Norwegian University of Life Sciences • Universitetet for miljø- og biovitenskap
Department of Ecology and Natural Resource Management
Philosophiae Doctor (PhD) Thesis 2012:54

Estimating forest biomass components by airborne and terrestrial laser scanning

Estimering av biomasse i skog ved hjelp av flybåren og bakkebasert laserskanning

Marius Hauglin

Estimating forest biomass components by airborne and terrestrial laser scanning

Estimering av biomasse i skog ved hjelp av flybåren og bakkebasert laserskanning

Philosophiae Doctor (PhD) Thesis

Marius Hauglin

Department of Ecology and Natural Resource Management
Norwegian University of Life Sciences

Ås 2012

Thesis number 2012:54
ISSN 1503-1667
ISBN 978-82-575-1090-9

PhD supervisors

Professor Erik Næsset
Department of Ecology and Natural Resource Management
Norwegian University of Life Sciences
P.O. Box 5003, NO-1432 Ås
Norway

Professor Terje Gobakken
Department of Ecology and Natural Resource Management
Norwegian University of Life Sciences
P.O. Box 5003, NO-1432 Ås
Norway

Evaluation committee

Research Forester Dr. Hans-Erik Andersen
USDA Forest Service PNW Research Station
University of Washington
Bloedel Hall 386
Box 352100
Seattle, WA 98195-2100
USA

Professor Paul M. Treitz
Department of Geography, Faculty of Arts and Science
Queen's University
Kingston, ON K7L 3N6
Canada

Professor Ole Hofstad
Department of Ecology and Natural Resource Management
Norwegian University of Life Sciences
P.O. Box 5003, NO-1432 Ås
Norway

Preface

This thesis is a partial fulfilment of my PhD studies at the Norwegian University if Life Sciences, Department of Ecology and Natural Resource Management. Together with a program of formal courses and a dissertation, it completes the requirements for the degree of Philosophiae Doctor.

I will like to thank my supervisors, professor Erik Næsset and professor Terje Gobakken, for providing prompt and constructive feedback and supervision throughout the project.
I will also thank Associate Professor Sorin Popescu who hosted an interesting stay at Texas A&M University.
I will further like to thank professor Tron Eid who supervised my master thesis and brought me into the realm of forest research in the first place, by hiring me as a research assistant.

My fellow PhD-students and my other colleagues have provided an open and inspiring environment for my PhD studies, I hereby thank all of you. A special thanks to Vegard Lien for adventurous journeys to our field plots in the hinterland of Aurskog-Høland, and to Dr. Hans Ole Ørka for helpful and inspiring discussions.

Finally I would like to thank my family, Live, Just, Agnes and Eskil for giving me the best inspiration and support I could wish for.

Ås, September 23, 2012.

Marius Hauglin

Contents

List of Papers

I. Hauglin, M., Gobakken, T., Lien, V., Bollandsås, O.M., Næsset, E., 2012. Estimating potential logging residues in a boreal forest by airborne laser scanning. *Biomass and Bioenergy* 36, 356–365.

II. Hauglin, M., Dibdiakova, J., Gobakken, T., Næsset, E. Estimating single-tree branch biomass of Norway spruce by airborne laser scanning. *Submitted.*

III. Hauglin, M., Astrup, R., Gobakken, T., Næsset, E. Estimating single-tree branch biomass of Norway spruce with terrestrial laser scanning. *Submitted.*

IV. Hauglin, M., Gobakken, T., Astrup, R., Ene, L., Næsset, E. Estimating single-tree branch biomass by airborne laser scanning using model training data with ground reference values obtained through terrestrial laser scanning. *Manuscript.*

Abstract

In this thesis forest inventory methods to estimate potential logging residues and branch biomass using both airborne and terrestrial laser scanning are explored. Forest inventories are essential for effective and sustainable management of forest resources. In the last ten years there has been an increased interest in the use of forest biomass for bioenergy purposes, and biomass from forests will most likely be one of several sources of energy that will have to replace fossil fuels in the future. One example of such use of forest biomass is the utilization of logging residues, biomass that would otherwise have been left in the forest during the logging. When logging residues become a commercial product from the forest, this resource should be quantified as part of the forest inventory to improve planning and management. Advanced remote sensing techniques, and in particular airborne laser scanning (ALS), have been subject to growing interest in the forest inventory community since the mid 1990s. The use of ALS data in forest inventories has transcended from an experimental to an operational state during the period. Two distinct methodologies using ALS data for forest inventories have emerged, commonly referred to as the *area-based* and the *single-tree* methods. In the former and most widely used procedure forest biophysical properties, such as mean stem volume, are predicted for area units. In the latter approach the aim is to derive properties, including position, of *each individual tree*. This single-tree method has not entered an operational state to the same degree as the area-based approach, and although promising results have been reported some issues are still not fully resolved.

The four studies in this thesis were carried out in a boreal forest in the south-eastern part of Norway, with Norway spruce and Scots pine as the dominant tree species. Several sets of field reference data were used, including destructive sampling of 50 spruce trees to accurately determine branch biomass. In Paper 1, a method to estimate potential logging residues on an area basis using ALS data of moderate density is presented. Logging residues was in this study defined as branches and tops and are “potential” in the sense that these are the tree-components that will become logging residues in the case of a final harvest. The methodology follows the *area-based* approach which has previously been successfully applied to derive estimates of e. g. mean stem volume. A strong relationship was found between field measured potential logging residues and ALS variables, with a mean prediction error of 22% at the stand level.

In the three subsequent papers in this thesis methods to estimate branch biomass at the single-tree level for Norway spruce were investigated. Norway spruce is one of the two main tree commercial species in the Nordic countries. In Paper 2, the relationship between ALS derived features and accurately measured branch biomass was found to be strong, with a mean prediction error of 35% for the best model. A major finding in this study was that the ALS data contained more information related to branch biomass than the actual field measured tree diameter and height.

The laser ranging principles used in ALS have also been applied in scanning from fixed positions on the ground, commonly known as terrestrial laser scanning (TLS). TLS is considered a potential tool for capturing information for forest inventories, possibly replacing time-consuming manual field registrations, or even better: capture structural information unobtainable through conventional field measurements. Extraction of information related to branch biomass from remote sensing data was further explored in Paper 3 where TLS data were used to estimate branch biomass. A strong relationship was found between TLS derived features and accurately measured branch biomass, with a mean prediction error of 32%. Although a strong relationship was found, the observed prediction accuracy was considered to be only moderate in light of the rather extensive data acquisition. In

Paper 4, data from ALS and TLS were combined; TLS derived branch biomass estimates were used as ground reference data in ALS based estimation of branch biomass. The aim of Paper 4 was to describe this methodological framework and assess the prediction accuracy using field and remote sensing data. Predictions with a mean error of 32% were obtained for single-tree branch biomass. The results show an improvement compared to an approach without the use of TLS data, and also suggest that branch biomass can be successfully included in a single-tree forest inventory.

Sammendrag

I denne avhandlingen er det beskrevet ulike metoder for modellbasert estimering av greinbiomasse ved hjelp av flybåren og bakkebasert laserskanning. For å kunne utnytte skogressursene på en lønnsom og bærekraftig måte er en skogbruksplantakst et viktig verktøy. I det siste tiåret har det vært en økt interesse for bruk av biomasse fra skog til bioenergiformål, og biomasse fra skog vil antakeligvis være en av flere kilder som vil måtte erstatte fossile energikilder i fremtiden. Et eksempel på skogsbiomasse brukt til bioenergi er hogstavfall, eller *greiner og topper* (GROT) . Når GROT blir et kommersielt produkt, vil det være hensiktmessig at denne delen av treet også blir kvantifisert og inkludert i en skogbruksplantakst.

Det har vært en økende interesse for bruk av digitale fjernmålingsteknikker, og spesielt flybåren laserskanning, innen skogtaksering i løpet av de siste 10-15 årene. Bruk av flybåren laserskanning har i løpet av denne perioden blitt kommersialisert, og i dag utføres mesteparten av all skogtaksering for skogbruksplanlegging i Norden ved hjelp av flybåren laserskanning. Det eksisterer to ulike metodikker for skogtaksering med flybåren laserskanning, *arealbaserte takster* og *enkelttretakster*. I arealbaserte lasertakster blir gjennomsnittstall, som for eksempel volum og middelhøyde predikert per arealenhet. Dette er den metodikken som i hovedsak er i bruk kommersielt. I enkelttretakster er målsettingen å predikere egenskaper på enkelttrenivå, inkludert nøyaktig posisjon. Enkelttretakster har ikke blitt tatt i bruk operasjonelt i samme grad som arealbaserte takster, og selv om metoden er lovende, er det fremdeles noen uløste utfordringer knyttet til den.

De fire studiene som inngår i denne avhandlingen, ble gjennomført i bartredominert skog i Aurskog-Høland kommune i Akershus. Flere datasett med feltregistreringer ble brukt, blant annet ett med nøyaktig vekt av greinbiomasse, bestemt etter felling av 50 grantrær. I artikkel 1 beskrives hvordan GROT kan estimeres i en arealbasert lasertakst. En sterk sammenheng ble funnet mellom GROT og data fra flybåren laserskanning, med en middelfeil på 22 % på bestandsnivå.

I de tre neste artiklene beskrives modellering av greinbiomasse av gran på enkelttrenivå, og i artikkel 2 ble det funnet en sterk sammenheng mellom data fra flybåren laserskanning og nøyaktig målt greinbiomasse, med en middelfeil på 35 % for den beste modellen. Et av funnene i studien beskrevet i artikkel 2 var at greinbiomasse ble mer nøyaktig modellert med flybåren laserskanning enn med feltmålt diameter og trehøyde og bruk av eksisterende allometriske biomassemodeller.

De samme prinsippene for avstandsmåling som ligger til grunn for flybåren laserskanning, brukes også i bakkebaserte laserskanning. Bakkebaserte laserskanning vurderes som et verktøy som kan erstatte tidkrevende manuelle feltregistreringer, eller muliggjøre registreringer av for eksempel strukturell informasjon, som det tidligere ikke har vært praktisk mulig å registrere. I artikkel 3 ble muligheten for å modellere greinbiomasse ved hjelp av bakkebaserte laserskanning undersøkt. En sterk sammenheng mellom data fra bakkebaserte laserskanning og nøyaktig målt greinbiomasse ble funnet, med en middelfeil på 32 % for den beste modellen. Nøyaktigheten til modellene vurderes likevel til å være moderat sett i lys av den relativt ekstensive datafangst og modelleringsprosessen som er påkrevet. I artikkel 4 presenteres en metodikk der bakkebaserte laserskanning blir brukt til å registrere feltdata som så inngår som del av treningsdataene i en flybåren laserskanningbasert modellering av greinbiomasse. Den beste modellen hadde i denne studien en middelfeil på 32 %, noe som var lavere enn for modeller der data fra bakkebaserte laserskanning ikke var brukt. Resultatet indikerer også at greinbiomasse kan inkluderes som egenskap i en laserbasert enkelttretakst.

Synopsis

1 Introduction

"[...] every method of reducing the labour involved in taking an inventory of the stock on hand has been welcomed by the forest manager."

Judson F. Clark, 1902. *Forestry Quarterly, 1*.

Forest inventories are essential for effective and sustainable management of forest resources. The need for efficient and reliable methods to derive forest information has been recognized for centuries, here illustrated by a quote from the very first issue of *Forestry Quarterly*, a predecessor of todays *Journal of Forestry*. The use of remote sensing – in particular airborne laser scanning (ALS) – has certainly reduced the labour of taking an inventory, and is currently changing the way forest inventories are conducted. It has enabled mapping of forests with an unprecedented level of detail, and has also expanded the range of possible features to include in an inventory. In this thesis I have investigated some of these possible features.

Emissions of CO_2 from combustions of fossil fuels are believed to be a major cause of the observed global climate change (IPCC, 2007). This has advanced the search for alternative sources of energy, including bioenergy from forests, and during the last 10 – 15 years there has been growing interest towards the use of forest biomass for energy purposes. In addition to the concern over climate change also the expected decrease in supply of fossil fuels is held as a reason to utilize more of the forest biomass. The reserves of fossil fuels on earth are inevitably decreasing and in a review study of future global supply of oil, Sorrell et al. (2010) suggest that "*there is a significant risk of a peak in conventional oil production before 2020*" and that forecasts with this peak beyond 2030 is too optimistic. Even if the expected output from unconventional fossil fuels such as shale gas is growing, it is by many still suggested that alternative sources of energy should be searched for (Hughes, 2011). Biomass from forests is thus likely to be one of several sources of energy that will have to replace conventional fossil fuels at some point in the future. The use of logging residues as fuel, directly or through chemical refinement processes, is one present example of the use of forest biomass for bioenergy purposes. In most countries only the tree stems have traditionally been harvested and processed by sawmills and the pulp and paper industries, while branches and tree tops have been left in the forest. Several studies emphasize however forest residues as a potential exploitable resource (Malinen et al., 2001; Gan and Smith, 2006; Mabee and Saddler, 2010), and the practice of not utilizing it is currently about to change.

Future changes in the energy market or in governmental policies may affect the price of energy from logging residues compared to energy from other sources. The European Commission reports a recent annual growth in the use of renewable energy in Europe that exceeds all other energy types, with bioenergy constituting a large part of the renewable energy (Eurostat, 2010). In some countries,

Abbreviations: ALS – Airborne laser scanning; TLS – Terrestrial laser scanning; DBH – Diameter at breast height; RMSE – Root mean square error.

such as Finland, harvesting of logging residues is already operational (Heinimö and Alakangas, 2006). This suggests that changes are taking place, which will give forest owners commercial incentives to harvest also logging residues.

Logging residues might be defined in several ways. In the Nordic countries usually the branches and the tree tops are included, but also stumps and roots may be considered to be logging residues. In the present thesis both logging residues, defined as branches and tops, and branch biomass alone are considered. It should be noted that in the present thesis the biomass of the needles are considered part of the branch biomass, and this definition is used consistently throughout the thesis. Others have used the term *crown biomass* for this – strictly speaking – combined biomass of branches and foliage (Zianis et al., 2005; Affleck and Turnquist, 2012).

Airborne laser scanning (ALS) is today used as an effective tool to collect data for forest inventories, and has transcended from an experimental to an operational state during the period from the early 2000s till today. Currently, a majority of the forest management inventories carried out in the Nordic countries are based on ALS data.

The research into the use of an airborne laser as a tool for registration of forest attributes goes back to the 1970s where works by V. I. Solodukhin at the Leningrad Scientific Research Institute of Forestry in the USSR, are considered to be the first published studies of the topic (Solodukhin et al., 1977, 1979). In these studies data from a profiling laser scanner was used to derive tree crown widths. Use of a profiling laser for forestry applications was further described in several studies from the 1980s and onwards (e.g. Nelson et al. 1984). ALS systems – with an emphasize on *scanning*, as opposed to *profiling* – became operational in the mid-1990s (Pfeifer and Briese, 2007), and experimental application in forestry for these scanning systems was shortly after described (Næsset, 1997). Several features make ALS attractive from a forest inventory perspective: Full area coverage data is being captured for all stands in a region or on a property, independent of factors such as proximity to forest roads that would have affected the costs of a manual field survey. This is however also true for aerial photos which through manual photo-interpretation in many places were the predecessor of ALS in forest inventories. But the three-dimensional structural data captured by ALS enables estimation of forest properties – such as single tree position – that is practically unobtainable through manual surveys or by interpretation of aerial photos. Previous studies also show that the use of ALS can give inventories with better accuracies and with lower costs, than aerial photo-interpretation (Eid et al., 2004). It shall however be noted that recent advances in photogrammetric methods could revive the use of digital aerial photos in forest inventory estimations, by providing automatically derived vegetation height measurements similar to those obtained with ALS (Næsset, 2002a; Baltsavias et al., 2008).

Two distinct methodologies using ALS data for forest inventories have emerged, commonly referred to as the *area-based* and the *single-tree* method. The former and most widely used procedure, requiring moderate to low point density ALS data is described by Næsset (2002b). In this approach forest properties such as mean stem volume are predicted for area units, typically of size 100 – 300 m^2. The predictions are based on statistical models, linking ALS derived features with field measurements of the desired biophysical properties from sample plots on a per-area basis. In a sampling context this is known as so-called small area estimation (Särndal et al., 2003). The single-tree method requires ALS data with a higher point density

and in this approach the aim is to derive properties, including position, of *each individual tree* (Hyyppä et al., 2001; Persson et al., 2002; Popescu and Wynne, 2004; Koch et al., 2006; Solberg et al., 2006; Wang et al., 2008; Ene et al., 2012). Interesting and promising results have been presented with the single-tree approach, but it must still be regarded as being under development. A future realization of a precise single-tree inventory coupled with a tailored management system could increase the effectiveness in the utilization of forest resources.

The same laser ranging principles (see section 2.1) as in the airborne scanning systems have also been applied in scanning from fixed positions on the ground, commonly known as terrestrial laser scanning (TLS). TLS is considered a potential tool for capturing information for forest inventories, possibly replacing time-consuming manual field registrations, or even better: capture structural information unobtainable through conventional field measurements. The use of TLS for forestry applications has not entered an operational state to the same degree as ALS, but TLS is subject to a growing focus from the forest inventory research community.

Forest owners, governmental institutions and other actors in the forest sector rely on forest inventories for planning and management of the forest resources on a property, regional or national level. The focus in forest inventories has traditionally been on timber volume for potential utilization in the forest industries. An increase in commercial harvesting of biomass components – such as tops and branches – should spark the demand for estimates of biomass components in forest inventories and resource management plans. The objective of this thesis was to describe methods to meet this demand, and more specifically to derived methods for estimation of potential logging residues using laser scanner data. The estimated logging residues are considered “potential” in the sense that these are components of the standing tree that will become logging residues when the tree is harvested. The aim of the thesis was pursued by conducting four studies covering different types of laser scanning and methodological approaches. Potential logging residues were estimated using ALS data with moderate density using an area-based approach (Paper 1). Focusing on branch biomass at the single-tree level, the relationship between accurately measured branch biomass (destructive sampling) and remote sensing data were assessed. This was done with both ALS data (Paper 2) and TLS data (Paper 3). Finally, a study with the combined use of TLS and ALS was conducted, with the aim of bringing single-tree inventory estimation of branch biomass closer to an operational stage (Paper 4). The relationship between the papers, the methodological approaches and the type of remote sensing data used are outlined in Fig. 1.

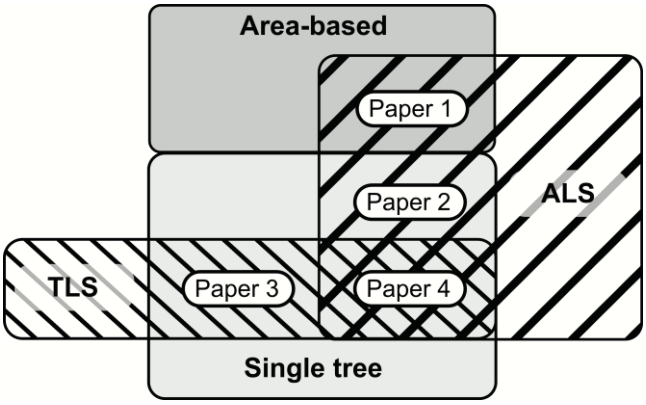

Fig. 1. The four papers in this thesis related to methodological approaches and type of remote sensing data used.

2 Background

2.1 Lidar technology

The principle behind the light detection and ranging (lidar) technology used in airborne and terrestrial laser scanning is to measure the time of travel of an emitted laser signal, from an instrument to a target and back (Wehr and Lohr, 1999). Combined with precise positioning of the instrument, an accurate position can be obtained for the points where the laser signal is reflected. Two different range measurement principles are commonly used in lidar instruments: A pulse ranging method and a phase-based ranging method (Wehr and Lohr, 1999). In the first method the laser is emitted in pulses, and the time of travel is recorded for each reflected echo. In the second method the phase-shift of the backscattered signal from a continuous laser beam is used to infer the time of travel. In ALS systems the pulse method is most commonly used, and in TLS systems both methods are applied, but the use of phase-based terrestrial scanners is increasing due to their ability to capture data at a higher rate than pulse scanners (Lichti et al., 2008). The higher capture rate comes however at the cost of a shorter maximum measuring distance, caused by phase ambiguity. In pulse range measurement systems, two methods of recording and storing the range data are commonly used today, namely discrete return and full-waveform (Pfeifer and Briese, 2007). In a discrete return system the returned echoes are determined from the backscattered laser signal in real-time, by using an algorithm with a threshold value (Fig. 2). Discrete return systems are often able to derive multiple returns from each backscattered laser pulse and only the range and position of these discrete echoes are stored. In a full-waveform system the entire backscattered signal is digitized, storing much more information than in discrete return systems. In these full-waveform systems discrete returns can be extracted during post-processing. ALS systems are sometimes characterized by referring to the size of the area at the target illuminated by the laser, i.e. by having a *large* or *small* footprint. Laser footprints with diameters < 1-2 m are typically considered small, while large footprints can range up to 70 m or more in diameter. In the present thesis small footprint discrete return ALS data and phase-based TLS data were used.

Spatial information derived from lidar is in many applications limited to the three-dimensional coordinates of the laser echoes. The path of travel of the laser signal, i.e., the vector from the lidar instrument to the target, does also contain spatial information, but this information is rarely considered. Given that all lidar measurements are in fact a result of a laser signal emitted in a *known direction* this information is – at least in principle – available from all sensors, both airborne and terrestrial. With a discrete return sensor the line along the path of travel of a laser signal from the sensor to the first recorded echo, and between any subsequent echoes, describes a space without enough matter to trigger a return. So when the laser *echoes* give information of where (biological) matter does exist, the laser signal *vectors* give information of where matter does not exist (at least not in sufficient quantities to trigger a return). The information from the laser signal vectors can therefore be used to distinguish between two distinct categories of areas with no recorded laser echoes, or seemingly *empty spaces* in the lidar data: Spaces that are empty because not enough matter exists there to trigger a return, and spaces that are empty because no laser signal has travelled through them. In discrete return systems, as those used in the ALS system applied in this thesis, certain sensor-specific parameter settings will however affect this laser signal vector information. These sensor-specific parameters include the echo-detecting algorithm and the energy threshold at which an echo is recorded, and in multiple return systems the length of the *blind zone* after each echo (Fig. 2). In this blind zone no echo is recorded, no matter how high the energy of the returned signal. Examination of the ALS data used in this thesis shows the blind zone in these particular acquisitions to be approximately 2 m after the first recorded echo, and about 4 m after the subsequent echoes. This corresponds with informal system characteristics given by the data vendor. An assessment of the effect of these sensor-specific parameters, which certainly

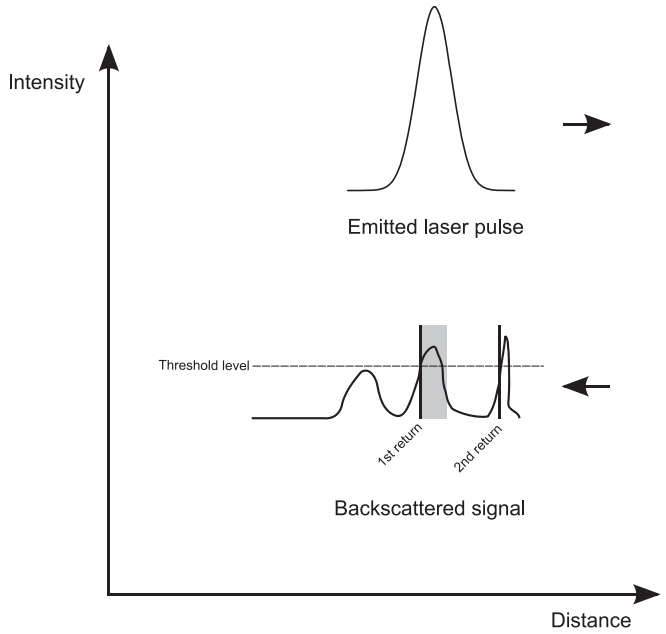

Fig. 2. Schematic representation of an emitted and backscattered laser signal. Echoes are registered at the points where the intensity of the reflected light exceeds a threshold level. The gray area indicates a 'blind zone', where no new echoes are recorded.

also affects the recorded *echoes*, was considered to be beyond the scope of this thesis. It can however be noted that, if a full waveform system had been used these effects could to a large extent be accounted for.

The spatial information inherent in the laser signal vectors was used in Paper 2 to identify laser pulses penetrating an interpolated crown surface constructed from the ALS echoes. In Paper 3 the laser signal vector information was used to distinguish between truly empty spaces and seemingly empty spaces caused by shadowing in the TLS data. The utilization of the laser signal vector information is further discussed and described in Paper 2 and 3.

2.2 Allometric biomass models

Single-tree biomass measurements are often needed as reference data in estimation of biomass at various spatial scales. Since the moisture content of trees vary, dry weight is typically preferred when quantifying forest biomass. Accurate measurements of the dry weight of biomass-components of a tree do however involve a time-consuming and work-intensive process, which include destructive sampling, weighing and drying. Consequently, the use of existing allometric models is common when obtaining single-tree forest biomass data from field measurements. These allometric models are typically species-specific models with diameter at breast height (DBH), and sometimes also tree height or other characteristics as explanatory variables. Allometric biomass models are usually derived using a set of destructively sampled trees, and there are studies that have collected and reviewed a large number of available regional and national allometric biomass models in e.g. Europe (Zianis et al., 2005) and North America (Ter-Mikaelian and Korzukhin, 1997; Jenkins et al., 2003, 2004). In addition to models for total above-ground biomass, also models for biomass of components such as branches, bark, roots, and foliage are available for certain species and regions. Allometric biomass models developed in Sweden, by Marklund (1988) were used in this thesis. Since Norway and Sweden are neighbouring countries and have similar forest types, the models by Marklund have been assumed to also apply to Norwegian conditions, and have been used in several recent studies in Norway (Næsset and Gobakken, 2008; Solberg et al., 2010; Næsset, 2011; Næsset et al., 2011; Breidenbach et al., 2012; Gobakken et al., 2012). The assumption of the applicability of Marklund's models to Norwegian conditions is reasonable, but a large-scale systematic validation of this assumption has not yet been carried out. Some limited investigations have however been carried out, and e.g. Bollandsås et al. (2009) did not find any significant bias when applying Marklunds models for branch biomass to mountain birch (*Betula pubescens* spp. *czerepanóvii*) in one region in Norway. The biomass models by Marklund are developed from a large set of trees systematically sampled from the whole of Sweden. Biomass models for separate estimation of stem wood, branches, foliage, stump, and roots are given for Norway spruce (*Picea abies* (L.) Karst.), Scots pine (*Pinus sylvestris* L.), and Birch (*Betula pubescens* Ehrh.). In total 1286 trees were sampled – 551 spruce trees, 493 pine trees and 242 birch trees. When compared to the material used by others to develop allometric biomass models this is a rather large dataset. In fact, of all the biomass model studies reviewed by Ter-Mikaelian and Korzukhin(1997), Jenkins et al.(2004), and Zianis et al. (2005) only one study used a larger dataset, and no single study with branch biomass models used a dataset comparable in size to the one used by Marklund. The errors associated with the biomass models as given by Marklund vary between the models for the different tree

Table 1. Mean percentage errors given for allometric biomass models by Marklund (1988). For models with DBH and tree height as explanatory variables, and models with DBH as the only explanatory variable (in parenthesis).

	Norway spruce	Scots pine	Birch
Stem wood	15.5% (24.7%)	19.3% (30.1%)	14.7% (24.8%)
Living branches	38.7% (40.2%)	48.1% (55.3%)	– (57.1%)

components (Table 1). Note that the errors associated with branch biomass predictions from these allometric models are larger than for predictions of e.g. stem wood. It means that branch biomass varies more than stem wood biomass for a given diameter and height, which is intuitively reasonable. This implies however that stem wood biomass can be more accurately predicted than branch biomass, using these models. It is also evident from the mean errors given in Table 1 that to add tree height information will increase the accuracy of the estimates more for stem wood than for branch biomass.

2.3 Forest biomass estimation with remote sensing data

The recent focus on the role of carbon in relation to climate change has increased the interest for estimation of biomass in forests. Forests are a large and renewable reservoir of carbon, and various remote sensing techniques have been proposed as means to measure the biomass in a particular region or area. Large area estimation of forest biomass has been proposed using spaceborne remote sensing data, such as satellite imagery (Fuchs et al., 2009), spaceborne laser (Boudreau et al., 2008) and synthetic aperture radar data (Solberg et al., 2010).

2.3.1 Biomass estimation with ALS

Several studies investigate the use of ALS to estimate forest biomass. Most of these studies follow an area-based approach, i.e. biomass estimates are derived at the plot, stand or regional level (Lim and Treitz, 2004; Hall et al., 2005; Næsset and Gobakken, 2008; Gonzalez et al., 2010; Næsset, 2011; Gleason and Im, 2012a). The general methodological approach in these studies is similar to that described by Næsset (2002b), and can be summarized as follows: The ALS data are first assigned to area units, typically raster cells or sample plots. The ALS echoes in each area unit are then analyzed, and a range of numerical features are derived. Usually the features are computed from the height distribution of the laser echoes, and typically include measures such as moments and order statistics of the distributions. Also features derived from the returned intensity of the echoes have been used. The ALS features are then related to ground reference values through regression models. The ground reference values are in the case of biomass usually derived with field measurements of e.g. DBH which is used with existing allometric models to get the ground reference biomass. Single-tree ground reference biomass is aggregated to get total biomass at plot level. The ALS regression models are then used to predict biomass for new area units. Most of the studies report total above ground biomass, but some – such as Lim and Treitz (2004) – also report separate estimates for biomass components, like foliage. Lately there have also been studies investigating the estimation of forest biomass over large areas using ALS as a sampling tool. In this approach only selected parts (a sample) of the area of interest are scanned with an airborne laser scanner. Statistical sampling strategies and appropriate estimators are then

used to derive biomass estimates for the entire area (Andersen et al., 2011; Gobakken et al., 2012; Stephens et al., 2012).

A few studies have proposed methods to estimate biomass at the single-tree level using ALS. Popescu (2007) estimated single-tree DBH from ALS data and used this ALS-derived DBH with allometric models to get total above-ground biomass. Räty et al. (2011) investigated the relationship between single-tree biomass components and ALS derived variables. Gleason and Im (2012a) estimated total above-ground biomass with ALS-derived variables at the plot and single-tree level.

With the exception of Räty et al. (2011) the studies referred to in this subsection all used ground reference values obtained with existing allometric biomass models, thus the actual biomass was not directly measured.

2.3.2 Biomass estimation with TLS

TLS has been proposed as a tool that might replace manual field measurements in forest inventories. Some studies aim at estimating biomass at the plot level, using TLS data (Garcia et al., 2011; Ku et al., 2012). At the single tree level several studies propose methods to derive DBH (Simonse et al., 2003; Bienert et al., 2007; Lovell et al., 2011; Moskal and Zheng, 2012). DBH and tree positions are typically derived by analyzing a horizontal slice of the TLS data, and the stems of the trees are then found by identifying half-circles or circle-like objects. An automatic circle-fitting algorithm is typically used and DBH and tree positions derived from the fitted circles. DBH estimated from the TLS data can then be used with existing allometric models to derive single-tree biomass (Yao et al., 2011; Vonderach et al., 2012). Holopainen et al. (2011) investigated the relationship between TLS-derived crown features and accurately measured above-ground biomass, obtained with destructive sampling.

Some studies analyze the TLS data by using voxels (Hosoi and Omasa, 2006; Bienert et al., 2010). Voxels – or *volumetric pixels* – are means to represent a given space using a three dimensional regular grid. The laser echoes can be allocated to individual voxels, and by inspecting the laser beam vectors, information can be derived for each voxel based on the intersecting laser beams. Detailed structural information related to biomass can be further derived from the aggregated voxel data.

3 Materials

3.1 Study area

The study area in this thesis is in Aurskog-Høland municipality (59°50'N 11°30'E, 120-390 m a.s.l.) located in the south-eastern part of Norway (Fig. 3). The total area of Aurskog-Høland is 960 km^2, with 670 km^2 productive forest. The forest type is boreal with Norway spruce and Scots pine as the dominant tree species.

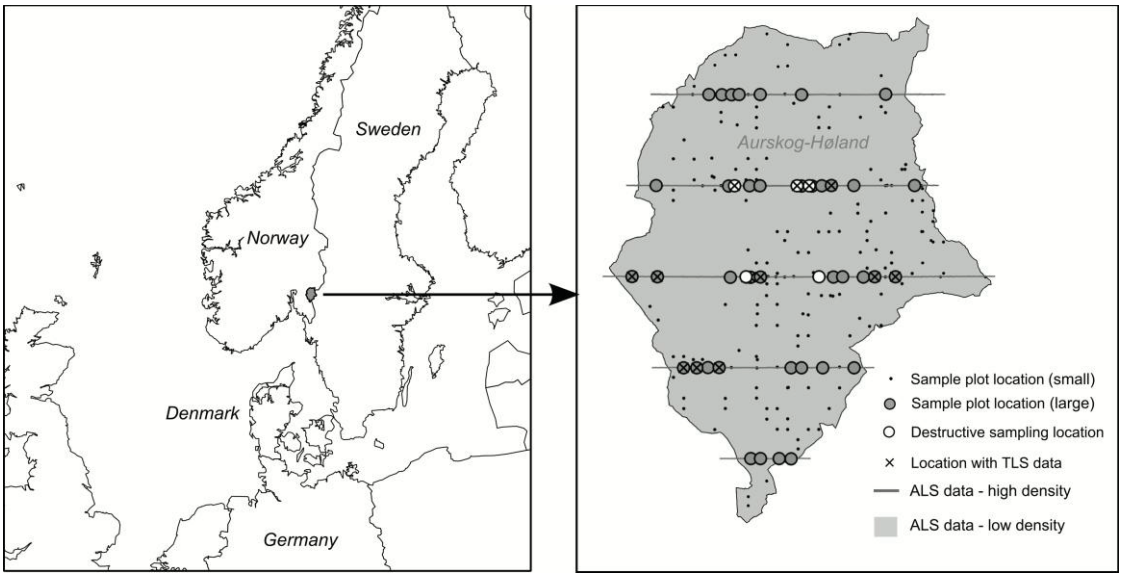

Fig. 3. The location of the study area (left), and the location of the sample plots and remote sensing data within the study area (right).

3.2 Forest inventory - stratification

An operational stand-based forest management inventory was carried out across the entire municipality prior to the field work, and this included a stratification of the productive forest. This stratification was utilized in collection of the field data used in the current thesis. The stratification was based on digital aerial photographs acquired in June 2005. Through digital stereo photogrammetry forest characteristics were interpreted manually by photo-interpretation of every stand, including stand borders, dominant tree species, site productivity and age class. The existing inventory was used as auxiliary information in the interpretation process. The H_{40} site index system was used for site productivity, where the index value is the dominant height at breast height age 40 years (Tveite, 1977; Braastad, 1980). In the present studies stands with photo interpreted H_{40} index value less or equal to 14 were defined to have poor site quality while stands with H_{40} index value higher than 14 were defined to have good site quality.

3.3 Field data

Field reference biomass was for this thesis obtained both directly by destructive sampling and weighing of individual tree components and indirectly by measuring individual tree DBH and (or) height and then deriving biomass through existing allometric models (Marklund, 1988). The DBH and tree heights were measured manually on field plots, and in one of the datasets (used in Paper 1) derived from automatically recorded harvester data. Single-tree biomass was aggregated up to plot or stand level in Paper 1.

3.3.1 Field plot data

Two sets of field plot data from the study area were used in this thesis:

1) One dataset collected during the fall of 2006 and comprising a systematic sample of 147 circular sample plots of size 200 m^2 selected in the mature productive forest of the study area. Within the plots, all trees with a DBH ≥ 4 cm were callipered. On each plot sample trees were selected with a

relascope, giving sample trees selected with probability proportional to stem basal area. Tree heights of the sample trees were measured with a hypsometer. This dataset was used in Paper 1.

2) A second dataset collected during the autumn of 2007 and the winter 2007-2008 consisting of initially 40 circular plots of size 1000 m^2 and 500 m^2, laid out in the mature productive forest of the study area. Within each of these plots all trees with DBH ≥5 cm were callipered and tree coordinates relative to the plot centre registered using a total station. On each plot 10-20 sample trees were selected according to procedures described in Maltamo et al. (2010), and on the sample trees height was measured with a hypsometer. Two different subsets of the plots in this second dataset were used in Paper 1 and Paper 4.

Coordinates of the plot centres in both datasets were obtained by differential Global Navigation Satellite Systems, using dual-frequency receivers observing pseudo-range and carrier phase of the Global Positioning System (GPS) and the Russian Global Navigation Satellite System (GLONASS). Based on the accuracy reported by the post-processing software and accuracies reported for comparable conditions in previous studies the positional accuracy is expected to be in the range of 0.01 – 0.8 m (Næsset, 1999; Hasegawa and Yoshimura, 2007; Andersen et al., 2009).

3.3.2 Harvester data

The dataset based on harvester derived DBH and tree height measurements used in Paper 1 was collected during spring and fall 2007 in 25 stands with an average size of 1.3 ha. These stands were clear-cut as part of a commercial harvesting operation. The length and diameter for each 10 cm section of all trees passing through the harvester head were measured and stored by the harvester computer. The registered diameter at 1.1 m was used as *DBH* for each stem assuming an average stump height of 20 cm. The total tree height was estimated based on the length of the measured stem and a local cut-diameter-to-top-length model.

The perimeter of each harvested stand was registered in the field after the harvest by real time kinematic positioning using GPS and GLONASS. In some stands a few individual trees were left standing inside the harvested stands (e.g. seed trees and dead trees). On these trees, height and DBH were measured in the field, and they were added to the list of harvester measured trees before the stand variables were calculated.

3.3.3 Destructive sampling

Field reference data based on destructive sampling of individual trees used in Paper 2, 3, and 4 were collected in June 2009. Five locations were chosen and from each location 10 trees were selected. The five locations were chosen to cover a range from poor to good site quality in mature forest. Within each location 10 trees of Norway spruce were selected according to the following procedure: First the diameter range of the stand was determined by measuring the DBH of the two smallest and two largest trees with a calliper. The diameter range was then divided into five classes, and two trees were selected in each class following procedures described in Paper 2. On all the 50 trees the crown projection was measured in the eight cardinal and intercardinal directions. The horizontal distance from the stem at breast

height to the vertical projection of the branch tip in the given direction was recorded.

The 50 trees were then felled, and the wet weight of the branches (including needles) of each tree was obtained by weighing the tree before and after the branches was cut off. The weighing was done with a mobile lift mounted on a truck. Samples of entire branches were selected among the living branches of each tree in order to determine the dry weight. From all the sampled branches there were taken sub-samples and these sub-samples were dried and the wet and dry weight of each was recorded. For each tree, height was measured with a hypsometer before the felling and accurately – using a measuring tape – after the felling. The total dry-weight biomass of the branches was derived based on determination of the dry-to-wet weight ratios in the selected branch samples.

The reader is referred to the individual papers for further details about the collection of the field data.

3.4 ALS data

Two sets of ALS data with unequal pulse density were used in this thesis. Both datasets were acquired with an Optech ALTM 3100 sensor on a fixed-wing aircraft. Up to four echoes were recorded for each pulse, and the planimetric coordinates and the ellipsoidal height values were by the vendor determined for all echoes. Classification of echoes into ground and vegetation echoes was carried out by iteratively fitting a triangular irregular network (TIN) from below as described by Axelsson (2000). The height above ground was calculated for all echoes by subtracting the respective TIN heights from the ellipsoidal heights. The low density data, used in Paper 1, were acquired in June 2005 with an average flying altitude of 1850 meter above ground. The pulse repetition frequency was 50 kHz, the scan frequency 71 Hz and the flight speed was 75 ms^{-1}. This resulted in an average point density on the ground of 0.7 m^{-2}. Only the first and last recorded echoes were used from this dataset. The high point density dataset used in Paper 2 and 4 was acquired in June 2006 with an average flying altitude of 800 meter above ground. The pulse repetition frequency for this acquisition was 100 kHz, the scan frequency 70 Hz and the scan angle ±5 degrees from nadir. This gave an average point density on the ground of 7–8 m^{-2}. In this dataset the corresponding position of the aircraft were provided for all echoes.

3.5 TLS data

Two sets of TLS data were used in this thesis, covering a subset of the destructively sampled trees (section 3.3.3) and a subset of the plots in the second field plot dataset (section 3.3.1).

The scans were acquired with a Leica HDS6000 phase-shift scanner in May and June 2009. The scanning was done with a horizontal and vertical angle increment of the laser measurements of 0.036 degrees. This corresponds to a point spacing of 15.9 mm at a 25 m distance from the scanner. The scanner had a beam diameter when leaving the instrument of 3 mm and a beam divergence of 0.11 mrad, resulting in an 8 mm footprint at 25 m. The scanner had a maximum measurement range of 79 m, and a full 360 x 310 degree scan (the scanners maximum field-of-view) was performed from each scanner position, i.e. both hemispheres excluding the ground directly beneath the scanner.

The scan positions were for the destructively sampled trees chosen so that each tree was scanned from at least two positions. Furthermore, the scanner position was chosen so that the top of each of the sample trees preferably was visible in at least

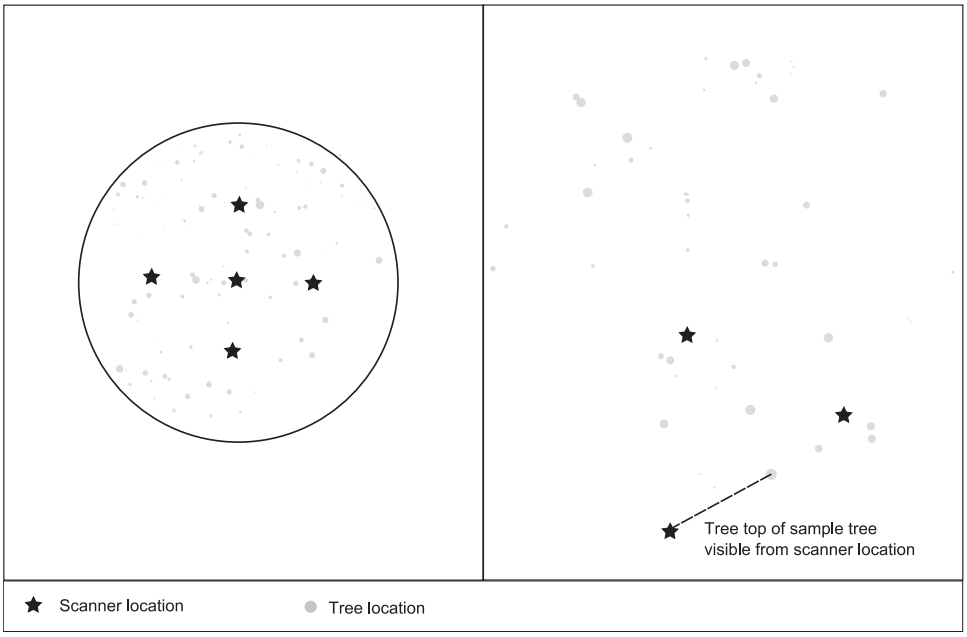

Fig. 4. Illustration of scanner locations at the sample plots (left) and on locations with destructively sampled trees (right).

one scan (Fig. 4). Scan targets were used in order to register (merge) the scans correctly. For the analysis in Paper 3 the data from each individual scan was kept separate after the registration process, and the gridded structure of the data retained. For the scanning of the sample plots five scans were performed at each plot. One scan in the plot centre and one scan in each of the cardinal directions (Fig 4).

4 Methods

4.1 Assigning laser echoes to individual trees

Several methods for automatic delineation of the ALS point cloud into single tree segments have been described (Hyyppä et al., 2001; Persson et al., 2002; Koch et al., 2006; Solberg et al., 2006; Ene et al., 2012; Gleason and Im, 2012b). Vauhkonen et al. (2012) found in a comparative study of segmentation algorithms a mean detection rate of around 70% and mean omission and commission errors of 49% and 23%, respectively. To avoid the errors introduced by automatic single-tree segmentation, the assignment of laser echoes to each individual tree was in Paper 2 based on field measured crown projections. An eight-sided polygon was formed from the crown projection measurements of each tree, and all echoes within the polygon were allocated to that tree. Since field measurements of individual tree crowns would be unavailable in an operational inventory setting, the assignment of laser echoes to individual trees was based on automatic single-tree segmentation of the ALS data in Paper 4. A marker-based watershed delineation algorithm was used (Ene et al., 2012).

4.2 Variables derived from the ALS and TLS data

In order to establish a connection between the desired biophysical property and the remotely sensed data, various features were extracted from the ALS and TLS data, and subsequently used in predictive regression models.

4.2.1 Area-based ALS-derived variables

In Paper 1, several height and density related variables were derived from the height information in the ALS data, following the methodological approach described by Næsset (2004). These variables were calculated from laser echoes spatially attributed to area units of 200 m^2.

The laser echoes with a height above ground ≥2 m were in Paper 1 considered to be canopy echoes. For these canopy echoes height and density related metrics were derived. The height metrics comprise heights at the 10^{th}, 20^{th}, .., 90^{th} percentiles of the echo height distribution. In addition the mean echo height and the corresponding coefficient of variation were computed. Density related metrics were derived by dividing the canopy (from the canopy threshold at 2 m to the 95 height percentile) into 10 height bins. Density variables were calculated as the ratio of the number of echoes above each height bin to the total number of echoes (including echoes below 2 m). A separate set of variables were derived from the first and last return ALS data.

4.2.2 Single-tree ALS-derived variables

In order to be able to estimate the biomass of the branches at a single tree level, the echoes from the crown of each individual tree were identified. The crown base height (CBH) was estimated from the ALS data for each individual tree and echoes above the CBH were denoted as crown echoes. In Paper 2, two methods of determining the CBH were considered, with the method described by Solberg et al. (2006) yielding the best results, and therefore used in the subsequent work. In this method the CBH is set at the height decile with the largest distance downwards to the next decile and with a minimum CBH at 0.85 m above ground

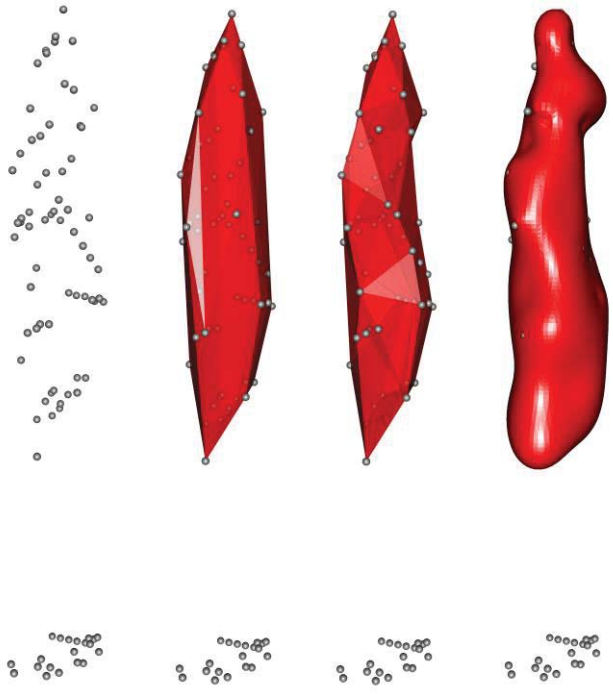

Fig. 5. Graphical representation of the ALS echoes assigned to one of the trees in the data material (left). Surfaces constructed from the crown echoes (from left to right): Three-dimensional convex hull, three-dimensional alpha-shape and radial basis function-derived surface.

(Solberg et al., 2006).

Several variables were then computed from the crown echoes assigned to each individual tree. These include height and density metrics and variables incorporating the intensity values, in addition to the volume variables described in the next paragraph. Further descriptions of the ALS derived variables are given in the individual papers.

An *a priori* assumption was that ALS derived variables related to a crown volume could be suited as predictor variables in branch biomass models. Although crown volume is an intuitive concept, it is not an unambiguously defined property of a tree. Since the presented methods do not rely on any strict definition, the term *crown volume* is – throughout this thesis – used as a reference to the *general concept,* rather than any specific technical definition. In Paper 2 and 4 several crown volume variables were derived from the crown echoes:

- The volume of a three-dimensional (3D) convex hull of the crown echoes.
- The volume of a 3D alpha shape surface constructed from the crown echoes.
- The volume of an interpolated surface of the crown echoes, calculated using radial basis functions.

The three crown volume methods are all based on procedures to construct a closed surface from scattered data points – which in this thesis would be the crown echoes (Fig. 5). A 3D convex hull of scattered points is described by e.g. Kovalevsky and Schultz (2004) and is straight-forward to compute, without any parameter values to be decided. The disadvantage of using a 3D convex hull to describe a tree crown surface is its lack of flexibility; it can only describe a strictly convex surface. The 3D alpha shape is related to the 3D convex hull, but allows for non-convex surfaces. A distance threshold parameter, called the *alpha* value must be set. The alpha value controls the concaveness of the surface, and data points beyond a distance from its neighbouring points corresponding to the alpha value will be excluded. The disadvantage of using a 3D alpha shape method to represent a crown surface is its actual exclusion of data points and the fact that it can yield multiple disconnected surface segments.

The surface constructed with radial basis functions is in contrast to the two former, a non-linearly interpolated surface. That is, the surface is not necessarily linear between two data points. The procedure applied in the present thesis forces the surface to go exactly through all the observed surface data points, but the surface is still non-linearly interpolated between these points. A radial basis function derived surface is very flexible, and can describe complex surfaces accurately, using all the observed surface points. A disadvantage is that the calculation of a radial basis function surface is complex, with several parameters and sub-algorithms that can be tuned or altered. Too few data points or an unfortunate placement of them might also lead to undesired artefacts in the interpolated surfaces. The algorithms used in the present thesis can for instance yield abnormally shaped surfaces if given fewer than five data points. The reader is referred to Edelsbrunner and Mücke (1994) for details on 3D alpha shapes and to Kato et al. (2009), Carr et al. (2001) and Paper 2 for a description of calculations of crown volume by the use of radial basis functions.

In Paper 2, an additional *corrected crown volume* was calculated, utilizing the information inherent in the laser pulse vectors. By inspecting the path of travel of the laser pulses together with the crown surfaces described above, any intersections between the path of the pulse and the surface can be

determined. The assumption was that whenever a pulse has travelled inside the crown surface, its path describes a space without enough biological matter to trigger a return. This space is defined by the crown surface at the intersection(s), the pulse vector and the radius of the return footprint perpendicular to the pulse vector. The volume of these penetrating pulses was subtracted to get a 'corrected' crown volume.

4.2.3 TLS-derived variables

Features were extracted from the TLS data for each individual tree following two distinct approaches. One based on crown dimensions measured in the unified TLS data and a second, more comprehensive voxel-based approach.

In the unified TLS data several measurements of crown dimension were carried out, and in the present thesis these were carried out manually. The measurements were done using horizontal and vertical slices of the TLS point cloud data and included measurements of crown widths and projected crown areas at given heights as well as the vertical crown length. The measurements were carried out solely using the TLS data which means that similar features could be extracted using an automated algorithmic approach.

In the voxel-based approach the unified TLS point cloud was first manually delineated into single-tree segments. Then a voxel space was created for the point cloud assigned to each individual tree. The individual voxels were then analyzed separately for each tree and scan position. By inspecting the laser beam vectors, the laser beams potentially hitting each voxel were classified into one of three categories:

i. Reflected before it entered the voxel (causing a shadow in the voxel).
ii. Reflected from within the voxel.
iii. Passed through the voxel.

From this information several features were extracted, including the ratio of entering laser beams to the number of beams reflected from within the voxel. This *reflection ratio* was taken as an indication of the amount of biological matter present within the voxel, and summed up for all voxels assigned to each individual tree. The reflection ratio for each voxel was calculated as a weighted mean of the ratios from the different scans. The number of entering beams was used as weights, giving more weight to the ratios from scans where the scanner was 'seeing' more of the voxel.

Further details about the ALS and TLS derived variables are described in the individual papers.

4.3 Predictive models

Linear least squares regression models are widely used in the literature when relating laser scanner-derived features to forest biophysical properties. Since the number of features derived from the remote sensing data can be large, some form of variable selection is often applied. The reduction of the number of variables is usually motivated by the need to get rid of highly correlated variables, and thus avoid multicollinearity, and to reduce the dimensionality of the data.

A stepwise variable selection – such as the *step* function in the R software (R Development Core Team, 2011) – is typically applied when a high number of remote sensing derived variables are used in predictive regression models. Although it is widely used, several statistical papers and textbooks advise to use this kind of procedure with caution (Babyak, 2004; Whittingham et al., 2006; Sheather, 2009). In the forest inventory research community there has in the last decade been applied several other regression techniques, in particular non-

parametric and machine learning methods (Næsset et al., 2005; Latifi et al., 2010; Vauhkonen et al., 2010; Zhao et al., 2011; Breidenbach et al., 2012; Gleason and Im, 2012a). In the present thesis random forest is tried out alongside linear least squares regression. Random forest has been found to perform well compared to other regression techniques, and with little tuning required (Hastie et al., 2003). Recent studies have reported successful use of random forest regression for forest inventory applications with remote sensing data (Latifi et al., 2010; Vauhkonen et al., 2010). The principles behind random forest regression as it is commonly used today were introduced by Breiman (2001), and in this machine learning algorithm a large number of binary trees are grown using bootstrap samples of the data. The predicted value is then the average of the predicted values from the individual binary trees. In the present thesis the implementation of random forest in the *randomForest* package in the statistical software R was used. Further details about the principles and use of the random forest algorithm can be found in Breiman (2001) and Hastie et al. (2003).

There are several reasons for using regression techniques other than linear least squares for prediction of forest properties from remote sensing data. Variable selection procedures typically applied in linear least squares regression can be avoided by using methods that are less affected by highly correlated variables and less sensitive to noise variables, such as random forest regression (Biau, 2012). Since selecting some variables, and thereby dropping others, can be seen as *throwing away* potential useful information this feature might be regarded as desirable. Alternative regression techniques might also be less dependent on assumptions about the data, such as error distributions. An alternative to both stepwise variable selection and the use of alternative regression techniques is to use linear least squares regression, and create or extract variables in such a way that no variable selection is necessary. This could be achieved by techniques such as principal component analysis (Johnson and Wichern, 2002) or simply by determining only one or a few variables a-priori and stick to these in the regression modeling. This second approach is suggested by e.g. Babyak (2004) and is reflected in this thesis by the use of the simple linear regression models in Paper 2 and 4.

4.4 Model validation – prediction accuracy

The prediction accuracy was in the present thesis assessed by computing a root mean square error (RMSE) from predicted and observed values, through a validation procedure. RMSE was computed as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(B_i - \hat{B}_i)^2}{n}}, \qquad (1)$$

where n is the number of observations, and B and $\hat{B}$ are the observed and predicted biomass of the ith observation, respectively. In the present thesis, RMSE is reported in percent of the mean observed value, i.e.

$$RMSE\% = \frac{RMSE}{\frac{\sum_{i=1}^{n} B_i}{n}} 100. \qquad (2)$$

Two validation procedures were applied: In Paper 1 and Paper 4 the models were used to predict new values in separate validation datasets. The prediction accuracies of the models were then assessed by comparing the predicted with the observed values. In Paper 2 and 3, cross-validation procedures were applied. In paper 2, the data from three out of five sample locations were chosen as training data, and the remaining two used for validation. This was repeated for

Table 2. Model statistics and prediction accuracies of the ALS-based models estimating potential logging residues at the plot and stand level.

				Plot validation		Stand validation	
Explanatory variables	Stratum	*n*[a]	R^2	*n*	RMSE%	*n*	RMSE%
ALS variables	All	147	0.86	120	27.6	25	22.2
ALS variables and tree species proportions	All	147	0.87	120	23.2	25	21.3
ALS variables	Poor sites	105	0.88	78	25.3	19	24.3
ALS variables	Good sites	42	0.77	42	22.5	6	17.6

[a] Number of observations in the model training data.

all combinations of selecting three out of five locations. In Paper 3, a Monte Carlo cross-validation procedure was used. In this procedure the dataset is repeatedly and randomly split in two, with one part used as training data for the model, and the other part used for validation. In both Paper 2 and 3 the reported prediction accuracies are the average for all the iterations. Monte Carlo cross-validation can be viewed as a simulation of repeated executions of a traditional data-splitting or holdout procedure. An advantage of Monte Carlo cross-validation over a traditional random split or holdout approach is that the results are not dependent on one single, potentially unfortunate split of the data. Other corresponding validation procedures do exist, such as *n*-fold cross-validation and validation using various forms of the bootstrap (Molinaro et al., 2005). Monte Carlo cross-validation was chosen in Paper 3 because it is straight-forward, yet with performance comparable to alternative methods (Xu and Liang, 2001; Molinaro et al., 2005; Kim, 2009).

5 Major findings

5.1 Area-based estimation of potential logging residues

A strong relationship was found between potential logging residues and ALS derived features at the plot level, with coefficient of determination (R^2) values of 0.77 – 0.88 (Table 2). When validating the models the RMSE ranged from 22.5% to 27.6% at the plot level and from 17.6% to 24.3% at the stand level. These accuracies observed through validation in two independent datasets were within the range reported by other studies estimating other forest stand characteristics by ALS (summarized by Næsset 2007). The mean errors of the estimates were however in the high end of this range, and larger than the reported mean errors for estimates of stand volume by e.g. Næsset (2007). The inclusion of plot-level species composition information did improve the prediction accuracy, whereas inclusion of stand-level species composition only led to a marginal improvement of prediction accuracy (Table 2).

5.2 Single-tree estimation of branch biomass with TLS

A strong relationship was found between branch biomass and TLS-derived features. The prediction accuracy was found to be higher than for estimates by conventional allometric models using field measured DBH and height. The best TLS model yielded in a Monte Carlo cross-validation an RMSE of 32%, compared to the corresponding RMSE for predictions with an existing allometric biomass model at 38% (Table 3). Two approaches were followed when deriving features from the TLS data, one based on manual crown measurements in the TLS data and one voxel-based approach. When comparing the individual TLS models, the two approaches yielded comparable results (Table 3). The prediction accuracy was found to decrease with decreasing voxel size, i.e. the model with variables derived with the smallest voxels yielded the poorest accuracy (Table 3).

5.3 Single-tree estimation of branch biomass with ALS

At the single-tree level the relationship between branch biomass and ALS-derived features was found to be strong. An RMSE of 35% was observed in the cross-validation of a model with ALS-derived crown volume as a single explanatory variable, and including other ALS features as explanatory variables did not improve the prediction accuracies (Table 4). Predicting branch biomass with field measured DBH and tree height and the use of allometric biomass models by Marklund (1988) corresponded in the present data material (the 50 destructively sampled trees) to an RMSE of 38%. The best ALS-based models yielded branch biomass predictions with higher accuracies than this (Table 4). The results in Paper 2 showed that inclusion of information derived from the laser pulse vectors improved the observed prediction accuracy.

With branch biomass obtain by TLS as response variable in the ALS-branch biomass models the RMSEs in the validation ranged from 32.4% to 38.8% (Table 4). Corresponding predictions derived with ALS-predicted DBH and existing allometric biomass models resulted in RMSEs of 41.3% and 44.5% (Table 4).

Table 3. Summary of the models and mean statistics (validation) for branch biomass estimation with TLS data. Corresponding statistics for estimation of branch biomass with field measured DBH and tree height.

	Explanatory variables	R^2	RMSE %
TLS models	TLS-derived crown measurement variables	0.74	34
	TLS-derived 0.1 m voxel variables	0.42	48
	TLS-derived 0.2 m voxel variables	0.66	36
	TLS-derived 0.4 m voxel variables	0.73	32
Existing allometric model	Field measured DBH and tree height	0.65	38

Table 4. Prediction accuracies observed in the validation of the single tree ALS-branch biomass models. Response variable used in the model fitting, explanatory variables in the models and regression technique used. In the case of the stepwise models, the *explanatory variables* refers to the initial set of variables entered into the stepwise procedure.

Response variable	ALS features – Explanatory variables[a]	Regression technique[b]	Accuracy of branch biomass predictions (RMSE%)
Branch biomass – destructive sampling	All	STEPWISE	45
	All	RF	38
	Crown volume	LS	35
Branch biomass – from TLS data	All	STEPWISE	39
	All	RF	39
	Crown volume	LS	32
DBH – field measured	All	STEPWISE	46[c]
	All	RF	41[c]

[a]Details are given in Paper 2 and 4.
[b]STEPWISE = stepwise linear least squares regression, RF = random forest regression, LS = simple linear least squares regression.
[c] Branch biomass predictions derived with ALS-predicted DBH and existing allometric biomass models.

6 Discussion

6.1 Area-based estimation of potential logging residues

The findings in this thesis show that good area-based estimates of potential logging residues (branches and tops) can be achieved by ALS in a boreal forest. The observed prediction accuracy was slightly poorer than the accuracies typically reported for corresponding stem volume predictions under comparable conditions (Næsset, 2007).

The assessment of the effect of tree species shows that tree species composition does influence on the relationship between the ALS derived variables and the biomass of potential logging residues and including this information improves the prediction accuracy of the model. This is in accordance with findings reported by e.g. Næsset and Gobakken (2008). The effect is however only pronounced when plot-level species proportion information is used. The use of stand level tree species proportion information at the computational cell level did not result in a similar improvement in prediction accuracy. One reason for this might be that the tree species composition in each cell may differ considerably from the overall stand species composition.

Straub and Koch (2011a) have proposed a method to estimate the stand-level bioenergy potential in terms of forest biomass by the use of ALS and multispectral aerial imagery. The proposed method relies on remote sensing estimation of stand volume, mean DBH, and mean age of the dominant trees. Although successful estimation of age is reported by Straub and Koch, age is by others (e.g. Maltamo et al., 2009) considered to be amongst the stand characteristics that are not

well suited for direct estimation by ALS. Straub and Koch did not quantify the prediction accuracy of the estimated biomass, but they report RMSEs from a leave-one-out cross-validation of the remote sensing-derived stand characteristics: 27.9%, 28.8% and 30.2% for volum, DBH and age, respectively. Based on these mean errors it is likely that potential logging residues were predicted with higher accuracies in the present thesis. Although there are slight differences in the definition of logging residues, a comparison between the approach in the present thesis and the approach presented by Straub and Koch (2011a) could be subject to further investigations.

6.2 Estimation of single-tree branch biomass with TLS

TLS differ from ALS in many aspects, although the same ranging principles are employed in both systems. One of the major differences from a forestry perspective is the TLS systems lack of ability to cover large areas. The collection of data over large areas is one of the great advantages of ALS, and this difference between TLS and ALS reveals the very different possible areas of application for the two systems. TLS does on the other hand provide highly detailed data at the single-tree and even sub-tree level. One obvious possible application of TLS is therefore to replace manual field measurements, either by increasing the efficiency of the data collection, or by enabling measurements of structural properties that would be unobtainable with manual registrations. In the present thesis, TLS predicted branch biomass was compared to branch biomass predictions obtained with allometric models using field measured DBH and tree height. The observed prediction accuracy was higher for the TLS based models than for predictions with existing allometric biomass models using DBH and tree height. In other words, better predictions of single-tree branch biomass were obtained with TLS data than through field measurements of DBH and tree height. The improvement was however relatively small, with the best TLS based model yielding an RMSE of 32%, a decrease of six percent points from the observed 38% RMSE of the predictions based on conventional field measurements. It should be noted that the errors observed for the existing allometric models by Marklund (1988) in the present data material corresponds well with the errors actually reported by Marklund, i.e. a mean error of 39% for branch biomass when using DBH and tree height as explanatory variables.

In the present thesis, TLS data were acquired through multiple scans at each location. The raw data were then post-processed before it was analyzed. The procedure involves thus both manual field work (setting up and administer the scanner equipment) and computer-based processing of the raw data. In the present thesis, several steps in the post-processing and the analysis also required manual work. In light of this rather extensive data collection and processing procedure, the gain in prediction accuracy over conventional field measurements seems to be small. A lack of adequate existing allometric models, or TLS data already collected for other purposes might nevertheless render the proposed method suitable. This requires however an available TLS branch biomass model. Since destructive sampling is costly and time-consuming, the transferability of such models from one region to another could be subject to further research. Artificially created training data might also in the future become an alternative to destructive sampling.

In the case of the voxel-based derivations of TLS features this thesis uncovered some methodological challenges, in particular related to the delineation of the point cloud into single tree crown segments. It

became evident that this delineation can be a difficult task with Norway spruce, mainly due to the combined effect of shadowing in the TLS data and close or overlapping crowns. Improving the delineation by the use of advanced segmentation algorithms utilizing the structural and distributional properties of the TLS data could be subject to further research.

6.3 Estimation of single-tree branch biomass with ALS

Few previous studies report comparable ALS-estimates of branch biomass validated against accurate field measurements. Räty et al. (2011) investigated correlation between ALS-derived features and biomass components. The highest correlation found for Norway spruce branch biomass was 0.71. A stronger relationship was found in the present thesis.

Some previous studies report ALS-based estimates of other single tree properties such as stem volume. Straub and Koch (2011b)report an RMSE of 24% in a study considering stem volume of Scots pine. This accuracy is higher than the accuracy found in the present study, and might be attributed to the fact that stem volume is more closely related to the tree height than branch biomass. Tree height is known to be well estimated by ALS (Persson et al., 2002). Other previous studies estimate total aboveground biomass at the single tree level (Popescu, 2007; Gleason and Im, 2012a). Compared to those studies higher accuracies were obtained for branch biomass in the present thesis. Popescu (2007) reported an RMSE of 47% when estimating total aboveground biomass of individual pine trees. Gleason and Im (2012a) reported RMSEs in the range of 68.1% – 119.6% for estimation of single-tree aboveground biomass in a mixed forest in New York State, the USA. The two studies are however not directly comparable to the present thesis since total rather than branch biomass were estimated. Moreover, accurate field measurements of biomass were not available in the material used by Popescu and Gleason and Im, thus validation was carried out using general allometric biomass models. The impact of using general models rather than accurate field measurements is however smaller than for estimation of branch biomass since the errors associated with branch biomass derived with general allometric models usually are larger than for total aboveground biomass (Marklund, 1988; Jenkins et al., 2003; Zianis et al., 2005).

Crown volume derived from the ALS data was in the present thesis found to be a good predictor variable for branch biomass. This is in contrast to findings by Räty et al. (2011) who reported a higher correlation between Norway spruce branch biomass and variables derived from the height distribution of the laser echoes, than between branch biomass and crown volume. A three dimensional convex hull crown volume was used by Räty et al. and the reported correlation with branch biomass was as low as 0.08. Findings in Paper 2 and 4 suggests that a more flexible crown volume algorithm might be better, and this could partly explain the weak relationship found in that study. Räty et al. did however find that for Scots pine the projected crown area was the ALS-derived feature with the highest correlation with branch biomass. Crown area is related to crown volume, and is also a geometrical derived feature. A single-tree crown volume calculated from an ALS-derived canopy height model have been proposed in several studies (Chen et al., 2007; Kwak et al., 2010; Jung et al., 2011; Gleason and Im, 2012a). Both Kwak et al. and Gleason and Im use this kind of crown volume as an explanatory variable when estimating single-tree aboveground

biomass. The lack of flexibility and the shortcoming in describing the true three-dimensional structure of a tree crown are disadvantages to that approach.

When comparing the predictions from the ALS models with branch biomass estimated by existing allometric models and actual field measured DBH and height, the ALS based models yielded a slightly better accuracy. In other words, the ALS data in the present thesis contained more information related to branch biomass than the actual field measured DBH and tree height.

In the data material used here the total weight measurements include dead branches. Since the fraction of dead branches was unknown, only samples from the live branches were used to calculate the wet-to-dry weight ratio. Consequently, a small error might have been introduced by this procedure, subject to the difference in moisture content of dead and live branches. The impact on the branch biomass figures is however assumed to be small.

6.4 Predictive models – regression techniques

In the present thesis different types of regression models were used and compared. These models can be divided into three categories: stepwise linear least squares models, simple linear models, and machine learning regression models (random forest). Although the studies in this thesis by no means were design to assess the performance and compare these regression techniques, some remarks will be given. In Paper 2 and 4 branch biomass was estimated by simple linear regression models with an ALS-derived crown volume as a single explanatory variable. These simple linear models performed well compared to the multiple regression models, in fact both in Paper 2 and in Paper 4 a simple linear model yielded the best prediction accuracy. This might be transferable to other types of remote sensing predictions but whereas crown volume is an intuitive variable for prediction of branch biomass, similar suitable single variables might be harder to find for other biophysical properties.

Accurate predictions from random forest regression require that the training data cover the entire range of the target population, and one feature of a random forest regression model is that it will not yield extrapolated values, i.e. values outside the feature space of the training data. This is to some extent a desirable property, since it ensures that only ‘sound’ values may be predicted. A consequence is that the predicted values at the extremes of the range are drawn towards the mean, thus introducing a bias in the predictions.

Random forest regression models and linear stepwise models were applied in Paper 2 through 4. The same set of initial explanatory variables was used in both regression modelling approaches. The prediction accuracy was in all the three studies similar or higher for the random forest model, compared to the stepwise linear model. Thus, random forest regression performed in the present thesis generally better than stepwise linear regression.

7 Conclusions

In this thesis the potential of estimating biomass components – in particular branch biomass – by airborne and terrestrial laser scanning was investigated. The results show that good predictions can be obtained both with airborne and terrestrial scanning. In a boreal forest, predictions of potential logging residues (branches and tops) were obtained from ALS data at the stand level, with accuracy comparable to other ALS-predicted stand attributes. The proposed procedure relies on

the same field and ALS data as those collected in many operational forest management inventories, and can be implemented with low or no additional costs – thus enabling the inclusion of potential logging residues in an area-based ALS forest inventory.

At the single-tree level prediction of Norway spruce branch biomass was investigated and good predictions obtained by using TLS data. The prediction accuracy was higher than with a conventional method using field measured DBH and existing allometric biomass models. This suggests that TLS can be a viable alternative to field measurements when collecting single-tree ground reference data. The gain, in terms of prediction accuracy must however be judged against the effort involved in acquiring and analyzing the TLS data.

Single-tree branch biomass was further successfully predicted with ALS data, showing the potential for including branch biomass in single-tree forest inventories. A methodological framework using both ALS and TLS was proposed, with TLS used to obtain ground reference values for ALS models. The approach yielded good predictions for Norway spruce branch biomass, suggesting that the methodology could be applied in ALS-based single-tree inventories.

The transferability of the findings to other tree species and forest types could be subject to further research. Also, the impact of using different scanner instruments and settings could be further investigated.

References

Affleck, D.L.R., Turnquist, B.R., 2012. Assessing the accuracy of crown biomass equations for the major commercial species of the interior northwest: study plan and preliminary results, in: McWilliams, Will; Roesch, Francis A. Eds. 2012. Monitoring Across Borders: 2010 Joint Meeting of the Forest Inventory and Analysis (FIA) Symposium and the Southern Mensurationists. e-Gen. Tech. Rep. SRS-157. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. pp. 247–254.

Andersen, H.-E., Clarkin, T., Winterberger, K., Strunk, J., 2009. An Accuracy Assessment of Positions Obtained Using Survey- and Recreational-Grade Global Positioning System Receivers across a Range of Forest Conditions within the Tanana Valley of Interior Alaska. Western Journal of Applied Forestry 24, 128–136.

Andersen, H.-E., Strunk, J., 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 Journal of Applied Forestry 26, 157–164.

Axelsson, P., 2000. DEM generation form laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing XXXIII, 110–117.

Babyak, M.A., 2004. What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models. Psychosomatic Medicine 66, 411–421.

Baltsavias, E., Gruen, A., Eisenbeiss, H., Zhang, L., Waser, L.T., 2008. High-quality image matching and automated generation of 3D tree models. Int. J. Remote Sens. 29, 1243–1259.

Biau, G., 2012. Analysis of a Random Forest Model. Journal of Machine Learning Research 13, 1063–1095.

Bienert, A., Queck, R., Schmidt, A., Bernhofer, C., Maas, H., 2010. Voxel space analysis of terrestrial laser scans in forests for wind field modeling. International Archives of Photogrammetry,Remote Sensing and Spatial Information Sciences XXXVIII, Part 5, Commission V, 92–97.

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. International Archives of Photogrammetry and Remote Sensing 3, 50–55.

Bollandsås, O.M., Rekstad, I., Næsset, E., Røsberg, I., 2009. Models for predicting above-ground biomass of Betula pubescens spp. czerepanóvii in mountain areas of southern Norway. Scandinavian Journal of Forest Research 24, 318–332.

Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., Kimes, D.S., 2008. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. Remote Sensing of Environment 112, 3876 – 3890.

Breidenbach, J., Næsset, E., Gobakken, T., 2012. Improving k-nearest neighbor predictions in forest inventories by combining high and low density airborne laser scanning data. Remote Sensing of Environment 117, 358–365.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32.

Braastad, H., 1980. Growth model computer program for Pinus sylvestris. Reports of the Norwegian Forest Research Institute 35, 272–359.

Carr, J.C., Beatson, R.K., Cherrie, J.B., Mitchell, T.J., Fright, W.R., McCallum, B.C., Evans, T.R., 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.

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, 1355–1365.

Edelsbrunner, H., Mücke, E., 1994. 3-Dimensional Alpha-Shapes. Acm Transactions on Graphics 13, 43–72.

Eid, T., Gobakken, T., Naesset, E., 2004. Comparing stand inventories for large areas based on photo-interpretation and laser scanning by means of cost-plus-loss analyses. Scandinavian Journal of Forest Research 19, 512–523.

Ene, L., Næsset, E., Gobakken, T., 2012. Single tree detection in heterogeneous boreal forests using airborne laser scanning and area-based stem number estimates. International Journal of Remote Sensing 33, 5171–5193.

Eurostat, 2010. Europe in figures - Eurostat yearbook 2010. Eurostat.

Fuchs, H., Magdon, P., Kleinn, C., Flessa, H., 2009. Estimating aboveground carbon in a catchment of the Siberian forest tundra: Combining satellite imagery and field inventory. Remote Sensing of Environment 113, 518 – 531.

Gan, J., Smith, C.T., 2006. Availability of logging residues and potential for electricity production and carbon displacement in the USA. Proceedings of the 4th annual workshop of IEA Bioenergy Task 31 “Sustainable Production Systems for Bioenergy: Forest Energy in Practice”. 30, 1011–1020.

Garcia, M., Danson, F.M., Riano, D., Chuvieco, E., Ramirez, F.A., Bandugula, V., 2011. Terrestrial laser scanning to estimate plot-level forest canopy fuel properties. International Journal of Applied Earth Observation and Geoinformation 13, 636–645.

Gleason, C.J., Im, J., 2012a. Forest biomass estimation from airborne LiDAR data using machine learning approaches. Remote Sensing of Environment 125, 80–91.

Gleason, C.J., Im, J., 2012b. A Fusion Approach for Tree Crown Delineation from Lidar Data. Photogrammetric Engineering and Remote Sensing 78, 679–692.

Gobakken, T., Næsset, E., Nelson, R., Bollandsås, O.M., Gregoire, T.G., Ståhl, G., Holm, S., Ørka, H.O., Astrup, R., 2012. Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. Remote Sensing of Environment 123, 443–456.

Gonzalez, P., Asner, G.P., Battles, J.J., Lefsky, M.A., Waring, K.M., Palace, M., 2010. Forest carbon densities and uncertainties from Lidar, QuickBird, and field measurements in California. Remote Sensing of Environment 114, 1561–1575.

Hall, S.A., Burke, I.C., Box, D.O., Kaufmann, M.R., Stoker, J.M., 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.

Hasegawa, H., Yoshimura, T., 2007. Estimation of GPS positional accuracy under different forest conditions using signal interruption probability. Journal of Forest Research 12, 1–7.

Hastie, T., Tibshirani, R., Friedman, J., 2003. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 1st ed. 2001. Corr. 3rd printing. ed. Springer. pp. 552.

Heinimö, J., Alakangas, E., 2006. Solid and Liquid Biofuels Markets in Finland - a study in international biofuels trade (Research report No. EN A-53). Lappeenranta University of Technology.

Holopainen, M., Vastaranta, M., Kankare, V., Räty, M., Vaaja, M., Liang, X., Yu, J., Hyyppä, H., Hyyppä, R., 2011. Biomass estimation of individual trees using stem and crown diameter TLS measurements. ISPRS Workshop on Laser scanning XXXIII.

Hosoi, F., 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, 3610–3618.

Hughes, D., 2011. Will natural gas fuel America in the 21st century? Post Carbon Institute Report. 64.

Hyyppä, J., Kelle, O., Lehikoinen, 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.

IPCC, 2007. Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC.

Jenkins, J., Chojnacky, D., Heath, L., Birdsey, R., 2003. National-scale biomass estimators for United States tree species. Forest Science 49, 12–35.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., Birdsey, R.A., 2004. Comprehensive database of diameter-based biomass regressions for North American tree species, General technical report. Department ofAgriculture, Forest Service., Newtown Square, PA: U.S.

Johnson, R.A., Wichern, D.W., 2002. Applied multivariate statistical analysis - Fifth Edition. Prentice Hall, Upper Saddle River, N.J. pp. 767.

Jung, S.-E., Kwak, D.-A., Park, T., Lee, W.-K., Yoo, S., 2011. Estimating Crown Variables of Individual Trees Using Airborne and Terrestrial Laser Scanners. Remote Sensing 3, 2346–2363.

Kato, A., Moskal, L.M., Schiess, P., Swanson, M.E., Calhoun, D., Stuetzle, W., 2009. Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. Remote Sensing of Environment 113, 1148–1162.

Kim, J.-H., 2009. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. Computational Statistics & Data Analysis 53, 3735–3745.

Koch, B., Heyder, U., Wein, H., 2006. Detection of individual tree crowns in airborne lidar data. Photogrammetric Engineering and Remote Sensing 72, 357–363.

Kovalevsky, V., Schulz, H., 2004. Convex hulls in a 3-dimensional space, in: Klette, R., Zunic, J. (Eds.), Combinatorial Image Analysis, Proceedings. Springer-Verlag Berlin, Berlin, pp. 176–196.

Ku, N.W., Popescu, S.C., Ansley, R.J., Perotto-Baldivieso, H.L., Filippi, A.M., King, B., Li, J., 2012. Assessment of available rangeland woody plant biomass with a terrestrial lidar system. PE&RS, Photogrammetric Engineering & Remote Sensing 78, 349–361.

Kwak, D.-A., Lee, W.-K., Cho, H.-K., Lee, S.-H., Son, Y., Kafatos, M., Kim, S.-R., 2010. Estimating stem volume and biomass of Pinus koraiensis using LiDAR data. Journal of Plant Research 123, 421–432.

Latifi, H., Nothdurft, A., 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, 395–407.

Lichti, D., Pfeifer, N., Maas, H., 2008. ISPRS Journal of Photogrammetry and Remote Sensing theme issue "Terrestrial Laser Scanning". ISPRS Journal of Photogrammetry and Remote Sensing 63, 1–3.

Lim, K.S., Treitz, P.M., 2004. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. Scandinavian Journal of Forest Research 19, 558–570.

Lovell, J.L., Jupp, D.L.B., Newnham, G.J., Culvenor, D.S., 2011. Measuring tree stem diameters using intensity profiles from ground-based scanning lidar from a fixed viewpoint. ISPRS Journal of Photogrammetry and Remote Sensing 66, 46–55.

Mabee, W.E., Saddler, J.N., 2010. Bioethanol from lignocellulosics: Status and perspectives in Canada. Bioresource Technology 101, 4806–4813.

Malinen, J., Pesonen, M., Määttä, T., Kajanus, M., 2001. Potential harvest for wood fuels (energy wood) from logging residues and first thinnings in Southern Finland. Biomass and Bioenergy 20, 189–196.

Maltamo, M., Bollandsås, O.M., Vauhkonen, J., Breidenbach, J., Gobakken, T., Næsset, E., 2010. Comparing different methods for prediction of mean crown height in Norway spruce stands using airborne laser scanner data. Forestry 83, 257–268.

Maltamo, M., Packalén, P., Suvanto, A., Korhonen, K., Mehtätalo, L., 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.

Marklund, L.G., 1988. Biomass functions for pine, spruce and birch in Sweden (Report No. 45). Swedish University of Agricultural Sciences, Umeå.

Molinaro, A.M., Simon, R., Pfeiffer, R.M., 2005. Prediction error estimation: a comparison of resampling methods. Bioinformatics 21, 3301–3307.

Moskal, L.M., Zheng, G., 2012. Retrieving Forest Inventory Variables with Terrestrial Laser Scanning (TLS) in Urban Heterogeneous Forest. Remote Sensing 4, 1–20.

Nelson, R., Krabill, W., MacLean, G., 1984. Determining forest canopy characteristics using airborne laser data. Remote Sensing of Environment 15, 201–212.

Næsset, E., 1997. Estimating timber volume of forest stands using airborne laser scanner data. Remote Sensing of Environment 61, 246 – 253.

Næsset, E., 1999. Point accuracy of combined pseudorange and carrier phase differential GPS under forest canopy. Canadian Journal of Forest Research 29, 547–553.

Næsset, E., 2002a. Determination of Mean Tree Height of Forest Stands by Digital Photogrammetry. Scandinavian Journal of Forest Research 17, 446–459.

Næsset, E., 2002b. 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-footprint airborne scanning laser. Scandinavian Journal of Forest Research 19, 164–179.

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., 2011. Estimating above-ground biomass in young forests with airborne laser scanning. International Journal of Remote Sensing 32, 473–501.

Næsset, E., Bollandsås, O.M., Gobakken, T., 2005. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. Remote Sensing of Environment 94, 541 – 553.

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.

Næsset, E., Gobakken, T., Solberg, S., Gregoire, T.G., Nelson, R., Ståhl, G., Weydahl, D., 2011. Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: A case study from a boreal forest area. Remote Sensing of Environment 115, 3599–3614.

Persson, A., Holmgren, J., Söderman, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. Photogrammetric Engineering and Remote Sensing 68, 925–932.

Pfeifer, N., Briese, C., 2007. Geometrical aspects of airbone laser scanning and terrestrial laser scanning. International Archives of Photogrammetry and Remote Sensing XXXVI, 311 – 319.

Popescu, S., Wynne, R., 2004. Seeing the trees in the forest: Using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. Photogrammetric Engineering and Remote Sensing 70, 589–604.

Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne lidar. Biomass and Bioenergy 31, 646–655.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. Vienna, Austria.

Räty, M., Kankare, V., Yu, X., Holopainen, M., Vastaranta, M., Kantola, T., Hyyppä, J., Viitala, R., 2011. Tree biomass estimation using ALS features, in: Proceedings of Silvilaser 2011. Presented at the Silvilaser 2011, Hobart, Australia.

Sheather, S.J., 2009. A Modern Approach to Regression with R. Springer, New York. pp. 392.

Simonse, M., Aschoff, T., Spiecker, H., 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. Umeå, Sweden, pp. 251– 257.

Solberg, S., Astrup, R., Gobakken, T., Næsset, E., Weydahl, D.J., 2010. Estimating spruce and pine biomass with interferometric X-band SAR. Remote Sensing of Environment 114, 2353–2360.

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, 1369–1378.

Solodukhin, V.I., Mazhugin, I.N., Zhukov, A.Y., Narkevitch, V.I., Popov, Y.V., Kulyasov, A.G., Marasin, L.E., Sokolov, S.A., 1979. Lazernaya aeros emka profilei lesa (Laser aerial profiling of forest). Lesnoe Khozyaistvo (Forest Management) 10, 43–45.

Solodukhin, V.I., Zhukov, A.Y., Mazhugin, I.N., Bokova, T.K., Polezhay, V.M., 1977. Vozmozhnosti lazernoi aeros emki profilei lesa (Potentialities of Laser Aerial Survey of the Forest Profiles). Lesnoe Khozyaistvo (Forest Management) 10, 53–58.

Sorrell, S., Miller, R., Bentley, R., Speirs, J., 2010. Oil futures: A comparison of global supply forecasts. Energy Policy 38, 4990–5003.

Stephens, P.R., Kimberley, M.O., Beets, P.N., Paul, T.S.H., Searles, N., Bell, A., Brack, C., Broadley, J., 2012. Airborne scanning LiDAR in a double sampling forest carbon inventory. Remote Sensing of Environment 117, 348–357.

Straub, C., Koch, B., 2011a. Enhancement of bioenergy estimations within forests using airborne laser scanning and multispectral line scanner data. Biomass and Bioenergy 35, 3561–3574.

Straub, C., Koch, B., 2011b. Estimating Single Tree Stem Volume of Pinus sylvestris Using Airborne Laser Scanner and Multispectral Line Scanner Data. Remote Sensing 3, 929–944.

Särndal, C.-E., Swensson, B., Wretman, J., 2003. Model Assisted Survey Sampling. Springer. pp. 716.

Ter-Mikaelian, M.T., Korzukhin, M.D., 1997. Biomass equations for sixty-five North American tree species. Forest Ecology and Management 97, 1–24.

Tveite, B., 1977. Site index curves for Norway spruce (Picea abies (L.) Karst.). Reports of the Norwegian Forest Research Institute 33, 1–84.

Vauhkonen, J., Ene, L., Gupta, S., Heinzel, J., Holmgren, J., Pitkanen, J., Solberg, S., Wang, Y., Weinacker, H., Hauglin, K.M., Lien, V., Packalen, P., Gobakken, T., Koch, B., Naesset, E., Tokola, T., Maltamo, M., 2012. Comparative testing of single-tree detection algorithms under different types of forest. Forestry 85, 27–40.

Vauhkonen, J., Korpela, I., Maltamo, M., 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.

Vonderach, C., Voegtle, T., Adler, P., Norra, S., 2012. Terrestrial laser scanning for estimating urban tree volume and carbon content. International Journal of Remote Sensing 33, 6652–6667.

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, 3938–3951.

Wehr, A., Lohr, U., 1999. Airborne laser scanning - an introduction and overview. ISPRS Journal of Photogrammetry and Remote Sensing 54, 68–82.

Whittingham, M.J., Stephens, P.A., Bradbury, R.B., Freckleton, R.P., 2006. Why do we still use stepwise modelling in ecology and behaviour? Journal of Animal Ecology 75, 1182–1189.

Xu, Q.S., Liang, Y.Z., 2001. Monte Carlo cross validation. Chemometrics and Intelligent Laboratory Systems 56, 1–11.

Yao, T., Yang, X., Zhao, F., Wang, Z., Zhang, Q., Jupp, D., Lovell, J., Culvenor, D., Newnham, G., Ni-Meister, W., Schaaf, C., Woodcock, C., Wang, J., Li, X., Strahler, A., 2011. Measuring forest structure and biomass in New England forest stands using Echidna ground-based lidar. Remote Sensing of Environment 115, 2965–2974.

Zhao, K., Popescu, S., Meng, X., Pang, Y., Agca, M., 2011. Characterizing forest canopy structure with lidar composite metrics and machine learning. Remote Sensing of Environment 115, 1978–1996.

Zianis, D., Muukkonen, P., Mäkipää, R., Mencuccini, M., 2005. Biomass and Stem Volume Equations for Tree Species in Europe. Silva Fennica 63.

Paper I

BIOMASS AND BIOENERGY 36 (2012) 356–365

Available online at www.sciencedirect.com

SciVerse ScienceDirect

http://www.elsevier.com/locate/biombioe

Estimating potential logging residues in a boreal forest by airborne laser scanning

Marius Hauglin, Terje Gobakken, Vegard Lien, Ole Martin Bollandsås, Erik Næsset*

Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, 1432 Ås, Norway

ARTICLE INFO

Article history:
Received 25 February 2011
Received in revised form
28 October 2011
Accepted 4 November 2011
Available online 25 November 2011

Keywords:
Logging residues
Forest inventory
Airborne laser scanning
Forestry
Harvesting

ABSTRACT

Regression models relating variables derived from airborne laser scanning to the amount of biomass of potential logging residues (PLR) were estimated for 147 sample plots (200 m^2) measured in mature boreal forest. Logging residues were in this study defined as branches and tops of all trees. The base model explained 86% of the variation, and when the data were stratified into two strata according to site quality, the stratum-specific models accounted for 88% and 77% of the variation in PLR on poor sites and on good sites, respectively. Effect of tree species composition were assessed by including the proportion of Norway spruce as potential explanatory variable and this extended model explained 87% of the variability. The estimated models were validated using two datasets, one comprising 120 sample plots (200 m^2), and the other consisting of 25 measured stands. The ground observations in the stand dataset were based on data collected by a harvester. The validation of the overall model gave an RMSE value of 27.6% and 22.2% of the mean value in the plot and stand data, respectively. The stratum-specific models gave RMSE values of 25.3 and 22.5% in the plot validation, and 24.3 and 17.6% in the stand validation. Including proportion of spruce in the model gave an RMSE of 23.2% in the plot validation and 21.3% in the stand validation. The study shows that PLR in a boreal forest can be estimated by airborne laser scanning with accuracy comparable to those obtained when estimating e.g. stand volume.

1. Introduction

Emissions of CO_2 from combustions of fossil fuels are believed to be a major cause of the observed global climate change [1]. This has advanced the search for alternative sources of energy, including bioenergy from forests, and during the last ten years there has been growing interest toward the use of forest biomass for energy purposes. The use of logging residues as fuel, directly or through chemical refinement processes, is one example of such use. In most countries only the tree stems have been harvested and processed by sawmills and the pulp and paper industries, while branches and tree tops have been left in the forest. Several studies emphasize however forest residues as a potential exploitable resource [2–4], and the practice of not utilizing it may change in the near future.

Future changes in the energy market or in governmental policies may affect the price of energy from logging residues compared to energy from other sources. The European Commission reports a recent annual growth in the use of renewable energy in Europe that exceeds all other energy types, with bioenergy constituting a large part of the renewable energy [5]. In some countries, such as Finland, harvesting of logging residues is already operational [6]. This suggests that changes are taking place, which will give forest owners commercial incentives to harvest also logging residues.

* *Corresponding author*. Tel.: +47 64965389; fax: +47 64968890.
E-mail address: marius.hauglin@umb.no (M. Hauglin).
0961-9534/$ – see front matter
doi:10.1016/j.biombioe.2011.11.004

The logging residues from a final cutting might be defined in several ways. In the Nordic countries usually the branches and the tree tops is included, but also stumps and roots may be considered to be logging residues. In the present study the potential logging residues (PLR) are defined as the branches and tops of stems for all trees in a given plot or stand. Logging residues from partly removal of trees, i.e. thinning operations are not considered.

Forest owners, governmental institutions and other actors in the forest sector rely on forest inventories for planning and management of the forest resources on the estate, regional and national levels. Many present forest inventories, however, focus mainly on timber volume for potential utilization in the forest industries. An increase in commercial harvesting of biomass components, such as tops and branches, will spark the demand for estimates of those biomass components in forest inventories and resource management plans.

Airborne laser scanning (ALS) is a technique widely used to collect topographical data. The simple principle behind this method is to measure the time of travel of an emitted laser pulse from an airplane to the ground and back. Combined with precise positioning of the airplane, an accurate position can be obtained for the point where the laser pulse is reflected. ALS systems are typically capable of recording 100,000 or more measurements per second, enabling collection of data with high spatial density. Since ALS measurements also can be used to estimate heights of objects like trees, this technique is currently being used in data collection for forest inventories in many countries [7]. Previous research have shown that accurate estimates for a range of forest stand characteristics such as tree height, tree volume and basal area can be derived from ALS data when combined with field sample plots (e.g. [8,9]). In the Nordic countries the majority of forest inventory surveys is currently carried out using ALS data, and most widely used is the two stage procedure developed by [9]. In this procedure mean forest characteristics per unit-area are estimated. No single-tree estimates are produced using this procedure. The unit-area estimates can be summed to total values for each forest stand. Procedures where the characteristics of each individual tree are estimated (e.g. [10]) have also been proposed. These methods are currently not as widely used as the area-based method and they are, to our knowledge, more expensive.

Methods for estimating the dry weight biomass of trees by ALS have been proposed in several studies. The focus has been at a tree level [11], a stand or plot level [12,13] or on a regional level [14,15]. All these studies report total above ground biomass, and [11,12,13] also estimate biomass for individual tree components.

When a forest is scanned by ALS, a majority of the laser pulses will echo from the canopy, i.e. the branches. This implies that the ALS data contain useful information on the canopy biomass. In fact, most of the information inherent in the ALS data will be directly related to the tree canopy. This in turn suggests that canopy biomass and hence potential logging residues (PLR) could be estimated by ALS with at least as high accuracy as timber volume, which is typically estimated with mean random errors at the stand level of 8–14 percent of the mean values [16].

Previous studies indicate that crown shape has an effect on the estimates of other forest characteristics derived by ALS [17]. Crown shapes differ between tree species, so the proportion of different tree species for an area is likely to have an effect on the ALS estimated PLR also.

Based on the documented suitability of ALS to estimate forest characteristics such as timber volume and the growing interest in the use of forest residues for bioenergy purposes, we stated three objectives for this study. The first was to estimate PLR by ALS, and to validate the results on a large scale independent dataset. The proposed methods should utilize the same type of ALS and field data as those presently collected for many operational forest inventories. The second objective was to assess the effect of tree species distribution on the relationship between PLR and ALS data. Finally, since utilization of forest trees for bioenergy purposes also may include trees too small to be considered for timber production, information about the biomass from such trees might be useful in a planning situation. Thus, this study was also aimed to derive separate estimates of PLR for small trees.

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 (*Picea abies* (L.) Karst.) and Scots pine (*Pinus sylvestris* L.) as the dominant tree species.

2.2. Field data

Three sets of field data were used in this study. A dataset comprising sample plots was used as the modeling data. In order to assess the accuracy of the proposed method two independent datasets were used for validation: One dataset with measured stands and one with sample plots.

2.2.1. Forest inventory - stratification

An operational forest inventory was carried out in the study area, including a stratification of the productive forest. This stratification was utilized in collection of the field data used in the present study. The stratification was based on digital aerial photographs acquired in June 2005 with a Vexcel UltraCam D camera. Through digital stereo photogrammetry forest characteristics were interpreted manually by photo-interpretation of every stand, including stand borders, dominant tree species, site productivity and age class. The H_{40} site index system was used for site productivity, where the index value is the dominant height at breast height age 40 years [18,19]. In the present study stands with photo interpreted H_{40} index value less or equal to 14 were defined to have poor site quality, and stands with H_{40} index value higher than 14 were defined to have good site quality.

2.2.2. Modeling data

The modeling dataset was collected during fall 2006 and comprised a systematic sample of 147 circular sample plots of

size 200 m^2 selected in the mature productive forest of the study area. Within the plots, all trees with a diameter at breast height (d_{bh}) $\geq$ 4 cm were callipered. On each plot sample trees were selected with a relascope, giving sample trees selected with probability proportional to stem basal area. Heights (h) of the sample trees were measured with a hypsometer. There were on average nine sample trees per plot.

The plot center coordinates were determined by differential Global Navigation Satellite Systems (dGNSS), using dual-frequency receivers observing pseudo-range and carrier phase of the Global Positioning System (GPS) and Global Navigation Satellite System (GLONASS). Previous research indicate that a mean accuracy of 0.3–0.5 m or better can be obtained with this method, conditional on the actual observation time, canopy density and satellite constellation [20].

2.2.3. *Plot level validation data*

The plot level validation data were collected during fall 2007 and winter 2007–2008 and consisted of 31 circular plots of size 1000 m^2 and 500 m^2, laid out in the mature productive forest of the study area. These plots were initially used in other studies, and some constraints were posed regarding the location of the plots (see [21] for details). Within each of these plots all trees with $d_{bh} \geq 5$ cm were callipered, and tree coordinates relative to the plot center registered using a total station. 10–20 sample trees were selected on each plot according to procedures described in [21]. On the sample trees h was measured with a hypsometer.

In the current study only the 800 m^2 (400 m^2 for the smaller plots) inner circle of the plots was used. These circular plots were in turn divided into 200 m^2 subplots, to correspond in size with the plots in the modeling data. The 200 m^2 subplots were formed by dividing the smaller plots (400 m^2) in two halves and the larger plots (800 m^2) in four quadrants along the cardinal directions. This division resulted in 120 subplots with size 200 m^2.

The plot center coordinates were determined by dGNSS, using dual-frequency receivers observing pseudo-range and carrier phase of the GPS and GLONASS. Based on the accuracy reported in the post-processing the expected positional accuracy was in the range 0.01–0.35 m.

2.2.4. *Stand level validation data*

The dataset used for the stand level validation was collected during spring and fall 2007 in 25 stands with an average size of 1.3 ha. These stands were clear-cut as part of a commercial harvesting operation. The length and diameter for each 10 cm section of all trees passing through the harvester head were measured and stored by the harvester computer. Since the entire tree did not pass through the harvester head (the tree tops were ignored), length and cut diameter on a sample of tops from the harvested stands were measured manually in field and used to estimate the total height for the harvested trees (see [21] for details). The registered diameter at 1.1 m was used as d_{bh} for each stem assuming an average stump height of 20 cm.

The perimeter of each harvested stand was outlined in the field after the harvest by real time kinematic positioning using GPS and GLONASS. Based on observation of the field controller during the operation, the mean accuracy of the measured nodes was assessed to be approximately 0.5 m. In some stands a few individual trees were left standing inside the harvested stands (e.g. seed trees and dead trees). Of these trees, h and d_{bh} were measured in the field, and they were added to the list of harvester measured trees before the stand PLR, volume, basal area and mean height were calculated.

Forest characteristics of the three datasets are summarized in Table 1.

2.3. *ALS data*

The ALS data were collected in June 2005 with an Optech ALTM 3100 sensor on a Piper PA31-310 fixed-wing aircraft. The average flying altitude was 1850 m above ground. The pulse repetition frequency was 50 kHz and the scan frequency 71 Hz. The flight speed was 75 ms^{-1}. This resulted in an average point density on the ground of 0.7 m^{-2}. The scan angle was $\pm 15^\circ$ from nadir, but laser pulses outside $\pm 13^\circ$ from nadir were discarded. Classification of echoes into ground- and vegetation echoes was carried out by the contractor, Blom Geomatics using the TerraScan software [22]. 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)

Table 1 – Forest characteristics of the three field datasets[a].

Characteristic	Modeling data		Plot validation data		Stand validation data	
	Mean	SD	Mean	SD	Mean	SD
Poor sites						
Volume (m^3 ha^{-1})	215.3	103.9	163.9	71.6	337.8	77.6
Basal area (m^2 ha^{-1})	26.0	10.5	21.4	8.0	31.1	6.9
Lorey's mean height (m)	16.9	2.1	15.5	2.1	22.6	2.2
PLR_{dw} (Mg ha^{-1})	27.4	15.1	24.4	12.2	41.3	9.4
Good sites						
Volume (m^3 ha^{-1})	341.4	141.9	317.0	105.7	234.5	57.6
Basal area (m^2 ha^{-1})	33.7	11.5	34.4	10.3	23.9	5.2
Lorey's mean height (m)	20.4	3.2	19.3	2.6	20.7	3.0
PLR_{dw} (Mg ha^{-1})	44.0	17.0	47.7	16.4	26.1	10.6
All						
Volume (m^3 ha^{-1})	251.3	128.9	217.9	112.0	303.3	89.7
Basal area (m^2 ha^{-1})	28.2	11.3	26.0	10.8	29.0	7.7
Lorey's mean height (m)	17.9	3.0	16.8	2.9	21.7	2.7
PLR_{dw} (Mg ha^{-1})	32.1	17.4	32.6	17.7	37.7	12.1

a Lorey's mean height is the basal area weighted mean height, SD is standard deviation and PLR_{dw} is dry weight PLR.

terrain model. The height above ground was calculated for all echoes by subtracting the respective TIN heights from the ellipsoidal heights.

The sensor used in this project records up to four echoes for each laser pulse, but only the first and last returned echoes were used in the current study.

2.3.1. ALS derived metrics

From the ALS data several height and density related variables were derived. These variables were in all the three datasets calculated from laser echoes spatially attributed to area units of 200 m^2. In the stand validation dataset a grid of 200 m^2 quadratic cells was used.

Only laser echoes with a height above ground $\geq$2 m were considered to be canopy echoes. This was done to eliminate echoes from the undergrowth and objects like large rocks, in accordance with previous research [23]. For these canopy echoes height and density related metrics were derived. The height metrics comprise heights at the 10th, 20th,…, 90th percentiles of the echo height distribution. In addition we used the mean echo height and the corresponding coefficient of variation. Density related metrics were derived by dividing the canopy (from the canopy threshold at 2 m to the 95 height percentile) into 10 equal height bins. (The height of the 95 percentile was chosen over the maximum canopy height because less variability is associated with it [15].) The density variables were calculated as the ratio of the number of echoes above height bin 0,1,2,…,9 to the total number of echoes (including echoes below 2 m). Separate variables were derived from the first and last return ALS data.

2.4. Calculations and analysis

2.4.1. Diameter-height models

Since tree height is needed as an independent variable in the allometric functions used in this study, tree heights were estimated for all trees in the data material. This was done using diameter-height models developed from the measured d_{bh} and h of the sample trees. Non-linear least squares models of the form

$$h = \alpha \cdot d_{bh}^{B}$$

were used, where B and α are coefficients in the model. The sample trees were allocated to one of six strata, and then a separate model was fitted for each stratum. Three groups of tree species were used, namely Norway spruce, Scots pine and broadleaved trees. Each group was subdivided into poor or good site quality, forming the six strata. Separate diameter-height models were developed for the three datasets. Thus, a total of 18 different models were estimated. The nls() function in the statistical software R (version 2.10.1) was used.

2.4.2. Potential logging residues

In this study PLR was defined as the branches and the top of the stem, cut at diameter = 7 cm. This cut diameter is in accordance with practice in previous research, and enables the use of existing allometric equations to get an estimate of the biomass of the top [24].

The dry weight of the branches and the dry weight of the top were calculated separately for each individual tree. The dry weight of the branches of each tree was calculated by species-specific allometric equations [25] with d_{bh} and h as explanatory variables.

The dry weight of the top was estimated for each tree as a given portion of the dry weight of the whole stem following these three steps: (i) First the dry weight of the stem was estimated by species-specific allometric equations [25] with d_{bh} and h as explanatory variables; (ii) in order to calculate the portion of the stem represented by the top, available volume equations were used. Both volume of the stem and the top were calculated by species-specific allometric equations [24,26–28] with h and d_{bh} as explanatory variables; (iii) a top to stem ratio was then calculated for the volume, and this ratio

Table 2 – Summary of the regression models. Relationship between observed PLR and ALS derived variables. Parameter estimates, R^2 and condition number (κ).

Model	Response variable	Stratum[a]	n	Predictive model[b]	R^2	κ[c]
A	ln PLR_{dw}	All	147	$2.167-2.847\ln h_{60.f} + 3.912\ln h_{70.f} + 0.471\ln d_{7.f} + 0.764\ln d_{1.l} - 0.301\ln d_{9.l} - 0.465\ln h_{50.l}$	0.86	448
B	ln PLR_{dw}	All	147	$2.702 + 0.399 prop_{spruce} - 0.452\ln h_{20.f} + 0.854\ln h_{90.f} + 0.638\ln d_{6.f} + 0.412\ln d_{1.l}$	0.87	93
C1	ln PLR_{dw}	Poor sites[d]	105	$4.219 + 4.286\ln h_{mean.f} + 1.043\ln h_{cv.f} - 3.69\ln h_{60.f} + 0.812\ln d_{8.f} - 0.287\ln d_{9.f} - 0.362\ln h_{cv.l} + 1.582\ln d_{1.l} - 0.918\ln d_{5.l}$	0.88	389
C2	ln PLR_{dw}	Good sites[d]	42	$2.161 + 1.033\ln d_{3.l} + 0.779\ln h_{max.f}$	0.77	70
D	ln PLR_{dw} ($d_{bh} < 10$ cm)	All	147	$3.801-0.65\ln h_{20.f} + 1.178\ln d_{0.f} - 0.599\ln h_{10.l}$	0.52	37

a Stratified according to data from the forest inventory.

b h_{10}, h_{20},h_{50}, h_{60}, h_{70}, and h_{90} are the heights of the 10th, 20th, 50th, 60th, 70th and 90th height percentiles of the canopy laser echoes. d_0, d_1, d_3, d_5, d_6, d_7, d_8 and d_9 are the ratio of echoes above canopy height bin 0,1,3,5,6,7,8 and 9 to the total number of echoes. h_{mean}, h_{max} and h_{cv} are the mean, maximum and coefficient of variance for the heights of the canopy echoes. The .f and .l subscript ending indicates first or last return dataset. $prop_{spruce}$ is PLR_{dw} proportion of Norway spruce.

c Condition number, see text (Section 2.4.4) for details.

d Good site quality: H_{40} site index > 14, poor site quality: H_{40} site index <= 14. See text (Section 2.2.1) for details.

was used to calculate the dry weight of the top based on the total stem dry weight obtained in the first step.

For trees with d_{bh} smaller than the designated top cut diameter (<7 cm) the PLR was set equal to the total above ground biomass of the tree. Biomass equations for birch were used for all broadleaved trees. Volume of the top for broadleaved trees was estimated using an equation for Scots pine since no equations for those species are available for Norway.

The dry weight of the PLR at each plot, subplot or grid cell was calculated as the sum of PLR for all trees on the plot or in the grid cell.

2.4.3. *PLR regression models*

The PLR calculated from the field observations in the modeling data and the corresponding ALS derived metrics were used as variables in linear regression models. All the ALS derived variables were used as potential explanatory variables in models of the form

$$\begin{aligned}\ln(\mathrm{PLR_{dw}}) = {} & \beta_1\ln(h_{10.f}) + \beta_2\ln(h_{20.f}) + \ldots + \beta_9\ln(h_{90.f}) \\ & + \beta_{10}\ln(d_{0.f}) + \beta_{11}\ln(d_{1.f}) + \ldots + \beta_{19}\ln(d_{9.f}) \\ & + \beta_{20}\ln(h_{\mathrm{mean}.f}) + \beta_{21}\ln(h_{\mathrm{cv}.f}) + \beta_{22}\ln(h_{10.l}) \\ & + \beta_{23}\ln(h_{20.l}) + \ldots + \beta_{30}\ln(h_{90.l}) + \beta_{31}\ln(d_{0.l}) \\ & + \beta_{32}\ln(d_{1.l}) + \ldots + \beta_{40}\ln(d_{9.l}) + \beta_{41}\ln(h_{\mathrm{mean}.l}) \\ & + \beta_{42}\ln(h_{\mathrm{cv}.l}) + \varepsilon,\end{aligned}$$

where the response variable $\mathrm{PLR_{dw}}$ is dry weight of PLR. The variables h_{10}, h_{20}, …h_{90} are the height of the 10th, 20th, …90th percentile of the canopy laser echo height distribution and d_0, d_1, …d_9 is the ratio of the number of echoes above canopy height bin 0, 1, …9 to the total number of echoes, respectively. h_{mean} is the canopy echo mean height and h_{cv} is the corresponding coefficient of variation. The *.f* and *.l* subscript denote first and last return, respectively. *e* is an error term assumed to be normally distributed and independent with mean zero and constant variance.

The transformation of the variables to the natural logarithmic scale was done because this has been found suitable for similar types of models by others [9], and to ensure homogeneous variance. Since such a transformation of the variables to the logarithmic scale will result in biased predictions when they are back-transformed, the back transformation of the predicted values was carried out with the corrections described by [29].

2.4.4. *Variable selection*

The step() function in the R software (R version 2.10.1) was used for variable selection which uses the Bayesian Information Criterion (BIC) as selection criterion. Both forward and backward selection was enabled, with the full model as the initial model. In cases where the number of observations was too small for the full model to be used, an empty model was used as the initial model.

Multicollinearity was assessed by computing the condition number for the selected model. The condition number was calculated with the kappa() function in R, and corresponds to the ratio of the largest to the smallest eigenvalue. We regarded a condition number larger than 1000 to indicate high multicollinearity. If the condition number was found to be higher than 1000 for any of the models the stepwise procedure was run again, but this time with a larger penalty for model complexity. This was repeated until the number of explanatory variables was reduced so that the condition number was smaller than 1000.

2.4.5. *Model validation*

In order to assess the validity of the regression models, PLR predictions were compared to observed PLR in the two independent datasets, i.e., one dataset comprising 120 subplots of 200 m^2 and the another set consisting of the 25 stands measured by the harvester.

For the stand level validation the models were used to predict $\mathrm{PLR_{dw}}$ for each grid cell in the stand validation dataset. For grid cells with no canopy echoes, the $\mathrm{PLR_{dw}}$ value was set to 0. $\mathrm{PLR_{dw}}$ in each stand was calculated from the weighted mean values for the grid cells in each stand, with the cell areas as weights. Since the estimated value is assumed to depend on the cell size [30], cells with area <140 m^2 were discarded

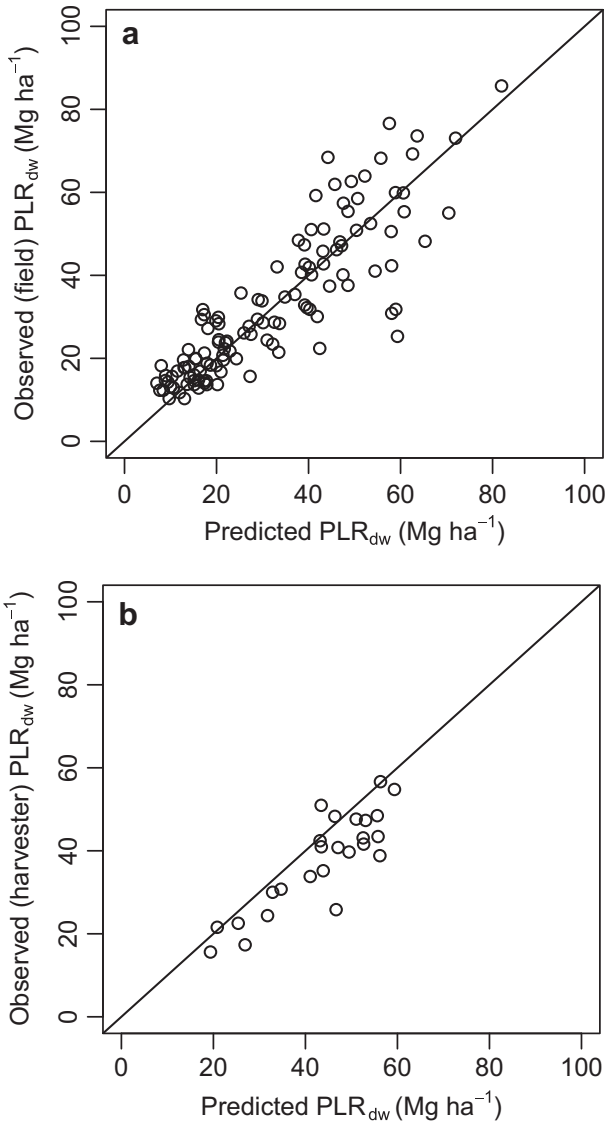

Fig. 1 – $\mathrm{PLR_{dw}}$ predicted by the base model (Table 2, model A) plotted against observed $\mathrm{PLR_{dw}}$ in the plot validation data (a) and observed $\mathrm{PLR_{dw}}$ in the stand validation data (b).

from the calculations. The chosen procedure is analog to that used in [9].

2.4.6. *Effects of tree species*

Crown shapes of Norway spruce and Scots pine differ. Thus, species proportion will most likely affect how the ALS echoes are distributed in the canopy. This will in turn have an influence on the derived ALS variables. To assess the effect of the proportion of the two main tree species on the relationship between the ALS derived variables and PLR, the proportion of spruce was included as a potential variable in the model. Proportion of Scots pine was then omitted because its correlation with the proportion of Norway spruce was close to one.

2.4.7. *Stratification*

Previous research has shown that good estimates of e.g. volume depend on stratified models (e.g. [9]). In the present study models were developed based on both unstratified and stratified data. Two strata were used: mature forest on poor sites and on good sites (see Section 2.2.1). Young forest was not considered in the present study since no final cutting is carried out in those areas. The stratification was based on the stand-wise information from the forest inventory in the study area.

2.4.8. *PLR of small trees*

The fraction of the PLR in a stand attributed to small trees would be useful information in a forest inventory. Both

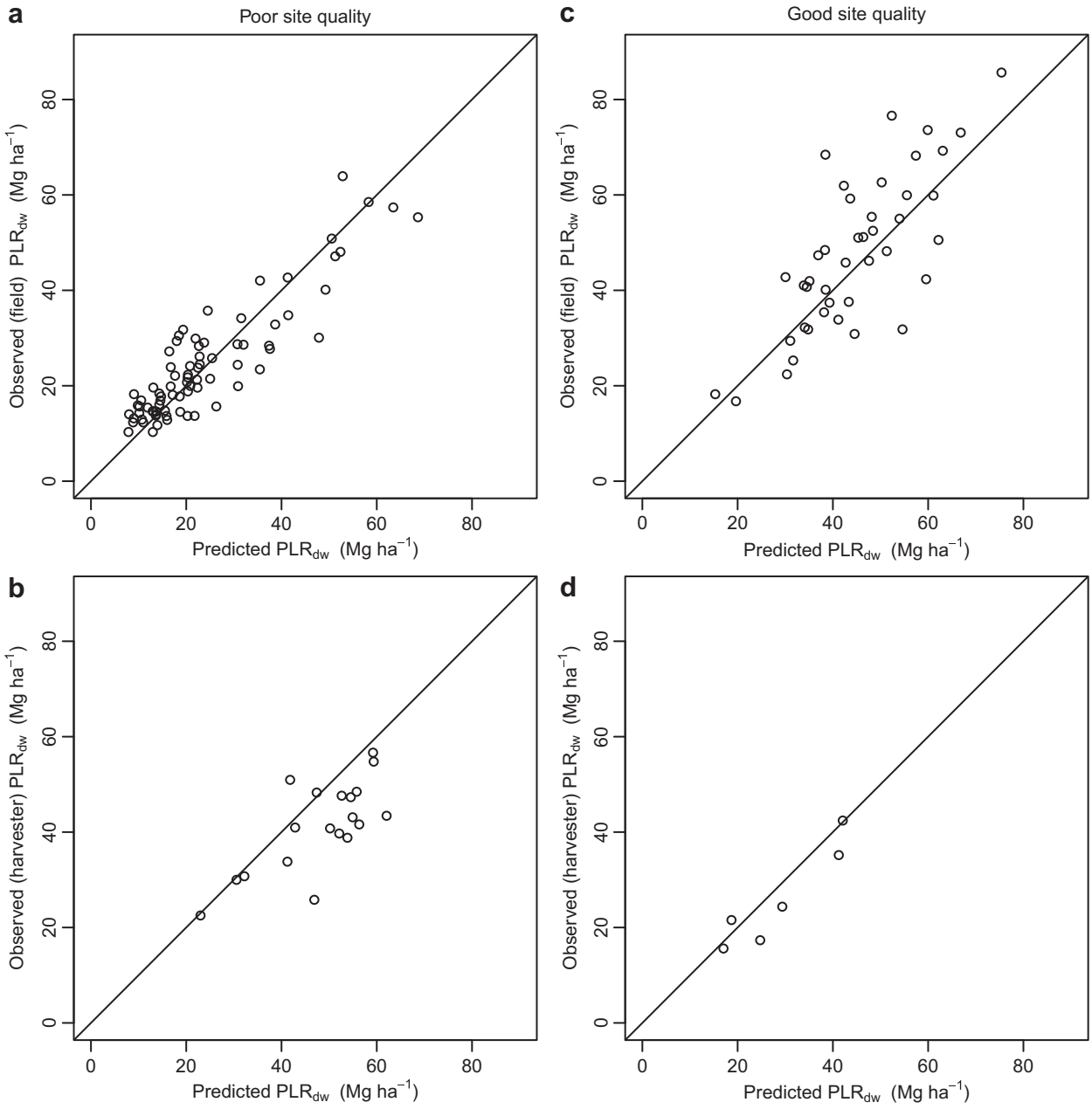

Fig. 2 – PLR$_{dw}$ predicted by the stratum-specific models (Table 2, model C1–C2) plotted against observed PLR$_{dw}$ in the plot validation data (a,c) and observed PLR in the stand validation data (b,d).

because this biomass potentially could be harvested, and because it, combined with an estimate of the total PLR, enables calculations of the amount of logging residues that would be available when harvesting the merchantable trees in the stand. A model with PLR_{dw} of trees with $d_{bh} < 10$ cm as response variable was therefore estimated. (The lower limit of d_{bh} of the trees considered was 4 and 5 cm in the modeling and the plot validation dataset, respectively.) Small trees were not registered in the stand validation dataset, and hence this particular model could not be validated using this dataset.

2.4.9. *Accuracy assessment*

The accuracy of the estimates was assessed by calculating the root mean square error (RMSE) of the model estimates. The RMSE was calculated as

$$\sqrt{\frac{\sum_{i=1}^{n}\left((\text{Observed PLR})_i - (\text{Predicted PLR})_i\right)^2}{n}}$$

where n is the number of plots or stands in a given dataset.

The differences between the predicted and observed PLR were tested for statistical significance by two-sided paired t-tests.

2.4.10. *Model labels*

The models were denoted by capital letters A – D as follows: A base model including all the modeling plots was denoted model A. The model with the proportion of spruce as an additional explanatory variable was denoted model B. Models with modeling data stratified according to site quality was denoted C1 and C2 for poor and good site quality, respectively. The model with PLR of small trees as response variable was denoted model D.

3. Results

3.1. *Model comparison – stratification*

First, all of the modeling sample plots were used to develop a base model (Table 2, model A). This overall model explained 86% of the variability in PLR_{dw}. The effect of tree species proportions was tested by including the proportion of Norway spruce as a potential explanatory variable in the model. Following the described procedure for variable selection, the proportion of Norway spruce was selected into the final model. This model explained 87% of the variability in PLR_{dw} (Table 2, model B).

Separate models were developed for two strata (Table 2, models C1 and C2). The variability explained by the stratum-specific models was 88% for poor sites and 77% for good sites. Compared to the base model these stratum-specific models explained more of the variability in PLR_{dw} on poor sites, but explained less of the variability on good sites.

In order to see if PLR of small and otherwise unmerchantable trees could be quantified separately, a model with PLR_{dw} of small trees ($d_{bh} < 10$ cm) as response variable was estimated. It turned out that 52% of the variability was accounted for by the model (Table 2, model D).

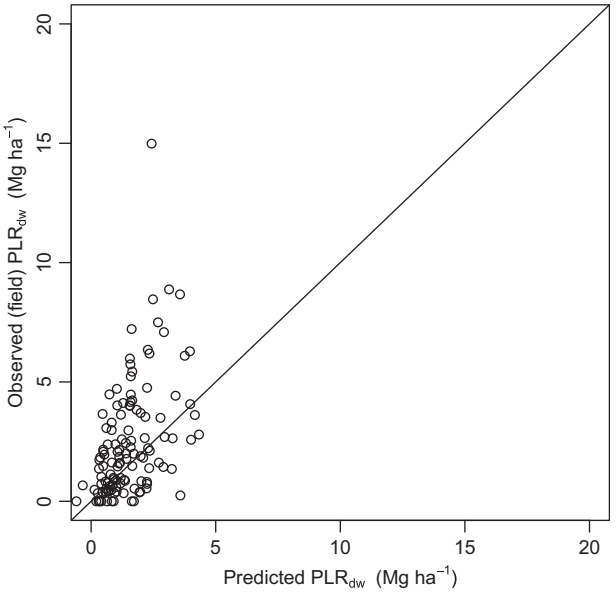

Fig. 3 – Predicted PLR_{dw} of small trees ($d_{bh} < 10$ cm) plotted against observed PLR_{dw} of small trees in the plot validation data (Table 2, model D).

3.2. *Model validation*

Model validations were carried out on two independent datasets. The models for total PLR_{dw} yielded good predictions in both the validation datasets (Figs. 1 and 2). Poorer predictions were observed for the model predicting PLR_{dw} of small trees only (Fig. 3).

3.2.1. *Plot level validation*

The plot level validation revealed an RMSE of 8.99 Mg ha^{-1} for the base model, which corresponds to 27.6% of the observed

Table 3 – Validation of the regression models. Observed mean PLR, mean difference between observed and predicted PLR, RMSE and RMSE as a percent of the observed mean PLR values.

Model[a]	n	Observed mean (Mg ha^{-1})	Mean difference (Mg ha^{-1})	(%)	RMSE (Mg ha^{-1})	(%)
Plot validation						
A	120	32.56	0.76 ns	2.3	8.99	27.6
B	120	32.56	1.14 ns	3.5	7.56	23.2
C1	78	24.43	0.43 ns	1.8	6.19	25.3
C2	42	47.67	2.98 ns	6.3	10.70	22.5
D	120	2.47	0.94***	37.7	2.31	93.7
Stand validation						
A	25	37.67	−5.88***	−15.6	8.35	22.2
B	25	37.66	−5.19***	−13.8	8.01	21.3
C1	19	41.33	−6.93***	−16.8	10.06	24.3
C2	6	26.08	−2.78 ns	−10.7	4.60	17.6

Significance levels: **$p < 0.01$, ***$p < 0.001$, ns: not significant ($p > 0.05$).
a Models described in Table 2.

mean PLR_{dw} (Table 3, model A). In the model where tree species proportion was included as an explanatory variable the corresponding RMSE value was 23.2% (Table 3, model B). The model for small trees had the largest error with an RMSE of 93.7% (Table 3, model D).

The mean difference between the field measured and the predicted PLR_{dw} in the plot validation data was 0.76 Mg ha^{-1} (2.3%) for the base model (Table 3, model A). The difference was not statistically significant. Also for the other models, no statistically significant differences could be observed, with the exception of the model predicting PLR_{dw} of small trees (Table 3, model D).

3.2.2. *Stand level validation*

When comparing observed and predicted PLR_{dw} in the stand validation data, the RMSE for the base model was 22.2% of the observed mean PLR_{dw} (Table 3, model A). For the model with tree species proportion included as an explanatory variable the corresponding RMSE was 21.3% (Table 3, model B). The two stratum-specific models had RMSE values of 24.3% for poor sites and 17.6% for good sites (Table 3, models C1–C3).

Mean difference between observed and predicted PLR_{dw} in the stand validation data was 5.88 Mg ha^{-1} (15.6%) for the base model (Table 3, model A). The difference was statistically significant, so the model overestimated the true PLR_{dw}. The model predicting PLR_{dw} of small trees (Table 2, model D) could not be validated at stand level due to lack of registrations of small trees in this dataset.

4. Discussion and conclusion

The reported accuracy of the estimates of PLR in this study was within the range reported by previous studies estimating forest stand characteristics by ALS (summarized in [16]) .The mean errors of the estimates in the present study were, however, in the high end of this range, and larger than the reported mean errors for estimates of stand volume in Ref. [16]. Ref. [13] reports an RMSE of top, foliage and branch biomass estimates on final cutting plots of 58.5%, 59.3% and 41.6%, respectively. The accuracy found in the present study is higher, and the reason is most likely the fact that [13] estimates biomass from an ALS derived diameter distribution, whereas the PLR is estimated directly from the ALS data in the present study. This issue is also discussed by [13], and they note that the former approach is more flexible in the sense that additional biomass equations can be applied without the need to fit a new regression model using the ALS data. Directly relating biomass to the ALS data is however assumed to yield a higher accuracy [13]. The results of the present study are in line with that assumption.

The stratum-specific models did not outperform the model based on the unstratified data. This is different from what is assumed to give best models for prediction of volume with similar methods [9] but in accordance with findings in predictions of above ground biomass [31]. Ref. [31] attributes this to the fact that the relationship between stem volume and crown properties differ for different forest types, and therefore requires separate models. Branch biomass, a large portion of PLR, is however more directly measured by ALS, and might be modeled accurately by one combined model. When examining the validation of the models, the stratum-specific models based on site quality overall performs slightly better than the combined model. This suggests that there might be gain in using stratum-specific models, but this gain is likely to be less pronounced than for models predicting volume.

The assessment of the effect of tree species shows that the proportion of the tree species does influence on the relationship between the ALS derived variables and PLR. The model where the proportion of one of the two dominant tree species was included as an explanatory variable performed better than the base model. This is in accordance with findings reported in [15]. In the stand level validation however, the model with the tree species proportion included performed only marginally better than the base model. One reason might be the fact that plot-specific species proportions were used in the fitting of the model, but predictions for the 200 m^2 cells were done with the stand level species proportions. Since the tree species composition in each cell may differ considerably from the overall stand species composition, this might be the cause of the poorer performance of this model in the stand level validation.

We were not successful in estimating a separate model with good predictions of PLR of small trees. There are however some measures that potentially could lead to improvements in the estimates of small tree PLR, such as changing the threshold dividing small and large trees to some other diameter value or include ALS variables from below the 2 m canopy threshold. The main obstacle in estimating characteristics of small trees is most likely the lack of information of those trees in the ALS data themselves. Since the trees are measured from an airborne sensor, small trees are prone to be shaded by larger trees, yielding few or no laser echoes from this part of the canopy. This is evident from previous studies as well, e.g. [32,33].

The cause of the overestimation of PLR in the stand validation data is not clear, but could possibly be attributed to errors in the harvester data, or the field registrations of the stands perimeter.

A direct comparison between the accuracy of PLR estimates and volume estimates in the present data material was not carried out. Both estimates rely on ground truth values that would be derived by the exact same underlying variables (tree heights and d_{bh}), hence limiting the usefulness of such a comparison. Some elements to consider regarding the possible differences in ALS estimated volume versus ALS estimated PLR will however be pointed out. Although most of the laser pulses are returned from the crown, there might be some aspects affecting the suitability of using ALS data to predict PLR. A dense upper canopy might prevent pulses from penetrating into the canopy, and give relatively few returned pulses from the lower part. This limits the information in the ALS data with respect to the total branch biomass, and compared to a situation where the pulses is more evenly spread out in the canopy, might result in lower accuracy in estimates of values related to canopy biomass, such as PLR. This might not be the case when estimating stem volume, which is more directly related to the tree heights.

In the procedure used in this study, as also described by [9], the estimation of stand characteristics by ALS is based on accurately positioned sample plots with ground observed values. Because measuring exact ground truth values for e.g. volume involves tedious (destructive) field work, one often derive ground truth values by using allometric equations, typically based on measured d_{bh} and tree height. Since true values are in fact not measured, errors in the ground truth data themselves are introduced. Random errors will occur due to individual trees deviating from the allometric relationship defined by the equation. Systematic errors may occur if tree allometry is systematically different from the allometry in the material used to derive the equations. This might for example happen if the equations are used in a new geographic area. A comparison of the reported random errors in two allometric equations for Norway spruce with h and d_{bh} as dependent variables for (i) stem volume [28] and (ii) dry weight of the branches [25] shows that the random errors are larger in relative terms for the estimate of the biomass of the branches compared to that of stem volume. The mean percentage error reported in [25] is 39% whereas it is 10% in [28]. As an example the corresponding mean percentage error on a 200 m^2 plot with 20 trees (the average in the material used in this study) would be 8.7% for biomass of the branches and 2.2% for stem volume. The difference in accuracy between the two equations suggests that canopy biomass varies more than stem volume for a given diameter and height, which sounds intuitively reasonable. The consequence of this is that ground truth values of PLR will be estimated with slightly lower accuracy than ground truth values for stem volume when using these allometric equations for Norway spruce. It is likely that the relationship between the allometric equations of volume and branch biomass exhibits the same tendency also for other species, such as Scots pine. These facts will in turn affect the ALS estimated PLR, and the random errors in the predictions. Due to the randomness of the errors, the effect on the aggregated predicted biomass at a stand level should be small. It is however interesting to note, that for a given area it might be so that the ALS data actually contain more information about the true branch biomass than the ground truth values obtained with allometric equations.

To conclude, the present study has shown that PLR in boreal forests can be estimated by ALS with accuracies in line with those obtained for other forest characteristics, such as volume and basal area, using operational procedures. The proposed procedure utilizes the same field and ALS data as are collected in many operational forest inventories, and can therefore easily be implemented with low or no additional costs.

Inclusion of information on the plot level tree species proportions improves the accuracy of the estimates. Adequate and separate estimates of PLR of small trees were not obtained in the present study.

The proposed method relies on allometric equations for ground truth values, and errors may be introduced due to this dependency. Further research should explore methods where ground truth values are more closely related to the true PLR, thus possibly utilizing more of the information of PLR inherent in the ALS data.

Acknowledgments

We wish to thank Prevista AS for giving us access to the field inventory data, Viken Skog BA for providing the harvester data and Blom Geomatics Norway for providing and processing the airborne laser data. Parts of the data used in the present study were funded by The Research Council of Norway, through the Forest Trust Fund and project #173911/I10: 'Assessing wood properties of forest resources'. We are also grateful to Mr. John Gunnar Dokk who took part in the field work. Finally we would like to thank for valuable comments and suggestions from the anonymous reviewers.

REFERENCES

[1] IPCC, Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Core Writing Team, Pachauri, R.K and Reisinger, A. (eds.) (2007).
[2] Mabee W, Saddler J. Bioethanol from lignocellulosics: Status and perspectives in Canada. Bioresour Technol 2010;101(13): 4806–13.
[3] Gan J, Smith C. Availability of logging residues and potential for electricity production and carbon displacement in the USA. Biomass Bioenergy 2006;30(12):1011–20.
[4] Malinen J, Pesonen M, Määttä T, Kajanus M. Potential harvest for wood fuels (energy wood) from logging residues and first thinnings in Southern Finland. Biomass Bioenergy 2001;20(3): 189–96.
[5] Eurostat. Europe in figures - eurostat yearbook 2010. Eurostat; 2010.
[6] Heinimö J, Alakangas E. Solid and liquid biofuels markets in Finland - a study in international biofuels trade, research report EN A-53. Lappeenranta University of Technology; 2006.
[7] Næsset E. 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 Environ 2009;113(1): 148–59.
[8] Holmgren J, Nilsson M, Olsson H. Estimation of tree height and stem volume on plots using airborne laser scanning. For Sci 2003;49(3):419–28.
[9] Næsset E. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sensing Environ 2002;80(1):88–99.
[10] Persson Å, Holmgren J, Söderman U. Detecting and measuring individual trees using an airborne laser scanner. Photogramm Eng Remote Sensing 2002;68(9):925–32.
[11] Popescu SC. Estimating biomass of individual pine trees using airborne lidar. Biomass Bioenergy 2007;31(9):646–55.
[12] Lim KS, Treitz PM. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. Scand J For Res 2004;19(6):558–70.
[13] Kotamaa E, Tokola T, Maltamo M, Packalén P, Kurttila M, Mäkinen A. Integration of remote sensing-based bioenergy inventory data and optimal bucking for stand-level decision making. Eur J For Res 2010;129:875–86.
[14] Boudreau J, Nelson RF, Margolis HA, Beaudoin A, Guindon L, Kimes DS. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. Remote Sensing Environ 2008;112(10):3876–90.

[15] Næsset E, Gobakken T. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. Remote Sensing Environ 2008;112(6): 3079–90.
[16] Næsset E. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. Scand J For Res 2007;22(5): 433–42.
[17] Nelson R. Modeling forest canopy heights: the effects of canopy shape. Remote Sensing Environ 1997;60(3):327–34.
[18] Braastad H. Growth model computer program for Pinus sylvestris. Rep Norwegian For Res Inst 1980;35:272–359 [In Norwegian with English summary].
[19] Tveite B. Site index curves for Norway spruce (Picea abies (L.) Karst.). Rep Norwegian For Res Inst 1977;33:1–84 [In Norwegian with English summary].
[20] Næsset E. Effects of differential single- and dual-frequency GPS and GLONASS observations on point accuracy under forest Canopies. Photogramm Eng Remote Sensing 2001; 67(9):1021–6.
[21] Maltamo M, Bollandsås OM, Vauhkonen J, Breidenbach J, Gobakken T, Næsset E. Comparing different methods for prediction of mean crown height in Norway spruce stands using airborne laser scanner data. Forestry 2010;83(3): 257–68.
[22] Terrasolid Ltd. TerraScan user's guide, tech. rep. Helsinki: Terrasolid Ltd; 2004.
[23] Næsset E. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. Scand J For Res 2004;19(2):164–79.
[24] Vestjordet E. Merchantable volume of Norway spruce and Scots pine based on relative height and the diameter at breast height or 2.5 m above stump level. Rep Norwegian For Res Inst 1968;25:411–59 [In Norwegian with English summary].
[25] Marklund LG. Biomassfuctions for pine, spruce and birch in Sweden, report 45. Umeå: Swedish University of Agricultural Sciences; 1988.
[26] Braastad H. Volume tables for Birch. Rep Norwegian For Res Inst 1966;21:265–365 [In Norwegian with English summary].
[27] Brantseg A. Volume functions and tables for Scots pine, South Norway. Rep Norwegian For Res Inst 1967;12:689–739 [In Norwegian with English summary].
[28] Vestjordet E. Functions and tables for standing trees. Norway spruce. Rep Norwegian For Res Inst 1967;22:539–74 [In Norwegian with English summary].
[29] Sprugel DG. Correcting for bias in log-transformed allometric equations. Ecology 1983;64(1):209–2010.
[30] Magnussen S, Boudewyn P. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. Can J For Res 1998;28(7):1016–31.
[31] Næsset E. Estimation of above- and below-ground biomass in boreal forest ecosystems. In: Thies M, Kock B, editors. Laser-scanners for forest and landscape assessment, International archives of photogrammetry, remote sensing and spatial information sciences. Freiburg, Germany: International Society of Photogrammetry and Remote Sensing; 2004. p. 145–8.
[32] Bollandsås OM, Næsset E. Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data. Scand J For Res 2007;22(1):33–47.
[33] Maltamo M, Eerikäinen K, Pitkänen J, Hyyppä J, Vehmas M. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sensing Environ 2004;90(3):319–30.

Paper II

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

Marius Hauglin[a]*, Janka Dibdiakova[b] , Terje Gobakken[a], Erik Næsset[a]

[a] *Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences.*
[b] *The Norwegian Forest and Landscape Institute.*

Abstract

The use of forest biomass for bioenergy purposes, directly or through refinement processes, has increased in the last decade. One example of such use is the utilization of logging residues. Branch biomass constitutes typically a considerable part of the logging residues, and should be quantified and included in future forest inventories. Airborne laser scanning (ALS) is widely used when collecting data for forest inventories, and procedures to estimate various forest characteristics at a single tree level already exists. Procedures for estimation of single-tree branch biomass of Norway spruce using features derived from ALS data are proposed in the present study. As field reference data the dry weight branch biomass of 50 trees were obtained through destructive sampling. Variables were further derived from the ALS echoes from each tree, including crown volume calculated from an interpolated crown surface constructed with a radial basis function. Spatial information derived from the pulse vectors were also incorporated when calculating the crown volume. Regression models with branch biomass as response variable were fit to the data, and the prediction accuracy assessed through a cross-validation procedure. Random forest regression models were compared to stepwise and simple linear least squares models. In the present study branch biomass was estimated with a higher accuracy by the best ALS-based models than by existing allometric biomass equations based on field measurements. An improved prediction accuracy was observed when incorporating information from the laser pulse vectors into the calculation of the crown volume variable, and a linear model with the crown volume as a single predictor gave the best overall results with a root mean square error of 35% in the validation.

Abbreviations: Dbh – diameter at breast height, RBF – radial basis function, ALS – airborne laser scanning

1 Introduction

Forest inventories are essential for optimal and sustainable management of forest resources. In the last ten years there has been an increased interest in the use of forest biomass for bioenergy purposes, directly or through refinement processes. Biomass from forests will most likely be one of several sources of energy that will have to replace fossil fuels in the future, and one example of the use of forest biomass for bioenergy purposes is the utilization of logging residues, i. e. biomass otherwise left in the forest during the logging. When logging residues become a commercial product from the forest, this resource should be quantified as part of the forest inventory to improve planning and management.

An increasing number of forest management inventories are based on data collected with airborne laser scanning (ALS) (Maltamo et al., 2011). While commercial and operational ALS-based forest inventories often are conducted according to the so-called area-based approach as described by Næsset (2002), also methods targeting single trees have been proposed (Hyyppä et al., 2001; Persson et al., 2002; 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 area-based information provided by the former. Although not as widely used as the area-based method at present, the use of single-tree forest inventories might in the future increase. This will depend on the ongoing technological and methodological research and development, and future costs for data acquisition. The potential of estimating individual tree characteristics by ALS

* Corresponding author: Marius Hauglin, Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences. Po. box 5003, 1432 Ås, Norway. Email adress: marius.hauglin@umb.no Tel.: +47 6496 5389.

has been investigated in several studies, including stem volume (Persson et al., 2002; Straub and Koch, 2011), stem diameter (Popescu, 2007), crown base height (Vauhkonen, 2010), leaf area index (Roberts et al., 2005) and biomass (Hauglin et al., 2011; Popescu, 2007; Räty et al., 2011).

When a tree is scanned by an airborne laser scanner, a majority of the laser pulses will be reflected from the crown, i.e. the branches. This suggests that the ALS data can contain useful information regarding 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. Logging residues are in this study defined as branches and tops, with branch biomass then constituting a considerable part. Since the biomass of the top will be highly dependent on the minimum cut diameter used in the harvest, estimates of branch biomass alone was in the present study chosen over biomass estimates of branches and top combined. Moreover, the biomass of the top is usually already accounted for in a forest inventory by being a part of the stem. Separate estimates of branch biomass can also be relevant for other applications such as canopy fuel assessments in wildfire behaviour modelling (e.g. Agca et al., 2011).

In a previous study on single-tree biomass estimation using ALS, a strong relationship between ALS data and branch biomass was reported (Popescu, 2007). The ground truth values in that study were obtained through allometric models, with field measured tree height (h) and diameter at breast height (d_{bh}) as explanatory variables. However, large errors are often associated with branch biomass estimates obtained at single-tree level with allometric models. As an example, one widely used allometric model for branch biomass of Norway spruce (*Picea abies* (L.) Karst.) in the Nordic countries reports a mean percentage error of 39% at the single tree level (Marklund, 1988). The strength of the method described by Popescu (2007) is that when d_{bh} and h are estimated, any suitable allometric model can be applied. Results from Hauglin et al. (2011) indicate however that a higher accuracy can be achieved by estimating branch biomass directly from the ALS data, rather than first estimating d_{bh} and h and then applying allometric equations.

In most studies the spatial information used in ALS-based forestry estimations is limited to the three-dimensional (x,y,z) coordinates of the laser echoes. The path of travel of the laser pulse, i.e. the vector from the aircraft to the echo, does also contain spatial information. This spatial information is however rarely considered. In the present study we utilized this pulse vector as additional information.

The first aim of the present study was to estimate single-tree branch biomass of Norway spruce by ALS-derived predictive models using accurately measured biomass obtained by destructive sampling . We also wanted to assess the predictive power and accuracy of the ALS-derived models. Promising results of using an ALS-derived crown volume as a predictor variable in branch biomass estimation has been showed in a previous pilot study (Hauglin et al., 2011). Thus, the second aim of the present study was to fully describe this approach and to see if predictions could be improved by incorporating information from the laser pulse vectors.

2 Materials and methods

2.1 Study area

The study area was in Aurskog-Høland municipality (59°50′N 11°30′E, 120-390 m a.s.l.) located in the south-eastern part of Norway. The total area of Aurskog-Høland is 960 km^2 with 670 km^2 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. Five locations were chosen, and from each location 10 trees were selected, resulting in a set of 50 trees to be destructively sampled. The five locations were chosen from potential sites in the intersections of the east-west-oriented strips of ALS data (see section 2.3) and forest roads, covering a range from poor to good site productivity in mature forest. Within each location sample trees of Norway spruce were selected according to the following procedure: First the diameter range of

the stand was determined by measuring the d_{bh} of the two smallest and two largest trees with a calliper. The diameter range was then divided into five classes, and two trees were selected in each class. The centerline of the ALS swath was located with a handheld GPS and from this and towards the north trees were successively selected until all the diameter classes were filled. If there were less than two trees in any diameter class when the edge of the ALS swath was reached, the procedure was repeated to the south from the centerline of the swath. In order to avoid edge effects from the forest road, trees with a distance >10 m to the forest road were preferred. Finally, due to the maximum operating distance of the truck-mounted mobile lift used, trees with a distance >30 m from the road were not selected.

On all the 50 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.

The 50 trees were then felled, and the wet 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 of 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: 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 (stem diameter < 5 cm). The diameter at the base of each of the sampled branches was recorded. From all the sampled branches there were taken three sub-samples at the inner, middle, and outer part of the branch, respectively. The sub-samples were dried and the wet and dry weight of each were recorded. The drying was done at 103±1 ˚C until constant mass. For each tree *h* was measured with a measuring tape after the felling. The characteristics of the 50 trees are summarized in Table 1.

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.

2.3 ALS data

ALS data were collected along two strips oriented in the east-west direction and located 9 km apart. The ALS 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 scan

Table 1. Summarized characteristics of the 50 trees in the data material. Maximum, minimum and mean of the field measured breast height diameter (d_{bh}), tree height (h) and dry weight of the branches (BR_{dw}) at each of the five sampling locations.

Location	d_{bh} (cm)	h (m)	BR_{dw} (kg)	n
1	9.7 – 30.4 (20.6)	10.9 – 21.2 (17.1)	8.9 – 128.6 (40.9)	*10*
2	12.3 – 35.1 (25.0)	8.2 – 26.3 (22.0)	10.5 – 152.3 (72.1)	*10*
3	11.2 – 39.8 (22.1)	12.3 – 23.2 (17.1)	11.2 – 137.0 (75.5)	*10*
4	10.0 – 40.3 (24.6)	8.7 – 26.9 (19.8)	11.6 – 163.2 (65.6)	*10*
5	16.6 – 27.8 (22.6)	15 – 22.4 (18.8)	30.9 – 115.2 (69.0)	*10*
all	9.7 – 40.3 (23.0)	8.2 – 26.9 (19.0)	8.9 – 163.2 (64.6)	*50*

angle was ±5 degrees from nadir. This gave an average point density on the ground of 7-10 m^{-2}. The beam divergence was 0.3 mrad. Up to four echoes were recorded for each pulse. The planimetric coordinates and the ellipsoidal height values were by the contractor determined for all echoes.

Classification of echoes into ground- and vegetation echoes was carried out by iteratively fitting a triangular irregular network (TIN) from below as described by Axelsson (2000). Points were in an iterative fashion added to the ground if they were within given threshold values. In the present study we used fixed threshold values: a maximum distance of 1.4 m and a maximum angle of 6 degrees to the TIN facet. Echoes classified as ground were used to construct a TIN surface. The height above ground was calculated for all echoes by subtracting the respective TIN heights from the ellipsoidal heights.

The corresponding position of the aircraft and the returned intensity were recorded for all echoes. The intensity values were normalized subject to the range from the aircraft as described by Ørka (2011).

2.3.1 Assigning laser echoes to individual trees

Several methods for automatic delineation of the ALS point cloud into single tree segments have been proposed (Persson et al., 2002; Koch et al., 2006; Solberg et al., 2006). Vauhkonen et al. (2012) found in a comparative study of segmentation algorithms a mean detection rate of around 70% and mean omission and commission errors of 49% and 23%, respectively. To avoid the errors introduced by automatic segmenting the laser echoes, we decided to assign each individual laser echo to the trees based on the field measured crown projections. An eight-sided polygon was formed from the crown projection measurements of each sample tree, and all echoes within the polygon were assigned to the targeted tree.

2.4 Calculations

2.4.1 Dry weight biomass of the branches

A wet to dry weight ratio was calculated for each branch sub-sample. For each tree a wet to dry weight ratio was calculated as the weighted mean of the ratios obtained from the samples. The diameters of the branches were used as weights. Finally, the total dry weight biomass of the branches for each tree, denoted as BR_{dw}, was calculated as the wet weight of the branches multiplied with the calculated tree-specific wet to dry weight ratio.

2.4.2 ALS-derived crown base height and crown echoes

Since the aim of this study was to estimate the biomass of the branches, the echoes from the crown of the trees had to be identified. The crown base height (CBH) was therefore estimated from the ALS data for each individual tree, and echoes (from all return categories) above the CBH were denoted crown echoes. Two methods for deciding the CBH were considered:

(1) Using echo height deciles as described by Solberg et al. (2006). The CBH is set at the height decile with the largest distance downwards to the next decile. The minimum CBH was set at 0.85 m above ground.

(2) Detecting the inflection point of a polynomial fit to echo height bins as described by Popescu and Zhao (2008). The Fast Fourier Transform smoothing was used. As proposed the polynomial was fit to the unsmoothed frequency profile when a numerical solution for the smoothed profile was unobtainable and the height of the 25th percentile used as the CBH when no numerical solution to the curve fitting procedure could be obtained.

2.4.3 ALS derived variables

In order to use the ALS data for prediction of branch biomass, variables were computed from the laser echoes allocated to each individual tree. Single-tree predictor variables derived from the ALS point cloud such as height percentiles and crown diameter have previously been used for prediction of e.g. tree height (Persson et al., 2002), d_{bh} (Popescu et al., 2003) and stem volume (Straub and Koch, 2011). An a priori assumption was that variables describing crown volume and density would be important in predicting of branch biomass. Also variables describing other characteristics could contribute, and in order to capture as much information as possible from the ALS data a large number of ALS derived variables were computed. The relative contribution from

each variable was then assessed in the next stage (see section 2.4.4).

We included variables computed from the heights and the height distribution of the laser echoes since they have been shown to correlate well with e.g. stem volume in previous studies (e.g. Straub and Koch, 2011). We further included variables derived from the echo intensity values, based on the fact that intensity originate not only as an effect of the *reflectance* of the target, but also by the size, or *silhouette area* of the target (Wotruba et al., 2005). This implies that density information could be captured by the intensity derived variables.

We also included variables computed based on echo categories and the spatial distribution of the echoes in the point cloud. The proportion of echoes in different categories is related to crown density, and the spatial distribution of echoes is related to both density and crown size and shape.

Since variables describing the crown volume was a main focus of the present study three crown volume variables were included and compared. The first was simply computed as the three-dimensional convex hull of the crown echoes using the *convhulln* function in the R package *geometry* (R version 2.14.0). The two other crown volume variables were computed as described in the next subsection. The complexity of the ALS derived variables used in the present study varies greatly, from simple statistical properties (e.g. mean echo height) to more extensive calculations (e. g. crown volume). The calculation of the crown volume variables are described in the next subsection, whereas the other ALS derived variables used in this study are defined in Table 2. An accurate and concise definition is given for each variable, and the reader is referred to Evans et al. (2009) for a further description of ALS derived features in general. Echoes from all return categories are used if not stated otherwise in the description. To account for non-linear relationship between BR_{dw} and the ALS derived variables, transformations of some selected variables were included (Table 2).

Radial basis function crown volume

We assumed that crown volume could be a potential important variable in estimating branch biomass. In order to create a crown surface that resembled the tree shape more closely than a three dimensional convex hull of the crown echoes, a crown surface was calculated for each tree by using a radial basis function (RBF) method. A crown volume was derived using this crown surface and information contained in the pulse vector of each echo.

Using an RBF method for calculating ALS derived crown volume was proposed by Kato et al. (2009). RBF methods are general purpose methods for functional approximations, especially suited for cases with scattered data points (Buhmann, 2009). In the case of the study by Kato et al. (2009) the function which was approximated was the (unknown) function describing the three-dimensional crown surface of a tree. Carr et al. (2001) describes more generally the methodological approach taken by Kato et al. (2009) and shows its application in reconstruction of three-dimensional surfaces from point observations. In the following we will outline the application of the RBF method in the present study, focusing on the parts that differ from e.g. Kato et al. (2009).

The crown surface was calculated from the crown echoes (from all return categories) allocated to each tree by using the RBF method as proposed by Kato et al. (2009). In order to preserve the true spatial relationship between the echoes the ellipsoidal heights were used, and not the above-ground heights. The method proposed by Kato et al. (2009) was modified to work with less dense ALS data, and the calculations followed these steps:

(1) Points to be on the surface of the crown were chosen by this procedure:

(i) The point cloud formed by the crown echoes was divided into *n* height bins, where *n* initially was set to 10. The height of each individual bin were set so that each bin contained (approximately) the same number of echoes.

(ii) If any of the bins contained less than three echoes, *n* was reduced so that each bin contained at least three echoes.

(iii) A two-dimensional convex hull was calculated in the horizontal plane for the echoes in each bin. The echoes at the border of the convex hull were marked as surface points, all others as being inside the crown (Fig. 1).

Table 2. Description of the ALS-derived variables. All echoes allocated to the specified tree and from all return categories are used if not stated otherwise. Transformations of selected variables. (Note: crown echoes are defined in the text, section 2.4.2).

Variable	Description
CV	Crown volume (Equation 3, described in section 2.4.3).
$crown_{ratio}$	Number of crown echoes to total number of echoes.
$first_{ratio}$	Number of first return echoes to total number of echoes.
$first_{ratio.crown}$	Number of first return crown echoes to total number of crown echoes.
cbh	Crown base height (see text, section 2.4.2).
$crown_{length}$	Crown length. The height from the CBH to the maximum echo height.
$D_{0.2}$	Density variables: Number of crown echoes above the height of the 0.2, 0.4, 0.6 and 0.8 fraction of the crown length, to the total number of crown echoes.
$D_{0.4}$	
$D_{0.6}$	
$D_{0.8}$	
$harea$	Convex hull area of the horizontal projection of the crown echoes.[a]
H_{20}	Heights of the 20, 40, 60, 80, and 95 percentile of the echo height distribution.
H_{40}	
H_{60}	
H_{80}	
H_{95}	
h_{max}	Maximum echo height.
h_{cv}	Coefficient of variation for echo heights.
h_{diff}	Distance between the mean height of first return crown echoes and second return crown echoes.
h_{mean}	Mean echo height.
$h_{mean.crown}$	Mean crown echo height.
i_{mean}	Mean intensity of the first return echoes.
$i_{mean.crown}$	Mean intensity of the first return crown echoes.
$i_{ratio.all}$	Fraction of crown echoes with above mean intensity values.
$i_{ratio.first}$	Fraction of first return crown echoes with above mean intensity values.
i_{sd}	Standard deviation of the echo intensity.
$kurtosis_{crown}$	Kurtosis of the crown echo height distribution.[b]
$pdensity$	Point density: Number of echoes divided by the horizontal convex hull area of the echoes.
$skew_{crown}$	Skew of the crown echo height distribution.[b]
$stemdist$	Mean horizontal distance to stem position (planimetric coordinates of the highest echo in the crown) for the crown echoes.
$stemdist_{cv}$	Coefficient of variation of horizontal distance from crown echoes to stem position.
$varea$	Mean of the two convex hull areas of the crown echoes in the x,z-plane and the y,z-plane, respectively.[a]
V_{chull}	Three dimensional convex hull volume of the crown echoes.[a]
V_{cs}	Volume of the crown surface (Equation 4, described in section 2.4.3).
transformations	
CV^2	Square transformation of the crown volume.
CV^3	Cubic transformation of the crown volume.
CV_{log}	Natural log transformation of the crown volume.
CV_{sqr}	Square root transformation of the crown volume.
$V_{chull.sqr}$	Square root transformation of the convex hull crown volume.[a]
$V_{cs.sqr}$	Square root transformation of the crown surface volume (Equation 4).

[a] The *convhulln* function in the R package *geometry* was used to compute the convex hull.

[b] Kurtosis and skew as defined by Evans et al. (2009).

(iv) the crown echoes with the largest and smallest height value were always marked as being on the crown surface.

(2) For each surface point two off-surface points were created in a given distance *d* from the surface point. The off-surface points were created on both sides of the surface point, on the horizontal line through the surface point towards the centroid of the echoes in each bin. Off-surface points for the topmost and lowermost crown echoes were created in a similar fashion but on a vertical line through the surface point. This resulted in three sets of points: Surface points, off-surface points *outside* the crown surface, and off-surface points *inside* the crown surface. Each point was given a value *f*, which was set to zero for the surface points, and *d* and *-d* for the outside and inside off-surface points, respectively:

$$f(x)\begin{cases} 0 & \text{if } x \text{ is a surface point} \\ d & \text{if } x \text{ is an off-surface point } \textit{outside} \text{ the crown surface} \\ -d & \text{if } x \text{ is an off-surface point } \textit{inside} \text{ the crown surface} \end{cases}$$

where d is the distance from the off-surface point to the corresponding surface point. In the present study the off-surface points were created with d= 0.3 m, and in the case of inconsistency between two off-surface points, one of them was removed.

(3) Following Carr et al. (2001) we calculated, through a linear system of equations, the weighting coefficient vector λ in an RBF approximation function of the form

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

where x is any point $\in \mathbb{R}^3$, N is the number of the surface and off-surface points and $\|\ \|$ is the Euclidian norm, i.e. the distance from x to point i.

(4) A crown surface was approximated by evaluating the function in Equation 1 for values of x in a three dimensional grid, and constructing a triangulated mesh surface where $f(x) = 0$.

(6) The volume inside the crown surface for each tree was calculated from the triangulated mesh by

$$V_{cs} = \sum_{i=1}^{M} V_s(t_i), \qquad (2)$$

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

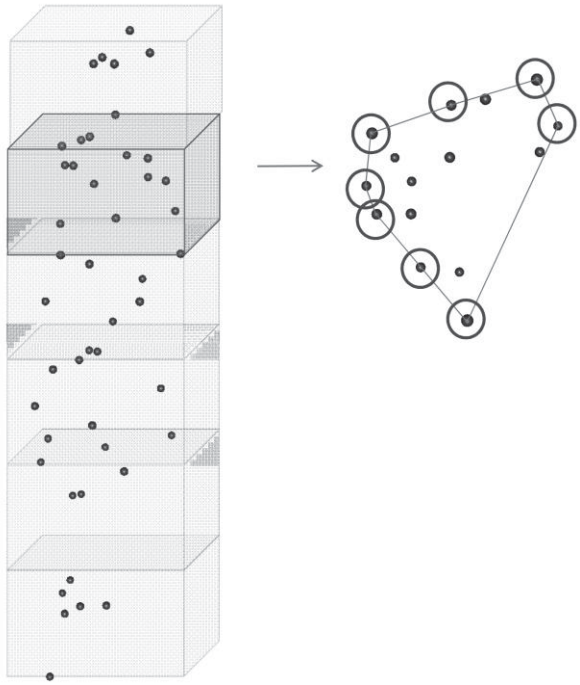

Fig. 1. Crown echoes from a single tree, and schematic representation of height bins (left). Top view of the echoes from one height bin (right). The 2D convex hull and the selected surface points are indicated by lines and circles, respectively.

Finally, the crown volume was calculated by including information from the vector of the laser pulses. By inspecting the path of travel of the pulses together with the crown surface described above, any intersections between the two can be determined. We assumed that whenever a pulse has travelled inside the crown surface without yielding an echo, its path describes a space without enough biological matter to trigger a return The minimum distance from an intermediate echo to a subsequent echo in the same pulse is a sensor-specific property (Ørka et al., 2010). Since no information from this part of the reflected signal can be known in a discrete return ALS system, this 'blind zone' was also regarded as a space without enough biological matter to trigger a return in the present study. This space is defined by the crown surface at the intersection(s), the pulse vector and the radius r of the laser footprint perpendicular to

the pulse vector (Fig 2). The radius *r* can be determined by the laser beam divergence and the length of the pulse vector, i.e., the range from the aircraft. The crown volume was then calculated for each tree as

$$CV = V_{cs} - [\sum_{i=1}^{P} V(pulse_i \cap CR)]k \,, \qquad (3)$$

where *P* is the number of pulses that intersects the crown, CR is the tree crown as given by the RBF crown surface, *V* is a volume function, and *k* is a factor, described later in this section. All nearby pulses where considered when *P* was determined, including pulses with echoes below the calculated CBH, and also pulses outside the extent of the trees crown projection, described in section 2.3.1.

The intersection of the crown surface and the pulses was determined by evaluating Equation 2 in points randomly scattered along the vector of each of the pulses that intersected the crown surface (Fig 3). The scattered points had a mean spacing *s*, and a distance to the vector of the given pulse ≤ *r*. The volume of the intersection between the pulse and the crown was then approximated for each pulse by

$$V(pulse \cap CR) = \left[\sum_{i=1}^{Q} \omega_i\right] \frac{1}{(\frac{1}{s})^3}, \qquad (4)$$

where *Q* is the number of scattered points with function values $f < 0$ (i.e. located *inside* the crown), denoted ω in Equation 4. In the present study we used $s = 0.1$ m, because we regarded this to yield a sufficient accuracy within feasible levels of computational burden. The above described procedure resembles in three dimensions the "dot count" method used to manually calculate areas in aerial photographs (e .g. Jensen, 2000, pp.176-178).

The volume of the intersections of the crown and the pulses were multiplied with a factor *k* to account for the fact that only parts of the crown surface was hit by the laser pulses. The area of the laser footprints hitting the crown surface projected onto a plane normal to the pulse vectors was computed (the vector of the pulse with the highest crown echo was used to determine the orientation of the plane). The factor *k* was defined as the ratio of this combined area of the laser footprints to the area of the crown surface projected onto the same plane (Fig. 3).

To be able to assess the effect of including information of the pulse vectors in the computation of the crown volume, the crown surface volume from Equation 2, denoted V_{cs}, was carried through as a separate variable to the regression models described in section 2.4.5.

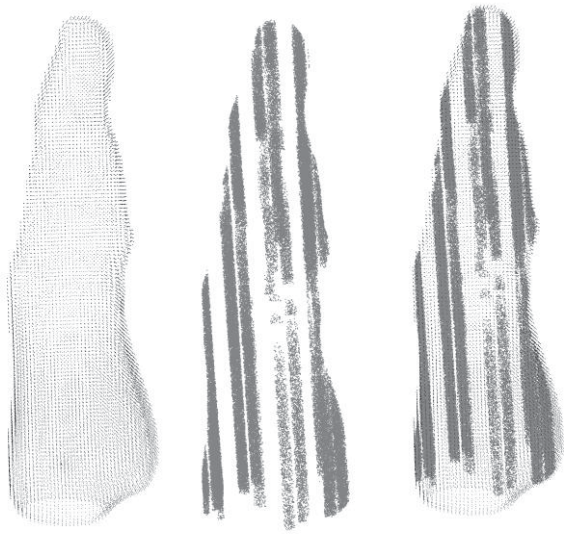

Fig. 2. Representation of the RBF crown surface (left), the intersection of the crown surface and the penetrating pulse vectors (middle) and a combination of the two (right), for one of the trees in the data material.

2.4.4 Variable comparison

In order to make good predictions of branch biomass from the ALS-data we wanted to find the most promising of the ALS-derived variables described in section 2.4.3 and Table 2. The relative importance of the ALS-derived predictor variables were therefore assessed by the following procedure:

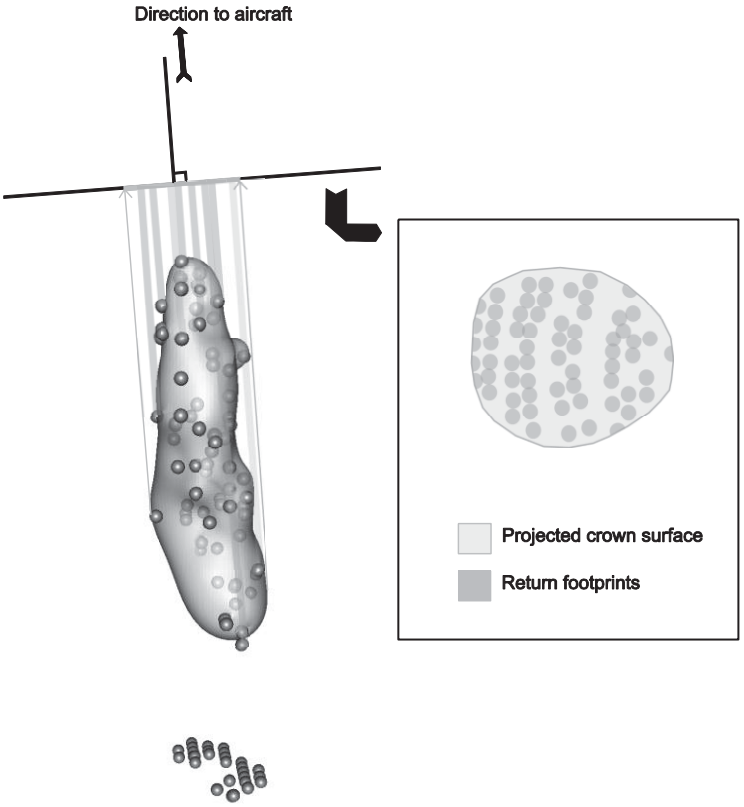

Fig. 3. The laser echoes and the RBF crown volume (left). Some of the pulse vectors are schematically indicated. The projection of the crown and the laser pulses on to a plane, normal to the pulse vectors (right).

(1) A linear regression model with stepwise variable selection was fit to the data. The *step* function in the statistical software R (version 2.14.0) was used. An empty model was set as the initial model, with both forward and backward selection enabled. The Bayesian information criterion was used as the initial selection criterion. To avoid strong correlation between predictor variables, the kappa value for the model was computed. We considered a kappa value exceeding 1000 to indicate strong multicollinearity, and in those cases the model was fit again, with a larger penalty for model complexity.

(2) The model-fitting described above was repeated 1 000 times with random samples of 25 trees from the 50 trees in the data material. The predictor variables were then ranked according to how many times they appeared in a final model. The rank of each variable was taken as an indicator of that variable's relative importance. The highest ranked predictor variables were subject to further investigations.

2.4.5 Predictive models

Linear least squares regression models are widely used in the remote sensing forestry literature when relating ALS-derived features to forest biophysical properties. Since the number of features derived from the remote sensing data can be large, some form of variable selection is often applied. The reduction of the number of variables is usually motivated by the need to get rid of highly correlated variables, and thus avoid multicollinearity and reduce the dimensionality of the data.

A stepwise variable selection – such as the *step* function in the R software – is typically applied when a high number of remote sensing derived variables are used in predictive regression models. Although it is widely used, this kind of procedure is in several statistical papers and textbooks advised

against (Babyak, 2004; Whittingham et al., 2006; Sheather, 2009). In the forest inventory research community there has in the last decade been applied several other regression techniques, in particluar non-parametric and machine learning methods (Næsset et al., 2005; Latifi et al., 2010; Vauhkonen et al., 2010; Zhao et al., 2011). In the present study, non-parametric random forest (RF) regression was chosen because it has been found to perform well compared to other regression techniques, and with little tuning required (Hastie et al., 2003). It is robust with respect to noise variables, which means that the variable selection deemed necessary in the least squares regression is avoided (Biau, 2012). Furthermore, other recent studies have reported successful use of RF for forest inventory applications with remote sensing data (Latifi et al., 2010; Vauhkonen et al., 2010). The principles behind random forest classification and regression as it is commonly used today were introduced by Breiman (2001). In this machine learning algorithm a large number of binary trees are grown from bootstrap samples of the data. The predicted value is then the average of the predicted values from the individual binary trees. Accurate predictions from random forest regression require that the training data covers the entire range of the population, and one feature of a random forest regression model is that it will not yield extrapolated values. This is to some extent a desirable property, since it ensures that only 'sound' values may be predicted. A consequence is that the predicted values at the extremes of the range are drawn towards the mean, thus introducing a bias in the predictions. In the present study the implementation of RF in the statistical software R (the *randomForest* package) was used. Further details about the principles and use of the RF algorithm can be found in Breiman (2001) and Hastie et al. (2003). In the present study a linear least squares model with a stepwise variable selection was also fit, as a reference.

An alternative to both stepwise variable selection and the use of alternative regression techniques is to create, or extract variables in such a way that no variable selection is necessary. This could be achieved by determining only one or a few variables a-priori and stick to these in the regression modeling. This approach is suggested by e.g. Babyak (2004)and can yield simple and interpretable models. In the present study simple linear regression models was fitted to single crown volume variables, and compared with models using the full set of ALS-derived variables.

In order to assess the predictive potential of the ALS derived variables, regression models were fit to the data with *BR* as response variable. Linear least squares, stepwise linear and RF regression were used and the models were repeatedly fit through a cross-validation procedure: First the data from three sampling locations (30 trees) were selected as training data and used to fit the regression models. Data from the remaining two locations (20 trees) were used for validation. This procedure was repeated once for all unique combinations of selecting three out of five locations, i.e. ten times. Model and validation statistics were then computed as the average over the ten runs.

The coefficient of determination (R^2) was inspected for an indication of the model fit,with R^2 values computed for the all the models as

$$\rho(fitted\ values, observed\ values)^2, \qquad (5)$$

where ρ is the Pearson's correlation.

The accuracy of the predictions was assessed by computing the root mean square error (RMSE) of the predicted values. The RMSE was computed as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Observed_i - Predicted_i)^2}{n}}, \qquad (6)$$

where *n* is the number of trees. In the present study the RMSE is reported in percent of the mean observed value.

2.4.6 Branch biomass estimated with existing allometric equations

The dry weight branch biomass of the 50 trees in the data material was also estimated with existing allometric equations, using field measured d_{bh} and *h* as explanatory variables (Marklund, 1988). The accuracy of the estimates derived with this allometric equation was assessed by computing the RMSE as described in 2.4.5. A paired t-test was carried out to check for bias in the estimates.

3 Results

3.1 Determining CBH from ALS data

Two methods of determining CBH from the ALS data were compared in the present study. The methods were compared by inspecting the goodness of fit and predictive performance of the regression models. Some differences between the methods were observed, and overall the models with ALS variables computed from crown echoes determined with the method described by Solberg et al. (2006) yielded best results (Table 3).

3.2 Variable comparison

Variables were derived from the ALS data as described in section 2.4.3 and Table 2. In order to find the most promising variables, they were compared and ranked subject to their performance in branch biomass models.

The ALS-derived variables described in Table 2 were ranked according to the number of occurrences for each variable in iteratively fit stepwise models as described in section 2.4.4 (variables derived with the CBH method of Solberg et al. (2006) was used). The crown volume variables outperformed the other variables and CV_{sqr} had the highest ranking and was selected as explanatory variable in 54% of the final models (Fig. 4). *CV* was ranked as number two and selected into 30% of the final models (Fig. 4). Among the other variables none were selected into more than 20% of the final models.

3.3 Branch biomass predictions

The dataset was repeatedly split in a training and validation dataset, as described in section 2.4.5. The RF and stepwise linear models were fit using all the ALS-derived predictor variables, and three models were fit with the crown volume variables CV_{sqr}, V_{cs} and V_{chull} as single predictor variables. Branch biomass was with all the models predicted in the validation dataset, and the predicted values compared to the field measured branch biomass. The overall best results were from models fit to data derived using the CBH method from Solberg et al. (2006), and only these results are given in the following. The reader is referred to Table 3 for the corresponding results using the other CBH method (described in section 2.4.2). In terms of model fit the stepwise model performed best, with a mean R^2 of 0.83. The results from the validation revealed that the linear model with CV_{sqr} as predictor variable had an average RMSE of 35% (Table 3). The models with V_{cs} and V_{chull} as single predictor variables yielded average RMSEs of 40% and 44%, respectively (Table 3). The gain in the prediction accuracy between the model with CV_{sqr} and V_{cs} as predictor variables corresponds to the effect of adding spatial information from the laser pulse vectors. The stepwise and RF models had average RMSEs in the validation of 45% and 38%, respectively (Table 3).

Table 3. Model and validation statistics from the cross-validation procedure. Mean R^2 from the model fitting and mean RMSE from the validation. Results with the two different methods used to derive CBH from the ALS data (see text, section 2.4.2).

		CBH: Solberg[a]		CBH: Popescu[a]	
Regression technique[b]	Explanatory variables	R^2	RMSE (%)	R^2	RMSE (%)
STEPWISE	All	0.83	45	0.76	51
RF	All	0.66	38	0.58	43
LS	CV_{sqr}	0.77	35	0.63	42
LS	V_{cs}	0.73	40	0.59	46
LS	V_{chull}	0.63	44	0.6	46

[a] Reference to CBH method. Solberg = Solberg et al. (2006), Popescu=Popescu and Zhao (2008).
[b] STEPWISE = Linear least squares stepwise regression, RF = random forest regression, LS = linear least squares regression.

3.4 Existing allometric equations

Allometric biomass equations by Marklund (1988) were used to predict branch biomass in the present data material, and the predicted values compared to the field measured branch biomass. The predicted branch biomass using these existing allometric equations and field measured d_{bh} and h corresponded to an RMSE of 36% in the present data material. No significant bias was observed.

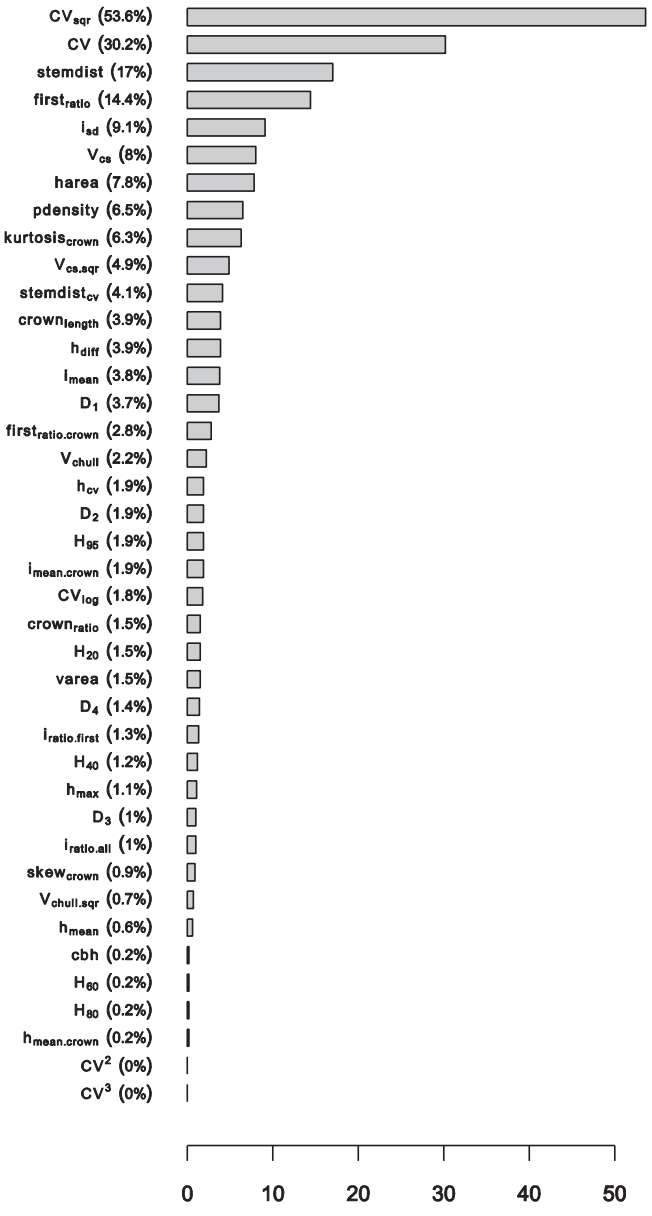

Fig. 4. Ranking of the ALS-derived variables based on iteratively fitting models to sub-samples of the data. The bar indicates the percentage of final models in which a variable was included.

4 Discussion and Conclusion

The results in the present study show that good estimates of single-tree branch biomass of Norway spruce can be obtained by estimating directly from the ALS data. The results are in line with the findings in Hauglin et al. (2011) regarding the RBF-derived crown volume as a good predictor for estimating branch biomass. Few previous studies report comparable ALS-estimates of branch biomass validated against accurate field measurements. Räty et al. (2011) report in their study the correlation between ALS-derived variables and biomass components in single trees of Norway spruce. The highest correlation for branch biomass (live branches) in their study was 0.71, corresponding to an R^2 in a single predictor linear model of 0.5. We found in the present study a higher accuracy for the branch biomass estimates, and we also found that using geometrically derived crown variables gave the best results. This is in contrast to Räty et al. who found that variables derived from the echo height metrics in each tree gave the best results for estimating branch biomass in spruce trees. For Scots pine however, Räty et al. found that crown area, a geometrically derived variable, had the highest correlation.

Some previous studies report ALS-based estimates of other single tree properties such as stem volume. Straub & Koch (2011) report an RMSE of 24% in a study considering stem volume of Scots pine. This is higher than the accuracy found in the present study, and might be attributed to the fact that stem volume is more closely related to the tree height than branch biomass. Tree height is known to be well estimated by ALS (Persson et al., 2002). Compared to Popescu (2007) the present study obtained a higher accuracy. Popescu (2007) reported an RMSE of 47% for estimates of total aboveground biomass of individual pine trees. The two studies are however not entirely comparable since no accurate field measurements of the biomass were available in the material used by Popescu.

When comparing the predictions from the ALS models with branch biomass estimated by existing allometric equations and actual field measured d_{bh} and h, the ALS based models yielded a slightly better accuracy. In other words, the ALS data contained in the present study more information related to branch biomass than the actual field measured d_{bh} and h. This suggests that the biomass of the branches might be more accurately estimated from the air than from field measurements of d_{bh} and h. If d_{bh} and h were to be estimated from the ALS data as by Popescu (2007), errors would be introduced. When these errors are taken into account we believe that the methods presented in the present study will yield a higher accuracy than using existing allometric equations with ALS-estimated d_{bh} and h, given that adequate training data is available. The actual gain in accuracy was however not assessed in this study, and could be subject to further research.

We have shown the use of geometric properties of the ALS echoes to be promising when estimating single-tree branch biomass, and in particular variables describing the crown volume. The inclusion of the spatial information contained in the laser pulse vectors gave an improvement, and we believe this improvement could be larger if the tree segments are automatically delineated. The potential errors in terms of inclusion of echoes from neighboring trees are in that case higher than with the procedure used in the present study. This could augment the importance of the correction derived from the pulses penetrating the RBF crown surface. We nevertheless believe the spatial information inherent in the pulse vectors should be utilized, especially when extracting features from ALS data at the singe tree level.

The crown volume derived from an RBF interpolated crown surface used in the present study is affected by the point density of the ALS data. A too small number of echoes allocated to a tree may even cause the proposed algorithm to fail. This was not observed in the present data material, but the properties of an RBF derived crown volume under different ALS point density conditions could be subject to further research. Furthermore, the transferability of the present approach to other tree-species and the effect of using other methods to determine CBH could be investigated further.

In the present study models fit with the machine learning regression technique of RF did not yield better prediction accuracy than the best linear

regression models, but the RF model did yield predictions with higher accuracy than the stepwise linear regression model, using the same set of variables. This suggests that RF might be a viable alternative to stepwise regression in single-tree variable estimation.

In the case of an implementation of the methods proposed in the present study practical means of acquiring training data must be considered. Destructive sampling is usually not an option due to the high costs, but the use of data obtained with terrestrial laser scanners, or artificially created training data might be viable alternatives. This could be subject to further research. A practical implementation of the presented method will also face the same challenges as other single-tree methods, namely those of segmentation errors. In the present study these were avoided by using field measured crown projections when constructing the single tree segments. Further research could investigate how these errors affect the biomass estimates derived by the proposed method.

In conclusion, the present study has revealed a strong relationship between ALS data and accurately measured branch biomass of Norway spruce at the single tree level. Crown volume derived from ALS data was shown to be a good predictor variable, and including information from the laser pulse vectors improved the results. More accurate branch biomass estimates were obtained from the ALS data than from field measurements of d_{bh} and h. The findings in the present study suggest that branch biomass can be satisfyingly estimated in an ALS-based single-tree forest inventory. The transferability of the findings to other tree species, forest conditions and ALS point densities should however be investigated further.

Acknowledgements

We wish to thank Blom Geomatics, Norway, for providing and partly processing the laser data. We are also grateful to Mr. Vegard Lien and Mr. Tord Storbækken (Norwegian University if Life Sciences); Mr. Arne Drømtorp, Mr. Leif Kjøstelsen and Mr. Espen Martinsen (The Norwegian Forest and Landscape Institute) who took part in the field work. Finally we would like to thank the anonymous reviewers for helpful and constructive comments and suggestions.

References

Agca, M., Popescu, S.C., Harper, C.W., 2011. Deriving forest canopy fuel parameters for loblolly pine forests in eastern Texas. Can. J. For. Res.-Rev. Can. Rech. For. 41, 1618–1625.

Axelsson, P., 2000. DEM generation form laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing XXXIII, 110-117.

Babyak, M.A., 2004. What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models. Psychosom Med 66, 411–421.

Biau, G., 2012. Analysis of a Random Forest Model. Journal of Machine Learning Research 13, 1063–1095.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32.

Buhmann, M.D., 2009. Radial Basis Functions: Theory and Implementations, 1st ed. Cambridge University Press. pp. 272.

Carr, J.C., Beatson, R.K., Cherrie, J.B., Mitchell, T.J., Fright, W.R., McCallum, B.C., Evans, T.R., 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.

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 1, 776–794.

Hasegawa, H., Yoshimura, T., 2003. Application of dual-frequency GPS receivers for static surveying under tree canopies. Journal of Forest Research 8, 103-110.

Hastie, T., Tibshirani, R., Friedman, J., 2003. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 1st ed. 2001. Corr. 3rd printing. ed. Springer. pp. 552.

Hauglin, M., Dibdiakova, J., Gobakken, T., Næsset, E., 2011. Estimating single-tree branch biomass of Norway spruce by airborne laser scanning, in: Proceedings of Silvilaser 2011. Presented at the Silvilaser 2011, Hobart, Australia.

Hyyppä, J., Kelle, O., Lehikoinen, 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.

Jensen, J.R., 2000. Remote Sensing of the Environment: An Earth Resource Perspective, 1st ed. Prentice Hall. pp. 544.

Kato, A., Moskal, L.M., Schiess, P., Swanson, M.E., Calhoun, D., Stuetzle, W., 2009. Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. Remote Sensing of Environment 113, 1148-1162.

Koch, B., Heyder, U., Wein, H., 2006. Detection of individual tree crowns in airborne lidar data. Photogrammetric Engineering and Remote Sensing 72, 357-363.

Latifi, H., Nothdurft, A., 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, 395–407.

Maltamo, M., Packalén, P., Kallio, E., Kangas, J., Uuttera, J., Heikkilä, J., 2011. Airborne laser scanning based stand level management inventory in Finland, in: Proceedings of Silvilaser 2011. Presented at the Silvilaser 2011, Hobart, Australia.

Marklund, L.G., 1988. Biomass functions for pine, spruce and birch in Sweden (Report No. 45). Swedish University of Agricultural Sciences, Umeå.

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., Bollandsås, O.M., Gobakken, T., 2005. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. Remote Sensing of Environment 94, 541 – 553.

Persson, A., Holmgren, J., Söderman, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. Photogrammetric Engineering and Remote Sensing 68, 925-932.

Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne lidar. Biomass and Bioenergy 31, 646-655.

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, 564-577.

Popescu, S.C., Zhao, K., 2008. A voxel-based lidar method for estimating crown base height for deciduous and pine trees. Remote Sensing of Environment 112, 767 - 781.

Roberts, S.D., Dean, T.J., Evans, D.L., McCombs, J.W., Harrington, R.L., Glass, P.A., 2005. Estimating individual tree leaf area in loblolly pine plantations using LiDAR-derived measurements of height and crown dimensions. Forest Ecology and Management 213, 54-70.

Räty, M., Kankare, V., Yu, X., Holopainen, M., Vastaranta, M., Kantola, T., Hyyppä, J., Viitala, R., 2011. Tree biomass estimation using ALS features, in: Proceedings of Silvilaser 2011. Presented at the Silvilaser 2011, Hobart, Australia.

Sheather, S.J., 2009. A Modern Approach to Regression with R. Springer, New York. pp. 392.

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, 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, 929-944.

Vauhkonen, J., 2010. Estimating crown base height for Scots pine by means of the 3D geometry of airborne laser scanning data. Int. J. of Remote Sensing 31, 1213-1226.

Vauhkonen, J., Ene, L., Gupta, S., Heinzel, J., Holmgren, J., Pitkanen, J., Solberg, S., Wang, Y., Weinacker, H., Hauglin, K.M., Lien, V., Packalen, P., Gobakken, T., Koch, B., Naesset, E., Tokola, T., Maltamo, M., 2012. Comparative testing of single-tree detection algorithms under different types of forest. Forestry 85, 27–40.

Vauhkonen, J., Korpela, I., Maltamo, M., 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.

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, 3938-3951.

Whittingham, M.J., Stephens, P.A., Bradbury, R.B., Freckleton, R.P., 2006. Why do we still use stepwise modelling in ecology and behaviour? Journal of Animal Ecology 75, 1182–1189.

Wotruba, L., Morsdorf, F., Meier, E., Nüesch, D., 2005. Assessment of sensor characteristics of an airborne laser scanner usinggeometric reference targets, in: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences.

Zhao, K., Popescu, S., Meng, X., Pang, Y., Agca, M., 2011. Characterizing forest canopy structure with lidar composite metrics and machine learning. Remote Sensing of Environment 115, 1978-1996.

Ørka, H.O., 2011. Improving forest inventory and monitoring by combining remotely sensed three-dimensional and spectral information (Ph.D. Thesis). Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences. Thesis 2011:22.

Ørka, H.O., Næsset, E., Bollandsås, O.M., 2010. Effects of different sensors and leaf-on and leaf-off canopy conditions on echo distributions and individual tree properties derived from airborne laser scanning. Remote Sensing of Environment 114, 1445–1461.

Paper III

Estimating single-tree branch biomass of Norway spruce with terrestrial laser scanning

Marius Hauglin[a]*, Rasmus Astrup[b], Terje Gobakken[a], Erik Næsset[a]

[a] *Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences.*
[b] *The Norwegian Forest and Landscape Institute.*

Abstract

Many remote sensing-based methods for estimation of forest biomass rely on allometric biomass models for field reference data. Terrestrial laser scanning (TLS) has emerged as a tool for detailed data collections in forestry applications, and methods have been proposed to derive e.g. tree-position, diameter at breast height and stem volume from TLS data. In the present study TLS derived features were related to destructively sampled branch biomass of Norway spruce at the single-tree level, and the results compared to conventional allometric models with field measured diameter and height. TLS features were derived following two approaches: one using manual crown measurements in the unified TLS point cloud and another more comprehensive voxel-based approach. Features were derived with voxels of size 0.1, 0.2 and 0.4 m, and the effect of voxel size was assessed. TLS-derived variables were used in regression models, and prediction accuracies assessed through a Monte Carlo cross-validation procedure. The model based on 0.4 m voxel data yielded the best prediction accuracy, with a root mean square error (RMSE) of 32%. The accuracy was found to decrease with an increase in voxel size, i.e. the model based on the 0.1 m voxels yielded the lowest accuracy. The model based on crown measurements had an RMSE of 34%. The accuracies of the predictions from the TLS-based models were found to be higher than from conventional allometric models, but the improvement was relatively small.

Abbreviations: Dbh – diameter at breast height, TLS – terrestrial laser scanning, RF – random forest

1 Introduction

The recent focus on forest carbon storage and the utilization of biomass for energy purposes has increased the need for mapping and monitoring of forest biomass. Several studies have described methods for estimation of forest biomass using remote sensing techniques, such as airborne laser scanning (Popescu, 2007; Boudreau et al., 2008; Næsset and Gobakken, 2008; Fuchs et al., 2009; Hauglin et al., 2011; Næsset et al., 2011; Gobakken et al., 2012). The spatial levels in these studies range from the regional level to estimates of individual tree biomass. All these studies rely on allometric biomass models for calculations of field reference data. The allometric models used typically estimate tree biomass from field measured diameter at breast height (d_{bh}), sometimes with the addition of tree height (h).

Zianis et al. (2005) and Jenkins et al. (2003) have compiled and reviewed allometric biomass models for Europe and the United States, respectively. Zianis et al. (2005) found that different allometric models developed for the same species gave different results. For Norway spruce (*Picea abies* (L.) Karst.) estimates of foliage biomass varied strongly between the various biomass models. Jenkins et al. (2003) emphasised the variability in biomass from one site to another and stated that it is difficult to estimate tree biomass accurately, even with a site-specific allometric model. In the Nordic countries the models developed by Marklund (1988) are widely used for biomass estimation of Norway spruce. The error associated

* Corresponding author: Marius Hauglin, Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences. Po. box 5003, 1432 Ås, Norway. Email adress: marius.hauglin@umb.no Tel.: +47 6496 5389.

with the single-tree predictions of spruce branch biomass is reported by Marklund (1988) to correspond to 39% of the mean value. Errors of aggregated predictions will be deflated, but bias will be introduced in the predictions if the allometry of the trees in question differs from that of the material used to develop the models. Findings reported by Hauglin et al. (2011) indicates that single-tree level airborne laser scanner data can actually contain more information regarding the branch biomass than the field measured d_{bh} and h. This suggests that remote sensing can be a viable addition to conventional allometric models. Terrestrial laser scanning (TLS) has in the last decade emerged as a tool for detailed data collection in forestry applications. The distance from an instrument to a target is inferred from the time of travel of emitted laser light when it is backscattered. The time of travel is derived either through direct timing of discrete laser pulses, or inferred from the phase-shift of the reflected light. Since the scanner is stationary, a fixed angle movement between the range measurements enables a spatially systematic (gridded) structure of the data. TLS scans from different positions in e.g. a field plot or a forest stand can be merged together by means of a registration process, usually involving artificial targets placed at the scanning location.

Several studies have described the estimation of single-tree characteristics such as d_{bh} and stem volume from TLS data (Simonse et al., 2003; Bienert et al., 2007; Lovell et al., 2011; Yao et al., 2011; Moskal and Zheng, 2012). The level of automatization and accuracy differ in these studies but they all aim at deriving tree positions and d_{bh} automatically. A full or partial reconstruction of the individual tree from TLS data has also been investigated (Gorte and Pfeifer, 2004; Pfeifer et al., 2004; Lefsky and McHale, 2008; Côté et al., 2011). Some studies present procedures to estimate total above-ground biomass from TLS data (Holopainen et al., 2011; Yao et al., 2011). Yao et al. (2011) estimated biomass by deriving individual tree diameters from the TLS data and then applied allometric models. Holopainen et al. (2011) manually derived individual crown diameters and other features from the TLS point cloud of each tree and related that to accurately measured biomass (destructive sampling). A strong correlation was found for Norway spruce (*Picea abies* (L.) Karst.) between the total above-ground biomass and TLS derived features. Holopainen et al. (2011) did not distinguish between individual biomass components, and the data material was in that study limited to 11 trees of Norway spruce.

One challenge when estimating tree characteristics from TLS data is the occurrence of shadowed areas caused by occlusion (Moskal and Zheng, 2012). A shadowed area is a three-dimensional space in which no laser beams have travelled. No information of the content of biological matter in this shadowed area can be directly derived from the TLS data. In a TLS system recording single echoes, a shadowed area will occur beyond the range of each recorded echo, along the line-of-sight of the scanner. Also multiple echo and full-waveform systems will adhere to this, although they will to some extent have the ability to penetrate through scattered matter. Solid objects like stems and dense branches or foliage will in any case cause completely shadowed areas. The presence of shadowed areas can to some degree be mitigated by combining scans from different positions, scanning the object or area of interest from different angles (Van der Zande et al., 2008). Shadowed areas will however still be present in a tree crown, caused both by neighbouring trees and also shadows caused by other parts of the crown itself. The upper part of the crown is usually more prone to shadowing due to a typical upward scanning direction. Due to the fact that the scanning is done from the ground, this part of the crown is also where the point density of the laser echoes is lowest. It should also be noted that the amount of shadowing will depend on the technical specification of the scanner in question (Ducey et al., 2012).

Several previous studies use a unified point cloud from multiple TLS scans as the data source for analysis. In such point cloud data there is no way to distinguish between a shadowed area, and a truly empty area. It follows from the description of a shadowed area above that the absence of laser echoes within a given space does not mean that this space is empty, if it lies within a shadowed area. Truly empty spaces, and shadowed areas which would appear similar in such a unified point cloud should ideally be treated differently in the analysis. The very nature of how many TLS scans

are acquired can however facilitate means to distinguish between the two. Since the scanner is stationary the path of the laser beam for every measurement is known, including measurements in directions without any echo or with echoes from beyond the range of the scanner. This additional information can be utilized when analyzing the TLS data, enabling the distinction between truly empty spaces and seemingly empty spaces caused by shadowing. When this beam vector information is taken into consideration, a laser scanner does not only give information of where biological matter exists, but also areas without enough matter to trigger an echo. Some previous studies have utilized this beam vector information. Danson et al. (2007) estimated gap fraction with a TLS system where only the echoes were recorded. In that study a 'scanner model' was used to estimate the direction of all emitted laser beams, and by comparing with the recorded echoes the direction of measurements with no echoes were inferred. Moorthy et al. (2008) applied a similar procedure when estimating leaf area index from TLS data of an artificial tree in a controlled experiment. Hosoi and Omasa (2006) constructed a high resolution voxel space from the TLS data and assigned values to the voxels conditional on their interaction with the laser beams. Voxels corresponding to the shadowed areas described above were marked. Bienert et al. (2010) used ray tracing from the actual echoes to infer the interaction between the laser beams and voxels, and classified the voxels in accordance with the state of this interaction.

Ground reference data are required in most forest inventory procedures using ALS (e.g. Hyyppä et al., 2001; Næsset, 2004). New and more accurate ground-based methods for single-tree biomass estimation could improve single-tree inventory procedures; and also be beneficial in other applications, such as crown fuel assessments and validation of existing allometric models. To our knowledge no previous study has related destructively sampled branch biomass to TLS data at the single-tree level. The aim of the present study was therefore to estimate branch biomass of Norway spruce by TLS using accurately measured reference data obtained with destructive sampling. We further wanted to assess if utilization of the information derived from the laser beam vectors could mitigate some of the effects of shadows in the TLS data, and thereby lead to improved branch biomass predictions.

2 Materials and Methods

2.1 Study area

The study area is in Aurskog-Høland municipality (59°50'N 11°30'E, 120-390 m a.s.l.) located in the south-eastern part of Norway. The total area of Aurskog-Høland is 960 km^2 with 670 km^2 productive forest. The forest type is boreal with Norway spruce and Scots pine (*Pinus sylvestris* L.) as the dominant tree species.

2.1.1 Selection of sample trees

Four sampling locations were chosen, and from each location sample trees were selected. The four locations were chosen from suitable locations in the proximity of forest roads and covered mature forest sites with both poor and good site productivity. Within each location 10 trees (9 trees in one of the locations) of Norway spruce were selected. First, the diameter range of the stand was determined by measuring the d_{bh} of the two smallest and two largest trees with a calliper. The diameter range was then divided into five classes, and two trees were selected in each class. In order to avoid edge effects from the forest road, trees with a distance >10 m to a forest road were preferred. Finally, due to practical reasons, trees with a distance >30 m from a road were not selected. The characteristics of the 39 trees are summarized in Table 1.

Table 1. Summarized characteristics of the 39 trees in the data material. Field measured breast height diameter (d_{bh}), tree height (h) and dry weight of the branches (BR).

	d_{bh} (cm)	h (m)	BR (kg)
min	9.7	7.9	8.9
max	39.8	26.6	152.3
mean	22.3	18.3	62.9

2.2 TLS data

TLS data were collected in May and June 2009 with a Leica HDS6000 phase-based scanner. The scanning was done with a horizontal and vertical angle increment of the laser measurements of 0.036 degrees. This corresponds to a point spacing of 15.9 mm at a 25 m distance from the scanner. The scanner recorded up to one echo and had a maximum measurement range of 79 m. A full 360 x 310 degree scan (the scanner's maximum field-of-view) was performed from each scanner position, i.e., both hemispheres excluding the ground directly beneath the scanner.

The scan positions were subjectively chosen so that each tree was scanned from at least two positions. Also, the scanner position was chosen so that the top of each of the sample trees preferably was visible in at least one scan. Scan targets were used in order to register (merge) the different scans correctly. The scans from each site were registered using the Leica Cyclone software (version 7.3.3). The average mean absolute error from the registration process reported by the software was 3.5 mm. For the subsequent analysis the data from each individual scan were kept separate but in the common coordinate system obtained through the registration process.

The relative position of each sample tree was recorded with a survey grade total station. A summary of the distances from the sample trees to the scanner positions is given in Table 2.

2.3 Destructive sampling of branch biomass

The 39 sample trees were felled, and the wet 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 of 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: (1) three branches from the lower part of the crown, (2) three from the middle part, and (3) three from the upper part of the crown. In addition, (4) two branches were taken from the top of the stem (stem diameter < 5 cm). The diameter at the base of each of the sampled branches was recorded. From all the sampled branches there were taken three sub-samples at the inner, middle, and outer part of the branch, respectively. The sub-samples were dried and the wet and dry weights of each were recorded. The drying was done at 103±1 ˚C until constant mass. For each tree d_{bh} and h were measured with a calliper and a Vertex III hypsometer, before the felling.

2.4 Calculations

2.4.1 Dry weight biomass of the branches

A wet to dry weight ratio was calculated for each branch sub-sample. For each tree a wet to dry weight ratio was calculated as the weighted mean of the ratios obtained from the samples. The diameters of the branches were used as weights. Finally, the total dry weight biomass of the branches (including needles) for each tree, denoted BR was calculated as the wet weight of the branches multiplied with the calculated tree-specific wet to dry weight ratio.

Table 2. Summary of the distances (m) from each individual sample tree to the scanner positions, order after their relative proximity to the sample trees. Number of sample trees (n) scanned from the indicated number of scanner positions (i.e. all 39 trees were scanned from at least three positions, and for 20 trees data from seven scanner locations were used).

Scanner position	Nearest	2nd	3rd	4th	5th	6th	7th
min	1.8	3.9	5.5	9.9	12.3	14.1	16.9
max	11.6	15.8	18.4	30.6	35.8	44.2	49.0
mean	5.1	8.3	11.4	16.6	22.5	26.9	32.7
n	39	39	39	34	24	20	20

2.4.2 TLS derived tree position and d_{bh}

Several studies have presented procedures to automatically derive tree position and d_{bh} from TLS point clouds (Simonse et al., 2003; Bienert et al., 2007; Yao et al., 2011; Moskal and Zheng, 2012). To avoid introduction of errors, position and d_{bh} were in the present study derived manually. A consistent procedure solely based on the TLS data were applied, which means that an accurate automatic algorithm should obtain similar results. The tree positions and d_{bh} were manually derived with the following procedure: First, the unified point cloud around each sample tree was classified into ground and non-ground by iteratively fitting a triangular irregular network (TIN) as described by Axelsson (2000). The ground echoes were then used to construct a TIN surface. The height above ground was calculated for all laser echoes as the vertical distance from the echo to the TIN surface. All echoes in the height interval of 1.25 – 1.35 m above ground were visually inspected and the tree position and d_{bh} determined by picking two pairs of diametrically opposed points at the brim of the identified tree stem (Fig. 1).

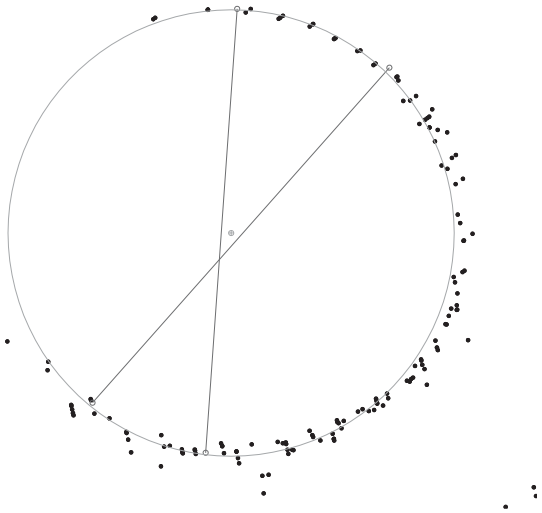

Fig. 1. Planimetric plot of the laser echoes corresponding to the breast height (1.25 – 1.35 m above ground) of one sample tree (black dots). Tree position and d_{bh} were extracted by manually selecting four points (open circles, the endpoints of the lines). Tree position and d_{bh} were calculated as the mean of the midpoints and lengths of the two lines, visualized by the crosshair symbol and the large circle, respectively.

2.4.3 TLS derived features

In order to relate the TLS data to the field measured biomass, features were extracted from the TLS data and used to create variables to be used in subsequent regression analysis. Two different approaches were followed when deriving single-tree TLS features. In the first approach a set of crown properties were manually measured in the TLS data. In the second, and more comprehensive approach a manual delineation of the TLS point cloud into single-tree segments was carried out, followed by a voxel-based feature extraction. A more detailed description of the two approaches is given below.

TLS features derived through crown measurements

Holopainen et al. (2011) found a strong relationship between TLS derived features and total above-ground biomass. Crown width and stem diameter measurements manually derived from the TLS data were used in that study. In the present study, a set of TLS features were derived from manual measurements of individual tree crowns in the unified TLS data. The manual measurements were performed using a 3D point cloud viewer with an additional 2D plot of a horizontal and a vertical slice of the data (Fig 2). The horizontal and vertical slices were of 1 m and 0.6 m thickness, respectively. In the 2D plot of the slices, measurements of the horizontal crown projection and the crown width could be done at any given height. The crown projection onto the horizontal plane was measured as a polygon and the crown width as the horizontal distance from one side of the crown to the other. Both measurements were performed at heights corresponding to 10, 20, 40, 60, and 80 percent of the crown length. The crown length was calculated as the vertical distance from the crown base height (CBH) and the highest recorded laser echo attributed to the given tree. The CBH was manually determined by visually inspecting the TLS data, and we defined it to correspond to the lowermost part of the lowermost living branch of the tree. The

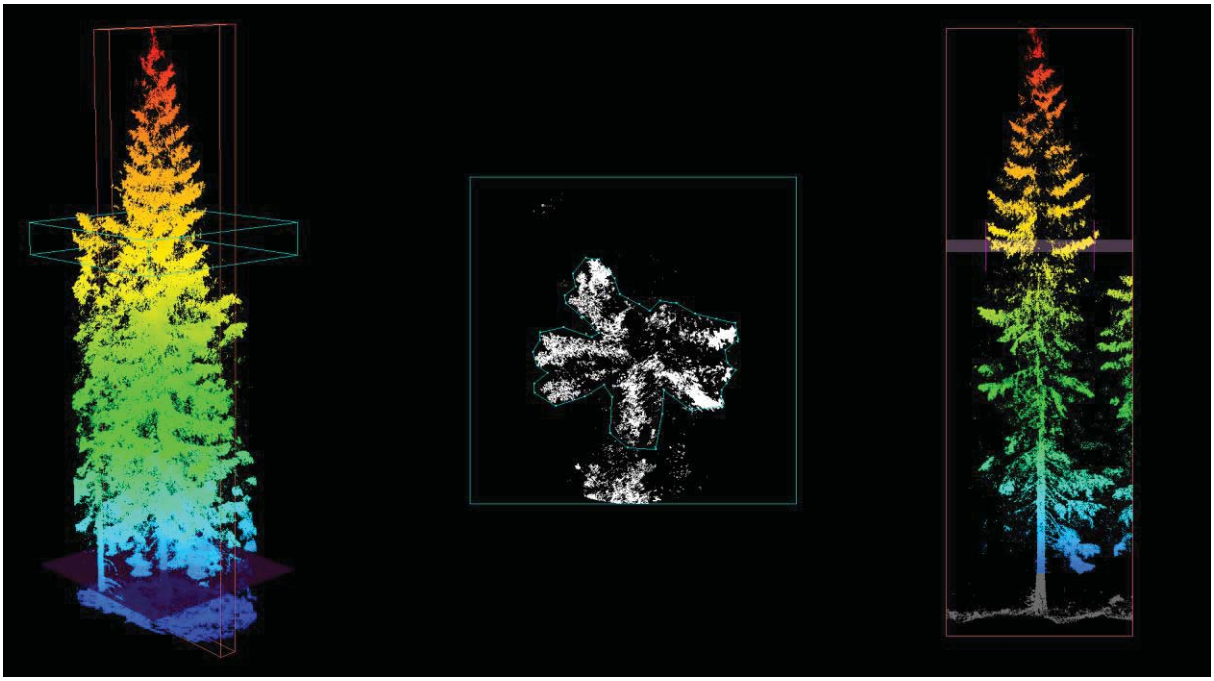

Fig. 2. Plot of the laser echoes from one of the sample trees. Overview (left), horizontal slice (middle) and vertical slice (right). The bounding boxes for the two slices are visualized on the overview plot. Manual registration of crown projection (middle) and crown width (right).

variables computed from the crown measurements are described and listed in Table 3.

Voxel-based TLS features

For each sample tree the data from each separate scan were analysed and combined to get the voxel-based TLS features. First, the spatial extent of each sample tree was manually determined by visually inspecting the unified point cloud in the surroundings of the tree position. All the laser echoes considered to belong to the sample tree in question were marked (Fig 3). A voxel space was created around the marked echoes and all voxels containing at least one marked echo were assigned to the specified tree. Separate analysis was carried out with cubic voxels of size 0.1 m, 0.2 m, and 0.4 m.

For each tree, the laser data from each of the separate scans were analysed for each voxel in the following steps: First, all the laser measurement beams "hitting" the voxel when viewed from the scanner position were identified (Fig 4) and assigned to one of three categories (*i*) reflected before the laser beam entered the voxel, (*ii*) reflected from inside the voxel, and (*iii*) reflected from beyond the range of the voxel (or with no

Fig. 3. Plot of the laser echoes from the surroundings of one of the sample trees. The echoes assigned to the sample tree appear in black, all other echoes are in grey.

Table 3. Description of the TLS-derived variables

Variable	Description
Crown measurement derived variables	
CR_{length}	Crown length. Vertical distance from the crown base height to the highest laser echo assigned to the tree.
CRA_{10} ,...,CRA_{80}	Crown projection area. Area of the crown projection measured at heights corresponding to 10, 20, 40, 60, and 80 percent of the crown length.
CRA_{sum}	Crown projection area sum. The sum of the area of the crown projection measurements at heights corresponding to 10, 20, 40, 60, and 80 percent of the crown length.
CRW_{10} ,...,CRW_{80}	Crown width. The crown width measured at heights corresponding to 10, 20, 40, 60, and 80 percent of the crown length.
CRW_{sum}	Crown width sum. The sum of the crown width measurements at heights corresponding to 10,20,40,60 and 80 percent of the crown length.
Voxel derived variables[a]	
VXR	Sum of voxel reflection ratios. The sum of the reflection ratios for all voxels assigned to the tree.
VXR_{inner} VXR_{mid} VXR_{outer}	The sum of the voxel reflection ratios for voxels assigned to the inner, middle and outer part of the crown, respectively.
VXN	Number of non-empty voxels assigned to the tree.
VXN_{inner} VXN_{mid} VXN_{outer}	The number of non empty voxels assigned to the inner, middle and outer part of the crown, respectively.

[a] Voxel variables were derived from voxels of size 0.1 m, 0.2 m, and 0.4 m (see section 2.4.3 for details).

recorded echo, indicating a reflection from beyond the maximum range of the scanner or a beam emitted into the sky) (Fig. 4). This categorization enabled us to identify those laser beams that actually entered into the voxel and how many of these were reflected from within the actual voxel. A similar approach was taken by Bienert et al. (2010) using ray tracing, but with no account of the reflections from beyond the maximum range of the scanner or beams with no recorded reflection.

Since the aim of the present study was to estimate the branch biomass, the laser beams assumed to be reflected from the stem was identified and treated as being in category *iii* described above. Previous studies have used TLS data to reconstruct the stem (Gorte and Pfeifer, 2004; Pfeifer et al., 2004; Côté et al., 2011). In the present study, stem position and shape was approximated by using the manually derived d_{bh} and stem position from the TLS data (see section 2.4.2). Two additional stem

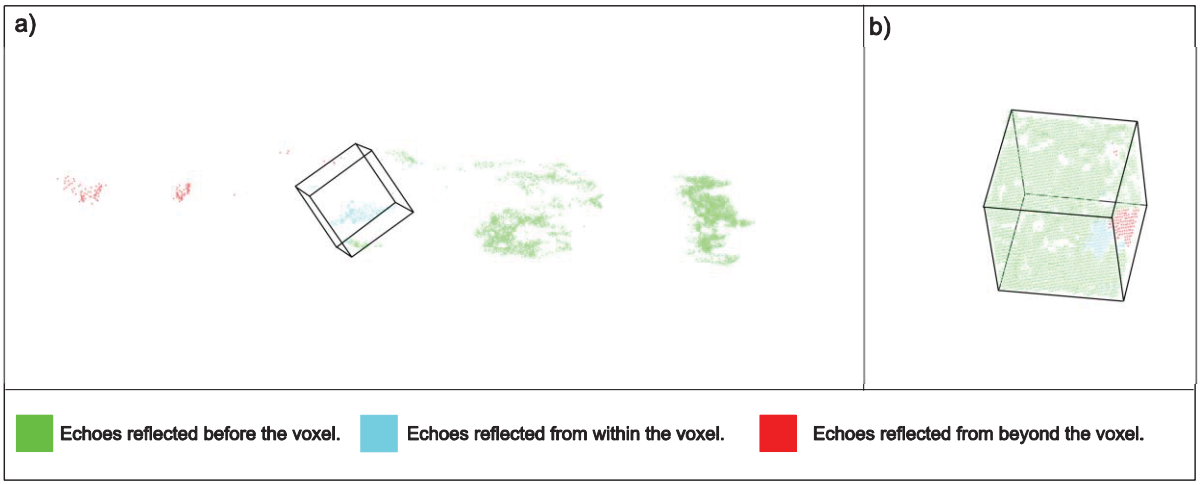

Fig. 4. Visualization of one voxel in the data material and all laser echoes (from one single scan) assigned to that voxel. Side view (a) and as viewed from the scanner (b). The three echo categories are indicated by different colors.

position and diameter measurements were derived in a similar fashion. The initial height of the two additional measurements was chosen to correspond to 30 and 50 percent of the tree height but was moved up or down if a satisfactory representation of the stem was unobtainable from the TLS data at those initial heights. The stem diameter and position were linearly interpolated between the three points of measurements, and above the highest stem measurement a vertical stem with a tapering of 1 cm per meter was assumed.

All echoes within a 3 cm distance from the approximated stem were considered to be reflected from the stem (Fig. 5). Only echoes up to the height corresponding to an approximated stem diameter of 7 cm were considered. This was done to minimize the errors caused by the non-vertical nature of the true stem, and also because it corresponds with a typical minimum cut diameter in a harvest operation.

Fig. 5. Stem echoes. Plot of the laser echoes for one of the sample trees. The laser echoes assigned to the stem of the sample tree appear in black, all other echoes are in grey.

The proportion of the entering laser beams that were reflected from within a voxel was considered to give an indication of the biomass present inside that voxel. This was calculated as the ratio between the number of recorded echoes to the number of entering beams. Thus, a value of 1 meant that all entering beams were reflected from within the voxel, and 0 that no echo was recorded within the voxel. For any given voxel this ratio would differ from one scan to another due to shadowing effects and different viewing angles. For each voxel a combined ratio was calculated as the weighted mean of the ratios for that particular voxel from each scan. The number of entering beams was used as weight, giving more weight to a scan with a position "seeing" more of the voxel.

In Norway spruce trees the biomass tends to be distributed unevenly along the length of a branch, with more needles towards the outer part of the branch. In order to incorporate this in the analysis we derived voxel features based on a classification of the voxels into three classes.

The voxels assigned to each tree were classified as belonging to the inner, middle or outer part of the crown. The distance to the approximated stem was used in the classification according to the following procedure: First, the voxels were assigned to vertical height bins of 2 m. The voxels assigned to each bin were further assigned to one of four quadrants. The voxels assigned to each quadrant were classified into three classes by dividing the maximum distance from any voxel to the stem into three equal parts (Fig 6).

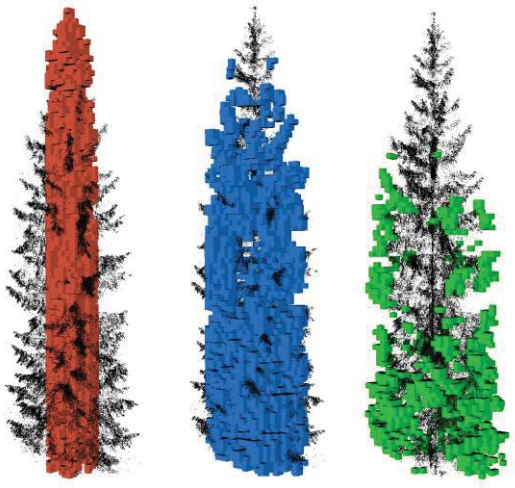

Fig. 6. Voxel classification. Visualization of the voxels assigned to the inner (left), middle (middle) and outer (right) part of the tree crown.

A range of features were extracted and computed from the voxel data for each individual tree.

All variables were computed for cubic voxels of size 0.1 m, 0.2 m, and 0.4 m. The assignment of the voxels to the inner, middle, and outer part of the crown were also utilized. The variables derived from the voxel data are listed and described in Table.

2.4.4 Predictive models

Linear least squares regression has been widely used to fit predictive models to remote sensing data and field observations. With the emergence of new machine learning algorithms such as support vector machine and random forest (RF), these and other regression techniques have been applied for remote sensing forestry predictions (Næsset et al., 2005; Vauhkonen et al., 2010; Zhao et al., 2011). In the present study, non-parametric RF regression was chosen because it has been found to perform well compared to other regression techniques, and with little tuning required (Hastie et al., 2003). It is robust with respect to noise variables, which means that the variable selection deemed necessary in the least squares regression is avoided (Biau, 2012). Furthermore, other recent studies have reported successful use of RF for forest inventory applications with remote sensing data (Latifi et al., 2010; Vauhkonen et al., 2010). The principles behind random forest regression as it is commonly used today were introduced by Breiman (2001). In this machine learning algorithm a large number of binary trees are grown from bootstrap samples of the data. The predicted value is then the average of the predicted values from the individual binary trees. In the present study the implementation of RF in the the *randomForest* package in the statistical software R was used (R Development Core Team, 2011). Further details about the principles and use of the RF algorithm can be found in Breiman (2001) and Hastie et al. (2003). In the present study a linear least squares model with a stepwise variable selection was also fit, as a reference.

In order to assess the predictive potential of the TLS derived variables, regression models were fit to the data with *BR* as response variable. Linear least squares and RF regression were used and the models were fit in a repeated sub-sampling validation procedure. In this procedure, also known as Monte Carlo cross-validation, the data is repeatedly split into a training set and a test set. A model is fit to the training set, and then the prediction accuracy is assessed by comparing the predicted values for the test set with the true values. The model and prediction accuracy statistics were averaged over all iterations. In the present study, 1000 iterations were used and 30 sample trees were each time selected into the training data set, leaving nine trees in the test data set.

In total seven distinct models were fit (Table 4). Since the voxel approach involved a more complex procedure compared to the approach using crown measurements, we wanted to quantify possible differences in prediction accuracy between the two methods. Two separate models were therefore fit to data derived with the two approaches. To assess the effect of voxel size, three separate models were fit with variables from the three voxel sizes used.

A linear stepwise model was fit as a reference, using an empty model as the initial model, and the full set of TLS derived variables as potential explanatory variables. The step function in the *stats* package of the statistical software R was

Table 4. Description of the models fit in the Monte Carlo cross validation.

Model name	Regression technique	Explanatory variables
lm-step-all	Stepwise linear least squares	all TLS-derived variables
RF-all	RF	all TLS-derived variables
RF-cr	RF	TLS-derived crown measurements variables
RF-vox10	RF	TLS-derived 0.1 m voxel variables
RF-vox20	RF	TLS-derived 0.2 m voxel variables
RF-vox40	RF	TLS-derived 0.4 m voxel variables
site-specific mod	Linear least squares	Field measured d_{bh} and h

used, with both forward and backward selection enabled. The order in which the variables were entered into the algorithm was randomly altered in each iteration.

In order to compare the TLS-based models to conventional site-specific allometric models also a model using the field measured d_{bh} and h was included. The linear regression model form as used for branch biomass by Marklund (1988) was applied in this site-specific allometric model, i.e.

$$\log(BR) = \beta_1 \left(\frac{d_{bh}}{d_{bh}+13}\right) + \beta_2 \log(h) + \beta_3 h + \epsilon, \quad (1)$$

where β_1,β_2 and β_3 are parameters to be estimated in the model and ε is an normally distributed error term. This model was included in the Monte Carlo cross-validation process, using the same training and test data.

In addition to the statistics from the seven regression models, also statistics from branch biomass estimates derived with existing allometric models in relation to the accurately measured branch biomass were computed in the validation iterations. Using the same training and test data, model fit and prediction accuracy were computed for the allometric model derived biomass predictions. Models by Marklund (1988) were used with field measured d_{bh} and h as explanatory variables. The model fit statistics were calculated by fitting a linear model with *BR* as response and the allometric model derived biomass as predictor using the training data. The prediction accuracy statistics were calculated by using the test data only, where biomass derived with allometric model and field measured d_{bh} and h was used as the predicted values.

2.4.5 Assessment of model fit and prediction accuracy

The coefficient of determination (R^2) where inspected and used as an indicator of model fit. Pseudo R^2 values were computed as

$$\rho(fitted\ values, observed\ values)^2, \quad (2)$$

where ρ is the Pearson's correlation.

The accuracy of the predictions was assessed by computing the root mean square error (RMSE) of the predicted values. The RMSE was computed as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(BR_i - \widehat{BR}_i)^2}{n}}, \quad (3)$$

where n is the number of trees and $\widehat{BR}$ is the predicted branch biomass. In the present study the RMSE is reported in percent of the mean observed value, i.e.

$$RMSE\% = \frac{RMSE}{\frac{\sum_{i=1}^{n} BR_i}{n}} 100. \quad (4)$$

Bias was measured as

$$bias = \frac{\sum_{i=1}^{n}(BR_i - \widehat{BR}_i)}{n}. \quad (5)$$

3 Results

The relationship between the TLS derived variables and the field measured branch biomass was found to be strong, with average R^2-values ranging from 0.42 to 0.88 for the tested models (Figure 7a). Overall, the voxel approach and the crown measurement approach yielded comparable R^2-values, but the results show that the R^2-values became lower with smaller voxel size. The model with the lowest R^2-value was the model based on variables derived from the 10 cm voxel data (Fig. 7a). The model with the highest R^2-value was the linear stepwise model, with an average R^2 of 0.88 (Fig 7a).

The cross-validation revealed that, except for the model based on the 10 cm voxel data, all the TLS based models yielded higher average prediction accuracies than the estimates derived with existing allometric models. The estimates derived with field measured d_{bh} and h using existing allometric models from Marklund (1988) had a mean RMSE of 38% (Fig. 7b). The best accuracy of the TLS-based models were obtained with the model using the 40 cm voxel data and the random forest model using the full set of TLS derived variables which both had a mean RMSE of 32% (Fig 7b).

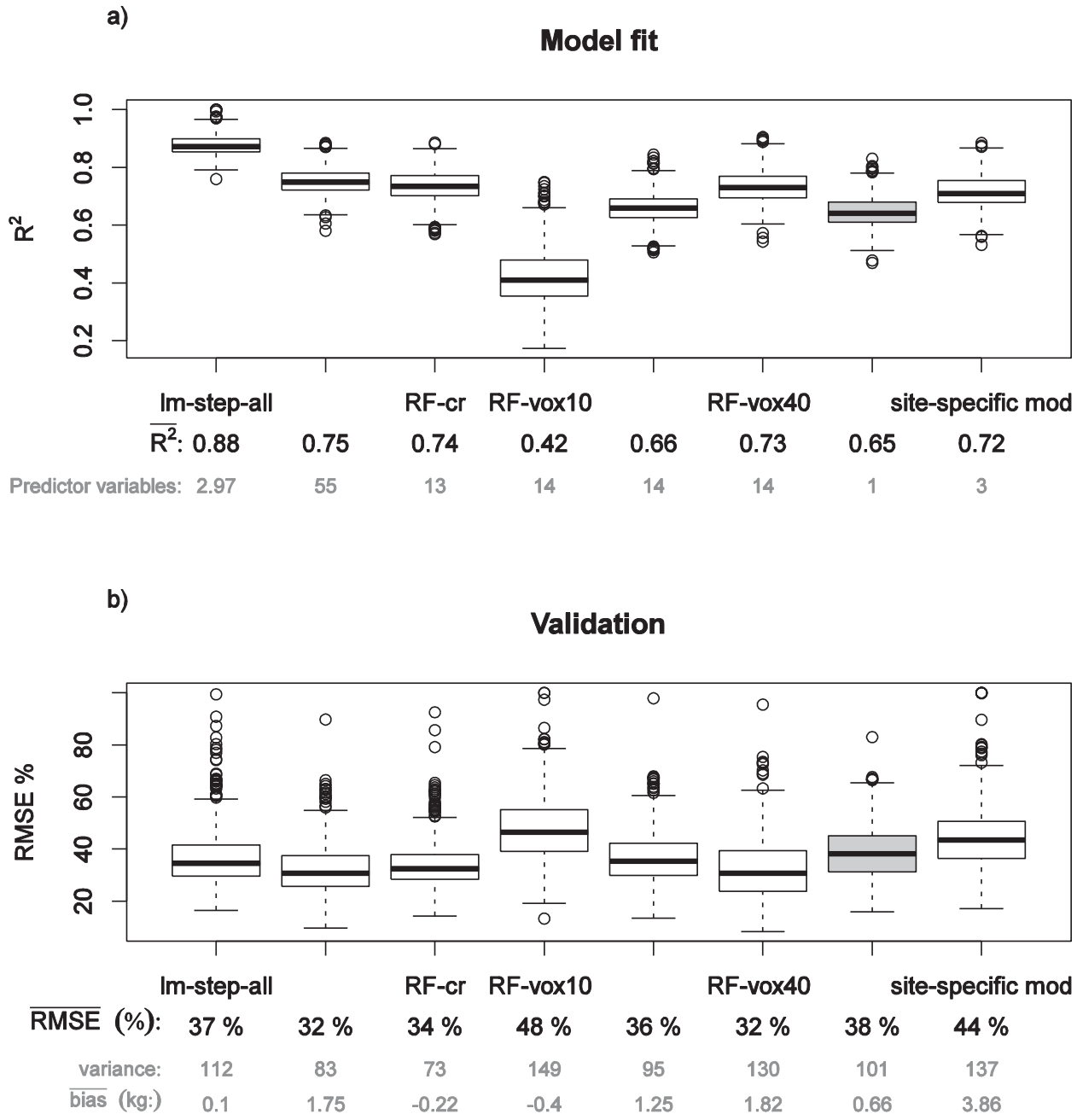

Fig. 7. Fit statistics and prediction accuracy of the regression models from the Monte Carlo cross-validation. In addition the corresponding statistics for prediction with an existing allometric model (marked with grey color). The mean R^2 and number of predictor variables in the model (a). The mean RMSE% and bias and the variance of the RMSE% for the model predictions (b). The description of the models are (starting with 1 at the left side of the figure, then counting to the right): (1) A linear stepwise model with all the TLS derived variables, (2) a RF model with all variables, (3) a RF model with crown measurement derived variables, (4-6) RF models with 0.1, 0.2 and 0.4 m voxel derived variables, (7) prediction with an existing allometric model, and (8) a linear model with field measured d_{bh} and h as predictor variables (see text, section 2.4.4 for details).

The validation results further showed that the stepwise model had a lower accuracy than the random forest model using the same initial set of variables (Fig. 7b). Only small biases were observed with the TLS-based models. When examining the variability (variance) among the RMSEs obtained over the 1000 iterations, the model with crown measurement variables was found to yield the most stable predictions (Fig. 7b). The most unstable model was found to be the site-specific allometric model (Fig. 7b).

4 Discussion and conclusion

A strong relationship was found between TLS data and field measured branch biomass, and the predictive power of the derived TLS variables was found to be better than both existing allometric models with field measured d_{bh} and h and a local allometric equation developed based on the same destructively-sampled data.

One hypothesis for the present study was that the models with the voxel-based TLS variables incorporating the beam vector information would yield estimates with a higher accuracy than models based on less comprehensive spatial analysis of the TLS data. The hypothesis was partly confirmed, but the differences between the models were small and the models based on smaller voxels yielded lower accuracy. One possible cause is the effect of the single-tree delineation. The voxel derived variables is highly dependent on the manual delineation of the point cloud into single-tree segments. It became evident during the course of the analysis that accurate crown delineation of Norway spruce can be difficult, even under canopy conditions with moderate density. The reason for this was the combined effect of overlapping crowns and shadowed areas in the TLS data. We nevertheless believe that a voxel-based approach could be pursued further, possibly utilizing the beam vector information to aid the delineation process. More research is however needed on this issue.

Although a strong relationship was found, the improvement in prediction accuracy was relatively low, given the time and effort spent on the data acquisition and analysis. Large amounts of data were collected for each scanned tree and a greater increase in prediction accuracy should be expected to justify this extensive fieldwork. In the present study a mere increase of six percent-points in the RMSE was obtained over conventional field measurements of d_{bh} and h. This require however the availability of an existing allometric model. For species or regions where such models are not readily available the use of TLS data could be an option, but training data still need to be obtained. Results in the present study show that a more accurate and stable model can be obtained using TLS data compared to a site-specific allometric model based on the local relationship of d_{bh} ,h and branch biomass, given the same set of training data. The performance of the model with the site-specific allometric model will most likely be better if a larger training dataset had been used. The performance and the high variance in the error statistic over the iterations for this model can, at least partly, be attributed to the small number of observations in the training dataset.

The reason for the apparent decrease in accuracy and a poorer model fit with decreasing voxel size is unclear. One reason could be that the shadowed areas were better represented within the larger voxel data since the voxels themselves were constructed solely from the point cloud of the laser echoes. A (small) voxel in a completely shadowed area would not be included in subsequent analysis, whereas this same space in the tree crown could be included in a larger voxel which happened to encompass some nearby laser echoes. This effect could be avoided by including seemingly empty spaces surrounding the trees when they are segmented, possibly utilizing the beam vector information. This will however further complicate the segmentation process.

In this study only a small gain in accuracy was obtained by the computationally-intensive voxel-based approach compared to the less intensive crown-dimension approach. It is often possible to obtain reasonable crown dimensions from a single TLS scan. Hence an efficient way of using TLS in biomass estimation may be to use a single-scan setup combined with the crown-dimension approach. In this context, it should also be noted that the derived crown dimensions would be expected to differ less between different scanning systems than with the voxel-based approach (Ducey et al. 2012).

The derivation of TLS features in the present study included several manual and semi-manual operations. These manual operations were conducted in such a way that they in principle could have been done automatically, solely using the TLS data. Methodology and specific algorithms to automatically perform these tasks are to our knowledge not readily available and expected errors and accuracy of such an automated approach should be further investigated.

To conclude, we found in the present study a strong relationship between TLS derived variables

and accurately measured branch biomass. The increase in prediction accuracy over the usage of existing allometric models with field measured d_{bh} and h was however relatively small. Several challenges were identified, one of them was to correctly assign the TLS data to single-tree segments under dense forest conditions. The use of the beam vector information did not substantially improve the predictions in the present study, but this could be subject to further research.

Acknowledgements

We wish to thank Geoplan 3D, Norway, for collecting and partly processing the laser scanner data, and Janka Dibdiakova at The Norwegian Forest and Landscape Institute for doing the lab work. We are also grateful to Arne Drømtorp, Leif Kjøstelsen and Espen Martinsen (The Norwegian Forest and Landscape Institute); Vegard Lien and Tord Storbækken (Norwegian University of Life Sciences) who took part in the field work.

References

Axelsson, P., 2000. DEM generation form laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing XXXIII, 110–117.

Biau, G., 2012. Analysis of a Random Forest Model. Journal of Machine Learning Research 13, 1063–1095.

Bienert, A., Queck, R., Schmidt, A., Bernhofer, C., Maas, H., 2010. Voxel space analysis of terrestrial laser scans in forests for wind field modeling. International Archives of Photogrammetry,Remote Sensing and Spatial Information Sciences XXXVIII, Part 5, Commission V, 92–97.

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. International Archives of Photogrammetry and Remote Sensing 3, 50–55.

Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., Kimes, D.S., 2008. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. Remote Sensing of Environment 112, 3876 – 3890.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32.

Côté, J.-F., Fournier, R.A., Egli, R., 2011. An architectural model of trees to estimate forest structural attributes using terrestrial LiDAR. Environmental Modelling & Software 26, 761–777.

Danson, F.M., Hetherington, D., Morsdorf, F., Koetz, B., Allgower, B., 2007. Forest Canopy Gap Fraction From Terrestrial Laser Scanning. IEEE Geoscience and Remote Sensing Letters 4, 157–160.

Ducey, M.J., Astrup, R., Seifert, S., Pretzsch, H.C., Larson, B.C., Coates, K.D., 2012. Comparison of forest attributes derived from two terrestrial LiDAR systems. Photogrammetric Engineering & Remote Sensing.

Fuchs, H., Magdon, P., Kleinn, C., Flessa, H., 2009. Estimating aboveground carbon in a catchment of the Siberian forest tundra: Combining satellite imagery and field inventory. Remote Sensing of Environment 113, 518 – 531.

Gobakken, T., Næsset, E., Nelson, R., Bollandsås, O.M., Gregoire, T.G., Ståhl, G., Holm, S., Ørka, H.O., Astrup, R., 2012. Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. Remote Sensing of Environment 123, 443–456.

Gorte, B., Pfeifer, N., 2004. Structuring laser-scanned trees using 3d mathematical morphology. International Archives of Photogrammetry and Remote Sensing 35, 929–933.

Hastie, T., Tibshirani, R., Friedman, J., 2003. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 1st ed. 2001. Corr. 3rd printing. ed. Springer.

Hauglin, M., Dibdiakova, J., Gobakken, T., Næsset, E., 2011. Estimating single-tree branch biomass of Norway spruce by airborne laser scanning, in: Proceedings of Silvilaser 2011. Presented at the Silvilaser 2011, Hobart, Australia.

Holopainen, M., Vastaranta, M., Kankare, V., Räty, M., Vaaja, M., Liang, X., Yu, J., Hyyppä, H., Hyyppä, R., 2011. Biomass estimation of individual trees using stem and

crown diameter TLS measurements. ISPRS Workshop on Laser scanning XXXIII.

Hosoi, F., 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, 3610–3618.

Hyyppä, J., Kelle, O., Lehikoinen, 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.

Jenkins, J., Chojnacky, D., Heath, L., Birdsey, R., 2003. National-scale biomass estimators for United States tree species. For. Sci. 49, 12–35.

Latifi, H., Nothdurft, A., 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, 395–407.

Lefsky, M., McHale, M.R., 2008. Volume estimates of trees with complex architecture from terrestrial laser scanning. Journal of Applied Remote Sensing 2, 023521.

Lovell, J.L., Jupp, D.L.B., Newnham, G.J., Culvenor, D.S., 2011. Measuring tree stem diameters using intensity profiles from ground-based scanning lidar from a fixed viewpoint. ISPRS Journal of Photogrammetry and Remote Sensing 66, 46–55.

Marklund, L.G., 1988. Biomass functions for pine, spruce and birch in Sweden (Report No. 45). Swedish University of Agricultural Sciences, Umeå.

Moorthy, I., Miller, J.R., Hu, B., Chen, J., Li, Q., 2008. Retrieving crown leaf area index from an individual tree using ground-based lidar data. Canadian Journal of Remote Sensing 34, 320–332.

Moskal, L.M., Zheng, G., 2012. Retrieving Forest Inventory Variables with Terrestrial Laser Scanning (TLS) in Urban Heterogeneous Forest. Remote Sensing 4, 1–20.

Næsset, E., 2004. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. Scandinavian Journal of Forest Research 19, 164–179.

Næsset, E., Bollandsås, O.M., Gobakken, T., 2005. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. Remote Sensing of Environment 94, 541 – 553.

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.

Næsset, E., Gobakken, T., Solberg, S., Gregoire, T.G., Nelson, R., Ståhl, G., Weydahl, D., 2011. Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: A case study from a boreal forest area. Remote Sensing of Environment 115, 3599–3614.

Pfeifer, N., Gorte, B., Winterhalder, D., 2004. Automatic Reconstruction of Single Trees from Terrestrial Laser Scanner Data. ISPRS - International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXV, 114– 119.

Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne lidar. Biomass and Bioenergy 31, 646–655.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. Vienna, Austria.

Simonse, M., Aschoff, T., Spiecker, H., 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. Umeå, Sweden, pp. 251– 257.

Van der Zande, D., Jonckheere, I., Stuckens, J., Verstraeten, W.W., 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, 526–538.

Vauhkonen, J., Korpela, I., Maltamo, M., 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.

Yao, T., Yang, X., Zhao, F., Wang, Z., Zhang, Q., Jupp, D., Lovell, J., Culvenor, D., Newnham, G., Ni-Meister, W., Schaaf, C., Woodcock, C., Wang, J., Li, X., Strahler, A., 2011. Measuring forest structure and biomass in New England forest stands using Echidna ground-based lidar. Remote Sensing of Environment 115, 2965–2974.

Zhao, K., Popescu, S., Meng, X., Pang, Y., Agca, M., 2011. Characterizing forest canopy structure with lidar composite metrics and machine learning. Remote Sensing of Environment 115, 1978–1996.

Zianis, D., Muukkonen, P., Mäkipää, R., Mencuccini, M., 2005. Biomass and Stem Volume Equations for Tree Species in Europe. Silva Fennica 63.

Paper IV

Estimating single-tree branch biomass by airborne laser scanning using model training data with ground reference values obtained through terrestrial laser scanning

Marius Hauglin[a], Terje Gobakken[a] ,Rasmus Astrup[b], Liviu Ene[a], Erik Næsset[a]

[a] *Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences.*
[b] *The Norwegian Forest and Landscape Institute.*

Abstract

Forest inventories are essential for sustainable management of forest resources. Several methods to conduct single-tree inventories using airborne laser scanning (ALS) have been proposed. Even *terrestrial* laser scanning (TLS) has recently emerged as a promising tool for collection of forest information. In the present study a novel methodological framework for an ALS and TLS-based inventory is proposed. Single-tree Norway spruce branch biomass was predicted using an ALS-model using training data with ground reference values obtained by TLS. ALS and TLS data were collected for a set of sample trees, including 68 trees with both ALS and TLS data. In total, 29 destructively sampled trees were used to fit a TLS branch biomass model, which then was used to predict branch biomass in a separate set of 68 trees. This dataset was subsequently used to fit ALS branch biomass models. When validating the models against accurately measured branch biomass obtained through destructive sampling of 17 trees the root means square errors ranged from 32.4% to 38.8% of the observed mean. Corresponding branch biomass predictions derived with ALS-predicted diameters and the use of conventional and existing allometric models resulted in root mean square errors of 41.5% and 44.5%. Thus, more accurate predictions were obtained using training data with ground reference values derived with TLS. The successful implementation in the present study suggests that the proposed framework can be used as a general approach when deriving ground reference values for single-tree ALS models.

Abbreviations: ALS – airborne laser scanning, TLS – terrestrial laser scanning, d_{bh} – diameter at breast height

1 Introduction

Forest inventories are essential for sustainable management of forest resources. An increasing number of forest management inventories are based on data collected with airborne laser scanning (ALS) (Maltamo et al., 2011). While commercial and operational ALS-based forest inventories most frequently are conducted according to the so-called area-based approach as described by Næsset (2002), methods targeting single trees have also been proposed (Hyyppä et al., 2001; Persson et al., 2002; Solberg et al., 2006; Ene et al., 2012). The latter methods usually require ALS data with higher resolution, but intend to give information on a single tree level, contrary to the area-based information provided by the former. Although not as widely used as the area-based method at present, the use of single-tree forest inventories might increase in the future. This will depend on the ongoing technological and methodological research and development, the need for single tree data, and future costs for data acquisition. The potential of estimating individual tree characteristics by ALS has been investigated in several studies, including stem volume (Persson et al., 2002; Straub and Koch, 2011), stem diameter (Popescu, 2007), crown base height (Vauhkonen, 2010), leaf area index (Roberts et al., 2005) and biomass (Popescu, 2007; Hauglin et al., 2011; Räty et al., 2011). Features derived from 3D alpha shapes of the ALS echoes have been used to

* Corresponding author: Marius Hauglin, Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences. Po. box 5003, 1432 Ås, Norway. Email adress: marius.hauglin@umb.no Tel.: +47 6496 5389.

predict diameter, tree height and volume (Vauhkonen et al., 2010). Hauglin et al. (2012b) showed variables describing the crown volume to be promising predictor variables when estimating branch biomass.

The same laser ranging principles as in the airborne scanning systems have also been applied in scanning from fixed positions on the ground, commonly known as terrestrial laser scanning (TLS). The use of TLS in forestry has not become operational to the same degree as the use of ALS, but several applications – such as estimation of tree position and diameter at breast height (d_{bh}) – have been investigated (Simonse et al., 2003; Bienert et al., 2007; Lovell et al., 2011; Yao et al., 2011; Moskal and Zheng, 2012). There are also studies presenting procedures to estimate e.g. total above-ground biomass from TLS data (Holopainen et al., 2011; Yao et al., 2011). Hauglin et al. (2012a) have described methods to estimate branch biomass using TLS.

ALS-based forest inventories are typically carried out with ground reference data obtained from manual field measurements of e.g. d_{bh} and tree height on selected sample trees collected in sample plot surveys. Depending on the ongoing technological and methodological research and development TLS might become an efficient and precise tool for collection of ground reference data in (ALS-based) forest inventories, possibly replacing manual registrations. The use of TLS can even enable registrations of features that are practically unobtainable through conventional field registrations, such as spatial structural properties.

In the last ten years there has been an increased interest in the use of logging residues for bioenergy purposes, i.e., biomass otherwise left in the forest during the logging (Hauglin et al., 2012c). When logging residues becomes a commercial product from the forest, the potential amount should be quantified as part of the forest inventory, to improve planning and management.

Estimation of forest biomass using ALS usually requires the use of field reference data obtained through allometric models, typically linking field measured d_{bh} and tree height with a given biophysical property (Popescu, 2007; Boudreau et al., 2008; Næsset and Gobakken, 2008; Hauglin et al., 2012c). For Norway spruce (*Picea abies* (L.) Karst.) – the tree species considered in the present study – the branch biomass can vary considerably in relation to d_{bh} and tree height, and predictions of Norway spruce branch biomass from allometric models has larger errors than corresponding predictions of e.g. stem volume (Marklund, 1988; Zianis et al., 2005). Hauglin et al. (2012a) showed that Norway spruce branch biomass estimates can be derived from TLS data with a higher accuracy than by conventional field measurements and existing allometric models, making estimation of branch biomass a suitable case for replacing manual field registrations with TLS.

Linear least squares regression models have been widely used to predict biophysical properties from remote sensing data. Other regression techniques such as random forest, support vector machine and partial least squares have also been applied in predictive models in forestry (Næsset et al., 2005; Vauhkonen et al., 2010; Zhao et al., 2011). Some recent studies report successful use of random forest regression for forest inventory applications with remote sensing data (Latifi et al., 2010; Vauhkonen et al., 2010). In the present study we used both linear least squares and random forest models.

We propose a procedure for ALS-based estimation of single-tree branch biomass using training data with ground reference values obtained through TLS. The use of TLS to obtain the ground reference data in an ALS-based inventory is to our knowledge a novel approach. An independently developed TLS-based model was used to obtain the ground reference data, and in order to resemble an operational approach, automatic single-tree segmentation was used in the assignment of ALS data to individual trees.

To test our proposed procedure with TLS as ground reference data in an ALS-based inventory, we estimated single-tree branch biomass of Norway spruce, and evaluated the proposed procedure with an independent dataset consisting of trees with accurately measured branch biomass obtained with destructive sampling. The prediction accuracies were compared to accuracies of branch biomass predictions obtained with ALS-predicted d_{bh} and the use of existing allometric models, i.e., predictions obtained without the use of TLS data.

2 Materials and methods

2.1 Study area

The study area is located in Aurskog-Høland municipality (59°50'N 11°30'E, 120-390 m a.s.l.) in the south-eastern part of Norway (Fig. 1, left). The total area of Aurskog-Høland is 960 km^2 with 670 km^2 productive forest. The forest type is boreal with Norway spruce and Scots pine (*Pinus sylvestris* L.) as the dominant tree species.

2.2 Forest inventory - stratification

An operational forest inventory was carried out in the study area, including a stratification of the productive forest. This stratification was utilized in collection of the field data used in the present study. The stratification was based on digital aerial photographs acquired in June 2005 with a Vexcel UltraCam D camera. Through digital stereo photogrammetry forest characteristics were interpreted manually by photo-interpretation of every stand, including stand borders, dominant tree species, site productivity, and age class, see further details in Næsset et al. (2011). The existing inventory was used as auxiliary information in the interpretation process.

2.3 Field data

Single-tree data from two field-campaigns in the study area were used in the present study. One dataset with single-tree measurements on 40 sample plots, and another with single-tree measurements obtained with destructive sampling on five sample locations. The 40 sample plot locations are marked by filled circles in Fig. 1 and described in subsection 2.3.1. The five destructive sampling locations are marked with hollow circles in Fig. 1 and described further in section 2.3.2. The data from the five locations with destructive sampling were in the present study split in two, resulting in the three datasets used in the analysis (Table 1).

2.3.1 Sample plot data

The sample plot field registrations were carried out in the fall 2007 and winter 2007-2008 on 40 circular plots of size 1000 m^2 and 500 m^2, laid out in the mature productive forest of the study area (Fig. 1, right). Single-tree data used in the present study were from a subset of 11 plots. The sample plots were initially used in other studies, and some constraints were posed regarding the location of the plots (see Maltamo et al. (2010) for details).

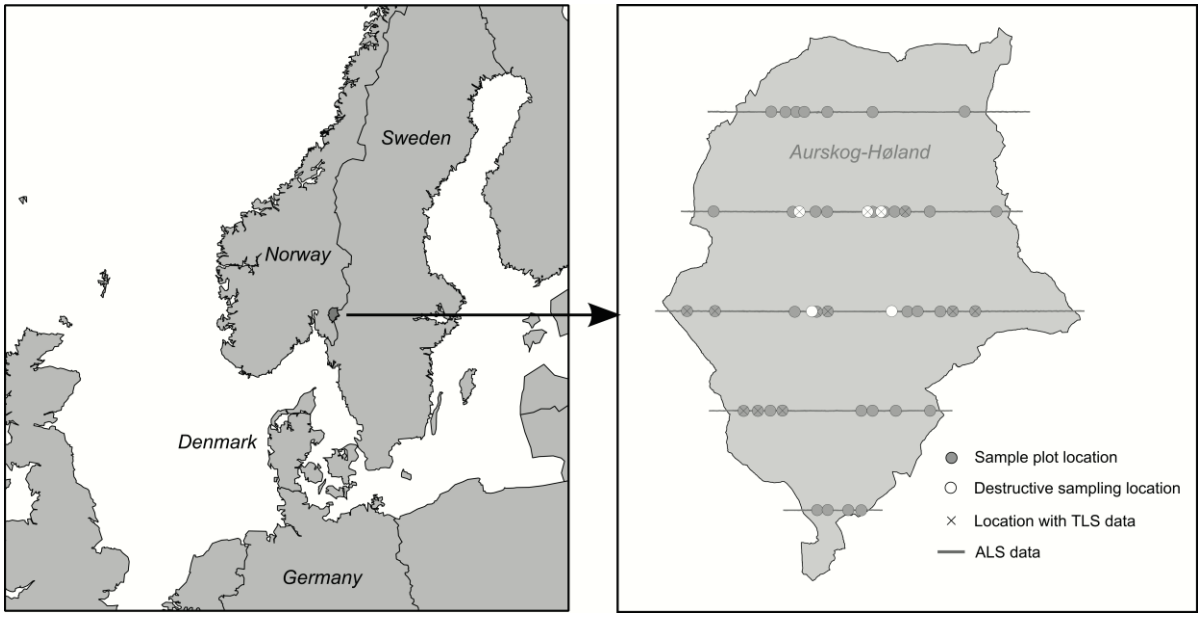

Fig. 1.. Location of the study area (left) and the locations of the field and remote sensing data within the study area (right).

Table 1. Characteristics of the three datasets used in the present study. Diameter at breast height (d_{bh}) and field measured dry weight branch biomass (*BR*).

	d_{bh} (cm)			*BR* (kg)			*n*
	min	max	mean	min	max	mean	
TLS model training dataset	9.7	39.8	22.2	8.9	152.3	60.7	*29*
ALS model training dataset	7.1	37.7	19.6	-	-	-	*68*
Validation dataset	16.6	40.3	25.4	28.6	163.2	75.7	*17*

Within each of these plots all trees with d_{bh} ≥5 cm were callipered, and tree coordinates relative to the plot center were registered using a total station. The plot center coordinates were determined by differential Global Navigation Satellite Systems (dGNSS), using dual-frequency receivers observing pseudo-range and carrier phase of the Global Positioning System (GPS) and the Global Navigation Satellite System (GLONASS). Based on the accuracy reported by the post-processing software and accuracies reported for comparable conditions in previous studies the positional accuracy is expected to be in the range of 0.01 – 0.8 m (Næsset, 1999; Hasegawa and Yoshimura, 2007; Andersen et al., 2009).

The observed number of trees on the sample plots were used to guide the segmentation algorithm (section 2.5.1), and a subset of the individual Norway spruce trees within 11 of the 40 plots (68 trees) were used in the present study (see section 2.4.1).

2.3.2 Destructive sampling data

Field data with accurately measured branch biomass were collected in June 2009. A total of 49 spruce trees were selected from five locations in the study. The five locations were chosen from potential sites in the intersections of the east-west-oriented strips of ALS data and forest roads, covering a range from poor to good site productivity in mature forest (Fig. 1, right). Within each location sample trees of Norway spruce were selected in a consecutive fashion. Further details about the selection procedure are given in Hauglin et al. (2012b). The 49 trees were 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 of 3000 kg was used. Samples of entire branches were selected among the living branches of each tree in order to determine the dry weight. Samples from the branches were dried and the raw and dry weights of each sample were recorded. The drying was done at 103±1 ˚C until constant mass. A wet to dry weight ratio was then calculated for each branch sub-sample. For each tree a wet to dry weight ratio was calculated as the weighted mean of the ratios obtained from the samples. The diameters of the branches were used as weights. Finally, the total dry weight biomass of the branches for each tree – denoted *BR* – was calculated as the wet weight of the branches multiplied with the calculated tree-specific wet to dry weight ratio. Further details about the determination of the dry weight are given in Hauglin et al. (2012b).

The coordinates of each tree in this dataset 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 dGNSS, using the same procedure as described for the sample plot dataset (subsection 2.3.1).

In the present study, this dataset with accurately measured branch biomass was split in two: first three locations (containing 29 trees) were selected and used as training data for a TLS branch biomass model (*TLS model training dataset* in Fig. 3). The remaining two locations (with 20 trees) were used for validation (*validation dataset* in Fig. 3). Not all trees at the locations were destructively sampled, and three of the initial 20 sampled trees were discarded from the validation dataset since their assigned crown segments also contained other trees (see section 2.5.1).

Characteristics of the trees in the data material used in the present study are summarized in Table 1.

2.4 Laser scanner data

2.4.1 ALS data

ALS data were collected along five strips oriented in the east-west direction and located 9 km apart (Fig 1, right). The ALS 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 scan angle was ±5 degrees from nadir. This gave an average point density on the ground of 7-8 m^{-2}. Up to four echoes were recorded for each pulse. The planimetric coordinates and the ellipsoidal height values were determined by the vendor for all echoes. The ALS data are further described in Breidenbach et al. (2010).

The classification of echoes into ground- and vegetation echoes was carried out by iteratively fitting a triangular irregular network (TIN) from below as described by Axelsson (2000). Points were in an iterative fashion added to the ground if they were within given threshold values. Echoes classified as ground were used to construct a TIN surface. The height above ground was calculated for all echoes by subtracting the respective TIN heights from the ellipsoidal heights. The range from the aircraft and the returned intensity were recorded for all echoes. The intensity values were normalized subject to the range from the aircraft as described by Ørka (2011).

2.4.2 TLS data

TLS data were obtained for 11 of the 40 sample plots (section 2.3.1) and at three of the five locations with destructively sampled trees (section 2.3.2 and Table 1). TLS data from multiple scans of 97 trees were used in the present study, and all scans were acquired with a Leica HDS6000 phase-shift scanner. The scanning was done with a horizontal and vertical angle increment of the laser measurements of 0.036 degrees. This corresponds to a point spacing of 15.9 mm at a 25 m distance from the scanner. The scanner had a beam diameter when leaving the instrument of 3 mm and a beam divergence of 0.11 mrad, resulting in an 8 mm footprint at 25 m.

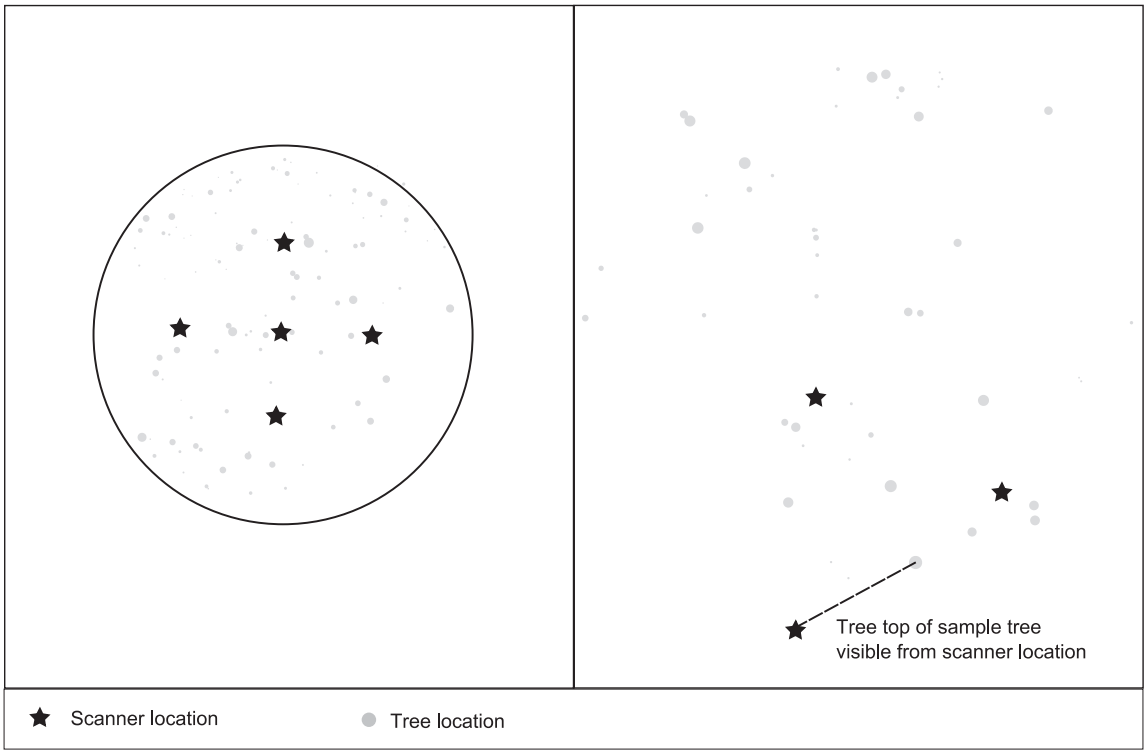

Fig. 2. Example of scanner locations at the sample plots (left), and on locations with destructive sampled trees (right).

The scanner had a maximum measurement range of 79 m. A full 360 x 310 degree scan (the scanners maximum field-of-view) was performed from each scanner position, i.e., both hemispheres excluding the ground directly beneath the scanner.

The scan positions were for the destructively sampled trees chosen so that each tree was scanned from at least two positions. Furthermore, the scanner position was chosen so that the top of each of the sample trees preferably was visible in at least one scan (Fig. 2, left). For the scanning of the sample plots five scans were performed at each plot. One scan from the plot center, and one scan from positions towards each of the cardinal directions from the center position (Fig. 2, right). Scan targets were used in order to register (merge) the scans correctly.

2.5 Calculations and analysis

Single-tree branch biomass predictions were derived using two approaches (outlined in Fig 3.). In the first approach a TLS–branch biomass model was used to predict ground reference values which were subsequently used in predictive ALS-branch biomass models (Fig. 3, right). This approach was compared to predictions made with ALS-predicted d_{bh} and the use of existing allometric biomass models (Fig. 3, left).

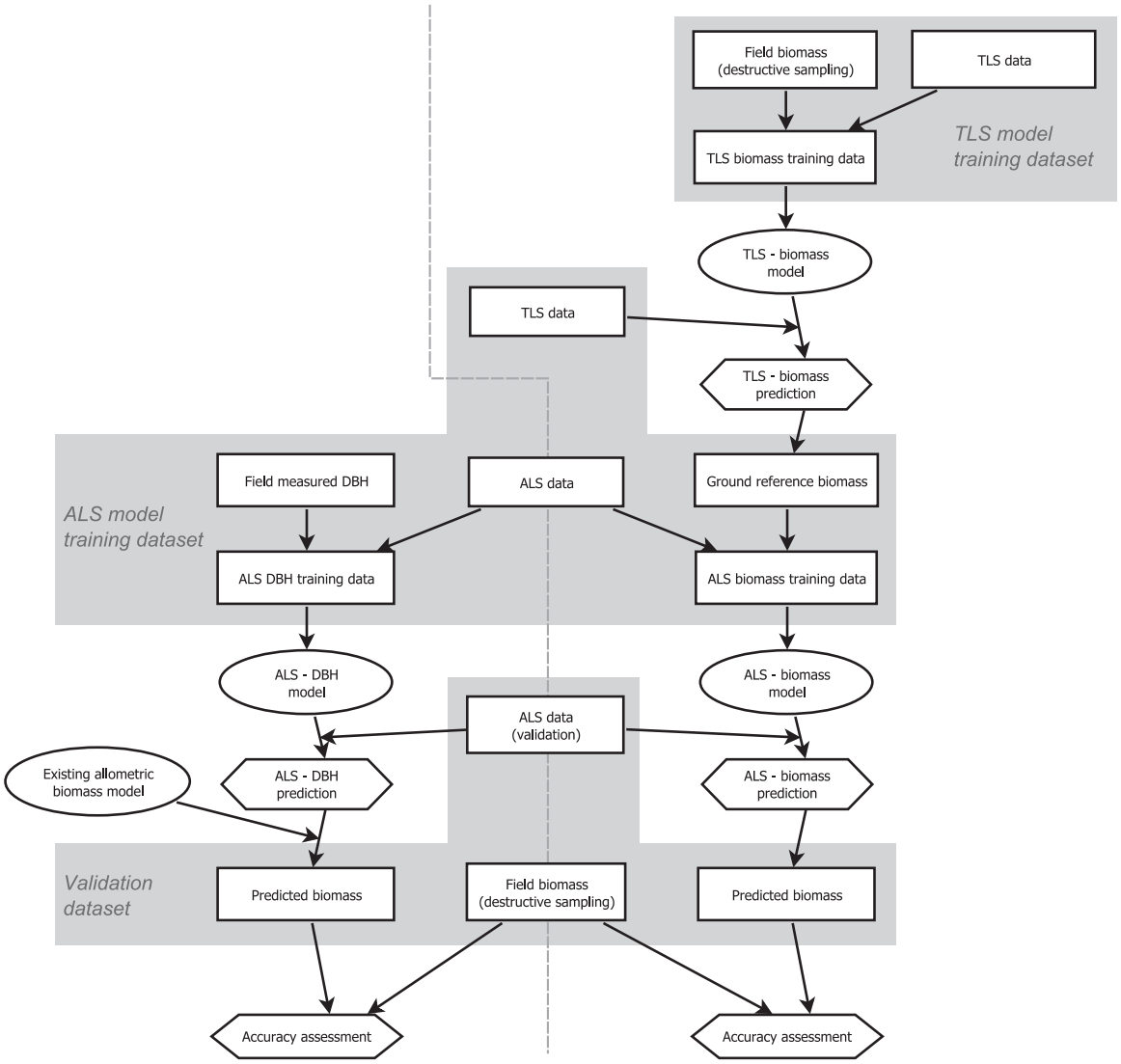

Fig. 3. Flowchart outlining the analysis. Rectangles represent data, ellipses models, and hexagons represent processes. Prediction of branch biomass with the use of TLS data (right side of dashed line). Prediction of branch biomass using ALS-predicted d_{bh} and existing allometric biomass models (left side of dashed line). The three datasets are indicated with a gray background (data from each set of trees are encompassed by a separate grey figure).

2.5.1 Single tree segmentation

In order to resemble an operational setting the ALS data were in the present study delineated into single tree segments using a fully automated segmentation algorithm (Ene et al., 2012). A marker-based watershed algorithm was used to delineate single tree segments from a canopy height model. An adaptive approach was used, utilizing a priori area-based stem number predictions to guide the delineation. Data from the sample plots (section 2.3.1) were used as training data in the prediction of stem numbers. The segmentation was carried out in a rasterized representation of the ALS data, but the crown segments used in the subsequent analysis was constructed as a convex hull in the horizontal plane of the ALS echoes within each raster segments (depicted in Figure 4). The reader is referred to Ene et al. (2012) for further details about the segmentation algorithm.

Automatic delineation of tree crowns will usually result in some erroneous segments (Vauhkonen et al., 2012). These can be caused by commission errors (segments of non-existing trees) or omission errors. Commission errors are typically caused by delineating large tree crowns into multiple segments, whereas trees omitted in the segmentation procedure are typically small or suppressed. Combining multiple adjacent crowns into one single segment will also lead to one or more trees being omitted. In the present study the ALS-derived segments were handled according to the following rules:

- Field measured trees were linked to ALS-derived crown segments if the field measured stem positions were inside the segment.
- Since the procedure allows for overlapping segments, a field tree with position inside more than one segment would be connected to the segment with the shortest distance from its centroid to the field tree position.
- ALS model training data: only trees with one single field measured tree position within a segment was used.
- Validation data: If more than one field measured tree was inside a segment the sum of the branch biomass for all the field trees positioned inside the segment was used as the reference branch biomass.
- Validation data: Segments encompassing trees that were not destructively sampled were excluded from the validation dataset since the reference biomass was not known for the whole segment.

Some of the rules are illustrated in Figure 4. In the present validation data material no segments did contain more than one destructively sampled tree, whereas three destructively sampled trees were within segments encompassing other trees, and these segments were therefore excluded from the validation dataset.

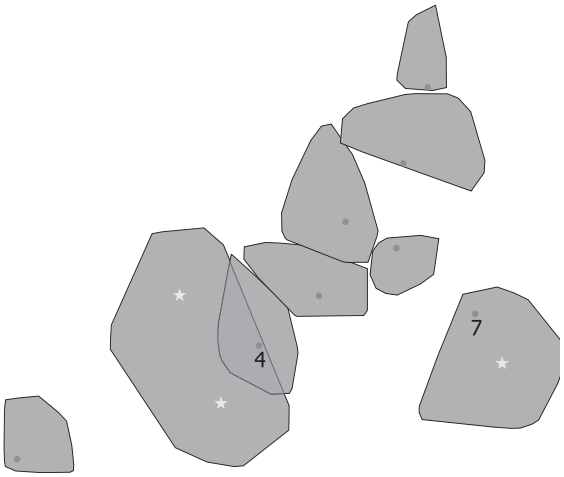

Fig. 4. Example of ALS-derived segments from the single tree segmentation for some of the sample trees. Field measured tree positions of destructively sampled trees (dots) and other trees (stars). Tree #4 was linked to the smallest of the two segments (see section 2.5.1). The segment containing tree #7 was excluded from the validation dataset since it also encompassed a tree without field measured branch biomass.

Table 2. Description of the ALS-derived variables. All echoes assigned to the specified tree and from all return categories are used if not stated otherwise. (Note: crown echoes are defined in the text, section 2.5.2).

Variable	Description
RBFvol	RBF-derived interpolated crown volume (see section 2.5.3).
alpvol	Three dimensional alpha shape volume of the crown echoes.[a]
chullvol	Three dimensional convex hull volume of the crown echoes.[b]
$crown_{ratio}$	Number of crown echoes to total number of echoes.
$first_{ratio}$	Number of first return echoes to total number of echoes.
$first_{ratio.crown}$	Number of first return crown echoes to total number of crown echoes.
cbh	Crown base height (see text, section 2.4.2).
$crown_{length}$	Crown length. The height from the CBH to the maximum echo height.
$D_{0.2}$	
$D_{0.4}$	
$D_{0.6}$	Density variables: Number of crown echoes above the height of the 0.2, 0.4, 0.6 and
$D_{0.8}$	0.8 fraction of the crown length, to the total number of crown echoes.
harea	Convex hull area of the horizontal projection of the crown echoes.[a]
H_{20}	
H_{40}	
H_{60}	Heights of the 20, 40, 60, 80, and 95 percentile of the echo height distribution.
H_{80}	
H_{95}	
h_{max}	Maximum echo height.
h_{cv}	Coefficient of variation for echo heights.
h_{diff}	Distance between the mean height of first return crown echoes and second return crown echoes.
h_{mean}	Mean echo height.
$h_{mean.crown}$	Mean crown echo height.
i_{mean}	Mean intensity of the first return echoes.
$i_{mean.crown}$	Mean intensity of the first return crown echoes.
$i_{ratio.all}$	Fraction of crown echoes with above mean intensity values.
$i_{ratio.first}$	Fraction of first return crown echoes with above mean intensity values.
i_{sd}	Standard deviation of the echo intensity.
$kurtosis_{crown}$	Kurtosis of the crown echo height distribution.[c]
pdensity	Point density: Number of echoes divided by the horizontal convex hull area of the echoes.
$skew_{crown}$	Skew of the crown echo height distribution.[c]
stemdist	Mean horizontal distance to stem position (planimetric coordinates of the highest echo in the crown) for the crown echoes.
$stemdist_{cv}$	Coefficient of variation of horizontal distance from crown echoes to stem position.
varea	Mean of the two convex hull areas of the crown echoes in the x,z-plane and the y,z-plane, respectively.[a]

[a] The *alphashape3d* package in R was used to compute the alpha shape volume.

[b] The *convhulln* function in the R package *geometry* was used to compute the convex hull.

[c] Kurtosis and skew as defined by Evans et al. (2009).

2.5.2 ALS-derived crown base height and crown echoes

For each individual tree segment the crown base height was estimated from the height of the ALS echoes, and echoes above the crown base height were denoted *crown echoes*. The method described by Solberg et al. (2006) was used: Crown base height was set at the height decile with the largest distance downwards to the next decile. The minimum crown base height was set at 0.85 m above ground.

2.5.3 ALS derived variables

In order to use the ALS data for prediction of branch biomass, variables were computed from the laser echoes allocated to each individual tree. In the present study three different crown volume variables were computed: (i) The volume of a three-dimensional (3D) convex hull of the crown echoes, (ii) a crown volume derived from the crown echoes by means of a radial basis function as described in by Hauglin et al. (2012b), and (iii) a volume derived from the 3D alpha shape of the crown echoes. The *alphashape3d* package in the statistical software R , version 2.14.0 (R Development Core Team, 2011) was used for the computations of the 3D alpha shape volume. A fixed alpha value of 0.5 m was used. The three crown volume variables were thus derived from different ways of constructing a surface from scattered data points. The reader is referred to Edelsbrunner & Mücke (1994) for a description of 3D alpha shapes and Kato et al. (2009) and Hauglin et al. (2012b) for a description of crown volume computed using radial basis functions.

The ALS derived features used as variables in the present study are listed and described in Table 2.

2.5.4 TLS derived variables

Features were extracted from the TLS data, and used to create variables included in a subsequent regression analysis. Hauglin et al. (2012a) found a strong relationship between TLS derived crown measurements and branch biomass. In the present study a corresponding set of TLS features were derived from manual measurements of individual tree crowns in the unified TLS data. The manual measurements were carried out using a 3D point cloud viewer with an additional two-dimensional plot of a horizontal and a vertical slice of the data. The horizontal and vertical slices were of 1m and 0.6m thickness, respectively. The crown projection onto the horizontal plane was measured as a polygon area, and the crown width as the horizontal distance from one side of the crown to the other. Both measurements were performed at heights corresponding to 10, 20, 40, 60, and 80 percent of the crown length. The crown length was calculated as the vertical distance from the crown base height and the highest recorded TLS laser echo attributed to the given tree. The crown base height was manually determined by visually inspecting the TLS data, and we defined it to correspond to the lowermost part of the lowermost living branch of the tree. The variables computed from the crown measurements are described and listed in Table 3.

Table 3. Description of the variables derived from crown measurements in the TLS data.

Variable	Description
CR_{length}	Crown length. Vertical distance from the crown base height to the highest laser echo assigned to the tree.
$CRA_{10,..,80}$	Crown projection area. Area of the crown projection measured at heights corresponding to 10, 20, 40, 60, and 80 percent of the crown length.
CRA_{sum}	Crown projection area sum. The sum of the area of the crown projection measurements at heights corresponding to 10, 20, 40, 60, and 80 percent of the crown length.
$CRW_{10,..,80}$	Crown width. The crown width measured at heights corresponding to 10, 20, 40, 60, and 80 percent of the crown length.
CRW_{sum}	Crown width sum. The sum of the crown width measurements at heights corresponding to 10,20,40,60 and 80 percent of the crown length.

2.5.5 Predictive models

In order to link branch biomass to the airborne and terrestrial laser scanner data, predictive regression models were fit to the training datasets. In the present study non-parametric random forest regression was applied alongside linear least squares models.

The principles behind random forest regression as it is commonly used today were introduced by Breiman (2001). In this machine learning algorithm a large number of binary regression trees are grown using bootstrap samples of the data. The predicted value is then averaged over the individual binary trees. So-called 'out-of-bag' prediction errors can be used to assess the prediction accuracy of a random forest model. In the present study the implementation of random forest in the *randomForest* package of the statistical software R was used, with default values for the *mtry* and *nodesize* parameters (number of variables/3 and 5, respectively). One thousand trees were grown for each model. Further details about the principles and use of the RF algorithm can be found in Breiman (2001) and Hastie et al. (2003).

In order to predict the ground reference branch biomass in the ALS training data, a TLS-based model was used. Following the approach described in Hauglin et al. (2012a) dry weight branch biomass was predicted using a random forest regression model with the TLS-derived explanatory variables described in Table 3. The TLS training data comprising 29 destructively sampled trees (section 2.3.2 and Table 1) were used to fit the TLS biomass model.

In a single tree ALS-based inventory a regression model or model chain is typically used when predicting biophysical properties for each individual tree. In the present study regression models were fit to the ALS training data to derive ALS biomass models. Several types of models and regression techniques have been used in previous studies, and in the present study random forest and linear least squares regression models were applied. The models were fit using the TLS estimated branch biomass as response variable (corresponding ALS – d_{bh} models were also fit, see next subsection).

2.5.6 Branch biomass predicted using existing allometric models

The proposed method was compared to an approach using ALS-predicted d_{bh} and existing allometric biomass models. First an ALS-model was used to predict d_{bh}, and then this ALS-predicted d_{bh} was used with existing allometric models to derive branch biomass (Fig.3, right). An existing species-specific allometric model by Marklund (1988) was used. The method described in this section corresponds to the approach used by Popescu (2007), predicting above-ground biomass. In the present study both a stepwise linear regression and a random forest regression model were used when predicting d_{bh} from the ALS-derived variables.

2.5.7 Final validation

The ALS models (section 2.5.5 and 2.5.6) were used to predict branch biomass for the 17 trees in the separate validation dataset. The predictions were then compared to accurate field measurements obtained by destructive sampling, and statistics computed to assess the prediction accuracy.

The coefficient of determination (R^2) of the models was inspected for an indication of the model fit. The R^2 values were computed for all the models as

$$R^2 = \rho(fitted\ values, observed\ values)^2, \quad (1)$$

where ρ is the Pearson's correlation.

The accuracy of the predictions was assessed by computing the root mean square error (RMSE) of the predicted values. The RMSE was computed as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(BR_i - \widehat{BR}_i)^2}{n}}, \quad (2)$$

where n is the number of trees and $\widehat{BR}$ is the predicted branch biomass.

Bias was quantified by

$$bias = \frac{\sum_{i=1}^{n}(BR_i - \widehat{BR}_i)}{n}, \quad (3)$$

and the significance of the difference between the observed and the predicted values was assessed with a paired t-test. In the present, study RMSE

and bias are reported in percent of the mean observed value, i.e.

$$RMSE\% = \frac{RMSE}{\frac{\sum_{i=1}^{n} BR_i}{n}} 100, \quad (4)$$

and

$$bias\% = \frac{bias}{\frac{\sum_{i=1}^{n} BR_i}{n}} 100. \quad (5)$$

3 Results

A random forest regression model linking branch biomass to TLS derived variables (Table 3) was fit to the TLS training dataset, consisting of 29 trees. The R^2 for this model was 0.80 and the out-of-bag prediction errors corresponded to an RMSE of 34.0%. The TLS model was used to predict ground reference branch biomass for the ALS model training dataset (Fig. 3, right).

Two models utilizing an initial set of all the ALS-derived variables were fit to the ALS training data using random forest and linear least squares stepwise regression techniques, respectively. In addition, three linear models with three different crown volume variables as single predictor variables were also fit. The R^2 values from the five models, ranging from 0.50 to 0.92 indicated moderate to good fit (Table 4 and Fig. 5).

Branch biomass was further predicted using the procedure described in subsection 2.5.6 where branch biomass was derived using ALS-predicted d_{bh} and existing allometric models. The R^2 values from the ALS – d_{bh} models indicated a good model fit, with 0.98 and 0.9 for the random forest and linear stepwise regression models respectively (Table 5 and Fig. 6).

Predictions from the ALS models were validated using the separate dataset consisting of 17 trees, whit branch biomass obtained through destructive sampling. In order to assess the prediction accuracy of the models, predicted branch biomass was compared to the observed values. The RMSE ranged from 32.4% to 38.8% for the ALS-models fit to training data with ground reference branch biomass derived with TLS (Table 4 and Fig. 7). For the corresponding predictions using existing allometric models and ALS-predicted d_{bh} the RMSEs were 41.5% and 44.5% when using the random forest and stepwise regression ALS-d_{bh} models, respectively (Table 5 and Fig. 8).

No significant bias could be observed for the predicted branch biomass in the present study (Table 4 and 5).

Table 4. Model and validation statistics for the ALS branch biomass models.
R^2 from the model-fitting (n=68), and RMSE% and bias% from the validation (n=17).

Explanatory variables	Regression technique[a]	Model fit	Validation	
		R^2	RMSE%	bias%
All	RF	0.92	38.8	13.3ns
All	STEPWISE	0.70	38.8	11.9ns
RBFvol	LS	0.50	32.4	13.5ns
alpvol	LS	0.57	38.3	16.7ns
chullvol	LS	0.53	34.2	12.3ns

Significance levels: ns: not significant ($p > 0.05$).
[a] RF = random forest regression, STEPWISE = stepwise linear least squares regression, LS = linear least squares regression

Table 5. Model and validation statistics for branch biomass predictions obtained with ALS-predicted d_{bh} and the use of existing allometric models. R^2 from the fitting of the ALS–d_{bh} models (n = 68), and RMSE% and bias% from the validation of the branch biomass predictions derived with the ALS-predicted d_{bh} and existing allometric models (n = 17).

Explanatory variables	Regression technique[a]	Model fit	Validation	
		R^2	RMSE%	bias%
All	RF	0.98	41.3	-12.0ns
All	STEPWISE	0.90	44.5	16.0ns

Significance levels: ns: not significant ($p > 0.05$).
[a] RF = random forest regression, STEPWISE = stepwise linear least squares regression

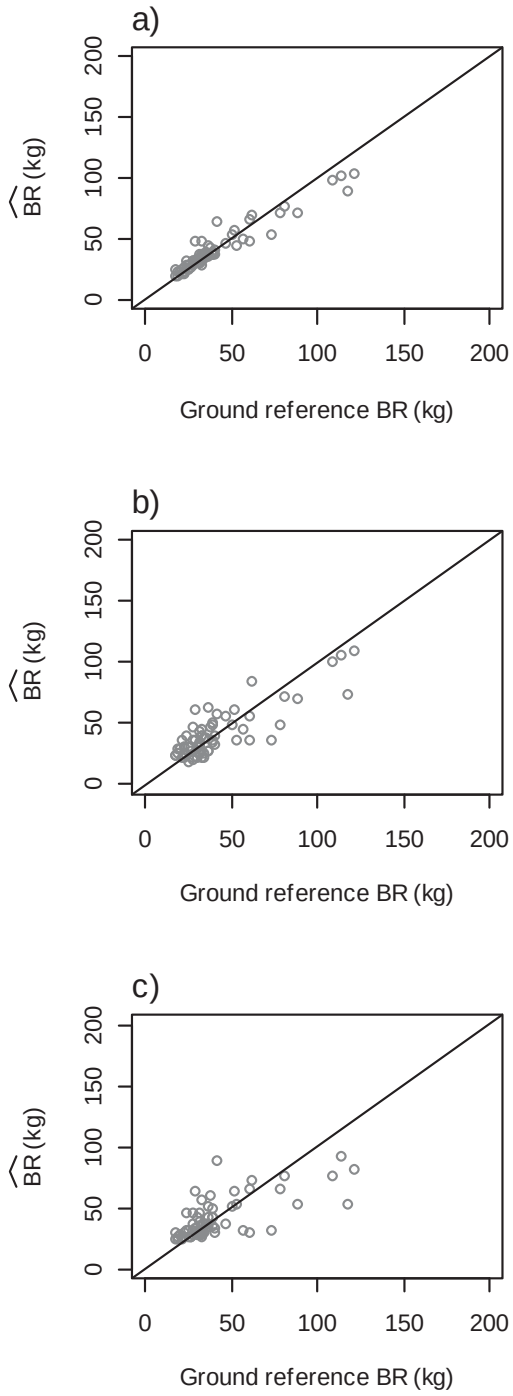

Figure 5. Predicted ($\widehat{BR}$) versus ground reference branch biomass (*BR*) for the ALS models in the model data with ground reference values obtained with TLS. Random forest regression (a), linear least squares stepwise regression (b), and linear least squares regression with radial basis functions-derived crown volume as a single explanatory variable (c).

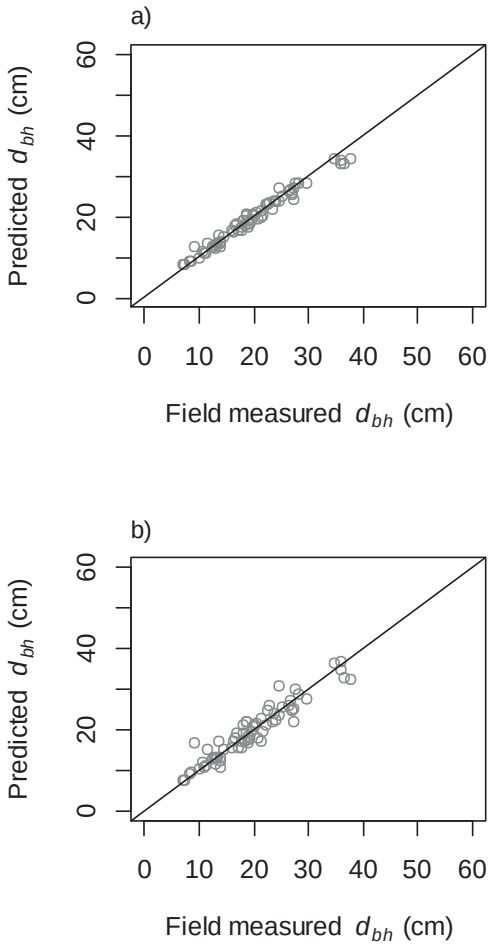

Figure 6. Predicted versus field measured d_{bh}. (a) Random forest regression and (b) linear least squares stepwise regression.

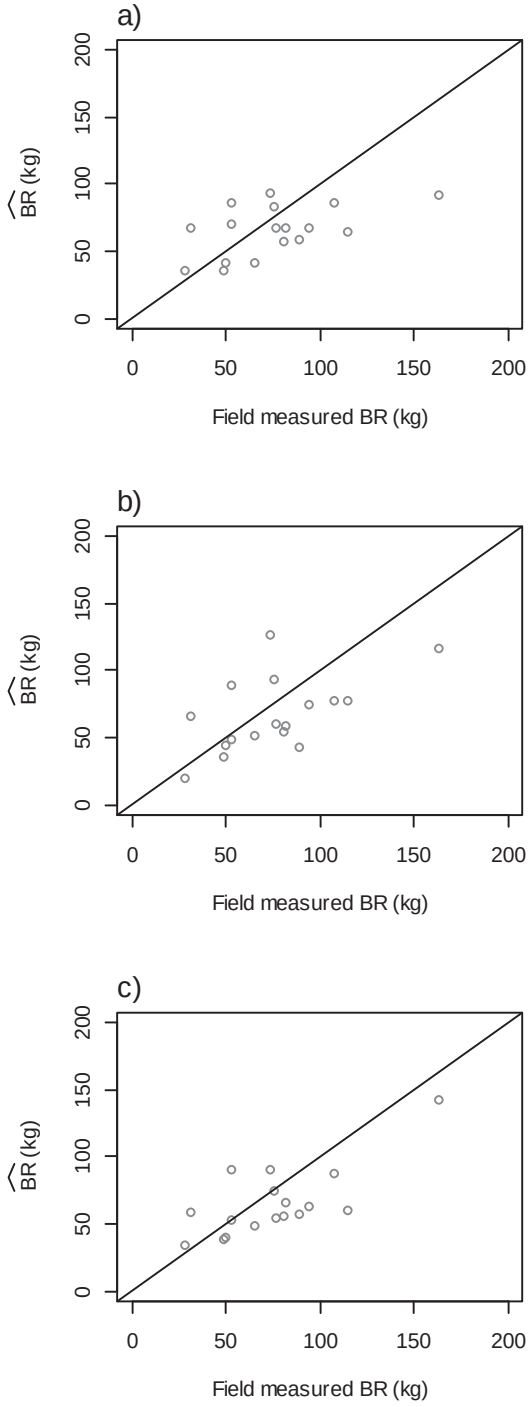

Figure 7. Predicted ($\widehat{BR}$) versus field measured branch biomass (*BR*) for the ALS branch biomass models in the validation dataset. (a) Random forest regression, (b) linear least squares stepwise regression, and (c) linear least squares regression with the radial basis function-derived crown volume as single explanatory variable.

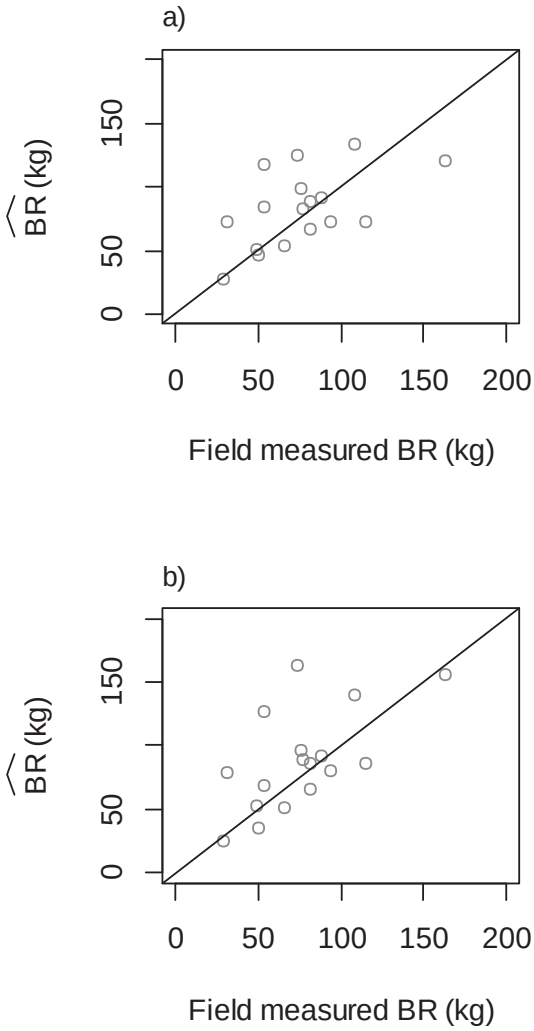

Figure 8. Predicted ($\widehat{BR}$) versus field measured branch biomass (*BR*) for the predictions derived with ALS-predicted d_{bh} and existing allometric models in the validation dataset (see text, section 2.5.6). (a) Random forest regression and (b) stepwise linear regression.

4 Discussion and conclusion

The present study investigated the prediction of single-tree branch biomass using ALS, with training data based on ground reference values obtained with TLS. The prediction accuracy was assessed using a dataset with branch biomass accurately measured through destructive sampling, and compared to predictions made without the use of TLS-derived ground reference values. The results shows that good single-tree branch biomass estimates of Norway spruce can be obtained from ALS models using training data with ground reference values derived with TLS.

TLS data were in the present study used to predict ground reference values of branch biomass. The accuracy of the predictions from the TLS-biomass model is in line with the results in Hauglin et al. (2012a).

ALS models fit to data with TLS-derived ground reference values yielded predictions with higher accuracies than predictions of branch biomass derived with existing allometric models and ALS-predicted d_{bh}. In other words, the use of training data obtained with TLS improved the ALS branch-biomass predictions. This result is in line with findings in Hauglin et al. (2012a) where TLS-based models were found to predict branch biomass with a higher accuracy than existing allometric models based on d_{bh} and tree height. The result is also in line with findings in Hauglin et al. (2012b) where ALS-models based on accurately measured branch biomass were found to provide predictions with higher accuracy than existing allometric models and actual field measurements of d_{bh} and tree height. Popescu (2007) estimated single-tree total aboveground biomass of pine trees using a two-step procedure with ALS-predicted d_{bh} and existing allometric models, corresponding to the one used in the present study (section 2.5.6). Popescu reported an RMSE of 47%. The prediction accuracy obtained for branch biomass in the present study is higher, but the two studies are not entirely comparable. They differ both in terms of tree species and in that different biomass components are estimated, i.e., total above-ground biomass versus branch biomass. Most importantly: Accurately measured biomass was not used for validation by Popescu (2007). Straub and Koch (2011) reported an RMSE of 24% when estimating stem volume in Scots pine using ALS. The higher accuracy than in the present study might be due to a strong relationship between stem volume and height, which is known to be closely related to ALS-derived variables (Persson et al., 2002). The relationship between d_{bh} and stem volume in Norway spruce is known to be stronger than the one between d_{bh} and branch biomass (Vestjordet, 1967; Marklund, 1988). The same might be true for height in relation to stem volume and branch biomass.

The existing allometric model used in the present study was developed from a large dataset covering similar forest-types, but from outside the study area (Marklund, 1988). This might lead to biased predictions, if the allometry in the study area deviates from that used to develop the allometric models. Since costly and time-consuming fieldwork often is involved when developing allometric models, the use of regional or national models is common (Jenkins et al., 2003; Zianis et al., 2005).

In the present study only one tree species was considered. In an operational forest inventory in a Nordic boreal forest typically two additional species or groups of species are considered, namely Scots pine and deciduous trees. In such a single tree inventory one is then faced with the task of classifying the single tree segments into one of several classes of species. This will lead to classification errors which would have affected the prediction accuracy. How these errors would affect the prediction accuracy of branch biomass following the approaches described in the present study could be subject to further investigations.

The derivation of TLS features in the present study included several manual and semi-manual operations, such as for example setting the crown base height and measuring the crown width (Hauglin et al., 2012a). These manual operations were conducted in such a way that they in principle could have been done automatically, solely using the TLS data. Methodology and specific algorithms to automatically perform these tasks are to our knowledge not readily available and expected errors and accuracy of such an automated approach should be further investigated. The TLS model in the present study was trained using data derived through destructive sampling. In an operational inventory high costs will prohibit the collection of such data. To use

existing TLS models might be a viable solution, and since the crown properties are measured directly, the model might be less sensitive to regional or local stand-level allometric differences. The transferability of TLS models between regions – possibly with data from different instruments – could thus be subject to further research.

In the present study, random forest regression was used alongside linear least squares regression. Random forest was used because it is relatively simple, yet has been found to perform well in comparison to other regression techniques, with little tuning required (Hastie et al., 2003). It is robust with respect to noisy variables and do not have problems with inter-correlated variables and multicollinearity, which means that the variable selection deemed necessary in the least squares regression is avoided (Biau, 2012). One feature of a random forest regression model is that it will not yield extrapolated values. This is to some extent a desirable property, since it ensures that only sound values may be predicted. A consequence is that the predicted values at the extremes of the range are drawn towards the mean, thus introducing a bias in those predictions. In the present study equal prediction accuracies were obtained with the random forest model and the stepwise linear regression model, using the same set of explanatory variables.

The number of trees on sample plots from the study area was used as auxiliary information in the segmentation process. Single-tree segmentation can be carried out without the use of this auxiliary data, if it is not available. This was however not done in the present study, and thus the effect on the accuracy of the predicted branch biomass with such a segmentation was not assessed.

In the data material used for validation, no segment contained more than one destructively sampled tree. The effect of predicting branch biomass for segments consisting of more than one tree was therefore not assessed, although such segments are likely to occur in an ALS-based single-tree inventory (Vauhkonen et al., 2012). With the approach described in section 2.5.6, the ALS model must in such multi-tree segments predict large d_{bh} values to produce accurate branch biomass estimates. If not, branch biomass will be underestimated for these segments. The ALS-derived crown volume of a multi-tree segment might to a larger extent maintain the same relationship to the actual branch biomass as in single-tree segments than the ALS-predicted d_{bh}. If this is the case, then the presence of multi-tree segments would increase the difference between the two approaches discussed in the present paper. More research is however needed to clarify this.

To conclude, we have shown that good predictions of single-tree branch biomass of Norway spruce can be achieved by using a combination of ALS and TLS data. The approach resembles a methodological framework where TLS is used to obtain the ground reference values, which is subsequently used as part of the training data for ALS-based predictive models. The successful implementation in the present study suggests that the proposed framework can be used as a general approach when deriving ground reference values for single-tree ALS models. The accuracy of the predictions were higher than for predictions derived with ALS-predicted d_{bh} and existing allometric models, in other words – the use of TLS to derive the ground reference branch biomass improved the prediction accuracies.

Acknowledgments

We wish to thank Blom Geomatics, Norway and Geoplan 3D, Norway for providing and partly processing the laser data. We further wish to thank Ms. Janka Dibdiakova at The Norwegian Forest and Landscape Institute for doing the lab work. We are also grateful to Mr. Arne Drømtorp, Mr. Leif Kjøstelsen, and Mr. Espen Martinsen (The Norwegian Forest and Landscape Institute); Mr. Vegard Lien, Mr. John Gunnar Dokk, and Mr. Tord Storbækken (Norwegian University of Life Sciences) who took part in the field work.

References

Andersen, H.-E., Clarkin, T., Winterberger, K., Strunk, J., 2009. An Accuracy Assessment of Positions Obtained Using Survey- and Recreational-Grade Global Positioning System Receivers across a Range of Forest Conditions within the Tanana Valley of Interior Alaska. Western Journal of Applied Forestry 24, 128–136.

Axelsson, P., 2000. DEM generation form laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing XXXIII, 110–117.

Biau, G., 2012. Analysis of a Random Forest Model. Journal of Machine Learning Research 13, 1063–1095.

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. International Archives of Photogrammetry and Remote Sensing 3, 50–55.

Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L., Kimes, D.S., 2008. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. Remote Sensing of Environment 112, 3876 – 3890.

Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., 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.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32.

Edelsbrunner, H., Mücke, E., 1994. 3-Dimensional Alpha-Shapes. Acm Transactions on Graphics 13, 43–72.

Ene, L., Næsset, E., Gobakken, T., 2012. Single tree detection in heterogeneous boreal forests using airborne laser scanning and area-based stem number estimates. International Journal of Remote Sensing 33, 5171–5193.

Hasegawa, H., Yoshimura, T., 2007. Estimation of GPS positional accuracy under different forest conditions using signal interruption probability. Journal of Forest Research 12, 1–7.

Hastie, T., Tibshirani, R., Friedman, J., 2003. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 1st ed. 2001. Corr. 3rd printing. ed. Springer. pp. 552.

Hauglin, M., Astrup, R., Gobakken, T., Næsset, E., 2012a. Estimating single-tree branch biomass of Norway spruce with terrestrial laser scanning.

Hauglin, M., Dibdiakova, J., Gobakken, T., Næsset, E., 2011. Estimating single-tree branch biomass of Norway spruce by airborne laser scanning, in: Proceedings of Silvilaser 2011. Presented at the Silvilaser 2011, Hobart, Australia.

Hauglin, M., Dibdiakova, J., Gobakken, T., Næsset, E., 2012b. Estimating single-tree branch biomass of Norway spruce by airborne laser scanning.

Hauglin, M., Gobakken, T., Lien, V., Bollandsås, O.M., Næsset, E., 2012c. Estimating potential logging residues in a boreal forest by airborne laser scanning. Biomass and Bioenergy 36, 356–365.

Holopainen, M., Vastaranta, M., Kankare, V., Räty, M., Vaaja, M., Liang, X., Yu, J., Hyyppä, H., Hyyppä, R., 2011. Biomass estimation of individual trees using stem and crown diameter TLS measurements. ISPRS Workshop on Laser scanning XXXIII.

Hyyppä, J., Kelle, O., Lehikoinen, 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.

Jenkins, J., Chojnacky, D., Heath, L., Birdsey, R., 2003. National-scale biomass estimators for United States tree species. Forest Science 49, 12–35.

Kato, A., Moskal, L.M., Schiess, P., Swanson, M.E., Calhoun, D., Stuetzle, W., 2009. Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. Remote Sensing of Environment 113, 1148–1162.

Latifi, H., Nothdurft, A., 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, 395–407.

Lovell, J.L., Jupp, D.L.B., Newnham, G.J., Culvenor, D.S., 2011. Measuring tree stem diameters using intensity profiles from ground-based scanning lidar from a fixed viewpoint. ISPRS Journal of Photogrammetry and Remote Sensing 66, 46–55.

Maltamo, M., Bollandsås, O.M., Vauhkonen, J., Breidenbach, J., Gobakken, T., Næsset, E., 2010. Comparing different methods for prediction of mean crown height in Norway spruce stands using airborne laser scanner data. Forestry 83, 257–268.

Maltamo, M., Packalén, P., Kallio, E., Kangas, J., Uuttera, J., Heikkilä, J., 2011. Airborne laser scanning based stand level management inventory in Finland, in: Proceedings of Silvilaser 2011. Presented at the Silvilaser 2011, Hobart, Australia.

Marklund, L.G., 1988. Biomass functions for pine, spruce and birch in Sweden (Report No. 45). Swedish University of Agricultural Sciences, Umeå.

Moskal, L.M., Zheng, G., 2012. Retrieving Forest Inventory Variables with Terrestrial Laser Scanning (TLS) in Urban Heterogeneous Forest. Remote Sensing 4, 1–20.

Næsset, E., 1999. Point accuracy of combined pseudorange and carrier phase differential GPS under forest canopy. Canadian Journal of Forest Research 29, 547–553.

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., Bollandsås, O.M., Gobakken, T., 2005. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. Remote Sensing of Environment 94, 541 – 553.

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.

Næsset, E., Gobakken, T., Solberg, S., Gregoire, T.G., Nelson, R., Ståhl, G., Weydahl, D., 2011. Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: A case study from a boreal forest area. Remote Sensing of Environment 115, 3599–3614.

Persson, A., Holmgren, J., Söderman, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. Photogrammetric Engineering and Remote Sensing 68, 925–932.

Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne lidar. Biomass and Bioenergy 31, 646–655.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. Vienna, Austria.

Roberts, S.D., Dean, T.J., Evans, D.L., McCombs, J.W., Harrington, R.L., Glass, P.A., 2005. Estimating individual tree leaf area in loblolly pine plantations using LiDAR-derived measurements of height and crown dimensions. Forest Ecology and Management 213, 54–70.

Räty, M., Kankare, V., Yu, X., Holopainen, M., Vastaranta, M., Kantola, T., Hyyppä, J., Viitala, R., 2011. Tree biomass estimation using ALS features, in: Proceedings of Silvilaser 2011. Presented at the Silvilaser 2011, Hobart, Australia.

Simonse, M., Aschoff, T., Spiecker, H., 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. Umeå, Sweden, pp. 251– 257.

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, 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, 929–944.

Vauhkonen, J., 2010. Estimating crown base height for Scots pine by means of the 3D geometry of airborne laser scanning data. Int. J. of Remote Sensing 31, 1213–1226.

Vauhkonen, J., Ene, L., Gupta, S., Heinzel, J., Holmgren, J., Pitkanen, J., Solberg, S., Wang, Y., Weinacker, H., Hauglin, K.M., Lien, V., Packalen, P., Gobakken, T., Koch, B., Naesset, E., Tokola, T., Maltamo, M., 2012. Comparative testing of single-tree detection algorithms under different types of forest. Forestry 85, 27–40.

Vauhkonen, J., Korpela, I., Maltamo, M., 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.

Vestjordet, E., 1967. Functions and tables for standing trees. Norway spruce. Reports of the Norwegian Forest Research Institute 22, 539–574.

Yao, T., Yang, X., Zhao, F., Wang, Z., Zhang, Q., Jupp, D., Lovell, J., Culvenor, D., Newnham, G., Ni-Meister, W., Schaaf, C., Woodcock, C., Wang, J., Li, X., Strahler, A., 2011. Measuring forest structure and biomass in New England forest stands using Echidna ground-based lidar. Remote Sensing of Environment 115, 2965–2974.

Zhao, K., Popescu, S., Meng, X., Pang, Y., Agca, M., 2011. Characterizing forest canopy structure with lidar composite metrics and machine learning. Remote Sensing of Environment 115, 1978–1996.

Zianis, D., Muukkonen, P., Mäkipää, R., Mencuccini, M., 2005. Biomass and Stem Volume Equations for Tree Species in Europe. Silva Fennica 63.

Ørka, H.O., 2011. Improving forest inventory and monitoring by combining remotely sensed three-dimensional and spectral information. Norwegian University of Life Sciences. PhD Thesis no. 2011:22.

ISBN 978-82-575-1090-9
ISSN 1503-1667

Norwegian University of Life Sciences
NO–1432 Ås, Norway
Phone +47 64 96 50 00
www.umb.no, e-mail: postmottak@umb.no