

**Dissertationes Forestales 104**

Estimating single-tree attributes by airborne laser  
scanning: methods based on computational geometry  
of the 3-D point data

Jari Vauhkonen  
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 in the auditorium AU100 of the University of Eastern Finland, Yliopistokatu 2, Joensuu, on 12<sup>th</sup> May 2010, at 12 o'clock noon.

*Title of dissertation:* Estimating single-tree attributes by airborne laser scanning: methods based on computational geometry of the 3-D point data

*Author:* Jari Vauhkonen

*Dissertationes Forestales 104*

*Thesis supervisors:*

Prof. Timo Tokola

School of Forest Sciences, University of Eastern Finland, Joensuu, Finland

Prof. Matti Maltamo

School of Forest Sciences, University of Eastern Finland, Joensuu, Finland

Dr. Ilkka Korpela

Department of Forest Sciences, University of Helsinki, Helsinki, Finland

*Pre-examiners:*

Prof. Juha Hyyppä

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

Prof. Christoph Kleinn

Department of Forest Inventory and Remote Sensing, Georg-August-University, Göttingen, Germany

*Opponent:*

Prof. Håkan Olsson

Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Sweden

ISSN 1795-7389

ISBN 978-951-651-297-9 (PDF)

(2010)

*Publishers:*

The Finnish Society of Forest Science

Finnish Forest Research Institute

Faculty of Agriculture and Forestry of the University of Helsinki

School of Forest Sciences of the University of Eastern Finland

*Editorial office:*

The Finnish Society of Forest Science

P.O. Box 18, FI-01301 Vantaa, Finland

<http://www.metla.fi/dissertationes>

**Vauhkonen, J.** 2010. Estimating single-tree attributes by airborne laser scanning: methods based on computational geometry of the 3-D point data. *Dissertationes Forestales* 104. 44 p. Available at <http://www.metla.fi/dissertationes/df104.htm>.

## ABSTRACT

Airborne laser scanning (ALS) has become a very common forest inventory data source during the 2000's. Previous research on single-tree interpretation of such data suggests limitations due to both undetected trees and inaccuracies in species recognition and allometric estimation of stem dimensions. This work examined reconstruction of tree crowns by means of computational geometry of the point data and techniques for turning the obtained crown shape and structure information into improved estimates of tree attributes.

Alpha shape metrics, i.e. a collection of various volume, complexity and area features derived from 3-D alpha shapes based on the point data, were found to have potential for describing species-specific allometric differences in the trees, while combining these metrics with features based on the height and intensity distributions in the data was beneficial with respect to the final accuracies. Nearest neighbor estimation proved efficient for making use of the high number of predictors available, but also for the simultaneous estimation of the attributes of interest, thus avoiding error propagation of an estimation chain. Random Forest, in particular, proved to be a flexible method with an ability to handle all available predictors with no need for their reduction. The classification of dominant to intermediate Scots pine, Norway spruce and deciduous trees showed an accuracy of 78%, and the estimates of diameter at breast height, tree height, and stem volume had root mean square errors of 13%, 3%, and 31%, respectively, when evaluated against separate validation data.

Less supervised tree detection and estimation resulted in unreliable tree-level descriptions of the test stands, being hindered by both inaccuracy in the tree attributes, especially in species identification, and errors in tree delineation. The need to acquire field reference data and a potential need for an auxiliary information source both place constraints on the applicability of the developed approach. On the other hand, it was shown that crown base height, which is an important measure of external quality of mature Scots pine trees, could be estimated with an RMSE of 20–30% solely by ALS data with a pulse density of 4 m<sup>-2</sup>. The results suggest focusing single-tree interpretation specifically towards detailed measurements on the dominant tree layer, thus presenting a further need to assess the tree-level production line with respect to obtainable information, alternative methods and their costs.

**Keywords:** Alpha shape; Delaunay triangulation; Forest inventory; LiDAR; Nearest neighbor; Random Forest

## ACKNOWLEDGEMENTS

Preparing this thesis was not a remarkably painful process – but quite the contrary! Everything went smoothly and I always seemed to have the best people around me, which now feels like a miracle. First, I had great supervisors, Prof. Timo Tokola, Prof. Matti Maltamo, and Dr. Ilkka Korpela, who co-authored the papers and commented my text and results but also many other things in life. The biggest 'thank you' goes to Prof. Tokola, without whom this work would possibly never have started, or at least not completed in the current form or pace.

In addition to my "official" supervisors, I learned a lot from Dr. Petteri Packalén, who also co-authored two papers included in this thesis. Mr. Juho Pitkänen and Dr. Kenneth Olofsson were co-authors of one paper, which I'm as well grateful of. I'm much obliged to the pre-examiners of this thesis, Professors Juha Hyypä and Christoph Kleinn, for their efficient review. Those directions important for individual studies are thanked in each paper.

I did most of this work at the Faculty of Forest Sciences of the University of Joensuu, under research projects funded by Academy of Finland and WoodWisdom-Net (WW-IRIS). I would like to thank my colleagues at Joensuu, but also at the Department of Forest Resource Management of the Swedish University of Agricultural Sciences, where I was lucky to spend a few months during the study. Finally, I thank my friends and relatives, but especially my brother Tero and my parents Pirjo and Veikko for their support throughout life.

Joensuu, April 2010

Jari Vauhkonen

## LIST OF ORIGINAL ARTICLES

This thesis is based on the following articles which will be referred to by their Roman numerals:

- I Vauhkonen, J., Tokola, T., Packalén, P. & Maltamo, M. 2009. Identification of Scandinavian commercial species of individual trees from airborne laser scanning data using alpha shape metrics. *Forest Science* 55(1): 37–47.  
<http://www.ingentaconnect.com/content/saf/fs/2009/00000055/00000001/art00004>
  
- II Vauhkonen, J., Tokola, T., Maltamo, M. & Packalén, P. 2008. Effects of pulse density on predicting characteristics of individual trees of Scandinavian commercial species using alpha shape metrics based on airborne laser scanning data. *Canadian Journal of Remote Sensing* 34(Suppl. 2): S441–S459.  
<http://pubservices.nrc-cnrc.ca/rp-ps/absres.jsp?jcode=cjrs&ftl=m08-052>
  
- III 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.  
[doi: 10.1016/j.rse.2010.01.016](https://doi.org/10.1016/j.rse.2010.01.016)
  
- IV Vauhkonen, J., Korpela, I., Pitkänen, J. & Olofsson, K. 2010. Estimating plot-level forest attributes per species by single-tree imputation using airborne laser scanning-based height, intensity, and alpha shape metrics. Manuscript.
  
- V Vauhkonen, J. 2010. Estimating crown base height for Scots pine by means of the 3-D geometry of airborne laser scanning data. *International Journal of Remote Sensing* 31(5): 1213–1226.  
[doi: 10.1080/01431160903380615](https://doi.org/10.1080/01431160903380615)

The articles I–III and V are reprinted with the kind permission of the publishers, while the study IV is the author version of the submitted manuscript.

Mr. Jari Vauhkonen was the main author and mainly responsible for all calculations and analyses, except for the stages involving aerial images in III–IV, and tree detection and delineation in V. The research ideas in I and II were developed jointly by the authors of these articles, whereas III–V were based solely on ideas by Mr. Vauhkonen. The co-authors contributed to various stages of the analyses and writing the articles, thereby improving the final quality of the papers.

## TABLE OF CONTENTS

|   |    |
|---|----|
| ABSTRACT .....  | 3  |
| ACKNOWLEDGEMENTS .....  | 4  |
| LIST OF ORIGINAL ARTICLES .....   | 5  |
| LIST OF ABBREVIATIONS .....   | 7  |
| 1 INTRODUCTION .....  | 9  |
| 1.1 Background .....  | 9  |
| 1.2 Tree-level inventory using airborne data .....                                  | 9  |
| 1.2.1 An overview .....   | 9  |
| 1.2.2 Tree detection and delineation .....  | 10 |
| 1.2.3 Feature extraction .....  | 11 |
| 1.2.4 Tree species recognition .....  | 12 |
| 1.2.5 Estimation of stem attributes .....   | 13 |
| 1.2.6 Validation of single-tree inventories .....                                   | 14 |
| 1.3 Objectives for the present work .....   | 15 |
| 2 MATERIALS AND METHODS .....   | 16 |
| 2.1 Study areas and data .....  | 16 |
| 2.2 Extraction of the per-tree ALS data .....                                       | 17 |
| 2.3 Estimation of tree-level attributes .....                                       | 18 |
| 2.3.1 An overview .....   | 18 |
| 2.3.2 Independent variables in the estimation .....                                 | 19 |
| 2.3.3 Estimation of the species and stem dimensions .....                           | 21 |
| 2.3.4 Variable reduction .....  | 21 |
| 2.3.5 Estimation of CBH .....   | 22 |
| 2.4 Simulation experiments .....  | 22 |
| 2.4.1 Effects of pulse density .....  | 22 |
| 2.4.2 Amount of reference data in NN imputation .....                               | 23 |
| 2.5 Estimation of plot-level attributes .....                                       | 23 |
| 2.6 Evaluation criteria and performance measures .....                              | 24 |
| 3 RESULTS .....   | 24 |
| 3.1 Tree-level assessment .....   | 24 |
| 3.1.1 A summary of the obtained accuracies .....                                    | 24 |
| 3.1.2 The properties and importance of the developed predictor variables .....      | 25 |
| 3.1.3 Effects of pulse density in the parametric prediction of species and DBH .... | 27 |
| 3.1.4 NN imputation of species and stem attributes .....                            | 27 |
| 3.1.5 CBH estimation .....  | 29 |
| 3.2 Area-level assessment .....   | 29 |
| 4 DISCUSSION .....  | 32 |
| REFERENCES .....  | 37 |

## LIST OF ABBREVIATIONS

|          |                                |
|----------|--------------------------------|
| 2-D; 3-D | 2-dimensional; 3-dimensional   |
| AIC      | Akaike information criteria    |
| ALS      | Airborne laser scanning        |
| CHM      | Canopy height model            |
| CBH      | Crown base height              |
| DBH      | Diameter at breast height      |
| LDA      | Linear discriminant analysis   |
| LRA      | Linear regression analysis     |
| MSN      | Most similar neighbor          |
| NN       | Nearest neighbor               |
| RF       | Random forest                  |
| RMSE     | Root mean squared error        |
| TIN      | Triangulated irregular network |



# 1 INTRODUCTION

## 1.1 Background

Different forest information systems require inventory data in varying resolutions. In Finland, for example, there are two operative inventory systems: national forest inventory for forest statistics and large-area planning, and stand-wise inventory for detailed forest management planning, but there also are interests towards highly specific inventories, such as pre-harvest measurements for timber procurement planning (e.g. Uusitalo 1995). Forest planning systems typically function at the level of single trees (e.g. Lämås and Eriksson 2003), and applications such as growth projections and simulated bucking would gain from a detailed description of stem dimensions and quality attributes, information that has traditionally not been collected at a required level of precision due to inefficient and laborious measurements involved (Uusitalo 1995). Since high-resolution remote sensing data allows tree-scale analysis (see e.g. Brandtberg and Warner 2006 for a review), remote sensing constitutes an interesting alternative for providing this information.

In particular, airborne laser scanning (ALS) has recently become an important technique for tree data acquisition. Due to its ability to measure three-dimensional (3-D) information, ALS data are usually regarded as having a greater potential for characterizing the canopy structure than other remote sensing materials (Koukoulas and Blackburn 2005; Magnusson 2006; Maltamo et al. 2006b; Uutera et al. 2006). ALS is starting to have an important role in practical forest inventories especially in Scandinavia, where Norway already has a tradition in ALS-based inventories since 2002 (Næsset et al. 2004). In Finland, an inventory system based on a combination of ALS data, aerial imagery and field sample plots is expected to be phased in during 2010–12 to replace the old field inventory for providing the data for management planning of private forests (Metsäkeskus 2009).

Most forestry applications of ALS are carried out as area-based estimation (Næsset 2002; Packalén 2009), although an alternative is to produce the attributes directly for single trees. Such an approach requires data in a high resolution, which currently entails higher data acquisition costs relative to area-based data. Obviously more interest will be shown towards single-tree methods also in practical forest inventories, however, since the data with a higher point density is expected to become more commonly available in the near future (Hyyppä et al. 2008a). Single-tree inventories carried out from the air inherently miss a portion of the smallest trees (e.g. Persson et al. 2002), which is a drawback, but the trees that are detected are highly representative of the dominant tree layer. However, prominent bias can also originate from inaccuracies in both species recognition and allometric, indirect estimation of the attributes of the detected trees (Korpela 2004; Korpela and Tokola 2006; Maltamo et al. 2007).

## 1.2 Tree-level inventory using airborne data

### 1.2.1 An overview

Single-tree remote sensing typically requires a ground resolution of at least 0.5 m (e.g. Lévesque and King 2003), somewhat depending on the tree size. The 0.6–1 m resolution of Ikonos and QuickBird satellite data has been found equally sufficient for tree delineation (e.g. Hirata et al. 2009), but airborne data is usually preferred due to better availability,

lower price, and the potential to obtain higher spatial resolution (Brandtberg and Warner 2006). Both spaceborne and airborne data are available in a digital format, which facilitates their automatic processing.

Tree-level interpretation of ALS data was initially proposed by Hyypä and Inkinen (1999) and Brandtberg (1999), later having become a popular research topic (see Hyypä et al. 2008a). Although three-dimensional (3-D) information can also be obtained from aerial images using photogrammetric techniques (Korpela 2004), the strength of ALS is its ability to directly reconstruct the target into a reliable 3-D point cloud. However, ALS data are based on a single laser wavelength band, while aerial photography has several bands sensitive to reflectance characteristics of different vegetation. In this sense, images have been favoured for tree species recognition (e.g. Holmgren et al. 2008b).

The interpretation of aerial images is, however, hampered by different spectral distortions caused by light fall-off effects and variations in atmosphere and view-illumination geometry (Lillesand et al. 2004). Aerial surveys of large areas must often be carried out under differing photographic conditions, which cause varying radiometric properties between the images and make their automatic interpretation more difficult (Mäkinen et al. 2006). The use of spectral images also complicates the inventory system and includes difficulties from the operational point of view (Packalén 2009), so that basing the inventory on ALS data alone forms a tempting alternative. Within this thesis, the estimation was based only on data acquired by small-footprint, discrete-return ALS systems (cf. Næsset et al. 2004).

Independent of the data source, tree-level inventory constitutes a chain of events, in which at least tree detection, feature extraction and estimation of tree attributes need to be considered (Talts 1977; Holmgren 2003; Korpela and Tokola 2006; Hirschmugl 2008). In Finland, practically any application requires timber estimates per species, so that species recognition is to be included in any case. The following presents the state-of-the-art in ALS-based single-tree inventory applicable to Scandinavian stand structure conditions, avoiding details, however, since there are several reviews and textbook chapters recently written on the topic (Hyypä et al. 2008a, b; Koch et al. 2008; Packalén et al. 2008a).

### *1.2.2 Tree detection and delineation*

In order to reduce the computational burden in processing mass points, the trees are usually detected from a 2.5-dimensional canopy height model (CHM) interpolated from the height data (Hyypä and Inkinen 1999; Persson et al. 2002; and many others). The cell values in the CHM represent the height difference between the top of the vegetation and the ground level, i.e. the canopy height, and local height maxima can be interpreted as tree top positions. Furthermore, tree height can be estimated as the values of these maxima, but other measurements require the tree crowns to be delineated from their surroundings. Mainly image analysis techniques are used also for that purpose, but the segmentation can equally be done by point-based techniques (e.g. Morsdorf et al. 2004; Wang et al. 2008).

An important aspect is that in most cases not all trees can be detected. Korpela (2004) analyzed the discernibility of trees in varying species and development classes by visually interpreting colour-infrared images with multiple views on the targets. The trees with heights of less than 40–60% relative to the dominant height were most probably missed, this proportion being dependent on forest structure and density. Most of the dominant trees, and thus 88–100% of the total volume could still be detected from the images. ALS-based studies have led to similar conclusions, as Persson et al. (2002), for example, detected 71%

of the stems, but 91% of their volume as measured in the field. Pitkänen et al. (2004), on the other hand, performed tree detection in a more heterogeneous forest, reporting a 40% detection rate for all trees, but that of 70% for the dominant trees.

Considering automatic interpretation, the algorithm has a major effect on the tree detection result (Karttinen and Hyypä 2008), which is often affected by the parameterization of the method (e.g. Solberg et al. 2006). In addition to omission errors caused by the undetected trees, also commission errors, i.e. segmentation of objects that are not trees, can occur. Solberg et al. (2006), for example, reported a 26% commission error rate in an inventory that found 66% of the field-measured trees. In this sense the conifers are less problematic than the deciduous trees, which often have multiple crowns of irregular shapes (Brandtberg et al. 2003; Koch et al. 2006).

As the area-level estimates are aggregated from single trees, their precision is a function of the errors in the tree detection phase. Two types of solutions for taking the tree detection errors into account have been presented. First, statistical approaches can be used for estimating the proportion of the undetected trees, and the tree detection result is then added to an estimate for those (Maltamo et al. 2004; Mehtätalo 2006; Flewelling 2008). Second, the estimation procedures can be modified to provide segments with a summation of field reference attributes rather than treating them as single trees (Lindberg et al. 2010; Breidenbach et al. 2010). Both of these approaches reduce the bias at the area-level, the latter being potentially able to also take the commission errors into account.

### *1.2.3 Feature extraction*

In order to perform the desired estimation task, the relevant information, i.e. geometric and radiometric properties with explanatory power for the tree attributes of interest, need to be extracted from the input data. The further estimation (section 1.2.5) combines direct measurements, species-specific properties that can be reconstructed from the data, and tree allometry, i.e. knowledge on dimensional relationships between plant parts.

Analogous to photogrammetric single-tree inventory (Talts 1977), tree height and different variables related to crown projection area (usually maximum crown width) have been the most common observations obtained from ALS data. Highly precise but underestimated tree height measurements are generally reported (Hyypä et al. 2008a). Crown width, on the other hand, is more difficult to determine (Persson et al. 2002; Popescu et al. 2003), since the result depends on the forest density and structure, and also on the tree delineation algorithm. For example, Persson et al. (2002) reported correlations of 0.99 and 0.76 for height and crown diameter, while the root mean square error (RMSE) was about 0.6 m for both. Pyysalo (2006), examining crown dimensions derived from ALS-based vector models, also reported underestimation of both vertical and horizontal dimensions, when the models were validated against side-view images of altogether 49 trees. According to Karttinen and Hyypä (2008), the applied pulse density has a minor effect on tree height estimation, but it can affect the crown delineation accuracy more severely (Pyysalo 2006; Goodwin et al. 2006).

In addition to tree height and crown 2-D characteristics, other geometric measurements and variables derived from the height and intensity values of the backscattered pulses can be used (e.g. Holmgren and Persson 2004). Return intensity value provides a measure of the amount of energy reflected from a target, the circumstances affecting these reflections from forest canopy being further discussed by Brandtberg (2007). Particularly, the intensity observations are affected by leaf size, orientation and foliage density (Korpela et al. 2010),

so that the intensity is not solely related to the reflectance properties of the vegetation (see also Moffiet et al. 2005). Height and intensity values form distributions, however, which are sources for further information.

Crown base height (CBH), on the other hand, is an attribute obtainable from ALS data that can also be verified against field observations. There have been active efforts to derive CBH from ALS data, since related field measurements are very time consuming. ALS-based approaches include analyzing structural properties of ALS point clouds (Pyysalo and Hyypä 2002; Holmgren and Persson 2004; Holmgren et al. 2008b; Popescu and Zhao 2008), direct analysis of the ALS height distribution (Morsdorf et al. 2004; Solberg et al. 2006), and regression analysis based on ALS variables (Maltamo et al. 2006a; Popescu and Zhao 2008; Maltamo et al. 2009b). The accuracy of estimating this attribute is not considered as high as parameters extracted from the upper crown. Usually an overestimation is reported, and a best-case RMSE of about 2 m (17%) was achieved by Popescu and Zhao (2008) by local regression models.

Considering the relatively short history of ALS-based single-tree measurements, the work that has been done in feature extraction appears insufficient. Only tree height and crown 2-D dimensions, which are directly obtainable from the segmented CHM, for example, are commonly used, even though ALS allows numerous variables to be extracted in addition to these. Studies on tree allometry (e.g. Mäkelä and Vanninen 2001; Kantola and Mäkelä 2006; Ilomäki et al. 2006) report a strong relationship between foliage mass and stem attributes, encouraging to develop variables quantifying the amount and allocation of foliage. On the other hand, by increasing the pulse density, also structural differences between coniferous and deciduous vegetation could possibly be pointed out.

#### 1.2.4 Tree species recognition

In Finland, remote sensing-based studies (e.g. Packalén 2009) attempt to separate commercial species groups of Scots pine (*Pinus sylvestris* L.), Norway spruce (*Picea abies* [L.] H. Karst.) and deciduous trees, the two conifers constituting more than 80% of the growing stock (Korhonen et al. 2006). The latter group consists of mainly birches (*Betula* spp. L.), but minor species such as aspen (*Populus tremula* L.), alders (*Alnus* spp. P. Mill.), willows (*Salix* spp. L.), and rowan (*Sorbus aucuparia* L.) are usually included in this group. High species recognition accuracy is crucial when the estimation is based on species-specific allometric dependencies. According to the simulations by Korpela and Tokola (2006), the entire estimation chain resulted in RMSEs of 30% and about 15% with species recognition accuracies of 75% and 80–90%, respectively, for the total volume of the sample stand. Considering ALS-based interpretation, Holmgren and Persson (2004) classified Scots pine and Norway spruce by their structural differences with >90% accuracy, later suggesting a similar accuracy to be obtained for the three species groups by including spectral mean values determined from aerial photographs (Holmgren et al. 2008b). The recent studies have, however, focused on deriving the species information solely from ALS data.

In Holmgren et al. (2008b), the strongest ALS-based predictors were a quantification of crown shape, obtained by the parameters of a parabolic model fitted to the ALS data, statistical measures derived from the proportions of first returns and the mean of intensity values. The distributions of intensity values were analyzed by Ørka et al. (2009) and Korpela et al. (2009b), the former reporting 88% accuracy of distinguishing dominant spruce and birch trees in Norway. Korpela et al. (2009b) examined more than 13 000 trees

in southern Finland, reporting accuracies of 81–85% of classifying pine, spruce and birch, and that of 91–93% for the conifer trees. Their later study (Korpela et al. 2010), however, indicates that even higher classification accuracies can be obtained using intensity variables normalized with reference to the scanning range and receiver gain settings. Certain deciduous species have been found deviant in terms of the backscatter properties (Säynäjoki et al. 2008; Korpela et al. 2009b; Kim et al. 2009).

Also leaf-off ALS data has been found useful in separating coniferous and deciduous vegetation (Brandtberg et al. 2003; Liang et al. 2007; Kim et al. 2009), a task in which Liang et al. (2007) obtained 89% accuracy in southern Finland by using the height differences between first and last returns within the tree crowns. Kim et al. (2009) examined multi-temporal data, reporting 83% and 73% accuracies in coniferous-deciduous classification using intensity variables derived from leaf-off and leaf-on data, respectively, the best result (91%) being obtained using their combination. This analysis was carried out in the temperate forest zone in southern U.S.A., but they also examined the discrimination between evergreen coniferous and broadleaved deciduous trees, i.e. species composition close to that of Scandinavia, in which case the previous accuracies were 97%, 63%, and 99%, respectively.

### 1.2.5 Estimation of stem attributes

A measurement and estimation chain that links photogrammetric single-tree measurements with allometric estimation of diameter at breast height (DBH) has motivated several studies in Scandinavia (Ilvessalo 1950; Jakobsons 1970; Talts 1977; Kalliovirta and Tokola 2005; Korpela and Tokola 2006; Maltamo et al. 2007). In Finland, Kalliovirta and Tokola (2005), for example, formulated national and regional species-specific models that used tree height and maximum crown width for predicting DBH. It is known, however, that various factors such as stand density and silvicultural history can affect the relationships between tree height, crown width and DBH (Korpela 2004; Maltamo et al. 2007; Kaitaniemi and Lintunen 2008). The accuracy of estimating DBH is restricted by the imprecision of the allometric relationships between measurable tree dimensions and the attributes of interest, being 10% in terms of RMSE in Finland (Korpela and Tokola 2006).

Stem total volumes and timber assortment volumes are commonly predicted by using DBH and height estimates based on airborne data in species-specific stem taper models (e.g. those by Laasasenaho 1982). The errors in the DBH estimates are compounded, however, when applied to stem taper models, which themselves also include inaccuracies. Maltamo et al. (2007), for example, simulated the accuracy of a single-tree inventory of 472 sample plots by predicting DBH from tree height, on the assumption that all the trees had been detected and both the tree height and species estimates were error-free. Despite the simplifying assumptions that could hardly be justified in a real-world application (cf. Korpela and Tokola, 2006), the simulated RMSE for the stem volume was about 23% at plot level, indicating a need for either additional predictors or an entirely novel estimation approach.

Takahashi et al. (2005) and Villikka et al. (2007), for example, used percentile variables based on the tree-level distribution of ALS height values for predicting the stem volume of sugi (*Cryptomeria japonica* D. Don.) and Norway spruce, respectively. Chen et al. (2007) introduced “canopy geometric volume”, defined as the area of a tree segment multiplied by its height (see also Nelson 1984; Hollaus 2006), to estimate tree-level basal area and biomass. All of these authors concluded that the ability to use additional variables will

improve the estimates for the attributes of interest relative to models based on tree height and crown diameter or area. The increased number of possible predictors requires caution in the estimation phase, however, as collinearity between the variables may cause a parametric model to be unstable. Also, normality and homoscedasticity assumptions need to be met in the case of linear regression models.

Recently, different non-parametric methods have been applied to producing tree attributes per species either by predicting theoretical diameter distributions (Packalén and Maltamo 2008; Peuhkurinen et al. 2008) or by estimating the attributes directly at the level of single trees (Maltamo et al. 2009b; Breidenbach et al. 2010). These studies have particularly focused on nearest neighbor (NN) search and imputation methods (e.g. Eskelson et al. 2009). As such approaches require no prior knowledge of the distribution of the data, their use may be highly relevant when non-linear and possibly diverse relationships exist between the independent and dependent variables. The cost is the need for *in situ* reference data, which can be largely avoided in the parametric estimation chain, although a local calibration will improve the accuracies (e.g. Kalliovirta and Tokola 2005).

The use of imputation methods places very high requirements on the extent of the reference data, however, as these should be representative of the entire phenomenon of interest. This means that variable imputation may seem problematic, especially at the level of single trees. Maltamo et al. (2009b) nevertheless used the *k*-Most Similar Neighbor (*k*-MSN) method (Moeur and Stage 1995) for predicting tree-level characteristics from a reference data set comprising only 133 trees. They found the *k*-MSN estimates to be generally more accurate than parametric sets of models constructed simultaneously by Seemingly Unrelated Regression, with tree-level RMSEs of 5%, 2%, and 11% for DBH, tree height and stem volume, respectively, in cross-validated reference data. The result was based on a local data set, however, and species identification was ignored, as the data applied to Scots pine only.

### 1.2.6 Validation of single-tree inventories

Considering species-specific estimation using single-tree methods, there appear to be only two studies reporting plot-level accuracies in Scandinavia (Korpela et al. 2007a; Breidenbach et al. 2010). First, Korpela et al. (2007a) tested allometric estimation for producing species-specific timber estimates. They used a semi-automatic method employing ALS data and aerial images for treetop positioning, height and crown width estimation, and species recognition, and used these observations to estimate stem dimensions with a species-specific allometric modeling chain (Kalliovirta and Tokola 2005). They reported a notable underestimation of 19% in the total volume, of which about 10% was accounted for omission errors and the rest for systematic errors in the estimation of DBH, the latter due to inaccuracy in the crown width measurements and the imprecision of the allometric models. Breidenbach et al. (2010), on the other hand, proposed a “semi-individual” tree detection method, in which the automatically produced crown segments were imputed by field attributes from segments considered to be nearest neighbors in terms of ALS and image features. This approach resulted in unbiased plot-level volume estimates with an RMSE of 17% of the total volume, for example, when evaluated by a cross validation procedure.

Ignoring species recognition, two otherwise interesting area-level aggregation results have been reported in Finland. First, Peuhkurinen et al. (2007) reported accurate DBH distributions to be obtainable for mature stands by single-tree interpretation of ALS data

and allometric DBH prediction, yet this result was validated on two pure spruce stands only. Second, Packalén et al. (2008) found both single-tree detection and the area-based method to result in equal accuracies in total volume and mean height, when the estimation was carried out on 41 sample plots. These accuracies were not validated at the tree level, but since stem number was considerably more underestimated with the single-tree method, certain imprecision can be expected in the tree-level attributes.

Finally, it should be adequately emphasized that tree-level data can alternatively be produced by predicting a theoretical set of trees using area-based estimation (Packalén and Maltamo 2008; Peuhkurinen et al. 2008). Since the high-density data required for actual tree detection is more expensive, single-tree analysis should either considerably improve the obtained accuracies or produce information that cannot be obtained from lower resolution data. Hypothetically, more detailed information is obtainable from direct measurements of dominant trees, while results by Korpela et al. (2007a), for example, indicate a need to refine the tree-level estimation. On the other hand, when attempting to validate saw-wood recovery estimates based on low density ALS data and aerial photographs, Peuhkurinen et al. (2008) concluded that the tree quality attributes affecting stem bucking (e.g. Uusitalo et al. 2004) could not be estimated from the height-diameter distributions generated from area-based data. Branch height properties (lowest living and dead branch) have been found to be the most essential quality attributes with respect to Scots pine (Uusitalo 1995), the results of Maltamo et al. (2009b) indicating these to be predictable by single-tree point cloud properties.

### 1.3 Objectives for the present work

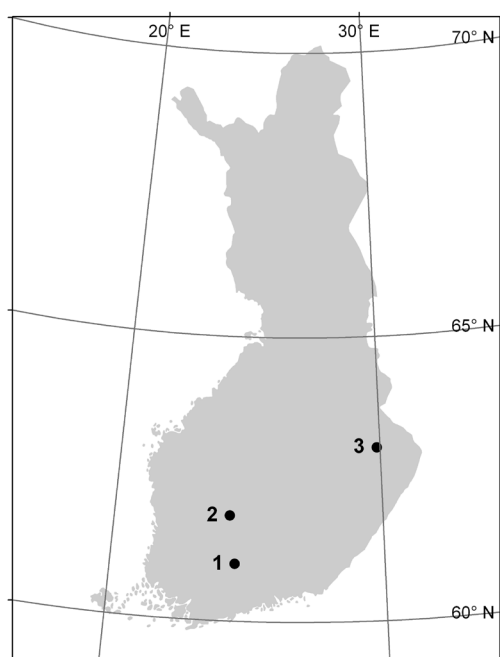
The aim of this work was to improve the estimation of single-tree attributes using ALS data. In particular, this work examined reconstruction of tree crowns by means of computational geometry of the point data and techniques for turning the obtained crown shape and structure information into improved estimates of species, stem dimensions, and CBH. The specific objectives for the studies reported in papers I–V were:

- I To develop 3-D structure-based features and examine species-specific differences in them relative to alternative ALS-based variables.
- II To test features corresponding to I in DBH prediction and to examine the effects of pulse density on the performance of these features in estimating both species and DBH.
- III To test nearest neighbor imputation in association with the features developed in I–II for the simultaneous estimation of tree species, DBH, height, and stem volume.
- IV To examine the accuracies of the techniques developed in III in an area-level timber inventory.
- V To develop adaptive methods for estimating CBH for Scots pine trees without a need for *in situ* reference data.

## 2 MATERIALS AND METHODS

### 2.1 Study areas and data

The experiments were carried out on three test sites in Finland (Figure 1). Harvoilanmäki data set was used in studies I and II, Hyytiälä in III and IV, and Koli in V. Tree species composition on each site consists of Scots pine, Norway spruce and to a lesser degree deciduous trees, mostly birch, but the Koli data was acquired from almost pure pine stands. The characteristics of the airborne data sets are described in Table 1.



**Figure 1.** Locations of the areas studied. 1 – Harvoilanmäki, 2 – Hyytiälä, and 3 – Koli.

**Table 1.** Main properties of the ALS data sets.

| Article(s)                     | I–II           | III–IV          | III – IV       | V               |
|--------------------------------|----------------|-----------------|----------------|-----------------|
| Instrument                     | TopEye MkII    | Optech ALTM3100 | Leica ALS50-II | Optech ALTM3100 |
| Acquisition date               | Sept. 19, 2004 | July 25, 2006   | July 4, 2007   | July 13, 2006   |
| Pulse density, m <sup>-2</sup> | 40             | 6–8             | 6–8            | 4               |
| Flying height, m               | 200            | 1000            | 930            | 900             |
| Footprint, cm                  | 40             | 25–28           | 17–18          | 24              |

The field measurements in the test sites were performed in 2007, 2007–2008 and 2006, respectively. In I–IV, the trees were mapped employing a photogrammetric-geodetic technique (Korpela et al. 2007b), in which the trees were first positioned on aerial images to serve as field control points for the positioning of the other targets by trilateration and/or triangulation. In V, the trees were positioned relative to GPS-positioned plot corners and projected onto the coordinate system of the ALS using the corner positions as reference points. The accuracy of positioning the corners was assessed to be approximately 1 m in the XY direction.

Except for study IV, only trees that were discernible in the images and/or visualized ALS data were included in the analyses (see section 2.2). The Hyytiälä data set (III–IV) consisted of three subsets of forest plots, a set of 59 circular, 0.04-ha plots, a set of 18 rectangular plots (0.08–0.24 ha, totaling 2.2 ha), and a set of four rectangular plots (0.27–1.00 ha, 2.43 ha). In III, the trees measured on the circular plots (N=1898) were used consistently as a reference data set throughout the study, while the rectangular plot data (N=1249) were used for validation. Study IV combined these for the reference data, and data for the four large-area plots (referred to as “stands” in the further text) served as validation data. Further properties of the data are given in each study.

In studies I, II and V, only field measurements were used in validation, while in III and IV some field attributes were modeled. The best available height observation was computed for each tree, being the field measurement, the height obtained in the treetop positioning, or an estimate derived from plot-level regression curve. Stem volumes were calculated using DBH and height in species-specific equations (III) or stem taper models (IV), both by Laasasenaho (1982), and in IV the same models were used for assessing the theoretical quantities of timber assortments by simulating stem bucking into logs of saw wood and pulp wood. The bucking algorithm used rules for allowable log lengths and minimum diameters, attempting to maximize saw wood proportion.

## 2.2 Extraction of the per-tree ALS data

In I–III, manual or semi-manual methods were used to directly link the ALS points to a tree, while IV and V included automatic crown delineation methods. In I and II, isolated trees with no branches overlapping with other trees and no undergrowth, as verified by visual examination in 3-D, were manually recorded using TerraScan software. A data set of 92 trees (53 pines, 30 spruces and 9 deciduous trees), which represent dominant or co-dominant trees, was generated in this manner.

In III and IV, the extraction of ALS data and derivation of variables was incorporated into a crown modeling procedure (Korpela 2007) in which a three-parameter curve of revolution is fitted to the ALS points near the treetop. In the method, local, species-specific regression models that predicted the crown width from DBH and tree height were first applied to initialize the three parameters defining the shape and scale of the crown envelope. The initial crown width was overestimated by multiplying by 1.2, and the resulting model was iteratively fitted to the ALS point cloud using weighted, non-linear least squares adjustment. The length of the crown model was fixed, and the CBH was always 40% down from the top. ALS points inside the envelope or within one RMSE of it were saved for feature computations. Returns below the 40% height were stored inside a

cylinder having a diameter equal to the maximum crown width and the RMSE of the fit. Most suppressed and intermediate trees with relative heights of less than 60% were rejected by this procedure. Both 2006 (ALTM3100) and 2007 (ALS50-II) data were used in the collection of the tree point data, but only 2007 data were included in the later analysis.

In studies IV and V, tree detection was based on a raster CHM at a resolution of 0.5 m, generated in different ways for each study. In IV, an initial triangulated irregular network (TIN) model of the canopy surface was created by taking the maximum first return height value in each 0.5 m cell, while the final CHM pixels were produced by linear interpolation from the overlapping TIN triangles. In V, the CHM was filled by first taking the maximum height value within a radius of 0.5 m. The final result was produced interpolating the empty cells by taking the average from a 3×3 window, this being successively repeated until every cell had a height value.

In the tree detection method (IV–V), the CHM was first low-pass filtered using Gaussian kernels with the size of the smoothing window increasing as a stepwise function of the heights of the CHM (Pitkänen et al. 2004). The crown segments were created around local height maxima in the filtered CHM using watershed segmentation with a drainage direction following algorithm (Pitkänen 2005). The algorithm requires the determination of the kernel widths ( $\sigma$ ) and the height classes for which the  $\sigma$  are applied. These were selected by visually comparing the number of the resulting local maxima against the initial CHM. The ALS data in the segments were assigned to trees by certain linking criteria. In IV, the linking algorithm optimized a graph of possible links weighted by Euclidean distances between the treetop candidates and the trees measured in the field (Olofsson et al. 2008). In V, a crown segment was linked to a field-measured tree if 1) only one field tree intersected the segment and 2) the difference between the maximum height value within the segment and the reference height was less than 2 m. Altogether 687 segments were considered as automatically detected tree candidates, but according to the linking criteria, only 185 mainly dominant trees were linked to crown segments.

## **2.3 Estimation of tree-level attributes**

### *2.3.1 An overview*

The main focus was on developing alpha shape metrics, i.e. various measures related to crown volume, shape and structure, to be used in estimation of tree attributes summarized in Table 2. These metrics were used in combination with alternative variables, i.e. mainly those based on the height and intensity distributions of the point data. In I and II, species classification and DBH estimation were performed using parametric, linear functions, whereas III and IV used NN search and imputation methods for the simultaneous estimation of species and stem dimensions. The independent variables used in the estimation are summarized in section 2.3.2, the estimation methods in section 2.3.3, and variable reduction related to them in 2.3.4. CBH estimation (study V; section 2.3.5) was based on the analysis of point cloud properties, being therefore fundamentally different from the other methods.

**Table 2.** A summary of the statistical estimation methods used within the study. Sp – species, h – height, v – stem volume, vs – saw-wood volume, vp – pulp-wood volume, LDA – linear discriminant analysis, LRA – linear regression analysis, RF – Random Forest.

| Article(s) | Objective variable    | Imputation method | Number of predictors | Variable reduction |
|------------|-----------------------|-------------------|----------------------|--------------------|
| I–II       | Sp                    | LDA               | 423                  | Yes                |
| II         | DBH                   | LRA               | 424                  | Yes                |
| III        | Sp, DBH, h, v         | <i>k</i> -MSN     | 1846                 | Yes                |
| III        | Sp, DBH, h, v         | RF                | 1846                 | Yes/No             |
| IV         | Sp, DBH, h, v, vs, vp | RF                | 1846                 | No                 |

### 2.3.2 Independent variables in the estimation

The alpha shape metrics were derived from 3-D alpha shapes computed from the point data. An alpha shape (Edelsbrunner and Mücke 1994) is based on the Delaunay triangulation of a point cloud such that each simplex of the triangulation is compared with the specified alpha value in the computation phase. Those simplices, which have an empty circumsphere with a squared radius larger than the defined alpha value, are removed. Thus, an alpha shape can be regarded as an alpha-weighted Delaunay triangulation (see Figure 2). The resulting shape depends on the parameter alpha: with small values, the shape reverts to the input point set and is the convex hull of it with very large values. The alpha shapes can contain cavities and holes and have disconnected parts.

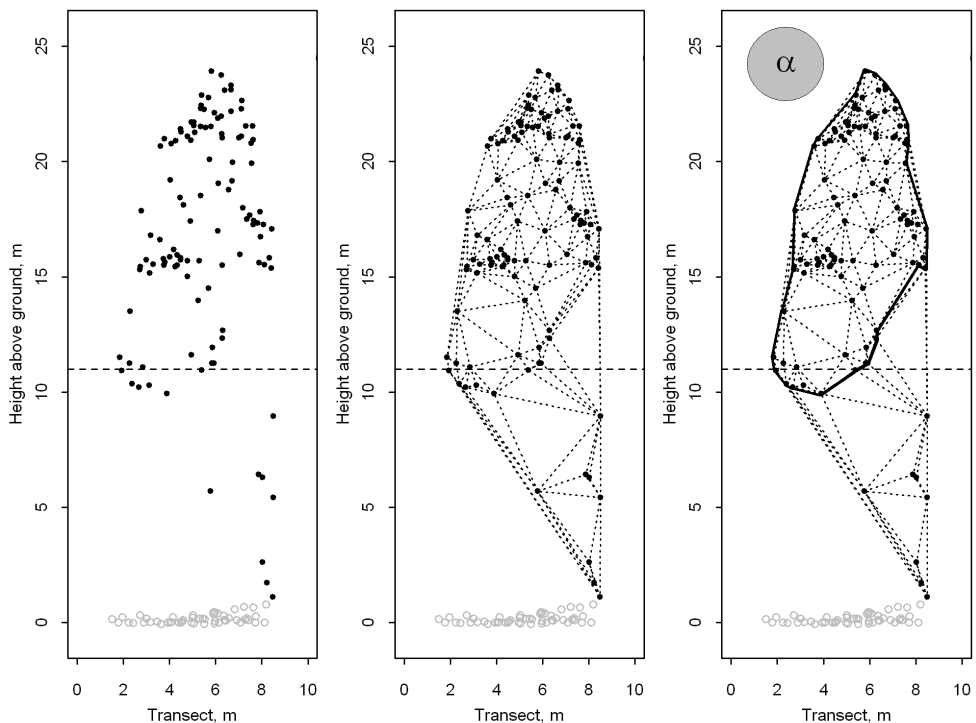
The 3-D variables used included volume and number of solid components, which indicates the number of separate components required to build the shape using the specified alpha value. The volume was computed with respect to interior and exterior of the alpha shape. The tetrahedra of the underlying Delaunay triangulation were classified as exterior when they did not belong to the alpha complex (i.e., to the boundary or interior of the alpha shape; see Edelsbrunner and Mücke 1994) and interior otherwise. These variables were calculated using different combinations of point data and alpha values in I–IV. The computations regarding the previous variables were carried out using the functionality of the open source library CGAL (Da and Yvinec 2007).

In addition to the 3-D variables, study II included a crown area estimate calculated as the 2-D convex hull of the point data. The crown profile analysis was further extended in III and IV by computing areas on different height levels. Studies III and IV also included estimates of crown height and length, calculated using a method described in study V.

From the height distribution variables, studies I–IV included percentiles and corresponding densities for 5, 10, 20, ..., 90 and 95% of the maximum height. Additionally, proportions of returns accumulated by these heights and basic descriptive variables were included in I–II and III–IV, respectively. The variables were calculated with respect to different echo categories, which were slightly different between I–II and III–IV. In addition to the tree-level variables, III and IV included the corresponding variables calculated at the plot level to describe the neighborhood of the trees.

Intensity variables were included in I–IV, but they were calculated in a different manner between I–II and III–IV, because of differences of processing the intensity values between the sensors. Intensity normalization (e.g. Höfle and Pfeifer 2007) was neither attempted, so that obtained intensity variables are sensor-specific. In I and II, the intensity variables were selected by an exploratory analysis of the species-specific differences in the obtained distributions. In III and IV, these were descriptive variables and percentiles, selected following Korpela et al. (2009b).

Studies I and II included texture analysis of a CHM at a resolution of 25 cm. In the analysis, the normalized gray-level co-occurrence matrix and features presented by Haralick et al. (1973) were tested. Here the CHM was generated by TIN interpolation and was used only for the texture analysis. Finally, statistical transformations, which included the natural logarithm and the square and cubic roots of the variables, were included in studies III and IV.



**Figure 2.** An example single-tree point cloud in the Koli data set (left) and the Delaunay triangulation based on it, illustrated in 2-D for ease of visualization. The right-hand figure shows the outer boundary of the highest connected component (solid line), determined using a predefined alpha value (filled circle). The field-measured CBH is illustrated using a dashed, horizontal line and ground hits using grey circles.

### 2.3.3 Estimation of the species and stem dimensions

The statistical estimation methods included linear discriminant analysis (LDA; e.g. Venables and Ripley 2002) for tree species classification (I–II), linear mixed-effects modeling (Searle 1971; Pinheiro and Bates 2000) for DBH prediction (II), and Most Similar Neighbor (MSN; Moeur and Stage 1995) and Random Forest (RF; Breiman 2001) methods applied to nearest neighbor search (Crookston and Finley 2008) for estimating all dependent variables simultaneously (III–IV). Both MSN and RF were tested in III, but only RF was used in IV.

In I and II, the prediction was obtained as a result of a linear function. In LDA, this function is based on discriminant scores created as linear combinations of the independent variables, attempting to maximally separate two or more classes. Mixed-effects modeling, on the other hand, basically extends linear regression analysis (LRA) with respect to taking into account the correlation structure in the data which consisted of two stages of sampling (sample plots, trees). In II, various transformations of the independent and dependent variables were tested to meet the normality and homoscedasticity assumptions of the linear modeling. In both I and II, separate functions were generated for the variable groups in order to find out the predictive power of each group.

In the NN methods (III–IV), the estimates for the attributes of interest are produced as weighted averages of the attributes of those reference observations that are similar in terms of a distance metric calculated in the predictor space formed by the independent variables. The MSNs are determined by distances computed in a projected canonical space (Moeur and Stage 1995), and  $k$ -MSNs (e.g. Maltamo et al. 2006b) are the  $k$  minima of those distances. RF, on the other hand, is basically a classification method, in which combinations of numerous classification trees are fitted from a random sample of reference data. The distance in the  $k$ -NN search is determined by “one minus the proportion of RF trees where a target observation is in the same terminal node as a reference observation” (Crookston and Finley 2008).

Studies I–III considered variable reduction (section 2.3.4) and formulated the models using the most essential predictors, but the ability of the RF algorithm to use all available variables (Breiman 2001) was also tested in III. In the case of NN methods, the user needs to decide either the size of the neighborhood, i.e. the value of the parameter  $k$ , or a maximum value for the distance metric (kernel methods). An increase in  $k$  will improve the precision of the imputation, but it will also shift the prediction towards the sample mean, thereby increasing the bias in the extreme values for the imputed variables (Eskelson et al. 2009). Study III tested values of  $k$  from 1 to 10. In IV, the estimation was carried out using RF with all available predictors and  $k=3$  on the grounds of the experience gained in III.

### 2.3.4 Variable reduction

Studies I and II used the accuracy ratio (Garczarek 2002) as the performance measure for adding individual variables to the discriminant functions. This ratio measures standardized Euclidean distances between scaled membership vectors and vectors representing the true class corners. In the selection, variables with the highest ratios were added to the models until the improvement in the performance measure was less than 1%. In II, the variables for the regression models were selected using the Akaike information criterion (AIC; Akaike 1974; Burnham and Anderson 1998; Venables and Ripley 2002). AIC measures the goodness of fit of a model, but includes a penalty for model complexity, the models giving the smallest AIC scores being the ones preferred.

In III, two variable reduction procedures based on internal importance measures applied to the RF algorithm were implemented, the purpose in both of them being to search for the best predictors by fitting RF separately to predict species and species-specific stem dimensions. As the first step, procedures adapted from Diaz-Uriarte (2009) and Hudak et al. (2008) were utilized, but instead of accepting the initial result, it was iterated 10 times, eventually retaining only the most frequent variables in the iterations. Finally, a sensitivity analysis was performed to find out effects the number of predictors had on the obtained results. In it, RF and  $k$ -MSN imputations were performed using predictors selected from the combined subset produced by the reduction strategies. Different numbers of predictors and groups with high and low inter-correlations were considered.

### 2.3.5 Estimation of CBH

Study V introduced two new methods for estimating tree-level CBH that employ the concepts of Delaunay triangulations and alpha shapes. The first method was based on detecting discontinuities in the 3-D triangulation in terms of large tetrahedral (cf. Figure 2). Two alternative methods were applied for classifying a tetrahedron as unacceptably large. In the first method, the highest 50% of returns were first triangulated and the volume of an average tetrahedron was used as this criterion. Second, a predefined alpha value was used for the same purpose. Efforts were made to link an alpha value with the tree size, but as the same result could be obtained using different alpha values, this was found troublesome. In the actual algorithm, the neighbors of the highest tetrahedron were traversed and if a tetrahedron was considered small by the given criterion, it was included in the 3-D structure modeling the tree crown. Its neighbors were similarly examined, this being repeated for as long as all connected cells meeting the given criterion had been traversed. The CBH was then defined as the height of the lowest vertex in the obtained structure.

The second method was based on extracting connected components from the lowest parts of an alpha shape generated with the full point data. An alpha value with one connected component was used as a starting point, and the alpha values were traversed in descending order until a new component was split or the minimum height value of the highest component was changed. The first split component was allowed to partly overlap the previous, but otherwise the removal was accepted only if the component was located below the current highest component. If not, the procedure was stopped and the CBH defined as in the previous paragraph.

The reference methods were based on analyzing the vertical profiles of the point clouds. The CBH estimation was based on analysis of return frequencies (Holmgren and Persson 2004; Solberg et al. 2006; Popescu and Zhao 2008), cross-sectional area (Holmgren et al. 2008b) and linear regression (Maltamo et al. 2006a; Popescu and Zhao 2008).

## 2.4 Simulation experiments

### 2.4.1 Effects of pulse density

Study II examined the effects of pulse density on the estimation of tree species and DBH by simulating thinning to the initial data of 40 pulses  $\text{m}^{-2}$ . The thinning procedure bears close resemblance to Magnusson et al. (2007). In it, altogether 15 thinning levels were defined by creating a corresponding number of grids with a systematically increasing cell size. For

each grid cell, the intersecting laser returns were removed except for a single randomly chosen one. Terrain elevation and, thus, the canopy height was estimated separately for each reduced data set, but the trees to be measured were not detected and delineated again from the thinned data. Instead, the returns belonging to each tree were identified by extracting the tree identifier that assigned each return to a certain tree from the full density data. The simulated data had  $12-0.5$  returns  $m^{-2}$  that had hit vegetation in the initial data. The performance of the models generated with the full density data was evaluated with the reduced data, these models were calibrated for each data set by estimating new coefficients, and completely new models were also constructed.

#### 2.4.2 Amount of reference data in NN imputation

Study III examined the sensitivity of the NN estimation to the amount of reference data by simulating thinned reference data sets at 50%, 25%, and 12.5% of the observations in the initial data set, generated by applying three selection strategies. The first corresponded to the manner of collecting reference data from randomly sampled field plots, in that entire plots were randomly selected until the required number of trees was obtained. In the second strategy, trees were selected randomly from the pooled tree set. In the third, it was assumed that the ALS data was acquired prior to the field-work, serving the role of an auxiliary information source for the selection of the reference data (cf. Hawbaker et al. 2009; Maltamo et al. 2009a). The trees were selected systematically from the initial reference data sorted by tree species and height, and within each species, the number of observations to be selected was determined by reference to the proportion of that species in the validation data.

## 2.5 Estimation of plot-level attributes

In study IV, the purpose was to test the aggregating of single-tree estimation (III) to area-level. The accuracies of total stem volume and timber assortments volumes, basal area and stem number were examined at levels of both stands and 10 m grid cells laid over these plots.

The data processing chain developed in this study will be referred to as *AutoLiDAR*. In it, tree crown segments were first delineated from ALS-based CHMs (see section 2.2). Second, these segments were produced with single-tree data using the RF imputation method tested in III. The reference data consisted of the two data sets in III, in which the point data were extracted by the crown modeling procedure (section 2.2).

For comparison, the corresponding estimates were produced using a semi-automatic, i.e. operator-assisted photogrammetric technique (*FotoLiDAR*) for mapping single-trees in images or in a combination of image and ALS data (Korpela et al. 2007a). It aims at treetop xyz positioning, height and crown width estimation, and species identification and converts these observations into DBH estimates using allometric models (Kalliovirta and Tokola 2005). Furthermore, stem taper curves (Laasasenaho 1982) are employed for the stem bucking and volume calculations. One difference relative to Korpela et al. (2007a) was that treetop xyz positioning was performed here using a faster monoplotted technique (Korpela et al. 2010).

## 2.6 Evaluation criteria and performance measures

The performance of the species classification was evaluated with the overall classification accuracy (%) and the kappa coefficient. In the case of all continuous variables, the accuracy measures were RMSE and bias:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n}}, \text{ and} \quad (1)$$

$$bias = \frac{\sum_{i=1}^n (x_i - \hat{x}_i)}{n}, \quad (2)$$

where  $n$  is the number of observations, and  $x_i$  and  $\hat{x}_i$  are the reference and estimated attributes, respectively, for the tree or grid cell  $i$ . The relative RMSEs were calculated by dividing the absolute RMSE values by the mean of the reference attribute.

In IV, tree detection was evaluated in terms of omission and commission error rates and by illustrating the area-level distribution of the estimated DBHs.

## 3 RESULTS

### 3.1 Tree-level assessment

#### 3.1.1 A summary of the obtained accuracies

The best-case accuracies obtained for tree attributes in I–V are given in Table 3. The accuracies of III are presented with respect to both leave-one-out cross-validation data and separate validation data. The cross-validation accuracies of species and DBH estimates were practically equal in I–III, but the accuracy considerably diminished in separate validation data. The main attention in Table 3 should therefore be focused on the accuracies obtained using separate validation data, i.e. studies III and IV.

When evaluated in separate validation data, species classification error of about 22% (accuracy of 78%) and RMSEs of 11%, 3% and 28% for DBH, height and stem volume, respectively, were reported in III. All tree attributes, especially the stem dimensions, were less accurate and included more bias, when they were produced using the *AutoLiDAR* method in IV. The *FotoLiDAR* method, on the other hand, produced better accuracies than the *AutoLiDAR* method, but also these were considerably lower than those obtained in III (Table 3). The accuracies of the individual studies are further examined in the following.

**Table 3.** Summary of the best-case tree-level accuracies obtained in I–V.  $N_{validation}$  column has the number of validation trees, while \*\* denotes cross-validated reference data. The errors are either classification error in overall accuracy or RMSE.

| Attribute                 | Study                     | Method                     | $N_{validation}$           | Error         | Bias          |            |
|---------------------------|---------------------------|----------------------------|----------------------------|---------------|---------------|------------|
| Species                   | I                         | LDA                        | 92**                       | 5 %           | -             |            |
|                           | III                       | RF <sub>1846</sub> , $k=1$ | 1898**                     | 6.70 %        | -             |            |
|                           |                           | RF <sub>1846</sub> , $k=1$ | 1176                       | 21.60 %       | -             |            |
|                           | IV                        | AutoLiDAR                  | 1495                       | 22.10 %       | -             |            |
|                           |                           | FotoLiDAR                  | 1467                       | 3.30 %        | -             |            |
| DBH (cm)                  | II                        | LRA                        | 53 pines**                 | 3.1 (7.5%)    | -0.1 (-0.3%)  |            |
|                           |                           | LRA                        | 30 spruces**               | 2.7 (9.7%)    | -0.3 (-0.5%)  |            |
|                           | III                       | RF <sub>130</sub> , $k=2$  | 1898**                     | 1.1 (6.4%)    | 0.1 (0.3%)    |            |
|                           |                           | RF <sub>130</sub> , $k=2$  | 1176                       | 2.0 (10.6%)   | 0.4 (1.9%)    |            |
|                           | IV                        | AutoLiDAR                  | 1495                       | 4.1 (18.6%)   | -1.1 (-5.0%)  |            |
|                           |                           | FotoLiDAR                  | 1467                       | 3.1 (13.8%)   | 1.0 (4.6%)    |            |
|                           | Height (m)                | III                        | RF <sub>1846</sub> , $k=4$ | 1898**        | 0.4 (2.5%)    | 0.1 (0.0%) |
|                           |                           |                            | RF <sub>1846</sub> , $k=4$ | 1176          | 0.4 (2.6%)    | 0.0 (0.0%) |
| IV                        |                           | AutoLiDAR                  | 1495                       | 1.4 (7.4%)    | 0.2 (1.0%)    |            |
|                           |                           | FotoLiDAR                  | 1467                       | 0.6 (3.2%)    | 0.5 (2.8%)    |            |
| Volume (dm <sup>3</sup> ) |                           | III                        | RF <sub>130</sub> , $k=3$  | 1898**        | 38.9 (17.4%)  | 2.3 (1.0%) |
|                           | RF <sub>130</sub> , $k=3$ |                            | 1176                       | 82.0 (27.5%)  | 14.1 (4.7%)   |            |
|                           | IV                        | AutoLiDAR                  | 1495                       | 152.1 (35.4%) | -41.5 (-9.7%) |            |
|                           |                           | FotoLiDAR                  | 1467                       | 126.8 (29.0%) | 50.1 (11.5%)  |            |
|                           | CBH (m)                   | V                          | LRA                        | 185**         | 1.44 (14.3%)  | 0.0 (0.0%) |

### 3.1.2 The properties and importance of the developed predictor variables

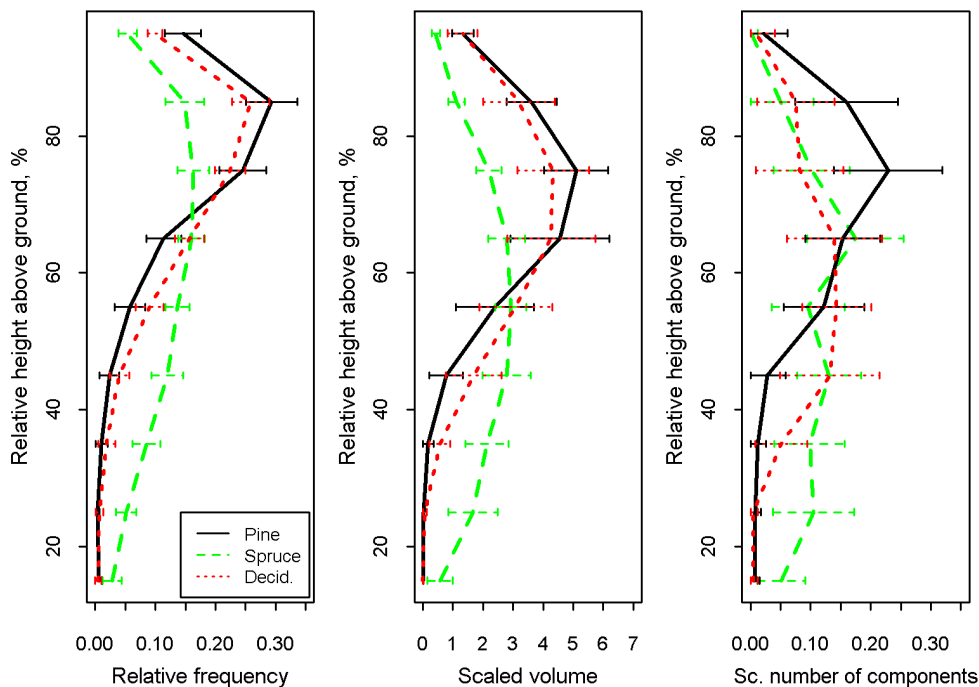
The species-specific differences in the developed variables were examined in study I. Figure 3 illustrates the crown profile obtained using either the developed volume and complexity metrics or variables based on the height value distribution. The profile based on the volume variables seems to differ slightly from the one based on the percentiles, when one is comparing pine with spruce, whereas the numbers of solid components are more distinctive than the height distribution-based profile with respect to pine with deciduous trees. However, the error levels were on a far lower level in the distribution-based profile (Figure 3).

The performance of individual variables in species classification was examined by quantifying them using kappa coefficients as performance measures. The highest kappas within the predictor groups were 0.72 for the predictor group of height distribution variables, 0.67 for crown volume variables, 0.59 for textural variables, and 0.38 for intensity variables. Plotting the most discriminative pairs of each group showed further

potential in separating coniferous species by height, texture and alpha shape metrics groups. The results corresponded to structural differences between these species as observed in the field. The intensity variables for deciduous trees differed slightly from those for the coniferous trees, but almost half of the deciduous trees were misclassified and no noticeable differences between the coniferous trees were found on the basis of these.

Height distribution variables, their combination with intensity variables, textural and intensity variables, alpha shape variables and a combination of these variable groups were further considered for species classification. Each discriminant function classified conifer trees fairly accurately (93–99%), so that the differences were obtained in the classification of the deciduous. A combination of the best variables from all the groups resulted in 95% of the trees in the study to be correctly classified with two deciduous trees misclassified as spruces. This discriminant function included two height distribution variables, three intensity distribution variables, and four alpha shape variables.

Study II formulated linear regression models from four different predictor groups, these being (1) tree height and crown area, (2) these and the height percentiles, (3) alpha shape metrics, and (4) a combination of these groups for DBH prediction. The models included 1–3 variables, which were alpha shape metrics except in the case of spruce, where one of the three variables was crown area. The best-case RMSEs for DBH were less than 10% (Table 3), and the differences in the performance of the model groups were minor, up to 4 percentage units for spruce.



**Figure 3.** Crown profiles of the tree species as described in terms of cumulative return frequencies (left), alpha shape volumes (middle), and numbers of solid components of alpha shapes (right). Error bars represent halves of the standard deviation values.

Study III involved variable selection, which also gives an impression of the importance of the variable groups in predicting the field attributes. Either 130 or 24 of the initial 1846 variables were preserved using the developed variable reduction strategies. Among the larger set, crown volume variables were most often involved (31 separate variables), followed by height distribution variables (9), intensity distribution variables (9), crown area variables (7) and one crown complexity and one crown length variable (Table 6). The other reduction procedure gave 4 crown volume variables, 3 height distribution variables and an intensity variable. In most cases, several statistical transformations of a predictor variable were included.

### *3.1.3 Effects of pulse density in the parametric prediction of species and DBH*

The effects of pulse density on the developed metrics were tested in study II. In the case of tree species, the performance of the models generated with the full-density data decreased rapidly as the pulse density was reduced. When new coefficients were estimated for these models, the decrease in the accuracy was less sharp, although there were slight deviations from the overall trend. Separate models generated for each thinning level maintained the accuracy rather well. All the methods used for predicting DBH, on the other hand, were less affected by the pulse density, and the accuracies could be virtually maintained until the lowest density levels by calibrating the model or constructing a new one.

The kappa coefficients measuring the accuracy of the species classification remained mostly above 0.4, and kappas of mostly around 0.8 were achieved with the density-specific models. The RMSEs obtained using density-specific models for DBH were up to two-fold relative to the initial accuracies. The performance reduction in estimating both species and DBH was usually most radical for the models based on alpha shape metrics only. Other variables were generally less sensitive to the pulse density, and the performance reduction was restricted by combining them with the alpha shape metrics.

### *3.1.4 NN imputation of species and stem attributes*

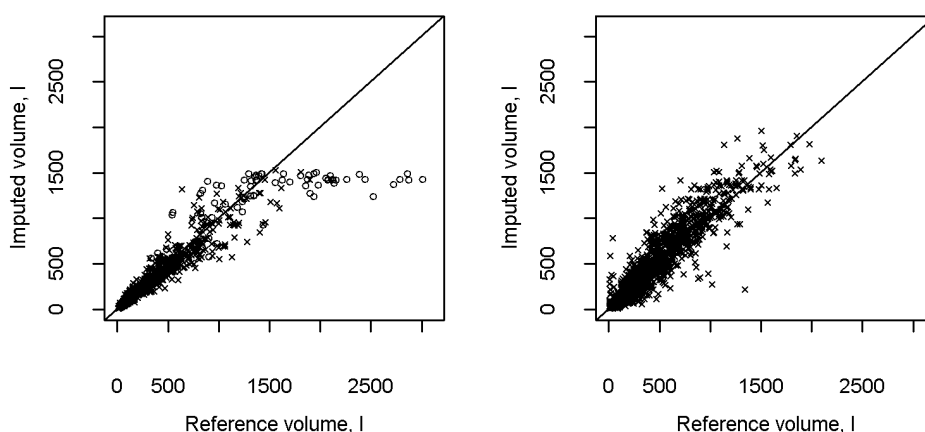
Study III used the ability of RF to employ all available predictors, but *k*-MSN and RF were compared only using the reduced sets of variables. The variable reduction was carried out using RF, so that the result cannot be considered optimal for the *k*-MSN method. However, the sensitivity analysis carried out in III indicated an in-optimality of about 2–4% only.

The best-case accuracies obtained in III were presented in Table 3, while Table 4 shows the differences between the imputation methods, when evaluated against separate validation data. Species classification accuracies of 70–79% and RMSEs of 30–36% were obtained for stem volume using *k*-MSN, whereas the corresponding figures for the RF method were 69–78% and 28–37%, respectively. Thus, *k*-MSN resulted in a slightly better accuracy with respect to predicting tree species, the model with 130 variables being the most accurate. The poorest *k*-MSN imputation was also slightly better than the poorest result obtained using RF imputation. On the other hand, RF produced both the best and the worst result in the estimation of stem volume. Rather than the method used for imputation, however, the number of predictor variables affected the results in the sense that better accuracies were mainly obtained by using a higher number of predictors, the difference being up to 10 percentage points.

**Table 4.** Ranges of the reliability characteristics of tree species (*Sp*) and stem volume (*v*) estimates when using different imputation alternatives. The value of *k* is given in parentheses, and the subscripts in the *Method* column indicate the number of predictor variables used.

| Method             | <i>Sp</i> accuracy, %  | <i>v</i> RMSE, %     | <i>v</i> bias, %   |
|--------------------|------------------------|----------------------|--------------------|
| RF <sub>1846</sub> | 74.2 (4) – 78.4 (1)    | 29.2 (5) – 30.8 (1)  | 5.3 (5) – 5.8 (10) |
| RF <sub>130</sub>  | 74.3 (2) – 75.7 (4)    | 27.5 (3) – 31.9 (1)  | 4.2 (1) – 5.9 (10) |
| RF <sub>24</sub>   | 68.5 (7) – 70.2 (1,3)  | 29.9 (8) – 36.8 (1)  | 1.8 (1) – 4.0 (10) |
| MSN <sub>130</sub> | 76.3 (1,2) – 78.9 (10) | 29.6 (10) – 33.9 (1) | 2.0 (1) – 4.5 (10) |
| MSN <sub>24</sub>  | 69.7 (1,2) – 72.8 (8)  | 31.4 (4) – 36.4 (1)  | 2.5 (1) – 4.0 (10) |

When evaluated by cross-validation in the reference data set itself, the estimates generally included errors of less than 10%, the RMSE for stem volume being about 17% (Table 3), but the accuracies were considerably lower in separate validation data. The figures in Tables 3 and 4 are presented in the validation data set consisting of those trees likely to have a similar observation within the reference data. The accuracies of DBH, tree height and stem volume were lower in other validation data sets, as the estimates saturated at the level of the largest reference observations (Figure 4). The accuracy of estimating tree species did not differ appreciably between the validation data sets. Tree species classification was usually successful in the case of the conifers, while less than 50% of the observed deciduous trees were correctly classified, being confused with both conifer species. Tree size had a minor effect on the success of tree species classification.



**Figure 4.** Reference vs. imputed stem volumes as obtained using the RF method with  $k=3$  in studies III (left) and IV. Circles in the left-hand figure represent imputations known to be extrapolated.

The accuracy of relating the ALS characteristics to the field attributes was affected by increasing stem density, causing inaccuracies in the derived characteristics due to interlaced tree crowns. The accuracy was related to plot-level basal area so that absolute inaccuracy in the imputed volume was higher for those trees that were located on plots with a basal area above 22 m<sup>2</sup>/ha.

When  $k=1$  was used, RF with all predictors was the best method in all cases except for DBH imputation, the reduced sets of variables generally resulting in higher accuracies when the value of  $k$  was increased (Table 4). Increasing  $k$  first sharply reduced the inaccuracy, which then stabilized and finally started to increase in some cases. The relative differences between the imputation methods remained approximately unaltered and the differences between the methods generally diminished with increasing values of  $k$ . The amount of reference data, on the other hand, had no clear effect on the imputation accuracy. Errors in species classification in particular remained at the level achieved with the full reference data.

In study IV, the accuracies of the imputations were considerably poorer (Table 3). The species recognition accuracy of *AutoLiDAR* was 78%, while the estimates for DBH, height and stem volume included overall RMSEs of 4.1 cm (19%), 1.4 m (7%) and 152 dm<sup>3</sup> (35%), respectively, with slight variations among the stands. The DBH and stem volume were overestimated by about 5% and 10%, respectively, whereas the height estimates included a positive bias of 1%. The scatter of these values shows more variation compared to estimates in III (Figure 4). The *FotoLiDAR* estimates were generally more accurate. Visual species recognition resulted in accuracies of 96–98%, while the RMSEs for DBH, height and stem volume were 3.1 cm (14%), 0.6 m (3%) and 127 dm<sup>3</sup> (29%), respectively (Table 3). As with *AutoLiDAR*, the DBH estimates included a 5% bias, but in the opposite direction. The bias in the stem volume estimates was somewhat larger.

### 3.1.5 CBH estimation

Study V reported correlations  $\geq 0.8$  between all estimates and the field-measured CBH, especially when the group of sawlog-sized trees alone was considered. With respect to the estimation accuracy, the developed methods resulted in RMSEs of approximately 3.5 and 3 m, when the estimation was carried out for all trees. The RMSEs produced by the reference methods were around 2 m, and combining one of them with the tree height estimate in linear regression resulted in the best accuracy with an RMSE 1.4 m (1.5 m for the trees of sawlog size). The estimation with fixed alpha value resulted in RMSE values comparable to the reference methods, but involved selecting the alpha parameter, which is dependent on the data density and also likely to be site specific.

The correlation coefficients as well as the estimation accuracies in terms of absolute RMSE values were usually slightly lower for the trees smaller than those of sawlog size. It should however be noted that the presence of the small trees was important for the modeling, since they provided information on the low CBH values, and thus improved the proportion of variability that was accounted for by the model.

## 3.2 Area-level assessment

More treetop candidates were linked to the field trees using the *AutoLiDAR* method than using the *FotoLiDAR* method, although the differences were small. Practically all the omission errors applied to trees shorter than 90% of the dominant height, as the probability of detection decreased with the height of the tree, and mainly the same trees were missed

using both the methods. The *AutoLiDAR* method produced considerably more commission errors (9–17% of the trees) than the semi-manual *FotoLiDAR* method ( $\leq 1\%$ ). Commission errors were slightly more common among the intermediate or co-dominant trees (60–90% of the dominant height), but no clear trend could be identified.

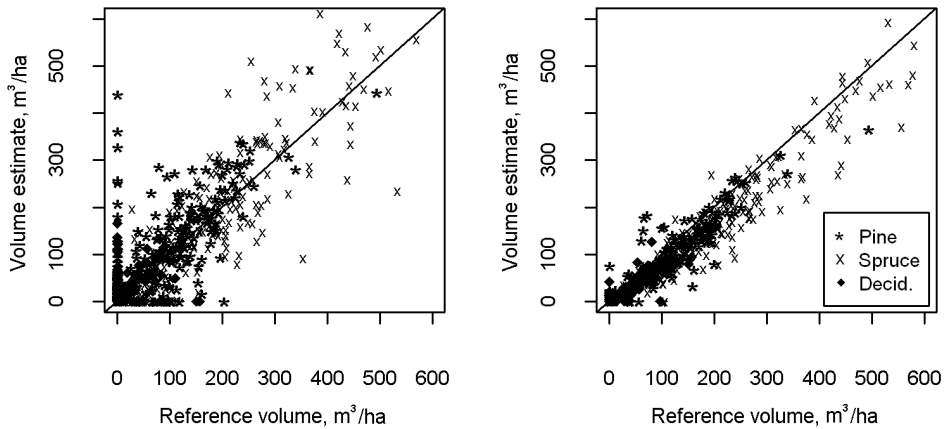
The accuracy of estimating the attributes in the 100 m<sup>2</sup> grid cells is presented by species and method in Table 5. The grid-level stem volume and basal area estimated by the *AutoLiDAR* method gained RMSEs of 25%, whereas the RMSE for the stem number was about 34%. The RMSEs for saw and pulp wood volumes were 35 and 27%, respectively. The basal area, total stem volume and saw wood volume were overestimated by 12–19%, whereas the stem number and pulp wood volume were underestimated by 6–9%. There were considerable stand-specific variations in predicting all these characteristics. The *FotoLiDAR* method resulted in better accuracy throughout, except in the case of stem number. The differences were minor, however, being 2.2 percentage points in stem volume and 0.8–5.5 percentage points in the other attributes. The *FotoLiDAR* estimates included more bias, all being underestimated by 17–23% except for pulp log volume, in which the bias was 7%.

**Table 5.** Accuracies of the forest attributes produced for the grid cells. V = stem volume, Vs = saw wood volume, Vp = pulp wood volume, G = basal area and N = number of stems.

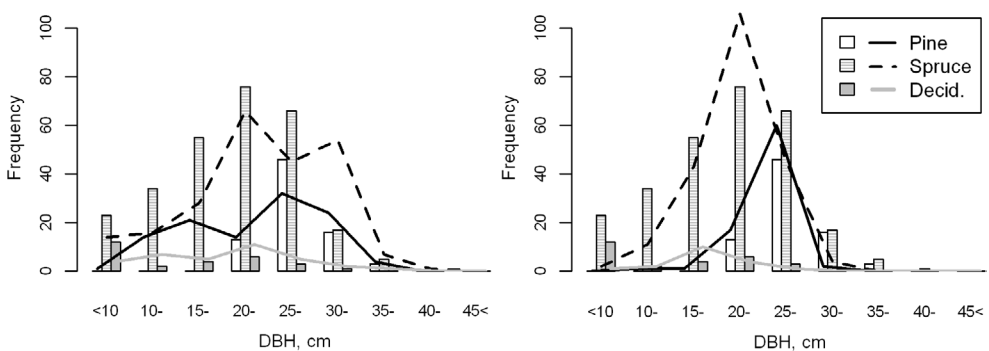
| Species | Attribute              | <i>AutoLiDAR</i> |                | <i>FotoLiDAR</i> |               |
|---------|------------------------|------------------|----------------|------------------|---------------|
|         |                        | RMSE             | Bias           | RMSE             | Bias          |
| Total   | V, m <sup>3</sup> /ha  | 67.9 (24.9%)     | -31.3 (-11.5%) | 61.9 (22.7%)     | 45.8 (16.8%)  |
|         | Vs, m <sup>3</sup> /ha | 67.6 (34.8%)     | -36.6 (-18.9%) | 56.8 (29.3%)     | 39.0 (20.1%)  |
|         | Vp, m <sup>3</sup> /ha | 20.3 (27.4%)     | 4.7 (6.4%)     | 17.1 (23.1%)     | 5.4 (7.2%)    |
|         | G, m <sup>2</sup> /ha  | 6.6 (24.8%)      | -3.3 (-12.4%)  | 6.4 (24.0%)      | 4.6 (17.3%)   |
|         | N / ha                 | 250.6 (33.8%)    | 67.5 (9.1%)    | 276.7 (37.3%)    | 168.7 (22.8%) |
| Spruce  | V, m <sup>3</sup> /ha  | 74.3 (40.9%)     | -3.1 (-1.7%)   | 50.3 (27.6%)     | 31.2 (17.1%)  |
|         | Vs, m <sup>3</sup> /ha | 66.9 (49.4%)     | -16.5 (-12.2%) | 43.3 (32.0%)     | 23.9 (17.6%)  |
|         | Vp, m <sup>3</sup> /ha | 24.6 (56.2%)     | 12.4 (28.3%)   | 14.6 (33.3%)     | 6.3 (14.4%)   |
|         | G, m <sup>2</sup> /ha  | 7.1 (42.3%)      | -0.3 (-1.5%)   | 5.0 (29.5%)      | 3.1 (18.4%)   |
|         | N / ha                 | 266.8 (58.9%)    | 131.7 (29.1%)  | 207.7 (45.8%)    | 123.5 (27.2%) |
| Pine    | V, m <sup>3</sup> /ha  | 77.0 (96.4%)     | -22.6 (-28.3%) | 31.0 (38.8%)     | 11.1 (13.9%)  |
|         | Vs, m <sup>3</sup> /ha | 69.6 (130.9%)    | -17.1 (-32.2%) | 31.5 (59.2%)     | 11.7 (21.9%)  |
|         | Vp, m <sup>3</sup> /ha | 20.7 (80.9%)     | -5.2 (-20.3%)  | 10.5 (41.2%)     | -0.8 (-3.0%)  |
|         | G, m <sup>2</sup> /ha  | 7.3 (85.4%)      | -2.4 (-28.5%)  | 3.1 (36.1%)      | 1.1 (12.6%)   |
|         | N / ha                 | 173.1 (76.9%)    | -54.3 (-24.1%) | 77.2 (34.3%)     | 17.7 (7.9%)   |
| Decid.  | V, m <sup>3</sup> /ha  | 31.8 (295.6%)    | -5.5 (-51.5%)  | 14.0 (130.2%)    | 3.6 (33.0%)   |
|         | Vs, m <sup>3</sup> /ha | 26.7 (478.4%)    | -3.0 (-54.2%)  | 13.9 (248.2%)    | 3.4 (61.7%)   |
|         | Vp, m <sup>3</sup> /ha | 9.6 (207.2%)     | -2.4 (-52.3%)  | 7.1 (153.2%)     | -0.2 (-3.8%)  |
|         | G, m <sup>2</sup> /ha  | 3.1 (252.5%)     | -0.6 (-50.0%)  | 1.5 (121.8%)     | 0.4 (34.8%)   |
|         | N / ha                 | 128.9 (203.5%)   | -9.9 (-15.6%)  | 106.0 (167.2%)   | 27.6 (43.5%)  |

The estimates for the species-specific attributes were less accurate than the totals (Table 5). The estimates for the attributes of the minor species were less accurate than those for the main species. The scatter of these values at the grid level is shown in Figure 5. The AutoLiDAR estimates include more variation and considerably more false zero-observations than the FotoLiDAR-estimates, whereas the latter display a clear trend underestimation.

Due to the bias in the tree-level estimates, species misinterpretation, and commission errors in tree detection, the estimated stand-level distributions of the *AutoLiDAR* method did not correspond very well with the reference (Figure 6). The predicted distributions emphasized the largest trees in the two mature stands, while the stem frequencies were clearly underestimated in the younger stands. The higher parts of the distributions fit well with the field reference in the *FotoLiDAR* method, but there are no observations of trees in the smallest and largest size classes.



**Figure 5.** Reference vs. estimated species-specific volumes per grid cells as obtained using the *AutoLiDAR* (left) and *FotoLiDAR* methods.



**Figure 6.** An example of observed and estimated DBH distributions as obtained using the *AutoLiDAR* (left) and *FotoLiDAR* methods. Bars – observed and lines – predicted distribution.

## 4 DISCUSSION

Recent advances in the use of ALS (e.g. Hyypä et al. 2008) have motivated attempts at the tree-level description of forest stands, which is of interest for applications to forest management and timber procurement planning, for example. The present study attempted to improve species recognition and allometric estimation of the stem attributes from single-tree ALS data. In particular, studies I–III developed variables and estimation techniques for using computational geometry in estimating these attributes. In IV, the accuracies of these were examined at the area level. Finally, V applied the corresponding techniques for estimating CBH, i.e. an important quality attribute for Scots pine timber.

Earlier, 90–96% cross-validation accuracies have been reported for species classification (Holmgren and Persson 2004; Holmgren et al. 2008b; Korpela et al. 2010). Korpela et al. (2007a) reported tree-level RMSEs of about 20%, 5%, and 46% for DBH, tree height and stem volume, respectively, following a similar estimation procedure to the *FotoLiDAR* method. On the other hand, Maltamo et al. (2009b) reported accuracies of 5%, 2%, and 11% for these stem dimensions in cross-validated reference data consisting of 133 pines. Popescu and Zhao (2008) obtained RMSEs of around 2 m for the CBH estimation with a very similar sample arrangement. Considering plot-level accuracies, Peuhkurinen et al. (2008), for example, reported best-case RMSEs of 22%, 78%, 65%, and 145% for total volume and volumes of spruce, pine and deciduous trees, respectively, in a Scots pine dominant area, using a slightly different validation plot design (254.5 m<sup>2</sup> circular plots) from the 100 m<sup>2</sup> grid cells used here. Considering the differences in the estimation approaches and the variation related to each field data set, the results obtained here (Tables 3 and 5) are well in line with the previous studies.

Alpha shape metrics, i.e. volume and complexity characteristics derived from the concept of 3-D alpha shapes (Edelsbrunner and Mücke 1994), were found to have potential for discriminating between tree species and describing the allometric differences in the trees, especially when computed from high-density data. The species-specific differences in the geometric arrangement of ALS returns could equally be quantified using either these volume characteristics, return frequencies at different relative heights, or textural features calculated from the CHM. The functions based on alpha shape metrics seemed to also provide information on deciduous trees. Both the height value distribution and the applied triangulations implicitly determine the crown profile, but since the triangulation-based approaches result in 3-D volume, those were found more strongly related to tree allometry.

Here an exploratory approach in which the alpha shape metrics were calculated with combinations of different data and alpha values was adopted. Considering the results, this technique proved efficient. This approach is also readily adaptable for further applications such as canopy reflection modeling (Rautiainen et al. 2008), estimating aboveground and component biomass (Popescu 2007) or mapping the defoliation of single trees (Solberg and Næsset 2007). For these purposes, however, it would be advantageous to develop an algorithm for an autonomous selection of the alpha value, leading to a single crown structure that could be validated against the field measurement (cf. Kato et al. 2009). An attempt to select an alpha value for generating a single 3-D structure would require defining the valid level of detail to be obtained (Zhu et al. 2008; Martynov 2008). Kato et al. (2009), on the other hand, reconstructed tree crowns by fitting a series of 2-D convex hulls to the outer point data. This “surface wrapping” technique produces 3-D crown structures without involving the selection of an alpha parameter. However, all of these attempts presume well-

defined targets, so that errors in the point data caused by undergrowth, for example, will contort the result.

In Maltamo et al. (2009b), the present author first estimated CBH using a 2-D technique also presented in V, and then computed the crown volume using a quasi-optimal alpha value selected so that the resulting alpha shape included the point data above the CBH within a single connected component (cf. Da and Yvinec 2007). This crown volume was included in the models for estimating various tree attributes, but its importance to the estimation was not separately assessed. The author's impression is, however, that the selection of the alpha value could be more integrally tied to the delineation of the crown point data. For example, the method that extracts connected alpha shape components from the crown base (V) is based on an alpha value search, the result of which could further be extended for deriving the 3-D crown silhouette.

Besides alpha shape and height distribution variables, study I employed textural features (Haralick et al. 1973), which were apparently extracted for the first time from the CHM. Principally similar information can be gained from the height value distributions, but the analysis based on the CHM texture can be carried out without extracting the point data for the crown segments. However, a more complex analysis would be required with respect to these, as suggested in study I. The intensity variables were in a smaller role, as these were considered noisy in the beginning of the study. The results of Korpela et al. (2010), however, indicate an ability to considerably reduce this noise by intensity normalization with reference to the scanning range and receiver gain control, thus improving the species recognition based on the sensor applied in III and IV by 6–8%. The intensity values are related to the amount of foliage within the tree crowns (Kim et al. 2009; Korpela et al. 2010), and more attention should be focused towards these in the future.

The estimations based on alpha shapes gained from including other predictors especially in the case of lower density ALS data. Furthermore, the dependencies between the predictors and the attributes of interest were rarely linear, indicating the choice of non-parametric estimation methods. Since the variable extraction procedure resulted in an impractical number of candidate variables, further demands were set for any imputation method that was able to utilize this information. The number of candidate variables could have been reduced by excluding some variables or groups on the basis of an expert opinion (cf. Korpela et al., 2009a), but the inclusion of the alpha shape metrics resulted in numerous, inter-correlated candidate variables, the rating of which for predictor importance would have been difficult. Selecting the most important out of a set of tens or hundreds of candidate variables is an ambiguous task independent of the variable reduction method.

In III, NN estimation was tested for making a better use of the high number of predictors available, but also avoiding the error propagation of an estimation chain. Here the ALS-based variables were used for estimating tree species, DBH, height and volume simultaneously by the  $k$ -MSN and RF imputation methods, which were selected based on experiences gained from previous studies (Maltamo et al. 2009b; Hudak et al. 2008). Only marginal differences between these could be pointed out when the estimation accuracies were considered, but the strength of the RF method was its ability to handle all predictors with no need for their reduction. This method with  $k=1$  showed an accuracy of 78% in species classification, and the estimates of DBH, height and volume had RMSEs of 13%, 3%, and 31%, respectively, when evaluated against separate validation data. Slightly higher accuracies could be achieved using the  $k$ -MSN method and different values for  $k$ , but overall, the previous estimates were very close to the best ones achieved.

Classification and modeling literature generally promotes approaches in which the result is obtained employing a sufficient minimum of predictor variables, appealing to the problem known as the curse of dimensionality (e.g. Theodoridis and Koutroumbas 2009). According to it, the increase in the number of predictor variables increases model noise without much gain in performance, leads to computational complexity, and weakens the generalization properties of a model. Due to its iterative training and evaluation procedure, RF algorithm is, however, considered more neutral to these problems (Breiman 2001; Crookston and Finley 2008). Here the robustness of RF was further verified, since altogether 1846 variables were used for constructing the RFs without discovering any instability due to the high number of predictors when validating the results against separate data. A high number of predictor variables apparently adds redundant information, but the ability to avoid a delicate variable selection process is an advantage as far as practical applications are concerned (see also Peuhkurinen et al. 2008). The variables were computed automatically, and although their number increases the processing time, practically no man-hours are involved no matter how high the number of possible predictors may be.

The need to acquire reference data is a crucial element to be considered with NN methods. The estimation in Korpela et al. (2007a), for example, was based on national-regional models formulated using measurements made on permanent sample plots used in the National Forest Inventory in Finland (Kalliovirta and Tokola 2005), so that the approach can basically be applied without a need for field work. Also, a single stand-specific observation may be enough to calibrate the estimates produced by those models to local conditions. Tree-level imputation places further demands on the reference data, as trees that are positioned in the field and corresponding ALS data are required. An inventory based on the current approach would therefore include the collection of reference data, which definitely places constraints on its applicability. According to the present results (III), only fairly small field reference data was required, but the potential for extrapolation to a geographically wider area, for example, is unknown.

In addition to the amount, care must be taken in collecting reference data of adequate quality. Study IV used an extensive body of reference data, 3147 observations, but the ALS points were extracted in a different manner from that used in collecting the validation data. This difference was expected to cause some inaccuracy in the nearest-neighbor search, and together with the segmentation errors, this was indeed observed in the form of overestimation. The 5–8% lower precision relative to III, a large portion of which can most likely be accounted for this difference, suggests that the point data sets for both reference and validation do indeed need to be extracted using the same method. This is a slight backlash, since the current point data collection technique would have been an excellent means for linking the tree attributes of known location to the point data. Segmented data, on the other hand, always include errors which affect the linking result. The results of IV also indicate a need to collect a local field reference data for each ALS data set, since similar inaccuracies can originate from sensor-specific differences, for example.

Study IV examined the aggregated, plot-level accuracy of single-tree remote sensing, as opposed to the evaluations at the tree level carried out in I–III. Using the *AutoLiDAR* method, the errors in tree delineation and estimating the tree attributes resulted in unreliable tree-level descriptions of the test stands. On the other hand, the semi-manual *FotoLiDAR* method was successful in species determination and locating the dominant trees, while the estimation based on it was hindered by inaccuracy in the crown width measurements and the imprecision of the allometric equations. In this analysis, the errors in tree detection were not taken into account in the aggregation phase (cf. Breidenbach et al. 2010). With the

*AutoLiDAR* method, the commission trees compensated for the number of those missed, which improved the accuracy of the total volume estimates but naturally detracted from that of the diameter distributions. Considerably more accurate distributions were produced by the *FotoLiDAR* method, but these were also averaged, on account of imprecision, the omission of small trees, and the averaging involved in the regression estimation.

The principal idea of single-tree interpretation is to produce tree lists for the target areas. Despite the fairly accurate grid-level totals, the automatically produced single-tree data need refinement. More operator intervention, in the form of calibration field measurements, for example, seems unavoidable. On the other hand, when semi-manual tree detection and point data extraction with aerial images as an auxiliary data source was involved (III), species recognition accuracy of 80% and stem volume estimates with an RMSE of about 30% were obtained. This analysis was based on only those trees that were discernible in the remote sensing data, i.e. those located on dominant and intermediate tree layers. The accuracy of single-tree attributes could have been higher also in IV if the interpretation was focused on the dominant trees, while also the differences in the training and validation data sets should be considered. Attempting to predict the full distribution of diameters would anyhow require modifications to the estimation procedure, in a manner presented by Breidenbach et al. (2010), for example.

In study V, the CBHs of dominant Scots pine trees were estimated by employing the concepts of Delaunay triangulations and alpha shapes. These new methods make use of 3-D triangulations of the point clouds, and in principle do not need any *a priori* knowledge for the estimation. Therefore the methods should adapt to the properties of the ALS data, yet this was tested within one data set only. The developed methods resulted in RMSEs of 3–4 m (20–30%), while an RMSE of less than 1.5 m (14%) was achieved by a local regression model. Still, considering the low level of correspondence between the field-measured CBH and the ALS point data of some trees, as confirmed visually, the technical definition of CBH was concluded to affect the result, when also the new methods may have approximated the living crown relatively well. Correspondingly to the collection of the modeling data, the bias in the CBH estimates could be removed by calibration field measurements.

Although aerial images were used in crown modeling in III and IV, the *AutoLiDAR* method, for example, did not involve any additional remote sensing data source. Furthermore, the payback of including aerial images only for the estimation can be questioned, considering the 85–90% species recognition accuracy obtainable solely from ALS (cf. Korpela et al. 2010). Still, the species recognition inaccuracy originated mainly from deciduous trees, the classification of which could be aided by image features (e.g. Holmgren et al. 2008b). An alternative means for acquiring the ALS data could be by systems capable of digitizing a full waveform of each backscattered pulse (see Mallet and Bretar 2009). Tests carried out in central Europe suggest gaining additional information for separating different types of vegetation, such as trees and bushes (Wagner et al. 2006), delineating trees (Reitberger et al. 2009) and distinguishing coniferous-deciduous species (Reitberger et al. 2008; Hollaus et al. 2009). Furthermore, waveform decomposition leads to an increase in the point density (Wagner et al. 2006), which would bring further possibilities for developing crown structural and complexity characteristics such as the number of solid alpha shape components.

There have been several reports of high tree detection rates and height estimation accuracies (see Hyyppä et al. 2008), but fewer authors have focused on the quality of the segmentation. Overall, rather little is known on the preconditions of single-tree

interpretation of ALS data, in terms of both appropriate ALS parameters and influence of the forest canopy structure. Here the CBH estimation was found feasible from data with 4 pulses  $\text{m}^{-2}$  (V), while sensitivity analysis performed in II suggested some 3 pulses  $\text{m}^{-2}$  to be required in alpha shape approach. However, the purpose of the latter analysis was particularly to verify the applicability of the metrics in data sets with lower density to be tested in further experiments, assuming accurate tree delineation independent of the pulse density, for example. A more complex analysis would be required to examine final effects of the data density by means of simulation (cf. Kukko and Hyypä 2009).

Although Kaartinen and Hyypä (2008) found only marginal improvements in tree detection and delineation accuracies using data with a density higher than 2 pulses  $\text{m}^{-2}$ , methods based on detailed point data would gain from higher density. Additionally, their conclusions were based on test sites overall suitable for single-tree detection, while stand structure is another important factor to be considered. Study III suggested the estimation accuracy to diminish in forests with a basal area above 22  $\text{m}^2/\text{ha}$ . This is only a suggestive figure of the effect of forest density, but shows the effect the clearly different crown geometries of interlaced and stand-alone trees, as can be seen by ALS, have to the estimation. Particularly crown coverage and spatial clustering could be attempted to be related to the accuracy of the estimation (cf. Falkowski et al. 2008). This information can be determined from ALS data (e.g. Holmgren et al. 2008a), and would thus give an idea of the expected performance prior to the actual analysis.

Furthermore, when considering the practical realization of an ALS-based inventory, the role of single-tree measurements should be clarified. Tree and area-based methods are not exclusive alternatives, and tree-level data can also be produced by predicting the diameter distribution from the area-based data (Peuhkurinen et al. 2008; Packalén and Maltamo 2008), for example. Assuming a well representative field reference data available, the area-based modeling approach is not as sensitive to the fundamental bias-problem in single-tree remote sensing, and although aiming at unbiased estimates is possible also by the latter (Breidenbach et al. 2010), the area-based approach will probably be preferred for many applications due to lower acquisition costs related to such data. Instead of attempting to develop single-tree methods further towards wall-to-wall mapping, it would be worthwhile to examine how area-based estimates could be complemented specifically by detailed measurements of the dominant tree layer. In this sense the ability to measure quality attributes such as CBH is essential.

Uusitalo (1995) proposed the information content collected for forest management planning to be improved from the wood procurement point of view by performing additional measurements for a representative selection of individual trees from those stands to be harvested. He found DBH, dead branch height, crown height, and tree height the most essential measurements with respect to the quality of Scots pine timber. The results of Maltamo et al. (2009b) indicate a possibility to produce these attributes from single-tree ALS data. It is reasonable to expect ALS data to be available nationwide in Finland, since such data will be required in private forest planning and topographic elevation modeling, yet the sampling density of the data is unlikely to be sufficient for single-tree methods. One important purpose for such data is the allocation of needs for detailed measurements. Whether the information content of the nationwide ALS data could be improved by additional airborne and/or field measurements should become apparent by comparing the efforts of acquiring additional data against gained values (cf. Kangas 2010). In all, it seems that further research should be focused on assessing the tree-level production line with respect to obtainable information, alternative methods and their costs.

## REFERENCES

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–723.
- Brandtberg, T. 1999. Automatic individual tree-based analysis of high spatial resolution remotely sensed data. Doctoral thesis. Swedish University of Agricultural Sciences, Centre for Image Analysis, Uppsala, Sweden. *Acta Universitatis Agriculturae Sueciae, Silvestria* 118. 155 p.
- 2007. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. *ISPRS Journal of Photogrammetry and Remote Sensing* 61(5): 325–340.
- & Warner, T. 2006. High-spatial-resolution remote sensing. In: Shao, G. & Reynolds, K.M. (eds.). *Computer applications in sustainable forest management*. Springer, pp. 20–41.
- , Warner, T.A., Landenberger, R.E. & McGraw, J.B. 2003. Detection and analysis of individual leaf-off tree crowns in small footprint, high density lidar data from the eastern deciduous forest in North America. *Remote Sensing of Environment* 85: 290–303.
- Breidenbach, J., Næsset, E., Lien, V., Gobakken, T. & Solberg, S. 2010. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment* 114: 911–924.
- Breiman, L. 2001. Random Forests. *Machine Learning* 45(1): 5–32.
- Burnham, K.P. & Anderson, D.R. 1998. *Model selection and inference: a practical information-theoretic approach*. Springer.
- Chen, Q., Gong, P., Baldocchi, D. & Tian, Y.Q. 2007. Estimating basal area and stem volume for individual trees from lidar data. *Photogrammetric Engineering & Remote Sensing* 73: 1355–1365.
- Crookston, N.L. & Finley, A.O. 2008. yaImpute: An R package for kNN imputation. *Journal of Statistical Software* 23(10). 16 p.
- Da, T.K.F. & Yvinec, M. 2007. 3D alpha shapes. In: CGAL Editorial Board (eds.). *CGAL user and reference manual, 3.3 edition*. Available online at [http://www.cgal.org/Manual/3.3/doc\\_html/cgal\\_manual/packages.html#Pkg:AlphaShapes3](http://www.cgal.org/Manual/3.3/doc_html/cgal_manual/packages.html#Pkg:AlphaShapes3).
- Diaz-Uriarte, R. 2009. varSelRF: Variable selection using random forests. R package version 0.7-1. Available online at <http://ligarto.org/rdiaz/Software/Software.html>.
- Edelsbrunner, H. & Mücke, E.P. 1994. Three-dimensional alpha shapes. *ACM Transactions on Graphics* 13: 43–72.
- Eskelson, B.N.I., Temesgen, H., Lemay, V., Barrett, T.M., Crookston, N.L. & Hudak, A.T. 2009. The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases. *Scandinavian Journal of Forest Research* 24: 235–246.
- Falkowski, M.J., Smith, A.M.S., Gessler, P.E., Hudak, A.T., Vierling, L.A. & Evans, J.S. 2008. The influence of conifer forest canopy cover on the accuracy of two individual tree measurement algorithms using lidar data. *Canadian Journal of Remote Sensing* 34(Suppl. 2): S338–S350.

- Flewelling, J. 2008. Probability models for individually segmented tree crown images in a sampling context. In: Hill, R.A., Rosette, J. & Suárez, J. (eds.). Proceedings of SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment and inventory. September 2008, Edinburgh, UK, pp. 284–294.
- Garczarek, U.M. 2002. Classification rules in standardized partition spaces. Doctoral thesis, University of Dortmund, Dortmund, Germany. 226 p. Available online at [hdl.handle.net/2003/2789](http://hdl.handle.net/2003/2789).
- Goodwin, N.R., Coops, N.C. & Culvenor, D.S. 2006. Assessment of forest structure with airborne LiDAR and the effects of platform altitude. *Remote Sensing of Environment* 103: 140–152.
- Haralick, R.M., Shanmugam, K. & Dinstein, J. 1973. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* 3: 610–621.
- Hawbaker, T.J., Keuler, N.S., Lesak, A.A., Gobakken, T., Contrucci, K. & Radeloff, V.C. 2009. Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. *Journal of Geophysical Research* 114(G00E04). 11 p.
- Hirata, Y., Tsubota, Y. & Sakai, A. 2009. Allometric models of DBH and crown area derived from QuickBird panchromatic data in *Cryptomeria japonica* and *Chamaecyparis obtusa* stands. *International Journal of Remote Sensing* 30: 5071–5088.
- Hirschmugl, M. 2008. Derivation of forest parameters from UltracamD data. Doctoral thesis, Institute of Remote Sensing and Photogrammetry, Graz University of Technology, Graz, Austria. 220 p.
- Höfle, B. & Pfeifer, N. 2007. Correction of laser scanning intensity data: Data and model-driven approaches. *ISPRS Journal of Photogrammetry and Remote Sensing* 62: 415–433.
- Hollaus, M. 2006. Large scale applications of airborne laser scanning for a complex mountainous environment. Doctoral thesis, Institute of Photogrammetry and Remote Sensing, Vienna University of Technology, Vienna, Austria. 111 p.
- , Mücke, W., Höfle, B., Dorigo, W., Pfeifer, N., Wagner, W., Bauerhansl, C. & Regner, B. 2009. Tree species classification based on full-waveform airborne laser scanning data. *SilviLaser 2009*, October 14–16, 2009, College Station, Texas, U.S.A.
- Holmgren, J. 2003. Estimation of forest variables using airborne laser scanning. Doctoral thesis summary. Swedish University of Agricultural Sciences, Department of Forest Resource Management and Geomatics, Umeå, Sweden. *Acta Universitatis Agriculturae Sueciae, Silvestria*, 278. 43 p.
- & Persson, Å. 2004. Identifying species of individual trees using airborne laser scanner. *Remote Sensing of Environment* 90: 415–423.
- , Johansson, F., Olofsson, K., Olsson, H. & Glimskär, A. 2008a. Estimation of crown coverage using airborne laser scanning. In: Hill, R.A., Rosette, J. & Suárez, J. (eds.). Proceedings of SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment and inventory. September 2008, Edinburgh, UK, pp. 50–57.
- , Persson, Å. & Söderman, U. 2008b. Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *International Journal of Remote Sensing* 29: 1537–1552.
- Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E. & Falkowski, M.J. 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment* 112: 2232–2245.
- Hyypä, J. & Inkinen, M. 1999. Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland* 16(2): 27–42.

- , Hyypä, H., Leckie, D., Gougeon, F., Yu, X. & Maltamo, M. 2008a. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing* 29: 1339–1366.
- , Hyypä, H., Yu, X., Kaartinen, H., Kukko, A. & Holopainen, M. 2008b. Forest inventory using small-footprint airborne LiDAR. In: Shan, J. & Toth, C.K. (eds.). *Topographic laser ranging and scanning: Principles and processing*. Taylor & Francis, pp. 335–370.
- Ilomäki, S., Nikinmaa, E. & Mäkelä, A. 2004. Crown rise due to competition drives biomass allocation in silver birch (*Betula pendula* L.). *Canadian Journal of Forest Research* 33: 2395–2404.
- Ilvessalo, Y. 1950. On the correlation between crown diameter and the stem of trees. *Publications of Forest Research Institute in Finland, Communicationes Instituti Forestalis Fenniae* 38(2). 32 p.
- Jakobsons, A. 1970. Sambandet mellan trädkronans diameter och andra träd faktorer, främst brösthöjdsdiametern: Analyser grundade på riksskogstaxeringens provträds material (Summary: The correlation between the diameter of the tree crown and other tree factors, mainly the breast-height diameter: Analysis based on sample-trees from the National Forest Survey). Department of Forest Survey, Royal College of Forestry, Stockholm, Sweden. *Rapporter och uppsatser* 14. 75 p. (In Swedish).
- Kaartinen, H. & Hyypä, J. 2008. EuroSDR / ISPRS Commission II Project "Tree Extraction" – Final report. Official publication n:o 53. EuroSDR, Frankfurt am Main, Germany. 60 p.
- Kaitaniemi, P. & Lintunen, A. 2008. Precision of allometric scaling equations for trees can be improved by including the effect of ecological interactions. *Trees* 22: 579–584.
- 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): 227–248.
- Kangas, A. 2010. Value of forest information. *European Journal of Forest Research*, in press.
- Kantola, A. & Mäkelä, A. 2004. Crown development in Norway spruce [*Picea abies* (L.) Karst.]. *Trees* 18: 408–421.
- Kato, A., Moskal, L.M., Schiess, P., Swanson, M.E., Calhoun, D. & Stuetzle, W. 2009. Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. *Remote Sensing of Environment* 113: 1148–1162.
- Kim, S., McGaughey, R.J., Andersen, H.-E. & Schreuder, G. 2009. Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sensing of Environment* 113: 1575–1586.
- Koch, B., Heyder, U. & Weinacker, H. 2006. Detection of individual tree crowns in airborne lidar data. *Photogrammetric Engineering & Remote Sensing* 72(4): 357–363.
- , Dees, M., Brusselen J. v., Eriksson, L., Fransson, J., Gallaun, H., Leblon, B., McRoberts, R.E., Nilsson, M., Schardt, M., Seitz, R. & Waser, L. 2008. Forestry applications. In: Li, Z., Chen, J. & Baltsavias, E. (eds.). *Advances in photogrammetry, remote sensing and spatial information sciences: 2008 ISPRS congress book*. Taylor & Francis, pp. 439–465.
- Korhonen, K.T., Heikkinen, J., Henttonen, H., Ihalainen, A., Pitkänen, J. & Tuomainen, T. 2006. Suomen metsävarat 2004–2005. *Metsätieteen aikakauskirja* 1B/2006: 183–221. (In Finnish).

- Korpela, I. 2004. Individual tree measurements by means of digital aerial photogrammetry. Doctoral thesis. University of Helsinki, Department of Forest Resource Management, Helsinki, Finland. *Silva Fennica Monographs* 3. 93 p.
- 2007. Incorporation of allometry in single-tree remote sensing with LiDAR and multiple images. In: Heipke, C., Jacobsen, K. & Gerke, M. (eds.). *Proceedings of ISPRS Hannover workshop 2007 on High Resolution Earth Imaging for Geospatial Information*. Hannover, Germany, May 29–June 1, 2007. IAPRS XXXVI Pt I/W51, 6 p.
- & Tokola, T. 2006. Potential of aerial image-based monoscopic and multiview single-tree forest inventory: A simulation approach. *Forest Science* 52: 136–147.
- , Dahlin, B., Schäfer, H., Bruun, E., Haapaniemi, F., Honkasalo, J., Ilvesniemi, S., Kuutti, V., Linkosalmi, M., Mustonen, J., Salo, M., Suomi, O. & Virtanen, H. 2007a. Single-tree forest inventory using lidar and aerial images for 3D treetop positioning, species recognition, height and crown width estimation. In: Rönnholm, P., Hyyppä, H. & Hyyppä, J. (eds.). *Proceedings of ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*, September 12–14, 2007, Espoo, Finland. IAPRS Vol. XXXVI, Part 3 / W52, pp. 227–233.
- , Tuomola, T. & Välimäki, E. 2007b. Mapping forest plots: An efficient method combining photogrammetry and field triangulation. *Silva Fennica* 41(3): 457–469.
- , Koskinen, M., Vasander, H., Holopainen, M. & Minkkinen, K. 2009a. Airborne small-footprint discrete-return LiDAR data in the assessment of boreal mire surface patterns, vegetation, and habitats. *Forest Ecology and Management* 258: 1549–1566.
- , Tokola, T., Ørka, H.O. & Koskinen, M. 2009b. Small-footprint discrete-return LIDAR in tree species recognition. In: Heipke, C., Jacobsen, K., Müller, S. & Sörgel, U. (eds.). *Proceedings of ISPRS Hannover workshop 2009 on High Resolution Earth Imaging for Geospatial Information*. Hannover, Germany, June 2009. IAPRS XXXVIII-1-4-7/W5.
- , Ørka, H.O., Maltamo, M., Tokola, T. & Hyyppä, J. 2010. Tree species classification using airborne LIDAR – Effects by stand and tree parameters, downsizing the training set, intensity normalization and sensor type. *Silva Fennica*, in press.
- Koukoulas, S. & Blackburn, G.A. 2005. Mapping individual tree location, height and species in broadleaved deciduous forest using airborne LiDAR and multi-spectral remotely sensed data. *International Journal of Remote Sensing* 26: 431–455.
- Kukko, A. & Hyyppä, J. 2009. Small-footprint laser scanning simulator for system validation, error assessment, and algorithm development. *Photogrammetric Engineering and Remote Sensing* 75(10): 1177–1190.
- Laasasenaho, J. 1982. Taper curve and volume function for pine, spruce and birch. *Publications of Forest Research Institute in Finland, Communicationes Instituti Forestalis Fenniae* 108. 74 p.
- Lämås, T. & Eriksson, L.O. 2003. Analysis and planning systems for multiresource, sustainable forestry: The Heureka research programme at SLU. *Canadian Journal of Forest Research* 33: 500–508.
- Lévesque, J. & King, D. 2003. Spatial analysis of radiometric fractions from high-resolution multispectral imagery for modelling individual tree crown and forest canopy structure and health. *Remote Sensing of Environment* 84: 589–602.
- Liang, X., Hyyppä, J. & Matikainen, L. 2007. Deciduous-coniferous tree classification using difference between first and last pulse laser signatures. In: Rönnholm, P., Hyyppä, H. & Hyyppä, J. (eds.). *Proceedings of the ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*, September 12–14, 2007, Espoo, Finland. IAPRS Vol. XXXVI, Part 3 / W52, pp. 253–257.

- Lillesand, T.M., Kiefer, R.W. & Chipman, J.W. 2004. Remote sensing and image interpretation. Wiley & Sons. 763 p.
- Lindberg, E., Holmgren, J., Olofsson, K., Olsson, H. & Wallerman, J. 2010. Estimation of tree lists from airborne laser scanning by combining single tree and area-based methods. *International Journal of Remote Sensing* 31: 1175–1192.
- Magnusson, M. 2006. Evaluation of remote sensing techniques for estimation of forest variables at stand level. Doctoral thesis summary. Swedish University of Agricultural Sciences, Umeå, Sweden. *Acta Universitatis Agriculturae Sueciae* 2006:85. 38 p.
- , Fransson, J.E.S. & Holmgren, J. 2007. Effects on estimation accuracy of forest variables using different pulse density of laser data. *Forest Science* 53: 619–626.
- Mäkelä, A. & Vanninen, P. 2001. Vertical structure of Scots pine crowns in different age and size classes. *Trees* 15: 385–392.
- Mäkinen, A., Korpela, I., Tokola T. & Kangas, A. 2006. Effects of imaging conditions on crown diameter measurements from high resolution aerial images. *Canadian Journal of Forest Research* 36: 1206–1217.
- Mallet, C. & Bretar, F. 2009. Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing* 64: 1–16.
- Maltamo, M., Eerikäinen, K., Pitkänen, J., Hyypä, J. & Vehmas, M. 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. *Remote Sensing of Environment* 90: 319–330.
- , Hyypä, J. & Malinen, J. 2006a. A comparative study of the use of laser scanner data and field measurements in the prediction of crown height in boreal forests. *Scandinavian Journal of Forest Research* 21: 231–238.
- , Malinen, J., Packalén, P., Suvanto, A. & Kangas, J. 2006b. Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data. *Canadian Journal of Forest Research* 36: 426–436.
- , Packalén, P., Peuhkurinen, J., Suvanto, A., Pesonen, A. & Hyypä, J. 2007. Experiences and possibilities of ALS based forest inventory in Finland. In: Rönnholm, P., Hyypä, H. & Hyypä, J. (Eds.). *Proceedings of ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, September 12–14, 2007, Espoo, Finland. IAPRS Vol. XXXVI, Part 3 / W52, pp. 270–279.*
- , Bollandsås, O.M., Næsset, E., Gobakken, T. & Packalén, P. 2009a. Different sampling strategies for field training plots in ALS inventory. *Proceedings of SilviLaser 2009 held in October 14–16, 2009, College Station, Texas, U.S.A.* 10 p.
- , 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): 507–521.
- Martynov, I. 2008. Computing the persistent homology of range images with alpha shapes. M. Sc. thesis, Department of Information Technology, Lappeenranta University of Technology, Lappeenranta, Finland. 67 p.
- Mehtätalo, L. 2006. Eliminating the effect of overlapping crowns from aerial inventory estimates. *Canadian Journal of Forest Research* 36: 1649–1660.
- Metsäkeskus. 2009. Metsävaratietoihin perustuvat tuotteet ja palvelut uudistuvat. 18 Dec 2009. Available at [http://www.metsakeskus.fi/web/fin/palvelut/Tulossa/Metsasuunnittelu\\_ja\\_tietopalvelut/etusivu.htm](http://www.metsakeskus.fi/web/fin/palvelut/Tulossa/Metsasuunnittelu_ja_tietopalvelut/etusivu.htm) [Accessed 2 March 2010]. (In Finnish).
- Moer, M. & Stage, A.R. 1995. Most Similar Neighbor: An improved sampling inference procedure for natural resources planning. *Forest Science* 41: 337–359.

- Moffiet, T., Mengersen, K., Witte, C., King, R. & Denham, R. 2005. Airborne laser scanning: exploratory data analysis indicates potential variables for classification of individual trees or forest stands according to species. *ISPRS Journal of Photogrammetry & Remote Sensing* 59: 289–309.
- Morsdorf, F., Meier, E., Kötz, B., Itten, K.I., Dobbertin, M. & Allgöwer, B. 2004. LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment* 92: 353–362.
- Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80: 88–99.
- , Gobakken, T., Holmgren, J., Hyypä, H., Hyypä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å. & Söderman, U. 2004. Laser scanning of forest resources: The Nordic experience. *Scandinavian Journal of Forest Research* 19: 482–499.
- Nelson, R., Krabill, W. & MacLean, G. 1984. Determining forest canopy characteristics using airborne laser data. *Remote Sensing of Environment* 15: 201–212.
- Olofsson, K., Lindberg, E. & Holmgren, J. 2008. A method for linking field-surveyed and aerial-detected single trees using cross correlation of position images and the optimization of weighted tree list graphs. In: Hill, R.A., Rosette, J. and Suárez, J. (eds.). *Proceedings of SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment and inventory*. September 2008, Edinburgh, UK, pp. 95–104.
- Ørka, H.O., Næsset, E. & Bollandsås, O.M. 2009. Classifying species of individual trees by intensity and structure features derived from airborne laser scanning data. *Remote Sensing of Environment* 113: 1163–1174.
- Packalén, P. 2009. Using airborne laser scanning data and digital aerial photographs to estimate growing stock by tree species. Doctoral thesis summary. University of Joensuu, Faculty of Forest Sciences, Joensuu, Finland. *Dissertationes Forestales* 77. 41 p.
- & Maltamo, M. 2008. Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. *Canadian Journal of Forest Research* 38: 1750–1760.
- , Maltamo, M. & Tokola, T. 2008a. Detailed assessment using remote sensing techniques. In: Gadow, K. & Pukkala, T. (eds). *Designing green landscapes. Managing Forest Ecosystems* 15, Springer, pp. 51–77.
- , Pitkänen, J. & Maltamo, M. 2008b. Comparison of individual tree detection and canopy height distribution approaches: A case study in Finland. In: Hill, R.A., Rosette, J. & Suárez, J. (eds.). *Proceedings of SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment and inventory*. Sept. 2008, Edinburgh, UK, pp. 22–29.
- Persson, Å., Holmgren, J. & Söderman, U. 2002. Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering & Remote Sensing* 68: 925–932.
- Peuhkurinen, J., Maltamo, M., Malinen, J., Pitkänen, J. & Packalén, P. 2007. Preharvest measurement of marked stands using airborne laser scanning. *Forest Science* 53: 653–661.
- , 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): 625–641.

- Pinheiro, J.C. & Bates, D.M. 2000. Mixed-effects models in S and S-PLUS. Springer.
- Pitkänen, J. 2005. A multi-scale method for segmentation of trees in aerial images. In: Hobbelstad, K. (ed.). Proceedings of the SNS-meeting at Sjusjøen – Forest Inventory and Planning in Nordic Countries, Norway, September 6–8, 2004. NIJOS-report 09/05, Norwegian Institute of Land Inventory, Oslo, Norway.
- , Maltamo, M., Hyypä, J. & Yu, X. 2004. Adaptive methods for individual tree detection on airborne laser based canopy height model. In: Theis, M., Koch, B., Spiecker, H. & Weinacker, H. (eds.). Proceedings of ISPRS working group VIII/2: "Laser-Scanners for Forest and Landscape Assessment". University of Freiburg, Germany, pp. 187–191.
- Popescu, S.C. 2007. Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy* 31: 646–655.
- & Zhao, K. 2008. A voxel-based lidar method for estimating crown base height for deciduous and pine trees. *Remote Sensing of Environment* 112: 767–781.
- , Wynne, R.H. & Nelson, R.F. 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing* 29: 564–577.
- Pyysalo, U. 2006. Single tree feature extraction from airborne laser scanner data applying vector models. Thesis for degree of Licentiate of Science in Technology. Department of Surveying, Helsinki University of Technology, Helsinki, Finland. 78 p.
- & Hyypä, H. 2002. Reconstructing tree crowns from laser scanner data for feature extraction. ISPRS Commission III Symposium, IAPRS XXXIV: 302–309.
- Rautiainen, M., Möttöus, M., Stenberg, P. & Ervasti, S. 2008. Crown envelope shape measurements and models. *Silva Fennica* 42(1): 19–33.
- Reitberger, J., Krzystek, P. & Stilla, U. 2008. Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. *International Journal of Remote Sensing* 29: 1407–1431.
- , Schnörr, C., Krzystek, P. & Stilla, U. 2009. 3D segmentation of single trees exploiting full waveform LIDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing* 64: 561–574.
- Säynäjoki, R., Packalén, P., Maltamo, M., Vehmas, M. & Eerikäinen, K. 2008. Detection of aspens using high resolution aerial laser scanning data and digital aerial images. *Sensors* 8: 5037–5054.
- Searle, S.R. 1971. Linear models. Wiley & Sons.
- Solberg, S. & Næsset, E. 2007. Mapping defoliation with lidar. In: Rönnholm, P., Hyypä, H. & Hyypä, J. (eds.). Proceedings of ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, September 12–14, 2007, Espoo, Finland. IAPRS Vol. XXXVI, Part 3 / W52, pp. 379–382.
- , Næsset, E. & Bollandsås, O.M. 2006. Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. *Photogrammetric Engineering & Remote Sensing* 72: 1369–1378.
- Takahashi, T., Yamamoto, K., Senda, Y. & Tsuzuku, M. 2005. Predicting individual stem volumes of sugi (*Cryptomeria japonica* D. Don) plantations in mountainous areas using small-footprint airborne LiDAR. *Journal of Forest Research* 10: 305–312.
- Talts, J. 1977. Mätning i storskaliga flygbilder för beståndsdatainsamling (Summary: Photogrammetric measurements for stand cruising). Department of Forest Mensuration and Management, Royal College of Forestry, Stockholm, Sweden. Research notes NR 6 – 1977. 102 p. (In Swedish).

- Theodoridis, S. & Koutroumbas, K. 2009. Pattern recognition. Fourth edition. Academic press, 961 p.
- Uusitalo, J. 1995. Pre-harvest measurement of pine stands for sawing production planning. Doctoral thesis. University of Helsinki, Department of Forest Resource Management, Helsinki, Finland. Publications 9. 96 p.
- , Kokko, S. & Kivinen, V.-P. 2004. The effect of two bucking methods on Scots pine lumber quality. *Silva Fennica* 38(3): 291–303.
- Utterä, J., Anttila, P., Suvanto, A. & Maltamo, M. 2006. Yksityismetsien metsävaratiedon keruuseen soveltuvilla kaukokartoitusmenetelmillä estimoitujen puustotunnusten luotettavuus. *Metsätieteen aikakauskirja* 4/2006: 507–519. (In Finnish).
- Venables, W.N. & Ripley, B.D. 2002. Modern applied statistics with S. Fourth edition. Springer, 495 p.
- Villikka, M., Maltamo, M., Packalén, P., Vehmas, M. & Hyypä, J. 2007. Alternatives for predicting tree-level stem volume of Norway spruce using airborne laser scanner data. *The Photogrammetric Journal of Finland* 20(2): 33–42.
- Wagner, W., Ullrich, A., Ducic, V., Melzer, T. & Studnicka, N. 2006. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitizing airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing* 60: 100–112.
- Wang, Y., Weinacker, H. & Koch, B. 2008. A lidar point cloud based procedure for vertical canopy structure analysis and 3D single tree modelling in forest. *Sensors* 8: 3938–3951.
- Zhu, C., Zhang, X., Hu, B. & Jaeger, M. 2008. Reconstruction of tree crown shape from scanned data. In: Pan, Z., Zhang, X., El Rhalibi, A., Woo, W. & Li, Y. (eds.). *Technologies for E-Learning and Digital Entertainment – Proceedings of Edutainment 2008*, Nanjing, China, June 25–27, 2008. *Lecture Notes in Computer Science* 5093/2008, pp. 745–756.