A method for using harvester data in airborne laser prediction of forest variables in mature coniferous stands

Jon Söderberg

Arbetsrapport 443 2015
Master thesis in Forest Management 30hp A2E
Master of Science in Forestry

Handledare:
Jörgen Wallerman

Sveriges lantbruksuniversitet
Institutionen för skoglig resurshushållning
901 83 UMEÅ
www.slu.se/srh
Tfn: 090-786 81 00

ISSN 1401–1204
ISRN SLU–SRG–AR–443–SE

A method for using harvester data in airborne laser prediction of forest variables in mature coniferous stands

Jon Söderberg

Keywords: raster, segmentation, ALS, area based method

Master thesis in Forest Management at the Dept. of Forest Resource Management
Master of Science in Forestry (SLU, Sweden), EX0768, 30 credits
Supervisor: Jörgen Wallerman, SLU, Dept. of Forest Resource Management, Remote sensing
Examiner: Mats Nilsson, SLU, Dept. of Forest Resource Management, Remote sensing
External supervisors: Anders Almäng, Sveaskog & Johan J. Möller, Skogforsk

Sveriges lantbruksuniversitet
Institutionen för skoglig resurshushållning
Utgivningsort: Umeå
Utgivningsår: 2015

ISSN 1401–1204
ISRN SLU–SRG–AR–443–SE

Preface

This study is a Master of Science thesis in Forestry, at the Department of Forest Resource Management, Swedish University of Agricultural Sciences (SLU), Umeå, Sweden. This study is supported by Sveaskog and Skogforsk. I would like to thank the following people for their assistance during this process;

Anders Almäng, Sveaskog
Johan J Möller, Skogforsk
Peder Axensten, SLU
Kenneth Olofsson, SLU
Jonas Bohlin, SLU
Mattias Nyström, SLU
Jonas Jonzén, SLU

I would like to express my gratitude to my supervisor, Jörgen Wallerman, for all his patience and guidance through this long process. Also I would like to thank Ljungbergs fonden for funding Ljungbergslaboratoriet, the new remote sensing lab at SLU, Umeå, without which this thesis would not have been possible.

Umeå, February 2015
Jon Söderberg

Sammanfattning

Behovet av tillförlitlig skoglig information för planering i skogsbruket har resulterat i en utveckling av olika inventeringsmetoder. Kostnaden styr ofta valet av inventeringsmetod och frekvensen av inventeringar, vilket påverkar kvaliteten på den information som finns tillgänglig för skogsförvaltare. Objektivt inventerade provytor har varit vanliga som referensdata för att skatta skogliga variabler från fjärranalysdata. Ett alternativ till fält inventering, som ännu inte har undersökts i nämnvärd utsträckning är skördardata.

Syftet med denna studie var att utveckla en metod för att skatta skogliga variabler med flygburen laserskanningsdata och skördaruppgifter och utvärdera skattningsnoggrannheten i metoden. Studien begränsades till volym/ha, aritmetisk medehöjd och aritmetisk medeldiameter i brösthöjd. Den area baserad metoden användes och skattningar med både regressionsanalys och k-MSN utvärderades.

Resultaten av denna studie var beståndsvisa skattningar med ett relativt medelfel för volym, höjd och diameter på 11%, 5% och 8% . Tidigare studier i Skandinavien har med hjälp av flygburen laserskanningsdata och fältytor, skattat bestånds volym, grundytevägd medelhöjd och grundytevägd medeldiameter med ett relativt medelfel på 11-14%, 3-6% och 9-13% för respektive variabel (Næsset 2007). Denna studie visar att bra skattningar kan uppnås utan kunskap om exakta trädpositioner genom att använda tillgängliga skördardata från befintliga avverkningar och flygburen laserskanningsdata från Lantmäteriets nationella projekt. Resultaten visar med andra ord att skördardata kan vara en användbar källa till referensdata för fjärranalystillämpningar.

Nyckelord: raster, segmentering, ALS, areametoden

Abstract

The need for reliable information for forestry planning has resulted in the development of different inventory methods. Cost often regulates the choice of inventory method and the frequency of inventories, which affects the quality of the information available to the forest managers. Objectively surveyed field plots have been common reference data for predicting forest variables with remote sensing data. An alternative, which has not yet been extensively explored, is data collected from mechanical harvesting operations.

The purpose of this study was to develop a method for predicting forest variables with airborne laser scanning (ALS) data and harvester data, and to evaluate the accuracy of the method. The study was limited to volume, arithmetic mean height and arithmetic mean diameter at breast height. The area based method was used in this study, prediction with both regression analysis and k-MSN was evaluated.

The results were stand level predictions with a relative RMSE of 11%, 5% and 8% for volume, height and diameter respectively. Previous Scandinavian studies have, using ALS data and training plots, predicted stand volume, basal area weighted height and basal area weighted diameter with a relative RMSE of 11-14%, 3-6% and 9-13%, respectively (Næsset, 2007).

Good predictions can be achieved without knowledge of precise tree positions by using available data from actual harvester operations and ALS data from Lantmäteriets national scanning. The results suggest that harvester data can be a useful source of training data for remote sensing applications.

Keywords: raster, segmentation, ALS, area based method

Table of Contents

1. Introduction

1.1 Background

Sustainable forest management is the driving force behind forest management planning (Jonsson *et al.*, 1993). The limited forest resource creates a conflict between present and future needs, forestry management is therefore, essentially an instrument to shape the future (Jonsson *et al.*, 1993; Duvemo & Lämås, 2006). Forest management planning in Swedish companies is usually divided into strategic, tactical and operational planning (Duvemo & Lämås, 2006; Bredström *et al.*, 2010). Strategic planning concerns long term policies for timber production over the entire forest holding (Jonsson *et al.*, 1993; Duvemo & Lämås, 2006), while tactical and operational planning concern application of silvicultural treatments (Jonsson *et al.*, 1993; Karlsson *et al.*, 2006). In tactical forestry planning, information about all stands within the management region is required (Duvemo *et al.*, 2014). Accurate information of the current state of the forest resource is important for planning silviculture measures (Barth, 2008; Nordström & Möller, 2009). Poor information leads to sub-optimal scheduling of silvicultural treatments, which frequently result in less economic revenue for the forest owner (Mäkinen *et al.*, 2010; Duvemo *et al.*, 2014). The need for reliable forest information for decision-making has led to the development of different inventory methods.

Inventory of forest estates are often a sound investment (Sonesson *et al.*, 2008). However, the acquisition of dependable forestry data to cover planning needs are associated with substantial cost (Duvemo & Lämås, 2006). On the other hand low quality information on a stand level has a large negative impact on the outcome of yield calculations and planning in final harvest operations (Nordström & Möller, 2009).
Forest companies must therefore balance the need for enhanced decision making and the costs of data acquisition (Duvemo & Lämås, 2006). Unfortunately, the cost of information collection is more tangible for forest owners than the benefits (Sonesson *et al.*, 2008). This often influences the regularity of inventories and the choice of inventory method, which in turn affects the quality of the information available to the forest managers.

Inventory methods are categorized as subjective or objective. With subjective inventory methods forest variables are estimated or measured by a surveyor in locations believed to be representative of the stand (Ståhl, 1992). The main reasons for using subjective field inventory methods is because they are cheap and fast, and the available objective field inventory methods are too expensive to cover all stands (Holmgren, 2004). Subjective inventories and estimates often have large systematic errors, the magnitude of these errors varies greatly between surveyors (Ståhl, 1992; Sonesson *et al.*, 2008). Systematic underestimation of stand volume up to 20% are common for data assessed with subjective methods (Sonesson *et al.*, 2008). Mean tree height estimations have standard errors of around 10%, mean tree diameter 10-12% and volume 15-25% with subjective methods (Ståhl, 1992). Objective field inventory methods are characterized by measurements taking place in objectively pre-selected field plots and following strict protocols (Ståhl, 1992). Ideally, objective methods produce results free from systematic errors and it is possible to evaluate the accuracy of the collected information (Ståhl, 1992). One way to cost-efficiently acquire objective data may be by using remote sensing.

The concept of remote sensing hinges on the ability to obtain information about an entity through analysis of data acquired by a sensor which is not in contact with the examined entity (Lillesand *et al.*, 2008). Commonly forest variables are predicted from remote sensing data using objective field samples called training data to model the relationships to the physical world (Holmgren, 2004; Barth, 2008; Lillesand *et al.*, 2008).

In forestry applications, many different technologies have been used for obtaining remote sensing data. Stereo-interpretation of aerial photography has been a historically important source of forest information (Lillesand *et al.*, 2008). Other methods rely on a more active approach, with radar and Light detection and ranging (LIDAR) technology the sensor emits energy pulses and receives return responses which are interpreted into information (Lillesand *et al.*, 2008). By mounting a scanning LIDAR in an airplane, i.e. airborne laser scanning (ALS), data from large areas can be acquired swiftly. There is also a ground-based system comparable to ALS, which is called terrestrial laser scanning (TLS). In 2009 Lantmäteriet, the Swedish national land survey, began a ALS project with national coverage, the ambition was to create a national digital elevation model (DEM) with a standard deviation of less than 0.5 meters and to have completed the project by 2015 (Lantmäteriet, 2014). As a byproduct, the ALS data collected by Lantmäteriet contains information regarding the state of the Swedish forests. With the emergence of the detailed DEM from Lantmäteriet, a whole new level of accuracy can be achieved with different sources of remote sensing data. Furthermore, advances in digital camera technology and image analysis algorithms have led to rising interest in digital photogrammetry, which provides opportunities for 3D modelling, similar to ALS (Maltamo *et al.*, 2006b; Bohlin *et al.*, 2012; Vastaranta *et al.*, 2013). ALS has received a lot of attention from the forest industry and the research community due to the wealth of information which can be gleaned from the data.

The forest company Sveaskog is Sweden’s largest forest owner, with over four million hectares, around 700 employees and a turnover of over 6 billion SEK annually (Sveaskog, 2014). Sveaskog is interested in using the data collected by their mechanical harvesting operations, in conjunction with ALS data, to improve predictions of upcoming harvesting operations, to improve harvesting planning, to simulate bucking results and to improve the control of the timber flow.

1.2 ALS data

LIDAR is an active remote sensing system which sends out light pulses and measure the time it takes for the reflected light pulse to return (Lillesand *et al.*, 2008). In ALS, between one and five reflections is usually recorded from each emitted light pulse (Means *et al.*, 2000). The first return is generally reflected from the top of the tree canopy and the last return from the ground. The millions of returns recorded from an ALS flight describe the forest and ground in three dimensions (3D) (Axelsson, 2000; Means *et al.*, 2000). Such 3D point clouds are referenced to the earth’s surface through a geo-referenced coordinate system. In Sweden the system SWEREF99 TM is commonly used. However, the height is measured in relation to the sea level (roughly) and to create useful metrics which actually describe the forest, the height must be related to the ground level (Bohlin *et al.*, 2012).

ALS is due to lower flight altitude, less affected by cloud cover, which frequently impairs satellite and aerial photography at boreal latitudes (Nilsson, 1997), nor does it require daylight to operate. Data from ALS have a shelf-life of five years for forestry planning purposes, according to Sonesson *et al.* (2008). It is more expensive than aerial photography due to the lower flight altitude required, meaning ALS cover less ground area per flight hour (Vastaranta *et al.*, 2013). With increased flight altitude, more ground can be covered but the density of returns decreases. This may reduce the accuracy of predictions significantly (Magnusson *et al.*, 2007), but with a density of 0.5-1 returns per m^2, as in the Lantmäteriet's scanning, the influence on prediction accuracy is minor (Holmgren, 2004; Maltamo *et al.*, 2006a).

The most common method used to predict forest variables with ALS data is the area based approach, which converts 3D point clouds into metrics useful as independent variables in predicting forest characteristics (Næsset, 2002; Næsset *et al.*, 2004). To be able to extract useful height information from the returns, the ALS data need to be converted to height relative to the ground surface, this process is called normalization (Næsset, 2002; Olofsson & Holmgren, 2014). The ground surface is identified from last returns using a complex algorithm and the returns receive a new height value relative to the ground (Axelsson, 2000; Næsset, 2002). A raster is applied to the area of interest, and then metrics related to the forests characteristics can be calculated for each cell. By combining raster cell metrics and field sampling data, models can be constructed and used to predict forest variables on a cell-level. These cell predictions are commonly aggregated to larger units, such as forest stands. An alternative to raster is segmentation, which partition an area based on pre-defined ALS metrics (Olofsson & Holmgren, 2014). This results in predictor units, i.e. segments, which have low internal variation in regard to the chosen metrics. Experiences across the boreal region show that tree height and volume predictions with the area based method performs equally, or better than any other remote sensing method (Hyyppä *et al.*, 2008; Maltamo *et al.*, 2011). With the area based method, even at low densities of light pulses (~1 pulse/m^2), ALS can be used predict stand characteristics such as; average volume, mean basal area, basal area weighed mean diameter and basal area weighed mean height (Næsset, 2002; Barth, 2008). One weakness of the area based method is the lack of reliable species specific predictions (Maltamo *et al.*, 2011).

Prediction of forest variables using data from ALS significantly improves the reliability of analysis and planning in forest resource management (Sonesson *et al.*, 2008; Barth *et al.*, 2012). ALS is useful for large area inventories at least in coniferous boreal landscape, while deciduous forest is more problematic (Næsset *et al.*, 2004). Gathering ALS data during leaf-off season is one approach to improve the accuracy of timber volume predictions, especially in deciduous and mixed stands (Næsset *et al.*, 2004). Another approach is to compliment the ALS with aerial photography to determine species mixture (Packalén & Maltamo, 2007; Tuominen & Haapanen, 2011).

Commonly, linear regression analysis is used for area based prediction for forest variables. An alternative to regression analysis is k-MSN, which is a non-parametric nearest neighbor method (Maltamo *et al.*, 2006a). In k-MSN prediction each raster cell or element will receive variable values based on the weighted average of a number (*k*) of most similar neighboring elements from the training data (Packalén & Maltamo, 2007; Nordkvist *et al.*, 2013). The similarity is measured through canonical correlation, which simplified, means that the similarity is measured in the modeled forest variables instead of ALS metrics (Moeur & Stage, 1995). For forest management, using a k-MSN approach in conjunction with remote sensing data can be an economical and objective way to get accurate predictions of stand characteristics (Packalén & Maltamo, 2007). ALS data is superior to other data sources for predicting stand volume with the k-MSN method (Maltamo *et al.*, 2006b).

The positional accuracy of the training data set is very important for accurate predictions with ALS data (Barth, 2008; Gobakken & Næsset, 2009). Gobakken & Næsset (2009) studied the effect of positional errors of field plots in regard to ALS predictions and found that larger plot sizes produced less deviance, the accuracy of height predictions were not severely affected by a positional error of less than 5 m, but basal area and volume predictions were more sensitive to positional errors. Equally important is knowledge of the position and orientation of the airborne remote sensing instrument (Hyyppä *et al.*, 2008). The instruments used to track the position and rotation of the sensor in relation to the ground are; differential global positioning system (DGPS) and inertia measurement unit (IMU) (Hyyppä *et al.*, 2008; Lillesand *et al.*, 2008). Using the area based method, stand total stem volume predictions with ALS has a RMSE of 11-14%, while basal area weighted mean tree height can be predicted with a RMSE of 3-6% and basal area weighted mean diameter with a RMSE of 9-13% in Scandinavian studies (Næsset, 2007).

1.3 Harvester data

The prediction of forest variables has to date been primarily dependent on field inventories, but forestry technology advances has introduced new tools. One new source of information comes from the harvester-machines, that record data for each processed log (Arlinger *et al.*, 2003; Rasinmäki & Melkas, 2005). This information is saved and uploaded through the mobile phone network to *Skogsbrukets Datacentral* (SDC), the forest industry's IT company, which daily receives data from 1400 harvesters all over Sweden (SDC, 2014). Harvester data collected in Sweden is recorded in accordance with the StanForD standard and records information about every bucked log (Arlinger *et al.*, 2012; Holmgren *et al.*, 2012). Data collected by mechanical harvesters have so far mainly been used for reporting daily production and controlling the timber flow (Arlinger & Möller, 2006; Hannrup *et al.*, 2011; Möller *et al.*, 2011). The SDC stores the harvester data for two years after harvesting (SDC, 2014).

One disadvantage of harvester data, from a forestry planning perspective, is the fact that the wealth of information acquired pertains to stands that no longer exist (Rasinmäki & Melkas, 2005). The solution is to combine the harvester data with remote sensing data, acquired before clear-felling (Rasinmäki & Melkas, 2005). In a limited study, Hannrup *et al.* (2011) found comparable accuracy for predicting stand volume from only harvester data, to the area based method with ALS data. One advantage for harvester data is that the cost of acquiring it for remote sensing purposes is negligible, since it already collected for production evaluation and timber flow analysis.

1.4 Harvester reference data for ALS-based predictions

Usually, field plots from forest inventories have been the reference data for predicting forest variables with remote sensing data. This has made it natural to use a raster approach with cell sizes comparable to the field plot size (Tuominen & Haapanen, 2011). When using harvester data as reference data, the irregular shape and size of a harvested stand may be better represented by a segmentation approach (Tuominen & Haapanen, 2011). Segmentation algorithms partition areas based on variance in pre-defined metrics, this usually creates a cumbersome number of tiny elements (Rasinmäki & Melkas, 2005). Therefore, a second step is commonly required to merge elements with similar variance until the selected minimum element size is achieved for all elements. Consequently, segmentation provides a method to partition into irregularly shaped elements of desired size, which are very homogenous in terms of forest variables (Tuominen & Haapanen, 2011).

Cells around the border of the stand may only partially cover the area of interest which reduces the accuracy of predictions (Rasinmäki & Melkas, 2005; Tuominen & Haapanen, 2011). However, Rasinmäki & Melkas (2005) found that the method for partitioning a stand when using harvester data was not influential on the accuracy of the predictions. Instead the size of the elements is the more influential factor (Rasinmäki & Melkas, 2005). Generally, at least when using field plot data, it is expected to have better accuracy in predictions based on smaller element size, since each element will contain less variation (Tuominen & Haapanen, 2011), but Rasinmäki & Melkas (2005) found the opposite to be true for predictions on stand level harvester data. This may, at least in part be due to the low accuracy of tree positions, the smaller the elements are, the larger probability that a tree will be assigned to the wrong element (Rasinmäki & Melkas, 2005; Gobakken & Næsset, 2009).

1.5 Aim

Scientific studies agree that predicting stand variables from ALS data and field plots give equal or better accuracy than most other inventory methods (Holmgren, 2003; Eid *et al.*, 2004). However, objective field inventories are a costly component of remote sensing predictions (Means *et al.*, 2000; Næsset, 2007). By using harvester data instead of field plot data the cost of field inventories would be eliminated, making the use of remote sensing a more attractive option for the forest industry (Rasinmäki & Melkas, 2005; Barth, 2012). The purpose of this study is therefore, to test the possibility to utilize the information provided from harvesters as training data for predicting forest variables in mature stands with ALS data metrics.

The goals of this study are to;

1. Develop a method to predict forest variables with ALS data using harvester data as training data
2. Evaluate the accuracy of the method for predicting different forest variables

This study is limited to mature coniferous stands ready for final-felling. Geographically, the data analyzed in this report cover central Sweden. The study was limited to three forest variables; stem volume, measured in cubic meters per hectare solid under bark (m^3ha^{-1} sub), arithmetic mean tree height (m) and arithmetic mean tree diameter in breast height, measured in millimeter over bark (mm ob). Stem volume, mean height and mean diameter will be referred to as volume, height and diameter, respectively in this study.

2. Materials

The study area is located in central Sweden, from the lake Siljan (60°50' N, 14°55' E) in the northwest to the lake Mälaren (59°30' N, 16°40' E) in the south east (Fig. 1). The most common tree species in this region are Norway spruce (*Picea Abies* (L.) Karst.), Scots pine (*Pinus Sylvestris* L.) and birch (*Betula spp.*)

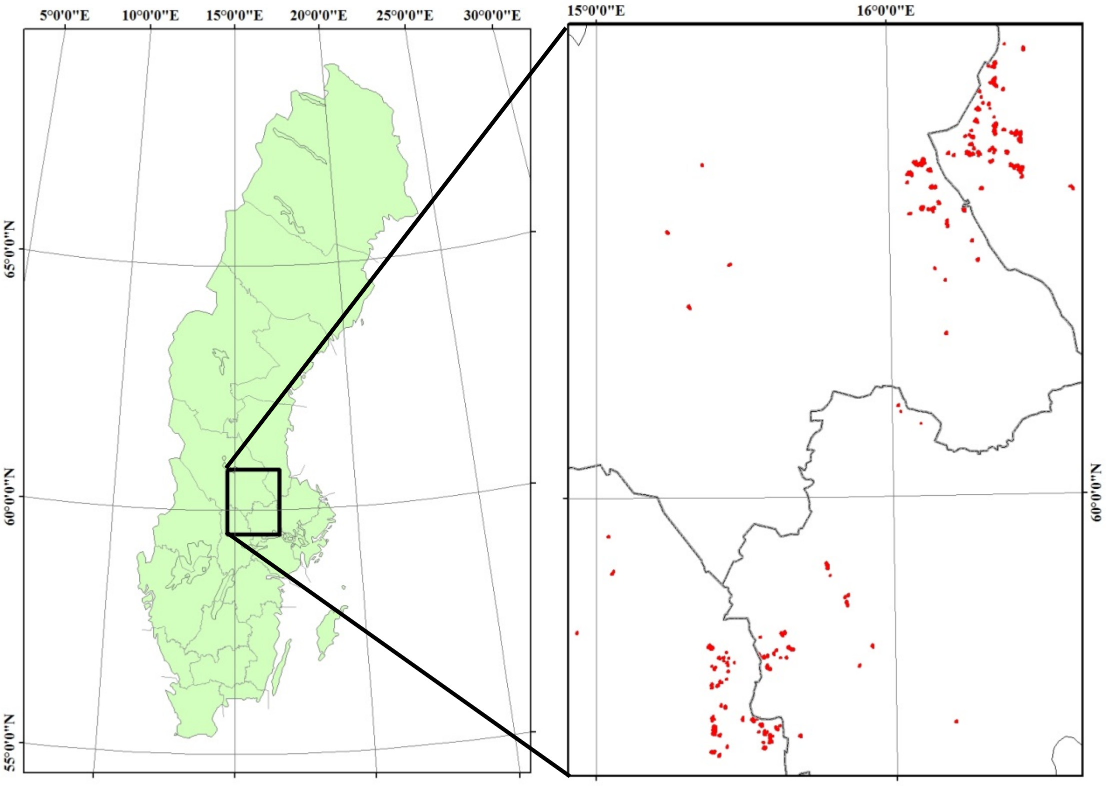

***Figure 1**. To the left, the extent of the study area located in central Sweden (59°30' N -60°50' N, 14°55' E-16°40' E). To the right, the examined stands (red) as distributed in the four counties Dalarna, Örebro, Västmanland and Gävleborg.*

2.1 Harvester data

The harvester data were recorded in the format of pri-files. Pri is an acronym for "production, individual" and is a file format for storage of harvester data in accordance with the StanForD standard developed during the 1980s (Arlinger *et al.*, 2003, 2012). Pri-files are created during logging operations and contain detailed data about all processed trees and individual logs, usually they also contain time-stamps and GPS-coordinates for the harvesters position when felling each tree (Arlinger *et al.*, 2003; Hannrup *et al.*, 2011). Today, there is a new standard, StanForD 2010, developed for communication between forestry machines (Arlinger *et al.*, 2012). As a part of StanForD 2010, a new format for recording harvester data called hpr, short for harvester production, has been developed (Arlinger *et al.*, 2012; Bhuiyan *et al.*, 2013). This is the successor to the pri format which will be replaced over the coming years (Bhuiyan *et al.*, 2013).

Harvester data was provided by Sveaskog's central market area office in Hedemora. The data were selected from harvested stands where the stand; 1) had been clear-felled, 2) was dominated by pine and spruce and 3) was harvested after Lantmäteriet had conducted the ALS in the area.

Choosing only clear-felled sites made analysis simpler, since thinned stands have a large unknown amount of remaining trees, which is unaccounted for in the harvester data. Trees left for nature conservation reasons will however still be unaccounted for in the harvester data from clear-felled sites. The motivation for selecting only pine/spruce stands was to reduce the influence of deciduous trees in the ALS data, since the deciduous trees reduce the accuracy of predictions of stand level characteristics (Næsset *et al.*, 2004; Villikka *et al.*, 2012). The average birch volume in the stands was 1.8% of the total volume, in line with criterion 2, but most deciduous trees were probably left for nature conservation reasons, so the actual level of compliance with criterion 2 is uncertain. To be able to use the harvester data, it needs to be connected to ALS data, which of course require that the ALS scan was conducted prior to the clear-felling, hence criteria 3.

Table 1. *Overview of collected harvester data stand level forest characteristics (n = 168)*

	Min	Mean	Max	SD
Volume (V), m^3ha^{-1} sub	87.6	200.1	353.9	46.8
Height (H), m	13.9	20.6	26.3	2.4
Diameter (D), mm	141.9	236.2	325.4	28.9

The harvester data (Table 1) consisted of, in total 510,001 logged trees from 168 stands, which cover an estimated area of 1160 hectares. These stands were divided into training and validation data sets using random selection. This resulted in 88 stands selected for training and 80 stands assigned for validation.

2.2 ALS data

The ALS data used in this project was collected through Lantmäteriet's national ALS project. For this project the flight altitude varied between 1700 and 2300 meters, the maximum scanning angle was limited to 20° and the flight path overlap was 20% (Lantmäteriet, 2014). The return density was between 0.5 to 1 per square meter using a footprint of 0.5 to 0.7 meters and the returns has been classified as either land, water, bridge or unclassified (Lantmäteriet, 2014).

The ALS data has also been processed at the Swedish University of Agricultural Sciences (SLU), as part of the joint project with Skogsstyrelsen to map forest variables for all of Sweden using ALS data from the Swedish Land Survey (Lantmäteriet) and field plots from the Swedish National Forest Inventory (Riksskogstaxeringen) as reference data (Skogsstyrelsen, 2014). The processing at SLU aimed to normalize the ALS data to an elevation model, to remove data anomalies and to eliminate overlaps from adjacent flight paths. To normalize the ALS data to ground level, the height of DEM created by Lantmäteriet was subtracted from the ALS data[1]. The data anomalies were corrected by removing returns above 50 meters and below -2 meters in relation to the DEM[2]. To eliminate overlap between flight paths, the data sets were compared with regards to scanning angle and only the ALS data from the flight path with the lowest scanning angle were kept[1].

[1] Peder Axensten, Research Engineer at SLU, pers. com. 2014-10-28
[2] Jonas Jonzén, Research Engineer at SLU, pers. com. 2014-09-02

3. Method

The harvester data required processing before it was serviceable for this study. The area based approach was chosen as the main focus for this study. However, to find the best way to predict forest variables with ALS data using harvester data as training data, several variants were examined. Two methods to handle positional information from the harvester data presented by Gobakken & Næsset (2009) were tested to examine the effect of positional errors. Two partitioning methods, raster and segmentation, were evaluated to examine the findings of Rasinmäki & Melkas (2005) and Tuominen & Haapanen (2011). When referring to raster cells and segments simultaneously, the term elements are used in this study. Three different element sizes for each partitioning method were also evaluated. Finally, two common methods of predicting relationships, multiple linear regression analysis and k-MSN, were tested and validated. All in all, 24 variants were evaluated for each of the forest variables; volume, height and diameter.

3.1 Data processing

The first step to process the pri-files was to convert them to the new hpr format. This was done using the experimental program "Hpr-analys" from Skogforsk. At this stage corrupt files which could not be converted and files without registered GPS-coordinates were discarded. The Hpr-analys software was used to calculate tree height and stem volume from the harvester data (Möller *et al.*, 2011). Tree heights were calculated by summing up the length of the logs for each stem and predicting the treetop using a function developed by Kiljunen (2002), which use several diameter measurements along the stem to predict the length of the top. The volume of a stem was calculated for 10 cm long stem sections using a mathematical formula for a truncated cone and transformed into volume solid under bark, using functions for bark thickness developed by Skogforsk[3]. The next step was to merge each set of hpr-files, the number of files per stand varied depending on stand size, into one comprehensive file for each stand. The stands were then examined in the Hpr-analys software and trees cut for main trails and landings were excluded, since they were not in the actual stands. Finally, the stands were exported as text-files which could be used for further analysis.

Each stand was scanned for duplicate stem entries, which were identified and removed. Single trees which lacked harvester GPS-position were assigned coordinates through interpolation. Each tree was then assigned an extra set of randomized GPS-coordinates by adding a random number, in the range ±9, to each of the East and North coordinates, a simple approximation of the reach of a harvester's boom was created (Hannrup *et al.*, 2011; Möller *et al.*, 2011). This was done to simulate the stem density and to test the positional effect on forest variable prediction, similarly to Gobakken & Næsset (2009).

To define the stand borders, stand polygons was created. The software Lastools was used to construct polygons from the original harvester GPS positions and an aggregation distance of 25 meters was used to generate the stand polygons (Hug *et al.*, 2004). A buffer of 10 meters was added to all stand polygons to account for uncertainty from the GPS positioning accuracy. The stand polygon area and the forest variables; volume, height and diameter was calculated for each stand. The percentage of the stand volume for each tree species was also calculated.

[3] Johan J. Möller, Researcher at Skogforsk, pers. com. 2015-01-26

3.2 Stand partitioning

To apply the area based method to harvester data, the data required to be broken down into smaller areas to base predictions on. To define prediction units, two methods were chosen for stand partitioning, raster and segmentation. In a study on tree volume prediction with harvester data, Rasinmäki & Melkas (2005) concluded that the method used for partitioning stands had marginal effect on the accuracy. However, the size of the elements may heavily influence the results, improving the accuracy of volume prediction with increasing element size (Rasinmäki & Melkas, 2005). Since the size of the elements are important for the overall prediction accuracy (Rasinmäki & Melkas, 2005; Tuominen & Haapanen, 2011), three different element sizes was selected to examine if element size affect the prediction accuracy. For the raster method, cells with 100, 400 and 1600 m^2 area were chosen. In the segmentation method minimum segment areas of 100, 300 and 900 m^2 respectively, were selected. These partitioning sizes are referred to as; small (S), medium (M) and large (L).

For both partitioning methods and for both original and randomized coordinates, the average values for all trees in each element were calculated for height and diameter. The sum of tree volume in each element was also calculated. Additionally, each elements area and polygon cover, defined as the percentage overlap of the stand polygon in the element, was calculated to identify borderline elements and adjust volume calculations by proportion of element area to polygon cover.

3.2.1 Raster method

The harvester measured variables were calculated on cell basis, as described above. This process was repeated for the six different raster variants (Table 2).

Table 2. *Overview of the six raster variants used in this study. Coordinates denote which set of GPS-coordinates were used to extract harvester data. Training cells and Validation Cells show how many elements were used to construct and test the prediction methods*

Partitioning variant	Coordinates	Cell area (m^2)	Training cells	Validation cells
rasHS	Harvester	100	46,222	36,820
rasHM	Harvester	400	13,497	10,556
rasHL	Harvester	1600	2,918	2,202
rasRS	Randomized	100	55,694	43,845
rasRM	Randomized	400	13,682	10,667
rasRL	Randomized	1600	2,922	2,209

3.2.2 Segmentation method

Segmentation was chosen as a partitioning method, because ALS data offer excellent opportunities to implement segmentation (Maltamo *et al.*, 2011). A segmentation software from SLU’s remote sensing section was used (Olofsson & Holmgren, 2014). However, the software used for this study was an older version, which based segmentation on ALS metrics from a raster, rather than polygons as described by Olofsson & Holmgren (2014). The raster cell size was set to 10 meters for calculating the metrics in the segmentation software.

The segmentation software created three metrics from ALS data, vegetation ratio (*VR*), average canopy height (*ACH*) and height percentile 95 (h_{95}). *ACH* was defined as the average height of first returns in 1 m raster cells, aggregated to chosen cell size, 10 m in this case. *VR* was calculated as the number of returns above the height cutoff of 2 m, also known as vegetation returns, divided by the total number of returns. The h_{95} metric was defined as the height where 95% of vegetation returns was found below.

The metrics were assembled into a three layered raster and were subsequently normalized by standard deviation to make comparison over differing scales meaningful and then used in the segmentation algorithm (Olofsson & Holmgren, 2014). The minimum segment size was set to 100 m^2, 300 m^2 and 900 m^2 respectively and the maximum size was set to 1,000,000 m^2 for all variants. The threshold value, measured in standard deviation, which was used to determine which elements should be merged, was set to 0.1.

Segmentation was done for all three defined segmentation sizes and variables from the harvester data were calculated for the six different segmentation variants (Table 3).

Table 3. *Summary of the segmentation variants examined in this study. Coordinates denote which set of GPS-coordinates were used to extract harvester data. Training segments and Validation segments show how many elements were used to construct and test prediction methods*

Partitioning variant	Coordinates	Min. segment area (m^2)	Avg. segment area (m^2)	Training segments	Validation segments
segHS	Harvester	100	302	17,835	13,940
segHM	Harvester	300	752	6,773	5,342
segHL	Harvester	900	2,237	1,946	1,526
segRS	Randomized	100	307	18,536	14,481
segRM	Randomized	300	760	6,831	5,376
segRL	Randomized	900	2,239	1,951	1,528

3.3 Calculation of independent variables

The software Lastools was used to extract metrics from the ALS data. This was repeated for three different raster with 10 m, 20 m and 40 m cell side respectively. These metrics were used directly for the respective raster variants. For the segmentation method the 10 m raster was the base for all segmentation variants and the metrics for each element was calculated as the average of all the raster cells located in the segment.

The metrics extracted from the ALS data were; height percentiles, average height, canopy cover, vegetation ratio and height count metrics. The height percentiles (h_{05}, h_{10}, h_{20}, h_{30}, h_{40}, h_{50}, h_{60}, h_{70}, h_{80}, h_{90}, h_{95}) provide the height, below which, the defined percentage of vegetation returns were located (Maltamo *et al.*, 2010). The average height (*avg*), defined as the average height of all returns above the height cutoff, the canopy cover (*cov*), calculated as the number of first returns above the height cutoff divided by the total number of first returns and the canopy density or vegetation ratio (*VR*), calculated as the number of returns above the height cutoff divided by the number of all returns, were also extracted from the ALS data. The height cutoff was set to 2 meters for all relevant metric calculations, defining all returns above the cutoff as vegetation returns (Nilsson, 1996; Maltamo *et al.*, 2010). Height count metrics (d_{00}, d_{01}, d_{02}, d_{03}, d_{04}, d_{05}) produced relative height density raster, which were calculated by dividing the number of returns in a defined height interval by the total number of returns and scaled to a percentage. This was done for the intervals; 2-5, 5-10, 10-15, 15-20, 20-25, 25-30 meters representing d_{00}, d_{01}, d_{02}, d_{03}, d_{04}, and d_{05}, respectively. The height count metrics were also transformed by the natural logarithm, ln(x) (lnd_{00}, lnd_{01}, ..., lnd_{05}), by inversion, 1/x ($invd_{00}$, $invd_{01}$, ..., $invd_{05}$), by the square root, $\sqrt{x}$ (sqd_{00}, sqd_{01}, ..., sqd_{05}) and by the square, x^2 ($x2d_{00}$, $x2d_{01}$, ..., $x2d_{05}$).

Two dummy variables were also defined, *spruce* and *pine*, to classify species dominance in stands. Stands which contained at least 70 percent of tree volume of one tree species was defined as dominated by that species. This resulted in 27 and 63 stands that were classified as pine and spruce dominated respectively.

The independent variables were examined for collinearity using variance inflation factor (VIF) (Equation 1) to identify metrics to exclude from regression modeling (Chatterjee & Hadi, 2006). VIF values higher than 10 are commonly viewed as a suspiciously high collinearity between independent variables (O'brien, 2007). Many metrics were found to be closely correlated and independent variables with a correlation of 0.9 or more were excluded from further analysis. For the six metrics chosen for building regression models; h_{95}, *cov*, lnd_{01}, lnd_{04}, *spruce* and *pine* were tested using VIF and no values higher than 5,15 were observed for any predictor data set which was deemed satisfactory.

$$VIF = \frac{1}{(1-R_i^2)} \qquad [1]$$

Where R_i^2 denotes the proportion of variance in the *i*:th independent variable in relation to the other independent variables.

3.4 Prediction of forest variables

Two methods were used to predict forest variables, multiple linear regression analysis and the non-parametric k-MSN method. The stand partitioning methods presented above does not conform to the irregular shape of forest stands, since they are both based on square elements, which makes prediction of forest variables in edge cells dubious. The approach chosen to reduce this influence was to only include elements that had at least 90% of its area inside the stand polygon, for predicting forest variables. Elements with missing data or ALS metrics with less than 2 meter average height were also excluded. Also the volume was transformed with the natural logarithm, to achieve a more linear relationship with the independent variables.

3.4.1 Linear regression

In multiple linear regression analysis (Equation 2) the variable of interest (Y) is called the dependent variable or target response and the variables which are thought to help explain variation in the dependent variable are known as predictor or independent variables (X) (Chatterjee & Simonoff, 2013). The coefficients (β) are unknown parameters and ε are random error terms (Chatterjee & Simonoff, 2013).

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_m X_m + \varepsilon \qquad [2]$$

The regression model is based on several assumptions that need to be fulfilled. Firstly, the linearity assumption, states that the model is assumed to relate the dependent variable (Y) and the independent variables (X) linearly through the coefficients (β) (Chatterjee & Hadi, 2006). Secondly, the errors (ε) are assumed to be independently and identically distributed (Chatterjee & Hadi, 2006). The independent variables are assumed to not be linearly dependent of each other (Chatterjee & Hadi, 2006). A forth assumption is that all observations have an equal role in influencing regression results (Chatterjee & Hadi, 2006).

Different regression models and independent variables were examined by Akaike Information Criterion (AIC) using a stepwise algorithm (Akaike, 1974). This algorithm starts with a basic model with one specified independent variable, h_{95}, and tests remaining independent variables stepwise in order of best correlation with the dependent variable, to determine the model (Venables & Ripley, 2002). From the stepwise models, recurring independent variables were selected to create the final regression models; h_{95}, lnd_{04}, *spruce* and *pine* for volume, and h_{95}, lnd_{01} and *spruce* for height and h_{95}, *cov*, lnd_{04}, *spruce* and *pine* for diameter, respectively. One advantage of using regression modelling is that the models are dependable for extrapolation outside the range of the data set (Chatterjee & Hadi, 2006).

3.4.2 k-MSN

All forest variables were predicted simultaneously with the k-MSN method, which is an adaptation of Moeur & Stage (1995) MSN method. This is a non-parametric nearest neighbor method which produces a weighting matrix used for choosing the *k* number of most similar neighbors from the reference data, using canonical correlation analysis (Packalén & Maltamo, 2007). Canonical correlation analysis (CCA) is a statistical approach to understand cross-covariance matrices. With two data sets, independent variables $X = (X_1, \ldots, X_n)$ and dependent variables $Y = (Y_1, \ldots, Y_m)$ CCA finds linear relationships between X_l and Y_l which maximizes the correlation between them (Moeur & Stage, 1995). This is called the first pair of canonical variables ($l = 1$). More canonical pairs may be found, but the number of pairs (l) is limited by the smallest data set, i.e. X_n or Y_m. The canonical correlation is largest for the first pair and drops for each additional pair, meaning that the predictive power of CCA is concentrated in the first few pairs of canonical variables (Moeur & Stage, 1995).

The k-MSN approach has the advantage over regression analysis, being able to predict many variables at the same time (Packalén & Maltamo, 2007; Crookston & Finley, 2008). With k-MSN the similarity between predicted elements and reference elements are measured in terms of forest variables and not arbitrary ALS metrics. Also the assumptions about normality and homoscedasticity are relaxed, compared with parametric methods such as regression analysis (Crookston & Finley, 2008). One significant draw-back of the k-MSN approach is the inability to extrapolate values, leading to overestimation of low values and underestimation of high values, outside the training data range (Maltamo *et al.*, 2011). A large set of training data is therefore required to cover the range of possible variation.

In this study, k was set to three, meaning the three most similar neighbors in terms of independent variables were used predict forest variables. The predicted value was calculated as weighted averages of the three nearest observations, with the weighting based on the reverse of the MSN distance (Packalén & Maltamo, 2007). The independent variables used in the k-MSN method were; *spruce, pine,* h_{95}, *cov,* d_{04}, lnd_{01} and lnd_{04}.

3.5 Model validation

For each prediction method the forest variables; volume, height and diameter was predicted for each element of the validation stands and the aggregated to stand level for evaluation. The results were validated at the stand level, using 80 randomly selected stands. The accuracy of the results were expressed in the form of root mean square error (RMSE) (Equation 3) and relative root mean square error (RMSE%) (Equation 4). To evaluate if any systematic errors was present in the predictions, bias (Equation 5) and relative bias (Equation 6) were calculated. The correlation coefficient, r, was also calculated to see how the predictions fit the measured values (Equation 7).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \quad [3]$$

$$RMSE\% = 100 * \frac{RMSE}{\bar{y}} \quad [4]$$

$$Bias = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n} \quad [5]$$

$$Bias\% = 100 * \frac{Bias}{\bar{y}} \quad [6]$$

$$r = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2} * \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad [7]$$

Where *n* is the number of validation forest stands, $\hat{y}_i$ is the predicted value for stand *i*, y_i is the observed value for stand *i*, $\bar{\hat{y}}$ is the mean of predicted values and $\bar{y}$ is the mean of the observed values.

4. Results

For each of the two prediction methods, a brief overview of the results for each predicted forest variable is presented, followed by a closer look at the best result for predicting each forest variable.

4.1 Prediction with regression method

The regression models which produced the best fit for each forest variable were; rasHS, rasHM and segRL for volume, height and diameter respectively (Table 4).

Table 4. *Properties of the regression models which most accurately predicted volume (ln(V)), height (H) and diameter (D) respectively*

Predicted variable	Independent variable	Coefficient	Std. Error	t value	p	R^2-adj
ln(V)	*(Intercept)*	3.848	0.046	83.338	0.000	0.076
	h_{95}	0.060	0.003	22.856	0.000	
	lnd_{04}	0.055	0.007	8.134	0.000	
	spruce	0.083	0.011	7.913	0.000	
	pine	-0.060	0.011	-5.233	0.000	
H	*(Intercept)*	6.957	0.191	36.392	0.000	0.544
	h_{95}	0.718	0.008	93.315	0.000	
	spruce	0.691	0.044	15.528	0.000	
	lnd_{01}	-1.086	0.027	-40.492	0.000	
D	*(Intercept)*	153.009	7.438	20.571	0.000	0.470
	h_{95}	8.287	0.452	18.328	0.000	
	cov	-1.289	0.054	-23.747	0.000	
	lnd_{04}	3.724	0.989	3.764	0.000	
	spruce	3.073	1.387	2.215	0.026	
	pine	9.854	1.454	6.776	0.000	

Predicting volume with regression analysis (Table 5) generally produced the best results using a medium element size of 300-400 m^2 with a relative RMSE of less than 11%. The bias was also generally lowest for this partitioning size. The correlation between ALS-predicted and harvester measured volume was high for all variants. The rasHS variant predicted volume with a RMSE of 21.0 m^3ha^{-1} sub at stand level, corresponding to 10.5% of stand volume (Table 4, 5). The regression model fitted to the rasHS dataset showed a low adjusted R^2 value, compared to the good fit between predicted and harvester measured values (Fig. 2).

The height was predicted with a RMSE of 1.1 m using the rasHM partitioning variant, corresponding to 5.1% of tree height (Table 4, 6). However all variants had similar results and the biases were low. The predicted height correlated well with harvester measured data with r = 0.97 for the rasHM variant (Fig. 3, Table 6). Stand 23 deviates from the rest, with a mean height of 13.9 m (Fig. 3). On closer inspection, the stand appears to be a multi-storied stand with tree heights ranging from 5.1 m to 25.1 m.

For diameter prediction, the RMSE ranged from 17.4 mm to 20.0 mm for different partitioning variants and tends to overestimate diameter by 1-3% (Table 7). The most accurate way to predict diameter was by the segRL variant (Fig. 4, Table 4). Again, the diameter of the multi-storied stand 23 is conspicuously lower than the rest of the data set, the measured stand mean diameter is 142 mm and the average of the entire data set is 236 mm. This leads to a clear overestimation of stand 23's mean diameter in the regression models.

Table 5. *Accuracy of predicted volume ($m^3 ha^{-1}$ sub) at stand level for regression variants*

Partitioning variant	RMSE	RMSE%	Bias	Bias%	r
rasHS	21.0	10.5	-3.2	-1.6	0.97
rasHM	21.5	10.7	-5.7	-2.9	0.97
rasHL	25.2	12.5	11.4	5.7	0.97
rasRS	34.8	17.4	-27.2	-13.6	0.97
rasRM	21.4	10.7	-4.8	-2.4	0.97
rasRL	25.1	12.5	10.8	5.4	0.97
segHS	25.2	12.6	-15.2	-7.6	0.98
segHM	21.3	10.6	3.2	1.6	0.98
segHL	29.0	14.4	15.2	7.6	0.98
segRS	24.3	12.1	-13.1	-6.5	0.97
segRM	21.4	10.7	3.3	1.7	0.98
segRL	26.9	13.4	13.1	6.5	0.98

Table 6. *Accuracy of predicted mean height (m) at stand level for all regression variants*

Partitioning variant	RMSE	RMSE%	Bias	Bias%	r
rasHS	1.2	5.8	0.1	0.7	0.98
rasHM	1.1	5.1	0.1	0.5	0.97
rasHL	1.1	5.2	0.2	1.0	0.97
rasRS	1.2	6.0	0.0	0.0	0.98
rasRM	1.1	5.1	0.1	0.3	0.97
rasRL	1.1	5.2	0.2	0.9	0.97
segHS	1.2	5.5	0.1	0.5	0.97
segHM	1.1	5.3	0.1	0.5	0.97
segHL	1.1	5.5	0.2	0.9	0.97
segRS	1.2	5.7	0.0	0.2	0.97
segRM	1.1	5.4	0.1	0.4	0.97
segRL	1.1	5.5	0.2	0.8	0.97

Table 7. *Accuracy of linear regression predicted mean diameter (mm ob) at stand level for different stand partitioning variants*

Partitioning variant	RMSE	RMSE%	Bias	Bias%	r
rasHS	20.0	8.4	7.3	3.1	0.93
rasHM	18.3	7.7	6.0	2.6	0.91
rasHL	18.5	7.8	5.3	2.2	0.89
rasRS	19.5	8.2	4.6	1.9	0.92
rasRM	18.1	7.7	4.9	2.1	0.91
rasRL	18.4	7.8	5.1	2.1	0.90
segHS	18.9	8.0	5.9	2.5	0.92
segHM	18.2	7.7	4.5	1.9	0.92
segHL	17.4	7.3	3.0	1.3	0.93
segRS	18.8	7.9	4.5	1.9	0.92
segRM	18.2	7.7	4.0	1.7	0.92
segRL	17.4	7.3	2.7	1.2	0.93

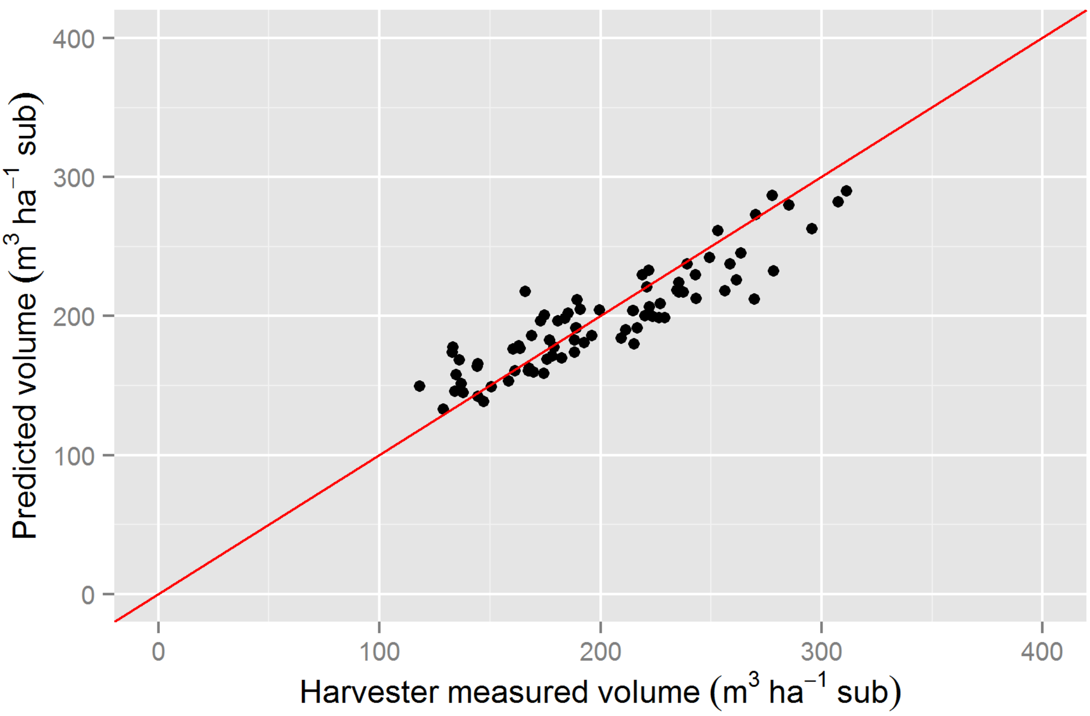

Figure 2. *Scatterplot of ALS-predicted volume and harvester measured volume using rasHS partitioning, aggregated to stand level for the 80 validation stands.*

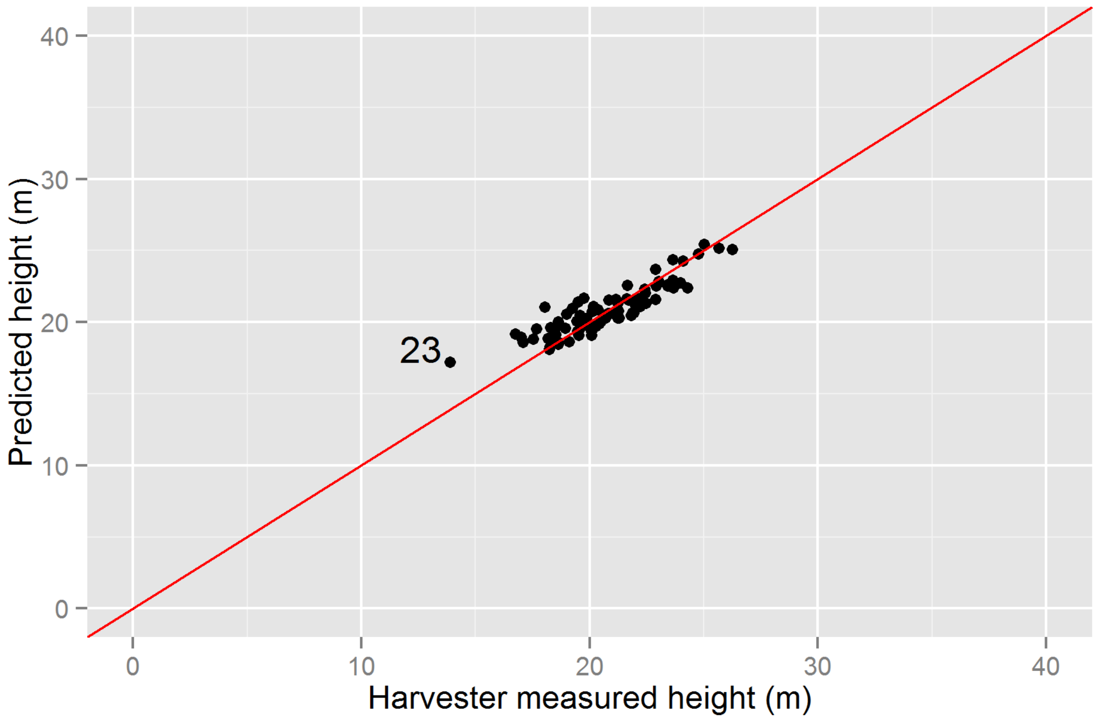

Figure 3. *ALS-predicted mean height plotted against harvester measured mean height for the rasHM partitioning variant, aggregated to stand level.*

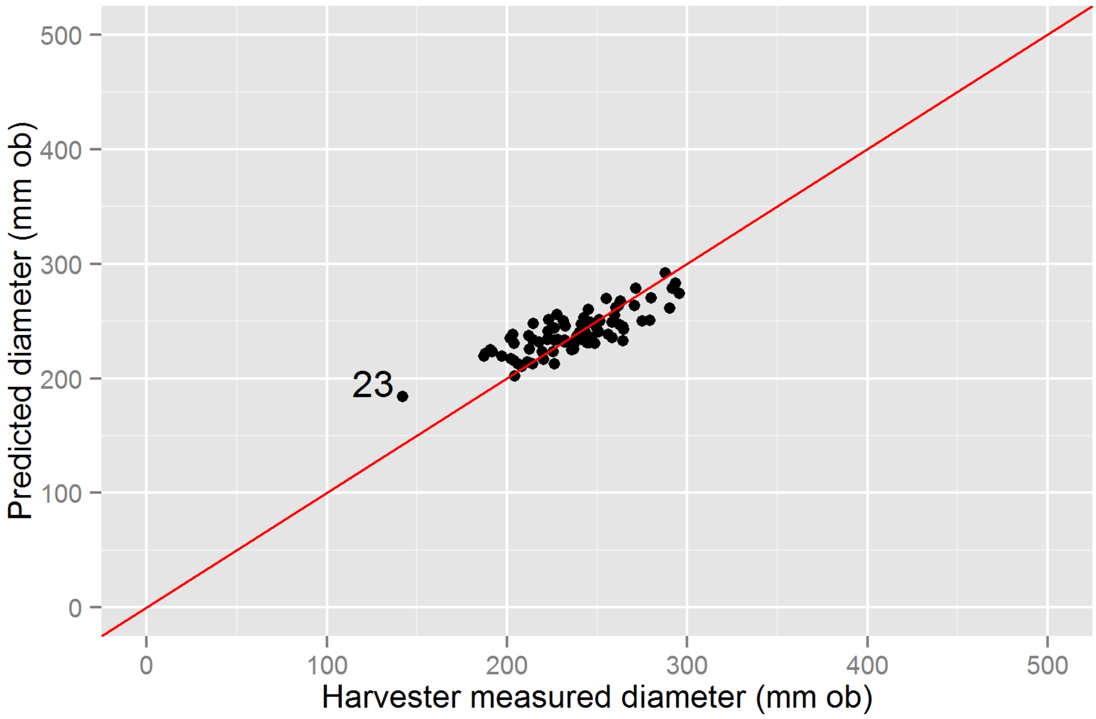

Figure 4*. ALS-predicted mean diameter using segRL partitioning, aggregated to stand level and plotted against harvester measured mean diameter.*

4.2 Prediction with k-MSN method

Predicting volume with the k-MSN method showed a RMSE of 22.1 m^3ha^{-1} sub for the best partitioning variant, segRS, and 35.8 for the worst, segHL (Table 8). The results indicate that segmentation with small elements perform best, especially with respect to bias. The segRS variant have virtually no bias and high correlation (r = 0.96) between predicted and measured stem volume at stand level (Fig. 5). The raster variants show best prediction accuracy for rasHM and rasRM, variants partitioned with medium sized elements (Table 8). Generally, volume predictions with regression analysis produce better accuracy than k-MSN.

Predicting height with the non-parametric k-MSN method yielded slightly better results compared to regression analysis, with a RMSE between 1.0 m to 1.2 m for the different variants (Table 9). Bias was also on average lower, less than 1% and the predicted values correlated well with the harvester measured heights, r = 0.98 for the most accurate variant rasHL, but there was a tendency to overestimate low heights (Fig. 6). As in the regressions, stand 23 sticks out, this is however more expected with k-MSN prediction.

All partitioning variants predict diameter with a RMSE of less than 20 mm (Table 10), with rasRM being the most accurate with a RMSE of 18.2 mm or 7.7% (Fig. 7). Regression models produce slightly better accuracy when diameter is concerned. The correlation coefficient (r) was slightly less for the k-MSN method compared with the regression predictions, while bias were on a similar level.

***Table 8**. Accuracy of predicted volume (m^3ha^{-1} sub) at stand level for different stand partitioning variants using k-MSN prediction*

Partitioning variant	RMSE	RMSE%	Bias	Bias%	r
rasHS	33.5	16.7	25.9	12.9	0.97
rasHM	23.6	11.8	8.6	4.3	0.96
rasHL	32.8	16.3	20.2	10.1	0.95
rasRS	23.3	11.6	-6.9	-3.5	0.97
rasRM	23.1	11.5	6.9	3.5	0.96
rasRL	30.3	15.1	18.7	9.3	0.96
segHS	23.8	11.9	5.9	2.9	0.97
segHM	27.4	13.7	12.5	6.2	0.97
segHL	35.8	17.8	22.3	11.1	0.96
segRS	22.1	11.0	-0.5	-0.3	0.96
segRM	25.6	12.7	9.2	4.6	0.97
segRL	34.7	17.3	19.8	9.9	0.96

Table 9. *Accuracy of predicted mean height (m) at stand level for different stand partitioning variants using k-MSN prediction*

Partitioning variant	RMSE	RMSE%	Bias	Bias%	r
rasHS	1.2	5.5	0.1	0.6	0.97
rasHM	1.0	4.9	0.1	0.3	0.98
rasHL	1.0	4.8	0.1	0.7	0.98
rasRS	1.2	5.7	0.0	-0.1	0.97
rasRM	1.0	4.9	0.0	0.0	0.98
rasRL	1.0	5.0	0.1	0.6	0.98
segHS	1.1	5.3	0.1	0.3	0.97
segHM	1.1	5.2	0.1	0.3	0.97
segHL	1.0	5.0	0.1	0.5	0.97
segRS	1.1	5.4	0.0	-0.2	0.97
segRM	1.1	5.3	0.0	0.1	0.96
segRL	1.0	4.9	0.1	0.5	0.98

Table 10. *Accuracy of predicted mean diameter (mm ob) at stand level using k-MSN for all the different stand partitioning variants*

Partitioning variant	RMSE	RMSE%	Bias	Bias%	r
rasHS	19.9	8.4	7.5	3.2	0.92
rasHM	18.6	7.9	5.6	2.4	0.91
rasHL	18.5	7.8	3.6	1.5	0.88
rasRS	19.4	8.2	4.7	2.0	0.92
rasRM	18.2	7.7	4.3	1.8	0.91
rasRL	18.9	8.0	3.5	1.5	0.87
segHS	19.2	8.1	5.5	2.3	0.89
segHM	18.9	8.0	3.8	1.6	0.89
segHL	18.6	7.9	2.6	1.1	0.90
segRS	18.9	8.0	3.9	1.6	0.90
segRM	19.2	8.1	3.1	1.3	0.87
segRL	18.7	7.9	2.0	0.9	0.90

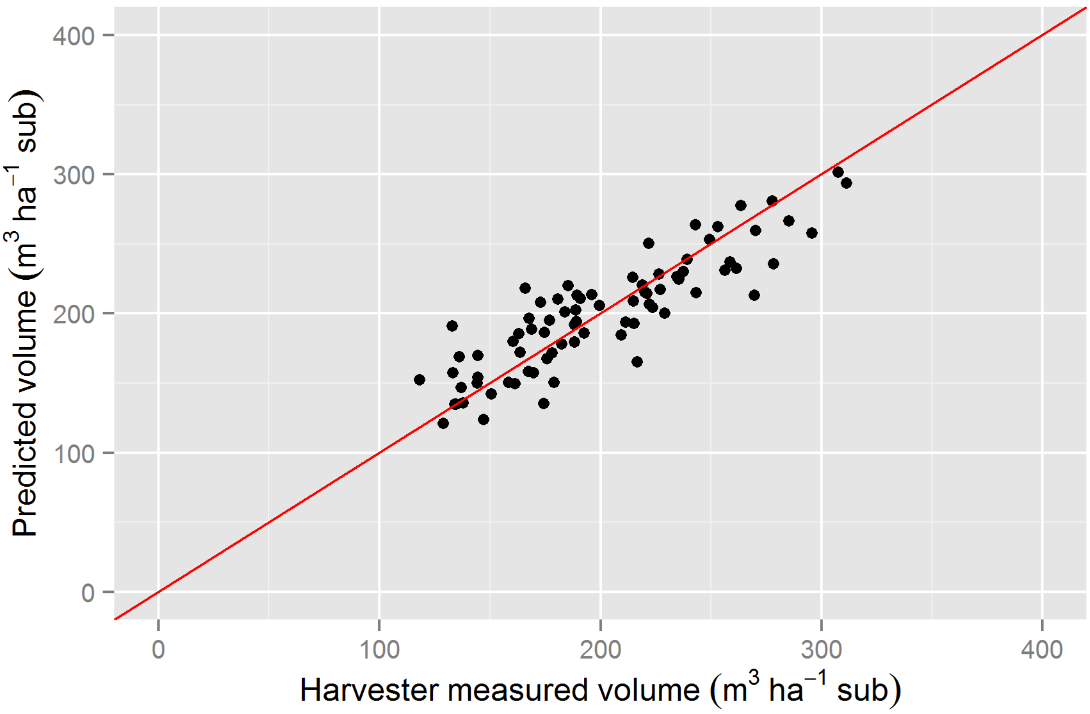

Figure 5. *ALS-predicted volume plotted against harvester measured volume using segRS partitioning, aggregated to stand level.*

Figure 6. *Predicted mean height using rasHL partitioning and k-MSN, aggregated to stand level and plotted against harvester measured height.*

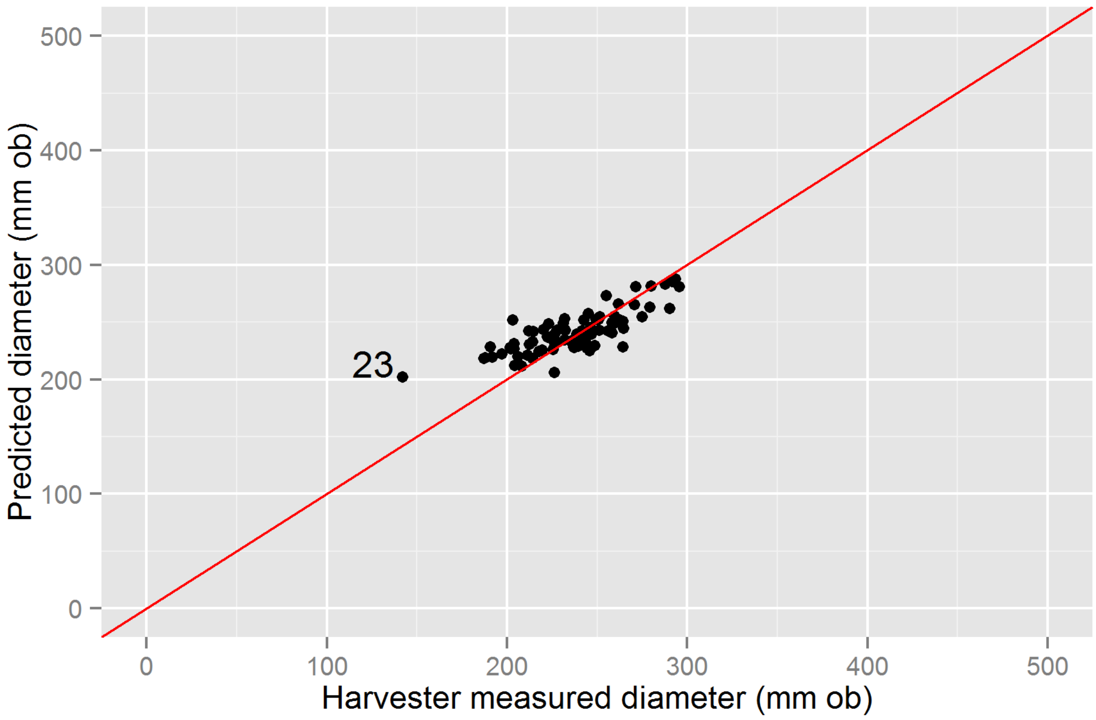

Figure 7. *ALS-predicted mean diameter using rasRM partitioning aggregated to stand level.*

5. Discussion

5.1 Results

The results presented in this study are comparable to the outcome of other papers using the area based method. Previous Scandinavian studies have shown, using ALS data and training plots, prediction accuracy in the range of 11-14% RMSE for stand volume while basal area weighted mean tree height has been predicted with a RMSE in the range of 3-6% and basal area weighted mean diameter with a RMSE of 9-13% (Næsset, 2007). A study using harvester data at plot level and high density ALS, predicted stem volume, mean tree height and mean diameter with a relative RMSE of 11%, 8% and 12%, respectively (Holmgren *et al.*, 2012). This study obtains similar results, 11%, 5% and 8% relative RMSE for stem volume, mean height and mean diameter respectively. The diameter predictions are in fact better than previous studies have reported, but this study used arithmetic, instead of basal area weighted mean height and diameter. However, the range of application for harvester data is clearly more limited, only being relevant to predict variables of mature coniferous forests. This study does examine predictions from harvester data on a large scale, compared with earlier studies. The data used in this study was also from actual forestry operations and not collected for research purposes.

The bias for height and diameter are consistently very low, absolute values ranged from 0% to 3%, but for volume predictions the biases are much larger, with absolute values ranging from 0% to 14%. However, the variants showing the best prediction accuracy also have very low bias, -0.3% and -1.6% for k-MSN and regression respectively. The regression method has quite low bias for predictions in medium sized elements, where the best predictions were found as well. The regression method commonly underestimates volume in small elements, while overestimating in large sized elements. For the k-MSN method the least biased variants have the smallest element size and bias then generally increase with increasing element sizes. Compared to subjective inventories, commonly used for forest management plans, this study produces better accuracies and less bias. The prediction accuracy in this study compared to subjective methods are; 11% vs. 15-25% for volume and 5% vs. 10% for height and 8% vs. 10-12% for diameter (Ståhl, 1992). The best predictions in this study underestimate volume with 0-2% compared to 0-20% for subjective methods (Sonesson *et al.*, 2008).

The differences between using the regression analysis method or the k-MSN method to predict forest variables are small in this study, both are viable. For volume and diameter prediction the regression method is slightly better than k-MSN, but k-MSN are marginally better for height predictions. Influence of the partitioning method, segmentation or raster, does not seem significant for predicting forest variables in this study. However, segmentation might not be the best way to partition stands when using harvester data, due to the frequently irregular shape of segments in combination with the uncertainty of the tree positions. The positional uncertainty in the harvester data should be taken into account when partitioning stands. The partitioning size, however, has discernable effects on prediction accuracy, especially for volume. When using k-MSN to predict volume, the trend for segmented stands is decreasing accuracy with increasing element size, this would support Tuominen & Haapanen's (2011) findings but contradict Rasinmäki & Melkas (2005). However, this is not the case when using regression, instead the best results are achieved with a medium element size of 300-400 m^2 regardless of all other partitioning factors. This does not support Rasinmäki & Melkas (2005) conclusion that larger elements produce better accuracy, but seems to fit well with the harvesters reach. Though, for mean height and mean diameter, the overall results indicate better prediction accuracy with increasing element size, as suggested by Rasinmäki & Melkas (2005). The differences between using the actual GPS positions from the harvester data and using simulated tree positions are minor for prediction accuracy. This statement would hold more weight if the tree simulation had been repeated 100 or 500 times, but this was not feasible within the scope of this study. Despite poor stand reconstruction accuracy and the resulting low explanatory power (R^2) of the ALS derived metrics, the current state of harvester data is still sufficient to produce predictions in line with current expectations of area based predictions using objectively measured field plots as reference data.

5.2 Limitations

There are several limitations of harvester data in this study that should be considered. By limiting the harvester data to final-felling stands, the data is bound to be fairly homogenous with regard to age, height and volume. Another limitation is that all the information about the forest gleaned from harvester data relates to stands already cut down, this creates an issue of stand reconstruction. The actual tree positions are not recorded in the harvester data, merely the position of the harvester is recorded for each tree felled and usually several trees are cut from the same position. In addition, the GPS-accuracy is lower in forests than in open ground due to distortions of the satellite signals from the tree canopy (Zheng *et al.*, 2005). However, the harvester cuts down the trees and creates a clearing around itself, in which the signal may be less disturbed. This creates uncertainty of the GPS positioning of the harvester. However, this is relatively small compared to the uncertainty created by the reach of the boom arm when considering tree positions. Considering the demands for accurate positioning of training data (Barth, 2008; Gobakken & Næsset, 2009), the results of this study are surprisingly accurate, and could potentially become even better with higher positional accuracy in the harvester data.

The nature of harvester data is quite different from field plots, concerning positional errors (Gobakken & Næsset, 2009). Each tree position is uncertain in relation to each other instead of just the center point of a field plot. However, Gobakken & Næsset (2009) shows that if the positional error increase from 5 m to 10 m, the field plot overlap of the true field plot area decrease from 60.8% and 72% to 25.8% and 45.5% for plot sizes of 200 m^2 and 400 m^2 respectively. Clearly larger field plots are more robust in relation to positional error than small ones. This also true for harvester data as Rasinmäki & Melkas (2005) discovered, the larger elements used reduce the likelihood of trees being assigned to the wrong element. Volume and diameter predictions are more sensitive to positional errors than height (Gobakken & Næsset, 2009). In this study, the prediction accuracy of height and diameter seem generally to support these findings. However, for volume prediction this does not seem to be the case, generally the best accuracy is found in medium sized elements of 400 m^2, corresponding well to the reach of a harvester.

Area measurement accuracy is vital in stand volume prediction accuracy (Hannrup *et al.*, 2011). Adding a buffer to harvester position will cover for positional uncertainty, but will inherently overestimate the harvested area and the errors are magnified in smaller stands (Hannrup *et al.*, 2011). Another important aspect is harvester measurement accuracy. Regular calibration and quality assurance is important to maintain a high accuracy in the harvester measurements (Hannrup *et al.*, 2011). Skogforsk have developed a method for quality assurance, that demands that log diameter and length measurements of the harvester does not deviate more than 4 mm and 2 cm respectively, from control measurements (Arlinger & Möller, 2006). This method has been implemented in the StanForD standard and is therefore readily available for all harvesters to utilize. From the harvester measurements several variables are calculated, including stem volume and tree height. The accuracy of height calculations in this study is unknown. The accuracy for height prediction of the tree tops was not directly evaluated by Kiljunen (2002), only dry mass logging residues.

Weaknesses in models are poor prediction accuracy for extreme values, even for regression models. Due to the homogenous data and averaged metrics, the models are limited to mature forest. Multi-storied stands are poorly represented in the data set in this study and one such stand was identified in the validation data, stand 23, which was clearly overestimated for height and diameter, while the volume was fairly accurately predicted. This was expected for the k-MSN method, which is known for overestimating values outside the range of the training data. But the regression method, known to be more robust for extrapolation, also similarly overestimated this stand. Possibly, there was not enough data for reliable k-MSN prediction, at least not for accurate predictions of multi-storied stands. Perhaps a separate function for multi-storied stands might be a way to improve prediction accuracy, but it would require a way to identify multi-storied stands from the site bank.

More ALS metrics could have been examined, this study was limited to height and density metrics which are known to be useful in prediction of forest variables, but there are many more metrics that can be calculated from ALS data. That is, some metrics with very good correlation to the examined forest variables may have been overlooked.

One aspect that has been ignored in this study is the temporal dimension, there is a gap between the time of the ALS flights and execution of the harvesting operations. For most of the study area ALS flights were conducted in 2010 and 2011, while the harvesting data was collected mainly during 2013 and 2014. This leaves between two to four growth seasons unaccounted for in the ALS data. Preferably the data should be collected in the same growth season, but an alternative is to simulate the tree growth in software, such as Heureka, and adjust the harvester data to conform to the season of the ALS data.

5.3 Future research and development

Implementation of records for individual tree positions in StanForD as suggested by Hannrup *et al.* (2011) would afford considerable improvement to the positional accuracy of the harvester data and it would be intriguing to examine these effects in terms of prediction accuracy of forest variables. This would also enable development of single tree methods using harvester data. Technology for registering the position of the tip of the boom arm in relation to the harvester are commercially available (Cranab, 2013). This suggests that increasing tree positioning accuracy substantially is currently feasible. Mounting TLS sensors on harvesters allows recording of the remaining stand, this enables objective follow-up and quality assurance in regards to thinning operations (Barth, 2012). This technique allows for automatic tree identification, positioning and measurement of breast height diameter (Barth, 2012). This could also be a method for documenting the stand for better reconstruction, which could be especially useful for repeating this study with data from thinning operations.

Harvesting planning data from forest companies may decrease or even eliminate the need to calculate stand areas. This would of course require the harvest planning data to be saved and stored with metadata which connects the planning data and the harvester data to facilitate efficient access and identification. In addition to the harvest planning data, information regarding planned nature conservation in the area could be useful to improve area calculation and volume predictions from harvester data.

This study was limited to only three forest variables; volume per hectare, mean height and mean diameter. However, the harvester data contains a lot more information, it would be feasible to predict more forest variables, such as basal area, Lorey's height, dominant height and it could also be interesting to evaluate prediction of diameter distribution in a stand. Perhaps even timber and pulpwood proportions could be estimated. It would also be of interest to test the effect of elevation and latitude in a larger scale project.

It would be interesting to repeat this study with data from thinning operations. This would require some way of obtaining information of remaining trees, perhaps predicted through thinning ratio. Skogforsk is currently conducting a study in which thinning ratio is predicted using boom angle for felled trees[4]. Thinning ratios could also be acquired in real time using a TLS-system on the harvesters (Barth, 2012). If successful this would widen the field of application for harvester data. This would likely be less accurate than in final-felling predictions, but could still be useful information for strategic and tactical planning purposes.

One obstacle for operational use of this proposed method is a time consuming bottle-neck in the use of the Hpr-analys software. The experimental software has a graphical user interface, which prevents full automation of the data processing. In this study the Hpr-analys software was a vital component for conversion of pri-files into hpr-files and in particular, the calculation of stem volume and top length. However, using pri-files directly should be possible but would require development of replacement functionality for the file conversion and calculation of tree variables.

The question of the future sources of remote sensing data needs to be considered, the national ALS project is almost completed and there are no indications of it being repeated, which require a new source of remote sensing data for operational use of this method. Smaller scale ALS projects could probably be supported by individual forest companies. However, if national coverage is desired, digital photogrammetry may be an alternative in the future. Aerial photography is cheaper than ALS and better at tree species identification but lack canopy penetration, which excludes density metrics (Bohlin *et al.*, 2012; Vastaranta *et al.*, 2013). Another alternative data source to evaluate is radar, which is showing potential (Fransson *et al.*, 2001; Persson, 2014). It is valuable to assess the possibility of these alternatives performing adequately as a substitute for ALS. These two remote sensing techniques are already operational on a national scale and data are readily available and frequently updated. This could be a way to lower costs for acquiring remote sensing data, which are required to use harvester data in prediction of forest variables.

[4] Johan J. Möller, Researcher at Skogforsk, pers. com. 2015-01-26

5.4 Conclusions

This study evaluates the utility of actual harvester data from operational forestry, and on a much larger scale than in previous scientific studies. Despite poor positional accuracy in the harvester data, this study shows that good predictions can be achieved without knowledge of precise tree positions by using available data from harvester operations and ALS data from Lantmäteriets national ALS scanning. The results suggest that harvester data can be a viable source of training data for remote sensing applications. This would be a very cost-effective way of obtaining information about the forest. There are also presently technologies available, which have the potential to improve the positional accuracy of harvester data and in extension increase prediction accuracy.

The prediction results are similar for both regression analysis and k-MSN and also for both segmentation and raster partitioning methods. However, some trends can be seen, stand volume prediction using regression analysis seem to perform best with an element size of 300 to 400 m^2. This may be related to the reach of the harvester. For k-MSN the optimal element size seems to be 100 m^2, particularly for segmentation partition. For mean diameter and mean height the predictions are similar for all variants, but seem to improve with increasing size of elements, generally producing the best accuracy with large elements of 900 to 1600 m^2.

6. References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), pp 716–723.

Arlinger, J. & Möller, J. J. (2006). *Kvalitetssäkring av skördarnas mätning*. Uppsala, Sweden: Skogforsk. (Resultat / Skogforsk; 2006:20).

Arlinger, J., Möller, J. J. & Sondell, J. (2003). *A description of pri-files : Background, structure and examples*. Uppsala, Sweden: Skogforsk. (Arbetsrapport / Skogforsk; 559).

Arlinger, J., Nordström, M. & Möller, J. J. (2012). *StanForD 2010 : Modern kommunikation med skogsmaskiner = StanForD 2010 : Modern communication with forest machines*. Uppsala, Sweden: Skogforsk. (Arbetsrapport / Skogforsk; 784).

Axelsson, P. (2000). DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing*, 33, pp 110–117.

Barth, A. (2008). *Flygburen laser gav bättre data om träden*. Uppsala, Sweden: Skogforsk. (Resultat / Skogforsk; 2008:15).

Barth, A. (2012). *Beståndsmätning med mobila sensorer i skogsbruket = Use of mobile sensors in forestry to measure stand properties*. Uppsala: Skogforsk. (Arbetsrapport / Skogforsk; 773).

Barth, A., Sonesson, J. & Skog, A. (2012). *Ny skogsstandard effektiviserar kommunikationen*. Uppsala, Sweden: Skogforsk. (Resultat / Skogforsk; 2012:7).

Bhuiyan, N., Arlinger, J. & Möller, J. J. (2013). *Kvalitetssäkring av beräkningsresultat från hprCM och konvertering av pri- till hpr-filer = Quality assurance of calculation results from hprCM and conversion of pri files to hpr files*. Uppsala, Sweden: Skogforsk. (Arbetsrapport / Skogforsk; 788).

Bohlin, J., Wallerman, J. & Fransson, J. E. S. (2012). Forest variable estimation using photogrammetric matching of digital aerial images in combination with a high-resolution DEM. *Scandinavian Journal of Forest Research*, 27(7), pp 692–699.

Bredström, D., Jönsson, P. & Rönnqvist, M. (2010). Annual planning of harvesting resources in the forest industry. *International Transactions in Operational Research*, 17(2), pp 155–177.

Chatterjee, S. & Hadi, A. S. (2006). *Regression analysis by example*. 4. ed. Hoboken, NJ, USA: John Wiley & Sons. (Wiley series in probability and statistics). ISBN 0471746967.

Chatterjee, S. & Simonoff, J. S. (2013). *Handbook of regression analysis*. Somerset, NJ, USA: John Wiley & Sons.

Cranab. *Pressrelease 2013-05-30*. [online] (2013). Available from: http://www.cranab.se/static/sv/281/. [Accessed 2015-01-30].

Crookston, N. L. & Finley, A. O. (2008). yaImpute: An R package for knn imputation. *Journal of Statistical Software*, 23(10), pp 1–16.

Duvemo, K. & Lämås, T. (2006). The influence of forest data quality on planning processes in forestry. *Scandinavian Journal of Forest Research*, 21(4), pp 327–339.

Duvemo, K., Lämås, T., Eriksson, L. O. & Wikström, P. (2014). Introducing cost-plus-loss analysis into a hierarchical forestry planning environment. *Annals of Operations Research*, 219(1), pp 415–431.

Eid, T., Gobakken, T. & Næsset, 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(6), pp 512–523.

Fransson, J. E. S., Smith, G., Askne, J. & Olsson, H. (2001). Stem volume estimation in boreal forests using ERS-1/2 coherence and SPOT XS optical data. *International Journal of Remote Sensing*, 22(14), pp 2777–2791.
Gobakken, T. & Næsset, E. (2009). Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 39(5), pp 1036–1052.
Hannrup, B., Bhuiyan, N. & Möller, J. J. (2011). *Utvärdering av ett system för beräkning och återföring av skördarbaserad information till skogliga register och planeringssystem = Evaluation of a system för calculation and feedback of harvester based information fo forestry planning systems*. Uppsala, Sweden: Skogforsk. (Arbetsrapport / Skogforsk; 757).
Holmgren, J. (2003). *Estimation of forest variables using airborne laser scanning*. Diss. Umeå, Sweden: Swedish University of Agricultural Sciences.
Holmgren, J. (2004). Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scandinavian Journal of Forest Research*, 19(6).
Holmgren, J., Barth, A., Larsson, H. & Olsson, H. (2012). Prediction of stem attributes by combining airborne laser scanning and measurements from harvesters. *Silva Fenn*, 46(2), pp 227–239.
Hug, C., Krzystek, P. & Fuchs, W. (2004). Advanced lidar data processing with LasTools. *Proceedings of XXth ISPRS Congress*, Istanbul, Turkey, 2004. pp 12–23. Istanbul, Turkey.
Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X. & Maltamo, M. (2008). Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, 29(5), pp 1339–1366.
Jonsson, B., Jacobsson, J. & Kallur, H. (1993). *The forest management planning package. Theory and application*. Uppsala, Sweden: Swedish Univ. of Agricultural Sciences. (Studia forestalia Suecica; 189). ISBN 91-576-4698-8.
Karlsson, J., Rönnqvist, M. & Frisk, M. (2006). RoadOpt: A decision support system for road upgrading in forestry. *Scandinavian Journal of Forest Research*, 21(S7), pp 5–15.
Kiljunen, N. (2002). Estimating dry mass of logging residues from final cuttings using a harvester data management system. *International Journal of Forest Engineering*, 13(1), pp 17–25.
Lantmäteriet. *Laserdata*. [online] (2014). Available from: http://www.lantmateriet.se/sv/Kartor-och-geografisk-information/Hojddata/Laserdata/. [Accessed 2014-09-09].
Lillesand, T. M., Kiefer, R. W. & Chipman, J. W. (2008). *Remote sensing and image interpretation*. 6. ed. New York, NY, USA: John Wiley & Sons. ISBN 9780470052457.
Magnusson, M., Fransson, J. E. S. & Holmgren, J. (2007). Effects on estimation accuracy of forest variables using different pulse density of laser data. *Forest Science*, 53(6), pp 619–626.
Maltamo, M., Bollandsas, O. M., Vauhkonen, J., Breidenbach, J., Gobakken, T. & Naesset, E. (2010). Comparing different methods for prediction of mean crown height in Norway spruce stands using airborne laser scanner data. *Forestry*, 83(3), pp 257–268.

Maltamo, M., Eerikäinen, K., Packalén, P. & Hyyppä, J. (2006a). Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry*, 79(2), pp 217–229.
Maltamo, M., Malinen, J., Packalén, P., Suvanto, A. & Kangas, J. (2006b). Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. *Canadian Journal of Forest Research*, 36(2), pp 426–436.
Maltamo, M., Packalén, P., Kallio, E., Kangas, J., Uuttera, J. & Heikkila, J. (2011). *Airborne laser scanning based stand level management inventory in Finland.* Hobart, Australia: SilviLaser.
Means, J. E., Acker, S. A., Fitt, B. J., Renslow, M., Emerson, L. & Hendrix, C. J. (2000). Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering and Remote Sensing*, 66(11), pp 1367–1372.
Moeur, M. & Stage, A. R. (1995). Most Similar Neighbor: An Improved Sampling Inference Procedure for Natural Resource Planning. *Forest Science*, 41(2), pp 337–359.
Mäkinen, A., Kangas, A. & Mehtätalo, L. (2010). Correlations, distributions, and trends in forest inventory errors and their effects on forest planning. *Canadian Journal of Forest Research*, 40(7), pp 1386–1396.
Möller, J. J., Arlinger, J., Barth, A., Bhuiyan, N. & Hannrup, B. (2011). *Ett system för beräkning och återföring av skördarbaserad information till skogliga register- och planeringssystem = A system for calculation and feedback of harvester based information to forestry planning systems*. Uppsala, Sweden: Skogforsk. (Arbetsrapport / Skogforsk; 756).
Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sensing of Environment*, 56(1), pp 1–7.
Nilsson, M. (1997). *Estimation of forest variables using satellite image data and airborne lidar*. Diss. Umeå, Sweden: Swedish University of Agricultural Sciences.
Nordkvist, K., Sandström, E., Reese, H. & Olsson, H. (2013). *Laserskanning och digital fotogrammetri i skogsbruket (2. uppl.)* [online]. 2. ed. Umeå, Sweden: Swedish University of Agricultural Sciences. Available from: http://pub.epsilon.slu.se/10956/7/__ad.slu.se_common_bibul_slub_epsilon_oppetarkiv_nordkvist_k_etal_140120.pdf. [Accessed 2014-11-20].
Nordström, M. & Möller, J. J. (2009). *Den skogliga digitala kedjan - fas 1*. Uppsala, Sweden: Skogforsk. (Arbetsrapport / Skogforsk; 676).
Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80(1), pp 88–99.
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(5), pp 433–442.
Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å. & Söderman, U. (2004). Laser scanning of forest resources: the nordic experience. *Scandinavian Journal of Forest Research*, 19(6), pp 482–499.
O'brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, 41(5), pp 673–690.
Olofsson, K. & Holmgren, J. (2014). Forest stand delineation from lidar point-clouds using local maxima of the crown height model and region merging of the corresponding Voronoi cells. *Remote Sensing Letters*, 5(3), pp 268–276.

Packalén, P. & Maltamo, M. (2007). The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment*, 109(3), pp 328–341.
Persson, H. (2014). *Estimation of forest parameters using 3D satellite data stereogrammetry, radargrammetry and interferometry*. Diss. Umeå, Sweden: Swedish University of Agricultural Sciences. (Acta Universitatis agriculturae Sueciae, 2014:84). ISBN 9789157681171.
Rasinmäki, J. & Melkas, T. (2005). A method for estimating tree composition and volume using harvester data. *Scandinavian Journal of Forest Research*, 20(1).
SDC. *SDC.se*. [online] (2014). Available from: http://www.sdc.se/. [Accessed 2014-11-20].
Skogsstyrelsen. *Skogliga skattningar från laserdata*. [online] (2014). Available from: http://www.skogsstyrelsen.se/Myndigheten/Projekt/Regeringsuppdrag/Skogliga-skattningar-fran-laserdata/. [Accessed 2014-10-17].
Sonesson, J., Arlinger, J., Barth, A., Eriksson, B., Frisk, M., Jönsson, P. & Möller, J. J. (2008). *Analys av potentiella mervärden i kedjan skog-industri vid användning av pulsintensiv laserscanning*. Uppsala, Sweden: Skogforsk. (Arbetsrapport / Skogforsk; 654).
Ståhl, G. (1992). *En studie av kvalitet i skogliga avdelningsdata som insamlats med subjektiva inventeringsmetoder : A study on the quality of compartmentwise forest data aquired by subjective inventory methods*. Umeå, Sweden: Swedish Univ. of Agricultural Sciences. (Rapport / Sveriges lantbruksuniversitet, Institutionen för biometri och skogsindelning, Avdelningen för skogsuppskattning och skogsindelning; 24).
Sveaskog. *Om Sveaskog - Sveriges största skogsägare*. [online] (2014). Available from: http://www.sveaskog.se/Om-Sveaskog/. [Accessed 2014-05-28].
Tuominen, S. & Haapanen, R. (2011). Comparison of Grid-Based and Segment-Based Estimation of Forest Attributes Using Airborne Laser Scanning and Digital Aerial Imagery. *Remote Sensing*, 3(5), pp 945–961.
Vastaranta, M., Wulder, M. A., White, J. C., Pekkarinen, A., Tuominen, S., Ginzler, C., Kankare, V., Holopainen, M., Hyyppä, J. & Hyyppä, H. (2013). Airborne laser scanning and digital stereo imagery measures of forest structure: comparative results and implications to forest mapping and inventory update. *Canadian Journal of Remote Sensing*, 39(5), pp 382–395.
Venables, W. N. & Ripley, B. D. (2002). *Modern applied statistics with S*. 4. ed. New York, NY, USA: Springer. (Statistics and computing (New York)). ISBN 0387954570.
Villikka, M., Packalén, P. & Maltamo, M. (2012). The suitability of leaf-off airborne laser scanning data in an area-based forest inventory of coniferous and deciduous trees. *Silva Fennica* [online], 46(1). Available from: http://www.silvafennica.fi/article/68. [Accessed 2014-09-17].
Zheng, J. Y., Wang, Y. H. & Nihan, N. L. (2005). Quantitative evaluation of GPS performance under forest canopies. *2005 IEEE Networking, Sensing and Control, ICNSC2005*. pp 777–782. New York, NY, USA: IEEE. ISBN 0-7803-8812-7.