as twice the mean radius at 1.4m. Volume is the sum of the volumes of each 0.1m slice assuming a circular cross section of radius r_{mean} , given in equation 4.

$$V = \sum_i (0.1\pi r_i^2) \quad (4)$$

Form factor f represents taper, and is essentially the ratio of space filled by the stem vs. the available space if it had no taper. We consider the usable stem from 0.4m to 25m. Schardt *et al.*, (2002) gives the following equation for f

$$f = \frac{4V}{\pi h d^2} \quad (5)$$

Sweep is initially defined at every point on the stem between 2.5m and 17.5m based on a 5m sliding window. Within each window position, two points at the top and bottom \mathbf{X}_1 and \mathbf{X}_2 define the average lean. The sweep for that window is then defined as the maximum deviation of any other centre point from that line. To calculate this we use the distance from a point to a line in 3D as given in Anton (2010).

$$d = \frac{|(\mathbf{X}_0 - \mathbf{X}_1) \times (\mathbf{X}_0 - \mathbf{X}_2)|}{|\mathbf{X}_1 - \mathbf{X}_2|} \quad (6)$$

The maximum of these values is retained and used to quantify the sweep in the tree.

The angle between the stem at 0.4m and 20m was used to define lean. Circularity is defined as the average circularity of the bottom 6 metres of the stem (before it is affected by branching).

2.6 Finding branch nodes

Internode distance is also an important metric of log quality. In *Pinus radiata* branches tend to occur in nodes, and the distance between these nodes is crucial for log grading.

Once we have mapped the stem we can orientate each horizontal slice on the centre line, effectively straightening the tree. After removing points on the stem, the remainder are due to branches and leaf litter (which accumulate on the stem near branch clusters). To find branch clusters we adapt the theory behind a Hough transform – which looks for straight lines in space. We know that all branches must radiate from the stem, and that they should occur in an angle greater than 0° and less than 90° from horizontal.

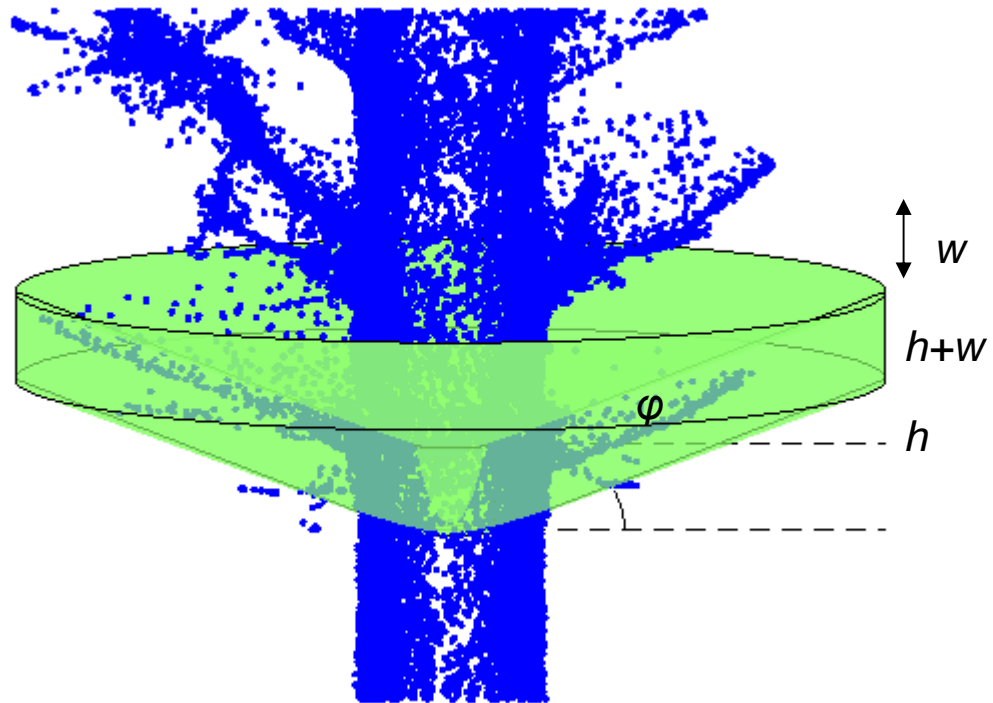


Figure 7 - Diagram of conical search for branches. Note returns from the stem would normally be removed prior to analysis

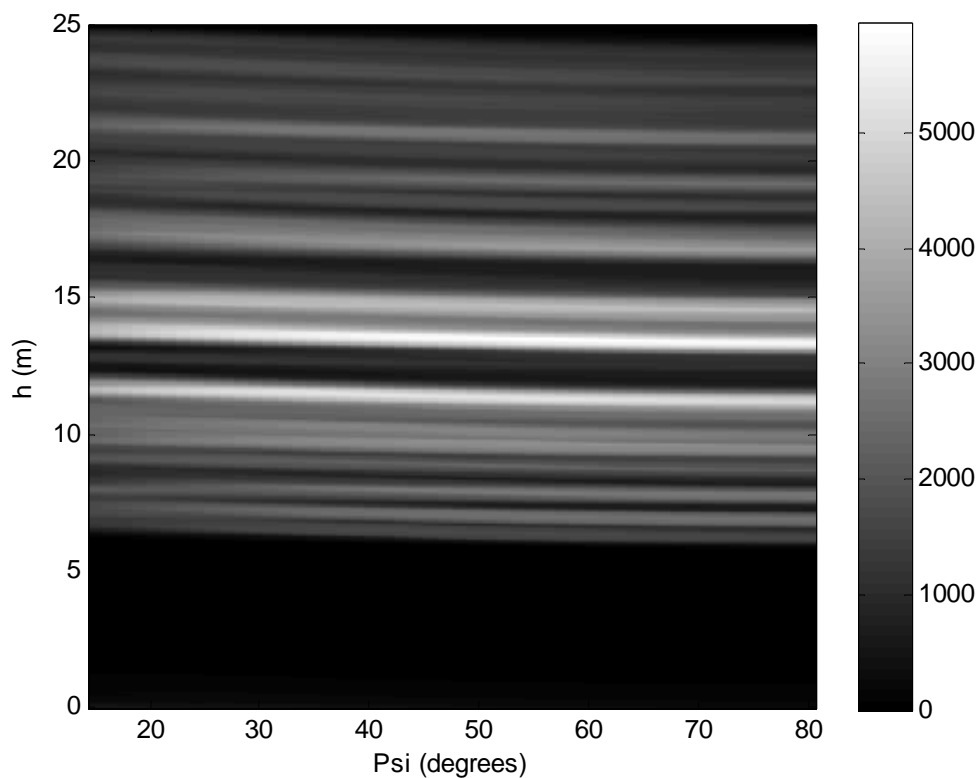


Figure 8 - Plot of $n(h, \phi)$, showing local maxima relating to branch clusters

So we construct a conical search around our stem, counting all points in the TLS point cloud that fall within a cone of width w and angle φ from a horizontal plane at height h (see figure 7). We sweep this cone through φ from 15° to 80° , and move h up the tree in 0.05m increments. In the end we have a value $n(h,\varphi)$ for every combination of h and φ . Local maxima in $n(h,\varphi)$ are likely locations of branch clusters, defined in terms of the height at which they intercept the stem and their average angle φ . A plot of $n(h,\varphi)$ is shown in figure 8. Note that maxima in h are much more apparent than in φ . This is because the branches in a cluster can have a range of φ , but all have a similar h .

If we define $N(h)$ as the sum of $n(h,\varphi)$ across all φ we get the plot shown in figure 9, shown alongside the centred point cloud for comparison. Locations with a high $N(h)$ are more likely to contain a branch cluster than locations with low $N(h)$. A simple first-derivative peak detection algorithm has been used to identify the peaks in $N(h)$, which can then be compared with the known branch locations (measured in the field) as shown in figure 10. In figure 10a the measured branches (shown in blue) are scaled in length and width by the average branch diameter. In figure 10b the detected branches (shown in red) are shown scaled in width and length by $N(h)$.

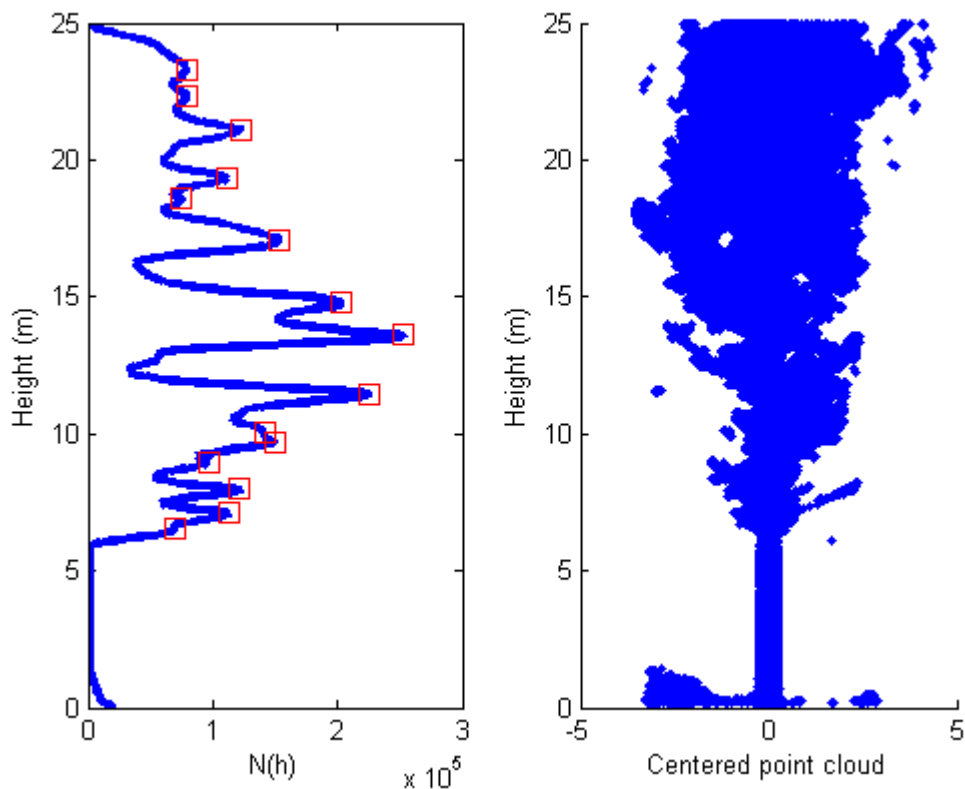


Figure 9 - Peak detection in $N(h)$ to detected branch clusters, and centred point cloud

In figure 10 it is apparent that the automatic detection does a good job of detecting the large branches that are lower in the tree. The algorithm is less dependable with higher and smaller branches. In this study we are interested in a single value to describe the general internode distance, so we take the median of the spacing between the detected clusters. In comparison, if we perform this median spacing approach on the real data (between all branches $>40\text{mm}$ in diameter) we get a comparable set of values, as shown in figure 11. This value is not meant to explicitly represent the exact internode distance – which varies up the tree and is largely dependent on what constitutes a branch – but is a comparable value that gives a general impression of log quality in terms of branch cluster spacing.

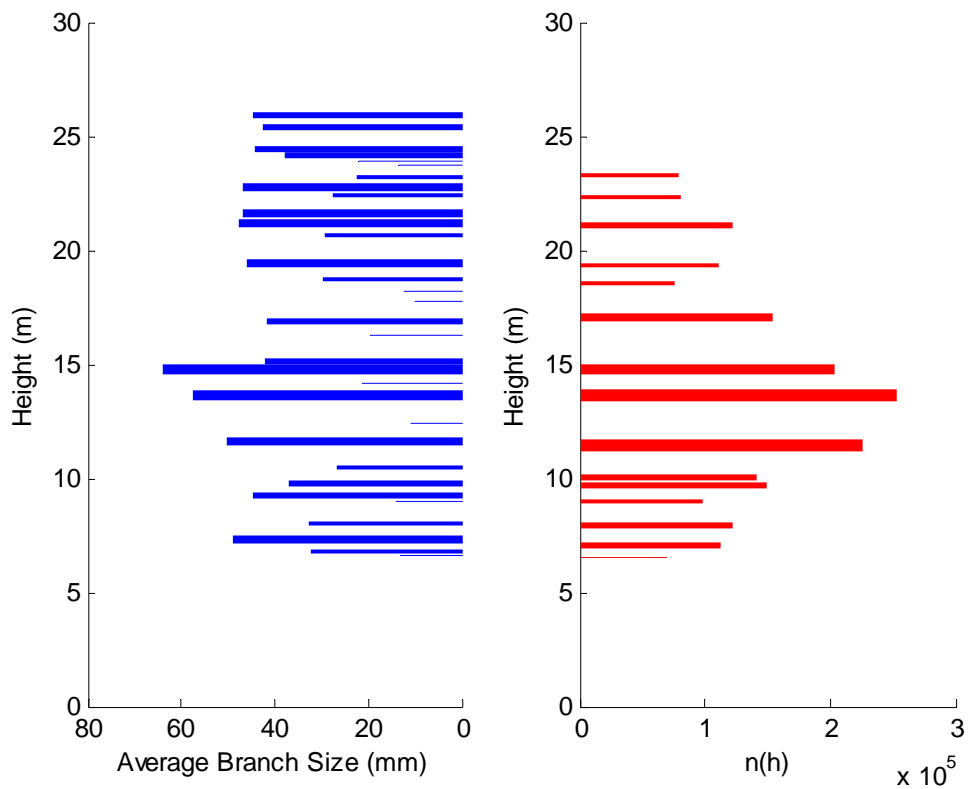


Figure 10 – a) Branches measured in the field and b) detected branches

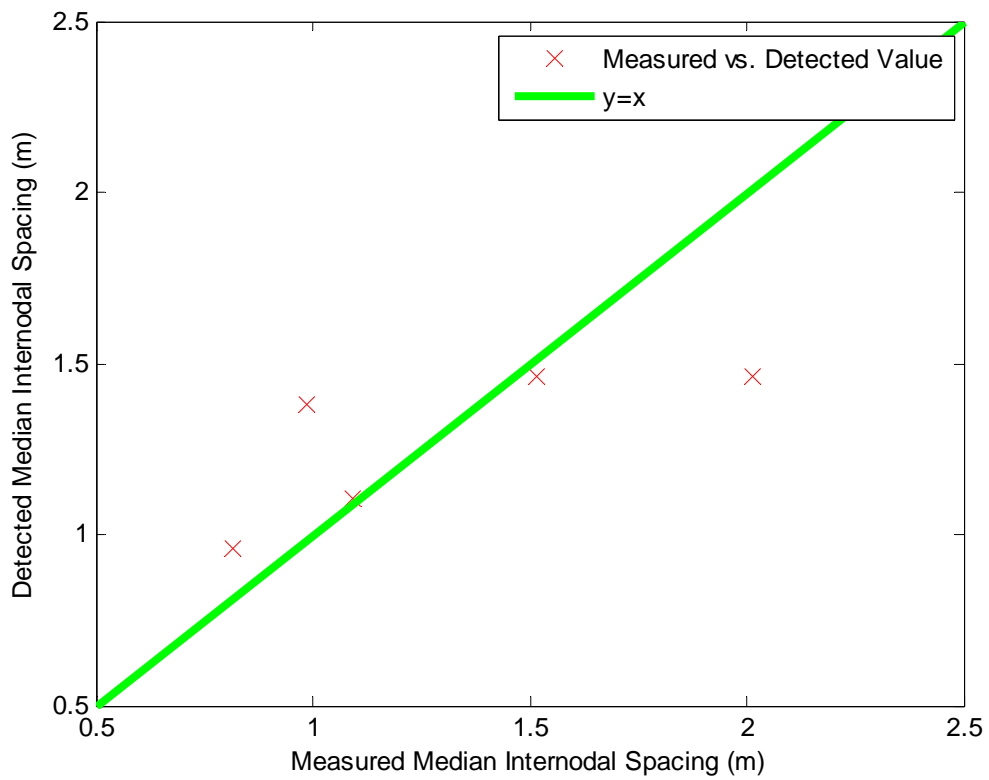


Figure 11- Comparison of median internode distance for five trees as measured on the tree and automatically detected in the TLS point cloud

2.7 Individual branch searching

To extend this work, it would be possible to retain θ as an independent variable, and count the number of points $n(h, \rho, \theta)$. Local maxima in $n(h, \rho, \theta)$ in 3D space would relate to individual branches. Individual branches could be identified, and potentially branch diameters determined. Figure 12 shows an example of this algorithm in operation. However, θ was not collected in the field for the branches on our study trees, so there was no way of verifying our findings. In addition, ALS will never pick up individual branches (except special cases in which the tree overhangs a road or compartment edge), so individual branch metrics were not used in this study. However, individual branch detection remains an interesting problem for future work.

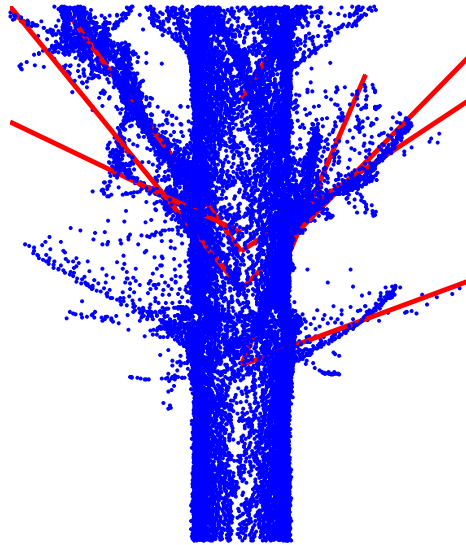


Figure 12 - Example of individual branch detection algorithm. Red lines show detected branches, blue dots show the point cloud. As there was no field data to verify the results this algorithm was not used to generate the final set of TLS metrics.

2.8 ALS data

For comparison with our TLS log quality metrics we derive a similar set from ALS to describe the structure of the crown.

2.9 Segmenting crowns from ALS data

Before the crowns could be analysed, they had to be segmented from the full set of ALS data. GPS points were taken with a Trimble ProXRT on each stump after felling to minimise the effect of canopy on GPS accuracy. All coordinates were differentially corrected and reported to be of sub-metre accuracy. Automatic segmentation is never perfect, so in this study an automatic segmentation was manually improved to give the best point clouds possible.

2.10 Crown area, volume, circularity and density

Once the crowns have been segmented a polygon was fitted to the 2D projection of the crown (figure 13a) and a convex hull to the canopy (figure 13b). To determine the 2D crown area, the centre point is set as the origin (defined as the mean of returns from the bottom $\frac{2}{3}$ of the tree), and the points converted to polar form. All points are then binned according to θ , and the maximum value of ρ is taken for each bin. Bins with no points are not included as vertices. Connecting these points results in our polygon which can be used to find the crown area A . Also the circumference d can be obtained and the circularity found as per equation 3.

$$S = \frac{1}{4\pi} \frac{d^2}{A} \quad (3)$$

The volume of the crown V is defined as the volume of a convex hull containing all crown returns. Convex hulls were created in Matlab using the `convhull` function, which is based on the Qhull algorithm in Barber *et al.*, (1996). Once V has been calculated the crown density is calculated as $\frac{N_{crown}}{V}$, where N_{crown} is the number of LiDAR returns in the crown.

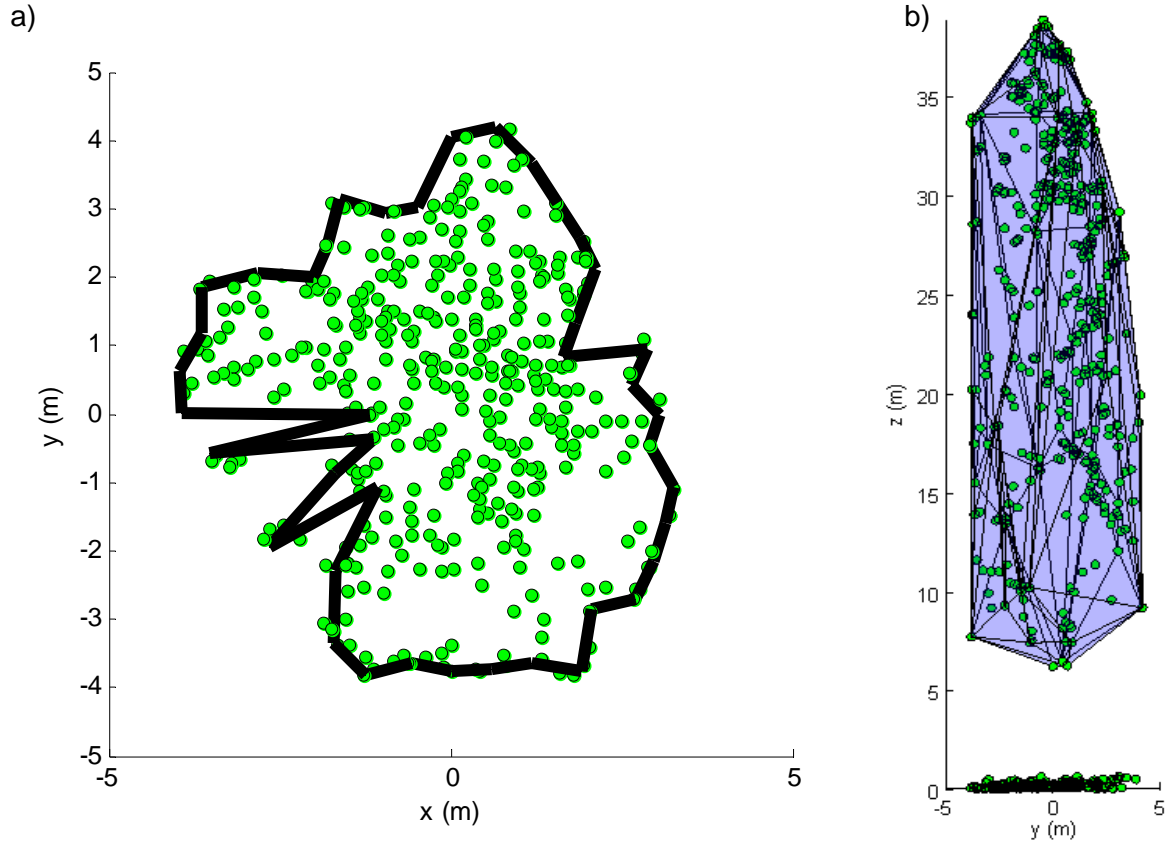


Figure 13 – a) Aerial view of crown with circumference defined (black line) and b) canopy with convex hull overlaid

2.11 Net bending moment

For this metric, we assume that every above-ground return relates to foliage or branch matter that is connected to the stem. This assumption is heavily dependent on good segmentation. If the assumption is true, then the mass (whatever it is) must exert a net bending moment on the stem. For example, if during the life of a tree its neighbour died opening up a canopy gap next to the tree, the tree is likely to grow towards the gap, perhaps straightening up again if the gap became filled. This would lead to the canopy being heavily weighted on one side of the stem, and would lead to compression wood on the side towards the gap. If we assume the net moment M is

$$M = |\sum_i \mathbf{r}_i \times \mathbf{F}_i| \quad (7)$$

where \mathbf{r}_i is the distance from the LiDAR return i to the point at which the stem joins the ground. \mathbf{F}_i is the force exerted by the mass at that point. Obviously we don't know the mass of the reflecting foliage – if it was foliage – but is a reasonable assumption to suggest that the foliage mass doesn't vary significantly across the regions in which it is detected. As we are only looking for relative measures, we can assign unit mass to each LiDAR point making

$$\mathbf{F}_i = m_i \mathbf{g} \approx \begin{bmatrix} 0 \\ 0 \\ 10 \end{bmatrix} \quad (8)$$

Where \mathbf{g} is the acceleration due to gravity. A larger assumption is to assume that the foliage detected is representative of all of the foliage on the tree. Many authors (Chasmer *et al.*, 2006; Hilker *et al.*, 2010) have noted that ALS tends to bias the upper canopy. We can justify this assumption by noting that under closed canopy conditions the lower canopy is hemmed-in, and any asymmetry is much more likely to become apparent in the upper canopy. In order to determine r_i , - the distance from each LiDAR return to the base of the stem - we must also make an assumption as to where this base actually is. As the path of the laser is always close to the zenith, the stems are generally obscured by the foliage and returns from the stem are rare and unreliable. So we make the assumption that the lower canopy is likely to be centred on the stem whilst the upper canopy is freer to deviate. Thus we determine our approximate stem-base location as the average of the returns from the lowest $\frac{2}{3}$ of the point cloud. This allows us to find r_i , and hence M .

2.12 Net water transport distance

If each LiDAR return in the canopy represents foliage, then there must be sufficient stem and branch infrastructure to supply it with water for photosynthesis to occur. The work of Pont (2003) shows that the stem diameter at any point can be related to distance above it that water must be transported to the foliage. Again, we do not know the exact path that the branches (and hence water) take, so we have to approximate a path that goes vertically from the stem base to the height of the return, and then radially outwards from there. If each return i is defined in cylindrical coordinates as (r, θ, z) , then W , the net water transport distance is

$$W = \sum_i r_i + z_i \quad (9)$$

3. Results

Results for the five trees are given in table 2. Table 3 shows a correlation coefficient matrix (\mathbf{R}) between the TLS metrics and the ALS metrics.

Table 2 – Metric results for five study trees

Method	Metric	Tree 1	Tree 2	Tree 3	Tree 4	Tree 5
TLS	DBH	0.472	0.485	0.516	0.399	0.474
	Volume	2.12	2.80	2.57	1.44	1.80
	Form value	0.492	0.617	0.498	0.468	0.414
	Sweep (m)	0.020	0.057	0.054	0.034	0.038
	Lean (°)	3.23	1.70	2.59	4.41	3.20
	Stem circularity	1.0081	1.0086	1.0148	1.0123	1.0048
	Median internode distance (m)	1.11	1.46	1.38	1.46	0.96
ALS	Crown area (m ²)	25.9	34.1	49.4	27.7	41.9
	Crown volume (m ³)	568	891	1117	720	1191
	Crown circularity	3.41	3.12	1.58	2.03	2.69
	Crown density (returns m ⁻³)	0.304	0.448	0.538	0.556	0.427
	Net bending moment (Nm)	2390	786	3610	1780	820
	Net water-transport distance	5220	10700	17800	11400	13500

Table 3 – Correlation coefficient matrix between TLS and ALS metrics.

		ALS					
		Crown Area	Crown Volume	Crown circularity	Crown density (returns m ⁻³)	Net bending moment (Nm)	Net water-transport distance
T L S	DBH	0.70	0.51	0.02	-0.22	0.33	0.36
	Stem volume	0.39	0.21	0.15	-0.12	0.17	0.17
	Form Value	-0.14	-0.23	0.30	-0.02	-0.12	-0.16
	Sweep (m)	0.65	0.62	-0.39	0.57	-0.03	0.70
	Lean (°)	-0.41	-0.34	-0.28	0.19	0.10	-0.16
	Stem circularity	0.24	-0.06	-0.76	0.68	0.77	0.46
	Median internode distance (m)	-0.08	-0.21	-0.41	0.65	0.20	0.22

4. Discussion

The purpose of this study is to investigate whether crown structure has any relationship to log quality, and if the structure can be quantified by aerial remote sensing. Although our small sample size of five trees is not sufficient for conclusive proof, it can suggest indicative TLS log quality metrics that can be approximated with ALS canopy metrics. DBH correlates moderately well with crown area ($R^2 = 0.48$). This is not surprising, and is similar to the work of Popescu *et al.*, (2003b). Crown volume also correlates well with crown area (not shown), with an R^2 of 0.84, although the DBH responds only weakly to crown area with an R^2 of 0.26.

It is surprising that stem volume does not correlate particularly well with crown volume ($R^2 = 0.04$), in contrast with Chen *et al.*, (2007) who obtained an R^2 of 0.78 with a similar linear relationship. Values for individual crown volume and area are extremely dependant on the segmentation, and it is likely this may have caused some errors. The fact that crown area was better than crown volume suggests that insufficient returns were obtained from the canopy base to reliably estimate the volume. In addition the good results from Chen *et al.* are based on summing individual crowns together to obtain a net volume for a stand. This technique is used as it makes up for poor segmentation, as omissions and commissions generally cancel out. It is very unlikely they would have got such a good result for individual trees. If timber volume is the principle aim, then empirical functions such as Stephens (2010) to determine volume over an area give better results than single trees analyses (Scharadt *et al.*, 2002).

Form factor does not show any promising correlations, perhaps due to its lack of variation across the five trees. Stem circularity also is very similar across our sample (the variation from maximum to minimum was only 2.5%). As a result, whilst this does show good correlations with net bending moment and crown density, it is unlikely that these models would hold on larger numbers of trees and samples with values for circularity ranging beyond those measured here.

Sweep exhibits moderate positive correlations with crown area, volume and density. This would imply that trees with more foliage are more prone to sweep. This agrees with the findings of Suarez *et al.* (2010) who found that stem straightness was inversely proportional to stand spacing. It is well known that canopy size and volume are proportional to stand spacing.

Lean did not correlate well with any ALS metrics, perhaps again because there was not a significant degree of variation within our sample. The median internode distance shows a moderate positive correlation with crown density, implying that as crowns become denser the branches become bigger and more spread out. This is feasible, as fewer larger branches would be likely to yield a greater number of LiDAR returns than many small branches due to the blind spot in discrete aerial LiDAR (see Reitberger *et al.*, (2008)).

Canopy metrics such as area, volume and density give better correlations than net bending moment and net water transport distance. This is probably due to the fact that the latter two metrics are extremely dependant on the supposed stem location – which simply cannot be inferred accurately from aerial LiDAR. The lack of correlation of lean with any metrics confirm that the canopy does not necessarily track symmetrically with stem, and thus our assumption of the stem base being the mean value of the returns from the bottom $\frac{2}{3}$ of the point cloud is likely to be false.

We have only been able to assess five trees. With a greater number of trees, and larger range in metric values, we may have been able to detect more correlations and achieve more meaningful correlations on those that we did pick up. Of the correlations we were able to detect, surrogate variables can be derived from ALS for DBH, sweep and internode distance. However the correlations were only moderate. There are two potential reasons for this:

- Crown structure is only moderately correlated to log quality
- ALS at 8pts per m² had insufficient resolution to determine canopy structure accurately enough

It is the author's opinion that the reality is a combination of the two, and more heavily weighted on the latter. By showing even moderate correlations between crown structure and log quality we know that canopy does - to some extent – indicate log quality. As remote sensing technology improves and becomes more affordable (higher point densities, greater resolution, improved canopy penetration etc.) it is extremely likely that these correlations will improve. Theoretically these correlations would asymptote to a value which expresses the 'complete knowledge' connection between crown structure and log quality, without any reduction due to remote sensing inaccuracies. The values shown in this study are more a statement about the ability of ALS (and subsequent algorithms) to describe crowns, than the actual correlation between crown structure and log quality. As technology improves we can get closer and closer to this theoretical 'complete knowledge' value, but we have shown that even with modest technology we can gain moderate inferences of practical use and value.

A subsequent part of this investigation will be to investigate links between ALS crown shape, TLS stem shape and internal wood properties.

5. Conclusion

Five mature *Pinus radiata* trees in Kaingaroa forest, New Zealand were flown with Aerial LiDAR (ALS) at 8pts per m² and scanned with terrestrial LiDAR (TLS). Novel algorithms were developed for the TLS data to extract log quality metrics for the trees. These were tree diameter at breast height (DBH), volume, taper, sweep, lean, stem circularity and internode distance.

The five study trees were manually segmented from the aerial LiDAR point cloud. Metrics were derived for each individual tree to describe the crown area, volume, density, circularity, net bending moment and water-transport distance.

Despite the small sample size, a promising relationship was found between ALS-derived crown area with DBH, and stem sweep. Crown density also showed potential as an indicator for internode distance. Net bending moment and net water transport distance did not show good correlations, most likely due to our inability to pinpoint the stem base in aerial LiDAR.

It is thought that the correlation between crown structure and log quality is greater than these results suggest, and the moderate strength correlations are due to poor resolution in the ALS. Higher point densities and technological improvements should increase the strength of these correlations.

Acknowledgements

This work was funded by Future Forests Research (FFR), and the Scion Capability Fund. Many thanks to Tim Farrier from New Zealand Aerial Mapping, Steve Smith from Aerial Surveys for the loan of the TLS, and David Pont, Peter Beets, Andrew Dunningham, Jonathan Harrington, Rod Brownlie, Mike Watt, Thomas Paul and Margaret Horner from Scion.

References

- Adams, T., Brack, C., Farrier, T., Pont, D., Brownlie, R., 2011. So you want to use LiDAR? - A guide on how to use LiDAR in forestry. *New Zealand Journal of Forestry* 55, 19–23.
- Anton, H., 2010. Elementary linear algebra. Wiley.
- Barber, C.B., Dobkin, D.P., Huhdanpaa, H., 1996. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)* 22, 469-483.
- Bienert, A., Scheller, S., Keane, E., Mohan, F., Nugent, C., 2007. Tree detection and diameter estimations by analysis of forest terrestrial laserscanner point clouds. In, *SilviLaser 2007*, pp. 50–55.
- Blair, J.B., Hofton, M.A., 1999. Modeling laser altimeter return waveforms over complex vegetation using high resolution elevation data. *Geophysical Research Letters* 26, 2509-2512.
- Bucksch, A., Fleck, S., 2009. Automated detection of branch dimensions in woody skeletons of leafless fruit tree canopies. In. *Citeseer*.
- Chasmer, L., Hopkinson, C., Treitz, P., 2006. Investigating laser pulse penetration through a conifer canopy by integrating airborne and terrestrial lidar. *Canadian Journal of Remote Sensing* 32, 116-125.
- Chen, Q., Gong, P., Baldocchi, D., Tian, Y.Q., 2007. Estimating basal area and stem volume for individual trees from lidar data. *Photogrammetric engineering and remote sensing* 73, 1355.
- Henning, J.G., Radtke, P.J., 2006. Ground-based Laser Imaging for Assessing Three Dimensional Forest Canopy Structure. *Photogrammetric engineering and remote sensing* 72, 1349.
- Hilker, T., van Leeuwen, M., Coops, N.C., Wulder, M.A., Newnham, G.J., Jupp, D.L.B., Culvenor, D.S., 2010. Comparing canopy metrics derived from terrestrial and airborne laser scanning in a Douglas-fir dominated forest stand. *Trees-Structure and Function*, 1-14.
- Hopkinson, C., Chasmer, L., Young-Pow, C., Treitz, P., 2004. Assessing forest metrics with a ground-based scanning lidar. *Canadian Journal of Forest Research* 34, 573-583.
- Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1999. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal on Optimization* 9, 112-147.
- Lefsky, M., Harding, D., Cohen, W., Parker, G., Shugart, H., USDA, F., 1999. Surface Lidar Remote Sensing of Basal Area and Biomass in Deciduous Forests of Eastern Maryland, USA.
- Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. Lidar remote sensing for ecosystem studies. *Bioscience* 52, 19-30.
- Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Espirito-Santo, F.D.B., Hunter, M.O., de Oliveira Jr, R., 2005. Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters* 32, L22S02.
- Lim, K., Treitz, P., Wulder, M., St-Onge, B., Flood, M., 2003. LiDAR remote sensing of forest structure. *Progress in Physical Geography* 27, 88.
- Maclean, G., Krabill, W., 1986. Gross-merchantable timber volume estimation using an airborne LIDAR system. *Canadian Journal of Remote Sensing* 12, 7-18.
- MAF, 2010. New Zealand Forest Industry Facts and Figures. In.
- Mannes, D., 2009. Operational Trial to Operational Reality - LiDAR at Forestry Tasmania. In, *ForestTech 2009*.
- MathWorks, I., 2000. MATLAB: the language of technical computing. Using MATLAB. MathWorks.
- Morsdorf, F., Kotz, B., Meier, E., Itten, K., Allgower, B., 2006. Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment* 104, 50-61.

- Morsdorf, F., Meier, E., Kotz, B., Itten, K.I., Dobbertin, M., Allgower, B., 2004. LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment* 92, 353-362.
- Murphy, G.E., Acuna, M.A., Dumbrell, I., 2010. Tree value and log product yield determination in 1 Radiata pine plantations in Australia: Comparison of terrestrial LIDAR with three other measurement systems.
- Nelson, R., Krabill, W., MacLean, G., 1984. Determining forest canopy characteristics using airborne laser data. *Remote Sensing of Environment* 15, 201-212.
- Nelson, R., Krabill, W., Tonelli, J., 1988. Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment* 24, 247-267.
- Pont, D., 2003. A model of secondary growth for radiata pine. *Masters thesis* University of Canterbury, Christchurch, New Zealand.
- Popescu, S., Wynne, R., Nelson, R., 2003a. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Can. J. Remote Sensing* 29, 564-577.
- Popescu, S.C., Wynne, R.H., Nelson, R.F., 2003b. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing* 29, 564-577.
- Reitberger, J., Krzystek, P., Stilla, U., 2007. Combined tree segmentation and stem detection using full waveform lidar data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36, 332-337.
- Reitberger, J., Krzystek, P., Stilla, U., 2008. Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *International journal of remote sensing* 29, 1407-1431.
- Riano, D., Chuvieco, E., Condés, S., González-Matesanz, J., Ustin, S., 2004. Generation of crown bulk density for *Pinus sylvestris* L. from lidar. *Remote Sensing of Environment* 92, 345-352.
- Riaño, D., Valladares, F., Condés, S., Chuvieco, E., 2004. Estimation of leaf area index and covered ground from airborne laser scanner (Lidar) in two contrasting forests. *Agricultural and Forest Meteorology* 124, 269-275.
- Schardt, M., Ziegler, M., Wimmer, A., Wack, R., Hyypä, J., 2002. Assessment of forest parameters by means of laser scanning. *International archives of photogrammetry remote sensing and spatial information sciences* 34, 302-309.
- Stephens, P.R., Kimberley, M.O., Bell, A., Brack, C., Searles, N. and Hagger, J., 2010. Airborne scanning LiDAR in a double-sample forest carbon inventory. *Report for Ministry for the Environment*, Wellington.
- Suarez, J.C., Gardiner, B.A., Luca, M.d., Goudie, J., Polsson, K., 2010. Consequences of stand structure on timber quality. In, *SilviLaser 2010*, Freiberg, Germany.
- Thies, M., Pfeifer, N., Winterhalder, D., Gorte, B.G.H., 2004. Three-dimensional reconstruction of stems for assessment of taper, sweep and lean based on laser scanning of standing trees. *Scandinavian Journal of Forest Research* 19, 571-581.
- van Leeuwen, M., Hilker, T., Coops, N.C., Frazer, G., Wulder, M.A., Newnham, G.J., Culvenor, D.S., 2011. Assessment of standing wood and fiber quality using ground and airborne laser scanning: A review. *Forest Ecology and Management*.
- Watt, P., Donoghue, D., 2005. Measuring forest structure with terrestrial laser scanning. *International journal of remote sensing* 26, 1437-1446.
- Wezyk, P., Koziol, K., Glista, M., Pierzchalski, M., 2007. Terrestrial laser scanning versus traditional forest inventory first results from the polish forests. pp. 12-14.

Towards automated and operational forest inventories with T-Lidar

A. Othmani¹, A. Piboule², M. Krebs³, C. Stolz¹ and L.F.C. Lew Yan Voon¹

¹ Laboratoire LE2I – UMR CNRS 5158, F-71200 Le Creusot, France
{ahlem.othmani, christophe.stolz, lew.lew-yan-voon}@u-bourgogne.fr

² ONF, R&D department, F-54000 Nancy, France, alexandre.piboule@onf.fr

³ ENSAM, Equipe Bois, F-71250 Cluny, France, michael.krebs@ensam.eu

Keywords: terrestrial laser scanning, forest inventory, tree detection, DBH.

Abstract

Forest inventory automation has become a major issue in forestry. The complexity of the segmentation of 3D point cloud is due to mutual occlusion between trees, other vegetation, or branches. That is why, the applications done until now are limited to the estimation of the DBH (Diameter at Breast Height), the tree height and density estimation. Furthermore other parameters could also be detected, such as volume or species of trees (Reulke and Haala) . . . This paper presents an effective approach for automatic detection, isolation of trees and DBH estimation. Tree isolation is achieved using an innovative approach based on a clustering methodology followed by a skeletonization step. The DBH of trees is then determined automatically. The efficiency of our algorithm is evaluated with comparison with ground data, measured by classical methods.

1. Introduction

Terrestrial laser scanner has become, in recent years, an important focus in forestry area. This technology allows to completely digitalizing a forest as a dense 3D point cloud. This gives a faithful and very accurate picture of a stand at a given time. The visual exploration of 3D point cloud gives lots of possibilities concerning the description of stem and branch structure of a tree. However, the challenge is to extract pertinent and correctly estimated forest inventory parameters from these raw data.

The absolutely needed parameters are the detection of trees and the estimation of the DBH. This has been already achieved in different works (Dassot et al 2010), with different degrees of success and robustness, but the real benefit in using the T-Lidar for forest inventory is due to determining others parameters which are very difficult or long to identify using traditional field methods, as stem volume, total tree volume, tree height, flexuosity, branchiness . . .

The “Computree” software presented in this paper, is developed by Office National des Forêts (French institute of forest management of public forests), with the goal of an operational approach, identifying all the parameters which can be detected by the T-Lidar. This approach has to put into consideration the important needs of an applied method:

- Robustness regardless different stand types and compositions.
- The approach must be the most computerized as possible, minimizing the ratio between time-consuming (ground and treatment) and quality of the data extracted.
- Try to be the most exhaustive as possible in tree detection and measurement, which requires us to correctly take into account the inter-trees occlusions.

This paper is organized as follows: section 2 presents a short review of recent methods to isolate tree and estimate the DBH. Section 3 details the methods used in Computree followed by a result analysis, a discussion and a conclusion. In the present paper the analysis of Computree's results will focus on DBH and localization. Indeed, the estimation methods of the other parameters are currently a work in progress.

2. State of the Art

The algorithms developed for forest inventory applications and reported in the literature usually require an estimation of the Digital Terrain Model (DTM) (extraction of ground profile), the detection of the position of the trees, the estimation of the height and the DBH, and sometimes other parameters like volume. In the subsequent sections, we will give a brief description of the common methods used to obtain these parameters.

2.1 Summary of the different methods used in forest inventory applications

2.1.1 Digital Terrain Model

All the applications start by the generation of the DTM, which serves as a height reference for the determination of the DBH. Different approaches have been proposed:

1. **Lowest Z-value:** the point cloud is separated into horizontal grid with a regular cell size. In each cell, the lowest Z-value is selected and specified as a ground height reference.
2. **Density allocation along the z-axis:** the ground reference height is determined by an analysis of the vertical density point repartition.
3. **TIN model:** a triangulated irregular network (TIN) model is iteratively created and used to isolate ground points.
4. **TerraScan and TerraModeler software:** the DTM is generated using TerraScan and TerraModeler softwares by company Terrasolid Ltd, which use a low point's routine. This process is not always a simple task because of the ghost points occurring below the ground, which should be reclassified.
5. **Searching for maxima:** It is similar to density allocation along z-axis, followed by a neighborhood consistency check and bilinear interpolation.

2.1.2 Tree detection method

1. **Hough transformation**
2. **Fitting circles :** Segmentation based on point cluster search, followed by a least square circle fitting
3. **Point density raster analysis:** two-step segmentation method based on point density raster analysis. The first segmentation step is a point cluster search in a cross section of the point cloud. In second step all clusters are analyzed to determine their point density inside a raster.
4. **Crescent Moon method:** creating a 'Crescent Moon' template by subtracting two, radially shifted circles. The points situated at the minimum, the mean and the maximum angle of the template are used to fit a circle. When the diameter is inside a certain range, then the position and diameter of the tree is recorded, otherwise not.
5. **Improved Hough transformation:** the trees stems are detected by using a Hough-transformation and measured accurately using a circle fitting and an ellipse fitting algorithm.
6. **Fitting cylinder:** Reconstructing trees from 3D point cloud automatically by fitting consecutive cylinders to the trunks. The reconstruction based on nonlinear least squares estimation is finalized as far as the RSME of the fitting exceeds a pre-defined value.
7. **Circle fitting statistics:** applying a least square circle fitting routine of the circular package of R.

2.1.3 Tree Height method

1. **Difference between lowest and highest point:** The tree height is defined as the difference between the lowest point (soil reference for the DBH) and the highest point inside of a cut cylinder.
2. **Difference between DSM and DTM:** Calculating the height as the difference between the DSM (Digital Surface Model) and DTM.
3. **Simple taper function:** Extrapolation of a simple taper function from DBH to estimate the height.

2.1.4 DBH method

1. **Fit a circle at 1.3m:** The breast height diameter is determined by cutting a slice of thickness d at a height of 1.30 meter above the representative terrain model point. An adjusting circle is fitted into the 2D projection of the points of the slice.
2. **Manual Mark:** DBH of each tree in the forest was determined and durably marked with reflecting tape before any measurements were carried out. As a fourth method of DBH derivation the intensity images of the sample plot were used to measure DBH in the scans manually.

2.1.5 Stem profiles estimation

Because of the difficulty of the segmentation of the 3D point cloud, only two approaches were found: one that uses segmentation in a layer with height close to 1.3 m above terrain while the other calculates a stem profile at different heights. We did not find any approach that isolate and extract the whole tree.

2.2 Synthetic view of the combinations of methods

The following table summarizes the state of the art

Table 1: state of the art

<i>Authors</i>	<i>DTM</i>	<i>Trees Detection</i>	<i>Trees Description</i>	<i>DBH</i>	<i>Tree height</i>
Simonse et al. 2003	Lowest Z-value + exclusion cone	Hough transform	Layer between 1.25 and 1.35 m	Fit circle at 1.3m	-
Bienert et al. 2006	Density allocation along the z-axis	Fitting circles	Layers at different heights	Fit circle at 1.3m	Difference between lowest and highest point
Bienert et al. 2007	Density allocation along the z-axis	Point density raster analysis	Layers at different heights	-	-
Maas and al 2008	Searching for maxima	Fitting circles	Cutting a vertical cylinder + Layers at different heights	Fit circle at 1.3m	Difference between lowest and highest point
Aschoff and Spiecker 2004	TIN model	Hough transform + fitting ellipses + fitting circles	Layers at different heights	Fit circle at 1.3m	-

Thies and Spiecker 2004	Lowest Z-value	Hough transform + fitting circle	Layers at different heights	Manual Mark	Taper function
Brolly and Kiraly 2008	TIN model	Fitting circles Crescent moon method	Layers at different heights	Fit circle at 1.3m	Difference between DSM and DTM Taper function
Klemmt et al. 2010	-	Circle fitting statistics	Layers at different heights	Fit circle at 1.3m	-

3. Proposed methodology

3.1 Material

The data set presented in this paper was acquired with the terrestrial laser scanner Faro Photon 120 (Figure 1.a). This scanner with an horizontal and vertical fields of view of respectively 360° and 320° has a range up to 153 m with a distance accuracy of +/- 2mm. Seven minutes scans were chosen as a good time /quality trade-off, which give a cloud of 44 million of points (Figure 1.b). The distance between two points is 6.3 mm at 10 m (it can reach 1.6 mm at 10 m at maximum resolution). The acquired data in native Faro files format are filtered and exported in Faro Scene as binary format XYB, where each point is defined by its three Cartesian coordinates (x, y, z) and information on the reflectance.



Figure 1: (a) laser scanner Faro Photon 120, (b) a 44 million points scan quick view

3.2. Computree algorithm

3.2.1 Soil extraction and DTM generation

This step consists in determining if each point belongs to soil or vegetation. This allows a size reduction of the analysed cloud by searching trees only in vegetation points, and the computation of the DTM used to have Z base reference for each stem.

An horizontal grid with a regular cell size of 50 cm x 50 cm is created and for each cell, the lowest Z value of points is stored in the “Zmin grid”.

The density of points between Zmin and (Zmin + 32 cm) of each cell is computed. Each cell with a density below 200 pts/m² is affected with a NA value (no soil points in this cell).

Each cell is compared with the 4x4 cells around. If at least one neighbour cell has a Zmin forming an angle with the observed cell upper than 45°, then a NA value is also affected to the observed cell.

The DTM is computed from the Zmin grid, interpolated with a Delaunay triangulation. Each cell with a NA value is estimated as the inverse distance weighted mean of its neighbours in the

triangulation. The soil points – between Z_{min} and $(Z_{min} + 32 \text{ cm})$ – are separated from vegetation points which are the only points that will be analyzed in the next step.

3.2.2 Clustering

The goal of this step is to obtain clusters of points. All of the points from one cluster have to belong to only one tree stem. The cloud of vegetation points, is sliced in horizontal Z-layer 1 cm thick. In each Z-Layer, points are grouped according to their proximity: two points belong to the same group if their distance is lower than 3 cm. Clusters with three or less points are dropped for next steps. For each points cluster, a circle is fitted by a least square routine (assuming trunks are approximately circular in horizontal plane). If the error of the circle fitting is too higher, the cluster is also dropped. Each fitted circle is compared to fitted circles of the ten Z-layers above and the ten Z-Layers below. If at least one circle in these layers has its centre horizontally included in the observed circle, and has a similar radius (between 0.67 and 1.5 times the radius of observed circle), it is kept else it is dropped. This test assumes that a stem should have a minimum degree of vertical continuity.

At the end of this step only the clusters of points that have some degree of circularity (crescent-shape if only one centre scan has been done) should remain.

3.2.3 Creation of virtual sections

In this step, 1 cm thick clusters are aggregated in vertically-oriented virtual sections. Two clusters are affected to the same virtual section if the bounding rectangles are horizontally intersected, and their Z-distance is less than 50 cm. The bounding rectangle of one cluster is the smaller XY oriented rectangle that contains all the points of the cluster. The sections containing less than 20 clusters are dropped. Previously, dropped clusters are tested again with each virtual section by the bounding rectangles intersection routine.

At the end of this stage a real tree could be represented by several virtual sections (figure 2A), due to occlusions.

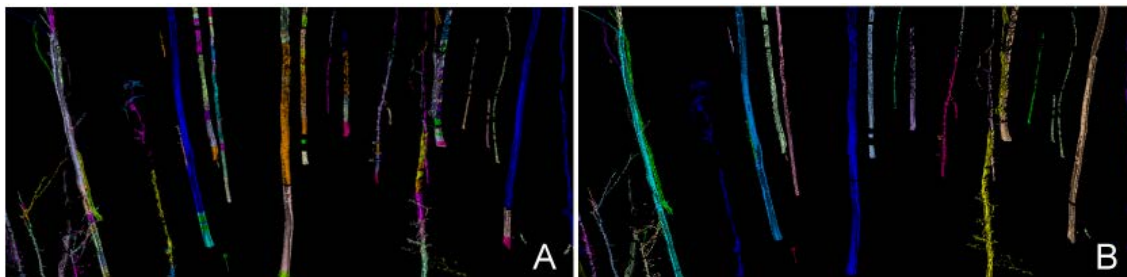


Figure 2 (A) trees represented by a mosaic of virtual sections, (B) trees after sections merging

3.2.4 Skeletonization

The goal of this step is to merge virtual sections belonging to the same tree. For each virtual section, a skeleton is created by merging the clusters of the section into skeleton slices, with a thickness of 10 cm. The nodes of the skeleton are defined as the barycenter of the successive skeleton slices. Two sections are merged together if they have approximately the same vertical extend (or one having its vertical extend included in the other ones), and having a maximum horizontal distance lower than 50 cm (figure 3A). The sections vertical continuity is tested. If the skeletons of two sections have their extremity in a reciprocal alignment, then the sections are merged in the same stem (figure 3B). Two sections are considered aligned if the directional extrapolation of the skeleton of each one is intersecting the other one's extremity, close to its node. The proximity criterion from node is defined by the horizontal extends of the extremity slice of the section. After merging the sections together the final skeleton is computed again and

smoothed (with a mobile-mean routine).

At the end of this step, it should be one virtual stem for each real tree stem or branch. The figure 2A shows the virtual sections, and the figure 2B shows the resulting virtual stems.

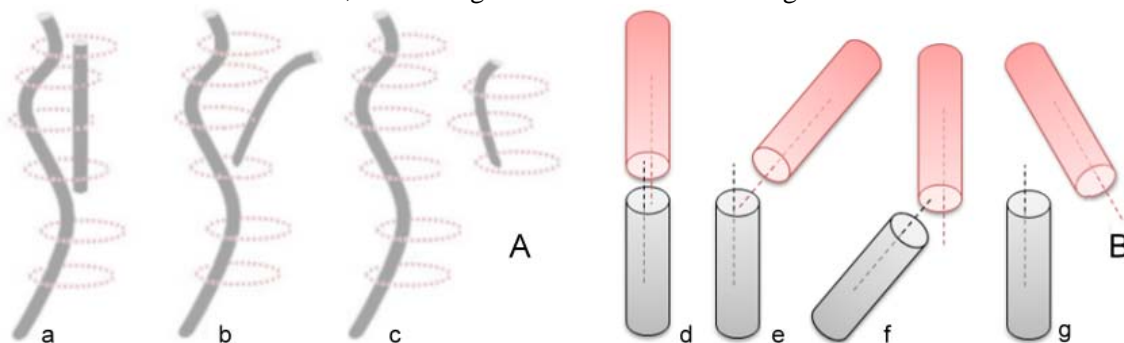


Figure 3: (A) virtual sections, (a) fusion, (b, c) no fusion, (B) segments, (d) fusion, (e,f,g) no fusion

3.2.4 Position and DBH calculation

This step fits a circle to each virtual stem at 1.30 m from the DTM. The centre of the circle gives the position of trees, and its diameter is given as an estimation of the DBH. For each skeleton node of a virtual stem, between 0 and 3.30 m from the DTM, a circle is fitted for the node's points, perpendicularly to the skeleton direction in this node. A linear regression is computed for the diameter of these circles against the height value. The outliers circles (from the Cook's distance criteria) are iteratively removed and the linear regression fitted again. The DBH is estimated from the fitted model.

4. Results and discussion

4.1 Tree detection

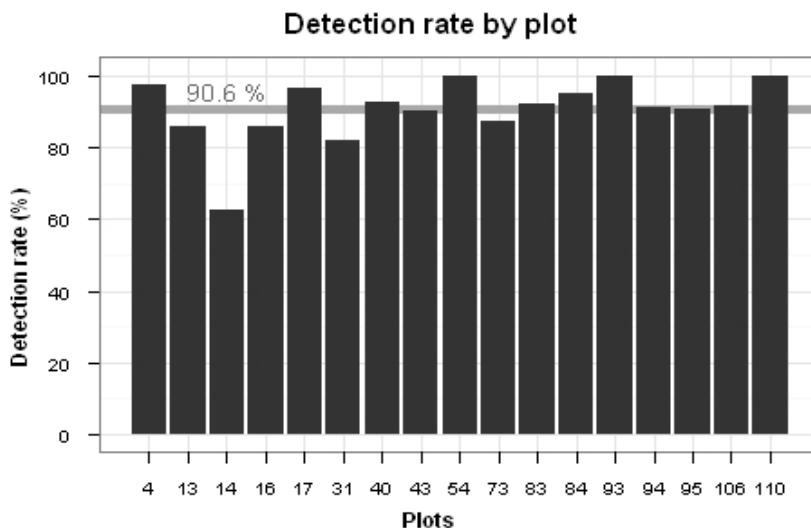


Figure 4 : Detection rate by plot

The test site was composed by 17 plots (of beech, with oak, hornbeam and sycamore maple) of mixed forest in state forest of Montiers-sur-Saulx, (Meuse, France) with a basal areas between 3 and 65 m²/ha and a density between 56 and 3600 stems/ha. The plots numbered (13, 14, 16, 17, 43, 84, 93, 95) are coppice with standards while the plots numbered (4, 31, 40, 54, 94, 106, 110)

are even-aged stands. The rest (73, 83) are pole stands. The plots were recorded from one laser scanner position. The figure 4 presents the detection rate by plot. Our algorithm has an average of detection rate of 90, 6% which shows excellent results especially considering the complexity of segmentation of 3D cloud point. Just one plot (14) presents a bad detection rate of 62,5%. Plot 14, has not many trees, far from the centre (between 13 and 15 m) with, between the centre and the trees, a lot of seedlings (3 to 6 m), which completely mask the background: there is a regeneration gap. The results of the detection and the isolation of trees done with Computree software are showed on figure 5.B.

Detection rate could probably be increased with a multi scans approach. It will be tested in the future, considering that our algorithm is low time-consuming, so testing a larger point cloud is not a matter. At the moment, the processing time for a 44 million point cloud is about 3 min on a computer with Intel © i5 processor. The time factor is extremely important for operational efficiency, it was taken into consideration throughout our development.

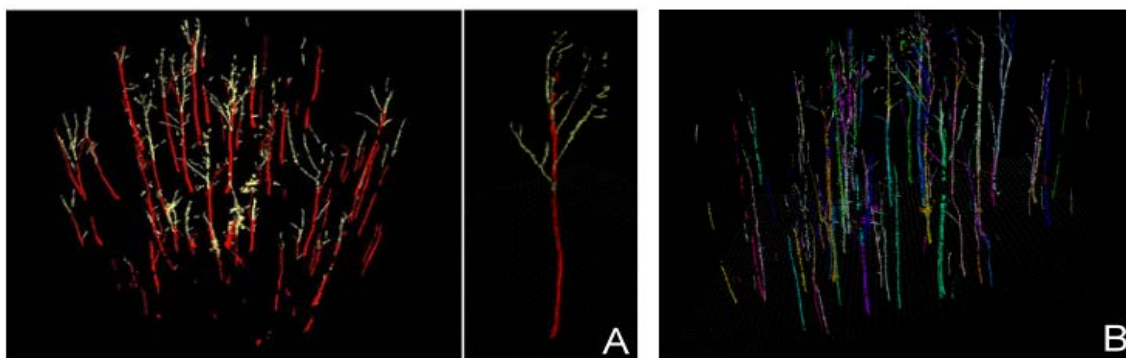


Figure 5: (A) isolated trunk (red color), branches (yellow color), (B) complete isolation of trees with Computree (a different color for each isolated tree)

4.2 DBH estimation

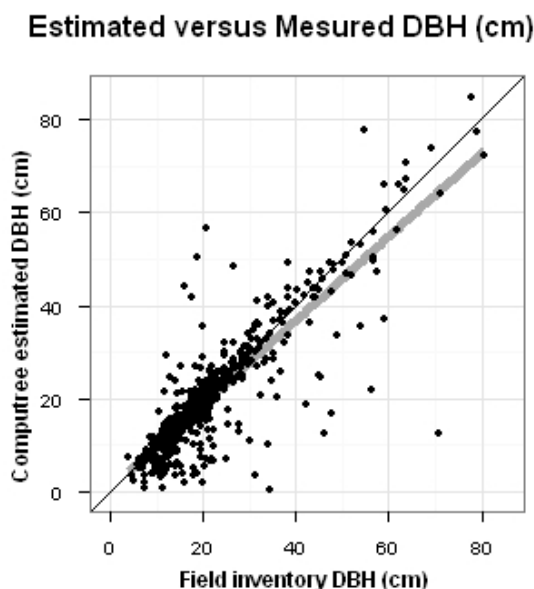


Figure 6 : Diameters computed from Computree Software in comparison with the diameters obtained manually. The identity line is figured as a black line, and the linear adjustment as a thick gray one

The dispersion between Computree estimation DBH and the DBH from the field inventory is presented in figure 6. The objective is to quantify the rate of data variability around the central value that is measured on the field. The algorithm is more powerful as the measures of

difference between the two data sets is optimal and our values are closer to the identity line. There is some little bias related to diameter. The estimation is globally good, 80 % of point having an absolute error below 5 cm. But it remains a little but significant proportion of very bad estimations. They are essentially due to some residuals problems in tree isolation, or to counter-productive effect of the Cook's Distance criteria (then it is more bad circle than good ones). One way of resolution would probably to improve clusters filtering before virtual section creation.

To well analyze the error estimation we study the error estimation by plot (figure 7).

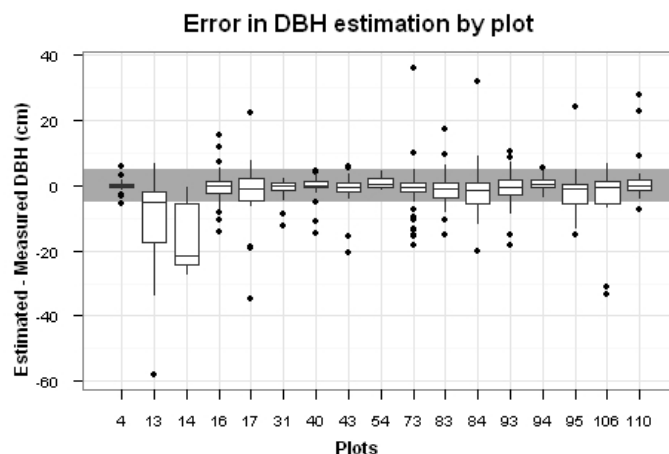


Figure 7: Error in DBH estimation

The figure 7 show the same results but splitted by plots. It gives for each plot the distribution of estimation error on the diameter. The gray area (median) represents an error of + - 5 cm. We can observe that the distribution is good and in the middle range for most of the plots, except for rare extreme outliers cited previously.

Two plots present a lower distribution: plot 13 and 14. For the plot 14 the result is a consequence of the structure of the plot which is a regeneration gap, containing a very important quantity of seedling, generating a lot of occlusions. The phenomenon is accentuated by the fact that the majority of tree in this plot are far away from the centre. The plot 13 is in a similar situation of occlusion by seedlings, but with a best spatial distribution of trees.

4. Conclusion and perspectives

The approach described provides an automatic method of detecting trees in terrestrial laser scans of forest sites and estimating their DBH. Our approach is very relevant and solves several problems of segmentation of 3D point cloud such as shadow effects and mutual occlusions. Our algorithm is very fast: a 3D point cloud of about 44 million points is processed in few minutes. The process of detection of trees gives a mean rate of 90% what is very accurate. In the future, it is planned compare simple scan versus multiple scans efficiency of Computree estimations, particularly for the case of plots with lots of seedlings. For the estimation of DBH, the results are very encouraging, but some problems have to be solved to avoid the small remaining proportion of outliers. The skeletonization approach seems to be a very interesting way of tree isolation. Several improvements are planned with a more comprehensive study of the ramification of the branches. We also have to implement and test other parameters calculations, like volume or branchiness. Finally a Ph. D. Work has been initiated for automatic species identification.

Acknowledgements

This work was initiated and financially supported by French ANR 2008 BIOE 003 Bio-energy EMERGE project and the ONF (Office National des Forêts). The Ph.D. of Ahlem Othmani on automatic species identification is funded by the Conseil Régional de Bourgogne and the ONF.

The authors wish also to thank particularly some EMERGE project contributors:

- Aurélie COLIN for her contribution in the evaluation of the algorithms developed and implemented in the Computree software.
- The LERFoB UMR INRA-AgroParisTech 1092, who supply the Faro 120 laser scanner.

References

- Reulke, R. and Haala, N., 2005. Tree species recognition with fuzzy texture parameters. *Lecture Notes in Computer Science. Springer Berlin 3322: 607-620.*
- Dassot T., Constant and M. Fournier, *Annals of Forest Science*, 2010, The use of terrestrial LiDAR technology in forest science: application fields, benefits and challenges *Official journal of the Institut National de la Recherche Agronomique (INRA)*
- Simonse, M., Aschoff T., Spiecker H. and Thies M., 2003, Automatic determination of forest inventory parameters using terrestrial laserscanning. *In Proceedings of the ScandLaser Scientific Workshop on Airborne Laser Scanning of Forests*. Umea, Sweden, pp. 251–257.
- Bienert A., Scheller S., Keane E., Mohan F., and Nugent C., 2007. Tree detection 700 and diameter estimations by analysis of forest terrestrial laser scanner point 701 clouds. *Laser scanning 2007 and Silvilaser 2007. ISPRS Commission III 702 Workshop 36, Part 3*, Espoo, Finland. Sept. 12–14, 2007
- Maas H.G., Bienert A., Scheller S. and Keane E., 2008. Automatic forest inventory parameter determination from terrestrial laser scanner data. *Int. J. Remote Sens.* 29 (5): 1579-1593.
- Aschoff T., Thies, M., Spiecker, H., 2004. Describing forest stands using terrestrial laser-scanning. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. XXXV, Comm. 5*, pp. 237- 241.
- Thies M., and Spiecker H., Evaluation and future prospects of terrestrial laser-scanning for standardized forest inventories. *In Proceedings of the ISPRS working group VIII/2 Laser-Scanners for Forest and Landscape Assessment, 2004.*
- Kiraly G., and Brolly G., (2008): Modelling single trees from terrestrial laser scanning data in a forest reserve. *The Photogrammetric Journal of Finland.* 21 (1): 37-50.
- Hans-Joachim K., Seifret T., Seifert S., Kunneke A., Wessel B. and Pretzsch H., Assessment of branchiness in a Pinus pinaster plantation by terrestrial laser scanner data as a link between exterior and interior wood properties, *SilviLaser 2010, University Freiburg, 79098 Freiburg.*

Revisiting the status of space-borne lidar missions for assessing structural and biophysical forest parameters in the context of sustainable management of Earth resources

S. Durrieu¹ and R. Nelson²

¹Cemagref, UMR TETIS, 500, rue J.F. Breton BP 5095, 34196 Montpellier Cedex 5, France,
sylvie.durrieu@teledetection.fr

²Code 614.4, Biospheric Sciences Branch, NASA/GSFC, Greenbelt, MD 20771 USA,
ross.f.nelson@nasa.gov

Abstract:

Assessing forest aboveground biomass at global scale is crucial to address the challenge of sustainable management of forest resources and to strengthen forest-based climate change mitigation. To achieve this goal relying on spaceborne lidar missions is acknowledged to be a highly relevant solution. However, if this is taken as a given from the measurement point of view, the premise that spaceborne observation is the most suitable solution to provide information for sustainable management of forest resources is worth discussing. In this paper we suggest to take a fresh look at measurement processes designed to support the monitoring of Earth resources. We discuss the sustainability of Earth observation from space considering (1) issues that call into question the assumption that Earth-orbiting platform will always be available to the civilian remote sensing community and (2) issues concerning environmental impacts of space activity on the Earth. This leads us to suggest some actions that could help to design future observation systems in a more sustainable way in order to strengthen the capacity of measurement processes to meet their stated functional goal, i.e. sustainable management of forest resources.

Keywords: space remote sensing, lidar, forest, biomass, sustainability, sustainable management

1. Introduction

Robust systems for measuring, assessing, and reporting key forest parameters, e.g. biomass, carbon, are needed to define adequate management practices and policies to address the challenge of sustainable management of forest resources and to strengthen forest-based climate change mitigation (Baker et al., 2010; Bernier and Schoene, 2009; Liu and Han, 2009; Thürig and Kaufmann, 2010). A spaceborne lidar that acquires samples of vegetation height and canopy closure measurements, used alone or in combination with optical and radar imagery, appears as the most promising way to estimate aboveground forest biomass and carbon at a global scale. Indeed such a solution combines beneficial measurement properties of spaceborne remote sensing and lidar technology. Spaceborne remote sensing greatly facilitates the acquisition of worldwide information consistent in both space and time. This information, supported by ground observations, is considered as the key to effective monitoring (DeFries et al., 2007; Fuller, 2006; Kleinn, 2002). The utility of lidar with respect to forest structure measurements and biomass estimation has been widely demonstrated (Lim et al., 2003; Næsset, 2004; Wulder et al., 2008) even in closed-canopy tropical areas supporting high biomass forests (>200 Mg ha⁻¹) (Kellner et al., 2009; Lefsky et al., 2005) where optical vegetation indices and volumetric radar measurements typically saturate (Castro et al., 2003). This led scientists to design and propose to space agencies, so far unsuccessfully, space lidar missions with forest measurement and monitoring as the primary scientific objectives (e.g. VCL, LVTM, Carbon-3D, DESDynI-Lidar, LEAF).

ICESat (2003-2009) was the first spaceborne lidar system designed to measure terrestrial surfaces. Despite the ice-centric design, several studies used ICESat/GLAS data to estimate forest structure and biomass at regional and global scales (Boudreau et al., 2008; Helmer et al., 2009; Lefsky et al., 2005; 2007; Lefsky, 2010; Nelson et al., 2009). However Nelson (2010) outlines some of the limitations associated with GLAS-based biomass estimates. ICESat-2, scheduled for launch in 2016, also includes among its secondary scientific objectives large-scale biomass and biomass change estimations (Nelson et al., 2010). But some simulations showed that the ground signal is expected to be lost at canopy closures exceeding ~95%, thus making calculation of canopy height impossible.

Even if the relevance of a spaceborne lidar mission designed to monitor global forest biomass is taken as a given from the measurement point of view, the premise that spaceborne observation is the most suitable solution to provide information for sustainable management of global forest resources should be discussed. Indeed the sustainability of Earth observation from space is not as obvious as it might seem.

The paper is structured to take a fresh look at measurement processes designed to support the monitoring of Earth resources and to promote discussion about the place of remote sensing within the context of sustainable management of these resources. Sustainability is defined as the capacity to endure and sustainable development as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs” (UN, 1987). Section 2 examines issues that call into question the basic assumption that Earth-orbiting platforms will always be available to the civilian remote sensing community. Section 3 focuses on the environmental impacts of space activity on the Earth and reports how these impacts affect sustainability. Section 4 suggests possible actions that might be taken to mitigate impacts associated with space missions by depending more on existing infrastructure to gather global observations.

2. Uncertainties about the future of Earth observation from space

2.1 Historical context, current state and future tendency of space activity

The development of space began in 1957 with the launch of Sputnik, the first artificial Earth-orbiting satellite. The total number of launches since 1957 exceeded 5000 during year 2009 (Figure 1) and the mean annual number of launches over the 10 past years has been slightly higher than 65 (McDowell, 2011).

The development of space has long been driven by the political and military aspirations of the two main actors of this sector, USA and Russia. With the end of the cold war in 1991, the rise of commercial interests in some applications – e.g. telecommunication and Earth observation - and the emergence of new space powers have led to a drastic transformation of the space sector. There are currently about 1000 active satellites, operated by 41 countries and several international consortiums (UCS, 2011). Figure 2 shows the distribution of satellites according to orbit classes and scientific/commercial disciplines. Of the 135 active Earth observation satellites, 120 are on LEO. The profound changes in the space sector led to a reduction in public investments that have weakened this sector (Pasco, 2003), at least as measured by launch activity. However, according to Pasco (2003), projects like the European Global Monitoring for Environment and Security (GMES) that bring space to society rather than the converse, could bolster this sector. Furthermore the development of satellite constellations and of micro-satellite technology, making space technology affordable for developing countries (Woellert et al., 2011), is assumed to result in an ongoing increase in the number of active satellites.

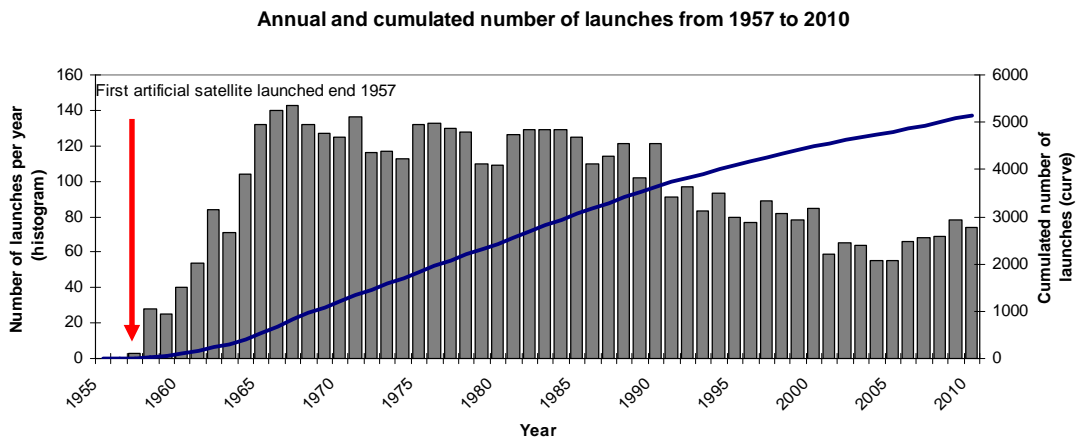
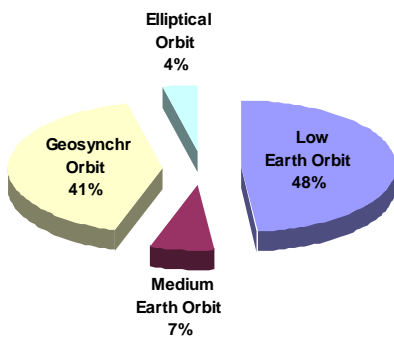


Figure 1: Evolution of the number of launches since 1957 estimated using the McDowell’s data base (McDowell, 2011). The histogram represents the number of annual launches (Y left axis) while the curve represents the cumulated number of launches since the launch of Sputnik (Y right axis).

(a) Distribution of active satellites according to orbit classes



(b) Distribution of active satellites according to the discipline in which the satellite is used

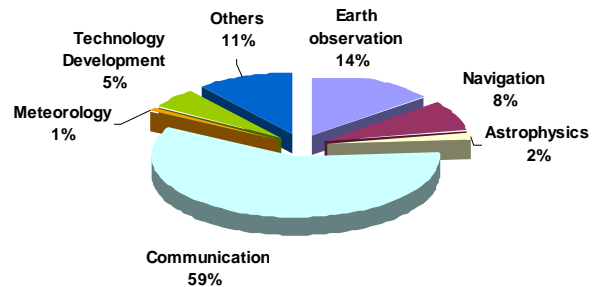


Figure 2: a) Distribution of the 957 current active satellites according to orbit class: Low Earth Orbit (LEO) refers to orbit with altitudes between 80 and 1,700 km; Medium Earth Orbit (MEO) to orbits between 1700 and 35,700 km ; Geosynchronous Orbits (GEO) are around 35,700 km; Elliptical orbits have a non constant altitude. b) Distribution of active satellites according to their use.

These figures result from analysis of the UCS database (UCS, 2011)

2.2 Space junk and its consequence on future activity

Space development has resulted in an increase in the amount of space debris, i.e. non-functional satellites, upper stages of launchers, functional debris (bolts, belts,...) and fragments originating from collisions, launcher upper stages, and spacecraft explosions, up to a point where orbital debris is currently a threat to spacecraft health and safety (Newman, 2010). The current number of catalogued objects, i.e. objects larger than 5 to 10 cm in Low Earth Orbit (LEO) and 30 to 100 cm at geostationary altitudes, that are tracked by the US Space Surveillance Network is higher than 13 000 and is increasing by several hundreds per year (Newman, 2010). Estimations of non-catalogued objects vary depending on the source; according to the CNES, there would be ~200 000 objects between 1 and 10 cm and ~35 millions between 0.1 and 1 cm (CNES, 2011).

The larger part of orbital debris population, ~40% of the total debris larger than 1 mm in size, resides in LEO (CNES, 2011). LEO space debris mitigation is a critical issue for space activity

sustainability. Up to now 4 accidental collisions have been identified (CNES, 2011) and 3 other are suspected (Flohrer et al., 2009). To prevent collisions with catalogued debris, alert systems for high-risk conjunction events have been developed by space agencies, permitting them to implement avoidance manoeuvres when necessary (Flohrer et al., 2009; Newman, 2010). Non-catalogued debris ranging from 1 to 10 cm can also generate very significant damage but the collision risk can only be studied statistically through analysis of impacts on dedicated experimental platforms or on launchers and large space debris that returns to Earth. Table 1 reports collision probabilities. A simulation model of LEO predicted that, even with no future launches, the critical point where the population of artificial space debris would grow at a faster rate than the natural decay rate could be reached in about 50 years (Bradley and Wein, 2009). The possibility that LEO could be made inaccessible for thousands of years by a chain reaction of debris collisions is underlined by several authors (Bradley and Wein, 2009; Weeden, 2011; Williamson, 2004).

Table 1 : Probability of collision over one year according to debris size for a satellite with a 20m² surface area at an orbital altitude similar to SPOT, i.e., 825 km (CNES, 2011).

Debris size	> 0.1 mm	> 1mm	> 1cm	> 10 cm
Probability of collision over 1 year	1	0.5	3×10^{-3}	2×10^{-4}

3. Space activity a source of pollution for the Earth environment

The risks linked to space activities that are most frequently discussed in the literature are on-orbit collision risks, which threaten the commercial exploitation of space, and risks to people on the ground during natural re-entries of debris, which is currently estimated to be lower than the risk associated with meteorite impact (CNES, 2011).

We discuss now a more seldom addressed topic: space activity as a source of pollution for the Earth environment. We focus on environmental impacts with respect to launch, life on-orbit, and end-of-life.

3.1 Launch and Orbit Insertion

The launch stage is responsible for two main kinds of pollution. The first one is the immediate return-to-Earth of the accelerator stage that separates from the launcher after fuel exhaustion and that is not systematically salvaged and seldom reused. The second one is related to propulsion system functioning. Most spacecraft depend on a rocket engine for propulsion. Approximate emissions for the main propellant types are given in table 2 and include, as for aircraft, greenhouse gases that directly add to radiative forcing and warming, as CO₂ and water vapour, and compounds that indirectly contribute to production or loss of green house gases such as ozone and methane (Ross et al., 2009). The amount of emitted gases is trivial compared to other sources. For example annual CO₂ emissions are estimated to be several kilotons compared to emissions of several hundred kilotons from aircraft, which is in turn between 2 and 3 % of the total emissions from all activities (Leary, 1999; Ross et al., 2009; Wilkerson et al., 2010). However, besides transient changes near the launch site that affect the lowermost troposphere, emissions may cause global lasting changes in the stratosphere where atmospheric circulation is characterized by an horizontal mixing of gases occurring faster than the vertical mixing (Ross et al., 2009).

Table 2: Approximate emissions for the four main propellant types (one solid and three liquid) given as mass fraction for each propellant. The total mass fraction exceeds unity because of the assumption that air mixed into the plume oxidizes CO and H₂ (source : Ross et al., 2011).

Propellant type	N ₂	CO ₂ +CO	H ₂ O + H ₂	ClO _x , HO _x , NO _x	HCl	Alumina soot
Solid (NH ₄ ClO ₄ /Al)	0.08	0.27	0.48	0.1	0.15	0.33
Cryogenic (LOX/H ₂ O)	-	-	1.24	0.02	-	-
Kerosene (LOX/RP-1)	-	0.88	0.30	0.02	-	0.05
Hypergolic (UDMH/N ₂ O ₄)	0.29	0.63	0.25	0.02	-	Trace

Local impacts of launch events are studied at the French Guiana Space Center (CSG). At each launch, 600 measurements are taken at several distances from the launch zone and include concentration measurements of hydrochloric acid, nitrogen dioxide, hydrazine and alumina. Analyses show that impacts are mainly localized nearby the launch area (<2.3 km) where high levels of HCl and Alumina concentrations are registered (see Table 3). Impacts were found to be low at intermediate distance (up to 8 km) and non-significant beyond. Impacts on water quality, vegetation and fauna are also monitored; up to now no significant negative impact has been noticed (CSG, 2011)

Table 3: Example of maximal concentrations of HCL and Alumina measured during an Ariane 5 launch (flight 185, 24-08-2008). Near field refers to a distance from launch site <2,4 km and far field from 2,4 to 24 km. Measures are compared to human toxicity thresholds (source: <http://www.ggm.drire.gouv.fr/>)

	Maximal near field concentration (mg/m ²)	Maximal far field concentration (mg/m ²)	Toxic limits defined for humans
Ion CL- (HCL)	5136.2	89.84 (measured at 4.350 Km)	90 mg/m ³ : irreversible effect after 30 mn exposure 700 mg/m ³ : lethal effect after 30 mn exposure
Alumina	94.68	3.49	Acceptable mean exposure value for workers = 10 mg/m ³ during 8h, 5 days/week

Part of the rocket combustion products are injected directly into the middle and upper stratosphere. The stratosphere is characterized by a low concentration in water vapour and includes the major part of the ozone layer. While climate response seems to be independent of where CO₂ emissions occur (<http://www.co2offsetresearch.org/aviation/DirectEmissions.html>), the increase in forcing due to water vapour emitted in the stratosphere is significant compared to a similar water vapour emission in troposphere (Leary, 1999). Furthermore emitted compounds contribute to ozone depletion in several ways. Some of them are highly reactive radicals - NO_x,HO_x,ClO_x - directly involved into catalytic cycles leading to an increase in the ozone removal rate while others contribute to increase the tropospheric radical reservoirs. As for emitted water vapour, it is the source gas for HO_x radicals and contributes to the formation of ice particles also responsible for ozone loss. Ozone loss from water vapour is highly nonlinear and difficult to predict (Ross et al., 2009). The ozone layer is protected by international agreements that limit the production of substances causing ozone depletion (i.e. the Montreal Protocol on Substances That Deplete the Ozone Layer) and Ross et al. (2009) demonstrated that if the Space Shuttle had met its original goal of weekly launches it would have been, alone, responsible for an ozone loss close to a quantity assumed by these authors to be the upper limit acceptable by the stratospheric protection community, even considering the unique value of space activity.

3.2 On-orbit life

Several compounds are also released in the upper atmospheric layers during the on-orbit lifetime of LEO satellites. First, the atmospheric drag in LEO causes orbital decay and the platform has to be repositioned occasionally. This is usually done using nozzle-based systems, most of them employing hydrazine as monopropellant. The highly exothermic catalytic decomposition of hydrazine produces jet of hot gas and thus thrust. The expelled gas is composed of ammonia (NH₃), dihydrogen (H₂) and nitrogen (N₂). Second, the presence of a diffuse atmosphere causes wearing effects on satellite platforms. Atomic oxygen, the predominant component in LEO atmosphere, is responsible for the degradation of thermal, mechanical, and optical properties of exposed materials (Liu et al., 2010). It interacts with hydrocarbon polymers (e.g., Kapton, Teflon, Mylar...) that are used to thermally insulate/protect parts of the satellite (Banks et al., 2011); important Kapton mass losses have already been observed attesting to the fact that volatile oxidation products have been released into the LEO domain.

3.3 End-of-life

The on-orbit lifetime of non-active satellites and other debris depends on the presence and density of the terrestrial atmosphere at a given altitude. Atmospheric drag slows down orbiting objects, making them return to Earth. During atmospheric re-entry objects are intensively heated and part of the material is sublimated. However large debris pieces can return to Earth. Most of them (> 70 %) will impact bodies of water (Johnson, 1999). As an example, during the Mir controlled re-entry, while the initial mass in orbit was around 140 tons, 30 tons of debris fell into the Pacific Ocean.

Another environmental threat from space activities comes from the use of nuclear reactors. Such reactors generate very high electrical power levels. Their use on military satellites drives the increased spatial resolving power of on-board radars and does away with the need for large solar sails. Decreasing the satellite cross-sectional area is paramount, making localization more difficult and lowering risks of hostile actions from anti-satellite systems (ASAT) (Finn, 1984). Nuclear generators also enabled missions such as Apollo Lunar surface experiments and interplanetary missions requiring to go where the sunlight intensity and the temperatures are low and the radiation belts very severe (Bennett, 2006). Since 1961, the USA and the former Soviet Union used nuclear energy to provide power for respectively 24 and 37 space systems (Bennett, 2006). Several nuclear-powered space vehicles are known to have fallen to Earth and were responsible for the release of radioactive elements either in atmosphere or on the Earth's surface (Finn, 1984), e.g. Cosmos-954 in 1978 or Transit satellite in 1964.

3.4 Impact of space activities on the Earth system

As a consequence of space activity in general and of LEO missions more specifically, debris, particles, and various chemical compounds are released within all the layers of the atmosphere, from the troposphere to the Exosphere (up to 10000 km). A portion of these elements falls back onto the Earth surface.

Our current atmosphere is the result of a long, progressive, and continuing evolution. The ocean-atmosphere envelope demonstrates non-linear dynamics, making relatively rapid changes in climatic patterns a likely feature of the future Earth. Forcing on the parameters that affect change (e.g., the greenhouse effect) may increase the speed and unpredictability of such changes (Adams, 2006). Historical and ongoing examples of the negative impacts associated with anthropogenically-induced atmospheric changes include acid rain, ozone depletion, and climate change. A unique characteristic of space activity is that it is the only human activity that releases elements in the upper atmospheric layers where concentrations of natural compounds are low. Consequently even the introduction of elements in small quantities can greatly affect atmospheric composition and chemistry; witness the effects of reactive radicals on the ozone

cycle (Ross et al., 2009). Impacts of space activity on complex processes occurring in upper atmospheric layers have been little studied until now and there is a clearly a lack of knowledge on the potential consequences of space activity on atmosphere composition and on radiative transfer. Environmental impacts and the amount of radioactivity that would reach the Earth surface in the case of the disintegration of a reactor core in the upper atmosphere during an accidental re-entry are also unknown (Finn, 1984). Based on current knowledge we can only say that, due to its special features, space activity brings with it a pollution capacity difficult to assess but presenting potentially high risks. Furthermore we only discussed in this section environmental impacts from launch to satellite end of life. With many products a large share of environmental impacts is not in the use of the product but in its production and transportation (Guinée et al., 2011). Therefore impacts associated with launcher, platform and instrument manufacturing and those impacts related to the functioning of ground segments should be also examined.

4. Toward sustainable Earth observation systems for forest monitoring

This last section discusses sustainable Earth observation networks or systems, i.e. systems involving measurement processes that are compatible with sustainability principles. In a first sub-section we will focus on elements concerning space activity in general while in a second sub-section we will consider actions that could be implemented when designing systems based on lidar technology for forest structure measurement.

4.1 Toward sustainable space activity

Two main problems hamper the development of space activity in a sustainable way: space debris, which threaten the activity itself, and potential negative impacts on both space and Earth environments. According to Williamson (2003) ethical and code policy for space should include protection of Earth orbit as a commercial and scientific resource by formalising debris mitigation measures. Space agencies, e.g., CNES, ESA, NASA, have already developed guidelines to mitigate space debris. At an international level, the United Nations Committee for the Peaceful Uses of Outer Space (UNCOPUOS) has already taken interest in space debris and in the use of nuclear energy in space. But, as a consultative body, it has no legislation power (Williamson, 2003; Williamson, 2004). Bradley and Wein (2009) demonstrated that achieving full compliance with the 25-year spacecraft deorbiting guidelines could maintain the lifetime risk from space debris at a sustainable level. However Taleb and al. (2009), who analysed common errors in risk management strategies, maintain that it is more efficient to reduce the impact of threat we cannot control rather than to focus on statistical predictions of low-probability high impact events. Following such advice would involve 1) doing all that is possible to reduce space junk by developing, for example, orbital debris removal operations as proposed by Weeden (2011), 2) making best efforts to mitigate future launch pollution and debris, keeping space as pristine as practically possible, and 3) thinking about alternatives to spaceborne solutions in case LEO becomes inaccessible.

We consider what could be done to protect the Earth environment from damages related to space activity, specifically addressing issues related to risk management and recycling. This discussion may require the reader to adopt non-traditional points of view. Concerning risk management, it would be necessary to shift from minimizing type-I error, i.e., rejecting the null hypothesis (or status-quo) when it is true, to minimizing type-II error, i.e., accepting the null hypothesis when it is false (Bergen et al., 2001). Usual scientific approaches tend to minimize type-I errors as a means to achieve high levels of confidence in the decision to throw out a null hypothesis and accept that some sort of change has occurred or that a “new” condition exists. When applied to environmental management, minimizing type I error means that we need to be near-certain that environmental or ecological damage has occurred (due to space activity in our

case) before we would accept the alternative hypothesis (Bergen et al., 2001). The net effect of relying on such an approach to determine action is that the damage would be done by the time the test tells us to act. With respect to prevention, mitigation, or remediation, this course of action is counterproductive.

Concerning recycling, it is worth noticing that, for the space activity sector, any object (satellite, piece of launcher...) is considered to be recycled when it has been destroyed during its re-entry in atmosphere or has fallen back on the earth surface. This is far from the notion of recycling in the context of sustainable development. The Earth's capacity to sequester human waste is limited (Adams, 2006) and recycling aims to reduce waste. Furthermore part of so-called "recycled" spacecraft is pollutants emitted into atmosphere or on the ground. To better grasp this issue research is needed to assess the behaviour of material released into the upper atmospheric layers by space activity. In addition to site impact studies realized on industrial sites, life cycle assessment (LCA) could be used to help quantify environmental impacts of Earth observation systems and to identify critical stages where these impacts might be mitigated. Environmental LCA are standardized (ISO 14040 and 14044) multi-criteria, quantitative approaches that permit to assess environmental impacts associated with all the stages of a product's life, from-cradle-to-grave (Guinée et al., 2011).

4.2 Designing systems based on lidar technology for forest structure measurement

Acquiring consistent and extensive spatial data for statistical reporting and mapping of forest resource is a technological issue required to address the challenge of sustainable management of forests. Integration of spaceborne and/or airborne lidar data with ground-based information is now widely advocated. However small steps, including some very simple actions, could help to optimize the way the several measurement processes are designed and used to provide information on forest structure and biomass.

Due to geopolitical constraints, satellite-based solutions seem to be currently the only way to acquire data globally. However, when designing space missions, actions could be done to make them more compatible with sustainability principles. A first step is to think about a mission as a contributor to space activity as a whole and therefore as a contributor to the problems discussed in this paper. Priority might be given to missions with long lifetimes in order to reduce the number of launches, space debris, and to lessen environmental damage. Priority might also be given to international missions in order to reduce mission duplication, e.g., multiple L-band radars, multiple 30m Landsat-like clones. Consideration should also be given to the use of existing space infrastructure. For instance, the International Space Station could host a lidar package to make measurements of the Earth's forests south of 51.6°N over the life of the ISS (current expectation ~10 year).

The relative contribution of each kind of measurement method, i.e. spaceborne, airborne acquisitions and field data, could be also cautiously compared considering a cost/benefit balance including economical, environmental, and societal aspects. For example, from a strictly economical point of view, space remote sensing is considered cost effective for end-users, especially given that much of the data are freely available. But this statement does not take into account all national funding invested into space activity. Consideration of the full compliment of expenditures might indicate that airborne solutions are more viable. Meanwhile, field data are often described as being costly and time consuming (Thuresson, 2002). But field campaigns, which are essential to develop models to assess forest parameters (calibration/validation steps), also provide information that cannot be acquired by other means, e.g. assessment of local biodiversity. Environmental impacts of complementary measurement processes could be compared using LCA in order to define the best way to combine them to reduce environmental impacts.

To end, we are going to suggest alternative solutions to current systems that might be worth exploring in the future. Traditional lidar acquisitions with topographic airborne systems are cost prohibitive, in particular for developing countries. Developing low cost, light systems designed specifically for forest resource assessment, e.g. PALS (Nelson et al., 2003) or LAUVAC systems (Allouis et al., 2010; Cuesta et al., 2010), might be one way to provide to developing countries a capacity to achieve accurate forest inventories. Thinking outside the box, can we embed light lidar systems on commercial jet aircraft? This could diminish both cost and environmental impacts. Figure 4 gives an idea of the distribution of the main commercial aircraft lines over the world and we can see that all terrestrial areas, except Antarctic, are covered, albeit at greatly varying flight line densities. Handling such data sets acquired on these commercial routes would certainly be challenging from a statistical standpoint.

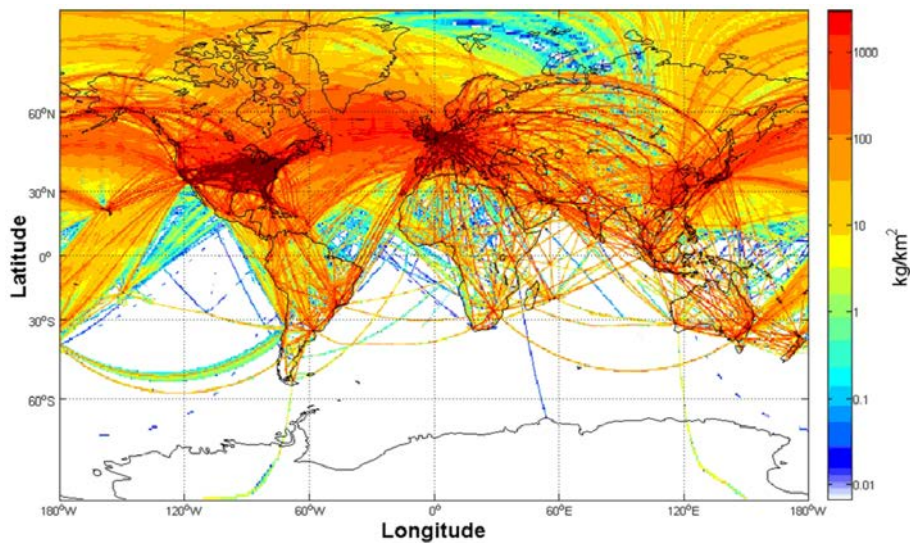


Figure 4: This figure presented in Wilkerson (2010) shows the total CO₂-C (kg/m²) emitted from commercial aviation for the year 2006 and indirectly gives an idea of commercial traffic density worldwide

5. Conclusion

There is a consensus to consider remote sensing from space as a technology of high value to provide worldwide measurements consistent in time and space. In this paper we proposed to change our view on measurements and to shift from a traditional view where measurement quality is assessed through metrology properties alone and where measurement processes are designed taking into account mainly economical and technological constraints to a more holistic view considering interactions between remote sensing observation systems and the Earth environment. This led us to suggest that regional and global forest measurements should be acquired in accordance with sustainability principles and with the willingness to reduce environmental impacts. Such goals strengthen the capacity of measurement processes to meet their stated functional goal, i.e., sustainable management of Earth resources. Currently, space remote sensing is mainly driven by economic and technical constraints, and is far from sustainability; there are notable uncertainties concerning the future of LEO accessibility and on the effects of pollution concomitant with space activity. We suggest some actions that could help to design future observation systems, in particular for measuring forest structure, in a more sustainable way. To that end, studies that cautiously assess environmental impacts of the several currently available measurement approaches should be considered.

References

- Adams, 2006. The Future of Sustainability: Re-thinking Environment and Development in the Twenty-first Century, IUCN, Gland, Switzerland, 18 p., http://cmsdata.iucn.org/downloads/iucn_future_of_sustainability.pdf, accessed May 2011
- Allouis, T., Durrieu, S., Cuesta, J., Chazette, P., Flamant, P.H. and Couteron, P., 2010. Assessment of tree and crown heights of a maritime pine forest at plot level using a fullwaveform ultraviolet lidar prototype. 2010 30th IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2010, Honolulu, HI, pp. 1382-1385.
- Baker, D.J., Richards, G., Grainger, A., Gonzalez, P., Brown, S., DeFries, R., Held, A., Kellndorfer, J., Ndunda, P., Ojima, D., Skrovseth, P.-E., Souza Jr, C. and Stolle, F., 2010. Achieving forest carbon information with higher certainty: A five-part plan. *Environmental Science & Policy*, 13(3): 249-260.
- Banks, B.A., Backus, J.A., Manno, M.V., Waters, D.L., Cameron, K.C. and de Groh, K.K., 2011. Prediction of Atomic Oxygen Erosion Yield for Spacecraft Polymers. *Journal of Spacecraft and Rockets*, 48(1): 14-22.
- Bennett, G.L., 2006. Space nuclear power: Opening the final frontier, Collection of Technical Papers - 4th International Energy Conversion Engineering Conference. 4th International Energy Conversion Engineering Conference, San Diego, CA, pp. 1433-1449.
- Bergen, S.D., Bolton, S.M. and L. Fridley, J., 2001. Design principles for ecological engineering. *Ecological Engineering*, 18(2): 201-210.
- Bernier, P. and Schoene, D., 2009. Adapting forests and their management to climate change: An overview. *Unasylva*, 60(231-232): 5-11.
- Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L. and Kimes, D.S., 2008. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. *Remote Sensing of Environment*, 112(10): 3876-3890.
- Bradley, A.M. and Wein, L.M., 2009. Space debris: Assessing risk and responsibility. *Advances in Space Research*, 43(9): 1372-1390.
- Castro, K.L., Sanchez-Azofeifa, G.A. and Rivard, B., 2003. Monitoring secondary tropical forests using space-borne data: Implications for Central America. *International Journal of Remote Sensing*, 24(9): 1853-1894.
- CNES, 2011. http://debris-spatiaux.cnes.fr/english/index_eng.html, accessed May 2011.
- CSG, 2011. <http://www.cnes-csg.fr/web/CNES-CSG-en/5678-weather-and-risk-management.php>, accessed May 2011.
- Cuesta, J., Chazette, P., Allouis, T., Flamant, P.H., Durrieu, S., Sanak, J., Genau, P., Guyon, D., Loustau, D. and Flamant, C., 2010. Observing the Forest Canopy with a New Ultra-Violet Compact Airborne Lidar. *Sensors*, 10(8): 7386-7403.
- DeFries, R., Achard, F., Brown, S., Herold, M., Murdiyarso, D., Schlamadinger, B. and de Souza Jr, C., 2007. Earth observations for estimating greenhouse gas emissions from deforestation in developing countries. *Environmental Science and Policy*, 10(4): 385-394.
- Finn, D.P., 1984. Nuclear satellites: why has the government downplayed their risks? *Environment*, 26(1): 14-20, 39.
- Flohrer, T., Krag, H. and Klinkrad, H., 2009. ESA's process for the identification and assessment of high-risk conjunction events. *Advances in Space Research*, 44(3): 355-363.
- Fuller, D.O., 2006. Tropical forest monitoring and remote sensing: A new era of transparency in forest governance? *Singapore Journal of Tropical Geography*, 27(1): 15-29.
- Guinée, J.B., Heijungs, R., Huppes, G., Zamagni, A., Masoni, P., Buonamici, R., Ekvall, T. and Rydberg, T., 2011. Life cycle assessment: Past, present, and future. *Environmental Science and Technology*, 45(1): 90-96.

- Helmer, E.H., Lefsky, M.A. and Roberts, D.A., 2009. Biomass accumulation rates of Amazonian secondary forest and biomass of old-growth forests from Landsat time series and the Geoscience Laser Altimeter System. *Journal of Applied Remote Sensing*, 3(1).
- Johnson, N.L., 1999. The reentry of large orbital debris. *Science and Technology Series*, 96: 285-293.
- Kellner, J.R., Clark, D.B. and Hofton, M.A., 2009. Canopy height and ground elevation in a mixed-land-use lowland Neotropical rain forest landscape. *Ecology*, 90(11): 3274.
- Kleinn, C., 2002. New technologies and methodologies for national forest inventories. *Unasylva*, 53(210): 10-15.
- Leary, N., 1999. Aviation and the Global Atmosphere-A Special Report of IPCC Working Groups I and III, IPCC Symposium, Tokyo, Japan, July 1, 1999, pp. p 7
- Lefsky, M.A., 2010. A global forest canopy height map from the moderate resolution imaging spectroradiometer and the geoscience laser altimeter system. *Geophysical Research Letters*, 37(15).
- Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Del Bom Espirito-Santo, F., Hunter, M.O. and de Oliveira Jr, R., 2005. Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, 32(22): 1-4.
- Lefsky, M.A., Keller, M., Pang, Y., de Camargo, P.B. and Hunter, M.O., 2007. Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing*, 1.
- Lim, K., Treitz, P., Wulder, M., St-Onge, B. and Flood, M., 2003. LiDAR remote sensing of forest structure. *Progress in Physical Geography*, 27(1): 88-106.
- Liu, G. and Han, S., 2009. Long-term forest management and timely transfer of carbon into wood products help reduce atmospheric carbon. *Ecological Modelling*, 220(13-14): 1719-1723.
- Liu, Y., Liu, X., Li, G. and Li, T., 2010. Numerical investigation on atomic oxygen undercutting of the protective polymer film using Monte Carlo approach. *Applied Surface Science*, 256(20): 6096-6106.
- McDowell, J., 2011. Jonathan's space report, N°640, April 2011, <http://www.planet4589.org/>, accessed 28 April 2011.
- Næsset, E., 2004. Practical large-scale forest stand inventory using small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, 19: 164-179.
- Nelson, R., Parker, G., Hom, M., 2003, A portable airborne laser system for forest inventory, *Photogrammetric Engineering and Remote Sensing*, 69 (3), pp. 267-273
- Nelson, R., 2010. Model effects on GLAS-based regional estimates of forest biomass and carbon. *International Journal of Remote Sensing*, 31(5): 1359-1372.
- Nelson, R., Boudreau, J., Gregoire, T.G., Margolis, H., Naesset, E., Gobakken, T. and Stahl, G., 2009. Estimating Quebec provincial forest resources using ICESat/GLAS. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 39(4): 862-881.
- Nelson, R., Neuenschwander, A., Ranson, K.J. and Cook, B., 2010. Current Characteristics of Two Space Lidars: ICESat-2 and DESDynI-Lidar, *Silvilaser 2010*, Freiburg, Germany, pp. 14-17.
- Newman, L.K., 2010. The NASA robotic conjunction assessment process: Overview and operational experiences. *Acta Astronautica*, 66(7-8): 1253-1261.
- Pasco, X., 2003. A new role for a new millennium? The changing nature of space activities. *Space Policy*, 19(1): 15-22.
- Ross, M., Toohey, D., Peinemann, M. and Ross, P., 2009. Limits on the space launch market related to stratospheric ozone depletion. *Astropolitics*, 7(1): 50-82.
- Taleb, N.N., Goldstein, D.G. and Spitznagel, M.W., 2009. The Six Mistakes Executives Make in Risk Management (October, pg 81, 2009). *Harvard Business Review*, 87(12): 123-128.

- Thuresson, T., 2002. Value of low-intensity field sampling in national forest inventories. *Unasylva*, 53(210): 19-23.
- Thürig, E. and Kaufmann, E., 2010. Increasing carbon sinks through forest management: a model-based comparison for Switzerland with its Eastern Plateau and Eastern Alps. *European Journal of Forest Research*: 1-10.
- UCS, 2011. Union of Concerned Scientists Satellite Database, Based on data from 1-31-11 database, <http://www.ucsusa.org/>, accessed 28 April 2011.
- UN, 1987. Report of the World Commission on Environment and Development: Our Common Future, UN Documents- Annex to document A/42/427- Chap.2, accessed 31 May 2011.
- Weeden, B., 2011. Overview of the legal and policy challenges of orbital debris removal. *Space Policy*, In Press, Corrected Proof.
- Wilkerson, J.T., Jacobson, M.Z., Malwitz, A., Balasubramanian, S., Wayson, R., Fleming, G., Naiman, A.D. and Lele, S.K., 2010. Analysis of emission data from global commercial aviation: 2004 and 2006. *Atmospheric Chemistry and Physics*, 10(13): 6391-6408.
- Williamson, M., 2003. Space ethics and protection of the space environment. *Space Policy*, 19(1): 47-52.
- Williamson, M., 2004. Protection of the space environment: The first small steps. *Advances in Space Research*, 34(11): 2338-2343.
- Woellert, K., Ehrenfreund, P., Ricco, A.J. and Hertzfeld, H., 2011. Cubesats: Cost-effective science and technology platforms for emerging and developing nations. *Advances in Space Research*, 47(4): 663-684.
- Wulder, M.A., White, J.C., Fournier, R.A., Luther, J.E. and Magnussen, S., 2008. Spatially Explicit Large Area Biomass Estimation: Three Approaches Using Forest Inventory and Remotely Sensed Imagery in a GIS Sensors, 8(1): 529-560.

Vegetation classification in the Swedish sub-arctic using a combination of optical satellite images and airborne laser scanner data

Karin Nordkvist*¹, Mattias Nyström¹, Heather Reese¹, Johan Holmgren¹ and Håkan Olsson¹

¹Swedish University of Agricultural Sciences, Department of Forest Resource Management, SE-90183 Umeå, Sweden.

Corresponding author. E-mail: karin.nordkvist@srh.slu.se

The aim of this pilot study was to investigate to which degree the accuracy of automated vegetation classification in the Swedish sub-arctic could be improved by combining optical satellite data with airborne laser scanner (ALS) data, compared to using satellite data only. This information is of interest in an ongoing discussion about the possible inclusion of the mountains in northern Sweden in the national laser scanning that started in 2009. A SPOT 4 scene and ALS data from an Optech ALTM Gemini scanner, both from 2010, were used in maximum likelihood classification. Data for training and validation was obtained from 279 plots with 20 m radius that were visited in field 2010. These plots were located near Abisko in northern Sweden (lat. 68° 23' N, long. 18° 53' E), on the north and south side of Lake Torne Träsk. A classification scheme with 7 classes based on the Swedish mountain vegetation map was used. Classification using only SPOT data gave an over-all accuracy of 75.6%, and the combination of SPOT data and ALS data increased the accuracy to 81.4%.

1. Introduction

Existing vegetation maps of the Swedish mountains were produced 30 years ago by manual air photo interpretation. There is currently no funding available for producing an updated version of the map with entirely manual methods.

Automated satellite image classification is widely used for large area land cover mapping, for example in the Swedish national version of the CORINE Land Cover data base (Hagner and Reese 2007); a state-wide land cover mapping of Wisconsin (Reese *et al.* 2002); the Land Cover Map of Great Britain (LCMGB) (Fuller *et al.* 1994); and MODIS Land Cover (Friedl *et al.* 2002), just to mention a few. In Sweden, images from the SPOT satellites are freely available from the SACCESS data base, which is updated yearly with a dataset covering the entire country during the vegetation period.

Airborne Laser Scanner (ALS) data has proved useful in mapping of certain vegetation types such as mires (Korpela *et al.* 2009). Several studies have shown the benefits of combining ALS data with different kinds of imagery, for example for estimations of forest variables (Erdody and Moskal 2010; Holmgren *et al.* 2008; Hyde *et al.* 2006), and for classification and mapping of forests (Dalponte *et al.* 2008; Hill and Thomson 2005; Ke *et al.* 2010), rangelands (Bork and Su 2007), and coastal and estuarine areas (Chust *et al.* 2008; Kempeneers *et al.* 2009). Starting in 2009, the Swedish National Land Survey (Lantmäteriet) is collecting laser scanner data for most of the country and it is presently discussed whether the mountain areas should be included in this national scanning. Although the main purpose is the production of a new national Digital Elevation Model (DEM), this data set could also be a resource for future vegetation mapping. In a previous study we used a combination of SPOT images and ALS data from the national laser scanning to classify vegetation in a boreal forest area in mid Sweden. Over-all accuracy was up to 16% higher than when using only SPOT data (Nordkvist *et al.* submitted). In this study, we have investigated if a similar method is useful for classification of

sub-arctic mountain vegetation in northernmost Sweden.

2. Material and methods

2.1 Study area

The 6 km × 23 km study area was located near Abisko in northern Sweden and centred around lat. 68° 23' N, long. 18° 53' E, altitude 340 to 1370 m a.s.l. (Figure 1). The vegetation on the south side of Lake Torne Träsk is predominantly heath above the tree line, and mountain birch forest (*Betula pubescens* ssp. *Czerepanovii*) below. The dominant vegetation on the north side is heath and a forest belt with a few large birch trees on tall herb meadows. The tree line is at approximately 650 m a.s.l. (Dahlberg *et al.* 2004).

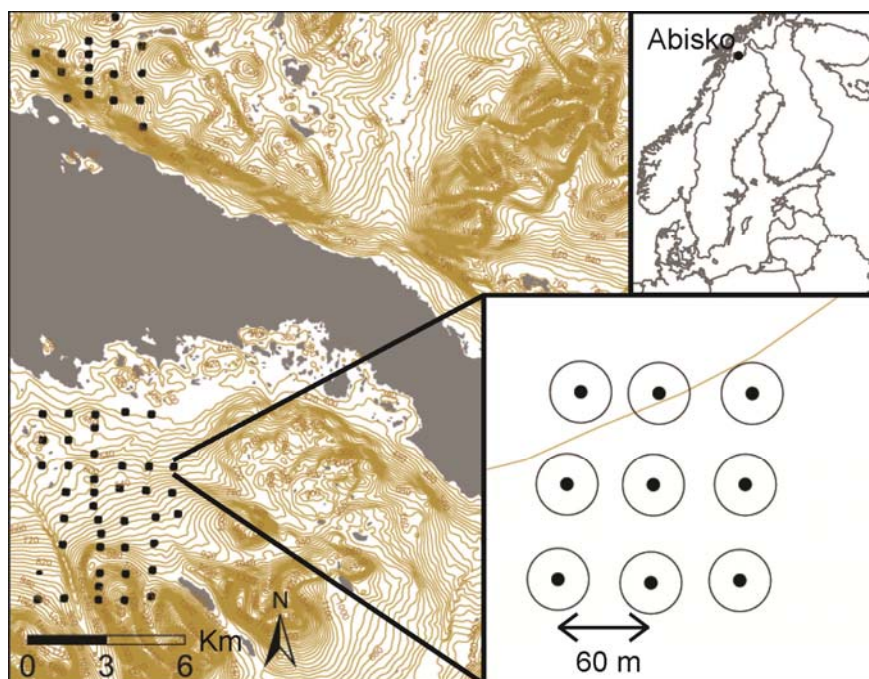


Figure 1. The map to the left shows the study area near Abisko in northern Sweden and the lower right image shows the distribution of plots within a cluster.

2.1.3 Vegetation reference data

Vegetation reference data was acquired at 532 circular sample plots with 20 m radius, grouped in clusters of 9 (Hedenås *et al.* submitted). The clusters were distributed on a systematic 1 km grid, with 60 m between the plots in each cluster. Plots were located on both the north and south side of Lake Torne Träsk (figure 1). The plots were visited in field during summer 2010 and classified according to the legend used in the Swedish mountain vegetation map. Plots where the dominating class covered less than 90% of the area, plots containing water and plots from classes with fewer than 8 members were excluded. The final training data set comprised 279 plots. After classification, extremely dry, dry and fresh heath were combined to form one single class called other heath.

Table 1:Vegetation classes used in the classification

Class	Short name	No. plots
-------	------------	-----------

Blocky areas and bedrock outcrops	Blocks and bedrock	15
Grass heath	Grass heath	26
Extremely dry heath	Other heath	8
Dry heath	Other heath	129
Fresh heath	Other heath	11
Alpine meadow with low herbs	Meadow	11
Moderate snow-bed vegetation	Snow-bed	12
Birch forest – heath type with mosses	Birch forest	42
Birch forest – sparsely grown	Sparse birch forest	25
Total		279

2.1.4 Remote sensing data

Optical satellite data used in the study was a multispectral SPOT 4 scene from August 20, 2010. The pixel size was $20 \text{ m} \times 20 \text{ m}$ and the image had been geo-rectified and co-registered to the Swedish grid system SWEREF 99 with an error of less than 0.5 pixels. ALS data was acquired on August 20, 2010 as a test flight for the national laser scanning. The scanner was an Optech ALTM Gemini scanner carried by a fixed wing aircraft. The flying altitude was approximately 1700 m above ground level and the average point density was $1.4 \text{ points m}^{-2}$.

2.1.5 Processing of remote sensing data

Laser returns were classified into ground and non-ground points using a progressive Triangulated Irregular Network (TIN) densification method implemented in the TerraScan software (Soininen 2011). A DEM was estimated by linear TIN interpolation with the laser returns classified as ground hits. In a first step, ground elevation was calculated in 1.0 m raster cells as the mean elevation of ground hits within each cell. Secondly, empty cells were assigned elevation values by TIN interpolation of the filled cells. The height value, h , of a laser return was computed as the height above the DEM. A height threshold of 0.5 m above the DEM was applied in order to separate canopy returns from returns of ground, stones and low vegetation. Laser data was extracted for the 20 m radius field plots, and several features were extracted based on the h distribution of the laser returns. The 10th, 20th, ..., 100th and 95th height percentiles (H_{10} , H_{20} , ...) were calculated for laser returns above the 0.5 m threshold. Mean and standard deviation of h were calculated both for returns above the threshold (\bar{h} and σ_h) and for all returns classified as non-ground (\bar{h}_{all} and $\sigma_{h,all}$). A vegetation ratio, V , was calculated as the number of returns above 0.5 m height divided by the total number of returns.

2.1.6 Classification

Image data from the SPOT scene was extracted using nearest neighbour sampling. Maximum likelihood classifications were done with the Minitab software, version 16.1.1. A first classification was done using the four SPOT bands as predictors, then different laser features were used as additional bands. The Normalized Difference Vegetation Index (NDVI) was also tested. It should be noted, however, that no reflectance calibration was done before calculating the NDVI. After classification, the three classes extremely dry, dry and fresh heath were combined into one single heath class, other heath, and the result was evaluated using cross validation.

3. Result

The three heath classes were very difficult to distinguish from each other, and were therefore combined to form one single heath class. Table 2 shows the classification results after the merging. The over-all accuracy when using only the four SPOT bands was 75.6%. The highest accuracy, 93.3%, was found for the blocks and bedrock class and the lowest, 28.0 %, for sparse birch forest. The error matrix (not shown) shows that the latter is classified as birch forest (heath type with mosses) in 60% of the cases. The highest over-all accuracy, 81.4%, was obtained when combining the four SPOT bands with mean height above ground of laser returns, \bar{h}_{all} . Accuracies for grass heath and other heath are improved and there is also an improvement in the two birch forest classes, although the confusion between the latter remains large. The combination of SPOT bands and $\sigma_{h,all}$ gives similar over-all accuracy as when using \bar{h}_{all} . Vegetation ratio V is the ALS feature that best discriminates the birch forest classes. Among different combinations of SPOT bands and NDVI, replacing SPOT band 3 (NIR) with NDVI gave the best result but did not give any improvement compared to using the four original bands. The use of height percentiles did not lead to any significant changes in the result.

Table 2: Over-all accuracy and classwise producer’s accuracies for different band combinations.

Bands†	Blocks, bedrock	Grass heath	Heath	Meadow	Moderate snow-bed	Birch forest	Sparse birch forest	Over-all
S1-S4	93.3	73.1	79.1	72.7	83.3	85.7	28.0	75.6
S1,S2,S4,NDVI	93.3	76.9	77.7	81.8	83.3	83.3	28.0	75.3
S1-S4, H_{10}	93.3	73.1	81.1	72.7	75.0	88.1	28.0	76.7
S1-S4, H_{95}	86.7	69.2	79.7	72.7	75.0	81.0	36.0	74.9
S1-S4, V	93.3	65.4	86.5	72.7	58.3	83.3	52.0	79.6
S1-S4, \bar{h}_{all}	93.3	80.8	85.1	72.7	83.3	88.1	44.0	81.4
S1-S4, $\sigma_{h,all}$	86.7	76.9	85.1	72.7	91.7	83.3	48.0	80.7

†S1-S4 are the four SPOT bands, NDVI is the normalized difference vegetation index, H_{10} and H_{95} are the 10th and 95th height percentiles, V is the vegetation ratio, \bar{h}_{all} is the mean height of laser returns and $\sigma_{h,all}$ is the standard deviation of h .

4. Discussion

The 5.8% improvement of classification accuracy when adding ALS data was small in the present study in the sub-arctic, compared to the results for a similar classification method applied in a boreal forest landscape reported by(Nordkvist *et al.* submitted). This could be explained by the relative height differences between classes being smaller above the tree line than in forest, and that some of the classes in the previous study were defined by height. A decision tree approach where the birch forest classes are separated from the others in a first step and then classified using vegetation ratio, which is correlated with canopy closure, might improve the result for these two classes. It should also be investigated to which degree texture features could improve discrimination between the remaining classes. Classification of laser returns into ground and non-ground is difficult in an undulating terrain with low vegetation.

This study is the first results from a two-year project about methods for large area vegetation mapping in the sub-arctic mountains of northern Sweden, using the combination of laser scanning and optical satellite data. Further studies will include additional use of the laser data and slope correction of the satellite data.

Acknowledgements

This ongoing study is financed by the Swedish National Space Board and the Swedish Environmental Protection Agency. The Swedish National Land Survey(Lantmäteriet) provided the laserdata.

References

- Bork, E. W. and Su, J. G., 2007. Integrating LIDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis. *Remote Sensing of Environment*, 111, 11-24.
- Chust, G., Galparsoro, I., Borja, A., Franco, J. and Uriarte, A., 2008. Coastal and estuarine habitat mapping, using LIDAR height and intensity and multi-spectral imagery. *Estuarine Coastal and Shelf Science*, 78, 633-643.
- Dahlberg, U., Berge, T. W., Petersson, H. and Vencatasawmy, C. P., 2004. Modelling biomass and leaf area index in a sub-arctic Scandinavian mountain area. *Scandinavian Journal of Forest Research*, 19, 60-71.
- Dalponte, M., Bruzzone, L. and Gianelle, D., 2008. Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 1416-1427.
- Erdody, T. L. and Moskal, L. M., 2010. Fusion of LiDAR and imagery for estimating forest canopy fuels. *Remote Sensing of Environment*, 114, 725-737.
- Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F. and Schaaf, C., 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83, 287-302.
- Fuller, R. M., Groom, G. B. and Jones, A. R., 1994. The Land Cover Map of Great Britain: an automated classification of Landsat Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing*, 60, 553-562.
- Hagner, O. and Reese, H., 2007. A method for calibrated maximum likelihood classification of forest types. *Remote Sensing of Environment*, 110, 438-444.
- Hedenås, H., Olsson, H., Jonasson, C., Bergstedt, J., Dahlberg, U. and Callaghan, T. V., submitted. Changes in tree growth, biomass and vegetation over a thirteen-year period in the Swedish sub-arctic.
- Hill, R. A. and Thomson, A. G., 2005. Mapping woodland species composition and structure using airborne spectral and LiDAR data. *International Journal of Remote Sensing*, 26, 3763-3779.
- Holmgren, J., Persson, Å. and Söderman, U., 2008. Species identification of individual trees by combining high resolution LIDAR data with multi-spectral images. *International Journal of Remote Sensing*, 29, 1537-1552.
- Hyde, P., Dubayah, R., Walker, W., Blair, J. B., Hofton, M. and Hunsaker, C., 2006. Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+ , Quickbird) synergy. *Remote Sensing of Environment*, 102, 63-73.
- Ke, Y., Quackenbush, L. J. and Im, J., 2010. Synergistic use of QuickBird multispectral imagery and LIDAR data for object-based forest species classification. *Remote Sensing of Environment*, 114, 1141-1154.
- Kempeneers, P., Deronde, B., Provoost, S. and Houthuys, R., 2009. Synergy of airborne digital camera and lidar data to map coastal dune vegetation. *Journal of Coastal Research*, SI 53, 73-82.
- Korpela, I., Koskinen, M., Vasander, H., Holopainen, M. and Minkinen, K., 2009. Airborne small-footprint discrete-return LiDAR data in the assessment of boreal mire surface patterns, vegetation, and habitats. *Forest Ecology and Management*, 258, 1549-1566.
- Nordkvist, K., Granholm, A.-H., Holmgren, J., Olsson, H. and Nilsson, M., submitted. Combining optical satellite data and airborne laser scanner data for vegetation

classificaton.

Reese, H. M., Lillesand, T. M., Nagel, D. E., Stewart, J. S., Goldmann, R. A., Simmons, T. E., Chipman, J. W. and Tassar, P. A., 2002. Statewide land cover derived from multiseasonal Landsat TM data - A retrospective of the WISCLAND project. *Remote Sensing of Environment*, 82, 224-237.

Soininen, A., 2011, TerraScan User's Guide. Terrasolid Ltd.

Lidar Data and Cooperative Research at Panther Creek, Oregon

James W. Flewelling¹ & George McFadden²

¹Seattle Biometrics, Seattle, WA, USA flewelling@seattlebiometrics.com

²Bureau of Land Management, Portland, OR, USA gmcfadde@blm.gov

Abstract

A 2,300 hectare forested watershed in the coastal mountain range of Oregon, USA is the subject of collaborative research with a principal objective of evaluating uses of lidar and other remotely sensed data for the development of detailed forest inventories. Panther Creek watershed (45° 18' N, 123° 21' W) is at an elevation of 100-700 m, about 57 km southeast of Portland. Major species are Douglas fir, western hemlock, western red cedar, grand fir, red alder and bigleaf maple; tree heights are up to 60 m. The Bureau of Land Management and other cooperators are using the watershed to test and develop methodology for detailed stand level forest inventories, the detailed mapping of soils and slope stability, and the assessment of other ecosystem functions.

Wall-to-wall discrete return lidar has been acquired under leaf-off conditions annually starting in 2007, and will continue through 2012. Leaf-on discrete return lidar was collected in 2007 and 2010 and will be collected in 2012. Surveys used Leica ALS50 Phase II or ALS60 lasers; pulse density is about 8 per m²; in 2010 selected areas received multiple passes, raising the density up to 50 pulses per m². Return intensities are being corrected for power output and camera-to-target distances. Full waveform lidar leaf-on data was acquired in 2010, as was 4-band color-infrared imagery using a Leica ADS40 camera. Also in 2010, hyperspectral data from a Hymap sensor was acquired. Eighty-four cadastral-surveyed 0.08 ha stem-mapped permanent plots were installed, mostly in 2009; measurement will be repeated after the 2012 growing season. Several other imagery sources are available.

A project goal is to compare and evaluate methodologies. All data are available to research groups wanting to participate. Data are well documented and organized, and include cut-outs of the remotely sensed data at each of the plot locations.

Key words: crown delineation, fused data, cooperative research.

1. Introduction

Wall-to-wall lidar is increasingly available for large forested areas in the western United States. Acquisition is often funded by public agencies, though private companies are also funding some acquisition campaigns. The State of Oregon has recently been acquiring lidar data at a rate of over 100,000 ha. per year (Oregon Department of Geology and Mineral Industries, 2009). Current campaigns have lidar densities averaging eight pulses per m². The expense of acquiring the data can often be justified without considering the use of the data for intensive forest inventory. However, public agencies including the U.S. Bureau of Land Management (BLM) are using the lidar data for stand delineation and estimation of per hectare attributes. The later are typically based on analysis techniques similar to those described by Næsset (2002).

There is a common interest in exploiting the lidar data and fused imagery to obtain detailed forest inventories. Of particular interest are inventories with better information on species distribution and the mapping of dominant and co-dominant trees. Additionally, inferences related to habitat, fire risk, down woody debris and numerous other landscape feature are of

interest. To research and evaluate suitable methods, the BLM, together with other governmental agencies and private parties have a cooperative research effort focusing on the Panther Creek watershed. The watershed has been the target of multiple efforts to collect remote sensing data, particularly airborne lidar data. In support of the inventory effort, stem-mapped forest plots have been measured, and will be remeasured after three years. The U.S. Environmental Protection Agency (EPA) is conducting an intensive soil survey of the area. Other data gathering efforts are ongoing, including the collection of terrestrial lidar data and meteorological data.

2. Panther Creek Watershed

The Panther Creek study area is a 2300 hectare forested watershed in the east side of the coastal mountain range of Oregon, USA. It is located 57 km southeast of Portland, Oregon, USA at 45° 18' N, 123° 21' W. The elevation ranges from 100 to 700 m. Annual precipitation is about 150 cm. The forests are mainly planted or natural stands of Douglas fir, with significant amounts of western hemlock, western red cedar, grand fir, red alder, bigleaf maple and several other species. Tree heights are up to 60 m. Management intensity throughout the watershed has been variable, with varying planting densities, and both thinned and unthinned regimes. The ecoregion classification is “Cascade mixed forest”(Bailey, 1995, p. 39-42).

3. Panther Creek Data

Much of the data for the Panther Creek project has already been collected. However, some data collection efforts, including soil sampling and meteorological data are ongoing, and other efforts are just getting started or are in the planning phase.

3.1 Lidar data

Six multi-point ALS data sets and one full waveform data set have already been collected (Table 1); two more lidar acquisitions are anticipated. All data acquisitions have similar requirements for off-nadir angle ($\pm 14^\circ$), flight line overlap (100%), and pulse density (≥ 8 pulses /m²). Ground densities average about 0.7 pulses per /m² in the leaf-off data sets, and 0.5 pulses per m² in the leaf-on data sets. Horizontal 1-sigma absolute accuracies for slopes < 20% are 30 cm or less. Line to line divergence is typically less than 10 cm. Additionally, the 2010.07.15 multi-point data set had some flight lines replicated so as to achieve very high pulse densities (≥ 50 pulses / m²) for portions of the study area. The Leica ALS60 unit recorded data from the automatic gain control; these can be used to improve normalized intensities.

Table 1: Present lidar data sets for Panther Creek.

Season	Date	Type	Instrument
2007 Leaf-on	2007.09.03	1-4 pts	Leica ALS50 - Phase II
2007 Leaf-off	2007.12.08	1-4 pts	Leica ALS50 - Phase II.
2008 Leaf-off	2009.03.28	1-4 pts	Leica ALS50 - Phase II
2009 Leaf-off	2010.03.29	1-4 pts	Leica ALS50 - Phase II
2010 Leaf-on	2010.07.15	1-4 pts	Leica ALS60
2010 Leaf-on	2010.07.15	Full wave	Leica ALS50 - Phase II + Digitizer
2010 Leaf-off	2011.04.17	1-4 pts	Leica ALS60

3.2 Other remotely sensed data

The other remotely-sensed data sets include imagery from public sources and acquisitions specific to Panther Creek (Table 2). Three-band (RGB) natural color imagery from the National Agricultural Imagery Program (NAIP) for 2005 and 2009 is available. Resolution is 1 m; horizontal accuracy is quoted as ± 6 m to true ground. Quickbird Imagery in 2008 includes Panchromatic (blue visible - NIR) with 0.6 m pixels, and color with 2.4 m pixels. High resolution 4-band imagery was acquired in March, 2010 with a Leica ADS-40 camera at 30 cm ground sample distance (GSD); one image with 30 cm pixels was ortho-rectified to a 20 m bare earth DEM and another was ortho rectified to a 0.3 m canopy surface model derived from the lidar data; an accuracy assessment in comparison with common linear features identified in the lidar data indicates spatial accuracy of about 40 cm. Hyperspectral data were collected with an Integrated Spectronics Hymap sensor. Four 32-band spectrometers were used with all data collected at a 3 m GSD. The hyperspectral data was collected without extensive ground survey pre-marks; in comparison with the lidar data there is a mean positional error of 6.3 m and a standard deviation of 4.5 m. Four flat black landscape targets (15 m \times 15 m) had been established under the canopy for establishing spectral signatures. GeoEye-1 satellite imagery was acquired in April, 2011 and July, 2011; the NIR image is 0.46 m GSD, and the RGB image is 1.9 m GSD; pixel size for the images is 0.5 m.

Table 2: Other remotely sensed data sets for Panther Creek.

Date	Type	Details
2005.06.28	RGB	NAIP, Leica ADS 40 (1 m pixel)
2008.06.29	NIR, RGB, pan	QuickBird (0.6 m panchromatic, 2.4 m color)
2009.06.23	RGB	NAIP, Leica ADS 40 (1 m pixel)
2010.03.19	NIR,RGB	Leica ADS 40 (30 cm GSD)
2010.07.30	Hyperspectral	Integrated Spectronics Hymap (3 m GSD)
2011.04.23	NIR, RGB, pan	GeoEye-1 (41 cm GSD pan, 165 cm GSD Color)
2011.07.06	NIR, RGB, pan	GeoEye-1 (41 cm GSD pan, 165 cm GSD Color)

3.3 Inventory sampling plan and tree measurements

An enhanced ability to create inventories is a project objective. However, the derivation of an inventory for Panther Creek is not a specific objective. Before starting to develop a sampling plan, the 2007 leaf-off lidar data and the 2005 NAIP imagery was used to delineate 144 stands. The primary ground data was to be a series of fixed-area stem-mapped plots. Sampling was to be limited to stands with cooperating public owners and cooperating large forestry landowners, excluding stands which had been recently clear-cut. There are 64 such stands with a combined area of 1451 ha. The 2007 leaf-off lidar data was used to derive several statistics for 30 m grid cells throughout the watershed; the grid cell statistics were aggregated to provide stand-level statistics. The statistics included HPCT90: the 90th percentile of the first returns above 2m; RH10: the ratio of HPCT10 to HPCT90; CC: the number of 1st returns greater than 2m divided by the total number of first returns. Additionally the NAIP photography was used to obtain a visual estimate of percentage of each stand in hardwood. The coniferous stands (> 80% conifer) were divided into three height groups with HPCT90 cut-points at 20.3 m and 31m. The tall stands (> 31 m) were further divided into three groups based on RH10, a measure related to depth of crown. The intermediate height stands (20.3 m - 31 m) were similarly sub-divided. The short stands were divided into three groups based on CC, a statistic assumed to be related to crown closure. Thus nine coniferous strata were defined.

A set of nine “modelling plots” was selected by choosing one stand in each of the nine strata, subjectively favoring the more extreme stratification statistics; within the nine selected stands,

sample locations were randomly selected; the first location which had lidar statistics similar to those for the average statistics for the whole stand was accepted.

A design based sample was also created for the coniferous plots. Within each strata one stand was selected with probability proportional to area. Within each selected stand, three plot locations were randomly selected. Plots selected this way can be used to make valid statistical inferences, where as the “modelling plots” would not be suitable for that purpose. Additionally two plots locations were randomly selected within two of the non-conifer stands (< 80% conifer); and two plot locations were arbitrarily selected in riparian area.

Yet another set of thirty-six plots were installed in conjunction with the soil survey; the location of these plots was not dependent upon the forest conditions. Three more plots were established by subjectively finding locations on the ground that appeared to be challenging due to the multi-story nature of the location; one more was located in a patch of red cedar, and two more were selected to capture specific edge condition between adjacent stands. All together, some eighty-four plot locations were selected.

Plot centers were established using GPS. Exact locations were subsequently determined within 0.25 m with a cadastral survey. Plots are circular, with all trees whose face is within 16.05 m of the center being measured. Effective plot area is slightly greater than 0.08 ha. Each tree was numbered with paint. Measurements included tree location (azimuth and direction), species, diameter (DBH), total height, and height to base of live crown. Live/dead status, broken-tops and extreme lean were noted. A subset of the trees were bored to obtain 5-year increment, sapwood area, and breast-height age. Tree measurements for seventy-eight plots were taken in the latter half of the 2009 growing season, or after the end of that growing season. Measurements for the final six plots were taken before the start of the 2011 growing season.

3.4 Other field data

In the summer of 2011, downed woody debris is being measured on all plots; large piece sizes anywhere on the full plot are being recorded; smaller pieces are measured on a smaller central portion of the full plot. Additionally hemispherical photos have been taken and ocular estimates of vegetation cover will be made. Terrestrial lidar data are being obtained at a majority of the plots during the 2011 growing season.

The EPA is conducting an intensive soil survey at Panther Creek in conjunction with efforts to estimate aboveground and below-ground carbon, characterize the hydro geomorphology, conduct a systematic terrain analysis, and establish a network of weather and soil monitoring stations within Panther Creek. The analysis at Panther Creek will guide the development of a soil-landscape-climate model for the coast mountain range. Completed work includes the description and sampling of thirty-five pedons, with soil samples being analyzed for physical and chemical properties.

3.5 Analyses

Analyses directed towards improved inventories will rely heavily upon already-developed techniques, with an objective of comparing the quality of results obtained from various combinations of data sources and methodologies. The first-named author of this paper is collaborating with others to apply per unit area techniques and individual tree crown delineation methods to lidar and imagery data sets. Per unit area analyses are being undertaken with the Fusion Software (McGaughey, 2009). Crown delineation using both lidar data and imagery will be undertaken with ITC Suite (Gougeon, 2010). Another of the initial efforts will be three-dimensional crown segmentation of the full waveform data using techniques developed by

Reitberger *et al.* (2010). These segmentation efforts will be followed by efforts to better co-locate the various data sources and the field data, and to cross identify stem mapped trees and delineated crown segments. Automated object-based fusion is a possibility.

Individual crown based methods are generally not expected to outperform area based methods of estimation for yield statistics (Peuhkurinen *et al.*, 2011). Indeed individual crown methods can be seriously biased unless coupled with rigorous sampling methodology and models which account for the inability to correctly isolate all crowns (Flewelling, 2008). However, the individual crown based methods have the potential to improve species prediction, particularly in mixed species stands. Similarly individual crown methods have the potential of more fully utilizing high density lidar data fused with high resolution imagery. The multiple sources of remotely sensed data provide an opportunity to test many combinations of methods and fused data sources.

4. Cooperative Research and Data Sharing

The Panther Creek cooperative research project currently involves over forty researchers and land managers representing federal, state and local agencies, landowners, a lidar provider, universities, and consultants. The project is loosely organized but does have a formal operating policy, a policy committee and a science committee.

A data sharing agreement exists, under which all of the airborne lidar data, other remotely sensed data, the field measurements for the stem mapped plots, and other selected data will be freely shared with any researchers who choose to participate. Additionally, much of the data will be made publicly available later; those data might not include some of the imagery which was licensed to the BLM but is not owned by the BLM. Most of the data relevant to forest inventory, including the stem-mapped plot data, the airborne lidar data and other remotely sensed data are available for sharing now. The tree data will subsequently be enhanced by “growing” the trees forward or backward to the dates of the lidar acquisitions. The data sharing mechanism is a password-protected web site, supplemented by the mailing of portable disks.

Persons or organizations wishing to participate in the data sharing are invited to contact the authors of this paper. Participants will be encouraged - but not required - to share a statement of their research objectives; and hopefully to share various results. Research on crown segmentation, object-based fusion, and species imputation is especially encouraged.

Acknowledgements

The financial support of the U.S. Forest Service through the Agenda 2020 Technology Alliance is gratefully acknowledged (FS Agreement 11-JV-11261989-036). Matt Boyd and Russ Faux of Watershed Sciences, Inc. have provided or helped acquire most of the remotely sensed data. David Marshall of Weyerhaeuser Company helped to develop all sampling plan and has been doing quality control on the field data; Connie Harrington of the U.S. Forest Service ensured that the sampling activities were contracted and performed in a timely manner. Steve Reutebuch of the U.S. Forest Service is acting as a project advisor. Francois Gougeon of the Canadian Forest Service, Hans-Erik Andersen of the U.S. Forest Service and Peter Krzystek of the University of Applied Sciences (München) have all offered to assist with various parts of the analysis.

References

- Bailey, R.G., 1995. *Description of the ecoregions of the United States*. USDA Forest Service. Misc. Pub. 1391, 108 p. + map.
- Flewelling, J.W. 2008. Probability models for individually segmented tree crown images in a sampling context. In Hill R.A., Rosette, J. and Suárez, J. (Eds). *Proceedings of the SilviLaser 2008 Conference*, Edinburgh, UK.
http://geography.swan.ac.uk/silvilaser/papers/oral_papers/Algorithm%20&%20Techniques%20Development/Flewelling.pdf [accessed 30 August 2011].
- Gougeon, F., 2010. *The ITC Suite Manual: A Semi-Automatic Individual Tree Crown (ITC) Approach to Forest Inventories*. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC. 92 p.
<http://cfs.nrcan.gc.ca/pubwarehouse/pdfs/31918.pdf> [Accessed 30 August, 2011]
- McGaughey, R. J. 2009. *FUSION/LDV: Software for lidar data analysis and visualization*. U.S. Forest Service. <http://www.fs.fed.us/eng/rsac/fusion/> [Accessed 30 August, 2011]
- Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80, 88-99.
- Oregon Department of Geology and Mineral Industries, 2009. Lidar Data Acquisition, Uses, & Partners. *Fact Sheet DOGAMI 01-23-09*. 2p.
<http://www.oregongeology.org/pubs/fs/lidar-leg-fact-sheet.pdf>
- Peuhkurinen, J., L. Mahtätalo, and M. Maltamo, 2011. Comparing individual tree detection and the area-based statistical approach for the retrieval of forest stand characteristics using airborne laser scanning in Scots pine stands. *Can. J. For. Res.* 41:583-598.
- Reitberger, J., M. Heurich, and P. Krzystek, 2010. Estimation of stem volume by using 3D segments derived from full waveform LiDAR data. *2010 Silvilaser proceedings*. Article 118 on distributed USB drive, 12 p.

LiDAR Estimation of Mean Dominant Height, Quadratic Mean Canopy Height and Stem Density in Native Sclerophyll Forests

Yadav P. Kandel¹, Julian C. Fox¹, Stefan K. Arndt¹ & Stephen J. Livesley¹

¹Department of Forest & Ecosystem Science, The University of Melbourne
e-mail: ykandel@student.unimelb.edu.au

Abstract

Plot level mean dominant height and quadratic mean canopy height were estimated quite accurately using the LiDAR data, and a new method for estimating the number of trees per hectare from LiDAR data was developed and used to predict the stem density for sample plots from two different types of native sclerophyll forests. R^2 of the regression model for the mean dominant height was 87.09 % for the Central Highlands Ash Regrowth (CHAR) and 92.1 % for the Black Range Mixed Species (BRMS) forest. Similarly, R^2 of the regression model for the quadratic mean canopy height was 48.4 % for the CHAR and 92.7 % for the BRMS forest. The number of trees were predicted with mean prediction error of - 64.12 trees per hectare for calibration plots and 105.29 trees per hectare for validation plots in CHAR forest which is a wet sclerophyll forest. In the BRMS forest which represents a dry sclerophyll forests, prediction error for number of trees was 79.99 trees per hectare for calibration plots and 4.96 trees per hectare for validation plots.

Key words: LiDAR, mean dominant height, quadratic mean canopy height, stem density.

1. Introduction

A forest stand is defined as a group of trees occupying a given area, which shares some characteristics such as species composition, size, or age (Newton, 2007). Oliver and Larson (1990) defined forest stand structure as “ the physical and temporal distribution of trees in a stand” which include the distribution of species, vertical and horizontal spatial patterns, size of trees or tree parts, tree age or combinations of these. Forests are three dimensional systems whose structure are a product of forest dynamics and biophysical processes and can be considered as a guide for biodiversity and ecosystem function (Spies, 1998).

A single value that represents the height of a stand is useful in describing the structure of a forest and can be used to estimate the stand volume or biomass using the volume or biomass functions developed for a particular forest stand. There are different ways of estimating the stand height from the heights of individual trees. Average height of trees, quadratic mean canopy height (QMCH) of trees, Lorey's mean height and mean dominant height (MDH) are some ways of expressing stand height. Stand density or number of trees per unit area is an important structural attribute which is mainly controlled by the size of the trees as they grow close together (Zeide, 1995). As size of trees increases, they use more resources as well as growing space. Trees in stands compete for resources and if resources are not sufficient for all trees and other components, some trees die and the number of trees per unit area decreases (Pretzsch and Biber, 2005). This trade off between the size and the number of trees suggests that the product of these quantities is subject to minor changes through stand development and that there may exist a combination of size and number of trees that is constant during the development of forest stands (Zeide, 1995). The relationship between the

size and number of trees is considered as a foundation of ecology (Zeide, 1987) and vital to the forest management (Vanclay, 2009; Zeide, 2005; West, 1982) as these relationships can be used to estimate stand density and stocking, optimal thinning intensity as well as to determine the scale of disturbance and self -thinning and other processes and functions of forest ecosystems (Zeide, 1995). There are various ways to relate size and number of trees in a forest stand.

Light Detection And Ranging (LiDAR) is relatively a new remote sensing technology that has advanced rapidly with advanced GPS and inertial navigation system and is capable of providing three dimensional structural information of global vegetation. LiDAR has been used to predict canopy height information, basal area and stem density as well as other structural information for different types of forest ecosystems (Dubayah et al., 2010; Duncanson et al., 2010; Chen et al., 2007; Coops et al., 2007; Andersen et al., 2006; Anderson et al., 2006; Bortolot and Wyne, 2005; Lefsky et al., 1999). Application of LiDAR technology, however, has not been fully tested for native sclerophyll forest of Central Highlands of Victoria, Australia which have among the largest forest biomass stocks in the world (Keith et al., 2009). The main objective of this study was to estimate the mean dominant height, quadratic mean canopy height and the number of trees per hectare using LiDAR data in two different types of native sclerophyll forests.

2. Materials and Methods

2.1 Study areas

This study concentrates on two different types of forest in the Central Highlands of Victoria which will be henceforth referred to as Central Highlands Ash Regrowth (CHAR) and Black Range Mixed Species (BRMS) forests. Both of these forests are part of Central Forest Management Area, one of the 14 forest management areas of Victoria. The Central Highlands of Victoria lies about 120 km north-east of the city of Melbourne and includes the foothills and mountains of the Great Dividing Range. The region covers approximately half a degree of latitude and one degree of longitude ($37^{\circ} 20' - 37^{\circ} 55' S$ and $145^{\circ} 30' - 146^{\circ} 20' E$) and covers an area of approximately 400, 000 ha. The elevation of the area ranges from 400 m to 1200 m above sea level. The region experiences mild, humid winters with occasional periods of snow with mean annual precipitation ranging from about 1200 to 2300 mm. Mean annual temperature of the area ranges from 7.8 deg to 13.4 deg

2.2 Field inventory data

Field inventory data for sampling plots of Central Highlands Ash Regrowth (CHAR) and Black Range Mix Species (BRMS) forests for the study area have been provided by the Victorian Department of Sustainability and Environment (DSE). Forest inventory data for 99 sampling plots for CHAR forest and 34 plots for BRMS forest were made available. All sampling plots in CHAR forest are of 0.04 ha (20 m × 20 m), while for BRMS there are 20 sampling plots of 0.04 hectares and 15 sampling plots of 0.25 hectares (50 m × 50 m). The location of each plot was recorded using differential GPS. In CHAR forest, diameter at breast height over bark (DBHOB) for every individual tree having diameter greater than 10 cm were recorded. In BRMS forest the DBHOB of all trees were recorded. Various plot level attributes were estimated from the inventory data provided by the DSE. Only trees having greater than 10 cm of diameter were used in this study.

2.2.1 Mean dominant height and quadratic mean canopy heights of the plots

The height of individual trees was not measured in the field, instead, mean dominant height and mean dominant diameter of the plots were provided. We used a height functions provided by DSE to estimate the height of individual trees from the mean dominant diameter and mean dominant height of the plots, together with the diameter of individual trees (Equation 1 for CHAR and Equation 2 for BRMS).

$$H = 10^{[\text{Log}_{10}(\text{MDH} \cdot 1.33) + (\text{MDD}/\text{DBHOB}) \cdot \{\text{Log}_{10}(\text{MDH}) - \text{Log}_{10}(\text{MDH} \cdot 1.33)\}]} \quad (1)$$

$$H = 10^{[\text{Log}_{10}(\text{MDH} \cdot 1.14394 + 5.36888) + (-1.5607 - 0.08073 \cdot \text{MDD}/\text{DBHOB})]} \quad (2)$$

where, H is height of trees in meter and

MDD = Mean Dominant Diameter (cm) (Average DBHOB of the 62 largest trees per ha)

MDH = Mean Dominant Height (m) (Average Height of the 62 largest DBHOB trees per ha)

$DBHOB$ = Diameter Breast Height (cm) for the subject tree

The DSE provided the mean dominant height for the plots which was estimated by taking the average height of 66 of the largest DBHOB trees per hectare. Quadratic mean canopy height (QMCH) is an index that can be used to describe the vertical structure of the canopy. This was developed by Lefsky et al. (1999b) and is defined as:

$$QMCH = \sqrt{\sum_i^{\text{max height}} CHP(i) \times i^2} \quad (3)$$

where, $CHP(i)$ is the fraction of total foliage at height at i and known as canopy height profile.

By comparing this equation with the allometric relationship between height and diameter for angiosperm "champion" trees developed and reported by Niklas (1994), Lefsky et al. (1999) stated that "the canopy height profile is being weighted by a factor that is proportional to the diameter required to support it, and conversely its average is transformed, by the square root to a variable that is proportional to height". This suggests QMCH represents the height of the tree with the average diameter and can be used in a similar way that quadratic mean diameter has been used in forestry (Lefsky, 1997). This implies that if average diameter of trees in plots is determined and converted to a height of tree having that diameter by using allometric equations between height and diameter of trees, it represents the quadratic mean canopy height of the plots. It was found that the quadratic mean canopy heights for plots determined using this approach almost have a one to one relationship with the quadratic mean height of trees of the plots. Therefore, in this study, quadratic mean height of trees in plots is considered as the quadratic mean canopy height of the plots and this was estimated directly from the heights of all individual trees in the sampling plots using the formula:

$$QMCH = \sqrt{\frac{\sum H^2}{n}} \quad (4)$$

where, n is the number of trees in the plots, H is the height of individual trees in meter and $QMCH$ is the quadratic mean canopy height for the plots in meter.

2.2.3 LiDAR data for the sampling plots

LiDAR data of the entire area of Central Highlands were provided by DSE which were acquired by AAMHatch Pty Ltd on behalf of DSE from a fixed wing aircraft flying at an approximate height of 1300 m between November 19th 2007 and January 10th 2008. The LiDAR system used in data acquisition was Optech ALTM3100EA having a 0.26 m laser footprint size. The data were taken with 25 % side overlap and swath width of 945 m. Ground and non-ground point clouds (LiDAR data) of the area were provided in 2 km × 2 km tiles separately. From these data tiles, LiDAR data for the sampling plots were extracted by using image processing software (ENVI, ITT Visual Information Solutions) and Lastools (free software) and the exact GPS coordinates of the sampling plots provided by the DSE.

2.2.4 LiDAR metrics

LiDAR data for sampling plots were processed using the software TiFFS (Toolbox for Lidar Data Filtering and Forest Studies) developed from Globalidar to generate digital elevation models (DEM) and canopy height models (CHM) and to derive various plot level LiDAR metrics such as mean heights, quadratic mean heights, percentile heights and standard deviation of all point heights of LiDAR data .

The quadratic mean height (QMH) of LiDAR points is given by:

$$QMH = \sqrt{\sum_i^{\max height} i^2 \times CHP(i)} \quad (5)$$

where, $CHP(i)$ is the fraction of laser points at height i .

2.2.5 LiDAR estimation of MDH, QMCH and stem density

Sample plots were divided into calibration and validation plots before developing the predictive models from regression analysis. 75 sample plots were randomly selected for calibration and the remaining 18 plots were used as the validation plots in CHAR. In BRMS, 20 plots were selected randomly for calibration and 14 plots were left for the validation purpose. Stepwise regression techniques as well as the best subset regression were used to find out the most suitable LiDAR metric as a predictor of the mean dominant height and quadratic mean canopy height using statistical software Minitab 16. Finally, the field measured/estimated mean dominant height and quadratic mean canopy height for the plots were regressed on the selected LiDAR metric to develop the predictive models.

Number of trees in sampling plots was estimated from LiDAR data in two steps. First, LiDAR estimated quadratic mean canopy heights of the sampling plots were converted to the average diameter of trees in sampling plots (quadratic mean height of trees represents the height of tree with average diameter) using the relationship between diameter and height of the trees developed from the regression analysis of the height and diameter data (Equation 6 for CHAR and Equation 7 for BRMS forests).

$$DBH = 0.0053 * H^{1.163} \quad (6)$$

$$DBH = 0.0017 * H^{1.584} \quad (7)$$

where, DBH is diameter at breast height and H is height of trees in meters.

The LiDAR derived average diameter of the plots was then used to estimate the number of trees per hectare in sampling plots using the relationship between the number of trees and average diameter developed for CHAR (Equation 8) and BRMS (Equation 9) as suggested by Zeide, (1995) which is the slight modification of Reineke's equation.

$$Ln(N) = 4.82 - 1.30 * Ln(DBH_{av}) \quad (8)$$

$$Ln(N) = 4.60 - 1.66 * Ln(DBH_{av}) \quad (9)$$

where, DBH_{av} is average diameter of trees in meters and N is the number of trees per hectare in the plots, Ln is natural log.

3. Results

Different structural attributes estimated from field inventory data show that there are 28 more trees per hectare in CHAR forest compared to the BRMS. Mean value of other attributes except mean dominant height are very close in both types of forests (Table 1). There was a very strong allometric relationship between diameter and height of trees in both types of forests. R^2 value of the regression equation for CHAR (Equation 6) was about 0.8 and for BRMS (Equation 7) 0.93 (Table 2). The relationship between number of trees per hectare and average diameter of trees in CHAR (Equation 8) was weaker ($R^2 = 0.52$). The relationship in BRMS (Equation 9), however, was very strong ($R^2 = 0.97$). The relationships are highly significant for both types of forests (p value < 0.0001) (Table 2).

Table 1. Different structural attributes for CHAR and BRMS forest

Structural attributes	CHAR	BRMS
Number of trees/ha	819	791
Mean dominant height (m)	39.8	33.6
Average height of trees (m)	26.6	28.2
Quadratic mean canopy height (m)	28.2	28.6
Mean dominant diameter (m)	0.53	0.55
Quadratic mean square diameter (m)	0.35	0.37

The regression model for the MDH for CHAR forest with 100th percentile height of LiDAR points as the predictor (Equation 10) had an R^2 of 87.1 % showing a very strong correlation between the field measured MDH and the 100th percentile height of the LiDAR points which is highly significant at 95 % confidence level (Table 2). Similarly, there was a very strong correlation between the field measured MDH and the 80th percentile height of LiDAR points for the BRMS forest as the R^2 of the predictive model (Equation 11) was 92.1 %.

$$MDH_F = -0.41 + 0.941 * LH_{100pc} \quad (10)$$

$$MDH_F = -2.27 + 1.17 * LH_{80pc} \quad (11)$$

where, MDH_F is the field-measured mean dominant height, LH_{100pc} and LH_{80pc} are 100th and 80th percentile height of LiDAR points in meters.

Regression analysis of field estimated quadratic mean canopy heights and the 70th percentile heights for the sample plots show that there was a good correlation between these two variables in CHAR as the R^2 value of the regression model (Equation 12) was 0.48 and the correlation between these variables in BRMS is very strong as the regression model (Equation 13) has an R^2 value of 0.93 (Table 2).

$$QMCH_F = 11.8 + 0.478 * LH_{70pc} \quad (12)$$

$$QMCH_F = -7.59 + 1.22 * LH_{70pc} \quad (13)$$

where, $QMCH_F$ is field estimated quadratic mean height and LH_{70pc} is 70th percentile height of LiDAR points in meter.

Scatterplots of the field estimated versus the LiDAR predicted values of MDH and QMCH for the calibration and the validation plots of the CHAR forest (Figure 1) and BRMS forest (Figure 2) show that the regression models were well fitted to the calibration data and they were also able to predict the MDH and QMCH quite accurately for new observations as well.

Table 2. Regression statistics for different regression models for CHAR and BRMS forest

Regression model	Regression statistics				
	R^2	RMSE (RMSE%)	MAE (MAE%)	p value	R^2 (Pred)
Equation 6	0.80	0.24 (15.2 %)	0.19 (11.8 %)	< 0.0001	0.79
Equation 7	0.93	0.15 (13.6 %)	0.12 (10.9 %)	< 0.0001	0.92
Equation 8	0.52	0.41 (6.3 %)	0.32(4.9 %)	< 0.0001	0.50
Equation 9	0.97	0.13 (2.1 %)	0.09 (1.5 %)	< 0.0001	0.96
Equation 10	0.87	3.8 (9.5 %)	2.6 (6.5 %)	< 0.0001	0.86
Equation 11	0.92	1.9 (6.0 %)	1.5 (4.6 %)	< 0.0001	0.90
Equation 12	0.48	4.9 (17.9 %)	3.5 (12.6 %)	< 0.0001	0.45
Equation 13	0.93	1.9 (7.4 %)	1.3 (4.9 %)	< 0.0001	0.91

When the method developed in this study was used to predict stem density, prediction bias produced for calibration plots in CHAR was - 64.12 trees per hectare (-7.6 %) which was very close to the prediction bias of stem density for calibration plots in BRMS (79.99 trees per hectare, 8.59 %) in terms of absolute error (about 8%). However, stem density in CHAR was underestimated whereas it was overestimated in BRMS. Prediction bias for validation plots was positive for both types of forests and it was significantly higher in CHAR (105.29 trees per hectare, 14.4 %) compared to that in BRMS (4.96 trees per hectare, 0.86 %) (Table 3).

Table 3. Prediction statistics for LiDAR predicted number of trees per hectare from LiDAR derived average diameter in CHAR and BRMS plots.

Prediction statistics	Number of trees per hectare			
	CHAR Forest		BRMS Forest	
	Calibration plots	Validation plots	Calibration plots	Validation plots
Mean predicted value	776.5	838.2	1010.49	584.39
Mean prediction bias	- 64.1 (7.6 %)	- 105.3 (14.4 %)	79.9 (8.6 %)	4.9 (0.86 %)
MSEP	165163.4	181133.6	87179.21	19442.89
The variance of the prediction error	101117.0	189365.2	85031.53	20911.99
Standard deviation of prediction error	317.9	435.2	291.60	144.60
Standard error of the estimated mean bias	36.7	99.8	65.20	32.32

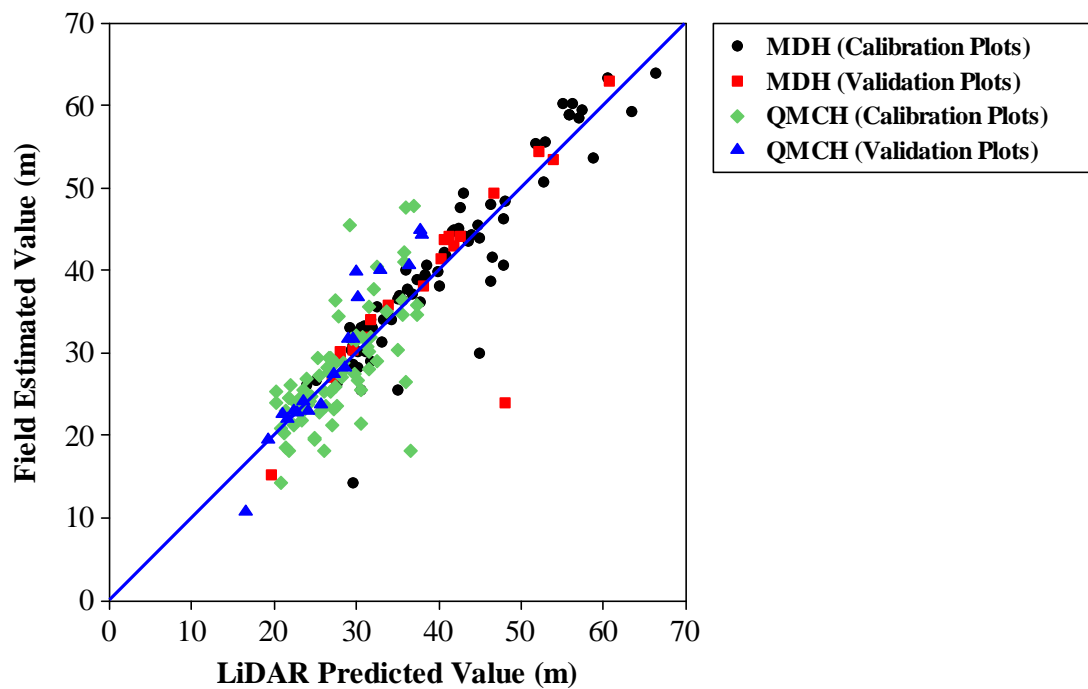


Figure 1. Scatterplot of the field estimated versus LiDAR predicted values for the MDH and QMCH for the calibration and validation plots of the CHAR forest

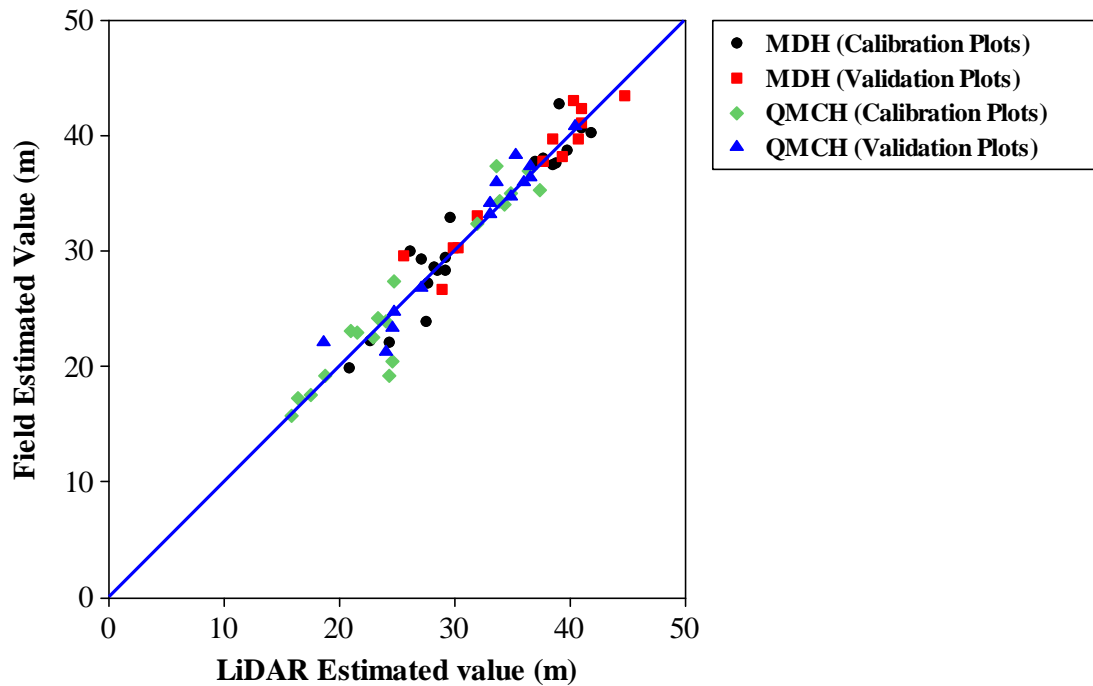


Figure 2. Scatterplot of the field estimated versus LiDAR predicted values for the MDH and QMCH for the calibration and validation plots of the BRMS forest

4. Discussion

Measure of stand height is useful in estimating stand volume or biomass from a stand volume or biomass function and to determine site index. There are different ways of measuring stand height. The arithmetic mean of the height of all trees in the stand is a useful measure of stand height in even-aged stands. However, it is not suitable for uneven aged stands and mixed species multi layered forest stands. Mean dominant height and Lorey's mean height are also used to measure the stand height both of these are influenced by the height of the few large trees. Quadratic mean canopy height (QMCH) is a new index developed by Lefsky et al. (1999b) which is more appropriate for the measure of the stand height in uneven-aged stands and mixed species multi layered forest stands. QMCH represents the height of the tree with the average diameter in the stand and therefore can be used to derive the average diameter of trees in stand from the allometric relationship between the height and the diameter of trees. This can also be used to develop more accurate stand volume or biomass functions as Lefsky (1997) found that QMCH of a stand had the highest correlation coefficient with the stand's basal area and biomass. QMCH is a weighted average of the canopy height profile and therefore is very useful in determining the two major characteristics of stand development: one distribution of foliage and the other size and number of stems. In light of the importance of QMCH to understand the stand dynamics, the results of this study in estimating QMCH for two different types of native sclerophyll forests using LiDAR data is very encouraging.

When the regression model for the MDH for the CHAR forest was used to predict the MDH from the 100th percentile height of LiDAR points for the 19 sampling plots which were randomly selected for validation purpose and were not used in the regression analysis, the mean value of the predicted MDH was 37.7 meters, which is extremely close to the mean value of the field-measured MDH (38.53 m) for the sampling plots. Similarly, regression model for the MDH for the BRMS forest was able to predict the MDH for the 14 validation plots with a mean prediction bias of - 0.31 m, less than one percent of the field measured MDH. When the regression models for the QMCH were used to estimate the QMCH for the validation plots, it was estimated with a mean prediction bias of - 3.5 m (- 11.9%) for CHAR plots and with - 0.39 m (- 1.2%) for BRMS plots. This demonstrated that LiDAR underestimated the MDH as well as the QMCH for both types of forests. However, the underestimation of these stand heights in the BRMS forest was significantly lower than the underestimation in the CHAR forest. Underestimation of mean dominant height or quadratic mean canopy height for both types of forests from LiDAR data is consistent with the results of other studies conducted in different types of forests in which LiDAR underestimated various canopy height indices (Naesset, 1997; Nilsson, 1996). The results of this study again confirmed that LiDAR estimation of canopy height indices are more accurate for uniform forests but the accuracy of estimation for mixed species multilayered forests is also good and can be used for many other purposes with a reasonable level of confidence.

Estimating stem density (number of trees per hectare) is very important to understand the overall structure and the dynamics of the stand which are directly related to the basal area, volume and biomass of forest stand. LiDAR data have been used to estimate stem density in different type of forests. Various kind of LiDAR metrics have been used directly to develop predictive models for stem density. LiDAR estimates of various attributes including stem density in coniferous forests are more accurate than that of in broad-leaved and mixed species forests. Stem density (number of trees per hectare) is the most difficult forestry attribute to estimate from remote sensing technology including LiDAR (Hurich and Thoma, 2008; Lefsky et al., 2001). When various LiDAR metrics were used directly to develop a regression model of stem density in CHAR and BRMS forests, the models developed had very low (less than 0.3) R^2 . Therefore, in this study, an indirect method of estimating stem density using LiDAR data was developed. The results of this study in estimating stem density in CHAR and BRMS forest using Reineke's formula and LiDAR derived average diameter of trees is encouraging. Since other studies estimating stem density from LiDAR data using this method are not available, the results of this study cannot be compared with the results of other studies directly. However, the overall accuracy of this method in estimating stem density from LiDAR data can be compared to the accuracy obtained in some other studies. Lee and Lucas (2007) developed a model to estimate stem density in mixed species woodlands and open forests near Injune, Queensland, Australia from a complex LiDAR metric called Height-Scale Crown Openness Index (HSCOI) with an R^2 of 0.82 and RMSE of 133 stems per hectare. When they applied the model to estimate stem density in similar forest of northeast Victoria, the R^2 was 0.19 with RMSE of 152 stems per hectare. As actual prediction bias was not reported, it is hard to compare the results with the results of this study. However, RMSE of prediction in northeast Victoria suggests that the prediction accuracy was not as accurate as the prediction of validation plots of BRMS forest of this study.

Hurich and Thoma (2008) used a model developed with LiDAR metrics to estimate the number of trees per hectare in a coniferous forest in Germany. The prediction bias they obtained was 29 trees per hectare which is certainly very small compared to the prediction bias produced for CHAR forest in this study, but it is very large compared to the prediction bias obtained for BRMS of this study. Næsset and Bjercknes (2001) developed a prediction model for number of stems per hectare with a LiDAR metric as the predictor for a forest of Norway spruce and Scots

pine in southeast Norway with an R^2 value of 0.42 and in cross validation observed a very large standard deviation of prediction errors (1209 stem per hectare) which is significantly greater than the standard deviation of prediction errors obtained for calibration or validation plots of both CHAR and BRMS forest of this study. However, this study was conducted in forests having the MDH of about 40 m whereas the study conducted by Næsset and Bjercknes (2001) was concentrated on young forest lower than 8 m of height which could be the reason for the larger prediction errors.

It was expected that the prediction of stem density for CHAR would be less accurate than that for BRMS because Reineke's formula was developed for even-aged fully stocked forest and CHAR is non-uniform forest compared to BRMS forest. The results of this study, however, demonstrate that the method developed and used in this study can estimate stem density with a reasonable level of accuracy even in comparatively non-uniform and mixed species multi layered forests such as CHAR forest.

5. Conclusions

Estimating MDH and QMCH for forest stands of two different type of forests using LiDAR data was quite accurate. QMCH is a new index of stand height measurement which represents the height of the tree with the average diameter of trees in stand and therefore can have wider application in sustainable forest management and in forest ecosystem studies. The results of estimates of stem density using LiDAR derived average diameter of trees in Reineke's formula can have a great impact on application of LiDAR as an operational tool in Australian forestry.

Acknowledgements

This research was conducted under the Ph.D. scholarships provided by The University of Melbourne. Field inventory data and LiDAR data for the study areas were provided by the Victorian Department of Sustainability and Environment. We would like to thank two anonymous reviewers for their comments and suggestions.

References

- Andersen, H. E., Reutebuch, S. E. and McGaughey, R. J., 2006. A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. *Canadian Journal of Remote Sensing* 32: 355-366
- Anderson, J., Martin, M. E., Smith, M. L., Dubayah, R. O., Hofton, M. A., Hyde, P., Peterson, B. E., Blair, J. B. and Knox, R. G., 2006. The use of waveform lidar to measure northern temperate mixed conifer and deciduous forest structure in New Hampshire. *Remote Sensing of Environment* 105: 248-261
- Bortolot, Z. J. and Wynne, R. H., 2005. Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data. *ISPRS Journal of Photogrammetry and Remote Sensing* 59: 342-360
- Chen, Q., Baldocchi, D., Gong, P. and Kelly, M., 2006. Isolating Individual Trees in a Savanna Woodland Using Small Footprint Lidar Data. *Photogrammetric Engineering & Remote Sensing* 72 (8):923-932
- Chen, Q., Gong, P., Baldocchi, D., and Tian, Y. Q., 2007. Estimating Basal Area and Stem Volume for Individual Trees from LiDAR data. *Photogrammetric Engineering & Remote Sensing* 73 (12): 1355-1365
- Coops, N. C., Hilker, T., Wulder, M. A., St-Onge, B., Newnham, G., Siggins, A. and Trofymow, J. A., 2007. Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. *Trees-Structure and Function* 21: 295-310
- Dubayah, R. O., Sheldon, S. L. Clark, ., D. B., Hofton, M. A, Blair, J. B., Hurtt, G. C. and Chazdon, R. L., 2010. Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *Journal of Geophysical Research-Biogeosciences* 115
- Duncanson, L. I., Niemann, K. O. and Wulder, M. A., 2010. Estimating forest canopy height and terrain relief from GLAS waveform metrics. *Remote Sensing of Environment* 114: 138-154
- Heurich, M. and Thoma, F., 2008. Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural European beech (*Fagus sylvatica*) and Norway spruce (*Picea abies*) forests. *Forestry* 81 (5):645-661
- Keith, H., Mackey, B. G. and Lindenmayer, D. B., 2009. Re-evaluation of forest biomass carbon stocks and lessons from the world's most carbon-dense forests. *Proceedings of the National Academy of Sciences of the United States of America*, 106(28): 11635-11640, (doi:10.1073/pnas.0901970106)
- Lee, A. C. and Lucas, R. M., 2007. A LiDAR-derived canopy density model for tree stem and crown mapping in Australian forests. *Remote Sensing of Environment* 111:493-518
- Lefsky, M. A., 1997. Application of Lidar Remote Sensing to the Estimation of Forest Canopy and Stand Structure. Ph. D. Dissertation, Department of Environmental Science, University of Virginia

- Lefsky, M. A., Cohen, W. B. and Spies, T. A., 2001. An evaluation of alternate remote sensing products for forest inventory, monitoring, and mapping of Douglas-fir forests in western Oregon. *Canadian Journal of Forest Research* 31: 78-87
- Lefsky, M. A., Harding, D. J., Cohen, W. B., Parker, G. G. and Shugart, H. H., 1999. Surface LiDAR remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of Environment*, 67:83-98
- Naesset, E., 1997. Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sensing of Environment* 61:246–253
- Næsset, E. and Bjercknes K. O., 2001. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. *Remote Sensing of Environment* 78:328-340
- Newton, A. C., 2007. *Forest Ecology and Conservation, A Handbook of Techniques*. Oxford University Press
- Niklas, K. J., 1994. *Plant Allometry: The Scaling of Form and Process*. The University of Chicago Press. Chicago and London
- Nilsson, M., 1996. Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sensing of Environment* 56(1): 1-7.
- Oliver, C. D. and Larson, B. C., 1990. *Forest Stand Dynamics*. McGraw-Hill Inc., New York
- Persson, A., Holmgren, J. and Sodermann, U., 2002. Detecting and measuring individual tree using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing* 68(9): 925-932
- Pretzsch, H. and Biber, P., 2005. A Re-Evaluation of Reineke's Rule and Stand Density Index. *Forest Science* 51(4):304-320
- Spies, T. A., 1998. Forest Structure: A Key to the Ecosystem. *Northwest Science*, 72 special issue No 2: 34-39
- Stone, J. N. and Potter, J. L., 1998. What is forest stand structure and how to measure it? *Northwest Science* 72, special issue No 2: 25-26
- Vanclay, J. K., 2009. Tree diameter, height and stocking in even-aged forests. *Annals of Forest Science* 66:702
- West, P. W., 1982. Comparison of stand density measures in even-aged regrowth eucalypt forest of southern Tasmania. *Canadian Journal of Forest Research* 13:22-31
- Zeide, B., 2005. How to measure stand density. *Trees* 19:1-14
- Zeide, B., 1995. A relationship between size of trees and their number. *Forest Ecology and Management* 72:265-272
- Zeide, B., 1987. Analysis of the 3/2 power law of self-thinning. *Forest Science* 33: 517-537

Using a Flux Footprint Model and Airborne LiDAR to Characterize Vegetation Structure and Topography Frequently Sampled by Eddy Covariance: Implications for MODIS Product Validation

L. Chasmer¹, N. Kljun², C. Hopkinson³, S. Brown¹, T. Milne³, K. Giroux¹,
A. Barr⁴, K. Devito⁵, I. Creed⁶, and R. Petrone¹

¹Cold Regions Research Centre, Wilfrid Laurier University, Waterloo ON Canada
Email: lechasme@yahoo.ca

²Dept. of Geography, College of Science, University of Swansea, Swansea, Wales UK

³Applied Geomatics Research Group, NSCC, Middleton NS Canada

⁴Environment Canada, National Water Research Institute, Saskatoon SK Canada

⁵Dept. of Biological Sciences, University of Alberta, Edmonton AB Canada

⁶Dept. of Biology, University of Western Ontario, London ON Canada

Abstract

Exchanges of CO₂ transported to eddy covariance instruments are often assumed to be representative of site average vegetation, understory, and topographical characteristics, regardless of the frequency with which these have been sampled. All sites have some degree of heterogeneity (e.g. an upland area, bog, area of dense understory, etc.), which could influence CO₂ exchanges if scalar fluxes from prevailing wind directions frequently sample these parts more than others. This could have implications for site representation, model evaluation, and remote sensing product validation and scaling.

The use of flux footprint models has improved our understanding of the spatial and temporal distribution of source/sink areas measured within the field of view of eddy covariance instrumentation (e.g. Schmid, 1994). The flux footprint is defined as the probability of flux contribution per unit area upwind of the eddy covariance instrumentation (Kljun *et al.* 2002, 2004). When a footprint is combined with remote sensing data, the probability density function of the weighted source/sink contribution to the eddy covariance instrumentation provides spatially contiguous information on vegetation structural and topographic influences on net ecosystem production (NEP) (Chasmer *et al.* 2008). Simple logic follows: if CO₂ fluxes originate from areas of higher biomass, then measurements of flux should indicate increased uptake (NEP) when compared with lower biomass areas (etc.), all else being equal.

Combining footprints with high resolution spatially continuous remote sensing data from airborne LiDAR, hyperspectral or spectral imagery provides a powerful tool for characterizing the areas sampled most frequently by eddy covariance. In this study, we use a 3D classification methodology to characterize vegetation structural and topographic attributes most frequently sampled by eddy covariance within 1) a homogeneous mature boreal aspen stand; and 2) a heterogeneous upland aspen/wetland complex using airborne LiDAR. The vegetation and topographic characteristics found within the areas most sampled at each site were then used to classify the larger region for evaluation of the MODIS gross primary production (GPP) product, i.e. choosing MODIS pixels that have similar attributes to those found within footprint most frequently sampled by eddy covariance.

The results of this study find that footprints from prevailing wind directions at the homogeneous mature aspen stand have, on average, taller trees (7%), greater effective LAI (30%), denser understory (5%), and fewer low-lying topographic depressions than secondary wind origins. At the heterogeneous aspen stand, footprints from prevailing wind directions have, on average,

shorter trees (-11%), lower effective LAI (-17%), and a greater proportion of topographic depressions. Classification of vegetation structure and topography within a 1 km radius of the homogeneous and heterogeneous stands indicated that 56% (homogeneous aspen) and 69% (heterogeneous aspen) were representative of vegetation and topographic attributes sampled by eddy covariance. Thus, prevailing wind directions may over- or under-sample some parts of the ecosystem more than others, which could result in over- or underestimates of NEP when compared with similar representative ecosystems.

When scaled to MODIS GPP, correspondence with GPP estimated using eddy covariance and meteorological methods improved by 13% when using LiDAR ‘classified’ pixels as opposed to those pixels most proximal to the tower. This illustrates that airborne LiDAR and footprint analysis can be used to link eddy covariance measurements of ecosystem exchanges between scales. This has important implications for assessment of spatial variability of vegetation/topography on NEP; identifying landscape features that are frequently sampled; classifying spatial heterogeneity; and scaling. More detail of this study is provided in Chasmer *et al.* (2011).

Key Words: CO₂ flux, flux footprint modelling, scaling, MODIS, eddy covariance representation, airborne LiDAR.

Chasmer, L., Kljun, N., Barr, A., Black, A., Hopkinson, C. McCaughey, H. and Treitz, P. 2008. Vegetation structural and elevation influences on CO₂ uptake within a mature jack pine forest in Saskatchewan, Canada. *Canadian Journal of Forest Research*, 38, 2746-2761.

Chasmer, L., Kljun, N., Hopkinson, C., Brown, S., Milne, T., Giroux, K., Barr, A., Devito, K., Creed, I., and Petrone, R. 2011. Characterizing vegetation structural and topographic characteristics sampled by eddy covariance within two mature aspen stands using LiDAR and a flux footprint model: Scaling to MODIS. *Journal of Geophysical Research-Biogeosciences, Special Issue on Advances in Upscaling of Eddy Covariance Measurements of Carbon and Water Fluxes*, 115, doi:10.1029/2010JG001567.

Kljun, N., Rotach, M.W., Schmid, H.P., 2002. A 3D Backward Lagrangian Footprint Model for a Wide Range of Boundary Layer Stratifications. *Boundary-Layer Meteorology*, 103, 205-226.

Kljun, N., Calanca, P., Rotach, M., and Schmid, H. 2004. A simple parameterisation for flux footprint predictions. *Boundary-Layer Meteorology*, 112, 503–523.

Schmid, H.P., 1994. Source areas for scalars and scalar fluxes. *Boundary-Layer Meteorology*. 67, 293-318.

** No full proceedings paper to follow (due to publication in JGR-B). Extended abstract only.

The significance of managed and natural vegetation on house survival during wildfires

Anders Siggins, Glenn Newnham & Raphaele Bianchi

¹CSIRO Land and Water, Anders.Siggins@csiro.au

²CSIRO Land and Water, Glenn.Newnham@csiro.au

³CSIRO Ecosystem Sciences, Raphaele.Bianchi@csiro.au

1. Introduction

The impact of wildfires at the urban interface is a major concern for people safety and property loss. Different models have been used to characterise the threat at the interface (e.g. Ahern Chladil 1999, Cohen 1995, Theobald and Romme 2007, Leonard 2009 Gill and Stephen 2009) taking into account some of the following parameters: vegetation, weather, topography, and urban vulnerability. Distance to forest has also been used to estimate the risk of house loss (Ahernad Chladil 1999). Most of the studies consider only the natural forest as part of the risk assessment, and managed vegetation such as isolated trees and wind breaks are only taken into account for detailed risk assessment at the individual house level (e.g. Leonard 2009). However, there is some evidence that managed vegetation plays a significant role in the fire spread leading to house ignition (Ramsey 1987, Leonard 2003, Cohen 2008 and Gill 2009).

The Australian forest fires of February 2009 resulted in the highest loss of life from forest fires in Australian history, and occurred in semi-rural and rural areas in Victoria. The most deadly of these fires occurred in the Kinglake region north of Melbourne. In this area there is generally no clear delineation of the urban interface. In this situation, where houses can be located within the natural forest, each house has its own unique forest boundary which must be defended in the event of a fire. Along with the natural forest, there are also small forest patches and isolated trees between houses and the forest that may provide a direct fuel path to the house for the flame front. These patches of vegetation are typically assumed to be ‘managed’ vegetation; areas from within which fire fighters can carry out suppression activities safely and effectively (Coleman 1995, Gill and Stephen 2009).

Due to the extent of the Kinglake fire, manual delineation of managed vegetation and natural forest would be a very time consuming process. Lidar data collected prior to the fire was used to generate maps of vegetation extent using a cover threshold method. Spatial patterns of vegetation extent were then used to discriminate managed vegetation from natural forest. The proximity and cover density of both managed vegetation and natural forest were then derived for each house, both destroyed and undamaged. The significance of both the managed and natural components of the vegetation extent in determining the probability of house loss was then explored.

2. Data and Methods

The Kinglake fire burned an area north of Melbourne of approximately 180,000ha, resulting in the loss of 1244 houses, and 120 fatalities. As part of a vegetation assessment program conducted by DSE in late 2008, a large proportion of this area had been flown with lidar data for natural resource management programs. Post-fire surveys found that of the 3656 homes recorded within the fire extent, 34% were destroyed by the fire. The location of the approximate centre of each house, along with the severity of any damage was collected as part of these surveys. Of the 3656 houses, 2116 fell within the region defined by the lidar data, 921 were

classified as destroyed, 357 received minor damage, and 838 were undamaged. For the purposes of this study, it was decided that the minor damage category was too ambiguous, and it was excluded from the analysis (Leonard et al 2009).

The lidar data was collected by fixed wing aircraft using an Optech ALTM3100EA instrument during 2008, with data collection finishing late in the year. The flying height was 1300m and the average point density for the entire area was 1.12 pulses per square metre. Data were provided in LAS format, with points pre-classified as either ground or non-ground. The ground point data were used to generate a two meter spatial resolution digital elevation model using an inverse distance weighted algorithm for all ground classified points within each grid cell:

$$GroundElevation = \sum_{i=1}^n w_i z_i \quad (1)$$

Where w_i is the weight of point i , and z_i is the elevation of point i . The weight w is calculated as:

$$w_i = \frac{d_i^{-p}}{\sum_{j=1}^n d_j^{-p}} \quad (2)$$

Where d_i is the distance from the point to the interpolation point, and p is the power parameter (set to 2).

All data were then corrected to height above terrain. Vertically projected vegetation cover fraction (C) was calculated for each cell using equation 2, but in order to ensure that there were sufficient data to produce a robust cover estimate, each cell included data from a five by five metre area centred on the output pixel. Cover was assessed for vegetation above two metres.

$$C = \frac{\# z_j | z_j > 2}{N} \quad (3)$$

Where N is the total number of lidar points within five metres of the pixel centre, and i_j is the height of the j^{th} point above the DEM (Lovell et al., 2003). A final binary forest/non-forest layer was produced by imposing a 20% threshold on the cover layer. Combining estimates of cover above two metres, with the 20% cover threshold, ensures that this forest/non-forest layer is consistent with the definition of forest as used in Australia's State of the Forest report (Montreal Process Implementation Group for Australia (2008)).

Managed vegetation includes patches of remnant forest, trees along roads and fence boundaries, and trees around homes. Tall shrubs (greater than two metres high) are also considered as part of the forest layer. These sources of managed vegetation can often connect to the natural forest, and can be difficult to remove without affecting the forest region delineation. An automated method of removal was employed using binary morphological erosion and dilation procedures. The details of this process can be found in Newnham et al (submitted). Two binary images were then generated; a natural forest/non-forest layer, and by subtracting this layer from the original lidar binary layer, a managed forest/non-forest layer.

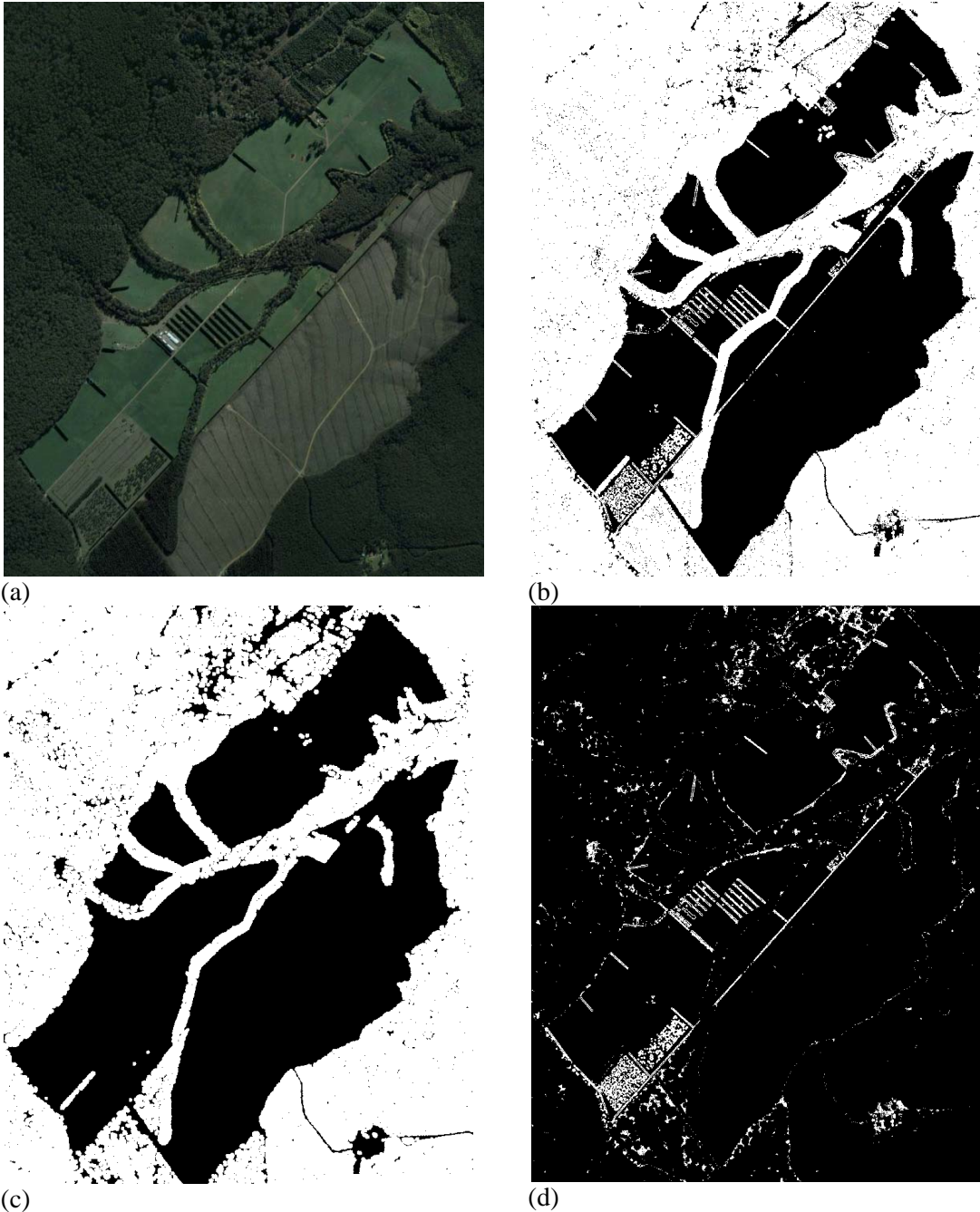


Figure 1: A subset of the fire affected area showing both managed and natural forest (a) Aerial photography provided by Google Maps, (b) the forest non-forest layer generated from the lidar data c) the natural forest layer and (d) the managed forest layer.

The managed and natural pixel closest to each house was determined using a radial search. A 200m radius region located around each house centre was then used to determine the total percentage covered for each of the vegetation classes. This resulted in four metrics associated with each individual house:

- Minimum distance to natural forest
- Cover fraction of natural forest within 200m
- Minimum distance to managed vegetation
- Cover fraction of managed vegetation within 200m

A binary-logistic regression model was then developed to explain the loss of homes in the fire based on these four metrics.

3. Result

The majority of managed vegetation occurred within 20 metres of the house, with no houses showing a minimum distance to managed vegetation of greater than 40m. The minimum distance to natural forest was always less than 200 metres, with the majority of houses being within 50 metres of natural forest. The cover statistics show that there is less managed vegetation surrounding houses in this area than the natural forest, though this managed vegetation generally occurs closer to the houses (see Figure2).

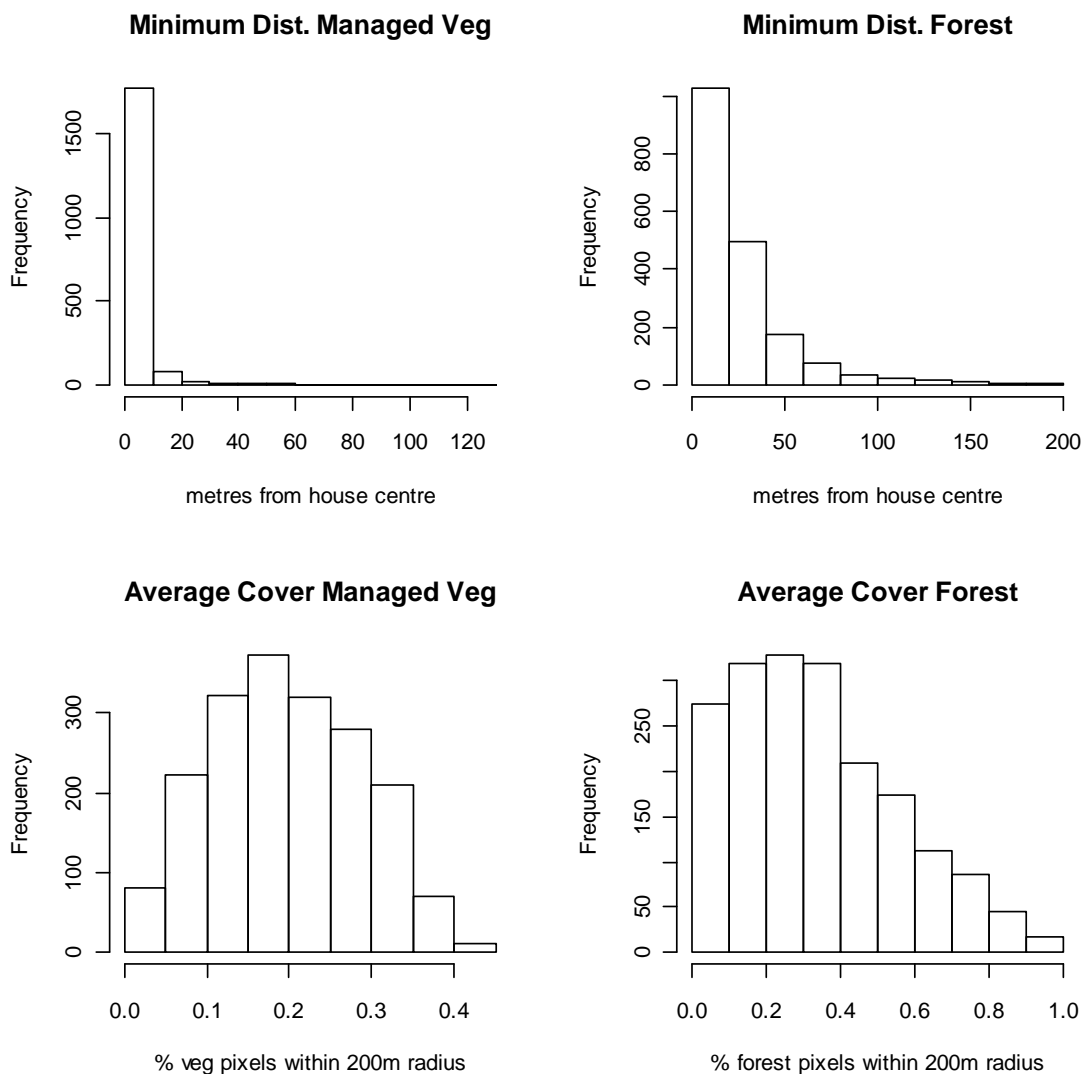


Figure 2: Histograms of lidar metrics across all houses for a) minimum distance to managed vegetation, b) minimum distance to natural forest, c) average cover of managed vegetation, b) average cover of natural forest

The binary logistic regression model showed that the significant factors in predicting the

destruction of houses in the Kinglake fire were the percentage of cover of natural vegetation ($p < 0.01$), the percentage of cover of the managed vegetation ($p < 0.01$) and the minimum distance to natural forest ($p < 0.01$). The minimum distance to managed vegetation was not found to be a significant predictor of the destruction of houses ($p = 0.19$). This may be due to the lack of variation in the distribution of distances to the managed vegetation for this particular fire affected area.

A new model was created with total cover around each house represented as the combination of the natural and managed cover results, which showed both the total vegetated cover and minimum distance to natural forest were significant ($p < 0.01$). The Akaike information criterion for this and the four parameter model was identical. Both models predict house destruction in approximately 55% of cases.

4. Discussion

The results of this lidar analysis suggest that the cover fraction and distance to natural forest were both significant variables in predicting the likelihood of a house being destroyed. The model predicted the destruction of houses correctly based on the characteristics of the surrounding vegetation in only 55% of cases. This is not surprising given the large number of factors not considered in this study that would have an effect on fire behaviour, including the topography of the land, wind speed and wind direction, as well as the vulnerability of the urban environment and the fire management processes employed during the fire. Some of these additional parameters may also be derived from the lidar, such as the digital terrain model, and used in conjunction with data from other sources to further refine the model.

Acknowledgements

We gratefully acknowledge the Department of Sustainability and Environment, Victoria, for providing the lidar data that was used in this study.

References

- Ahern, A., & Chladil, M. (1999). How far do bushfires penetrate urban areas. In, *Proceedings of the 1999 Australian Disaster Conference* (pp. 21–26). Chen, K.P., & McAneney, J. (2004). Quantifying bushfire penetration into urban areas in Australia. *Geophysical Research Letters*, 31, L12212.
- Cohen, J. D. (2008). The wildland-urban interface fire problem. A consequence of the fire exclusion paradigm. *Forest History Today* (Fall 2008): 20-26.
- Cohen, J. D. (1995). Structure ignition assessment model (SIAM). *Biswell Symposium: fire issues and solution in urban interface and wildland ecosystems*. D. R. Weise and R. F. Martin. Albany, CA, (USDA Forest Service, Washington, DC): 85-92.
- Gill, A. M. and S. L. Stephens (2009). Scientific and social challenges for the management of fire-prone wildland-urban interfaces. *Environmental Research Letters* 4(3): 034014.
- Haralick, R.M., Sternberg, S.R., & Zhuang, X. (1987). Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 532-550.
- Leonard, J , Blanchi, R, Leicester, R, Lipkin, F , Newnham, G, Siggins, A, Opie, K, Culvenor, B, Cechet, B, Corby, N, Thomas, C, Habili, N, Jakab, M , Coghlan, R,

- Lorenzin, G, Campbell, D, Barwick, M (2009). Building and Land use planning research after the 7th February 2009 Victorian bushfires. *Preliminary findings, Interim report* USP2008/018 - CAF122-2-12
- Leonard, J., R. Blanche, et al. (2009). Profiling urban interface vulnerability. *Fire and Material 2009 11th Conference*. 26-28 January 2009, San Francisco, USA.
- Leonard, J. E. (2003). People and property – a researcher's perspective. In *Australian Burning: Fire Ecology, Policy and Management Issues*. G. Cary, D. Lindenmayer and S. Dovers. Melbourne, Vic, CSIRO Publishing.
- Lovell, J.L., Jupp, D.L.B., Culvenor, D.S., & Coops, N.C. (2003). Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. *Canadian Journal of Remote Sensing*, 29, 607-622.
- Montreal Process Implementation Group for Australia (2008). *Australia's State of the Forests Report 2008*. Bureau of Rural Sciences, Canberra.
- Newnham, G.J., Siggins, A.S., Blanche, R., Culvenor, D.S., Leonard, J., Mashford, J. Assessing Vulnerability at the wildland-urban interface using airborne lidar. (*in preparation*).
- Ramsay, G. C., McArthur, N.A. Dowling, V.P. (1987). Preliminary results from an examination of house survival in the 16 February 1983 bushfires in Australia. *Fire and Materials* 11: 49-51.
- Serra, J.P. (1982). *Image analysis and mathematical morphology*. London ; New York: Academic Press.
- Theobald, D. M. and W. H. Romme (2007). Expansion of the US wildland-urban interface. *Landscape and Urban Planning* 83(4): 340-354.

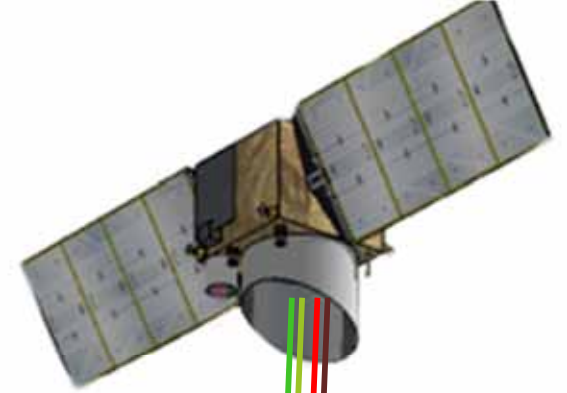
Silvilaser 17-19th Oct, 2011, Tasmania

Looking forward to LiDAR's colourful future

Dr Iain H Woodhouse

i.h.woodhouse@ed.ac.uk

With thanks to: Caroline Nichol, Felix Morsdorf, Genevieve Patenaude, John Moncrieff, Jim Jack, Emal Rumi, David Henry, Mal MacDonald, Antonio Delussu, Jackie Rosette





The value of forests



Radar



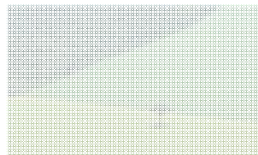
Multispectral Lidar (air)



Multispectral Lidar (space)



The value of forests



Radar

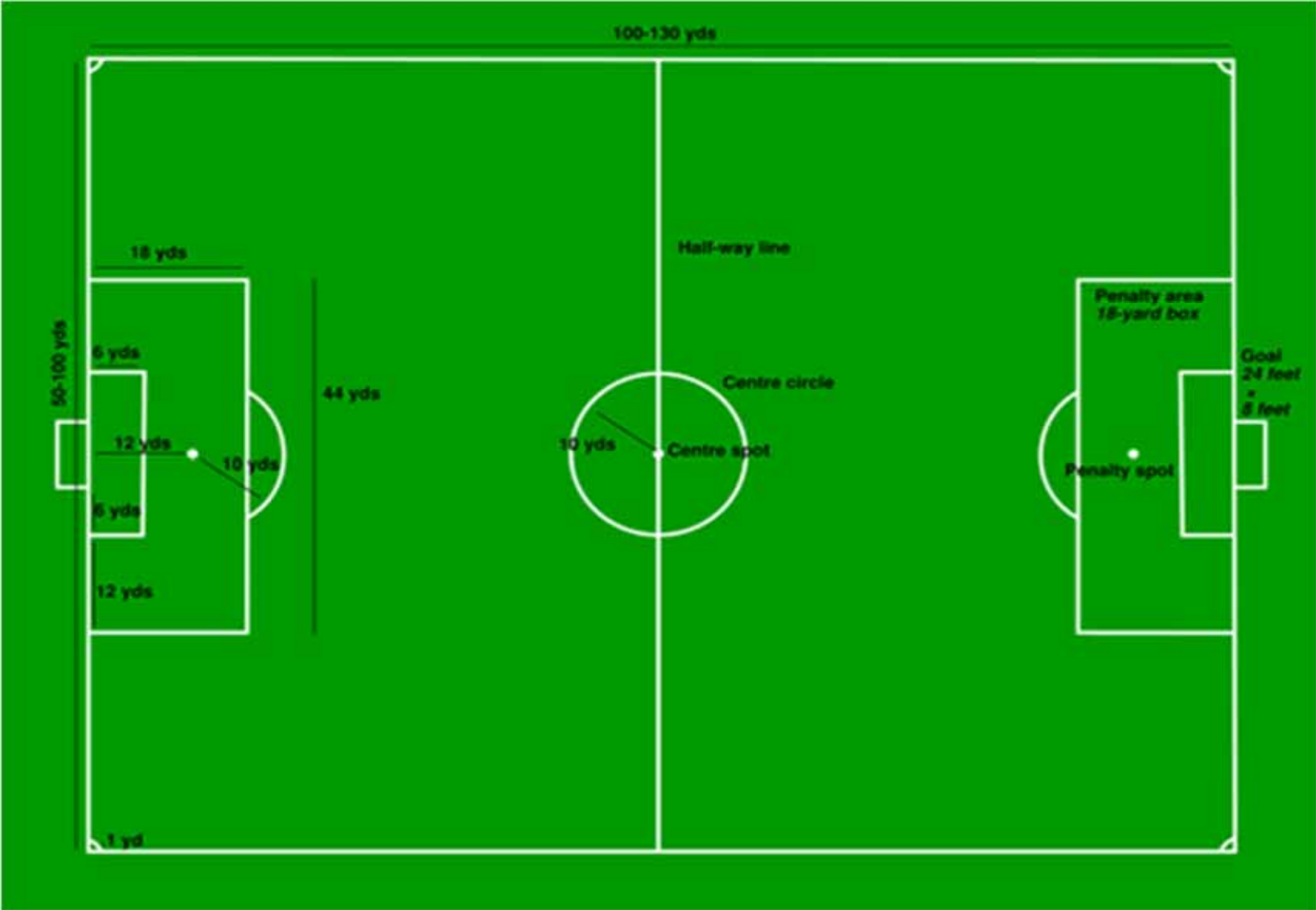


Multispectral Lidar (air)



Multispectral Lidar (space)

Global deforestation rate:



THE INDEPENDENT

Five trillion dollars – the cost, per year, of vanishing rainforest

And that's just the start, says a
trailblazing UK study

By Matt Chorley
POLITICAL CORRESPONDENT

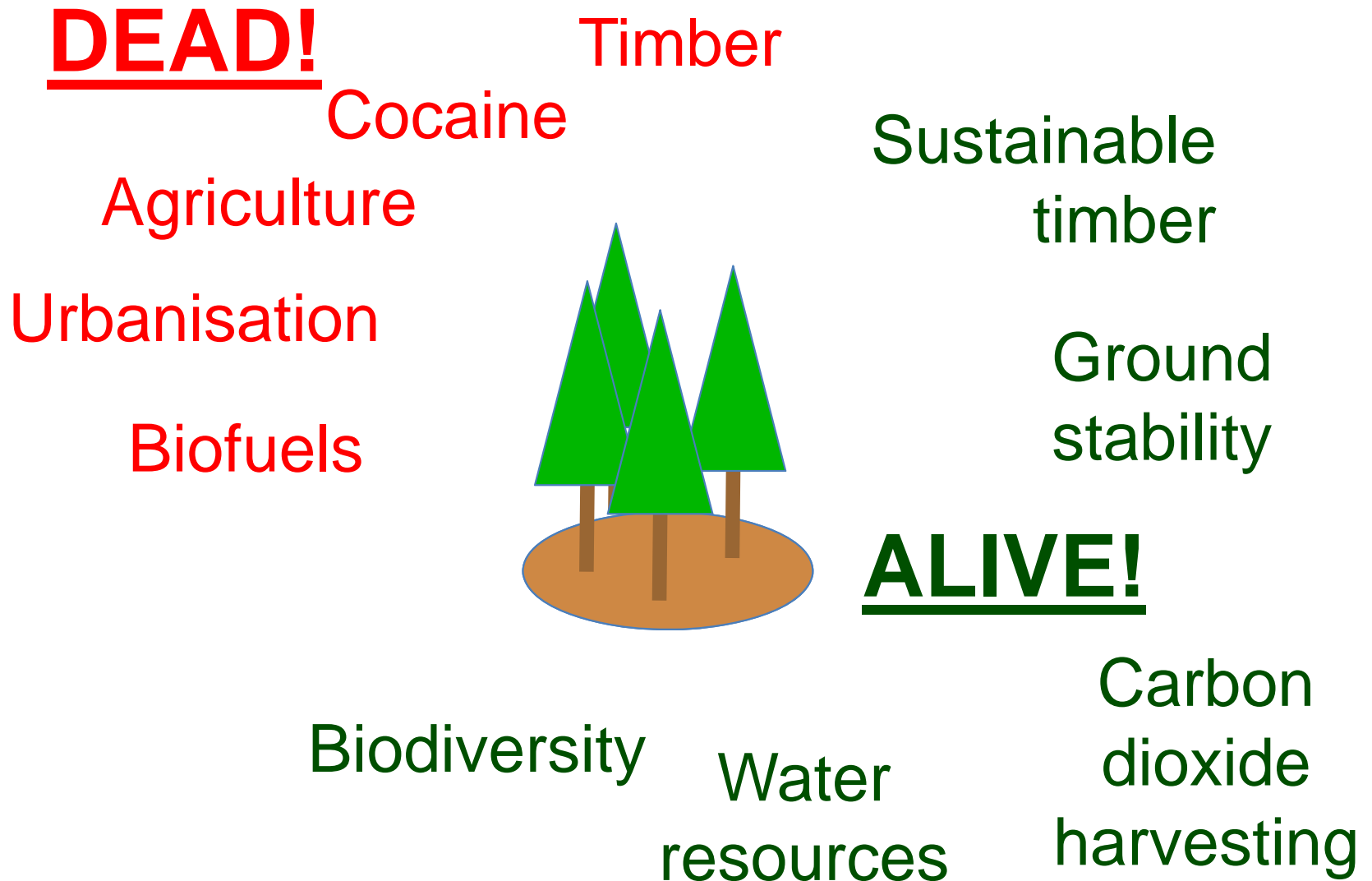
British scientific experts have announced a major breakthrough in the fight to save the natural world from climate change, leading to an international effort to safeguard a global system worth at least \$5 trillion a year to the world economy.

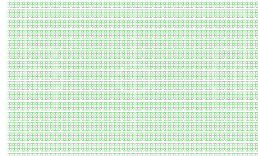
Groundbreaking new research by a former banker, Pavan Sukdolak, has placed a price tag on the value of the global network of environmental services. The study triggered an international agreement to halt the destruction of rainforests, wetlands and coral reefs.

With experts warning that the world is on a collision course with disaster to stem the loss of biodiversity, the study is two decades behind the global climate change agenda, the United Nations says. The World Bank and many other countries almost every government in the world can afford to believe that they will be unaffected by the alarm. The study, which species are disappearing, is part of the Convention on Biological Diversity, which was signed in Nagoya, Japan, later this year. The shift from solely ecological concerns to a hard-headed assessment



The pressure on forested land:

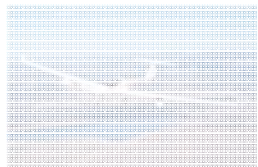




The value of forests



Radar



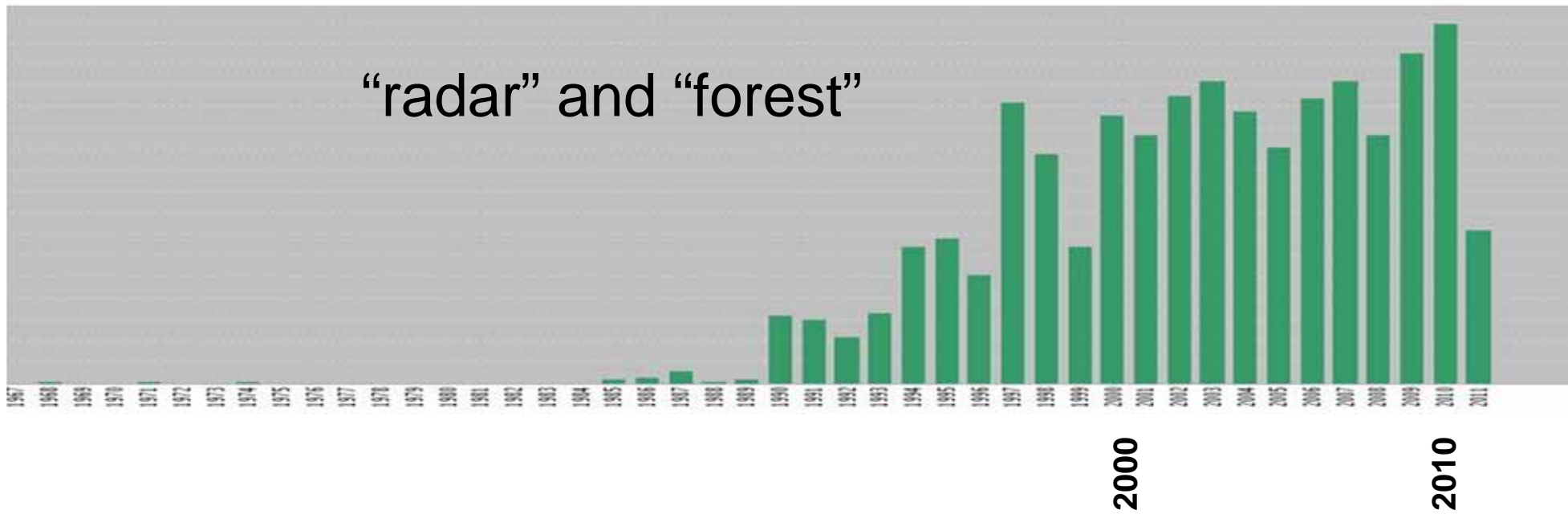
Multispectral Lidar (air)



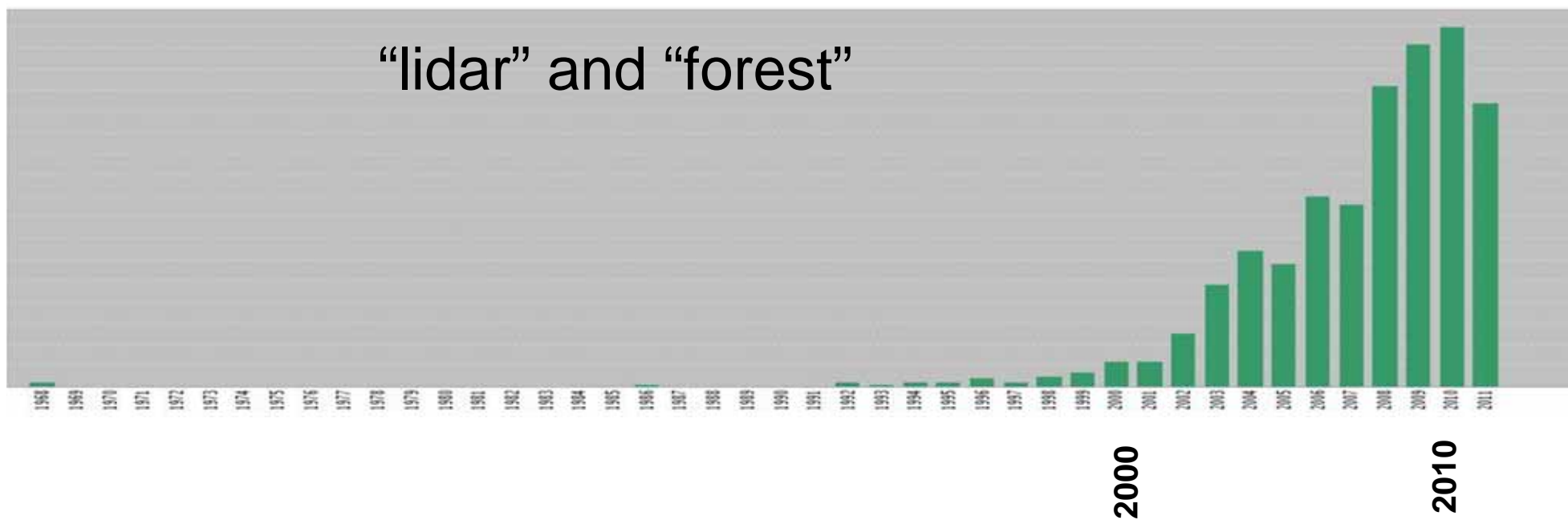
Multispectral Lidar (space)

Total publications:

“radar” and “forest”



“lidar” and “forest”



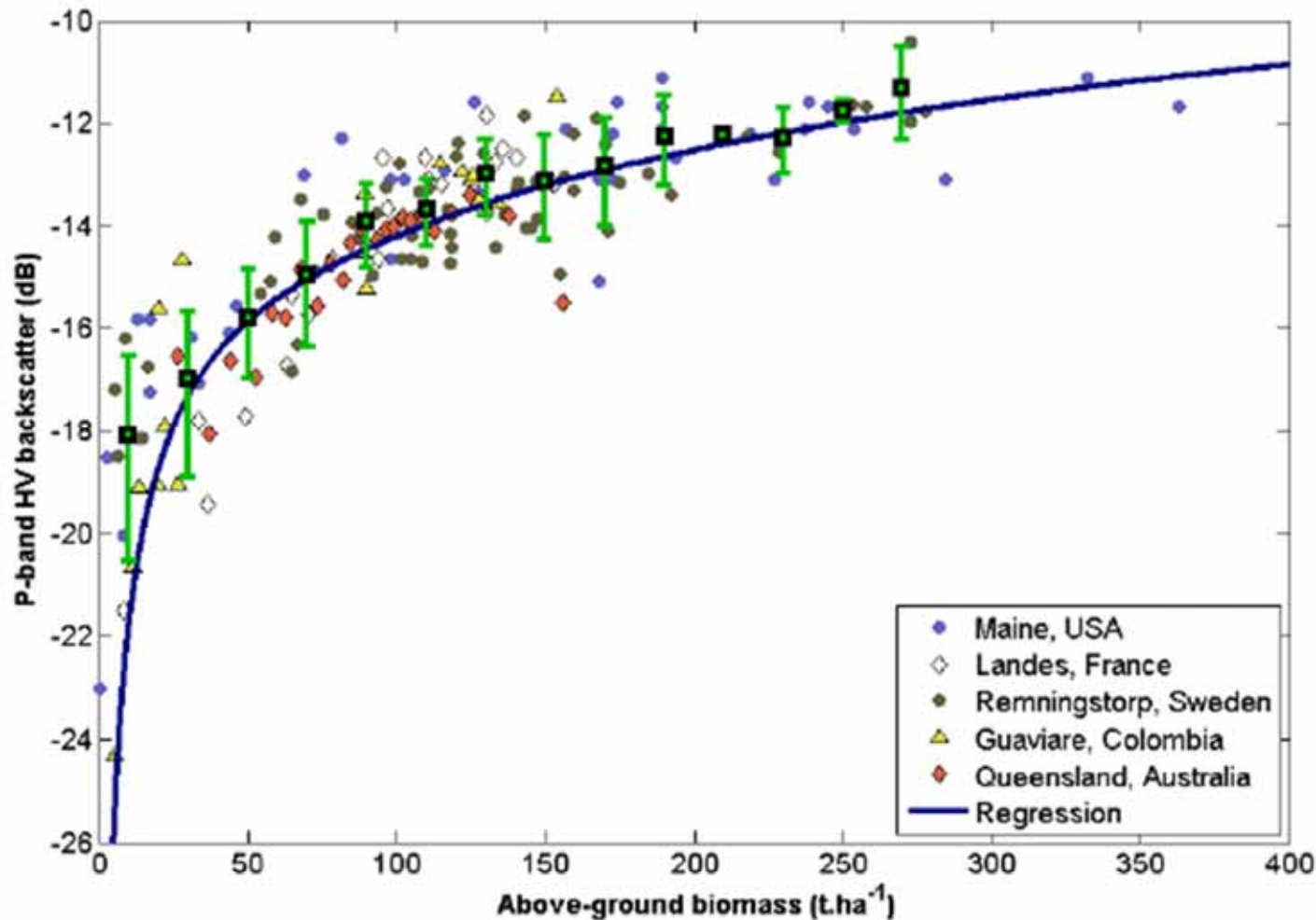
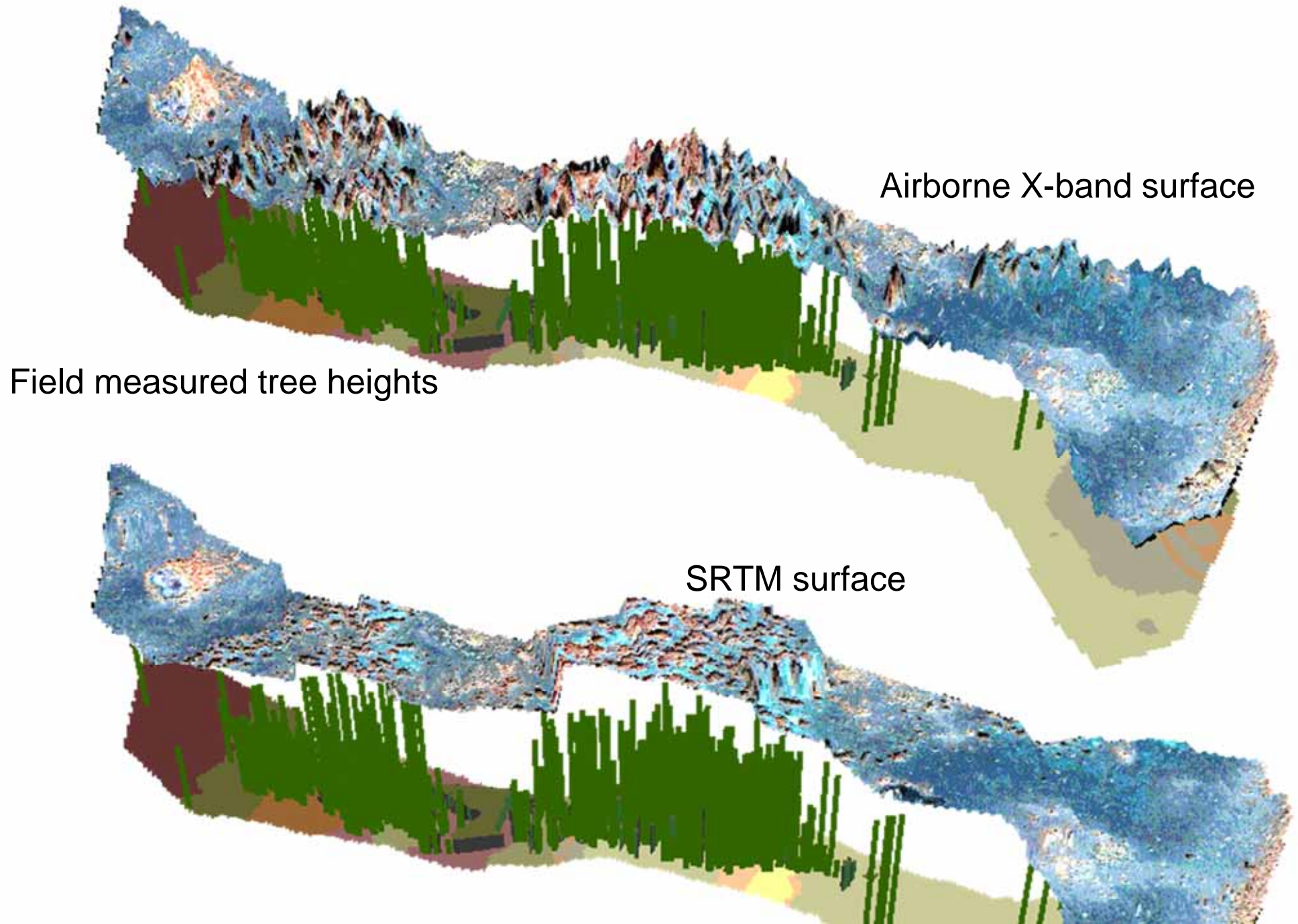


Figure 2.10: P-band HV backscattering coefficient plotted against above-ground biomass from experiments conducted at five different forests. The backscattering coefficient is expressed as $\beta^0 = \sigma_{HV}^0 / \cos \theta$ where θ is the incidence angle to provide a first order correction for differences due to the different incidence angles in the various datasets. The green points with error bars represent the mean value and standard deviation of all points falling within a biomass bin of $\pm 10 \text{ t ha}^{-1}$. The line is a regression curve applied to the full dataset. The corresponding RMSE in biomass is 51.6 t ha^{-1} and the coefficient of determination $r^2 = 0.67$. [Credits: Le Toan]

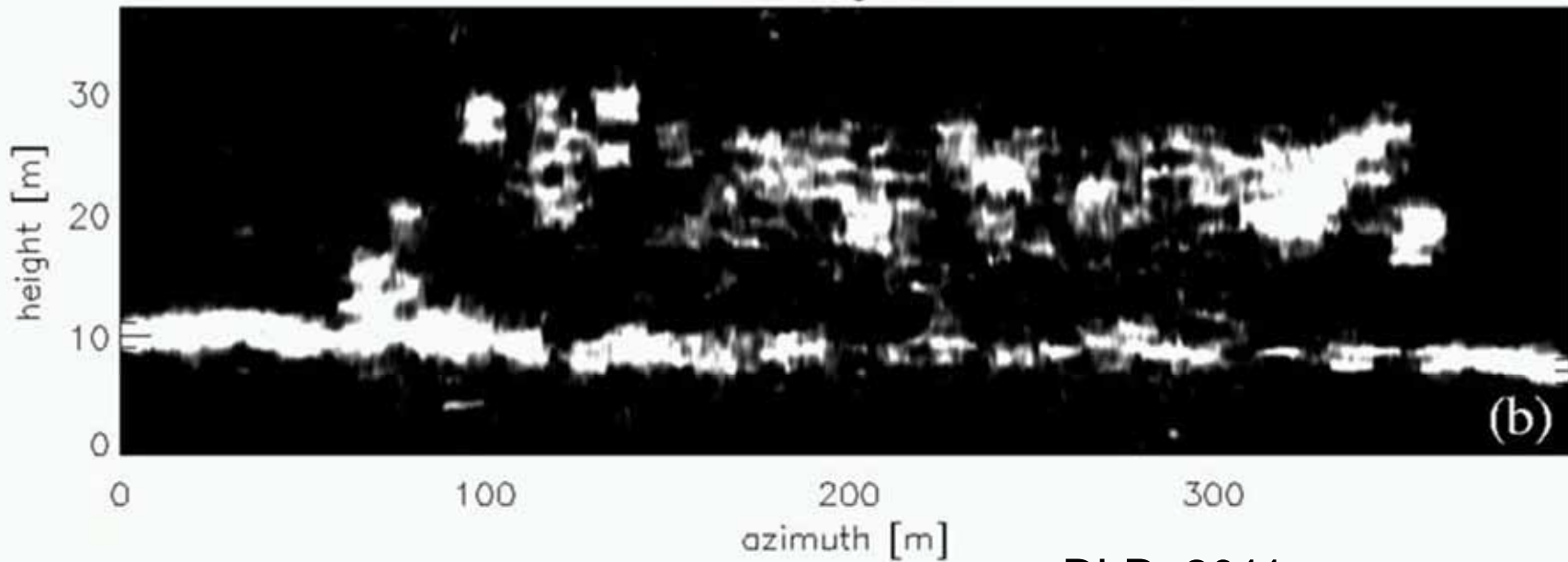
Canopy height from InSAR in savanna in Belize



SAR tomography

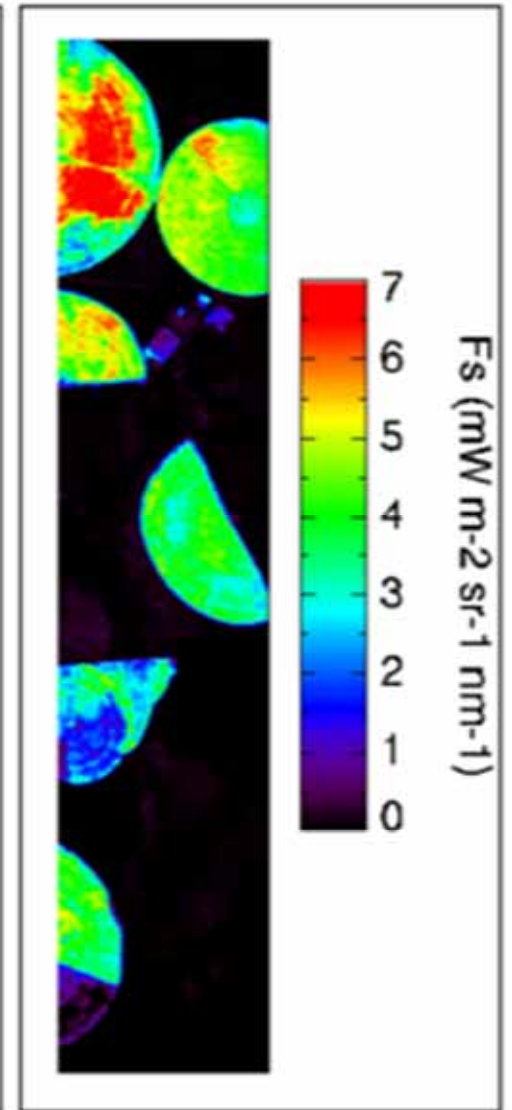
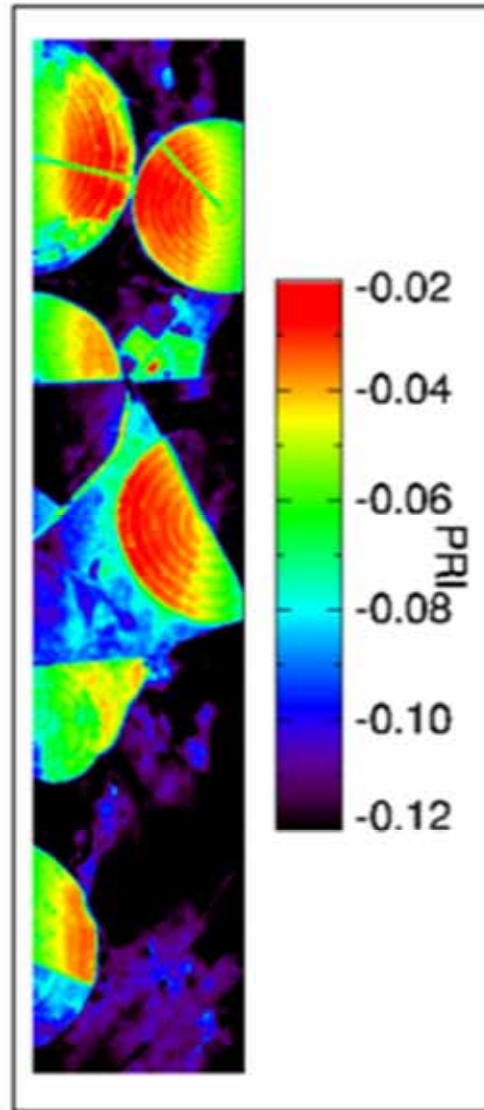
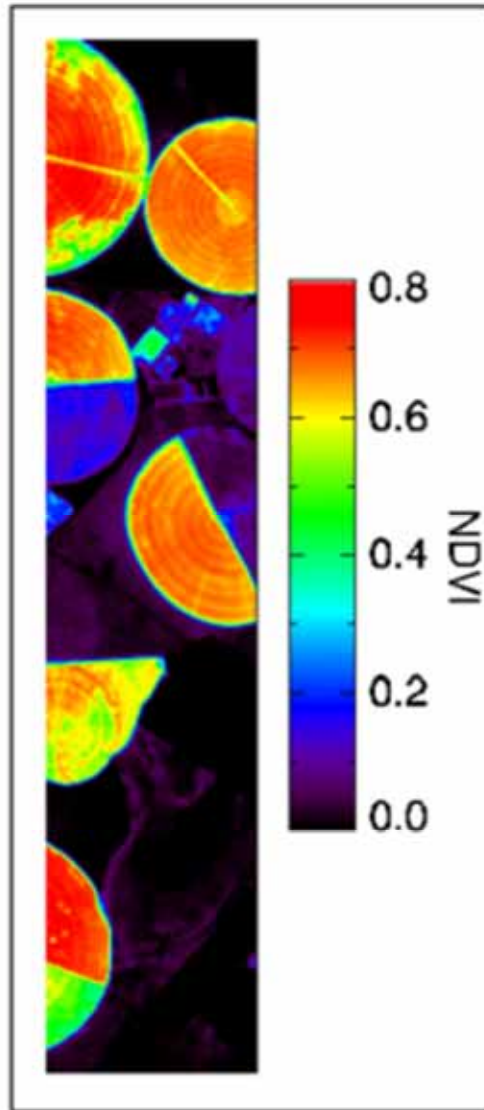
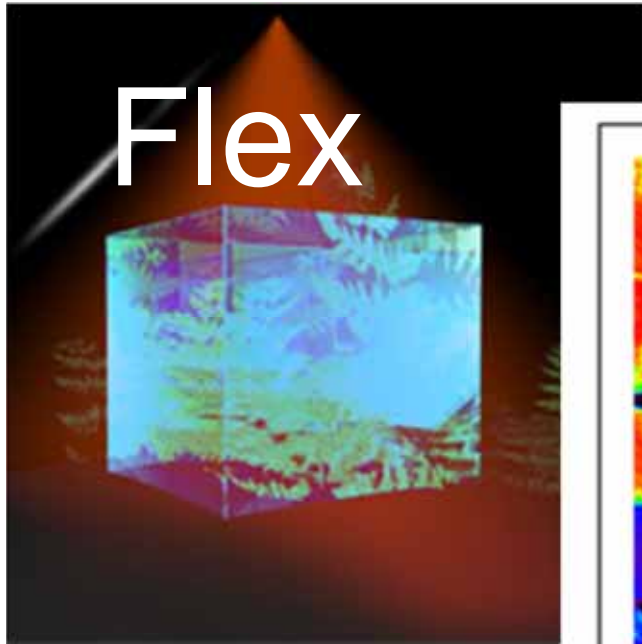


Tomogram

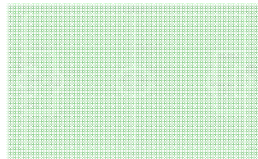


From statics to process?

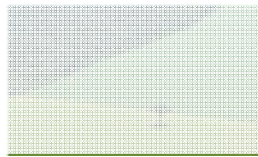
FLEX – Fluorescence Explorer
HypIRI



From FLEX MAG
Report, ESA



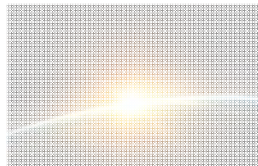
The value of forests



Radar



Multispectral Lidar (air)



Multispectral Lidar (space)

University of Edinburgh
Eco-Dimona



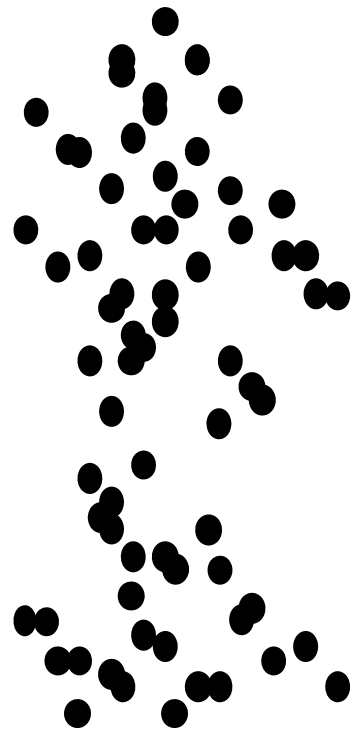
Parallel network
of non-linear
optical
wavelength
conversion
channels excited
by a common
laser



www.carbomap.com

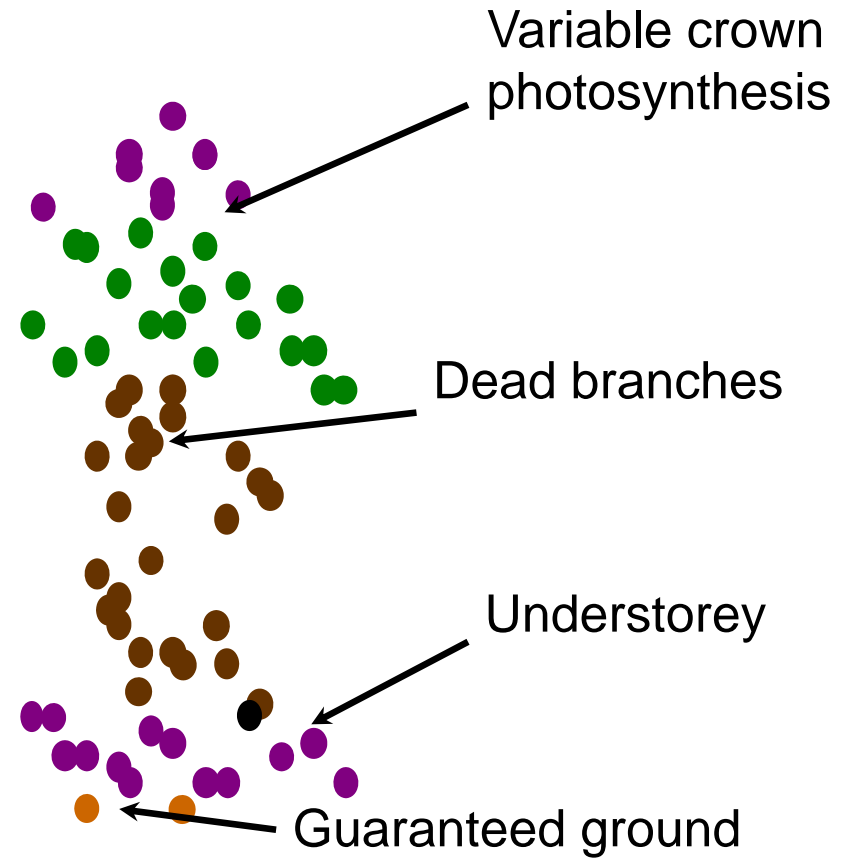
Multispectral lidar

Existing systems

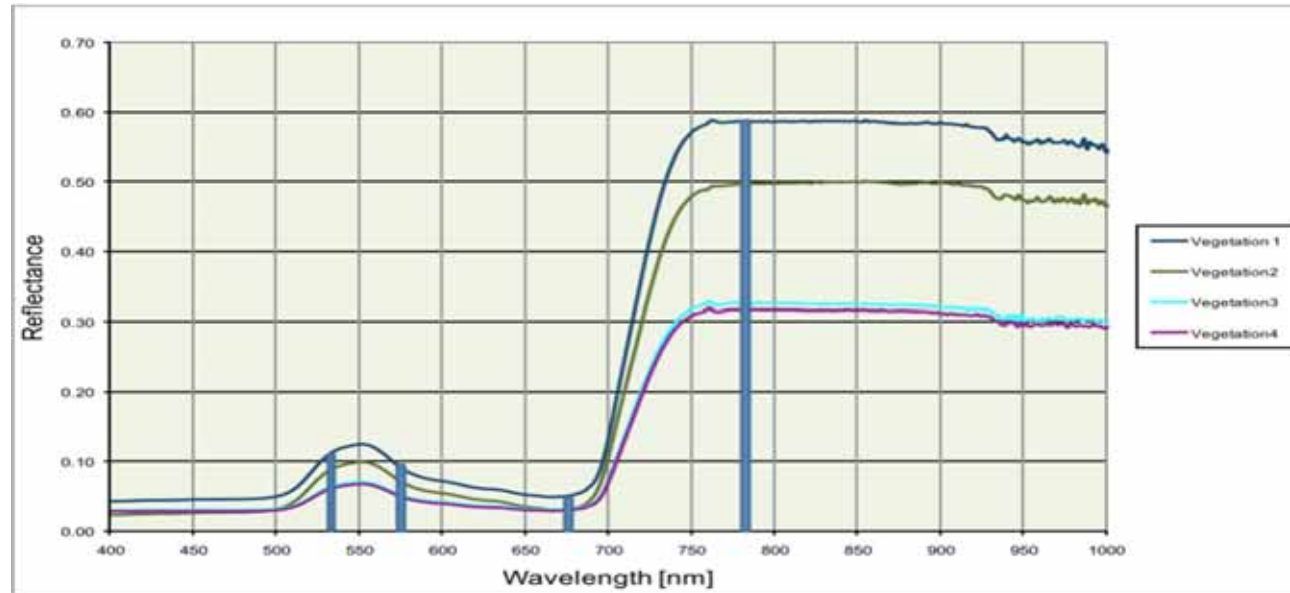


Tree height

Multispectral



Key Parameters for GPP Retrieval



For absorbed photosynthetic active radiation (APAR):

NDVI : Normalized Difference Vegetation Index

$$NDVI = \frac{\rho_{780} - \rho_{670}}{\rho_{780} + \rho_{670}}$$

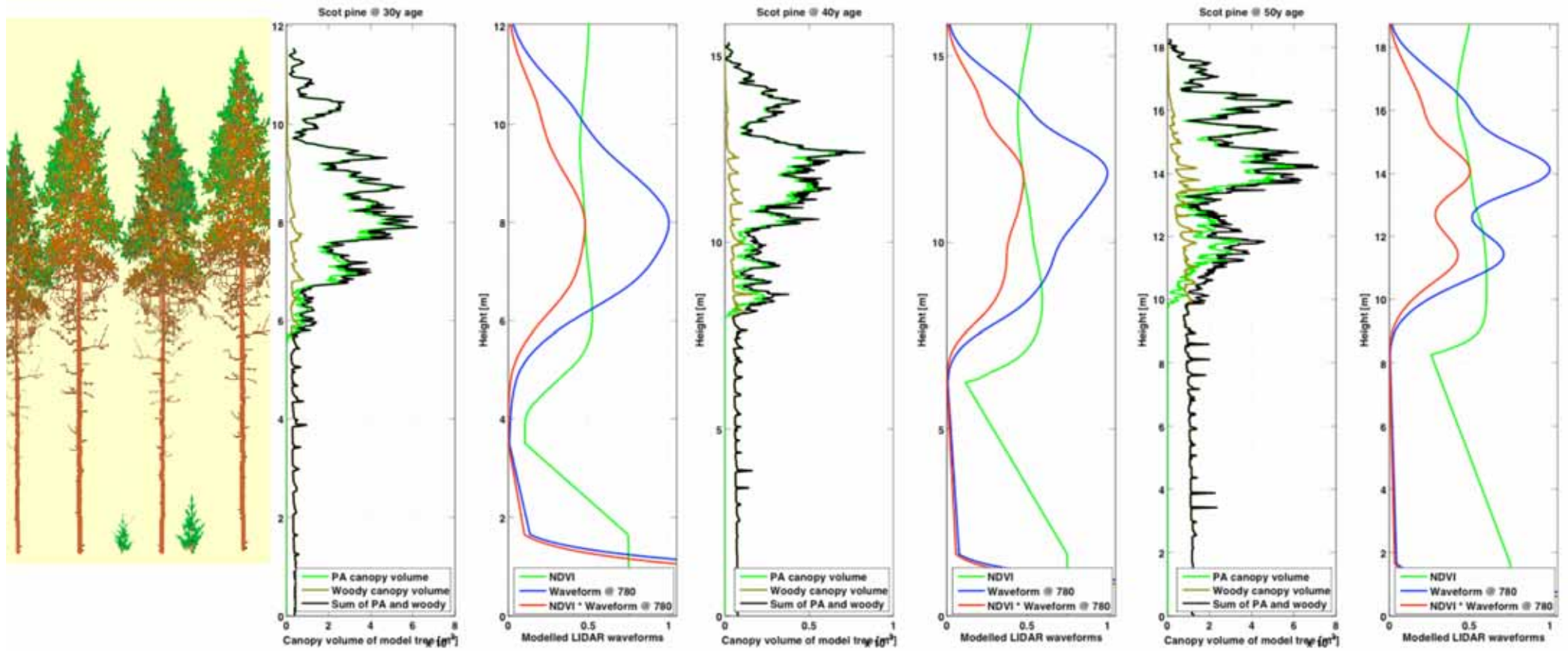
For Light Use Efficiency (LUE):

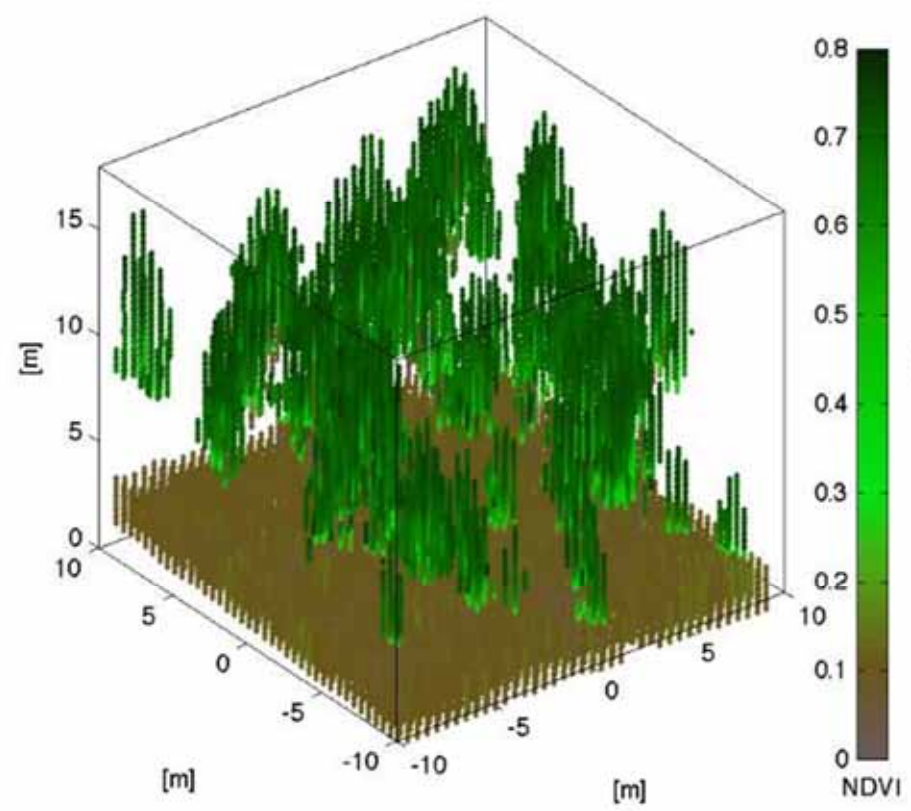
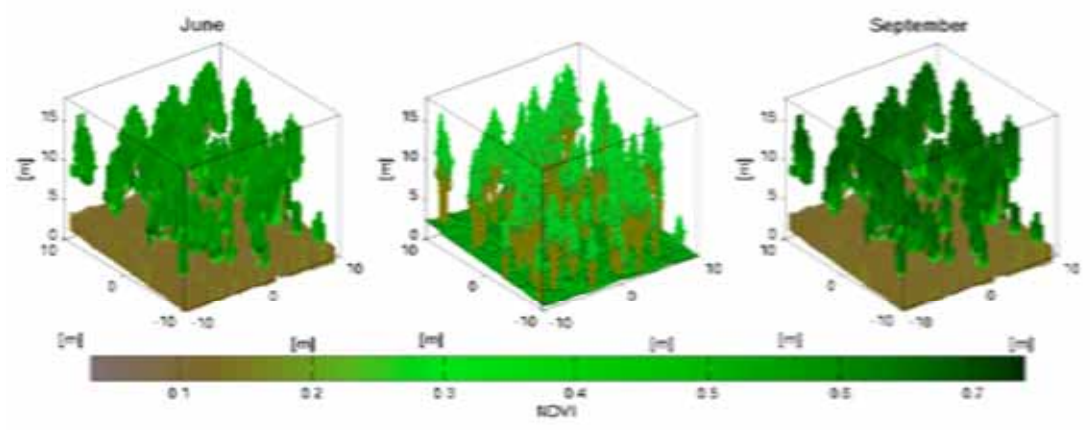
PRI : Photochemical Reflectance Index

$$PRI = \frac{\rho_{531} - \rho_{570}}{\rho_{531} + \rho_{570}}$$

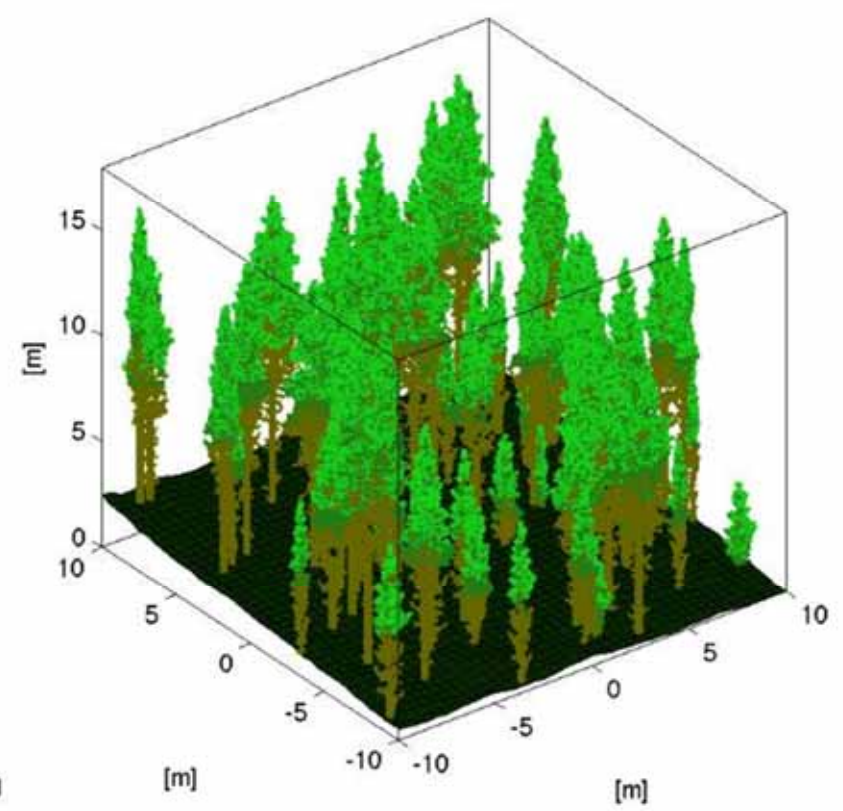
Model multispectral lidar waveform (Morsdorf et al, 2009)

- Tree architecture from Treegrow
- PROSPECT model to compute reflectance and transmission
- POVRAY to simulate lidar profile



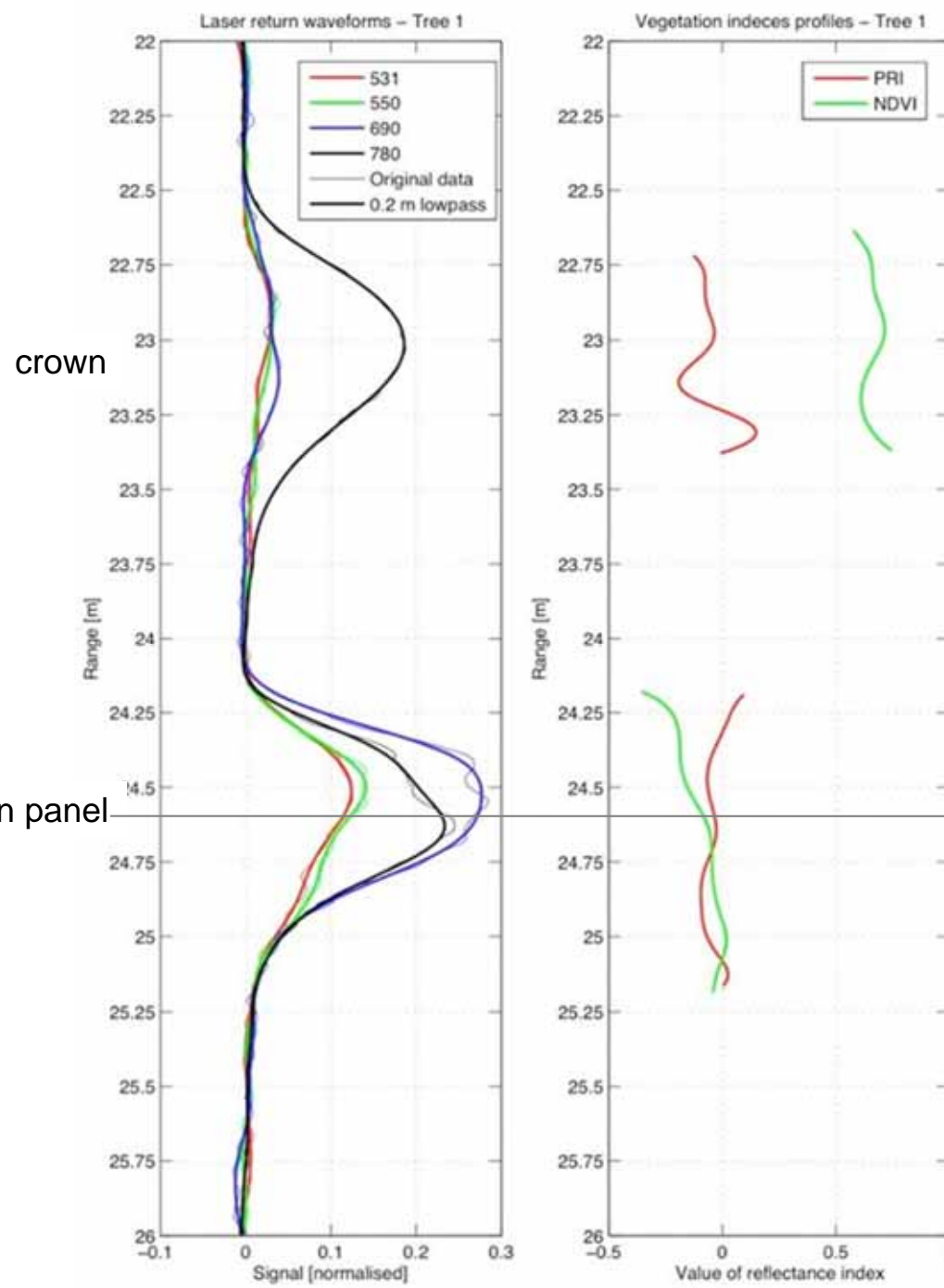


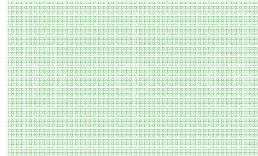
Simulated columnar returns



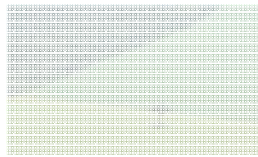
TreeGrow model

Woodhouse et al, GRSL, 2011

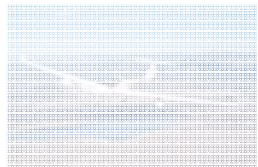




The value of forests



Radar

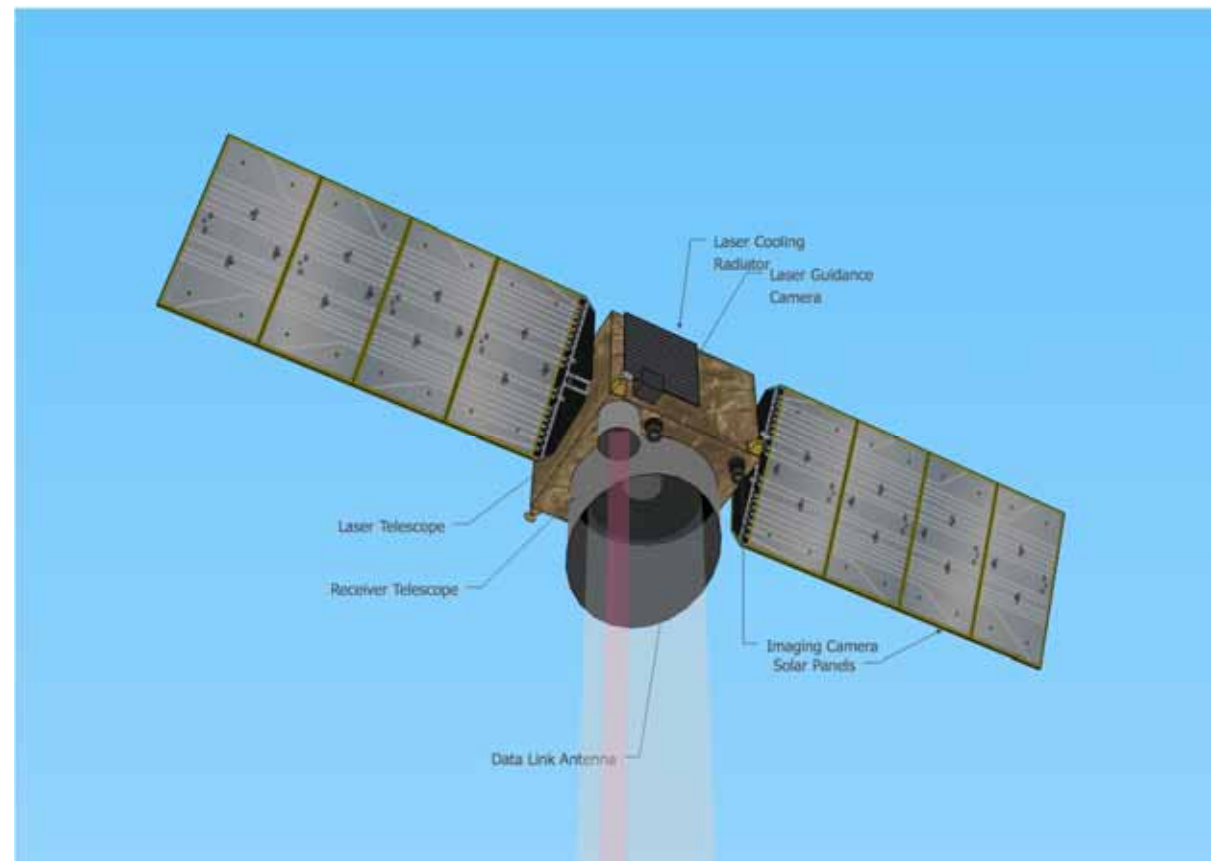


Multispectral Lidar (air)



Multispectral Lidar (space)

Spaceborne Multi-Spectral Canopy LiDAR “SpeCL”



The SpeCL mission concept

- Large footprint (30m) waveform lidar
- Max 1 km grid sampling
- Revisit same footprint every 90 days
- 4 wavelengths (actually, 6!)
 - PRI: 532nm (44mJ) and 570nm (44mJ)
 - NDVI: 660nm (70mJ) and 780nm (27mJ)
 - Additional: 1320nm and 1569nm
- Vertical information: structure and process



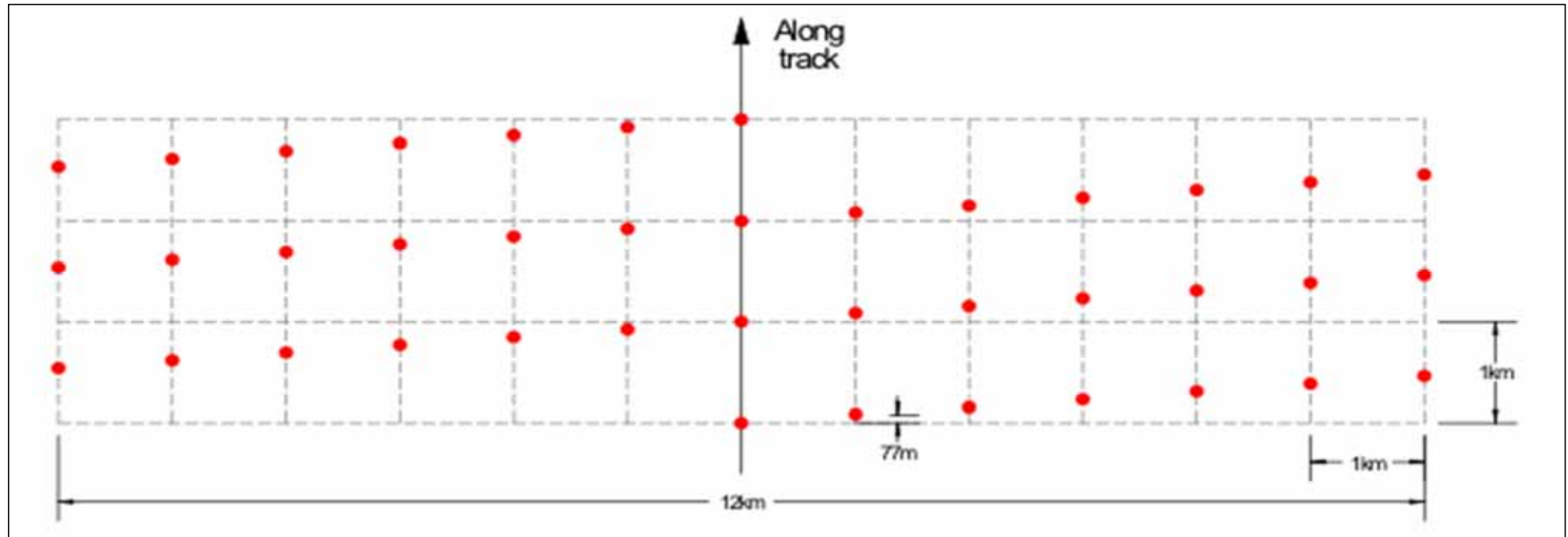
ALADIN on ADM-Aeolus

Source: diode-
pumped Nd:YAG
400mJ, 100ns,
1064nm, 100Hz

1.5m primary

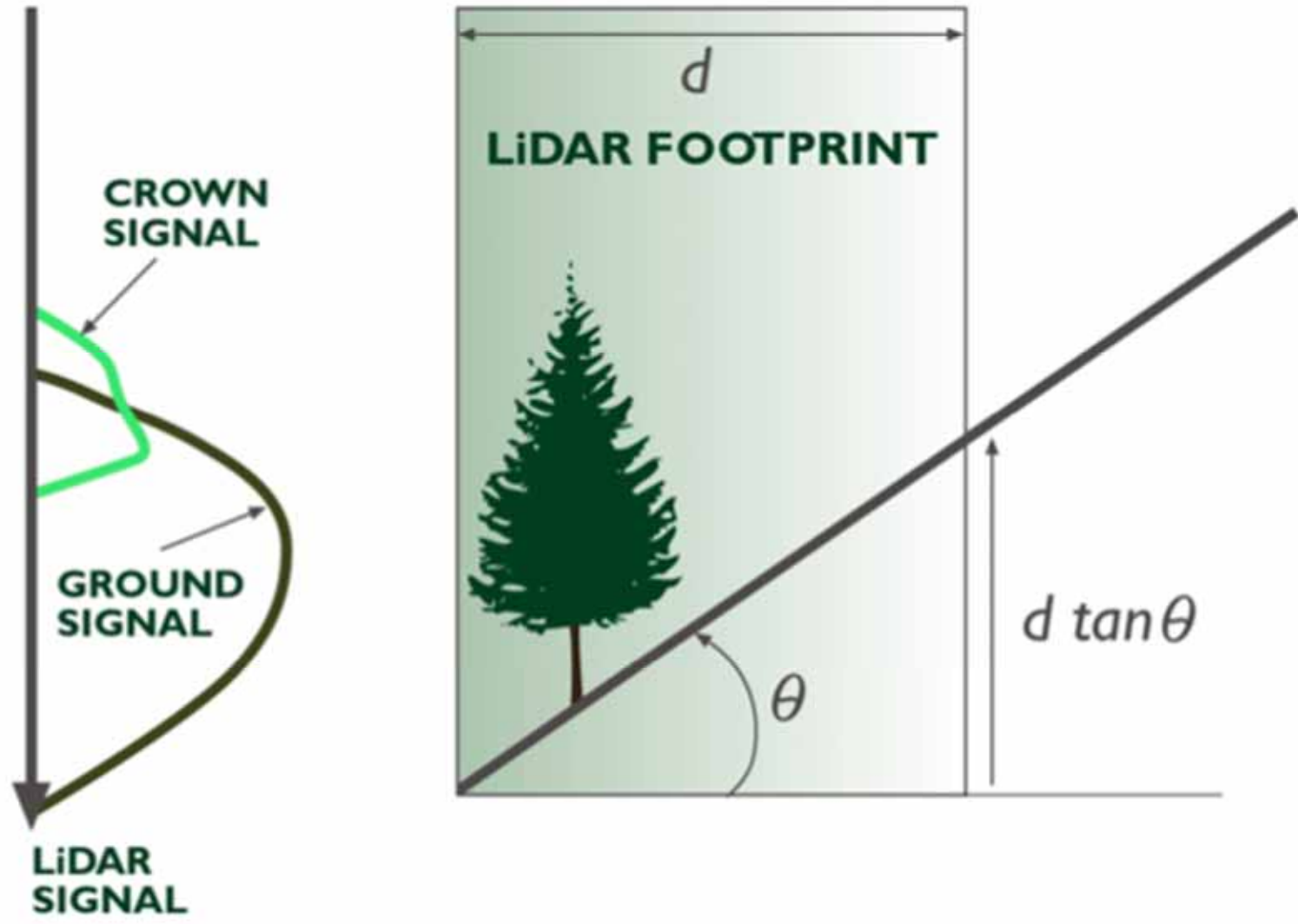
(will be) mature,
space based,
efficient pulse
emission for
532nm

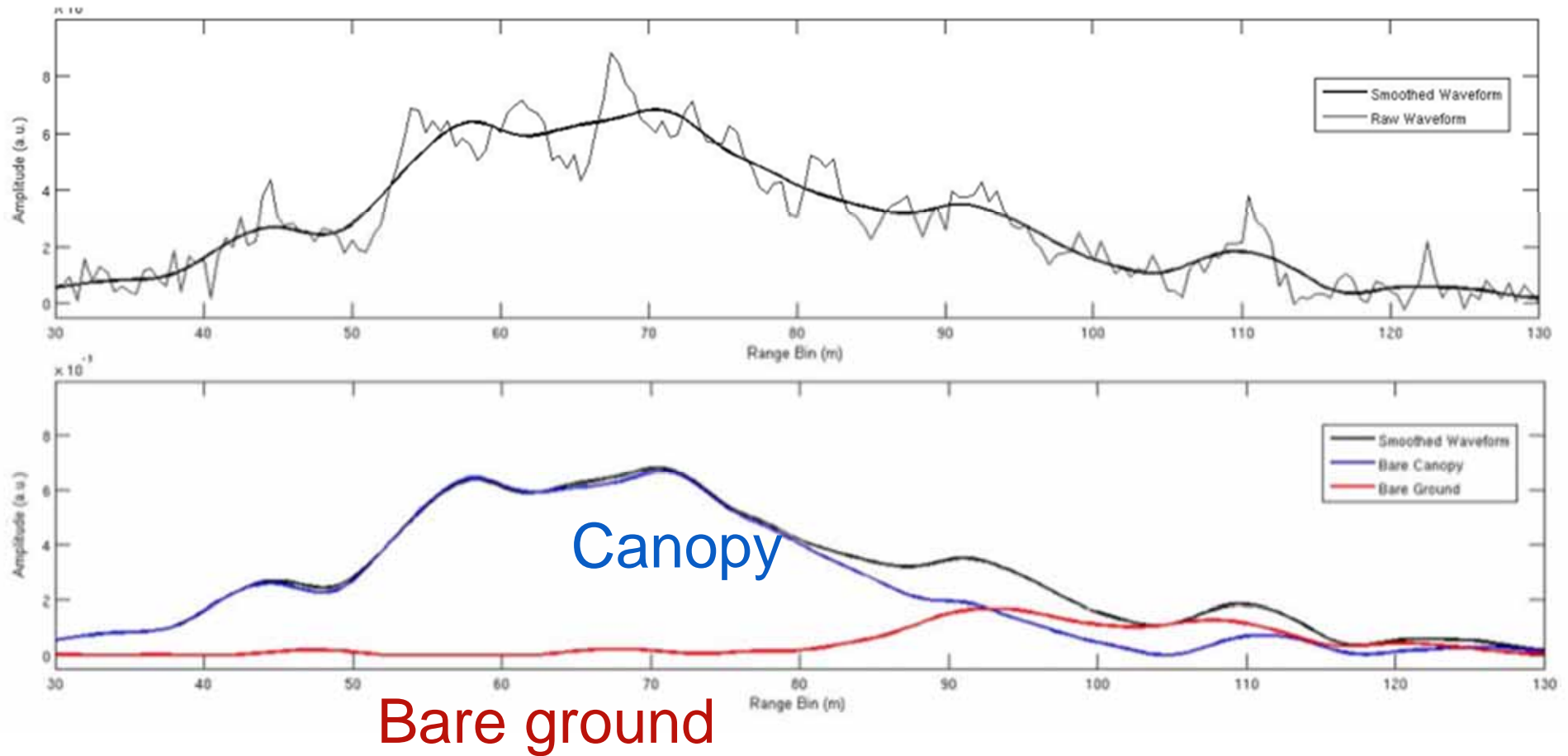
Scan Pattern



- Laser scan on 1km x 1km grid
- 13 shots per swath -> 12km wide swath
- Forward ground speed = 7.2 km/sec
- Time to travel 1km = 139 msec
- Required time between shots = $139/13 = 10.692$ msec
- Required PRF = 93.525 Hz
- Along track distance between shots = $7.2 \times 10.692 = 77$ m

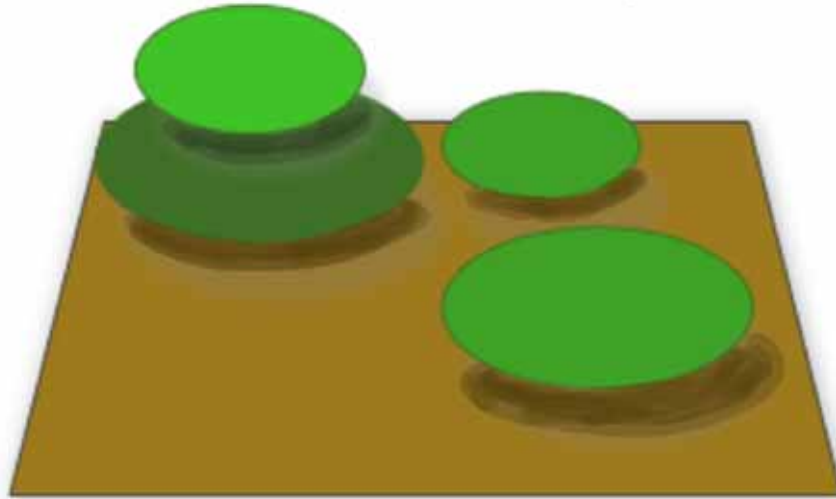
Using multispectral info for slope



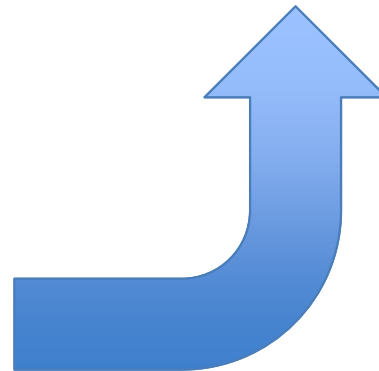
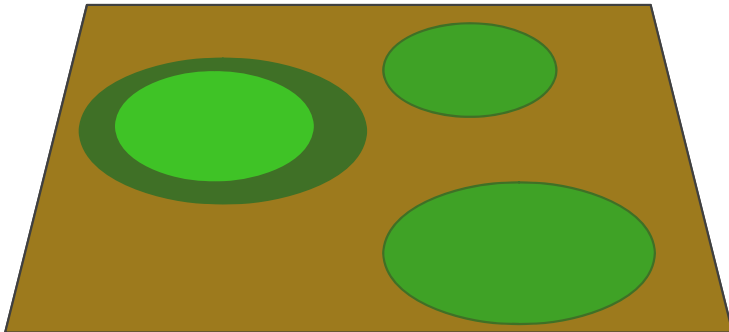
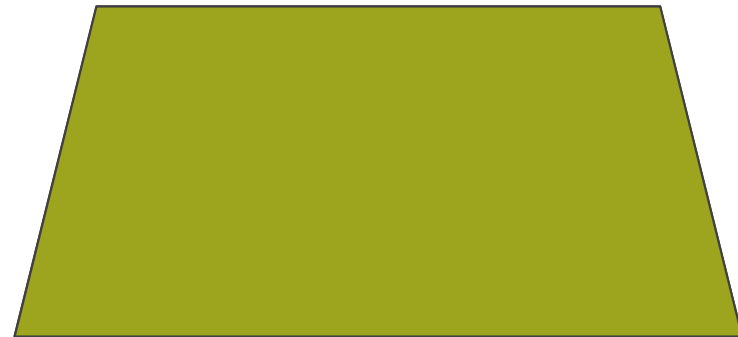


Mixed pixel calibration

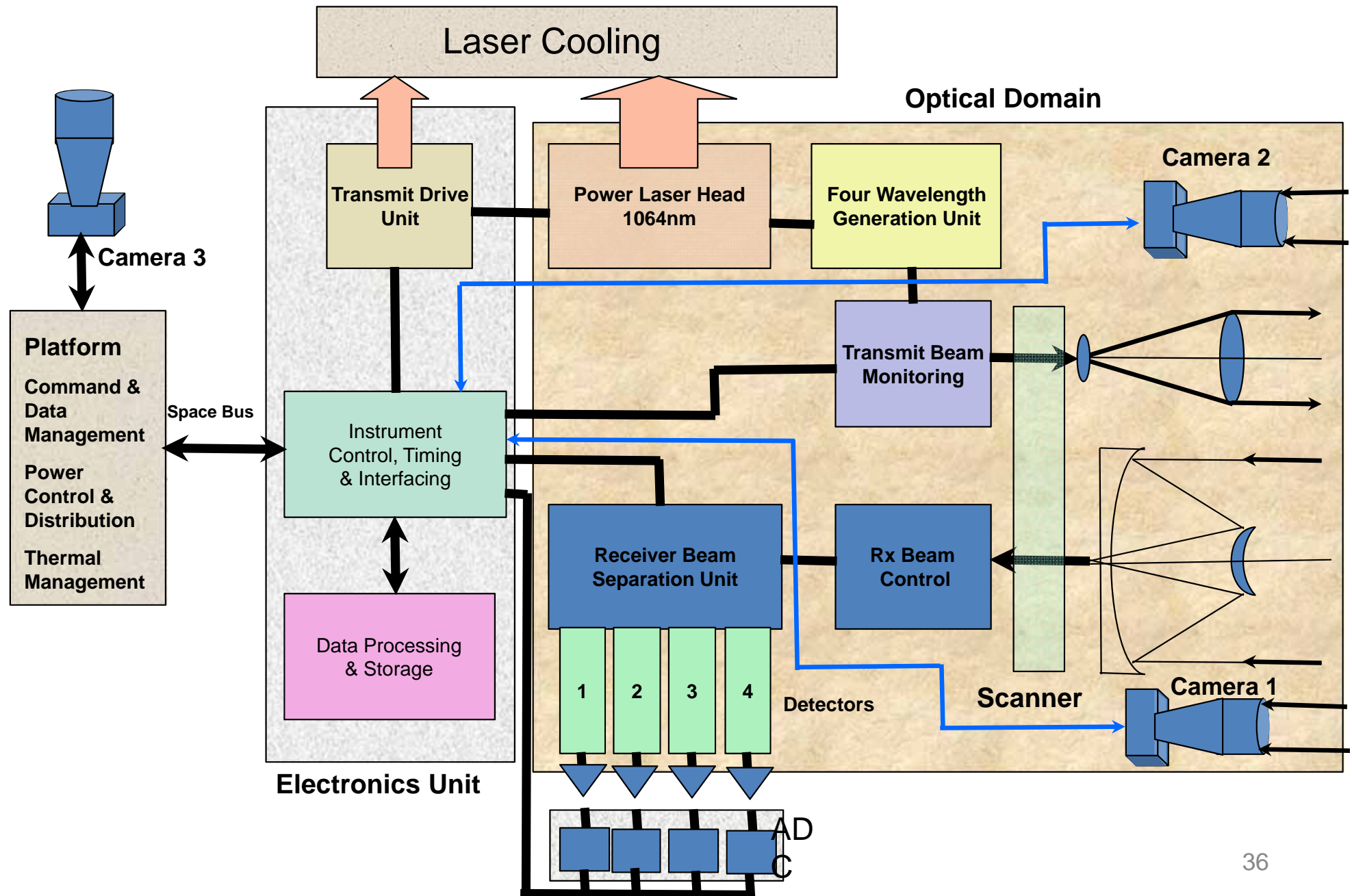
Multispectral lidar footprint



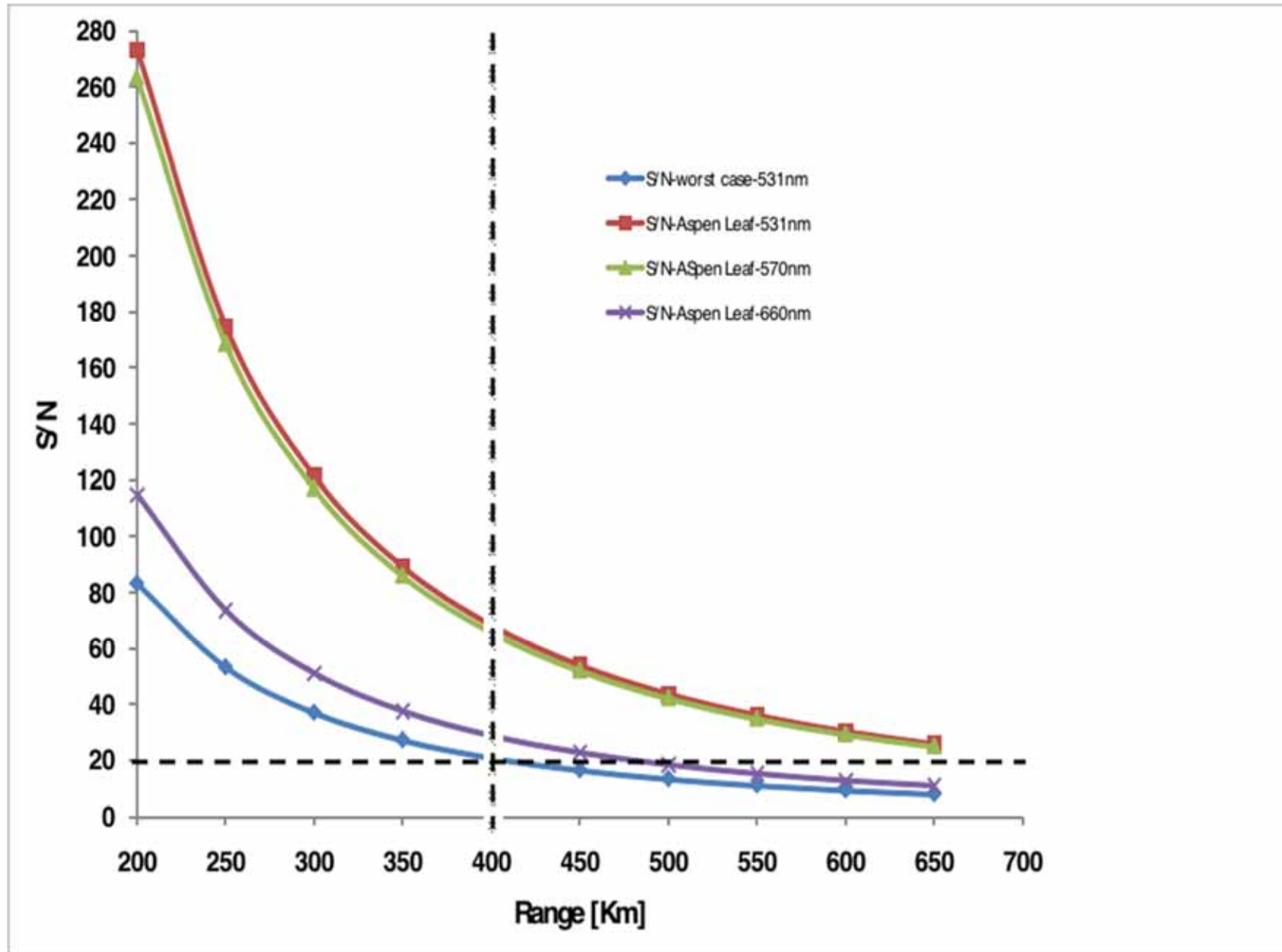
Multispectral "mixed" pixel



System Design Concept

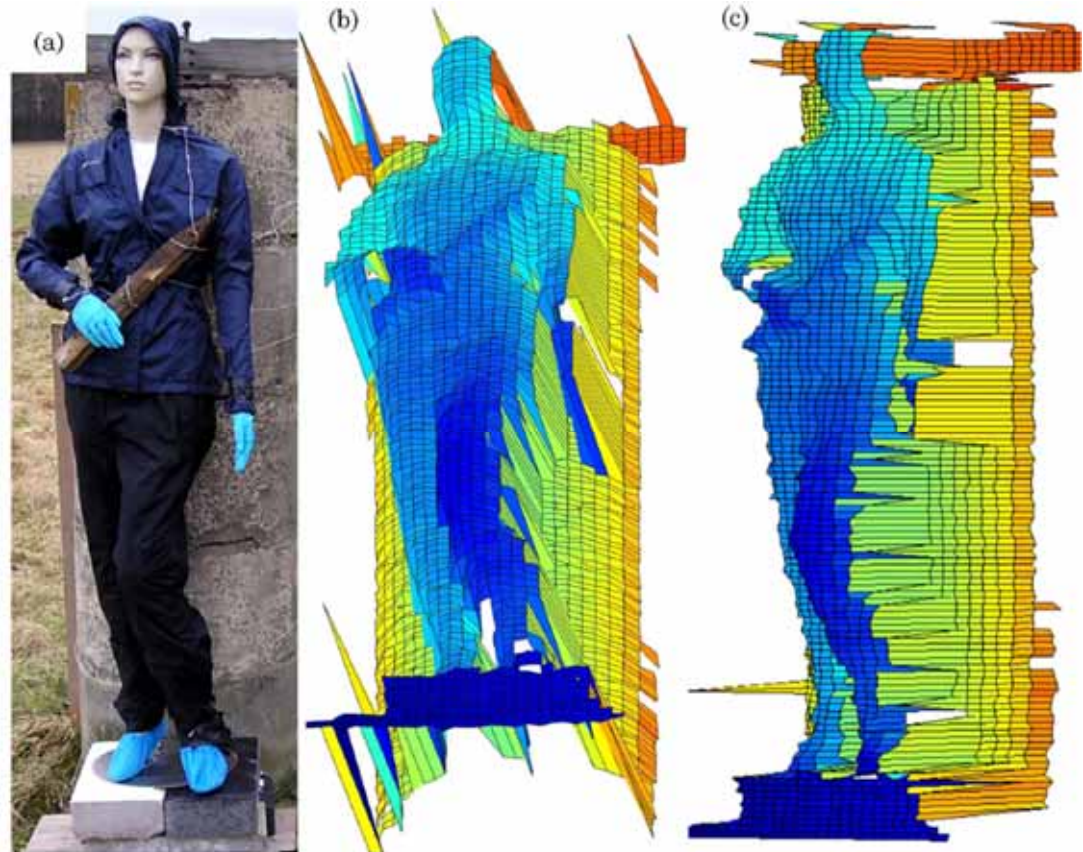


System Performance

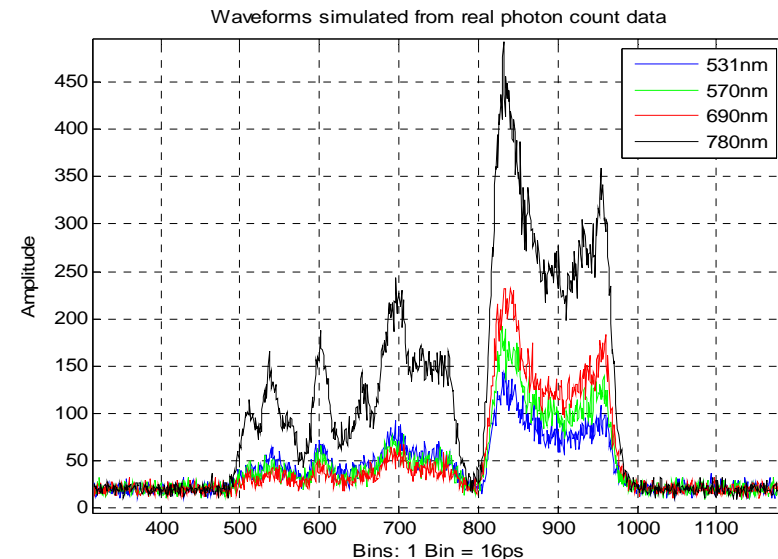
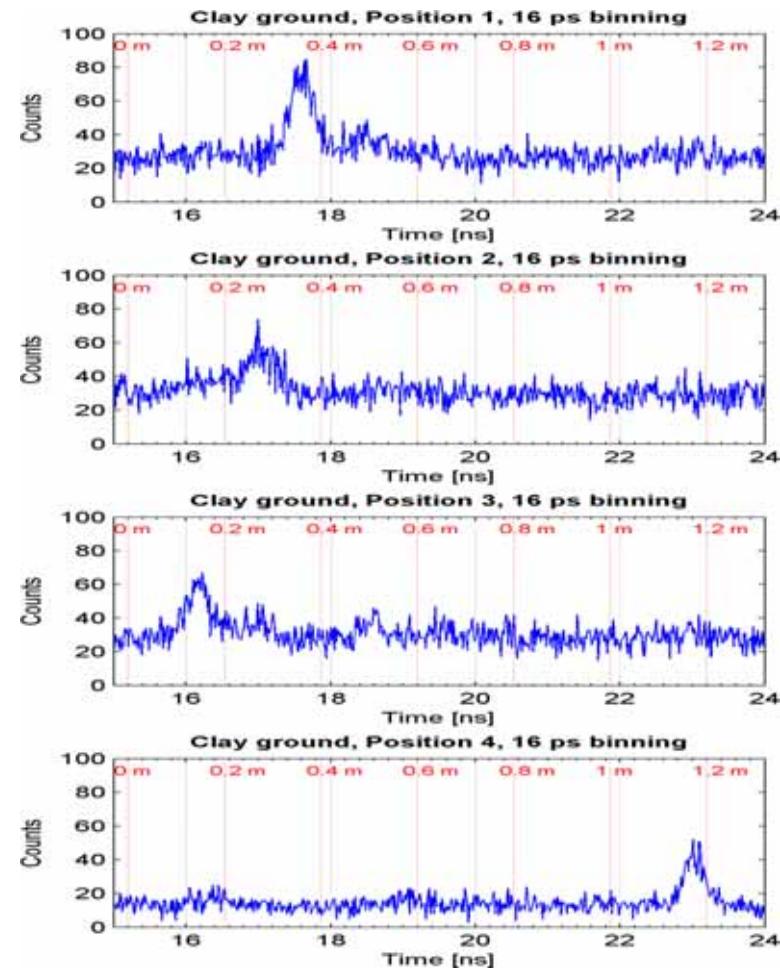


Next: hyperspectral lidar on a cherry picker

A long-range time-of-flight scanning sensor based on high-speed time-of-flight scanning



Four examples of measured LiDAR data at 850nm, from a pixel at the tree apex, going outwards towards the ground (bottom) with two intermediate pixels shown from points lower on the tree.



SpeCL: A multispectral Canopy Lidar

- proposal was submitted to ESA Earth Explorer - EE8 call, June, 2010
- Dec 2010 Earth Science Advisory Committee (ESAC) selected 2 missions for Phase A, and advised that 11 out of 33 have scientific potential - SpeCL was one of them.
- *“a very innovative mission concept, but at present it is technically very immature.”*
- *“ESAC recommends that studies be conducted ..
“(1) to develop and demonstrate the observation technique, supported by airborne campaigns with a prototype instrument; “*



Summary

- Multi- or Hyper-spectral LiDAR will come one day.
- It can work from space (if we need it).
- Final Report for SpeCL will be concluded in Oct, 2011 (please email me for a copy)
- Cherry-picker field trials to be completed March 2012 (please email me if you want to be kept in touch)
- Need to have an airborne demonstrator.

Thank you for your attention. 😊

Dr Iain H Woodhouse

i.h.woodhouse@ed.ac.uk

<http://forestplanet.wordpress.com>

www.carbomap.com



A new photon-counting lidar system for vegetation analysis

Jacqueline Rosette¹, Christopher Field², Ross Nelson¹, Phil DeCola², Bruce Cook¹

¹NASA Goddard Space Flight Center, Biospheric Sciences, Greenbelt, MD, 20771, USA.
jacqueline.rosette@nasa.gov; ross.f.nelson@nasa.gov; bruce.cook@nasa.gov

²Sigma Space, 4600 Forbes Blvd, Lanham, MD 20706, USA;
christopher.field@sigmaspace.com; phil.decola@sigmaspace.com

Abstract

This paper considers the potential of a new scanning photon-counting system for vegetation analysis. The 3D Mapper sensor was developed by Sigma Space Corporation and is being tested within NASA's Carbon Monitoring System (CMS) project (NASA, 2010). The sensor is able to map 60 km² per hour using less than 150 mW of 532 nm green light with about 30 cm between measurement points. While this area coverage rate is already several orders of magnitude higher than can be achieved by conventional lidar, substitution of higher power lasers would permit significantly higher mapping rates with the same resolution or much higher spatial resolution at the current rates. Data were collected for a test site to the west of Fredericksburg, Virginia, USA and demonstrated the capability with a low powered laser, of relatively high density data collection, and good penetration through the canopy, despite high canopy fractional cover and a hazy atmosphere at the time of flight. This preliminary study supports the potential of this emerging technology for vegetation analysis. Further research is required to develop algorithms to exploit the capabilities of such systems and to provide a greater understanding of the interactions with vegetated surfaces. Studies of this nature will inform future photon-counting satellite lidar sensors such as NASA's ICESat II, which is due for launch at the beginning of 2016.

Keywords: Photon-counting, green wavelength, ambient noise, signal detection algorithms, Carbon Monitoring System (CMS)

1. Introduction

1.1. NASA's Carbon Monitoring System project

This research is being carried out within the context of NASA's Carbon Monitoring System (CMS) initiative (NASA, 2010; Suárez *et al.*, submitted this edition). One component of CMS is the local-scale mapping of biomass using datasets which are expected to be readily available within the US (these include the US Geological Survey National Land Cover Data (NLCD) and lidar data). The aim is to determine a methodology for county-level analysis which could be extended and applied at regional, State or national scales. Since photon-counting lidar data offers prospects for future large area mapping for vegetation analysis, the potential of such sensors is being considered as part of this project. Further details of the CMS initiative are presented in Suárez *et al.*, submitted this edition.

1.2. Photon-counting lidar systems

The emerging technology of photon-counting lidar offers the potential for low energy expenditure and potential high altitude operation allowing extended laser lifetime and large area coverage. This newest type of lidar technology is currently generally operated at green wavelengths (532 nm). For some airborne systems, this is due to a greater efficiency of the detector and in the case of NASA's ICESat II, it is as a result of technical readiness for space flight.

Low laser energy output ensures eye safety of these instruments despite operating at a visible wavelength. A high pulse repetition rate and photon detection probability produces a high point density even whilst flying at greater altitudes whilst a narrow pulse duration (<1ns) allows photons to be located with greater vertical precision.

One significant factor resulting from the green wavelength is that photons returned from the emitted pulse cannot be distinguished from detected photons resulting from ambient noise. A small detector field of view and narrow optical band-pass filter are two important elements to reducing the background noise as much as possible. Much of the remaining background can be eliminated by coincidence filtering. Acquiring data at dusk or night would further reduce the background noise.

1.2.1. Micro-Altimeter

To date, few photon-counting systems have been developed and perhaps among the earliest of these was the Micro-Altimeter which flew several times in early 2001. It was a four channel, conical scanning instrument with a 7 to 20 mW micro-chip laser. The system produced profiles and terrain maps from 6 to 12 km including through heavy fog and under shallow bays (Degnan *et al.*, 2001). Some further examples of profiling and scanning photon-counting lidar systems, from both airborne and spaceborne platform are outlined below.

1.2.2. SIMPL

The Slope Imaging Multi-polarisation Photon-counting Lidar (SIMPL) is an example of an airborne small footprint photon-counting profiling lidar which operates at both 1064nm and 532nm wavelengths (Dabney *et al.*, 2010). A single pulse is emitted which is split into four beams, each with four channels for green and NIR wavelengths each of which at parallel and perpendicular polarisations. The two polarisations respectively identify photons which have been reflected from a single surface or which have undergone multiple scattering. The four beams are distanced approximately 5 metres apart, producing four profile 'slices' through the canopy. The laser repetition rate of 11.4kHz and an aircraft speed of 100m/second may be expected to produce 5-15 detected pulses per square metre.

Using SIMPL, Harding *et al.*, in press 2011, have explored the influence of lidar wavelength on the ability to determine standard waveform metrics which may be employed to predict biomass. By aggregating detected photons over a distance along the transect, the authors calculated a cumulative height distribution (such as that used for waveform or discrete return analysis). Height of median energy (HOME) and canopy cover metrics were compared and little difference was found between the two wavelengths, suggesting that lidars using 532nm could produce comparable biomass estimates to those obtained by current 1064nm systems.

1.2.3.ICESatII and MABEL

NASA's forthcoming ICESat II mission is due for launch in early 2016 (GSFC, 2011a). In contrast to ICESat I, its successor will carry a medium footprint, photon-counting profiling lidar operating at 532nm wavelength. This instrument is named ATLAS, the Advanced Topographic Laser Altimeter System.

The current planned configuration is for a single emitted pulse which is split into six beams, arranged as three adjacent pairs. Each pair will have a stronger and a weaker beam (100 μ J and 25 μ J respectively) which aims to address issues of detector dynamic range when alternating between bright and dark surfaces such as ice and water. A distance of 3.3km is anticipated between each pair and members of the pair will be separated by 90m. The high repetition rate of 10 kHz from an altitude of ~496km will produce overlapping footprints of 10m diameter which will be distanced at 0.7m intervals.

1-3 photons are anticipated to be detected per footprint and, although the spatial location of photons within the footprint will be unknown, the aggregation of returns along the ground tracks will allow a vertical profile to be created. Like its predecessor, the primary objective of ICESat II is not the retrieval of vegetation, one of its science objectives is measuring vegetation height as a basis for estimating large-scale biomass and biomass change (GSFC, 2011a). This new technology will offer a new perspective of the world and open opportunities for different approaches to global vegetation analysis.

Prospects for data collection in preparation of ICESat II are being tested using NASA's high-altitude simulator, the Multiple Altimeter Beam Experimental Lidar, MABEL (GSFC, 2011b). MABEL is a demonstrator instrument for the ICESat II mission, flying above the atmosphere at an altitude of 20km on NASA's ER-2 aircraft. It has been flown in December 2010 and again in March-April 2011 at different times of day, producing different levels of solar background which can be used to test signal detection algorithms. Data have been made available online by NASA at GSFC, 2011b.

1.2.4.Sigma Space 3D Mapper

For the study presented in this paper, the prototype 3D Mapper photon-counting, scanning lidar developed by Sigma Space Corporation, USA, is being tested to assess the potential of this sensor for vegetation analysis within the context of NASA's Carbon Monitoring System initiative (NASA, 2010).

The instrument measures approximately 60 km²/hour with 30 cm postings using about 150 mW of green light, emitting a 532nm, 0.7ns (700 picosecond) laser pulse at a repetition rate of 20 kHz. A 10x10 array of beamlets is produced on the ground and two time-of-flight cards, each with 50 channels, record the data collected. Photomultipliers were selected as the detectors for this system because of their very fast recovery time. The short system dead time, about 1.6 ns, means that each channel is armed and ready for the next event within 20 cm of a previous detection. This permits high resolution vertical mapping of forest canopies. The system uses 532 nm light because the detector has much higher quantum efficiency for green light than for the infrared. The green may have an additional advantage for vegetation measurements in that the lower reflectivity (relative to IR) significantly reduces multiple scattering and so improves the measurement fidelity.

Scanning patterns using this prototype instrument can be conical or near-linear. For a conical scanning pattern, the half-width angle is typically 9°, producing up to a 3-second difference between forward and backward views of the same location on the ground. For this study, a near-linear scan was used at a 45° angle to the flight path producing a swath width of approximately 300m.

2. Data collection

A test site at Fredericksburg, Virginia, USA was flown on 12th May, 2011, using the 3D Mapper photon-counting sensor. The data presented were acquired during a single pass flying at circa 280 km/hour at an altitude of 1km. This prototype system, in this mode, can capture data over 6,000 hectares per hour. Flight time was at 3pm, meaning that the sun angle and atmospheric haze observed during the flight produced challenging conditions to test the performance and capabilities of the instrument.

The site is located at a distance of 15 km to the west of Fredericksburg, and approximately 80 km SSW of Washington DC. The area contains predominately deciduous forest shown in lighter green in Figure 1 (below), of mixed height composition and density as well as residential and retail land use. The transect used for illustration of the sensor capabilities below crosses an area of evergreen (coniferous) forest with dense canopy cover (darker green, figure 1, below).



Figure 1. Test site near Fredericksburg, Virginia, USA. Above: GoogleMaps image with a 800m section outlined in red; Below: National Land Cover Data (US Geological Survey). Mid green represents deciduous forest; evergreen forest is indicated in darker green, light green is mixed forest, pale pink is developed/open space, and beige is shrub/scrub. Some misclassification can be observed which would have implications for biomass mapping using this dataset for stratification of the landscape.

3. Initial observations and results

The area illustrated within Figure 2 is 800 metres across and shows a horizontal breadth through the canopy of 10 metres. The tree canopy and ground are clearly visible against the solar background above and below. At times, haze prevented visibility to the ground surface, however, even under these conditions, adequate signal photons were received to enable the ground and canopy surfaces to be differentiated from background noise.

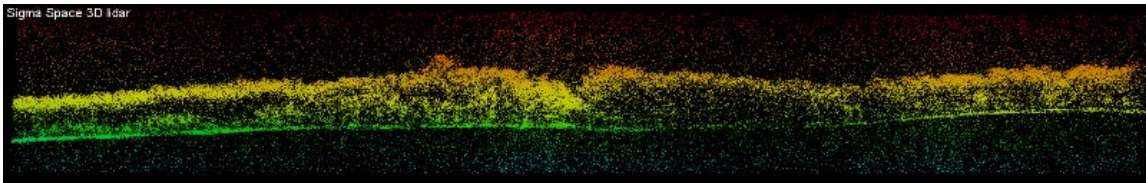


Figure 2. Illustration of vertical profile showing signal and noise above and below the intercepted surfaces.

Figure 3 shows a close-up view of the vegetation canopy. Emergent and suppressed crowns can be visually identified. This suggests the potential for methods to be developed to distinguish understorey vegetation beneath an upper canopy.

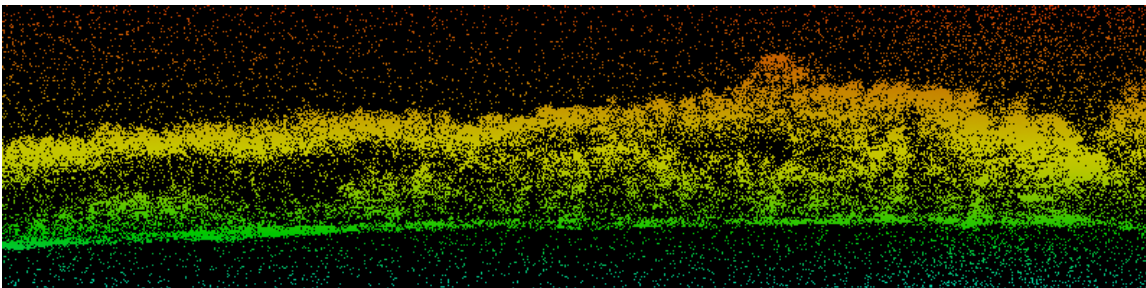


Figure 3. Observation of understorey vegetation beneath a dense canopy.

4. Next steps

4.1. Research flights

A research-focused photon-counting lidar campaign is planned during Summer 2011 over the Jug Bay Wetlands Sanctuary (Jug_Bay, 2011). This is an area largely of mixed broadleaf forest, located approximately 32km to the southeast of Washington DC. Field data were collected between 2003 and 2005 by a team of volunteers of the Wetland Sanctuary, for 300 10x10m plots arranged at 100m intervals on the UTM 18N grid. Discrete return lidar data were also acquired across this area of Maryland State in 2004. Additionally, some further field measurements were taken in 2011 as part of NASA's CMS project (NASA, 2010; Suárez *et al.*, submitted this edition).

This will enable a comparison of photon-counting data with field measurements as well as conventional discrete return lidar data. The time difference in lidar data collection along with repeat field measurements would allow growth to be observed and will assist in the assessment of biomass mapping as part of the CMS project.

5. Discussion

5.1. Capabilities and potential applications

Initial tests suggest that promising results may be obtained from small footprint photon-counting sensors for the generation of vegetation products. The high density of the point cloud which is produced, in excess of that which is typically collected by discrete return airborne lidar data, aims to improve the characterisation of vegetation canopies and offers the opportunity for established analysis techniques to be applied to this new technology.

The potential to maintain a high detection rate of photons from a high altitude, could reduce data acquisition costs and permit more economical inventory and mapping of broad areas. The observation of understorey vegetation (Hill, 2007), which is difficult to achieve using more conventional lidar systems, could improve the capabilities of identifying invasive species and accounting for over-reached trees in canopy metric-based statistical equations for biomass mapping. This also presents the opportunity for improvements in the detection of small and suppressed trees using individual tree-level lidar inventories (Suárez, 2010).

A greater number of returned photons are observed for more reflective surfaces (e.g. painted lines on roads) and so changes to photon density could offer the possibility of direct crown-width detection in open canopies by observations of shadowing effects on the ground. Additionally, data acquisition during leaf-off conditions, and the improved vertical precision anticipated from the short duration pulse, may reveal potential applications for forestry such as timber quality (number and location of branches and tree architecture) seen in figure 4.

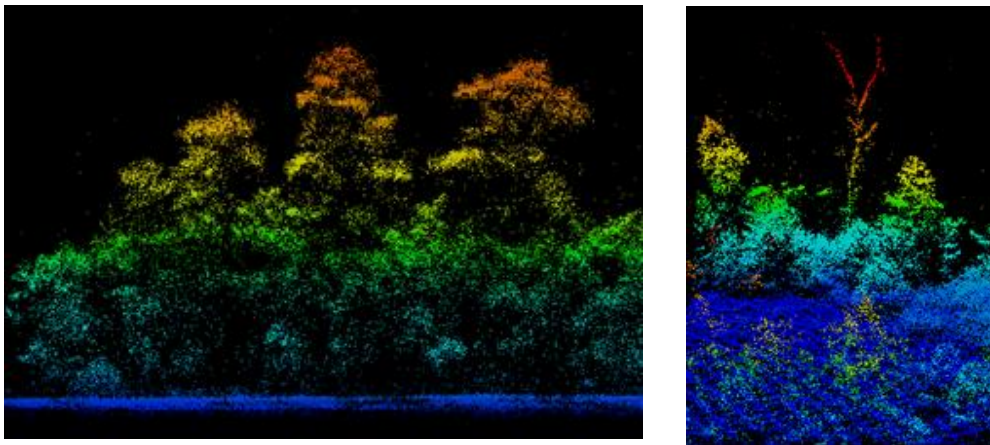


Figure 4. Canopy profile and branch architecture from an earlier test flight at the Smithsonian Environmental Research Center (SERC), Maryland, USA using the Sigma Space prototype photon-counting sensor.

5.2. Challenges of green wavelength photon-counting

The principal challenge of emerging photon-counting systems at green wavelengths is the ability to distinguish signal from noise. Photons returned from an intercepted surface cannot be differentiated from those originating from ambient noise and, whilst for planar surfaces such as roofs and bare earth, these can be classified reasonably easily, vegetation will require more careful consideration.

This is likely to be particularly important when identifying rough-surfaced transition zones such as the top of canopy. Additionally, noise photons cannot be discriminated within the canopy region and therefore adaptations to percentile-based methods of statistical analysis will be required, which take into account the random but uniformly-distributed photons beyond the true intercepted surfaces in order to adjust for spatially-varying ambient noise.

5.3. Prospects for photon-counting lidar

NASA's ICESat II mission is planned to be a green-wavelength, photon-counting, profiling system (GSFC, 2011a). As the only currently-planned satellite lidar sensor, the subject of ambient noise and the development of canopy/ ground-finding algorithms are issues which will need to be addressed. Data from MABEL (GSFC, 2011b) gathered at an altitude of ~20,000 km aboard the NASA ER-2 aircraft are providing an opportunity to develop algorithms in advance of the launch of the satellite. Possible future high-altitude airborne scanning photon-counting systems producing wider swaths would improve capabilities to map at landscape scales. For example a next generation version of the 3D Mapper flying at an altitude of 5-7000 km could cover circa 25,000 hectares per hour.

Research using the FLIGHT and DART lidar simulation models (North, 1996; North *et al.*, 2010) aims to further inform the understanding of photon-counting lidar and their interactions with vegetation canopies. As a newly-emerging technology this will be subject to ongoing research and evaluation to extract the full potential and to develop new methods to process data which aim to offer new capabilities.

6. Conclusion

This paper has illustrated preliminary results of a test flight at Fredericksburg, Maryland, USA, using the Sigma Space prototype 3D Mapper photon-counting system. Initial findings suggest great promise can be offered by such systems for vegetation analysis which may improve current capabilities offered by discrete return lidar systems. Further work is required to address issues of ambient noise and to determine characteristics of photon interactions with the canopy. A lidar campaign is planned for Summer 2011 over the Jug Bay Wetlands Sanctuary which will permit more in-depth analysis of the prospects of green-wavelength, photon-counting lidar for forest assessment.

Acknowledgements

This work forms part of the NASA Carbon Monitoring System (CMS) initiative funded by the US Congress. Partners within the local biomass component of this project include NASA Goddard Space Flight Center, University of Maryland College Park, Sigma Space Corporation, US Forest Service, SGT and Geodigital.

The 3D Mapper photon-counting lidar sensor used in this study was developed and flown by Sigma Space Corporation, Lanham, MD, USA. Volunteers and staff at the Jug Bay Wetlands Sanctuary are acknowledged for the collection of field data and for generously making this available for use within the CMS project. The research has also been supported by resources from Swansea University and Forest Research, UK.

References

- Dabney, P., Harding, D., Abshire, J., Huss, T., Jodor, G., Machan, R., Marzouk, J., Rush, K., Seas, A., Shuman, C., Sun, X., Valett, S., Vasilyev, A., Yu, A. and Zheng, Y., 2010. The Slope Imaging Multi-polarization Photon-counting Lidar: development and performance results, IEEE International Geoscience and Remote Sensing Symposium, 11686732, DOI 10.1109/IGARSS.2010.5650862, pp. 253-256.
- Degnan, J., McGarry, J., Zagwodzki, T., Dabney, P., Geiger, J., Chabot, R., Steggerda, C., Marzouk, J. and A., C., 2001. Design and performance of an airborne multikilohertz, photon-counting microlaser altimeter. International Archives of Photogrammetry and Remote Sensing, XXXIV - 3/ W4: 9-16. Annapolis, MD, 22-14 Oct. 2001.
- GSFC, 2011a. ICESat-2. Available online at: <http://icesat.gsfc.nasa.gov/icesat2/index.php>.
- GSFC, 2011b. ICESat 2 - MABEL simulator. Available online at: <http://icesat.gsfc.nasa.gov/icesat2/mabel.php>.
- Harding, D., Dabney, P. and Valett, S., in press 2011. Polarimetric, two-color, photon-counting laser altimeter measurements of forest canopy structure, SPIE Proceedings LIDAR and RADAR 2011, Nanjing, China.
- Hill, R.A., 2007. Going Undercover: Mapping Woodland Understorey from Leaf-On and Leaf-Off Lidar Data. In: P. Rönholm, H. Hyyppä and J. Hyyppä (Editors), Proceedings of the ISPRS Workshop 'Laser Scanning 2007 and SilviLaser 2007'. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Espoo, Finland.
- Jug_Bay, 2011. Jug Bay Wetlands Sanctuary. Available online at: <http://www.jugbay.org/> Accessed June 2011.
- NASA, 2010. NASA Carbon Monitoring System Initiative. Available online at: <http://cce.nasa.gov/cce/cms/index.html>.
- North, P.R.J., 1996. Three-Dimensional Forest Light Interaction Model Using a Monte Carlo Method. IEEE Transactions on Geoscience and Remote Sensing, 34(4): 946-956.
- North, P.R.J., Rosette, J.A.B., Suárez, J.C. and Los, S.O., 2010. A Monte Carlo radiative transfer model of satellite waveform lidar. International Journal of Remote Sensing 31(5): 1343-1358.
- Suárez, J., Pinto, N., Rosette, J., Fatoyinbo, T.E., Nelson, R., Dubayah, R. and Cook, B., submitted this edition. A Carbon Monitoring System (CMS) for US counties using Airborne LiDAR Data.
- Suárez, J.C., 2010. An Analysis of the Consequences of Stand Variability in Sitka Spruce Plantations in Britain using a combination of airborne LiDAR analysis and models, PhD, University of Sheffield, Sheffield, 295 pp.

Error assessment and mitigation for hyper-temporal UAV-borne LiDAR surveys of forest inventory.

Luke Wallace¹, Arko Lucieer¹, Darren Turner¹ & Christopher Watson¹

¹School of Geography and Environmental Studies, University of Tasmania
Luke.Wallace@utas.edu.au

Abstract

Remotely sensed LiDAR data has become an important tool in the management of modern forest inventories. Monitoring the high frequency changes within forests with this data has been restricted by the cost and intermittent nature of LiDAR surveys. The use of Unmanned Aerial Vehicles (UAVs) as a remote sensing platform is a rapidly developing field and is capable of allowing highly dynamic environmental changes to be monitored. As such recent studies presented in the literature highlight the potential of UAV systems for forest monitoring. This study further investigates the potential of UAVs by examining the achievable accuracy of a newly developed UAV-borne LiDAR system in comparison to a traditional full scale system. The major contributions to the error budget of a UAV-borne LiDAR system are constrained through the use of a novel UAV specific processing workflow. Central to this workflow is the fusion of observations from a low cost Inertial Measurement Unit, a GPS receiver and a high definition video camera with a Sigma-Point Kalman Smoother allowing for highly accurate estimates of orientation. We found that using this workflow and under certain flying conditions accuracies similar to a modern full-scale system are achievable from this low-cost platform.

Key Words: Unmanned Aerial Vehicles, UAV, LiDAR, Accuracy, Error Propagation.

1. Introduction

Airborne LiDAR remote sensing has become an important tool in the management of modern forest inventories (Hyypä et al. 2008). Ongoing research into the processing and analysis of LiDAR data has allowed for the development of an extensive array of LiDAR derived data products from which a wide range of forest metrics can be derived (Akay et al. 2009). Stand metrics and tree-level statistics derived from LiDAR have provided forest managers with significantly richer information about their forests (Lim et al. 2003; Morsdorf et al. 2009). It is however evident that the full potential of LiDAR technology for forest measurement and management is yet to be reached. Prohibitive factors such as high survey costs and short flying seasons, in many areas, have meant that assessing factors such as forest health, defoliation and rate of canopy closure are not feasible from the current intermittent LiDAR surveys utilised by forest managers.

Recently, improvements in small scale positioning technology have enabled the use of Unmanned Aerial Vehicles (UAVs) as a remote sensing platform offering a distinctive combination of very high resolution data capture at a significantly lower survey cost to traditional platforms. Jaakkola et al. (2010) provided the first example of the potential of this technology for use within the forest industry. By deploying a rotor wing UAV equipped with a number of positioning sensors, in combination with two on-board LiDAR sensors, Jaakola et al. (2010) produced high-resolution data sets capable of individual tree level mapping. Both Jaakkola et al. (2010), and more recently, Lin et al. (2011), have shown that due to the improved density of a UAV captured LiDAR point cloud several metrics (in particularly individual tree

heights) can be measured at a finer scale and with higher precision using already developed processing algorithms when compared to traditional LiDAR platforms. Because of their high spatial and temporal resolution, together with low operation costs, UAVs can provide a more targeted approach to forest monitoring and allow for the use of multi-temporal surveys such as forest health and canopy closure monitoring. Studies such as Jakkola et al. (2010) and Lin et al. (2011) suggest that the combination of low cost, high resolution data capture, UAV platforms are likely to be the next tool of choice for optimising detailed small area surveys within forests.

Current research into the use of UAVs as a 3D data-capture platform includes application specific use in a variety of different fields ranging including for agriculture crop monitoring (Hunt Jr et al. 2010; Berni et al. 2009) to archaeology surveys (Eisenbeiss and Zhang 2006) for example. Despite significant developments into the use of UAVs for 3D mapping, a rigorous analysis of the error structure present within a UAV LiDAR based platform, and how these errors propagate into the final 3D measurements has yet to be undertaken. Such an analysis of error is necessary for use in forestry due to the use of UAV platforms for multi-temporal surveys and the need to distinguish small scale change from error. This paper presents an analysis of the propagation of error based on the stochastic error model of a UAV-borne LiDAR system developed at the University of Tasmania using lightweight, low-cost sensors. In this analysis, we make use of the well-known error propagation techniques used for full-scale traditional LiDAR systems to determine the achievable accuracy under a standard LiDAR processing algorithm and highlight the major sources of error. One of the novel contributions of this paper is in the development of an accurate position and attitude determination framework based on the use of a low cost Inertial Measurement Unit, High Definition video and a Sigma-Point Kalman Smoothing (SPKS).

2. Methods

2.1 Equipment

The low-cost multi-rotor UAV (Droidworx/Mikrokoopter AD-8) currently under development by the TerraLuma research group at the University of Tasmania will be used for this study (Figure 1). In comparison to other UAV platforms multi-rotor UAVs offer increased stability and decreased vibration making it the ideal UAV for LiDAR mapping. However, the main drawback of multi-rotor UAVs is their limited payload. In the case of the TerraLuma UAV this is 2.8 kg, which when the primary sensor (Ibeo LUX Automotive Laser Scanner, 1 kg), batteries and logging equipment are taken into account allows for only a minimal payload for position and attitude sensors. Based on these requirements (and a desire to minimise cost) a lightweight sensor suite has been designed consisting of a Microstrain 3DM-GX2 MEMs based Inertial Measurement Unit (IMU) (50 g), a Novatel OEMV-1DF dual-frequency GPS receiver (21.5g + 113 g antenna) and a ContourGPS digital video camera (150 g).

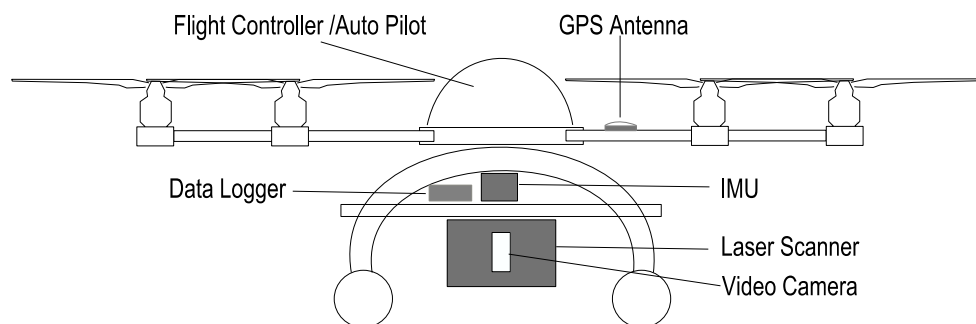


Figure 1. Schematic of the TerraLuma UAV remote sensing platform under development at the University of Tasmania.

The sensor payload is mounted on the UAV through a rigid sensor framework designed such that the lever arm offsets between the LiDAR, IMU, GPS and video camera are minimised and constant. The frame also allows for an adequate sky-view for the accurate operation of both the GPS antenna and GPS enabled video camera. The properties of each sensor have been rigorously tested in order to independently determine their standalone accuracy values under flight conditions. The results from these tests are displayed in Table 1 along with a brief explanation of the methods used to determine the estimates of instantaneous accuracy. Each sensor has been chosen to minimise both cost and weight while providing sufficient resolution and accuracy. All data logging and time synchronisation is performed using an on-board miniaturised computer (Gumstix Verdex pro). All other processing is completed offline.

Table 1: Description of the sensors mounted on-board the TerraLuma UAV and the method used to determine the accuracy of the sensor.

Equipment	Description and Accuracy Determination	Data Rate (Hz)	Standalone Accuracy
Dual-Frequency GPS receiver	Provides 3D positional observation of the helicopter operating under typical short baseline (<3 km) configurations. The GPS receiver has been bench marked against a geodetic grade GPS receiver.	20	Horizontal +/- 0.03 m Vertical +/- 0.05 m
Microstrain 3DM-GX2 IMU	A light weight - IMU consisting of MEMs based accelerometer, gyroscope and magnetometer triads. An independent calibration procedure following the methods of Zhang et al. (2010) has been performed to verify the calibration and accuracy of the sensor.	100	Orientation +/- 2.0° (+ bias instabilities and noise)
ContourGPS HD video camera	A high definition video camera equipped with a GPS antenna allowing accurate time synchronisation with the other sensors. The video camera has been calibrated and lens distortions removed using the procedure described by Bouquet (2010). The accuracy of the orientation as determined by video observation has been quantified using dense ground control as reference.	30	Orientation +/- 0.2 - 0.5°
Ibeo LUX laser Scanner	The Ibeo LUX sensor measures points in four scanning layers and in doing so can record up to 22000 returns/sec. The scanner has a measurement range of up to 200 m. The beam divergence of the Ibeo LUX laser scanner is 0.08° horizontally and 1.6° vertically.	880 (per scan layer)	Range +/-0.1 m

2.2 Airborne LiDAR Error Propagation

The calculation of the ground coordinates from the UAV-borne LiDAR system observations follow the same methodology as used for full scale traditional platforms. Coordinates of points that reflect the outbound laser pulse can be calculated directly from the range measurement from the LiDAR sensor, combined with data from the Positioning and Orientation System (POS) on board the UAV using the well known “LiDAR equation” (Eqn 1) (Baltsavias, 1999).

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + R_b^m [R_s^b r^s + a^b] \quad (1)$$

Where:

$[X \ Y \ Z]^T$	is the position vector as measured by the POS system expressed in the Earth Centred, Earth Fixed (ECEF) cartesian frame;
R_b^m	is the attitude matrix as measured by the POS and parameterized by the pitch, roll and yaw angles;
R_s^b	is the boresight matrix describing the angular offset between the body frame and the LiDAR frame;
r^s	is the observation vector from the Ibeo LUX system and consists of a range observation as well as an encoder angle; and
a^b	is the lever arm offset between the origin of the POS frame and the LiDAR frame.

There has been significant and ongoing research into the various factors that affect the accuracy of coordinates derived from a LiDAR system (Schaer et al., 2007). Individual LiDAR systems will also contain unique factors that affect the overall error (see May & Toth, 2007 and Morin, 2002 for an overview). However, these errors can be summarised into 17 error components which will occur in every system. These error components and can be described as:

- 3 errors existing in the measurement of the absolute position (σ_x , σ_y and σ_z);
- 3 errors existing in the measurement of aircraft orientation (σ_ω , σ_ϕ and σ_κ);
- 6 errors caused by the inaccurate calibration of the system affecting the boresight angles (σ_{ω_b} , σ_{ϕ_b} and σ_{κ_b}) and lever arm offset (σ_{x_L} , σ_{y_L} and σ_{z_L});
- 3 Internal LiDAR System errors occur in measurements of range (σ_r) and the two encoder angles (σ_β and σ_γ) measured from the UAV; and
- 2 errors due to divergence of the laser beam which propagate in the horizontal direction (σ_{B_h}) and elevation angle measurements within the laser scanner reference frame (σ_{B_e}). These errors will be modelled as one quarter of the quoted beam divergence of the laser scanner following Lichti and Gordon (2004) and Glennie (2007).

These error components can be propagated through the functional model of the LiDAR system equation enabling the magnitude of the error in the final coordinates of a point to be determined. Propagation for an individual LiDAR strike can be performed by linearising equation 1, through the truncation of the Taylor series expansion after the 1st term and assuming that each of the error sources are uncorrelated (Schaer et al., 2007). This enables the determination of the 3x3 point covariance matrix C_{xyz} , using equation 2 as follows:

$$C_{xyz} = \begin{bmatrix} C_x & C_{xy} & C_{xz} \\ C_{yx} & C_y & C_{yz} \\ C_{zx} & C_{zy} & C_z \end{bmatrix} = F_{ll} C F_{ll}^T \quad (2)$$

Where:

F_{ll}	is the Jacobian matrix of the linearised functional model; and
C	is the stochastic model given by a diagonal matrix containing the magnitude of the 17 summarised error sources.

This covariance matrix can be used a-priori to a LiDAR survey in order to determine the best

and worst case point positioning accuracy. The analysis in this report will consider a scenario based on typical UAV flying heights (e.g. 30 – 120 m Above Ground Level), with the aircraft flying a flat northern path (i.e. ω , ϕ and $\kappa = 0^\circ$). Furthermore, it will be assumed that the observation of a LiDAR measurement has coincided directly with observations of position and orientation from which the expected accuracy of a single LiDAR measurement from this system can be simulated.

2.3 Data Processing and Calibration

The UAV's position and orientation system consists of three sensors (IMU, video, and GPS), providing observations at variable rates up to 200Hz. The rate of the LiDAR sensor, however, is significantly faster, measuring up to 22,000 returns per second. Therefore, the observations from these sensors are required to be interpolated in order to determine the position and orientation of the LiDAR sensor at the instant of each range observation. The most commonly used algorithm for this purpose is the Extended Kalman Filter (EKF), which is limited by its complexity and in its accuracy by the inclusion of a first order linearisation of the functional model (Van Der Merwe & Wan, 2004). To overcome these limitations, the fusion of the positional data for the TerraLuma UAV is completed with the use of a loosely coupled Sigma-Point Kalman smoother (SPKS). It has been shown that variants of SPKS consistently outperform the EKF in terms of correctness, robustness and ease of implementation (Kelly & Sukhatme, 2009; Van Der Merwe & Wan, 2004). Therefore, the use of a loosely coupled SPKS with a potentially dynamic sensor set such as that on-board the UAV provides an ideal filtering option.

To further improve the accuracy of the orientation estimates, the SPKS will be augmented with a novel algorithm that has been developed to include observations of orientation through the use of high definition video footage. The process used involves processing downward looking video footage using an algorithm that is similar to the Structure from Motion (SfM) technique often used in robotics or close range photogrammetry (Barazzetti et al. 2010). A post-processing based strategy allows for the optimal solution to be found through the use of a full bundle adjustment (Nagai et al., 2008). This inclusion of high definition video observation adds further redundancy and through direct observations of relative orientation provides greater accuracy in comparison to a traditional GPS/IMU positioning system.

The output of the GPS/IMU/Video SPKS is the position and orientation of the origin of the body frame within the ECEF Cartesian system. Considering this information, the next step within the LiDAR workflow requires the position and orientation of the laser system within the ECEF system to be known, which is reliant on the calibration parameters. Within a UAV LiDAR system the effect of errors within the boresight angles are minimised due to the low flying height. For instance, a boresight misalignment of 0.01° , which results in an error of 1.31 m at a flying height of 700 m, will result in a horizontal error of only 0.005 m at typical UAV flying heights (30 - 120 m). The errors in the measurement of the lever arm offset, however, propagate directly into the accuracy of point position and need to be minimised by an appropriate calibration strategy. Conventional procedures for determining calibration parameters require a periodic survey of a well observed site (i.e. often an air field). However, the modularity of a UAV system suggests that the calibration parameters are likely to be significantly more dynamic and change between individual surveys. Therefore, a specific calibration procedure has been designed for this UAV. The procedure follows the three traditional stages of calibration for a LiDAR system outlined by Habib et al. (2010). For the purposes of this study the accuracy of the system calibration of the UAV will be considered to be equal to a full scale system based on simulation of the designed calibration strategy.

3. Results

The propagation of error into LiDAR point clouds is usually based on the flying conditions and error expected within a state-of-the-art LiDAR system at the time of publication. For example, Goulden & Hopkinson (2010) reports on the error within a LiDAR system based on the Optech 3100 scanner. The conditions of a UAV survey are, however, significantly different to full-scale surveys due to factors such as reduced flying heights and the greater inaccuracies of the miniaturised sensors. A comparison between the error contribution of each of the 17 components within a TerraLuma system and typical Optech 3100 setup (Table 2) highlights the inaccuracy associated with the measurement of angular quantities in low cost systems. Furthermore, table 3 outlines a comparison of other key variables often used as a measure for data quality in forest surveys, highlighting the exceptionally high point density (1000 pts/m²) of the system.

Table 2. A-priori standard deviation values of parameters within TerraLuma LiDAR System in comparison to a full scale system (adapted from Goulden & Hopkinson, 2010 and based on an Optech ALTM 3100 scanner)

Parameter	TerraLuma Value (1 σ)	Full Scale Value (1 σ)
σ_x, σ_y	0.03 m	0.03 m
σ_z	0.05 m	0.05 m
σ_ω and σ_ϕ	0.2 - 0.5 °	0.005 °
σ_κ	0.2 - 0.5 °	0.010 °
$\sigma_{\omega_b}, \sigma_{\phi_b}$ and σ_{κ_b}	in $\sigma_\omega, \sigma_\phi$ & σ_κ	in $\sigma_\omega, \sigma_\phi$ & σ_κ
$\sigma_{x_L}, \sigma_{y_L}$ and σ_{z_L}	0.01 m	0.01 m
σ_r	0.10 m	0.015 m
σ_β	0.125 °	0.003 °
σ_{B_h}	0.020 °	0.014 °
σ_{B_e}	0.400 °	0.014 °

Table 3. A comparison of the key LiDAR variables between the TerraLuma UAV and an Optech ALTM 3100 scanner.

Variable	TerraLuma UAV	Full Scale System
Typical Flying height	50 m	1100 m
Scan Angle Range	-60 – 50	-25 – 25
Swath width at 25°	47 m	1300 m
Maximum Swath width	146 m	1300 m
Point Density	up to 1000 pt/m ²	3-15 pt/m ²
Laser footprint	0.07x1.40 m	0.27x0.27 m
Max. angle of incidence (flat terrain)	60	25

One of the key constraints within a low-cost system such as the TerraLuma UAV is the accuracy of the IMU. The IMU is essential to the system, providing the highest rate of measurement, tracking the position and orientation of the system in between the observations of accurate sensors. The IMU used within the TerraLuma system has been determined to have standalone short-term orientation accuracies in agreement to the manufacturer’s specifications of 2° . This accuracy is accompanied by significant bias instabilities and noise that are characteristic of a MEMS based IMU (which if uncorrected can result in a drift of up to 0.2° per second). The result of considering observations of orientation made by a drift-free IMU ($\sigma\omega$, $\sigma\phi$ and $\sigma\kappa = 2^\circ$) on the accuracy of the LiDAR measurements can be seen in Figure 2. This figure demonstrates that the error due to poor estimates of orientation contributes up to 98% of the total error budget.

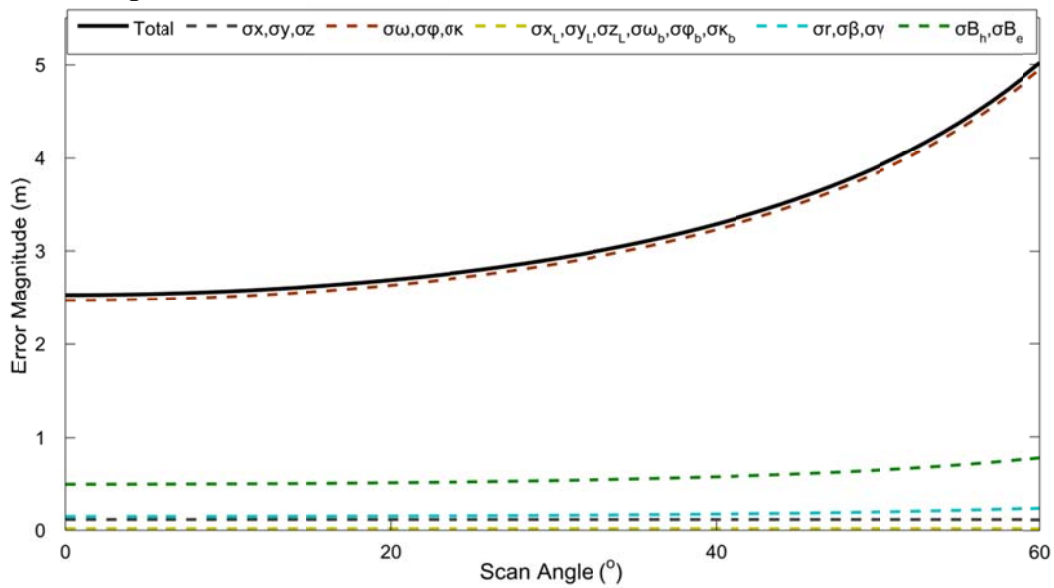


Figure 2. The effect of all error sources within the TerraLuma UAV LiDAR system on the point positioning accuracy of the LiDAR measurements, considering the worst case system errors and a flying height of 50 m.

The conventional method for constraining the error within the observations of orientation is to make use of additional observations from a GPS receiver. The GPS receiver on board the UAV makes high accuracy observation of position at a lower rate (20 Hz) than the IMU. Therefore, the use of a well designed integration strategy that constrains the IMU errors at each GPS observation can provide improved orientation estimates at the higher rate of the IMU observations. Under such a strategy, it has been shown that the errors in the observations of pitch and roll ($\sigma\omega$ and $\sigma\phi$) can be reduced to 0.5° and to 1.5° in the yaw observation ($\sigma\kappa$) (Du, 2010). Although this resulting in a significant reduction in error, orientation is still the primary source of error (as shown in Figure 3).

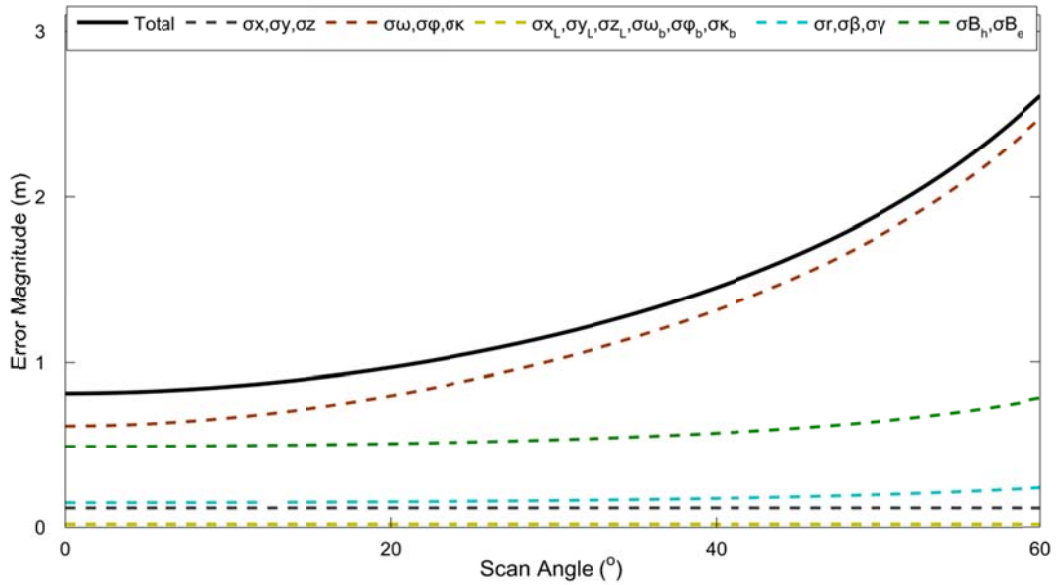


Figure 3. The result of a GPS/IMU filtering strategy constraining orientation error, to 0.5° in pitch and roll and 1.5° in Yaw, at a flying height of 40 m resulting in a higher overall LiDAR point positioning accuracy.

Orientation estimates from the outlined SfM routine can be used to further constrain orientation errors. Initial experiments of the novel algorithm developed for this research has shown significant improvements in the estimates of yaw and smaller improvements in the estimate of pitch and roll. Based on these initial tests, orientation estimates are expected to be reduced to within a range of 0.2 to 0.5° . Figure 4 shows that the contribution of error from orientation is now within a similar range to the contributions of other system components. The primary error contributions can now be attributed to both the vertical beam divergence of the Ibeo LUX system and the orientation errors. It is noted that the Ibeo LUX has a high vertical beam divergence to ensure full coverage of the scanned area for its intended short range application in automation and this error component is fixed.

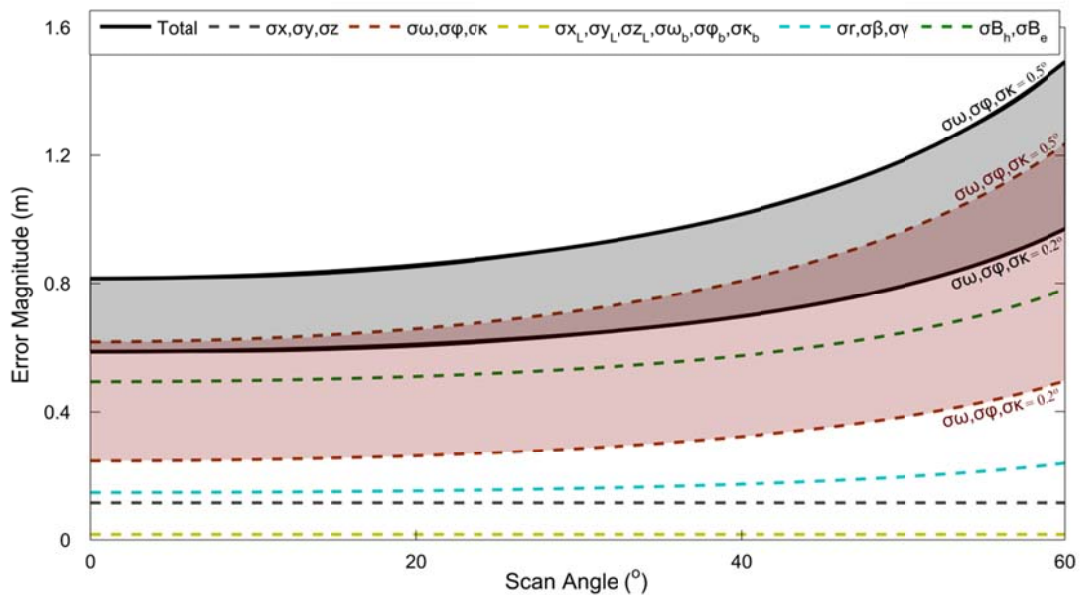


Figure 4. The expected result of improved orientation measurements due to the inclusion of HD video and Kalman filtering within the on-the-point positioning accuracy of the TerraLuma LiDAR system.

The point positioning error present within the observations made by the TerraLuma platform has been shown to increase with scan angle (Figures 2-4). Considering the high beam divergence of the Ibeo LUX scanner, the other dominant variable affecting the accuracy of the LiDAR observations is flying height (demonstrated in Figure 5). The magnitude of horizontal error towards the edges of the scan angle range increases by an order of magnitude between flying heights of 30 and 120 m. This effect is least pronounced at the centre of the scan suggesting that if increased flying heights are to be used the scan angle of the LiDAR system should be constrained. Furthermore, increased flying heights also result in significant decreases in the density of the measured point cloud.

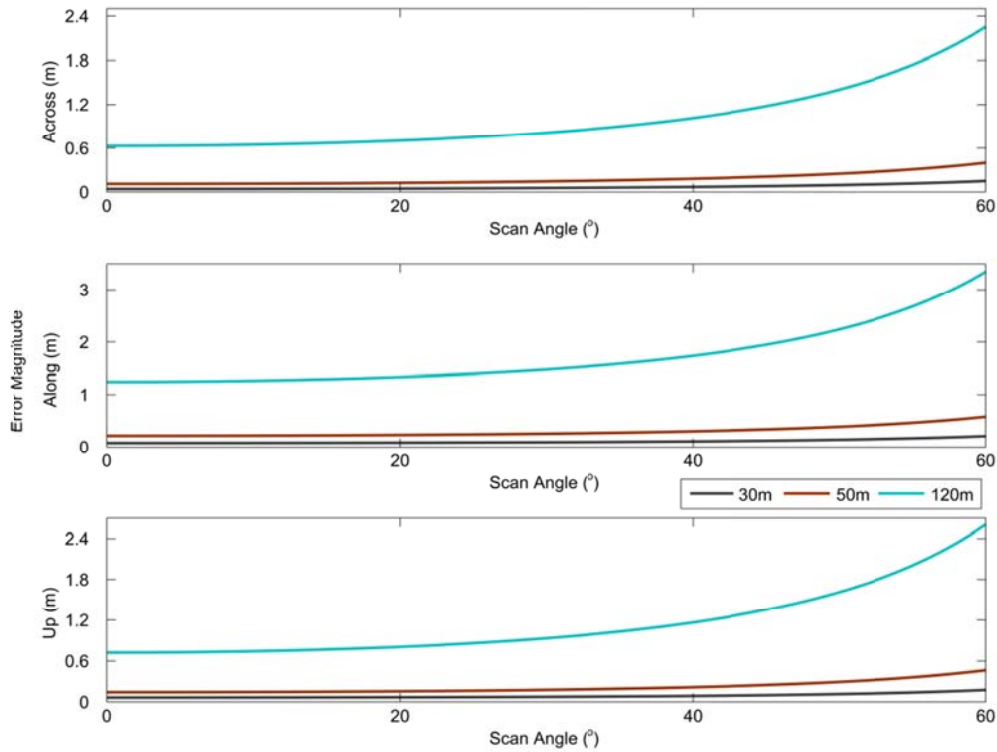


Figure 5. The accuracy of the TerraLuma LiDAR system based on the error statistic presented in Table 1 (σ_ω , σ_ϕ and $\sigma_\kappa = 0.35^\circ$) and at flying heights of 30 m, 50 m and 120 m.

The cumulative effect of the errors, as a function of scan angle, within the TerraLuma system (flying at 50 m) in comparison to a typical full scale system flying at 1100 m is illustrated in Figure 6. The full scale system modelled is the Optech ALTM 3100 scanner (properties given in Table 2 and 3) which has a scan angle range of $\pm 25^\circ$, within which the accuracy of the UAV system is within an acceptable range of the full scale system.

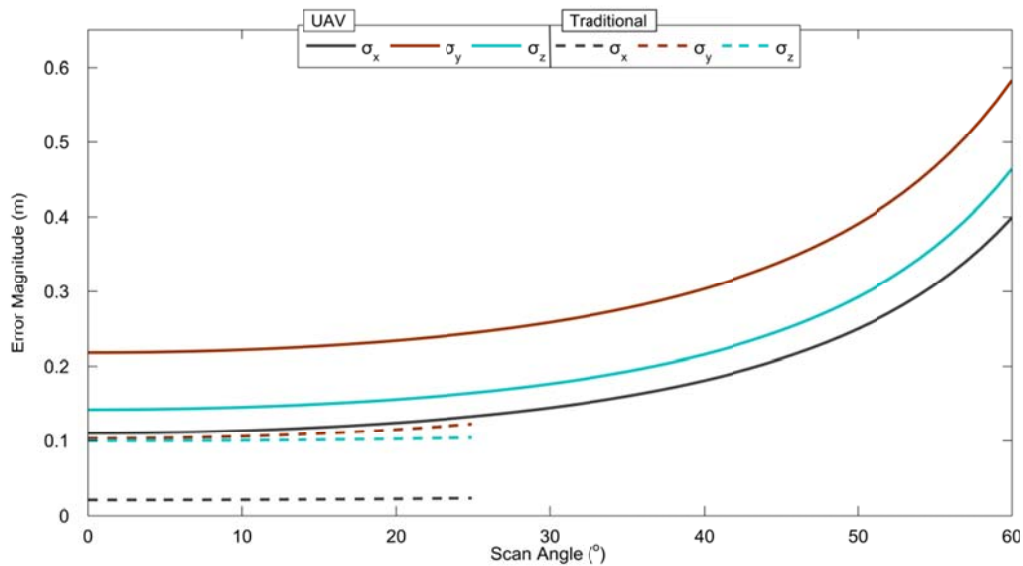


Figure 6. Accuracy of the TerraLuma LiDAR system ($\sigma\omega$, $\sigma\phi$ and $\sigma\kappa = 0.35^\circ$) flying at 50 m in comparison to the accuracy of a traditional full scale system flying at 1100m (full scale system properties adapted from Goulden and Hopkinson, 2010).

4. Discussion

The use of a GPS/IMU/Video SPKS has allowed for a significant reduction in the error budget of the TerraLuma UAV LiDAR system. At a typical flying height of 50 m UAV-borne LiDAR can be expected to produce point clouds with sufficient accuracy for forest mapping. This accuracy, coupled with the increased point density achieved by the TerraLuma LiDAR system, presents a significant advantage for the fine-scale mapping of forests. The use of such a system will most likely result in a reduction in the underestimation of tree height measurements due to the highly dense point cloud being produced. Moreover, the modularity and low cost of the a TerraLuma LiDAR system will allow surveys to be conducted at a higher temporal frequency, allowing canopy closure rates and forest health to be monitored closely.

Several limitations of the use of a UAV-borne LiDAR system for forest measurement can also be determined from the results. The restricted flying height means that ground coverage is going to be significantly reduced and that the angle of incidence of a measurement towards the edge of the scan on the canopy will be increased. Although limited ground coverage is not necessarily a large restriction in the type of survey that would be designed for UAV platforms, increases in the angle of incidence would result in poor estimates of forest health indicators such as canopy closure and fraction cover. In order for UAVs to be used for this measuring this metric empirical correction based on other field measurements would be required.

The next stage in the development of the TerraLuma UAV is the collection and analysis of data over a forested test area. This will allow the capabilities of the system for forest mapping to be fully assessed and a set of survey designs to be developed with the aim of determining key forest indicators with potentially high temporal change such as forest health, bio-mass and canopy closure rates. Other future improvements to the system are reliant on advances in LiDAR technology, however, increased uptake of current technology could result in demand for an adaption of automotive laser scanners into lightweight scanners with smaller beam divergence for specific mapping purposes.

5. Conclusion

This study has demonstrated that a UAV-borne LiDAR using low-cost, lightweight sensors can produce point clouds with only slightly worse accuracies than full-scale traditional systems but with much higher point densities. It has been shown that improvements in the estimates of system orientation, produced through the use of a novel video augmented SPKS, has allowed for a low cost, light weight sensor suite to produce a point cloud of adequate accuracy for forestry mapping. Furthermore, by showing the achievable accuracy of a UAV derived point cloud, the implications and potential of such a platform for hyper-temporal resolution forest surveys, especially in the areas of change detection have been highlighted in this study.

References

- Akay, A.E., Oğuz, H., Karas, I.R., & Aruga, K., 2009. Using LiDAR technology in forestry activities. *Environmental monitoring and assessment*, 151, 1, 117-25.
- Baltsavias, E.P., 1999. Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54, 2, 199 - 214.
- Barazzetti, L., Remondino, F., Scaioni, M., & Brumana, R., 2010. Fully automatic UAV image-based sensor orientation. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII, Part 5 Commission V Symposium*. Newcastle upon Tyne, UK.
- Berni, J. A. J., Zarco-tejada, P. J., Suárez, L., & Fereres, E., 2009. Thermal and Narrowband Multispectral Remote Sensing for Vegetation Monitoring From an Unmanned Aerial Vehicle. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 3, 722-738.
- Bouget, J., 2010. Camera Calibration Toolbox for Matlab®. Retrieved from http://www.vision.caltech.edu/bouguetj/calib_doc/.
- Du, S., 2010. Integration of Precise Point Positioning and Low Cost MEMS IMU. Unpublished masters dissertation, *University of Calgary, Calgary, Canada*.
- Eisenbeiss, H., & Zhang, L., 2006. Comparison of DSMs generated from mini UAV imagery and terrestrial laser scanner in a cultural heritage application, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVI-Part5*, 90-96.
- Glennie, C., 2007. Rigorous 3D error analysis of kinematic scanning LIDAR systems, *Journal of Applied Geodesy*, 1, 147-157.
- Goulden & Hopkinson, 2010. The forward propagation of integrated system component errors within airborne LiDAR data, *Photogrammetric Engineering and Remote Sensing*, 76, 5, 589 – 601.
- Habib, A., Bang, K.I., Kersting, A.P., & Chow, J., 2010. Alternative Methodologies for LiDAR System Calibration. *Remote Sensing*, vol. 2, No. 3, pp. 874-907.

- Hunt Jr, E.R., Hively, W.D., Fujikawa, S.J., Linden, D.S., Daughtry, C.S.T., & McCarty, G.W., 2010. Acquisition of NIR-Green-Blue Digital Photographs from Unmanned Aircraft for Crop Monitoring. *Remote Sensing*, vol. 2, No. 1, pp. 290–305.
- Hyypä, J., Hyypä, H., Leckie, D., Gougeon, F., Yu, X., & Maltamo, M. 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, vol. 29, No. 5, pp. 1339-1366.
- Jaakkola, A., Hyypä, J., Juha, K., Kukko, A., Yu, X., Xiaowei, Kaartinen, H., Lehtomäki, M. & Lin, Y., 2010. A low-cost multi-sensoral mobile mapping system and its feasibility for tree measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, No. 6, pp. 514-522.
- Kelly, J., & Sukhatme, G. S., 2009. Visual-inertial simultaneous localization, mapping and sensor-to-sensor self-calibration. *2009 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, Daejeon, Korea: 360-368.
- Lichti, D.D., and Gordon, S. J., 2004. Error Propagation in Directly Georeferenced Terrestrial Laser Scanner Point Clouds for Cultural Heritage Recording, *Proceedings of FIG Working Week 2004*, Athens, Greece.
- Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M., 2003. LiDAR remote sensing of forest structure. *Progress in Physical Geography*, 27, 1, 88-106.
- Lin, Y., Hyypä, J., & Jaakkola, A., 2011, Mini-UAV-Borne LIDAR for Fine-Scale Mapping. *Geoscience and Remote Sensing Letters, IEEE*, 8, 99, 426–430.
- May, N., & Toth, C., 2007, Point positioning accuracy of airborne LiDAR systems: A rigorous analysis. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, 36, 107–111.
- Morin, K. W., 2002, Calibration of airborne laser scanners, Unpublished Masters dissertation, *University of Calgary, Calgary, Canada*.
- Morsdorf, F., Nichol, C., Malthus, T., & Woodhouse, I.H., 2009. Assessing forest structural and physiological information content of multi-spectral LiDAR waveforms by radiative transfer modelling. *Remote Sensing of Environment*, 113, 10, 2152-2163.
- Nagai, M., Chen, T., Ahmed, A., & Shibasaki, R., 2008. UAV Borne Mapping by Multi Sensor Intergration. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Part b1 XXXVII*, Beijing, China: 1215-1222.
- Schaer, P., Skaloud, J., Landtwing, S., & Legat, K., 2007. Accuracy estimation for laser point cloud including scanning geometry. *5th International Symposium on Mobile Mapping Technology (MMT2007)*, Padua, Italy.
- Van Der Merwe, R., & Wan, E. A., 2004. Sigma-point Kalman filters for integrated navigation. *Proceedings of the 60th Annual Meeting of the Institute of Navigation (ION)*, Ohio, USA: 641–654.

SilviLaser 2011, Oct. 16-19, 2011– Hobart, Aus

Zhang, H., Wu, Y., Wu, M., & Hu, X., 2010, Improved multi-position calibration for inertial measurement units, *Measurement Science and Technology*, 21.

Sorted Pulse Data (SPD) Format: A new file structure for storing and processing LiDAR data

Pete Bunting^{1,2}, John Armston³ Richard Lucas¹ & Daniel Clewley¹

¹Institute of Geography and Earth Sciences, Aberystwyth University, UK. pete.bunting@aber.ac.uk

²Landcare Research, Palmerston North, NZ

³ Joint Remote Sensing Research Program, Centre for Spatial Environmental Research, School of Geography, Planning and Environmental Management, University of Queensland
j.armston@uq.edu.au

Abstract

This paper presents a new generic method and format for storing and processing airborne and terrestrial LiDAR pulse data within a HDF5 file. The format is specifically designed to support both traditional discrete return and full waveform data, uses a pulse (rather than point) based data model and has been developed and applied successfully using a wide range of disparate airborne and terrestrial LiDAR datasets. The format is proposed as an alternative to existing solutions as it includes support for full waveform data, explicit pulse based data structures and flexible spatial indexing using cartesian, spherical and polar coordinate systems and projections. The HDF5 format supports compression but in part due to the more complex data structures used the amount of compression that can be achieved is limited. However, it compares favourably with the file size of uncompressed LAS files and is able to accommodate a much wider range of LiDAR datasets.

Keywords: LiDAR, Storage, File Format, Waveform, Discrete Return, Pulses

1. Introduction

The uses and availability and application of Light Detection and Ranging (LiDAR) has grown significantly over the last 10-15 years (Shan and Toth, 2009) to the point that regular acquisitions across large areas and over time are commonplace. However, new methods of data storage, including indexing and compression are needed because of the large quantities of data acquired as well as algorithms that take advantage of these. New formats are also needed to support the use of full waveform LiDAR systems and Terrestrial Laser Scanners (TLS) data alongside the more traditional discrete return airborne datasets.

Currently, LiDAR data are stored either in an ASCII format (often without a standard definition of the format) or American Society for Photogrammetry & Remote Sensing (ASPRS) LAS format, which has been the standard for storing discrete return LiDAR data as a binary file (ASPRS, 2011). Extensions for the LAS 1.X format have recently been proposed and implemented (although not as part of the ASPRS standard) for indexing and compressing the LAS format (Isenburg, 2011). However, the LAS format does have several current limitations:

- a) A suitable standard for storing full waveform data is lacking. Although LAS 1.3 has been proposed and the specification approved, only a partial solution for waveform data has been presented and the structures for all the necessary data are not provided. Additionally, it is worth noting that current open source libraries that read/write LAS 1.3 currently ignore the waveform data.
- b) The waveform is also just viewed as an extension to the discrete return datasets for which LAS was designed. Only a limited number of returns are permitted and not all attributes of discrete returns derived from waveform data can be stored. LAS 2.0 may

completely overhaul the LAS data structure to change this, but would likely result in a file format that was compatible with code written for LAS 1.X files, therefore making it equivalent to any other newly proposed file format with regards to software support.

- c) Data are viewed as discrete points rather than pulses. There is no set mechanism to store information that can be associated with the transmitted as well as the received components of the laser pulse (e.g., pulse origin coordinate and shape). Such information is particularly important for waveform analysis.

As a result of difficulties in using full waveform data, particularly in relation to efficiently storing pulses rather than individual returns, this research aimed to develop a new LiDAR format that focused on the efficient storage of pulses rather than individual returns but allowed processing of both. The format was designed to allow processing of all LiDAR datasets including those acquired by discrete return and full waveform airborne, spaceborne and terrestrial laser scanners. The format was designed specifically for post-processing, rather than for use during acquisition, and to allow data to be efficiently accessed (through random spatial access procedures) such that a suite of data processing methods could be applied. The following sections provide an overview of the format.

2. Method

To store the LiDAR data, from pulsed laser systems where power output takes the form of pulses of light on a time scale, a new set of data structures stored within the hierarchical data format version 5.0 (HDF5) file format was proposed. The HDF5 format presents a number of advantages, including generic readability across a range of platforms and architectures, and support within common remote sensing software (e.g., ITTVIS ENVI) and programming languages (e.g., ITTVIS IDL, C++, Python, Java, C#). The following sections therefore outline the pulse and point data representations and their relationship to the header file meta-data and the methods of data storage within the HDF5 file, including spatial indexing and compression.

2.1 Data Representation

The primary data types for LiDAR data are pulses and points (Figure 1). A pulse represents a single measurement unit from a pulsed laser system and all the information common to the transmitted and received pulses of light is stored within this data type. In addition to storing a list of points, the pulse also references two data blocks that store the transmitted and received waveform data when present, with these viewed as two samples from the same measurement. The transmitted and received waveforms are viewed as two samples from the one measurement. By contrast, a point represents a single discrete return from the transmitted pulse and can be measured directly at sensor or subsequently derived from the digitised waveform data. Where points are derived from a waveform, the waveform offset fields (the distance/time along the waveform) are used to connect the point with the waveform.

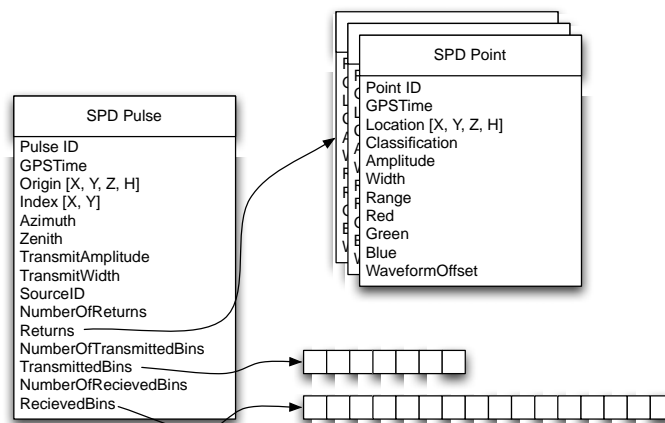


Figure 1. The structure of an SPD Pulse and Point

A wide range of parameters have been associated with a pulse (Table 1), some of which (e.g., the scan angle rank and scan direction flag) map directly onto LAS version 1.3 data fields. However, in the new format, other parameters (e.g., origin, index point, azimuth and zenith angle) are uniquely defined. The format stores the waveform as arrays of unsigned integer values (either with 32, 16 or 8 bits) where gains and offsets are defined for each pulse. The X and Y values define the spatial location of the pulse when data are indexed; these are required as a pulse can be indexed in a number of ways (e.g., based on first or last return or maximum waveform amplitude), depending on whether discrete or waveform and the level of pre-processing applied. The origin, zenith, azimuth and the range to the start of the waveform are required to locate the start of the received waveform in cartesian coordinates. In addition to the X and Y and also Z origin parameters, the SPD pulse includes a height field that can be populated with the height of the origin above the ground surface following its identification. Azimuth and zenith fields can be defined for all datasets but are used mainly for TLS data which, when projected using a spherical coordinate system, can provide useful information on, for example, rates of river bank erosion.

Table 1: Attributes defined for an SPD Pulse.

Description	Data Type	Units
Pulse ID – unique identifier for each pulse.	Unsigned 64 bit Integer	N/A
GPS Time – pulse acquisition time	64 bit float	Nano Seconds
Origin (X, Y, Z)	64 bit float	Dependent of Coordinate System
Origin Height – height above ground surface	32 bit float	Dependent of Coordinate System
Index Point (X, Y) – point used to index the pulse	64 bit float	Dependent of Coordinate System
Azimuth	32 bit float	Radians
Zenith	32 bit float	Radians
Number of returns	Unsigned 8 bit Integer	N/A
List of Points (Return)	SPD Point	N/A
Number of transmitted waveform values	Unsigned 16 bit Integer	N/A
List of values for transmitted waveform	Unsigned 32 bit Integer	N/A
Number of received waveform values	Unsigned 16 bit Integer	N/A
List of values for received waveform.	Unsigned 32 bit Integer	N/A
Range to the start of the waveform	32 bit float	Distance
Amplitude of the transmitted pulse (intensity)	32 bit float	Power

Width of the transmitted pulse	32 bit float	FWHM; Nano seconds
User field – can be used to store additional information.	Unsigned 32 bit Integer	N/A
Pulse source ID	Unsigned 16 bit Integer	N/A
Edge of flight line flag.	Unsigned 8 bit Integer	N/A
Scan direction flag	Unsigned 8 bit Integer	N/A
Scan angle rank.	32 bit float	Degrees
Waveform Noise Threshold	32 bit float	Power
Transmitted Waveform Gain	32 bit float	N/A
Transmitted Waveform Offset	32 bit float	N/A
Received Waveform Gain	32 bit float	N/A
Received Waveform Offset	32 bit float	N/A

The SPD pulse can also reference a list of individual returns associated with a pulse and represents these using a point structure (Table 2). The majority of the fields are aligned with the LAS 1.3 specification but additional parameters such as the height of the point above the ground surface, range of the pulse from the origin (used for spherical coordinate systems) and the offset of the return within the waveform (if present) are included.

Table 2: Attributes defined for an SPD Point

Description	Data Type	Units
Return ID – unique identifier for each return within pulse.	Unsigned 8 bit Integer	N/A
GPS Time – return acquisition time	64 bit float	Nano Seconds
X, Y, Z	64 bit float	Dependent of Coordinate System
Height – height above ground surface	32 bit float	Dependent of Coordinate System
Range – range from origin defined in pulse (spherical coordinates)	32 bit float	Dependent of Coordinate System
Amplitude of the return (intensity)	32 bit float	Power
Width of the return	32 bit float	FWHM; Nano seconds
Red, Green and Blue values for visualization	Unsigned 16 bit Integer	N/A
Classification of return	Unsigned 8 bit Integer	N/A
User field – can be used to store additional information.	Unsigned 32 bit Integer	N/A
Offset within waveform – in time and array index	Unsigned 32 bit Integer	N/A
Model key point flag	Unsigned 8 bit Integer	N/A
Low point flag	Unsigned 8 bit Integer	N/A
Within overlapping region flag	Unsigned 8 bit Integer	N/A
Ignore return flag.	Unsigned 8 bit Integer	N/A

Representing the data as pulses presents a number of advantages over other point-based formats (e.g., LAS) in that:

- a) These correspond with the sensor-technology and process of data acquisition by pulsed laser systems, allowing all information associated with each individual pulse to be stored for processing.
- b) Pulses which did not result in a return are stored, with this being useful for applications

such as quantifying vegetation canopy structure (e.g., gap fraction) and cover from TLS data.

- c) Returns are explicitly connected and ordered within the pulse, allowing for easy processing.
- d) Waveform rather than only individual discrete returns are represented. Returns are directly connected to the location within the waveform from which they were extracted, if decomposed.

2.2 Header Meta-Data

The SPD header attributes allow meta-data and other dataset parameters associated with the dataset to be stored within the HDF5 file. Other advantages of the HDF5 format are that the dataset parameters can be easily edited within generic HDF5 software, such as HDFView, and other attributes can be added at any time without affecting the existing software (although new values would not be read or used). For reference, the SPD format expects to find fields for the parameters listed in Table 3.

Table 3: SPD File header parameters

Parameter	Description
Spatial reference	A variable length string data field which stores the projection of the dataset as a WKT string
Index type	Specifies how the file is a spatially indexed (not used for unindexed data).
File type	Specifies the file type and whether the data is unordered (i.e., no index) or written sequentially or non-sequentially if an index is used.
Discrete points defined	Specifies whether fields are defined for discrete return data.
Decomposed points defined	Specifies whether fields are defined for data which has been decomposed from a waveform.
Transmitted Waveform defined	Specifies whether the transmitted waveform data is stored within the file.
Received Waveform defined	Specifies whether the received waveform data is stored within the file.
Version	The major and minor version numbers of the SPD format used.
Pulse Version	Specifies the version of the Pulse data type, in the future multiple pulse data types could be defined as technology changes.
Point Version	Specifies the version of the Point data type, in the future multiple point data types could be defined as technology changes.
Generating software	The software used to generate the file.
System identifier	The system being used to generate the file.
Date and time of creation	The date and time the file was created.
Date and time of capture	The date and time the data stored within the file was captured.
Number of points	The number of points stored within the SPD file.
Number of pulses	The number of pulses stored within the SPD file.
User meta data	A variable length string field which can be used by the user to store further meta data.
Bounding volume	The minimum and maximum X, Y and Z bounds of the dataset.
Bounding sphere	The minimum and maximum azimuth, zenith and range bounds of the dataset (when a spherical coordinate system is defined).
Bin size	The size of the bins (units depending on index type and coordinate system) used for indexing the data (not used for unindexed data).
Number of bins in the X axis	The number of bins within the x axis of the spatial index (not used for unindexed data).

Number of bins in the Y axis	The number of bins within the y axis of the spatial index (not used for unindexed data).
RGB defined	Whether the RGB values for each point have been defined.
Pulse compression block size	The size of the blocks used to compress the pulses written to the file.
Point compression block size	The size of the blocks used to compress the points written to the file.
Received waveform compression block size	The size of the blocks used to compress the received waveform written to the file.
Transmitted waveform compression block size	The size of the blocks used to compress the transmitted waveform written to the file.
Temporal bin spacing	The time in nano seconds between the waveform bins
Waveform Bit Resolution	The number of bits (i.e., 8, 16 or 32) used for storing each element of the waveform.
Origin defined	Specifies whether the origin and therefore the azimuth and zenith values have been defined.
Method of pulse indexing	Specifies how the pulse index X and Y fields are defined (i.e., if the first, last return or maximum intensity was used).
Height defined	Specifies whether the height fields within the SPD pulse and point have been defined.
Synthetically generated return numbers	Specifies whether the return numbers are not given in order for the pulse. For example, first and last returns maybe distributed separately with no means of reconnecting them so a pulse will be created for each point resulting in pulses containing a single return within a return number greater than 1.

In addition, the following fields for defining the sensor and acquisition parameters are included within the SPD header: laser wavelength; pulse repetition frequency; beam divergence; sensor height; laser footprint; maximum scanning angle; sensor speed; scan rate; point density; pulse density; across track spacing; along track spacing; angular spacing in the azimuth; and angular spacing in the zenith.

2.3 Storage on Disk

To represent the data within the HDF5 file, a hierarchical directory structure is created within the top level of the HDF5 file (Figure 2). This structure separates the LiDAR data, header attributes, spatial index (if provided) and overview image (only available if the file has a spatial index) into discrete sections, thereby allowing easier human navigation. Within the data directory, separate lists (arrays) are used to store the received and transmitted pulses, points and waveforms. The waveforms are stored as lists of integers, while the points and pulses are stored as the predefined types defined in Tables 1 and 2. Header values are stored as individual datasets within the header directory (Table 3). If a spatial index is used, two 2-dimensional arrays (i.e., images) are defined within the index directory, which contain the pulse list offsets and the number of pulses within each grid cell. These allow the pulse list to be accessed randomly. Where a spatial index is provided, a ‘quicklook’ overview image of the dataset is also stored, with this being a single band image stored within the quicklook directory.

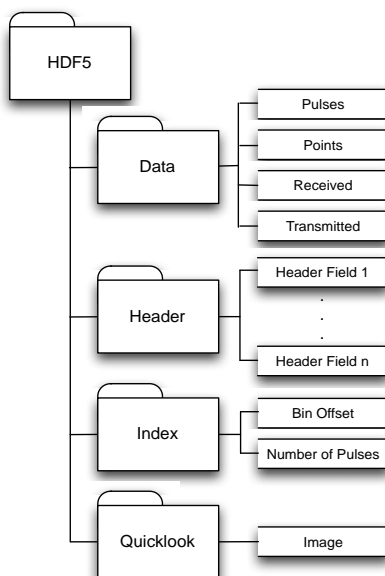


Figure 2. HDF5 file internal directory structure. The index and quicklook directories are only available when a spatial index is defined.

2.3.1 Spatial Indexing

When storing the pulse data within the HDF5 file, two options are provided for ordering the list of pulses. The first is unsorted (pulse data; UPD) as no specific ordering is forced on the data. In practice, however, this is usually time-sequential and therefore equivalent to a standard LAS file. The second is through the use of a spatial index (sorted pulse data; SPD), which defines (and is referenced to) the order of the pulses within the list. There are a number of methods for spatial indexing which could be used, such as quad-trees (Finkel and Bentley 1974). However, in this case, a simple grid-based index was adopted, whereby all the pulses within a grid cell are written consecutively within the pulses list. This method allows easy and rapid spatial access to the pulse data at the index resolution and the data can be mapped directly onto image pixels, the most commonly derived output. Image-like processing methodologies for controlling memory usage can also be used. The main disadvantage of this method is that the data file consequently has a native scale, defined by the index, while quad-tree and similar indexing approaches are scale independent. However, if one of these indexing methods were used, the pulses could not be written to the file consecutively, which may require the reader to jump to a large number of locations within the input file when a region of data is selected. This can be an expensive operation in terms of time and memory, particularly if compression is deployed.

When indexing the file, two options for ordering the pulses within the list are provided (Figure 3).

- a) Pulses are written in order from the top of the scene down in rows and across each row from left to right. Whilst the data has to be written and, in some cases processed in this order, fast read performance is provided as all pulses within a row (or continuous part of a row) can be read in a single operation because they are consecutively ordered.
- b) Only the pulses within a cell are written consecutively within the pulses list. This increases the flexibility of the writing operation, as cells can be randomly written to the output file. However, the reading performance is reduced as multiple read operations, potentially from different parts of the file, would be required to select pulses from a number of cells.

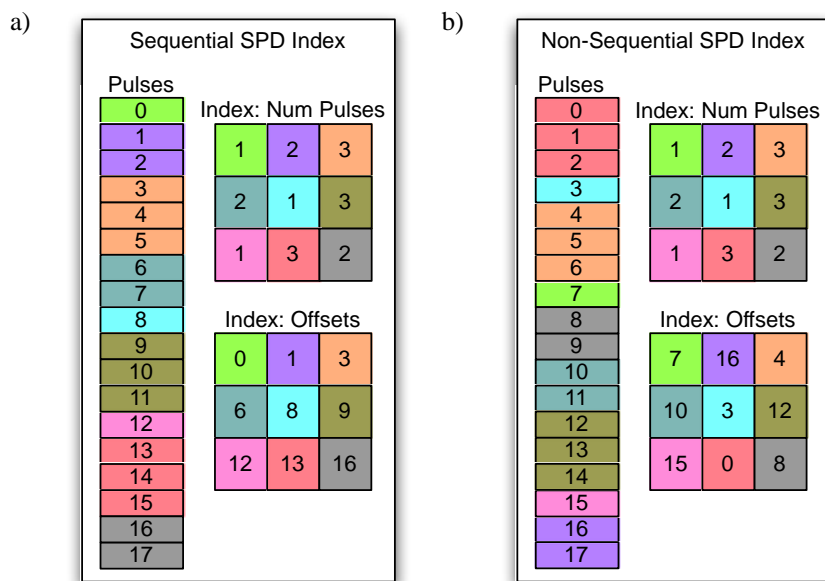


Figure 3. The spatial indexing structure of the SPD file, where the colours indicate the index cell the pulses are associated with and where the corresponding point reside in the pulses list. a) sequentially written and b) non-sequentially written.

The SPD formats spatial indexing currently supports cartesian, spherical and polar coordinate systems and projection (Figure 4). For airborne laser scanning data, the cartesian coordinate system is generally the most appropriate for indexing but for TLS data, the index representation is expected to be application dependent. The option used for any specific file is defined within the SPD file header.

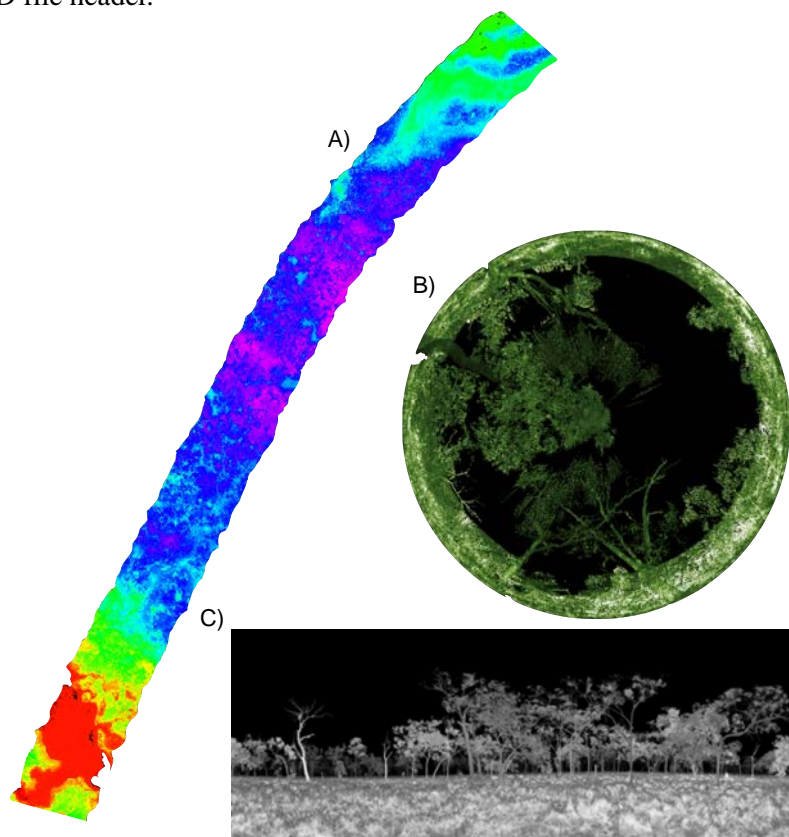


Figure 4. Overview images for a) Cartesian-indexed airborne LiDAR scene, b) polar-indexed terrestrial LiDAR scene and c) a spherically-indexed terrestrial LiDAR scene.

2.3.2 Compression

Another advantage of defining this format within a HDF5 file is that it provides support for compression, using the zlib deflate algorithm (IETF 1996). The pulse data, pulses, points and waveforms are each stored as separate lists and are compressed separately into blocks. Whereas a single list of pulses, points, and waveform values are defined in the file, on disk they are compressed in blocks that are by default, 250 objects in size, although this is customisable and defined in header. If the block size is increased, the memory requirements of the application using the format will also increase but the resulting file size may become smaller. Decreasing the block size has the opposite effect. In addition, if only small selections of data are retrieved from the file at any one time or from many different parts of the file, the large block sizes will result in a slower performance as the whole block needs to be decompressed for the data to be read. Following experimentation, a value of 250 for the block sizes for all the lists was found to be most appropriate.

3. Discussion

3.1 Observations

Versions of this file format have now been in active use for over two years and a number of observations have been made during this time. First, the spatial indexing of the LiDAR data provides memory efficient processing allowing datasets larger than the memory available to be processed efficiently and in a timely manner. Spatial indexing is a key component of containing the memory requirements of any given algorithm as only the specific parts of the dataset need to be read into memory at any one time, while avoiding the need to search the whole dataset. Additionally, storing the index data structures within the same file as the data ensures all the information is kept together and file management is simplified. Second, the pulse based storage format, as opposed to points, has created a significantly more flexible system. This is particularly relevant for waveform data and any future sensor developments where the data structures used within the file structure directly map onto the process by which the data are acquired. Furthermore, by explicitly connecting discrete returns associated with a single transmitted pulse, further information and assumptions can be made when attempting to classify the ground surface and assessing vegetation structures. Finally, the use of the grid-based index significantly simplifies many data processing problems. This is particularly the case when integrating LiDAR and other datasets (e.g., optical, radar or other LiDAR datasets) as the data can be indexed to the exact grid on which the other data has been acquired, thereby avoiding the need for resampling and losing information from one of the datasets.

3.2 Limitations

While the grid based indexing has been very successful and has a number of performance benefits in terms of reading the data from the file when processing it at the index resolution, there are a number of cases where being able to access the file at multiple resolution more easily would be useful. Functionality for generating binned and interpolated products at resolutions that are a multiple of factor of the bin size are currently being tested. Although, the grid-based index can be used to significantly reduce the search space when extracting regions of data that do not directly map onto the index resolution or grid, a quad-tree (or similar) index could be beneficial. The addition of a quad-tree index is therefore being considered for inclusion into future versions of the file format, in addition to the existing grid based index.

When using the sequentially written SPD file, which results in the best reading performance, the requirement to write the file in a specific order (i.e., top to bottom, left to right) can cause

problems for very wide datasets, as each row has to be completely written before the next one. Although, this can be solved by using the non-sequentially binned SPD file structure, the read performance is penalised.

Finally, when compared to other binary file formats (e.g., LAS), the SPD format results in a large relative file size owing to the increased data structure complexity and the inclusion of the spatial index within the file. Unlike LAS, compression is used by default and the resulting SPD file size is approximately 5-10 % smaller for an SPD file when compared to an equivalent uncompressed LAS file. If LASzip (Isenburg 2011) is used, the resulting compressed LAS file is up to 75 % smaller than the equivalent SPD file. The variation depends on the structure of the data being stored. If each pulse only has a single discrete return then the SPD file will be storing two pieces of information (i.e., a pulse and point) per return compared to the one (i.e. point) within the LAS file. As the number of returns associated with the pulses increases, this difference correspondingly decreases. However, it should be recognised that the SPD files are storing data that cannot be stored in or derived from data contained within a LAS 1.X file.

3.3 Extensions for future data types

The next development within LiDAR sensors is anticipated to be the use of multiple wavelength scanners. These sensors will transmit pulses of different wavelengths (e.g., 1064 nm and 1550 nm) and the corresponding waveform or returns recorded. Such systems are expected to increase discrimination between foliage and woody materials within vegetated environments. Currently, the SPD format does not explicitly support these data types, other than as separate files per wavelength (a header parameter defines the laser wavelength). In the future, a further field could be added to the pulse data type defining the wavelength of the pulse allowing these data to be stored within a single file.

4. Conclusions

This paper has presented a new file format for the storage and processing of LiDAR within a HDF5 file. The format is specifically designed to support both traditional discrete return and waveform data, using a pulse (rather than point) based data model. Additionally, the format explicitly supports the storage and manipulation of terrestrial, airborne and spaceborne datasets. Finally, the format also contains flexible spatial indexing allowing easy, image-like, random spatial access to the pulse data through a number of coordinate systems and projections and linking with other remote sensing datasets from imaging sensors. It is expected that this format could replace or offer an alternative to existing methods of data storage for laser scanning data, particularly for full waveform datasets.

References

- American Society for Photogrammetry & Remote Sensing (ASPRS; 2010) LAS Specification version 1.3, October 2010.
- Finkel, R. and Bentley, J.L. (1974). Quad Trees: A Data Structure for Retrieval on Composite Keys. *Acta Informatica* 4 (1): 1–9
- Isenburg, M. (2011) LAsTools : converting, filtering, viewing, processing, and compressing LIDAR data. <http://www.cs.unc.edu/~isenburg/lastools/> [Last access 13th June 2011]
- Shan, J. and Toth C. K. (2009) Topographic and laser ranging and scanning: principles and processing. CRC Press. ISBN 978-1-4200-5142-1
- The Internet Engineering Task Force (IETF; 1996) RFC 1951: DEFLATE Compressed Data Format Specification version 1.3. <http://tools.ietf.org/html/rfc1951> [Last access 13th June 2011]

Tree detection, delineation, and measurement from LiDAR point clouds using RANSAC

Peter Tittmann¹, Sohail Shafii², Bruce Hartsough³, Bernd Hamann⁴

¹Department of Geography, UCD (University of California, Davis),
ptittmann@gmail.com

²Department of Computer Science, UCD, ssshafii@ucdavis.edu

³Department of Biological and Agricultural Engineering,
UCD, brhartsough@ucdavis.edu

⁴Department of Computer Science, UCD, hamann@cs.ucdavis.edu

Abstract

As Light Detection And Ranging (LiDAR) (point) data sets increase in resolution, earth scientists become more interested in detecting and delineating trees using LiDAR. The majority of conventional methods that detect and delineate trees convert point data into gridded surfaces. Unfortunately, this conversion process has the potential to introduce error. We improve a point-based geometric model fitting strategy based on “RANdom Sample Consensus” (RANSAC), known as *StarSac*, and compare the method’s results against field data. The analysis demonstrates that *StarSac* produces similar results to field data, and is a strong alternative to conventional methods.

Keywords: geometric, RANSAC, model fitting

1. Introduction

As the capabilities of aerial Remote Sensing (RS) technologies such as Light Detection And Ranging (LiDAR) increase in precision, the potential to directly measure vegetation characteristics has increased as well. The majority of methods using LiDAR data for individual tree detection and delineation emphasize the conversion of point data into gridded surfaces, and the application of algorithmic tools widely used for terrain surface analysis to identify and delineate individual trees. While these methods have been shown to be effective under a range of circumstances, the interpolation of points to a gridded surface followed by watershed, valley-following, or other such methods have the potential to introduce error from both steps. To reduce the impact of such errors, methods are often parameterized by field data (species, canopy height, etc.) and as a result can be quite accurate in tree detection and delineation. If extensive field data collection is required to parameterize the algorithm, the efficiency of the inventory effort is compromised.

We have developed a “RANdom SAMple Consensus” (RANSAC)-based (Fischler and Bolles 1981) program, henceforth referred to as *StarSac*, which uses a geometric model fitting strategy to identify individual tree crowns directly from point data. Basal area is then calculated using a regression, relating Diameter at Breast Height (DBH, 1.37 m) to total height. We identify the location, and height of all trees within the scene without parameterization from field data using the point data directly and avoid the compounding error problem described above. Field data is used to parameterize the height-DBH regression for estimation of DBH.

1.1 Study site

The study site is the *van Eck forest*, consisting of 879 ha of mixed conifer forestland in Humboldt County, California. The van Eck forest contains a great deal of structural and successional diversity, including riparian areas, selectively logged second growth mixed conifer

stands, and stands which have not been harvested since initial clear cut in the early part of the century. The forest is divided in to four tracts, ranging from 130 ha to 315 ha (See Figure 1). It is a well-stocked second growth forest with an average timber volume of $170 \text{ m}^3 \text{ ha}^{-1}$.

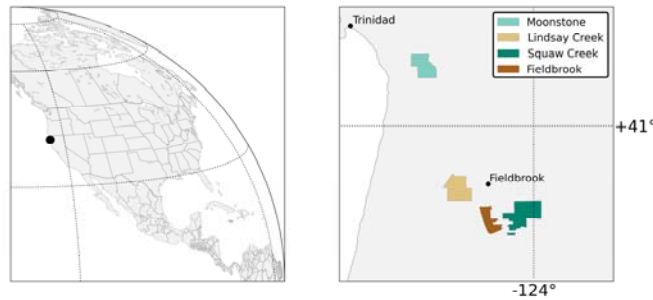


Figure 1: van Eck Study area location and management units.

2. Background

2.1 Detection and delineation

The majority individual tree identification methods using LiDAR data combine the creation of a Digital Surface Model (DSM) and Canopy Height Model (CHM) with Local Maxima (LM) filtering to identify tree locations (Jensen et al. 1987; Kaartinen et al. 2008; Lee et al. 2010; McCombs et al. June 2003; Tesfamichael et al. 2009, 2010). The efficacy of methods relying on LM filtering for tree detection depends on the determination of an analysis window size that reflects the crown area of the trees being identified. Popescu et al. (2002) tested an approach using a variable window defined by stand specific field data. However, this approach is dependent on field observation of crown radii to calibrate the LM window size.

Several methods have been tested to estimate tree crown and bole parameters (radius, bole volume, etc) once tree location has been established. Region growing methods such as the watershed delineation adapted from terrain analysis are common (Hyypä et al. 2001; Schardt et al. 2002; Ziegler et al. 2000). Geometric models of tree crown shapes have also been used to delineate individual trees in Airborne Laser Scanning (ALS) data. Abstract tree crown form was first described by Horn (1971). Shapes were further modified by Pollock (1996). Geometric shapes have been used by others to measure tree crowns from aerial photography and LiDAR (Gong et al. 2002; Holmgren et al. 2003; Pollock 1996; Sheng et al. 2001; Wolf and Heipke 2007, Persson 2001, Persson et al 2002, Andersen et al. 2002, Popescu 2003, Wack et al. 2003, Falkowski et al. 2006, Wolf and Heipke 2007, Heurich 2008, Kaartinen et al. 2008).

2.2 RANSAC, *StarSac*

RANSAC (Fischler and Bolles 1981) is a paradigm for fitting experimental data to a mathematical model. RANSAC has notable advantages over other canopy-fitting approaches as it iteratively determines the best set of points fitting a model within the point cloud. It has successfully been applied to the detection of objects from point clouds (Bretar and Roux 2005; Fontanelli et al. 2007; Forlani et al. 2003, 2006; Reitberger et al. 2007, 2009; Schnabel et al. 2007; Tarsha-Kurdi et al. 2007). We have revised *StarSac* (Shafii et al. 2009), a program was developed using RANSAC and Oliver Kreylos' Virtual Reality Toolkit (VRUI), and verified the results against those of a field survey. Unlike other projects, we used a modified version of RANSAC based on a preliminary maxima filter to find and measure tree canopies.

3. Methods

3.1 Field data collection

The forest consists of four tracts, which were further divided into twenty-one stands ranging in size from four to forty ha. Stand inventory was taken for standing live trees greater than 15.24 cm Diameter at Breast Height (1.37 m) (DBH). The primary objective of the inventory was to estimate total biomass and by extension total forest carbon. The variable plot method outlined in Dilworth and Bell (1963) was used. A Basal Area Factor (BAF) was selected for each stand prior to sampling to produce an average of six to eight “in” trees per plot. Plots were spaced across each tract on a 50m x 100m grid. A total of 660 measure plots were installed.

3.2 Regression models

A regression model of the allometric relationship between height and DBH derived from field data was used to predict DBH from LiDAR derived tree heights. A general non-linear regression model for all species was used to establish DBH from height using measured trees from plots within the same stand. The equation used for regression analysis takes the standard form of:

$$DBH = a * H^b \quad (1)$$

Where H is total tree height; a and b are spatially variable regression coefficients.

3.3 LiDAR data collection

LiDAR data were collected for 1796 ha on March 17th, 2008, conducted with an Optech 3100 sensor mounted in a Cessna Caravan 208B, with specifics shown in Table 1. Instrumentation was set to yield an average native pulse density of ≥ 6 pls/m² over terrestrial surfaces. The TerraScan® software suite was used to classify ground and non-ground points (Soninen 2004).

Table 1: Data Collection statistics.

Sensor	Optech 3100
Survey Altitude (AGL)	900 m
Pulse Rate	> 71 kHz
Pulse Mode	Single
Mirror Scan Rate	52 Hz
Field of View	28° (\pm from nadir)
Overlap	100% (50% Side-lap)

3.4 StarSac

The current RANSAC algorithm is summarized in pseudo-code below, with references to sections that explain key parts of the algorithm in greater detail.

- 1) For all locally maximal LiDAR points Max_i for $i=1 \dots n$: (Section 3.4.1)
 - a) Find set of points around Max_i to create a fixed-size window, one large enough to contain most canopies.
 - b) For iterations $j=1 \dots T$: (Section 3.4.2)
 - i) Randomly select a subset of window points to create model M_j with Max_i as its peak, reject if shape is inappropriate. (Section 3.4.3)
 - ii) Create consensus set C_j for M_j , determine radius. (Section 3.4.4)

- iii) If M_j has at most a (predefined) ratio of outliers to inliers, grade and compare it with the best model M_{best} . Otherwise, ignore it. Keep track of M_{best} . If no previous model was found then M_j is chosen as M_{best} assuming that its ratio of outliers to inliers is appropriate. (Section 3.4.4)
 - iv) If M_{best} found, mark it.
- 2) Visualize canopy-approximating models as shaded surfaces for delineation. (Figure 6)
 - 3) Calculate a height for each canopy-approximating model. (Section 3.4.5)

3.4.1 Local Maxima

Maximum points Max_i are first return points (classified as non-ground) identified during pre-processing. Each maximum is selected based on the fact that it is higher than points inside of a $1.5m \times 1.5m$ window centered on the maximum (see Figure 2). This small box is large enough to capture most peaks in our test data sets.

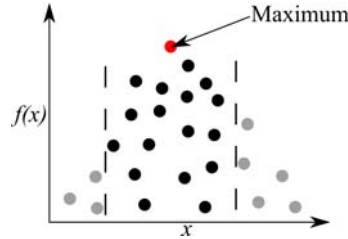


Figure 2: An artificial sketch of a LiDAR point that has been identified as a maximum where each point is defined by the one-dimensional function $f(x)$. The maximum is higher than the other points inside of the box indicated by the dashed lines.

3.4.2 Number of Model Iterations

We calculate the number of RANSAC iterations T according to Schnabel et al. (2007), based on the probability that an appropriate model is found. If a window consists of N points, k points are used to instantiate our model. A good consensus set (i.e., inliers) consists of at least c points. The probability of finding an appropriate model in a single pass is:

$$P(c) = \frac{\binom{c}{k}}{\binom{N}{k}} \approx \left(\frac{c}{N}\right)^k \quad (2)$$

After picking s poor models, the probability of detecting an appropriate model is calculated by evaluating $P(c,s)$:

$$P(c,s) = 1 - (1 - P(c))^s \quad (3)$$

If we were to solve for s , we can calculate the number of candidates T required to detect shapes of size c with probability $P(c,T) \geq p_i$ as:

$$T \geq \frac{\ln(1 - p_t)}{\ln(1 - P(c))} \quad (4)$$

The denominator of (44) can be approximated by the Taylor series $\ln(1 - P(c)) = -P(c) + O(P(c)^2)$ and (44) can be written as:

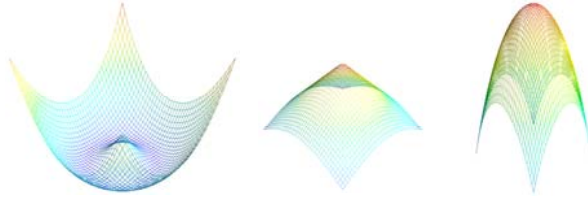
$$T \approx \frac{-\ln(1 - p_t)}{P(c)} \quad (5)$$

We use (55) to define our number of iterations per window.

3.4.3 Model Creation

Our algorithm is currently capable of fitting a “shape-shifter” to data points. Each shape has a parameter defining its central peak (x_c, y_c, z_c) (defined by Max_i used to create the current window) and one parameter α which interpolates between a cone and a paraboloid. If α is zero, the shape is similar to a cone; if it is one, it resembles a paraboloid (Figure 3).

$$f(x, y) = -\alpha * ((x - x_c)^2 + (y - y_c)^2) + (\alpha - 1.0) * \sqrt{((x - x_c)^2 + (y - y_c)^2)} + z_c \quad (8)$$



(a) Shape-shifter with $\alpha=-1.0$. (b) Shape-shifter with $\alpha=0.0$. (c) Shape-shifter with $\alpha=1.0$.

Figure 3: Renderings of the shape-shifter (88) using various values of α . The valid range for α is 0-1.0, as an α value of -1.0 creates an unusual shape as shown in Figure 3(a).

3.4.4 Consensus Set and Model Grading

Each model’s consensus set C_j is created by selecting inlier points within the window. We assemble C_j by computing the $f(x, y)$ value for each window point and comparing that value with the point’s z coordinate. If the difference is smaller than a pre-defined error metric ϵ , the point is added to C_j . From C_j , one can then calculate the radius by calculating the two-dimensional, (x, y) distance between the model’s central peak and the inlier point furthest from the peak. In order to compare against other models, each model is graded based on the number of inliers. The model with the best grade (most inliers) is chosen.

3.4.5 Tree height

To calculate the height for each tree, we use Hardy’s (1971) multiquadric method to reconstruct the ground surface beneath the tree crown. We then subtract the elevation of the hardy surface at the (x, y) location of the crown apex from the height (z -coordinate) of the crown apex.

3.5 LiDAR inventory

StarSac was run in a batch process over all LiDAR data in stands where field data was collected. The classified LAS files were subset into blocks containing $\approx 150,000$ points. Each block was further subset into ground-only points and points classified as first return and vegetation. LiDAR data was processed using the LibLAS (Loskot 2008) command line tools and application programming interface (API). *StarSac* output was collected in a PostgreSQL (The PostgreSQL Global Development Group 2005) relational database with the PostGIS (Holl and Plum 2009) spatial object extensions.

To test the accuracy of the tree identification method outlined above, “in” trees were identified from the LiDAR-derived trees based upon DBH, BAF, and distance to the nearest plot center used in the field inventory. Stand-specific, non-linear regression coefficients derived from field data were used to regress LiDAR-derived tree height to tree DBH. Once DBH was modeled, the status of all trees with regard to the BAF was assessed based upon the distance between the tree and the plot center.

4. Results

4.1 Field inventory

The BAF selected for each plot resulted in between 5 and 8 trees per plot. Basal area ranges between $36 \text{ m}^2\text{ha}^{-1}$ to $74 \text{ m}^2\text{ha}^{-1}$ and generally varies with DBHq though with greater magnitude.

4.1.1 Regression models

The LiDAR methods (*StarSac*) employed here do not differentiate between species, thus regression coefficients were developed using all trees within a given stand. Visual inspection of Figure 4 indicates similarity between species, justifying the application of *StarSac*. Figure 5 represents the results of the regression of DBH and height for all field-measured trees.

4.2 LiDAR analysis

Individual tree identification using the RANSAC method was effective in most cases. Visual inspection of the consensus sets (Figure 6) identified was used extensively in testing the impact

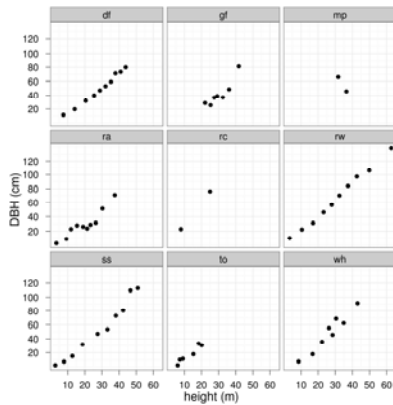


Figure 4: Average DBH within height bins. df, Douglas fir; gf, Grand; mp, big-leaf maple; ra, Red alder; rc, Western red cedar; rw, Coast redwood; ss, Sitka spruce; to, Tan oak; wh, Western hemlock.

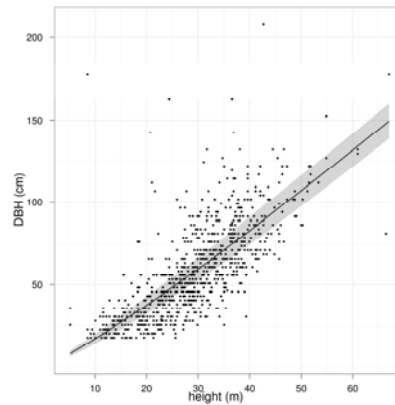
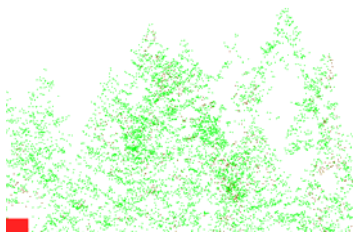
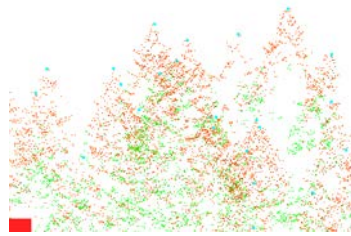


Figure 5: Height-DBH regression for all trees.

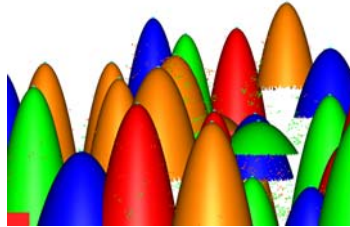
of changes to the approach.



(a) Raw LiDAR points.



(b) Raw LiDAR points with consensus sets and maximum points.



(c) Raw LiDAR points with rendered graphs.

Figure 6: An example of three-dimensional graphs that depict the models used to identify the canopies in StarSac. Each graph is created by using the model's equation and radius. The raw LiDAR rendering is showing in Figure 6(a). In Figure 6(b), central peak points are rendered as thick, cyan points and consensus set points are colored red. The resulting graphs of the canopies are shown in Figure 6(c).

4.3 LiDAR inventory

The LiDAR inventory method was compared with field methods for the determination of basal area. Stand-aggregated basal area estimates were derived using the basal area calculated for each plot. Figures 7 and 8 show the basal area and tree count comparison between methods.

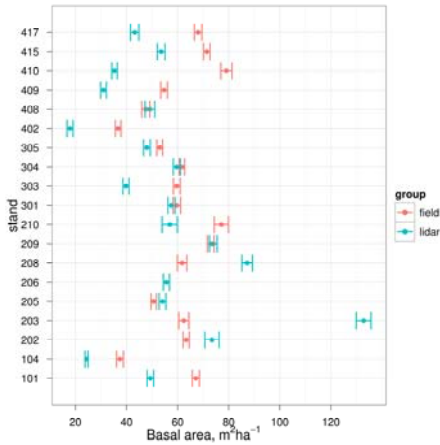


Figure 7: Average plot basal area comparison between LiDAR and field methods with 95% confidence interval whiskers.

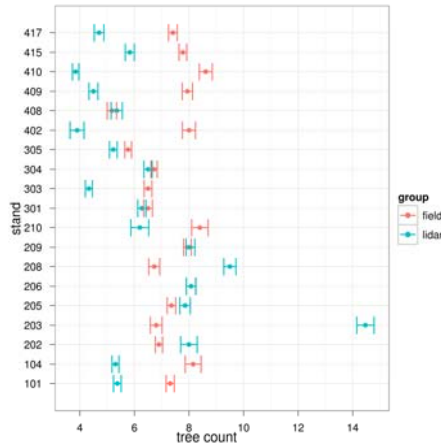


Figure 8: Average plot tree count comparison between LiDAR and field methods with 95% confidence interval whiskers.

Figure 9 indicates that the LiDAR method results in greater basal area estimation in the 50 cm to 125 cm DBH range while field methods estimate greater basal area in the 12 cm to 50 cm and 150 cm to 350 cm ranges.

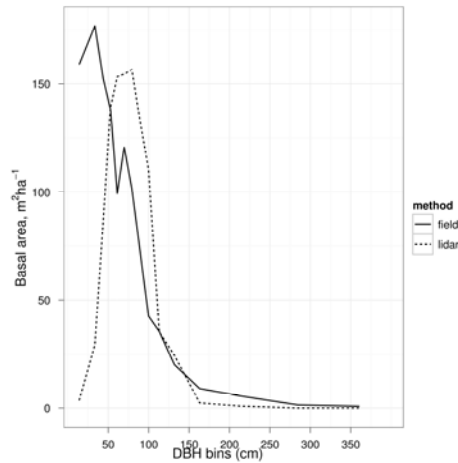


Figure 9: Density plot of basal area by DBH histogram bins for LiDAR and field methods.

The results show strong agreement in many stands while a few stands show some disparity. In general, field methods resulted in a higher tree count and greater basal area.

Table 2 shows tree count/plot averaged across all plots. While there are significant differences in the count and basal area between the LiDAR and field methods for given stands, the combined statistics indicate that the results of field and LiDAR methods are within a RMSD of less than $3.5 \text{ m}^2\text{ha}^{-1}$ (see Table 3).

Table 2: Average plot tree count and basal area for all plots.

LiDAR count	Field count	(average) LiDAR basal area, m^2ha^{-1}	(average) Field basal area, m^2ha^{-1}
6.64	7.17	57.07	60.33

Table 3: Paired t test and RMSD for tree count and basal area for all plots.

df	Tree count t	Tree count P	BA ^a t	BA ^a P	Count RMSD	BA ^a RMSD
301	-1.807	0.072	-1.278	0.202	0.530	3.261

^a basal area

Paired t test between LiDAR and field methods indicate that the means for all paired plots were not significantly different for tree counts ($P=0.072$) or basal area ($P=0.202$).

4.4 Factors affecting agreement between LiDAR and field measures

While overall differences between the methods are not significant, there is obviously some variation between the means for LiDAR and field methods for both tree count and basal area (Figures 7 and 8). As the van Eck forest stands were classified by canopy density and DBH classification, we can assess the impact of these generalized stand characteristics on the differences between the two measurement methods. This is accomplished using ANalysis Of

Variance (ANOVA) for the regression of the stand characteristics (*predictor*) and the log-transformed ratio of basal area estimates (*response*) from the two basal area estimation methods

A two-way ANOVA reveals that neither the interaction of canopy cover and DBH, or DBH alone has significant effects, but that canopy cover alone has an impact upon the variation between measures (Table 4). A one-way ANOVA (Table 5) reveals that the variance in estimated basal area between methods is significant in stands classified in the 40-60% canopy coverage range. Levene’s test reveals that the variance is homogeneous between canopy cover levels (Table 6), validating the assumptions in the one-way test.

Table 4: Two-way ANOVA test for the influence of canopy cover and DBH classification on variance between estimation methods.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dbh class (dbh)	1	0.00	0.00	0.00	0.9972
canopy density class (dens)	1	5.81	5.81	8.59	0.0036
dbh:dens	1	0.95	0.95	1.41	0.2361
Residuals	298	201.39	0.68		

Table 5: One-way ANOVA test for the influence of canopy cover on variance between estimation methods.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.1984	0.1390	-1.43	0.1547
40-60% cover	-0.7076	0.2949	-2.40	0.0170
60-80% cover	0.0702	0.1482	0.47	0.6359

Table 6: Levene’s test of one-way ANOVA residuals from the influence of canopy cover on differences between basal area measures.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.6413	0.0879	7.29	0.0000
40-60% cover	-0.0336	0.1865	-0.18	0.8571
60-80% cover	-0.0074	0.0937	-0.08	0.9375

Figure 10 shows the effect of canopy density on tree count and basal area estimation by LiDAR and Field methods. Mid-density stands (40%-60%) show significant variation in basal area estimates. Diameter classes did not strongly influence differences in tree counts or basal area estimation by the two methods (Table 4). Figure 11 shows the variability between methods across the range of DBH classes and suggests that variation between basal area methods is greatest in the 0 cm to 20 cm DBH and 61 cm to 81 cm DBH classes.

5. Conclusions, Future Research

This research gives insight into forest inventory from aerial LiDAR data across forest stands that are heterogeneous with regard to management history, species mix, and site characteristics. Tree detection and height estimation is accomplished without the use of regression models or gridded data. The replication of a variable plot method was used so that results of field and RANSAC-based LiDAR methods are comparable. The comparisons in Section 4 indicate that the tree identification and delineation-based LiDAR inventory method, applied to dense, mixed-species stands on variable terrain,

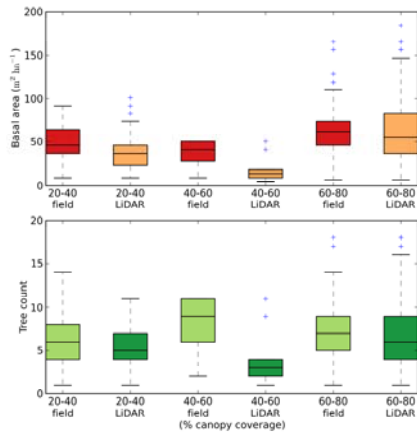


Figure 10: Variation between LiDAR and field based tree count and basal area estimates by percent canopy cover classes. Box extends from the lower to upper quartile values of the data, with a line at the median.

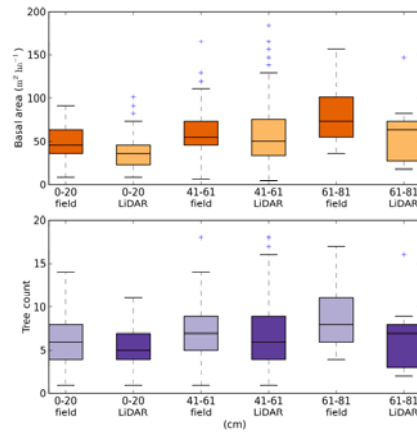


Figure 11: Variation between LiDAR and field based tree count and basal area estimates by diameter classification. Box extends from the lower to upper quartile values of the data, with a line at the median.

yields similar tree count and basal area estimates to the field inventory method. The small difference between average basal area estimated by the two methods indicates that though some trees are not identified using the LiDAR method, they tend to be smaller trees and contribute less to the aggregate statistics.

There exist multiple directions for future research. We intend on refining the RANSAC method such that LM filtering may be eliminated from the algorithm and the sub-canopy vegetation characteristics can be assessed as well. Additionally, we will investigate ways to incorporate methods identified in (Schnabel et al. 2007) for preliminary selection of model parameters using point normals. We hope to improve *StarSac* so that it may provide information about a range of structural measures relating to habitat, fire behavior, and forest health.

Acknowledgements

The authors would like to thank Oliver Kreylos for providing thoughtful commentary on the manuscript.

References

- H.E. Andersen, S.E. Reutebuch, and G.F. Schreuder. Bayesian object recognition for the analysis of complex forest scenes in airborne laser scanner data. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 34(3/A):35–41, 2002. ISSN 1682-1750.
- F. Bretar and M. Roux. Hybrid image segmentation using LiDAR 3D planar primitives. In *Laser scanning 2005*, pages 12–14. ISPRS, 2005.
- J.R. Dilworth and J.F. Bell. *Variable plot cruising*. OSU Book Stores, 1963.
- M.J. Falkowski, A.M.S. Smith, A.T. Hudak, P.E. Gessler, L.A. Vierling, and N.L. Crookston. Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of lidar data. *Canadian Journal of Remote Sensing*, 32(2): 153–161, 2006.
- Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*,

- 24:381–395, June 1981. ISSN 0001-0782. URL <http://doi.acm.org/10.1145/358669.358692>.
- D. Fontanelli, L. Ricciato, and S. Soatto. A fast ransac-based registration algorithm for accurate localization in unknown environments using lidar measurements. In *IEEE International Conference on Automation Science and Engineering, 2007. CASE 2007*, pages 597–602, 2007.
- G. Forlani, C. Nardinocchi, M. Scaioni, and P. Zingaretti. Building reconstruction and visualization from LIDAR data. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY REMOTE SENSING AND SPATIAL INFORMATION SCIENCES*, 34(5/W12):151–156, 2003. ISSN 1682-1750.
- G. Forlani, C. Nardinocchi, M. Scaioni, and P. Zingaretti. Complete classification of raw LIDAR data and 3D reconstruction of buildings. *Pattern Analysis & Applications*, 8(4): 357–374, 2006. ISSN 1433-7541.
- P. Gong, Y. Sheng, and G.S. Biging. 3D Model-based tree measurement from high-resolution aerial imagery. *Photogrammetric engineering and remote sensing*, 68 (11):1203–1212, 2002. ISSN 0099-1112.
- Rolland L. Hardy. Multiquadric equations of topography and other irregular surfaces. *J. Geophys. Res.*, 76(8):1905–1915, 1971. ISSN 0148-0227. 10.1029/JB076i008p01905. URL <http://dx.doi.org/10.1029/JB076i008p01905>.
- Marco Heurich. Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the bavarian forest national park. *Forest Ecology and Management*, 255(7):2416 – 2433, 2008. ISSN 0378-1127. URL <http://www.sciencedirect.com/science/article/B6T6X-4S02JX3-1/2/453e6ff86a62cf2643f94d558cfaff71>. Large-scale experimentation and oak regeneration.
- Stephan Holl and Hans Plum. PostGIS. *GeoInformatics*, 03/2009:34–36, April 2009. URL <http://fluidbook.microdesign.nl/geoinformatics/03-2009/?page=34>.
- J. Holmgren, M. Nilsson, and H. Olsson. Estimation of tree height and stem volume on plots using airborne laser scanning. *Forest Science*, 49:419–428(10), 2003. URL <http://www.ingentaconnect.com/content/saf/fs/2003/00000049/0000003/art00009>.
- H.S. Horn. *The adaptive geometry of trees*. Princeton Univ Pr, 1971. ISBN 0691023557.
- J. Hyypä, O. Kelle, M. Lehikoinen, and M. Inkinen. A segmentation-based method to retrieve stem volume estimates from 3-d tree height models produced by laser scanners. *Geoscience and Remote Sensing, IEEE Transactions on*, 39(5):969–975, May 2001. ISSN 0196-2892.
- John R. Jensen, Michael E. Hodgson, Halkard E. Mackey, and William Krabill. Correlation between aircraft mss and lidar remotely sensed data on a forested wetland. *Geocarto International*, 2(4):39–54, 1987.
- H. Kaartinen, J. Hyypä, X. Liang, P. Litkey, A. Kukko, X. Yu, H. Hyypä, and M. Holopainen. Accuracy of automatic tree extraction using airborne laser scanner data. *SilviLaser*, 2008:8th, 2008.
- Oliver Kreylos. Vrui vr, 2011. URL <http://idav.ucdavis.edu/~okreylos/ResDev/Vrui/index.html.v2.0>.
- H. Lee, KC Slatton, BE Roth, and WP Cropper. Adaptive clustering of airborne LiDAR data to segment individual tree crowns in managed pine forests. *International Journal of Remote Sensing*, 31 (1):117–139, 2010. ISSN 0143-1161.

- Mateusz Loskot. Introduction to asprs las data processing with liblas. In *FOSS4G 2008*, 2008.
URL
<http://www.osgeo.org/ocs/index.php/foss4g/2008/paper/view/362>.
- J.W. McCombs, S.D. Roberts, and D.L. Evans. Influence of fusing lidar and multispectral imagery on remotely sensed estimates of stand density and mean tree height in a managed loblolly pine plantation. *Forest Science*, 49:457–466(10), June 2003. URL
<http://www.ingentaconnect.com/content/saf/fs/2003/00000049/0000003/art00013>.
- A. Persson, J. Holmgren, and U. Soderman. Detecting and measuring individual trees using an airborne laser scanner. *PE & RS- Photogrammetric Engineering & Remote Sensing*, 68(9):925–932, 2002. ISSN 0099-1112.
- Åsa Persson. Extraction of individual trees using laser radar data. Master’s thesis, Chalmers University of Technology, Department of Signals and Systems, Göteborg, Sweden., 2001.
- Richard James Pollock. *The automatic recognition of individual trees in aerial images of forests based on a synthetic tree crown image model*. PhD thesis, University of British Columbia, 1996.
- S.C. Popescu, R.H. Wynne, and R.F. Nelson. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, 29(5): 564–577, 2003.
- Sorin C. Popescu, Randolph H. Wynne, and Ross F. Nelson. Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. *Computers and Electronics in Agriculture*, 37 (1-3):71 – 95, 2002. ISSN 0168-1699. URL
<http://www.sciencedirect.com/science/article/B6T5M-478RN2H-6/2/55e9624eb4068f98ed9e04011f1d74fc>.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- J. Reitberger, P. Krzystek, and U. Stilla. Combined tree segmentation and stem detection using full waveform lidar data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36:332–337, 2007.
- J. Reitberger, C. Schnorr, P. Krzystek, and U. Stilla. 3D segmentation of single trees exploiting full waveform LIDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6):561–574, 2009.
- M. Schardt, M. Ziegler, A. Wimmer, R. Wack, and J. Hyypä. Assessment of forest parameters by means of laser scanning. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 34(3/A.):302–309, 2002. ISSN 1682-1750.
- R. Schnabel, R. Wahl, and R. Klein. Efficient ransac for point-cloud shape detection. *Computer Graphics Forum*, 26(2):214–226, 2007. ISSN 1467-8659. URL
<http://dx.doi.org/10.1111/j.1467-8659.2007.01016.x>.
- S. Shafii, B. Hamann, R. Hutchison, O. Kreylos, and J. Viers. Tree Detection and Delineation of the Consumnes River Preserve. In B.G. Lees and S.W. Laffan, editors, *10th International Conference on GeoComputation*, Sydney, November-December 2009. University of New South Wales.
- Y. Sheng, P. Gong, and GS Biging. Model-based conifer-crown surface reconstruction from high-resolution aerial images. *PE & RS- Photogrammetric Engineering and Remote Sensing*, 67(8):957–965, 2001. ISSN 0099-1112.
- A. Soninen. *TerraScan User’s Guide*. Terrasolid, 2004.

- F. Tarsha-Kurdi, T. Landes, and P. Grussenmeyer. Hough-transform and extended RANSAC algorithms for automatic detection of 3D building roof planes from Lidar data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(3/W52):407–412, 2007.
- S.G. Tesfamichael, F. Ahmed, J.A.N. van Aardt, and F. Blakeway. A semi-variogram approach for estimating stems per hectare in eucalyptus grandis plantations using discrete-return lidar height data. *Forest Ecology and Management*, 258(7):1188 – 1199, 2009. ISSN 0378-1127. URL <http://www.sciencedirect.com/science/article/B6T6X-4WMD2J9-4/2/allab87a82802ef1a61ccde32fe2ea43>.
- S.G. Tesfamichael, J.A.N. van Aardt, and F. Ahmed. Estimating plot-level tree height and volume of eucalyptus grandis plantations using small-footprint, discrete return lidar data. *Progress in Physical Geography*, 34(4): 515–540, 2010. URL <http://ppg.sagepub.com/content/34/4/515.abstract>.
- The PostgreSQL Global Development Group. PostgreSQL 8.1.4 documentation, 2005. URL <http://www.postgresql.org/docs/manuals/>.
- P. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78 (1):138–156, 2000.
- R. Wack, M. Schardt, U. Lohr, L. Barrucho, and T. Oliveira. Forest inventory for eucalyptus plantations based on airborne laserscanner data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2003.
- Bernd-Michael Wolf (né Straub) and Christian Heipke. Automatic extraction and delineation of single trees from remote sensing data. *Machine Vision and Applications*, 18:317–330, 2007. ISSN 0932-8092. URL <http://dx.doi.org/10.1007/s00138-006-0064-9>. 10.1007/s00138-006-0064-9.
- Michaela Ziegler, Harald Konrad, Johannes Hofrichter, Andreas Wimmer, Georg S. Ruppert, Mathias Schardt, and Juha M. Hyypä. Assessment of forest attributes and single-tree segmentation by means of laser scanning. In Gary W. Kamerman, Upendra N. Singh, Christian Werner, and Vasyl V. Molebny, editors, *Topography, Altimetry, and Vegetation Monitoring Systems*, volume 4035, pages 73–84. SPIE, 2000. 10.1117/12.397780. URL <http://link.aip.org/link/?PSI/4035/73/1>.

Mobile terrestrial laser scanning in urban tree inventory

Markus Holopainen¹, Mikko Vastaranta¹, Ville Kankare¹, Tuula Kantola¹, Harri Kaartinen², Antero Kukko², Matti Vaaja³, Juha Hyyppä², Hannu Hyyppä³

¹University of Helsinki, Finland, first.last@helsinki.fi

²Finnish Geodetic Institute, first.last@fgi.fi

³Aalto University School of Science and Technology, first.last@tkk.fi

Abstract

In this study we evaluated an accuracy of mobile terrestrial laser scanning (MLS) measurements in urban tree inventory. The MLS data were collected in August 2010 with the FGI Roamer mobile mapping system, consisting of a Faro LS 880 laser scanner and a NovAtel HG1700 SPAN58 INS system mounted in a car. Study areas were divided to park and urban forest located in Seurasaari, Helsinki, Finland. Studied inventory characteristics were tree location and –diameter-at-breast-height (*dbh*). Reference measurements consisted altogether from 201 trees, which locations were measured using RTK GPS and static terrestrial laser scanning while tree *dbhs* were measured using steel callipers. Tree mapping and *dbh* measurements were done from the MLS point clouds manually. Tree location accuracy of MLS measurements were 0.72 m in park and 0.47 cm in forest conditions as root mean squared error in *dbh* were 7.0% and 12.5%, respectively. MLS can be used in urban tree inventory in targets with a good visibility and relatively dense network of roads or paths.

Keywords: Urban forestry, mapping, diameter-at-breast height, monitoring

1. Introduction

The urban forests and parks are highly important for many reasons. Urban forest are used for recreation, they provide scenery values and maintain biodiversity. In the city of Helsinki, there are 902 ha of managed parks and 4020 ha of urban forests. The City of Helsinki Park and Garden Department (HPGD) maintains digital tree register (appr. 20 000 trees) including species, height, diameter-at-breast-height (*dbh*), location, and vitality status of the tree. Tree register data is used in town and environmental planning, locating hazardous (for citizens) old trees and biodiversity monitoring.

Trees in the register are located, using tachymeter measurements or manually placing trees to the large-scale city planning map. Requirements for up-to-date map data also in city parks and forests are increasing and an important question is how to keep the digital databases up-to-date for the various applications. Traditional updating procedures such as visual interpretation of digital aerial images or field measurements using tachymeter are either inaccurate or very expensive.

Recently, development of laser scanning technology has opened new possibilities for tree mapping. Small-footprint airborne laser scanning (ALS) is a method based on laser range measurements from an aircraft and the precise orientation of these measurements between a sensor (the position of which is known by using a differential-GPS technique) and a reflecting object, the position of which (*x, y, z*) is to be defined. The ALS gives the georeferenced point cloud. By analyzing the point cloud by using pattern recognition methods, it is possible to locate

individual trees and determine individual tree heights among many other tree variables (Hyypä and Inkinen 1999).

Fixed-position (mounted on a tripod) terrestrial laser scanners offer a high potential for 3D mapping of smaller areas with detail. The principle of terrestrial laser scanning (TLS) is simple – a highly collimated laser beam scans over a predefined solid angle in a regular scan pattern and measures the time of flight of the laser signal. The scanning range of the middle-range terrestrial system allows distance measurements between 2 and 800 meters. TLS has been mainly used for detailed modelling of individual trees and canopies (Vastaranta *et al.* 2009).

A mobile-based terrestrial laser scanning (MLS) system is a modification of the ALS. It resembles ALS by having a laser scanner, a GPS receiver, an IMU and preferably camera(s), but it is operated from the top of the moving ground vehicle, such as car or a harvester, and it is used for shorter distances. Due to shorter operating distances, it can more easily have a higher pulse rate than an ALS. At present, it is possible to use software and methods developed for TLS and ALS, but due to different scanning geometry, changing point density as a function of range and the fast processing needed, algorithms for MLS data processing need to be developed separately. The MLS also provides new possibilities to monitor city forests. Currently, there is work towards characterisation of the quality of city trees with the MLS (Fig. 1).



Figure 1. Representation of a point cloud collected with MLS from study area. Trees are well characterised.

RTK (real-time kinematic) is an accurate global positioning system (GPS)-based method for measuring absolute positions to centimetre accuracy in open areas (e.g. Bilker and Kaartinen 2007). However, dense canopy often disturbs GPS measurements. In practice, it is inefficient to measure every tree position with GPS.

The City of Helsinki Park and Garden Department commissioned the University of Helsinki, Department of Forest Sciences, Finnish Geodetic institute and Aalto University, School of science and technology to evaluate an accuracy and efficiency of mobile laser measurement method that can be utilized in city tree mapping and monitoring. In this paper we describe our first results of the study.

2. Method

2.1 Study area

Seurasaari is a popular outdoor recreation area, located approximately 5 km from the Helsinki city center. It was made a public park in 1890 and quickly became a popular place for recreation activities. Seurasaari receives hundreds of thousands of visitors a year. Seurasaari is a wooded island with rocks, hills, wetlands and herb-rich forests covering about 46 hectares. Our study area in Seurasaari consists of: 1) an urban park (approximately 1.2 ha) (Fig. 2) with a mainly old oaks growing far apart and only grass as a understory vegetation, 2) urban forest area (approximately 0.8 ha) which can be classified as semi-natural old-growth forest. In the both areas there is a dense network of man-made outdoor paths that are driveable also with a vehicle.

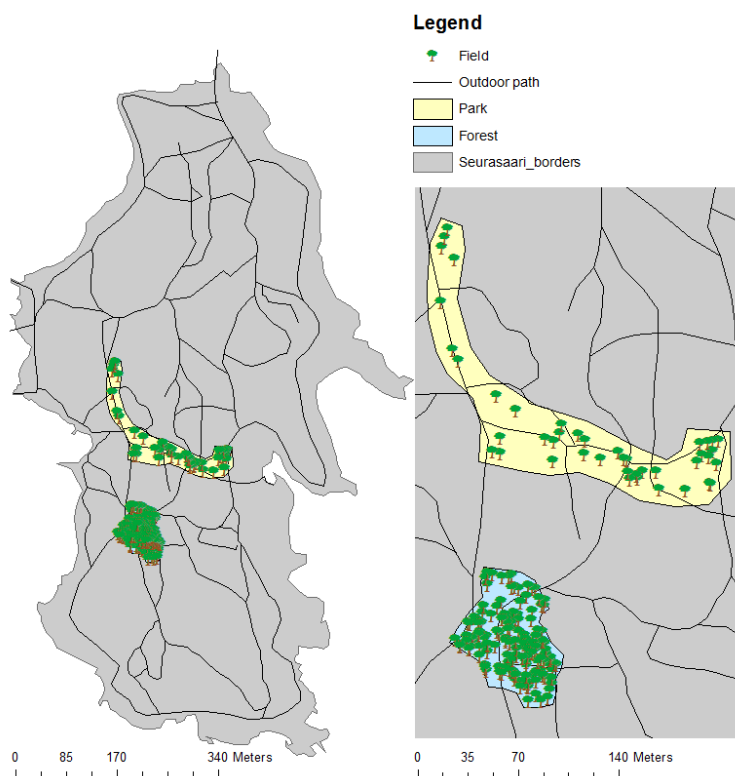


Figure 2. Study areas in Seurasaari.

2.2 GPS and tree diameter measurements

The field measurements were carried out in July 2010. Altogether 201 trees, including nine different tree species were mapped in the field using direct GPS measurements. The GPS device used for measurements was GEOXT 2008 (Trimble Navigation Ltd., Sunnyvale, CA, USA). The *dbh* and tree species were determined for trees with a *dbh* larger than 50 mm with steel callipers. General statistics from field trees based on field measurements are presented in table 1 and 2. GPS measurements were used to form a preliminary tree map.

Table 1. Dbh statistics from field measurements in park area.

Species	PARK				
	n	min	mean	max	stdev
Acer platanoides	1	356	356.0	356	0.0
Alnus sp.	1	357	357.0	357	0.0
Betula sp.	2	491	495.5	500	4.5
Picea abies	2	451	455.5	460	4.5
Pinus sylvestris	6	434	478.2	500	30.9
Quercus robur	23	138	408.9	500	100.1
Sorbus aucuparia	4	67	157.0	250	68.5
Tilia cordata	2	500	500.0	500	0.0
Ulmus sp.	3	500	500.0	500	0.0
Total	44	67	415.3	500	111.4

Table 2. Dbh statistics from field measurements in forest area.

Species	FOREST				
	n	min	mean	max	stdev
Acer platanoides	7	78	219	460	138.5
Alnus sp.	17	84	247.7	415	87.8
Betula sp.	12	56	229	500	131.5
Picea abies	50	50	233	500	109.3
Pinus sylvestris	26	51	299	500	159.2
Populus tremula	13	50	241	347	90.9
Salix caprea	2	363	373	382	13.4
Sorbus aucuparia	30	58	104.5	317	55.1
Total	157	50	231.0	500	128.8

2.3 Terrestrial laser scanning

TLS data was collected in September 2010, using a Leica HDS6000 terrestrial laser scanner. The scanner uses phase-shift measurements of continuous waves to measure the distances, with a data acquisition rate of 500 000 points per second. The point spacing is 6.3 mm at the distance of 10 m. For this study, several scannings were done from the area and point clouds were georeferenced. Tree detection was done manually, using TerraScan software. The scanner positions were measured with a RTK (real-time kinematic) GPS and reference locations for trees were determined using TLS.

2.4 Mobile terrestrial laser scanning

The MLS data were collected in August 2010 with the FGI Roamer mobile mapping system developed at the Finnish Geodetic Institute (Fig. 3). The Roamer consists of a Faro LS 880 laser scanner with a measurement frequency of 120 kHz and a NovAtel HG1700 SPAN58 INS system. With slightly modified hardware for the standard FARO LS, it provides so-called tunnel mode, or profile measurements, synchronized with external positioning and data logging

systems. This information is needed to derive the position and attitude information for each 3D point produced by the laser scanner. The mirror rotation frequency, or scan rate of the scanner on the Roamer can be set to 3-30 Hz, thus giving a vertical angular resolution of 0.0096-0.096 degrees (0.17-1.7 mrad), respectively. Corresponding point spacing at a typical scanning range of 15 metres in road mapping is thus 2.5-25 mm in the scanning plane. Tree detection was done manually from the MLS point clouds using TerraScan software.



Figure 3. MLS measurement unit, ROAMER

2.5 Evaluation of tree location and *dbh* accuracy

Tree location and *dbh* measurement accuracy for MLS was evaluated. Reference used for *dbh* was measured in the field using steel callipers. Traditionally the City of Helsinki Park and Garden Department produces tree maps using tachymeter measurements. Based on previous studies the location accuracy of the manual measurement from the static TLS data is comparable to tachymeter measurements (e.g. Mechelke et al. 2007). Therefore, manual measurements from TLS data were used as a reference for tree location accuracy. Statistical measures used were bias and root mean squared error (RMSE). Accuracies were studied in urban park and forest environment separately.

3. Result

3.1 Tree mapping accuracy

Location of trees detected from the MLS data were compared to the TLS (+RTK) locations. In park environment 97.7% of the trees were detected with location accuracy of 0.72 m. The respective stats in urban forest area were 68.2% and 0.54 cm (Table 3).

Table 3. Tree mapping accuracy.

Area	Detection, %	Location, m
Park	97.7	0.72
Forest	68.2	0.47
Total	72.6	0.54

Overall, *dbh* was measurable from 64.4% of the trees (Table 4). Shrubs, dense understory and branches disturbed the visibility to stem. Measurement accuracy for *dbh* was 7.0% and 12.5% in urban park and forest, respectively. In both environments *dbh* was underestimated.

Table 4. Measurement accuracy of *dbh*.

Area	n	<i>dbh</i> detection, %	RMSE	RMSE%	BIAS	BIAS%
Park	34	79.1	18.5	7.0	5.0	2.3
Forest	60	58.3	23.8	12.5	14.7	3.9
Total	94	64.4	26.0	10.1	8.6	2.8

4. Discussion

In this study mobile mapping and monitoring method was tested in urban forest environment. In urban park environment, the tested method produced tree detection accuracy over 97% (Table 3). In urban forest conditions, shrubs, dense understory and branches disturbed the visibility to stem and the tree detection accuracy retained low (68.2%). Location accuracy of the detected trees varied from 0.47 metres to 0.72 meters.

Our *dbh* measurement accuracy results are somehow comparable to previous studies of camera and laser devices. The accuracy in measuring *dbh* have varied from 8 mm to 16 mm with the laser-relascope (Kalliovirta *et al.*, 2005) and from 8.8 mm to 14.3 mm with laser-dendrometers (Skovgaard *et al.*, 1998; Parker and Matney, 1999). With camera-based systems the accuracies obtained have varied from 7.0 mm to 9.9 mm (Ashley and Roger 1969; Bradshaw 1972; Varjo *et al.*, 2006) The accuracies achieved in Vastaranta *et al.* (2009) were 8.3 mm with the TLS, 8.5 mm with the laser-camera and 14.3 mm with the laser-relascope. Relative accuracy of the *dbh* has been around 5-10%. In this study, relative RMSE in the park was 7% and in urban forest 12.5%. MLS is far more practical for rapid updating tasks than TLS or laser-camera.

MLS provide means for monitoring trees growing near roads or paths. Tree mapping, change detection and tree health monitoring would be the main applications for MLS in urban parks and forests. TLS is suitable for areas where the positional accuracy of the needed tree maps is high priority and additional information from the trees is needed. Large TLS campaigns are justifiable when there is intensive data need for several applications. It is hard to justify TLS measurements only for tree mapping, although it is still more cost-efficient than traditional tachymeter measurements. Compared to TLS measurements, MLS provides faster data collection and almost as accurate positional accuracy (Holopainen *et al.* 2011).

This study showed feasibility of the MLS measurements in urban tree inventory. MLS can be used in urban tree inventory in targets with a good visibility and relatively dense network of roads or paths. However, our results are preliminary, since the number of trees was rather low and measurements from point clouds were manual. We are going to expand our study area in city of Helsinki and focus on automatize of point cloud processing.

Acknowledgements

This study was made possible by financial aid from the Finnish Academy for the projects “Improving the Forest Supply Chain by Means of Advanced Laser Measurements” and “Science and Technology Towards Precision Forestry”. We also thank the city of Helsinki, park and garden department for co-operation.

References

- Ashley, M. and Roger, R. E., 1969. Tree heights and upper stem diameters. *Photogram. Eng.* 35(6), 136-146.
- Bilker, M. and Kaartinen, H., 2001. The Quality of Real-Time Kinematic (RTK) GPS Positioning. Reports of the Finnish Geodetic Institute.
- Bradshaw, F. J., 1972. Upper stem diameter measurements with the aid of 35 mm photographs. *Aust. For. Res.*, 6(1), 17-20.
- Holopainen, M., Vastaranta, M., Kankare, V., Hyypä, J., Liang, X., Litkey, P., Yu, X., Kaartinen, H., Kukko, A., Kaasalainen, S., Hyypä, H., Vaaja, M. & Jaakkola, A., 2011. The use of ALS, TLS and VLS measurements in mapping and monitoring urban trees. In: Stilla, U., Gamba, P., Juergens, C. & Maktav, D. (Eds) JURSE 2011 - Joint Urban Remote Sensing Event - Munich, Germany, April 11-13, 2011.
- Hyypä, J. and Inkinen, M., 1999. Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland*, 16:27-42.
- Kalliovirta J., Laasasenaho J., and Kangas A., 2005. Evaluation of the Laser-relascope. *Forest Ecology and Management*, 204(2-3), 181-194.
- Mechelke, K., Kersten, T. P., and Lindstaed, M., 2007. Comparative investigations into the accuracy behaviour of the new generations of terrestrial laser scanning systems. *Optical 3-D measurement techniques VIII*, Gruen/Kahmen, Zurich, pp. 319-327.
- Parker, P. C. and Matney, T. G., 1999. Comparison of optical dendrometers for prediction of standing tree volume. *Southern J. Appl. Forestry*, 23(2), 100-107.
- Skovgaard, J. P., Johannse, V. K., and Vanclay, J. K., 1998. Accuracy and precision of two dendrometers. *Forestry* 71(2), 131-139.
- Varjo, J., Henttonen, H., Lappi, J., Heikkonen, J., and Juujärvi, J., 2006. Digital horizontal tree measurements for forest inventory. Working Papers of the Finnish Forest Research Institute 40.
- Vastaranta, M., Melkas, T., Holopainen, M., Kaartinen, H., Hyypä, J. and Hyypä, H., 2009. Laser-based field measurements in tree-level forest data acquisition. *The Photogrammetric journal of Finland*, 1/2009:51-61.

Another dimension from LiDAR – Obtaining foliage density from full waveform data

Thomas Adams¹, Peter Beets², Christopher Parrish³

¹Scion, 49 Sala Street, Rotorua, New Zealand. thomas.adams@scionresearch.com

²Scion, 49 Sala Street, Rotorua, New Zealand. peter.beets@scionresearch.com

³NOAA/NGS, JHC-CCOM, 24 Colovos Road, Durham, NH, USA. chris.parrish@noaa.gov

Abstract

LiDAR tells the user *where* surfaces are, not *what* they are. In this study we investigate the potential for waveform LiDAR to provide more information on the nature of the returns over forestry. Waveform LiDAR was acquired for ten *Pinus radiata* plots in a New Zealand plantation, along with comprehensive leaf area sampling in 2m vertical bands. The decay rate of each waveform peak was shown to be a useful tool for estimating foliage density, and has potential for identifying regions containing ground and understorey.

Leaf Area Density (LAD) is an expression of foliage density per unit height, and a relationship between waveform decay rate and LAD was developed with an R^2 of 56%. Incorporating the proportion of discrete LiDAR that fell in that band (which itself has an R^2 of 50%) improves this model to explain 69% of the variation in LAD. This is a good result, especially given the costs and difficulties in measuring leaf area directly. As foliage density varies dramatically on a fine scale it was not possible to differentiate the nature of every single LiDAR return – but by averaging over a small area local variation in LAD could be easily mapped.

Ground returns could be distinguished as having short decays, and broad leafed understorey typically had values between those of the canopy and ground, although surface roughness and slope make it impossible to robustly identify single returns. This study produced a useful model for estimating LAD in *Pinus radiata* which could easily be extended to other coniferous species.

Keywords

Waveform LiDAR, foliage density, Leaf Area Density, waveform shape, deconvolution

1. Introduction

1.1 Foliage measurement

Foliage is the engine that drives tree growth. Foliage may be used as a good measure of tree growth and carbon sequestration, but also of tree form and fire risk. Knowledge of the 3D distribution of foliage can be used for many applications, such as carbon accounting (Stephens 2007), growth and yield forecasting (Naesset 1997), fire modelling (Morsdorf *et al.* 2004) and biodiversity estimation (Hofton *et al.* 2006).

Direct measurement of foliage location for individual trees is time consuming, costly and not overly representative at a stand scale (Bongers 2001). More spatially integrated measures of foliage distribution have emerged with simpler formulations which can be easily included in further analyses. The most common of these are canopy cover, Leaf Area Index (LAI) and Leaf Area Density (LAD). Canopy cover is the projected area of canopy – as viewed from above - per unit area of ground. The maximum canopy cover value of 1 indicates that the canopy completely obscures the ground, or is ‘closed’. Leaf Area Index goes further, and is the one-sided area of leaf tissue per unit of ground surface area (Watson 1947). Leaf Area Density extends this once more, and is defined as the total leaf area per unit ground area per unit height. Essentially, LAD is the vertical distribution of leaf area that in summation will equal the LAI.

LAI and LAD can only be measured exactly through destructive sampling of single trees. LAI can be estimated through litter collection and known tree-allometry, ground-based hemispherical photography and LiDAR, although none of these methods offer high levels of accuracy or have been used to resolve LAI into LAD (Forster and Nairn 2010). Of these methods LiDAR appears to offer the greatest potential because it provides a three dimensional characterisation of canopy (Lefsky *et al.* 1999).

1.2 LiDAR

LiDAR – or Light Detection and Ranging - has long been used for mapping large, continuous surfaces such as ground and buildings. Often aerial imagery is acquired at the same time to assist in surface identification. Recently LiDAR has found a home in forestry, where returns from the canopy add valuable information on the vertical structure of the forest as well as the ground beneath. For a thorough explanation of the function and application of LiDAR in forestry the reader is directed to Hyypä *et al.* (2004), Lim *et al.* (2003) or Adams *et al.* (2011).

LiDAR systems can be separated into two classes: discrete-return and full-waveform. Discrete-return systems use hardware-based subsystems (e.g., a constant fraction discriminator and time interval meter) to extract and record ranges and intensities in real time for a few individual returns per transmitted pulse (typically less than 5). Full-waveform systems on the other hand, digitize the backscattered laser echo sampled over a set period (typically around 1ns), providing a complete record of received signal amplitude over time. These digitized waveforms are stored for subsequent processing and analysis, thereby enabling additional information to be extracted. This information can include not only additional ranges, but also metrics related to surface roughness, slope and other characteristics. Figure 1 gives an example of a digitized waveform. Note how the discrete returns occur slightly before the waveform peaks; this is due to the fact that the discrete-return ranging is based on identifying a point on the leading edge at which the amplitude is a constant fraction of the peak (e.g. the 50% point). In both discrete-return and full waveform systems, georeferenced point clouds can be produced using the extracted ranges, sensor position and orientation data, scan angle data, sensor model, and associated calibration parameters.

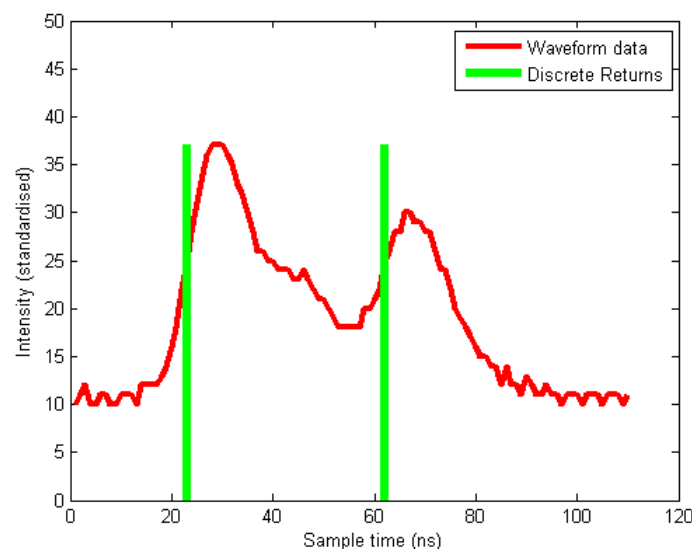


Figure 1 – An example of waveform and discrete-return LiDAR data

Canopy is not a flat impermeable surface like the ground, and it is this fact that allows us to collect information throughout its depth. Unlike in urban environments, where each return results from a single large surface, the scale of foliage is much less than that of the beam width (~0.2m) and vertical sampling interval (0.15m). Thus each return is due to the integrated reflections from a large cylinder of foliage. If the waveform digitizer has a nanosecond sampling period, then the

waveform at that point is due to the reflections from a biscuit tin sized volume of foliage of around 0.005m^3 . It should be noted that the return signal from each ‘biscuit tin’ depends not only on the foliage in that volume, but also on the attenuation of the beam by foliage above it. Multiple scattering events have been shown to be of negligible importance (Blair and Hofton 1999).

Some authors have attempted to investigate foliage density by considering the proportion of discrete returns per unit height. Morsdorf (2006) found a correlation between LAI measured by hemispherical photography and the proportion of discrete returns that hit the ground with an R^2 of 69%. This correlation was also found by Riaño (2004) on two sites in Spain. However, these studies give no information on the vertical distribution of the foliage, or LAD. Vertical canopy information from discrete LiDAR always suffers from the fact that the transmitted laser pulse width can be greater than 10ns, and there is typically a ‘blind-spot’ between returns of 1.2 - 3m during which no other returns can be counted (Reitberger, Krzystek, and Stilla 2008).

1.3 Full waveform LiDAR analysis

Full-waveform LiDAR can greatly reduce this blind spot by post-processing the waveform data to identify proximal peaks which would otherwise be treated as one (Parrish *et al.* 2011). Chauve *et al.* (2009) found 40-60% additional points in an Alpine coniferous forest. The most common approach is to approximate the waveform as a series of Gaussians, fitted by a non-linear least-squares approach (Hofton, Minster, and Blair 2000; Reitberger, Krzystek, and Stilla 2006), or expectation-maximisation (Persson *et al.* 2005). Wagner *et al.* (2006) found that fitting Gaussian peaks to the data could account for 98% of waveform shapes, although this was over an urban environment and he notes that Gaussians may not always be appropriate over vegetation.

Additional canopy points are not the only benefit of working with full-waveform data. The signal shape itself may yield information on the nature of the target. Whilst discrete LiDAR returns are given an intensity, the value is a function of transmitted power, range, atmospheric transmittance, system transmittance, receiver aperture, beam divergence, surface roughness, relative geometry between the sensor and target, and any attenuation due to foliage higher in the canopy. Studies have corrected for range and scan angle to derive a back-scattering cross section (Wagner *et al.* 2006), and for species identification between *Lodgepole Pine* and *Sitka Spruce* in Scotland (Donoghue *et al.* 2007). Reitberger used peak width and integral as clustering parameters to segment individual trees and group them according to species. The study achieved 80% accuracy in identifying deciduous and coniferous species (Reitberger, Krzystek, and Stilla 2006). Lin and Mills (2010) found pulse width less noisy than intensity for identifying surface roughness and slope. Like Wagner (2006), Chauve *et al.* (2009) note that the waveform is often skewed, and that a Gaussian may not be appropriate in all cases. Experimentation with a lognormal distribution provided a better fit for some pulses over a forest, although a standard Gaussian was found to provide the best fit in 99.3% of cases. Generalised Gaussians were also used, which theoretically should provide a better fit than standard Gaussians, although this didn’t occur due to the extra degrees of freedom and limitations of the optimisation algorithm.

Whilst other authors have noted this skewness of the waveform, in this study we use it to categorise returns. If we assume that the needles and branches within a sampled volume are sufficiently small and randomly orientated, then - as a simple model - the canopy can be imagined as a volume of semi-transparent gas. Whilst this is not a physical reality, it allows us to employ some modelling simplifications that allow a quantitative insight into the canopy’s average properties. We can then assume that the canopy has – as a gas would – a reflectivity R , transmissivity T , and absorbance A . Thus transmission through the canopy can be modelled with the Beer-Lambert law

$$T = \frac{I_T}{I_0} = e^{-\alpha x} \quad (1)$$

Where I_T is the transmitted intensity, I_0 is the initial intensity, α is the absorption coefficient and x is the distance travelled through the gas. Note that whilst transmissivity and reflectivity are both

functions of wavelength, lasers are fundamentally limited to a single wavelength so we can ignore this and define our parameters at a wavelength of 1064nm. For an interesting investigation into multi-spectral LiDAR the reader is directed to Morsdorf *et al.* (2009).

At a depth x into the canopy, a constant proportion (R) of the incident light will be reflected. We assume Lambertian reflectance, and that this is further attenuated by the foliage-gas as it leaves. I_R at the surface is then given by

$$I_{R(\text{surface})} = I_0 R e^{-2ax} \quad (2)$$

Thus we see that if the canopy is of a constant density and distribution, we can expect an exponential decay in intensity characterised by a decay constant of $2a$. For simplicity this decay constant will be called λ .

I_R as measured at the receiver is also a function of range as defined by the radar equation

$$I_{R(\text{receiver})} = \frac{I_0 \sigma F^4}{(4\pi)^2 (\rho+x)^4} \quad (3)$$

Where σ is the target scattering coefficient, F is the propagation factor which accounts for atmospheric loss, and ρ is the distance to the start of the foliage. So we may also see a fall off in I_R in the form of x^{-4} . However our study site was flown at a height of $\rho \sim 1000\text{m}$, whereas $x \sim 2\text{m}$, so $x \ll \rho$ and it is reasonable to ignore this effect.

Jutzi and Stilla (2006) note that the return waveform $y(t)$ is a convolution of the surface response $v(x,y,z)$, the transmitted waveform $o(t)$, the response of the measurement unit $m(t)$ and the spatial beam distribution $p(x,y)$. If the pulse is timed and we know the plane position and scan angle in three dimensions (i.e. we can describe x,y and z as a function of t) then we can describe $v(x,y,z)$ as $v(t)$. Furthermore, based on our simple isotropic gas model of the canopy locale, we can ignore $p(x,y)$ but add in a background noise $n(t)$ leaving

$$y(t) = v(t) * o(t) * m(t) + n(t) \quad (4)$$

Comparison of the outgoing waveform $o(t)$ (sampled as standard with some waveform digitisers) with the return waveform $y(t)$ for large flat surfaces (a pond on a windless day) shows that $m(t)$ is negligible compared to $o(t)$. Thus we have

$$y(t) \approx v(t) * o(t) + n(t) \quad (5)$$

The surface response $v(t)$ is what we are interested in determining, and equation (5) may be solved by a Wiener deconvolution where we estimate $g(t)$ to minimise the error in our estimation of $v(t)$ so that

$$\hat{v}(t) = g(t) * y(t) \quad (6)$$

The Weiner deconvolution filter can be used to find $g(t)$, shown here in the frequency domain

$$G(f) = \frac{O^*(f)S(f)}{|O(f)|^2S(f) + N(f)} \quad (7)$$

Where $S(f)$ is the mean power spectral density of the outgoing signal $o(t)$. $N(f)$ is the power spectral density of the noise $n(t)$. This may be solved to obtain $\hat{v}(t)$, which will from hereon in be referred to as the deconvolved waveform.

The skewness of waveform profiles over foliage has been noted by several authors (Wagner *et al.* 2006; Chauve *et al.* 2009; Jutzi and Stilla 2006). In this study we define this skewness as an

exponential decay, which will be more apparent in the deconvolved waveform than the original as the blurring function of the outgoing pulse has been removed. In this study we test the hypothesis that the decay constant of the deconvolved waveform will correlate with the local foliage density, expressed as an LAD.

2. Method

2.1 Field data collection

The field data collection for this project was shared with that of Beets *et al.* (2011). Ten 0.16ha field plots of nine year-old *Pinus radiata* were selected for study. Each plot was planted with the same stocking in 1997, and had received the same silviculture. Five representative sample trees per plot were felled from 21st August - 8th September, 2006. Tree crowns were weighed fresh in the field by 2m height zone, measured from the base of each tree. Fifty needle fascicles were sampled randomly from each height zone and stored in polythene bags with water. Sample branches from each 2m height zone were weighed fresh in the field, and partially dried to aid with separation of needles from branches. Needles were oven dried to constant weight at 65 degrees C and weighed. Leaf area of the 50 fascicles per 2m height zone was determined on an all-surfaces basis from fascicle length and volume, using methods described in Beets and Lane (1987), and then oven-dried and weighed. LAD by 2m height zone was calculated for each zone by multiplying the fascicle dry weight by the specific leaf area of the 50 fascicles. The leaf area index per plot was obtained using the basal area ratio method, where the sum of zonal leaf areas of sample trees per plot is multiplied by the plot basal area divided by the sample tree basal area (Madgwick and Service 1981).

2.2 LiDAR data collection

The study site was flown by New Zealand Aerial Mapping on 9th June 2007, and measured with an Optech ALTM 3100EA LiDAR system and waveform digitiser with a sampling rate of 1ns. Raw GPS data and discrete LiDAR information was processed with REALM software into a Corrected Sensor Data (CSD) file. The waveform data was measured at 1ns intervals and provided as five swathes in Optech's NDF binary format with an IDX index file. Matlab code was adapted from Parrish (2007) to suit the updated NDF format and to georeference the waveforms according to New Zealand's NZGD2000 coordinate system.

2.3 Waveform analysis

Much work has been done on identifying and resolving peaks in waveform data (Chauve *et al.* 2009). In this work we are investigating the decay of the waveform, so separating closely-spaced peaks is of no use. For efficiency we use the discrete LiDAR dataset to identify peaks, and the waveform LiDAR to determine the decay rate. This process is detailed below.

Step 1

Read and georeference LiDAR data. Using adapted code from Parrish (2007) each sample in the waveform is assigned an xyz coordinate and compared with the field plot locations. Waveforms – or sections of waveforms – that fall within the field plot are written to a new file with their respective locations. In comparison with Parrish, the waveform is not deconvolved at this point, and all points within the sample areas are written to a text file. In addition the CSD file is read and the discrete LiDAR point cloud is extracted for the whole of Puruki forest.

Step 2

Subtract ground surface from LiDAR data. By subtracting the ground surface from the data we determine heights above ground instead of heights above sea-level, which is standard practice for determining information on canopy as opposed to terrain. In this paper we refer to this process as

‘degrounding’. The discrete LiDAR data for the whole forest is used with FUSION’s *GroundFilter.exe* function (McGaughey 2010) to determine a ground surface on a 1m grid. This gives a better result than using the clipped data for individual plots. This grid is then used to linearly interpolate a ground height for every waveform sample saved in the text file produced in step 1, and every discrete return that falls within the plot areas. For every plot a georeferenced degrounded waveform file is produced, along with a corresponding degrounded discrete point cloud.

Step 3

Deconvolve waveform. Each degrounded waveform file is read, and each waveform line individually analysed. A Wiener deconvolution (see introduction) is used, using the outgoing waveform as the deconvolution kernel. Note that the deconvolution increases the amount of noise, particularly away from peaks.

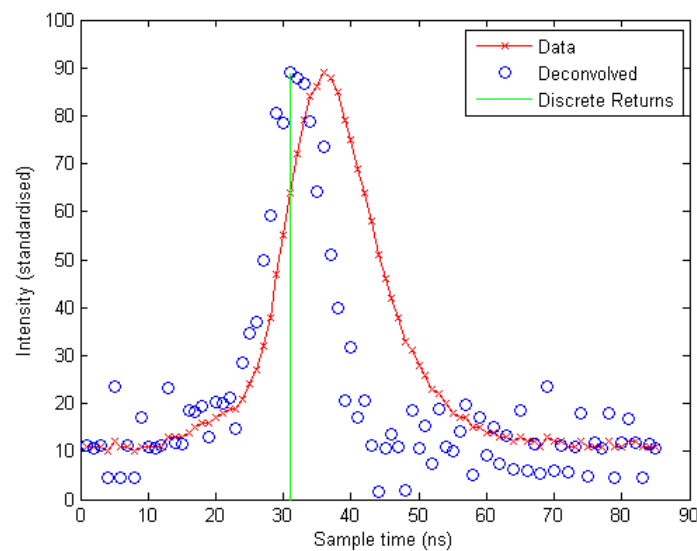


Figure 2. Original waveform and deconvolved waveform.

Step 4

Separate peaks and fit exponential decay. Peaks are identified as the nearest local maxima to each discrete return. As the discrete return is determined on the leading edge of a pulse, the local maxima should always be just after this point. In the instances where a waveform contains multiple peaks, the data between peaks is sampled from local maxima to local minima to find the exponential decay. The decay is found using a weighted least squares approach, where samples are weighted by their height (to encourage curve fitting to match the peaks as opposed to the background noise). Figure 3 shows a waveform with two returns and two fitted exponential decays. Note that the x axis is measured in time, not in height. If the decay were measured in height, waveforms travelling non-perpendicularly to the ground would appear compressed and hence have artificially increased decay constants (λ). Once the decay constants are found they are added to a point cloud output that includes xyz coordinates, intensity and λ .

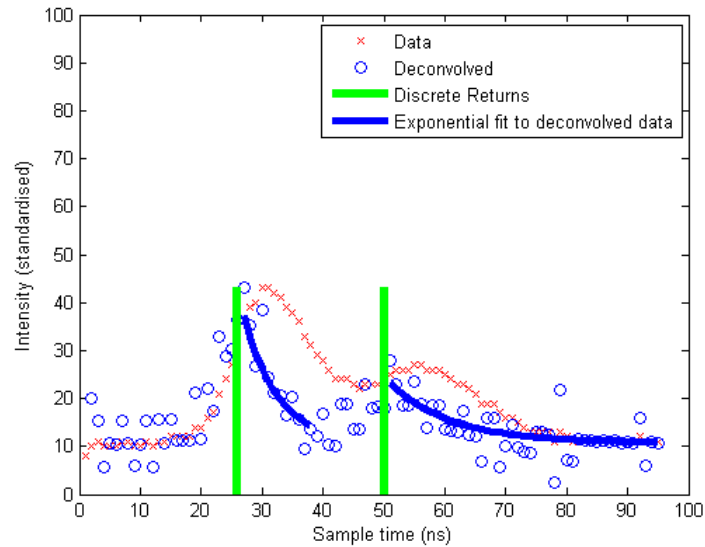


Figure 3. Example of multiple peak waveform with two exponential decays fitted.

3. Results

3.1 LAD vs. proportion of returns.

Extending the work of Morsdorf *et al.* (2006), we initially use the degrounded discrete LiDAR data to compare LAD with the proportion of LiDAR returns that occurred within each height band. Figure 4 shows the results provide a reasonable model (R^2 of 50%). This R^2 is not as good as the 69% found by Morsdorf *et al.*, although their model considered net LAI as opposed the LAD at different height bands. Our model is also flawed in that it does not go through the origin – a LAD of 0 should lead to no returns.

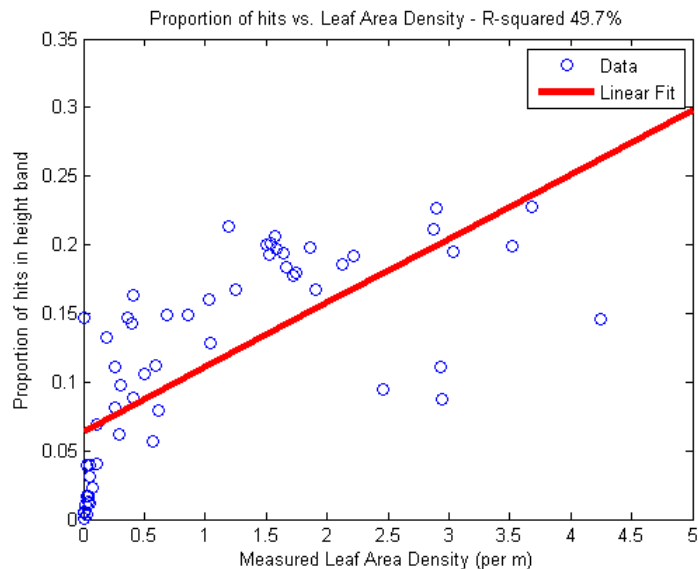


Figure 4 – Leaf Area Density vs. proportion of hits per height band

3.2 LAD vs. average decay rate

Within each height band there are many LiDAR returns. If our approximation of foliage being an isotropic gas were true, then we would expect all of the returns to exhibit the same decay. However this is only a simple model, and some pulses will hit dense areas of vegetation whereas others will find a gap and only receive a glancing blow. So each individual decay rate is not of

much use, but the average over the whole height band should be informative. Experiments with the mean and median showed that the mean is slightly more related to LAD, with an R^2 of 55.8% as opposed to 54.1% for the median. The correlation for the mean is illustrated in figure 5.

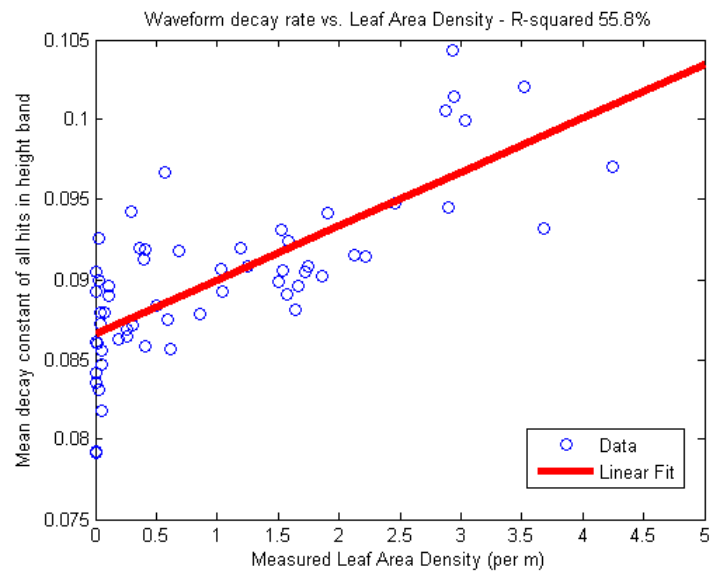


Figure 5 – Deconvolved waveform decay rate vs Leaf Area Density

This is a good result and shows that waveform decay rates are representative of LAD. It also yields the question whether this model could be improved by using the proportion of LiDAR returns as well as the mean decay constant. Assuming a simple linear model

$$L = \beta \bar{\lambda} + \gamma p + c \quad (8)$$

Where L is LAD, $\bar{\lambda}$ is mean decay constant, p is proportion of hits, c is a constant and β and γ are constants of proportionality. Solving this for a least-squares solution, we obtain the model shown in Figure 6, with a R^2 of 69%. This is a great improvement in accuracy, although it is apparent that the relationship is not perfectly linear, leading to underestimation at lower LAI values.

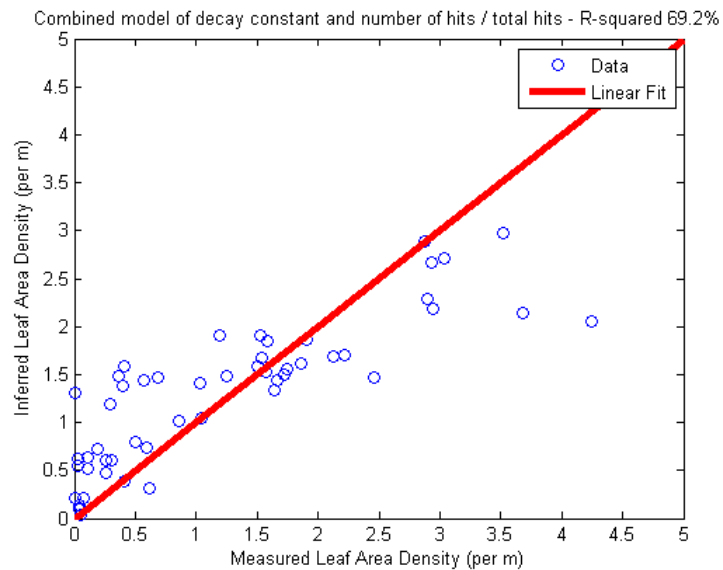


Figure 6 – Model for Leaf Area Density based on average decay constant and proportion of hits.

4. Discussion

The results show that modelling foliage as a semi-transparent gas and applying the Beer-Lambert law for transmittance enables reasonable estimates of the parameters of interest. A good correlation between Leaf Area Density and average decay rate $\bar{\lambda}$ was found, with an R^2 of 56%. Unfortunately, this simple model does not work for individual pulses. If it did, the decay rate could easily be used to differentiate returns from foliage to those from branches, the stem, understorey and the ground. Unfortunately - just like intensity - decay constant is still a function of slope, roughness and size. For each surface type there are so many variations and ways in which the light can hit it that only the mean of a large number of hits is useful. For example, light hitting a particularly dense patch of foliage would yield a shorter decay than light hitting hard ground at an oblique angle. The crucial difference between decay constant and intensity is that decay constant is not affected if the light beam has been attenuated prior to hitting the target. Figure 7 shows an attempt to correlate LAD with mean intensity within a height band, and the R^2 value of 2% shows that there is little or no correlation.

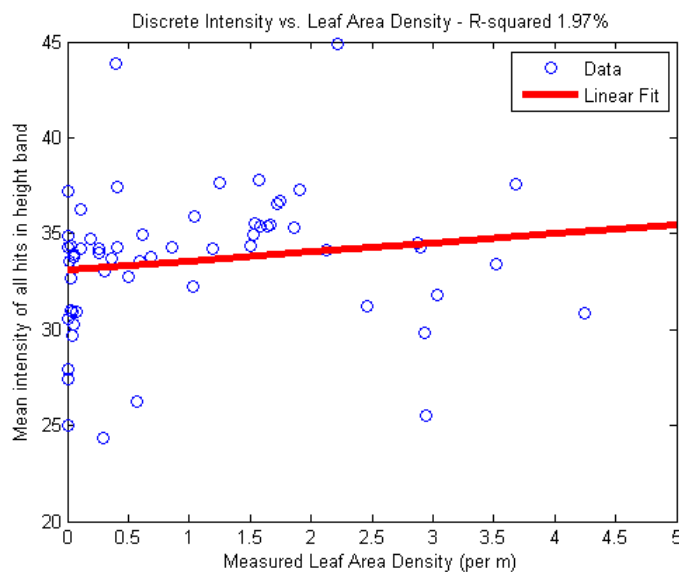


Figure 7 – Discrete return leaf intensity vs. leaf area density.

When considering the average decay constant within a height band, there is a distinct change between foliage and ground. Figure 8 shows how LAD inferred from $\bar{\lambda}$ varies with height for plot 10, in 1m height bands, along with the measured LAD. Two trends are clear: The LiDAR method underestimates LAD at the very top of the tree, and grossly overestimates it at the base. The underestimation at the top of the tree is likely due to the fact that LiDAR does not sample everywhere, and the beam footprints are prone to missing the very tops of trees (Hyypä *et al.* 2004). The overestimation at the base (where the measured LAD was 0) is due to the fact that understorey and ground returns also feature. Due to the broad-leafed nature of the understorey (ferns, grasses, and broad-leafed shrubs), these are likely to attenuate incident light faster than dispersed needles, and have a higher decay rate. The ground will attenuate light almost instantly, so can be expected to have an even higher decay rate. Understorey was recorded for plot 10 as approximately 70% cover up to a height of 4m. This result would be consistent with the increase in $\bar{\lambda}$ from this height onwards.

Unfortunately, as we have already mentioned, for each height band there exists a wide range of λ due to its dependence on surface roughness and slope. We cannot expect to separate out individual returns from understorey as distinct from canopy foliage, but an increase in $\bar{\lambda}$ can be used as an indicator of where understorey begins. Higher values of $\bar{\lambda}$ could also be used to identify ground returns, but it is unlikely this method would be as effective as the well-researched surface fitting methods employed for digital terrain model extraction (see Hyypä *et al.* (2004) for a history of

methods). Further research is necessary for applying this method in mixed species forests – it may for example be possible to gauge the species mix by the range in λ , particularly if it is a mix of coniferous and deciduous species.

This study has determined that – like intensity – waveform decay is a useful addition to a discrete LiDAR point set. Whilst each individual value does not determine the surface type for each individual return, when taken as a population over an area they can be useful – along with geometry and aerial photography – for identifying surface traits. A model for foliage density from LiDAR will be of use for carbon accounting, growth and yield, and fire risk modelling.

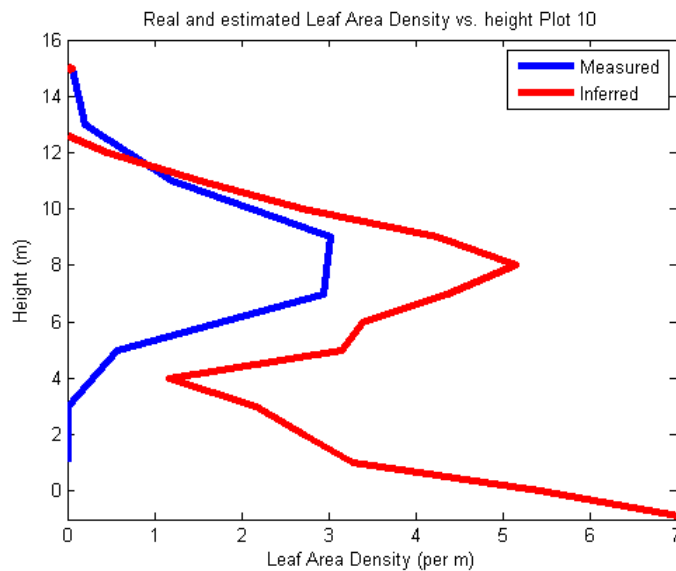


Figure 8 – Leaf Area Density (measured and inferred from mean decay model) vs. height in plot 10.

5. Conclusion

Foliage density is a useful characteristic for describing forests. It can be used for measuring growth and yield, carbon sequestration, health, biodiversity and fire risk. Directly measuring foliage density by destructively sampling trees is expensive and unrepresentative over large areas. Photographic methods – such as aerial and hemispherical photography – provide reasonable estimates of Leaf Area Index (LAI), but are generally unable to resolve these data into a vertical distribution.

Aerial LiDAR gives a full 3D representation of the canopy and can provide this information for large areas in a short time. Waveform and discrete LiDAR over a New Zealand radiata pine forest was obtained and compared with Leaf Area Density (LAD) measurements made in 2m height intervals across ten 0.16ha sample plots. Returns in foliage were generally characterised by a skewed shape, and the hypothesis that this skew was related to foliage density was tested.

By deconvolving the waveform with the outgoing pulse shape, an expression of the surface response can be found (Jutzi and Stilla 2006). A simple model of the canopy as a semi-transparent gas was assumed, which dictates that reflectance would follow an exponential decay as per the Beer-Lambert law. By fitting exponential decays to each deconvolved waveform peak, values for the decay constant λ were obtained. In this study we tested the hypothesis that λ will characterise the density of the foliage at that location.

λ is also a function of surface slope and roughness, as is the case for intensity. λ is preferential to intensity for surface identification since it is not affected by beam attenuation higher in the canopy. Although λ cannot be used to identify each individual return, the mean value for λ shows a strong relationship to surface type when averaged over a volume. Highest values were obtained for areas of predominately ground, medium values for understory and lower values for canopy. Within the

canopy, λ can be used to estimate LAD with an R^2 of 56%. The proportion of LiDAR returns to fall in each height band can describe LAD with an R^2 of 50%, and a model utilising both decay rate and proportion of returns described 69% of the variation in LAD. This model can be used for remotely sensing foliage density in *Pinus radiata* forests above the understorey, but could be calibrated for other species and potentially even mixed species forests.

Acknowledgements

The work in this paper was funded by the Scion Capability Fund and used data collected by the Ministry for the Environment. Many thanks to Nigel Searles from MfE for use of the data, and to Andrew Dunningham, Jonathan Harrington and David Pont from Scion for their help and knowledge. Particular thanks must also go to New Zealand Aerial Mapping, in particular Tim Farrier, for providing an excellent LiDAR service and for assistance with countless queries and demands.

References

- Adams, T., Brack, C., Farrier, T., Pont, D., and Brownlie, R. (2011). So you want to use LiDAR? - A guide on how to use LiDAR in forestry. *New Zealand Journal of Forestry*, 55 (4): 19–23.
- Beets, P., and Lane, P. (1987). Specific leaf area of *Pinus radiata* as influenced by stand age, leaf age, and thinning. *New Zealand Journal of Forestry Science*, 17 (2): 283-291.
- Beets, P.N., Reutebuch, S., Kimberley, M.O., Oliver, G.R., Pearce, S.H., and McGaughey, R.J. (2011). Leaf area index, biomass carbon and growth rate of radiata pine genetic types and relationships with LiDAR. *Submitted to Forests*.
- Blair, J.B., and Hofton, M.A. (1999). Modeling laser altimeter return waveforms over complex vegetation using high resolution elevation data. *Geophysical Research Letters*, 26 (16): 2509-2512.
- Bongers, F. (2001). Methods to assess tropical rain forest canopy structure: an overview. *Plant Ecology*, 153 (1): 263-277.
- Chauve, A., Vega, C., Durrieu, S., Bretar, F., Allouis, T., Deseilligny, P., and Puech, W. (2009). Processing full-waveform lidar data in an alpine coniferous forest: assessing terrain and tree height quality. *International journal of remote sensing*, 30 (19): 27.
- Donoghue, D.N.M., Watt, P.J., Cox, N.J., and Wilson, J. (2007). Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data. *Remote Sensing of Environment*, 110 (4): 509-522.
- Forster, W.A., and Nairn, J. (2010). *Literature Review of Methods for Describing Plant Canopies. Contract report written for FRST Contract No. LVLX0901: Protecting NZ's environment from pesticide exposure.*
- Hofton, M., Blair, J., Rabine, D., Dubayah, R., and Greim, H. (Eds.). (2006). *Using Lidar-derived 3-D Vegetation Structure Maps to Assist in the Search for the Ivory-billed Woodpecker.*
- Hofton, M.A., Minster, J.B., and Blair, J.B. (2000). Decomposition of laser altimeter waveforms. *Geoscience and Remote Sensing, IEEE Transactions on*, 38 (4): 1989-1996.
- Hyypä, J., Hyypä, H., Litkey, P., Yu, X., Haggrén, H., Rönnholm, P., Pyysalo, U., Pitkänen, J., and Maltamo, M. (2004). Algorithms and methods of airborne laser-scanning for forest measurements. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (Part 8): 1682-1750.
- Jutzi, B., and Stilla, U. (2006). Characteristics of the measurement unit of a full-waveform laser system. *Revue Française de Photogrammétrie et de Télédétection*, 182 (2006-2): 17–22.
- Lefsky, M.A., Cohen, W., Acker, S., Parker, G.G., Spies, T., and Harding, D. (1999). Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment*, 70 (3): 339-361.
- Lim, K., Treitz, P., Wulder, M., St-Onge, B., and Flood, M. (2003). LiDAR remote sensing of forest structure. *Progress in Physical Geography*, 27 (1): 88.

- Lin, Y.C., and Mills, J.P. (2010). Factors influencing pulse width of small footprint, full waveform airborne laser scanning data. *Photogrammetric engineering and remote sensing*, 76 (1): 49-59.
- Madgwick, H., and Service, N.Z.F. (1981). *Above-ground dry-matter content of a young close-spaced Pinus radiata stand*: New Zealand Forest Service.
- McGaughey, R. (2010) FUSION. US Forest Service, Pacific Northwest Research Station, <http://www.fs.fed.us/eng/rsac/fusion/>.
- Morsdorf, F., Kotz, B., Meier, E., Itten, K., and Allgower, B. (2006). Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment*, 104 (1): 50-61.
- Morsdorf, F., Meier, E., Kotz, B., Itten, K.I., Dobbertin, M., and Allgower, B. (2004). LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment*, 92 (3): 353-362.
- Morsdorf, F., Nichol, C., Malthus, T., and Woodhouse, I.H. (2009). Assessing forest structural and physiological information content of multi-spectral LiDAR waveforms by radiative transfer modelling. *Remote Sensing of Environment*, 113 (10): 2152-2163.
- Naesset, E. (1997). Estimating timber volume of forest stands using airborne laser scanner data* 1. *Remote Sensing of Environment*, 61 (2): 246-253.
- Parrish, C.E. (2007). Vertical object extraction from full-waveform lidar data using a 3d wavelet-based approach. PhD, Civil and Environmental Engineering, Wisconsin-Madison, Wisconsin.
- Parrish, C.E., Jeong, I., Nowak, R.D., and Smith, R. (2011). Empirical Comparison of Full-Waveform Lidar Algorithms: Range Extraction and Discrimination Performance. *Photogrammetric Engineering & Remote Sensing (in press)*.
- Persson, Å., Söderman, U., Töpel, J., and Ahlberg, S. (2005). Visualization and analysis of full-waveform airborne laser scanner data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (part 3): 103-108.
- Reitberger, J., Krzystek, P., and Stilla, U. (2006). Analysis of full waveform lidar data for tree species classification. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (Part 3): 228-233.
- Reitberger, J., Krzystek, P., and Stilla, U. (2008). Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *International journal of remote sensing*, 29 (5): 1407-1431.
- Riaño, D., Valladares, F., Condés, S., and Chuvieco, E. (2004). Estimation of leaf area index and covered ground from airborne laser scanner (Lidar) in two contrasting forests. *Agricultural and Forest Meteorology*, 124 (3-4): 269-275.
- Stephens, P.R., Watt, P. J., Loubser, D., Haywood, A., and Kimberley, M.O.,(Ed.) (Eds.). (2007). *Estimation of carbon stocks in New Zealand planted forests using airborne scanning LiDAR*. Finland.
- Wagner, W., Ullrich, A., Ducic, V., Melzer, T., and Studnicka, N. (2006). Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60 (2): 100-112.
- Watson, A. (1947). Saligia. *Journal of the Warburg and Courtauld Institutes*, 10: 148-150.

The Sorted Pulse Data Software Library (SPDLib): Open source tools for processing LiDAR data.

Pete Bunting^{1,2}, John Armston³ Daniel Clewley¹ & Richard Lucas¹

¹Institute of Geography and Earth Sciences, Aberystwyth University, UK. pete.bunting@aber.ac.uk

²Landcare Research, Palmerston North, NZ

³ Joint Remote Sensing Research Program, Centre for Spatial Environmental Research, School of Geography, Planning and Environmental Management, University of Queensland
j.armston@uq.edu.au

Abstract

SPDLib is a new set of tools that allow processing and analysis of the full range of LiDAR data from terrestrial, airborne and spaceborne systems, including both discrete return and waveform datasets. The software provides an implementation of the SPD file format that allows efficient and flexible storage of these datasets largely through the inclusion of spatial indexing and pulse (rather than point) based data structures. A visualisation tool (SPD Points Viewer), which builds on top of SPDLib and the SPD file format, has also been developed. The software and source code have recently been made freely available and can be accessed online through open source code repositories. Future developments will focus on the development of new waveform processing functionality and optimizing performance. The software and documentation can be obtained from <http://www.spdlib.org>.

Keywords: LiDAR, Software, Processing, Full Waveform, Discrete Return, Pulse

1. Introduction

As the availability and volume of Light Detection and Ranging (LiDAR) data have increased and its use expanded beyond traditional surveying, software needs to be made available to allow users to efficiently process and exploit these data across a wide range of applications. Most of the LiDAR data available for users has been acquired by airborne systems, with the post spacing often > 50 cm and coverage often exceeding hundreds or even thousands of square kilometers. However, terrestrial laser scanning (TLS) systems are also increasingly being used with these providing very high post spacing (1-10 cm) data over comparatively small areas (often < 1 ha). LiDAR systems also have varying specifications, from single and multiple return discrete return recording sensors to sensors that digitize full transmitted and received waveforms (Mallet *et al.*, 2009).

From these datasets, a wide range of data products have been generated. For example, within the vegetation community, common products derived from airborne LiDAR have included Digital Terrain Models (DTM), Digital Surface Models (DSM), Canopy Height Models (CHM) and structural measures such as gap fraction, canopy openness, leaf area index, height percentiles and apparent foliage profiles. From these measures, studies have estimated a wide range of parameters, relating to carbon stocks (Lee and Lucas, 2007), timber volume (Næsset, 2002), biodiversity (Hill and Thomson 2005) and forest health (Müller and Brandl, 2009). However, in each case, different software have been used and often developed in-house such that it is not available for wider use.

The development of software for LiDAR process is not trivial, particularly when dealing with large datasets that cannot fit into the memory of many computers used for processing. Whilst a number of commercial software products are available (e.g., Terrasolid's Terra product line,

PointTools and ESRI's ArcGIS), these are limited to processing discrete return data and do not support the ingestion or analysis of waveform data. Sensor manufacturers do provide tools for processing the data from their systems, such as Riegl's RiAnalyze, RiProcess and RiWorld or Leica's Cyclone software, but these are mainly concerned with the pre-processing steps directly following acquisition, such as registration to world coordinates or translating waveform data into discrete returns for use within existing software packages. A large body of software has also been generated through academic research, some of which has been released into the public domain under a range of licensing terms. Such software includes OPALS (from the Vienna University of Technology), which supports a wide range of data processing tasks including the processing of waveform data. BCAL LiDAR tools (Idaho State University) and RSC LAS Tools (University of Queensland) provide utilities that are built on top of IDL/ENVI, for processing and visualising point cloud LiDAR data. LASTools (University of North Carolina) has functionality for processing very large discrete return LiDAR datasets, focusing particularly on filtering and producing elevation models. More specialised tools include MCC-LiDAR (US Forest Service), which only provides an implementation of the ground return classification algorithm of Evans & Hudak (2007), and libLAS which provides a general purpose software library for reading and writing the ASPRS LAS file format data.

The specific nature of most software means that many of the often disparate datasets used by research and natural resource management agencies cannot be ingested, analyzed and processed in a simple, standardized and integrated manner. Presented with the difficulty of using discrete return and waveform LiDAR and TLS data acquired by different ground-based, airborne and/or spaceborne sensors over sites in Australia and Wales, UK, this study aimed to develop a new set of tools for the storage and processing of LiDAR data, specifically with support for terrestrial, airborne and spaceborne discrete return and full waveform data. Key objectives were to provide commands for complete end-to-end processing of single wavelength LiDAR data following receipt of the data from the provider to products, support the analysis and interpretation of large datasets and to facilitate batch processing. The software was also designed to provide a reference implementation of the sorted pulse data (SPD; Bunting *et al.*, 2011) file format, built on the HDF5 format. The SPD file format supports the storage of waveform and discrete return data while providing spatial indexing of the data for efficient data processing, all of which the SPDLib software takes full advantage. In parallel, the development of high quality visualisations of the LiDAR data for validation, implementation and testing purposes was envisaged.

2. Method

2.1 Sorted Pulse Data (SPD) Format

The software library and tools presented in this paper are built on top of the SPD file format (Bunting *et al.*, 2011) which has been designed specifically for the storage of LiDAR waveform and discrete return data acquired by terrestrial, airborne and spaceborne sensors. The format uses a pulse based structure as opposed to a solely point based structure, where pulses contain all the information associated with a transmitted pulse from the sensor, including transmitted and received waveforms and the discrete returns, determined by Gaussian decomposition of digitised waveforms or by the sensor hardware using proprietary methods (Bunting *et al.*, 2011). The SPD format also supports 2D spatial indexing of the pulses, where pulses can be referenced using cartesian, spherical or polar coordinate systems and projections. These indexes can be used to significantly speed up data processing whilst allowing the data to be appropriately projected and are particularly useful when analysing and interpreting TLS data. The format is defined within a HDF5 file, which provides a number of benefits including broad support across a wide range of platforms and architectures and support for file compression.

2.2 The Sorted Pulse Data Software Library (SPDLib)

SPDLib has been implemented within C++ and provides support for a wide range of platforms, although has predominately been tested on UNIX and Linux systems. The software is written with an object-orientated modular design throughout, allowing for new functionality to be added using the C++ application programming interface (API) with relative ease. In addition, a more user-friendly Python binding interface to SPDLib has been provided allowing easy access to the raw pulse data and spatial index for new functionality to be developed more easily. Finally, a wide range of tools has been developed for common tasks (e.g., ground return classification, data management and the calculation of metrics) associated with LiDAR data processing.

2.2.1 Tools and Workflow

The recommend workflow for airborne LiDAR data is shown in Figure 1. Where multiple input files for a study are available (i.e., associated with a number of flight lines), the first step is to create a single data file without a spatial index, referred to as an Unsorted Pulse Data (UPD) file. If waveform data are the input, it is recommended that Gaussian decomposition is applied, generating a file with both waveforms and discrete returns that are linked to one another.

Up until this point, the data are processed without a spatial index (as one is not required). However, the data need to be converted to the SPD format for the remainder of the processing chain using the commands provided. Once the SPD file has been created, using an application specific index resolution and origin, then the ground returns are classified. From the classified data DTM, DSM and CHM raster products can be generated and the above-ground height attributes within the SPD file (i.e., on pulses and points) can be defined. The SPD format contains, within a single file, both a topographic elevation and above-ground height attribute for each pulse origin and discrete return. After defining the above-ground height field, a range of metrics commonly applied to LiDAR data can be calculated. However, rather than implementing each metric individually a set metric primitives and mathematical operations between metrics have been defined. Therefore, all metrics can be calculated through any combination of these primitives and operations, an XML file is used to define these combinations. The XML interface allows the above-ground height, sensor-target range, Gaussian amplitude (or intensity) and width fields to be used and combined, facilitating calculation of a large number of existing canopy, terrain and statistical metrics and the creation of new metrics as required.

2.2.1.1 File format conversion

A number of common LiDAR formats are supported for conversion to the SPD format and new formats are being progressively integrated. These include a range of ASCII formats, although these tend to differ between data providers, and the LAS binary format, which provides a standard interpretable format for discrete return data. It should be noted that some data providers and software packages do not always populate all the available data and header fields within the LAS file and this can limit the usefulness of some datasets. For example, if the points are not associated with the correct return numbers, the pulses cannot be reconstructed when the SPD file is built. Users are therefore recommended to consider the fields they require, ideally prior to acquisition of the data, to ensure they request all the data they might need.

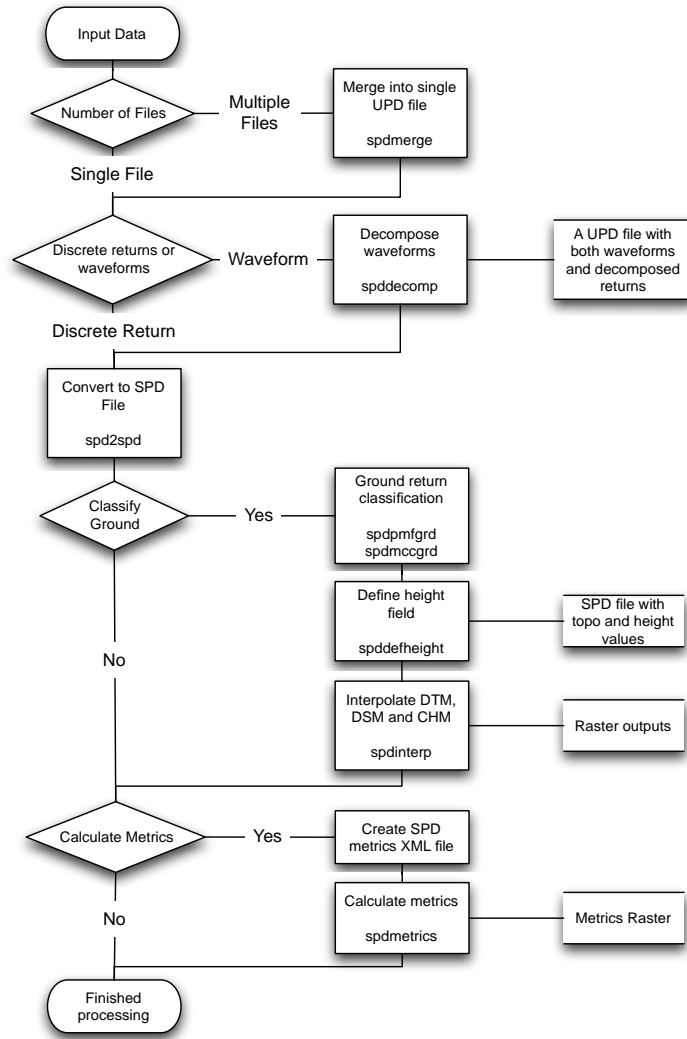


Figure 1. SPDLib airborne LiDAR processing chain.

2.2.1.2 Decomposition of full waveform data

The SPDLib software supports Gaussian decomposition (Wagner *et al.*, 2006) of received waveforms (P_i) to retrieve discrete returns, which are linked to the waveforms within the SPD file for each pulse by the time (t_i). The zero-crossings of the waveform first derivative above a nominal noise threshold are identified and used as the starting values for the N Gaussian amplitude (\hat{P}_i) and time (t_i) parameters (Equation 1). Bounds are also placed on the pulse width parameter (s_p). The Levenberg-Marquardt method for non-linear least-squares is used (Equation 1) to solve for these parameters.

$$P_r(t) = \sum_{i=1}^N \hat{P}_i e^{-\frac{(t-t_i)^2}{2s_i^2}} \quad (1)$$

Decomposing the waveform allows existing algorithms for analysing discrete return data to be used (e.g., for ground return classification) while retaining the waveform for later processing steps. This is achieved as the discrete returns and waveforms are linked through the pulse based architecture.

2.2.1.3 Ground return classification and interpolation

Three ground return classifiers are available within SPDLib. The first is a simple plane-fitting classifier that is best applied to scenes with relatively flat terrain and uses the lowest elevation returns within each grid cell (of the index) to identify ground returns. The classification process applies an iterative plane fitting approach for each grid cell, using a window of cells around the central cell for context. This process is iterative with a plane fitted at each step with returns above the plane progressively removed until either there are no points remaining in the central cell or all the remaining points fit onto the plane. The second algorithm is an implementation of the progressive morphological filtering approach by Zhang *et al.* (2003), which can be applied to landscapes with variable terrain and supports the removal of a wide variety of features such as vegetation and buildings. The third algorithm is the multiscale curvature classification algorithm of Evans and Hudak (2007), which is optimized for vegetated environments.

To define the above-ground height field for each return and pulse, two options are available. A raster DTM can be used with the terrain elevation for each point defined using a nearest neighbor algorithm. Alternatively, values can be directly interpolated from the classified ground returns. The second approach is recommended, especially for datasets with large ground post spacing. Several interpolation algorithms are available for identifying terrain height values from the classified ground returns. The recommended (Bater and Coops 2009) and default option is the natural neighbor algorithm (Sibson 1981). The same approach is used and recommended for generating DTM, DSM and CHM products.

2.2.1.4 Calculation of LiDAR metrics

A large body of literature (e.g., Hall *et al.*, 2005, Coops *et al.*, 2007) deals with the definition and formulation of metrics that can be calculated from 3D LiDAR data, particularly for vegetated environments. These can be simple statistical moments, percentiles of the canopy height or return amplitude or count ratios (often linked to canopy cover). More advanced metrics can include the parameters of probability density functions (e.g. Weibull; Coops *et al.*, 2007) or biophysical metrics such as foliage projective cover (Armston *et al.*, 2009).

To provide an interface to allow these metrics to be calculated with SPDLib, a list of metric primitives (Table 1) has been defined for above-ground height, topographic elevation, return amplitude, return width and range. The returns classification (e.g., ground, vegetation etc.), minimum and maximum values for the attribute being calculated (e.g., above-ground height) and return number (e.g., first, last etc.) can be used to restrict the points and pulses used for the calculations.

Table 1: List of the metric primitives available within SPDLib.

Metric	
Number of Returns	Absolute Deviation
Sum	Coefficient Of Variation
Mean	Percentiles
Median	Skewness
Mode	Person Mode
Minimum	Person Median
Maximum	Kurtosis
Standard Deviation	Count ratio
Variance	

A list of mathematical operators (Table 2) which can be applied to either another mathematical operator or metric primitive are then available to allow a range of LiDAR metrics to be derived. The interface therefore provides a highly flexible tool for subsequent analysis of LiDAR data.

Table 2: List of the metric operators available in SPDLib.

Operator	Inputs
Add	Two metrics
Minus	Two metrics
Multiply	Two metrics
Divide	Two metrics
Power	One metric
Absolute	One metric
Square Root	One metric
Sine	One metric
Cosine	One metric
Tangent	One metric
Inverse Sine	One metric
Inverse Cosine	One metric
Inverse Tangent	One metric
Log (base 10)	One metric
Natural Log	One metric
Exponential	One metric
Percentage	Two metrics
Add constant	One metric and one constant value
Minus constant from metric	One metric and one constant value
Minus metric from constant	One metric and one constant value
Multiply by constant	One metric and one constant value
Divide metric by a constant	One metric and one constant value
Divide constant by metric	One metric and one constant value
Metric to the power of constant	One metric and one constant value
Constant to the power of metric	One metric and one constant value
Number of returns above metric value	One metric
Number of returns below metric value	One metric

To interface with this metrics command, an XML file is defined with a hierarchical list of metrics and operators. These can either be applied to the LiDAR data to produce a raster output or to a polygon shapefile of polygons where the calculated values for each polygon are added to the shapefiles attribute table.

2.2.2 Python Binding

The SPDLib software also provides a Python binding to the C++ library to allow SPD (both indexed and unsorted non-indexed) files, including waveform data, to be read and written directly from python. This allows new and more advanced features and functions to be implemented and gives access to the functionality Python/SciPy offers (e.g., Matplotlib, plotting; Figure 2). In addition, the Python binding can be used to convert unsupported

external formats to SPD files more easily than through the use of C++, therefore allowing other SPDLib tools or the SPD Point Viewer to be used.

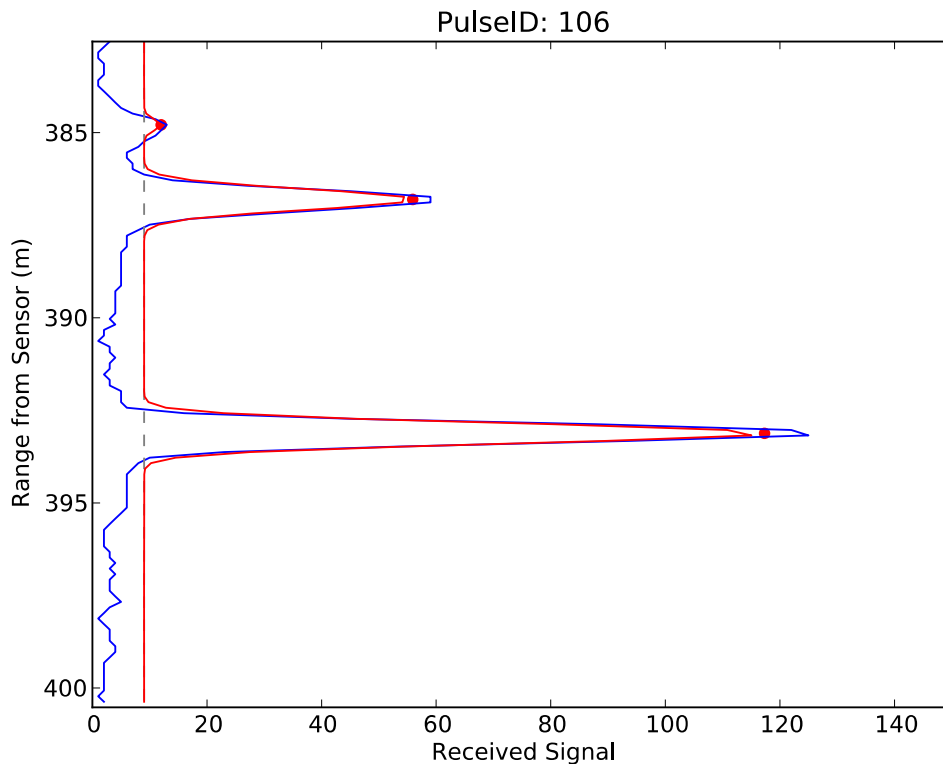


Figure 2. A received waveform (blue line) and decomposed discrete-returns (red dots) plotted using Matplotlib from a Python script using the SPDLib Python binding. The noise threshold is shown as the dashed line and the reconstructed waveform, from the fitted Gaussians, is shown in red.

2.3 Visualisation

Visualisation of the LiDAR point cloud and associated data (e.g., aerial photography, hyperspectral imagery) is important to many users. For example, the results of an applied algorithm (e.g., a ground return classification) or a simple overview of the distribution of returns can provide new insights into the information content of the data. For this reason, the SPD Points Viewer application has been created using C++ and QT4 to provide a cross platform visualisation tool built on top of the SPDLib software library.

The viewer first provides a window displaying the SPD files overview ('quicklook') image (Figure 3a), which directly maps onto the SPD files spatial index and allows the user to select a region of interest. A second view (Figure 3b) displays the selected region as 3D points. The 3D points can be coloured by a number of variables including return amplitude, width, classification, RGB values (e.g., from co-registered optical imagery), height and elevation, while the z component of the points can be either the topographic elevation or height field. The colouration of points by variables provides a high degree of flexibility in visualisation that allows data to be better understood and interpreted.

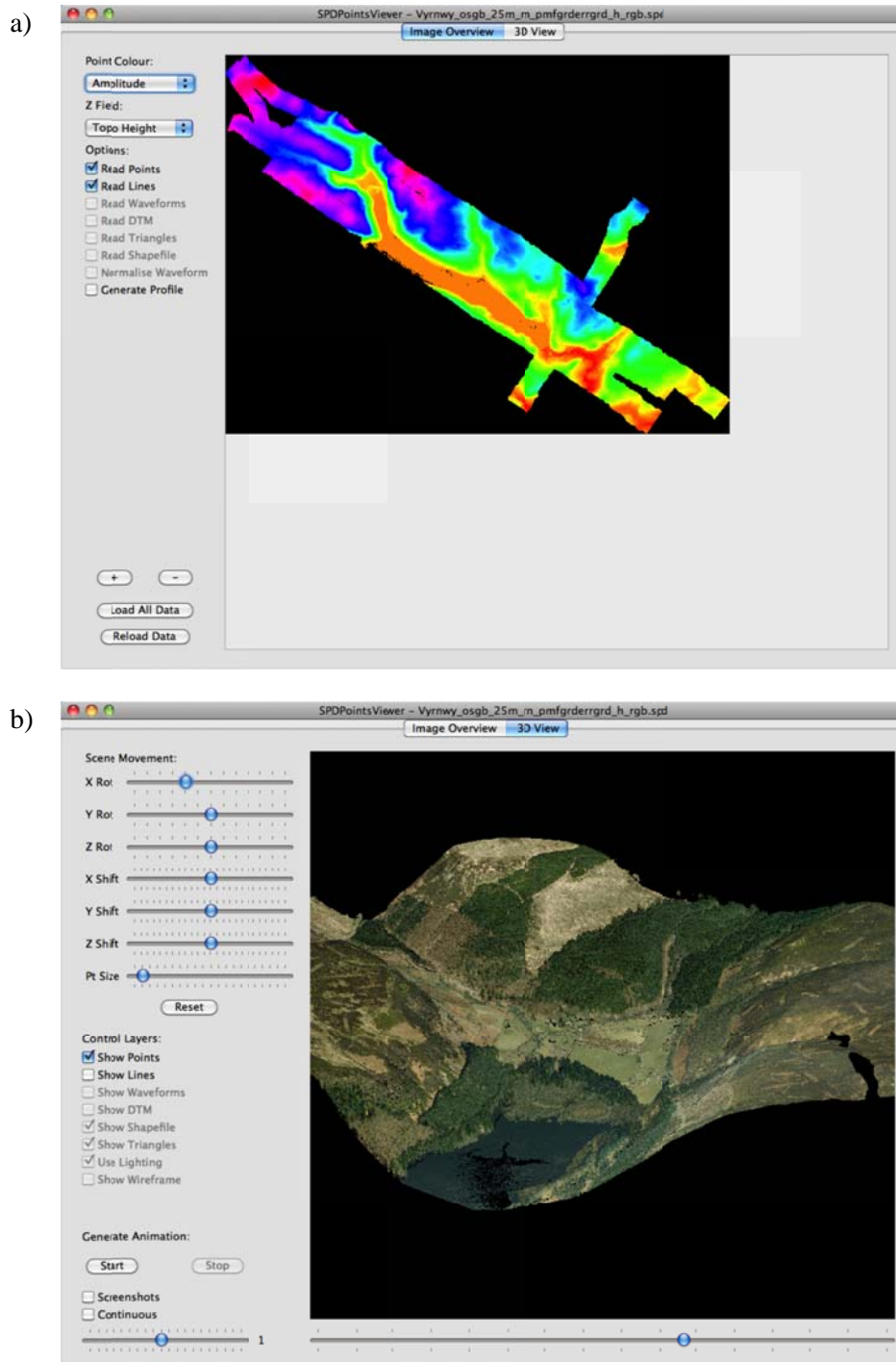


Figure 3. The user interface to the SPD Points Viewer. a) The overview image (Mean elevation; SPD file index), from which a region of the dataset can be selected and b) the 3D view of the selected region. In this case, the points are coloured using the RGB values from a true colour aerial photograph.

To visualise the full waveform data, the individual bins of the received waveform are displayed as 3D points coloured by their intensity value (Figure 4a) where a noise threshold (as shown in Figure 2) is used to remove waveform bins containing only noise. Taking advantage of the pulse-based nature of the SPD file format, the viewer also provides an option to display the vectors of the pulses (Figure 4b), which gives insight into the direction of pulse transmission and interaction with 3D structures and the ground surface and assists interpretation of the LiDAR data.

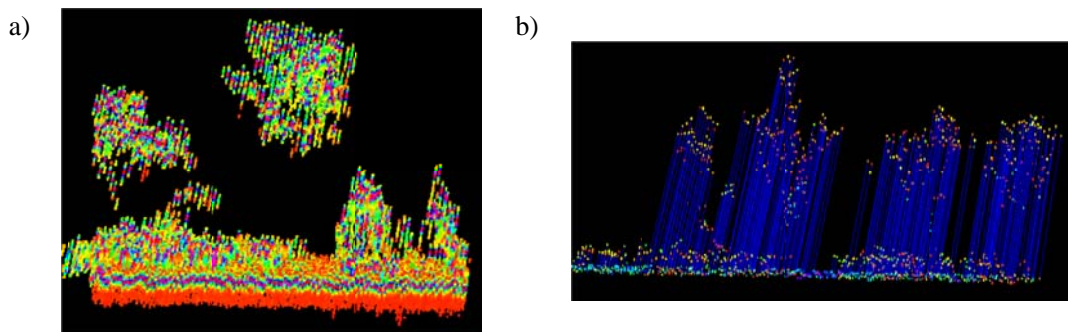


Figure 4 a) Full waveform data displayed within the SPDPointsViewer (red is low intensity and purple is high). b) The pulse vectors as displayed within the SPD Points Viewer.

3. Licensing and future developments

3.1 Licensing

The software is split into two parts for licensing. The main library is released under a General Public License (GPL) version 3 license (GNU General Public License, 2007) whilst the software to read and write SPD files is released under the MIT license (Open Source Initiative, 2011). Whilst provided completely free of charge, it is without a warranty or promise of support. The GPL license is a so called viral license, meaning that any works derived from this software also have to be covered by the GPL license and therefore distributed openly. Therefore, any community improvements or changes to the algorithms have to be fed back allowing everyone to benefit from this work, as they benefitted from the original. The SPD file readers and writers are released under the MIT license to allow support for this format to be included within other software packages (e.g., commercial) without the restrictions of the viral license.

3.2 Supported Systems

Currently the software is developed, built and tested on UNIX and Linux platforms but all the libraries and methods used support the Windows platform. Therefore, the library, tools and viewer are considered fully cross platform. In the future, it is the intention to also provide full support for the Windows platform and to make binary versions available.

3.3 Future Developments

One of the more significant future developments will be official support for Windows alongside UNIX and Linux platforms. In principle, this should be relatively simple to achieve, but will require full testing. Currently, the main weakness in the software is the writing performance of the SPD file writer. This can be solved through the introduction of buffering and further optimisations, which will be implemented in the near future. In terms of data processing, future developments are expected to concentrate on further algorithms for processing waveform data and specifically those that support the extraction of vegetation metrics. Finally, further binary version of the software are expected to be released, which will provide a much simpler installation process as, for platforms other than Mac OSX, the user is currently required to compile the software themselves.

4. Summary and conclusions

In many areas of research and application, terrestrial, airborne and spaceborne LiDAR have been acquired but their use has often been limited by the lack of capacity to store, visualise, interpret and process these data. In many cases, the software available has not been sufficient to address all of the requirements of the users and many have had to develop their own in-house, which has often not been available to others. SPDLib has been developed for these same reasons but the software has been made freely available, from <http://www.spdlib.org>, and is now open source. Furthermore, the software is easy to use by those with very little prior knowledge of LiDAR and will provide new options and insights for those that are more experienced.

The tools make use of the SPD file format, which provides fast data access through a spatial index and pulse based data structures, allowing large datasets to be efficiently processed. The software provides good capability for storing and processing waveform and discrete return LiDAR data from a range of platforms. The SPD Points Viewer also provides a unique opportunity to explore and understand discrete return and waveform data and allows integration with optical or even radar data. Development is ongoing, largely to support more data types and applications, improve performance and integrate new algorithms. However, the software provides basic and advanced processing and analysis capability that is robust and builds on knowledge and understanding from the scientific and user community. As such, its use in investigating and exploiting the enormous amounts of information contained with the diverse range of LIDAR data is encouraged.

References

- Armston, J., Denham, R., Danaher, T., Scarth, P. and Moffiet, T. 2009. Prediction and validation of foliage projective cover from Landsat-5 TM and Landsat-7 ETM+ imagery for Queensland, Australia. *Journal of Applied Remote Sensing*, 3: 033540.
- Bater, C., & Coops, N. (2009). Evaluating error associated with lidar-derived DEM interpolation. *Computers and Geosciences*, 35(2), 289–300.
- Bunting, P., Armston, J., Lucas R. M., and Clewley D. (2011) Sorted Pulse Data (SPD) Format: A new file structure for storing and processing laser scanning data. Submitted to *Silvilaser* 2011.
- Coops, N., Hilker, T., Wulder, M., & St-Onge, B. (2007). Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. *Trees-Structure And Function*.
- Evans, J. S., & Hudak, A. T. (2007). A multiscale curvature algorithm for classifying discrete return LiDAR in forested environments. *IEEE Transactions on Geoscience and Remote Sensing*, 45(4), 1029–1038.
- GNU General Public License (2007) <http://www.gnu.org/licenses/gpl.html>
- Hall, S., Burke, I., Box, D., Kaufmann, M., & Stoker, J. (2005). Estimating stand structure using discrete-return lidar: an example from low density, fire prone ponderosa pine forests. *Forest Ecology And Management*, 208(1-3), 189–209.
- Hill, R., & Thomson, A. (2005). Mapping woodland species composition and structure using airborne spectral and LiDAR data. *International Journal Of Remote Sensing*, 26(17), 3763–3779.

- Lee, A., & Lucas, R. (2007). A LiDAR-derived canopy density model for tree stem and crown mapping in Australian forests. *Remote Sensing Of Environment*, 111(4), 493–518.
- Mallet, C. and Bretar, F. (2009). Full-waveform topographic LiDAR: State of the art. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 1-16.
- Müller, J., & Brandl, R. (2009). Assessing biodiversity by remote sensing in mountainous terrain: the potential of LiDAR to predict forest beetle assemblages. *Journal of Applied Ecology*, 46(4), 897–905.
- Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing Of Environment*, 80(1), 88–99.
- Open Source Initiative OSI (2011). The MIT License (MIT)
<http://www.opensource.org/licenses/mit-license.php> [Last accessed June 2011]
- Sibson, R. (1981). A brief description of natural neighbor interpolation (Chapter 2). In V. Barnett. *Interpreting Multivariate Data*. Chichester: John Wiley. pp. 21–36
- Wagner, W., Ullrich, A., Ducic, V., Melzer, T., & Studnicka, N. (2006). Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(2), 100–112.
- Zhang, K., Chen, S., Whitman, D., Shyu, M., Yan, J., & Zhang, C. (2003). A progressive morphological filter for removing nonground measurements from airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 41(4), 872–882.

Comparison of Point Cloud Data Reduction Methods in Single-Scan TLS for Finding Tree Stems in Forest

Paula Litkey, Eetu Puttonen, & Xinlian Liang

Finnish Geodetic Institute, paula.litkey@fgi.fi, eetu.puttonen@fgi.fi, xinlian.liang@fgi.fi

The point density in a single-scan terrestrial laser scanner (TLS) point cloud is very dense close to the scanner and gets sparser as the distance from the scanner increases. A full circular scan can contain tens of millions of points, which is impractical for most algorithms that work on point data. The number of points can be reduced by taking a sample of the original data. We have studied what influence different sampling methods have on the number of points that falls on tree stems. We propose that the number of points available on a far-away tree can be increased with a smart data reduction scheme. The data reduction favours far-away points over the densely located points close to the scanner. The main findings of this study are that removing ground points before sampling gives a great advantage in data reduction and that a point selection using only horizontal distances (2D Cartesian, xy-plane) favours low points.

Keywords: terrestrial laser scanner, data reduction, ground model

1. Introduction

Finding stems from a TLS point cloud of a forest plot is essential task for several applications: tree map generation, forest inventory, change detection and biomass estimation. Many studies have already presented methods for stem modelling (Thies *et al.* 2004), canopy structure (Fleck *et al.* 2007) and tree modelling (Gorte and Winterhalder 2004, Hosoi and Omasa 2006, Xu *et al.* 2007). More automatic plot scale forest methods have also been presented for tree location (Liang *et al.* 2009), and volume estimation (Bienert *et al.* 2006). The problem of unequal point distribution in range has been mainly noted when only single scan data were used (Liang *et al.* 2009). In multi-scan methods, the combined point clouds of the trees have more equal sizes than in single scan, and thus the distance dependent point distribution has not lowered the detectability of objects far away from the scanner. The automation chain, however, is broken in most approaches in the registration of the point clouds. Henning and Radtke (2008) introduced a registration method for multiple forest range-images using ground points and stem centres as tie points is presented. According to their results the accuracy of the stem diameter estimates is generally dependent on the number of the surface points representing tree stems and this value varies between studies. We believe that a considerable gain in efficiency and performance can be achieved with a data reduction method that is designed for forest data. Both single- and multi-scan methods will benefit from smart data reduction, especially when the automatization level of TLS forest methods increases.

Laser scanner point cloud data reduction has been studied more in the field of reverse engineering. Lee *et al.* (2001) present uniform and non-uniform grid methods for surface data reduction. In several of the methods presented in (Lee *et al.* 2001), either point locations or their depth values are medians of points that lie on a grid cell. Combined normals of a triangulation inside a grid cell were also used. This kind of approach assumes that the points lie on a smooth surface. Another area where the TLS point cloud data reduction has been studied are registration applications (Madow *et al.* 2010). Many common registration methods are related to Iterative Closest Point ICP (Besl and McKay 1992) and suffer from dense point clouds. Madow *et al.* (2010) classify data reduction into two types: range-dependent and range-independent. In the range-dependent methods, each scan is processed to find feature points with special characteristics. These can be representative points from sequences with similar ranges in the same 2D slice, points from salient geometrical regions, mesh vertices or octree cube centres.

The range-independent data reduction can be based on random or uniform selection strategies. Spherical selection strategy is introduced in (Mandow *et al.* 2010). The data reduction problem, however, is different in forest than either in built environment or in reverse-engineering: the number and distribution of trees is arbitrary and there are no surfaces on which the points are ordered.

In this study, different data reduction methods were tested on a single-scan point cloud measured in a forest plot. Data ranging up to 20 meters from the scanner were used. It was of interest to find such a method that would preserve the stem points as well as possible. We wanted to process the whole scan as one piece of data, since if the data were cut into pieces then the edges would have to be separately checked for stems that had been split between two pieces of data. Point distance from the scanner or single point location information was used to form as even distance-based point distribution as possible. The distance computation added to the total processing complexity. On the other hand, consecutive processing was simpler because the point distribution was more equal on objects that lie on different distances from the scanner. Data reduction by picking every n^{th} point was used as a reference method.

We tested the effect of different data reduction schemes with two different scenarios:

1. Sampling was performed from all data points
2. Ground points were removed before the sampling, the number of points that were selected was the same as in 1.

The text is divided as follows: The data acquisition and original point cloud properties are described in chapter 2, the used data reduction strategies and ground detection method are described in chapter 3, results on point distance distribution and available points on tree stems are given in chapter 4 and discussion in chapter 5.

2. Data

The data was scanned in Evo, Finland (61.19°N, 25.11°E) using a Leica HDS6000 scanner (Leica Geosystem AG, Heerbrugg, Switzerland) in 2008. The scanner uses phase-shift measurements of continuous waves to measure the distances. The scanning rate was 500,000 points per second and the angle increment in both directions was 0.036° , point spacing at 10m distance was 6.3 mm. The spot size at exit was 3 mm (based on Gaussian definition) and the beam divergence 0.22 mrad. The scanner angle space was $360^\circ \times 320^\circ$. The scanner automatically filters the raw point cloud of approximately 40.4 million points. The clear-sky noise points, points with low intensity and all points closer than 0.8m from the scanner were pre-filtered during the data transportation. The pre-filtered data consist of a single scan point cloud of 31.5 million points. The data are in local Cartesian xyz-space with the scanner set in the origo. The range distribution of the point cloud is illustrated in Figure 1.

In Figure 1, the top left plot (a), illustrates the point distance distribution. The blue patch describes the ground points and the green one the vegetation points. The cyan and magenta part visualize the ring that is highlighted in the plot b. It can be seen that most of the points lie inside 10m horizontal distance from the scanner. In more detail: 50% of the points lie within 2.5m, 80% within 6m, 89% within 10m, 95% within 14m, and 98% within 20m from the scanner. The top right plot (b) of Figure 1 shows how the area of a one-meter wide ring increases at each distance. The locations of example trees are plotted with stars. The bottom left plot (c) shows the average point density in each distance ring in blue (curve with top on the left) and the area of the rings in green (ascending line). The cyan lines show the cyan ring of plot b. The point density is not monotonically decreasing with increasing distance, because vertical objects (trees) introduce local peaks in the point density and in the closest distances, pre-filtering has been

applied. The bottom right plot (d) shows the effect of beam divergence compared to the growing size of the resolution cell as the distance increases. It can be seen that in distances $<7\text{m}$ the points illuminating a planar surface are overlapping. Thus, the data are redundant in close range surfaces further emphasizing the need for data reduction.

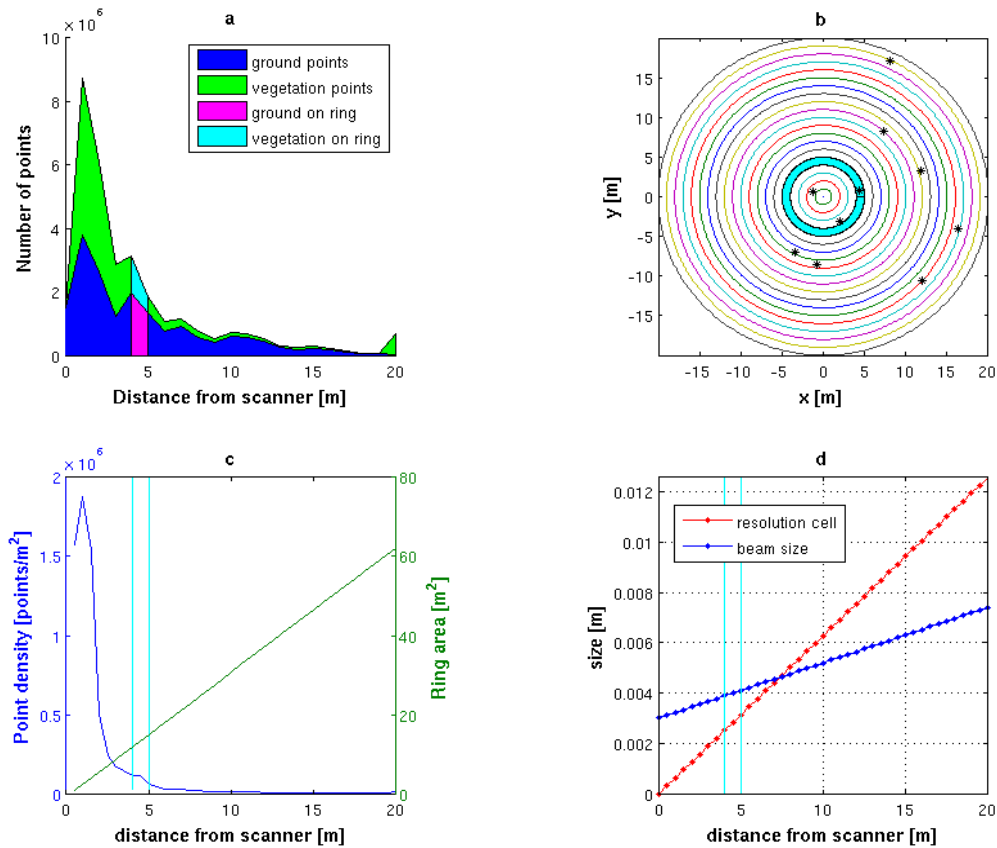


Figure 1: Point distribution of the pre-filtered point cloud. a) Distribution of all points and the proportion of ground points in different distances. b) Visualization of the reference trees and sizes of one-meter rings at different distances from the scanner. c) blue - Point densities on the one meter rings and green - the ring area. d) The beam diameter and resolution cell diameter at different distances.

2.1 Reference data

Stem point clouds were manually delineated from the pre-filtered point cloud within the horizontal distance of 20m from the scanner. Tree specimens of similar size were selected as example cases from different distances from the scanner. The stems and their properties that were computed from the pre-filtered point cloud are listed in Table 1 and the respective full point clouds are visualized in Figure 4, top plot.

Table 1: reference stem data from full point cloud.

Tree Id	Distance (m)	Number of points	Stem height (m)	DBH estimate (cm)	Average point density (1/m ²)
1	1.4	858907	7.6	27.6	431326.3
2	3.7	115991	9.8	14.8	86325.5
3	4.4	62080	9.1	18.8	37823.4
4	7.8	49030	12	26.5	15924.4
5	8.6	41410	9.7	23	19211.5
6	11	19800	9.9	25	8236.7
7	12.3	18707	11	36.5	4807.6
8	16	6027	11.1	21.4	2618.5
9	16.8	5176	9.7	22.6	3911.5
10	19	4530	12.1	17.1	2978.3

3. Sampling methods

We made two basic assumptions about the collected point cloud. The assumptions were based on the spherical scanning geometry. The first assumption was that the point density of the cloud was expected to decrease as r^{-n} in general, where r is the point distance from the scanner. The second assumption was that the angular distribution of the point cloud was considered to be equal in all directions. In a real scanning situation these assumptions do not strictly hold, as there will be returns from ground and surrounding objects. Also, the scanning angle was limited to the downward direction. However, the random point selection was used because it did not require any preceding knowledge of the scanned area. We used two different point selection approaches: distribution based and grid based selection.

Random point selection, with every n^{th} point picked was used as a reference. No distance information were used. This sampling preserved the form of the original point distance distribution.

3.1 Distribution based point selection

The total of six different sampling methods were tested:

Weighted random bin sampling (WRBS) that was carried out in two steps. First, the points were divided into a distance histogram with a pre-set number of bins. Then, a random bin was selected. The selection was weighted with the number of points in the bin. The bin selection was controlled with a user-set threshold value. The threshold value was set so that the least populated distance bin was selected with 90% acceptance rate. This was done to equalize the point distribution. The acceptance rate for other populated bins was normalized according to this value. After a bin was selected, a single point in it was picked randomly. This sampling was performed with two different variations: with the absolute point numbers in a bin and with the logarithm of the absolute point numbers in a bin. The logarithm made the final sampled point distribution more even, but it also gave additional sampling efficiency by lowering the rejection rate of the most populated bins during the first randomization.

The 'cumulative sum'-value sampling method resembled the Case 2, but with a difference that the cumulative sum value (CSV) of each bin was used as a selection parameter instead of the absolute point number. In the 'CSV' point sampling, all points were first binned into a distance histogram like in the Case 2. Then, the bins were arranged in a descending order according to their point number ratio. Next, a bin-wise cumulative sum of the point ratios was calculated. After this, the calculated bin-wise cumulative sum value was given to all individual points within the bins. After CSV assignment all points with a CSV higher than a user-defined threshold were added to the sampled point list. Finally, the remaining points were picked by randomly selecting one and then comparing its CSV value against another random number. A point was accepted, if its CSV was higher than the generated random number.

In addition to the previous three selection methods the following three methods were also tested. Their point selection algorithm was the following:

1. Calculate single point distances from the scanner with a selected distance metric (2D / 3D)
2. Arrange a point list according to the distances in an ascending order
3. Pick a random point from the list using a selected sampling method (Selections 1-3, below)
4. Add the picked point into a reduced point list and mark it as used in the original list.
5. Repeat steps 3 and 4 until a wanted number of points have been selected.

Selection 1: Random square root point picking from two-dimensional point distance distribution (2D, $r^{1/2}$ horizontal distances in the Cartesian scanner space, xy plane). The point sampling was carried out by using a square rooted random point distribution that emphasized the chance to pick a point located far-away from the scanner. Also, the square rooted random distribution produces an even point distribution in cases where the point density decreases as r^{-2} .

Selection 2: Random square root point picking from three-dimensional distances (3D, $r^{1/2}$). The random point selection was carried out as in the Case 3, but in this case a 3D point distance distribution was used instead of the horizontal distance distribution.

Selection 3: Random cubic root point picking from horizontal distances (3D, $r^{1/3}$). The random point picking was carried out with a cubic rooted random point distribution. The cubic rooted random point distribution was used as it produces a uniform point density in a three-dimensional spherical case.

3.2 Grid based point selection

The grid based point distribution was based on computing a horizontal grid and a grid cell membership for each point. The total number of points N (computed from the wanted percentage) that was selected from the cells cell was equally divided between all populated cells whose number was N_{pop} . The number of points that was selected from each cell, N_c , was then $N_c = \text{round}(N/N_{pop})$. There were cells with less points than N_c , the balance between N_c and cells, that were completely included was found iteratively. Points within the cells with more than N_c points were uniformly random sampled to form a cell-wise subset of N_c points. Extra points that were caused by rounding were cut off with random selection to set the number of selected points exactly to the wanted percentage. The grid based selection was computed with a square grid of one-meter cell size and with a polar grid with 2m horizontal distance steps and $\pi/20$ radian angular steps. In the polar grid, the cell size increased as the distance from the scanner grew.

3.3 Ground points

Ground point removal from the data was performed by transforming first the whole point cloud into polar coordinates in Cartesian xy-plane. Then, the data were binned with a polar histogram. A polar histogram was used as it suited better for the hemispherical scanning geometry than the Cartesian one. The ground level was set by going through every bin in the polar histogram and computing a height histogram with the points within them. This was performed by comparing the absolute height of the lowest point in a single polar bin against the height of the most populated height bin in the same polar bin. The smaller height value in the comparison was then selected as the ground height. The polar histogram bins were connected to each other with Delaunay triangulation after their heights had been set.

The number of ground points found with the described method was 13.9 M leaving 17.6M points to be sampled from (the whole data were 31.5M points).

4. Result

The sampling algorithms were studied in cases where the data has been reduced to 25% - 5% of the original points. Cases where the ground was removed were sampled so that the number of selected points was the same as in cases that included the ground. In the following results the 5% level is shown, because the results were most visible in that level. Figure 2 illustrates the pre-filtered point cloud distribution and distributions from different data reduction schemes.

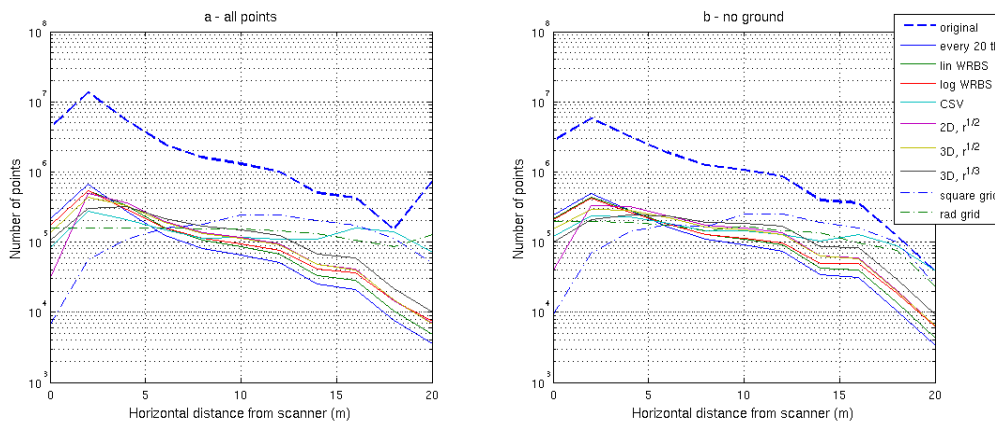


Figure 2: Original point distribution and distribution after the point cloud was reduced to 5% using different methods, a) data reduction from all points, b) from points that are not ground

Figure 2 shows that most of the distribution based sampling methods (linear WRBS, logarithmic WRBS, 2D square root, 3D square root and, 3D cubic root) have similar general form to the original point cloud, whereas the grid based data reduction schemes have their own shape. The logarithmic y-scale diminishes the differences between the distribution based sampling methods.

To compensate size differences of the example trees, an estimated diameter at breast height (DBH) and stem point cloud height were used to compute average point densities for stems. The average point densities on tree stems with different data reduction schemes are listed in Table 2.

Table 2: Average point densities on reference stems after data reduction to 5% from original, ground included

Method	Stem ID									
	1	2	3	4	5	6	7	8	9	10
pre-filtered	431326	86326	37823	15924	19212	8237	4808	2619	3912	2978
Every 20th	21594	4312	1895	791	961	418	232	135	195	147
Lin WRBS	22013	5145	2309	1009	1247	549	315	167	252	209
Log WRBS	19441	4390	2036	1070	1326	599	363	206	359	263
CSV	3498	3039	1493	779	978	477	277	462	2347	2978
2D, $r^{1/2}$	11701	5576	2654	1333	1674	746	437	259	375	297
3D, $r^{1/2}$	5555	4886	2409	1257	1573	717	452	243	364	279
3D, $r^{1/3}$	3167	4178	2357	1496	1857	968	607	346	531	437
Square grid	368	483	640	588	518	465	370	362	626	887
Polar grid	2156	1020	669	452	539	392	266	288	692	1014

The figures in Table 2 confirm the similarity of the distribution between both WRBS- and n^{th} root methods. With the exception of 3D methods, they are not efficiently reducing points in the area closest to the scanner.

In the square grid method, the number of points close to the scanner is low, because the of grid cells is small. In the case with ground points included, the number of points on the closest stem declines so that the original stem height is not preserved (Figure 4, bottom). In the polar grid, the number of grid cells is the same at all distances.

It should be noted, that the figures of Table 2 are merely single realizations of random processes on single data and thus minor differences between the results cannot be used to rank the methods. Table 2 should instead be taken as directional tool to figure out which methods work and why.

Six data reduction methods were selected for comparison on individual stem level based on these results. The selected methods were: every 20th (reference method), 2D square root selection (representative of all similar distributions), 3D cubic root selection, CSV selection and the square- and polar grid methods.

Figure 3 illustrates that the ground point removal increased the number of the stem points on all distances. To illustrate the problem with too large point clouds we modelled the stems using modelling algorithm described in Litkey *et al.* (2008), it fitted circles to single stem height slot data. Due to ineffective data reduction in the closest range, the processing time in the every 20th and 2D sqrt schemes remained high. The ground removal made the processing even slower.

The model extent measure visualized in the bottom row of Figure 3 was computed as the ratio between the number of rings fitted to the stem point cloud in the original and the reduced point clouds. If the effective height of the stem is reduced too much, it might not be found in the stem detection. Also, in Henning and Radtke (2008) tie points from different heights of the stems were used in referencing, so the number of possible tie points would be reduced if the stem shortened in data reduction.

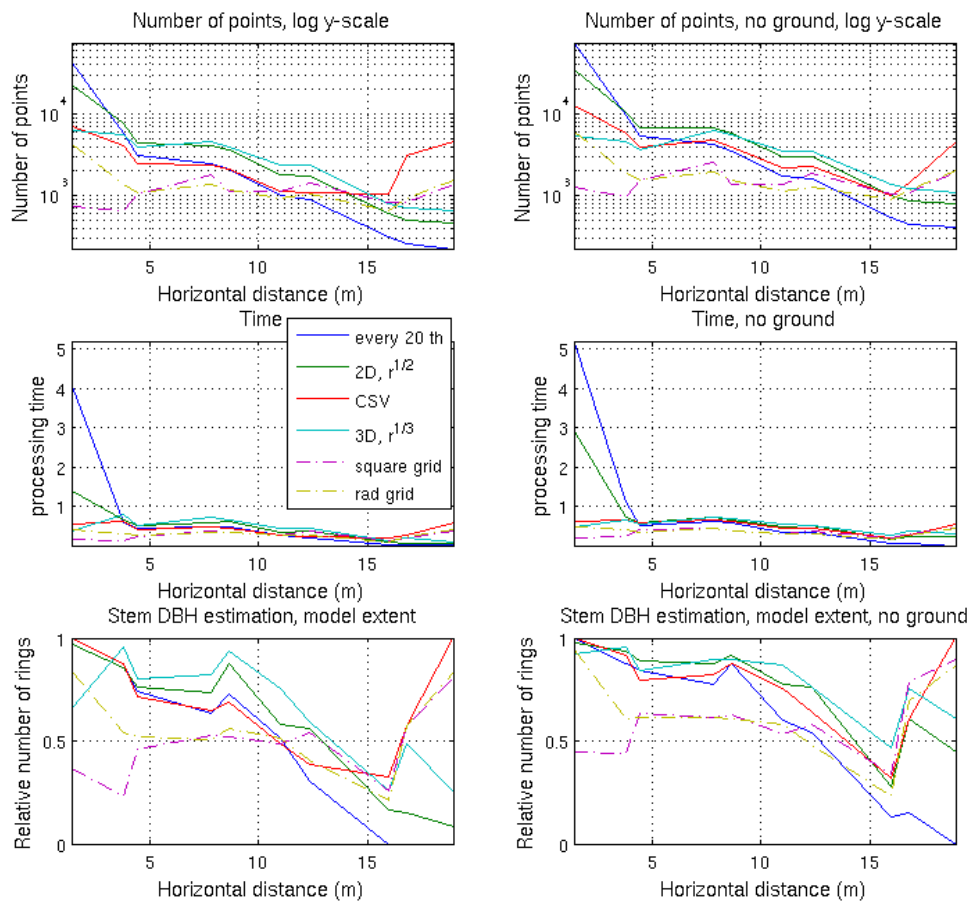


Figure 3: Comparison of selected data reduction schemes on stem level. Left - all points, right - no ground points. Top – number of points on the reference stems by distance, middle – time required for stem modelling. Bottom – the number of modelled height slots divided by the number of modelled height slots in original point cloud.

The stem point clouds at different horizontal distances from the scanner in original, every 20th and square grid samplings are plotted in Figure 4. The stem point clouds were shattered into pieces in the distance mainly because of branch shadows.

Overall, in this data, in distances up to 10m, with all sampling levels, all stems presented in Table 1 remained detectable and their stem shapes were preserved. In the ranges where beam overlap exists (up to 7m, Figure 1d), data reduction is needed to cut the large stem point clouds for efficient processing. Further away, the visibility of the stems can be secured if as many points as possible are included.

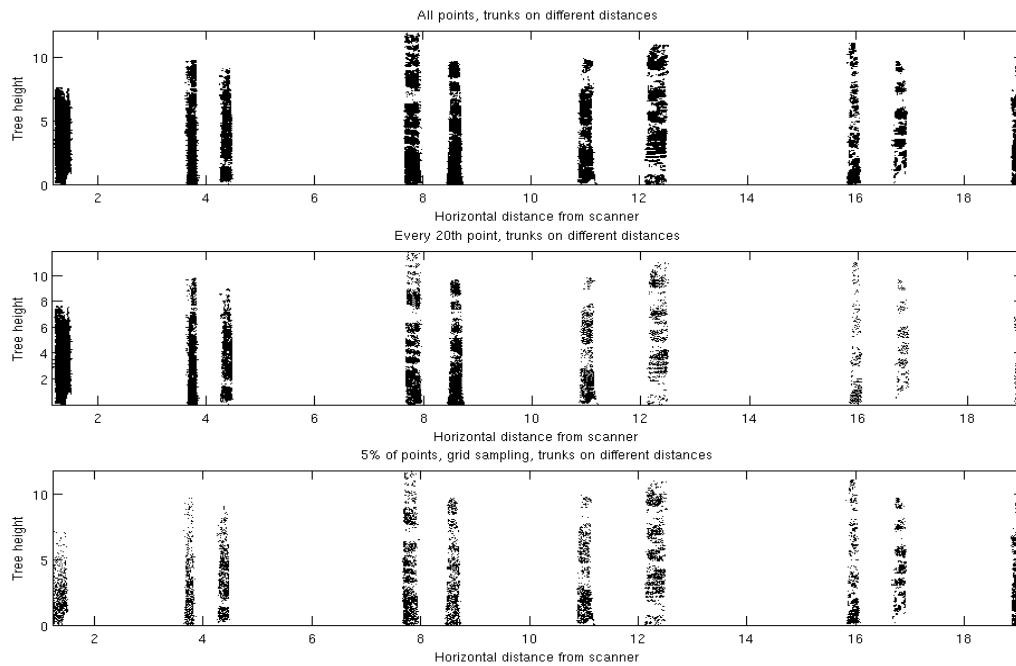


Figure 4: Tree stem point clouds on different horizontal distances, ground points included. Top - original, middle - every 20th sampling, bottom - grid sampling with square grid.

5. Discussion

The results of this study show that stem data can be observed from increased range with reduced point cloud if the data reduction scheme is selected carefully. The role of the ground points is significant, since they represent almost half of the data on the close ranges (Figure 1, distances below 3m). The data reduction scheme is essential not only to increase the stem detection range, but also to decrease the size of unnecessarily dense stem point clouds in the close range. The ground removal alone is not sufficient, since in the close range the individual stem point densities increase if no other data reduction is used.

Based on this study, we suggest the following roadmap for data reduction in a single-scan TLS point cloud:

1. ground point removal.
2. saving all the points in the furthest distance of interest (not reducing their number).
3. selecting individual points in distance bins or grid cells or sampling so that the local point distribution is accounted for.

The results of the study imply that the every n^{th} sampling is not optimal for forest data. Also the use of 2D distance measure leads to selecting unnecessarily many ground points when they are included.

The presented data reduction schemes can be readily used to reduce the point number in different applications where several TLS scans are merged to form large datasets. In Liang *et al.* (2010) the possibility to detect tree stems from several different scans and then georeferencing only the stem points instead of full point clouds has been studied. In another study (Lehtomäki *et al.* 2011, submitted) used stop-and-go collected laser scanner data as a reference for a mobile laser scanning application. The data reduction scheme proposed in this study would allow a significant reduction of point cloud size making several-scan studies more manageable while retaining their inherent accuracy.

Acknowledgements

We would like to thank anonymous reviewers for their helpful comments on the beam overlap. And Harri Kaartinen for the field TLS measurement data and Antero Kukko for details on the scanning parameters.

References

- Besl, P. J. and McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (2), 239-256.
- Bienert, A., Scheller, S., Keane, E., Mullooly, G. and Mohan, F., 2006. Application of terrestrial laser scanners for the determination of forest inventory parameters. In: *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, Vol. 36-5.
- Fleck, S., Obertreiber, N., Schmidt, I., Brauns, M., Jungkunst, H.F. and Leuschner, C., 2007. Terrestrial lidar measurements for analysing canopy structure in an old-growth forest. In: P. Rönnholm, H. Hyypä, J. Hyypä (Eds.). *Proceedings of the ISPRS Workshop 'Laser Scanning 2007 and SilviLaser 2007'*. Espoo.125-129.
- Gorte, B., Winterhalder, D., 2004. Reconstruction of Laser-Scanned Trees using Filter Operations in the 3D-Raster Domain. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* Vol. XXXVI, Part 8/ W2: 39-44.
- Henning J. G. and Radtke P. J., 2008. Multiview range-image registration for forested scenes using explicitly-matched tie points estimated from natural surfaces. *ISPRS Journal of Photogrammetry & Remote Sensing* (63): 68 – 83.
- Hosoi, F. and Omasa, K., 2006. Voxel-based 3-D modeling of individual trees for estimating leaf area density using high-resolution portable scanning lidar. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 12, 3610-3618.
- Lee, K. H. , Woo, H. and Suk, T, 2001. Data reduction methods for reverse engineering. *International Journal of Advanced Manufacturing Technology* 17: 735:743.
- Lehtomäki M., Jaakkola A., Hyypä J., Kukko A., and Kaartinen H., 2011. Extraction of Poles and Tree Trunks from Mobile Laser Scanning data, submitted to *ISPRS Laser Scanning*.
- Liang, X., Litkey,P., Hyypä, J., Kaartinen, H.,Vastaranta, M. & Holopainen, M., 2009. Automatic tree location mapping using TLS for plot-wise forest inventory. *Proceedings of Silvilaser 2009*, 14-16 October, College Station, Texas USA.
- Liang, X., Litkey,P., Hyypä, J., Kaartinen, H., & Holopainen, M., 2010. Automatic stem location mapping using several single-scan TLS for plot-wise forest inventory. *XXIII IUFRO World Congress*, 23-28 August, 2010, Seoul, Korea.
- Litkey, P., Liang, X, Hyypä, J., Kukko, A., Kaartinen, H. and Holopainen, M., 2008. Single-scan TLS methods for forest parameter retrieval. *Proceedings of Silvilaser 2008*, 17- 19 September 2008, Edinburgh, UK.
- Mandow, A., Martínez, J. L., Reina, A. J. and Morales, J., 2010. Fast range-independent spherical subsampling of 3D laser scanner points and data reduction performance evaluation for scene registration. *Pattern Recognition Letters* 31:1239-1250.
- Thies M., Pfeifer N., Winterhalder D., and Gorte B. G. H., 2004. Three-dimensional reconstruction of stems for assessment of taper, sweep and lean based on laser scanning of standing trees. *Scandinavian Journal of Forest Research*, 6(19): 571 – 581
- Xu, H., Gosset, N. and Chen, B., 2007. Knowledge and heuristic-based modeling of laser scanned trees. *ACM Transactions on Graphics*, Vol 26, No 4, Article 19.

Automated Log Counting: Proof of Concept Algorithm

Hamish Marshall¹

¹ Interpine Forestry Ltd (hamish.marshall@interpine.co.nz)

Abstract

Improving log inventory is a key area where the New Zealand forestry industry could significantly improve its supply chain performance. Although the process of counting logs seems relatively simple; in reality it is a difficult and labour intensive job. This is particularly significant to the New Zealand log export industry which is required to count and barcode every log (excluding pulp) that is exported. The fluctuating nature of export markets means that automated methods of counting logs hold significant potential. This paper investigates the accuracy of log counts for logs in pile/stacks using 3-dimensional (3D) point cloud data obtained from a ground based LiDAR scanner. In the past there have been a number of attempts to develop an automatic log counting system, the majority of these have used 2-dimensional photographic images. It was hypothesised that using 3D point data would overcome some of the problems that these approaches have encountered in the past. The validation study carried out on the algorithm showed that logs can be accurately counted and log diameters can be measured. Further work would be required to develop the algorithm into a commercial product and to determine the most cost effective hardware required to collect the 3-dimensional data required by the algorithm.

1. Introduction

Exact log counts are required at marshalling points and at shipside. Present systems rely on truck drivers and stevedores to accurately count the number of logs both on truck and in piles (bunks) on the ground at the wharf prior to loading. Incorrect log counts incur a cost through additional labour to correct mistakes identified at a later stage, or from shipping unticketed logs where errors are not identified. New approaches are being considered to improve on the current manual system.

Automated log counting using standard photography and digital image processing has been trialed in the past however accuracy levels were not high enough for this combination to be considered practical. There is one example of stereo photography being used, sScale^(TM) developed by Danish Company (Dralle Ltd, www.dralle.dk, Accessed on 3 July 2009) who have commercialised a system to both count and measure the volume of logs. To overcome some of the problems of standard photography in this application, ground based LIDAR has been suggested. A Chilean company has developed Logmeter⁴⁰⁰⁰ (www.woodtechms.com, Accessed on 3 July 2009) that uses LiDAR (Light Detection and Range) to measure and scale volume of logs on a truck.

Whether using standard, stereo photography, or LiDAR to capture the data, processing algorithm are required to automatically extract the information. LiDAR produces a dataset of 3D points where the x,y and z coordinates are known for each point on the object that the laser strikes. This type of data has several potential advantages for counting objects in an image over standard photography. There are at least two general approaches that could be used in this application:

- The point data is turned into an image which is then used in a traditional image

- processing algorithms (thresholding and watershed) to count the logs;
- Use the raw point data and develop object recognition algorithm (subject of this report).

The objective of this project was to develop and test a prototype log counting algorithm that uses the 3D point cloud dataset generated by a ground-based LiDAR scanner. This paper details the methodology behind the algorithm and two validation exercises that were design to establish proof of concept.

2. Methodology

2.1 Algorithm Development

The LiDAR Log Counting algorithm is made up of three basic components: searching, filtering and stopping. This algorithm has been developed only to use the x, y, and z data from the ground based LiDAR scanner. The number of user inputs into the algorithm has been kept to a minimum. The users only have to estimate the allowable maximum and minimum diameters of the logs in the stack. This is as simple as entering the maximum LED (large end diameter) and minimum SED (small end diameter) for the log grade being counted. Figure 1 gives an overview of the algorithm.

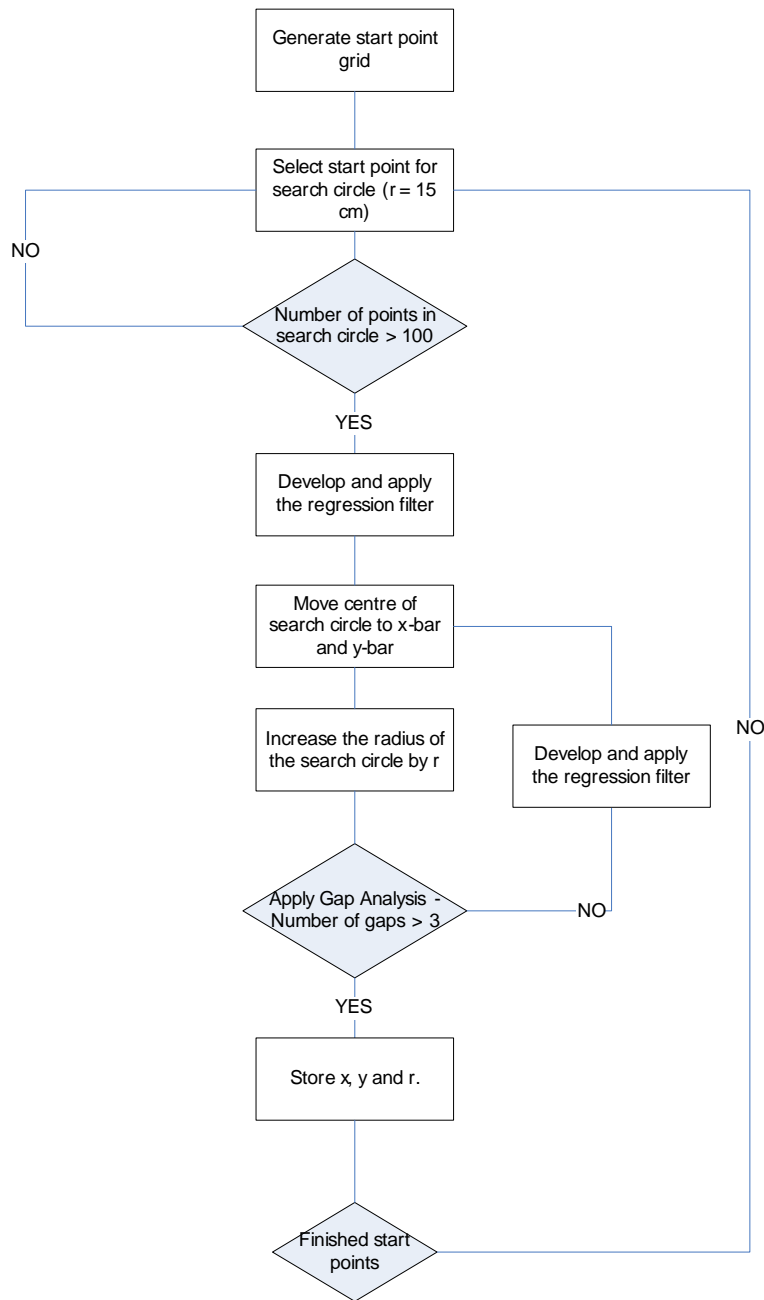


Figure 1. A flow chart of the LiDAR Logging Counting (LLC) algorithm

2.1.1 Searching Component

The searching component of the algorithm uses a systematic grid of starting points to find good locations to start the localised search. A “GOOD” search start point is defined as having at least 100 points within a search circle of 150 mm. Once a “GOOD” start point is found the search circle centre is moved so that it is centred on the mean x, y co-ordinates from the old circle’s location (Figure 2).

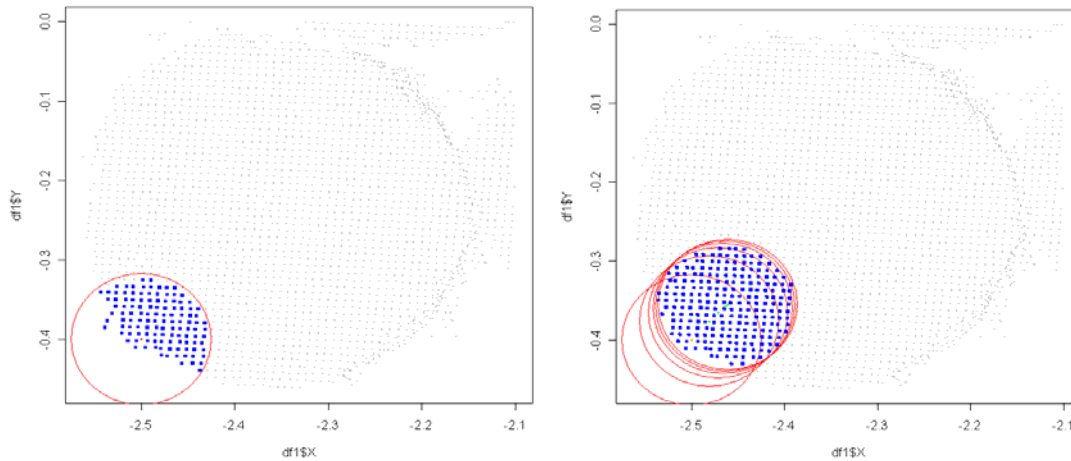


Figure 2. The demonstration of how the search circle is moved.

The search circle centre is continually shifted until the distance that the search circle moves is less than 1 mm. Once this point is reached the radius of the search circle is increased in a linear fashion. The radius of the search circle grows at a constant 1 cm each iteration (Figure 3).

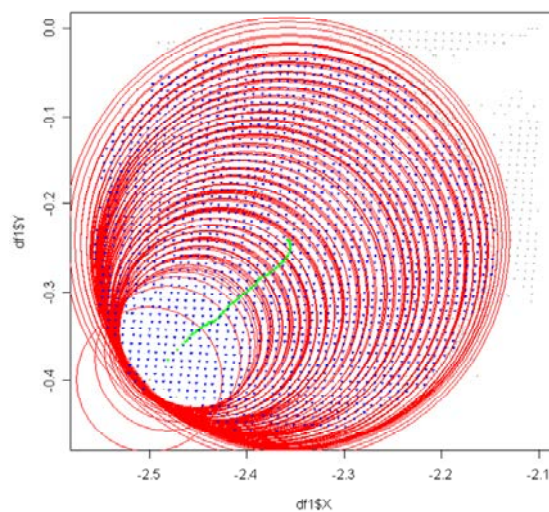


Figure 3. The search pattern of the algorithm

2.1.2 Filtering Component

This algorithm is primarily looking for flat circular surfaces of points that have reflected off the log ends. However the raw LiDAR dataset contains points that not only have reflected off the log end but also off other objects and the sides of the logs. The algorithm uses an adaptive filter to eliminate as many points that have not reflected off the current log end that is being searched. The algorithm utilises multiple linear regression to develop a model of the log end surface. This relies on the following characteristics of the LiDAR dataset:

- The log ends are normally cut with a straight face and hence can be modelled using a linear regression;
- Neighbouring logs are really on the same plane.

The points within the search circle are used to develop a linear function relating the x and y values to the z values. This function is then applied to the whole LiDAR dataset any point with a residual value (actual – predicted) less than 0.0075 is removed from the dataset until the current log has been isolated.

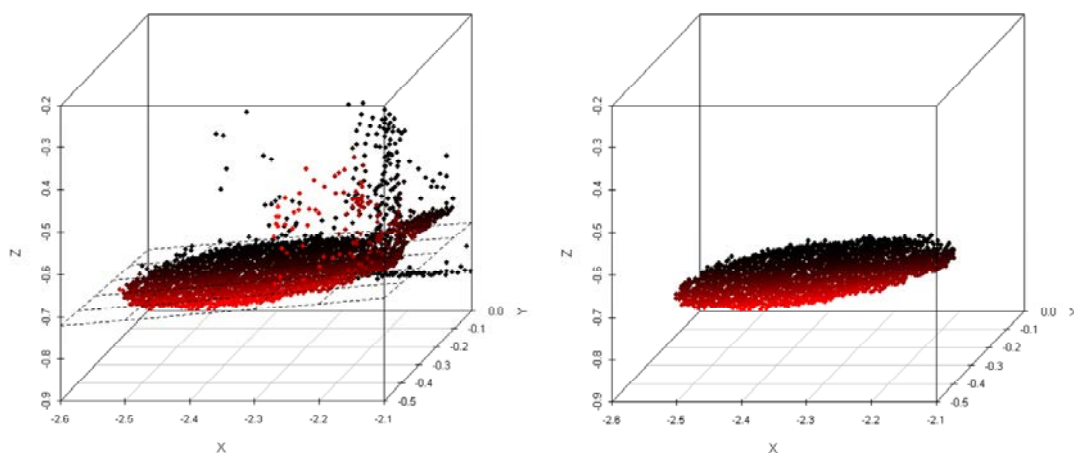


Figure 4. The effects of the regression filter.

Figure 4 shows the impact of the filter in eliminating points within the dataset that do not represent part of the log surface.

2.1.3 Stopping Component

Each time the diameter increases the stopping criteria is calculated; the criteria simply looks at the number of gaps in the outer 20 mm of the search circle. The angle from a horizontal axis is calculated for each of the points in this outer circle. The points are then sorted according to this angle. For the stopping criteria to be met there has to be at least three valid gaps. The rules that are used to count the number of gaps are as follows:

- Gaps in the direction that the search circle is moving are not counted;
- If the angle (radians) is greater than 0.6 a gap is counted;
- If there is only one point between two gaps then the two gaps are considered to be only one;
- If the gap comprises an angle greater than 2 radians, then that is considered to be made up of 3 gaps.

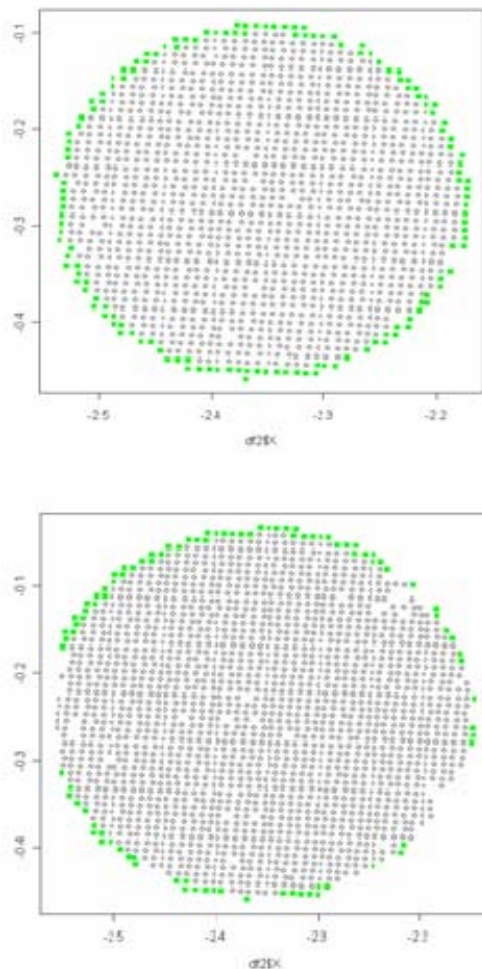


Figure 5. An illustration of the stopping criterion.

Once the stopping criterion is reached the search circle diameter (an approximation of the log diameter) is checked against the maximum and minimum allowable diameters for the log pile. If the target diameter is outside the allowable diameter range it is rejected as a log and not added to the count. If the log is inside the allowable diameter range it is added to the count and the points in the image that made up that log are removed.

2.2 Data Collection

The source data for this project was acquired by scanning the end of the log pile with a ground-based LiDAR scanner. The LiDAR scanner collects a 3 dimensional point dataset of the log faces. The scanner used in this trial was capable of 4000 points per second, with a maximum density of 1 per mm^2 . For this application, it seems that a scanning density of 1 per cm^2 (measured at maximum range) seems most suitable. On average, a scan at this density took no more than 5 minutes to scan a bunk of logs. The LEICA Scan Station (www.leica-geosystems.com) used in this trial was also capable of collecting the intensity of the reflected laser beam as well as the RGB (red, green blue) value for each point. These were not used in the algorithm to assist in the counting as they are not collected by cheaper LiDAR scanners.

Figure 6 gives an example of data collected during the scanning of the end of a bunk of logs. This dataset can be filtered to remove unwanted data such as returns from the log bunk and other objects in the background.

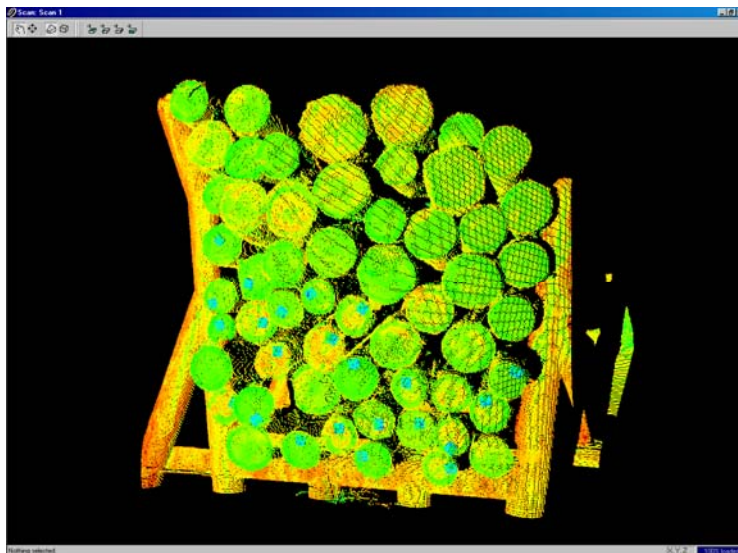


Figure 6. A view of the raw LiDAR data

2.3 Validation

A validation dataset was collected at the Port of Tauranga using the LIECA Scan Station. In total 16 scans were collected. It would have been beneficial to take a more systematic approach to collecting this data to cover a wider range of different log piles, log diameters and scanning locations. However project financial constraints combined with the high cost of hiring the scanning equipment meant that these scans were largely taken in largely an advantageous matter. Of the 16 scans; 6 were log piles in bunks with the remainder being log piles on the ground. Figure 7. Three examples of the range of the log stacks included in the validation dataset.



Figure 7. Three examples of the range of the log stacks included in the validation dataset.

In this study the scanner location was largely left to the third party scan operator. In most cases the location was no less than 2 metres away and not greater than 10 metres way from the pile/stack. The

scanner was also placed directly in front of log pile and as close to the centre of the pile as possible.

2.3.1 Log Count Validation

The algorithm applied to each scan. For each scan the algorithm user input variables (minimum diameter/ maximum diameter) were changed to match the grade specification of the logs in the pile. The algorithm log counts were compared to human counts taken off the images. Human counts were repeated several times to verify that the manual counts were accurate.

2.3.2 Log Diameter Validation

When the original scans were taken there was insufficient time to measure diameters in the field. However given the 6 millimetre positional accuracy of the LiDAR scanner it was deemed that measuring the diameters manually from the images would give an equally accurate representation of the true log diameters as field measurement. A R script was written to allow a human to measure the log diameter from the LIDAR scans.

Figure 10 shows how the R script allowed the manual measurement to be made from the LiDAR scans on a computer screen. Figure 10 was created from the LiDAR; the LIECA Scanning Station used to collect this data contained not only returns the x,y and z position but also the red, green and blue value and the laser return intensity.

3. Results

3.1 Log Count Validation Study

Table 1 shows the actual (manual) log counts compared to the counts produced by the algorithm. In Scan O only an approximate count could be carried out due to poor image quality.

Table 1. Manual vs Algorithm Log Count

Scan	Actual Count	Algorithm Count	Percentage Accuracy
A (Bunk)	14	14	100 %
B (Bunk)	35	35	100 %
C (Pile)	31	30	96.6 %
D (Bunk)	64	64	100 %
E (Pile)	209	204	96 %
F (Bunk)	24	24	100 %
G (Bunk)	24	23	96 %
H (Pile)	14	14	100 %
I (Pile)	31	31	100 %
J (Pile)	44	45	97 %
K (Pile)	58	56	96.6 %
L (Pile)	75	75	100 %
M (Pile)	79	81	97.5%
N (Pile)	~ 101	101	100 %
O (Pile)	181	175	96.7%

The least accurate counts were all from log pile such as those in the centre and left images in Figure 7. It seems that algorithm performs better when the edges of the log piles are well defined by the

sides of a bunk rather than those which formed part of a larger log stack.

The LEICA Scan Station scanner used in this study uses a whisker broom scanner; a mirror scans across and reflects light into a single detector which collects one data point at a time. In this application the disadvantage of this type of scanner is that the laser only hits the log ends perpendicular to the exact centre of the scan. Logs further away from the location of the scanner get scanned at increasingly acute angle, this leads to increased chance of shadowing that can affect the accuracy of the counts. Other units use “push broom scanning” which may be more suitable for this application and could lead to more accurate counts even on log piles that are not in bunks as the laser beam hits the log perpendicular more often (Figure 8).

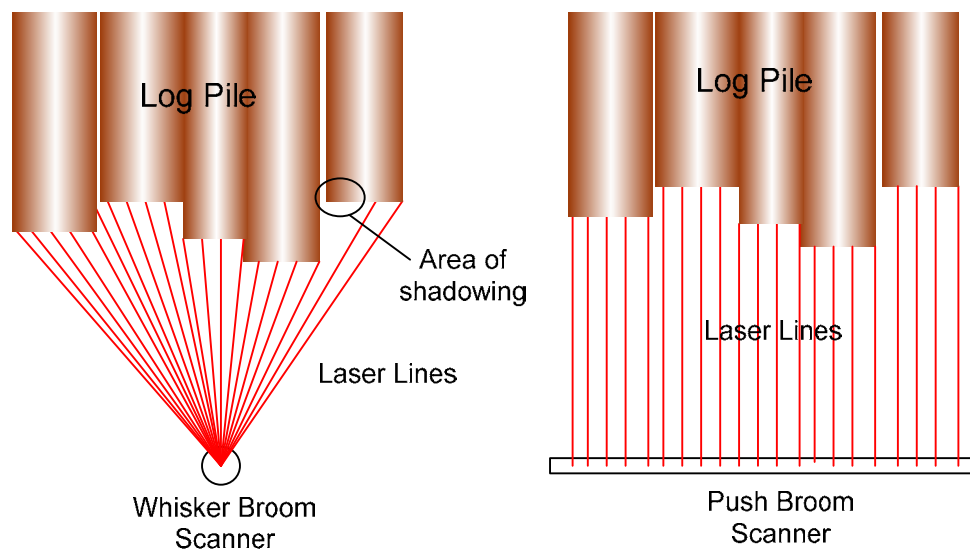


Figure 8. Different scanning options and their impact on counting accuracy.

3.1. Diameter Determination Validation Study

It should be noted that this algorithm was not developed to measure the diameter of the log. As part of the methodological approach of the algorithm the diameter of each log is estimated. However the algorithm is not optimised to accurately measure the diameter of the logs in a pile. In total 397 log diameters (measured vs predicted) were compared. The diameters in the validation set ranged from 165 mm to 760 mm.

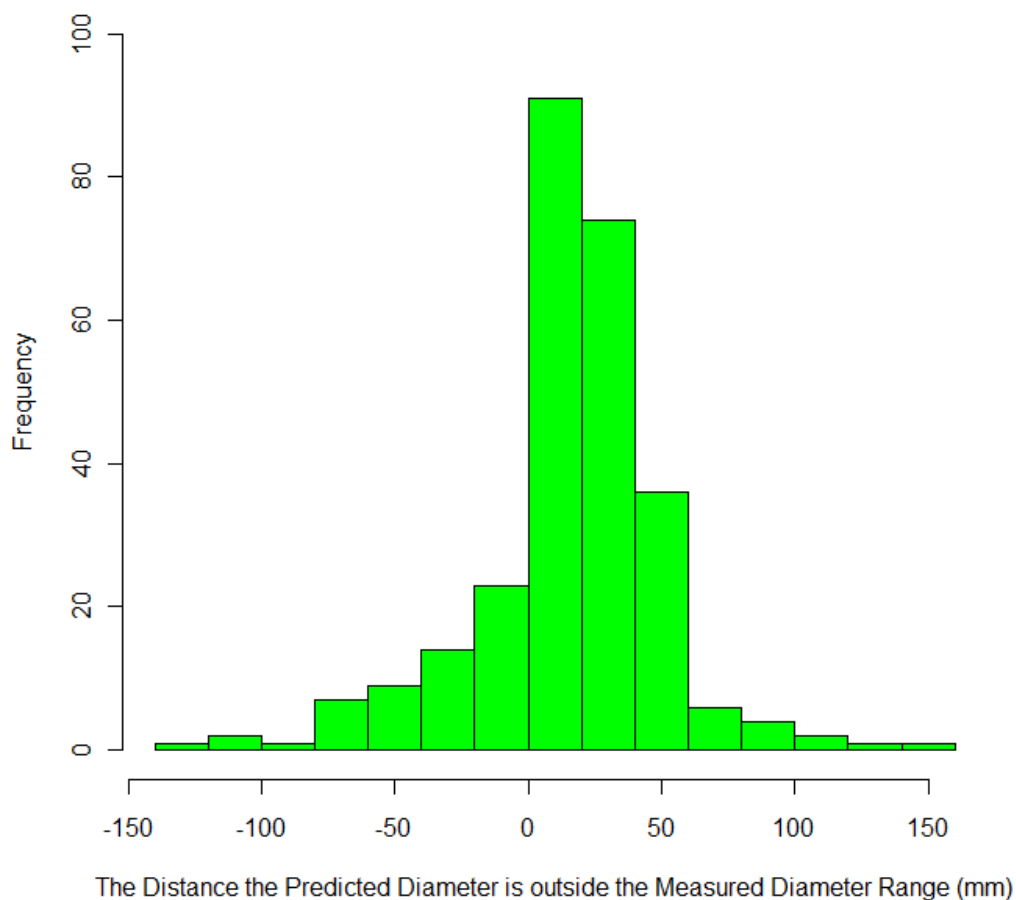


Figure 9. Distribution of the algorithm predicted diameters that are outside the measured diameter range.

Figure 9 shows the distribution of the residuals of the algorithm predicted diameters and the measured diameters. This graph does not include the 124 algorithm predicted diameters that were inside the manually measured diameter range. This means that 69% of the algorithm diameters were outside the manually measured range. In the majority of the cases the algorithm slightly over predicts the diameter as compared to manually measured diameter.

Paired t-tests were used to determine the measured and algorithm predicted diameters were significantly different. The tests showed that there was no statistical significant difference (p-value = 0.2258) between the algorithm predicted diameter and the maximum manually measured diameter (diameter 1). However there was a statistical significant difference (p-value < 0.000) between the algorithm predicted diameter and the minimum manually measured diameter (diameter 2).

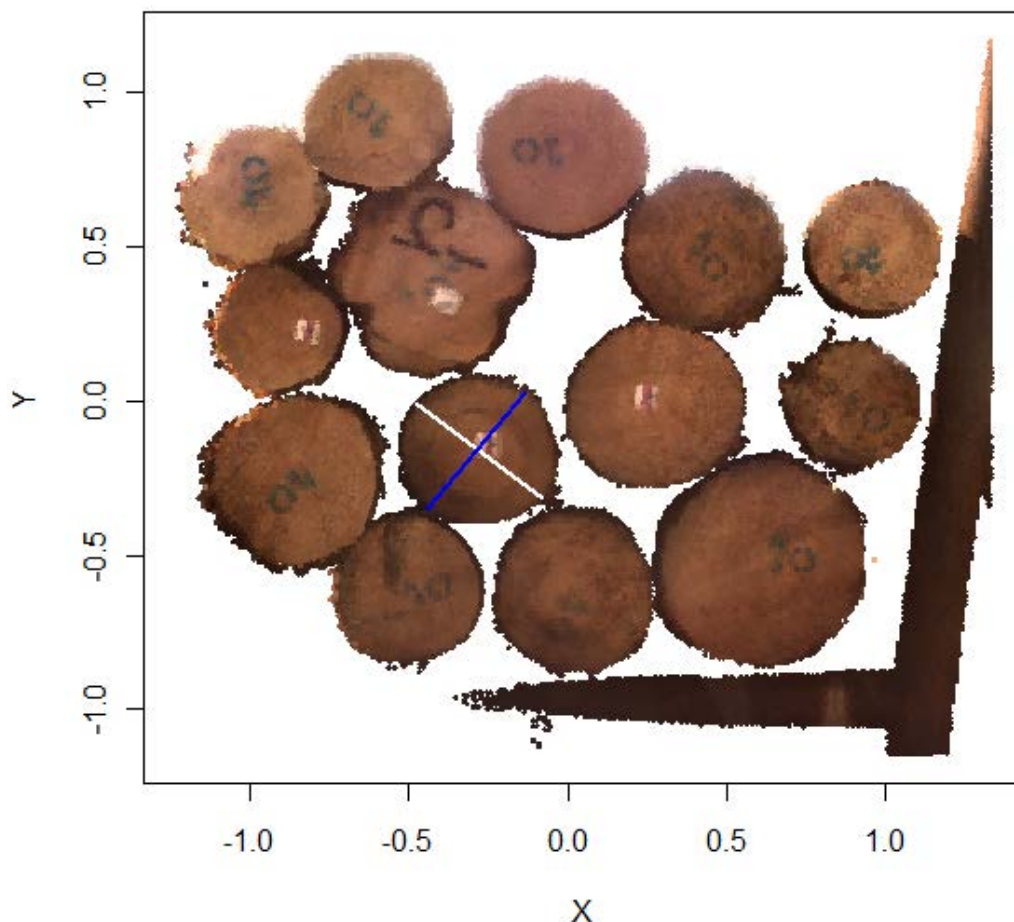


Figure 10. A typical image used to manually measure the diameters of the logs. The white and blue lines demonstrate how the diameters of log were measured from the image.

Both the largest diameter (diameter 1) as well as the diameter (diameter 2) at right angles to that largest diameter was measured on each log. These diameter measurements were all carried out by a technician with practical scaling experience. No scaling log diameter reductions were done to either the manual or algorithm measured diameters.

The irregular shape of most log ends mean that the exact diameter measurement of a log is always somewhat subjective. This makes comparing the accuracy of diameter measurements made by different measurement techniques difficult.

4. Discussion

The original idea for the development of this algorithm came from the need to accurately count logs in pile/stacks particularly at the port before export. Past attempts have been made in New Zealand including one that was used commercially in the early 1990s. A Danish company (Dralle Ltd) currently markets a system called sScale^(TM) used for both counting log in piles and measuring volume (www.dralle.dk). Dralle Ltd's website claims that the accuracy of their system in measure volume is within 2% at a cost of €0.5 per cubic metre (~\$(NZ) 1.04). Using standard photography has some advantages in term of capital cost of equipment but is

disadvantageous in terms of lighting requirements.

The goal of this project was to investigate and proof the concept that a 3d point cloud of x,y, and z co-ordinates could be used to count logs in a log stack. Ground-based LiDAR scans are being used in a number of industries, including mining and film making to collect 3d point clouds of objects. By using a 3d point cloud to count logs in pile/stack, issues around lighting are no longer problematic; in fact the images could be taken at night in complete darkness. The Logmeter developed in Chile uses LiDAR to scan a whole truck from a above to estimate diameter, length, volume and a range quality characteristics (<http://www.woodtechms.com> , Accessed on 3 July 2009).

This report outlines the methodology used to proof the concept that the number of logs in a pile can be counted using a 3d point cloud representation of the end of the pile of logs. A small validation study showed that the algorithm can routinely produce counting accuracy of greater than 96%. The accuracy improves to 100% if the logs are contained in a log bunk such as the image on the far right of Figure 7. It is likely further work on the algorithm such as implementing improved segmentation techniques could improve the counting accuracy. The target has to be to obtain 100 % accuracy 100 % of the time for this algorithm to be a commercial success.

The algorithm outlined in this study was not designed to measure the diameter of the logs that were being counted. However as part of the counting methodology an estimate of the diameter is obtained. Due to the irregular nature of the log diameter it is difficult to accurately qualify the accuracy of one diameter measurement against another measured using a different methodology. In this study 31 % of the algorithm measured logs were within manual measurement range (two manually measurement were made per log).

In an operational situation performance (time to count a pile of logs) of the algorithm will be important. The original prototype was developed in the R statistical language, porting it across to C# showed that significant improvements in the processing performance can be achieved. Further performance improvements are likely to be available by both improving the design of the algorithm as well as utilising programming technologies such as multi-threading and parallel processing.

The research covered in this report did not cover hardware. LiDAR scanners are expensive, the LIECA Scan Station costs approximately \$(US) 150,000 however it has numerous features that are not utilised in this application. There are other cheaper LiDAR options such as those manufactured by SICK (a Swedish company) that cost around \$(US) 10,000. 3D point cloud data of objects can be generated from stereo photography, which simply requires two high quality cameras such as used by Dralle Ltd in the sScale^(TM) product.

A report on the commercialisation of log counting technology was produced by Seltec Advisory Limited (Anderson 2009). They reported discussions with a number of potential customers in the market. This report found that simply carrying out log counts on piles/stacks would not be enough to create a commercially successful product. It seems that the original demand for automated log counting disappeared as export demand has shifted from smaller K grade to larger A grade products, meaning that fewer logs need to be counted per cubic metre of volume. Anderson (2009) indicated that if the concept could be extended to include automatic volumetric scaling to export and domestic scaling rules then there may be a market for such technologies. To turn the current algorithm into a commercially viable product would require additional development work and due to the qualitative nature of some scaling rules that development work would be non-trivial.

5. Conclusion

This report outlines a proof of concept for a log counting algorithm using 3D point cloud data generated from a ground based LiDAR scan of the end of a log stack. The accuracy of log counts is upwards of 96% for the validation data sets. Although not originally designed to measure log diameter, a validation study carried as part of this project showed that the algorithm described in this report could be improved to accurately measure the diameter of all the logs in a pile/stack.

To be fully commercialised this algorithm would need further work to improve its accuracy both of log counting and log diameter measurements. There are numerous image processing techniques as well as computer programming techniques that could be investigated to improve overall performance. Any additional work on this project should focus on developing automated methods for identifying defects that affect the scaled volume of logs. Additional research into the optimal hardware to capture the 3D point cloud would also need to be undertaken before a cost effective tool could be released.

A business case carried out by Seltec Advisory Limited on the market size for log stack scanning and automated log counting concluded that the market is relatively small. The study did identify some niche markets that would benefit from the concept of log counting. This market would dramatically increase if the algorithm could automatically scale logs based on New Zealand's scaling rules however from algorithm development point of way this is a non-trivial matter.

From a research perspective, this project has highlighted the potential of using ground based LiDAR scanning technology and 3d point cloud data for determining log characteristics. The lessons learnt and techniques developed as part of this will hopefully be utilised in further research and development into this or other applications.

6. Acknowledgements

Future Forest Research for providing the funding for this project.
Bruce Robinson (Geosurvey Limited) for assisting in the scanning.
James Anderson (Seltec Advisory Limited) for providing commercialisation insights.

7. References

Anderson J. 2009. Log Stack Scanning: Business Case. Prepared for Scion and Future Forests Research. Seltec Advisory Limited. Confidential Report. Seltec Advisory Limited.

LogMeter⁴⁰⁰⁰ <http://www.woodtechms.com/logmeter1.php>, Accessed on 3 July 2009.

sScale^(TM) www.dralle.dk, Accessed on 3 July 2009.

Generating an automated approach to optimize effective leaf area index by Canadian boreal forest species using airborne LiDAR

Heather Morrison^{1,2}, Chris Hopkinson^{2,1}, Laura Chasmer³ & Natascha Kljun⁴

¹Acadia University, Wolfville, Nova Scotia, Canada morrison.h@gmail.com

²Applied Geomatics Research Group, NSCC Annapolis Campus, Nova Scotia, Canada

³Cold Regions Research Centre, Wilfrid Laurier University, Waterloo, Ontario, Canada

⁴Department of Geography, Swansea University, Swansea, Wales, United Kingdom

Abstract

Obtaining forest structure data to compute leaf area index (LAI) can be a challenge in remote areas like the Canadian boreal forest. Light ranging and detection (LiDAR) data provides a 3-dimensional view of the forest that can be calibrated with minimal field data requirements relative to other remote sensing data. Our objective is to develop an automated method for combining a limited amount of field data with LiDAR to generate estimates of LAI. To accomplish this we used geographic information system (GIS) tools to expand upon a physically-based gap fraction model by incorporating a process for optimizing extinction coefficient by forest species. In this paper we demonstrate a simple, efficient method for optimizing remote sensing-based estimates of canopy attributes from limited field data. We were able to reduce the RMSE in modelled effective leaf area index by an average of 0.48 across all species. Combining such simple model optimisation approaches with other automated LiDAR-based canopy attribute extraction procedures shows promise as we move towards ever greater levels of LiDAR forestry operationalisation.

Keywords: LiDAR, leaf area index, optimization, extinction coefficient, Boreal forest

1. Introduction

1.1 Rational

Leaf area index (LAI), which is defined as half of the total leaf area per unit ground area (Chen *et al.*, 2006), is an important input parameter used within biogeochemical, biomass, and ecological models. Accurate estimates of LAI are therefore important, as small deviations or biases in could result in sometimes compounded errors within these models. Several studies have used plot-based measurements of gap fraction (used to derive effective LAI (LAI_e) and LAI) when scaling to lower resolution spectral imagery (e.g. Fernandez *et al.*, 2003; Fernandez *et al.*, 2004). However, plot measurements often do not represent the full range of vegetation characteristics found within ecosystems, and can be time consuming to acquire. Airborne Light Detection and Ranging (LiDAR) data offers an alternative method for continuously mapping LAI at high resolution. LiDAR provides a three dimensional representation of the canopy, understory, and ground surface topography measured using reflected laser pulses. The basic rationale for LiDAR-based LAI mapping is that the vertical distribution of laser pulse returns within the canopy is related to the foliage profile (Magnussen and Boudewyn, 1998) such that if only ground-level returns occur in a given area then the likelihood of overlaying leaf area is low. Conversely, a greater density of above ground (or canopy level) returns in a given area indicates a higher leaf area. From this basic understanding,

LAI can be estimated directly as a function of the canopy gaps thus observed (e.g. Solberg *et al.*, 2006). However, gap fraction-based estimates of LAIe (which must further take into account the canopy clumping, woody to total leaf area ratio, and needle to shoot area ratio in order to estimate true LAI) requires an estimate of extinction coefficient (k). The objective of this study is to investigate model parameter optimization of k to improve LAIe estimates within three boreal forest ecosystems: mature black spruce, a jack pine chronosequence of four sites, and a mature aspen stand.

1.2 LiDAR-based LAI models

A number of LiDAR-based LAIe models have been developed that employ range and echo data provided by discrete-return airborne systems. These include mean return elevation methods (e.g. Lim *et al.*, 2003), fractional canopy return methods (e.g. Riaño *et al.*, 2004; Solberg *et al.*, 2006), and the examination of canopy volume (e.g. Lefsky *et al.*, 1999). Models were developed and tested for a specific forest type but often require calibration. The intensity-based gap fraction (or fractional cover) model of Hopkinson and Chasmer (2007), is one LiDAR-based model that has been shown to require minimal or no calibration. The model divides LiDAR returns into four echo classes (first, single, intermediate, last) and generates grids of intensity by summing returns within a cell. It then accounts for a two-way power transmission loss by intermediate and last return hits using a square root function. First and single hits at and below 1.3 m from the ground surface are subset to represent below-canopy (ground) hits. A ratio of total returns intensity to this below-canopy subset is used to estimate gap fraction:

$$P = \frac{\left(\frac{\sum I_{GroundSingle}}{\sum I_{Total}}\right) + \sqrt{\frac{\sum I_{GroundLast}}{\sum I_{Total}}}}{\left(\frac{\sum I_{First} + \sum I_{Single}}{\sum I_{Total}}\right) + \sqrt{\frac{\sum I_{Intermediate} + \sum I_{Last}}{\sum I_{Total}}}} \quad (1)$$

Subscripts indicate the echo class and subset of each return. This model has been tested on a variety of study locations across Canada resulting in estimates comparable to ground based (DHP) measures of gap fraction (or fractional cover) without requiring calibration (Hopkinson and Chasmer, 2009). This study uses the intensity-based model of Hopkinson and Chasmer (2007) and an automated plot-based optimisation routine to create a more accurate model of LAIe that can be applied to a broad range of boreal forest types.

2. Study Area

The study area is located in the Boreal forest of Saskatchewan, Canada (Fig. 1) on a number of sites being monitored as part of Fluxnet-Canada (2002-2007) and the Canadian Carbon Program (2007-2011) networks. A variety of stand types were sampled including a three stage chronosequence of jack pine (mature ~95 years old, harvested in 1975, harvested in 1994); a mature aspen stand and a mature black spruce stand (Table 1). The total number of plots examined within each stand type were randomly divided into training and testing categories for modeling and validation.



Figure. 1. Map showing location of study area within Canada

Table 1. Forest plot descriptions and stand type

Stand	Description	LAIe/DHP	
		Training Plots	Testing Plots
JP	All Jack Pine Sites	75	56
OJP	Old Jack Pine	25	27
HJP75	Jack Pine harvested in 1975	25	17
HJP94	Jack Pine harvested in 1994	25	12
OBS	Old Black Spruce	20	8
OA	Old Aspen	20	11

3. Methods

3.1 DHP collection and analysis

Ground-truth data were collected August 10-20, 2005 and July 29-August 3, 2008. Five digital hemispheric photos (DHP) were collected per geographically located plot (dGPS), one at the center and four located 11.3 m from the center in each cardinal direction (N, E, S, W) using a compass bearing and tape measure. All images were captured using a Nikon Coolpix 8.0 Megapixel camera positioned 1.3 m off the ground (at mature sites, 0.5 m at HJP94), facing north, fitted with a 180° fisheye lens with the exposure set one ‘f stop’ lower than normal exposure to improve contrast between foliage and sky. DHPs were processed using CAN_EYE software (http://www.avignon.inra.fr/can_eye/) which utilizes user enhanced automated image classification to calculate gap fraction and LAIe from two-tone images.

3.2 LiDAR data collection and preparation

LiDAR data were collected by the Applied Geomatics Research Group (AGRG) coincident with DHP collection on August 12, 2005 and August 2, 2008 using an ALTM 3100 laser scanner. The 2005 LiDAR data collection was flown at a height of 950 m a.g.l., with a laser pulse repetition frequency (PRF) of 70 kHz, and a scan angle of $\pm 19^\circ$ (with 50% overlap of scan lines). The 2008 LiDAR data collection was flown using the same sensor at a height of 700 m a.g.l., with a PRF of 70 kHz and a scan angle of $\pm 20^\circ$ (with 50% overlap of scan lines). The point data were classified using *Terrascan* (Terrasolid, Finland) into ground, canopy and echo code classes then gridded using *Surfer 8* (Golden Software Inc., USA) by assigning summed intensity values to each cell based on points that fell within 2.5 m to generate 1 m resolution grids. Classification and gridding was also performed by the AGRG in preparation for modeling (Hopkinson and Chasmer, 2009).

3.3 Optimization process

Gap fraction (P) grids were calculated using the intensity-based model published by Hopkinson and Chasmer (2007) described above. If the canopy is assumed to be a turbid medium with randomly distributed foliage then the Beer-Lambert Law can be applied:

$$LAIe = -\ln(P) / k \quad (2)$$

where extinction coefficient (k) is a function of leaf angle distribution, radiation type and direction, and canopy structure and clumping (Bréda, 2003). Initially, a mid-value k of 0.5 is used in this study because it represents a spherical (random) projection coefficient for leaves of any shape, (Chen *et al.*, 1997) and is an accepted alternative to species specific values (Richardson *et al.*, 2009).

$$LAIe_{LiDAR} = -\ln(P) / 0.5 \quad (3)$$

The k term in equation (3) is then optimized for each species by rearranging the general equation (2) using measurements of LAIe from captured DHPs ($LAIe_{DHP}$), to train new estimates for k based on species (k_{NEW}):

$$k_{NEW} = LAIe_{LiDAR} / (2 * LAIe_{DHP}) \quad (4)$$

LAIe raster layers were generated for the entire study area by equation (3) using an automated GIS-based tool as a baseline for optimization. Mean LAIe values were extracted for 11.3 m radius plots at the geo-located photo positions and the training subset were compared to coincident LAIe measured using DHPs to generate k_{NEW} for each species (4). LAIe raster layers were then regenerated using the P layers and substituting k_{NEW} for 0.5 in equation (3). Model quality was determined using the testing subset of plots for each species (Table 1).

4. Results

LAIe estimated using a generic 0.5 extinction coefficient (equation 3) resulted in means that were significantly different from DHP LAIe ($p < 0.05$) across all species. The generic model underestimated LAIe for both conifer species while overestimating the broad-leaved aspen compared with measurements gathered in the field (Fig. 2). These results are comparable to those published by Bréda (2002) who indicated that coniferous stands trended towards extinction coefficients less than 0.5. The need for a more specific k for predicting LAIe from LiDAR is also highlighted by these results.

Including k_{NEW} improved LAIe model fit, reducing the RMSE by an average of 0.48 across all species (Fig. 2, Table 2). The greatest improvement was observed in the OBS model which translates to a shift in average LAIe values from 1.17 to 2.42. The lowest RMSE occurred across the JP stands at 0.35 after optimization. The average absolute shift in LAIe across all species is 0.79, which is greater than the difference observed between coniferous and deciduous species, signifying a change in canopy structure.

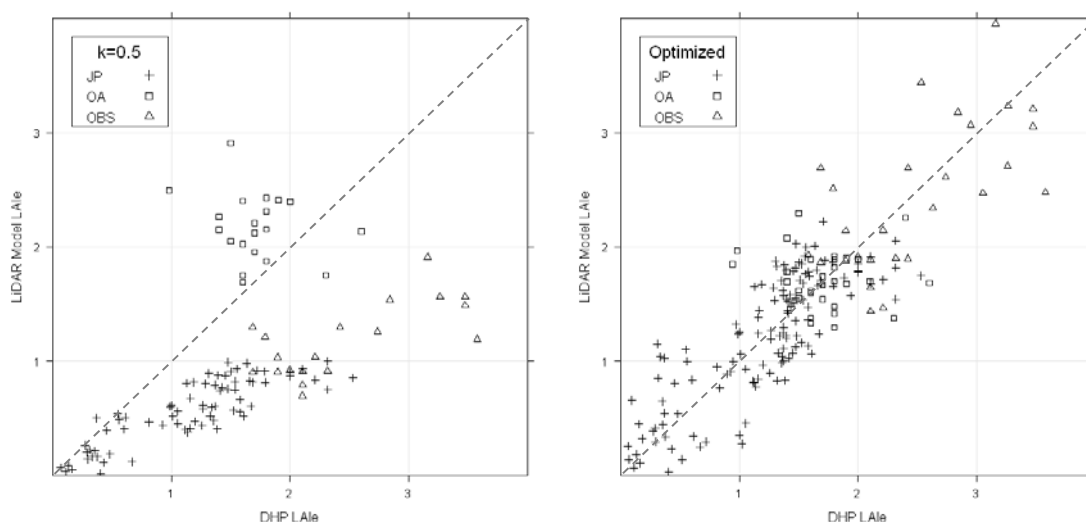


Figure 2. LAI_{LiDAR} compared to LAI_{DHP} for all data (n=190). $k=0.5$ represents pre-optimization results (left) and optimized (right) represents estimates using k_{NEW} . Dashed line is 1:1.

LAI_e for plots set aside for testing calculated using k_{NEW} revealed no significant difference at the 95% confidence level between LiDAR modelled and DHP measured LAI_e for all species, indicating a significant improvement over the generic model.

Table 2. Pre and Post-optimization statistics

Stand	Pre-optimized	Post-optimized	k_{NEW}	RMSE	RMSE	RMSE
	mean	mean		pre-optimized	post-optimized	post-optimized
	LAI _{eLiDAR}	LAI _{eLiDAR}		Training plots	Training plots	Testing plots
JP	0.57	1.17	0.24	0.71	0.35	0.32
OBS	1.17	2.42	0.24	1.34	0.49	0.56
OA	2.17	1.71	0.63	0.67	0.44	0.42

5. Conclusion

An automated optimization model such as the one presented here creates opportunities to gain knowledge of forest structure over large areas using limited field data. The adjustment of k when modelling LAI_e from LiDAR intensity data will improve results that will be reflected in environmental applications based on remote sensing data. Further investigation including more species and age classes would benefit from model optimization of this type including investigating optimal intensity gridding parameters (Morrison *et al.*, 2011). This work is part of a larger effort to operationalize forest structure modelling routines through the generation of automated tools.

Acknowledgements

Dr. Harry McCaughey, Queen’s University and the Canadian Carbon Program and Fluxnet-Canada for subsidizing LiDAR data collection and field work in 2005, Chris Beasy and James Churchill for assistance with DHP data collection in 2005 and 2008. This project was also partially funded by a UK Natural Environment Research Council grant (NE/G000360/1) awarded to Dr. Natascha Kljun at the University of Swansea and Trevor Milne provided guidance with Python coding. The Canada

Foundation of Innovation is acknowledged for funding Dr. Hopkinson's LiDAR laboratory at the AGRG.

References

- Bréda, N., Soudani, K., and Bergonzini, J.C., 2002. *Mesure de l'indice foliaire en forêt*. Paris: ECOFOR.
- Bréda, N.J.J., 2003. Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. *Journal of Experimental Botany*, 54(392), 2403-2417.
- Chen, J.M., Rich, P.M., Gower, S.T., Norman, J.M., and Plummer, S., 1997. Leaf area index of boreal forests: Theory, techniques, and measurements. *Journal of Geophysical Research*, 102(D24), 29,429-29,443.
- Chen, J.M., Govind, A., Sonnetag, O., Zhang, Y., Barr, A., and Amiro, B., 2006. Leaf area index measurements at Fluxnet-Canada forest sites. *Agricultural and Forest Meteorology*, 140, 257-268.
- Fernandez, R., Butson, C., Leblanc, S., and Latifovic, R., 2003. Landsat 5 TM and Landsat 7 ETM+ based accuracy assessment of leaf area index products for Canada derived from Spot-4 VEGETATION data. *Canadian Journal of Remote Sensing*, 29(2), 241-258.
- Fernandez, R., Miller, J.R., Chen, J.M., and Rubinstein, I., 2004. Evaluating image-based estimates of leaf area index in boreal conifer stands over a range of scales using high resolution CASI imagery. *Remote Sensing of Environment*, 89, 200-216.
- Hopkinson, C. and Chasmer, L., 2007. Modelling canopy gap fraction from LiDAR intensity. *ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*. Espoo, Finland: 190-194.
- Hopkinson, C. and Chasmer, L., 2009. Testing LiDAR models of fractional cover across multiple forest ecozones. *Remote Sensing of the Environment*, 113, 275-288.
- Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A., and Harding, D., 1999. LiDAR remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment*, 70(3), 3399-361.
- Lim K., Treitz, P., Baldwin, K., Morrison, I., and Green, J., 2003. LiDAR remote sensing of biophysical properties of northern hardwood forests. *Canadian Journal of Remote Sensing*, 29(5), 658-678.
- Magnussen, S. and Boudewyn, P.A., 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research*, 28, 1016-1031.
- Morrison, H., Hopkinson, C., Wulder, M.A., 2011. Optimal LiDAR gridding parameterization for effective leaf area estimation in the boreal forest Yukon Territory, Canada. *Proceedings of the SilviLaser 2011 Conference*, Oct. 16-20, Hobart, Tasmania
- Riaño, D., Valladares, F., Condes, S., and Chuvieco, E., 2004. Estimation of leaf area index and covered ground from airborne laser scanner (LiDAR) in two contrasting forests. *Agricultural and Forest Meteorology*, 124(3-4), 269-275.
- Richardson, J.J., Moskal, L.M., and Kim, S-H., 2009. Modelling approaches to estimate effective leaf area index from areal discrete-return LiDAR. *Agricultural and Forest Meteorology*, 149, 1152-1160.
- Solberg, S., Næsset, E., Hanssen, K.H., and Christiansen, E., 2006. Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sensing of Environment*, 102(3-4), 364-376.

Developing a Regional Canopy Fuels Assessment Strategy using Multi-Scale Lidar

Birgit Peterson¹ & Kurtis Nelson²

¹ ASRC Research and Technology Solutions contractor to the U.S. Geological Survey (USGS), Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD 57198, bpeterson@usgs.gov

² USGS, EROS, Sioux Falls, SD 57198, knelson@usgs.gov

Abstract

Accurate assessments of canopy fuels are needed by fire scientists to understand fire behavior and to predict future fire occurrence. A key descriptor for canopy fuels is canopy bulk density (CBD). CBD is closely linked to the structure of the canopy; therefore, lidar measurements are particularly well suited to assessments of CBD. LANDFIRE scientists are exploring methods to integrate airborne and spaceborne lidar datasets into a national mapping effort. In this study, airborne lidar, spaceborne lidar, and field data are used to map CBD in the Yukon Flats Ecoregion, with the airborne lidar serving as a bridge between the field data and the spaceborne observations. The field-based CBD was positively correlated with airborne lidar observations ($R^2 = 0.78$). Mapped values of CBD using the airborne lidar dataset were significantly correlated with spaceborne lidar observations when analyzed by forest type ($R^2 = 0.62$, evergreen and $R^2 = 0.71$, mixed). Though continued research is necessary to validate these results, they do support the feasibility of airborne and, most importantly, spaceborne lidar data for canopy fuels assessment.

Keywords: GLAS, FlamMap, canopy bulk density, fuels mapping, fire behavior modeling

1. Introduction

Wildland fire can have a significant impact on ecosystems and human populations. To better understand wildland fire occurrence and fire behavior, it is critical to have a thorough understanding of the distribution of fuels in the landscape. Fire behavior models are driven by inputs that describe the three-dimensional distribution of fuels. Canopy fuels are defined as the live and dead biomass located within tree crowns (Keane *et al.* 2001) and are characterized in part by canopy bulk density (CBD), which is the mass of available canopy fuel per unit canopy volume. The CBD parameter is used by fire modeling systems, such as FlamMap (Finney 2006), and is a key determinant of whether fire will spread from crown to crown as an active canopy fire.

The LANDFIRE project (Rollins 2009; <http://www.landfire.gov>) has mapped fire behavior model inputs, including CBD, for the entire United States using a set of field data, Landsat imagery, digital elevation models (DEMs) and derivatives, climate gradients, and other ancillary data. The first set of LANDFIRE products was completed in 2009. LANDFIRE is developing a strategy to regularly provide improved and updated data. The updating plan includes the exploration and integration of lidar data for characterizing canopy structure and fuels to address a demonstrated need for better estimates of LANDFIRE fuels products (Krasnow *et al.* 2009).

Traditionally, fuel parameters are obtained through field observations of vegetation structure. More recently, lidar has also been used to quantify vegetation canopy structure (Dubayah and Drake 2000; Lefsky *et al.* 2002), which is particularly useful for mapping the fuels complex

(Riaño *et al.* 2004; Morsdorf *et al.* 2004; Andersen *et al.* 2005; Peterson *et al.* 2007; Skowronski *et al.* 2007). While previous studies using lidar to assess fuels have used airborne lidar data, recent work has shown the applicability of the spaceborne Geoscience Laser Altimeter System (GLAS) data for estimating forest canopy structure (Lefsky *et al.* 2007; Sun *et al.* 2008; Nelson *et al.* 2009). Because canopy fuels are defined largely by canopy structure, we wish to explore the utility of GLAS data for estimating CBD, which would promote more accurate mapping of CBD, especially where field data are scarce. The incorporation of GLAS data into a regional fuels mapping effort, while leveraging airborne data acquisitions, will provide detailed vegetation structure metrics over larger areas to inform fuels mapping.

2. Objectives

The objective of this study was to explore how integrating lidar data collected at different scales can contribute to a regional-to-national scale canopy fuels mapping effort. This investigation was conducted in the Yukon Flats Ecoregion (YFE) of interior Alaska (Gallant *et al.* 1995). We used two lidar data sets: an airborne, small-footprint, discrete-return dataset collected for a subset of the study area; and transects of GLAS data for the entire YFE. Using these two datasets, along with field and ancillary data, we modeled and mapped CBD then used the different CBD products as inputs into FlamMap.

3. Methods

3.1 Study Area

The YFE (Figure 1) is located in interior Alaska and encompasses approximately 33,400 km². Much of the vegetation is short-statured boreal forest. Common tree species are black spruce (*Picea mariana*), white spruce (*Picea glauca*), quaking aspen (*Populus tremuloides*), alder (*Alnus* sp.), and willow (*Salix* sp.). Forest stands rarely exceed 20 m except in riparian areas and floodplains. The terrain of the YFE is flat, with 95% of the area at $\leq 3^\circ$ slope as indicated by the National Elevation Dataset (NED). Access to the area is most practically achieved by float plane.

3.2 Field data

Vegetation sampling was conducted in 2010 in a 2-m-wide transect running north-south through the plot center. Within this transect, we recorded diameter at breast height (DBH), tree height and height to the base of the live crown, and species. We sampled 24 plots at two lake sites. We then generated estimates of CBD for each plot using the tree list data and FuelCalc (Reinhardt *et al.* 2006). FuelCalc does not calculate CBD for broadleaf species because canopy fire propagation is rare in those species. Six of our plots were in pure willow, alder, or aspen stands, so CBD was not calculated for those sites.

3.3 Airborne Lidar

The airborne lidar data were collected in summer 2009 for a sub-area (2605 km²) of the YFE (Figure 1). The data were collected by Aero-Metric, Inc. with an airborne Optec ALTM Gemini. The dataset has a horizontal accuracy of 1.15 m, with a nominal point spacing of 2.3 m and a vertical positional accuracy of 0.10 m. The raw data were processed and delivered as 2.5-m resolution raster datasets that include the bare-Earth digital surface model (DSM) and first-return DSM. Height above ground (HAG) was derived by differencing the bare-Earth DSM and the first-return DSM.

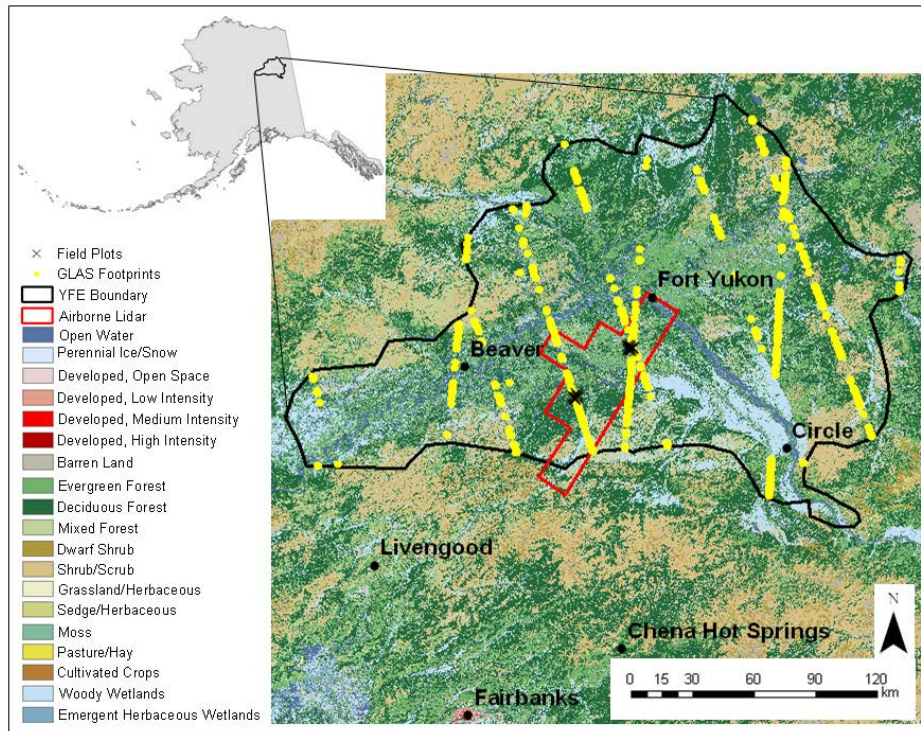


Figure 1: Map showing location of YFE study area, airborne lidar acquisition boundary, GLAS footprint locations, and field plot locations. The NLCD land cover map is shown in the background.

Various metrics were derived from the airborne lidar dataset to predict CBD. The maximum, mean, and minimum HAG were calculated for each plot. All other airborne lidar metrics were derived from a 10×10 m grid of the point cloud data using the vertical distribution of return elevations within the grid cell. For each 10×10 m grid cell at six different height thresholds (>1 m, >2 m, >3 m, >4 m, >5 m, and >6 m), the ratio of canopy returns above the threshold to the total number of canopy returns was calculated. The maximum, minimum, and mean of these ratios were calculated for each plot, which described the vertical distribution of canopy material between and the lower, middle, and upper portions of the canopy.

We used a stepwise linear regression procedure in *R* (<http://www.r-project.org/>) to identify models for estimating CBD. Our goal was to identify a parsimonious yet predictive regression model. The final model was then applied to the entire set of airborne lidar data where the National Land Cover Database (NLCD; Selkowitz and Stehman 2011) indicated evergreen or mixed forest. Focal statistics (maximum, minimum, and mean) were generated from the HAG layer and 10×10 m ratio data. The resulting map was gridded to 30 m to match the spatial resolution of LANDFIRE products. Also in accordance with LANDFIRE methods, the CBD for hardwood forests was set to 0.01 kg m^{-3} . For all other NLCD classes, the CBD was set to 0.

3.4 GLAS

We obtained GLA01 (waveform data) and GLA14 (land/canopy elevation data and footprint locations) GLAS products (<http://nsidc.org/data/icesat/index.html>) from the Laser L3F acquisition (release 31) for the entire YFE. Included in the GLA14 product is a set of metrics describing Gaussian curves fit to the waveform, including number of peaks, elevation, width, and amplitude of each Gaussian (Harding and Carabajal 2005). L3F acquisitions occurred during leaf-on conditions (May 24 through June 26, 2006). The GLAS footprints are nominally 65 m in diameter and 172 m apart along-transect.

The GLA01 and GLA14 products were processed to derive quartile and decile heights of waveform energy following Sun *et al.* (2008), as well as canopy depth, total waveform energy (Peterson *et al.* 2007), and canopy height, which is the difference between the ground elevation and the elevation of the signal beginning. We used these derived metrics plus the GLA14 Gaussian metrics as independent variables in a regression analysis.

We filtered the GLAS returns based on four criteria: we eliminated footprints 1) which were located on slopes $> 3^\circ$ as indicated by NED; 2) where the NLCD land cover was not uniform; 3) which were cloud contaminated according to GLA14; and 4) with waveforms for which only a single Gaussian was identified assuming that no distinct canopy could be inferred. This filtering resulted in 718 usable footprints for the YFE with 115 falling in the airborne lidar subset area.

None of the GLAS footprints were coincident with the field plot locations. Therefore, we used the CBD values from the airborne lidar-based map as our dependent variable in the regression. For each of the 115 GLAS footprints, we extracted the CBD value at the footprint center. In R a stepwise linear regression procedure was used to identify a model for estimating CBD. This model was applied to the GLAS footprints in the YFE to estimate CBD for those locations.

Because the GLAS data are sampled at discrete locations, we used a regression-tree approach for mapping CBD from GLAS. We used seven Landsat Thematic Mapper spectral bands and the DEM, slope, aspect, and NLCD land cover as independent variables. These values were extracted at each footprint center. We used Cubist (<http://www.rulequest.com>) to develop the regression tree and a spatial applier to map CBD for the YFE. NLCD land cover was used to assign CBD values to hardwood forests and non-forested pixels.

3.5 FlamMap Fire Behavior Modeling

To assess the impact of mapping CBD using airborne lidar and GLAS data, the CBD maps were used to conduct fire behavior analyses. Analyses were completed at two scales, the entire YFE and the sub-area covered by the airborne lidar data. FlamMap was used with input layers from LANDFIRE and the lidar-derived CBD maps. In the sub-area, a baseline model run was completed using the LANDFIRE CBD layer. The airborne- and GLAS-derived CBD layers were each then substituted for the LANDFIRE CBD, and FlamMap was re-run keeping all other data and settings the same. In the full YFE run, the baseline LANDFIRE data were again used as a baseline, and the CBD layer was then replaced with the GLAS-derived CBD layer, keeping all other data and settings constant. Thus, a total of five FlamMap runs were made.

Each FlamMap run produced flame length (FL), rate of spread (ROS), and crown fire activity (CFA) output layers. FL and ROS are continuous outputs measured in meters and meters per minute, respectively. The CFA layer separates the landscape into unburned (water, barren, etc.), surface fire only, passive crown fire (individual tree torching), and active crown fire (fire spreading from tree to tree) classes. The CBD layer is used only for calculating crown fire properties so that the amount of unburned and surface fire areas are consistent between runs, only the amount of active versus passive crown fire varies in the CFA outputs. The amount of active and passive crown fire for each run is summarized, as are descriptive statistics of the FL and ROS outputs for each model run.

4. Results

4.1 Airborne Lidar

The final model identified through stepwise regression analysis for estimating CBD from the airborne lidar is shown in Equation 1.

$$\text{CBD} = 0.11 + 0.05 * \text{hagsd} + 1.19 * \text{meanrat1} - 0.93 * \text{meanrat2} - 0.26 * \text{maxrat2} - 0.26 * \text{maxrat4} + 0.26 * \text{maxrat5}, \quad (1)$$

where hagsd is the standard deviation of the HAG; meanrat1 and meanrat2 are the mean ratios at 1 and 2 m above ground, respectively; and maxrat2, maxrat4, and maxrat5 are the maximum ratios at 2, 4, and 5 m, respectively. This model had a coefficient of determination (R^2) of 0.78, an adjusted R^2 of 0.67, and a residual standard error (RSE) of 0.05 kg m^{-3} and was used to generate the airborne lidar-derived map of CBD.

4.2 GLAS

The model generated using all 115 GLAS footprints in the YFE sub-area resulted in a weak relationship, with an R^2 of 0.30. To develop a model with more predictive power, we split the 115 GLAS footprints into two sets by NLCD forest class: evergreen forest or mixed evergreen/deciduous forest. We also eliminated footprints where the standard deviation of the mapped CBD value within a 60 m radius of a footprint center was $> 0.05 \text{ kg m}^{-3}$.

The final regression model for estimating CBD from GLAS data for footprints falling within evergreen stands ($N = 30$) is shown in Equation 2.

$$\text{CBD} = 0.156 - 0.003 * \text{r90} - 0.079 * \text{r40} + 0.063 * \text{r30} + 0.068 * \text{r20} - 0.053 * \text{r10} + 0.00002 * \text{tenergy} - 0.298 * \text{numpeak} + 0.009 * \text{height}, \quad (2)$$

where r10, r20, r30, r40, and r90 are the decile heights at 10, 20, 30, 40, and 90% of waveform energy, respectively, tenergy is the total waveform energy, numpeak is the number of Gaussian peaks, and height is the canopy height. This model had an R^2 of 0.61, an adjusted R^2 of 0.46 and an RSE of 0.03 kg m^{-3} and was used to generate CBD values for the GLAS footprints in evergreen forest.

The final regression model for estimating CBD from GLAS waveform data for footprints falling within mixed stands ($N = 29$) is shown in Equation 3.

$$\text{CBD} = 0.161 - 0.012 * \text{r90} - 0.028 * \text{r70} + 0.051 * \text{r40} + -0.138 * \text{r25} + 0.212 * \text{r20} - 0.091 * \text{r10} + 0.023 * \text{depth} + 0.104 * \text{numpeak}, \quad (3)$$

where r25 and r70 are the 25th and 70th percentile heights of waveform energy, respectively, and depth is the canopy depth. This model had an R^2 of 0.80, an adjusted R^2 of 0.72, and an RSE of 0.06 and was used to generate CBD values for the GLAS footprints in mixed forest stands.

The Cubist regression-tree modeling initially resulted in a correlation coefficient (R) of 0.31. This was caused by an underrepresentation of the tails of the CBD distribution. To improve the predictive power of the model, we adjusted the sample to ensure an even distribution of CBD values. This resulted in a stronger model with an R of 0.90, and predicted values that represented the entire range of CBD values in the training data.

4.3 FlamMap

In the study sub-area, there were three runs using LANDFIRE, airborne lidar-derived, and GLAS-derived CBD. There were 61923 ha of surface fire and 77948 ha of crown fire. Using the LANDFIRE CBD, 22% of the crown fire was active, whereas when using airborne lidar- and GLAS-derived CBD layers, 74% and 79% of the crown fire was active, respectively. The

ROS using the LANDFIRE CBD ranged from 1 to 47 m/min with a mean of 21 m/min. When using both airborne lidar- and GLAS-derived CBD, the ROS ranged from 1 to 49 m/min with a mean of 23 m/min. FL ranged from 1 to 26 m with a mean of 11 m using the LANDFIRE CBD. With the airborne lidar-derived CBD, the FL ranged from 1 to 46 m with a mean of 15 m. The FL from the GLAS-derived CBD ranged from 1 to 38 m with a mean of 14 m. Both the increased FL and ROS using the lidar-derived CBD reflect the increase in active crown fire modeled from these data.

For the full YFE runs, there were 1,774,107 ha of surface fire and 1,321,264 ha of crown fire. Using the LANDFIRE CBD, 23% of the crown fire was active, whereas 64% of the crown fire was modeled as active using the GLAS-derived CBD layer (Figure 2). The FL using the LANDFIRE CBD ranged from 1 to 21 m with a mean of 9 m, compared to a range of 1–34 m with a mean of 10 m using the GLAS-derived CBD. The ROS using both CBD maps ranged from 1 to 39 m/min with a mean of 16 m/min.

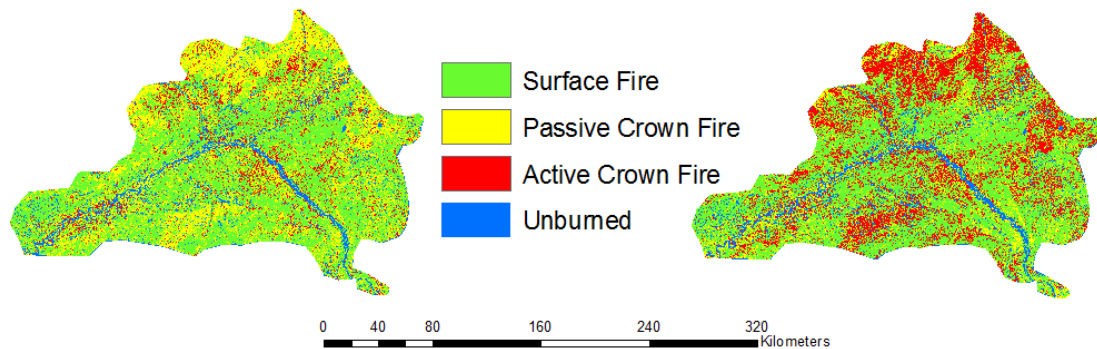


Figure 2: FlamMap Crown Fire Activity output maps using LANDFIRE CBD (left) and GLAS-derived CBD (right)

5. Discussion

The regression model derived from the airborne lidar and field-based CBD explains nearly 80% of the variability in the dataset. This is somewhat lower than those reported by others, using more parsimonious models (Andersen *et al.* 2005; Riaño *et al.* 2004; Skowronski *et al.* 2007). Several factors likely affected these results: 1) the allometries that FuelCalc uses to derive CBD are not well developed for some boreal species and could produce invalid results, 2) our limited plot sample is not representative of the full range of CBD values present within the YFE, and 3) the landscape in the YFE is very heterogeneous and the vegetation type changes across relatively small scales (10s of meters). The other studies conducted their work in more homogeneous landscapes of evergreen forest, which simplifies both the calculation of field-based CBD and the derivation of lidar metrics.

The results of the GLAS-based CBD prediction are comparable to previous work using airborne waveform data to derive CBD in the Sierra Nevada Mountains of California (Peterson *et al.* 2007). Once we split the GLAS footprints into forest classes, we were able to predict 61% (evergreen), and 72% (mixed) of the variability using GLAS metrics. The splitting into forest types likely resulted in a stronger relationship because of the way CBD is assigned to hardwood species. Two waveforms may be described by similar parameters, but one falling in an evergreen stand will have a much higher CBD associated with it than one falling into a mixed stand with a large percentage of hardwoods.

Both of the lidar-derived maps indicate CBD values for the YFE that are higher than those produced by LANDFIRE. Given the lack of field data, the LANDFIRE CBD layers for much of interior Alaska are based on expert opinion of expected fire behavior under historical weather conditions. In contrast, the airborne lidar-derived CBD map is based on field measurements. Because this map is then used to train the GLAS-derived CBD map, it is also related to the field data. By using field measurements, the resultant CBD maps capture more of the range of values and spatial variability of the landscape. Higher CBD values are expected from the lidar-derived CBD maps because the CBD calculated from the field data was generally higher than the coincident LANDFIRE CBD values. The LANDFIRE CBD values ranged from 0 to 0.15 kg m⁻³ for the field plot locations, while the calculated values ranged from 0.02 to 0.29 kg m⁻³. The average CBD value was 0.05 kg m⁻³ higher when calculated from the YFE field data than the corresponding LANDFIRE CBD value.

The FlamMap outputs confirmed the significance of the higher CBD values in the lidar-derived maps with the higher incidence of active crown fire and corresponding increases in FL and ROS. This also confirms the sensitivity of fire behavior models to changes in CBD since there were significant changes in the FlamMap outputs and other fire behavior modeling systems such as FARSITE (Finney 1998) are based on the same underlying fire behavior models as FlamMap. The outputs of these models are used operationally for both strategic and tactical resource management decisions which have significant societal, ecological, and financial consequences. Therefore, the model inputs must be based on the best available source data and methods, which are constantly evolving, to ensure the most reasonable outputs.

6. Conclusions

This study highlights a multi-scale approach to regional canopy fuels mapping using airborne and spaceborne lidar data. These methods can leverage all available lidar collections from across the United States, and their inherent ability to characterize vegetation structure, to map canopy fuel parameters. While canopy fuels have been previously mapped with airborne lidar, we have demonstrated a novel approach to mapping canopy fuels with GLAS data. There are some issues to consider prior to adapting this approach nationally. For example, the flat terrain of the YFE allowed us to reasonably ignore the effects of terrain on the GLAS waveform. For other regions with steeper slopes, this effect needs to be addressed (Lefsky *et al.* 2007). Additionally, while GLAS data are no longer being collected, spaceborne lidar data are scheduled to continue. NASA is scheduled to launch the Advanced Topographic Laser Altimeter System (ATLAS) in 2016, which will provide additional global lidar observations.

Acknowledgments

This project was sponsored by LANDFIRE. Additional support was provided by the USGS Climate Effects Network. We wish to thank Dr. Bruce Wylie, USGS EROS, for access to the lidar data and enabling the field data collection. Dr. Peterson's work was performed under USGS contract number G08PC91508. The use of any trade, product or firm name is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

- Andersen, H.-E., McGaughey, R., and Reutebuch, S., 2005. Estimating forest canopy fuels parameters using LIDAR data. *Remote Sensing of Environment*, 94, 441-449.
- Dubayah, R. and Drake, J., 2000. Lidar remote sensing for forestry. *Journal of Forestry*, 98, 44-46.

- Finney, M., 1998. *FARSITE: Fire Area Simulator – model development and evaluation*. Research Paper RMRS-RP-4, Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 1-47.
- Finney, M., 2006. An overview of FlamMap fire modeling capabilities. In: P. Andrews and B. Butler (Eds.). *Fuels Management – How to measure success: Conference Proceedings. 28-30 March 2006, Portland, OR*. USDA Forest Service Proceedings RMRS-P-41, Ft. Collins, CO, 213-220.
- Gallant, A.L., Binnian, E.F., Omernik, J.M., and Shasby, M.B., 1995. *Ecoregions of Alaska*. U.S. Geological Survey Professional Paper 1567, 1-73.
- Harding, D.J. and Carabajal, C.C., 2005. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters*, 32, L21S10: 1-4.
- Keane, R.E., Burgan, R. and van Wagtenonk, J., 2001. Mapping wildland fuels for fire management across multiple scales: Integrating remote sensing, GIS, and biophysical modeling. *International Journal of Wildland Fire*, 10, 301-319.
- Krasnow, K., Schoennagel T andVeblen, T.T., 2009. Forest fuel mapping and evaluation of LANDFIRE fuel maps in Boulder County, Colorado, USA. *Forest Ecology and Management*, 257, 1603-1612.
- Lefsky, M.A., Cohen, W.B., Parker, G.G. and Harding, D.J., 2002. Lidar remote sensing for ecosystem studies. *BioScience*, 52, 19-30.
- Lefsky, M.A., Keller, M., Pang, Y., de Camargo, P.B. and Hunter, M.O., 2007. Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms. *Journal of Applied Remote Sensing*, 1, 013537: 1-18.
- Morsdorf, F., Meier, E., Kötz, B., Itten, K.I., Dobbertin, M. and Allgöwer, B., 2004. LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment*, 92, 353-362.
- Nelson, R., Ranson, K.J., Kimes, D.S., Kharuk, V. and Montesano, P., 2009. Estimating Siberian timber volume using MODIS and ICESat/GLAS. *Remote Sensing of Environment*, 113, 691-701.
- Peterson, B., Dubayah, R., Hyde, P., Hofton, M., Blair, J.B. and Kaufman, J., 2007. Use of LIDAR for forest inventory and forest management application. In: R. McRoberts, G. Reams, P. Van Deusen, and W. McWilliams (Eds.). *Proceedings of the Seventh Annual Forest Inventory and Analysis Symposium, October 3-6, 2005, Portland, ME*. General Technical Report WO-77, 193-200.
- Reinhardt, E., Lutes D. and Scott, J., 2006. FuelCalc: A method for estimating fuel characteristics. In: P. Andrews and B. Butler (Eds.). *Fuels Management – How to measure success: Conference Proceedings. 28-30 March 2006, Portland, OR*. USDA Forest Service Proceedings RMRS-P-41, Ft. Collins, CO, 273-282.
- Riaño, D., Chuvieco, E., Condés, S., González-Matesanz, J. and Ustin, S.L., 2004. Generation of crown bulk density for *Pinus sylvestris* L. from lidar. *Remote Sensing of Environment*, 92, 345-352.
- Rollins, M., 2009. LANDFIRE: A nationally consistent vegetation, fire, and fuel assessment. *International Journal of Wildland Fire*, 18, 235-249.
- Selkowitz, D.J. and Stehman, S.V., 2011. Thematic accuracy of the National Land Cover Database (NLCD) land cover for Alaska. *Remote Sensing of Environment*, 115, 1401-1407.
- Skowronski, N., Clark, K., Nelson, R., Hom, J. and Patterson, M., 2007. Remotely sensed measurements of forest structure and fuel loads the Pinelands of New Jersey. *Remote Sensing of Environment*, 108, 123-129.
- Sun G., Ranson, K.J., Kimes, D.S., Blair, J.B. and Kovacs, K., 2008. Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data. *Remote Sensing of Environment*, 112, 107-117.

LiDAR-based estimation of forest floor fuel loads using a novel distributional approach

Jan A.N. van Aardt¹, Mary Arthur², Gretchen Sovkoplas² & Tyson Lee Swetnam³

¹Rochester Institute of Technology, Rochester, NY, USA; vanaardt@cis.rit.edu

²University of Kentucky, Lexington, KY, USA; marthur@email.uky.edu & gretchen.sovkoplas@uky.edu

³University of Arizona, Tucson, AZ, USA; tswetnam@arizona.edu

Abstract. Light detection and ranging (LiDAR) has seen significant application across a range of forest structural assessment applications, ranging from forest volume and biomass assessment, to ecological applications such as leaf area and fuel load modelling. However, quantification of sub-canopy structure remains a challenge, especially when considering downed coarse woody debris (CWD) near the ground surface. This is true because the LiDAR signal attenuates through the canopy, LiDAR systems can be set to record the last of many returns, which is often the ground itself, and there is a system-specific vertical resolution that influences detection of structure in-between returns. We applied a LiDAR distributional approach to CWD modeling that included both above-ground and theoretical “below-ground” returns, with the latter being attributed to multiple scattering effects. This was done for oak dominant forests in central Appalachia, Kentucky, USA. Medium-fast (10h) and medium-slow (100h) CWD fuel loads exhibited the best results; e.g., an adjusted $R^2 = 0.99$ and a root mean square error value of 0.111 Mg/ha (4.7% of the mean) were achieved for 100h CWD fuel loads. Independent variables included a balanced set from both the above- and below-ground distributions. Results hint at the significant potential of extending distributional approaches to CWD estimation.

Keywords: LiDAR, Coarse woody debris, fuel load, distributional analysis.

1. Introduction

The assessment of forest structure, e.g., volume, biomass, leaf area index (LAI), etc., remains a priority for forest managers and ecologists for reasons ranging from taking stock inventory, performing carbon assessment, and investment planning, to tracking invasive dynamics. One important motive for structural assessment, however, revolves around the need to better understand fire dynamics as a function of fuel load, among other things. For example: Forest managers increasingly are applying prescribed fire as a management tool in the central and southern Appalachian hardwood regions of the United States (Brose *et al.*, 2001). This use of fire as a tool, instead of a reactive management response, is based on a better understanding of the historic (McEwan *et al.*, 2007) roles of fire in this region. In fact, many management plans of national forests throughout the region already include fire as part of their toolkit, e.g., the USDA Forest Service at Daniel Boone National Forest, Kentucky. This increased use of prescribed fires necessitates the need for improved predictions of future burn behavior and improved assessment of the impacts on forest structure, especially at the landscape scale. This is only one of the reasons that light detection and ranging (LiDAR) extensively has been used to characterize forest structural attributes. Not only does such a remote sensing approach enable a synoptic assessment, but LiDAR has emerged as an accurate and precise tool for this purpose.

LiDAR-based assessment of vegetation structure has focused primarily on the assessment of forest volume and biomass and has met with reasonable success (R^2 values typically > 0.80), with studies dating back to the mid-1980s (e.g., Nelson *et al.*, 1988). More recent LiDAR studies have mostly dealt with forest volume and biomass assessment at the plot- and

stand-level (e.g., Lefsky *et al.*, 1999; Næsset, 2002; Popescu *et al.*, 2004; van Aardt *et al.*, 2006). Differentiation also is made between individual tree vs. per-unit-area height distributional approaches (Means *et al.*, 2000; van Aardt *et al.*, 2006). Distinct challenges to such approaches include the scalability of results due to huge data volumes, resource costs, as well as limited extension of work to sub-canopy structure and fuel load applications. Examples of the latter two topics include leaf area index estimation (e.g., Zheng and Moskal, 2009) and fire effects on forest structure (e.g., Hall *et al.*, 2005). However, comprehensive fuel load studies, or detailed sub-canopy structural assessments, remain scarce.

Riaño *et al.* (2003) demonstrated that LiDAR can be used to extract forest fuel distributions from LiDAR data in forests dominated by coniferous and deciduous tree species, but Skowronski *et al.* (2007) highlighted the breakdown of LiDAR-based fuel load assessment for larger areas. Mutlu *et al.* (2008) attempted to improve on LiDAR-only approaches by showing that fuel models can be calibrated using a remote sensing fusion approach that hedges on LiDAR and Quickbird imagery. Beneath the canopy, Seielstad and Queen (2003) have also used small footprint LiDAR data to measure fuel loads of coarse woody debris on the forest floor. This was done by performing "obstacle density" assessment within 2m of the forest floor. We want to expand on such methods by applying a LiDAR distributional approach, similar to those used by Means *et al.* (2000), van Aardt *et al.* (2006), and Pesonen *et al.* (2008). The first two studies focused on forest volume/biomass estimation, while Pesonen *et al.* (2008) highlighted the potential of a distributional approach for coarse woody debris assessment. We propose to improve forest floor fuel load assessment by expanding such distributional approaches to include negative digital elevation model (DEM) residual distributions.

We hypothesize that (i) above-ground LiDAR height distributions are indicative of the downed coarse woody debris (CWD) structure, either through correlations with canopy level structure or via scattering of laser pulses close to the ground and (ii) that so-called negative height residuals, or LiDAR heights that are apparently negative after subtraction of the DEM, potentially contain signal that can be associated with multiple scattering effects. Our objectives therefore are to (i) determine if hypothesized multiple scattering effects and their inclusion in LiDAR distributional analysis can contribute to modeling CWD fuel loads at 1h, 10h, 100h, and 1000h fuel weights and (ii) to assess the relative impact of an increase in DEM spatial resolution, i.e., an increase in "below-ground" returns on the first objective's outcomes. We believe that the multiple scattering effect, caused by a delayed return signal due to vegetation structural complexity (Wu *et al.*, 2009), can be linked to variation in forest floor CWD (Figure 1).

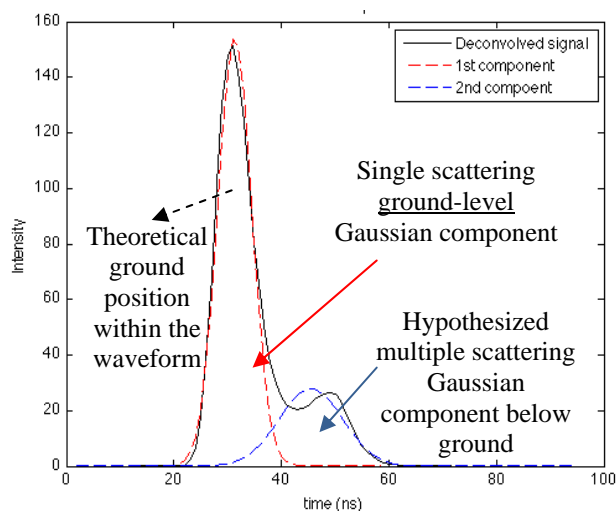


Figure 1: An example of hypothesized multiple scattering effects, i.e., a delayed LiDAR signal that exhibits as being below-ground. This example is from a savanna environment, using an Optech small-footprint waveform LiDAR (0.56 mrad beam divergence). This multiple scattering (blue Gaussian) was attributed to scattering due to dense herbaceous biomass (Wu *et al.*, 2009).

2. Methods

2.1 Study area

The study area is located in the Cumberland district, Daniel Boone National Forest, Kentucky, USA (Figure 2). This central Appalachian region is characterized by upland oak-dominated forests, e.g., scarlet oak (*Quercus coccinea* Muench.) and chestnut oak (*Q. prinus* L.), with some black oak (*Q. velutina* Lam.) and white oak (*Q. alba* L.), where fire-sensitive red maple (*Acer rubrum* L.) and eastern white pine (*Pinus strobus* L.) are gaining dominance.

2.2 Plot establishment

A total of ninety-three plots (10 x 40 m; Figure 3), part of two of the longest running prescribed fire studies on the Cumberland Plateau (Blankenship and Arthur, 2006), have been control-burned at various frequencies. These include a control set, 2003/2009 (infrequent), and 2003/2004/2006/2008 (frequent) burn plots. Plots run 10m uphill and 40m perpendicular to the slope and include January-March, 2009 pre-burn measurements for overstory (trees >10cm diameter-at-breast-height; DBH), midstory includes trees (<10cm DBH in first quadrant), and CWD in terms of 1h, 10h, 100h, 1000h fuels at the fuel transects (pre-burn). The CWD values correspond to 0-0.635cm diameter (1h), 0.635-2.54cm diameter (10h), 2.54-7.62cm diameter (100h), and 7.62 cm diameter (1000h) (Brown *et al.*, 1982). However, our analysis was constrained to 17 pre-burn, frequent burn treatment plots due to LiDAR coverage.

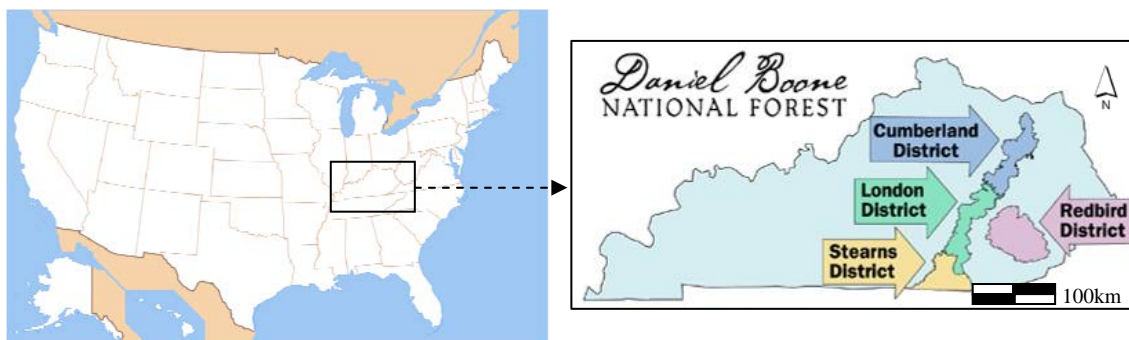


Figure 2: The study area is located in the Cumberland District of the Daniel Boone National Forest.

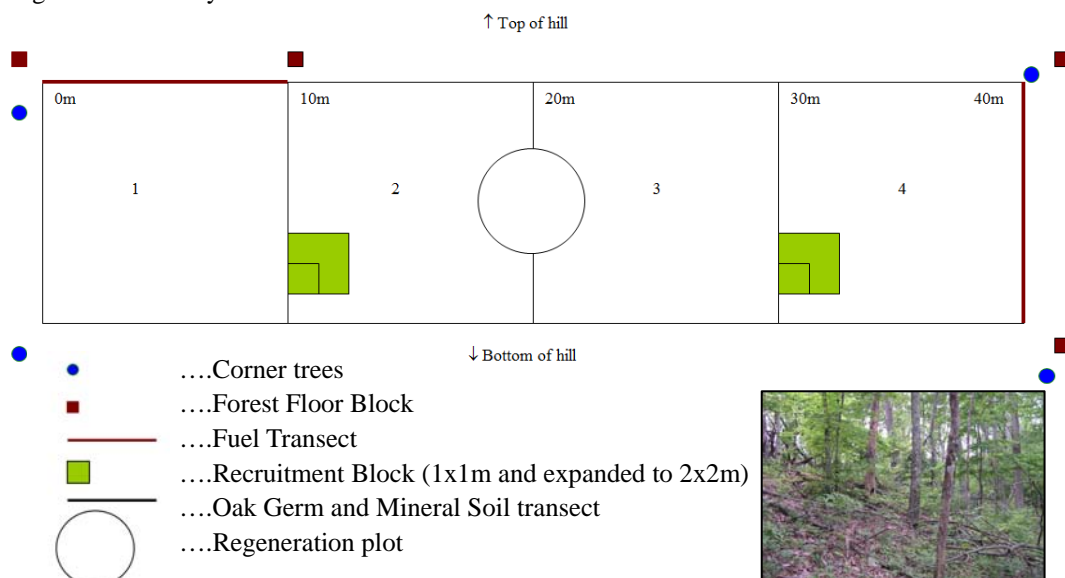


Figure 3: Plot layout for assessment of various forest structural parameters, including a line transect for measurement of CWD. The inset shows an example of a 100h fuel load CWD plot.

2.2 LiDAR data description and analysis

A Leica ALS50 (64KHz pulse rate; 43Hz scan rate; 4 returns/pulse; >5 hits/m²; 1.4m footprint; 1,655m swath) was used to collect data for the study area during February 2009, prior to the burn treatment. LiDAR returns were classified into ground vs. non-ground returns by the vendor using the well-known Terrasolid slope-radius algorithm (van Aardt *et al.*, 2006). The LiDAR returns subsequently were processed using the following workflow:

1. *DEM generation*: Kriging and nearest neighbor interpolations (Surfer V. 9.0) were used to generate 1m and 10m DEMs of the area. Two interpolation algorithms and spatial resolutions were used to assess the impacts of a theoretically smoother vs. rougher DEM and to determine if resolution impacts derivation of height residuals, respectively. Stated differently, both the interpolation approach and the varying spatial resolutions allowed us to partition the above- and below-ground LiDAR distributions differently.
2. *Residual calculation*: Residuals (height above DEMs) were calculated using Surfer (V. 9.0) in the case of each interpolation algorithm and spatial resolution for the LiDAR point cloud for 25m radius plots, centered on the field plots. We maintained negative residuals, or theoretical below-ground LiDAR returns, which was hypothesized to be related to multiple scattering effects due to scattering by CWD. This is different from most previous approaches, e.g., van Aardt *et al.* (2006), where such returns are set to zero or discarded, since they are assumed to be due to either misclassified ground returns that negatively impact DEM derivation or sensor error.
3. *Modeling*: We generated LiDAR height distribution statistics (see Means *et al.*, 2000; van Aardt *et al.*, 2006) for positive and negative residuals and their associated LiDAR intensities separately. Both sets of distribution statistics were used as inputs to stepwise linear regression (PROC STEPDISC in SAS V. 9.2), with $\alpha = 0.10$ for variable entry to result in ≤ 10 variables for all CWD models; most models contained < 4 independent variables. CWD models were assessed based on their adjusted R² and root mean square error (RMSE) values. These two metrics were considered robust, since they penalize models for overfitting and provide a precision estimate, respectively.

3. Results

Figure 4 shows an example of the below-ground LiDAR return distribution for all the plots based on Kriging at a 10m spatial resolution. While one could argue that such a seemingly normal distribution could be indicative of random system error, it remains striking that the distribution mimics a bell shape for limited multiple scattering close to the ground surface, i.e., fine matter, and a spread-out distribution tail that are theoretically due to fewer, but coarser woody debris objects. Table 1 shows the results of the analysis workflow described above. Results are given in terms of DEM spatial resolution, interpolation method, CWD level, the two model metrics, and the independent variables that were selected as significant to each model.

Table 1: CWD modeling results for all spatial resolutions, interpolation approaches, and CWD fuel levels

Cell	Grid	Fuel	Adj. R ²	*RMSE (Mg/ha)	**Variables
1m	Kriging	1h	0.22	0.102	Canopy02P
		10h	0.38	0.606	StdMeanRefN; StdMeanVegP
		100h	0.41	1.117	P_VegP_10
		1000h	0	N/A	None
	Nearest Neighbor	1h	0.51	0.081	MinVegN; P_VegN_90
		10h	0.46	0.566	P_VegN_90

Cell	Grid	Fuel	Adj. R ²	*RMSE (Mg/ha)	**Variables
10m		100h	0.74	0.734	P_VegP_90; CVVegN; StdRefN; Canopy01P
		1000h	0	N/A	None
	Kriging	1h	0.75	0.058 (16.7%)	StdMeanVegP; MaxVegN; Canopy02P
		10h	0.98	0.11 (9.6%)	MaxVegP; MinVegP; StdMeanVegP; P_VegP_10; MaxVegN; P_VegN_90; MaxRefN; RangeRefN; Canopy02P; Canopy08P
		100h	0.99	0.111 (4.7%)	MeanVegP; CVVegP; SkewnessVegP; StdVegP; MedianVegP; P_VegP_70; CVRefP; P_VegN_75; KurtosisRefN; SkewnessRefN
		1000h	0.69	4.59 (47%)	MinVegP; P_VegP_40; P_VegP_90; Canopy10P
	Nearest Neighbor	1h	0.58	0.075	RangeVegN; Canopy04P
		10h	0.28	0.653	StdMeanRefN
		100h	0.31	1.21	RangeRefN
		1000h	0	N/A	None

*% indicate RMSE as percentage of average fuel weight

***Veg* = vegetation height return; *Ref* = intensity; *P* = positive & *N* = negative residual

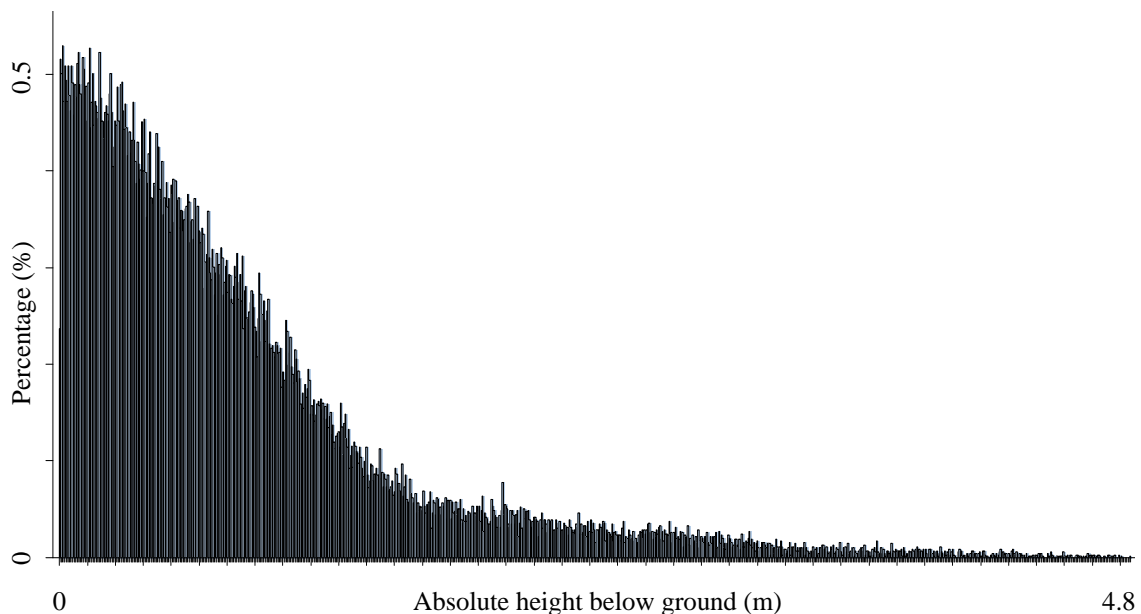


Figure 4: Below-ground return distribution for all plots (Kriging at 10m spatial resolution).

Figures 5 and 6 show the observed vs. predicted model plots for Kriging at 10m spatial resolution and 1h and 10h fuels, respectively. Finally, Figure 7 shows an example of the LiDAR return distribution for the plot that exhibited the smallest RMSE (-0.0001 Mg/ha) for Kriging at 10m and a 100h CWD fuel load. This CWD model included both positive (above-ground) and negative (below-ground) height and intensity residuals. For instance, it is interesting to note that the 75th negative percentile is included as an independent variable - this speaks to the delayed returns, due to hypothesized multiple scattering, that the "coarse" nature of 100h CWD structures would cause.

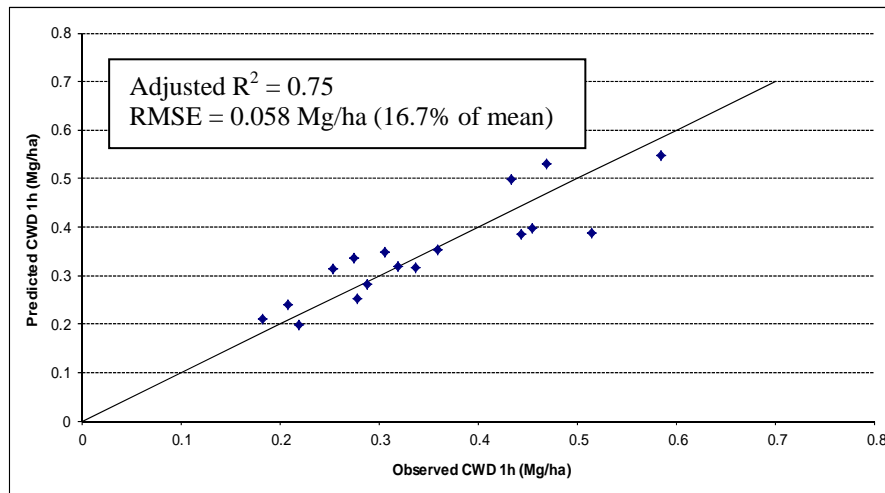


Figure 5: The observed vs. predicted plot for CWD modeling based on Kriging (10m) for 1h fuels.

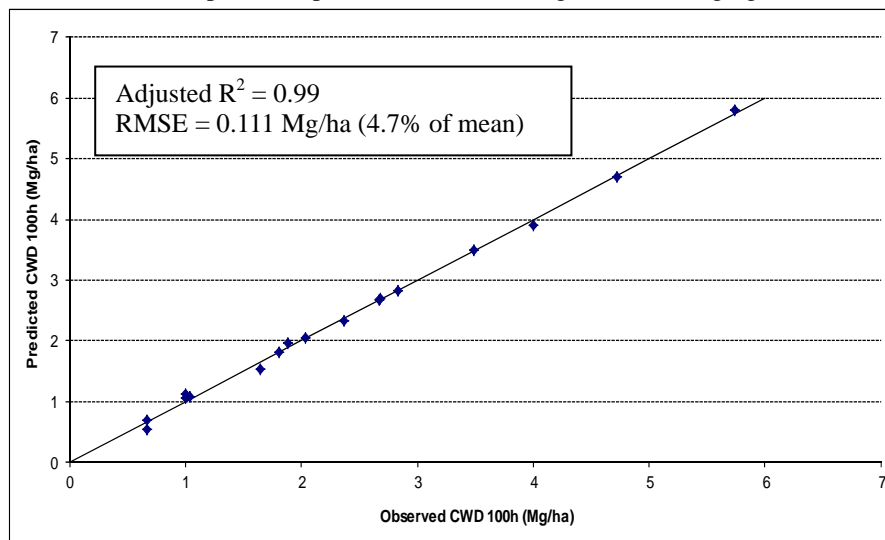


Figure 6: The observed vs. predicted plot for CWD modeling based on Kriging (10m) for 100h fuels.

4. Discussion

The best CWD model fits had high adjusted R² values of between 0.77-0.99, although models exhibited a range of adjusted R² and RMSE values. The models based on Kriging at 10m spatial resolution performed best, especially in the case of the intermediate CWD fuel loads (10h and 100h). Models with more independent variables performed better, as expected; however, these models also included a more balanced mix of negative residual (“below-ground”) and positive (above-ground) distributional variables when compared to models with fewer independent variables. These independent variable sets included specifically percentile, intensity, maximum, and range metrics, similar to findings by Means *et al.* (2000) and van Aardt *et al.* (2006). These two studies focused on above-ground forest volume and biomass estimation and it is interesting to note that the approach could potentially be extended to CWD modeling, as indicated by the results.

The selection of Kriging and a 10m spatial resolution as best performing models was attributed to the smoother, accurate interpolation due to Kriging (van Aardt *et al.*, 2006) and the manner in which 10m DEM grid cells partition the LiDAR height distribution; this latter aspect needs to be explored more fully in future studies. The low RMSE values (<10% of the mean) for medium-slow and medium-fast burning fuels were encouraging, while the high RMSE in the case of 1000h CWD fuel weight (47% of the mean) was attributed to the small sample and number of independent variables. Finally, we concluded that, depending on the application, there is evidence that previously considered “erroneous” return values could be useful as indicators of fine-scale structure close to the ground surface. This especially was evidenced by inclusion of negative residual intensity and range values, which show that such delayed signals are useful for modeling structures close to the ground surface. However, this approach requires significant validation across regions, since industry experts (e.g., Joe Liadsky, Optech system engineer, *personal communication*) expressed doubts that multiple scattering could be detected at narrow beam divergence angles (0.56 mrad), even though he acknowledged that there is evidence of “structure” in the signal.

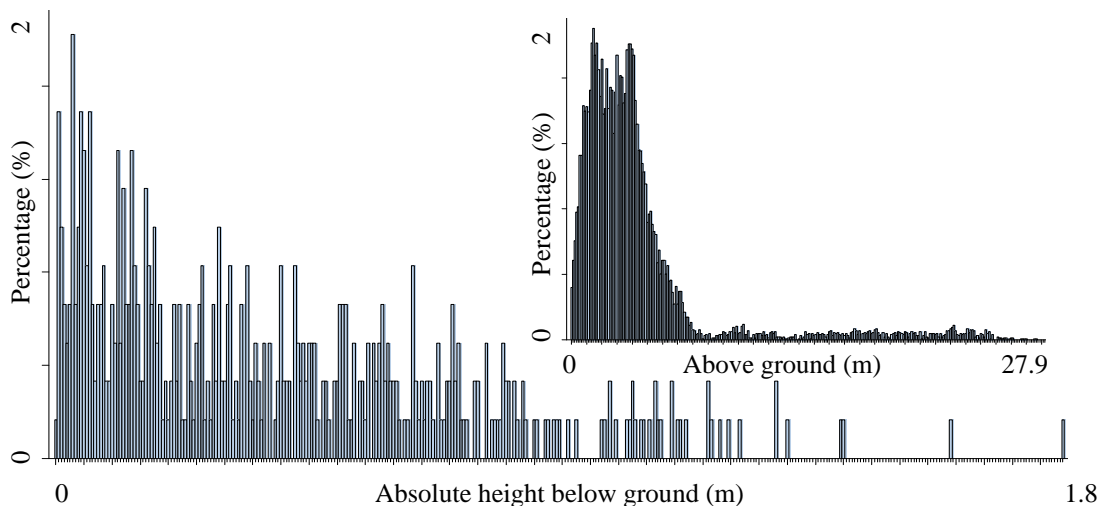


Figure 7: Below-ground and above-ground (inset) LiDAR return distributions for the plot that exhibited the smallest RMSE for CWD modeling based on Kriging (10m) for 100h fuels.

4. Conclusions

We have shown that consideration of close-to-ground multiple scattering in CWD fuel load modeling is justified. This was corroborated by the inclusion of negative LiDAR return DEM residuals as part of the independent variable set in most CWD models. Fusion of above-ground and scattered/delayed returns appeared to work best in the case of medium-fast (10h) and medium-slow (100h) burning fuels. This was attributed to lack of a distinct signal for fine 1h fuels and very coarse 1000h fuels, resulting in limited or dampened LiDAR responses, respectively. Future research will focus on validating the approach by increasing the sample size and the LiDAR point density, and testing across diverse regions.

Acknowledgements

We would like to thank Kucera International for the LiDAR data collection. This work was supported in part by the National Science Foundation’s Partnerships for Innovation Program (# 0917839) and the United States Department of Agriculture (USDA; #3049024176-11-419).

References

- Blankenship B.A. and M.A. Arthur, 2006. Stand structure over nine years in burned and fire-excluded oak stands on the Cumberland Plateau, Kentucky. *Forest Ecology and Management* 225: 134-145.
- Brose P., T. Schuler, D. Van Lear and J. Berst, 2001. Bringing fire back: the changing regimes of the Appalachian mixed-oak forests. *Journal of Forestry* 99: 30-35.
- Brown J.K., R.D. Oberheu, and C.M. Johnston, 1982. Handbook for inventorying surface fuels and biomass in the interior West. *General Technical Report INT-129*. Ogden, UT: USDA Forest Service, Intermountain Forest and Range Experiment Station.
- Hall S.A., I.C. Burke, D.O. Box, M.R. Kaufmann, and J.M. Stoker, 2005. Estimating stand structure using discrete-return LiDAR: an example from low density, fire prone ponderosa pine forests. *Forest Ecology and Management* 208:189-209
- Lefsky M.A., W.B. Cohen, S.A. Acker, G.G. Parker, T.A. Spies, and D. Harding, 1999. LIDAR remote sensing of the canopy structure and biophysical properties of Douglas-fir/western hemlock forests. *Remote Sensing Environment* 70, 339–361.
- McEwan, R.W., T.F. Hutchinson, R.P. Long, R.D. Ford, and B.C. McCarthy, 2007. Temporal and spatial patterns of fire occurrence during the establishment of mixed-oak forests in eastern North America. *Journal of Vegetation Science* 18: 655-664.
- Means J.E., S.A. Acker, B.J. Fitt, M. Renslow, L. Emerson, and C.J. Hendrix, 2000. Predicting forest stand characteristics with airborne scanning LiDAR. *Photogrammetric Engineering and Remote Sensing* 66(11):1367-1371.
- Mutlu M., S.C. Popescu, C. Stripling, and T. Spencer, 2008. Mapping surface fuel models using LiDAR and multispectral data fusion for fire behavior. *Remote Sensing of Environment* 112: 274–285.
- Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80:88-99.
- Nelson, R., W. Krabill, and J. Tonelli, 1988. Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment* 24:247-267.
- Pesonen, A., M. Maltamo, K. Eerikäinen, and P. Packalén, 2008. Airborne laser scanning based prediction of coarse woody debris volumes in a conservation area. *Forest Ecology and Management* 255: 3288–3296.
- Popescu, S.C., R.H. Wynne, and J.A. Scrivani, 2004. Fusion of small-footprint LiDAR and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, U.S.A. *Forest Science* 50(4):551-565.
- Riaño D., E. Meier, B. Allgower, E. Chuvieco, and S.L. Ustin, 2003. Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sensing of Environment* 86: 177–186.
- Seielstad C. A. and L.P. Queen, 2003. Using airborne laser altimetry to determine fuel models for estimating fire behavior. *Journal of Forestry*, 101, 10-15.
- Skowronski N., K Clark, R. Nelson, J. Hom, and M. Patterson, 2007. Remotely sensed measurements of forest structure and fuel loads in the Pinelands of New Jersey. *Remote Sensing of Environment* 108: 123–129.
- van Aardt, J.A.N., R.H. Wynne, and R.G. Oderwald, 2006. Forest Volume and Biomass Estimation Using Small-Footprint LiDAR-Distributional Parameters on a Per-Segment Basis. *Forest Science* 52 (6):636-649.
- Wu J., J.A.N. van Aardt, G. P. Asner, R. Mathieu, T. Kennedy-Bowdoin, D. Knapp, K. Wessels, B.F.N. Erasmus, and I. Smit, 2009. Connecting the dots between laser waveforms and herbaceous biomass for assessment of land degradation using small-footprint waveform lidar data. *Proceedings of the 2009 IEEE IGARSS*, July 13-17, 2009, South Africa, 4p.
- Zheng G. and L.M. Moskal, 2009. Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors. *Sensors* 9(4):2719-2745.

Using Airborne Survey to map Stream Form and Riparian Vegetation Characteristics across Victoria

Nathan Quadros¹, Rick Frisina² & Paul Wilson³

¹Information Services Branch, DSE Victoria, nathan.quadros@dse.vic.gov.au

²Information Services Branch, DSE Victoria, rick.frisina@dse.vic.gov.au

³Office of Water, DSE Victoria, paul.wilson@dse.vic.gov.au

1. Introduction

The State of Victoria has established a methodology for providing river health data for regional planning. The index of stream condition (ISC) is a baseline dataset used in river investment and planning. The ISC evaluates the environmental conditions of the major rivers and tributaries across Victoria. The ISC integrates the results from five key components of river health; river hydrology, water quality, riparian vegetation, stream form and aquatic life. These components and their metrics are shown in Table 1.

Table 1: Structure of the Index of Stream Condition (ISC)

Hydrology	Stream Form	Riparian Vegetation	Water Quality	Aquatic Life
Low flows	Bank condition	Vegetation width	Total phosphorus	AUSRIVAS
High flows	Artificial instream barriers	Fragmentation	Turbidity	SIGNAL
Zero flows	Large instream wood	Vegetation overhang	Salinity (EC)	EPT Index
Seasonality		Large trees	pH	Number of
Variability		Weeds (Willows)		Families
		Foliage cover		
		Structure		

Previous ISC assessments were successfully undertaken for all five components in 1999 and 2004. In these assessments the in-stream components of hydrology, water quality and aquatic life were sampled at fixed sites within the streams. The field sites were usually at the downstream end of a river reach (a uniform section of river). This typically ensured that all processes which act upon the reach are accounted for in the measurement sample. However, the stream form and riparian vegetation components of the ISC used random field site selection for the assessment. The random sample data was collected within reaches and the results used to infer the condition of the whole reach.

The random sampling for the stream form and riparian vegetation components of the ISC were frequently not at locations where there had been on-ground works programs to improve stream condition. Therefore the ISC could not be used to assess the impact of Government investment to improve stream condition. This has led DSE to investigate and develop the ability to assess these components of the ISC using remote sensing techniques. Thus avoiding the need for random site selection and to mobilise field crews to assess some 2500 sites across the State. Instead a census approach could be adopted that would collect data from 27,000km of stream length across the state. The results of the study have shown that LiDAR and aerial photography data can be used to accurately map various riparian vegetation and stream form metrics for the ISC.

The ISC assessment undertaken during 2009-2010 has included a significant investment in the use of LiDAR and aerial photography to assess the riparian vegetation and river form components of ISC stream network. Fugro Spatial Solutions (FSS) have been performing the surveying and metric production components of the project.

The use of remote sensing will result in a more accurate and comprehensive assessment of riparian vegetation and stream form which will in turn provide significant benefits when prioritising and undertaking on-ground works to improve river condition.

2. Measuring Riparian Vegetation and Stream Form Using Airborne Surveys

The health of riparian vegetation and stream form is assessed by measuring a number of metrics. Each of the metrics can be measured by either or combination of LiDAR and/or aerial photography. As vegetation measurements are required for assessing riparian health the surveys have to be designed to provide enough measurements within the vegetation canopy. Preliminary investigations outlined the key specification requirements for producing riparian metrics as requiring:

- Four point discrete echo LiDAR data
- Along track average point spacing of 0.5m
- Across track average point spacing of 0.5m
- An evenly distributed average point density of 4 points/m²
- A laser footprint size of 0.3m
- A maximum scan angle of 30 degrees; no more than 15 degree half angle
- An absolute vertical accuracy to AHD of $\pm 20\text{cm}$ at 1σ for the ISC rivers
- A horizontal accuracy of $\pm 30\text{cm}$ at 1σ
- A relative vertical accuracy of $\pm 5\text{cm}$ at 2σ
- Data covering at least 300m either side of the stream banks

The aerial photography was collected at a 15cm resolution, and as near as possible to the time of the LiDAR data collection. Field data was also collected at 200 sites across the State for verification purposes within a similar timeframe.

3. The 2009-10 Victorian ISC Survey

The 2009-10 Victorian ISC Survey included 27,000km of major rivers and tributaries from across the state. The project management involved dividing the survey into the ten river catchments shown in Figure 2. This generally resulted in each catchment being surveyed over a one month period and the processing of data for each catchment occurring independently.

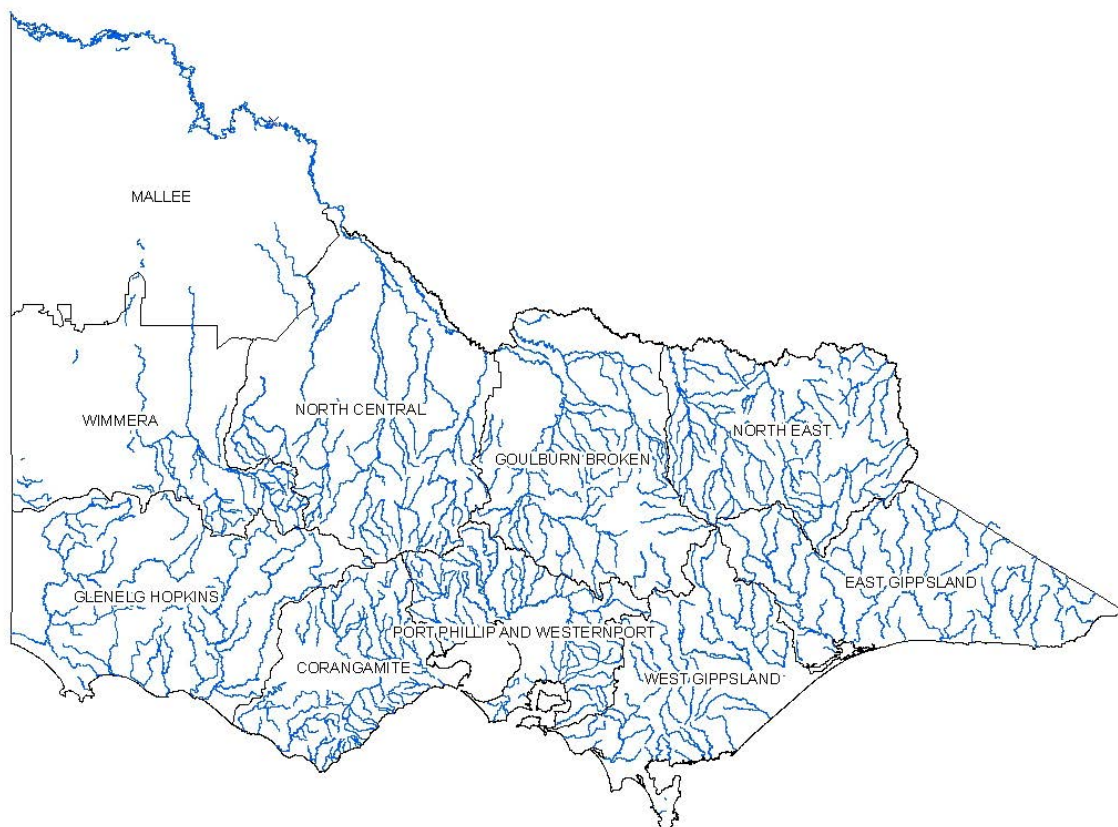


Figure 21: The major rivers and catchments within the Victorian Index of Stream Condition (ISC)

The project produced four generic spatial product types that depict both stream form and riparian vegetation characteristics: aerial photography, standard LiDAR, non-standard LiDAR and ISC Metric products.

The imagery captured during the project was 15cm RGBI (colour visible and infra red) aerial photography. The standard LiDAR products consisted of ground surface DEMs and classified LAS files, the production of which are typical and well understood.

The non-standard LiDAR products are less typical, while the ISC metric products have specifically been developed for this project. These products are less well understood and warrant further description. The non-standard and ISC metric products delivered as part of the project include:

Non-Standard LiDAR products used to derive the ISC metrics:

- Canopy height model
- Ground slope model
- Fractional cover counts (a measure of the amount of vegetation present)
- LiDAR intensity images

ISC metric products used to assess river condition:

- Stream form metrics: streambed width, bankfull width, river centrelines, water bodies, bare ground, bank condition, large in-stream wood and artificial instream barriers.
- Riparian vegetation metrics: width of vegetation, foliage cover, structure, vegetation overhang, fragmentation, tree weeds (willows) and large trees.
- River health sub-index: riparian sub-index and stream form sub-index

In addition, field products were used for calibration, validation and/or quality assurance of all the above products.

The development of the non-standard LiDAR products and the ISC metrics used for the assessment of river health are outlined in this paper.

4. Non-Standard LiDAR Products used as Input to ISC Metrics

The non-Standard LiDAR products described in this paper are the fractional cover count, canopy height model and slope products. These products are used derive a number of the ISC metrics. For these products, automatic classification and operator visual checks and corrections have been undertaken at FSS to remove built structures and non-vegetation returns from the LAS point cloud before product development.

4.1 Canopy Height Model (CHM)

The canopy height is the vertical distance between the ground and top of vegetation as computed from the LAS point cloud. As shown in Figure 2, the canopy height model has been produced for each 2m cell up to a distance of 300m from the bank full width for each stream. This distance from the stream is generally the limit of riparian vegetation.

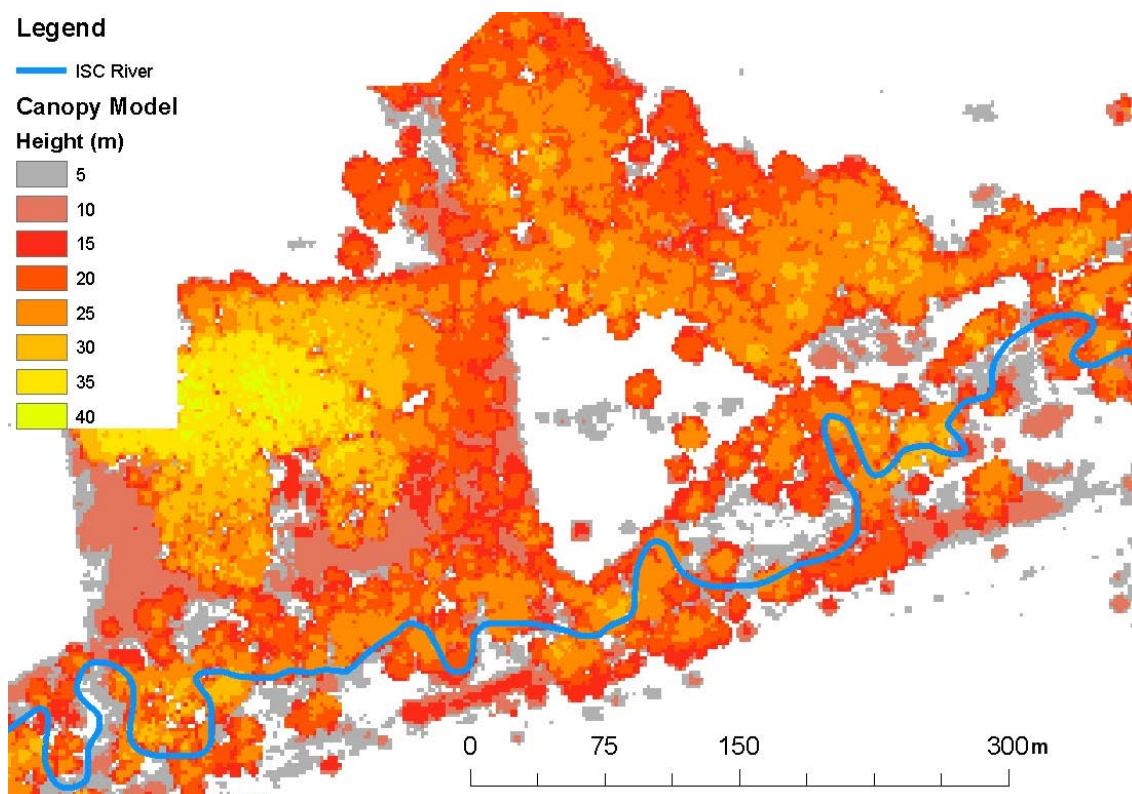


Figure 2: The canopy height model (CHM) surrounding a stream in the ISC

4.2 Fractional Cover Count (FCC)

The fractional cover count measures the density of vegetation in a particular area. This is expressed as the number of LiDAR points classified as vegetation as a percentage of all LiDAR points for the particular area. Two types of FCC are calculated: height class FCC (used in

foliage cover and structure) and overall FCC (used in fragmentation, vegetation overhang and vegetation width).

4.2 Ground Slope and LiDAR Intensity

The ground slope is the rate of change in height of the ground. It is computed in degrees from the DEM as the change in height across a particular cell. See examples of the ground slope as a background in the examples shown for the streambed and bankfull widths.

The LiDAR intensity image displays the strength of the return LiDAR pulse. More reflective surfaces give a higher value than surfaces which absorb the LiDAR pulse.

5. ISC Metrics Used to Measure Stream Condition

Other than stream form and riparian vegetation the ISC metrics can also be divided into referenced and unreferenced metrics. Several metrics are scored based on their comparison to a reference condition. This can be thought of as relative scored ISC metrics, as opposed to an absolute score.

The reference condition is determined from a presumed natural, or pristine, condition of a stream. The ISC compares the expected pristine condition against the observed current condition to infer the condition (ie little difference between expected and observed equals good condition, while a large change between expected and observed equals poor condition). The use of reference conditions allows comparison of different reaches across Victoria. Therefore, comparisons are not only restricted to streams in the same geographical zones, more general comparisons about improvement or deterioration of health can be made between catchments. This statewide comparison is an important feature of the ISC. For the riparian vegetation metrics the reference condition uses the Ecological Vegetation Class (EVC) benchmark. These have been mapped for the whole of Victoria. However, they needed to be remodelled, due to the level of accuracy required for comparison with the LiDAR products. The new EVC benchmarks have been used as the expected condition for the riparian vegetation metrics. For the stream form, the reference condition had to be specifically derived for this project.

The descriptions of the ISC metrics are split into stream form and riparian vegetation metrics, within each of these sections they are scored as unreferenced or referenced metrics.

5.1 Stream Form ISC Metrics

The stream form is the study of the shape of the land and the processes that are maintaining or changing that shape today. In rivers, it describes and explains processes of erosion and deposition and their relationships with different substrates and flow patterns. The stream form metrics in the context of the ISC are the streambed width, bankfull width, river centrelines, water bodies, bare ground, bank condition, large in-stream wood and fish barriers.

Only the last three stream form metrics are used in the ISC scoring for river health. The first five stream form metrics are used as inputs to the remaining stream form and riparian vegetation metrics.

5.1.1 Streambed Width (SBW)

Streambed width is the distance between the toes of the bank (where the bed of the stream meets the stream bank) on each side of the river. Where the toe of the bank is not visible due to water it is taken as the water level.

The pre-existing (but low accuracy) ISC river network supplied as part of the original specifications is used to guide the approximate location of the streambed. The majority of SBW is mapped from the non-standard LIDAR ground slope layer as shown in Figure 3. However, sections have been mapped from the aerial imagery where the LiDAR data coverage is not adequate. The slope layer (or occasionally the aerial imagery) is used to screen digitise the toe of the river bank for both sides of the river. The SBW is identified either by the wetted perimeter (water in the channel extending from bank to bank) or by a clear break towards the bottom of the slope (which is shown in white below).

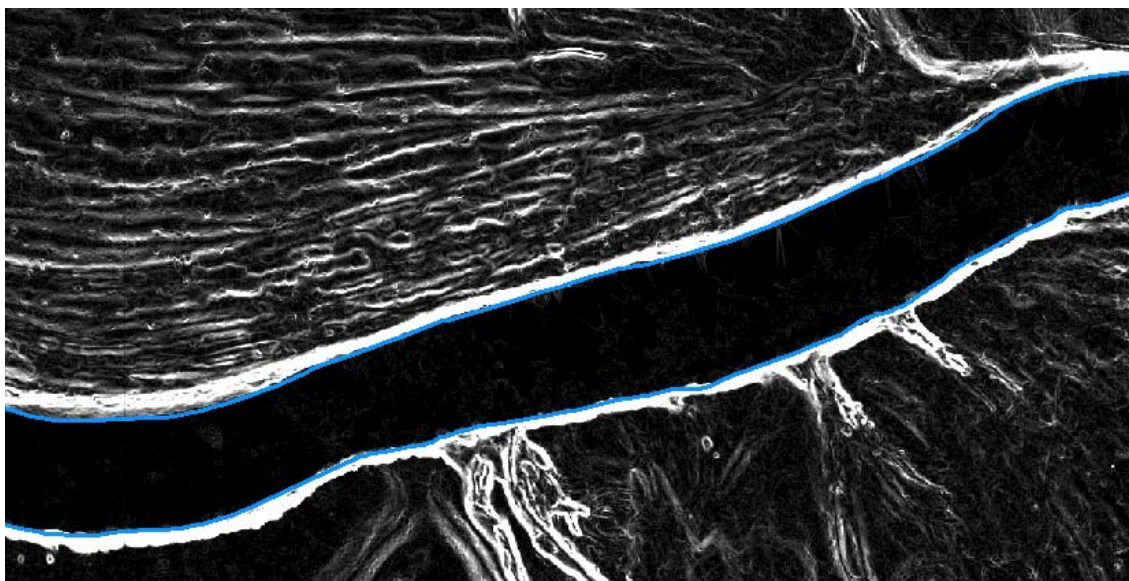


Figure 3: The streambed width (in blue) is delineated at the bottom of the slope

An automated process was tried for mapping SBW (and bankfull width), however it did not produce a consistently accurate product across all rivers within Victoria. It was subsequently replaced with screen digitising.

An automated process developed by FSS divides the SBW line work into reaches. The reaches are sections of streams that are typically 10 – 30km in length and fairly homogeneous in terms of hydrology, vegetation and geomorphology. The new river centre lines are created from the SBW; reach breaks are first introduced to the centrelines and then applied to SBW using a line which is generally perpendicular to the direction of the stream.

5.1.2 Bankfull Width (BFW)

Bankfull width represents the width between the top of the lowest bank and the corresponding height on the opposite bank as computed on both sides of the river. It sits outside the streambed and is effectively the width of the stream before it breaks its banks. The BFW is shown in Figure 4 over the top of the slope layer.

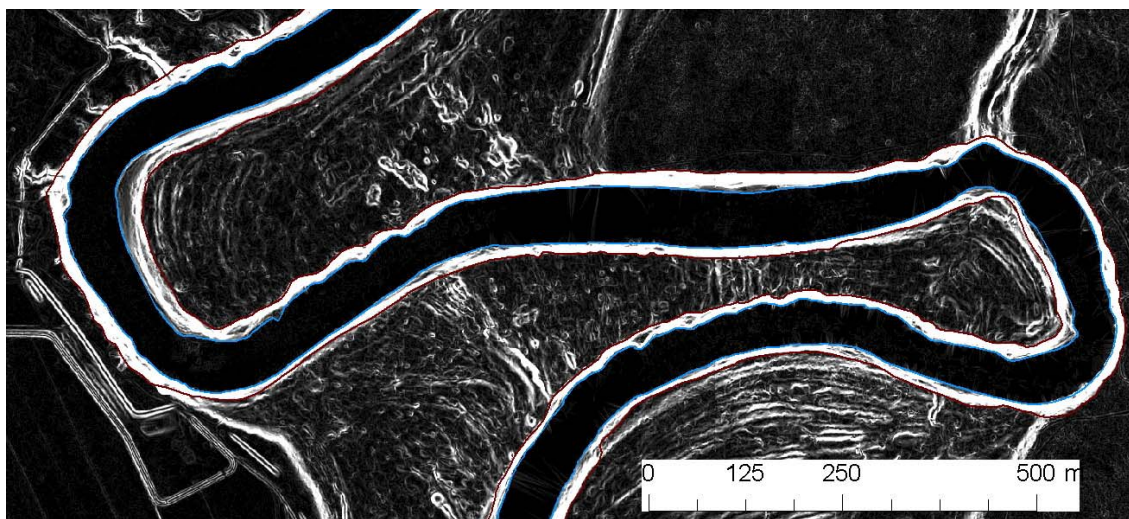


Figure 4: The bankfull width (in brown) is delineated at the top of the slope.

Like the SBW, the pre-existing ISC river network is used to guide the approximate location of the BFW. The BFW uses the same inputs as the SBW, and is mapped at the same time using screen digitising to map the top of bank for both sides of the stream. An automated in-house process which was developed by FSS is then used to level both sides of the bank. At 5m intervals along the river a transect is created across the river near-to-perpendicular to the centreline. The lowest bank is then identified and a node is created on the opposite bank at the same height as the lowest bank on the same transect. Adjacent nodes for the left and right bank along the river are subsequently joined to create level BFW lines.

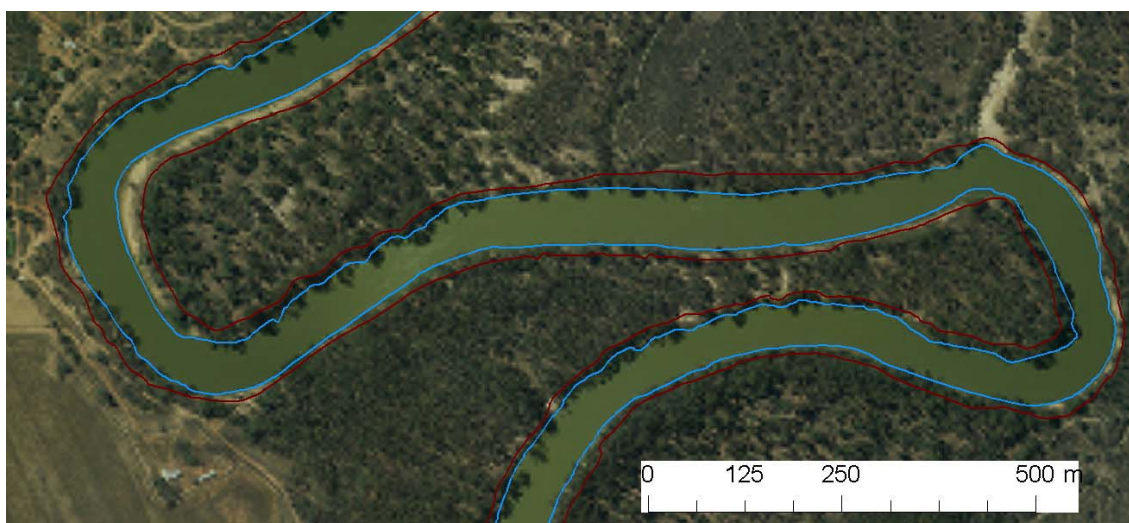


Figure 5: The bankfull width (in brown) is shown with the streambed width (in blue) over aerial imagery.

5.1.3 River Centrelines (RC)

The river centrelines define the middle of the river and are located at the midpoint of the streambed width. The RC are mapped from streambed width layer. However, sections without a streambed width have been mapped from the aerial imagery to complete the river network. These unmapped areas relate to large bodies of water like lakes which did not have their banks mapped.

An automated vector process is used to find the midpoint of the streambed width at every 1m

along the network. Part of this process relates the new river reaches to the old river reaches and aligns the reach breaks. This process also divides the river centreline network into 100m sections. It should be noted that the last section in a reach is between 50 to 150m long, so that the new reach begins with a new 100m section.

Due to the large amounts of data produced for the ISC it is necessary to develop a geodatabase to store the data. The central element of the database is the river centrelines network which is divided into 100m segments, within the ISC stream reaches (approx 1200 reaches, which are typically 10 – 30km in length). It is important to note that the river centreline network is the core product within the geodatabase for spatially linking the ISC attributes, statistics, scoring and reference scores for all reach and 100m sections of river.

5.1.4 Water Bodies (WB)

Water bodies are defined as areas of the within the bankfull width which contain water at the time of survey. The water bodies are mapped from the aerial imagery using a water index to classify areas of water within the streambed. The water bodies are not used in the ISC scoring however will be used within ISC follow up work.

5.1.5 Bare Ground (BG)

The bare ground is defined as areas of no grass or vegetation on the bank face. The bare ground is mapped by selecting areas on the bank face which are not classed as either vegetation or water from the imagery.

The bare ground is not scored within the ISC, however is used as an input into bank condition.

5.1.6 Bank Condition (BC)

The BC has been scored based on the proportion of bank face that has been classified as eroding above the expected rates. First of all, confined streams are identified in the ISC river network and removed from the BC calculation process. Confined segments are not able to move laterally and so are given the best condition score.

Stream bank types (SBT) are then manually identified for each 100m section. This is because the sediment composition of the streambed and bank affect the stability of the stream. All streams are classified as one of eight SBTs which each have particular reference erosion thresholds.

For the BC each cell on the bank face is classed as eroding or not eroding. This is based on the predefined threshold value for the slope and whether the cell is identified as bare ground. Each SBT has different slope thresholds. The erosion scores are aggregated into the 100m sections. For every 100m section, the % number of pixels classified as eroding is calculated.

The mapped erosion percentages are compared to reference erosion thresholds for each 100m section. The reference is based on the SBT and the planform of the river (straight, curved, tortuous). The results are recorded for each 100m section as either being as expected for erosion or having too much erosion.

The erosion scores for all 100m sections are aggregated and a percentage of eroding sections is calculated for each reach. The confined segments are added back in at this stage as non-eroding segments. The final % eroding sections is computed into a final score for the ISC.

5.1.7 Large In-Stream Wood and Artificial Instream Barriers

The large in-stream wood and artificial in-stream barriers were manually assessed using the 15cm aerial photography. Artificial in-stream barriers include such things as bridges, road crossings, and weirs. The in-stream large wood (ie. logs) is assessed based on their size and complexity. On the smaller streams where there is a dense canopy of overhanging vegetation, it is not possible to assess if any in-stream large wood is present. In these situations, the vegetation overhang component will be used and recorded as potential in-stream large wood.

5.2 Riparian Vegetation ISC Metrics

The riparian vegetation is the vegetation next to streams, also known as the riparian zone. The riparian width is defined by the vegetation itself, since plants requiring wet soil characteristics are usually different to those adjacent to the stream. Consequently, the riparian zone is any land that is next to, or is directly influenced by a stream. The riparian vegetation metrics in the context of the ISC are the vegetation width, fragmentation, vegetation overhang, large trees, foliage cover and structure.

5.2.1 Vegetation Width (VW)

Vegetation width is the measurement from the toe of the bank to where the overall vegetation cover is less than 20% FCC. VW is measured perpendicular to the stream direction.

The overall FCC is filtered for values greater than 20% and within 200m of the SBW. A polygon feature class is created representing the vegetation beside the river. Transects perpendicular to the stream direction are created every 5m to compute the VW from the stream.

The VW transects are stored as an ESRI line feature class within the ISC geodatabase. Using the transects to compute the VW width, the measurements and scores for each reach and 100m section of river are stored in the geodatabase.

5.2.2 Fragmentation

Fragmentation is a measure of how patchy the riparian vegetation is. Gaps in the riparian vegetation are defined by woody vegetation with less than 20% FCC and covering an area greater than 10x10m. The searches for gaps within the riparian vegetation are limited to within 40m of the SBW. The percentage area of fragmentation within each reach is then used to calculate the fragmentation score within the ISC.

5.2.3 Vegetation Overhang (VO)

Vegetation overhang is the length of woody vegetation along the stream overhanging the streambed. The overall FCC is used to identify areas with greater than 20% vegetation cover and a polygon representing vegetation cover is used as an input for the computation of the VO. The vegetation polygons are intersected with the SBW lines to produce a VO line feature class representing VO over the streambed. The percentage overhang on both sides of the stream are then used to compute the VO ISC score for each reach.

5.2.4 Large Trees (LT)

The large trees are defined as vegetation canopies which exceed a predefined height, relating to the most common canopy tree species to occur in an EVC. The large trees use the EVCs to provide referential scoring for the ISC. The Victorian ecological vegetation class (EVC) spatial

layer is used to apply benchmark scores to each riparian area. The EVC reference scores are based on the average characteristics of a mature and apparently long-undisturbed stand of the same vegetation community.

For the LT the EVC raster grid (within a 60m SBW buffer) is converted so that each cell represents the EVC LT canopy height obtained from a look-up table which lists the most common canopy tree species for each EVC. The canopy height model is then filtered for large trees within each of the EVCs. A “window” which matches the minimum crown size of the typical LT within the relevant EVC is used to group pixels in the CHM within each EVC. Groups of pixels which have 75% of their pixels larger than the reference EVC height are identified as large trees. An example of a large tree raster grid before the “window” filtering is shown in the figure below.

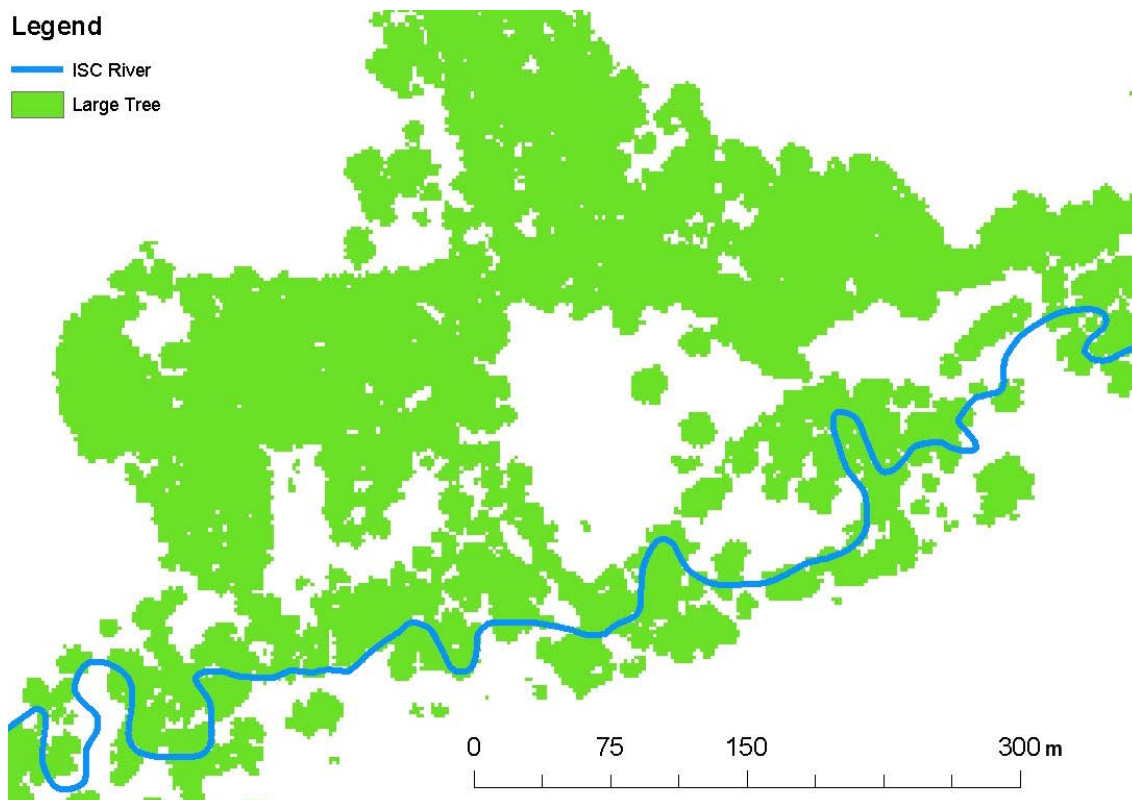


Figure 6: The LT raster grid before filtering for crown size (same location as Figure 2)

The scoring for LT is based on the percentage difference in area of LT between the expected and observed for each EVC within each reach.

5.2.5 Foliage Cover

The foliage cover is defined as the percentage vegetation cover by leaves, branches and tree trunks for three height categories:

1. 0.3 – 1.5m (represents ground cover)
2. 1.5 – 5m (represents shrubs)
3. > 5m (represents trees)

The foliage cover filters each of the height classes above for fractional cover counts with values greater than 20% cover. Each of the three height classes are then given a presence or absence of vegetation for each cell. The percentage area of cover for each of the height classes are

computed for each EVC. The expected percentage area cover for each of the three layers, which is determined from reference locations, is then computed against the observed percentage area cover for each of the three layers. Three ISC scores are then computed for each reach based on each of the three height classes.

5.2.6 Structure

The structure metric is a measurement of the presence or absence of vegetation within each of six fractional cover height intervals. For each of the height classes listed below fractional cover counts are tested for values greater than 20%.

- 1.5 – 5m
- 5 – 10m
- 10 – 15m
- 15 – 20m
- 20 – 25m
- >25m

Each of the six height classes are given a presence or absence of vegetation for each cell. Cells assigned the value of 1 have vegetation present in the particular height class and cells assigned 0 have no vegetation present. The presence and absence values are added together to give the total number of height classes present for each cell.

The next step involves referencing the observed structure against the expected structure based on the EVCs. The difference between the observed mean value and the expected number of height classes are recorded for all pixels within each EVC. Scores are then aggregated for each reach and input into the ISC.

5.3 ISC Metric Sub-Indices

The ISC metric sub-indices amalgamate the reach scores for the riparian vegetation and stream form metrics into a single score for each sub-index. The sub-indexes will be combined with the other sub-indices shown in Table 1 to provide an assessment of health for each river reach in Victoria.

6. Conclusions

Compared to the results from prior research, the preliminary results have shown that riparian vegetation and stream form metrics can be measured to varying degrees of success for the Victorian ISC. Although, it should be noted that the results are a significant increase in detail, coverage and accuracy from the previously available information and the products are still being tested and validated in various catchments across the state.

More significantly, the final results will provide a much needed input into the Victorian Strategy for River, Estuaries and Wetlands and more specifically will help to set management targets. The datasets will provide a key source of information for the regional priority setting process which identifies sites where on-ground works should be undertaken to either protect river values or to mitigate against risks to their values.

Acknowledgements

The authors wish to thank:

Fugro Spatial Solutions for the surveying, development and production of the standard, non-standard and ISC metric products.

The *Centre for Remote Sensing and Spatial Information Science, School of Geography, Planning and Environmental Management, University of Queensland* for the trial work on the ISC metric production.

The *Department of Sustainability and Environment, Arthur Rylah Institute for Environmental Research* for the EVC development and production.

The *University of Melbourne, Department of Resource Management & Geography, Dr James Grove* for support on the development of the ISC metric products.

Full Waveform LiDAR for Assessment of River Health

David Moore¹, Alys Wall², Thomas Wilson³.

¹Terranean Mapping Technologies Pty. Ltd. david.moore@terranean.com.au

²Bureau of Meteorology. A.Wall@bom.gov.au

³Terranean Mapping Technologies Pty. Ltd. thomas.wilson@terranean.com.au

1 Introduction

The Murray-Darling Basin covers 1,000,000 square kilometres; 13% of the Australian continent and produces 39% of Australia's total agricultural output. However, thirteen of the Basin's 23 valleys are categorised as being in 'very poor' health and a further seven are in 'poor' health (Davies et al 2008). The Murray-Darling Basin Authority (MDBA) is responsible for reporting the health of the Basin's rivers and for this purpose the Sustainable Rivers Audit (SRA) was established to collect and analyse data to determine the environmental condition of the rivers by objective and repeatable methods.

The first SRA assessed river health under three indicator themes; 1) Fish, 2) Macroinvertebrates and 3) Hydrology. Two additional themes were included in the second SRA; Physical Form (of the river channels) and Vegetation. This project describes how these two additional themes were measured using full-waveform LiDAR. Highly automated workflows developed in the TNTmips™ Spatial Modelling Language were used to extract information from the LiDAR for statistical analysis. A semi-automated tool was developed to measure a range of river channel and foliage attributes by objective and repeatable means. More than 50 measurements are generated for each of 1610 sites stratified across the Basin's 23 river valleys. Extensive field survey was used to verify the accuracy of the data.

Full waveform LiDAR represents an extremely rich source of environmental information. This was exploited by the Sustainable Rivers Audit (SRA) program as a means of taking diverse measurements of environmental attributes. The project methodology was highly innovative and required the development of new algorithms and processes to extract the required information from LiDAR point clouds.

The data processing produced five successive levels of data, each more refined than the previous. These levels are:

1. Raw waveform LiDAR instrument data, airborne GPS/IMU and field GPS
2. Processed LiDAR point cloud; classified LiDAR tiles.
3. Derived data: terrain surfaces, vegetation surfaces, foliage density rasters, contours
4. Channel features; top and bottom banks, centrelines, riparian zones
5. Measurements; Data matrices containing channel and vegetation measurements such as channel depth, width, bank angles, stream power and foliage density by strata.

Levels 1 to 3 are common to most full wave LiDAR projects, except that the foliage density rasters are not standard outputs from a LiDAR project. The generation of Level 4 datasets was the most innovative aspect of this project and involved the development of algorithms and software tools, with significant input from geomorphologists and ecologists. Having produced the Level 4 data, the data matrices (Level 5) were relatively straightforward to generate, but the matrices include a significant number of measured variables that normally would not be available for environmental assessments.

2 Method

2.1 Data Collection

A stratified sampling design was employed, 70 sites were randomly generated on the river networks within each of the 23 valleys in the Basin, giving a total of 1610 sites. Within each valley the sites were stratified against elevation zone and broad vegetation type. An additional 109 check sites were also selected to measure the spatial accuracies achieved in different types of vegetation and topography.

Each sampling site was a 2000 metre by 700 metre rectangle aligned to the primary river channel. Within the rectangular site, a one kilometre stretch of river channel and adjacent riparian zone were analysed.

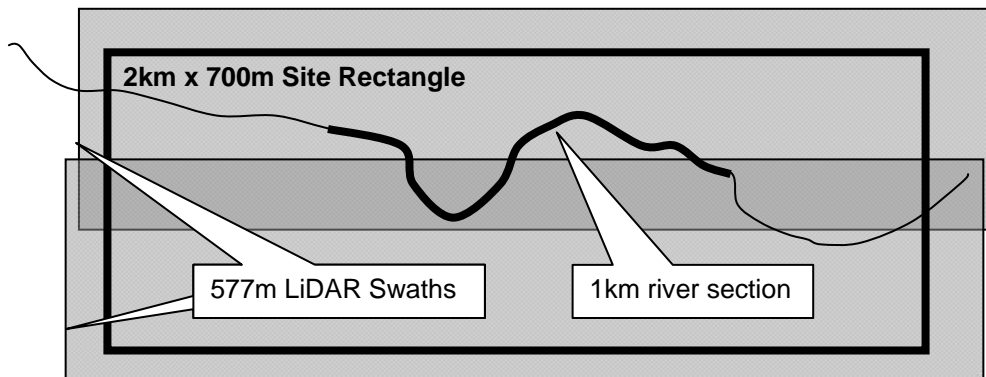


Figure 1. Each 2000 m x 700 m site is covered by two 577 m wide LiDAR swaths with 35% overlap. A 1km section of river within each site is analysed.

The aerial surveys were constrained to ensure that data was not captured when water was overflowing the river channels and the LiDAR and Vexcel imagery were captured within two weeks of each other. These constraints and the need to survey field check points ahead of the aerial survey introduced some logistical complexities that were exacerbated by extensive flooding that occurred early in the project (December 2009 through March 2010).

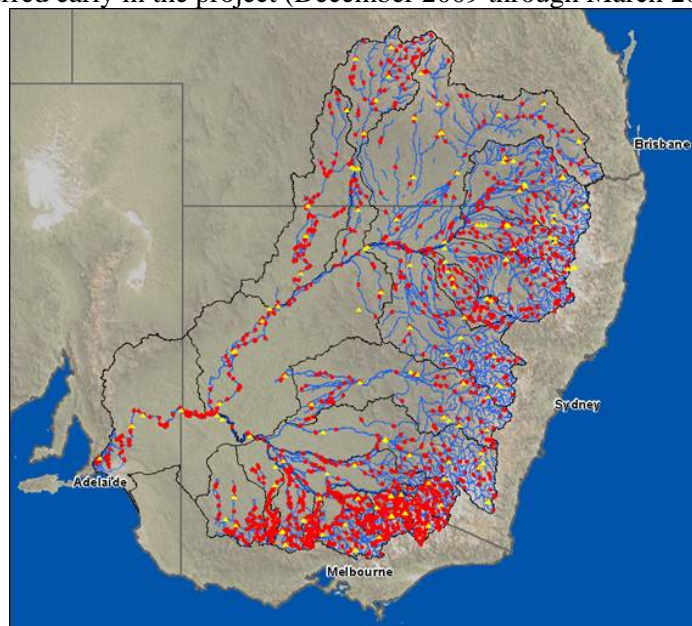


Figure 2. Channel sites (red) and field check sites (yellow triangles) overlaid on the 23 major river valleys that make up the Murray-Darling Basin.

The LiDAR was recorded using a Trimble Harrier 56 and 68 LiDAR instruments. Both systems were operated with the following parameters:

- Flying Height 500 m above ground.
- Flying Speed: 205 km/hr
- Scanning Angle 60 degrees
- Overlap: 35%
- Swath Width: 577 metres
- Scan Rate: 76 Hz
- Pulse Rate: 200 kHz
- Point Spacing:
 - Along Track 0.5 m
 - Across Track 0.5 m
- Ratio: 1 : 1 (along track : across track)
- Capture Point Density: 4.05 per square metre within swath
- Average Point Density: 6.89 per square metre
- Spot Footprint: 0.25 m

2.2 Primary Processing

Standard LiDAR processing techniques were employed to generate point clouds from the instrument data, combine the LiDAR strips for each site, adjust to control and classify the LiDAR points into the classes: 1) Ground / Water, 2) Vegetation and 3) Built Structures.

An automated batch process was developed in the object oriented TNTmips Spatial Modelling Language (SML) to produce the primary datasets (rasters and contours) from the classified LiDAR points. TNTmips is well suited to this task due to its ability to integrate LiDAR LAS files with GIS data (ESRI shape & Grid) and image formats (GeoTIFF, JP2000). A single tool developed in TNTmips automatically sorts batches of LAS files associating each file spatially with a site and assigns a name based on the Valley, Site and capture date. The same tool then generates all the raster surfaces, foliage density layers, contours and other primary datasets, then outputs these to a variety of formats and re-imports them for validation. The TNTmips job queue manager proved highly effective in managing and optimising the computationally intensive processing across multiple networked computers to produce the large number of output files for all 1610 sites. The single batch process produced the following primary datasets from the classified LAS files:

Table 1. The primary data sets.

No.	Name	Description	format
1a	Raw LiDAR	One file per swath, named according to Valley, Site ID, capture date, sortie, and UTM zone	LAS
1b	Classified LiDAR	One file per site, trimmed to site polygon, named according to Valley, site, zone.	LAS
2	Ground Points	Ground points without built structures	ascii text
3	Ground + Building Points	Ground points with built structures	ascii text
4	Vegetation Points	Vegetation Points	ascii text
5	DEM	Raster terrain surface without built structures – 1 metre grid spacing	Arc Grid
5	DTM	Raster terrain surface with built structures – 1 metre grid spacing	Arc Grid

7	CEM	Vegetation height above ground – 1 metre grid spacing	Arc Grid
8	PLR	The percentage of LiDAR returns within 17 height ranges above the ground: <ul style="list-style-type: none"> • 0.0 – 0.1 m • 0.1 – 0.5 m • 0.5 – 1.0 m • 1 – 2 m • 2 – 5 m • 5 – 10 m • 5 – 12 m • 10 – 20 m • 20 – 35 m • > 2 m • > 3 m • > 5 m • >10 m • > 12 m • > 20 m • > 35 m 	Arc Grid
9	Contour	0.25 m interval contour	ESRI Shape
10	AOI	Site rectangle polygon extracted from GIS layer of all site rectangles	ESRI Shape

The TNTmips SML script also generated ANZLIC compliant metadata, obtaining information, such as capture date and extent of coverage, from the input data and other GIS layers.

2.3 Secondary Processing

The primary datasets were used as the inputs to secondary processes that extracted river channel features. Two steps were involved: 1) mapping of water surfaces in the river channel, and 2) extraction of channel features and site zones.

A ‘Variable Extraction Toolkit’ was developed, again using the TNTmips object oriented Spatial Modelling Language, to map channel and riparian features (secondary datasets) and also measure channel and riparian vegetation attributes (measurement datasets). The Variable Extraction Toolkit has a graphical user interface, and contains a number of interactive and automated feature extraction tools organised into a simple workflow.

The first step of the feature extraction process is to map the channel centreline. The approximate path of the river channel is traced quickly over the relief shaded DEM. The variable extraction tool then maps the channel centreline as a line of best fit through the lowest part of the channel. Where the channel contains water, the centreline passes equidistantly through the centre of each water body. The channel centre line is automatically trimmed to a length of one kilometre.

Flow direction is determined visually by inspection of the longitudinal profile of the channel centreline and if necessary by reference to a basin-wide drainage network. The channel must be oriented correctly in order to assign the left and right banks relative to the flow direction.

Nineteen transects are semi-randomly generated at right angles to the channel centreline. On each transect, points representing the left and right top banks and left and right bottom banks are generated. The top bank points are initially located at the first Riley Bench Index maxima. The Riley Bench Index is calculated as the change in channel width over the change in channel slope. Maxima in the Riley Index occur on the inside edge of horizontal surfaces (Pickup, 1976).

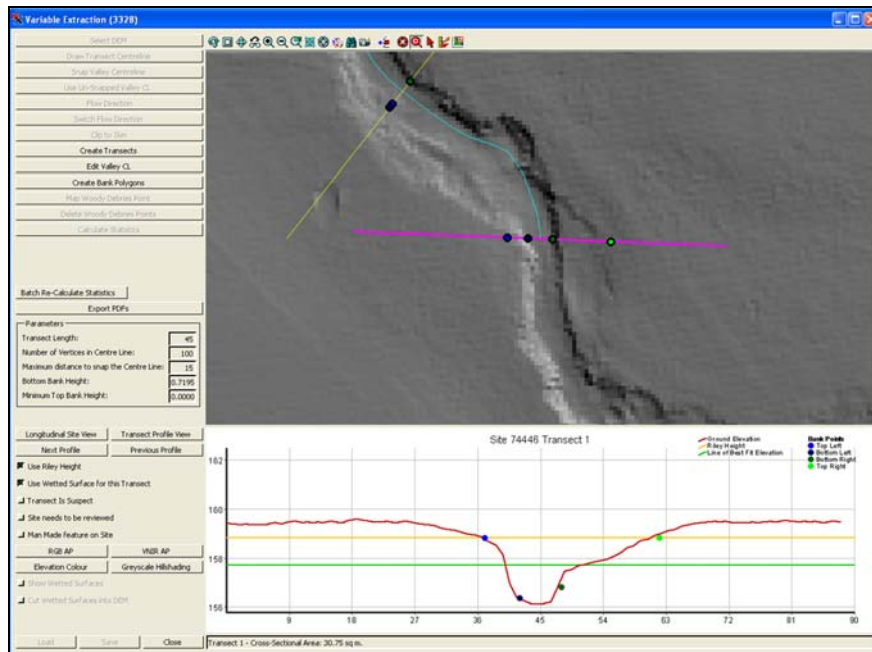


Figure 3. Transects and top-bank bottom-bank points generated in the Variable Extraction Tool.

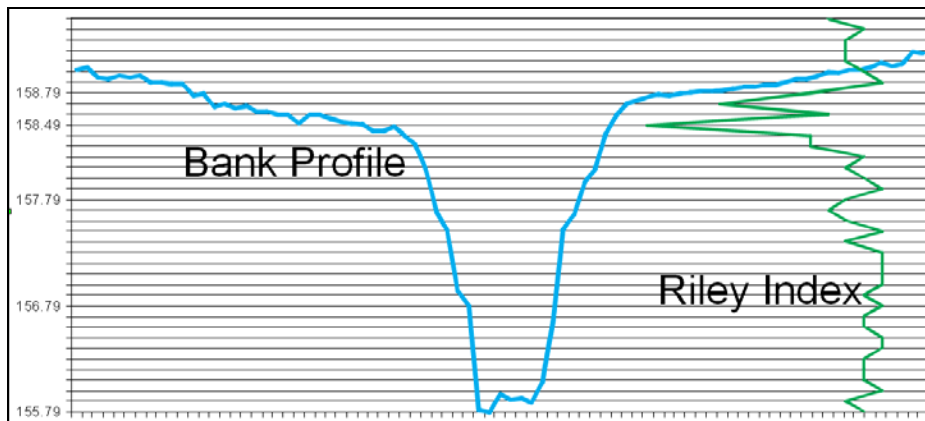


Figure 4. Riley Index plotted with bank profile.

The aim is to find the top and bottom of the active channel, which must be distinguished from minor channels within the active channel and wider incised channels. It was necessary to provide sufficient flexibility to enable the operator to guide the program to choose the correct channel, while minimising the subjectivity of the process. This was achieved by allowing only three discretionary inputs: 1) choosing either the Riley Bench Index maxima or the mean bank height for the site, 2) specifying a minimum bank height for all the transects on a site and 3) discarding transects that are unsuitable for the analysis (e.g. transects on confluences).

Operators can view the channel in 3D using anaglyph glasses, switch between two different visualisations of the DEM surface and the false colour and natural colour Vexcel images of the site. Thus information, such as the presence of exposed sediments and vegetation, is taken into account when assessing the bank lines. Outliers are evident when bank lines are viewed in profile with the channel centre line and the mean height of the top bank above the channel centreline. As the operator adjusts the settings for each transect, the top bank profile generally converges with the mean bank height.

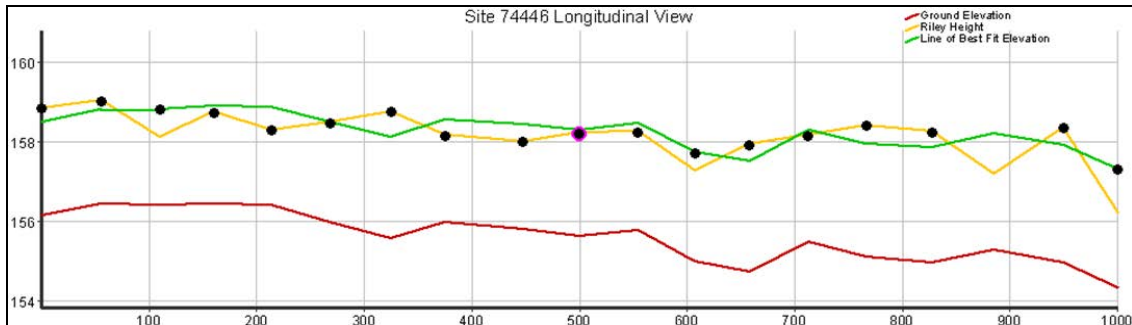


Figure 5. Longitudinal profile of a channel site, showing the Riley Index (yellow), The mean bank height (green) and the channel centreline (red).

After the bottom bank and top bank points have been generated for each of the 19 transects, top and bottom bank lines are interpolated between the points using a path following algorithm that applies, in order of priority, the following criteria:

1. Minimise change of slope along the bank line
2. Minimise the length of the bank line
3. Follow the zone of maximum terrain surface inflexion.

The accuracy of the bank lines is less critical for the purposes of this project, than the accurate placement of bank points on transects. This is because the physical form measurements apply directly to the transects, while the bank lines are used only for segmenting the site into channel, banks and floodplains for the purpose of measuring vegetation structure in these 'Site Zones'. Therefore no interactive control was provided for the mapping of bank lines.

After generating the bank lines, the site is segmented into the Site Zones listed in Table 2 and illustrated in Figure 6.

Table 2. Site zone polygons for vegetation measurements

Zone	Description
Left bed	Between channel centreline and left bottom bank
Right bed	Between channel centreline and right bottom bank
Left bank	Between left bottom bank and left top bank
Right bank	Between right bottom bank and right top bank
25LB / 25RB	25 metre buffer from left / right top bank
50LB / RB	50 metre buffer from top bank to 25 metre buffer from top bank
50LP / RP	From 50 metre buffer to site boundary

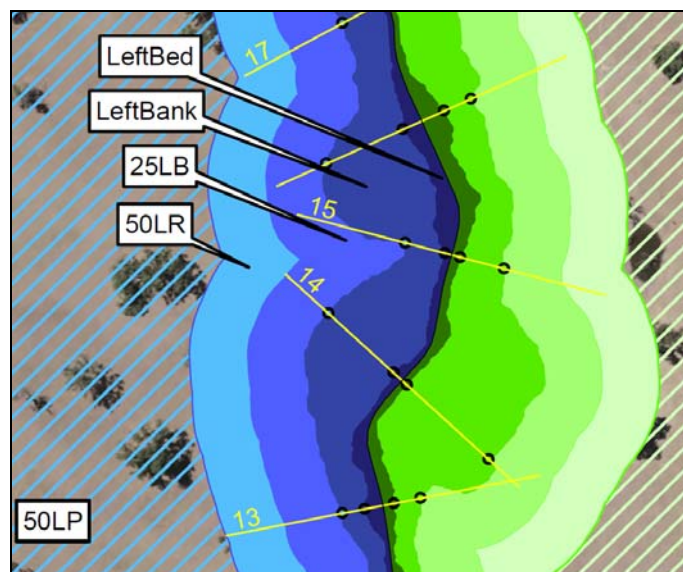


Figure 6. Site polygons segmenting channel and adjacent riparian zone.

The secondary data sets produced by these methods are listed in Table 3. Each data set includes ANZLIC compliant metadata.

Table 3. Site zone polygons for vegetation measurements

No.	Name	Description	Format
1	ChannelCL	Computed channel centreline -lowest path or mid path through water.	Shape + TNTmips
2	Transects	Maximum 19 transects for a site	Shape + TNTmips
3	BankPts	Left and right top and bottom points for all transects on a site	Shape + TNTmips
4	BankLines	Interpolated left and right, top and bottom bank lines for a site.	Shape + TNTmips
5	BankPolygons	Polygons generated from bank lines and perpendicular end line.	Shape + TNTmips
6	SiteZones	Site zone polygons described in table 2.	Shape + TNTmips

2.4 Data Measurement

The final data generated for the project consist of two data matrices: 1) a Vegetation measurement matrix, and 2) a Physical Form measurement matrix. These matrices provided the basis for statistical analyses and modelling that were performed in subsequent phases of the project.

The spatial units for the vegetation measurements were generated by intersecting the Site Zone polygons with vegetation polygons a pre-1750 vegetation map that had been compiled for the MDBA for this purpose. For each resulting polygon, representing the intersection of a site zone polygon with a vegetation polygon, the mean of the Percentage LiDAR Returns for each of the 17 strata listed in Table 1 are calculated as polygon attributes. These data represent the vertical distribution and density of foliage by vegetation type across the zones associated with each river channel.

Intersecting the vegetation polygons with the Site Zone polygons, calculating the mean PLR for each stratum as polygon attributes and writing the results to a data matrix in CSV format is performed as a single process within the Variable Extraction Tool.

Similarly, a range of Physical Form measurements is generated from the transect profiles, bank points and channel centreline. These attributes, listed in Table 4, are calculated as attributes of the GIS line features and are exported to a data matrix for each site as an automated process.

Table 4. Physical Form Measurements

Input Variable	Description
ChnlCent	Channel centreline; length in metres
ValCent	Valley centre; length in metres
MaxElev	Maximum elevation of channel centreline
MinElev	Minimum elevation of channel centreline
ElevDiff	Elevation Range of the channel centreline
TranMin	Minimum elevation of transect
TBnkAng(L&R)	Left and right bank top angles
TBnkHt(L&R)	Left and right bank top height
BBnkHt(L&R)	Left and right bank bottom height
TBnkArea	Cross sectional area below top bank level
AngLn(L&R)	Length of profile from bottom bank to top bank – left and right
SAngLnR	Length of profile from spill height to bottom bank
Convex(L&R)	Cross sectional area of the bank profile above the bank angle line.
Concave(L&R)	Cross sectional area of the bank profile below bank angle line.
Inflect(L&R)	Number of times bank lines crosses bank angle line
Length(L&R)	Length of bank lines for the site – left and right banks
WetBed	Does transect intersect water? Yes or No
WetWdth	Width of water at transect from wetted area layer
BedWdth	Distance between left and right bottom banks
ChnlWdth	Distance between left and right top banks
ChnlDpth	Channel Depth; height of spill level above bottom bank
AreaChan	Channel area – between left and right top banks
AreaBnk(L&R)	Area between top and bottom banks for the site – left and right
StrmPow	Stream Power; (max elev minus min elev) / valley centreline length
ManMade	Evidence of manmade channel or features, or water control

3 Discussion

Full waveform LiDAR was found to be a very effective source of data for taking detailed and precise measurements of the river channels and riparian vegetation. The minimum density of four outgoing laser pulses per square metre produced a ground point density of more than one point per square metre, even in complex vegetation, generating extremely accurate and detailed ground surface models. The high pulse density, full waveform LiDAR, returned up to 20 points per square metre in areas of complex vegetation enabling the distribution of foliage to be mapped in three dimensions with a high level of detail.

The TNTmips software is an extremely effective geospatial modelling environment for establishing automated workflows to produce consistent and repeatable GIS output data from the classified LiDAR LAS files. The TNTmips system includes a Job Queue Manager which greatly facilitated management of the data processing by optimising the use of networked multiple core computers. It also provided detailed information on the status of the various tasks that had to be performed in sequence for more than 1610 individual sites. TNTmips also proved to be a highly effective geospatial modelling environment for the development of the Variable Extraction Tool, due to its ability to integrate a wide range of data types and formats, including LiDAR LAS files, and its advanced geospatial analysis functions.

The data from this project were generated for statistical analysis in what is intended to be a long term program for monitoring the health of the MDB rivers. Consequently there is a requirement for methodologies that minimise subjectivity and bias by eliminating user interaction as much as practical. Variable extraction was performed entirely by GIS operators with no training or experience in geomorphology. Throughout development and implementation of the variable extraction methodology, expert geomorphologists assessed the quality of the data and provided advice on opportunities for improvement of the algorithms and work flows. Protocols were established for obtaining advice from professional geomorphologists where decisions could not be made with confidence by the GIS operators. The primary skill the GIS operators needed to learn in order to perform the variable extraction was the ability to distinguish alluvial systems from other types of river sections that do not apply to this project. Similarly, confluences and built structures such as bridges and dam walls can make transects or entire sites unsuitable and the operators were trained to recognise these and flag them for discarding or checking.

The methodology and algorithms are designed for alluvial systems and do not operate effectively on river systems that are flat, topographically constrained or contain large boulders. Such river sections are difficult sites from which to extract useful information for both geomorphologists and automated methods. For this reason, the project was designed with significant sampling redundancy; of the 70 sites surveyed in each valley, only 50 are required for the subsequent analysis based on a biostatistical power analysis.

An independent assessment of the data by professional geomorphologists reported that on less than 10 percent of sites, bank lines were not correctly assigned to the active channel and between 16 and 35 percent of sites have minor errors in the alignment of bank lines which will not adversely affect the site measurements. As discussed above, the interpolation of bank lines is less critical than the correct assignment of bank points on a profile. The results suggest that between 58 and 75 percent of sites were mapped as would be expected from an experienced geomorphologist. It was noted that the manual identification of active channels and delineation of bank lines is a subjective process and that different geomorphologists would be likely to produce different results. Regardless of the extent to which the bank lines mapped by the methods described here are consistent with the definitions generally accepted by geomorphologists, the fact that they are generated through a repeatable and objective process lends this data to statistical analysis.

The Percentage LiDAR Returns (PLR) was calculated as a surrogate measure of relative foliage density. PLR was defined as the percentage of total LiDAR points from an area, reflected from within each specified stratum. It was found that this is not the optimal method for calculating relative foliage density as explained below.

1. The number of LiDAR points returned from within a vegetation stratum for a specified area will be determined by the number of LiDAR pulses emitted into the specified area. The number of points recorded from all stratum is proportional to the emitted pulse density, so this will not affect the PLR for any stratum.

2. The density of foliage intercepting and reflecting laser pulses from within that stratum across the specified area. Increased foliage density will tend to increase the reflected signal and hence the number points recorded and the PLR.
3. Attenuation of the laser signal energy through foliage above the specified strata. This attenuation reduces the number of points detected from lower strata and hence the calculated PLR of lower strata is reduced by the foliage of higher strata. This causes compounding between the PLR of different strata and therefore there is not a simple relationship between PLR and foliage density. Further work is required to disentangle the interactions between strata.
4. Compounding this effect, an increase in foliage density within vegetation above and below the stratum being measured, will increase the total number of points returned from the subject area. This has the effect of reducing the percentage of LiDAR points from that stratum and hence the calculated PLR. This effect is more easily overcome than the previous, simply by calculating PLR as a percentage of the total emitted laser pulses falling on an area rather than as a percentage of the total reflected points. Total Emitted Pulses is independent of foliage and is closely approximated by the number of 1st echo points recorded. Subsequent investigation suggests that this is a more robust measure of foliage density.

This project has shown that full waveform LIDAR is intensely information rich and provides a basis for measuring the Physical Form of the Basin river channels and associated vegetation with a higher degree of rigour and repeatability that has previously been possible. Future monitoring and evaluation of the Murray-Darling Basin would most likely combine remotely sensed data with field calibration and coincident collection of field data, such as channel pebble size, channel depth where water is present, and also vegetation recruitment.

4 References

- Davies, P.E., Harris, J.H., Hillman, T.J., and Walker, K.F., 2008, SRA Report1: A Report on the Ecological Health of Rivers in the Murray-Darling Basin, 2004-2007. Prepared by the Independent Sustainable Rivers Audit Group for the Murray-Darling Basin Ministerial Council.
- Davies, P.E., Harris, J.H., Hillman, T.J., and Walker, K.F., 2010, The Sustainable Rivers Audit: assessing ecosystem health in the Murray-Darling Basin, Australia. *Marine and Freshwater Research*, 61, 764-777.
- Johansen, K., Arroyo, L.A., Ashcraft, E., Hewson, M., and Phinn, S., 2009, Assessing the potential for remotely sensed statewide mapping of streamside zone and physical form metrics in Victoria, Australia, 2 April 2009. Prepared for Paul Wilson, Department of Sustainability and Environment, State of Victoria.
- Pickup G., 1976, Alternative Measures of River Channel Shape and their Significance. *Journal of Hydrology (N.Z.)*, 15, 9-16

Gearing Toward the Potential of LIDAR Application in Malaysian Forestry

Mohd Hasmadi Ismail¹ and Mohamad Sam Manaf²

¹Forest Surveying and Engineering Lab., Faculty of Forestry
Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
email: mhasmadi@putra.upm.edu.my

²RS&GIS Consultancy Sdn. Bhd.
J-3A-12, Solaris Mont Kiara, Jalan Solaris, 50480 Mont Kiara, Kuala Lumpur, Malaysia
email: sam@resgis.com.

Abstract

Information on forest properties have grown over time and will continue crucially in the future. The focus on timber for commercial trade in early 1960's in Malaysia has been changed towards multi function forestry, supported by multi resources survey. Starting with high demand of the latest data and accurate information, and cost effective monitoring system, application of various technology of sensing system is applied into forestry. The introduction of precision forestry concept is not new but in Malaysia is still at infancy stage. It deals with advanced sensing technologies and analytical tools to support site-specific economic, environmental, and sustainable decision making for the forest management and development. The key discipline is highly relying on accurate, timely and detailed forest inventory characterization and structural information. This is possible by utilization of accurate measurement forestry data and information to improve operations and processes. Despite of the current use of high resolution satellite and airborne sensing, LiDAR is a promising alternative tool to be used in forestry sector. LiDAR can be used in forest engineering for terrain mapping and road planning, and tree/stand measurement for tropical forest. This paper gives a synopsis of LiDAR sensing technology application and its potential to Malaysian forestry.

Key words: precision forestry, LiDAR, forest management.

1. Introduction

The adoption of precision forestry concept is not new but in Malaysia is still at infancy stage. It deals with advanced sensing technologies and analytical tools to support site-specific economic, environmental, and sustainable decision making for forest management and development (Mohd Hasmadi *et al.*, 2007). Precision forest management is a business of geoinformatics technology. The use of geoinformatics for precision forest management has become popular in developing country but at different stages. Precision

forestry leverages advanced technology and tools to support sustainable decision making, where it is highly reliant on accurate, timely, repeatable, detailed and spatially explicit forest inventory characterizations and structural information (Jason *et al.*, 2002). Forest applications that are based solely, mainly or partly on remote sensing technology are structured by data type. Beginning with basic to in-depth description of the user requirements lead to an increased demand for up to date, reliable and applicable on forest information for effective monitoring system. Forest measurement through traditional surveying has gained the development of new measurement tools. The rapid development of the image processing has made possible to combine digital data so that the aerial or even satellite imagery can be used to find the tops of the trees, estimate tree lengths, dbh sizes and thus volumes. The s of height measurement and vertical structure of forest structure opens a wide application in the forest ecosystem especially in combination with other optical data. In addition images does not necessarily be taken from the spaceborne or airborne. The different tree species , their location, characteristic , etc can be identifird and analysed by using radar scanning tecnology such as LiDAR.

LiDAR is an innovative technology that assists the assessment of forest conditions as well as establishing a viable approach to long-term monitoring to support the management of natural forests. The LiDAR ability is rapidly collect highly accurate three-dimensional information of the forest and it ecosystem and make a significant impact in overcoming the challenges faced by government agencies and non-governmental and organizations (NGOs) to solve new challenge in reducing emissions from deforestation and degradation initiatives. Meanwhile image processing software and techniques supports the advance data, providing greater analytical capabilities, thus improved knowledge, than was previously possible. LiDAR have been effectively demonstrated and reported in precision forestry applications, including forest height inventory assessment (Andersen *et al.*, 2006), multiple resource inventory (Reutebuch *et al.* 2005), ecosystem studies (Lefsky *et al.*, 2002) and stand value estimates (Murphy 2008). The ability of LiDAR is expanding for the assessment of ecosystem services (Richardson *et al.*, 2009) and biomass estimation (Popescue *et al.*, 2004). This paper gives a synopsis of LiDAR sensing technology application and its potential to Malaysian forestry. The current state of application and challenges toward possible direction of forest resources management also highlighted. This demonstrates the use and need of future precision forestry application of LiDAR technologies. Thus we believe that LiDAR technology integrated with geoinformatic procedures and efficient field sampling technique could provide a fundamental data for ecologically, socially and economically sustainable forest management in the future.

2. Information needs in precision forestry

Forest information requirements are increased over time and will continue in the future. Information is changing where the data is not just as before most of the focus is on the timber resources for commercial use. In modern forest management the society has move from an economic focus toward multi-purpose or multi-functional forestry. In order to support the spirit of multi-functional forest multi-resources, an inventory is essential. Precision data reflects more attention by the society and policy for understanding the forest as a complex ecosystem that supporting country's development and human life. Precision forestry uses high technology sensing and analytical tools or software to support site specific assessment. Adequate quality information derived from precise data helps in maximizing economic return for good decision making. Any information gained from surveying processes should provide sufficient data to support biodiversity assessment and other environmental resources including maintaining the quality of the environment. Precision forestry is also refers to site specific management. Site specific forest management may be refers to precise information of the tree stand such as measured tree volume from ground measurement and correlates with remote sensing data and GIS technique and/or and may be soil information of the specific forest land. The combination of field checking, aerial photo interpretation/remotely sensed data and data interpretation will refine details for the planning process.

In precision forestry the quality of information is depend to the quality of data acquisition. Nohr (2001) stated that information quality can be defined as the sum of all requirements expected from information (data) in order to fulfil specific information needs. Olson (2003) defined data quality as two consentient aspects: first, the dependence of perceived quality on the user's needs; second, the so-called "fitness for use", which is the ability to satisfy the requirements of intended use in a specific situation. Among the most important criteria about the quality of information are:

Accuracy: Accurate information describes properties or the state of relevant objects according to the reality. If the entity can be described with measurable variables the degree of accuracy can be described easily.

Reliability: The user of information should be convinced that any information available is correct in the widest sense. Even if a high accuracy is given for any variable it might be less reliable because of the measurement procedure used.

Relevancy and process orientation: Information of high quality should meet objectively given information needs. Although this criterion seems to be quite understandable, it is one of the more difficult

ones. In particular in forest management decision making is often based on individual approaches of information use.

Timeliness: Information that is not available in time is useless and therefore of very low quality.

Completeness: Incomplete information which misses several and crucial parts may be misleading. Normally decisions are based on complex sets of information which add up to a comprehensive picture of the situation to be considered.

Presentation: This criterion deals with the fact that information needs to be presented in a suitable manner. As information needs to be interpreted in order to prepare decisions appropriate presentation is an important quality issue. Although the criteria mentioned above are not listed according to a hierarchy, they will be of different importance in different enterprises.

The types of tools applied in precision forestry are varies. For measurement and monitoring tools; LiDAR, remotely sensed data can be used to develop highly forest canopy characteristic, digital elevation model (DEM) and digital surface model (DSM) which are useful to determine stream lines and topography under canopy. In electronic mapping the global positioning system (GPS) and inertial navigation system (INS) were used for navigation under forest canopy. Then GIS is common software which uses for data management and recently web based data management also play a role in huge data storage and management. Next the decision support system (DSS) takes place by optimization of spatial data and simulates the scenario. The capability of ICT based operation research techniques in supporting DSS were proven by the simulation and network analysis (to name a few). The multi criteria decision analysis (MCDA) is widely used and recognized as a solution for the DSS in decision making process.

3. LiDAR at a glance

LiDAR stands for Light Detection and Ranging and is very similar to the radar. LiDAR is a better choice than radar because it has a greater ability to reflect images, making more objects visible. The principle of LiDAR system is a laser ranging. According to Young (1986) a high directional optical light could be created with laser process, thus yielding the high collimation and high optical power required for ranging. The advantages of laser was it demonstrated the high energy pulse that can be realized in short intervals and short wavelength light which it can be highly collimated using small aperture. Laser light has a much shorter wavelength that make it possible to accurately measure much smaller object such as cloud particle and creates wave form. The ability of laser sensing the points over the path or swath define by the

instruments scan angle and altitude of the platform. Equipped with receiver and scanning system the distribution of points cloud using previous profiling system were scanned to the along track path of the aircraft or vehicles. The use of GPS with LiDAR is only deployed in the 1980s to allow precise positioning of aircraft. LiDAR is categorised as an “active” remote sensing because the sensor both emits and records the radiation signal in the form of frequent, short-duration laser which is not record radiation reflected by the surface from a source external (such as sun) to the sensor. In general LiDAR technology has application in geomatics, archaeology, geography, geology, geomorphology, seismology, forestry, remote sensing and atmospheric physics (Cracknell and Hayes, 2006).

As comparison LiDAR uses waves ten to one hundred thousand times shorter than radar waves. This is why LiDAR are able to collect much more information. The elapsed time from when a laser is emitted from a sensor and intercepts an object can be measured using either pulsed ranging or continuous wave. Pulsed ranging is where the travel time of a laser pulse from a sensor to a target object is recorded, meanwhile continuous wave is where the phase change in a transmitted sinusoidal signal produced by a continuously emitting laser is converted into travel time (Wehr and Lohr, 1999). The development of LiDAR technology is parallel through applications of GPS and inertia navigation system (INS) also referred to as inertial measurement units (IMU). LiDAR technology for terrestrial applications differ in (1) whether they record the range to the first return, last return, multiple returns, or fully digitize the return signal, (2) footprint size (from a few centimeters to tens of meters), and (3) sampling rate/scanning pattern.

LiDAR has some advantages. The advantages of LiDAR are able to carry out direct sampling for measuring vertical and horizontal of forest structure, flexibility in operation (day and night capture), fast delivery times. High resolution and accuracy (tree height measurement and superior DEM generation), ease of GIS integration, cost effective for large scale project and provide wall-to-wall coverage. Other advantages are ability to define terrain under vegetation, typical data vertical accuracy of 0.15m rms, typical point spacing of 1m, and acquiring over 500 million data points per hour. Most commercial airborne LiDAR systems are low-flying, small-footprint (5–30 cm diameter), high pulse rate systems (1000–10,000 Hz). In addition, most commercial LiDAR systems record the range to the highest, and/or lowest, reflecting surface within the footprint, and are not fully imaging, using instead many laser returns in close proximity to each other to recreate a surface.

5. Overview of LiDAR applications in Malaysia

Malaysia is aware the emerging of LiDAR in market to be applied in various sector that can benefit users in managing natural resources. In MAPMalaysia conference on April 2011 in Sabah the LiDAR session Emerging Trends, LiDAR and 3D was attracted by many attendances from delegates. Successful implementation of precision forestry depends on numerous factors, including the extent to which conditions within a field are known and manage, the adequacy of input recommendation and the degree of application. The enabling technologies of precision forestry can be integration of five major components: remote sensing, GIS, global positioning system (GPS), computers and application.

Malaysia is developing country where any development of the forestland for other purposes is subject to environmental impact assessment report and guidelines by the government. The use of LiDAR for surveying and monitoring residential development project was tested in Bukit Tinggi area in Pahang, Malaysia. The project was carried out by the RS & GIS Consultancy Sdn. Bhd. The aim of the project beside to develop the area for low density residential, it also maintain the greenness of the area by minimizing the environmental impact. The big trees were conserved and very minimum slope cutting was done. The rationale is to maintain the ecosystem of the area. Figure 1 showing LiDAR “point clouds” of residential home at nearby Bukit Tinggi, Pahang, Malaysia surrounded by natural forest. Apart from the LiDAR terrain survey, this company also offers advanced geospatial support services to develop the necessary information layers to meet the landscape design and planning initiatives in the area of interest. Besides generating high-resolution Digital Terrain Models (DTM), they also offer to develop detailed contour and slope classification layers, support vital analysis of drainage, road, and culvert design, and analyze layouts and view-sheds of various development options to be explored. These are important to meets the goal in biodiversity conservation, enhancement and sustainable development.

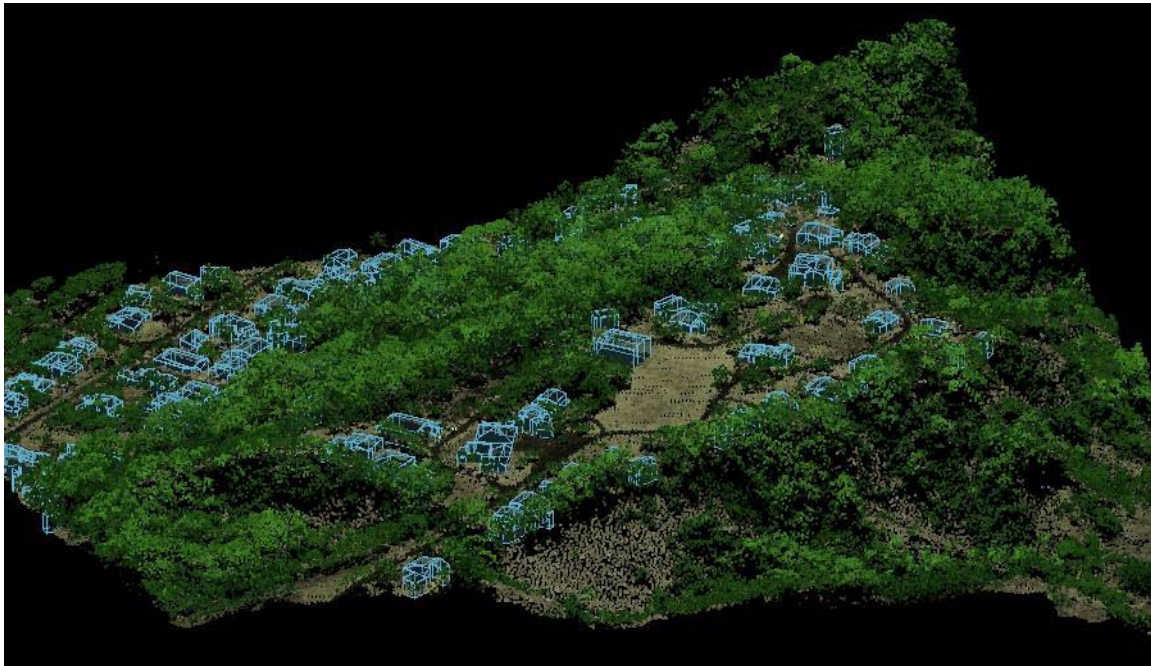


Figure 1: LiDAR “point clouds” of residential home at Bukit Tinggi, Pahang, Malaysia surrounded by natural forest.

In Kuala Lumpur city an 88-storeys PETRONAS Twin Towers (PTT) was completed in 1998. The PTT is an integral part of the Kuala Lumpur City Centre (KLCC) project. The PTT became the tallest buildings in the world on the date of completion (Sebestyen, 1998). Although PTT is surpassed by the Taipei 101 in 2004, but it still remain as the tallest “Twin building’ in the world (Palmer, 2008). It rises to 451.9 m in height and surrounded by public park namely KLCC park which spanned about 10 ha. below the building. Figure 2 shows the KLCC with twin tower by using LiDAR data. The color coded data is capable to portray the scenes with amenities available such as jogging and walking paths, a fountain with incorporated light show, wading pools, and a children's playground.



Figure 2: LIDAR color coded point cloud of the KLCC showing world's tallest twin building in Kuala Lumpur (451.9 m).

Other experience was a collaboration study between Universiti Putra Malaysia (Faculty of Engineering and Faculty of Forestry with Limitless Company in Dubai). The aim of the study was to monitor and locating tree in city of Dubai along the overhead transmission cables and cut them whenever trees are overcrossing the cables. A Limitless Dynamic Laser Scanning System owned by Limitless LCC, a Dubai World Company was used for data acquisition (Figure 3). The system structure is basically equipped with laser scanner mounted on top of land cruiser and connected to GPS for positional floating navigation and IMU (Inertial Measurement Unit) to track the system orientation movements. The integration of GPS and IMU unit are the bases in developing dynamic laser scanning in providing an accurate location and orientation of the scanned features (Mahmoud *et al.*, 2010).



Figure 3: Dubai World, Limitless Dynamic Laser Scanning System equipped with IMU and GPS.

Two dynamic laser scanning missions were conducted in two locations (Figure 4). The first mission conducted in Dubai, Limitless LLC, Zone one, where the positional extent of the study area within Latitude $58^{\circ} 24' 44''\text{N}$ to $58^{\circ} 24' 41''\text{N}$ and Longitude $55^{\circ} 05' 34''\text{E}$ to $55^{\circ} 05' 36''\text{E}$. The vehicle speed was set between 10 to 15 km/h. This range is adequate to collect the the intensity of the point cloud that could reflects the shape and location of the trees, although slowing down the vehicle speed would slightly enhance the data. The data were procesed to positioning the accurate location, filtration process to remove unwanted point clouds. The ground observation was conducte and subjected to post processing . The surveyed laser scanning data, post processing and LAS file generation were conducted using the Applanix equipments and the related POSpac software, where the color coded 3D surface generation was conducted using Quick Terrain software.



Figure 4. Location of study area (Limitless LLC, Zone One)

The output of the study area is shown in Figure 5. Laser scanning system is capable in monitoring the overall city features including the city trees. The color reflecting the features elevations where cyan is reflecting the ground surface then green, yellow, orange and red. The black objects are either invisible features or it is non-reflective surface such water or asphalt pavement. This scene also show the importance of the accessibility of the concerned city trees in order to avoid any dark areas and to reflect the physical conditions. The high brightness along the road is generated due to the high reflectivity of the features which also reflects the shape resolution of the features. Results showed that the laser scanning was capable to recognize the trees height and shape/extension. The schematic map in Figure 6 imposed on the color coded image showing the selected tree height, where this case the tree height was 6.5 m and it extension was 18 m. The results are showing the feasibility of conducting the dynamic laser scanning as part of the regular overhead transmission lines maintenance. Meanwhile Figure 7 shows the tree shape and extent from second LiDAR mission of the study area.

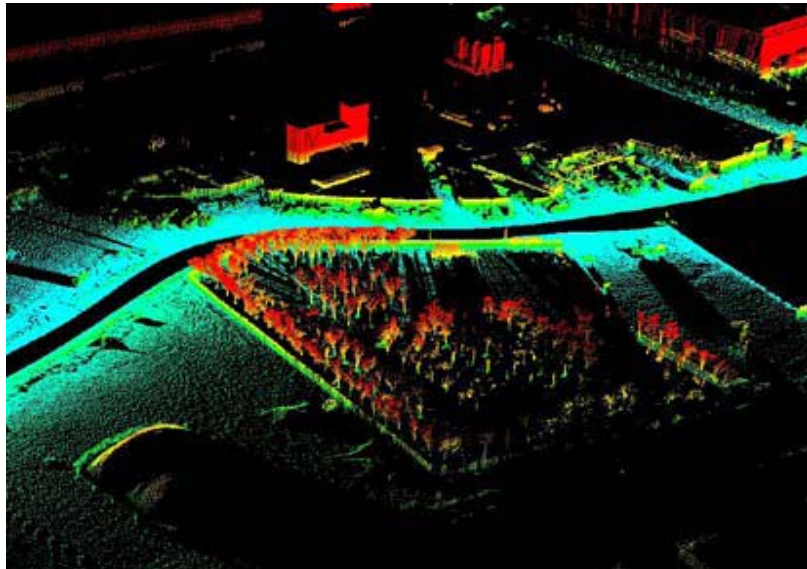


Figure 5: Dynamic LiDAR scanning outcomes of the study area

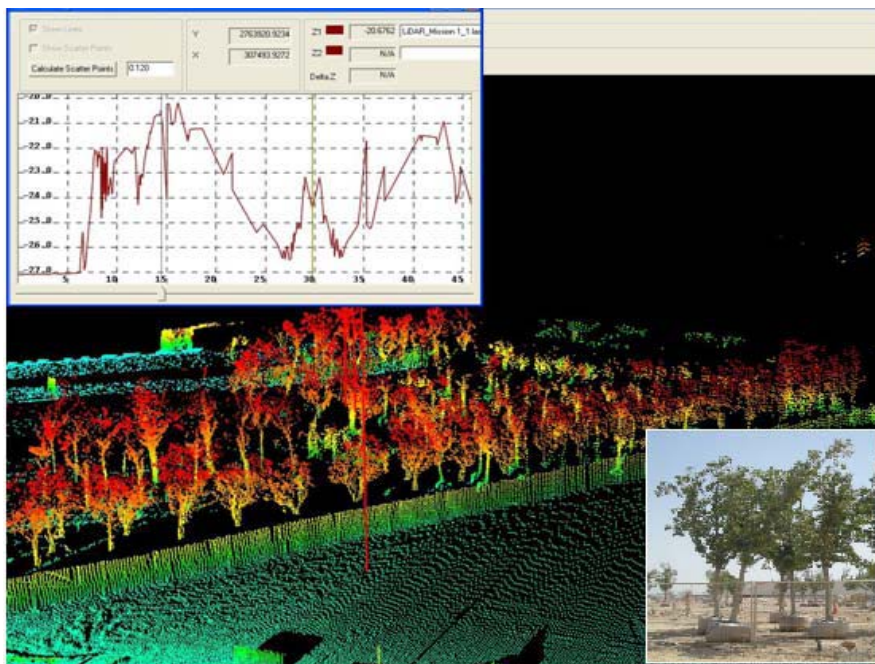


Figure 6: Calculating the trees height and shape/extension from LiDAR data

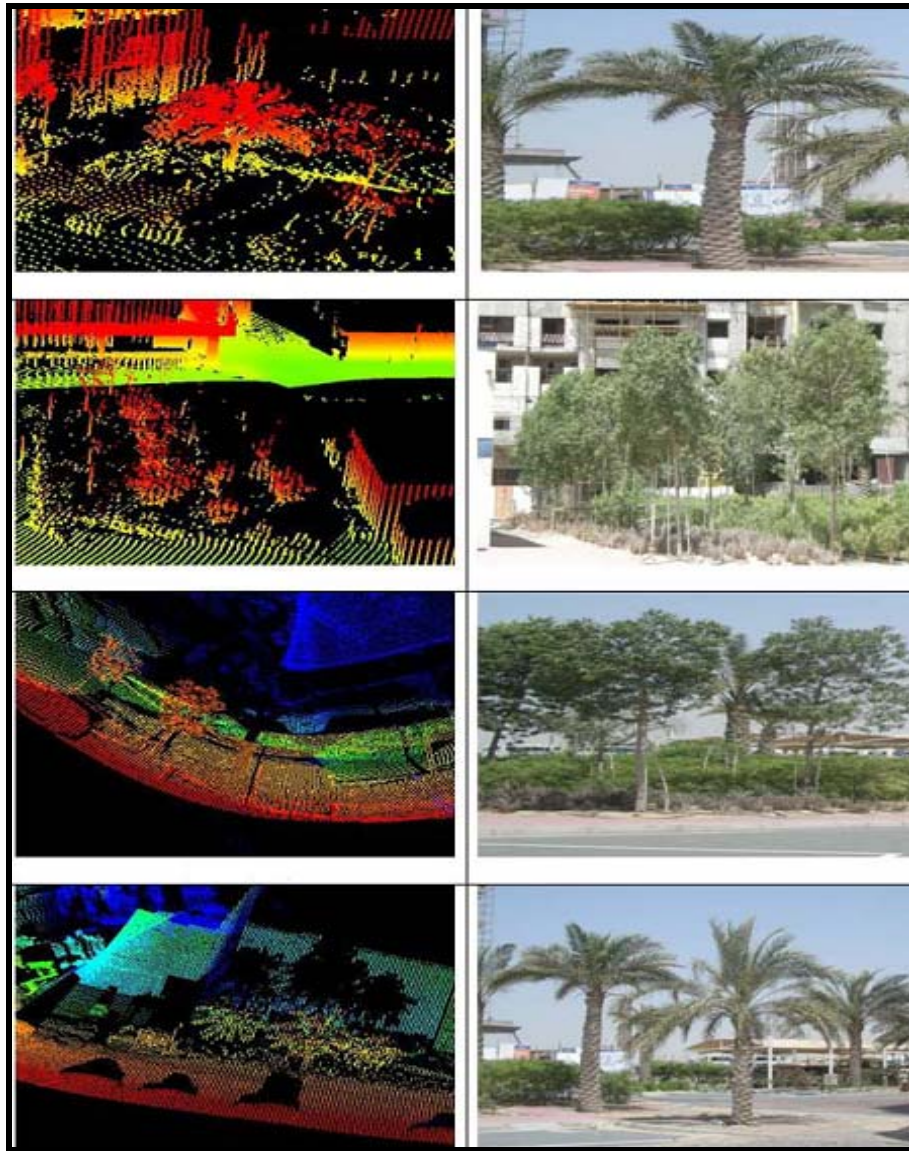


Figure 7: A color coded extracted trees locations and shapes from the second LiDAR Mission

The dynamic laser scanning is very effective in reflecting city trees location and acceptable outcome in terms of trees shapes. The positional accuracy is very encouraging in both vertical and horizontal directions. The horizontal coordinates are re-projected to the local datum in UAE-Dubai (DLTM), where the vertical coordinates are ellipsoidal heights. The dynamic laser scanning is well performed in maintaining the overhead cables from any damages caused by the trees due to the ability of calculating the trees height in sufficient accuracy (up to 10 cm). The outcomes are comparable with aerial LiDAR outcomes where the coverage is less but the accuracy is more and the operational cost is less.

6. Conclusion

There are huge potential of using LiDAR technology for precision forestry in many developing countries including Malaysia. The adoption of LiDAR technology in precision forestry also depends on product reliability, the support provides by manufacturers and the ability to show the benefits. The potential of LiDAR in forestry application in Malaysia is wide and require knowledge and techniques to process LiDAR data. However it is clear that LiDAR applications in forestry will continue to increase. We believe the theoretical understanding of the relationships that exist between forest structure and LIDAR response is still incomplete. In general, the most important where the LiDAR can play a significant role is in some of the research area such as canopy and tree height estimation, LiDAR for forest structure and biomass and volume. Other aspect of a LiDAR need to be emphasized is how much the intensity component of the laser return signal is adequate as a source information to forestry application. On the other hand, data fusion between LiDAR and other remote sensing images is becoming a topic in itself. LiDAR technology will become integrated with digital cameras and also by effective fusion techniques with photogrammetry and multispectral information. Finally, by integrating LiDAR systems with imaging sensors, more advance techniques will emerge, thereby satisfying the wide range of data requirements for forestry application at local and regional scales.

References

- Cracknell, A.P., Hayes, L., 2006. Introduction to remote sensing (2 ed.). London: Taylor and Francis, UK.
- Hunter, G., Cox, C., Kremer, J., 2006. Development of a commercial laser scanning mobile mapping system – StreetMapper. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 36; Available online at: <http://www.isprs.org/proceedings/XXXVI/1-W44/www.pegasus4europe.com/>
- Jason, B. D., Ralph, O., Dubayaha, D.B., Clark, B.C., Robert, G. K., Bryan,B.,Michelle, A. H., Robin L. C., John, F. W., Stephen, D. P., 2002. Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment*, 79, 305– 319.
- Lefsky, M.A., Cohen, W.B., Parker, G.G.,Harding, D.J., 2002. Lidar remote sensing for ecosystem studies. *Bioscience*, 52(1), 19-30.
- Mahmoud, F.A.,Ahmad,R., Mohd Hasmadi,I., Alias,M.S., 2010. Utilization of the dynamic laser scanning technology for shielding the overhead electrical cables by developing a smart monitoring for city tree progress. *Int. J. of Information Processing and Management*, 2(1), 148-159.

Paper presented at the 11th Int. Conference on LiDAR Applications for Assessing Forest Ecosystems. SilviLaser 2011, 16-20 October 2011, University of Tasmania, Hobart, Australia.

- Mohd Hasmadi, I., Kamaruzaman, J., Pakhriazad, H.Z., Frisco, N., 2007. Geoinformatics for better forest management. *Poster paper presented at the National Conference on the Management & Conservation of Forest Biodiversity in Malaysia: Forest Biodiversity for Better Life*, 20-21 March, 2007, Marriot Hotel, Putrajaya, Malaysia. 12p.
- Murphy, G.E., 2008. Determining stand value and log product yields using terrestrial lidar and optimal bucking: a case study. *Journal of Forestry*, 106(6), 317-324.
- Nohr, H., 2001, Management der Informationsqualität, Nr. 3/2001, Fachhochschule Stuttgart, Stuttgart.
- Olson, J.E, 2003. Data quality - The Accuracy Dimension, Morgan Kaufmann Publishers, San Francisco.
- Palmer, A.L., 2008. Historical dictionary of architecture: The Scarecrow Press, Inc.
- Popescu, S. C., Wynne, R. H., Scivani, J. A., 2004. Fusion of small footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, U.S.A. *Forest Science*, 50(4), 551-565.
- Reutebuch, S.E., Andersen, H.E., McGaughey, R.J., 2005. Light detection and ranging (LIDAR): An emerging tool for multiple resource inventory. *Jour. For. Sept.*, 286-292.
- Richardson, J., Moskal, L. M., Kim, S., 2009. Modeling approaches to estimate effective leaf area index from aerial discrete-return LIDAR., *Agricultural and Forest Meteorology*, 149, 1152-1160.
- Sebestyen, G., 1998. Construction: Craft to industry. London: Taylor & Francis, UK.
- Wehr, A., Lohr, U., 1999. Airborne laser scanning – an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54, 68–82.
- Young, M., 1986. Optics and lasers: Including fibers and optical waveguide. Berlin, Springer Verlag.

Laser scanning by echo signal digitization and waveform processing

Andreas Ullrich¹ & Martin Pfennigbauer²

¹RIEGL Laser Measurement Systems GmbH, Riedenburgstraße 48, 3580 Horn, Austria

aullrich@riegl.co.at

²RIEGL Research FGmbH, Riedenburgstraße 48, 3580 Horn, Austria

mpfennigbauer@rieglresearch.co.at

1. Introduction

For more than one decade LIDAR technology is widely used to acquire 3D mass data in a variety of applications. The devices used are frequently addressed as laser scanners and the acquisition of 3D data by employing this kind of LIDAR technology is known as laser scanning. Three distinctive fields of applications are usually categorized:

- terrestrial laser scanning (TLS) makes use of so-called 3D laser scanners, often mounted on tripods, performing measurements in three dimensions (ranging and two angular measurements) and are based on the time-of-flight measurement principle with either pulsed laser radiation or continuous-wave modulated laser radiation,
- airborne laser scanning (ALS), where the laser scanning device is mounted aboard any kind of airborne vehicle, e.g., fixed-wing aircrafts or rotary aircrafts,
- mobile laser scanning (MLS), where the laser scanning devices are mounted on ground-based vehicles, e.g., cars or boats.

Usually, so-called 2D laser scanners are used in ALS and MLS, where the laser beam is deflected by a scanning mechanism performing a line scan and just one scan angle per laser measurement is acquired. The line scan may produce a nearly straight line on the target's surface, but may also describe a circular line scan pattern or any other 1-dimensional curve. In order to gain again 3D data, both the ALS and MLS system have to be complemented by an integrated IMU/GNSS system (inertial measurement unit / global navigation satellite system) providing precise information on the position and orientation of the laser scanner device over time in order to transform the laser scanner data in post-processing into a geo-referenced coordinate system.

The mere point cloud, usually a huge number of points in 3D representing the accessible surfaces of the objects surveyed, is the primary data product of any scanning LIDAR in TLS, ALS or MLS applications. However, additional attributes to every point of the point cloud provide essential and valuable information on the surveyed objects, like the estimated reflectance of the target's surface at the laser wavelength.

Airborne laser scanning systems employing echo digitization and full waveform analysis (FWA) became commercially available with the *RIEGL LMS-Q560* in 2004 (Hug *et al.* 2004; Wagner *et al.* 2004; Mallet and Bretar 2009). These systems do not instantaneously provide 3D data with high precision and accuracy, as they store the digitized echo signals and scan parameters on a data recorder and the precise laser ranging is done by the so-called full waveform analysis (FWA) in post-processing off-line. Such instruments have been classified as so-called small-footprint full-waveform ALS systems in contrast to echo-digitizing systems operated from space with large diameter laser footprints on the earth's surface. A typical laser footprint of the above-mentioned system is usually less than 0.4 meters from typical operating heights of about 1000 m above ground.

Since its first introduction there has been a continuous improvement in laser scanner hardware and thus data acquisition with respect to measurement rate and measurement range, but also in data processing with respect to classification, surface model extraction, and radiometric measurements (Ullrich *et al.* 2007; Wagner 2010). Numerous publications on full waveform analysis are based on data from the *RIEGL LMS-Q560* laser scanner and its successor, the *RIEGL LMS-Q680i* (*RIEGL* 2011).

Beside research and academic investigation, these laser scanners are widely used for real-life large-scale data production, covering applications in corridor mapping, large-scale area mapping, data acquisition in mountainous regions, and even on glaciers. The instruments are regarded as highly-reliable long-time stable workhorses for ALS in general. Together with the laser scanner hardware, *RIEGL* also offers a comprehensive software suite for managing, processing, analyzing, and visualizing data acquired with ALS systems or MLS systems in large-scale commercial projects. Within the software suite, *RiANALYZE* (*RIEGL* 2011) performs the FWA according to selectable algorithms.

In addition to FWA based on digitized and stored echo signals in off-line processing, *RIEGL LMS* has introduced series of commercial scanning systems, the V-Line, in 2008, (Pfennigbauer and Ullrich 2010), offering also echo digitization but on-line waveform processing, yielding similar results compared to full waveform analysis with even higher accuracy and precision, but with limitations with respect to multi-target resolution as explained below. V-Line laser scanners are offered as 3D laser scanners for TLS, but also as 2D laser scanners for ALS (e.g. the *RIEGL VQ-580*) and MLS (e.g. the *RIEGL VQ-250*). Figure 1 shows images of the *RIEGL LMS-Q680i* and the *RIEGL VQ-250*.



Figure 1: Commercial airborne laser scanners employing waveform digitization, the *RIEGL LMS-Q680i* for FWA in ALS (left) and the *RIEGL VQ-250* with online waveform processing for MLS systems (right).

Subsequently we will discuss the challenges in LIDAR technology related to multiple-pulse processing. As there might be some confusion about the term “full waveform data” or plain “waveform data” we will propose a classification of waveform data associated to laser scanning systems. We will briefly address different approaches on full-waveform analysis. The benefits of FWA with respect to mere analog signal processing will be discussed and we will provide an outlook on future developments.

2. Multi-target challenges in LIDAR technology

The technique of choice for long-distance ranging is time-of-flight measurement based on short laser pulses. Although the principle is simple and straight forward – emitting a short laser pulse in a collimated beam, receiving the echo pulses originating from backscattering of the emitted laser pulse on targets, and measuring the time between emitting and receiving, i.e., the time of flight – there are challenges in designing, manufacturing, and operating such instruments, at least when pushing the capabilities of the technology to its limits. Laser scanners are characterized by numerous features ranging from laser wavelength (Pfennigbauer and Ullrich 2011), maximum target distance, measurement speed, scanning range and speed, scan pattern, measurement accuracy and precision, to physical size, power supply requirements, and laser safety class, to name only a few. Additionally, compactness, reliability, short- and long-term stability of the internal and external calibration parameters are crucial to the use of such LIDAR-based systems. Multi-target resolution, as addressed in detail below, is especially important in applying LIDAR technology in, e.g., forestry, as the user of the final data may not only be interested in the uppermost parts of the canopy and the terrain itself, but also of all the layers of vegetation in between.

As the laser beam, although usually collimated to a divergence of less than 1 mrad, may hit not just a single target object, it is beneficial from a user's point of view, to get all the ranges to the targets the laser pulse has interacted with in a way, that the respective echo signal exceeds the detection threshold of the receiver. Providing more than just one target range per laser pulse is usually addressed as multi-target capability. Laser range finders based on the pulsed time-of-flight principle are capable of providing multiple targets per laser pulse, whereas phase-based cw (continuous wave) measurement schemes widely used in TLS for near range 3D data acquisitions are not on principle. However, there are fundamental limits to the multi-target capability: the laser pulse width and the system bandwidth limit the power to resolve echo pulses from nearby targets, as the finite pulse width of the laser pulse will lead to merging of the target echoes if the temporal difference is less than the pulse width. The capability to resolve two nearby targets is described by the multi-target resolution (MTR), stating the minimum target distance that can be resolved. In order to improve MTR, laser pulse width has to be reduced and system bandwidth has to be increased. There are limits imposed by the current state-of-the-art in laser technology, receiver technology and also system bandwidth, and these system parameters have also be traded-off against other system parameters like maximum range and laser safety.

Figure 2 below shows example waveforms illustrating multi-target situations. In (a) the multiple targets are separated in time, so that no influence of the early target to the late target return is to be expected. In (b) the signals already merged significantly, but still can be identified as superimposed targets as local maxima can be seen. In (c) the targets lie so close that by merging no individual local maxima can be found while the shape of the echo signal differs significantly from that of a single target situation.

Usually, accuracy and precision of a LIDAR system are stated for single-target test conditions. However, a first echo signal in the receiver may have some impact on the subsequent targets of the same laser pulse due to effects in the receiver electronics and the impact will increase the nearer the targets are. Echo digitization with waveform processing provides a significantly improved accuracy and precision in multi-target environments compared to LIDARs relying on mere analog signal detection and processing, addressed frequently as direct detection LIDARs, as it is possible to decompose, i.e. reconstruct, the superimposed signals to determine the individual ranges and amplitudes.

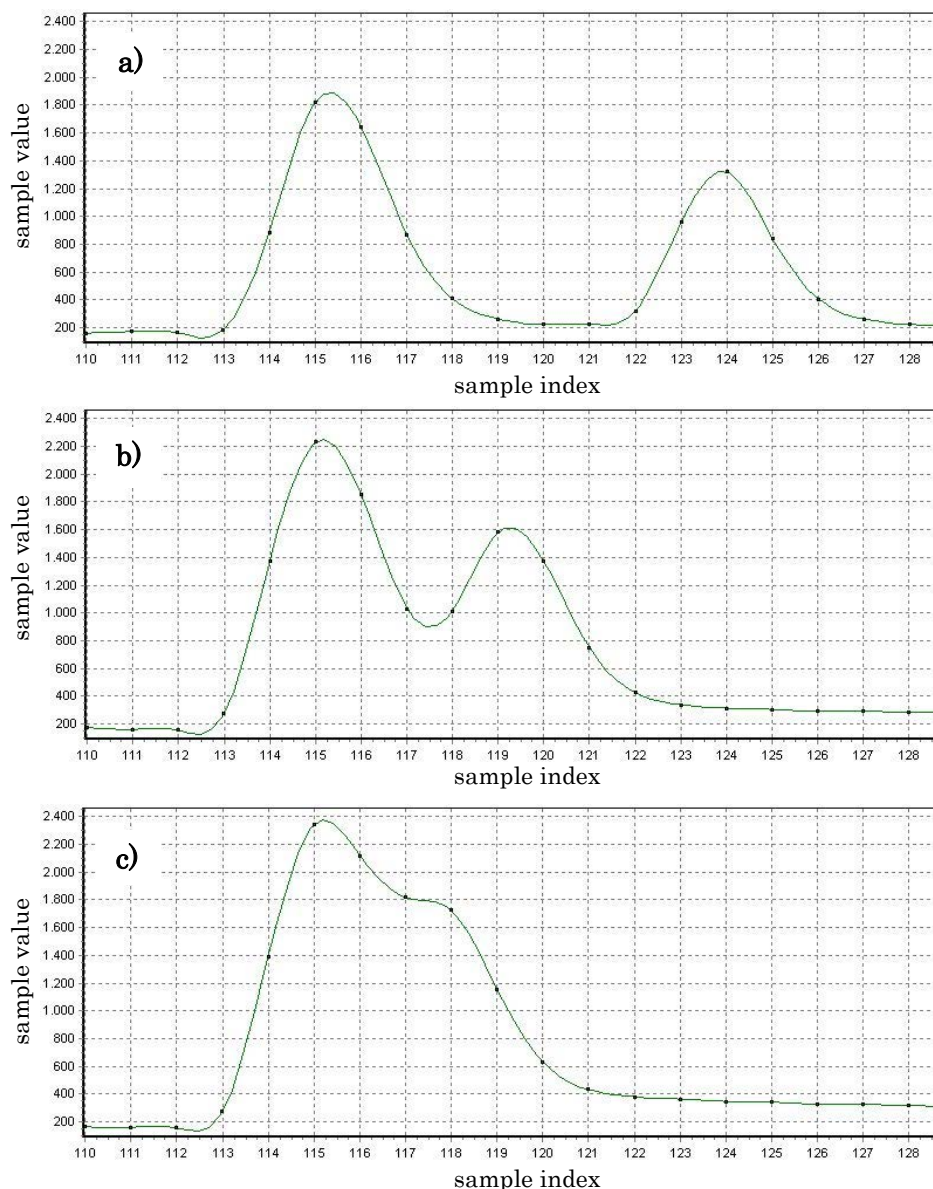


Figure 2: Waveforms for different target situations: targets sufficiently apart to deliver separate echo pulses (a), nearby targets with merging echo pulses (b), and targets so close that merged echo pulses have no separate maxima. Note that the green curves connecting the actual sampling values are obtained by employing a cubic spline just for improved visualization. Waveform data have been acquired with *RIEGL* VZ-400 with an optional waveform output.

3. Echo signal digitization with digital signal processing

In any LIDAR system a photodetector converts the optical echo signals into electrical signals. Within this paper we restrict the discussion to photodetectors operated in so-called linear mode, in which the amplitude of the electrical signal of the detector output is proportional to the optical signal power over a wide dynamic range and we do not discuss Geiger-mode receivers, which do not provide any radiometric information on the targets. In all practical LIDAR systems used for the applications mentioned above, the process of conversion is described as direct detection as in contrast to homodyne or heterodyne detection, a scheme widely used in the longer wavelength range of the electro-magnetic spectrum and in communications technology. Common to both, echo digitizing systems and discrete return systems is that the electrical

signals are amplified before further processing.

In echo-digitizing systems, the signals are sampled at a sufficiently high sampling rate and converted to a digital representation before target detection. This conversion is done by so-called analog-to-digital converters (ADCs). All further processing is then done in the digital regime, either on-line or off-line, after storing the sample data to and retrieving from a data recorder for off-line full waveform analysis.

Tasks to be carried out in digital signal processing are target detection, i.e., the discrimination of echo signals against noise, and parameter estimation for each detected target, with parameters usually including the temporal position of the target yielding finally the range to the target, the amplitude of the target signal yielding an estimate for the target's laser cross-section, and parameters allowing to estimate the backscatter profile of the target along the beam axis, like e.g. the pulse width.

In contrast to an echo-digitizing system, an analog discrete return system has to accomplish target detection and time-of-arrival estimation in real time by means of analog electronics. A separate analog amplitude estimator may guess the signal amplitude of the analog electrical target pulse, usually with a lot of shortcomings. Time-of-arrival estimation may be based on schemes like constant-fraction detection, analog differentiation with zero-crossing detection, or similar, all originating decades in the past in RADAR technology and all showing the effect of trigger walk, i.e., the estimated time-of-arrival depending on the amplitude of the electrical target signal. Especially in target constellations leading to signals as sketched in Figure 2 (b), the analog estimators usually yield significant ranging errors for the second and further targets and for signal as sketched in Figure 2 (c) analog means completely fail to retrieve further targets.

Echo digitization and waveform analysis is most beneficial in critical target situations, as sketched in Figure 3. In case the laser beam (sketched with an exaggerated high beam divergence) hits just a single plane target perpendicular to the laser beam axis, also the discrete return system may give accurate results. However, with slanted targets (as the roof of the building) and especially with complex multi-target situations when measuring into vegetation the echo-digitization / waveform analysis systems will provide clearly more precise and more detailed point cloud data.

4. Classifying Waveform Data Types

Echo signal digitization is the prerequisite to perform waveform analysis. The *RIEGL* LMS-Q560, introduced in 2004, was the first commercial laser scanner for ALS with all derived data products relying on the digitized echo signals only. Other products appeared on the market, offering echo digitization as an option, but with ranging still relying on analog ranging as in discrete return systems. In 2008, *RIEGL* introduced the V-Line, instruments for all three categories TLS, ALS and MLS, also based on echo-digitization but on on-line waveform processing. Subsequently, we attempt to classify waveform data the user can find on the market into different categories (cf. **Table 1** for an overview).

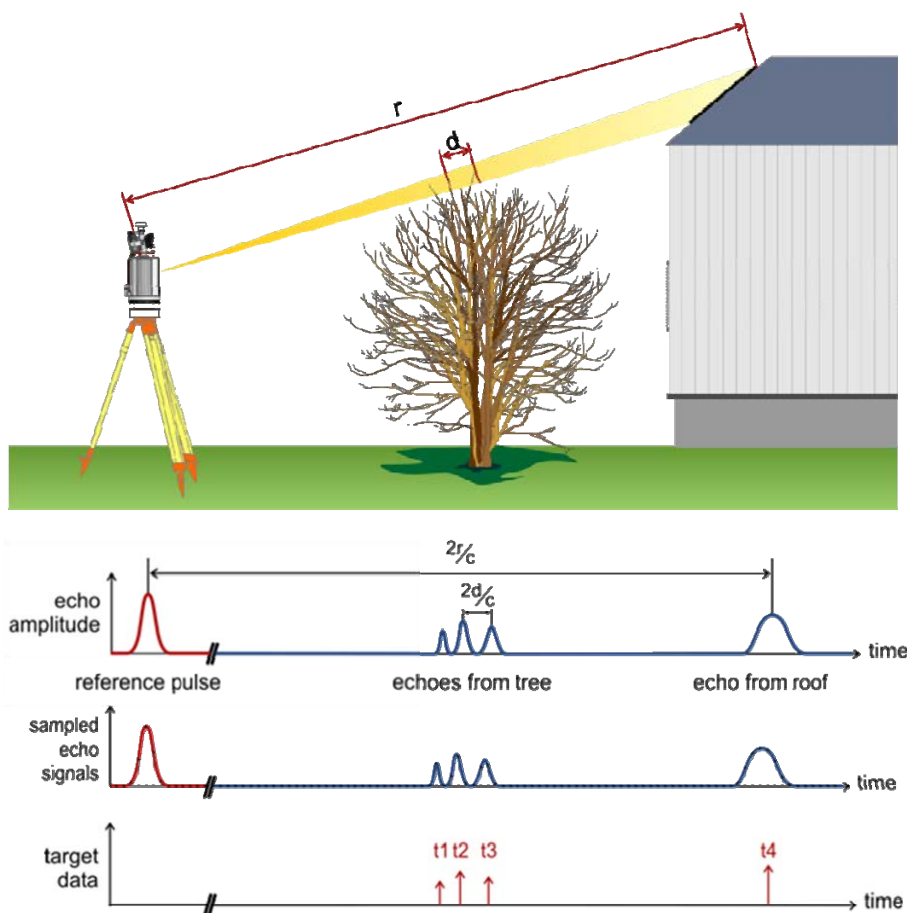


Figure 3: Illustration of the interaction of the laser pulse with different targets, the digitization process, and target extraction by FWA.

Full waveform data: These classical waveform data include the digitized echo signals and also data on a replica of the emitted pulse. All data products can be derived from the waveform data by means of a full-waveform analysis (FWA). In case the system pulse shape is nearly Gaussian, the Gaussian decomposition yields excellent results with high precision and accuracy. The waveform data also contain additional information for each laser shot with respect to time stamping to an external time regime like UTC and scan angle. With an appropriate sensor model, the ranges and attributes obtained by FWA are subsequently converted into a point cloud in the scanner's coordinate system with point attributes like amplitude, pulse width, and time stamp. In order to measure beyond the unambiguity range according to the pulse repetition rate of the laser, a precise time stamp related to each laser pulse has to be available, as e.g., in the *RIEGL* LMS-Q680i.

Echo waveform data: these data contain digitized echo signals on the target echoes only and no waveform data on the emitted pulse. Therefore, additional information on the precise emission time for each laser pulse has to be present to perform ranging in waveform analysis. Again the data set is complemented by external time stamping and scan angle.

Tightly-coupled echo waveforms: these data are optionally provided by LIDAR instruments with ranging based on echo digitization and online waveform processing. The waveforms are exactly the same data employed by the on-line waveform processing. The term tightly coupled refers to the fact that there is no additional ADC for just deriving some waveform data. An example is the *RIEGL* VZ-400 with its waveform option. Whether or not waveform data is

provided for a laser shot can be determined by the user via thresholds. For example, if all the target return received for a laser shot show the expected system response, there is no need to pass the waveforms for post-processing. On the other hand, in case of merging echo pulses, the waveforms are provided and computational expensive algorithms may derive more comprehensive and more accurate results in off-line waveform analysis.

Loosely-coupled signal samples: these data are delivered optionally by discrete return LIDARs with ranging based on analog electronics. The collection of waveform data by a separate digitizer is not related to the derived point cloud as different signal chains are used, therefore the term loosely coupled. These waveforms have merely an illustrative character to the points of the point cloud and the waveform's usability for improving the data quality of the discrete return system is very limited. The concept of the loosely-coupled waveforms is the one the LAS 1.3 format is propagating. The limited use of such data may be the reason for the very limited spread of the waveform option in the LAS 1.3 format.

Table 1: Comparison of the different waveform data types

	data content	range derivable from waveform	ADC coupling	user selectability of content
full waveform data	emitted pulse, all echo waveforms	yes	identity	no
echo waveform data	all echo waveforms	yes	identity	no
tightly-coupled echo waveforms	echo waveforms only	yes	tight	yes
loosely-coupled signal samples	fixed number of samples per laser shot	no	loose	no

Storing the waveform data of a replica of the transmitted pulse, which makes the difference between the first two categories, would be of significant advantage, in case the stability of the laser power and/or the laser pulse shape is questionable. In a well-designed system stability of the laser is sufficiently high and the waveforms on the transmitter pulse do not provide additional information compared to the precise emission time for each laser pulse. If one is especially interested in the system pulse shape for a special FWA algorithm, it is always recommended to derive that from real echo signals from single-point-targets or flat perpendicular targets, which are almost always found in each data set.

5. Challenges in full-waveform analysis

In multi-target environments a laser pulse interacts with numerous targets along the laser beam axis. As long as the targets have geometrical cross-sections smaller than the laser footprint at the target, there is a chance, that a fraction of the laser beam not obscured by early targets, may hit other targets. At each target, the laser pulse is partly absorbed and partly reflected. If the reflected or backscattered part of the pulse is received at the LIDAR's receiver with an amplitude exceeding the detection threshold, the range to this target can finally be determined by the LIDAR. For all but the first target, the responses of the targets are not only given by the respective laser radar cross section but also by the attenuation of the laser pulse by the preceding targets. It is worth noting, that attenuation by a target cannot be retrieved from the amount of

backscattering. Thus, only the laser radar cross section of the first target can be estimated accurately.

For the further discussion on FWA, it is advantageous to describe the interaction of the laser beam with the targets along the axis the laser pulse is travelling on as a one dimensional backscatter profile.

Assuming the backscatter profile is known, the optical signal over time at the receiver's aperture can be derived as the convolution of the laser pulse with the backscatter profile. If we further assume, that the LIDAR's receiver is linear, which is usually the case for small electrical signals, the electrical signal over time prior to AD conversion is given as the convolution of the system pulse response, as introduced earlier, with the backscatter profile with some noise added by the optical signal itself and receiver electronics. And, if we further assume that the sampling is done at a sufficiently high sampling rate, the digitized signal is an exact replica of the electrical receiver signal with some digitization noise added. However, it should be noted, that for larger signals outside the linear regime of the receiver, superposition takes place in a more complicated form as summation, signal compression, and bandwidth limitation take place in an intermingled form.

Generally speaking, the aim of FWA is to reverse the convolution of the system response with the backscatter profile and to find the backscattering identities along the laser beam axis with their respective parameters.

Numerous different approaches have been proposed to actually extract the backscattering properties of the targets from the digitized echo signals. Two different classes of analysis approaches can be seen: rigorous approaches aiming at the deconvolution (e.g. Roncat *et al.* 2011) and approaches based on modeling the digitized echo waveforms by means of basis functions (e.g. Wagner *et al.* 2006, Roncat *et al.* 2008). Deconvolution is prone to noise in the waveform, and there will always be noise in a well-designed LIDAR system. This noise will lead to backscatter artifacts and thus a “noisy” final point cloud, if no further precautions are implemented.

The most popular and widely used approach for FWA is the Gaussian decomposition. The underlying assumption is that the system response is at least nearly Gaussian, the backscattering contributions are also nearly Gaussian, the Dirac delta function can be well approximated by a very narrow Gaussian pulse, and, as the convolution of two Gaussian pulses is again a Gaussian pulse, also the digitized echo signal is the sum of Gaussian pulses – again assuming that superposition and linearity applies. Actual implementations of Gaussian decomposition rely on the following steps: find target candidates, i.e., Gaussian pulses in the waveforms, usually local maxima above a certain threshold, determine three parameters for each target candidate, i.e., position on the time axis, amplitude, and Gaussian pulse width, in order to fit the actual waveform in a least square sense. The pulse width of the target's backscatter is then the difference of the actual pulse width of the model pulse in the electrical regime and the pulse width of the system response.

This modeling approach can further be improved by not just using an approximate model for the system response such as a Gaussian pulse, but the actual system response of the system, as applied in RIEGL's online waveform processing in the V-Line. This approach gives the utmost accuracy and precision which can be achieved in an echo-digitizing LIDAR system and also perfectly accounts for effects imposed by non-linear signal compression. However, online waveform processing has its limitations when superposition of signals from nearby targets is present. Due to the lack of computational power in real-time processing the rigorous approach of LSQ-Fitting of numerous superposing responses cannot be applied. However, in this case,

online waveform processing at least informs the user about the merging of target responses by providing information on the deviation of the actual target's pulse shape from the expected pulse shape (Pfennigbauer and Ullrich 2010).

6. Benefits gained from Full Waveform Analysis

Sampling, digitizing, and storing the electrical receiver signals in a LIDAR system, the waveforms, provide the solid basis for a thorough insight into the interaction of the laser pulse with the targets hit by the laser beam. The waveforms contain all the available information “gained” by the laser pulse in an accessible way. The information is accessed by means of algorithms in the full waveform analysis and the standard parameters are retrieved such as range and amplitude, but also additional parameters like pulse width in case of Gaussian decomposition or pulse shape deviation in case the decomposition makes use of the actual system pulse response. In contrast, the discrete return LIDAR just provides ranges and maybe amplitudes for each target and all the information contained in, e.g., the shape of the echo pulses is lost and can never be recovered by post-processing.

The additional parameters from FWA are especially beneficial to the task of point cloud classification, i.e., assigning every point to a specific class like terrain/ground, vegetation, man-made objects, and similar. It has been demonstrated that the accuracy of classification of low vegetation can be significantly improved by making use of the estimated pulse width (Ullrich et. al. 2007).

Multi-target resolution and multi-target accuracy are limited by the system bandwidth. It is straight forward in FWA by, e.g., Gaussian decomposition, to identify all target echoes which are separated in a way that each echo leads to a local maximum in the waveform. However, it has been demonstrated that it is possible to even discriminate targets that are closer with the presumption that the waveform does not originate from a volume backscatterer or a slanted target (Roncat 2008).

Pulse width or pulse shape deviation can be used to clean up point clouds in a straightforward way before applying ICP (iterative closest point) algorithms for point cloud registration. Cleaning up is done by deleting all points with questionable reliability, i.e., measurements into vegetation or measurements on the edges of objects before a nearby background object. The iterative registration process will significantly converge more reliable and faster with “clean” point clouds. Especially small steps in depth below the multi-target resolution can be detected and false points can be deleted, as at least the pulse width and the pulse shape deviation give hints on such critical target constellations (Pfennigbauer *et al.* 2009; Pfennigbauer and Ullrich 2010).

Algorithms for FWA are numerous and the selection of the algorithm and tuning of it can be optimized for certain applications in TLS, ALS and MLS. The user of full waveform data can trade off for example detection threshold against false alarm rate by tuning the detection threshold in the echo detection process in FWA, or the user can tackle flaws in the analog signal processing chain resulting, e.g., in ringing after large echo signals.

Full waveform data is ideal for radiometric calibration of ALS data as demonstrated in detail in (Wagner 2010). Echo-digitization with online waveform processing as implemented in the *RIEGL* V-Line instruments forms the basis for the calibrated reflectance reading for each measurement (Pfennigbauer and Ullrich 2010).

7. Summary and Outlook

Echo signal digitization with subsequent online waveform processing or off-line full waveform analysis has established itself as the measurement technique of choice in state-of-the-art laser scanning devices for TLS, ALS and MLS applications, as it delivers accurate, low-noise, rich-in-detail point clouds with additional attributes to improve post-processing and the potential to straightforward radiometric calibration. These laser scanners have found widespread use and the interest in waveform analysis is not restricted to research and academic institutions, but is nowadays frequently found as the “ranging engine under the hood” of laser scanners in everyday commercial use in mass data production.

With the availability of new laser sources, more powerful electronics in the field of signal conversion, with the steady increase in on-board computational power, it can be expected, that multi-target resolution will further increase by utilizing shorter laser pulses and higher sampling rates with higher digitization depths. The improvements in data storage devices and the increase in data transmission speed enable even higher measurement rates, even at higher sampling rates. Online waveform processing of the future may reach the power of off-line from today, so that powerful online multi-target processing would provide the point clouds as rich in details and attributes as those of today but in real time.

References

- Doneus, M., Briese, C., Fera, M. and Janner, M., 2008. Archaeological prospection of forested areas using full-waveform airborne laser scanning. *Journal of Archaeological Science*, 35(4), 882–893.
- Hug, C., Ullrich, A., Grimm, A., 2004. Litemapper-5600. A waveform-digitizing LIDAR terrain and vegetation mapping system. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (Part 8/W2), 24 - 29.
- Mallet, C., Bretar, F., 2009. Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 1-16.
- Pfennigbauer, M., Rieger, P., Studnicka, N., and Ullrich, A., 2009. Detection of concealed objects with a mobile laser scanning system. In *SPIE Laser Radar Technology and Applications XIV*, Orlando, 7323, 732308-1 - 732308-9.
- Pfennigbauer, M., and Ullrich, A., 2010. Improving quality of laser scanning data acquisition through calibrated amplitude and pulse deviation measurement. In *SPIE Laser Radar Technology and Applications XV*, Orlando 7684, 76841F-1 - 76841F-10.
- Pfennigbauer, M., and Ullrich, A., 2011. Multi-wavelength airborne laser scanning. In *International lidar mapping forum, ILMF*, New Orleans.
- RIEGL Laser Measurement Systems GmbH, 2011. www.riegl.com.
- Roncat, A., Wagner, W., Melzer, T. and Ullrich, A., 2008. Echo detection and localization in full-waveform airborne laser scanner data using the averaged square difference function estimator. *The Photogrammetric Journal of Finland*, 21(1), 62–75.

- Roncat, A., Bergauer, G., and Pfeifer, N., 2011. B-spline deconvolution for differential target cross-section determination in full-waveform laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 418–428.
- Ullrich, A., Hollaus, M., Briese, C., 2007. Utilization of full-waveform data in airborne laser scanning applications. In *SPIE Laser Radar Technology and Applications XII*, Orlando, 6550, 65500S-1 - 65500S-12.
- Wagner, W., Ullrich, A., Melzer, T., Briese, C. and Kraus, K., 2004. From single-pulse to full-waveform airborne laser scanners: potential and practical challenges. In: *International Archives of Photogrammetry and Remote Sensing*, XXXV, Istanbul, Turkey.
- Wagner, W., Ullrich, A., Ducic, V., Melzer, T. and Studnicka, N., 2006. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(2), 100–112.
- Wagner, W., 2010. Radiometric calibration of small-footprint full-waveform airborne laser scanner measurements: Basic physical concepts. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65, 505–513.

Crown coverage calculation based on ALS data

Lothar Eysn¹, Markus Hollaus¹, Klemens Schadauer² and Andreas Roncat¹

¹ Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Gußhausstraße 27-29, 1040 Vienna, Austria, <le, mh, ar @ipf.tuwien.ac.at>

² Department of Forest Inventory at the Federal Research and Training Center for Forests, Natural Hazards and Landscape, Seckendorff-Gudent-Weg, 1130 Vienna, Austria, <klemens.schadauer@bfw.gv.at>

Abstract

The objective of this paper is to present and evaluate a new geometrically unambiguously defined approach to calculate forest canopy cover, also known as crown coverage (CC) from airborne laser scanning (ALS) data based on national forest inventory (NFI) data. The CC is defined as the proportion of the forest floor covered by the vertical projection of the tree crowns. Most forest definitions lack in precise geometrical definitions for the calculation of CC and therefore, the results of common calculation methods differ and tend to be incomparable. To demonstrate the effect of such an unclear defined, common CC calculation method, CC maps, generated from moving window algorithms using different kernel shapes and sizes, are calculated and analyzed for three study areas in Tyrol, Austria. The new unambiguously approach, the tree triples method, is based on defining CC as a relation between the sum of the crown areas of three neighbouring trees at a time and the area of their convex hull. The approach is applied for the same study areas and is compared with forest masks that are generated from moving window algorithms using different kernel shapes and sizes.

Keywords: forest definition, canopy cover, forest border delineation, vegetation mapping, LiDAR

1. Introduction

The delineation as well as the classification of forests has a long tradition in remote sensing. Considering different forest definitions (e.g. Austrian forest law, FAO) forested land can for example be composed of tree crowns, forest gaps, forest streets or harvested areas. It is often difficult to derive this complex land use class “forest” from remotely sensed data in a reliable and comprehensible way. In different forest definitions the criterion of crown coverage (CC) is a fundamental and obligatory parameter for classifying forested areas. For example the international forest definition of the United Nations Food and Agricultural Organization (FAO) defines a forest as land of at least 0.5 ha with a potential tree height of at least five meters and a CC greater than 10% (FAO/FRA, 2000). CC, also known as canopy coverage or forest canopy cover, is defined as the proportion of the forest floor covered by the vertical projection of the tree crowns (Jennings et al., 1999). In (Korhonen et al., 2011) vertically measured crown cover is referred as vertical canopy cover (VCC). The current paper considers VCC. An unclear defined detail is the treatment of gaps within the projected tree crowns itself. The traditional definition of canopy cover includes an “outer edge” or “envelope” of a crown, inside of which the cover is thought to be continuous, but in practice the “outer edge” is sometimes very difficult to observe (Korhonen et al., 2006). For the current paper those crown gaps are not considered.

To evaluate the amount of CC for an area, in-situ measurements or remote sensing techniques can be used. In-situ measurements are time consuming and are mainly operated for sample plots while remote sensing techniques overcome the limitation of plot-wise sampling and provide the possibility to analyze large areas. As terrestrial measurements deliver the ground truth for most

of the remote sensing techniques, in-situ samples are a fundamental input for cross validations. A comparison of common terrestrial measuring techniques can be found in (Korhonen et al., 2006). An often applied method for assessing an area's CC is the manual interpretation of orthophotos. This technique is however costly, limited by shadowing effects and the quality of the results are dependent on the interpreter. It is therefore difficult to obtain objective quantitative measurements that are suitable for comparisons with remotely based CC measures (Holmgren et al., 2008). A different approach is to define the amount of CC as a relation between two trees. Depending on the threshold of CC, the tree species and the tree crowns size a maximum distance between two trees can be determined (Hauk and Schadauer, 2009). This method, which is originally based on the work of Hasenauer (Hasenauer, 1997), is currently used for the manual delineation of forested areas at the Department of Forest Inventory at the Federal Research and Training Center for Forests, Natural Hazards and Landscape (BFW) in Austria.

As an alternative to the manual photo interpretation the technique of airborne laser scanning (ALS) was established for assessing an area's CC (Holmgren et al., 2008; Korhonen et al., 2010). ALS, as an active remote sensing technique, is not influenced by shadowing effects or different sun illumination conditions, is able to deliver reliable information even for small forest gaps and is well suited for estimating CC. The normalized digital surface model (nDSM), calculated by subtracting the digital terrain model (DTM) from the digital surface model (DSM) provides an excellent data source for calculating the CC. Using the nDSM, a height threshold can be applied to decide whether a pixel is covered by tree crowns or not. In a next step the CC can be calculated by dividing the reference area by the tree crown covered area. As reference area forest stands or moving windows with user defined circular or squared kernel shapes are commonly in use. Unfortunately, due to the lack of precise geometric descriptions of the CC (i.e. reference size and -shape) the derived results are often not comparable and make the CC to a doubtful criterion. Therefore, this study aims at defining a novel, geometrically clear defined method for an automatic calculation of CC based on ALS and NFI data. In this approach CC is defined as a relation between the sum of the crown areas of three neighbouring trees at a time and the area of their convex hull. The new method is applied for three study areas in Tyrol, Austria considering the forest definition of the Austrian national Forest inventory (NFI). This study is part of the research project "LASER-WOOD" funded by the Klima- und Energiefonds in the framework of the program "NEUE ENERGIEN 2020". As LASER-WOOD is an ongoing project this paper describes first results of the ongoing investigations.

The remaining parts of this paper are organized as follows: Section 2 describes the selected study areas and the used data. In Section 3 the methodology and implementation is explained. Section 4 shows results and their discussions whereas in Section 5 concluding remarks are given.

2. Study area and dataset

2.1 Study area

In this contribution three different study areas in Austria are investigated. The study areas are located in the "Zillertal" which is located in the eastern part of the federal state of Tyrol. Each study area covers an area of 2.5 x 2.5 km and shows different structures and amounts of forested land (Figure 1). Study area 1 consists of a loose stocked forest at the upper timberline (Figure 1a) with elevations from 1800 to 2000 m above sea level (a.s.l.). Study area 2 consists of a fragmented forest with patch-wise forest stands on the hillside (Figure 1b) with elevations from 600 to 1600 m a.s.l. Study area 3 consists of a mainly dense forest with different age classes (Figure 1c). The elevations for study area 3 reach from 700 to 1500 m a.s.l. The dominant tree species in all three study areas are coniferous trees. Beside the forested areas buildings and power lines can be found in the study areas.

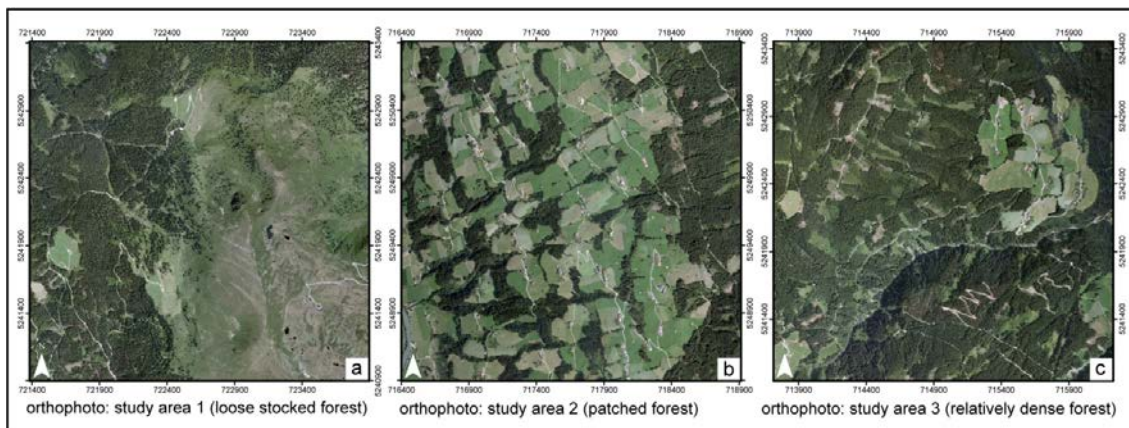


Figure 1: Orthophotos of the study areas. (a) Study area 1 shows a loose stocked forest at high elevations. (b) Study area 2 shows a patched, fragmented forest. (c) Study area 3 shows a relatively dense forest with different age classes.

2.2 ALS data

The used ALS data was acquired using an Optech Inc. ALTM 3100 laser scanner during multiple flight campaigns in 2008 under leaf-off and leaf-on canopy conditions. The mean point density is about 4 echoes / m² for study area 1, 5 echoes / m² for study area 2 and 9 echoes / m² for study area 3. Further details can be found in (Eysn et al., 2010a).

2.3 Derived base products

The ALS data has been processed and filtered using the hierarchic robust filtering approach (Kraus and Pfeifer, 1998) to obtain DTM's. For the processing of the DSM a land cover dependent derivation approach (Hollaus et al., 2010) was chosen. By subtracting the DSM from the DTM a normalized digital surface model (nDSM) was created as a fundamental base product for calculating the CC and delineating forested areas. Additionally a slope adaptive echo ratio (sER) map (Höfle et al., 2009), as a measure for local transparency and roughness of the top-most surface, was derived. To eliminate buildings and other artificial objects, the sER map was corrected with morphological operations and thresholding to a so called "vegetation mask". Further information on this correction can be found in (Eysn et al., 2010b). The spatial resolution of the derived products is 1 x 1 m².

3. Methodology and Implementation

3.1 Moving window approach

As described in the Introduction the automatic inspection of the criteria CC is crucial and unfortunately not clearly defined. Especially for larger scale applications like the automatic delineation of forested areas based on ALS data the moving window approach leads to varying results. To demonstrate the effect of different parameters for kernel shapes and -sizes on the resulting CC maps, multiple variations of these two parameters have been analyzed. To be able to compare the results, the sum of areas fulfilling different CC thresholds are compared with each other.

As a basis for these calculations a combination of a height thresholded nDSM and the vegetation mask is chosen. Pixels with a nDSM value greater than 2.0 m and an sER value less than 85% are assumed to be crown covered and are set to one. Pixels not fulfilling these criterions are set to zero. The height threshold is set to consider the minimum height criterion of the Austrian

NFI. Based on the derived binary, “preliminary vegetation map” the CC values are calculated with a circle- and square-shaped kernel with different kernel sizes using the software OPALS (OPALS, 2011). The used kernel sizes are defined as a radius (in pixels) from 1 to 40 Pixels. For example a kernel radius of 3 the square shaped kernel is a 7x7 matrix. The centre pixel of the kernel is calculated by the mean of all pixels covered by the kernel and represents the CC for this Pixel. To be able to check the results of these calculations against different forest definitions, several CC thresholds are applied to the calculated CC maps. For this study CC thresholds from 10 % to 100 % with steps of 10 % were chosen. The results of this processing step are binary maps which represent so called potential forest masks. Finally, the potential forest mask’s size within the study area is determined. For visualisation purposes those areas are plotted against the kernel sizes corresponding to the selected CC thresholds (see Figure 3).

3.2 Tree triples approach

The developed method for the calculation of CC aims at defining the criteria of CC with a clear geometrical definition which is based on ALS data and NFI data. The basic idea is to express CC as a relation between the sum of the crown areas of three neighbouring trees at a time and the area of their convex hull (see Figure 2).

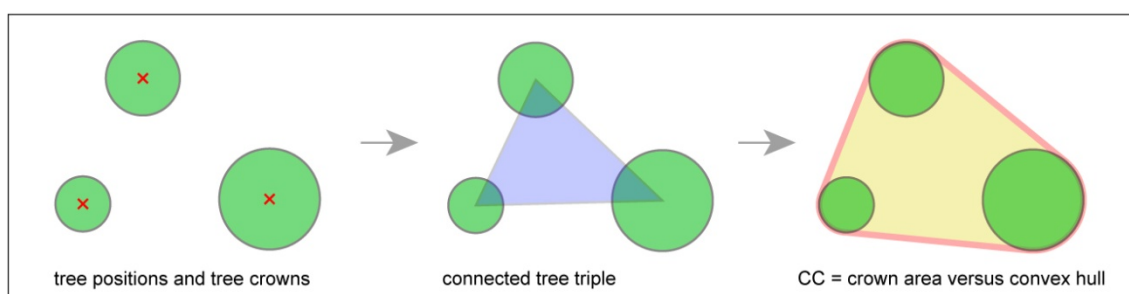


Figure 2: tree triples approach: three trees at a time are connected. The amount of CC is the relation between the area covered by crowns and the area of the convex hull.

As described in (Eysn et al., 2010b) the tree positions are detected with a local maxima filter based on the nDSM and the vegetation mask. To consider the minimum height criterion of the Austrian NFI a height threshold of 2.0 m is applied to the maxima search. The crown diameters are assessed using empirical functions, which act as a relationship between tree height and crown radius. These functions are calibrated based on measurements of crown radii from the Austrian NFI, whereas for this study the function was assimilated for trees near the timberline. Further details can be found in (Eysn et al., 2010b). To find the tree triples for calculating the CC, a Delaunay triangulation is applied to the detected local maxima. The Delaunay triangulation is calculated using libraries of the Open Source software CGAL (CGAL, 2011). In a next step the sum of the crown areas A_{cr} of three neighbouring trees at a time and the area of their convex hull A_{hull} is calculated for each tree triple. For this purpose a tool was implemented in Python (PYTHON, 2011) which imports a triangulation, calculates the parameters A_{cr} and A_{hull} and returns a CC value for each tree triple. For overlapping tree crowns within a tree triple the intersected crown area is used for A_{cr} . Tree triples respectively their triangles are removed if the selected CC threshold is not fulfilled. The result of these calculations is a potential forest mask which considers the minimum height criterion as well as the minimum CC criterion. As the exported result is a triangulation with triangles fulfilling the CC criterion and not the convex hulls of the tree triples, the borderlines of the derived potential forest mask represent the tree stem axes. For this reason the resulting map is buffered by the half of the maximum available

crown diameter found in the study area. In order to prevent an overestimation of the derived potential forest mask the buffered area is intersected with the vegetation mask.

3.3 Comparison of the two approaches

To be able to compare the results of the two different approaches, a final forest mask, based on the potential forest mask, is derived for both methods. This is necessary because the moving window approach delivers raster based information, the tree triples approach delivers triangle based information which makes a direct comparison difficult.

For both methods the final forest mask is calculated considering the geometrical aspects of the forest definition of the Austrian NFI. For the moving window approach different final forest masks are calculated because of the different kernel sizes and –shapes. The minimum height (set to 2.0 m) and the minimum CC (set to 30 %) is already handled in the potential forest mask. The minimum area criterion is applied by vectorizing the potential forest mask and by deleting single polygons or filling forest gaps with an area less than 500 m². In a next step the minimum width criterion (set to 10 m) is applied by morphological operations. As areas might have changed due to deletion according to the minimum width criterion, the minimum area criterion is checked a second time after this step.

4. Results and Discussion

4.1 general considerations

The definition of CC claims a strictly vertical projection of the tree crowns. In ALS the laser beam vectors are inclined in most instances (except at nadir) and the criteria of a vertical projection is not strictly maintained. However, in typical ALS surveys the off-nadir angles are at maximum 20°, so this effect should remain relatively small. In addition, however, the penetration ability of an ALS pulse may be limited through small canopy gaps, and vary somewhat with the technical acquisition settings (Korhonen et al., 2010). The crowns in the derived base products of ALS tend to be overestimated because the base products are widely raster based and the exact size of the modelled crowns depend on the spatial resolution of the models.

In the following sections the results of the previous calculations are presented and discussed:

4.2 Moving window approach

The results of the moving window method for the three study areas are presented in Figure 3. For each study area the results are separated by the used kernel shape and the different CC thresholds (colored curves). The vertical axis represents the sizes of the resulting areas or potential forest masks in relation to the whole extent of the study area while the horizontal axis represents the different kernel sizes. The values on the vertical axis are normalized between 0 % and 100 %. For example, a kernel size of 0 (which means just one pixel) results in a potential forest mask similar to the vegetation mask. If the window size is not correlated with the resulting areas of fulfilled CC thresholds, all curves of the different CC threshold should be strictly horizontal. If the study area would be covered by a forest by 100%, all curves should be strictly horizontal lines which overlap at 100%.

The CC threshold curves for study area 1 are wide spread compared to the results of the other study areas which seems to be a cause of the loose stocked forest pattern. For study area 2 and study area 3 the curves are narrower. This findings show, that a decrease of the forest density leads to an increased effect of different CC thresholds on the found area.

The result of study area 1 shows a strong variation of the resulting areas between kernel size 1 and kernel size 8 for most of the selected CC thresholds. For kernel sizes from 9 to 40, the resulting forest masks seem to be more independent on the kernel size. In study area 2 (patched forest) a strong variation of the resulting areas is given for a larger range of kernel sizes compared to study area 1. It can be deduced, that the gradient and the curvature of the different curves reflect the kernel size dependency of the resulting potential forest masks for different selected CC thresholds.

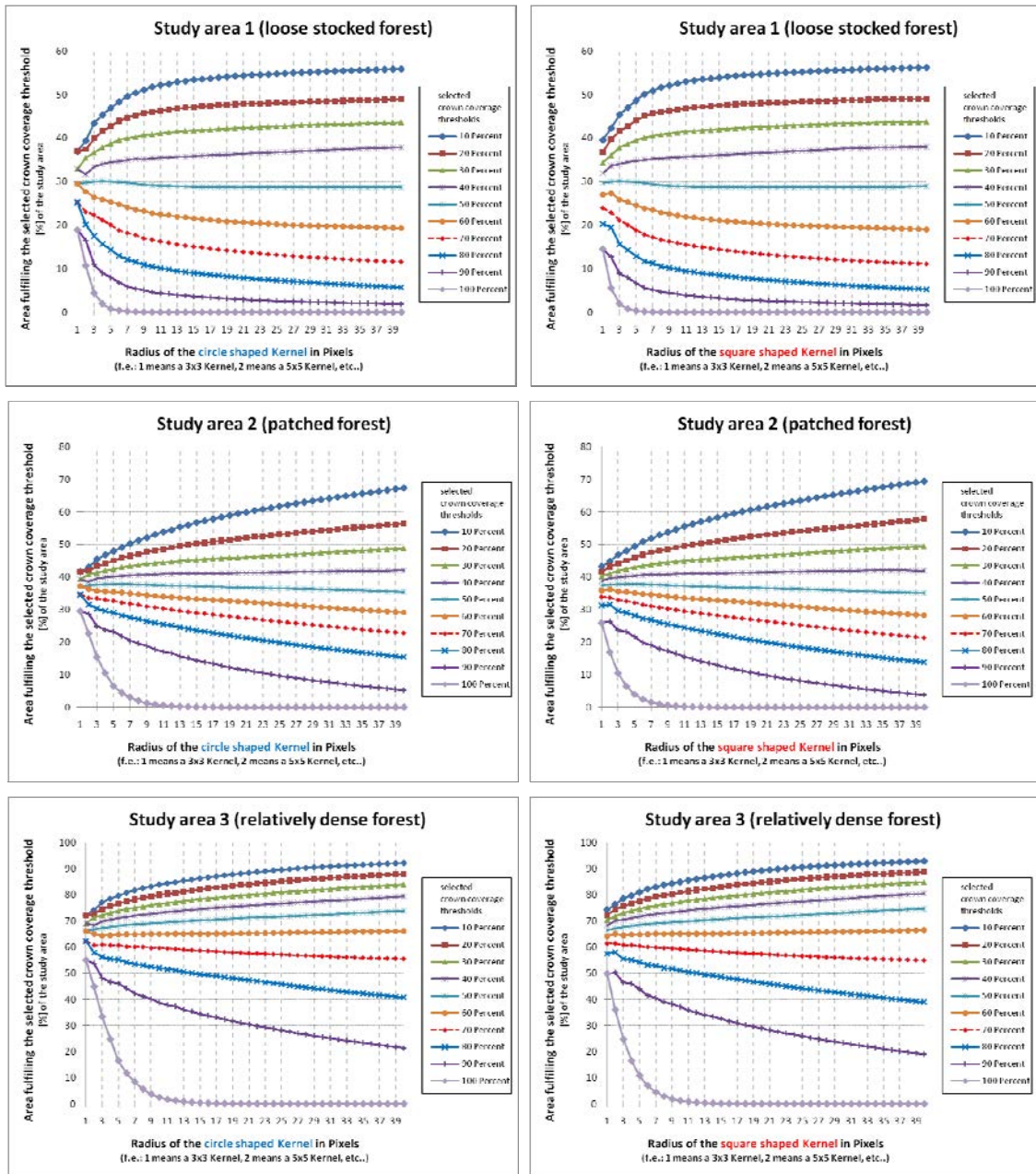


Figure 3: Resulting sizes of the potential forest masks of three different study areas using the moving window approach with varying kernel sizes. In the left column the results for a circle shaped kernel with different CC thresholds is shown for all three study areas. In the right column the results for a circle shaped kernel with different kernel sizes and different CC thresholds are shown.

4.3 Tree triples approach

A manual inspection of the automatically detected potential tree positions based on the nDSM and the vegetation mask shows suitable results (Figure 4b). Due to the limitation of the maxima search using the vegetation mask mainly maxima in vegetated areas are found. Because of the small kernel size of 5 x 5 pixels multiple local maxima are sometimes found within the area of single tree crowns. Especially within dense forested areas the detected local maxima do not represent the exact tree stem positions and the tree detection rate can be low. Single and clear separable trees in loose stocked areas are correctly detected in most instances. It can be assumed that the amount of detected local maxima is highly correlated with the chosen kernel size. In relation to the inspection of CC, not exact or non detected tree positions within a dense forest play a minor role since the criterion of CC is most critical for sparse, loose stocked forest areas where primary single, clearly separable trees are present.

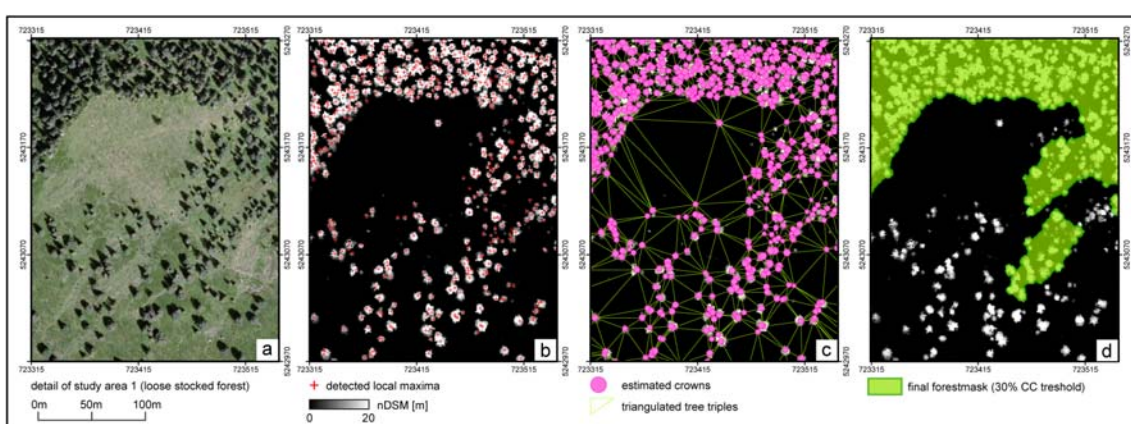


Figure 4: Intermediate results of the tree triple approach; a) orthophoto of a subset of study area 1 b) detected local maxima c) detected tree triples plus estimated crowns d) final forest mask fulfilling the criteria of the Austrian NFI.

For the detected local maxima the corresponding tree crowns were calculated based on the calibrated formulas determined from NFI data. To validate the estimated crowns, the derived crown areas are compared to the source map. The source map for the calculations is the height thresholded (nDSM > 2 m) combined with the vegetation map. In the source map, all pixels fulfilling the selected thresholds are assumed to represent a crown pixel. For each study area the sum of these pixels represent the amount of land covered by tree crowns. The areas of the estimated crowns are also summed up. The comparison of these resulting sums (see Table 1) shows a good estimation of the tree crowns (see Figure 5b) for a loose stocked forest (delta = 3,9 %) while the estimation is worse (see Figure 5d) for a relatively dense forest (delta = 21,8 %). It can be assumed, that the overall smaller estimated sum of crown areas for the relatively dense forest can be explained by limitations of the local maxima search or because the calibration of the relation tree height versus tree crown was performed for mainly trees at the upper timberline.

Table 1: Validation of the estimated tree crowns in comparison to the source map

	\sum crowns source map [%]	\sum estimated crowns [%]	delta [%]
study area 1 (loose stocked forest)	28,5	24,6	3,9
study area 2 (patched forest)	36,5	24,3	12,2
study area 3 (rel. dense forest)	64,9	43,1	21,8

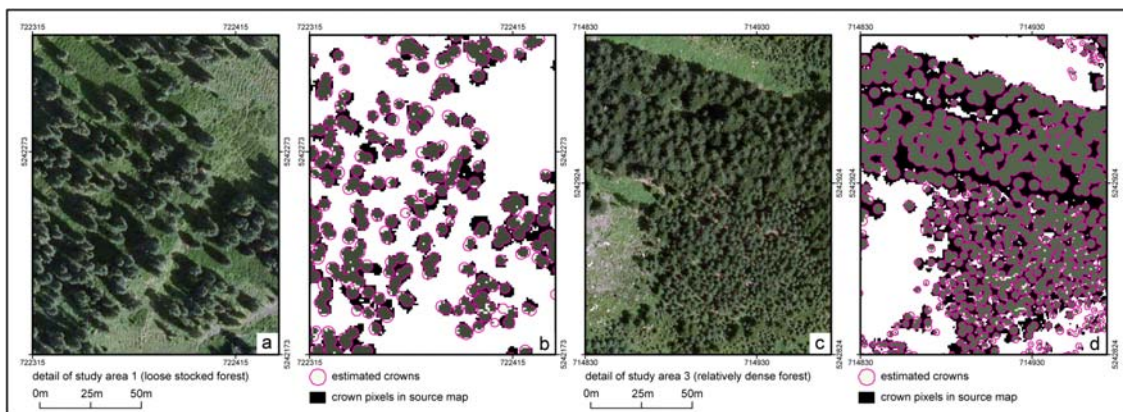


Figure 5: comparison of estimated crowns with the source map; a) orthophoto of a subset of study area 1 b) estimated tree crowns in a loose stocked forest c) orthophoto of a subset of study area 3 d) estimated tree crowns in a relatively dense forest

The Delaunay triangulation of the potential tree positions shows conclusive results for the connection of tree triples (see Figure 4c). The derived tree triples are reliable filtered and eliminated depending on the selected CC threshold and provide, especially at loose stocked areas at the forests timberline, suitable results for the potential forest mask. This mask is a fundamental input for the delineation of forested areas based on a forest definition and therefore, the less detail of this “CC map” at relatively dense forested areas plays a minor role since the focus is on loose stocked areas.

The final forest masks for the comparison of the different approaches are calculated based on the forest definition of the Austrian NFI. Due to the applied minimum area criterion small forest patches with an area less than 500 m² are removed and forest clearings with an area less than 500 m² are assigned to the forest area. Narrow forest areas are eliminated by applying the minimum width criterion. The results of the calculated final forest masks are plotted in Figure 6. The results of study area 1 show almost similar curves for the circle- and square-shaped kernel while the study area 3 shows differing curves with increasing kernel size. It can be assumed that an increasing density of a forested area combined with an increasing kernel size leads to more different results in the resulting final forest mask. Compared to the tree triples approach, the results for the patched and relatively dense forest show almost similar results if a kernel size of 9 to 12 m is chosen, while the results differ for the loose stocked forest. The results show big differences with de- and increasing errors at increasing kernel sizes. Those big differences reflect the limitations of the moving window approach since the results are high correlated with the kernel size.

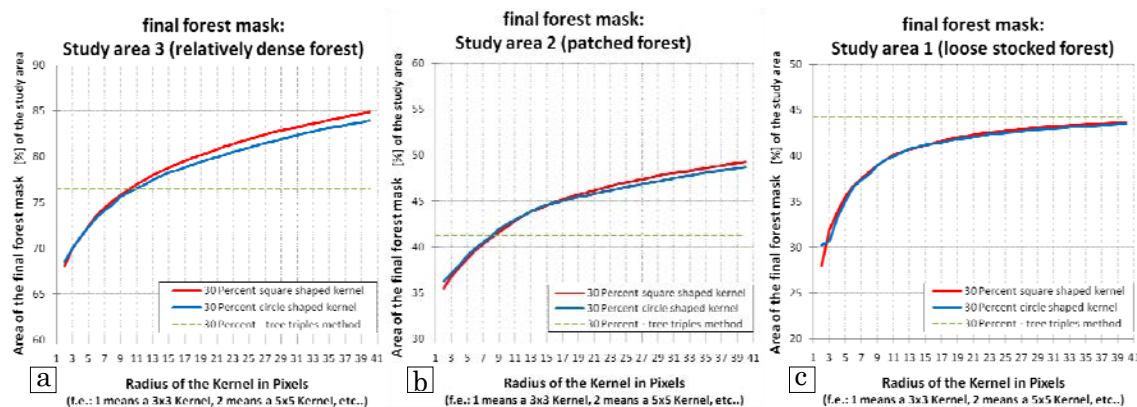


Figure 6: comparison of the resulting forest masks for the moving window approach and the tree triples approach; a) loose stocked forest b) patched forest c) relatively dense forest

5. Conclusion

This study shows the high potential of ALS data for assessing CC and consequently for deriving a forest mask for large areas. A clear geometrical definition for the calculation of CC is necessary since CC is a fundamental criterion in most forest definitions and the results of the moving window method are differing due to its high dependency on the kernel size. It could be shown, that a decrease of the forest density leads to an increased effect of different CC thresholds on the found forest area. Especially at the upper timberline, different kernel sizes and CC thresholds lead to different results. The tree triples method can overcome the limitations of the moving window approach especially at loose stocked forests. The local maxima detection works reliable for such forest areas. The local maxima detection could be improved, especially for dense forests, by applying a more complex detection method. The estimation of tree crowns based on the tree height shows consistent results at the, related to CC, critical area at the upper timberline. The estimation of crowns could be improved by a local calibrated transfer function. In future studies, the method will be firstly investigated for mixed and deciduous forests and secondly a validation with forest inventory will be performed. However, acquiring reference measurements from field data for large areas as well as the manual orthophoto interpretation is still challenging and therefore a reliable method for calculation CC from ALS data is a big effort.

Acknowledgements

The ALS data and orthophotos for the test areas Zillertal were kindly provided by the department of Geoinformation of the Tyrol state government (Amt der Tiroler Landesregierung, Gruppe Landesbaudirektion, Abteilung Geoinformation). A great deal of gratitude goes out to our colleague Christian Ginzler at the Federal Institute for Forest, Snow and Landscape Research (WSL) for the constructive discussions on this topic. The forest inventory data is kindly provided by the Department of Forest Inventory at the Federal Research and Training Center for Forests, Natural Hazards and Landscape, Vienna, Austria. This study is done within the project LASER-WOOD (822030), funded by the Klima- und Energiefonds in the framework of the program "NEUE ENERGIEN 2020".

References

- CGAL Computational Geometry Algorithms Library, <http://www.cgal.org/>, last aCCess: 20.05.2011
- Eysn, L., Hollaus, M., Mücke, W., Vetter, M. and Pfeifer, N., 2010a. Waldlückenerfassung aus ALS Daten mittels alpha-Shapes. In: *Dreiländertagung - 30. Wissenschaftlich-Technische Jahrestagung der DGPF*, Vienna, Vol. Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band 19 (2010): 552 - 560.
- Eysn, L., Hollaus, M., Vetter, M., Mücke, W., Pfeifer, N. and Regner, B., 2010b. Adapting alpha-shapes for forest delineation using ALS Data. In: *10th International Conference on LiDAR Applications for Assessing Forest Ecosystems (Silvilaser 2010)*, Freiburg, Germany, Vol.: 10 p.
- FAO/FRA, 2000. Definitions of forest and forest change, Forest Resources Assessment Programme, Rome 2000, 15.
- Hasenauer, H., 1997. Dimensional relationships of open-grown trees in Austria. *Forest Ecology and Management*, 96, 197-206.
- Hauk, E. and Schadauer, K., 2009. Instruktion für die Feldarbeit der Österreichischen Waldinventur 2007 – 2009 Vienna, 201 p.
- Höfle, B., Mücke, W., Dutter, M., Rutzinger, M. and Dorninger, P., 2009. Detection of building regions using airborne LiDAR – A new combination of raster and point cloud based GIS methods. *GI_Forum 2009 - International Conference on Applied Geoinformatics, Salzburg*, 66-75.

Hollaus, M., Mandlbürger, G., Pfeifer, N. and Mücke, W., 2010. Land cover dependent derivation of digital surface models from airborne laser scanning data. In: *ISPRS Commission III Symposium PCV2010*, Saint-Mandré, France, Vol. Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Volume XXXVIII: 6-6.

Holmgren, J. et al., 2008. Estimation of crown coverage using airborne laser scanning. In: *SilviLaser 2008, 8th international conference on LiDAR applications in forest assessment and inventory*, Heriot-Watt University, Edinburgh, UK, Vol.: 50-57.

Jennings, S.B., Brown, N.D. and Sheil, D., 1999. Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. *Forestry*, 72(1), 59-59.

Korhonen, L., Kaartinen, H., Kukko, A., Solberg, S. and Astrup, R., 2010. Estimating vertical canopy cover with terrestrial and airborne laser scanning. In: *10th International Conference on LiDAR Applications for Assessing Forest Ecosystems (Silvilaser 2010)*, Freiburg, Germany, Vol.

Korhonen, L., Korpela, I., Heiskanen, J. & Maltamo, M., 2011. Airborne discrete-return LiDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sensing of Environment* 115(4): 1065-1080.

Korhonen, L., Korhonen, K.T., Rautiainen, M. and Stenberg, P., 2006. Estimation of forest canopy cover: a comparison of field measurement techniques. *Silva Fennica*, 40(4), 577-588.

Kraus, K. and Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 53(4), 193-203.

OPALS Orientation and Processing of Airborne Laser Scanning Data, <http://www.ipf.tuwien.ac.at/opals/>, last aCCess: 18.05.2011

PYTHON Python Programming Language, <http://www.python.org/>, last aCCess: 05.05.2011

Modelling light conditions in forests using airborne laser scanning data

Werner Mücke* and Markus Hollaus

Institute of Photogrammetry and Remote Sensing, Vienna University of Technology,
Gußhausstraße 27-29, 1040 Vienna

Abstract

The amount of available sunlight in vegetated areas is an important factor influencing species composition, plant morphology and natural succession. It is therefore a significant parameter in forestry, ecology and other sciences dealing with biodiversity relevant studies. Research indicates a strong correlation between the quality and quantity of sunlight and the vegetation structure, both in horizontal and vertical direction. Due to the high complexity and variability of the canopy architecture, continuous area-wide data collection of light conditions in the understorey is needed for accurate modelling of light transmission. However, conventional ground based measurement methods are pointwise and time consuming, therefore not feasible for data acquisition of large areas.

The ability of small-footprint airborne laser scanning (ALS) to penetrate small canopy gaps makes this remote sensing method especially suitable for vegetation studies. Geometric information of the vegetation structure can be derived directly from the 3D point cloud. This allows for modelling of the distribution of sunlight-absorbing or intercepting parts of the foliage, which consequently cast shadows on the surrounding understorey vegetation or the ground. Light transmission through the canopy can therefore be described in a very direct way by employing this 3D structural information.

In this paper a methodology for modelling light conditions in forests using ALS data is proposed. The approach is based on a modified version of photogrammetric monoplottting. The parallel sun rays from variable sun positions act as projection rays being traced through the 3D point cloud (i.e. laser echoes) that represents the canopy. A defined size is assigned to each individual laser echo which casts a shadow of the respective size and shape. Shadowed areas are then derived by intersecting these projection rays with a digital terrain model and by rasterizing the projected point cloud. By employing ALS data from different acquisition times (leaf-on and leaf-off) the influence of vegetation phenology is explored. The derived shadow raster maps describe where a shadow is cast and how many intercepting parts of the canopy contribute to it. Consequently, these maps provide an excellent input for modelling the amount of available sunlight in vegetated areas, considering canopy gaps in arbitrary directions and also the seasonal variability of vegetation. The first results show that ALS is a time- and cost- efficient means for area-wide analysis of sunlight condition for forest floors, as well as for different understorey layers.

Keywords: *LiDAR, light transmission, vegetation structure, monoplottting, vegetation phenology*

*Corresponding author. Email: wm@ipf.tuwien.ac.at

1. Introduction

The amount and quality of available sunlight in vegetated areas is an important factor having critical effects on sub-canopy air temperature, photosynthesis and soil condition (e.g. soil moisture). It therefore influences species composition, plant morphology and natural succession, all of which are significant parameters for biodiversity relevant studies in order to assess, evaluate and monitor the vegetation's current condition. The scattering process and the interaction of sunlight within the foliage are of a highly complex nature, dependent on biophysical and geometrical features of the canopy [Kimes and Smith, 1980; Woolley, 1971]. It was found that the vertical and horizontal structure of vegetation, as well as the existence and size of canopy gaps have great impact on the distribution of light in vegetated areas [Canham *et al.*, 1990; Comeau and Heineman, 2003; Lieffers *et al.*, 1999; Whitmore *et al.*, 1993]. Knowledge about the canopy architecture, meaning the spatial composition of trees or bushes and the arrangement of their branches and leaves or needles, is therefore of critical importance for the modelling of light transmission.

Conventional methods of measuring the photosynthetically active radiation (PAR), describing the visible spectrum of sunlight that is intercepted (IPAR) or transmitted (TPAR) by the vegetation, are ground based (e.g. directly with quantum sensors or indirectly with hemispherical photographs) [Hardy *et al.*, 2004]. These measurements are time consuming, thus not feasible for area wide acquisition and sometimes requiring frequent repetitions due to seasonality effects [Oshima *et al.*, 1997; Romell *et al.*, 2009]. Modern remote sensing techniques provide a time- and cost-efficient way of data acquisition, enabling wide-area analysis of locations otherwise very hard to reach for ground based inventories. Especially small-footprint laser scanning, also referred to as light detection and ranging (LiDAR), is an observation technique well suited for the derivation of geometric information on the canopy architecture. Recent studies have investigated and successfully applied airborne or spaceborne laser scanning data for the determination of light transmission through the canopy [Jochem *et al.*, 2009; Lee *et al.*, 2009; Parker *et al.*, 2001; Todd *et al.*, 2003].

In this paper, a methodology for the modelling of light conditions in forests employing high-density full-waveform airborne laser scanning (ALS) data is proposed. A point cloud based approach is used to predict patterns of direct sunlight and shadow in a deciduous forest depending on the existence and location of canopy gaps in every part of the vegetation stratum. The proposed procedure is capable of deriving maps showing the distribution of shadowed areas for arbitrary sun positions. By analysing ALS data from different acquisition times the influence of vegetation phenology on the light and shadow distribution can be observed.

2. Study area and data set

The study site is located in the federal state of Burgenland in Austria. The area comprises a forest consisting mainly of deciduous trees and bushes of different stages of succession. For the purpose of this study, an area of 150 x 250 m² with loosely distributed vegetation of different height was chosen in order to have unobstructed areas where the results of the modelled vegetation shadows could be visually observed (see Figure 1). The analysed ALS data were acquired during two flight campaigns in 2010, which were kindly provided by the company *RIEGL Laser Measurement Systems GmbH* within the research project *TransEcoNet* [2011]. For the purpose of vegetation phenology studies the same area was scanned twice: in February under leaf-off and in June under leaf-on conditions. The mean point (i.e. echo) density in open areas was 12.5 pts/m², in overgrown areas sometimes twice this amount and more. A digital surface model (DSM) was derived from the first echoes by selection of the highest points within grid cells of 0.5 x 0.5 m². The digital terrain model (DTM) was calculated based on the last

echoes using hierarchic robust filtering [Briese *et al.*, 2002]. Additionally, a normalized DSM (nDSM) was calculated by subtracting the DTM from the DSM. All three models had a grid width of 0.5 m. Based on the DTM the normalized heights of all echoes were derived and only the echoes above 0.25 m were selected as vegetation echoes for the subsequent step of shadow projection.

For the modelling of light conditions also the position of the sun is needed. The MIDC SOLPOS calculator [2011] was employed to compute azimuth and elevation of the sun's location for hourly intervals and the period of a whole year (2010) for the location of the study site.

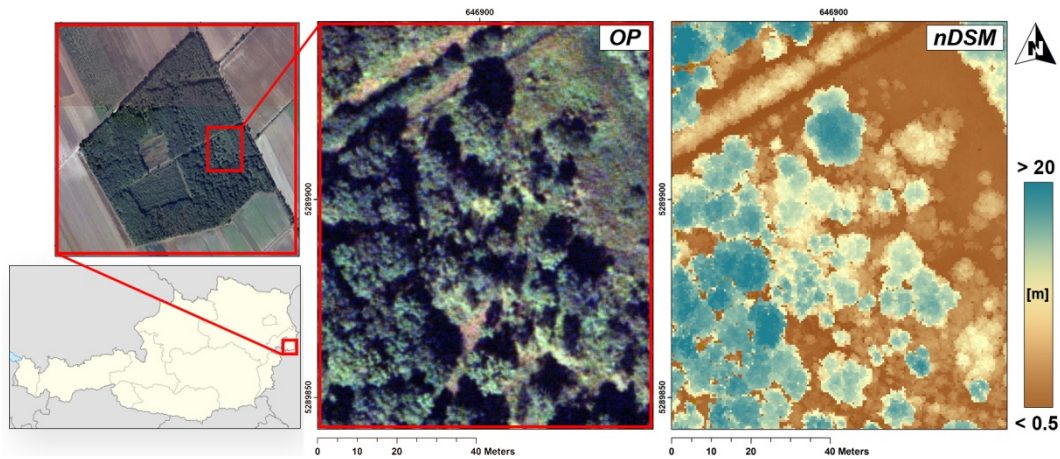


Figure 1: Location and overview of the study area. Central image shows the true-colour orthophoto (OP), the right image the normalized digital surface model based on the leaf-on data (nDSM = DSM - DTM).

3. Method

3.1. Basic concept

The advantage of ALS, as an active measurement system, compared to airborne and spaceborne imagery, is that the laser beams can pass through small gaps in the foliage. It can therefore retrieve information about overgrown smaller vegetation or objects and, in most cases, the laser beams also reach the ground itself. The author's hypothesis is that what is true for laser beams has to apply for sun rays, as well. Every gap the laser can penetrate, also can the sunlight and if the laser beam is intercepted by an object, e.g. a tree trunk or a branch, also is the sunlight. Consequently, a shadow is cast by the intercepting object. If this object is big enough to be covered by the entire footprint of the laser beam (a so-called *extended target*), the total emitted energy contributes to the backscattered signal. However, due to the conical shape of the laser beam, getting wider as it travels through space, it is likely that more than one object is hit with one shot. This is especially true in vegetated areas, where multiple echoes per emitted laser pulse are the norm (so-called *non-extended targets*) [Wagner, 2005]. As a result, the emitted energy is shared among the different scatterers. Due to the complex nature of the scattering process in vegetation we do not exactly know to which part each of the scatterers contributes, meaning we do not exactly know what spatial extent the intercepting objects have. This circumstance also influences the casting of shadows according to our hypothesis.

The basic concept is that a sun ray, considered as a projection ray, travels through space, meets a laser point with a defined spatial extent, which is subsequently projected onto the ground represented by a DTM. A shadow in the shape and size of his projection is cast onto the DTM. The definition of the point's size then relates to the before described problem of not knowing the extent of single scatterers. To deal with this fact it is assumed that for the case of an extended target the point size has to be equal to the size of the entire footprint area. The size of

non-extended targets is to be reduced according to the number of echoes in the shot. For example, if the shot produces three echoes, than the respective footprint area is divided by three and the size of the point is only a third, as is the shadowed area correspondingly. Hence, an extended target casts a bigger shadow than a non-extended, which also finds its trivial equivalent in nature: e.g. bigger branches cast wider shadows. Extensions of this straightforward assumption are presented in the outlook at the end of this paper (see section 5). For the task of projecting the laser points onto the DTM we decided to apply a monoplotting approach, which is explained in detail in the following section.

3.2. The monoplotting approach

The term *monoplotting* in photogrammetry refers to the analytical analysis of curved object surfaces using a single orthorectified image (orthophoto). It is based on the methods of projective geometry, considering the orthorectified image as a central projection of the terrain surface. The approach can be described by three main steps: (1) selecting a pixel of the orthophoto, (2) defining the projective ray through the projection centre and the selected pixel and (3) intersecting this ray with the surface model in order to interpolate the 3D surface coordinate [Kraus, 2007].

For the purpose of shadow projection in this study this monoplotting concept was utilized, however some adaptations to the usage with ALS data had to be made. The projection centre, analogous to the perspective centre of the lens in photogrammetry, in this case is the sun. Because of the large distance from sun to Earth the projection rays appear nearly parallel, compared to the more conical shape in the standard case. Furthermore, instead of pixels in a digital image, the 3D laser points are used to define the projection rays. They originate from the sun, go through each echo and project it down to the DTM (see Figure 2). According to section 3.1, the size of each projected echo is defined by the number of consecutive echoes resulting from an emitted laser shot and the respective footprint size of the laser beam. Calculated from the beam divergence of 0.5 mrad of the employed laser scanner [RIEGL, 2011] and an average flying altitude of 500 m above ground level during this campaign, the diameter of the footprint on horizontal surfaces results in 0.25 m and its area is 0.049 m². Even for the largest trees in the study area (around 21 m), the differences in footprint size due to differences in range are insignificant and therefore ignored. Hence, the same footprint size applies for the calculation of the size of all of the projected echoes.

3.3. Implementation and processing

The monoplotting algorithm was implemented in MATLAB [2011]. As input an ALS point cloud (column-wise in ASCII format including information on echo number) and a DTM (GeoTIFF format), as well as a set of desired sun positions (as described in section 2) for which the projection should be carried out, have to be provided. To be able to observe and discuss the influences of vegetation phenology on the resulting shadowed areas, the leaf-off data set together with the averaged sun positions from February and the leaf-on data set together with the sun positions from June were used. Altogether, the monoplotting had to be done for 10 hourly sun positions in February and for 14 in June. Although SOLPOS returns positions before dawn and after dusk, only azimuths between sunrise and sunset and above 10° of elevation were considered.

The desired accuracy of the interpolation can also be defined as an additional input parameter. For the current study it was set to 0.25 m, which is half the grid width of the DTM. The result of the MATLAB processing is one file for each sun position containing the 3D coordinates of the projected points (x,y,z_{DTM}) and their assigned sizes in m² in a four column ASCII format.

Subsequently, the ASCII files were imported into GRASS GIS [2011] using the module *r.in.xyz*, creating raster data sets with 0.5 m grid size where all the projected point sizes within one grid cell were accumulated. To reduce salt and pepper effects the resulting raster maps were smoothed with a mode filter using the module *r.neighbors* and a square kernel with a size of 3 x 3 pixels. Finally, the georeferenced shadow raster maps were exported with the module *r.out.gdal* to GeoTIFF format and visualized (see Figure 3 and Figure 4).

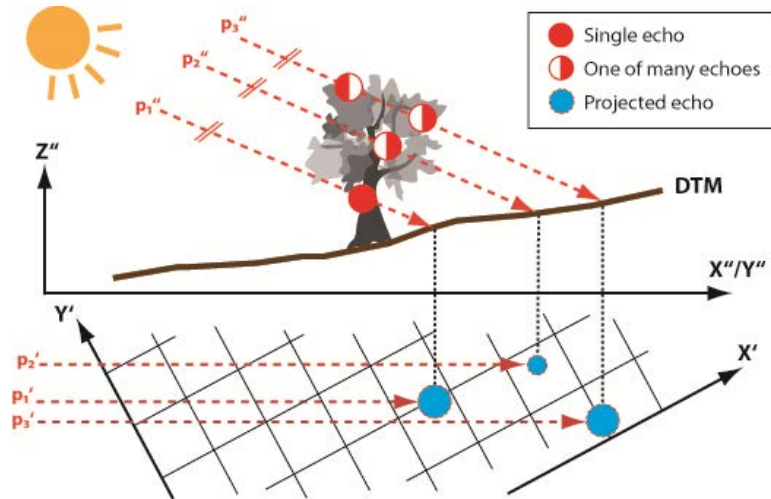


Figure 2: Schematic depiction of the proposed monoploting approach in horizontal (X'/Y') and upright projection (X''/Y''). Parallel sun (projection) rays $p_{1,2,3}$ projecting the ALS points onto the DTM, casting shadows of different size according to the number of echoes within the respective laser shot.

4. Results and discussion

For leaf-on and for leaf-off conditions the shadows were computed for the months February and June, respectively. Figure 3 and Figure 4 each show three examples of the resulting shadow raster maps at three times of day. The informative value of the proposed methodology had to be verified visually, as no reference measurements, e.g. with a quantum sensor, lux meter or a similar instrument, were at hand. The maximum sun elevation angle calculated with SOLPOS for February was roughly 30° , which was around noon. At all other times of day, the sun irradiates in rather flat elevations angles, therefore creating very long shadows, especially for high trees. This can be seen in Figure 3a and c correspondingly, where the high tree in the centre of the image (red number 1) casts his shadow far away (marked as red ellipse), whereas the lower ones cast shorter shadows (compare tree heights in Figure 1). In June the sun elevation angles are steeper, therefore the shadows are generally shorter. This is true for the results shown in Figure 4, where the shadowed areas are smaller and at all times nearer to the shading object. It can also be observed that the differences in acquisition time influenced the result. In Figure 4 the shadows appear generally darker than in Figure 3, meaning that more or “bigger” echoes contributed to the respective shadowed area on the ground. Also Figure 4 creates the impression as if the tree tops cast the darker shadows. The obvious explanation is the fact that during leaf-on conditions the majority of laser echoes come from the top canopy, whereas during leaf-off conditions more penetration takes place and the echoes are more equally distributed over all height levels. For the present study area, 38 % of the vegetation echoes from the leaf-off data set were located above the 80th height percentile, while for the leaf-on data this number increased to 65 %. The sensitivity to the different input data sets suggests that the proposed method could be used for a seasonality dependent modelling of light conditions, thus considering the vegetation phenology in a functional relationship.

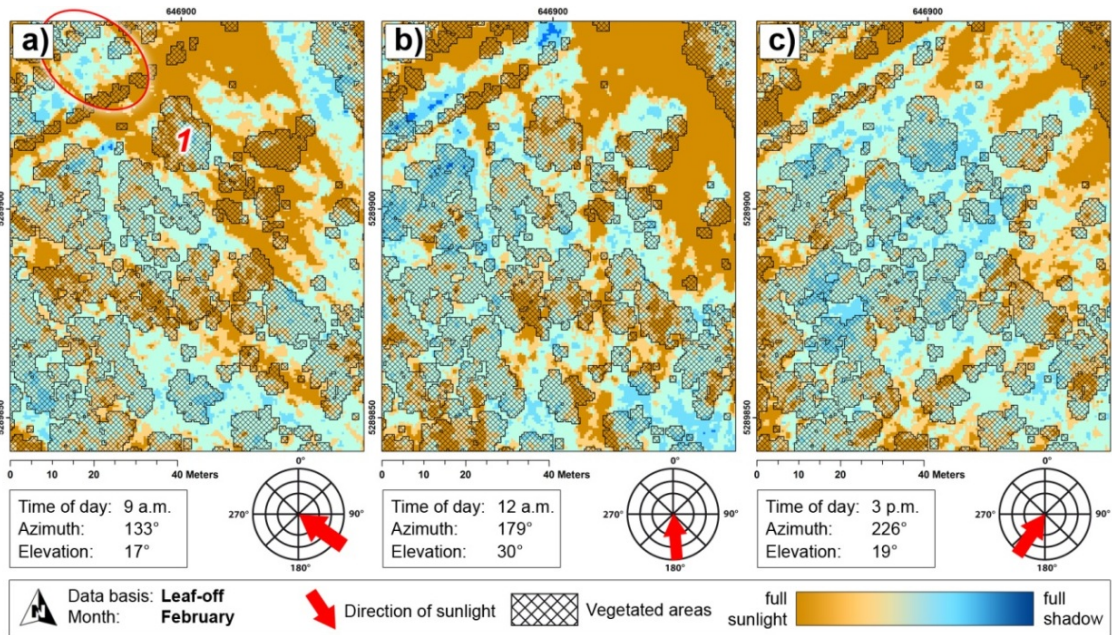


Figure 3: Shadow raster maps based on the leaf-off data (selected vegetation echoes only) and the averaged sun positions from February. Yellow means no shadow is cast, blue means the area is shadowed. The darker the blue, the more or “bigger” echoes contribute to the shadow.

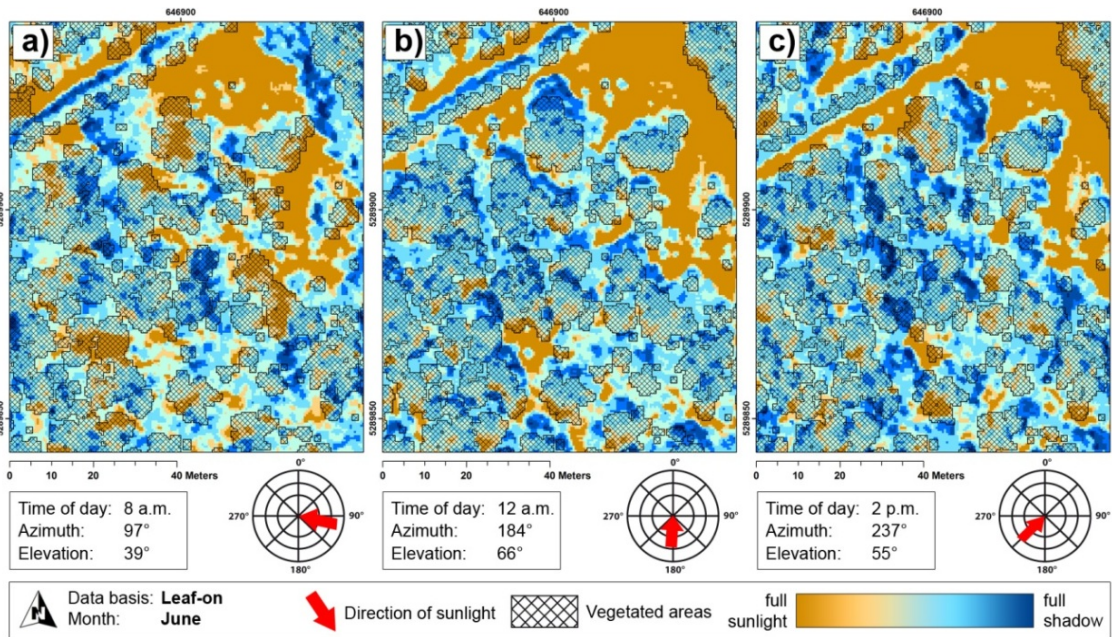


Figure 4: Shadow raster maps based on the leaf-on data (selected vegetation echoes only) and the averaged sun positions from June. The cast shadows are generally darker compared to Figure 3 due to the more dense distribution of the laser echoes in the tree crowns in the summer flight.

The proposed algorithm of shadow projection bears analogies to ray tracing, which was invented in 3D computer graphics for visible surface determination [Agoston, 2005]. As well as ray tracing, monoplotting is a computationally extensive procedure, especially as it was applied in this study for the 3D ALS point cloud. Therefore, the runtime of the MATLAB program with respect to the different employed data sets shall be noted. The leaf-off data consisted of 242595 echoes and the computation of the shadow raster maps took between 2.5 and 3.5 minutes for a single sun position. For the leaf-on data with 361119 echoes computing times ranged from 3.5 to 5 minutes. The differences in processing time within one data set resulted from the fact that

the interpolation step, as it is currently implemented, tends to converge slower with shallow sun elevation angles. To speed up computation, binary input and output could be employed, which would speed up the I/O tasks in MATLAB significantly.

5. Conclusions and outlook

Methods for sunlight and shadow prediction have to consider the fact that shadows cast in overgrown areas are not continuous and of the same intensity or darkness. We applied a point cloud (i.e. vegetation echoes) based method which solves this task effectively and, as shown in the resulting shadow raster maps, quite successfully. The simplified assumption of the projected point sizes being dependent on the number of consecutive echoes in a laser shot (see section 3.1) can be further extended to better reproduce the natural conditions. According to Wagner et al. [2006] the backscattering cross-section, a measure which can be derived from full-waveform ALS data during the task of data calibration, can be referred to as the effective area of collision of the laser beam and an object. Therefore, the cross-section can be employed in order to derive the point size according to the proposed method, representing a much more adequate depiction of the intercepted area of the laser beam. Due to the currently used method of creating the raster maps (accumulating all the projected point sizes within a grid cell; see section 3.3), strongly inclined parts of the DTM would accumulate more points and thus produce darker shadows. This is also the fact for strongly varying point densities in case of heterogeneously distributed scan lines or strip overlaps. Inclined surfaces could be compensated by the introduction of the surface slope and aspect to the algorithm. Varying point densities may be considered by normalisation of the shadow values using a point density map. Furthermore, the topographic shadows, which are currently not considered, could be integrated by applying a line of sight analysis on the DTM. In this way, parts of the surface which are shadowed by others can be found. The total cast shadow then has to be the sum of the topographic and the vegetation shadow. However, these limitations are not significant for the presented study site, as the area is rather flat with low relief energy. On the other hand, the usage of a surface model as projection surface opens up interesting opportunities for the modelling of light conditions in different vegetation strata. A DSM comprising the herbaceous or shrub layer could be created based on selected laser echoes from the respective height levels. Consequently, the light situations in these different parts of the vertical structure could be examined, which provide valuable input for various biodiversity analyses.

Acknowledgements

This study was partly funded by the TransEcoNet project implemented through the CENTRAL EUROPE Program co-financed by the ERDF.

References

- Agoston, M. K. (2005), *Computer graphics and geometric modelling: implementation and algorithms*, Springer.
- Briese, C., N. Pfeifer, and P. Dorninger (2002), Applications of the robust interpolation for DTM determination, *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 34(3/A), 55–61.
- Canham, C. D., J. S. Denslow, W. J. Platt, J. R. Runkle, T. A. Spies, and P. S. White (1990), Light regimes beneath closed canopies and tree-fall gaps in temperate and tropical forests, *Canadian Journal of Forest Research*, 20(5), 620–631.
- Comeau, P. G., and J. L. Heineman (2003), Predicting understory light microclimate from stand parameters in young paper birch stands, *Forest Ecology and Management*, 180(1-3), 303-315.

- GRASS GIS (2011), GRASS GIS - The World Leading Free Software GIS, Available from: <http://grass.fbk.eu/> (Accessed June 2011)
- Hardy, J. P., R. Melloh, G. Koenig, D. Marks, A. Winstral, J. W. Pomeroy, and T. Link (2004), Solar radiation transmission through conifer canopies, *Agricultural and Forest Meteorology*, 126(3-4), 257-270.
- Jochem, A., B. Höfle, M. Hollaus, and M. Rutzinger (2009), Object Detection in Airborne Lidar Data for Improved Solar Radiation Modelling in Urban Areas, *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII(Part 3/W8).
- Kimes, D. S., and J. A. Smith (1980), Simulation of solar radiation absorption in vegetation canopies, *Applied Optics*, 19(16), 2801-2811.
- Kraus, K. (2007), *Photogrammetry. Geometry from Images and Laser Scans*, Gruyter.
- Lee, H., K. C. Slatton, B. E. Roth, and W. P. Cropper (2009), Prediction of forest canopy light interception using three-dimensional airborne LiDAR data, *International Journal of Remote Sensing*, 30, 189–207.
- Lieffers, V. J., C. Messier, K. J. Stadt, F. Gendron, and P. G. Comeau (1999), Predicting and managing light in the understory of boreal forests, *Canadian Journal of Forest Research*, 29(6), 796-811.
- Mathworks (2011), MATLAB - The Language Of Technical Computing, Available from: <http://www.mathworks.com/> (Accessed June 2011)
- MIDC SOLPOS (2011), Solar Codes & Algorithms: Solpos, Available from: <http://rredc.nrel.gov/solar/codesandalgorithms/solpos/> (Accessed June 2011)
- Oshima, K., Y. Tang, and I. Washitani (1997), Spatial and seasonal patterns of microsite light availability in a remnant fragment of deciduous riparian forest and their implication in the conservation of *Arisaema heterophyllum*, a threatened plant species, *Journal of Plant Research*, 110(3), 321-327.
- Parker, G. G., M. A. Lefsky, and D. J. Harding (2001), Light transmittance in forest canopies determined using airborne laser altimetry and in-canopy quantum measurements, *Remote Sensing of Environment*, 76(3), 298–309.
- RIEGL (2011), RIEGL - LMS-Q680 Productdetails, Available from: <http://www.riegl.com> (Accessed May 2011)
- Romell, E., G. Hallsby, and A. Karlsson (2009), Forest floor light conditions in a secondary tropical rain forest after artificial gap creation in northern Borneo, *Agricultural and Forest Meteorology*, 149(6-7), 929-937.
- Todd, F. Csillag, and P. M. Atkinson (2003), Three-dimensional mapping of light transmittance and foliage distribution using lidar, *Canadian Journal of Remote Sensing*, 29(5), 544-555.
- TransEcoNet (2011), TUD - TransEcoNet - Landscapes without borders, Available from: <http://www.transeconet.eu/> (Accessed June 2011)
- Wagner, W. (2005), Physical principles of airborne laser scanning,
- Wagner, W., A. Ullrich, V. Ducic, T. Melzer, and N. Studnicka (2006), Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner, *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(2), 100–112.
- Whitmore, T. C., N. D. Brown, M. D. Swaine, D. Kennedy, C. I. Goodwin-Bailey, and W.-K. Gong (1993), Use of Hemispherical Photographs in Forest Ecology: Measurement of Gap Size and Radiation Totals in a Bornean Tropical Rain Forest, *Journal of Tropical Ecology*, 9(2), 131-151.
- Woolley, J. T. (1971), Reflectance and Transmittance of Light by Leaves, *Plant Physiology*, 47(5), 656-662.

Estimation of Leaf Area Index based on airborne laser scanning and imaging spectroscopy

Pyare Pueschel¹, Henning Buddenbaum¹ & Joachim Hill¹

(1) Department of Environmental Remote Sensing and Geoinformatics, Trier University, 54286 Trier, Germany
(p.pueschel@uni-trier.de)

Key Words: Forestry, LAI, laser scanning, imaging spectroscopy

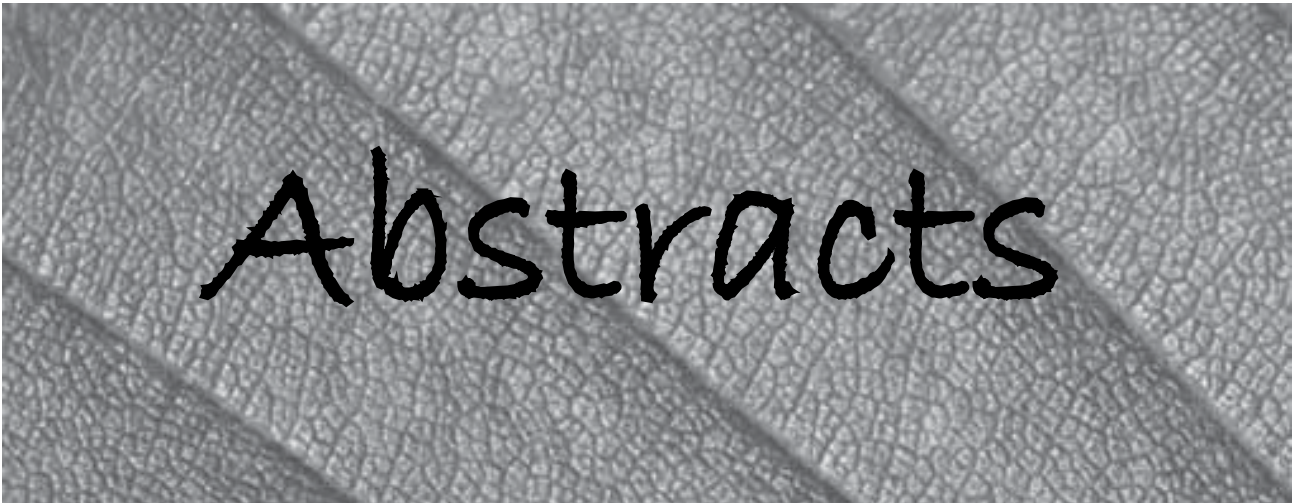
Airborne laser scanning has become an established method for deriving structural forest parameters like tree height, crown length or crown cover. While some parameters are easily retrieved with laser scanning data and imaging spectroscopy, others are not, - the reason for which are model and (ground-based) measurement uncertainties. This is especially true for the estimation of the Leaf Area Index, which practically only allows for indirect measurements in forest environments. Since (a) the LAI is an important input parameter for process-based forest growth models and (b) there is yet no clear preference for using laser scanning or imaging spectroscopy for its retrieval, unlike other structural forest parameters, this study compares the LAI estimation both from airborne laser scanning and imaging spectroscopy and discusses related difficulties.

During an extensive field campaign at one of the EnMAP (Environmental Mapping and Analysis Program) core forest research sites (Merzalben, Pfaelzerwald, Germany) in August 2010, a total of 25 plots in mixed beech-oak forest stands were sampled for LAI. A laser scanning dataset collected almost contemporarily to the HyMap image acquisition served for establishing fractional cover maps with two different methods (1). Corresponding LAI maps were generated based on the formulation proposed by Verger et al. (2) and according to the classic LAI formulation (3). The results were compared to the mean ground-based LAI of the 25 plots. To test the transferability of the LAI retrieval from laser scanning a comparison was made with a second laser scanning dataset from another forest region (4). The LAI retrieved from the laser scanning was then compared to LAI derived from empirical regressions with selected hyperspectral vegetation indices.

Results show that the individual empirical models all provide a good estimation of Leaf Area Index with R^2 values ranging from 0.50 to 0.92. By stratifying according to tree species and stand structure as well as adjusting the LAI-model applied to the laser scanning data, the performance of these empirical models can be greatly improved for both the laser scanning and the HyMap-based vegetation indices. However results also show that both approaches suffer equally from the main limitation of the empirical approach - which is the lack of transferability. For the laser scanning data the estimation is influenced by the combined effects of variable laser scanner system parameters and site-specific canopy variables such as clumping and leaf angle distribution. Further research is needed to assess these effects, particularly with regard to the different laser scanning based gap fraction models. For this purpose we will apply terrestrial laser scanning to validate both the airborne laser scanning derived gap fraction and the ground-based measurements.

References

- 1 Hopkinson C & L Chasmer, 2009. Testing LiDAR models of fractional cover across multiple forest ecozones. Remote Sensing of Environment, 113: 275–288
- 2 Verger A, F Camacho, FJ García-Haro & J Meliá, 2009. Prototyping of Land-SAF leaf area index algorithm with VEGETATION and MODIS data over Europe. Remote Sensing of Environment, 113: 2285–2297
- 3 Monsi M & T Saeki, 2005. On the factor light in plant communities and its importance for matter production. Annals of Botany, 95: 549-567
- 4 Buddenbaum H & J Hill, 2010. Retrieval of LAI from Airborne Hyperspectral and Airborne Laser Scanner Data using a Canopy Reflectance Model. ESA Hyperspectral Workshop 2010, 17–19 March 2010, Frascati, Italy.



Abstracts

A national review of airborne lidar application in Australian forest agencies

Russell Turner ¹, Nicholas Goodwin ², Jeremy Friend ³, David Mannes ⁴, Jan Rombouts ⁵, Andrew Haywood ⁶

¹ Forest Science Centre, Department of Primary Industries New South Wales, Sydney, NSW

² Remote Sensing Centre, Department of Environment and Resource Management, Ecosciences Precinct, Dutton Park. QLD

³ Planning, Environment and Silviculture, Forest Products Commission, Bunbury, WA

⁴ Resource Information, Forestry Tasmania, Hobart, TAS

⁵ Resource Planning, Forestry South Australia, Mount Gambier, SA

⁶ Resource Planning, Department of Sustainability and Environment, VIC

This paper provides a narrative of airborne lidar application across Australian forest agencies. It includes a brief history of early lidar research and operational trials, as well as current programs and future directions on a state by state basis. This review demonstrates a diverse range of lidar applications and increasing adoption of lidar technology within state agencies across Australia.

Airborne LiDAR based forest inventory in Bangladesh for REDD plus MRV: scope and potentiality

Parvez Rana* ^{1,2}, Hanna Holm ³, Tuomo Kauranne ³

¹ School of Forest Sciences, University of Eastern Finland, Finland

² Department of Forest Resource Management, Swedish University of Agricultural Sciences, Sweden

³ Arbonaut Ltd. Kauppakatu 21, Finland

Nowadays, the accurate measurements of carbon stock for carbon trading in REDD plus (Reducing Emissions from Deforestation and Forest Degradation in Developing Countries) countries are going highly demanding. IPCC (Intergovernmental Panel on Climate Change) Tier 3 level accuracy for estimation of emissions from deforestation and forest degradation requires detailed national inventory of key carbon stocks, repeated measurements and modeling. Present study has been carried out to know scope and potentiality of the airborne LiDAR based forest inventory in Bangladesh for REDD plus MRV (monitoring, reporting and verification). Here we supposed a hybrid method where the integration of airborne LiDAR data with satellite imagery and ground truth data based forest inventory in Bangladesh. As the forest of Bangladesh is highly dynamic and inaccessible due to hilly and mountainous area, this method will give an accountable and transparent report of carbon stock. We also highlighted the limitation of this approach in a developing country like Bangladesh due to poor economic and technical condition. Till now there is no record of application of airborne LiDAR system for forest inventory in Bangladesh. Finally, we recommended that the Forest Department of Bangladesh with financial and technical help from international organization can do a pilot project in Sundarban Mangrove Forest.

Stand level inventory of eucalypt plantations using small footprint LiDAR in Tasmania, Australia

Robert Musk

Models derived using Brieman's Random Forests algorithm have been identified in past studies as having greater predictive accuracies than those derived using nearest neighbour imputation approaches. This is attributed to the algorithms ability to model complex interactions among predictor variables and its resistance to overfitting. These two properties are of particular value in modelling LiDAR-derived variables where strong colinearity is a common feature. In this study, the random forest algorithm is applied to a large inventory dataset to generate mapped estimates of forest stand structure. The ability of the algorithm to identify an optimal set of candidate variables is assessed by means of an iterative model fitting procedure. The study area comprises a eucalypt hardwood plantation estate in northern Tasmania, Australia. Model pseudo R2 values were 74.6% for basal area, 96.0% for mean dominant height, 64.2% for stocking and 83.9% for merchantable stand volume respectively.

Ground based and airborne lidar - a natural combination

David Jupp

When, at the end of the 20th century, a group from CSIRO (Australia) was evaluating support (aka seeking funding) for their ideas about combined airborne and ground based Lidar, the ground based system they proposed was seen as an adjunct for on-ground checking. But their common experience with forestry and environmental companies was lack of interest in airborne and amazing enthusiasm for ground based. Perhaps the idea of a measurement tool that was not remotely operated by someone else but was in the forest and operated by them was what was attractive. Whatever the reason, that group has been focused on ground based Lidar ever since. I am sure this is not an uncommon experience, but perhaps ground based systems are not yet fulfilling that observed interest to become an ubiquitous component of forest measurement.

One possible reason for this is that ground based (GB) and airborne (AB) systems need each other and that will be the somewhat “rhetorical” topic of this talk. A major combined use that has been suggested is that GB can be used to “calibrate” or “validate” AB. “Calibrate” means setting parameters for data interpretation of AB such that it generates biomass or something else of interest over a wide area. “Validate” means to ensure that such wide area information is staying sufficiently close to the truth to be useful. If this were not enough, people may suggest that GB samples an area and AB can extend the sampling to a wide area containing the samples. But this is not too different from calibration. Again, in tall and dense forests the two systems may provide “handshaking” between information about the upper and lower parts of the forest.

It is tempting to look at GB and AB as the same activity from different “platforms”, but this is not quite correct. GB systems are really not necessarily fixed to the “ground” but rather work from a fixed reference coordinate system allowing multiple angles and multiple volumes sampling from a number of locations – even from above the canopy. I believe AB is characterised by the dynamical nature of its collection system and is much more similar to space-borne (SB) systems than to GB systems. Accepting this definition, we can consider measurement strategies for GB systems that will be called “extensive” strategies and “tomographic” strategies. The first aims to collect as much information as possible from a single point of reference at a site and sample as many different (but possibly not overlapping) sites within a wide area as possible. The second aims to use multiple points of reference to sound the same volume to maximise information about the volume – which is usually necessarily a single site. Each of these must deal with the complex combinations of gaps, hits and occlusion within a forest and they tend to do it in different ways.

Within this framework, the talk will use the experiences of scientists mostly at CSIRO (Australia), Boston University, City University of New York and University of British Columbia to discuss the potential for successful integration of AB and GB systems. It involves issues of data compatibility (such as the use of intensity information, waveforms and calibration), sampling versus tomography and consistency of processing and interpretation models. It will be proposed that ultimately, the realisation of the integration involves deriving a “transfer function” between measurements in the two systems. This function must, by the nature of the systems, be statistical and can be empirical or model based. Of all the possibilities, model based transfer is the one that probably interests scientists most and the talk will illustrate how the groups above have progressed towards this end. Finally, it should be pointed out that there may be more answers to the questions posed here in the presentations at the conference than there are in the talk. Hopefully, that will be the case. The main objective is to address the issues and look to the work of the many scientists now working in AB and GB Lidar to make it happen.



Harvesting productivity analysis using LiDAR

Muhammad Alam*, Martin Strandgard, Mark Brown, Julian Fox

Department of Forest and Ecosystem Science, The University of Melbourne

Mechanised harvesting operations are common in Australia because of their increased productivity and efficiency, improved worker safety and reduced cost of operations. Most research has found that the productivity and efficiency of a mechanised harvesting system is affected by a number of factors including forest stand characteristics (tree size or piece size, stand density, undergrowth), terrain variables (slope, rocks, woody debris), operators' skill and machinery limitations. The purpose of the study was to use remote sensing technology to quantify these forest stand and terrain factors (particularly slope) and hence derive relationships to predict harvester productivity from remote sensing data. A case study was conducted in mature radiata pine (*Pinus radiata*) plantation at Mount Burr Reserve Forest, South Australia (37.61° S, 140.44° E). LiDAR (Light Detection And Ranging) flown in 2007 was used to identify and quantify stand and terrain factors (particularly tree size). A time and motion study conducted during final harvest was used to estimate the impact of each factor (tree size and slope) on harvester productivity. Tree size estimates derived from the LiDAR data were grown to the point of harvest using empirical growth models. The point of harvest tree size estimates were ground-truthed against harvester measurements of the same trees. Empirical models were then developed to enable the LiDAR-derived estimates of tree size to be used to estimate productivity of harvesting equipment. The robustness of these relationships will be tested by applying the model to areas not used in the development process.

Scaling plot to stand-level lidar to province in a hierarchical approach to map forest biomass in Nova Scotia

Chris Hopkinson, David Colville*, Danik Bourdeau, Suzanne Monette, Robert Maher

This paper presents a study that used lidar transect, plot and wide area polygon sample data collected across Nova Scotia, Canada from 2005 to 2010 to calibrate and extrapolate above ground forest biomass from permanent sample plots (PSPs) to forest stand polygons to the entire Province. The whole tree dry biomass estimate for the total forest resource inventory (FRI) database in Nova Scotia is ~ 373x 106tonnes ±39%. Where lidar coverage exists, biomass is modelled at the 25 m grid cell resolution, which is a great improvement over the previous eco region level estimates, allowing for more effective operational stand management. Given the large spatio-temporal domain of the data sources, one of the major challenges faced in this study was temporal latency between coincident field, lidar and GIS data inputs, which was a significant contributor to the overall level of uncertainty in the result.

Estimating stand volume from nonparametric distribution of airborne LiDAR data

Doo-Ahn Kwak*¹, Taejin Park², Jong Yeol Lee², Woo-Kyun Lee²

¹Environmental GIS/RS Centre, Korea University, South Korea

²Division of Environmental Science and Ecological Engineering, Korea University, South Korea

This study was performed to estimate stand-level volume using the characteristics of vertical and horizontal distribution of airborne Light Detection And Ranging (LiDAR) data. It is found that the height distributional parameters, such as percentile, of LiDAR data reflected on-and in-canopy in a stand have the relationship with stand volume in previous research. However, we assumed that the nonparametric height distribution form of canopy LiDAR returns would be obviously related with the stand volume directly. Nonparametric height distribution was presented to be a continuous line according to the frequency of LiDAR returns by the height. Thereafter, the sum of each height of all canopy returns, which means the area below the continuous line, was compared to stand volume using National Forest Inventory (NFI) data. In addition, for verifying the volume of test stands, the similarity which is the overlapping ratio between the height distribution curves of sample and test stand was calculated. The relationship between the height sum and stand volume was relatively high to be $R^2=0.83$. Based on such relationship, the maximum similarity of each test stand was computed as compared sample stands. As a result, mean similarity and root mean square error (RMSE) of estimated stand volumes were 82% and 34.96m³/ha respectively. However, supplementary indices, for non-overlapping part in similar distribution of canopy returns of sample and test stand, are needed to reduce such errors.

A method for linking TLS- and ALS-derived trees

Andreas Fritz*, Holger Weinacker, Barbara Koch

University of Freiburg, Department of Remote Sensing and Landscape Information Systems

Within the past decade progress towards automatic recognition of individual trees and their parameters was made in both TLS and ALS-data based algorithms. In this paper we present an approach to combine single trees derived from ALS and TLS-data in order to gain a higher level of information. Therefore, two data sets are used: 1. a set of 3D-stemfiles generated by the algorithm described in Bienert *et al.* 2007 and 2. a set of detected single trees for the corresponding area of the data set 1 based on the algorithm described in Gupta *et al.* 2010. The 3D-stemfiles include position, information regarding sweep and diameter in 10cm height intervals. The ALS-tree description covers the position, maximum crown diameter and length as well as tree top height. This information is used for a hierarchic approach of linking ALS and TLS-derived trees based on three different initial matching algorithms. The estimated position error is taken into account to generate an initial list of matching candidates. The 2D-distance based initial linking method linked 41% of the TLS-trees. It was found that 3D-estimation of the tree top based on sweep information of the TLS-trees led only to minimal more imputations than the 2D-approach. A possible reason is seen in the linear models chosen, which do not reflect the tree shape invariably. Future work focuses on the integration of species information and the quantification of false linkage, which could not be evaluated within this study.

Reducing extrapolation bias of area-based k-nearest neighbour predictions by using individual tree crown approaches in areas with high density airborne laser scanning data

Johannes Breidenbach*^{1,2}, Erik Næsset², Terje Gobakken²

¹Norwegian Forest and Landscape Institute

²Norwegian University of Life Sciences

K-nearest neighbour (kNN) approaches are popular statistical methods for predicting forest attributes in airborne laser scanning (ALS) based inventories. Their main upsides are the simplicity to predict multivariate response variables and their freedom of distributional assumptions on the conditional response. One of their largest draw-backs is that predictions outside the range of the reference data inherently result in an under-or overestimation. This property of kNN approaches is known as extrapolation bias and aggravates with an increasing number of neighbours (k) used for the prediction. This study presents one possibility to reduce extrapolation biases of predictions based on the area-based approach (ABA) by using individual tree crown (ITC) approaches within those specific areas of a low density ALS acquisition where the point density might be sufficiently high for using ITC methods. In the proposed strategy, additional (or artificial) reference plots augmented field measured plots. Artificial plots were created by applying ITC segmentation to a canopy height model derived from high density ALS data. The response variable biomass per hectare was predicted for every segment following a semi-ITC approach. The segment predictions were aggregated at the artificial plot level. The artificial plots were then treated in the same way as the original reference data to make predictions in areas with low density ALS data based on the ABA. It was hereby assumed that the predicted plot level response on the artificial plots is equivalent with the observed plot level response on the original reference data. The data consisted of 110 reference plots with a smaller data range than the 201 independent validation plots. Considerable extrapolation bias was visible if only the reference plots were used for the prediction. Almost no extrapolation bias was found if the prediction was based on reference plots augmented by artificial plots. The root mean squared error (RMSE) of the biomass predictions based on the reference plots was 39.1%. The RMSE reduced to 29.8% if the reference plots were augmented by artificial plots.

Tree biomass estimation using ALS features

Minna Rätty*¹, Ville Kankare¹, Xiaowei Yu², Markus Holopainen¹, Mikko Vastaranta¹, Tuula Kantola¹, Juha Hyypä², Risto Viitala³

¹Department of Forest Sciences/University of Helsinki, Finland

²Finnish Geodetic Institute

³HAMK University of Applied Sciences, Finland

Today the estimation of biomass and detection of changes in biomass in large areas is based on coarse remote sensing data and field measurements, which are time consuming, expensive and, above all, in local level inaccurate. The recent development of techniques has offered opportunities to develop new methods, e.g. laser scanning. Airborne laser scanning (ALS) derived features could be used to estimate the total biomass of standing trees. The objective of this study was to make preliminary investigations between accurately measured biomasses in the field and ALS derived features. Study material consisted of 38 sample trees: 19 Scots pines (*Pinus sylvestris*) and Norway spruces (*Picea abies*), which biomasses were accurately measured. ALS derived segments representing the field trees were matched and features for trees were extracted from ALS points within segments. Correlations between biomasses and ALS features were calculated and simple regression models were formulated. The relative residual errors were 21% for Scots pine and 40% for Norway spruce. More empirical tests are needed for ALS based tree biomass estimations.

Stand level species classification and biomass estimation using LiDAR height, intensity, and ratio parameters

Taejin Park*¹, Doo-Ahn Kwak², Woo-Kyun Lee¹, Jong-Yeol Lee¹

¹Department of Environmental Science and Ecological Engineering, Korea University, Korea

²Environmental GIS/RS Centre, Korea University, Seoul, South Korea

In this study we use airborne LiDAR to classify tree species and estimate volume at the stand scale using multiple linear discriminant analysis and multiple linear regression analysis. This involved the extraction of 38 independent variables from LiDAR data including height, intensity, and ratio metrics. In stand species classification, the 90 percentile of height (*HC,90*), standard deviation of the intensity (*IC,std*) and vegetation intensity ratio (*VIR*) were the most suitable variables for explaining each stand species. Hit ratio represented by accuracy in discriminant analysis was 81.7% in stand species classification. Afterward, the regression models were estimated using each variable, with the best model then selected using the corrected Akaike's Information Criterion (AICc). *HC,90*, mode of intensity (*IC,mode*) and standard error of mean of intensity (*IC,se*) were applied to optimally explain the stand volume of Japanese Larch (*Larix leptolepis*), with an $R^2=0.83$. With the mean of height (*HC,mean*), mode of height (*HC,mode*), standard deviation of intensity (*IC,std*) and range of intensity (*IC,range*) could be used to predict the stand volume of Japanese red pine (*Pinus densiflora*), with an $R^2=0.79$. Finally, the 80th height percentile (*HC,80*), *IC,mode* and the kurtosis of intensity distribution (*IC,kurt*) were applied to predict the stand volume of Oaks (*Quercus* spp.) with an $R^2=0.68$.

Effect of scan coverage on stem diameter measurement using terrestrial LiDAR

Akira Kato*¹, L. Monika Moskal², Tatsuaki Kobayashi³

¹GraduateSchool of Horticulture, Chiba University, Japan

²Precision Forestry Cooperative, College of the Environment, School of Forest Resources, University of Washington, USA

³GraduateSchool of Horticulture, Chiba University, Japan

This paper presents a new approach to measure stem diameters based on the data acquired by multiple scanning by terrestrial lidar. Recent terrestrial lidar (Riegl VZ400) has wider coverage and is able to efficiently

provide the highest point density data. Stem diameter derived from terrestrial lidar was compared with field measured diameter at breast height (d.b.h) of 42 sample trees. Stem returns of d.b.h were extracted and used to identify the approximated stem centre using principal component analysis. Various scan coverage of stem returns was used in the algorithm developed in this study to assess which is the most appropriate to measure stem diameter. The results show that more than 40% scan coverage of stem returns can produce stem diameter with an error of 5 cm or less using the algorithm. The applied technique can also assess the quality of wood by estimating straightness of stems from the alignment of stem centres at several heights. Furthermore, stem volume which is the most important variable to estimate the amount of carbon can also be measured directly using this technique.

Stem curve measurement using terrestrial laser scanning

Xinlian Liang*¹, Juha Hyyppä¹, Ville Kankare², Markus Holopainen²

¹Finnish Geodetic Institute, Department of Remote Sensing and Photogrammetry, Masala, Finland

²University of Helsinki, Department of Forest Resource Management, Finland

Terrestrial laser scanning (TLS) has been shown to be a promising technology for the accurate forest inventory on the sample plots. The advantages of applying TLS can be improving the accuracy and efficiency of the field measurements. In addition, TLS data have the possibility to provide more tree parameters than what are commonly accepted and employed at the moment. This paper discusses the automatic measurement of the stem curve using TLS. A pine and a spruce were used in the experiment. The stem curve estimated from point cloud was compared to the field measurements. The experiment shows that the estimation of the stem curve from single-scan and merged point clouds are comparable to each other. This result indicates that TLS data has the potential to automatically estimate the stem curve.

Estimating single-tree branch biomass of Norway spruce by airborne laser scanning

Marius Hauglin*, Janka Dibdiakova, Terje Gobakken, Erik Næssetc

Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences

Dry weight of the branches of 20 trees of Norway spruce was obtained through destructive sampling. Airborne laser scanning data from the same trees were used to calculate crown volume for each tree. The crown volume was derived by using the crown laser echoes with a radial basis function to construct a crown surface. A regression model was fitted to the data, with the crown volume as explanatory variable and the dry weight of the branches as response. The model revealed a strong relationship between the two, with $R^2 = 0.80$. A leave-one-out cross-validation gave a root mean square error of 34%.

Airborne laser scanning-based stem volume imputation in a managed, boreal forest area: a comparison of estimation units

Jari Vauhkonen*¹, Petteri Packalén¹, Juho Pitkänen²

¹University of Eastern Finland, School of Forest Sciences, Joensuu, Finland

²Finnish Forest Research Institute, Joensuu Research Unit, Joensuu, Finland

In typical airborne laser scanning (ALS)-based inventories, the forest is aggregated from initial estimation units, for which the attributes are produced using variable imputation techniques. The initial units vary in size and shape, being usually either regular grid cells or segments derived from the ALS data. This study compared small grid cells and segments of trees or tree groups as initial estimation units in an ALS-based estimation of species-specific, plot-level volume. The experiments were carried out in a managed, boreal forest area in Eastern Finland, where pine was the dominant species, and spruce and deciduous trees formed the other species groups. The field data consisted of 79 sample plots (400–900 m² in area) and the ALS data had a density of about 12 pulses/m². The estimation was overall very accurate, resulting in best-case root mean squared errors

Applying terrestrial LiDAR to derive gap fraction distribution time series during bud break

Kim Calders*¹, Jan Verbesselt¹, Ham Bartholomeus¹, Martin Herold¹

¹Laboratory of Geo-Information Science and Remote Sensing, Wageningen University

The scientific community is witnessing a significant increase in the availability of different global satellite derived biophysical data sets. However, the use of such data is currently not supported by accurate in-situ biophysical measurement (e.g. canopy structure) in both a research and operational context for the monitoring of forest and land dynamics. Consequently, there is an urgent need for methods to measure in-situ canopy structure accurately and better integrate with improved and innovative remote sensing approaches. This paper explores the use of a ground-based, upward looking LiDAR instrument, combined with a fully automated analysis method to retrieve the gap fraction distribution. Traditional inventory methods for the assessment of forest structure are less objective or based on a 2D approach. We compare the seasonal dynamics of gap fraction distribution from hemispherical photographs and terrestrial LiDAR measurement during bud break.

Preliminary analysis shows that gap fraction distributions derived from terrestrial LiDAR were consistently lower than the values obtained from hemispherical photography. This might indicate that the LiDAR scans at the centre position of the plot are not representing the plot scale variation. However, the LiDAR based methodology is fully automated, requires no operator interference and is more objective, whereas the analysis of hemispherical photographs requires a large number of operator decisions (e.g. thresholding). Further improvements of this LiDAR-based method can still be achieved by (i) a better understanding of scanner settings and data resolution on the derived gap fraction and (ii) integration of target intensity in the analysis. This paper highlighted the high potential and need for a robust method to derive gap fraction distributions to monitor seasonal dynamics in forests.

Foliage profiles from ground based waveform and discrete point LiDAR

Jenny Lovell*¹, David Jupp², Eva van Gorsel², Jose Jimenez-Berni², Chris Hopkinson³, Laura Chasmer⁴

¹CSIRO Marine and Atmospheric Research, Hobart, Australia

²CSIRO Marine and Atmospheric Research, Canberra, Australia

³Applied Geomatics Research Group, Lawrencetown, Canada

⁴Wilfrid Laurier University, Waterloo, Canada

Terrestrial lidar systems provide a means to characterise the structure of a forest canopy. Their use to measure foliage area volume density depends on the ability to account for sampling effects and intensity calibration of the instrument. This paper presents a theoretical framework for the unbiased calculation of foliage amount using a waveform recording lidar instrument to simulate point cloud data. The method is initially based on the hemispherical scan configuration of the instrument, but is generalised to be applied to point cloud data in a generic coordinate system. The theory is tested with the simulated point cloud data as well as data from a commercial instrument. Foliage profiles from the terrestrial lidar instruments and airborne lidar are compared.

Generating an automated approach to optimize effective leaf area index by Canadian boreal forest species using airborne LiDAR

Heather Morrison*², Chris Hopkinson¹, Laura Chasmer³, Natascha Kljun⁴

¹Acadia University, Wolfville, Nova Scotia, Canada

²Applied Geomatics Research Group, NSCC Annapolis Campus, Nova Scotia, Canada

³Cold Regions Research Centre, Wilfrid Laurier University, Waterloo, Ontario, Canada

⁴Department of Geography, Swansea University, Swansea, Wales, United Kingdom

Obtaining forest structure data to compute leaf area index (LAI) can be a challenge in remote areas like the Canadian boreal forest. Light ranging and detection (LiDAR) data provides a 3-dimensional view of the forest

that can be calibrated with minimal field data requirements relative to other remote sensing data. Our objective is to develop an automated method for combining a limited amount of field data with LiDAR to generate estimates of LAI. To accomplish this we used geographic information system (GIS) tools to expand upon a physically-based gap fraction model by incorporating a process for optimizing extinction coefficient by forest species. In this paper we demonstrate a simple, efficient method for optimizing remote sensing-based estimates of canopy attributes from limited field data. We were able to reduce the RMSE in modelled effective leaf area index by an average of 0.48 across all species. Combining such simple model optimisation approaches with other automated LiDAR-based canopy attribute extraction procedures shows promise as we move towards ever greater levels of LiDAR forestry operationalisation.

Change detection of mountain vegetation using multi-temporal ALS point clouds

Mattias Nyström *, Johan Holmgren, Håkan Olsson

Section of Forest Remote Sensing, Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Sweden

Multi-temporal laser scanner data to be used in change detection studies will most likely be acquired with different sensors, flying altitudes, and system parameters. Therefore, calibration is probably needed in order to make laser returns from vegetation comparable between two laser data acquisitions. In this study, two ALS point clouds were acquired with different sensors and flying altitudes. The first data set had 11.5pointsm⁻² and was obtained in 2008 with a Top Eye MKII scanner and the second with a density of 1.1pointsm⁻² was obtained in 2010 with an Optech ALTM Gemini scanner. The test site was located in Abisko in northern Sweden with forest dominated by mountain birch. Six meter radius sample plots were placed in the forest-tundra ecotone and assigned one of the following treatments: (1) reference with no removal of trees, (2) removal of 50% of the total number of stems above 1.5m, and (3) removal of 100% of the total number of stems above 1.5m. Histogram matching was used to calibrate the two data sets and sample plots were then classified into the three treatments. The overall classification accuracy was 82% using only the proportion of vegetation returns from the canopy as explanatory variable. Features created from gridded laser data had overall higher classification accuracy than laser features created directly from the point cloud. Histogram matching made the two data sets comparable by reducing the difference between them. These early results show how changes can be detected even with different sensors, flying altitudes, and system parameters.

Stability of LiDAR-derived raster canopy attributes with changing pulse repetition frequency

Allyson Fox* ^{1,2}, Chris Hopkinson ^{1,2}, Laura Chasmer ³, Ashley Wile ²

¹Acadia University, Wolfville, Nova Scotia, Canada

²Applied Geomatics Research Group, NSCC Annapolis Campus, Nova Scotia, Canada

³Cold Regions Research Centre, Wilfrid Laurier University, Ontario, Canada

Laser pulse characteristics (pulse emission rate and inherent pulse properties) influence the representation of forest canopy structure using LiDAR data. As the use of LiDAR-derived models for large scale forest canopy characterization increases, there is a need to optimize flight configuration settings to achieve this efficiently, and to ensure that changes observed in multi-temporal growth studies are due to forest change and not flight configuration influences. Using an Optech Inc. ALTM 3100 airborne LiDAR sensor pulse repetition frequency (PRF) was systematically varied over seven flights, in a three hour period, over Acadian mixed-wood forest plots in Nova Scotia, Canada in July 2005. Canopy height and fractional cover models were created and analysed to determine if differences in sensor configuration settings influence typical LiDAR-derived raster representations of canopy structure. Preliminary findings for both canopy height and fractional cover models are evaluated and discussed.

Comparison of the spatial pattern of trees obtained by ALS based forest inventory techniques

Petteri Packalén*³, Jari Vauhkonen¹, Eveliina Kallio¹, Jussi Peuhkurinen¹, Juho Pitkänen², Inka Pippuri¹, Matti Maltamo¹

¹Faculty of Science and Forestry, University of Eastern Finland, Joensuu, Finland

²Finnish Forest Research Institute, Joensuu Research Unit, Joensuu, Finland

³College of Forestry, Oregon State University, USA

The spatial pattern of trees in a forest can be defined as the locations of the trees in relation to each other. The spatial arrangement of a point (e.g. tree) pattern may be random (Poisson), clustered or regular. In this study the spatial pattern of trees was determined at the plot level by using L function, which is a square root transformation of Ripley's K function. The spatial pattern of tree was summarized in to three classes: regular, random and clustered. The study was carried out with 79 sample plots located in a managed forest area in eastern Finland. Tree maps were produced with the individual tree detection (ITD) and semi-individual tree detection (Semi-ITD) and spatial patterns of trees were calculated from the tree coordinates. The spatial pattern of trees was also predicted directly by using patch metrics calculated from the canopy height model as explanatory variables (AREA). The low resolution airborne laser scanning (ALS) data was used in the AREA and the high resolution data in the ITD and Semi-ITD. The Kappa value for the ITD was almost zero, which indicates virtually random classification. The AREA and Semi-ITD methods were clearly more accurate than the ITD. Kappa values for the Semi-ITD and AREA were 0.34 and 0.24, respectively, which nevertheless cannot be considered to be very good. However, determining the spatial pattern of trees by ALS is somewhat unexplored field of study. It should be studied how well the spatial pattern of trees can be determined in different type of forests.

Fusion of airborne LiDAR and WorldView-2 MS data for classification of depth to permafrost within Canada's sub-Arctic

Laura Chasmer*¹, Chris Hopkinson², Heather Morrison², Richard Petrone¹, William Quinton¹

¹Cold Regions Research Centre, Wilfrid Laurier University

²Applied Geomatics Research Group, NSCC, Lawrencetown

The discontinuous permafrost zone of north-western Canada is characterised by a heterogeneous landscape of tree-covered permafrost plateaus that rise 0.5 m to 2.0 m above the surrounding fens and bogs. The depth to permafrost or "frost table" is influenced to some extent by vegetation canopy cover, which drives complex feedbacks related to permafrost thaw. Spectral remote sensing offers the possibility of large area mapping of canopy and ground surface characteristics that may be used as a proxy for permafrost thaw within remote northern areas. However, this depends on whether or not spectral band scan be used to identify slight variations in vegetation characteristics. The following study compares vegetation and topographic characteristics obtained using airborne Light Detection And Ranging (LiDAR) with high spatial resolution WorldView-2 spectral bands and *in situ* transect measurements of the depth to frost table. The results of this study indicate that the depth to the frost table is related to above ground vegetation cover and tree height, yet relationships are complicated by canopy and under story characteristics, topographic derivatives, and the position of the measured frost-table transect within the fragmented plateau. Comparisons between vegetation structural characteristics and WorldView-2 spectral bands are also examined so that confidence can be applied to depth of frost table estimates from WorldView-2 based on canopy characteristics. Discrete WorldView-2 pixels are related to depth to frost table (bands red, near infrared1,2) and canopy metrics/topography obtained from airborne LiDAR. Variability is due, in part to absorption of near infrared by shadow fractions observed within WorldView-2 pixels, and spectral reflectance of ground vegetation visible within mixed pixels. High resolution spectral imagery, therefore, provides a link to processes controlling spatial variability of the depth to frost table.

Using high density ALS data in plot level estimation of the defoliation by the Common pine sawfly

Tuula Kantola*¹, Paivi Lyytikäinen-Saarenmaa¹, Mikko Vastaranta¹, Ville Kankare¹, Xiaowei Yu², Markus Holopainen¹, Mervi Talvitie¹, Svein Solberg³, Paula Puolakka⁴, Juha Hyyppä²

¹University of Helsinki, Finland

²Finnish Geodetic Institute

³Norwegian Forest and Landscape Institute, Norway

⁴Finnish Forest Research Institute, Vantaa, Finland

The climate change has been related to the increase of forest insect damages in the boreal zone. The prediction of the changes in the distribution of insect-caused forest damages has become a topical issue. The common pine sawfly (*Diprion pini* L.) is regarded as a significant threat to boreal Scots pine (*Pinus sylvestris* L.) forests. Efficient and accurate methods are needed for monitoring and predicting changes in insect defoliation. In this study, the field work has been carried out in 2009 in Eastern Finland, where *D. pini* has caused considerable damage in managed Scots pine forests. Altogether 95 sampling plots were used in the analysis. A high density ALS data was acquired simultaneously with the field work. The aim of the present study was to test the accuracy of the plot level needle loss predictions determined from the area based and single tree ALS features separately. The Random Forest method (RF) was utilized in the estimation. The best classification accuracy for the test set was 67.4% (area based features). The best plot level accuracy using the tree-wise features was 60.6%, respectively.

Assessing spatial variation for tree and non-tree objects in a forest-tundra ecotone in airborne laser scanning data

Nadja Thieme*, Ole Martin Bollandsås, Terje Gobakken, Erik Næsset

Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, Norway

Changing climate is expected to have a significant impact on temperature-sensitive ecosystems like the forest-tundra ecotone. In Norway, this ecotone constitutes a large proportion of the total land area and effective monitoring techniques are required. It has been indicated that height and intensity data from airborne laser scanning may hold potential for monitoring of small trees. In the present study, Voronoi polygons and variograms were employed in order to assess the spatial patterns of trees and non-tree objects located in the forest-tundra ecotone. Patterns both for trees and non-tree objects could be recognised using Voronoi polygons in combination with height and intensity values. Furthermore, variograms and cross-variograms revealed different characteristics for trees and non-tree objects, however, limited to large individual objects located on flat terrain.

Exploring horizontal area-based metrics to discriminate the spatial pattern of trees using ALS

Inka Pippuri*, Eveliina Kallio, Matti Maltamo, Petteri Packalén, Heli Peltola

Faculty of Science and Forestry, University of Eastern Finland, Finland

Airborne Laser Scanning (ALS) data can be used to accurately determine tree and stand characteristics. We hypothesize here that three-dimensional ALS data can also be used for characterizing the horizontal forest structure like the spatial pattern of trees. This kind of information is of primary interest in forest management. The objectives of this study were (1) to identify ALS point cloud metrics and horizontal texture and landscape metrics, which can be used to determine the spatial pattern of trees and (2) to study how well the clustered spatial pattern of trees can be separated from others. The field data consisted of 28 microstands, of which 11 were clustered and 17 random or regular. Linear discriminant analysis was used to classify the microstands by means of the metrics calculated from ALS data. The best ALS metrics to determine the spatial pattern of trees were determined by the best overall accuracies (OA) and kappa-values (k) and based on the significance tests of models and the correlation matrices of metrics.

The classification of the spatial pattern of trees succeeds well based on ALS metrics, with the overall accuracy being 0.89 and kappa-value 0.77. Especially the calculated landscape metrics were found good predictors of

Airborne lidar sampling of the Canadian boreal forest: Planning, execution & initial processing

Chris Hopkinson¹, Laura Chasmer*, Michael Wulder², Nicholas Coops³, Trevor Milne¹, Allyson Fox¹, Christopher Bater³

¹Applied Geomatics Research Group, Lawrencetown, Nova Scotia, Canada

²Pacific Forestry Centre, Canadian Forest Service, Victoria, British Columbia, Canada

³Department of Forest Resource Management, UBC, Vancouver, British Columbia, Canada

During the summer of 2010, a transcontinental aerial survey mission was performed to acquire 24,000 line km of lidar transects covering >15,000 km² representing all ecozones within Canada's boreal forest. The coverage equates to ~21 million 'lidar plots' at the 25 m grid cell resolution. Each 'plot' contains the position and intensity of 1000 to 2000 laser points, which describe the terrain surface and 3D canopy structure, which will be used to predict forest inventory attributes and to support calibration of wide area satellite-based imagery. Furthermore, in similar fashion to geo-located permanent sample plots, the lidar transect flight path from 2010 can be re-surveyed in the future to facilitate monitoring of forest development and change in a consistent and quantifiable manner. The paper describes the mission planning criteria, survey logistical considerations and customised transect data processing routines.

Assessing the accuracy of GLAS topography estimation by using airborne Light Detection And Ranging (LiDAR) measurements

Han Meng*, Bernard Devereux, Gabriel Amable

Geography Department, the University of Cambridge, UK

Topography estimation is a key factor in forestry studies. The accurate prediction of topography underneath tree canopies will certainly improve the subsequent forest bio-physical characteristics estimation such as tree height, stem volume, biomass/carbon stocks. Thus, the assessment of the accuracy of GLAS topographical estimation is essential before the data can be used for forest bio-physical characteristics prediction. This study proposes the use of airborne LiDAR measurements to assess GLAS ground elevation estimates in a mixed woodland and arable site in south-east England near Thetford, UK, at 52.4N, 0.81E, given that airborne LiDAR measurements have already been validated using 'ground-truth' data. GLAS full waveforms are decomposed into up to six Gaussian modes and different indices, such as waveform centroid position (GLA14 position) and GLA01 last peak position, are calculated based on the peak positions of these Gaussian modes. Elevations estimated from these indices are compared with airborne LiDAR elevation estimates for assessment purpose and optimal estimates will be selected based on the results.

Four comparison models are introduced in this study. From these, model 1 (the comparison between GLA14 elevation and non-filtered airborne last return pulses elevation) and model 4 (the comparison between GLA01 last mode elevation and filtered airborne last return pulses elevation) have the best performance with R-squared values of 0.89 and 0.87, respectively, and RMSE values of 3.82 and 4.69, respectively. After removal of outliers for model 4, the R-squared value improves to 0.99 and the RMSE value reduced significantly to 0.66. A simplified experiment is implemented in this study in order to investigate the impacts on biomass/carbon stock estimates arising from use of different models, with the assumption that there is a uniform average tree height of 20 meters and uniform stem density through the study site.

Characteristics of satellite LiDAR waveform in tropical rain forests from the comparison with canopy condition derived from high resolution satellite data

Yasumasa Hirata

Climate Change Office, Forestry and Forest Products Research Institute, Tsukuba, Japan

This study aims to investigate characteristics of satellite LiDAR waveform in tropical forests by comparing with canopy structure derived from high resolution satellite data. Study area is located in the Tangkulap Forest

Reserve, Sabah, Malaysia, which is managed by the Sabah Forestry Department under the Deramakot Forestry District. ICESat GLAS data for the study area were prepared and provided by the National Snow and Ice Data Center. GLA01data and GLA14 data were used. Footprints of laser pulse from ICESat GLAS were identified on the QuickBird image and stand structures in the footprints were estimated from crown information. First, canopy closure within a footprint was calculated using the generated mask. Distribution of individual crown areas within a footprint was investigated for all footprints in the study area. Grade of degradation due to historical selective logging was decided from these two factors, namely, canopy closure and crown size structure. Waveform in each footprint was extracted from ICESat GLAS data and the relationship between stand structure, which was estimated from crown information, and the waveform was investigated. In addition, waveform in oil palm plantation, which was outside of the extent of high resolution satellite data, was also investigated. The condition of the area was identified using Google Earth. Stand structure was estimated from waveform of satellite LiDAR data. Length of waveform almost indicated maximum tree height. Peak position of waveform indicated the height of canopy layer. The height and position of peak of waveform indicated the grade of forest degradation. Further studies are required for identify the relationship between waveform of a shot of laser pulse from satellite LiDAR and canopy condition such as canopy closure and distribution of crown area quantitatively.

Model development for the estimation of aboveground biomass using a lidar-based sample of Canada's boreal forest

Christopher W. Bater^{1,*}, Michael A. Wulder^{2,*}, Nicholas C. Coops^{1,*}, Chris Hopkinson³, Samuel B. Coggins⁴, Eric Arsenault⁵, André Beaudoin⁶, Luc Guindon⁶, R.J. Hall⁵, Philippe Villemaire⁶, & Murray Woods⁷.

¹Integrated Remote Sensing Studio, Department of Forest Resources Management, University of British Columbia.

²Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada.

³Applied Geomatics Research Group, Centre of Geographic Sciences.

⁴Nisga'a Lisims Government.

⁵Northern Forestry Centre, Canadian Forest Service, Natural Resources Canada.

⁶Laurentian Forestry Centre, Canadian Forest Service, Natural Resources Canada.

⁷Southern Science and Information Section, Ontario Ministry of Natural Resources.

The northern forested areas of Canada are largely unmanaged and not subject to inventories with the same level of detail or regularity as southern forested regions. In an effort to augment monitoring and inventory activities, airborne light detection and ranging (lidar) has been employed to obtain plot-level information over a sample of Canada's northern forests. During the summer of 2010, a series of 34 transects were flown over a total length of more than 24,000 km, spanning the width of the Canadian landmass from Nova Scotia to the Yukon, and crossing eight ecozones and 13 UTM zones. Following data acquisition, a suite of plot-level lidar vegetation metrics were calculated. To develop estimates of forest attributes such as biomass, however, field data were required from the range of conditions found across the region. To that end, datasets were acquired from Quebec, Ontario and the Northwest Territories. In this paper we describe the development of regression models for large area estimates of various tree aboveground biomass components using field and lidar datasets of uncommon provenance, with significant differences both in terms of the environments in which they were collected, and the characteristics of the field and lidar surveys. The equations developed are deemed suitable for application and extrapolation across the national series of lidar transects.

Early assessment of industrial needs: harvesting and allocation decisions supported by ALS and TLS

Gero Becker*, Thomas Smaltschinski, Martin Opferkuch, Holger Weinacker

Institute of Forest Utilization and Work Science & Department of Remote Sensing and Landscape Information Systems*). University of Freiburg

Context: To plan their own supply with regard to product demand, the wood industry needs precise requirements towards the qualitative properties of its purchased logs (volume, species, dimension, length, diameter, knottiness, taper, sweep and the absence of stem defects). Early information of the prospective wood quality from the designated harvestable stands is therefore required. Traditional inventory data are in most cases out-dated and not detailed enough to fulfil these information needs. In many forests (e.g. small private holdings) inventory data do not exist at all. To conduct up-to-date pre-harvest ground inventory is time consuming and costly. In this situation remote sensing with laser technology is a promising alternative.

Methodology: Aerial laser scanning (ALS) covers big areas and provides primarily height data, from where terrain information as well as canopy information can be extracted, so that single tree recognition and modelling of the crown shape is possible. Broadleaved and coniferous trees can be distinguished, but species identification is still difficult (Spectral aerial photography may help to solve this problem). From the tree height and the crown shape, important quality information like branchiness can be derived based on well established allometric functions, but information (diameter, shape, bark features) of the stem below the crown (which represents the most valuable part of the tree) is difficult to obtain via ALS.

Terrestrial Laser Scanning (TLS) provides information on the below-crown part of the stand. Depending on the type of laser scanner and the stand density, within a circle with a radius of $r \sim \pm 15\text{m}$ the exact position and detailed dimension and quality data of every (visible) stem can be obtained already with one instrument set up. Theoretically, a total coverage of the stand would be possible, with a sufficient number of instrument set ups, but this would not be economically feasible for industrial application.

The suggested solution is a combination of total (crown-) assessment via ALS and sampling of stem data via TLS. After ALS scanning of the respective stand, circular sample plots are defined and located. Number and radius of the plots are derived from stand characteristics (variation of stand density and tree heights) based on statistical considerations. For all merchantable trees within these circular plots, their (foot) position and crown shape is 3-D modelled from ALS points. TLS scanning of the stem of all (merchantable) trees is then conducted from the centre of these sample plots. Georeferenced data of ALS and TLS positions allow modelling the full tree by composing the respective crown and stem data.

Remotely sensed crown structure as an indicator of wood quality: A comparison of metrics from aerial and terrestrial laser scanning

Thomas Adams*, David Pont, Jonathan Harrington

Scion, 49 Sala Street, Rotorua, New Zealand.

Aerial LiDAR offers a fast and efficient means to estimate wood quantity, but there has been little work to date on wood quality. In this study we investigate the hypothesis that remotely sensed crown structure from Aerial Laser Scanning (ALS) can be used as an indicator of log quality at an individual tree level.

A New Zealand *Pinus radiata* forest was flown with aerial LiDAR at 8 pts per m². Five trees from within the forest were scanned with a terrestrial laser scanner (TLS) to determine external signs of log quality. These measurements were diameter at breast height (DBH), volume, taper, sweep, lean, circularity and average internode distance. In this study we develop a series of metrics from ALS point clouds for each tree to describe the crown structure, which are then correlated against the TLS data. To derive these metrics, novel algorithms were developed for TLS data which extend the level of detail previously obtainable. These algorithms are also detailed in this paper.

As only five trees were studied, the results are proof-of-concept more than outright proofs. The purpose of this paper is to document techniques which will be employed in the future over a much greater sample, proving the preliminary findings presented here. In this small sample we found that crown area from ALS had a moderately strong correlation with DBH and sweep. Crown density from ALS was also moderately correlated to average internode distance. The correlations show that there is at least a moderate connection between

crown structure and log properties, and that at higher LiDAR pulse densities and a larger sample size we can expect to describe this connection with greater certainty.

In further studies we also hope to correlate ALS and TLS metrics with internal wood properties, as found from destructive sampling.

Developing lidar interpretation software for wood resource inventory in Forests NSW

Russell Turner*, A Farjad, J Trinder, S Lim

¹Forest Science Centre, Department of Primary Industries, Sydney, NSW

²School of Surveying and Spatial Information Systems, UNSW, Sydney, NSW

Forest inventory programs are traditionally based on very limited field sampling data which is then extrapolated across the entire forest estate. One of the major weaknesses of this approach is that the limited number of plots often covers less than 2% of the total forest area and this can influence how representative the data may be of forest variation. Instead of relying solely on field plots sampled at around 1 per 400 to 1000 ha, future resource inventory programs could utilise high sampling density full-waveform lidar to conduct on-screen manual interpretation. The premise is that one analyst using lidar full-waveform data onscreen could potentially manually interpret 50 lidar plots per day compared to two field crewmembers measuring 4 to 6 plots per day. This involves a paradigm shift from 100% field survey dependent forest sampling to a mix of lidar plot interpretation with significantly fewer field plot samples. If feasible, this innovative resource assessment approach has the potential to provide significant savings in future resource inventory programs. The strength of the new generation of full waveform lidar systems lies in the enormous amount of structural data that can be rapidly collected. However, this strength is also their weakness for two reasons. Firstly, these systems generate extremely large volumes of data that demand exceptional data storage capacity (i.e. terabytes of space). And secondly, there is a scarcity of commercial software capable of processing the data in a way customised specifically for forestry purposes. Forests NSW (FNSW) and the University of New South Wales (UNSW) have developed a new lidar processing platform that offers the visualisation of point cloud data viewed in 2D and 3D displays and a suite of manual tools to add markers, measure stem and crown parameters and tag key attributes such as form quality, species and growth stage for each tree. In addition, a series of automated plot statistics can also be extracted such as point percentile counts at nominate height thresholds, common descriptive statistics (e.g. max, mean, mode, median, standard deviation etc.), and canopy cover percentage. A prototype of the new software should be ready for field testing in late 2011.

Towards automated and operational forest inventories with T-Lidar

Othmani Ahlem* ¹, Piboule Alexandre ², Krebs Michael ³, Stolz Christophe ¹, Lew-yan-voon Lew ¹

¹Laboratoire LE2I –UMR CNRS, France

²ONF,R&D department, F-54000 Nancy, France

³ENSAM,Equipe Bois, F-71250 Cluny, France

Forest inventory automation has become a major issue in forestry. The complexity of the segmentation of 3D point cloud is due to mutual occlusion between trees, other vegetation, or branches. That is why, the applications done until now are limited to the estimation of the DBH (Diameter at Breast Height), the tree height and density estimation. Furthermore other parameters could also be detected, such as volume or species of trees (Reulke and Haala)...

This paper presents an effective approach for automatic detection, isolation of trees and DBH estimation. Tree isolation is achieved using an innovative approach based on a clustering methodology followed by a skeletonization step. The DBH of trees is then determined automatically. The efficiency of our algorithm is evaluated with comparison with ground data, measured by classical methods.

3-D modelling of forest structure for parameterization of radiative transfer models

Martin Van Leeuwen*, Nicholas Coops, Glenn Newnham, Thomas Hilker, Darius Culvenor, Michael Wulder

Reconstructions of individual trees and their complex canopy structure provide an important means for studying a range of physiological processes including photosynthesis, respiration and resource use efficiencies and for assessing the effects of competition and crown structure on tree functioning.

However, measuring and registering detailed descriptions of tree and canopy structure has for long been challenging due to the laborious nature of data acquisition and subjectivity of taking field measurements in complex forest scenes. This study investigates the potential of ground-based, time-of-flight laser scanners for use in the 3D explicit reconstruction of forest structure and parametrization of radiative transfer models.

Evaluation of nonlinear equations for predicting diameter from tree height for *Pinus radiata* (D. Don) in an airborne laser scanning-based plantation inventory

Huiquan Bi*^{1,2}, Julian Fox², Yun Li³, Yuancai Lei⁴, Yong Pang⁴

¹ Forest Science Centre, Science and Research Division, New South Wales, Department of Industries and Investment, Beecroft, NSW

² School of Forest and Ecosystem Science, University of Melbourne

³ School of Information Science and Technology, Beijing Forestry University, Beijing, China

⁴ Institute of Forest Resources Information Techniques, Chinese Academy of Forestry, Beijing, China

More than 30 height-diameter equations in the forest biometrics literature were evaluated to select candidates for deriving equation forms for predicting diameter from tree height in support of LiDAR based forest inventory. The evaluation was based on four criteria: (1) the height-diameter function is inversable, (2) the inverse function is continuous and monotonically increasing over a specified working range of total tree height, (3) DBH is equal to zero at breast height in the inverse function, and preferably (4) the inverse function has an inflection point that is consistent with biological expectations. A total of 12 candidate equation forms were derived, which included 5 two-parameter and 7 three-parameter equations. The estimation properties and predictive performance of these 12 equation forms were further evaluated and compared through repeated sampling and fitting using data from 3581 trees destructively sampled for taper measurements from *Pinus radiata* plantations across New South Wales, Australia. Three equation forms, including the constrained Richards, Weibull and the combined power and exponential function, displayed superior prediction accuracy and estimation properties, and so were recommended as the primary equation forms for developing diameter-height equations. The remaining equation forms were marred by either lower prediction accuracy or poorer estimation properties or both. The three recommended equation forms should only serve as basic deterministic specifications, upon which other tree and stand variables should be incorporated as predictors to further improve their predictive performance.

Revisiting the status of space-borne lidar missions for assessing structural and biophysical forest parameters in the context of sustainable management of Earth resources

Sylvie Durrieu*¹, Ross Nelson²

¹ Cemagref, UMR TETIS, 500, rue J.F. Breton BP 5095, 34196 Montpellier Cedex 5, France

² Biospheric Sciences Branch, NASA/GSFC, Greenbelt, USA

Assessing forest aboveground biomass at global scale is crucial to address the challenge of sustainable management of forest resources and to strengthen forest-based climate change mitigation. To achieve this goal relying on spaceborne lidar missions is acknowledged to be a highly relevant solution. However, if this is taken as a given from the measurement point of view, the premise that spaceborne observation is the most suitable solution to provide information for sustainable management of forest resources is worth discussing. In this paper we suggest to take a fresh look at measurement processes designed to support the monitoring of Earth resources. We discuss the sustainability of Earth observation from space considering (1) issues that call into question the assumption that Earth-orbiting platform will always be available to the civilian remote

sensing community and (2) issues concerning environmental impacts of space activity on the Earth. This leads us to suggest some actions that could help to design future observation systems in a more sustainable way in order to strengthen the capacity of measurement processes to meet their stated functional goal, i.e. sustainable management of forest resources.

Vegetation classification in the Swedish sub-arctic using a combination of optical satellite images and airborne laser scanner data

Mattias Nyström*, Karin Nordkvist, Heather Reese, Johan Holmgren, Håkan Olsson

Swedish University of Agricultural Sciences, Department of Forest Resource Management, Sweden.

The aim of this pilot study was to investigate to which degree the accuracy of automated vegetation classification in the Swedish sub-arctic could be improved by combining optical satellite data with airborne laser scanner (ALS) data, compared to using satellite data only. This information is of interest in an ongoing discussion about the possible inclusion of the mountains in northern Sweden in the national laser scanning that started in 2009. A SPOT 4 scene and ALS data from an Optech ALTM Gemini scanner, both from 2010, were used in maximum likelihood classification. Data for training and validation was obtained from 279 plots with 20 m radius that were visited in field 2010. These plots were located near Abisko in northern Sweden (lat. 68° 23' N, long. 18° 53' E), on the north and south side of Lake Torne Träsk. A classification scheme with 7 classes based on the Swedish mountain vegetation map was used. Classification using only SPOT data gave an over-all accuracy of 75.6%, and the combination of SPOT data and ALS data increased the accuracy to 81.4%.

Lidar data and cooperative research at Panther Creek, Oregon

James Flewelling*¹, George McFadden²

¹Seattle Biometrics, Seattle, WA, USA

²Bureau of Land Management, Portland, OR, USA

A 2,300 hectare forested watershed in the coastal mountain range of Oregon, USA is the subject of collaborative research with a principal objective of evaluating uses of lidar and other remotely sensed data for the development of detailed forest inventories. Panther Creek watershed (4518' N, 12321' W) is at an elevation of 100-700 m, about 57 km southeast of Portland. Major species are Douglas fir, western hemlock, western red cedar, grand fir, red alder and big leaf maple; tree heights are up to 60 m. The Bureau of Land Management and other cooperators are using the watershed to test and develop methodology for detailed stand level forest inventories, the detailed mapping of soils and slope stability, and the assessment of other ecosystem functions. Wall-to-wall discrete return lidar has been acquired under leaf-off conditions annually starting in 2007, and will continue through 2012. Leaf-on discrete return lidar was collected in 2007 and 2010 and will be collected in 2012. Surveys used Leica ALS50 Phase II or ALS60 lasers; pulse density is about 8 per m²; in 2010 selected areas received multiple passes, raising the density up to 50 pulses per m². Return intensities are being corrected for power output and camera-to-target distances. Full waveform lidar leaf-on data was acquired in 2010, as was 4-band color-infrared imagery using a Leica ADS40 camera. Also in 2010, hyperspectral data from a Hymap sensor was acquired. Eighty-four cadastral-surveyed 0.08 ha stem-mapped permanent plots were installed, mostly in 2009; measurement will be repeated after the 2012 growing season. Several other imagery sources are available. A project goal is to compare and evaluate methodologies. All data are available to research groups wanting to participate. Data are well documented and organized, and include cut-outs of the remotely sensed data at each of the plot locations.

LiDAR estimation of quadratic mean canopy height and stem density in native sclerophyll forests

Yadav Prasad Kandel*, Julian Fox, Stefan Arndt, Stephen Livesley

Department of Forest & Ecosystem Science, The University of Melbourne

LiDAR, relatively a new active remote sensing technology, capable of providing three-dimensional structural information of forest stands as well as individual trees has already been established as an operational tool in European and North American forestry. LiDAR estimates of various structural and biophysical parameters are

more accurate for pine forests than that for the broad-leaved and mixed species multi-story forests. In this study, plot level mean dominant height and quadratic mean canopy height were estimated quite accurately using the LiDAR data from two different types of native sclerophyll forests. R^2 of the regression model for the mean dominant height was 87.09 % for the Central Highlands Ash Regrowth (CHAR) and 92.1 % for the Black Range Mixed Species (BRMS) forest. Similarly, R^2 of the regression model for the quadratic mean canopy height was 48.4 % for the CHAR and 92.7 % for the BRMS forest. Stem density (number of trees per hectare) is the most difficult forestry attribute to estimate from remote sensing technology including LiDAR. When various LiDAR metrics were used directly to develop a regression model of stem density in the CHAR and BRMS forests, the models developed had very low (less than 0.3) R^2 . Therefore, in this study, an indirect method of estimating stem density using LiDAR data was developed. Using this new indirect method, the number of trees was predicted with mean prediction error of -64.12 trees per hectare for calibration plots and 105.29 trees per hectare for validation plots in CHAR forest, which is a wet sclerophyll forest. In the BRMS forest, which represents a dry sclerophyll forest, prediction error for number of trees, was 79.99 trees per hectare for calibration plots and 4.96 trees per hectare for validation plots.

Using a flux footprint model and airborne LiDAR to characterize vegetation structure and topography frequently sampled by Eddy Covariance: Implications for MODIS product validation

Laura Chasmer*¹, N Kljun², Chris Hopkinson³, S Brown¹, T Milne³, K Giroux¹, A Barr⁴, K Devito⁵, I Creed⁶, Richard Petrone¹

²Dept. of Geography, College of Science, University of Swansea, Swansea, Wales UK

³Applied Geomatics Research Group, NSCC, Middleton NS Canada

⁴Environment Canada, National Water Research Institute, Saskatoon SK Canada

⁵Dept. of Biological Sciences, University of Alberta, Edmonton AB Canada

⁶Dept. of Biology, University of Western Ontario, London ON Canada

Exchanges of CO₂ transported to eddy covariance instruments are often assumed to be representative of site average vegetation, understory, and topographical characteristics, regardless of the frequency with which these have been sampled. All sites have some degree of heterogeneity (e.g. an upland area, bog, area of dense understory, etc.), which could influence CO₂ exchanges if scalar fluxes from prevailing wind directions frequently sample these parts more than others. This could have implications for site representation, model evaluation, and remote sensing product validation and scaling. The use of flux footprint models has improved our understanding of the spatial and temporal distribution of source/sink areas measured within the field of view of eddy covariance instrumentation (e.g. Schmid, 1994). The flux footprint is defined as the probability of flux contribution per unit area upwind of the eddy covariance instrumentation (Kljun *et al.* 2002, 2004). When a footprint is combined with remote sensing data, the probability density function of the weighted source/sink contribution to the eddy covariance instrumentation provides spatially contiguous information on vegetation structural and topographic influences on net ecosystem production (NEP) (Chasmer *et al.* 2008). Simple logic follows: if CO₂ fluxes originate from areas of higher biomass, then measurements of flux should indicate increased uptake (NEP) when compared with lower biomass areas (etc.), all else being equal. Combining footprints with high resolution spatially continuous remote sensing data from airborne LiDAR, hyperspectral or spectral imagery provides a powerful tool for characterizing the areas sampled most frequently by eddy covariance. In this study, we use a 3D classification methodology to characterize vegetation structural and topographic attributes most frequently sampled by eddy covariance within 1) a homogeneous mature boreal aspen stand; and 2) a heterogeneous upland aspen/wetland complex using airborne LiDAR. The vegetation and topographic characteristics found within the areas most sampled at each site were then used to classify the larger region for evaluation of the MODIS gross primary production (GPP) product, i.e. choosing MODIS pixels that have similar attributes to those found within footprint most frequently sampled by eddy covariance. The results of this study find that footprints from prevailing wind directions at the homogeneous mature aspen stand have, on average, taller trees (7%), greater effective LAI (30%), denser understory (5%), and fewer low-lying topographic depressions than secondary wind origins. At the heterogeneous aspen stand, footprints from prevailing wind directions have, on average, shorter trees (-11%), lower effective LAI (-17%), and a greater proportion of topographic depressions.

Classification of vegetation structure and topography within a 1 km radius of the homogeneous and heterogeneous stands indicated that 56% (homogeneous aspen) and 69% (heterogeneous aspen) were representative of vegetation and topographic attributes sampled by eddy covariance. Thus, prevailing wind

directions may over- or under-sample some parts of the ecosystem more than others, which could result in over- or underestimates of NEP when compared with similar representative ecosystems.

When scaled to MODIS GPP, correspondence with GPP estimated using eddy covariance and meteorological methods improved by 13% when using LiDAR 'classified' pixels as opposed to those pixels most proximal to the tower. This illustrates that airborne LiDAR and footprint analysis can be used to link eddy covariance measurements of ecosystem exchanges between scales. This has important implications for assessment of spatial variability of vegetation/topography on NEP; identifying landscape features that are frequently sampled; classifying spatial heterogeneity; and scaling. More detail of this study is provided in Chasmer *et al.* (2011).

Satellite vs. airborne lidar estimates of aboveground biomass and forest structure metrics at footprint scale
Sorin Popescu ¹, Kaiguang Zhao ¹, Amy Neuenschwander ², Chinsu Lin ³

¹ Spatial Sciences Laboratory, Department of Ecosystem Science and Management, Texas A&M University, United States

² Centre for Space Research, University of Texas

³ Department of Forestry, National Chiayi University, Taiwan

Small footprint airborne lidar, sometimes referred to as airborne laser scanning (ALS), provides the best measurement accuracy of terrain elevation and vegetation heights, even on sloped terrain or for dense forests. However, large footprint, full waveform satellite lidar data, such as data provided by the Geoscience Laser Altimeter System (GLAS) aboard the Ice Cloud and land Elevation Satellite (ICESat), proved to have the potential for assessing vegetation parameters at unprecedented scales, from regional to continental and global extents. The overall goal of this study was to compare biomass estimates and height metrics obtained by processing GLAS waveform data and spatially coincident discrete-return airborne lidar data over forest conditions in east Texas, which are characteristics of much of the south-eastern United States. The study area includes pine plantations in various developmental stages, old growth pine stands, and upland and bottomland hardwoods. Since biomass estimates are derived from waveform height metrics, we also compared ground elevation measurements and canopy parameters. More specific objectives were to compare the following parameters derived from GLAS and airborne lidar: (1) ground elevations; (2) maximum canopy height; (3) average canopy height; (4) percentiles of canopy height; and (5) above ground biomass. We used the elliptical shape of GLAS footprints to extract canopy height metrics and biomass estimates derived from airborne lidar. Individual tree parameters, including tree height, crown width and tree locations, were estimated from the ALS-derived canopy height model using an individual-tree isolation method and were related to diameter-at-breast-height (dbh) measurements and dbh-based general biomass equations for pine and mixed hardwood to compute above-ground biomass. The resultant biomass map derived at individual tree level was used as the dependent variable in our investigations of deriving biomass at footprint scale using GLAS variables and linear regression models.

Results indicated a very strong correlation for terrain elevations between GLAS and airborne lidar, with an *r* value of 0.98 and a root mean square error of 0.78 m. GLAS height variables were able to explain 80% of the variance associated with the reference biomass derived from airborne lidar, with an RMSE of 37.7 Mg/ha. Most of the models comparing GLAS and airborne lidar height metrics had R-square values above 0.9.

The significance of managed and natural vegetation on house survival during wildfires

Anders Siggins, Glenn Newnham & Raphaele Bianchi

¹CSIRO Land and Water

²CSIRO Land and Water

³CSIRO Ecosystem Sciences

The impact of wildfires at the urban interface is a major concern for people safety and property loss. The Australian forest fires of February 2009 resulted in the highest loss of life from forest fires in Australian history, and occurred in semi-rural and rural areas in Victoria. The most deadly of these fires occurred in the Kinglake region north of Melbourne. In this area there is generally no clear delineation of the urban interface. In this situation, where houses can be located within the natural forest, each house has its own unique forest

Error assessment and mitigation for hyper-temporal UAV-borne LiDAR surveys of forest inventory.

Luke Wallace*, Arko Lucieer, Darren Turner, Christopher Watson

School of Geography and Environmental Studies, University of Tasmania

Remotely sensed LiDAR data has become an important tool in the management of modern forest inventories. Monitoring the high frequency changes within forests with this data has been restricted by the cost and intermittent nature of LiDAR surveys. The use of Unmanned Aerial Vehicles (UAVs) as a remote sensing platform is a rapidly developing field and is capable of allowing highly dynamic environmental changes to be monitored. As such recent studies presented in the literature highlight the potential of UAV systems for forest monitoring. This study further investigates the potential of UAVs by examining the achievable accuracy of a newly developed UAV-borne LiDAR system in comparison to a traditional full scale system. The major contributions to the error budget of a UAV-borne LiDAR system are constrained through the use of a novel UAV specific processing workflow. Central to this workflow is the fusion of observations from a low cost Inertial Measurement Unit, a GPS receiver and a high definition video camera with a Sigma-Point Kalman Smoother allowing for highly accurate estimates of orientation. We found that using this workflow and under certain flying conditions accuracies similar to a modern full-scale system are achievable from this low-cost platform.

A new photon counting lidar system for vegetation analysis

Jaqueline Rosette*¹, Christopher Field², Ross Nelson¹, Phil DeCola², Bruce Cook¹

¹NASA Goddard Space Flight Center, Biospheric Sciences, Greenbelt, USA

²Sigma Space, Lanham, USA

This paper considers the potential of a new scanning photon-counting system for vegetation analysis. The 3D Mapper sensor was developed by Sigma Space Corporation and is being tested within NASA's Carbon Monitoring System (CMS) project (NASA, 2010). The sensor is able to map 60 km² per hour using less than 150 mW of 532 nm green light with about 30 cm between measurement points. While this area coverage rate is already several orders of magnitude higher than can be achieved by conventional lidar, substitution of higher power lasers would permit significantly higher mapping rates with the same resolution or much higher spatial resolution at the current rates. Data were collected for a test site to the west of Fredericksburg, Virginia, USA and demonstrated the capability with a low powered laser, of relatively high density data collection, and good penetration through the canopy, despite high canopy fractional cover and a hazy atmosphere at the time of flight. This preliminary study supports the potential of this emerging technology for vegetation analysis. Further research is required to develop algorithms to exploit the capabilities of such systems and to provide a greater understanding of the interactions with vegetated surfaces. Studies of this nature will inform future photon-counting satellite lidar sensors such as NASA's ICESat II, which is due for launch at the beginning of 2016.

Sorted Pulse Data (SPD) Format: A new file structure for storing and processing LiDAR data.

Peter Bunting*^{1,2}, John Armston³, Daniel Clewley¹, Richard Lucas¹

¹Institute of Geography and Earth Sciences, Aberystwyth University, UK

²Landcare Research, Palmerston North, NZ

³Joint Remote Sensing Research Program, Centre for Spatial Environmental Research, School of Geography, Planning and Environmental Management, University of Queensland

This paper presents a new generic method and format for storing and processing airborne and terrestrial LiDAR pulse data within a HDF5 file. The format is specifically designed to support both traditional discrete return and full waveform data, uses a pulse (rather than point) based data model and has been developed and applied successfully using a wide range of disparate airborne and terrestrial LiDAR datasets. The format is proposed as an alternative to existing solutions as it includes support for full waveform data, explicit pulse based data structures and flexible spatial indexing using cartesian, spherical and polar coordinate systems and projections. The HDF5 format supports compression but in part due to the more complex data structures

Another dimension from LiDAR - Obtaining foliage density from full waveform data

Thomas Adams*¹, Peter Beets¹, Christopher Parrish²

¹Scion, 49 Sala Street, Rotorua, New Zealand

²NOAA/NGS, JHC-CCOM, Durham, NH, USA

LiDAR tells the user *where* surfaces are, not *what* they are. In this study we investigate the potential for waveform LiDAR to provide more information on the nature of the returns over forestry. Waveform LiDAR was acquired for ten *Pinus radiata* plots in a New Zealand plantation, along with comprehensive leaf area sampling in 2m vertical bands. The decay rate of each waveform peak was shown to be a useful tool for estimating foliage density, and has potential for identifying regions containing ground and understorey. Leaf Area Density (LAD) is an expression of foliage density per unit height, and a relationship between waveform decay rate and LAD was developed with an R2 of 56%. Incorporating the proportion of discrete LiDAR that fell in that band (which itself has an R2 of 50%) improves this model to explain 69% of the variation in LAD. This is a good result, especially given the costs and difficulties in measuring leaf area directly. As foliage density varies dramatically on a fine scale it was not possible to differentiate the nature of every single LiDAR return – but by averaging over a small area local variation in LAD could be easily mapped. Ground returns could be distinguished as having short decays, and broad leaved understorey typically had values between those of the canopy and ground, although surface roughness and slope make it impossible to robustly identify single returns. This study produced a useful model for estimating LAD in *Pinus radiata* which could easily be extended to other coniferous species.

The Sorted Pulse Data Software Library (SPDLib): Open source tools for processing LiDAR data.

Peter Bunting*^{1,2}, John Armston³, Daniel Clewley¹, Richard Lucas¹

¹Institute of Geography and Earth Sciences, Aberystwyth University, UK

²Landcare Research, Palmerston North, NZ

³Joint Remote Sensing Research Program, Centre for Spatial Environmental Research, School of Geography, Planning and Environmental Management, University of Queensland

SPDLib is a new set of tools that allow processing and analysis of the full range of LiDAR data from terrestrial, airborne and spaceborne systems, including both discrete return and waveform datasets. The software provides an implementation of the SPD file format that allows efficient and flexible storage of these datasets largely through the inclusion of spatial indexing and pulse (rather than point) based data structures. A visualisation tool (SPD Points Viewer), which builds on top of SPDLib and the SPD file format, has also been developed. The software and source code have recently been made freely available and can be accessed online through open source code repositories. Future developments will focus on the development of new waveform processing functionality and optimizing performance. The software and documentation can be obtained from <http://www.spdlib.org>.

Comparison of point cloud data reduction methods in single-scan TLS for finding tree stems in forest

Paula Litkey, Puttonen Eetu, Liang Xinlian*

Finnish Geodetic Institute

The point density in a single-scan terrestrial laser scanner (TLS) point cloud is very dense close to the scanner and gets sparser as the distance from the scanner increases. A full circular scan can contain tens of millions of points, which is impractical for most algorithms that work on point data. The number of points can be reduced by taking a sample of the original data. We have studied what influence different sampling methods have on the number of points that falls on tree stems. We propose that the number of points available on a far-away tree can be increased with a smart data reduction scheme. The data reduction favours far-away points over the densely located points close to the scanner. The main findings of this study are that removing ground points before sampling gives a great advantage in data reduction and that a point selection using only horizontal distances (2D Cartesian, xy-plane) favours low points.

Automated log counting: Proof of concept algorithm

Hamish Marshall

Interpine Forestry Ltd

Improving log inventory is a key area where the New Zealand forestry industry could significantly improve its supply chain performance. Although the process of counting logs seems relatively simple; in reality it is a difficult and labour intensive job. This is particularly significant to the New Zealand log export industry which is required to count and barcode every log (excluding pulp) that is exported. The fluctuating nature of export markets means that automated methods of counting logs hold significant potential. This paper investigates the accuracy of log counts for logs in pile/stacks using 3-dimensional (3D) point cloud data obtained from a ground based LiDAR scanner. In the past there have been a number of attempts to develop an automatic log counting system, the majority of these have used 2-dimensional photographic images. It was hypothesised that using 3D point data would overcome some of the problems that these approaches have encountered in the past. The validation study carried out on the algorithm showed that logs can be accurately counted and log diameters can be measured. Further work would be required to develop the algorithm into a commercial product and to determine the most cost effective hardware required to collect the 3-dimensional data required by the algorithm.

Optimal LiDAR gridding parameterization for effective leaf area estimation in the boreal forest Yukon Territory, Canada

Heather Morrison* ^{1,2}, Chris Hopkinson ^{2,1}, Michael Wulder ³

¹Acadia University, Wolfville, Nova Scotia, Canada

²Applied Geomatics Research Group, NSCC Annapolis Campus, Nova Scotia, Canada

³Pacific Forest Centre, Canadian Forest Service, Victoria, British Columbia, Canada

The increased availability of LiDAR-based forestry models raises questions about fundamental procedural steps undertaken before published models begin. The processing stage being investigated in this study is the parameterization of routines used for gridding inputs for forestry models. Grids are a valuable format for modelling as they organize scattered point clouds into manageable pixels for sophisticated processing. Our objective was to examine the effect of grid cell resolution and circular search radii for gridding on resulting leaf area index (LAI) data layers. LAI is employed in models at various scales and was therefore of interest for a range of study objectives. We generated 16 gridded estimates of LAI using unique combinations of cell resolution and search radius for comparison with values measured in the field. Our results determined that cell resolution was not important, allowing for more application flexibility without introducing bias, while search radius was critical for obtaining the most accurate estimates. This type of scale sensitivity analysis is important for any modelled variable that will be applied in a variety of spatial contexts.

Notes:

Developing a regional canopy fuels assessment strategy using multi-scale LiDAR

Birgit Peterson*¹, Kurtis Nelson²

¹ ASRC Research and Technology Solutions contractor to the U.S. Geological Survey (USGS), Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD

² USGS, EROS, Sioux Falls, SD

Accurate assessments of canopy fuels are needed by fire scientists to understand fire behaviour and to predict future fire occurrence. A key descriptor for canopy fuels is canopy bulk density (CBD). CBD is closely linked to the structure of the canopy; therefore, lidar measurements are particularly well suited to assessments of CBD. LANDFIRE scientists are exploring methods to integrate airborne and spaceborne lidar datasets into a national mapping effort. In this study, airborne lidar, spaceborne lidar, and field data are used to map CBD in the Yukon Flats Eco region, with the airborne lidar serving as a bridge between the field data and the spaceborne observations. The field-based CBD was positively correlated with airborne lidar observations ($R^2 = 0.78$). Mapped values of CBD using the airborne lidar dataset were significantly correlated with spaceborne lidar observations when analysed by forest type ($R^2 = 0.62$, evergreen and $R^2 = 0.71$, mixed). Though continued research is necessary to validate these results, they do support the feasibility of airborne and, most importantly, spaceborne lidar data for canopy fuels assessment.

LiDAR-based estimation of forest floor fuel loads using a novel distributional approach

Jan Van Aardt*¹, Mary Arthur², Gretchen Sovkoplak², Tyson Lee Swetnam³

¹Rochester Institute of Technology, Rochester, NY, USA

²University of Kentucky, Lexington, KY, USA

³University of Arizona, Tucson, AZ, USA

Light detection and ranging (LiDAR) has seen significant application across a range of forest structural assessment applications, ranging from forest volume and biomass assessment, to ecological applications such as leaf area and fuel load modelling. However, quantification of sub-canopy structure remains a challenge, especially when considering downed coarse woody debris (CWD) near the ground surface. This is true because the LiDAR signal attenuates through the canopy, LiDAR systems can be set to record the last of many returns, which is often the ground itself, and there is a system-specific vertical resolution that influences detection of structure in-between returns. We applied a LiDAR distributional approach to CWD modeling that included both above-ground and theoretical “below-ground” returns, with the latter being attributed to multiple scattering effects. This was done for oak dominant forests in central Appalachia, Kentucky, USA. Medium-fast (10h) and medium-slow (100h) CWD fuel loads exhibited the best results; e.g., an adjusted $R^2=0.99$ and a root mean square error value of 0.111Mg/ha (4.7% of the mean) were achieved for 100h CWD fuel loads. Independent variables included a balanced set from both the above-and below-ground distributions. Results hint at the significant potential of extending distributional approaches to CWD estimation.

Using airborne survey to map stream form and Riparian vegetation characteristics across Victoria

Nathan Quadros*¹, Rick Frisina¹, Paul Wilson²

¹ Information Services Branch, DSE Victoria

² Office of Water, DSE Victoria

The State of Victoria has established the index of stream condition (ISC) methodology for providing a river health assessment. The index of stream condition (ISC) is a baseline dataset used in river investment and planning. The ISC evaluates the environmental conditions of the major rivers and tributaries across Victoria. The ISC assessment undertaken during 2009-2010 has included a significant investment in the use of LiDAR and aerial photography to assess the riparian vegetation and river form components of ISC stream network. The health of riparian vegetation and stream form is assessed by measuring a number of metrics. Each of the metrics can be measured by either or combination of LiDAR and/or aerial photography. The scores for all the ISC metrics are amalgamated into a single sub-index score for both the river form and riparian vegetation.

Deploying LiDAR applications - Gearing toward the potential of LiDAR application in Malaysian forestry

Mohd Hasmadi Ismail¹ and Mohamad Sam Manaf²

¹ Forest Surveying and Engineering Lab., Faculty of Forestry University Putra Malaysia, Selangor, Malaysia

² RS&GIS Consultancy Sdn, Kuala Lumpur, Malaysia

Information on forest properties have grown over time and will continue crucially in the future. The focus on timber for commercial trade in early 1960's in Malaysia has been changed towards multi function forestry, supported by multi resources survey. Starting with high demand of the latest data and accurate information, and cost effective monitoring system, application of various technology of sensing system is applied into forestry. The introduction of precision forestry concept is not new but in Malaysia is still at infancy stage. It deals with advanced sensing technologies and analytical tools to support site-specific economic, environmental, and sustainable decision making for the forest management and development. The key discipline is highly relying on accurate, timely and detailed forest inventory characterization and structural information. This is possible by utilization of accurate measurement forestry data and information to improve operations and processes. Despite of the current use of high resolution satellite and airborne sensing, LiDAR is a promising alternative tool to be used in forestry sector. LiDAR can be used in forest engineering for terrain mapping and road planning, and tree/stand measurement for tropical forest. This paper gives a synopsis of LiDAR sensing technology application and its potential to Malaysian forestry.

Laser scanning by echo signal digitization and waveform processing

Martin Pfennigbauer*, Andreas Ullrich*

RIEGL Laser Measurement Systems GmbH, Austria MARTIN

LIDAR technology based on time-of-flight ranging with short laser pulses enables the acquisition of accurate and dense 3D data in form of so-called point clouds. The technique is employed from different platforms like stable tripods in terrestrial laser scanning or aircrafts, cars, and ships in airborne and mobile laser scanning. Historically, these instruments used analogue signal detection and processing schemes with the exception of instruments dedicated for scientific research projects or bathymetry. In 2004, a laser scanner device for commercial applications and for mass data production, the RIEGL LMS-Q560, was introduced to the market, making use of a radical alternative approach: digitizing the echo signals received by the instrument for every laser pulse and analysing these echo signals off-line in a so-called full waveform analysis in order to retrieve almost all information contained in the echo signal using transparent algorithms adaptable to specific applications. In the field of laser scanning the somewhat unspecific term "full waveform data" has since been established. We attempt a classification of the different types of the full waveform data found in the market. We discuss the challenges in echo digitization and waveform analysis from an instrument manufacturer's point of view. Furthermore, the benefits to be gained by using this technique, especially with respect to the multi-target capability of LIDAR instruments employing echo digitization and the possibilities for applications in forestry assessment are addressed.

New methods and algorithms - Crown coverage calculation based on ALS data

Lothar Eysn¹, Markus Hollaus¹, Klemens Schadauer², Andreas Roncat¹

¹ Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Austria

² Department of Forest Inventory at the Federal Research and Training Center for Forests, Natural Hazards and Landscape, Austria

The objective of this paper is to present and evaluate a new geometrically unambiguously defined approach to calculate forest canopy cover, also known as crown coverage (CC) from airborne laser scanning (ALS) data based on national forest inventory (NFI) data. The CC is defined as the proportion of the forest floor covered by the vertical projection of the tree crowns. Most forest definitions lack in precise geometrical definitions for the calculation of CC and therefore, the results of common calculation methods differ and tend to be incomparable. To demonstrate the effect of such an unclear defined, common CC calculation method, CC maps, generated from moving window algorithms using different kernel shapes and sizes, are calculated and

analyzed for three study areas in Tyrol, Austria. The new unambiguous approach, the tree triples method, is based on defining CC as a relation between the sum of the crown areas of three neighbouring trees at a time and the area of their convex hull. The approach is applied for the same study areas and is compared with forest masks that are generated from moving window algorithms using different kernel shapes and sizes.

LiDAR estimation of quadratic mean canopy height and stem density in native sclerophyll forests

Yadav Prasad Kandel*, Julian Fox, Stefan Arndt, Stephen Livesley

Department of Forest & Ecosystem Science, The University of Melbourne

LiDAR, relatively a new active remote sensing technology, capable of providing three-dimensional structural information of forest stands as well as individual trees has already been established as an operational tool in European and North American forestry. LiDAR estimates of various structural and biophysical parameters are more accurate for pine forests than that for the broad-leaved and mixed species multi-story forests. In this study, plot level mean dominant height and quadratic mean canopy height were estimated quite accurately using the LiDAR data from two different types of native sclerophyll forests. R² of the regression model for the mean dominant height was 87.09 % for the Central Highlands Ash Regrowth (CHAR) and 92.1 % for the Black Range Mixed Species (BRMS) forest. Similarly, R² of the regression model for the quadratic mean canopy height was 48.4 % for the CHAR and 92.7 % for the BRMS forest. Stem density (number of trees per hectare) is the most difficult forestry attribute to estimate from remote sensing technology including LiDAR. When various LiDAR metrics were used directly to develop a regression model of stem density in the CHAR and BRMS forests, the models developed had very low (less than 0.3) R². Therefore, in this study, an indirect method of estimating stem density using LiDAR data was developed. Using this new indirect method, the number of trees was predicted with mean prediction error of -64.12 trees per hectare for calibration plots and 105.29 trees per hectare for validation plots in CHAR forest, which is a wet sclerophyll forest. In the BRMS forest, which represents a dry sclerophyll forests, prediction error for number of trees, was 79.99 trees per hectare for calibration plots and 4.96 trees per hectare for validation plots.

Modelling light conditions in forests using airborne laser scanning data

Werner Mücke, Markus Hollaus

Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Vienna

The amount of available sunlight in vegetated areas is an important factor influencing species composition, plant morphology and natural succession. It is therefore a significant parameter in forestry, ecology and other sciences dealing with biodiversity relevant studies. Research indicates a strong correlation between the quality and quantity of sunlight and the vegetation structure, both in horizontal and vertical direction. Due to the high complexity and variability of the canopy architecture, continuous area-wide data collection of light conditions in the understorey is needed for accurate modelling of light transmission. However, conventional ground based measurement methods are pointwise and time consuming, therefore not feasible for data acquisition of large areas. The ability of small-footprint airborne laser scanning (ALS) to penetrate small canopy gaps makes this remote sensing method especially suitable for vegetation studies. Geometric information of the vegetation structure can be derived directly from the 3D point cloud. This allows for modelling of the distribution of sunlight-absorbing or intercepting parts of the foliage, which consequently cast shadows on the surrounding understorey vegetation or the ground. Light transmission through the canopy can therefore be described in a very direct way by employing this 3D structural information. In this paper a methodology for modelling light conditions in forests using ALS data is proposed. The approach is based on a modified version of photogrammetric monoplottting. The parallel sun rays from variable sun positions act as projection rays being traced through the 3D point cloud (i.e. laser echoes) that represents the canopy. A defined size is assigned to each individual laser echo which casts a shadow of the respective size and shape. Shadowed areas are then derived by intersecting these projection rays with a digital terrain model and by rasterizing the projected point cloud. By employing ALS data from different acquisition times (leaf-on and leaf-off) the influence of vegetation phenology is explored. The derived shadow raster maps describe where a shadow is cast and how many intercepting parts of the canopy contribute to it. Consequently, these maps provide an excellent input for modelling the amount of available sunlight in vegetated areas, considering canopy gaps in arbitrary directions and also the seasonal variability of vegetation. The first results show that ALS is a time- and cost- efficient means for area-wide analysis of sunlight condition for forest floors, as well as for different understorey layers.

