

461

# Predicting the occurrence of large-diameter trees using airborne laser scanning

Lauri Korhonen, Christian Salas, Torgrim Østgård, Vegard Lien, Terje Gobakken, and Erik Næsset

Abstract: Large-diameter trees are important for both ecological and economic reasons, but they have become increasingly rare. Thus, there is an interest in easily locating such trees, and for this purpose, the use of airborne laser scanning (ALS) seems suitable. Our objective was to assess the accuracy of area-based ALS estimation in predicting the number of large-diameter Scots pines (*Pinus sylvestris* L.). A sample of 856 trees with a diameter >35 cm were measured from 1109 sample plots located in eastern Norway. We fitted negative binomial and zero-inflated negative binomial models for predicting large-diameter tree counts. ALS-derived and external variables were used as predictors when fitting the models. The accuracy was assessed based on the weighted kappa coefficient and cross validation. Our best model was based on three ALS height distribution variables, one horizontal ALS variable, and plot elevation. Its overall accuracy was 65.8% and the weighted kappa was 0.55. Although there was a clear relationship between the response and the proposed predictor variables, fairly large errors in the predicted large-diameter tree counts.

Key words: large trees, LiDAR, area-based method, negative binomial regression, old growth.

**Résumé :** Les arbres de grand diamètre sont importants, à la fois pour des raisons écologiques et économiques, mais ils sont devenus de plus en plus rares. Il y a donc un intérêt à localiser facilement ces arbres et l'utilisation du balayage laser aéroporté (BLA) semble appropriée pour ce faire. Notre objectif était d'évaluer la précision d'une estimation basée sur la surface obtenue par BLA pour prédire le nombre de pins sylvestres (*Pinus sylvestris* L.) de grand diamètre. Un échantillon de 856 arbres, d'un diamètre > 35 cm, a été mesuré à partir de 1109 placettes échantillons situées dans l'est de la Norvège. Nous avons ajusté des modèles binomiaux négatifs et des modèles binomiaux négatifs à excès de zéros pour prédire le nombre d'arbres de grand diamètre. Des variables obtenues par BLA et des variables externes ont été utilisées comme prédicteurs lors de l'ajustement des modèles. La précision a été évaluée sur la base du coefficient kappa pondéré et par validation croisée. Notre meilleur modèle était basé sur trois variables de distribution de hauteur obtenues par BLA, une variable horizontale obtenue par BLA et l'altitude de la placette. Sa précision globale était de 65,8 % et le coefficient kappa pondéré était de 0,55. Bien qu'il y ait une relation claire entre la variable de réponse et les variables prédictives proposées, d'assez grandes erreurs dans la prédiction du nombre d'arbres de grand diamètre étaient courantes. [Traduit par la Rédaction]

Mots-clés : grands arbres, LiDAR, méthode d'estimation basée sur la surface, régression binomiale négative, forêt ancienne.

# 1. Introduction

Large trees are rare in many forests and difficult to find. They are important for both economic and ecological reasons. Large trees dominate the structure, dynamics, and function of many temperate and tropical forests (Lutz et al. 2012), and their occurrence controls a series of ecological processes. As pointed out by Franklin et al. (1987), a falling large tree often kills other trees, thus controlling stand mortality, influencing regeneration, and mixing forest soil by uprooting other trees. Large standing-dead trees (i.e., snags) are a major resource for the ecosystem because of the large array of organisms present in the decaying log and because they provide shelter for animal species (Franklin et al. 1987). Large trees are key features of old-growth forests, and as an ecological rule, old-growth forests are less common than young forests (Oliver 1980).

Intensive management of commercial forests in the Nordic countries means that trees are usually cut before the forest reaches the old-growth stage. Also, on a worldwide scale, oldgrowth forests are rare (Burrascano et al. 2013), and even then, the number of large trees in them may be relatively small (e.g., 12–20 trees·ha<sup>-1</sup>, Salas et al. 2006). Therefore, finding large trees during an inventory can be methodologically challenging, and from an economic perspective, finding suitable timber is sometimes cost prohibitive. For instance, carpenters working with the restoration of old wooden churches and other buildings of great historical interest need saw logs with exceptionally large dimensions, and these are not found in commercial short-rotation forests (e.g., Godal 1996).

Airborne laser scanning (ALS) has the potential to replace traditional inventories as a main data source for forest management planning (Næsset 2014; Maltamo and Packalen 2014). The increasing availability of ALS data opens new possibilities for ecological studies. In particular, ALS can provide wall-to-wall measurements of the ecologically important properties of the canopy structure, which are otherwise impossible to obtain using field-based methods. Müller and Vierling (2014) provide an extensive review of the

Received 6 October 2015. Accepted 1 January 2016.

L. Korhonen.\* University of Eastern Finland, School of Forest Sciences, P.O. Box 111, 80101 Joensuu, Finland.

C. Salas. Universidad de La Frontera, Laboratorio de Biometría, Departamento de Ciencias Forestales, Av. Francisco Salazar 01145, Temuco, Chile.

T. Østgård, V. Lien, T. Gobakken, and E. Næsset. Norwegian University of Life Sciences, Department of Ecology and Natural Resource Management, P.O. Box 5003, NO-1432 Ås, Norway.

Corresponding author: Lauri Korhonen (email: lauri.korhonen@uef.fi).

<sup>\*</sup>Previous address: University of Helsinki, Department of Forest Sciences, P.O. Box 27, 00014 University of Helsinki, Finland.

application of ALS data for biodiversity studies, especially the assessment of animal and plant habitats. ALS has also been applied to detect snags (Martinuzzi et al. 2009; Wing et al. 2015), other types of deadwood (Maltamo et al. 2014), and cavity trees (Eskelson et al. 2009). However, we are not aware of any studies that consider the ALS-based inventory of living trees with large diameters, although some of the methods applied in snag inventories may also be applicable to the inventory of large-diameter trees.

ALS-based forest inventories are usually performed using the area-based method (Næsset 2002). The inventory area is covered wall to wall by ALS data. A sample of field plots is distributed across the area, often following commonly adapted sampling designs such as systematic and stratified sampling. Volume and other stand variables are either measured or estimated at the plot level. The plots are positioned with Global Navigation Satellite System (GNSS) receivers, and laser echoes from within the plot borders are used to calculate the ALS variables that describe the height and density of the canopy. These ALS variables are used as predictors in statistical models that predict a response variable of interest such as the timber volume. These models can then be used to predict forest variables over the whole inventory region.

Tree size can be quantified based on diameter at breast height (DBH, 1.3 m), height, or volume. In this study, we specifically focus on trees with a large diameter, because it is the easiest dimension to be measured in the field, and it represents both commercial and ecological interests. The original initiative for this research came from a specialized sawmill that wanted to enhance the procurement of sawlogs with very large dimensions. From a remotesensing perspective, the quantification of tree size based on DBH is more difficult than a quantification based on height. ALS sensors can measure the tree height directly with only a slight underestimation (Gaveau and Hill 2003), and the tree height correlates with the DBH. However, tree height-diameter curves are often asymptotic, i.e., old, dominant trees tend to grow little in height but continue to grow in diameter up to very great ages. Thus, if only the height is known, it is difficult to predict the DBH of a dominant tree. However, the DBH is also correlated with crown diameter (Ilvessalo 1950), and the crown diameters estimated from ALS data may improve the predictions of stem volume (Popescu et al. 2003).

The prediction of stand-level variables such as mean diameter (Næsset 2002) and diameter distributions (Gobakken and Næsset 2004) are well-known ALS applications. However, studies focusing on large trees are conducted mainly as long-term experiments using field measurements (e.g., the study of snags by Ganey and Vojta (2005)), but research on methods for predicting the occurrence or density of large trees using remote sensing is lacking. Indeed, the estimation of the occurrence of individual large trees is more difficult than the estimation of aggregated properties such as the mean diameter of all trees. Aggregated properties may not be adequate for the purposes of large-diameter tree inventory, because stands may have only a few trees with very large diameters, and therefore, their contribution to the estimate means may be small. Thus, to produce reliable information for large trees, direct modeling might be a viable alternative to the extraction of large-tree information from existing mean DBH or diameter distribution databases.

The area-based method can be applied using relatively low pulse densities (<1 m<sup>-2</sup>), which makes it cost effective and applicable for large areas. With low pulse density, individual trees cannot be detected. It is possible, however, to interpolate canopy height models (CHMs) that describe the height of the canopy surface above ground level using low pulse density ALS data. Using 1 m resolution CHMs, tree groups can be distinguished, although their area may be overestimated due to the low resolution. Pippuri et al. (2012) applied ALS data with a nominal pulse density of 0.6 m<sup>-2</sup> to calculate horizontal area-based variables based on a CHM. These included texture (Haralick et al. 1973) and landscape (McGarigal and Marks 1995) variables such as the mean size of canopy patches and the mean area of connected canopy pixels above a specified height threshold. Such variables might also correlate with the presence of large trees.

Linear regression models commonly applied in area-based ALS estimation are not suitable for modeling count (i.e., non-negative integer) variables (Consul 1989). Examples of studies where the response variable is a count include tree mortality (Affleck 2006) and also the number of seedlings in a regeneration plot (Zhang et al. 2012). It is common that real-world count data have more zeroes than a statistical randomization would imply. Such data are commonly called zero inflated. To fit models for zero-inflated data, a generalized linear model (GLM) framework must be used. GLMs extend linear statistical modeling to response distributions that belong to the exponential family (Schabenberger and Pierce 2002). A GLM consists of three components: a probability density function, a linear predictor, and a link function (McCulloch and Searle 2001). Several models have been proposed to handle zeroinflated count data (Famoye and Singh 2006). Among the most used are the zero-inflated Poisson and the zero-inflated negative binomial (ZINB) models. For example, Zhang et al. (2012) compared different negative binomial (NB) models to predict tree recruitment and obtained their best results with ZINB models. Further details on zero-inflated models can be found in Consul (1989) and Affleck (2006). As an alternative, nonparametric methods such as the k-nearest neighbor can be used, because they do not require assumptions concerning the distribution of the response variable (Eskelson et al. 2009). However, Eskelson et al. (2009) recommended NB models instead of nonparametric methods for the prediction of stand-level cavity tree and snag abundance.

The objective of the present study was to estimate the number of large-diameter trees at the plot level using an area-based ALS approach and GLMs applicable to the count data and to evaluate the accuracy of the predictions using different types of predictor variables. In addition to the commonly used canopy height and density variables (Næsset 2002), we also tested variables that describe the horizontal canopy structure (Pippuri et al. 2012). Additionally, we considered external plot-level variables such as elevation, because the occurrence of large trees may be affected by the previous harvest history and with less focus on the utilization of more marginal and less accessible forests at higher elevations.

## 2. Materials and methods

#### 2.1. Data

The study area is situated in the municipality of Tolga, located in Hedmark County, Eastern Norway (62°23'N, 11°02'E; Fig. 1). Boreal forests in Tolga were intensively harvested until the late 19th century to supply the nearby mines, but some areas remained untouched. Scots pine (*Pinus sylvestris* L.) forests occur in the valleys between the treeless fells, with Norway spruce (*Picea abies* (L.) Karst.), downy birch (*Betula pubescens* Ehrh.), and goat willow (*Salix caprea* L.) occurring as minor species. In addition, pure birch forests are common, especially in the peatlands, close to water bodies, and at higher elevations near the alpine tree line.

ALS data from five separate east-west oriented flight lines (strips) were obtained from the region (Fig. 1). The strips were spaced 6 km apart. The ALS acquisition was performed on the 6th and 8th of August 2006 as a part of the Hedmark County ALS sample survey (Gobakken et al. 2012). Two different Optech ALTM3100 scanners were used, but only one strip was scanned with the second sensor. The ALS data were acquired with similar settings across all strips: altitude, 800 m above ground level; flight speed, 75 m·s<sup>-1</sup>; pulse repetition frequency, 100 kHz; scan frequency, 55 Hz; half scan angle, 17°. The resultant pulse density was



Fig. 1. Location of the study area and the placement of survey strips within it. The black dots indicate field plots within the scanned strips.

2.8 m<sup>-2</sup>, and the footprint size was 0.68 m. The data was delivered by the contractor (Blom Geomatics AS) and contained the XYZ coordinates of the echoes relative to ground level.

Field plots 32 m × 32 m in size were placed directly below the flight lines, and the plots on a given line always bordered the previous one. The fieldwork was conducted from the 5th to the 21st of October 2009 by walking along the lines with a GNSS navigator. The selection of the sections of the sample strips to be surveyed was subjective and based on local knowledge so that the resultant data set would be as large and diverse as possible (Fig. 1). A total of 1109 plots were sampled. Trees with a DBH > 35 cm (n = 855) were calipered and positioned using real-time differentially corrected GNSS. The positioning accuracy was typically better than 1 m, but this could occasionally decrease if the real-time differential correction data was unavailable. One tree in a plot corresponds to 9.8 trees·ha<sup>-1</sup>. Approximately every fifth tree (n = 145) was also measured for height. In our sample, 94% of the trees are pines, followed by spruce (5%) and deciduous species.

The distribution of the large-diameter trees (Fig. 2) shows that the majority of the plots (66%) had no large trees. Only a small percentage of plots (6.1%) had more than three large-diameter trees. In addition, there was no clear relationship between the height and diameter of the large-diameter trees (Fig. 3). It should be noted that we did not measure trees with a DBH < 35 cm.

## 2.2. Predictor variables

Three types of predictor variables were used for modeling purposes: variables derived from the canopy height distributions, CHM variables, and external variables. A complete list of the predictors and their abbreviations is given in Table 1. Canopy height distribution variables were calculated at the plot level using only echoes above a 1.3 m height threshold. We computed several height percentiles ( $h_0$ ,  $h_{10}$ , ...,  $h_{90}$ ) from this distribution. Furthermore, we derived several measures of canopy density. The calculation of point variables representing canopy density was carried out for 10 different vertical fractions of equal thickness (Næsset 2004). The thickness of each fraction was defined as one-tenth of the distance between the 95 percentile and the lowest canopy height, i.e., 1.3 m (Gobakken and Næsset 2008). Canopy densities

Fig. 2. Observed tree counts per plot within the 1109 plots.



were then computed as the proportions of laser echoes above fraction 0, 1, ..., 9 to the total number of echoes and denoted as  $d_0$ ,  $d_1$ , ...,  $d_9$ , respectively. In addition, the maximum, mean, and coefficient of variation (cv) of the echo heights were calculated. Each of these variables was calculated twice using the first and last echoes as two separate data sets. We respectively use lowercase letters "f" and "1" to indicate the variables derived from the first and last echoes.

ALS variables that describe the horizontal canopy structure were calculated from the CHMs created with 0.5 m resolution. The CHM was created by assigning each grid cell the value of the highest ALS echo within a 1 m radius from the center of the cell. This method was selected because it also enables the creation of smooth CHMs from low-density ( $1 < echoes \cdot m^{-2}$ ) data (Mustonen

Fig. 3. Height-diameter scatterplot for large-diameter sample trees (n = 145)



Table 1. List of all predictor variables and their abbreviations.

Abbreviation	Definition
Height distribu	tion variables
max, mean, cv	Maximum, mean, and coefficient of variation of echo heights above 1.3 m (m)
$h_0, h_{10}, \dots, h_{90}$	Height percentiles for echoes above 1.3 m (m)
$d_0, d_1, \dots, d_9$	Density variables for echoes above 1.3 m (%)
Horizontal CHM	1 variables
N <sub>cp</sub> , A <sub>cp</sub>	The number of canopy patches and their total area
Amean <sub>cp</sub> , Asd <sub>cp</sub>	Mean and standard deviation of the area of the canopy patches
Hmean <sub>cp</sub> , Hsd <sub>cp</sub>	Mean and standard deviation of the height of the canopy patches
Nap	Number of all (ground and canopy) patches
Amean <sub>ap</sub> , Asd <sub>ap</sub>	Mean and standard deviation of the area of all patches
euler	Euler number of the canopy patches
CHM texture va	riables
con, cor, ene, hom	Contrast, correlation, energy, and homogeneity
External variab	les
along ganget	Slope separt and elevation of the plot

slope, aspect,	Slope, aspect, and elevation of the plot
elev	
rdist sinder	Distance to road and site index

Note: For height distribution variables, Subscripts f and 1 indicate if the variable was calculated using first-and-only or last-and-only echoes, respectively. For horizontal canopy height model (CHM) variables, subscripts h and l indicate if the variable was calculated using high (0.5 m) or low (1 m) resolution CHM, respectively; the numeric subscript indicates the height that was used to divide the patches into ground and canopy. For CHM texture variables, the number subscript indicates the offset distance.

et al. 2008; Pippuri et al. 2012) and, therefore, is commonly applied in practical inventories to create CHMs for stand delineation. As our initial ALS pulse density of 2.8 m<sup>-2</sup> was slightly higher than that commonly used in practical inventories (0.5-1.0 m<sup>-2</sup>), we reduced it to 1 m<sup>-2</sup> by randomly selecting one echo from each 1 m<sup>2</sup> cell and then created an alternative low-resolution CHM using a 1 m resolution and a 2 m search radius (Mustonen et al. 2008). The ALS height distribution variables were not recalculated with the lower pulse density, as density reduction does not significantly influence the estimation results (Maltamo et al. 2006).

The CHMs were used to calculate horizontal and texture variables for each plot. Horizontal variables were calculated similar to Pippuri et al. (2012). The image was first divided into canopy and ground patches (separate groups of eight-connected pixels) using height threshold values of 10, 12.5, 15, and 17.5 m. The number of patches, the mean area of patches, and the standard deviation of the area of patches were then calculated separately using canopy and ground-and-canopy patches as two separate data sets. In addition, for canopy patches, we also calculated the total area, mean pixel height, standard deviation of the pixel heights, and the Euler number (the number of canopy patches minus the total number of holes in these patches). Haralick's texture metrics (Haralick et al. 1973) of contrast, correlation, energy, and homogeneity were also calculated from the CHMs. In the texture calculation, the CHM was reclassified into eight gray-level classes, and the variables were calculated as the mean of all the directions using three and six pixel offset values.

Finally, external variables were obtained from different sources. The terrain elevation, slope, aspect, and distance to the nearest road were extracted from topographic map data for every plot. In addition, a site productivity index was obtained from a forestry database. These variables helped us to quantify the effects of different growth conditions and management histories that occurred within the sample area.

#### 2.3. Modeling techniques

When using GLMs, the data are assumed to follow a prespecified distribution, which must be defined in the model fitting. As our data consist of counts, the normal distribution is not a good option, but Poisson or NB distribution models can be considered instead. Firstly, we tested Poisson regression as the simplest alternative, but the model diagnostics indicated an inadequate fit due to overdispersion. NB regression performed better, as it extends the Poisson regression by adding a dispersion parameter that is independent of the mean (Affleck 2006). However, if the data are clearly zero inflated, then the NB model may not be flexible enough. In such cases, ZINB models may be appropriate.

To specify a GLM, we have to define both the probability density function of the random variable and the link function, which merge the regression coefficients and the covariates of the model. The NB distribution's probability density function can be written as

(1) 
$$P(Y = y) = \frac{\Gamma(y + \theta^{-1})}{\Gamma(y + 1)\Gamma(\theta^{-1})} \left(\frac{\theta^{-1}}{\theta^{-1} + \lambda}\right)^{\theta^{-1}} \left(\frac{\lambda}{\theta^{-1} + \lambda}\right)^{y}$$

where  $\lambda$  is the mean parameter,  $\theta$  is the dispersion parameter, and y is the random variable. The NB model can be seen as a generalization of the Poisson model in which the variance (characterized by the dispersion parameter,  $\theta$ ) can differ from the expected value. Thus, the mean of the NB distribution is  $\lambda$ , as in the Poisson model, but the variance is  $\lambda + \theta \lambda^2$  (Zhang et al. 2012). When  $\theta$  is zero, the NB distribution simplifies to the Poisson distribution. NB regression is based on fitting the NB distribution (eq. 2) to the data, so that the mean parameter  $\lambda$  (equal to  $\mu$ ) is estimated as a function of the predictor variables. Furthermore, the dispersion parameter  $\theta$  is estimated as a constant. To obtain  $\lambda$  from the regression equation, an inverse transformation of the selected link function must be applied. The default is the logarithmic link function, i.e., the  $\lambda$  is obtained as exp( $\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}$ ).

ZINB regression extends the NB regression by introducing an additional binomial component that enables estimation that is even more flexible. ZINB models are appropriate when the data can be seen to be generated by two stochastic processes, one of which only contributes zeroes (Welsh et al. 1996). Our data can be seen to have such an interpretation, in that some of the plots are candidates for having large trees, whereas others could immediately be seen as not having large trees. The minimum field-measured height for a large tree was 7 m, and thus, we hypothesized that the plots where all of the echo heights were smaller than this limit would



Fig. 4. Two examples from the ALS strips with some large trees (white triangles) located in field. Darker points indicate high laser echoes.

contribute only zeroes. The remaining plots either could or could not have large trees in them.

In accordance with Zhang et al. (2012), the probability mass function of the ZINB mixture model with binomial and NB components is

(2) 
$$P(Y = y) = \begin{cases} \pi + (1 - \pi) \left(\frac{\theta^{-1}}{\theta^{-1} + \lambda}\right)^{\theta^{-1}} & y = 0\\ (1 - \pi) \frac{\Gamma(y + \theta^{-1})}{\Gamma(y + 1)\Gamma(\theta^{-1})} \left(\frac{\theta^{-1}}{\theta^{-1} + \lambda}\right)^{\theta^{-1}} \left(\frac{\lambda}{\theta^{-1} + \lambda}\right)^{y} & y > 0 \end{cases}$$

where  $\pi$  is a parameter representing the probability that the observation is drawn from the zero-inflation component,  $\lambda$  is the mean parameter,  $\theta$  is the dispersion parameter, and *y* is the random variable. When  $\theta = 0$ , the model is similar to a zero-inflated Poisson model. In a ZINB model, both  $\lambda$  and  $\pi$  are estimated as linear combinations of the covariates. If necessary, different covariates can be used for  $\lambda$  and  $\pi$ . For  $\pi$ , the logistic link function is typically used. In this case, the expected value of the whole mixture distribution (i.e., the model prediction) is not directly one of the parameters but depends on both  $\pi$  and  $\lambda$ , i.e.,  $E[Y] = \mu = \lambda(1 - \pi)$  (Affleck 2006).

We tested both the NB and ZINB models but found that the ZINB models provided a slightly better prediction accuracy. Thus, we employed ZINB as the main model, but we also report the best NB model for comparison purposes because NB models are simpler. The ZINB models were fitted using the pscl package (Zeileis et al. 2008) in R statistical software (R Core Team 2013). The NB models were also fitted in R, using the glm.nb function from the external library MASS (Venables and Ripley 2002). The Akaike information criterion (AIC) (Akaike 1974) was calculated based on the estimated likelihoods and the number of model parameters so that smaller values indicate a better model fit.

# 2.4. Variable selection and accuracy assessment

First, we created separate models based on each group of predictors reported in Table 1. An exhaustive search was performed to select the two or three best predictor variables from each group. These variables were candidates for the general model, and an exhaustive search was again made to find the two or three most influential variables for the general model. These variables were fixed into the model, and manual experimentation was finally applied to test different supplementary variables and their transformations for the final model.

The different models were evaluated based on a leave-strip-out cross validation. In leave-strip-out cross validation, the five original strips were split into nine substrips based on the terrain (52–204 plots per substrip). Each substrip was left out of the data one at a time, and predictions for that strip were produced using a model that was fitted using the remaining eight substrips. In this way, the predictions were not influenced by plots in the geographical vicinity. Error matrices were used to quantify the model accuracy. Each tree count with 0–8 trees had its own category, but all of the plots with >8 large trees were pooled so that the error matrix had 10 categories.

The main criterion for the prediction accuracy was the weighted kappa coefficient (wк) (Cohen 1968). The original kappa coefficient was developed for nominal variables, and therefore, it is not ideal for ordinal variables such as counts because it assigns equal weight to small and large errors (Næsset 1996). The weighted kappa assigns observations on the diagonal of the error matrix weight one, one step away from the diagonal; weight two, two steps away from the diagonal, etc. - so that large distances from the diagonal also have large weights. Apart from this change, the interpretation is similar to the ordinary kappa, i.e., a value of zero indicates a random classification and a value of one a perfect classification. Ordinary kappa (κ), overall accuracy (OA, the percent of all observations classified correctly), producer's accuracy (PA, the percent of correct classifications in each class observed in the field), and user's accuracy (UA, the percent of correct classifications in each predicted class) were also computed.

# 3. Results

Visual inspection indicated that it was possible to detect many large crowns directly from the point cloud (Fig. 4) because the echo density was larger than seen in most area-based applications. However, it was common that even though there was a large crown visible in the data, no large trees were measured in the same field location. An obvious reason could be that the diameter of the tree was less than 35 cm, and therefore, it was not included in our field data. Secondly, there was a 3 year difference in time between the ALS data acquisition (2006) and the field inventory (2009). Thus, harvests and windthrow could have removed some trees after the ALS acquisition, although we saw no evidence that any such large-scale events had taken place when we undertook the fieldwork. Sometimes there were offsets in the field-measured

**Table 2.** Best models by variable group.

Variable group	Predictors	AIC	OA	к	Wκ
Height distribution	$\max_{1}, d_{101}, d_{801}$	1924	61.8	0.33	0.46
Horizontal (0.5 m)	Amean <sub>ap12.5 h</sub> , Hmean <sub>cp17.5 h</sub> , Amean <sub>ap17.5 h</sub>	2081	57.4	0.29	0.37
Horizontal (1 m)	Amean <sub>ap12,51</sub> , Hmean <sub>cp12,51</sub> , euler <sub>101</sub>	2088	58.5	0.29	0.37
Texture	$cor_{3 h}$ , $hom_{3 h}$ , $hom_{6 h}$	2207	54.7	0.24	0.31
External	elev, slope	2414	35.6	0.11	0.14
All variables (ZINB model)	$d_{6f}, d_{11}, h_{901}, \log(\text{elev}), \text{Amean}_{ap12.5 h}$	1845	65.8	0.37	0.55
All variables (NB model)	$d_{6f}, d_{1l}, h_{90l}, \log(\text{elev}), \operatorname{Amean}_{ap12.5 h}$	1907	63.6	0.33	0.52

**Note:** Variable abbreviations are explained in Table 1. AIC, Akaike information criterion; OA, overall accuracy;  $\kappa$ , kappa coefficient; w $\kappa$ , weighted kappa coefficient; ZINB, zero-inflated negative binomial; NB, negative binomial.

tree positions and the crown echoes which related to them, which might have been caused either by errors in the GNSS positioning or by tilted stems. Because of these differences, high echoes could sometimes be located outside the plot on which a large tree was positioned in the field or vice versa. Some of these obvious errors were corrected manually, but some may still have remained undetected even though all plots were visually inspected. These inconsistencies between the data sets contributed to the somewhat low accuracy in the prediction of the plot-wise tree counts.

The best ZINB models obtained using variables from each group (Table 1) are shown separately in Table 2. The results show that the standard height distribution variables provided better accuracy ( $w\kappa = 0.46$ ) than any of the other variable groups. Horizontal variables such as the number of CHM regions above a specific height threshold and their mean area had a good correspondence with the visual impression of the point cloud (Fig. 5) but did not reach as good an accuracy as height distribution variables ( $w\kappa = 0.37$  for both resolutions). The texture variables only reached  $w\kappa = 0.31$ . Of the external variables, plot elevation and slope were slightly correlated with the presence of large trees. The prediction accuracy using only these variables remained poor ( $w\kappa = 0.14$ ), but this is reasonable as these variables are not directly related to tree size.

The accuracy assessment of the final ZINB model including all variables is shown in Table 2, and its parameter estimates are shown in Table 3. The importance of the height distribution variables was also reflected in the final model, which contained five predictor variables: two density variables ( $d_{6f}$  and  $d_{1l}$ ), one height percentile  $(h_{901})$ , the logarithm of the plot elevation (elev), and the mean area of all horizontal patches obtained using a 12.5 m height threshold (Amean<sub>ap12.5h</sub>). The summary information about these variables is given in Table 4. This model reached  $w_{K} = 0.55$ , which is clearly larger than the values obtained with any of the single variable groups. Notably, the height distribution variables selected in the final model were different from the variables selected in the model that had only height distribution variables, i.e., they functioned better with elev and  $\rm Amean_{ap12.5h}$  than when alone. Some of the ZINB model's count model variables were not significant in the zero-inflation part of the model, where only  $d_{6f}$ ,  $h_{901}$ , and log(elev) were retained.

The best NB model is also reported in Tables 2 and 3. Based on the same variables as used in the ZINB count model, it reached a  $w_{\rm K} = 0.52$ . Thus, it can be seen that using the ZINB model with an additional zero-inflation component improved the  $w_{\rm K}$  by three percentage points. Furthermore, three variables were needed in the zero-inflation part to gain an improvement over the NB model, indicating that the zero-inflation was not directly related to tree height as we assumed. If only one height variable was included in the zero-inflation model, the  $w_{\rm K}$  decreased to 0.51, i.e., there was no improvement over the simpler NB model.

There were many alternative models that provided nearly the same accuracy, but two variables were present in all of them, namely the  $h_{901}$  height percentile and plot elevation. Large-diameter trees are obviously more common in stands with large heights, so the inclusion of a height quantile was expected. The marginal distribution of log(elev) was nonlinear and peaked at the elevation

of 760 m, which is slightly above the mean of all plots. The reason for this could be that the lower elevations have been harvested more frequently and that trees at high elevations seldom reach the lower limit of 35 cm diameter. Leaving the elevation out of the final model decreased the w $\kappa$  to 0.43, which is smaller than what was obtained using only height distribution variables. Thus, the influence of elevation on the overall variable selection was considerable.

Because the coefficients shown in Table 3 were normalized by subtracting the mean and dividing by the standard deviation of the original variables, the larger coefficients indicate a larger weight in the final prediction. It can be seen that the only height variable present in the models had the largest normalized coefficient in both NB and ZINB count models, i.e., the presence of large-diameter trees was related to tree height. The two density percentiles  $d_{6f}$  and  $d_{11}$  had opposite signs in both models. These variables were correlated (R = 0.75), but  $d_{1f}$  was calculated using first-and-only echoes and  $d_{61}$  was calculated using last-and-only echoes. First echo variables typically represent the canopy surface, and last echo variables represent the canopy somewhat below its surface. Both variables increase with the canopy density, but if the pulse penetration depth is small, then  $d_{11}$  will have a large value in relation to  $d_{6f}$ . Thus, the interpretation of these opposite signs could be that large-diameter trees occurred in places where the penetration into the canopy was small. This corresponds well with typical old-growth pine stands where the trees have rounded tops with dense needle layers.

The final variable in the model was mean area of all ground and canopy patches using a 12.5 m threshold value and 0.5 m CHMs (Amean<sub>ap12.5h</sub>). It had a negative sign in both count models, i.e., a small mean patch area indicated more large trees. This variable had its maximum value in plots that had no tree tops reaching this height (12.5 m) and a smaller value if there were many of them. Omitting Amean<sub>ap12.5h</sub> from the model only decreased the w $\kappa$  to 0.52, meaning that the estimation could also be done without the horizontal variables.

The error matrices for the ZINB and NB models are shown in Tables 5 and 6, respectively. It is easy to observe that although there was a clear correlation between the observed and predicted values, the individual estimates frequently had fairly large errors. For example, there was a plot that was predicted to have at least eight large-diameter trees, but in reality, there was only one largediameter tree. Similarly, two plots that had at least nine large-diameter trees were predicted as having only one or two large-diameter trees. The standard errors of the model coefficients were also fairly large (Table 3).

# 4. Discussion

The results of the area-based estimation indicated that it is possible to predict the occurrence of large trees in a plot, but the individual predictions are often inaccurate. Landis and Koch (1977) characterized  $\kappa$  values ranging from 0.4 to 0.6 as a moderately successful classification, and our observed w $\kappa$  = 0.55 falls within this range. **Fig. 5.** An example of horizontal CHM variables. White color represents a CHM area above a 12.5 m height threshold. A large crown with maximum ALS height = 15.1 m is clearly visible with both 0.5 m and 1.0 m resolutions, but no tree with a DBH > 35 cm was recorded in the field for this location.



**Table 3.** Parameter estimates of the final negative binomial (NB) and zero-inflated negative binomial (ZINB) models.

	ZINB mode	el	NB model			
	Estimate	SE	Estimate	SE		
Count mode	1					
Intercept	-0.8201	0.089	-1.2084	0.070		
d <sub>6f</sub>	-0.4666	0.076	-0.6745	0.071		
d <sub>11</sub>	0.5973	0.060	0.6123	0.062		
h <sub>901</sub>	1.2210	0.093	1.2755	0.092		
log(elev)	0.8997	0.074	0.7402	0.069		
Amean <sub>ap12.5</sub>	-0.4215	0.103	-0.6915	0.095		
$\log(\theta)$	1.9536	0.344	1.5385	0.215		
Zero-inflatio	n model					
Intercept	-5.4574	1.218				
d <sub>6f</sub>	1.9011	0.366				
h <sub>901</sub>	-2.5997	1.024				
log(elev)	5.6865	1.503				

**Note:** The ZINB model has two components: the count model that uses the negative binomial pdf and logarithmic link and the zero-inflation model that uses the binomial pdf with a logit link.  $\theta$  is the model dispersion parameter. The response variables were normalized to follow the N(0,1) distribution. All variables were statistically significant at p < 0.05. SE, standard error.

**Table 4.** Minimum, mean, maximum, and standard deviation (SD) values for the variables that appear in the models.

	Minimum	Mean	Maximum	SD
d <sub>6f</sub> (%)	0.0	18.5	65.9	14.9
$d_{11}$ (%)	0.0	22.9	65.8	14.0
h <sub>901</sub> (m)	1.4	9.8	21.6	4.2
elev (m)	492	725	989	95.4
Amean <sub>ap12.5</sub> (m <sup>2</sup> )	41.3	709	1156	473

It is likely that the inaccuracy of individual predictions is mainly due to the problems mentioned above. We consider that the trees that were just below the 35 cm diameter limit were the main source of error, because they could not be separated from the trees that were just above the limit. On the other hand, measuring and positioning all trees above, for example, a 30 cm limit would have increased the amount of fieldwork and thus decreased the size of the modeling data set. A later check of any uncertain trees would have helped, but this was not possible in our case. A better alternative might have been an interpretation of possible large trees based on the CHM before the field campaign and then measurement of these trees during the campaign. Especially on higher elevations, there were some trees that had very small heights in relation to their DBH (Fig. 3). It is therefore likely that introducing a height criterion into the definition of large trees would have improved the prediction accuracy, because such outliers would have been excluded. However, we only measured height for every fifth tree, so we were unable to test a height criterion in practice. Other possible error sources included the occasional inaccuracy of tree positioning and the 3 year time difference between the ALS data acquisition and the fieldwork.

Standard ALS-based canopy height distribution variables seem to suffice as predictor variables, as only one CHM-based variable was selected into the final model, and its influence on the predictions was fairly small. The 0.5 and 1 m resolution CHM variables provided equally good results (Table 2). External topographic variables may provide useful additional information. In our case, the plot elevation was a crucial variable in the model and leaving it out resulted in a major decrease in accuracy. Other external variables did not improve the model.

Although the overall accuracy obtained with regressions was not very high, the models were capable of taking into account the large amount of zeroes and uncovering the relationships between the variables, despite the noisy nature of the data. It is likely that the low degree of accuracy is related to the data itself and not the modeling technique. ZINB models performed slightly better than the simpler NB models, but based on the available data, we cannot be certain if this difference is real or only a result of overfitting the model. The results above were derived from a leave-strip-out cross validation, and the w $\kappa$  values were typically a few percent smaller than in the training data. This decrease was considerably smaller than what we observed while testing the *k*- nearest neighbor method with random forest distance metric in an initial phase of the analyses.

Affleck (2006) recommended the use of NB models without a separate zero-inflation component, as they are simpler to interpret and less prone to overfitting. However, Eskelson et al. (2009) assumed that ZINB models might be better in cases where the data showed evidence of two separate distributions, which was not the case in their data of snag abundance. Zhang et al. (2012) confirmed this assumption in their study on tree recruitment prediction. Our results indicated that the ZINB regression was better than the other parametric or nonparametric methods that were initially tested, but the differences were fairly small. It would also have been possible to convert the tree counts into trees per hectare and use linear regression analysis, but we consider count modeling methods to be better justified, as any such transformations would not change the discrete nature of the data.

	Observed											
	0	1	2	3	4	5	6	7	8	9+	Sum	UA (%)
Predicted												
0	609	62	10	3	0	0	0	0	0	0	684	89.0
1	109	80	44	10	7	4	0	0	0	1	255	31.4
2	13	18	25	12	7	6	3	1	0	1	86	29.1
3	5	5	11	10	7	4	3	1	0	0	46	21.7
4	1	2	1	2	2	1	3	1	0	0	13	15.4
5	0	0	3	0	1	0	0	1	0	0	5	0.0
6	0	0	1	2	0	2	0	0	0	1	6	0.0
7	0	0	0	2	1	0	0	2	1	1	7	28.6
8	0	0	0	0	0	1	0	0	0	0	1	0.0
9+	0	1	0	0	1	0	0	1	1	2	6	33.3
Sum	737	168	95	41	26	18	9	7	2	6	1109	
PA (%)	82.6	47.6	26.3	24.4	7.7	0.0	0.0	28.6	0.0	33.3		OA = 65.8

 Table 5. Error matrix of the final zero-inflated negative binomial (ZINB) model (Table 3) derived from leave-strip-out cross validation.

Note: The weighted kappa coefficient for the model was 0.55. OA, UA, and PA, overall accuracy, user's accuracy, and producer's accuracy, respectively.

Table 6. Error matrix of the final negative binomial (NB) model (Table 3) derived from leave-strip-out cross validation.

	Observed											
	0	1	2	3	4	5	6	7	8	9+	Sum	UA (%)
Predicted												
0	601	70	12	2	0	0	0	0	0	0	685	87.7
1	112	71	43	11	9	3	1	0	0	2	252	28.2
2	16	19	20	14	7	8	2	1	0	0	87	23.0
3	5	5	14	8	4	3	3	1	0	0	43	18.6
4	2	2	3	2	2	1	3	0	0	0	15	13.3
5	1	0	3	1	3	0	0	2	1	0	11	0.0
6	0	0	0	2	0	2	0	0	0	2	6	0.0
7	0	0	0	0	0	0	0	2	0	0	2	100.0
8	0	1	0	1	0	1	0	0	0	1	4	0.0
9+	0	0	0	0	1	0	0	1	1	1	4	25.0
Sum	737	168	95	41	26	18	9	7	2	6	1109	
PA (%)	81.5	42.3	21.1	19.5	7.7	0.0	0.0	28.6	0.0	16.7		OA = 63.6

Note: The weighted kappa coefficient for the model was 0.52. OA, UA, and PA, overall accuracy, user's accuracy, and producer's accuracy, respectively.

One of the advantages of the area-based method (which in statistical sampling terminology is a small-area estimation technique based on model predictions) compared with sample-based inventories that do not use auxiliary data with complete geographical coverage is the ability to create maps. Maps depicting the frequency of large-diameter trees should be valuable to guide, for example, timber procurement efforts, because they would guide the user to locations with large trees, although the exact number of trees would be uncertain. If the objective is to establish a reliable inventory of large-diameter trees, then the model predictions could be used as prior information to guide the placement of a limited number of sample plots into locations where large-diameter trees are likely to be found (Pesonen et al. 2009).

It can be argued that predicting the mean diameter of each stand and selecting the sites where predictions indicate trees larger than 35 cm in diameter could be used to find stands with large-diameter trees. However, individual large-diameter trees can also exist in stands in which the mean diameter is smaller than this limit. If the whole diameter distribution could be precisely predicted, then separate models for large trees would not be needed. The diameter distributions derived from remotely sensed data tend to be rough approximations and not very accurate for extreme values (for example, large-diameter trees). Therefore, to directly model the variable of interest is a suitable and reliable alternative.

Finally, the use of individual tree detection methods (e.g., Vauhkonen et al. 2012) would probably enable a better estimation of the dominant tree layer features than an area-based approach. Thus far, high pulse density (>5 m<sup>-2</sup>) ALS data acquisitions needed for the reliable detection of individual tree crowns have usually been viewed as too expensive for most practical applications. The use of ALS sensors that utilize multiple-pulses-in-air technology enables the acquisition of denser point clouds at reasonable costs, thus making individual tree detection applications more feasible. Nevertheless, in most cases, managers will probably choose to minimize the cost, which means that low-density ALS data sets will be more commonly available than data which is suitable for single-tree segmentation. In such cases, these methods can be applied in the mapping of, for example, large-diameter trees, snags, or other count parameters of interest. Maps of sites with large-diameter trees should prove valuable for managers that need to locate large trees either for protection or harvesting.

In conclusion, our results show that GLM methods are suitable for the area-based estimation of large-diameter tree counts. However, the predictions achieved were approximate and could have fairly large errors, which were mostly related to problems in the field data. ZINB models showed a better fit than NB models, but the difference was fairly small. ALS height quantiles together with plot elevation sufficed for prediction, but horizontal canopy variables may also improve the models.

# Acknowledgements

L. Korhonen was supported by the strategic funds of the University of Eastern Finland and the Academy of Finland. C. Salas was supported by Dirección de Investigación y Desarrollo, Universidad de La Frontera, Chile, DIUFRO Project No. DI11-0018. The ALS data acquisition was funded by the Research Council of Norway, and we thank Blom Geomatics, Norway, for its collection and processing. We also thank Dr. Lauri Mehtätalo (University of Eastern Finland) and the anonymous reviewers for their helpful comments.

# References

- Affleck, D.L.R. 2006. Poisson mixture models for regression analysis of standlevel mortality. Can. J. For. Res. 36(11): 2994–3006. doi:10.1139/x06-189.
- Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control, 19(6): 716–723. doi:10.1109/TAC.1974.1100705.
- Burrascano, S., Keeton, W.S., Sabatini, F.M., and Blasi, C. 2013. Commonality and variability in the structural attributes of moist temperate old-growth forests: a global review. For. Ecol. Manage. 291: 458–479. doi:10.1016/j.foreco.2012.11. 020.
- Cohen, J. 1968. Weighed kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol. Bull. 70(4): 213–220. doi:10. 1037/h0026256.
- Consul, P.C. 1989. Generalized Poisson distributions: properties and applications. Vol. 99. Statistics: textbooks and monographs. Marcel Dekker Inc., New York.
- Eskelson, B.N.I., Temesgen, H., and Barrett, T.M. 2009. Estimating cavity tree and snag abundance using negative binomial regression models and nearest neighbor imputation methods. Can. J. For. Res. 37(9): 1749–1765. doi:10.1139/ X09-086.
- Famoye, F., and Singh, K.P. 2006. Zero-inflated generalized Poisson regression model with an application to domestic violence data. J. Data Sci. 4: 117–130. Franklin, J.F., Shugart, H.H., and Harmon, M.E. 1987. Tree death as an ecological
- process. BioScience, **37**: 550–556. doi:10.2307/1310665. Ganey, J.L., and Vojta, S.C. 2005. Change in snag populations in Northern Ari-
- Ganey, J.L., and Vojca, S.C. 2005. Change in snag populations in Northern Arzona mixed-conifer and ponderosa pine forests 1997–2002. For. Sci. 51(5): 396–405.
- Gaveau, D.L.A., and Hill, R.A. 2003. Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data. Can. J. Remote Sens. **29**: 650–657. doi:10.5589/m03-023.
- Gobakken, T., and Næsset, E. 2004. Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. Scand. J. For. Res. 19(6): 529–542. doi:10.1080/02827580410019454.
- Gobakken, T., and Næsset, E. 2008. Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. Can. J. For. Res. 38(5): 1095–1109. doi:10.1139/X07-219.
- Gobakken, T., Næsset, E., Nelson, R., Bollandsås, O.M., Gregoire, T., Ståhl, G., Holm, S., Ørka, H.O., and Astrup, R. 2012. Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. Remote Sens. Environ. **123**: 443–456. doi:10.1016/j.rse.2012.01.025.
- Godal, J.B. 1996. Tre til laft og reis: gamle hus fortel om materialbruk ("Timber for log and frame constructions: old houses tell the story of material use"). Landbruksforlaget, Oslo, Norway.
- Haralick, R.M., Shanmugam, K., and Dinstein, I. 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. 3(6): 610–621. doi:10.1109/ TSMC.1973.4309314.
- Ilvessalo, Y. 1950. On the correlation between the crown diameter and the stem of the trees. Comm. Inst. For. Fenn. **38**(2).
- Landis, R.J., and Koch, G.G. 1977. The measurement of observer agreement for categorical data. Biometrics, **33**: 159–174. doi:10.2307/2529310.
- Lutz, J.A., Larson, A.J., Swanson, M.E., and Freund, J.A. 2012. Ecological importance of large-diameter trees in a temperate mixed-conifer forest. PLoS One, 7(5): e36131. doi:10.1371/journal.pone.0036131.
- Maltamo, M., and Packalen, P. 2014. Species-specific management inventory in Finland. *In* Forestry applications of airborne laser scanning. Concepts and case studies. Managing forest ecosystems 27. *Edited by* M. Maltamo, E. Næsset, J. Vauhkonen. Springer Science + Business Media Dordrecht, Netherlands.
- Maltamo, M., Eerikäinen, K., Packalén, P., and Hyyppä, J. 2006. Estimation of stem volume using laser scanning based canopy height metrics. Forestry, 79: 217–229. doi:10.1093/forestry/cpl007.
- Maltamo, M., Kallio, E., Bollandsås, O.M., Næsset, E., Gobakken, T., and Pesonen, A. 2014. Assessing dead wood by airborne laser scanning. In Forestry applica-

tions of airborne laser scanning. Concepts and case studies. Managing forest ecosystems 27. *Edited by* M. Maltamo, E. Næsset, and Vauhkonen, J. Springer Science + Business Media Dordrecht, Netherlands.

- Martinuzzi, S., Vierling, L.A., Gould, W.A., Falkowski, M.J., Evans, J.S., Hudak, A.T., and Vierling, K.T. 2009. Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. Remote Sens. Environ. 113: 2533–2546. doi:10.1016/j.rse.2009.07.002.
- McCulloch, C.E., and Searle, S.R. 2001. Generalized, linear, and mixed models. Wiley-Interscience, New York, U.S.A.
- McGarigal, K., and Marks, B.J. 1995. FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure. General Technical Report PNW/ Pacific Northwest Research Station.
- Müller, J., and Vierling, K. 2014. Assessing biodiversity by airborne laser scanning. In Forestry applications of airborne laser scanning. Concepts and case studies. Managing forest ecosystems 27. Edited by M. Maltamo, E. Næsset, and J. Vauhkonen. Springer Science + Business Media Dordrecht, Netherlands.
- Mustonen, J., Packalén, P., and Kangas, A. 2008. Automatic segmentation of forest stands using a canopy height model and aerial photography. Scand. J. For. Res. 23(6): 534–545. doi:10.1080/02827580802552446.
- Næsset, E. 1996. Use of the weighted Kappa coefficient in classification error assessment of thematic maps. Int. J. Geogr. Inf. Syst. 10: 591–604. doi:10.1080/ 02693799608902099.
- Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80(1): 88–99. doi:10.1016/S0034-4257(01)00290-5.
- Næsset, E. 2004. Practical large-scale forest stand inventory using a smallfootprint airborne scanning laser. Scand. J. For. Res. 19: 164–179. doi:10.1080/ 02827580310019257.
- Næsset, E. 2014. Area-based inventory in Norway from innovation to an operational reality. *In* Forestry applications of airborne laser scanning. Concepts and case studies. Managing forest ecosystems 27. *Edited by* M. Maltamo, E. Næsset, and J. Vauhkonen. J. Springer Science + Business Media Dordrecht, Netherlands.
- Oliver, C.D. 1980. Forest development in North America following major disturbances. For. Ecol. Manage. 3: 153–168. doi:10.1016/0378-1127(80)90013-4.
- Pesonen, A., Leino, O., Maltamo, M., and Kangas, A. 2009. The comparison of field sampling methods and the use of airborne laser scanning as auxiliary information for assessing coarse woody debris. For. Ecol. Manage. 257: 1532– 1541. doi:10.1016/j.foreco.2009.01.009.
- Pippuri, I., Kallio, E., Maltamo, M., Peltola, H., and Packalén, P. 2012. Exploring horizontal area-based metrics to discriminate the spatial pattern of trees and need for first thinning using airborne laser scanning. Forestry, 85(2): 305– 314. doi:10.1093/forestry/cps005.
- Popescu, S.C., Wynne, R.H., and Nelson, R.F. 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Can. J. Remote Sens. 29(5): 564–577. doi:10.5589/m03-027.
- R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org/.
- Salas, C., LeMay, V., Núñez, P., Pacheco, P., and Espinosa, A. 2006. Spatial patterns in an old-growth Nothofagus obliqua forest in south-central Chile. For. Ecol. Manage. 231(1-3): 38-46. doi:10.1016/j.foreco.2006.04.037.
- Schabenberger, O., and Pierce, F.J. 2002. Contemporary statistical models for the plant and soil sciences. CRC Press, Boca Raton, Florida, USA.
- Vauhkonen, J., Ene, L., Gupta, S., Heinzel, J., Holmgren, J., Pitkänen, J., Solberg, S., Wang, Y., Weinacker, H., Hauglin, K.M., Lien, V., Packalén, P., Gobakken, T., Koch, B., Næsset, E., Tokola, T., and Maltamo, M. 2012. Comparative testing of single-tree detection algorithms under different types of forest. Forestry, 85(1): 27–40. doi:10.1093/forestry/cpr051.
- Venables, W.N., and Ripley, B.D. 2002. Modern applied statistics with S. 4th edition. Springer.
- Welsh, A.H., Cunningham, R.B., Donnelly, C.F., and Lindenmayer, D.B. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecol. Model. 88: 297–308. doi:10.1016/0304-3800(95)00113-1.
- Wing, B.M., Ritchie, M.W., Boston, K., Cohen, W.B., and Olsen, M.J. 2015. Individual snag detection using neighborhood attribute filtered airborne lidar data. Remote Sens. Environ. 163: 165–179. doi:10.1016/j.rse.2015.03.013.
- Zeileis, A., Kleiber, C., and Jackman, S. 2008. Regression models for count data in R. J. Stat. Softw. 27(8): 1–25. doi:10.18637/jss.v027.i08.
- Zhang, X., Lei, Y., Cai, D., and Liu, F. 2012. Predicting tree recruitment with negative binomial mixture models. For. Ecol. Manage. 270: 209–215. doi:10. 1016/j.foreco.2012.01.028.