Prediction of diameter distributions in boreal forests using remotely sensed data

Janne Räty
School of Forest Sciences
Faculty of Science and Forestry
University of Eastern Finland

Academic dissertation

To be presented, with the permission of the Faculty of Science and Forestry of the University of Eastern Finland, for public criticism via an online Lifesize conference, on 5th June 2020, at 12 o'clock noon

Title of dissertation: Prediction of diameter distributions in boreal forests using remotely sensed data

Author: Janne Räty

Dissertationes Forestales 294

Thesis Supervisors:
Professor Matti Maltamo
University of Eastern Finland, Faculty of Science and Forestry, School of Forest Sciences, Finland

Professor Petteri Packalen
University of Eastern Finland, Faculty of Science and Forestry, School of Forest Sciences, Finland

Pre-examiners:
Docent Kalle Eerikäinen
Metsähallitus Forestry Ltd., Finland

Doctor Michele Dalponte
Research and Innovation Center, Fondazione Edmund Mach, Italy

Opponent:
Doctor Jean-Pierre Renaud, Office National des Forêts, France

ABSTRACT

Diameter distributions are usually characterized in forest management inventories using probability density functions (PDF). Depending on the inventory method, the PDF parameters are derived using either predicted or assessed forest attributes. The application of PDF is not essential for forest inventories that rely on remotely sensed data, because the diameter distributions can be predicted using empirical tree lists via the nearest neighbor (NN) approach. This thesis comprises three objectives. The general aim is to investigate NN-based prediction of diameter distributions in Finnish forest inventories. Firstly, the response configurations of the NN approach were examined in the prediction of species-specific diameter distributions. Secondly, different remote sensing datasets were utilized in the prediction of diameter distribution for logwood-sized trees. For example, bitemporal and multispectral airborne laser scanning (ALS) datasets were compared to the Finnish forest inventory standard in which unispectral ALS and aerial images are used. Thirdly, two approaches that fuse an area-based approach (ABA) and individual-tree detection (ITD) in the prediction of diameter distributions were proposed. The results showed that the standard response configuration used in NN imputation is suboptimal if diameter distributions are of interest. The findings also indicate that the multispectral ALS dataset performs poorly in the prediction of logwood volumes by tree species. Instead, the use of bitemporal ALS (leaf-off and leaf-on) data provide almost comparable error rates with the use of ALS data and aerial images in the prediction of logwood volumes by tree species. The ABA-ITD fusion of diameter distributions provided slight improvements in the mean error index values associated with the predicted diameter distributions. It should be noted, however, that ITD is more sensitive to errors than ABA, for example, in forests with a bimodal or descending diameter distribution. Structural analysis of forests using ALS data is a possible indicator for the selection of prediction approach. The pulse density of the national ALS data will be increased in the 2020s, which opens up the possibility to apply the ABA-ITD fusion approach in practical applications.

ACKNOWLEDGEMENTS

Joensuu, 21st April 2020

LIST OF ORIGINAL ARTICLES

This thesis comprises three research articles. They are referred to Roman numerals in the thesis. The papers are reproduced with permission provided by the publishers.

I Räty J., Packalen P., Maltamo M. (2018). Comparing nearest neighbor configurations in the prediction of species-specific diameter distributions. Annals of Forest Science 75(26): 1–16. https://doi.org/10.1007/s13595-018-0711-0

II Räty J., Packalen P., Maltamo M. (2019). Nearest neighbor imputation of logwood volumes using bi-temporal ALS, multispectral ALS and aerial images. Scandinavian Journal of Forest Research 34(6): 469–483. https://doi.org/10.1080/02827581.2019.1589567

III Räty J., Packalen P., Kotivuori E., Maltamo M. (2020). Fusing diameter distributions predicted by an area-based approach and individual-tree detection in coniferous-dominated forests. Canadian Journal of Forest Research 50: 113–125. https://doi.org/10.1139/cjfr-2019-0102

Janne Räty was the first and corresponding author in all research articles. He also implemented the analyses, wrote the manuscripts and participated in the planning of the study design. Together with Prof. Matti Maltamo and Prof. Petteri Packalen, Janne Räty also developed the initial research ideas. Profs. Maltamo and Packalen commented and refined the manuscripts as supervisors. Eetu Kotivuori contributed to the analyses carried out in study **III** and provided comments on the manuscript of study **III**.

TABLE OF CONTENTS

ABBREVIATIONS

ABA	Area-based approach
ALS	Airborne laser scanning
CHM	Canopy height model
DBH	Diameter at breast height
DDSC	Diameter distribution shape class
DGM	Diameter of basal area median tree
F.BEST	Fusion approach, theoretical (best possible)
F.PRED	Fusion approach based on predicted weights
F.REPL	Fusion approach based on a replacement
GCAE	Gram-Charlier A-series expansion
GNSS	Global navigation satellite system
GSD	Ground sampling distance
HGM	Height of basal area median tree
IMU	Inertial measurement unit
ITD	Individual tree detection
LiDAR	Light detection and ranging
MD	Mean difference
MLR	Multinomial logistic regression
MSN	Most similar neighbor
M–ALS	Multispectral airborne laser scanning data
M–CH2–ALS	Airborne laser scanning data (the 2nd channel of M–ALS)
NN	Nearest neighbor
PDF	Probability density function
RMSE	Root mean square error
SA	Simulated annealing
S11–ALS	Airborne laser scanning data collected in 2011
S16–ALS	Airborne laser scanning data collected in 2016
TLS	Terrestrial laser scanning
UAV	Unmanned aerial vehicle

ERRATA

In paper **I**, the caption of Figure 5 belongs to Figure 6 and vice versa.

1 INTRODUCTION

1.1 Tree size distributions in Finnish forest management

The principal goal of a forest management inventory is to comprehensively describe forests for the purposes of forest management planning. Typically, forest attributes (such as volume basal area and dominant height) are employed in the planning of silvicultural operations in homogenous forest units, namely stands, in a forest estate. These forest attributes can be easily seen as the most valuable outputs of forest management inventories, although tree size distribution provides a more comprehensive description of a forest. Tree size distributions can be characterized using various attributes, for example, diameter at breast height (DBH), height, or volume. The use of DBH is usually justified by the tree-level predictive models that employ DBH as the primary predictor variable (Laasasenaho 1982; Repola 2009). Tree size distribution described using DBH can be considered as a multivariate forest attribute from which individual forest attributes originate. In this thesis, the focus is on tree size distributions characterized using DBH (i.e. diameter distributions).

The characterization of diameter distributions is simpler in a managed forest than in a natural forest because the silvicultural operations simplify forest structures (Rouvinen and Kuuluvainen 2005). Managed Finnish forest stands can be even-aged monocultures, although several tree species can also be grown simultaneously in distinct canopy layers. Therefore, the diameter distribution of Finnish managed forests usually has up to two distribution peaks (without considering tree species). Still, in many cases, the shape of the diameter distribution in a managed even-aged forest stand is unimodal, i.e., Gaussian-shaped distribution, due to the rapid regeneration after a clear-cut (Rouvinen and Kuuluvainen 2005). In natural conditions, Gaussian-shaped diameter distributions are rare because of the disturbance events that usually affect only a small proportion of trees in a forest at a time. Therefore, the forest structures are irregularly disturbed, both spatially and temporally, thereby leading to more complex diameter distributions than in managed forests (Kuuluvainen 2002). For example, descending (reverse J), multimodal (e.g. bimodal) and irregular are possible shapes for diameter distributions in forests that are not under active silvicultural management. Moreover, bimodal-shaped diameter distributions are common in uneven-aged managed forests.

Diameter distributions by tree species are required in Finnish forest inventories due to timber procurement requirements, and growth and yield modeling that requires tree species-specific information. The separation of forest attributes by tree species is possible due to the small number of commercial tree species. Scots pine (*Pinus sylvestris* (L.)), Norway spruce (*Picea abies* (L.) Karst.), and the deciduous species group are separated in Finnish forest management inventories. The deciduous species group mainly consists of silver birch (*Betula pendula* Roth), and downy birch (*Betula pubescens* Ehrh.), as well as aspen (*Populus tremula* (L.)) and alder (*Alnus* spp.).

Diameter distribution is a key component in the forest simulators used in forest planning. Forest simulators include the models required to simulate, for example, future growth. Finnish forest simulators, such as MELA and MOTTI, usually use tree-level growth models (Hynynen et al. 2002). It is evident that errors in the initial diameter distributions may multiply during the planning period. Therefore, the errors may lead to inoptimality losses because of the suboptimal schedule of the silvicultural treatments. Diameter distributions are also important descriptors of the structural characteristics associated with a forest. For

example, the layered vertical structure of a forest and the variability of tree sizes are indicative of biodiversity values (Esseen et al. 1997). Forest management inventories rarely focus on the characterization of the vertical structure of forests. Diameter distributions correlate with the height distribution of a forest (Loetsch et al. 1973), and therefore diameter distributions can be used as an indicator for biodiversity values.

1.2 Development of diameter distribution modeling techniques

The comprehensive description of diameter distribution in a forest stand requires that DBH is measured for each tree. It is a common practice to omit the smallest trees in forest inventories (Keränen et al. 2015). In Finland, diameter distribution is typically truncated using a minimum DBH limit of 5 cm. Despite the exclusion of the smallest trees, diameter distributions are laborious to measure in the field. Several techniques for the determination of diameter distributions without laborious field measurements have been developed (e.g. Päivinen 1980; Kilkki et al. 1989). In Finland, the planning of forest management has strongly relied on inventories by forest stands (i.e. inventories by compartments or stand-level inventories) that have also established guidelines for the development of diameter distribution modeling. Inventories by forest stands were the principal forest inventory approach in Finland from the beginning of the 20th century to the 2010s. Inventories by forest stands were carried out by visiting each forest stand of a forest estate. The inventories were based on parceling out the forest estates according to homogenous forest stands. This approach required a land survey, until around the mid-20th century when aerial images were available for the delineation of forest stands (Koivuniemi and Korhonen 2006). The field assessments were based on visual assessments and height measurements. Later, a relascope was used to measure angle count samples, which enabled more objective stem density and basal area measurements than the visual assessments. The volume of growing stock was estimated by means of stand volume tables, and height and basal area measurements (Nyyssönen 1954). The stand volume tables were applied up to the 1980s, when theoretical diameter distributions were introduced for the total growing stock (Päivinen 1980). Later in the 1990s, the theoretical diameter distributions were also augmented, so that each tree species and all canopy layers were considered separately in the inventories by forest stands (Mykkänen 1986; Kilkki et al. 1989; Siipilehto 1999). Hence, the inventories in each forest stand were based on angle-count samples, the measurements of mean attributes, and visual assessments (Kangas et al. 2004). The measured forest attributes, such as basal area, mean diameter and age, were used in the construction of a theoretical diameter distribution.

The theoretical diameter distribution model is usually based on a probability density function (PDF) that is controlled by two or three parameters. Various PDF have been applied in the modeling of diameter distributions. In Finland, Cajanus (1914) was the first to apply a PDF in the description of diameter distributions; he applied a Gram-Charlier distribution, which is a sophisticated version of the Gaussian distribution. Later, several PDFs have been proposed for diameter distributions, such as Weibull (Bailey and Dell 1973; Kilkki et al. 1989), Johnson S_b (Hafley and Schreuder 1977; Siipilehto 1999) and Beta (Loetsch et al. 1973; Päivinen 1980). The capability to describe the various shapes of the diameter distributions depends on the function (Wang and Rennolls 2005). Generally, PDF can only describe unimodal diameter distributions, although the flexibility to describe ascending (J-shaped) and descending (reverse-J-shaped) distributions differs between PDF. A two-parameter Weibull function has emerged as the most commonly used PDF in the modeling

of diameter distributions (Siipilehto 1999). The two-parameter Weibull can describe unimodal (Gaussian-shaped), and ascending and descending diameter distributions as well. The multimodality of diameter distributions can be dealt with separate modeling of canopy layers (Lönnroth 1925; Cao and Burkhardt 1984) or with mixture models (Liu et al. 2002).

The PDF parameters are usually predicted or recovered by employing easily measurable forest attributes as predictor variables (Kilkki and Päivinen 1989; Siipilehto 1999; Siipilehto and Mehtätalo 2013). The parameter prediction approach requires predictive models for function parameters. For example, Siipilehto (1999) and Kilkki et al. (1989) have proposed models for the parameters of the Weibull function. Those models apply forest attributes as predictor variables. Instead, the parameter recovery approach utilizes the mathematical relationships between forest attributes and the parameters of a PDF. The recovery equations can involve forest attributes (Siipilehto and Mehtätalo 2013), moments (Burk and Newberry 1984) or percentiles (Bailey and Dell 1973). Studies have also proposed approaches that are not tied to the application of a PDF, e.g., a percentile (Borders et al. 1987; Kangas and Maltamo 2000a) or a nearest neighbor (NN) approach (Haara et al. 1997). The latter is referred to as an imputation method.

Modeling an unweighted diameter distribution is referred to as a stem-frequency distribution. Diameter distributions can also be weighted. The angle-count sampling applied in the inventories by forest stands is based on the measurement of basal area. Therefore, the diameter distributions have usually been weighted by basal area in order to be consistent with the angle count samples. The usage of basal area weighted diameter distribution is justified by the assumption that the basal area weighting places more emphasis on logwood-sized trees than unweighted distributions (Päivinen 1980). Later, Maltamo et al. (2007) showed that the benefit of the basal area weighting is negligible in terms of the error associated with timber volume compared to unweighted (i.e. stem frequency) diameter distribution. Therefore, the focus of this thesis is on stem-frequency diameter distributions.

1.3 Era of remote sensing-based forest inventories

Remote sensing was permanently integrated into boreal forest inventories during the 21st century (Næsset 2014; Maltamo and Packalen 2014). The most remarkable technology from the point of view of modern forest inventories has been the light detection and ranging (LiDAR), which can measure the distance between a target and the LiDAR sensor (Wehr and Lohr 1999). The scanning LiDAR sensors can be operated from airborne vehicles equipped with a global navigation satellite system (GNSS) receiver, inertial measurement unit (IMU) and computational unit with storage capacity. Due to technological developments in GNSS and IMU, the scanning airborne LiDAR system can accurately measure georeferenced 3D point clouds (Nelson 2013). Airborne LiDAR is henceforth referred to as airborne laser scanning (ALS) in this thesis. The acquisition of ALS data for forestry purposes (in Finland) is carried out over large areas covering several thousands of forest estates at a time (100,000 – 500,000 ha), which has improved the cost-efficiency of forest management inventories (Maltamo and Packalen 2014). In this thesis, forest inventories that rely on remote sensing are referred to as ALS-based forest inventories, since ALS data are typically the most notable remote sensing data source.

The ALS-based forest inventories can be divided into two fundamental categories: area-based approach (ABA; see Nilsson 1996; Næsset 1997), and individual-tree detection (ITD; see Brandtberg 1999; Hyyppä and Inkinen 1999). Approaches that combine characteristics

from ABA and ITD have also been proposed (Breidenbach et al. 2010; Lindberg et al. 2010; Packalen et al. 2015). Globally, the most applied approach to predict forest attributes is ABA. The ITD approach relies on the detection of trees and, therefore, requires high-density (> 5 pulses · m⁻²) ALS data. To date, the application of ITD in most operational forest inventories has been rare.

The ALS-based forest inventories that rely on ABA have been conducted operationally for many years, for example, in Finland and Norway (Maltamo and Packalen 2014; Næsset 2014). The ALS-based forest inventories typically employ field sample plots that are used as training observations for the construction of predictive models. The features derived from remotely sensed datasets are required as predictor variables in the predictive models. The prediction relies on the statistical relationships between field measurements and remotely sensed data. For example, in Finland, the predictive models that include predictor variables derived from ALS data and aerial images are applied to predict forest attributes for 16×16 m grid cells, following the wall-to-wall principle, in the inventory area (Maltamo and Packalen 2014). In Finnish inventories, the inventory areas typically cover hundreds of thousands of hectares and are inventoried during two sequential years. Finally, the grid cells can be used to aggregate predicted forest attributes for larger units, such as forest stands. Field measurements and additional visits are sometimes needed, especially in seedling and sapling stands. Nevertheless, cost savings of 60 % are assumed when the traditional inventories by forest stands are replaced with ALS-based forest inventories in Finland (Maltamo and Packalen 2014). The most important reason for such a cost saving is the reduced amount of fieldwork compared to the traditional stand-level inventory procedure. Inventories by forest stands require a visit to each forest stand (e.g. 100,000 stands per inventory area), whereas only a sample of field plots are required in the ALS-based forest inventories. Moreover, the ALS-based forest inventory achieves smaller or similar error rates associated with forest attributes than the inventories by forest stands when the tree species are not considered (see Haara and Korhonen 2004; Wallenius et al. 2012). The remote sensing-based recognition of tree species is an issue in mixed forests, in which the inventories by forest stands usually outperform the ALS-based forest inventories in terms of the prediction errors associated with the minor tree species.

1.4 ALS-based forest inventories by tree species

Height-related features derived from ALS data are the most applied features in the prediction of diameter distributions and forest attributes. However, height-related features do not provide adequate discrimination between tree species in ALS-based forest inventories. Modern ALS devices also record echo intensities that have been applied in the classification of tree species (Korpela et al. 2009; Vauhkonen et al. 2014). Echo intensities are usually determined from the amplitude of return echo. The amplitude describes the power of the laser echo returned to the ALS system (Wagner et al. 2006). However, previous studies have indicated that the compound of intensity and height-related features may not provide satisfactory error rates associated with the prediction of species-specific forest attributes (Räty et al. 2016; Kukkonen et al. 2019a).

Remote sensing data sources other than ALS data (height and intensity features) can improve the predictive performance of species-specific models (Packalén and Maltamo 2006; Dalponte et al. 2014; Vauhkonen et al. 2014; Kukkonen et al. 2018). Multispectral aerial images are most typically employed in species-specific forest inventories (Næsset 2014;

Maltamo and Packalen 2014) because reflectance recorded in the near-infrared region of the spectrum differs significantly between coniferous and deciduous forests. Moreover, the combination of ALS data and hyperspectral aerial images have been proved to discriminate between tree species (Dalponte et al. 2013; Dalponte et al. 2014). The joint usage of ALS data and aerial images essentially requires two separate data acquisitions, which increases costs compared to the collection of ALS data alone. Thus, a single sensor solution would significantly reduce the expenses related to data acquisitions. The most promising single sensor solution has been a multispectral ALS system (e.g. Optech 2020), which typically carries out LiDAR measurements simultaneously, with separate sensors operating at different wavelengths (e.g. 1064, 1550, 532 nm). To date, multispectral ALS data have been examined for the recognition of tree species (e.g. Axelsson et al. 2018; Kukkonen et al. 2019b) and for the prediction of forest attributes (Dalponte et al. 2018; Kukkonen et al. 2019a). In general, the results suggest that the predictive performance associated with multispectral ALS data outperforms that of unispectral ALS data when tree species are considered. However, multispectral ALS data do not outperform the combination of unispectral ALS data and aerial images in tree species-specific predictions.

Satellite imagery and bitemporal ALS datasets can also provide the predictive power required in species-specific forest inventories. Kukkonen et al. (2018) observed that Sentinel-2 satellite images provide almost similar predictive performance in ALS-based species-specific forest inventories to multispectral aerial images. Repeated ALS measurements have been successfully utilized, for example, in the monitoring of forest growth (e.g. Zhao et al. 2018). Bitemporal ALS have not been applied in species-specific forest inventories prior to this thesis. The conditions under which ALS datasets are collected may also influence the predictive performance in species-specific inventories. Villikka et al. (2012) showed that ALS data collected under leaf-off conditions are better able to distinguish between coniferous and deciduous tree species than leaf-on ALS data.

1.5 Diameter distributions in ALS-based forest inventories

1.5.1 Area-based approach

The modeling of diameter distributions in ALS-based forest inventories has relied on the techniques inherited from the era of inventories by forest stands. The PDF parameters can be predicted using similar principles in the ALS-based forest inventories to those employed in the inventories by forest stands. The measured forest attributes of the parameter models are replaced with forest attributes predicted by the ALS-based forest inventory. The parameter prediction models are not inventory-specific, but the purpose has been to fit general models that could be applied to the whole country (Kilkki and Päivinen 1989; Siipilehto 1999). Therefore, the locality of the inventory area cannot be adequately addressed. Forest attributes predicted using ALS data can also be used in the recovery of parameters associated with a PDF (Mehtätalo et al. 2007). However, prediction errors associated with the forest attributes may induce convergence problems (Mehtätalo et al. 2007; Siipilehto and Mehtätalo 2013; Maltamo et al. 2018). It is evident that convergence problems are more common when tree species are considered in the recovery. Moreover, the moments or percentiles needed in the parameter recovery can be predicted using ALS data (Gobakken and Næsset 2004; Cosenza et al. 2019). A system of percentiles of a diameter distribution can also be predicted using ALS data, which means that a PDF is not needed at all (Gobakken and Næsset 2005).

The PDF parameters can also be predicted directly using ALS data (Gobakken and Næsset 2004). The prediction of parameters is typically difficult, since ALS features are strongly related to the vertical structure of the forest but not explicitly to the diameters of trees. The prediction of parameters can be accommodated by fitting models separately for the stratified data, e.g. by development class (Gobakken and Næsset 2004) or thresholds in forest attributes (Thomas et al. 2008). However, consideration of tree species or canopy layer increases the number of parameters that must be predicted, and the stratification should be known for the targets as well. The simultaneous prediction of numerous parameters is possible, although the parameters of the density functions do not strongly correlate with the ALS features. For example, Thomas et al. (2008) successfully predicted bimodal diameter distributions applying a mixture of Weibull models and multiple linear regression.

The prediction of diameter distributions has also been investigated from the perspective of advanced statistical inference (Magnussen et al. 2013) and sampling theory (Magnussen and Renaud 2016). Magnussen et al. (2013) predicted diameter distributions using the Gram-Charlier A-series expansion (GCAE) of a PDF and the cumulants of ALS-based canopy height distributions. Their approach covered a wider range of shapes associated with diameter distributions beyond PDF alone. The computationally complex GCAE provided minor improvements to the errors associated with diameter distributions compared to the simpler approach based on the prediction of distribution deciles. Magnussen and Renaud (2016) applied multidimensional scaling and first-return ALS canopy heights for the model-assisted estimation of diameter distribution. They proved that multidimensional scaling can be used to link the relative frequency distribution of canopy heights to an observed diameter distribution. The approach has advantages because it can be applied at different scales regardless of, for example, plot sizes.

The fieldwork associated with ALS-based forest inventories substantially differs from the principles applied in the inventories by forest stands, since ALS-based inventories require comprehensive field measurements (tree-level, empirical tree lists) from sample plots. Therefore, the application of the modeling techniques developed for the inventories by forest stands are suboptimal in ALS-based forest inventories when the trees are aggregated to the stand-level. The prediction of diameter distribution using measured trees, i.e. empirical tree lists and the NN approach, has superseded methods that rely on PDF in Finnish ALS-based forest inventories. The option to apply tree lists will be provided in the Finnish operational forest inventories in the 2020s.

Non-parametric approaches for the prediction of diameter distributions usually refer to the NN approach (Packalén and Maltamo 2008). The NN approach is the most often used approach and has proved to be flexible compared to the approaches that are based on PDF. Here, the flexibility implies that the shape of the predicted diameter distribution is not dependent on the characteristics associated with a specific PDF. Packalén and Maltamo (2008) showed that the prediction of diameter distributions using NN imputation outperforms the approach based on the Weibull distribution when tree species were considered in the prediction. Peuhkurinen et al. (2008) and Maltamo et al. (2009a) also successfully predicted diameter distributions using the NN approach. The most significant advantage of the NN approach in ALS-based forest inventories is that the predictions are constructed using a local field dataset (per inventory area). The localization of the diameter distribution predictions can better consider the regional variation in forest attributes than, for example, diameter distributions predicted using national-level parameter models.

1.5.2 Individual tree detection

The critical issues related to the prediction of diameter distributions using the ITD approach are: (1) the detection of suppressed trees, and (2) the prediction of DBH associated with detected trees. After that, the construction of diameter distributions from predicted DBH is a straightforward process. The prediction of diameter distributions using ITD has been studied (Lindberg et al. 2010) but operational applications that apply ITD have been rare.

A common ITD approach is to use a canopy height model (CHM) that is interpolated using ALS echoes returned from the forest canopy. Subsequently, the CHM is smoothed, and individual treetops are detected from the CHM using a local maxima algorithm (e.g. Persson et al. 2002). Finally, the tree crowns are delineated using image processing techniques, such as a watershed algorithm. The delineation of individual trees can also be implemented from ALS data without the image processing techniques and CHM (Reitberger et al. 2009; Duncanson et al. 2014; Kansanen et al. 2019a). Despite the segmentation algorithms, detection of suppressed trees is a major challenge as laser pulses emitted from the ALS device cannot properly penetrate dense canopy layers (Persson et al. 2002). Moreover, a group of trees can be difficult to detect as distinct trees (Packalen et al. 2013). The errors related to the detection of suppressed trees may not have a critical effect on the errors associated with predicted volumes but are surely critical shortcomings from the point of view of diameter distributions (Persson et al. 2002; Vauhkonen 2020).

Error rates associated with the prediction of DBH have a significant effect on the goodness of diameter distribution predicted using ITD. The DBH of logwood-sized trees are more difficult to predict than pulpwood-sized trees, since the increase in tree height diminishes as a tree matures. Moreover, DBH values are affected by numerous factors, such as site fertility and silvicultural activity. In many cases, those factors are difficult to incorporate into the models due to a lack of data. The ITD approach also requires that tree species are reliably predicted because the DBH-height relationships depend on tree species. Mixed-effects models are commonly used in the prediction of DBH (Kalliovirta and Tokola 2005), and NN imputation (Maltamo et al. 2009b) and copulas (Xu et al. 2019) have also been successfully applied. To overcome the challenges related to the prediction of DBH, the features associated with the characteristics of crown segments can be used in conjunction with traditional ALS-based height and intensity features (Vauhkonen et al. 2010). However, the error rates associated with predicted DBH in the previous studies have been relatively large (at lowest, errors of 2–3 cm).

1.5.3 Fusion of area-based approach and individual-tree detection

Several studies have investigated the fusion of ABA and ITD for the prediction of diameter distributions. Typically, the principle has been that the trees in the dominant canopy layer are predicted using ITD and the trees below the dominant canopy layer are predicted using the principles of ABA. Maltamo et al. (2004) predicted diameter distributions using only ITD information, although there is a risk of large errors associated with the stem numbers in the smallest diameter classes. The shortest detected tree by ITD was used as a cut-off point for the height distribution of detected trees. The cut-off point divided the height distribution into the Weibull-based predicted diameter distribution part (left tail) and the ITD part (right tail). Finally, the height distribution was transformed into a diameter distribution using the DBH models. They showed that the fusion of ITD and the Weibull-based prediction provided smaller error rates associated with timber volume and stem number than using ITD alone. As

observed in Maltamo et al. (2004), ITD underestimates the stem number and, therefore, ABA-based forest attributes, such as stem number, are used to calibrate ITD-based diameter distributions (Lindberg et al. 2010; Ene et al. 2012). Lindberg et al. (2010) presented a calibration framework that reduced the large bias associated with stem numbers predicted using ITD. They calibrated the ITD-based diameter distribution to be consistent with the ABA-based diameter.

Several studies have indicated that ABA outperforms ITD in the prediction of understory trees (Xu et al. 2014; Hou et al. 2016; Shin and Temesgen 2018). Xu et al. (2014) followed the idea of the cut-off point presented in Maltamo et al. (2004) and applied both a replacement approach and histogram matching to fuse the diameter distributions predicted using ABA and ITD. Here, the replacement approach generally means that the right tail of the ABA distribution is replaced with the ITD distribution. Later, Hou et al. (2016) also adapted the idea of a cut-off point to the prediction of species-specific diameter distributions. In general, the studies showed that the fusion of ABA and ITD decreased the error rates associated with the predicted diameter distributions. Hou et al. (2016) suggested that the logwood-sized trees can be predicted with smaller error rates using ITD than ABA. Similar findings concerning the errors associated with predicted logwood volumes were also reported by Peuhkurinen et al. (2011). Shin and Temesgen (2018) also fused the diameter distributions predicted using ABA, ITD, and a cut-off point. They focused on assessing the effect of parameters, such as different height-related cut-off points, CHM resolutions, and the magnitude of CHM smoothing, on the predictive performance of the ABA-ITD fusion. Their main finding was that the parameters are dependent on the forest characteristics, and that fusion parameters, therefore, should be selected separately for each target forest.

1.6 Evaluating the goodness of predicted diameter distributions

The evaluation of predicted diameter distribution against observed diameter distribution is not as straightforward as in the case of forest attributes, such as volume or basal area. The performance of the diameter distribution prediction method is typically carried out using forest attributes derived from the diameter distribution (Packalén and Maltamo 2008; Peuhkurinen et al. 2008), error indices (Reynolds et al. 1988; Gobakken and Næsset 2004; Packalén and Maltamo 2008), or statistical tests (Poudel and Cao 2013; Strunk et al. 2017). However, there is no consensus as to which one is the best measure. The selection of measure depends on the purpose for which the predicted diameter distribution will be used. The error rates associated with timber assortment volumes, namely logwood and pulpwood volumes, are valid measures if the economic value of a forest is deemed to be of interest. However, the shape of diameter distributions may not be measured sufficiently with the error rates associated with predicted timber assortment volumes. Error indices or statistical tests may be more robust to assess the goodness of predicted diameter distribution when the shape of distribution is of interest. An issue related to the simultaneous use of several measures is that they can be discordant.

The error indices and statistical tests are easy to apply and interpret, since they are typically based on stem numbers associated with fixed diameter classes. Reynolds et al. (1988) proposed the error index for the evaluation of the goodness-of-fit of diameter distribution models. They applied the error index that sums all the errors associated with the stem frequencies of all diameter classes. However, they also suggested that the errors in stem numbers could be computed as a proportion of the total stem number when the prediction

errors associated with stem number are ignored. The proportional formulation of the error index has been adapted and modified in the context of ALS-based forest inventories (Gobakken and Næsset 2004; Packalén and Maltamo 2008). The proportional form of the error index is at a specific range, which facilitates the interpretation of index values.

1.7 Objectives

The general aim of this thesis is to enhance the NN approaches used in the prediction of diameter distributions. The NN distributions can be easily adapted to ALS-based forest management inventories in Finland, since the required field data have already been measured. Topics concerning the configurations of the NN imputations, remotely sensed data, and the fusion of ABA and ITD approaches are covered in this thesis. The specific objectives of this thesis are as follows:

(1) To examine how the response configurations of NN imputations affect the predictive performance associated with species-specific diameter distributions (**I**).
(2) To calibrate the predicted diameter distributions with predicted total volume and to evaluate the effect of calibration on the errors associated with diameter distributions (**I**).
(3) To evaluate the predictive performance of NN imputation when different combinations of remote sensing data are used in the prediction of diameter distributions of logwood-sized trees. The most interesting data combinations to evaluate are multispectral ALS, and bitemporal ALS datasets (**II**).
(4) To examine the fusion of diameter distributions predicted using ABA and ITD. The specific aim is to examine forest characteristics as criteria for the selection of ABA, ITD, or fusion approaches in the prediction of diameter distributions (**III**).

2 MATERIALS AND METHODS

2.1 Study site and field data

The study site used in this thesis covers about 43,000 hectares and is located in eastern Finland (Figure 1). The study site represents a typical Finnish managed forest dominated by coniferous tree species. The main tree species are Norway spruce, Scots pine, Silver birch, and Downy birch. All deciduous species are considered as a single tree species group. This study only considers young, middle-aged, and mature forests, whereas seedling or sapling forests were not included in the analyses.

Two plot samples were established in the study site. The first sample consisted of 424 circular plots. The second sample consisted of square-shaped plots, which are larger than the circular plots and represent forest stands. The circular sample plots were used in papers **I** and **II**. The square-shaped sample plots (hereafter 30×30 m plots) were used for validation purposes in paper **II** (n = 105). In paper **III**, the analyses were implemented using only coniferous dominated 30×30 m plots (n = 92). The coniferous dominated plots were selected because the restriction provided simplified circumstances to evaluate the performance of the proposed approaches.

The circular plots were measured between June and September 2016. The radius of circular plots was either 9 m (71 % of plots) or 12.62 m (29 % of plots). The radius of 12.62 m was selected if the stem number inside the plot was less than 20. Most of the circular plots were distributed over the study area using a systematic cluster sampling design. The distance between adjacent clusters was 1200 m. The centers of the circular sample plots were accurately positioned by means of a GNSS receiver. The coordinates of the plot centers were post-corrected using reference stations. The study site is a part of the operational forest inventory operated by the Finnish Forest Centre, and so 27 % of the circular plots were distributed in the inventory area using different sampling designs. Although the sampling design employed by the Finnish Forest Centre is different to the systematic sampling design employed here, the actual field measurements were quite similar in all the circular plots (Suomen metsäkeskus 2016).

Field measurements in the 30×30 m plots were carried out between June and October 2017. The plots were sampled from the systematic sampling design by means of *a priori* knowledge concerning the development classes and the proportions of the dominant tree species observed in the set of circular sample plots. The development classes were estimated using an ALS dataset collected in summer 2016, and the dominant tree species were fetched from the Multi-source National Forest Inventory data (Natural Resources Institute Finland 2013; see Tomppo and Halme 2004). In the field, the XY coordinates were determined for each measured tree using CHM (resolution 0.5 m) and the triangulation approach proposed by Korpela et al. (2007). The XY locations of the trees were used when the 30×30 m plots were divided into 15×15 m subplots (hereafter subplots; n = 420). The analyses were implemented at the subplot-level, but the results were eventually aggregated to the 30×30 m level (forest stand).

In each plot, DBH and tree species were measured for each tree with a DBH ≥ 5 cm. Tree height was also measured for each tree, except in the plots measured by the Finnish Forest Centre, where the tree heights were only measured for a subset of trees and those trees were used in the calibration of the random parts of the height model of Eerikäinen (2009). Stem volumes were computed for each tree as a function of DBH and tree height using the models presented in Laasasenaho (1982). Logwood and pulpwood volumes were computed using taper curves (Laasasenaho 1982) with bucking parameters that were shown in paper **I**. The logwood and pulpwood volumes are theoretical, i.e. quality reductions were not considered. Volumes were computed by tree species, although the birch model was used for all deciduous species. Dead trees were excluded since they usually do not have a prominent role in the dominant tree layer in Finnish managed forests. Only plots located entirely within a forest stand were used in this study. Finally, the attributes of individual trees were aggregated to the plot-level and multiplied out to the hectare-level. Means and standard deviations of forest attributes associated with the field data are presented in Table 1.

In paper **III**, the shapes of the diameter distributions were visually determined for each 30×30 m plot using the empirical diameter distributions (bin width: 4 cm). The empirical diameter distributions were assigned to the following shape classes (hereafter DDSC): Gaussian, reverse J, and bimodal. For representative examples of DDSC, please refer to Figure 2 in paper **III**.
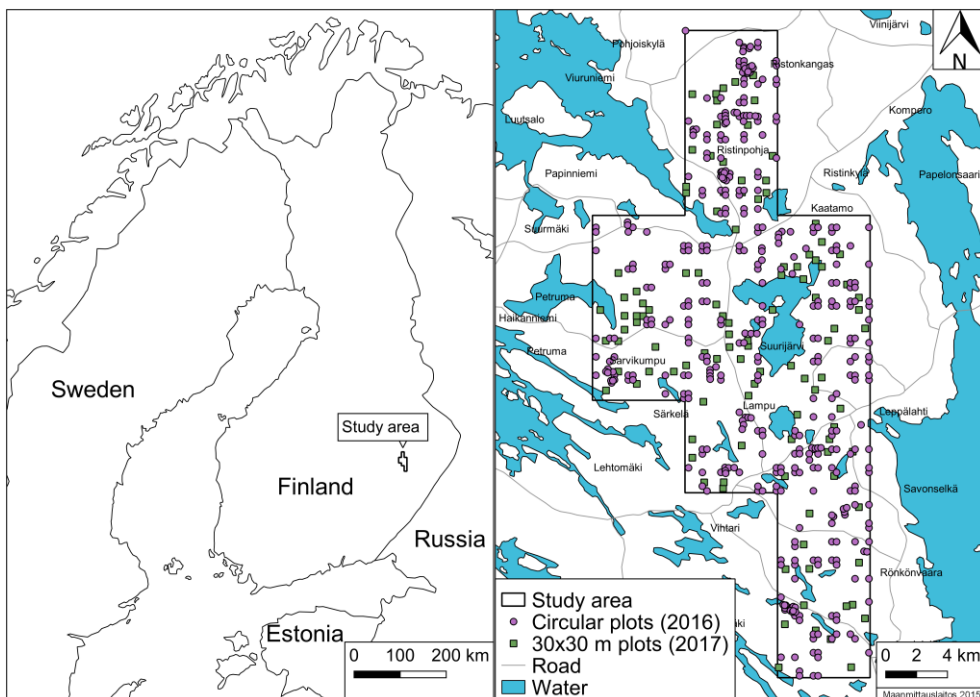
Figure 1. Location of the study area and sample plots in Finland.

Table 1. Means and standard deviations (sd) of the main forest attributes in the field datasets. DGM = the diameter of basal area median tree, HGM = the height of basal area median tree.

		Circular plots		15 x 15 m plots		30 x 30 m plots	
Attribute	Population	Mean	Sd	Mean	Sd	Mean	Sd
Volume (m³ · ha⁻¹)	Scots pine	76.0	83.4	77.1	90.7	77.1	85.1
	Norway spruce	87.7	109.1	87.5	109.9	87.5	103.5
	Deciduous	22.8	36.1	41.0	62.2	41.0	58.4
	Total	186.4	100.4	205.6	94.6	205.6	86.1
Basal area (m² · ha⁻¹)	Scots pine	8.79	8.8	8.9	9.7	8.6	9.2
	Norway spruce	9.9	10.5	9.8	10.3	9.8	9.7
	Deciduous	2.9	4.2	4.4	5.9	4.4	5.6
	Total	21.6	8.2	22.8	7.7	22.8	6.8
DGM (cm)	Scots pine	21.0	6.5	22.2	7.7	22.6	6.9
	Norway spruce	17.1	8.4	17.8	9.8	17.7	9.2
	Deciduous	14.8	7.5	16.8	8.5	17.4	8.6
HGM (m)	Scots pine	17.6	4.8	18.6	4.4	18.9	5.0
	Norway spruce	14.8	6.5	15.0	7.1	15.1	7.0
	Deciduous	15.3	5.5	17.0	6.5	17.4	6.4

2.2 Remotely sensed material

Four different remote sensing datasets were used in this thesis: multispectral leaf-on ALS dataset (M–ALS data), two unispectral leaf-off ALS dataset (S11–ALS and S16–ALS data), and aerial images. In paper **I**, the S16–ALS dataset was used in conjunction with the aerial images. In paper **II**, all abovementioned remote sensing datasets were used. In paper **III**, M–ALS dataset was only used.

The M–ALS dataset was collected under leaf-on conditions in June 2016 at an altitude of 850 m above ground level using a fixed-wing airplane. The airplane was equipped with a Teledyne Optech Titan laser scanning system that can capture up to four range and intensity measurements per emitted pulse. The Titan laser scanning system has three different sensors that operate at the wavelengths of 1550 nm (first channel), 1064 nm (second channel, M–CH2–ALS data), and 532 nm (third channel). The scanning half-angle was fixed at 20 degrees. The parameters used in the data acquisition resulted in an average density of 4.8 pulses per square meter for the first and second channel, and 3.7 pulses per square meter for the third channel for each flight line. Since the lateral overlap was 55 %, the observed pulse densities are at least twice as dense as the aforementioned pulse densities.

The unispectral S16–ALS dataset was acquired under leaf-off conditions between 30 April 2016 and 3 May 2016 at an altitude of 2400 m above ground level. The dataset was collected using a fixed-wing airplane that was equipped with a Leica ALS60 laser scanner system. The Leica ALS60 can capture up to four range and intensity measurements per emitted pulse. The scanning half angle was set at 20 degrees and the lateral overlap was 20 %. The parameters of the data acquisition resulted in a nominal pulse density of 0.8 pulses per square meter.

The unispectral S11–ALS dataset was collected under leaf-off conditions between 25 April 2011 and 26 April 2011 at an altitude of 2200 m above ground level. The dataset was collected from an airplane equipped with the Leica ALS60 laser scanning system. The parameters of the data acquisition were broadly similar to those that were used in the acquisition of S16–ALS. The parameters of the data acquisition resulted in a nominal pulse density of 0.9 pulses per square meter. Since the S11–ALS dataset was collected approximately five years before the field measurements, it was essential to detect the implemented silvicultural operations between the data acquisitions. These silvicultural operations were detected by comparing the S16-ALS and S11-ALS datasets. The detection approach was explained in paper **II**.

Returned ALS echoes were assigned to three echo categories according to the return order: first, last, and intermediate. If only one echo was captured per emitted pulse, the echo was assigned to both the first and last category. The ground echoes were identified by the method proposed by Axelsson (2000). The ground echoes were used to interpolate a digital terrain model (DTM) using a Delaunay triangulation. The normalized (i.e. aboveground) heights were computed by subtracting DTM from the initial orthometric echo heights. Intensity values associated with the M–ALS and S–16–ALS datasets were normalized for the range as presented in Korpela et al. (2010).

The aerial images were captured with a DMC Z/I Intergraph (01-0128) digital camera on 23 and 24 May 2016. The aerial images were captured from an airplane flying at an altitude of 4100 m above ground level using a nominal lateral and longitudinal overlap of 30 and 80 %, respectively. The camera has a focal length of 30 mm and is capable of recording four spectral bands: red, green, blue, and near-infrared. The camera has 3456 × 1920 pixels in

multispectral bands, which resulted in a ground sampling distance (GSD) of about 160 cm. The aerial images were not orthorectified nor pansharpened.

2.3 Feature extraction

Several statistical features were extracted from the height and intensity measurements of the ALS data. The set of features included: mean, standard deviation, maximum, minimum, kurtosis, skewness, the percentiles of echo height measurements and intensities, and densities with fixed height values. Echo proportions by echo categories were also computed. In papers **II** and **III**, the ratio features were computed between the height and intensity features extracted from the different channels of the multispectral ALS data. The features were computed by echo categories, and a height cut-off was set at 1.3 m to avoid the effect of ground echoes. The densities were always computed without the cut-off. For the ABA inventories, the features were computed at the plot-level, whereas for ITD the features were computed at the level of the segmented tree crowns.

The features extracted from the aerial images comprised mean, standard deviation, minimum, and maximum. The features were computed by the spectral bands. The spectral values were fetched from the pixels of aerial images for the first echoes of ALS data by projecting them over the aerial images by means of external and internal orientation. Due to the overlapping aerial images, several pixel values were assigned to an individual ALS echo. To accommodate this, the mean of the overlapping pixel values was used.

2.4 Nearest neighbor approach

In the literature, predictions using the NN approach are usually referred to as NN imputations (Vauhkonen et al. 2010; Packalen et al. 2012). Henceforth, the NN approach is also called NN imputation in this thesis. The NN imputation with ABA was applied to predict diameter distributions (**I**, **II**, and **III**) and to predict the DBH of trees detected by ITD (**III**).

The NN imputation searches for the most similar references from training data for a target. The similarity between the target and reference observations is measured using a distance or similarity metric. Several distance metrics have been studied in the context of forest inventories (McRoberts et al. 2017), but the most similar neighbor (MSN) distance (Moeur and Stage 1995) has performed well in species-specific forest inventories (Packalén and Maltamo 2007). The MSN distance applies canonical correlation analysis, which means that the similarity measurements between target and reference plots are not only based on the predictor variables, but that the response configuration has an effect too. On the one hand, this characteristic is advantageous when predictions are carried out with a multivariate response configuration, but on the other, a user must consider the selection of a response configuration. The MSN distance is computed as follows:

$$d^2{}_{uj} = \frac{(\boldsymbol{x}_u - \boldsymbol{x}_j)\ \Gamma\Lambda^2\Gamma^T\ (\boldsymbol{x}_u - \boldsymbol{x}_j)^T}{(1 \times p)\quad (p \times p)\quad (p \times 1)}$$
$$(1 \times 1)$$

(1)

where $d^2{}_{uj}$ is the squared distance, $\boldsymbol{x}_u$ and $\boldsymbol{x}_j$ are row vectors comprising predictors (p refers to the number of predictors) from a target (u) and reference (j) plot, Γ is the matrix of canonical coefficients of predictor variables, and Λ is the diagonal matrix of squared canonical correlations.

In addition to the selection of a distance metric, the user of the NN imputation has to decide the number of neighbors (please refer to section 2.5), the selection routine for predictor variables (please refer to section 2.9), and the weighting scheme for the NN. In this thesis, the squared MSN distances were inversed in order to use them as weightings for the NNs.

2.5 Prediction of diameter distributions using the area-based approach

The prediction of diameter distribution using the NN imputation is technically similar to the prediction of univariate forest attributes. Instead of a set of forest attributes, empirical tree lists are retrieved from reference plots and assigned to a target. The interesting viewpoint in the diameter distributions predicted by NN imputation lies in the simultaneous prediction of diameter distributions and forest attributes. Due to the multivariate response configurations in NN imputations, the species-specific prediction is straightforward to implement, since separate models by tree species are not necessarily needed. Especially in species-specific imputations, it should be noted that the NN imputation is incapable of extrapolation outside the training data, and the training data, therefore, must represent all possible combinations of tree species compositions and tree sizes.

Here diameter distributions were predicted using both simultaneous NN imputation and separate NN imputation by tree species. The simultaneous NN imputation uses one model to predict species-specific diameter distributions, whereas separate NN imputation by tree species uses separate models for each tree species (group) of interest. Different response variable configurations were established for both imputation methods. The motivation for the examination of the predictive performance associated with the different response configurations originates from the MSN distance and multivariate NN imputation. The multivariate response configuration optimized for the prediction of specific forest attributes may not be an optimal alternative for the prediction of diameter distributions. Table 2 shows all the response configurations that were tested in the NN imputation of diameter distributions. The response configurations typically included forest attributes, but attributes related to the empirical diameter distributions, such as Weibull parameters, were also tested. The parameters of the Weibull distributions (c = shape and b = scale) were estimated by maximizing a likelihood function in the empirical data. In addition to the response configurations, the NN imputation is also affected by a hyperparameter k. The k parameter determines the number of neighbors and was fixed at 5 in this thesis, excluding the imputation of DBH in paper **III** (k fixed at 1). The prediction using the NN imputation is also dependent on the number of predictor variables. Depending on the preliminary runs, the number of predictor variables was fixed at 10 (**II** and **III**) or 11 (**I**).

Table 2. Response configuration used in the nearest neighbor (NN) imputation in paper **I**. SET_2 was also used in the species-specific prediction of logwood volumes presented in paper **II**. V = volume; G = basal area; N = stem number; DGM = the diameter of basal area median tree; HGM = the height of basal area median tree; Vlog = logwood volume; Vpulp = pulpwood volume; D10, D20, …, D90 = percentage points (10, 20 % etc.) computed from empirical diameter distribution; F6, F8, …, F34 = empirical stem frequencies (bin width: 2 cm) by diameter classes; b = the scale parameter of Weibull distribution; c = the shape parameter of Weibull distribution.

	Attributes of response configuration	Abbreviation of response configuration
Simultaneous NN imputation	V, G, N, DGM, HGM	SET_1
	G, N, DGM	SET_2
	G, N, DGM, HGM	SET_3
	V, N, DGM, HGM	SET_4
	G, N, DGM, D30, D80	SET_5
	G, N, DGM, c, b	SET_6
	V, Vlog, Vpulp	SET_7
	Vlog, N/G	SET_8
Separate NN imputation by tree species	V, G, N, DGM, HGM	SET_{sep1}
	G, N, DGM	SET_{sep2}
	D10, D20, ..., D90, N	SET_{sep3}
	D10, D20, ..., D90	SET_{sep4}
	V, Vlog, Vpulp	SET_{sep5}
	F6, F8, …, F34	SET_{sep6}

2.6 Calibration of diameter distribution with total volume

The response configuration does not usually include total volume when species-specific forest attributes are of interest, which negatively affects the errors associated with total attributes. Therefore, it makes sense to calibrate the diameter distributions predicted using a multivariate NN imputation with separately predicted total volume (**I**). The calibration with total volume resembles the calibration estimation approach (Kangas and Maltamo 2000b), but here only one attribute is used in the calibration procedure.

A linear regression model with three predictor variables was fitted for total volume. The predicted total volume, and the sum of predicted species-specific volumes (from NN imputation), were used to compute a calibration factor for each plot. The frequencies of each tree in a predicted diameter distribution of a plot were multiplied by a calibration factor. The formulation of the calibration factor can be seen in Equation 4 in paper **I**.

2.7 Prediction of diameter distributions using individual tree detection

The ITD approach applied in paper **III**, follows the approaches presented by Persson et al. (2002), Pitkänen et al. (2004) and Packalén et al. (2008). First, CHM was interpolated using the maximum echo heights within the cells (resolution 0.5 m). CHM was then smoothed to reduce the information. The smoothing was based on the Gaussian filter, and the level of smoothing depended on the ALS-measured height of a forest. Individual treetops were detected from CHM using a local maxima algorithm. Finally, the detected crowns were segmented using a watershed algorithm (Gauch 1999) with a drainage direction (Narenda and Goldberg 1980).

The diameters of detected trees were predicted by applying the NN imputation. The detected trees were linked to the observed trees only when the linkage could be done reliably. The linkage was based on the Euclidean distance and the height difference between the observed and detected tree. Those linked trees were used as training data, whereas all the detected trees were the targets of prediction in the NN imputation. A leave-30×30 m plot-out cross validation was applied to predict the DBH of the detected trees.

The ITD was only applied in coniferous-dominated forests, and diameter distributions were not predicted by tree species. However, a mixture of deciduous and coniferous species exists on several plots. Therefore, it is necessary to consider the species-specific height-diameter relationships in the prediction of diameters. To overcome issues related to the differences between tree species, the MSN distance and multivariate response configuration were applied in the NN model. The multivariate response configuration was comprised of species-specific responses (D_{pine}, D_{spruce}, and $D_{deciduous}$). For each tree, the multivariate response included one species-specific DBH value, and the other two species-specific DBH values were set to zero. Therefore, one DBH value was assigned to each tree because the number of nearest neighbors (k value) was fixed at 1. Finally, the diameter distributions were constructed using 4 cm diameter classes for each 30×30 m plot.

2.8 Fusion of diameter distributions predicted by ABA and ITD

Frameworks for the fusion of ABA and ITD diameter distributions were established in paper **III**. Initially, the fusion of diameter distributions predicted by ABA and ITD was evaluated by finding an optimal fusion parameter (F.BEST) utilizing empirical diameter distributions. The theoretical framework was adapted to practice by presenting fusion based on predicted weightings (F.PRED). In addition, a fusion based on replacement (F.REPL) was proposed that applies the pre-determination of forest structures in the selection of ABA or ITD.

F.BEST and F.PRED apply a linear weighting function for ITD and ABA diameter distributions by diameter classes for each plot. The weightings associated with the diameter classes can be adjusted using the slope parameter of the weighting function. The weighting function is fitted separately for each plot. The slope parameters of F.BEST were optimized for each plot by minimizing the error associated with the fused diameter distribution. The stem frequencies of fused diameter distribution were computed as a weighted average of ABA and ITD stem frequencies in each diameter class (bin size: 4 cm). The slope parameter could have negative and positive values at a fixed range. The intercept of the weighting function was fixed on the y-axis at the point where the ABA and ITD diameter distributions are equally weighted. Therefore, the horizontal line (slope 0) gives equal weighting to ABA and ITD for each diameter class of diameter distribution. Instead, the minimum and

maximum slope values imply 100 % ABA and 100 % ITD diameter distributions, respectively. The slopes were converted to angles (β) at the range of [0°, 90°] where 45° implies the fusion with equal weightings for ABA and ITD. The principle of the weighting scheme is illustrated in Figure 4 in paper **III**. The F.PRED approach is based on a similar weighting scheme as the F.BEST approach. The prediction of β values was implemented using a regression model with two predictor variables. The β values that were optimized in the F.BEST approach were used as a training dataset. The predictor candidates comprised ALS features, as well as forest attributes derived from the ABA and ITD predictions.

The F.REPL approach originates from the hypothesis that the predictive performance of ITD varies according to the forest structure. Therefore, the F.REPL approach aims to select the optimal method (ABA or ITD) for each target of interest. Thus, the targets must be pre-classified according to DDSC. Multinomial logistic regression (MLR) modeling was used to predict DDSC for each plot using features derived from ALS data and forest attributes predicted by ABA and ITD. The classification was implemented applying a leave-30×30 m plot-out cross validation.

2.9 Selection of predictor variables

The predictor variables of the NN imputations were selected using an optimization-based algorithm. The same algorithm was also modified for an MLR model in paper **III**. The optimization-based variable selection was implemented using a simulated annealing (SA) algorithm (Kirkpatrick et al. 1983). Packalén et al. (2012) proposed the SA-based algorithm for the selection of predictor variables in NN imputation. The algorithm searches for the combination of a fixed number of predictor variables that minimizes the loss function. In the case of NN imputation, the loss function was computed as the mean of root mean square error (RMSE) values associated with the predicted response variables. In classification (MLR), an inversed Cohens´ kappa was used as a loss function. The optimization starts with a random guess and the iteration is stopped when a predetermined number of iterations has been met.

An iterative search algorithm was applied to the linear regression models fitted for total volume (**I**) and for the fusion parameter (**III**). The predictor variables for the regression models were selected using an algorithm that searches for the optimal combination of a fixed number of predictor variables with respect to RMSE using total enumeration.

2.10 Performance assessments

Prediction errors of logwood and pulpwood (i.e. timber assortments) volumes by tree species were used to evaluate the goodness of the diameter distributions. In paper **II**, logwood volumes were also computed separately for dominant and minor species. The predictions associated with timber assortment volumes were evaluated using relative RMSE (RMSE%, referred to as the error rate) and mean difference (MD%) in papers **I** and **II**:

$$RMSE\% = \frac{\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}}{\bar{y}} \times 100 \qquad (2)$$

$$MD\% = \frac{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n}}{\bar{y}} \times 100 \qquad (3)$$

where $\hat{y}$ and y refer to a predicted and an observed forest attribute, respectively, n is the number of field plots, and $\bar{y}$ is the mean value associated with the observed forest attribute.

Two different variants of error indices were also used to evaluate the goodness of diameter distributions. Both error indices used in this thesis originate from the error index described in Reynolds et al. (1988). In paper **I**, the error index (hereafter ErrorIndex1) was computed as the remainder between observed and predicted stem frequencies by diameter classes, and the errors in diameter classes are divided by the observed stem number. This error index is applied in the ALS context, for example, in Gobakken and Næsset (2004) and Maltamo et al. (2018). In paper **III**, the variant of the Reynold's error index (hereafter ErrorIndex2) proposed by Packalén and Maltamo (2008) was used. The values of ErrorIndex2 are within the range [0, 2] in which the minimum implies total agreement and the maximum total disagreement between distributions. ErrorIndex2 only measures the similarity in the shape of distributions, whereas ErrorIndex1 also considers the error associated with stem numbers. The error indices were computed as follows:

$$ErrorIndex1 = \sum_{i=1}^{m} 100 \times \left| \frac{f_i - \hat{f}_i}{N} \right| \qquad (4)$$

$$ErrorIndex2 = \sum_{i=1}^{m} \left| \frac{f_i}{N} - \frac{\hat{f}_i}{\widehat{N}} \right| \qquad (5)$$

where f_i and $\hat{f}_i$ are the stem frequencies associated with the observed and predicted diameter distributions in diameter class i, m is the number of diameter classes, and N and $\widehat{N}$ refer to the stem number of observed and predicted diameter distributions, respectively.

The classification using an MLR model was applied in the prediction of DDSC in paper **III**. The classification performance of the MLR modeling was evaluated using Cohen's kappa (κ) and overall accuracy (OA).

3 RESULTS

3.1 Response configurations in nearest neighbor imputation (I & II)

3.1.1 Simultaneous NN imputation

The results indicated that the response configuration used in NN imputation affects the error rates associated with predicted diameter distributions. In general, it can be stated that the configuration applied in Finnish operational forest inventories (SET_1) is a suboptimal alternative for the prediction of diameter distributions (Table 3). The results described in paper **I** showed that SET_2 generally outperformed the other response configurations that were used with the simultaneous NN imputation. Compared to the error rates associated with logwood volumes predicted using SET_1, SET_2 achieved 6.3, 10.5, and 15.8 % improvements in the error rates for pine, spruce, and deciduous species, respectively. This indicates that the forest attributes that are strongly associated with diameter distributions (e.g. basal area, mean diameter and stem number) are appropriate response variables, whereas attributes with a weaker association with diameter distribution may induce an increase in error measures. For example, SET_2 provided smaller error rates associated with the predicted diameter distributions than those that included the height attribute (SET_3 or SET_4). ErrorIndex1 values were generally in agreement with the error rates and indicated that SET_2 outperformed SET_1 in the prediction of diameter distributions by tree species (see Figure 4 in paper **I**). The MD% values associated with the coniferous species were minor but were greater for the deciduous species.

3.1.2 Simultaneous NN imputation versus separate NN imputation by tree species

The findings in paper **I** showed that the smallest error rates associated with predicted species-specific diameter distributions were achieved using SET_{sep1} (Table 3). The configuration of SET_{sep1} resulted in a mean RMSE value of 78.2 % (species-specific timber assortment volumes). The corresponding mean RMSE values of the simultaneous imputation were 95.0 % and 87.3 % with SET_1 and SET_2, respectively. The results indicated that SET_{sep1} provides smaller error rates associated with spruce and deciduous logwood volumes, and for all pulpwood volumes, compared to SET_2. However, the error rates associated with predicted total and pine logwood volumes were slightly smaller for SET_2 than SET_{sep1}. The mean ErrorIndex1 values related to the logwood-sized trees were, however, contradictory to the error rates (see Table 6 in paper **I**), and the results indicated that the shape of the deciduous diameter distribution may not be predicted as accurately using SET_{sep1} (compared to SET_2).

Table 3. Relative root mean squared error (RMSE%) and mean difference (MD%) (in parenthesis) associated with predicted species-specific logwood and pulpwood volumes. For abbreviations, please refer to Table 2.

	SET_1	SET_2	SET_{sep1}
$Vlog_{tot}$	36.4 (-1.3)	35.0 (-0.5)	35.5 (1.0)
$Vlog_{pine}$	67.1 (-1.5)	62.9 (-2.6)	66.8 (1.4)
$Vlog_{spruce}$	69.7 (0.8)	62.4 (2.8)	57.2 (1.4)
$Vlog_{decid}$	211.0 (-21.1)	177.6 (-15.8)	155.9 (-5.3)
$Vpulp_{tot}$	38.2 (2.2)	36.0 (1.3)	33.9 (-0.2)
$Vpulp_{pine}$	60.9 (2.1)	62.3 (0.5)	56.3 (-1.0)
$Vpulp_{spruce}$	64.9 (4.2)	62.4 (3.1)	54.4 (-0.2)
$Vpulp_{decid}$	96.7 (-1.2)	96.1 (-0.6)	78.3(1.1)

In paper **II**, species-specific diameter distributions were only considered for logwood-sized trees. The results showed that if only logwood-sized trees are of interest, it is reasonable to employ simultaneous NN imputation with species-specific logwood volumes as response variables. However, the results reported in paper **II** were in line with the results of paper **I**, where separate NN imputation by tree species provided smaller error rates associated with deciduous species compared to simultaneous NN imputation.

3.2 Calibration of diameter distributions with predicted total volume (I)

The prediction of total volume using the regression model provided a RMSE value of 17.2 %. The results showed that the calibration with total volume generally decreases the error rates and ErrorIndex1 values associated with predicted diameter distributions. The improvements were especially observed in errors associated with species-specific and total logwood volumes. The error rates associated with predicted pulpwood volumes increased in some cases. The effect of the total volume calibration (using SET_2 and SET_{sep1}) on RMSE values associated with predicted logwood and pulpwood volumes are presented in Figure 2.
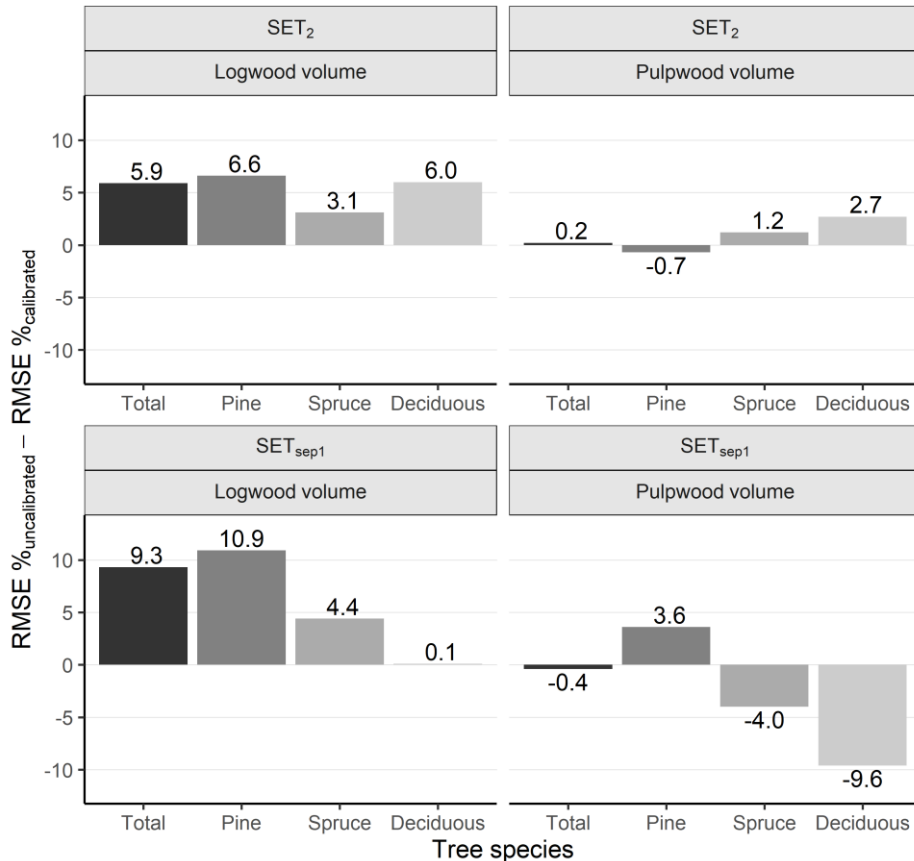
Figure 2. Effect of the calibration with total volume (**I**). The upper row refers to the results computed using the SET$_2$ response configuration, whereas the lower row refers to the results obtained with SET$_{sep1}$. For the abbreviations of response configurations, please refer to Table 2. Negative values indicate that the calibration increased the root mean squared error (RMSE%) value.

3.3 Remotely sensed data in the prediction of logwood volumes (II)

3.3.1 Multispectral ALS data

The results showed that the NN imputation using the M–ALS dataset achieved slightly smaller error rates associated with species-specific logwood volumes than the M–CH2–ALS dataset alone. In general, the M–ALS dataset provided smaller error rates associated with minor species than the M–CH2–ALS dataset. Instead, the S16–ALS dataset (acquired under leaf-off conditions) outperformed the M–ALS dataset in the species-specific predictions. The combination of the M–CH2–ALS or S16–ALS datasets and aerial images clearly outperformed the M–ALS dataset in the prediction of logwood volumes. This is evident in the error rates associated with species-specific, dominant, and total logwood volumes. The error rates associated with the dominant and total logwood volumes are presented in Figure 3. For detailed presentation of results please refer to paper **II**.
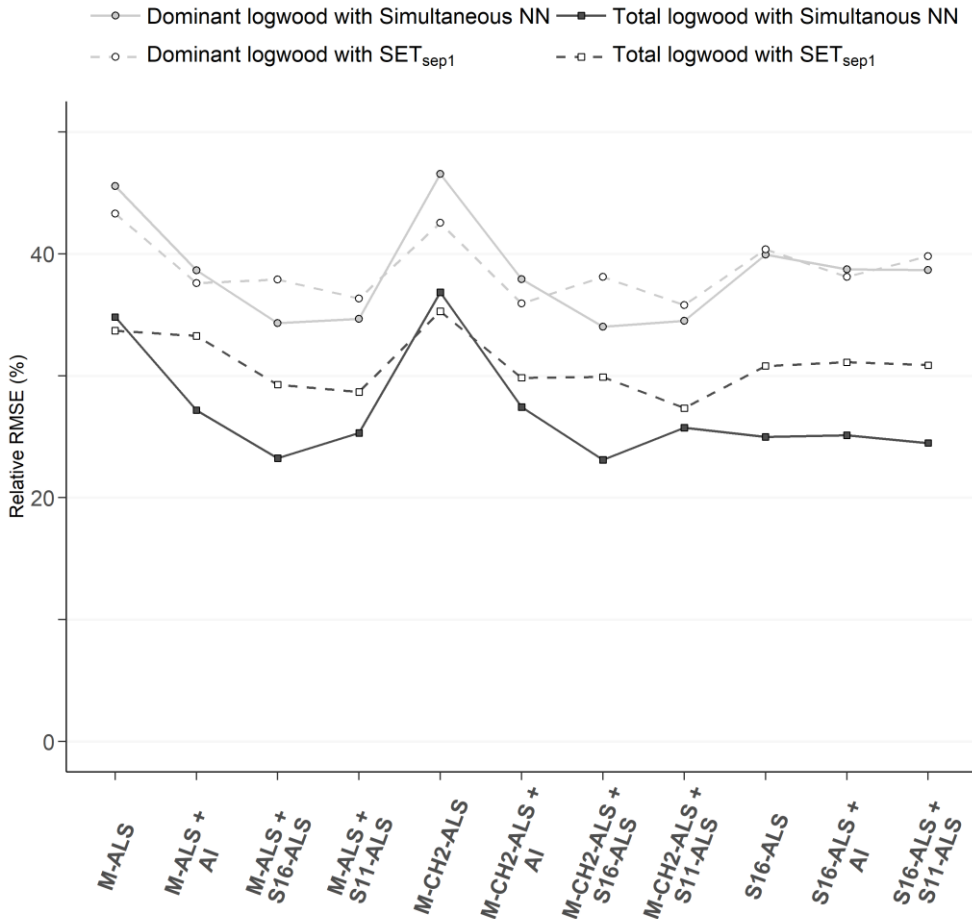
Figure 3. Relative root mean squared error (RMSE%) values associated with predicted dominant and total logwood volumes by remote sensing datasets (**II**). The dashed line shows the error rates associated with the separate nearest neighbor (NN) imputation by tree species, whereas the solid line refers to the simultaneous NN imputation in which species-specific logwood volumes are used as response variables. AI = aerial images. For the abbreviations of the airborne laser scanning datasets please refer to section 2.2.

3.3.2 Bitemporal ALS data

The simultaneous usage of M–CH2–ALS and S16–ALS datasets outperformed the combination of the M–CH2–ALS dataset and aerial images, in the prediction of logwood volumes by tree species. In general, the decreases in the error rates were most significant with deciduous logwood volumes but clearly visible with coniferous species as well (please refer to Figure 4 in paper **II**). The combined use of the M–CH2–ALS and S16–ALS datasets also provided the smallest error rates associated with (using simultaneous NN imputation) dominant and total logwood volumes (Figure 3). The combination of the recent leaf-on and

older leaf-off ALS datasets (M–CH2–ALS + S11–ALS) also outperformed the combination of the M–CH2–ALS dataset and aerial images, in the prediction of logwood volumes.

3.4 Fusion of diameter distributions predicted using ABA and ITD (III)

3.4.1 Evaluation of ABA, ITD, and F.BEST

The findings showed that ABA outperformed ITD in forests with reverse-J or bimodal diameter distributions. Instead, ITD resulted in smaller ErrorIndex2 values than ABA in forests with Gaussian-shaped diameter distributions. The results suggested that bimodal-shaped diameter distributions are insufficiently described using ITD because it may not detect small trees in the leftmost peak of diameter distribution. The mean ErrorIndex2 values of 0.30 and 0.46 were achieved using ABA and ITD, respectively.

The F.BEST fusion outperformed ABA and ITD resulting in a mean ErrorIndex2 value of 0.21. The performance benefits achieved using F.BEST were clearly dependent on DDSC, which is demonstrated using example plots in Figure 4. The findings suggested that the F.BEST approach would not provide improvements on the errors associated with predicted diameter distributions in forests with Gaussian-shaped diameter distribution but the fused distribution is exactly the ITD distribution (left-hand column in Figure 4). The findings indicated that the fusion may not provide improvements (compared to ABA-based diameter distributions) in reverse-J forests due to the poor performance of ITD in the small diameter classes (middle column in Figure 4). In many cases, the ITD predicted trees in the dominant canopy layer more reliably than ABA (right-hand column of Figure 4). In this example, the distribution associated with the dominant canopy layer is better characterized using ITD (or the fusion of ABA and ITD) than ABA. The leftmost peak of the distribution is typically underpredicted with the ITD approach, and the fusion with ABA can, therefore, be recommended in bimodal forests.
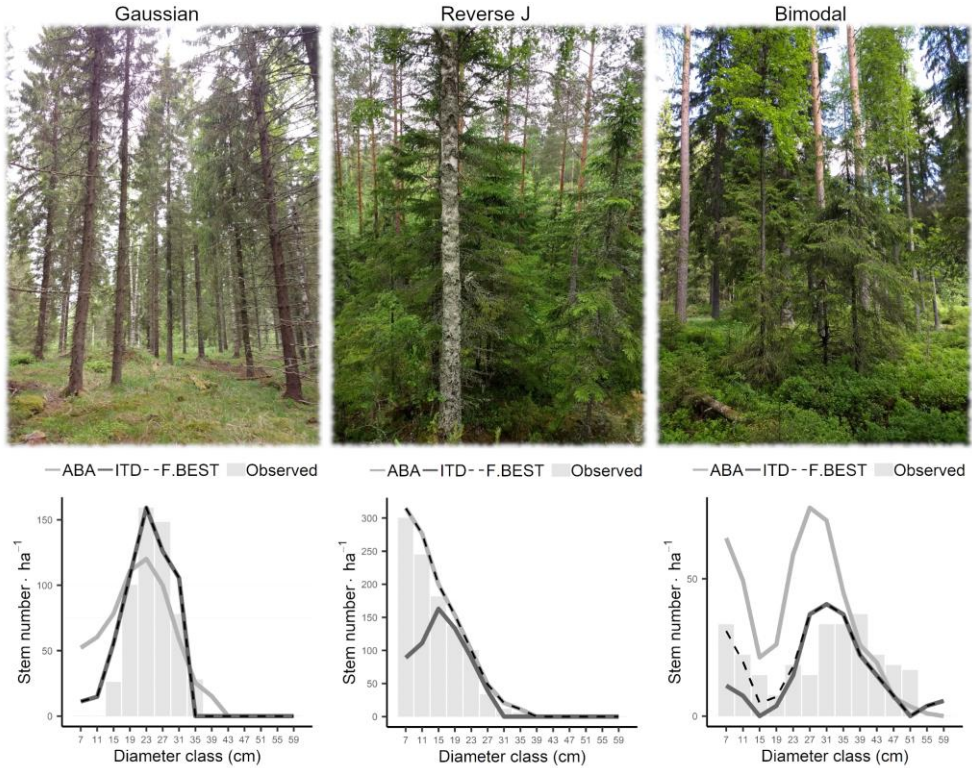
Figure 4. Examples of the performance of area-based approach (ABA), individual tree detection (ITD), and theoretical ABA-ITD fusion (F.BEST) in different forest structures.

3.4.2 Practical evaluation of the fusion approaches

The results indicated that the successful application of F.PRED is difficult in practice. In general, $\beta_{F.BEST}$ did not correlate well with the predictor variable candidates. For the model fitted using the training data, the coefficient of determination was 0.24 and the standard error was 30.2°. Although, the model fit was poor, the mean ErrorIndex2 values computed from the diameter distributions (predicted by F.PRED) was 0.28. The F.PRED approach achieved slightly smaller mean ErrorIndex2 values than ABA (cf. 0.28 and 0.30).

The F.REPL approach requires the pre-classification of the plots according to DDSC. The MLR model predicted the DDSC with kappa = 0.75 and OA = 83.70 %, when the leave-30 × 30 m plot-out cross validation was applied. The F.REPL approach achieved a mean ErrorIndex2 value of 0.29, which implies a minor improvement compared to ABA.

4 DISCUSSION

4.1 Major findings of the thesis

The results reported in paper **I** showed that the response configuration, which included species-specific forest attributes (SET$_1$), is a suboptimal choice for the prediction of species-specific diameter distributions. The findings also indicate that the separate NN imputation by tree species achieves smaller error rates than simultaneous NN imputation in general. The findings show that optimal configurations comprised basal area (G), stem number (N) and diameter of basal area median tree (DGM), whereas volume (V) was not included, which means that the errors associated with volume attributes are not minimized in the feature selection. This is somewhat problematic from the point of view of operational inventories, since total and species-specific volumes are important attributes in forest planning. In general, the smallest error rates associated with, for example, species-specific volumes could be achieved by setting them as response variables. This effect results from the canonical correlation analysis applied in the computation of the MSN distances and the loss function of the feature selection. The calibrations can be used to decrease the error rates associated with total attributes (such as total volume), when they are not response variables.

The calibration with total volume decreased the error rates related to the volumes by tree species. However, the effect of the calibration on the error rates associated with the predicted timber assortments is not straightforward, since the relationship with total volume is ambiguous. For example, the calibration used here assumes the same calibration factor for each tree in a plot. Therefore, pulpwood and logwood volumes are calibrated using a similar calibration factor, which may not always be the optimal choice. To deal with this, the calibration could also be applied with any attribute that can be predicted with smaller errors using a univariate model, rather than a multivariate NN model. For example, the calibration could be implemented in terms of total logwood or pulpwood volume. All in all, the calibration with total volume, in particular, decreased the error rates associated with the logwood volumes. This calibration approach has similarities with the calibration estimation (Deville and Särndal 1992), which is typically used to calibrate theoretical diameter distribution (based on PDF) models to match with the separately predicted forest attributes.

The error rates associated with the prediction of species-specific diameter distributions are highly dependent on the success of species recognition. The results of paper **II** clearly show that the multispectral ALS data (1064, 1550, and 532 nm) do not outperform the combination of unispectral ALS data and aerial images in Finnish remote sensing-based forest inventories. Kukkonen et al. (2019a) also concluded that the error rates associated with species-specific volumes are noticeably larger using a multispectral ALS system than the combination of unispectral ALS and aerial images. In general, single sensor solutions for species-specific forest inventories are attractive because the operation of several sensors always increases data acquisition costs and complicates the data collection. The main issue related to the typical data acquisition for Finnish forest inventories is that ALS data and aerial images should be collected as simultaneously as possible, although simultaneous data collection with different sensors is not straightforward since the successful acquisition of aerial images, for example, is sensitive to weather and light conditions (Packalen et al. 2009). Studies have indicated that aerial images could be replaced with high-resolution satellite imagery in certain circumstances. Kukkonen et al. (2018) showed that satellite imagery provides almost comparable predictive power to aerial images in species-specific forest

inventories. The application of satellite imagery results in cost savings, but the limited availability of cloud-free images may be an issue in operational practices. The likelihood of acquiring cloud-free images would improve if the revisiting times associated with satellite instruments were shorter. However, it must be noted that forest operators may still prefer aerial images to satellite imagery for other purposes, such as stand delineation.

The results presented in paper **II** indicate that the bitemporal ALS dataset performs equally well in the species-specific prediction of logwood volumes as the combination of unispectral ALS data and aerial images. This finding may also be extended to the prediction of species-specific forest attributes in general. The results, however, indicate that the greatest benefit of a bitemporal dataset for species-specific predictions is achieved when ALS datasets are collected under leaf-off and leaf-on conditions. This finding indicates that the improved predictive performance of species-specific attributes is affected by the leaf conditions of deciduous forests under which the ALS dataset was acquired. Finnish forest inventories typically apply ALS datasets that are collected under leaf-on conditions. The collection of leaf-off data is a challenge due to the narrow acquisition season in Finnish conditions (Villikka et al. 2012).

The availability of bitemporal and multitemporal ALS datasets will increase in the near future due to the repetition of national ALS data acquisitions. The usage of older datasets would be reasonable from the point of view of economic efficiency. Few studies have proposed approaches that utilize repeated ALS data in forest inventories. One interesting approach is data assimilation, which can involve a time series of remotely sensed data in the prediction process (Nyström et al. 2015). The aim of the data assimilation technique is to update forest inventory predictions using the time series of remotely sensed data and field measurements. In general, the data assimilation technique requires more than two data acquisition points, which means that bitemporal data are not adequate for analyses. Bitemporal ALS data have been applied, for example, in the detection of growth and change (e.g. Næsset and Gobakken 2005; Bohlin et al. 2017). The approach presented in paper **II** requires that the ALS features were computed from a bitemporal dataset, and any further features related to, for example, the differences in ALS datasets were not considered. The application of the approach itself is straightforward but bitemporal ALS datasets may involve areas where forests have been thinned or clear-cut. The silvicultural activity that occurs between the data acquisitions may distort the prediction routines. In paper **II**, the silvicultural activities were detected using the changes in density features derived from ALS datasets, but the approach was not validated owing to a lack of ground truth data.

Previous studies have indicated that the increment in pulse density itself does not significantly decrease the error rates associated with the ALS-based forest inventories (Gobakken and Næsset 2008; Strunk et al. 2012). An examination of the results presented in paper **II** also indicates that pulse density does not explicitly affect the error rates related to the predicted diameter distributions. Instead, the high-density ALS dataset enables the application of the ITD approach. The findings in paper **III** generally showed that ITD has several shortcomings in the prediction of diameter distributions in boreal conditions. The main issue that makes the application of ITD such a challenge for the prediction of diameter distribution is the inaccurate detection of trees in dense young or multi-layered forests. However, trees in a dominant canopy can be, in general, reliably detected using ITD, which makes the fusion of ABA and ITD compelling, when compared to the use of ABA alone.

The findings described in paper **III** suggest that the shape of the diameter distribution will be indicative of which prediction approach (ABA, ITD, or ABA-ITD fusion) performs best. The results reported in paper **III** show that the prediction of diameter distributions in bimodal

and Gaussian forests benefits from the utilization of ITD. The findings indicated that the application of weighting-based fusion (F.PRED) is a challenge, since the prediction of the fusion parameter ($\beta_{F.PRED}$) is difficult. Instead, the fusion, which was based on the replacement of ABA distributions with ITD distributions in Gaussian forests (F.REPL), could be adapted in operational forest inventories. The requirement of the F.REPL approach is that DDSC must be predicted with a sufficient level of accuracy, since the F.REPL approach applies ABA and ITD explicitly according to DDSC. Therefore, the misclassifications may lead to severe errors in the prediction of diameter distributions when diameter distribution predicted by ITD is assigned to, for example, a bimodal plot.

4.2 Future research

The application of bitemporal ALS data requires further research in regard to the features extracted from the ALS datasets and the time difference between data acquisitions, for example. Features that involve the relationships between ALS datasets could provide information that correlate with forest characteristics, such as forest growth. Here, the features were broadly similar to those used in Finnish ABA inventories. The time difference between the acquisitions was five years. However, the difference may be greater or less than five years in practical cases where nation-wide ALS datasets are applied.

There are also other new remote sensing data sources that could be considered in the prediction of diameter distributions in the future. For example, interest in the application of unmanned aerial vehicles (UAVs) has increased substantially in recent years (Colomina and Molina 2014). The UAVs applied for forestry purposes are typically equipped with digital cameras (multi- or hyperspectral) but LiDAR sensors are also used (Sankey et al. 2017). The UAV-based forest inventories cannot cover large areas, which effectively means high costs per area unit. However, the use of UAVs in the prediction of diameter distributions in forest management planning and timber procurement should be considered as, in some cases, field visits to forest stands are needed anyway. For example, the utilization of operational forest inventory data and an UAV inventory at the level of individual forest stands (e.g. mature stands marked for cutting) may decrease the errors associated with species-specific diameter distributions. In specific forest structures and types, the determination of DBH from UAV data may also be feasible (Kattenborn et al. 2018; Puliti et al. 2020). Smaller error rates associated with diameter distributions are especially important for timber procurement.

Terrestrial laser scanning (TLS) could be used in the collection of field samples in the future (Pitkänen et al. 2019). In order to replace current field measurements, TLS measurements should provide at least as accurate data as the present methods, and the acquisition costs should be affordable. It is evident that TLS could measure some attributes better than the present measurement approaches, for example, TLS data can be used to derive stem curves (Liang et al. 2014). However, the direct measurement of tree height may be an issue due to the lack of visibility in a forest (Liang et al. 2016). The visibility issue also means that the TLS point cloud measured from a single point cannot be used to detect all trees as such, since trees can be concealed behind other trees (Kansanen et al. 2019b). Difficulties in recognizing tree species from TLS point clouds may also induce errors into the TLS measurements (Othmani et al. 2013). For these reasons, TLS may not be the definitive technique for the measurement of diameter distributions in forest inventories.

It is also possible to collect LiDAR data from forests during field visits or while conducting silvicultural operations. For example, a person carrying a LiDAR system, or a

harvester equipped with a LiDAR sensor, might be attractive alternatives for data acquisition. These alternatives could be more reasonable than TLS systems because the movement related to data acquisition itself does not cause additional costs.

Forthcoming shifts in remote sensing-based forest inventories may be strongly related to the management and exploitation of different data sources in an optimal way. It is critical to know which datasets should be involved in the inventory process in order to achieve the smallest errors, for example, at the scale of individual forest stands, or at the national-level. In addition to the utilization of data sources, methodological shifts may be possible and could provide smaller prediction error rates for forest attributes. For example, the forthcoming shift from techniques that rely on PDF to NN-based tree lists is a step forward in the prediction of diameter distributions in Finnish forest inventories. Future studies may also pursue the incorporation of deep learning approaches to the field of forest inventories (Niska et al. 2010; Garcia-Gutiérrez et al. 2016). Nevertheless, the predictive performance, robustness and simplicity of the multivariate NN approach are superior characteristics for Finnish forest inventories. It should be noted that deep learning approaches are not as transparent as the NN approach from the point of view of the user. On the one hand, it is evident that investigations related to the application of deep learning in ALS-based forest inventories are welcome. Yet, on the other hand, the errors associated with species-specific forest inventories are more likely to be reduced by the selection of optimal remote sensing and field data rather than by introducing new modeling approaches.

5 CONCLUSIONS

The response configuration used in the NN imputation of species-specific diameter distributions influences the error rates associated with the predicted diameter distributions. In general, the results in this thesis showed that the Finnish operational standard (15 species-specific forest attributes as response variables) resulted in larger error rates associated with the predicted diameter distributions compared to the response configuration in which the species-specific volumes and heights were omitted. The findings also indicate that the separate NN imputation by tree species outperforms simultaneous NN imputation, especially in the prediction of minor tree species.

The multispectral ALS data provided smaller errors associated with predicted logwood volumes than leaf-on unispectral ALS data alone. However, the errors were greater compared to those achieved with the combined use of unispectral ALS data and aerial images. The bitemporal datasets provided comparable predictive performance for logwood volumes than the combination of ALS data and aerial images.

The results indicate that ITD is more prone to errors than ABA, in forests with reverse-J and bimodal diameter distributions. However, the ITD approach outperformed ABA in the prediction of Gaussian-shaped diameter distributions. Two approaches were proposed to fuse the diameter distributions predicted using ABA and ITD, in a practical setting. Both approaches resulted in a minor improvement in mean error index values compared to the ABA approach, which indicates that ITD could provide additional information to ABA in the prediction of diameter distributions in boreal conditions.

REFERENCES

Axelsson P. (2000). DEM Generation from laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing 33(4): 110–117.

Axelsson A., Lindberg E., Olsson H. (2018). Exploring multispectral ALS data for tree species classification. Remote Sensing 10(2): 183. https://doi.org/10.3390/rs10020183

Bailey R.L., Dell T.R. (1973). Quantifying diameter distributions with the Weibull function. Forest Science 19(2): 97–104. https://doi.org/10.1093/forestscience/19.2.97

Bohlin I., Olsson H., Bohlin J., Granström A. (2017). Quantifying post-fire fallen trees using multi-temporal lidar. International Journal of Applied Earth Observation and Geoinformation 63: 186–195. https://doi.org/10.1016/j.jag.2017.08.004

Borders B.E., Souter R.A., Bailey R.L., Ware K.D. (1987). Percentile-based distributions characterize forest stand tables. Forest Science 33(2): 570–576. https://doi.org /10.1093/forestscience/33.2.570

Brandtberg T. (1999). Automatic individual tree based analysis of high spatial resolution aerial images on naturally regenerated boreal forests. Canadian Journal of Forest Research 29(10): 1464–1478. https://doi.org/10.1139/x99-150

Breidenbach J., Næsset E., Lien V., Gobakken T., Solberg S. (2010). Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. Remote Sensing of Environment 114(4): 911–924. https://doi.org/10.1016/j.rse.2009.12.004

Burk T.E., Newberry J.D. (1984). A simple algorithm for moment-based recovery of Weibull distribution parameters. Forest Science 30(2): 329–332. https://doi.org/10.1093/forestscience/30.2.329

Cajanus W. (1914). Ueber die Entwicklung gleichaltriger Waldbestände. Acta Forestalia Fennica 3: 142.

Cao Q.V., Burkhart H.E. (1984). A segmented distribution approach for modeling diameter frequency data. Forest Science 30(1): 129–137.

Colomina I., Molina P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 92: 79–97. https://doi.org/10.1016/j.isprsjprs.2014.02.013

Cosenza D.N., Soares P., Guerra-Hernández J., Pereira L., González-Ferreiro E., Castedo-Dorado F., Tomé M. (2019). Comparing Johnson's SB and Weibull functions to model the diameter distribution of forest plantations through ALS data. Remote Sensing 11(23): 2792. https://doi.org/10.3390/rs11232792

Dalponte M., Orka H.O., Gobakken T., Gianelle D., Naesset E. (2013). Tree species classification in boreal forests with hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 51(5): 2632–2645. https://doi.org/10.1109/TGRS.2012.2216272

Dalponte M., Ørka H.O., Ene L.T., Gobakken T., Næsset E. (2014). Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. Remote Sensing of Environment 140: 306–317. https://doi.org/10.1016/j.rse.2013.09.006

Dalponte M., Ene L.T., Gobakken T., Næsset E., Gianelle D. (2018). Predicting selected forest stand characteristics with multispectral ALS data. Remote Sensing 10(4): 586. https://doi.org/10.3390/rs10040586

Deville J.-C., Särndal C.-E. (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association 87(418): 376–382. https://doi.org/10.1080/01621459.1992.10475217

Duncanson L.I., Cook B.D., Hurtt G.C., Dubayah R.O. (2014). An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems. Remote Sensing of Environment 154: 378–386. https://doi.org/10.1016/j.rse.2013.07.044

Eerikäinen K. (2009). A multivariate linear mixed-effects model for the generalization of sample tree heights and crown ratios in the Finnish national forest inventory. Forest Science 55(6): 480–493. https://doi.org/10.1093/forestscience/55.6.480

Ene L., Næsset E., Gobakken T. (2012). Single tree detection in heterogeneous boreal forests using airborne laser scanning and area-based stem number estimates. International Journal of Remote Sensing 33(16): 5171–5193. https://doi.org/10.1080/01431161.2012.657363

Esseen P.-A., Ehnström B., Ericson L., Sjöberg K. (1997). Boreal Forests. Ecological Bulletins (46): 16–47.

García-Gutiérrez J., González-Ferreiro E., Mateos-García D., Riquelme-Santos J.C. (2016). A preliminary study of the suitability of deep learning to improve LiDAR-derived biomass estimation. In: Martínez-Álvarez F., Troncoso A., Quintián H., Corchado E. (ed.) Hybrid Artificial Intelligent Systems. Springer International Publishing, Cham. p. 588–596. https://doi.org/10.1007/978-3-319-32034-2_49

Gauch J.M. (1999). Image segmentation and analysis via multiscale gradient watershed hierarchies. IEEE Transactions on Image Processing 8(1): 69–79. https://doi.org/10.1109/83.736688

Gobakken T., Næsset E. (2004). Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. Scandinavian Journal of Forest Research 19(6): 529–542. https://doi.org/10.1080/02827580410019454

Gobakken T., Næsset E. (2005). Weibull and percentile models for lidar-based estimation of basal area distribution. Scandinavian Journal of Forest Research 20(6): 490–502. https://doi.org/10.1080/02827580500373186

Gobakken T., Næsset E. (2008). Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. Canadian Journal of Forest Resesearch. 38(5): 1095–1109. https://doi.org/10.1139/X07-219

Haara A., Korhonen K. (2004). Kuvioittaisen arvioinnin luotettavuus. Metsätieteen aikakauskirja 2004(4): 489–508. https://doi.org/10.14214/ma.5667

Haara A., Maltamo M., Tokola T. (1997). The K-nearest-neighbour method for estimating basal-area diameter distribution. Scandinavian Journal of Forest Research 12(2): 200–208. https://doi.org/10.1080/02827589709355401

Hafley W.L., Schreuder H.T. (1977). Statistical distributions for fitting diameter and height data in even-aged stands. Canadian Journal of Forest Research 7(3): 481–487. https://doi.org/10.1139/x77-062

Hou Z., Xu Q., Vauhkonen J., Maltamo M., Tokola T. (2016). Species-specific combination and calibration between area-based and tree-based diameter distributions using airborne laser scanning. Canadian Journal of Forest Research 46(6): 753–765. https://doi.org/10.1139/cjfr-2016-0032

Hynynen J., Ojansuu R., Hökkä H., Siipilehto J., Salminen H., Haapala P. (2002). Models for predicting stand development in MELA System. Metsäntutkimuslaitos.

Hyyppä J., Inkinen M. (1999). Detecting and estimating attributes for single trees using laser scanner. Photogrammetric Journal of Finland 16: 27–42.

Kalliovirta J., Tokola T. (2005). Functions for estimating stem diameter and tree age using tree height, crown width and existing stand database information. Silva Fennica 39(2): 386. https://doi.org/10.14214/sf.386

Kangas A., Maltamo M. (2000a). Percentile based basal area diameter distribution models for Scots pine, Norway spruce and birch species. Silva Fennica 34(4): 619. https://doi.org/10.14214/sf.619

Kangas A., Maltamo M. (2000b). Calibrating predicted diameter distribution with additional information. Forest Science 46(3): 390–396. https://doi.org/10.1093/forestscience/46.3.390

Kangas A., Heikkinen E., Maltamo M. (2004). Accuracy of partially visually assessed stand characteristics: a case study of Finnish forest inventory by compartments. Canadian Journal of Forest Research 34(4): 916–930. https://doi.org/10.1139/x03-266

Kansanen K., Vauhkonen J., Lähivaara T., Seppänen A., Maltamo M., Mehtätalo L. (2019a). Estimating forest stand density and structure using Bayesian individual tree detection, stochastic geometry, and distribution matching. ISPRS Journal of Photogrammetry and Remote Sensing 152: 66–78. https://doi.org/10.1016/j.isprsjprs.2019.04.007

Kansanen K., Packalen P., Maltamo M., Mehtätalo L. (2019b). Horvitz-Thompson-like estimation with distance-based detection probabilities for circular plot sampling of forests. Preprint. Available from: https://arxiv.org/abs/1910.14647

Kattenborn T., Hernández J., Lopatin J., Kattenborn G., Fassnacht F.E. (2018). Pilot study on the retrieval of DBH and diameter distribution of deciduous forest stands using cast shadows in UAV-based orthomosaics. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences IV-1: 93–99. https://doi.org/10.5194/isprs-annals-IV-1-93-2018

Keränen J., Peuhkurinen J., Packalen P., Maltamo M. (2015). Effect of minimum diameter at breast height and standing dead wood field measurements on the accuracy of ALS-based forest inventory. Canadian Journal of Forest Research 45(10): 1280–1288. https://doi.org/10.1139/cjfr-2015-0103

Kilkki P., Maltamo M., Mykkänen R., Päivinen R. (1989). Use of the Weibull function in estimating the basal-area diameter distribution. Silva Fennica 23: 311-318. https://doi.org/10.14214/sf.a15550

Kirkpatrick S., Gelatt C.D., Vecchi M.P. (1983). Optimization by simulated annealing. Science 220(4598): 671–680. https://doi.org /10.1126/science.220.4598.671

Koivuniemi J., Korhonen K. (2006). Inventory by Compartments. In: Kangas A., Maltamo M. (ed.) Forest Inventory. Managing Forest Ecosystems, vol 10. Springer, Dordrecht.

Korpela I., Dahlin B., Schäfer H., Bruun E., Haapaniemi F., Honkasalo J., Ilvesniemi S., Kuutti V., Linkosalmi M., Mustonen J. (2007). Single-tree forest inventory using lidar and aerial images for 3D treetop positioning, species recognition, height and crown width estimation. In Proceedings of ISPRS workshop on laser scanning 2007 and SilviLaser 2007, Espoo, Finland, 12–14 Sept 2007, vol XXXVI, Part 3/W52. p. 227–233.

Korpela I., Tokola T., Ørka H.O., Koskinen M. (2009). Small-footprint discrete-return LIDAR in tree species recognition. Proceedings of ISPRS Hannover workshop 2009 on High Resolution Earth Imaging for Geospatial Information IAPRS XXXVIII-1-4-7/W5: 6.

Korpela I., Ørka H., Maltamo M., Tokola T., Hyyppä, J. (2010). Tree species classification using airborne LiDAR – effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. Silva Fennica 44(2):156. https://doi.org/10.14214/sf.156

Kukkonen M., Korhonen L., Maltamo M., Suvanto A., Packalen P. (2018). How much can airborne laser scanning based forest inventory by tree species benefit from auxiliary optical data? International Journal of Applied Earth Observation and Geoinformation 72: 91–98. https://doi.org/10.1016/j.jag.2018.06.017

Kukkonen M., Maltamo M., Korhonen L., Packalen P. (2019a). Comparison of multispectral airborne laser scanning and stereo matching of aerial images as a single sensor solution to forest inventories by tree species. Remote Sensing of Environment 231: 111208. https://doi.org/10.1016/j.rse.2019.05.027

Kukkonen M., Maltamo M., Korhonen L., Packalen P. (2019b). Multispectral airborne LiDAR data in the prediction of boreal tree species composition. IEEE Transactions on Geoscience and Remote Sensing 57(6): 3462–3471. https://doi.org/10.1109/TGRS.2018.2885057

Kuuluvainen T., Mäki J., Karjalainen L., Lehtonen H. (2002). Tree age distributions in old-growth forest sites in Vienansalo wilderness, eastern Fennoscandia. Silva Fennica 36(1): 556. https://doi.org/10.14214/sf.556

Laasasenaho J. (1982). Taper curve and volume functions for pine, spruce and birch. Communicationes Instituti Forestalis Fenniae 108: 1–74. Available from: http://urn.fi/URN:ISBN:951-40-0589-9

Liang X., Kankare V., Yu X., Hyyppä J., Holopainen M. (2014). Automated stem curve measurement using terrestrial laser scanning. IEEE Transactions on Geoscience and Remote Sensing 52(3): 1739–1748. https://doi.org/10.1109/TGRS.2013.2253783

Liang X., Kankare V., Hyyppä J., Wang Y., Kukko A., Haggrén H., Yu X., Kaartinen H., Jaakkola A., Guan F., Holopainen M., Vastaranta M. (2016). Terrestrial laser scanning in forest inventories. ISPRS Journal of Photogrammetry and Remote Sensing 115: 63–77. https://doi.org/10.1016/j.isprsjprs.2016.01.006

Lindberg E., Holmgren J., Olofsson K., Wallerman J., Olsson H. (2010). Estimation of tree lists from airborne laser scanning by combining single-tree and area-based methods. International Journal of Remote Sensing 31(5): 1175–1192. https://doi.org/10.1080/01431160903380649

Liu C., Zhang L., Davis C.J., Solomon D.S., Gove J.H. (2002). A finite mixture model for characterizing the diameter distributions of mixed-species forest stands. Forest Science 48(4): 653–661. https://doi.org/10.1093/forestscience/48.4.653

Loetsch F., Zöhrer F., Haller K.E. (1973). Forest inventory 2. Munique, B LV-VerlagsgeselIschaft.

Lönnroth E. (1925). Untersuchungen über die innere Struktur und Entwicklung gleichaltriger naturnormaler Kiefernbestände: basiert auf Material aus der Südhälfte Finnlands. Acta Forestalia Fennica 30(1): 7196. https://doi.org/10.14214/aff.7196

Magnussen S., Renaud J.-P. (2016). Multidimensional scaling of first-return airborne laser echoes for prediction and model-assisted estimation of a distribution of tree stem diameters. Annals of Forest Science 73(4): 1089–1098. https://doi.org/10.1007/s13595-016-0581-2

Magnussen S., Næsset E., Gobakken T. (2013). Prediction of tree-size distributions and inventory variables from cumulants of canopy height distributions. Forestry 86(5): 583–595. https://doi.org/10.1093/forestry/cpt022

Maltamo M., Packalen P. (2014). Species-specific management inventory in Finland. In: Maltamo M., Næsset E., Vauhkonen J. (ed.) Forestry applications of airborne laser scanning: Concepts and case studies. Springer Netherlands, Dordrecht. p. 241–252. https://doi.org/10.1007/978-94-017-8663-8_12

Maltamo M., Eerikäinen K., Pitkänen J., Hyyppä J., Vehmas M. (2004). Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sensing of Environment 90(3): 319–330. https://doi.org/10.1016/j.rse.2004.01.006

Maltamo M., Suvanto A., Packalén P. (2007). Comparison of basal area and stem frequency diameter distribution modelling using airborne laser scanner data and calibration estimation. Forest Ecology and Management 247(1): 26–34. https://doi.org/10.1016/j.foreco.2007.04.031

Maltamo M., Næsset E., Bollandsås O.M., Gobakken T., Packalén P. (2009a). Non-parametric prediction of diameter distributions using airborne laser scanner data. Scandinavian Journal of Forest Research 24(6): 541–553. https://doi.org/10.1080/02827580903362497

Maltamo M., Peuhkurinen J., Malinen J., Vauhkonen J., Packalén P., Tokola T. (2009b). Predicting tree attributes and quality characteristics of Scots pine using airborne laser scanning data. Silva Fennica 43(3): 203. https://doi.org/10.14214/sf.203

Maltamo M., Mehtätalo L., Valbuena R., Vauhkonen J., Packalen P. (2018). Airborne laser scanning for tree diameter distribution modelling: a comparison of different modelling alternatives in a tropical single-species plantation. Forestry 91(1): 121–131. https://doi.org/10.1093/forestry/cpx041

McRoberts R.E., Chen Q., Domke G.M., Næsset E., Gobakken T., Chirici G., Mura M. (2017). Optimizing nearest neighbour configurations for airborne laser scanning-assisted estimation of forest volume and biomass. Forestry 90(1): 99–111. https://doi.org/10.1093/forestry/cpw035

Mehtätalo L., Maltamo M., Packalén P. (2007). Recovering plot-specific diameter distribution and height-diameter curve using ALS based stand characteristics. ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, September 12-14, Espoo, Finland.

Moeur M., Stage A.R. (1995). Most Similar Neighbor: An improved sampling inference procedure for natural resource planning. Forest Science 41(2): 337–359. https://doi.org/10.1093/forestscience/41.2.337

Mykkänen R. (1986). Weibull-funktion käyttö puuston läpimittajakauman estimoinnissa. Master's thesis, University of Eastern Finland, Faculty of Forestry. p. 80. (In Finnish) Natural Resources Institute Finland. (2013). Volume of pine, spruce and birch (data version 15). Available from http://kartta.luke.fi/opendata/valinta-en.html. [Accessed 27 April 2017].

Næsset E. (1997). Estimating timber volume of forest stands using airborne laser scanner data. Remote Sensing of Environment 61(2): 246–253. https://doi.org/10.1016/S0034-4257(97)00041-2

Næsset E. (2014). Area-based inventory in Norway – From innovation to an operational reality. In: Maltamo M, Næsset E., Vauhkonen J. Forestry applications of airborne laser scanning: Concepts and case studies. Springer Netherlands, Dordrecht. p. 215–240. https://doi.org/10.1007/978-94-017-8663-8_11

Næsset E., Gobakken T. (2005). Estimating forest growth using canopy metrics derived from airborne laser scanner data. Remote Sensing of Environment 96(3): 453–465. https://doi.org/10.1016/j.rse.2005.04.001

Narendra P.M., Goldberg M. (1980). Image segmentation with directed trees. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-2(2): 185–191. https://doi.org/10.1109/TPAMI.1980.4766999

Nelson R. (2013). How did we get here? An early history of forestry lidar. Canadian Journal of Remote Sensing 39(sup1): S6–S17. https://doi.org/10.5589/m13-011

Nilsson M. (1996). Estimation of tree heights and stand volume using an airborne lidar system. Remote Sensing of Environment 56(1): 1–7. https://doi.org/10.1016/0034-4257(95)00224-3

Niska H., Skon J.-P., Packalen P., Tokola T., Maltamo M., Kolehmainen M. (2010). Neural networks for the prediction of species-specific plot volumes using airborne laser scanning and aerial photographs. IEEE Transactions on Geoscience and Remote Sensing 48(3): 1076–1085. https://doi.org/10.1109/TGRS.2009.2029864

Nyström M., Lindgren N., Wallerman J., Grafström A., Muszta A., Nyström K., Ståhl G., Olsson H. (2015). Data assimilation in forest inventory: First empirical results using ALS data. In: Durrieu, S., Vega C. (ed.) Proceedings of SilviLaser 2015, 28–30 September 2015, La Grand Motte, France: Irstea-UMR TETIS, IGN-LIF and Société Française dePhotogrammétrie et de Télédétection (SFPT), p. 174-176.

Nyyssönen A. (1954). Metsikön kuutiomäärän arvioiminen relaskoopin avulla. Summary: Estimation of stand volume by means of the relascope. Communicationes Instituti Forestalis Fenniae 44(6): 31.

Optech. (2020). Optech Titan multispectral lidar system - high precision environmental mapping. Available from: http://info.teledyneoptech.com/acton/attachment/19958/f-02b6/1/-/-/-/-/Titan-Specsheet-150515-WEB.pdf. [Cited 2 January 2020]

Othmani A., Lew Yan Voon L.F.C., Stolz C., Piboule A. (2013). Single tree species classification from terrestrial laser scanning data for forest inventory. Pattern Recognition Letters 34(16): 2144–2150. https://doi.org/10.1016/j.patrec.2013.08.004

Packalen P., Maltamo M. (2006). Predicting the plot volume by tree species using airborne laser scanning and aerial photographs. Forest Science 52(6): 611–622. https://doi.org/10.1093/forestscience/52.6.611

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): 328–341. https://doi.org/10.1016/j.rse.2007.01.005

Packalén P., Maltamo M. (2008). Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. Canadian Journal of Forest Research 38(7): 1750–1760. https://doi.org/10.1139/X08-037

Packalén P., Pitkänen J., Maltamo M. (2008). Comparison of individual tree detection and canopy height distribution approaches: a case study in Finland. In: Hill R, Rosette J., Suárez J. (ed.) Proceedings of SilviLaser 2008, 17–19 September, 2008, Edinburgh, UK.

Packalén P., Suvanto A., Maltamo M. (2009). A two stage method to estimate species-specific growing stock. Photogrammetric Engineering & Remote Sensing 75(12): 1451–1460. https://doi.org/10.14358/pers.75.12.1451

Packalén P., Temesgen H., Maltamo M. (2012). Variable selection strategies for nearest neighbor imputation methods used in remote sensing based forest inventory. Canadian Journal of Remote Sensing 38(5): 557–569. https://doi.org/10.5589/m12-046

Packalen P., Vauhkonen J., Kallio E., Peuhkurinen J., Pitkänen J., Pippuri I., Strunk J., Maltamo M. (2013). Predicting the spatial pattern of trees by airborne laser scanning. International Journal of Remote Sensing 34(14): 5154–5165. https://doi.org/10.1080/01431161.2013.787501

Packalen P., Strunk J.L., Pitkänen J.A., Temesgen H., Maltamo M. (2015). Edge-tree correction for predicting forest inventory attributes using area-based approach with airborne laser scanning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8(3): 1274–1280. https://doi.org/10.1109/JSTARS.2015.2402693

Persson A., Holmgren J., Soderman U. (2002). Detecting and measuring individual trees using an airborne laser scanner. Photogrammetric Engineering and Remote Sensing 68(9): 925–932.

Peuhkurinen J., Maltamo M., Malinen J. (2008). Estimating species-specific diameter distributions and saw log recoveries of boreal forests from airborne laser scanning data and aerial photographs: a distribution-based approach. Silva Fennica 42(4): 237. https://doi.org/10.14214/sf.237

Peuhkurinen J., Mehtätalo L., Maltamo M. (2011). Comparing individual tree detection and the area-based statistical approach for the retrieval of forest stand characteristics using airborne laser scanning in Scots pine stands. Canadian Journal of Forest Research 41(3): 583–598. https://doi.org/10.1139/X10-223

Pitkänen J., Maltamo M., Hyyppä J., Yu X. (2004). Adaptive methods for individual tree detection on airborne laser based. International Archives of Photogrammetry. Remote Sensing and Spatial Information Sciences 36(8): 187–191.

Pitkänen T.P., Raumonen P., Kangas A. (2019). Measuring stem diameters with TLS in boreal forests by complementary fitting procedure. ISPRS Journal of Photogrammetry and Remote Sensing 147: 294–306. https://doi.org/10.1016/j.isprsjprs.2018.11.027

Poudel K.P., Cao Q.V. (2013). Evaluation of methods to predict Weibull parameters for characterizing diameter distributions. Forest Science 59(2): 243–252. https://doi.org/10.5849/forsci.12-001

Puliti S., Breidenbach J., Astrup R. (2020). Estimation of Forest Growing Stock Volume with UAV Laser Scanning Data: Can It Be Done without Field Data?. Remote Sensing 12(8): 1245. https://doi.org/10.3390/rs12081245

Päivinen R. (1980). Puiden läpimittajakauman estimointi ja siihen perustuva puustotunnusten laskenta. Folia Forestalia 442: 1–28. Available from: http://urn.fi/URN:ISBN:951-40-0461-2

Reitberger J., Schnörr Cl., Krzystek P., Stilla U. (2009). 3D segmentation of single trees exploiting full waveform LIDAR data. ISPRS Journal of Photogrammetry and Remote Sensing 64(6): 561–574. https://doi.org/10.1016/j.isprsjprs.2009.04.002

Repola J. (2009). Biomass equations for Scots pine and Norway spruce in Finland. Silva Fennica 43(4): 184. https://doi.org/10.14214/sf.184

Reynolds M., Burk T., Huang W. (1988). Goodness-of-fit tests and model selection procedures for diameter distribution models. Forest Science 34(2): 373–399.

Rouvinen S., Kuuluvainen T. (2005). Tree diameter distributions in natural and managed old Pinus sylvestris-dominated forests. Forest Ecology and Management 208(1): 45–61. https://doi.org/10.1016/j.foreco.2004.11.021

Räty J., Vauhkonen J., Maltamo M., Tokola T. (2016). On the potential to predetermine dominant tree species based on sparse-density airborne laser scanning data for improving subsequent predictions of species-specific timber volumes. Forest Ecosystems 3: 1. https://doi.org/10.1186/s40663-016-0060-0

Sankey T., Donager J., McVay J., Sankey J.B. (2017). UAV lidar and hyperspectral fusion for forest monitoring in the southwestern USA. Remote Sensing of Environment 195: 30–43. https://doi.org/10.1016/j.rse.2017.04.007

Shin J., Temesgen H. (2018). Generating tree-lists by fusing individual tree detection and nearest neighbor imputation using airborne LiDAR data. Open Journal of Forestry 8(4): 500–531. https://doi.org/10.4236/ojf.2018.84032

Siipilehto J. (1999). Improving the accuracy of predicted basal-area diameter distribution in advanced stands by determining stem number. Silva Fennica 33(4): 650. https://doi.org/10.14214/sf.650

Siipilehto J., Mehtätalo L. (2013). Parameter recovery vs. parameter prediction for the Weibull distribution validated for Scots pine stands in Finland. Silva Fennica 47(4): 1057. https://doi.org/10.14214/sf.1057

Strunk J., Temesgen H., Andersen H.-E., Flewelling J.P., Madsen L. (2012). Effects of lidar pulse density and sample size on a model-assisted approach to estimate forest inventory variables. Canadian Journal of Remote Sensing 38(5): 644–654. https://doi.org/10.5589/m12-052

Strunk J.L., Gould P.J., Packalen P., Poudel K.P., Andersen H-E., Temesgen H. (2017). An examination of diameter density prediction with k-NN and airborne lidar. Forests 8(11): 444. https://doi.org/10.3390/f8110444

Suomen metsäkeskus. (2016). Kaukokartoitusperusteisen metsien inventoinnin koealojen maastotyöohje, versio 1.7. Maastotyöohje. (In Finnish)

Thomas V., Oliver R.D., Lim K., Woods M. (2008). LiDAR and Weibull modeling of diameter and basal area. The Forestry Chronicle 84(6): 866–875. https://doi.org/10.5558/tfc84866-6

Tomppo E., Halme M. (2004). Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: a genetic algorithm approach. Remote Sensing of Environment 92(1): 1–20. https://doi.org/10.1016/j.rse.2004.04.003

Vauhkonen J. (2020). Effects of diameter distribution errors on stand management decisions according to a simulated individual tree detection. Annals of Forest Science 77: 21. https://doi.org/10.1007/s13595-020-0918-8

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(6): 1263–1276. https://doi.org/10.1016/j.rse.2010.01.016

Vauhkonen J., Ørka H.O., Holmgren J., Dalponte M., Heinzel J., Koch B. (2014). Tree Species recognition based on airborne laser scanning and complementary data sources. In: Maltamo M., Næsset E., Vauhkonen J. (ed.) Forestry applications of airborne laser scanning: Concepts and case studies. Springer Netherlands, Dordrecht. p. 135–156. https://doi.org/10.1007/978-94-017-8663-8_7

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 46(1): 68. https://doi.org/10.14214/sf.68

Wagner W., Ullrich A., Ducic V., Melzer T., Studnicka N. (2006). Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. ISPRS Journal of Photogrammetry and Remote Sensing 60(2): 100–112. https://doi.org/10.1016/j.isprsjprs.2005.12.001

Wallenius T., Laamanen R., Peuhkurinen J., Mehtätalo L., Kangas A. (2012). Analysing the agreement between an airborne laser scanning based forest inventory and a control inventory – a case study in the state owned forests in Finland. Silva Fenn. 46(1): 69. https://doi.org/10.14214/sf.69

Wang M., Rennolls K. (2005). Tree diameter distribution modelling: introducing the logit–logistic distribution. Canadian Journal of Forest Research 35(6): 1305–1313. https://doi.org/10.1139/x05-057

Wehr A., Lohr U. (1999). Airborne laser scanning—an introduction and overview. ISPRS Journal of Photogrammetry and Remote Sensing 54(2): 68–82. https://doi.org/10.1016/S0924-2716(99)00011-8

Xu Q., Hou Z., Maltamo M., Tokola T. (2014). Calibration of area based diameter distribution with individual tree based diameter estimates using airborne laser scanning. ISPRS Journal of Photogrammetry and Remote Sensing 93: 65–75. https://doi.org/10.1016/j.isprsjprs.2014.03.005

Xu Q., Li B., Maltamo M., Tokola T., Hou Z. (2019). Predicting tree diameter using allometry described by non-parametric locally-estimated copulas from tree dimensions derived from airborne laser scanning. Forest Ecology and Management 434: 205–212. https://doi.org/10.1016/j.foreco.2018.12.020

Zhao K., Suarez J.C., Garcia M., Hu T., Wang C., Londo A. (2018). Utility of multitemporal lidar for forest and carbon monitoring: Tree growth, biomass dynamics, and carbon flux. Remote Sensing of Environment 204: 883–897. https://doi.org/10.1016/j.rse.2017.09.007