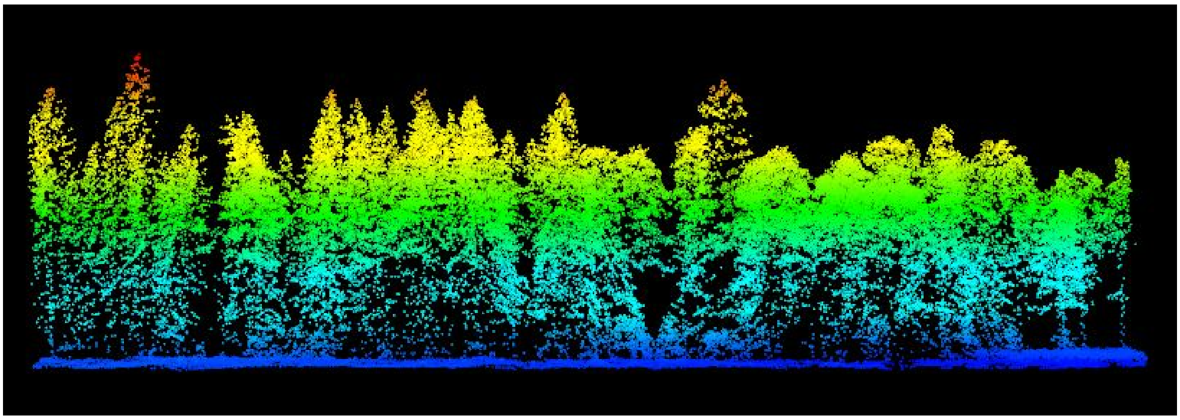
Evaluating inventory methods for estimating stem diameter distributions in micro stands derived from airborn laser scanning



Anders Lundholm

Arbetsrapport 409 2014
Examensarbete 30 ECTS A2E
Master of Science in Forestry (SLU, Sweden),
Master of Environmental Management (UNB, Canada)

Handledare:
Johan Holmgren

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–409–SE

Evaluating inventory methods for estimationg stem diameter distributions in micro stands derived from airborn laser scanning

Anders Lundholm

Keywords: kMSN, imputation, transect inventory, guided sampling, ALS

Master thesis in Forest Management at the dept of Forest Resource Management
Master of Science in Forestry (SLU, Sweden), Master of Environmental Management (UNB, Canada)
EX0768, A2E, ENVS6005 (UNB)
Supervisor: Johan Holmgren, SLU, Institutionen för skoglig resurshushållning, Remote sensing
Examiner: Håkan Olsson, SLU, Institutionen för skoglig resurshushållning, Remote sensing

Sveriges lantbruksuniversitet
Institutionen för skoglig resurshushållning
Utgivningsort: Umeå
Utgivningsår: 2014

ISSN 1401–1204
ISRN SLU–SRG–AR–409–SE

Abstract

A lot of research has focused on which laser metrics and which k-Nearest Neighbour (kNN) distances give the most accurate estimations. Most studies suggest that the kNN distance k-Most Similar Neighbour (kMSN) is the most accurate for estimating forest variables from local training data. However, little research has focused on how to best acquire training data for estimating forest variables. The aim of this study was (1) to evaluate if a faster and simpler inventory method can be used to estimate a similar accuracy for diameter distribution as unbiased sample plot inventory, by using Airborne Laser Scanning (ALS) data and kMSN imputation. (2) To reduce the sample size in order to find the threshold for minimum training data size without reducing the accuracy. Three different sampling methods were compared; circular plot inventory Sample 1 (S1) and two transect inventories Sample 2 and Sample 3 (S2 and S3). S1 use plots with 5 m radius. Both S2 and S3 were located between the S1 circular plots, using approximately a 2 m sampling width. For S2 a transect start between two circular plots *i* and *i+1*, and ends between two circular plots *i+1* and *i+2*. For S3 a new transect was created approximately every 10th tree, which on average created shorter transects. The field data from the three inventory methods was linked with ALS data to create three sets of reference data, with ALS data being extracted from the centre of each plot and transect. The reference data was then used for a kMSN imputation on a set of validation plots. Twenty five plots with 40 m radius were used to validate the kMSN estimates and all trees on the validation plots were callipered. The three sets of imputed trees were compared to the measured variables (ground truth) on the validation plots to assess the accuracy. The imputed mean diameter had a relative Root Mean Square Error (RMSE) of 15.4%, 16.4%, and 17.4% for S1, S2, and S3 respectively. The absolute error index means were 84.4, 91.9, and 82.2 for S1, S2, and S3 respectively. The relative error index means were 0.38, 0.33, and 0.33 for S1, S2, and S3 respectively. The results showed that training data from transect inventory can be used to estimate diameter distributions (both absolute and relative) with similar accuracy as training data from circular plot inventory. The transect stem density need to be measured with higher accuracy to get reliable estimates of forest variables when using training data from transect inventory. The results also showed that a fairly small set of training data (100-150 plots) can be used without reducing the accuracy much. An attempt at imputing stands was made but the estimates were not very accurate. K was set to one and the training data consisted of 90 micro stands, it would be preferable to use a higher k and a larger set of training data.

Introduction

Diameter distribution is important for forest management planning since it determines the proportions of merchantable and total volume, it can also be used for estimating the proportion of saw logs and pulp of the merchantable timber. Diameter distribution is also important for conservation and biodiversity since larger diameter trees tend to harbour a larger amount of species (Fries et al. 1997). Tree diameter also has an effect on harvesting and thinning revenue as harvesting takes roughly the same amount of time independent of tree diameter, i.e. harvesting large diameter trees cost as much as harvesting small diameter trees but the latter result in smaller revenue.

Having good information about the diameter distribution is thus an important decision basis for forest management. In the perfect world every tree would be measured to know the exact condition of the forest but due to obvious reasons a total inventory is very time consuming, costly and impractical. Even in cases where individual properties are small, it is often not realistic to measure all stands. Measuring a sample of the forest provides good estimations of the current forest condition. Therefore forest variables, such as tree diameter and tree height, have to be estimated for the stands that not are inventoried. One option is to use an area based ALS method where a small field sample is combined with ALS data that cover the entire forest holding. The field sample and ALS data is linked to create training data and it is then possible to estimate forest variables for the stand where there is only ALS data and no field inventory data. Unbiased inventory methods in forestry are often time consuming and costly. To capture the true diameter distribution in a homogenous area it might not be necessary to calliper all trees in circular plots. A linear inventory with a small width (i.e. a transect) could capture not only a sample of the trees in a circular plot but several nearby trees, resulting in a sampling from a larger area. In circular plot inventory time is spent measuring trees in the plot and walking between the plots. During transect inventory the sampling could be done as the inventory taker walks from one point to the other.

Laser Scanning

The first successfully operational lasers were built in 1960 in the United States of America and in the 1970s the technology was tested for various natural resource measurements, but in the early works trees were considered a noise (i.e. disturbing the measurements) (Nelson, 2013). With the advancements in computer technology (smaller computers, high performance and high storage capacity) and the use of Internal Measurement Units (IMU) reliable airborne laser scanning of forests became possible. Laser scanners emit pulses that reflect back when they hit an object. Using the time of flight of the emitted laser and the speed of light the distance from the laser scanner to the object can be calculated (Lim et al. 2003). The IMU keeps track of the airplanes navigational angles in roll, pitch, and yaw, a high precision GPS is needed to keep track of the location of the airplane (Lim et al. 2003). When millions of laser returns have been recorded they can be visualised as a three dimensional point layer resembling the scanned forest area (Lim et al. 2003).

Light Detection and Ranging (LiDAR) refers to the device that emits lasers pulses and receives them, the usage of LiDAR is thus independent of platform (terrestrial and airborne)

(Wulder et al. 2013). ALS is the combination of a LiDAR device, a scanner, an IMU, a GPS, and a computer for data processing and data storage, aboard an airplane or a helicopter (Wulder et al. 2013).

The vertical sampling refers to the maximum number of samples that can be recorded for each emitted pulse; horizontal sampling refers to the area of the footprint and the number of footprints per unit of area (Lim et al. 2003). The LiDAR footprint, or circular sampling area or each emitted pulse, equals height multiplied by divergence; the higher above ground the scanner is the larger the footprint is going to be (Lim et al. 2003). There are two types of pulse range data used for forestry applications; discrete return LiDAR and full waveform LiDAR. Discrete return LiDAR captures at least first and last return and usually has a smaller footprint, making it possible for the pulses to reach the ground. Capturing multiple returns is important if the goal is to measure the true vertical structure and most commercial systems can capture between two and five returns (Lim et al. 2003). With a small footprint it is difficult to penetrate all levels of vegetation which results in first and last returns capturing only a portion of the canopy and ground surface (Lim et al. 2003). Full waveform LiDAR emits pulses and records how much of the pulse is returned to the sensor for a series of equal time intervals and manages to capture information all throughout the vertical forest structure from the tree top, inside the canopy, understory, bushes and finally the ground (Lim et al. 2003). In other words full waveform manages to capture the entire backscattered signal of every pulse. Recording the intensity (magnitude) of the returned pulse can be used to determine what kind of surface it reflected on (Kim et al. 2008). The method is based on different surfaces (or even tree species) have different reflectivity and in theory, information about the target could be derived from intensity data. Kim et al. (2008) found that it could mainly differentiate between conifers and deciduous, which can already be done with near infrared aerial photos.

Methods for Estimating Forest Variables

There are mainly two methods for estimating forestry variables; single tree method and area based method. Single tree methods rely on high density data to delineate individual tree crown and use tree crown characteristics and dimensions to extract forest variables (Persson et al. 2002; Popescu et al. 2003; Gupta et al. 2010a; Gupta et al. 2010b; Reitberger et al. 2010; Kaartinen et al.2012; Holmgren & Lindberg, 2013; Ko et al. 2013). Both high and low density data can be used for area based methods. In area based methods plots or entire stands are measured, the relationship between field measured variables and ALS metrics is used to predict the forest variables on other areas with similar ALS metrics. To predict forest variables either a parametric method like regression analysis or a non-parametric method like kNN algorithms can be used. Regression analysis builds a model that use ALS metrics and beta coefficients (beta coefficients are created based on the relationship between field measured variables and ALS metrics) to predict forest variables. A kNN algorithm can be applied to both single trees methods and area based methods when predicting forest variables. When using kNN, there is a need to collect a comprehensive set of field data. The location of the field data is then linked to remote sensing data (or in this case of this study, ALS metrics) of that location to create the training data. The kNN algorithm takes the ALS metrics of the target areas (areas not field inventoried) and groups them with the training data based on similarities in the ALS metrics. Field measured variables (or more so, a list of trees) from training data are then imputed to the target areas. Using a relatively small set of field data, forest variables for a large forested area can be estimated. There are several different distances that can be used to find the nearest neighbour. kMSN has been used to accurately predict forest attributes in coniferous forest on plot and stand level by using the mean canopy height (Maltamo et al. 2006, Packalén & Maltamo, 2007, 2008; Hudak et al. 2008). The difference between kNN and kMSN is in how the variables are transformed. kMSN uses canonical correlation analysis which means that new pairs of linear correlated variables (called canonical variables) are created from the predictor (ALS metrics) and response variables (field measured variables) (Moeur & Stage, 1995). Predictor variables that are highly correlated with the response variable are thus given a larger weight in the canonical variables compared to the predictor variables that are not highly correlated with the response variable (Crookston et al. 2002). Over-fitting the kMSN model is not a vital problem since it has proven to be a robust method for predictions, whether the whole data set or cross validating is used (Johnson & Wichern, 2007).

Using ALS and kMSN, Maltamo et al. (2006) managed to predict stand attributes with higher accuracy than field inventory (relative RMSE of predicted stand volume was 6%). Latifi et al. (2010) estimated forest attributes in a German mixedwood forest using Random Forest (RF) and three kNN distances; Mahalanobis distance, kMSN, and Euclidean distance. They found RF to predict the most accurate forest attributes, though the results of the kMSN method were almost as accurate as RF (Latifi et al. 2010). Similar results were achieved by Vauhkonen et al. (2010), they found small difference in accuracy when imputing single tree variables using kMSN and RF.

Diameter Distribution

In a study by Bollandsås et al. (2013) seemingly unrelated regression and kMSN was used to compare diameter distributions in south eastern Norway. Their conclusion was that both the parametric and non-parametric approach resulted in good diameter distribution estimates, but the kMSN approach was more reliable in estimating the diameter of the large trees (Bollandsås et al. 2013). Species specific diameter distributions were estimated with kMSN and Weibull distribution, the estimation used a combination of predictor variables from ALS data and aerial photographs. The results showed that the kMSN performs better than Weibull distribution (Packalén & Maltamo, 2008). Maltamo et al. (2011) tested four different plot selection methods for kMSN estimation; random selecting, random selecting within forest type strata, selection of plots according to geographic location, or selection using ALS data. They concluded that using ALS data to guide the selection of plots increased the accuracy for the MSN estimations (volume and stem number) and that the data set could be reduced to 100 plots without reducing accuracy (Maltamo et al. 2009, 2011).

Comprehensive Training Data

Many of the studies that use ALS data to estimate forest variables on stand and plot level focus on selecting predictor variables (Packalén & Maltamo, 2008; Hudak et al. 2008; Latifi et al. 2010; Maltamo et al. 2009, 2010). Little attention has been given to improving methods of gathering training data for ALS estimations. Forests and single trees can automatically be delineated with high accuracy (Koch et al. 2009). Forest stands can automatically be segmented into micro stands using ALS data with as good accuracy as by a human interpreter when segmenting is done with regards to volume, mean diameter and height (Mustonen et al. 2008). The drawback of micro stands, compared to traditional forest management stands used in Scandinavian forestry, is that they are often too small to be operational due to extremely detailed segmentation. Parameters can be set to avoid creating too small stands but this also increase the risk that the segmentation is too general. Traditional forest management stands are often based on previous management actions, they are delineated subjectively, require correction after field visit and may thus be too generalized and they are not entirely based on ecological homogeneity (Mustonen et al. 2008; Tuominen & Haapanen, 2011). Micro stands are done objectively without requiring time consuming work and are done at no additional cost since the laser data is already acquired for estimating forest variables. Even if the individual micro stand is small, operational planning could be done so that several micro stands close together are subjected to the same management operation at the same time (Leppänen et al. 2008). The drawback of automatic ALS segmentation is delineating stand with regards to species composition, but using ALS data in a combination with colour or colour-infrared images has managed to improve the stand delineation with regard to species composition (Leppänen et al. 2005; Mustonen et al. 2008). Using small, homogenous micro stands when planning field work for collecting training data helps in determining whether a comprehensive enough set of training data will be measured, thus avoiding measuring unnecessary plots. Using ALS as a priori information to guide the sampling when planning the field work could help in measuring training data that includes all kinds of forest types (Ståhl et al. 1997). Guided line sample transect was developed to inventory of sparse, scattered objects (Ståhl et al. 1997) and could thus be used in order to capture all kinds of forest types. In a report by the Forestry Research Institute of Sweden, diameter distribution measured from sample plot inventory and transect inventory was compared. They found no difference between measured diameter distributions or time consumption between the methods (Hansson, 1999).

Aim of the Present Study

The aims of this study were (1) to evaluate if a faster and simpler inventory method can be used to estimate a similar accuracy for diameter distribution as unbiased sample plot inventory, by using ALS data and kMSN imputation. (2) To reduce the sample size to find the threshold for minimum training data size that can be used without reducing the accuracy.

Material and Method

Airborne laser scanning data was used to create a raster of the forest area. The raster was used as input in an algorithm that automatically segmented the forest into micro stands. Circular sample plots were placed in the micro stands and the circular plots were used to guide the location of the transect inventory. The three sets of training data were used in a kMSN imputation on validation plots. Imputation was done with training data on plot/transect level and training data on stand level. Diameter distribution was the main forest variable of interest and the estimated diameter distributions were compared to the ground truth diameter distribution, using relative error index and absolute error index. The size of the training data was reduced to find the threshold for smallest training data that can be used without reducing the accuracy.

Test Site

The test site is the Remingstorp property (58º45′ N, 13º66′ E), a 1100 hectare (ha) forest estate in southern Sweden. This forest property has been used widely for remote sensing studies; such as ALS, multispectral aerial images, and radar. The forest is dominated by Norwegian spruce (*Picea abies* (L.) Karst.), but there is also some Scots pine (*Pinus sylvestri* L.) and occurrence of broadleaf species. The property is intensively managed and consists of stands in every phase of the compartment management system.

Laser Data

The laser data was acquired with a Reigl LMS-Q560 laser scanner onboard a helicopter flying at 400 m altitude. The area was scanned in September 2010 (leaf-on conditions), and the average point density was 30 returns/m^2. The scanner used four rotating mirror surfaces to get a parallel line scan pattern. The pulse repetition frequency (PRF) was up to 240 kHz, but the PRF was 160 kHz since a 60 degree scan angle was used. The scan frequency was between 10 to 160 Hz and the beam divergence was 0.5 mrad. The laser wave length was 1550 nm, the pulse length about 3 ns. The scan width across the flight direction was ±30 degrees, and the scan width along the flight direction was 0 degrees.

Micro Stands

Using the ALS data, a map with 10x10 m raster cells was created and the forest was automatically segmented into 1836 homogenous micro stands. The size of the micro stands depended on the homogeneity of the forest; the more homogenous the forest, the larger the micro stand and vice versa. To ensure that comprehensive training data was collected, the Local Pivotal Method (LPM) (Grafström et al. 2012) was used for selecting both which micro stands to inventory and the location of the circular sample plots within the micro stands. Which micro stands should be inventoried was decided based on how commonly occurring that kind of micro stand was in the whole set of micro stands. Micro stands were sampled for inventory to ensure that the micro stand sample comprehended all kinds of stands located on the property. A total of 90 of the micro stands were selected for inventory. The amount and location of the circular sample plots depended on the homogeneity of the forest structure in each micro stand. If the micro stand was very homogenous, the minimum amount of sample plots (four) were assigned to the micro stand. With increasing heterogeneity, more sample plots were assigned to the micro stand. The field plots were located to comprehend as much of the difference in each micro stand as possible. The micro stands vary from a couple of hundred m^2 to a few ha in size, and had between four and 27 sample plots in them.

Field Data for Training

Three sampling methods were used; S1 which was a circular plot inventory collected in July and August 2013. For S1 a total of 872 sample plots in 90 micro stands were measured. These plots were 5 m radius and all trees larger than 40 mm Diameter at Breast Height (BRH) were callipered. ALS metrics for S1 plots were extracted in the center of the circular plot (Fig. 1). S2 and S3 are transecting inventory methods and were measured in September 2013. The location of transects was decided based on the location and order of the circular sample plots. From the start plot (most northern and western plot in the micro stand) the field worker walked in a straight line toward the closest, not yet visited sample plot and did this until the field worker has passed through the location of every circular sample plot in the stand. All trees above 40 mm DBH that could be reached when the person taking the inventory extends his arms (roughly 2 m width) were callipered. The micro stand transect is divided into several transects (think of them as linear plots). For S2 a total of 829 transects in 90 micro stands were measured. A new S2 transect start between two circular plots *i* and *i+1*, and ends between two circular plots *i+1* and *i+2*. ALS metrics for a S2 transect were thus extracted from the center of the circular plot located in the transect (Fig. 1). For S3 a total of 746 transects in 90 micro stands were measured. A new S3 transect was created every 10th tree and ALS metrics for a S3 transect were extracted from the location of the 5th tree in the transect (Fig. 1).

The first transect in S2 and S3, and the last transect in S2 were made only half as long as they should be according to the sampling method. In the case of S2, there is neither direction nor distance from which to approach the S1 plot in first transect or proceed from the S1 plot in the last transect. In the case of S3, the first 5 trees were assigned to the GPS coordinate of the first S1 plot in the micro stand and the following 10 trees assigned to the first created GPS coordinate etc. The S3 method was done because of distance between two S1 plots is sometimes quite long, in some cases up to 300 m. The laser metrics in sample plot *i* might not be representing for the callipered trees standing 150 m away and S3 reduces the distance between the location of callipered tree and location for ALS data extraction. It was assumed that the stem density around a circular plot is similar to the stem density inside a circular plot, thus no counting plots were deployed for the transect inventories. The S1 plots were used as counting plots for the transect inventories.

There should be an even number of plots/transects for S1 and S2 (Table 2). Some micro stands were assigned one circular sample plot per raster cell, these micro stands were bogs or brooks where most trees were very small. This resulted in transects without any trees, transects which might have been removed in the data processing due to missing values. Another reason why there are less S2 plots is due to the fact that it was extremely hard to keep track of which plot the inventory taker was in when the plot centers are only 10 m apart. Add to this a GPS with a positioning accuracy of ±10 m that continuously made notifications when the inventory taker arrived at a new plot for every step the inventory taker took. S3 have a lower number of transects since a new transect was created every 10th tree rather than creating a new transect for every corresponding circular sample plot.

A normal GPS was used during the collection of the field samples (S1, S2, and S3). S1 used a differential GPS for part of the inventory. But due to errors with the differential GPS during the field work the uncorrected coordinates for S1 were used for the study. A visual comparison of the three sample methods plot/transect location and where ALS data is extracted for each plot/transect can be seen in (Fig. 1).

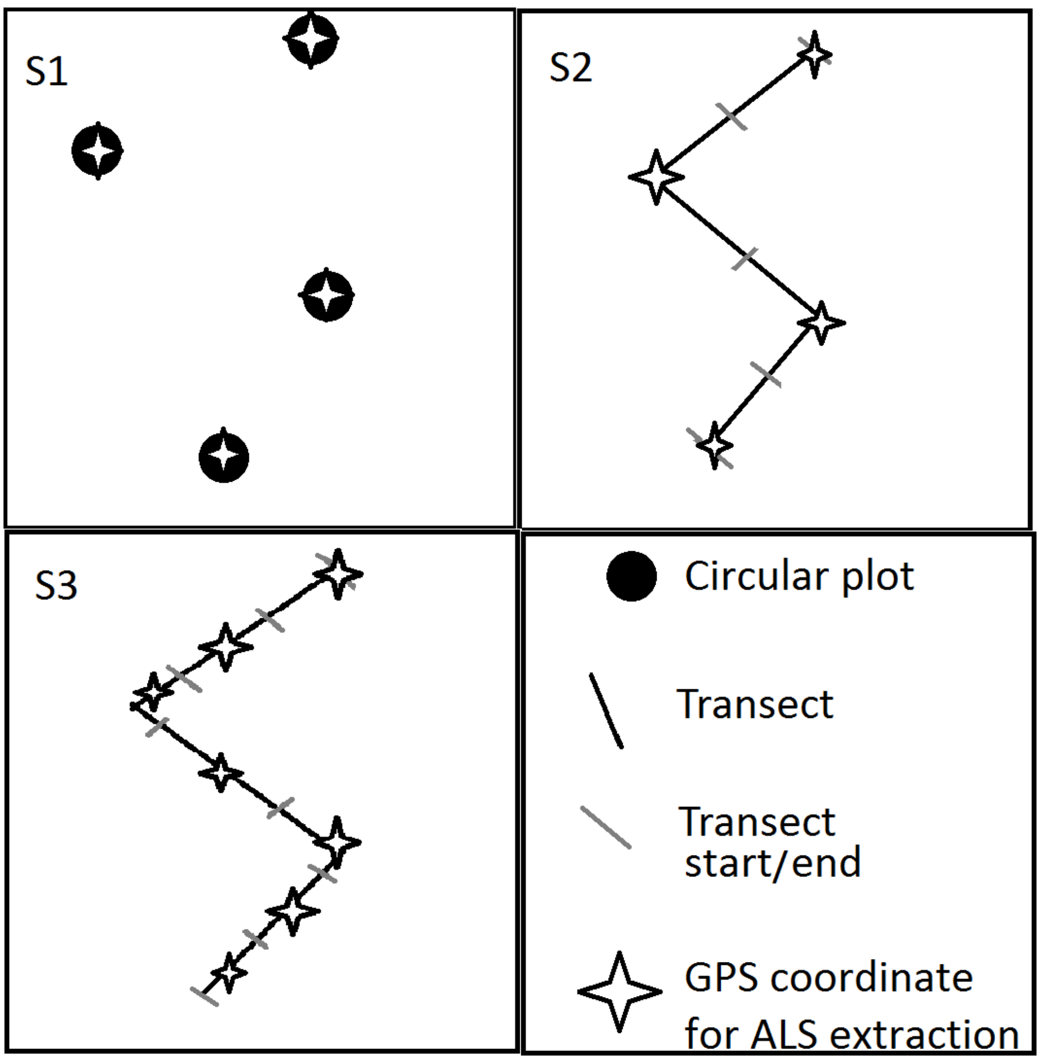


Figure 1. *Location of plots/transect and GPS coordinate for ALS data extraction for the S1, S2, and S3. S1 plots have a radius of 5 m. ALS data for S2 were extracted from the same coordinate and S1. S2 transect start between two circular plots i and i+1, and ends between two circular plots i+1 and i+2. In each S3 transect there are 10 trees and the 5th tree is the location for ALS data extraction.*

Validation Plots

The validation data consist of 25 circular plots with a 40 m radius in which all trees above 40mm DBH were callipered. These large plots (5026 m^2) were considered to be micro stands and were used for validation. Using the training data from S1, S2, and S3, three set of trees were imputed to the validation plots. The performance of S1, S2, and S3 was examined by comparing the imputed arithmetic mean diameter, imputed basal area weighted mean diameter, imputed basal area and imputed diameter distribution to the measured variables on the validation plots. The minimum, mean, maximum, and standard deviation values for number of stems, basal area, DBH, and Lorey's mean height for the validation plots can be seen in (Table 1).

Table 1. *Minimum, mean, maximum, and standard deviation for stems per ha, basal area per ha, breast height diameter, and Lorey's mean height (basal area weighted) for the 25 control plots*

	N, ha^{-1}	**G, m^2, ha^{-1}**	**DBH, cm**	**h_{Lorey}, m**
Min	347.0	20.4	12.8	16.2
Mean	696.3	31.8	22.8	23.4
Max	1096.1	45.1	34.3	30.1
Stdev	26.4	5.6	4.8	4.8

kMSN Estimates

The kNN distance kMSN (Moeur & Stage, 1995) was used to estimate the diameter distribution on the validation plots. The kMSN algorithm takes the laser metrics from the target data and finds which reference data (training data) is the most similar by using canonical correlation analysis (Moeur & Stage, 1995). The field measurements from the training data are then imputed to the target raster cell. In other words, training data (list of trees) from one raster cell is copy-pasted onto a target raster cell based on the similarities between the two raster cells ALS metrics. All data processing, imputations, and analyses were done in R Studio computing program (http://www.r-project.org/) using the package yaImpute (Crookston & Finley, 2008).

Seven predictor variables extracted from the ALS data were used in the analyses; average height, standard deviation height, vegetation ratio, crown height average, the 10^{th} height percentile, the 50^{th} height percentile, and the 90^{th} height percentile. The response variable used was arithmetic mean diameter on plot level.

Three different sets of training data were used for the kMSN imputation; one for the circular sample plots (S1) and two for the transect inventory (S2 and S3) (Fig. 1). Since the micro stands were created to be homogenous and because the forest is autocorrelated, it was assumed that the forest looks the same where the ALS metrics for the transect were extracted as it does throughout the rest of the transect. This assumption was used to add tree representation to the callipered trees in S2 and S3. Since the area of the transect was unknown (no fixed transect width), it was assumed that the stem density in the transect was the same as the stem density in the closest circular sample plot in the same micro stand.

Stand Imputation

Micro stands were imputed with the kMSN algorithm with k set to one. Arithmetic mean diameter in the micro stand was used as response variable and the mean of the seven ALS metrics as predictor variables. The response variable mean was achieved by calculating the means of all callipered trees in a micro stand. The predictor variable means were achieved by taking the mean of all laser metrics in a micro stand. For the imputation, the means of all laser metrics in a target stand (validation plot) was used. The representation per ha of each imputed tree in the validation plot was calculated by taking the mean stem density in each inventoried micro stand. The mean stem density was then divided by the total number of inventoried trees in each that micro stand. Finally each tree stem representation per ha was scaled to the validation plot area in ha. Since stand imputation use the micro stand mean values only two sets of training data was used; circular (S1) and transect (S2/S3). S2 and S3 has the same list of trees (only divided into transects of different lengths), so in a stand imputation where mean ALS metrics from micro stands are used S2 and S3 are identical.

Mean Diameter and Diameter distribution

Mean diameter was calculated using both arithmetic mean diameter and basal area weighted mean diameter (Equation 1). Means were calculated for the imputed trees, from all three sampling methods and compared to the validation plots.

$$D_{BA_{\mathrm{k}}} = \frac{\sum_{i=1}^{n} d_{ki}^{3}}{\sum_{i=1}^{n} d_{ki}^{2}} \qquad (1)$$

$D_{BA_{\mathrm{k}}}$ is the basal area weighted mean diameter in validation plot *k*.d_{ki} is the diameter of tree *i* in validation plot *k.* The basal area weighted mean diameter gives a larger weight to large diameter trees, or more so, weighting down small diameter trees that grow in the understory. The sum of the diameter multiplied with the basal area is divided by the sum of all basal areas. Since π is used both in the numerator and the denominator it can be broken out of the equation.

The diameter distribution of the imputed trees was calculated for the three sampling methods and compared to the diameter distribution in the validation plots. One cm diameter classes were used for calculating diameter distribution. The smallest diameter class were for trees <40 mm DBH, and the largest diameter class were for trees >500mm DBH. There were 48 diameter classes in total. The area representation for the trees from the transect inventories was calculated by assuming that the stem density was the same throughout the transect as it was in the closest circular sample plot. Multiplying the number of trees in the circular sample plot with 10,000 divided by the plot size gives trees per ha. Dividing trees per ha with the number of trees in the transect results in the tree representation each transect tree have on one ha. Transect tree representation per ha multiplied with 100 divided by 10,000 produces the tree representation each transect tree has on a 10x10 m raster cell. See Equation 2.

$$\xi_i = \left(\frac{K_i * \frac{10,000}{79}}{\kappa_i} \right) * \frac{100}{10,000} \tag{2}$$

Where ξ_i is the tree representation per raster cell for all line inventoried trees in sample plot i, K_i is the number of callipered trees in circular sample plot *i*, κ_i is the number of callipered trees in transect *i*, 10,000 is a hectare in m^2, 79 is the circular plot area in m^2, and 100 is the area of a raster cell in m^2.

Accuracy Assessment

To assess the accuracy of the imputation using S1, S2, and S3 as training data, the variables arithmetic mean diameter, basal area weighted mean diameter, and basal area from the kMSN imputations were compared to the field measured variables in the validation plots. Relative RMSE (Equation 3) and relative Bias (Equation 4) were used. The relative values were used as it makes the accuracy of the estimates comparable between variables in this study, i.e. arithmetic mean diameter, basal area weighted mean diameter, basal area, and the two error indices. Relative RMSE and relative bias also make the accuracies of this study comparable to the accuracies achieved in other studies.

$$\text{Relative RMSE} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}}{y_{mean}} * 100 \tag{3}$$

$$\text{Relative Bias} = \frac{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n}}{y_{mean}} * 100 \tag{4}$$

Where n is the number of validation plots, y_i and $\hat{y}_i$ is field measured value and predicted value respectively for validation plot i, y_{mean} is the validation plot mean value of the variable.

To assess the accuracy of the diameter distribution, two different methods were used; relative error index based on Packalén & Maltamo (2008), and absolute error index based on Reynolds et al. (1988). Relative error index compensates for any error in stem estimation because it focuses on the shape of the diameter distribution (Equation 5). Absolute error index includes the number of predicted stems when comparing diameter distributions (Equation 6) and gives an overall explanation of the absolute deviations between estimated and observed number of trees in all diameter classes (Bollandsås & Næsset, 2007). Relative diameter distribution thus excludes any error in stem density estimation. Relative diameter distribution is a comparison between transects and circular plots that only takes into account which method captures the most accurate diameter distribution shape. Absolute diameter distribution includes errors in estimating the number of stems.

$$\text{Relative Error Index} = \sum_{1=i}^{n} 0.5 \left| \frac{f_i}{N} - \frac{\hat{f}_i}{\hat{N}} \right| \qquad (5)$$

Where f_i is the true and $\hat{f}_i$ the imputed number of stems in the diameter class i, n is the number of diameter classes, and N is the true and $\hat{N}$ the imputed total number of stems in the micro stand. 0.5 is a leverage that makes the error index be between 0 and 1; where 0 indicate a perfect fit and 1 that there is no distribution overlap at all.

$$\text{Absolute Error Index} = \sum_{1=i}^{n} 100 * \left| \frac{f_i - \hat{f}_i}{N} \right| \qquad (6)$$

Relative error index was calculated from the relative diameter distribution and absolute error index from the absolute diameter distribution (using the calculated tree area representation). Error indices were calculated for each validation plot by comparing the imputed diameter distribution from S1, S2, and S3 with the measured diameter distribution in the validation plots.

Sample Reduction

The size of the training data from each micro stand was reduced to 75%, 50%, 25% and one plot/transect from each micro stand (called 10%) to examine the threshold for minimum training data size. In all cases, except the one plot/transect per micro stand, the size of the training data from each stand was not reduced to contain fewer than four samples. The number of plots and number of trees from each sample reduction for each sample method is presented in (Table 2). Relative and absolute error indices were also calculated for the reduced sample sizes. Only one sampling was performed for each analysis on the reduced training data. The mean value, minimum value, maximum value was then calculated for the diameter distribution error indices.

Table 2. *The number of plots/transects and number of trees used for training data for the three sample methods S1, S2, and S3 when reducing the sample size from 100% to 75%, 50%, 25%, and 10%. Due to the samples having different number of plots/transects to begin with the relative sample size is an approximate value of the sample size used*

	S1 Plots	**S1 Trees**	**S2 Transects**	**S2 Trees**	**S3 Transects**	**S3 Trees**
100%	872	7585	829	7394	746	7394
75%	656	5828	625	5583	565	5619
50%	441	3748	420	3877	382	3810
25%	219	1940	207	1789	204	2081
10%	90	827	90	795	90	940

Results

Stand Imputation

The arithmetic mean diameter of the imputed micro stands was estimated with relative RMSE 22.4% and relative bias 2.2% for the circular inventory (S1 trees), relative RMSE 28.3% and relative bias 18.5% for the transect inventory (S2/S3 trees). The imputed basal area was estimated with a relative RMSE 55.5% and relative bias 33.2% for the circular inventory, relative RMSE 48.0% and relative bias 5.4% for the transect inventory.

Mean Diameter and Basal Area Weighted Mean Diameter

The inventory methods were used to estimate arithmetic mean diameter for the impute trees. Relative RMSE of the arithmetic mean diameter was 15.4%, 16.4%, and 17.5% for S1, S2, and S3 respectively (Fig. 2). Relative bias for arithmetic mean diameter was -4.8%, 6.3%, and 6.7% for S1, S2, and S3 respectively. The inventory methods were used to estimate basal area weighted mean diameter with relative RMSE 18.2%, 18.5%, and 20.9% for S1, S2, and S3 respectively. Relative bias for basal area weighted mean diameter was 10.7%, 12.8%, and 13.3% for S1, S2, and S3 respectively.

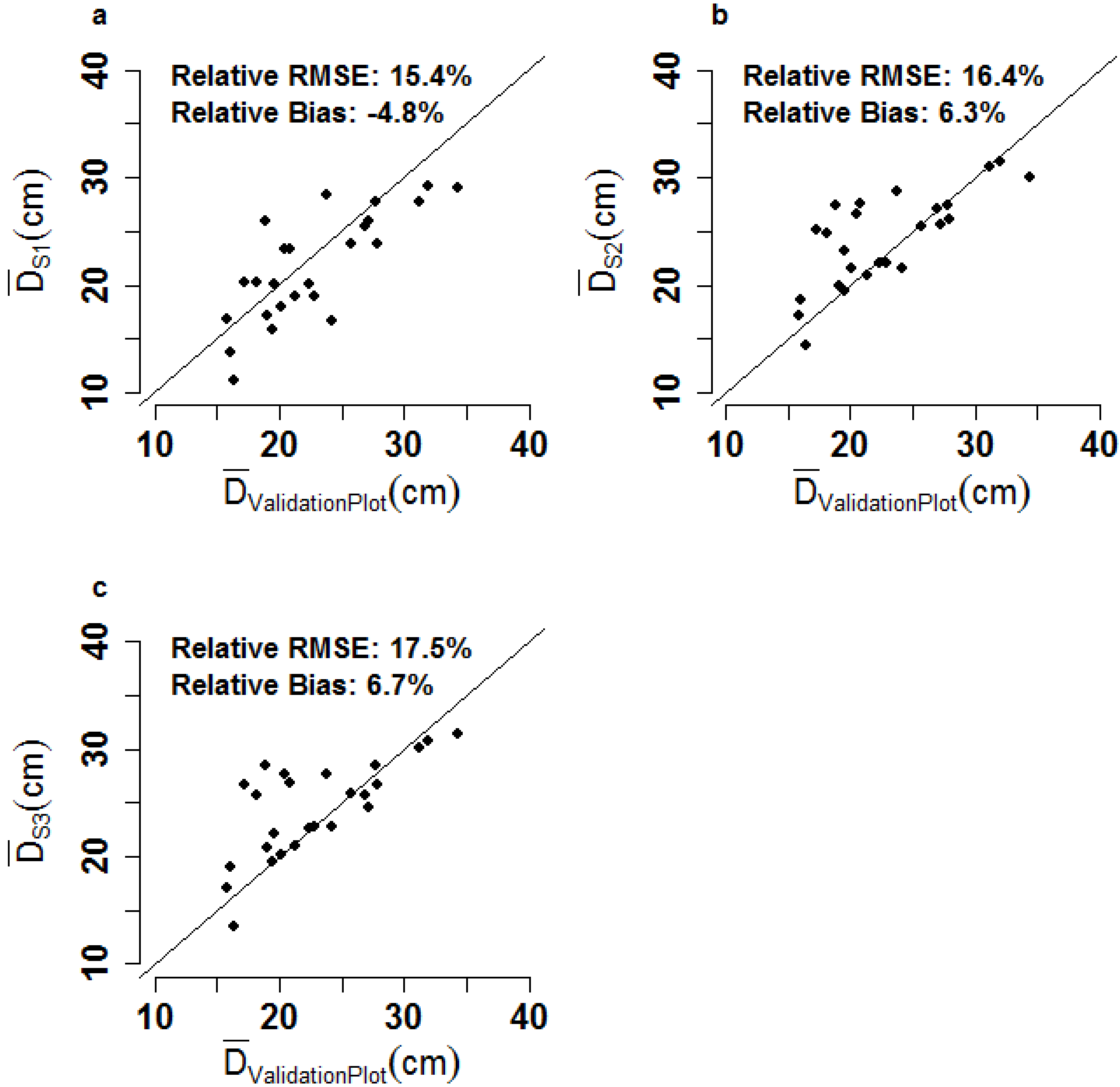


Figure 2. *Imputed arithmetic mean diameter on the validation plots, using training data from S1 (a), S2 (b), and S3 (c) plotted against arithmetic mean diameter on the validation plots. The number of validation plots (N) is 25, the straight line starts at the origin (0,0) and has the incline (1:1).*

Basal Area

Basal area was calculated from 1 cm diameter classes from imputed trees, the relative RMSE was 17.5%, 42.2%, and 34.2% for S1, S2, and S3 respectively (Fig. 3). The relative bias was 5.3%, 33.8%, and 29.2% for S1, S2, and S3 respectively.

.

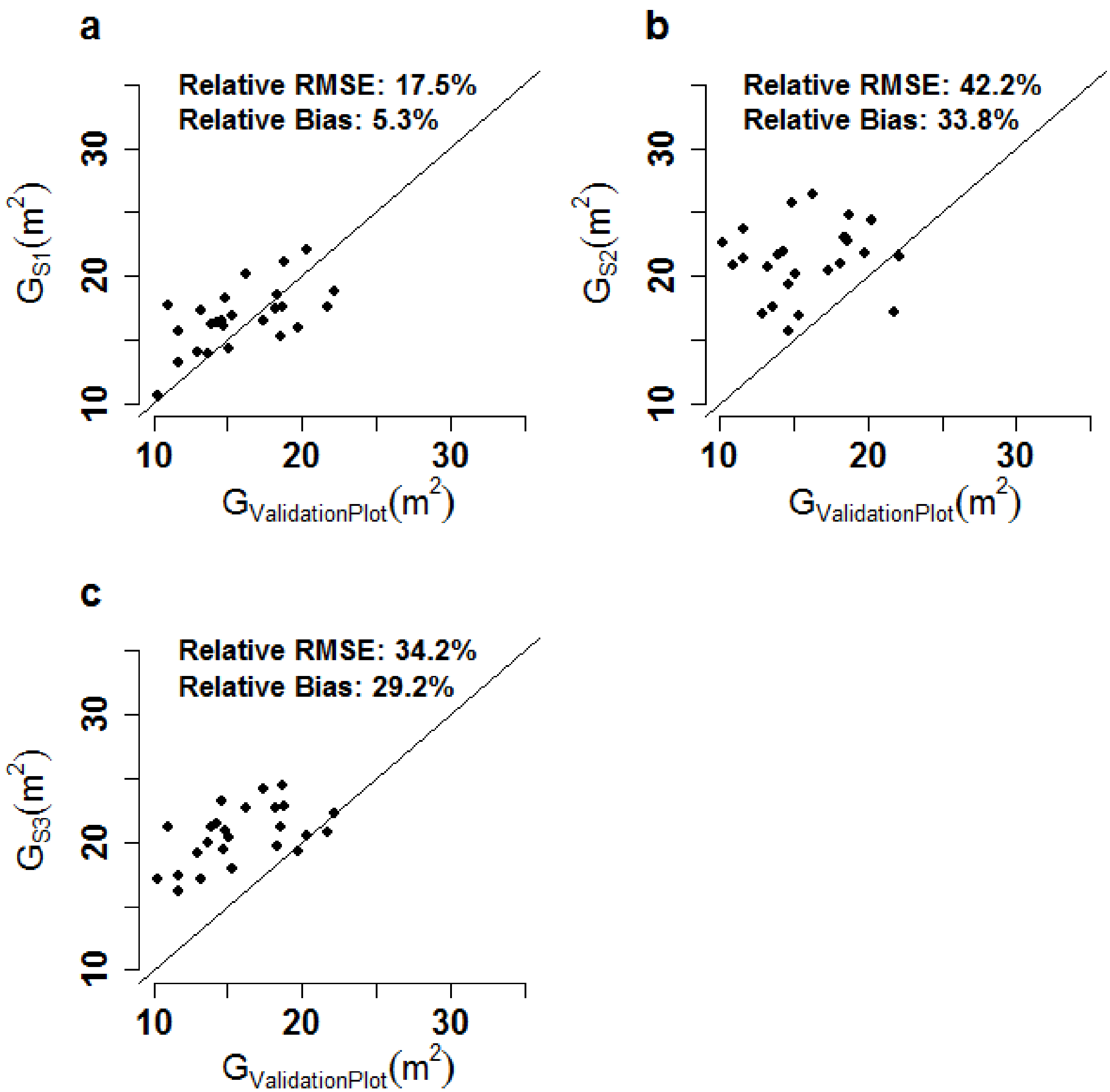


Figure 3*. Total basal area calculated from 1 cm diameter classes of the imputed trees using S1 (a), S2 (b), and S3 (c), plotted against the measured basal area on the validation plots. The number of validation plots (N) is 25, the straight line starts at the origin (0,0) and has the incline (1:1).*

Diameter Distribution and Sample Size Reduction

Estimated and validation plot diameter distribution was compared using absolute Error Index and was calculated for S1, S2, and S3 using the full set of training data, 75% of the training data from each micro stand, 50% of the training data from each stand, 25% of the training data from each stand, and one training plot from each stand (roughly 10% of the total training data) (Fig. 4). The absolute error index means for S1, S2, and S3 was 84.4, 91.9, and 82.2 respectively when the whole set of training data was used. The mean, minimum, and maximum error index was calculated for each sample size for each of the three sampling methods.

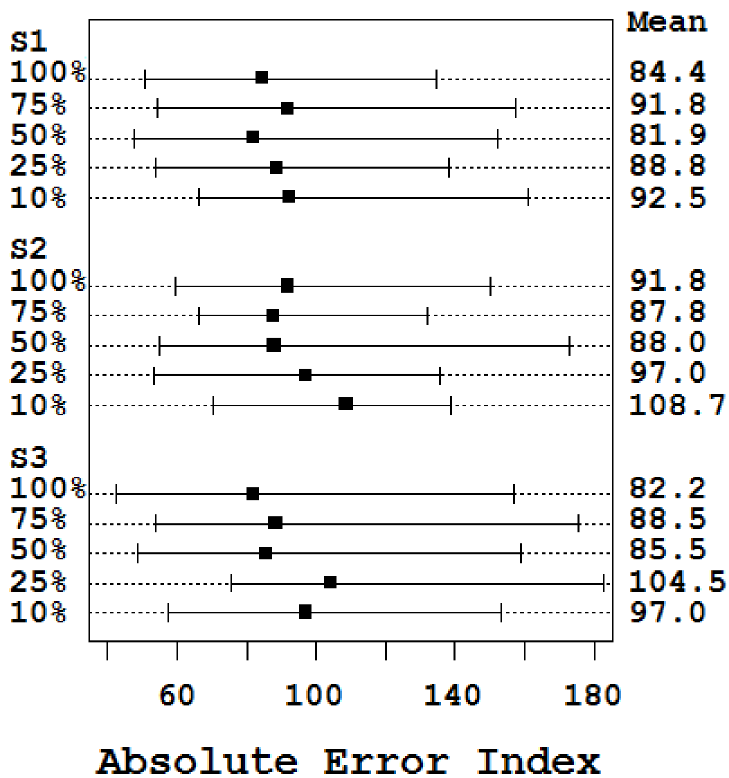


Figure 4. *Absolute Error Index for S1, S2, and S3. The diameter distribution was estimated using the full set of training data, 75%, 50%, 25% and 10% of the training data. Mean, min, and max values are presented in the graph to show the range of the error indices. The mean values are written to the right of the figure since it can be hard to read their exact value due to the scale of the x-axis. The number of validation plots (N) is 25.*

The relative diameter distribution was estimated since there was some uncertainty of how reliable the area representation was for S2 and S3. The relative error index means for S1, S2, and S3 was 0.38, 0.33, and 0.33 respectively when the whole set of training data was used. The estimated relative diameter distribution of the three sampling methods was compared to the observed relative diameter distribution (Fig. 5). For each sampling method the full set of training data was used and then reduced to 75% of the training data from each micro stand, 50% of the training data from each stand, 25% of the training data from each stand, and one training plot from each stand (roughly 10% of the total training data). The mean, minimum, and maximum error index was calculated for each sample size for each of the three sampling methods.

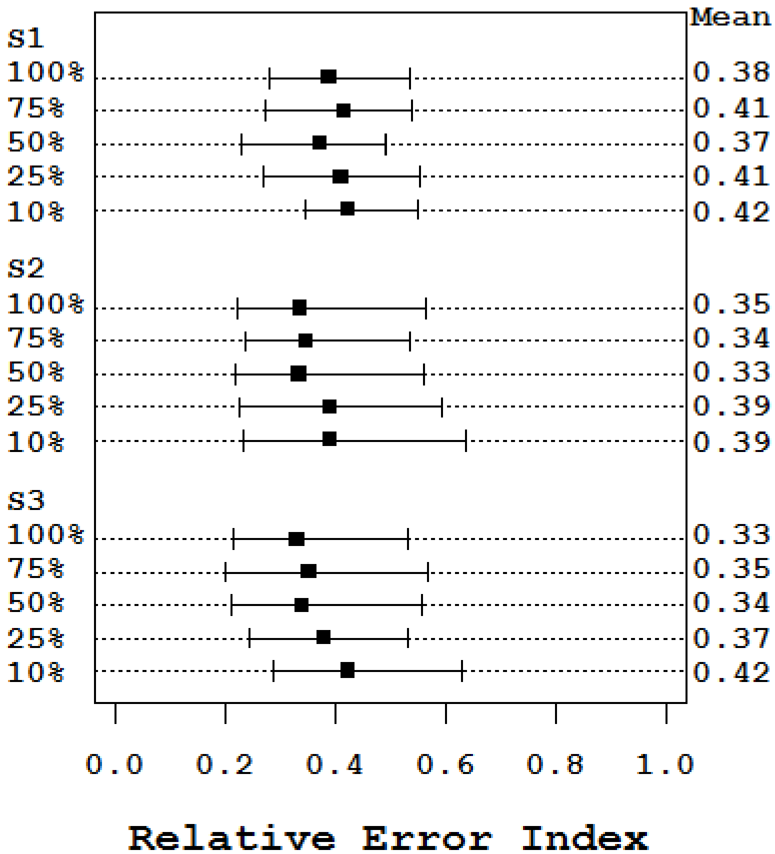


Figure 5. *Relative error index for S1, S2, and S3. The diameter distribution was estimated using the full set of training data, 75%, 50%, 25% and 10% of the training data. Mean, min, and max values are presented in the graph to show the range of the error indices. The error index varies between 0 and 1, where 0 is a perfect overlap of diameter distribution and 1 means there is no overlap at all. The mean values are written to the right of the figure since it can be hard to read their exact value due to the scale of the x-axis. The number of validation plots (N) is 25.*

Discussion

The micro stand imputation was not a very successful method in this study; there are many factors that account for this. The k was set to 1, which does not leave room for a lot of flexibility in terms of which trees are imputed. Maltamo et al. (2009) used k larger than one and gave the imputed trees a weight by inversing the kMSN distance, thus trees from more similar plots were given a higher weight that trees from a less similar plots. A similar weighing method should be used when imputing stands. The plot imputation used k set to one, but since one plot is imputed for every 10x10m raster cell in the validation plot the average of all imputed trees in a target plot could be seen as a comprehensive set of data similar to using a large k since there are 56 raster cells in each micro stand. There was also a small set of training data (90 micro stands), and the transect inventories (S2 and S3) did not have sufficient measurements of the stem density. Therefore it was decided to keep using plot imputation. There were also deviating parts of the forest in the reference stands which probably had an effect on using the mean values of the ALS metrics in a micro stand for the imputation.

This study focused on comparing transect inventory with traditional circular plot inventory in order to find a simpler and faster way of acquiring training data for estimating diameter distribution, using ALS data and kMSN imputation. There were little difference between the three sampling methods when the imputed arithmetic mean diameter and imputed basal area weighted mean diameter was compared to the validation plots. Mean diameter had slightly lower relative RMSE than basal area weighted mean diameter for all three samples. This is probably due to there being many large diameter trees in the training data that were imputed to the control (the largest tree in S2 and S3 had a diameter of 140 centimetres and trees with DBH larger than 50 cm were not uncommon). Arithmetic mean diameter was used as response variable for the imputation, with hindsight perhaps basal area weighted mean diameter would have been a better choice since it better represents the trees that are interesting for forest management. The relative RMSE for mean diameter and basal area weighted mean diameter were slightly higher than what has been achieved in other studies using kMSN imputation (Packalén & Maltamo, 2007, 2008; Maltamo et al. 2009; Vauhkonen et al. 2010). The higher relative RMSE is probably a result of this study not using algorithms to find the best predictor variables.

Accurate mean diameter can be estimated regardless of the number of trees used. However, assuming that the stem density was the same inside the circular plots (S1) as around them was not a very successful method the tree representation each imputed tree has, which can be seen in the basal area graphs. S2 and S3 (Fig. 3, b and c) have a very high relative RMSE and high relative bias for basal area and the graphs reveal no correlation pattern. S1 managed to estimate a somewhat similar basal area as that of the validation plots. When comparing the imputed basal area of S2 and S3 it can be seen that using shorter transects (S3) is a better method than using longer transects (S2).

It did not work to assume that the stem density was autocorrelating throughout the micro stand. This was probably due to how the locations of the S1 plots were decided with the Least Pivotal Method. S1 plots were positioned within raster cells where the laser metrics were deviating from the other raster cells in that micro stand. Small groups of deviating raster cells were included in the micro stand because they were too small to form a micro stand of their own. The stem density might be deviating only in that circular plot, but the stem density of that circular plot is assumed to be occurring throughout the entire transect that pass through the circular plot with deviating forest structure. If such a transect is also located in the homogenous parts of that stand, the trees in the homogenous part will be given a stem density that is calculated from a deviating part of the micro stand. Many deviating cells were populated by several small diameter trees (usually broadleaves). Using the stem density from these sites would cause each tree in the transect to have a higher representation factor, thus overestimating the imputed number of trees and imputed basal area. Another kind of deviating cells are small gaps in mature forests; such as a wet hollows, old forest roads, groups of wind thrown, dead trees due to insects and fungi etc. Whether the lower stem density is glaring or not it will give the trees in the transect a lower representation rate, thus underestimating the imputed number of trees and imputed basal area. The LPM is a good method for sampling on the entire spectrum of different forest types. It was a drawback of this study that the S1 plots (often positioned in deviating parts of the micro stand) were used to guide the transect location and that the assumption was made that the stem density is autocorrelating.

The radius of the S1 plots was set to 5 m which may have resulted in few trees per S1 plot in micro stands with large trees. In such forest it is possible that many trees are only partly inside the plot or very close to the plot. In micro stands with large trees this would result in a lowered stem density, resulting in underrepresentation of the transect trees representation factor. Thus, underrepresentation probably occurred in homogenous forest even if the counting plot was not located in a deviating part of the stand. However the results of basal area (Fig. 3) suggest overrepresentation was more commonly occurring in the transect inventory than underrepresentation. This could be explained with few trees being callipered in a transect due to the trees being spaced out, i.e. mature forest. If few trees are callipered in the approximate 2 m transect width the captured trees will have a large representation factor. This probably had a lot of influence on the results since most of the forest in the validation plots is mature forest. While the validation stands are mostly mature forest with some thinning forest, the micro stands where the training data was collected cover the whole spectrum of forest (broadleaf and conifer forests in all different development classes, bog forests, and copse wood). It is unclear if this would affect the results since the training data cover the type of forests that occur in the validation stands as well as other kinds of forests.

S3 was not collected properly. For S3, coordinates for ALS metrics extraction were supposed to be created every 10th tree. Over the whole dataset new GPS coordinates were created on an average of every 10.7th tree, with a range from 4 trees per coordinate to 23 trees per coordinate. These calculations were done on a micro stand level and trees were assigned to a coordinate based on the average number of trees per coordinate in that micro stand. Thus trees have been assigned to coordinates that they should not have belong to. If done properly S3 could be a very viable method. Though there were no exact comparison of time consumption it can be said that gathering the field data for S1 and S2/S3 took 4 weeks and 2 weeks respectively. S1 had a struggling differential GPS that delayed the fieldwork, S2 and S3 did not have to deploy any counting areas. It can still be assumed that transect inventory method is faster than circular plot inventory. The shorter time consumption will probably make transect inventories more interesting in the future, especially for large scale forest inventories.

In the report done by (Hansson, 1999) the criteria for using transect inventory was that each transect in each stand should be at least 1000 m and consist of 10 counting plots. Using a similar counting plot density, the distance between callipered tree and ALS extraction point is never more than 50 m. The study focused on measuring diameter distribution prior to regeneration harvest and used counting plots with a 7.98 m radius ($200m^2$) since larger trees tend to be more spaced out. In the planning phase of the field work it was determined that using such a high number of counting plots would not result in a faster inventory method. Hansson (1999) reported that both inventory methods were equal in time consuming. It should be pointed out that Hansson (1999) wanted to make clear estimates on stand level to estimate the diameter distribution using conventional field inventory. This study examines faster methods of gathering comprehensive enough training data to make accurate estimates for the rest of the forest. However, a larger number of counting plots than what was used in this study is needed to accurately estimate stem density. The counting plot radius could be adjusted according to forest type, i.e. smaller radius in forest that has never been thinned and larger radius in forest close to regeneration harvest.

Figure 4 shows that the method used in this study managed to estimate similar accuracy on the true diameter distribution regardless of which training data was used, even if the tree representation of the transect trees was calculated poorly. The error indices (both absolute and relative) achieved in this study are almost identical to those achieved in a Norwegian study (Maltamo et al. 2009). Maltamo et al. (2009) achieved absolute error indices in the range of 77.85 - 93.3 and relative error indices in the range of 0.35-0.39 when cross validating the data for plots stratified into four kinds of forests. Similar ranges of error indices have been achieved in this study (Fig. 4, 5). When comparing the absolute error indices (Fig. 4) there is only a small difference between S1, S2, and S3. The ranges for the absolute error indices are similar for all three sampling methods (Fig. 4). The transect inventory methods managed to capture a relative diameter distribution that is closer to the truth (Fig. 5) than the circular inventory method. However, the relative error indices from the transect inventories have a larger range than the range of the relative error indices from circular plot inventory (Fig. 5).

Reducing the sample size did not affect the accuracy until the sample size became very small (10% of training data, i.e. 90 plots/transects). This is similar to the result of other studies that have examined the effect on the accuracy of the estimates by reducing the size of the training data (Maltamo et al. 2009; Vauhkonen et al. 2010; Maltamo et al. 2011). These studies suggest that a set of training data should not be smaller than 100 plots. Since only one sampling was made, the sampling is subjected to the chance of an unfortunate sampling. In a study by Maltamo et al. (2009) the sampling was repeated 100 times for each sample reduction. They presented the mean value of all accuracies to eliminate the chance of sampling a bad sample.

When estimating the diameter distribution all three samples managed to impute a lot of small diameter trees (<7 cm DBH) that did not exist in the control plots. It is likely that the small trees were not captured in the ALS data and were imputed along with the trees that were actually representative for the most similar ALS metrics. Removing the small trees would likely have resulted in smaller error indices.

Future research and improvement of the transect inventory

To improve the applicability of transect inventory as a method for collecting training data it would be preferred to use shorter transects. With shorter transects the distance between callipered trees and extracted ALS metrics is reduced. If the transect is in a homogenous forest and there are many counting plots for accurate stem density there is probably less need to have short distances between callipered tree and extracted ALS metrics. It would be a great benefit to use ALS data as a priori information to guide the location of the transect inventory. In this study the ALS data was used to guide the locations of the S1 plots only. Ståhl et al. (1997) proposed a guided transect sampling to inventory sparse, geographically scattered populations. A similar method could be applied for transect inventory to ensure that a comprehensive enough set of training data is collected. The results could also be improved by using a larger number of ALS metrics or even transforming the already present metrics; this study used only seven laser metrics. Latifi et al. (2010) applied an algorithm that is used in medical science to find the genes with the highest correlation for selected attributes. This gene algorithm was used to select which predictor ALS variables to use based on their correlation with the response variables. Other studies (Maltamo et al. 2006; Hudak et al. 2008) use optimization algorithms that stepwise removes independent variables to minimize the relative RMSE of the results. It is also possible to increase the number of field measured response variables such as basal area, height, stem number, and volume. Maltamo et al. (2009) found that lower error indices were achieved using more response variable. They also used transformed variables (squared volume) and percentiles of the diameter but the diameter percentiles did not improve the predictions as they have done in other studies (Maltamo et al. 2009).

Conclusion

This study evaluated different inventory methods for collecting training data for kMSN imputation of diameter distribution. The results showed that training data from transect inventory can be used to achieve similar accuracy of diameter distribution (both absolute and relative) as training data from circular plot inventory. Transect inventory is a faster method than circular plot when collecting training data. The calculation of transect tree representation (from the stem density in representing counting plots) has to be refined in order to improve the accuracy of the method. Since small sets of training data (100 plots/transects, compared to the 800 plots/transects used in this study) can be used without reducing the accuracy, more time could be spent on calculating transect tree representation more accurately by collecting more counting plots.

Acknowledgements

The field work for this thesis was a part of a project funded by The Swedish Forest Society Foundation. I would like to thank all the people who contributed to this work and helped me finish writing my thesis. First of all I would like to thank my supervisors; Johan Holmgren for being extremely helpful in all aspects throughout the work of this thesis. Jörgen Wallerman for helping me out with some essential programming I had issues with, Anna Ringvall for her comments on the linear inventory method, prof. Leblon for her suggestions of how to improve the thesis. Thanks to Håkan Olsson and Heather Reese for all the insightful and interesting conversations. I would also thank my father, Erik Lundholm, for reviewing my thesis and commenting from the perspective of an academic who does not work with ALS.

Thanks to all my friends who helped me clear my mind during coffee breaks and provided good laughs, you really helped making this a fun time. Thanks to my family for always being supportive and helping me out. Last but not least, thanks to my loving fiancé Laurissa Wilson for encouraging me, brighten up my day, and always keeping me determined to keep on working with my thesis.

References

Bollandsås, O.M. & Næsset, E. (2007) Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data. *Scandinavian Journal of Forest Research*, 22, 33-47.

Bollandsås, O.M., Maltamo, M., Gobakken, T. & Næsset, E. (2013) Comparing parametric and non-parametric modeling of diameter distributions on independent data using airborne laser scanning in a boreal conifer forest. *Forestry*, 86, 493-501.

Crookston, N. L., Moeur, M. & Renner, D. (2002) Users Guide to the Most Similar Neighbor Imputation Version 2. 2002:General Technical Report RMRS-FTR-96. Rocky Mountain Research Station, US Department of Agriculture, Forest Service, Fort Collins, Colorado, USA.

Crookston, N.L. & Finley, A.O. (2008) yaImpute: An R Package for kNN Imputation. *Journal of Statistical Software*, 23 (10), 1-16.

Fries, C., Johansson, O., Pettersson, B. & Simonsson, P. (1997) Silvicultural models to maintain and restore natural stand structures in Swedish boreal forests. *Forest Ecology and Management*, 94, 89-103.

Grafström, A., Lundström, N. & Schelin, L. (2012) Spatially Balanced Sampling through the Pivotal Method. *Biometrics*, 68(2), 514-520.

Gupta, S., Koch, B. & Weinacker, H. (2010a) Single tree modeling using full waveform lidar data in a complex forest. In: Wagner W. & Székely, B. (Eds.) ISPRS TC VII Symposium – 100 Years ISPRS, Vienna, Austria, July 5–7, 2010, IAPRS, Vol. XXXVIII, Part 7B. 249-254.

Gupta, S., Koch, B. & Weinacker, H. (2010b) Single tree detection using full waveform laser scanner data. In: Koch, B., Kändler, G. & Teguem, C. (Eds.) Silvilaser 2010 conference, 14th-17th Sept., Freiburg, Germany. 368-376.

Hansson, F. (1999) Inventering före avverkning – metodval och resursåtgång. (Eng: Inventory before regeneration felling - choice of method and resource usage). 1999:15. Skogforsk, Uppsala, Sweden.

Holmgren, J. & Lindberg, E. (2013) Tree crown segmentation based on a geometric tree crown model for prediction of forest variables. *Canadian Journal of Remote Sensing*, 39 (S1), 86-98.

Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E. & Falkowski, M.J. (2008) Nearest neighbor imputation of species-level. plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment*, 112, 2232-2245.

Johnson, R. A. & Wichern, D. W. (2007) Applied Multivariate Statistical Analysis, 6th edition. Upper Saddle River: Pearson Education Inc.

Kaartinen, H., Hyyppä, J., Yu, X., Vastaranta, M., Hyyppä, H., Kukko, A., Holopainen, M., Heipke, C., Hirschmugl, M., Morsdorf, F., Næsset, E., Pitkänen, J., Popescu, S., Solberg, S., Wolf, B.M. & Wu, J.-C. (2012) An International Comparison of Individual Tree Detection and Extraction Using Airborne Laser Scanning. *Remote Sensing*, 4, 950-974.

Kim, S., McGaughey, R.J., Andersen, H.E. & Schreuder, G. (2008) Individual Tree Species Identification Using LiDAR Intensity Data. In: Proceedings of 2008 ASPRS Annual Conference, Portland, Oregon.

Ko, C., Sohn, G. & Remmel, T.K. (2013) Tree genera classification with geometric features from high-density airborne LiDAR. *Canadian Journal of Remote Sensing*, 39 (S1), 73-85.

Koch, B., Straub, C., Dees, M., Wang, Y. & Weinacker, H. (2009) Airborne laser data for stand delineation and information extraction. *International Journal of Remote Sensing*, 30(4), 935-963.

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, 4, 395-407.

Lim, K., Treitz, P., Wulder, M., St-Onge, B. & Flood, M. (2003) LiDAR remote sensing of forest structure. *Progress in Physical Geography*, 27 (1), 88-106.

Leppänen, V. J., Tokola, T., Maltamo, M., Mehtätalo, L., Pusa, T. & Mustonen, J. (2008) Automatic Delineation of forest stands from LiDAR data. In: *Conference presentation in GEOBIA.*

Maltamo, M., Bollandsås, O.M., Næsset, E., Gobakken, T. & Packalén, P. (2011) Different plot selection strategies for field training data in ALS-assisted forest inventory. *Forestry*, 84 (1), 23-31.

Maltamo, M., Malinen, J., Packalén, P., Suvanto, A. & Kangas, J. (2006) Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. *Canadian Journal of Forest Research*, 36, 426-436.

Maltamo, M., Næsset, E., Bollandsås, O.M., Gobakken, T. & Packalén, P. (2009) Non-parametric prediction of diameter distributions using airborne laser scanner data. *Scandinavian Journal of Forest Research*, 24, 541-553.

Moeur, M. & Stage, A.R. (1995) Most Similar Neighbor: An Improved Sampling Inference Procedure for Natural Resource Planning. *Forest Science*, 41 (2), 337-359.

Mustonen, J., Packalén, P. & Kangas, A. (2008) Automatic segmentation of forest stands using a canopy height model and aerial photography. *Scandinavian Journal of Forest Research*, 23, 534-545.

Nelson, R. (2013) How did we get here? An early history of forestry lidar. *Canadian Journal of Remote Sensing*, 39 (S1), 6-17.

Packalén, P. & Maltamo, M. (2008) Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. *Canadian Journal of Forest Research*, 38, 1750-1760.

Packalén, P. & Maltamo, M. (2007) The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photography. *Remote Sensing of Environment*, 109, 328-341.

Persson, Å., Holmgren, J. & Söderman, U. (2002) Detecting and Measuring Individual Trees Using an Airborne Laser Scanner. *Photogrammetric Engineering and Remote Sensing*, 68 (9), 925-932.

Popescu, S.C., Wynne, R.H. & Nelson, R.F. (2003) Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, 29, 5, 564-577.

R Development Core Team. (2013) 2014-01-12-last update, R: A Language and Environment for Statistical Computing. [Homepage of R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.] [Online]. Available at: http://www.R-project.org. [Retrieved: 2014-01-12].

Reitberger, J., Heurich, M. & Krzystek, P. (2010) Estimation of stem volume by using 3D tree segments derived from full waveform LiDAR data. In: Koch, B., Kändler, G. & Teguem, C. (Eds.) Silvilaser 2010 conference, 14th-17th Sept., Freiburg, Germany.

Reynolds, M.R., Jr., Burk, T.E. & Huang, W.-C. (1988) Goodness-of-Fit Tests and Model Selection Procedures for Diameter Distribution Models. *Forest Science*, 34 (2), 373-399.

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, 945-961.

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.

Wulder, M.A., Coops, N.C., Hudak, A.T., Morsdorf, F., Nelson, R., Newnham, G. & Vastaranta, M. (2013) Status and prospects for LiDAR remote sensing of forested ecosystems. *Canadian Journal of Remote Sensing*, 39 (S1), 1-5.