2 June 2017

Kriging prediction of stand level forest information using mobile laser scanning data adjusted for non-detection

Svetlana Saarela[1,3*], Johannes Breidenbach[1], Pasi Raumonen[2], Anton Grafström[3], Göran Ståhl[3], Mark J. Ducey[4] & Rasmus Astrup[1]

[1]*Norwegian Institute of Bioeconomy Research, NIBIO P.O. Box 115, NO-1431 Ås, Norway.*

[2]*Laboratory of Mathematics, Tampere University of Technology, TUT P.O. Box 553, FI-33101 Tampere, Finland.*

[3]*Department of Forest Resource Management, Swedish University of Agricultural Sciences, SLU Skogsmarksgränd, SE-90183 Umeå, Sweden.*

[4]*Department of Natural Resources and the Environment, University of New Hampshire, James Hall, 56 College Road, Durham, New Hampshire 03824, U.S.A.*

Abstract

This study presents an approach for predicting stand level forest attributes utilizing mobile laser scanning (MLS) data collected as a non-probability sample. Firstly, recordings of stem density were made at point locations every 10th metre along a subjectively chosen MLS track in a forest stand. Secondly, Kriging was applied to predict stem density values for the centre point of all grid-cells in a 5×5 m lattice across the stand. Thirdly, due to non-detectability issues a correction term was computed based on distance sampling theory. Lastly, the mean stem density at stand level was predicted as the mean of the point-level predictions multiplied

[*]*Corresponding author:* Svetlana Saarela
Department of Forest Resource Management, Swedish University of Agricultural Sciences, SLU Skogsmarksgränd, SE-90183 Umeå, Sweden
svetlana.saarela@slu.se

with the correction factor, and the corresponding variance was estimated. Many factors contribute to the uncertainty of the stand level prediction; in the variance estimator we accounted for the uncertainties due to Kriging prediction and due to estimating a detectability model from the laser scanning data. The results from our new approach were found to correspond fairly well with estimates obtained using field measurements from an independent set of 54 circular sample plots. The predicted number of stems in the stand based on the proposed methodology was 1366 with a 12.9 % relative standard error. The corresponding estimate based on the field plots was 1677, with a 7.5 % relative standard error.

Keywords: Covariogram; Detectability function; Forest management; Model-based inference.

1. Introduction

Traditionally, stand level forest inventories have been based on field surveys where the surveyor allocates sample plots either subjectively or by random sampling to assess key stand characteristics such as stem density, age, and growing stock volume (Kangas and Maltamo 2006). This information typically is acquired for supporting forest management decisions. However, due to high costs of field data collection, substantial research has been attributed to replacing, or improving the cost-efficiency of, field surveys using remotely sensed data during the last decades. Today, laser scanning has become a widely-used technique to support forest inventories (*e.g.,* Næsset 2002, Boudreau *et al.,* 2008; Andersen *et al.,* 2011). For practical applications, airborne laser scanning (*e.g.,* Naesset 2002, Wulder *et al.,* 2012; Neigh *et al.,* 2013) is currently the most common approach. For variables, such as stand height, volume and biomass it can provide stand level estimates that are as precise as, or even more precise, than those obtained from traditional stand level inventories (*e.g.,* Hyyppä & Hallikainen, 1996; Hyyppä & Inkinen, 1999; Næsset, 2002; Ørka *et al.,* 2016). Several studies also point at the potential of laser scanning for providing auxiliary data for improving large-area forest surveys (*e.g.,* Gobakken *et al.,* 2012; Saarela *et al.,* 2015; Ene *et al.,* 2017).

Also, terrestrial laser scanning (TLS) has been investigated in many studies (*e.g.,* Watt *et al.,* 2005; Ducey *et al.,* 2013; Ducey & Astrup, 2013; Liang *et al.,* 2016; Vaaja *et al.,* 2016; Olofsson & Holmgren, 2016) and it has the potential to become a competitive alternative or adjunct to traditional sample plot inventories. With TLS, currently, a normal procedure is to mount a laser scanning device at several locations in a stand whereby detailed tree level information can be obtained from the measurements. Ducey *et al.* (2013) report that stand conditions and scanner attributes affect how well tree stems can be identified and measured and that none of the scanners evaluated in their study provided sufficiently reliable diameter measurements to substitute field measurements. On the other hand, Olofsson *et al.* (2014) suggest a method which can measure stem diameters with relative root-mean-square error (RMSE) of 14%. However, with the laser scanner mounted at fixed positions (i.e. single scan mode) there is a problem that some of the trees are hidden by trees located close to the measurement device and thus are not detected. To overcome potentially negative bias due to this, Ducey & Astrup (2013) and Astrup *et al.* (2014) developed procedures to adjust for non-detection and thus make the estimators (almost) unbiased.

Terrestrial laser scanning is rapidly developing and modern measurement devices are becoming smaller, cheaper and more accurate. They can also be combined with satellite-based positioning systems, so that all the trees in the vicinity of the measurement device automatically receive a position. Terrestrial laser devices need not be stationary during measurements and, thus, mobile laser scanning (MLS) has emerged as a special branch of TLS. In this case the lasers are mounted on all-terrain vehicles (ATVs), on unmanned aerial vehicles (UAVs) (*e.g.,* Jaakkola *et al.,* 2010; Glennie *et al.,* 2013, Forsman *et al.,* 2016), or they may be held by or be attached to a surveyor walking through the forest (e.g., Liang *et al.,* 2014; Rönnholm *et al.,* 2016; Lehtola *et al.,* 2016). For MLS, promising results have been obtained by Rönnholm *et al.* (2016), who evaluated the quality of backpack laser scanning data by comparing them with UAV laser scanning data. Jaakkola *et al.* (2010) performed both automatic and manual tree finding, height determination, and automatic measurement of diameter at breast height (DBH). The tree height bias was 2 cm for manual measurement and -15 cm for automatic measurement. The measurement of DBH was obtained with a root mean square error of 2.1 cm. Forsman *et al.* (2016) reported a root-mean-square error of 14% (3.7 cm) in DBH estimation by MLS data using a line-wise intensity-based clustering method.

With MLS it is not obvious how estimation procedures should be developed in order to produce precise and unbiased estimators at the level of stands (or larger areas). There are at least three main problems that need to be addressed. These are (i) non-detection errors of similar kind as with ordinary TLS, (ii) difficulties to obtain probability samples and thus unbiased estimators, since accessibility (at least for ATVs) typically is limited in many parts of forest stands, and (iii) measurement errors at the level of single trees due to inexact laser-based determination of key features such as height and diameter.

The objective of this study was to develop and evaluate procedures for predicting stand level attributes from MLS data, based on previous experiences with handling non-detection in TLS surveys (Astrup *et al.,* 2014). The developed procedures constitute one way to account for the first two issues outlined above, i.e. adjusting for non-detection errors and handling non-probability samples. The study was performed in a 4.7 ha large study area in southern Norway, based on data collected from a mobile laser scanner mounted on an ATV. The estimates obtained were compared with estimates from an independent set of field sample plots from the same area. We focus on the prediction of stem number of mid- and overstory trees as a characteristic that is relatively directly available from MLS data without a need for further models for assessing characteristics such as stem volume and biomass.

2. Material and method

2.1. Material

2.1.1. Field data

The study area was the 4.7 ha Frydenhaug forest in Ås municipality, approximately 30 km south of Oslo, Norway. The forest is owned by the Norwegian University of Life Sciences and is mainly used for educational and recreational purposes. Frydenhaug is dominated by Norway spruce and Scots pine but is rather heterogeneous with some areas that are dominated by non-indigenous tree species and broadleaves. In 2014 parts of Frydenhaug were thinned resulting in some areas with very low stand densities. Frydenhaug can be viewed as a very small forest or as a large heterogeneous stand; in this work Frydenhaug is viewed as a single heterogeneous stand.

As baseline data for comparison, trees on 54 sample plots on a systematic 20 × 40 m grid were measured in September-October 2015 according to the field protocol of the Norwegian National Forest Inventory for temporary plots (Landsskogtakseringen 2008). The circular sample plots had a size of 250 m^2 (radius of 8.92 m) and all trees with DBH ≥ 5 cm were recorded (but note that for comparing with the MLS survey a 10 cm DBH threshold was applied in the analyses). The measured variables included DBH measured with a calliper, tree height measured using a Vertex hypsometer, and species. The frequency of DBH measurements in the field plots is shown in Figure 1. The basal area weighted mean diameter was found to be 40.2 cm and the stem density 357 trees ha^{-1}.

Figure 1 about here

2.1.2. MLS data

MLS data were acquired on 19 November 2014 with a Trimble MX2 system mounted on an ATV. The system consists of two rotating Dynascan S250 scanning devices with a wavelength of 905 nm, a scanning rate of 20 Hz, and a pulse rate of 36 kHz. The sensor's location is determined by two global navigation satellite system (GNSS) antennas with 220 channels coupled with an inertial navigation unit (Applanix IMU-42). Images from a Ladybug camera system consisting of five 6-megapixel cameras have been used to colorize the point cloud that was derived using Trimble's generic Trident software.

In Figure 2 the location of the study area in Norway, the MLS track, identified tree locations and the field sample plots are presented.

Figure 2 about here

2.2. Method

2.2.1. Overview

An overview of the procedure is provided in Figure 3. The procedure comprises four main components. Firstly, an algorithm developed by Raumonen *et al.* (2015) was applied to recognize individual trees from the MLS data and measure the number of stems within 10 × 50 m sample plots; the width (10 m) was taken along the MLS track and the length (50 m) follows from measuring 25 m on both sides of the MLS track (Figure 4).

Secondly, to overcome the problem that probability sampling was not applied (the ATV track was chosen from the point of view of where the ATV could easily pass while measuring a fair portion of the stand) we applied Kriging (*e.g.,* Cressie, 1990; Thompson, 1992) for predicting stem number at stand level, i.e. model-based inference was applied.

Thirdly, to overcome the problem of some trees being non-detected, distance sampling theory (Buckland *et al.,* 2005; Ducey & Astrup 2013) was applied to estimate a correction factor. Lastly, Kriging predictions corrected for non-detection were averaged across the stand to obtain a stand level prediction of stem density and the corresponding uncertainty. The details of the four main components of the procedure are provided below.

Figure 3 about here

2.2.2. Obtaining tree level and sample plot data from MLS data

Stem bases were automatically extracted from the scanned point cloud using the method presented in Raumonen *et al.* (2015). In this method, the point cloud is partitioned into small subsets and for a layer near the ground the surface normals of the subsets are estimated. Potential stems are identified by selecting groups of subsets whose surface normals are nearly horizontal, indicating a vertical surface. Expanding these groups with neighbouring subsets completes the stems. Then, cylinders are fitted into sections around 1.3 m from the ground to estimate DBH. In this way, the locations of trees and their corresponding DBHs were estimated. The DBH frequency from the MLS measurements is presented in Figure 1. We applied a DBH threshold of 10 cm in the study to focus on mid- and overstory trees, but also to avoid partially unresolved problems of identifying small understory trees with the tree segmentation algorithm. The number of trees identified with the tree segmentation algorithm within the 10 × 50 m plots were counted and the corresponding stem density estimate was assigned to the centre point of each plot (Figure 4). In case a plot was located close to the stand boundary, only the 10 × 25 m subplot located entirely within the stand was used. Thus, stem density values from the tree segmentation algorithm were assigned to 87 points along the MLS track, at 10 m intervals.

Figure 4 about here

2.2.3. Kriging prediction of stem densities

One problem with ATV-based MLS inventories of forest stands is that it would be very difficult to select probability samples and thus to apply design-based inference for estimating target parameters, such as stem density and growing stock volume, without bias. The solution we propose in this study is to allow selection of non-probability samples and use of model-based inference (*e.g.,* Gregoire 1998; McRoberts, 2010; Magnussen, 2015; Ståhl *et al.*, 2016) for predicting the target quantities. With model-based inference we rely on certain modelassumptions. In this case our assumption, in general terms, is that values at non-sampled locations can be adequately predicted through spatial interpolation. Thus, the next step of our approach was to apply Kriging for predicting MLS-derived stem densities for each centre point in a dense (5×5 m) square lattice across the stand. Using Kriging, each cell in the lattice obtains a predicted value through spatial interpolation of the measurements so that the variance of the predictor is minimised. In our study, we applied ordinary Kriging (*e.g.,* Thompson 1992) based on a covariogram model estimated from the 87 observations of MLS-derived stem density along the ATV track. A negative exponential covariogram model, assuming isotropic conditions (i.e. that the spatial autocorrelation is the same in all directions), was applied, i.e.

$$Cov_{y_{(t+d)},y_t}(d) = \beta e^{\left(-\frac{|d|}{\alpha}\right)} \quad (1)$$

Where $Cov_{y_{(t+d)},y_t}(d)$ is the covariance, as a function of distance, between two stem density values $y_{(t+d)}$ and y_t separated by the distance d; α and β are model parameters. A simple covariance estimator was employed in this study

$$Cov_{y_{(t+d)},y_t}(d) = \frac{1}{n_d}\sum_{t=1}^{n_d}\left(y_{(t+d)} - \bar{y}\right)\left(y_t - \bar{y}\right)$$

where the summation is over the distinct pairs of stem density values from locations along the scanning track, that are distanced d apart from each other; and n_d is the number of such pairs. Then, the covariogram model (Eq. 1) was fitted using nonlinear least squares. The statistics of the fitted model are presented in Table 1.

Table 1 about here

The fitted model and the observed data are presented graphically in Figure 5. In the figure it can be seen that the covariance is close to zero at distances greater than 60 m. Using this covariogram we applied Kriging to predict (MLS-derived) stem densities for each centre point in the square lattice. The Kriging predicted value was obtained as a weighted average across the $n_{MLS} = 87$ original observations as

$$\hat{y}_t = \sum_{i=1}^{n_{MLS}} a_i \, y_i \qquad (2)$$

where $\hat{y}_t$ is the predicted stem density at location *t*, y_i is the predicted stem density value at the i^{th} location along the scanning track, and a_i is the corresponding weight. The weights $a_1, a_2, \dots, a_{n_{MLS}}$ are obtained as (*cf.* Cressie 1990)

$$\boldsymbol{a} = \boldsymbol{G}^{-1}\left[\boldsymbol{h} + \boldsymbol{1}\left(\frac{1 - \boldsymbol{1}^T \boldsymbol{G}^{-1} \boldsymbol{h}}{\boldsymbol{1}^T \boldsymbol{G}^{-1} \boldsymbol{1}}\right)\right] \qquad (3)$$

where the vectors $\boldsymbol{a}$ and $\boldsymbol{h}$ are

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n_{MLS}} \end{pmatrix}, \ \boldsymbol{h} = \begin{pmatrix} c_{1t} \\ c_{2t} \\ \vdots \\ c_{\boldsymbol{n_{MLS}}t} \end{pmatrix} \qquad (4)$$

with $c_{1t}, c_{2t}, \dots, c_{\boldsymbol{n_{MLS}}t}$ being covariances between the t^{th} location and the different locations with observations along the scanning track. The covariance values are estimated using the negative exponential covariogram model (Eq. 1). The matrix $\boldsymbol{G}$ is

$$\boldsymbol{G} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n_{MLS}} \\ c_{21} & c_{22} & \cdots & c_{2n_{MLS}} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n_{MLS}1} & c_{n_{MLS}2} & \cdots & c_{n_{MLS}n_{MLS}} \end{pmatrix}_{n_{MLS} \times n_{MLS}} \qquad (5)$$

each element of which is a covariance value between different locations along the track (the diagonal contains variances).

Figure 5 about here

2.2.4. Adjusting for non-detection

In a standard case of applying Kriging for predicting stem density at stand level an average over the lattice level predictions are computed and the corresponding uncertainty assessed (*e.g.*, Cressie 1990; Thompson 1992). (Note that we speak of predicting rather than estimating stem density at stand level, since in model-based inference this quantity is assumed to be a random variable rather than a fixed parameter.) However, in the MLS-based approach several trees remain non-detected within the rectangular sample plots mainly because they are at least partly hidden by other trees. Thus, there is a need to adjust for non-detection to obtain an unbiased predictor of stand level stem density (in the model-based sense). We applied distance sampling theory (Buckland *et al.,* 2005) to compute a correction factor, which was subsequently applied to all the MLS-derived values that were predicted according to the description in section 2.2.3. Distance sampling in this case is based on an assumption that the probability that an object of interest will be detected decreases with the distance from the survey line (or point) according to a detectability function. The detectability function is assumed to be well defined within certain conditions, such as those within a forest stand. Although ideally data should be acquired through probability sampling when distance sampling methods are applied (Buckland *et al.,* 2005), we used the MLS-derived data from the 87 rectangular sample plots along the ATV track to estimate a detectability function for our study stand. A half-normal model was applied, i.e. we used the model

$$g(d) = e^{\left[-\frac{1}{2}\left(\frac{d}{\sigma}\right)^2\right]} \quad (6)$$

where $g(d)$is the detection function, σ is a scale parameter, and d is the perpendicular distance from a detected tree to the ATV track. The maximum likelihood estimator (MLE) for the parameter σ is

$$\hat{\sigma}^2 = \frac{1}{M}\sum_{i=1}^{M} d_i^2 \quad (7)$$

where M is the number of detected trees obtained by pooling the data from the 87 10 m line transects, and d_i is the perpendicular distance from the i^{th} detected tree to the corresponding line transect. In Figure 6, the observations and the fitted detection function is shown.

Figure 6 about here

Using the detection function, a correction factor to be multiplied with each MLS-derived stem density estimate is

$$Q = \frac{w}{\mu} \tag{8}$$

where $w = 25$ m is the half-length of our MLS plots, and μ is the effective half-length, which is estimated as (*e.g.,* Buckland *et al.,* 2005)

$$\hat{\mu} = \sqrt{\hat{\sigma}^2 \frac{\pi}{2}} \tag{9}$$

2.2.5. Prediction of stand level stem number

As the final step the stand level stem number was predicted as the average predicted (non-adjusted) stem density across the $N = 1875$ points in the square lattice, multiplied by the factor Q and by the stand area, i.e.

$$\hat{\tau}_{MLS} = \hat{Q}\frac{A}{N}\sum_{t=1}^{N}\sum_{i=1}^{n_{MLS}} a_{ti}\, y_i = \hat{Q}\frac{A}{N}\sum_{t=1}^{N}\hat{y}_t = A\hat{Q}\bar{\hat{y}} \tag{10}$$

where $\hat{Q}$ is an estimated correction factor based on Eqs. (8) and (9), $A = 46963$ m^2 is the area of the stand in square metres, $\hat{y}_i$ is the predicted stem density for the t^{th} point in the square lattice using Eq. (2), and $\bar{\hat{y}}$ is the average of predicted stem densities across the N points.

While the prediction of stem number is straightforward once the values for the lattice centre points are predicted using Kriging, and corrected for non-detection, the uncertainty of the prediction is more difficult to assess. The reason is that it involves several sources of uncertainty: uncertainty due to the Kriging prediction and due to adjusting for non-detection (estimation of the correction factor Q). In our derivation of the variance of the predictor of stem number we accounted for these two uncertainty sources, but not for the uncertainty related to the estimation of the covariogram model, partly since including the latter would lead to substantial difficulties, partly since it is common practice in Kriging applications to assume that this source of uncertainty is small (*e.g.,* Thompson 1992). Thus, the variance of $\hat{\tau}_{MLS}$ is

$$V(\hat{\tau}_{MLS}) = V\left(A\hat{Q}\bar{\hat{y}}\right) = A^2 V\left(\hat{Q}\bar{\hat{y}}\right)$$

Using the formula for the variance of the product of two independent random variables (*e.g.,* Goodman 1960), we can develop the expression further

$$V(\hat{\tau}_{MLS}) = A^2\left[\bar{\hat{y}}^2 V\left(\hat{Q}\right) + \hat{Q}^2 V\left(\bar{\hat{y}}\right) + V\left(\hat{Q}\right)V\left(\bar{\hat{y}}\right)\right] \tag{11}$$

By replacing the variances of $\hat{Q}$ and $\bar{\hat{y}}$ with their corresponding estimators we obtain an estimator of $V(\hat{\tau}_{MLS})$.

We first derive the variance estimator $\hat{V}\left(\bar{\hat{y}}\right)$. Recall that $\bar{\hat{y}} = \frac{1}{N}\sum_{t=1}^{N}\sum_{i=1}^{n_{MLS}} a_{ti}\, y_i$; thus, the variance of $\bar{\hat{y}}$ can be written as

$$V\left(\bar{\hat{y}}\right) = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N} Cov\left(\sum_{i=1}^{n_{MLS}} a_{ki}\, y_i, \sum_{j=1}^{n_{MLS}} a_{li}\, y_i\right) = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{i=1}^{n_{MLS}}\sum_{j=1}^{n_{MLS}} a_{ki}a_{lj} Cov(y_i, y_j) \tag{12}$$

The covariances $Cov(y_i, y_j)$ are estimated using the covariogram function based on distance; by replacing $Cov(y_i, y_j)$ with its estimator $\widehat{Cov}(y_i, y_j)$ in Eq. (12) we obtain the estimator $\hat{V}(\bar{\hat{y}})$.

The variance of $\hat{Q}$ can be written as

$$V(\hat{Q}) = V\left(\frac{w}{\hat{\mu}}\right) = w^2 V\left(\frac{1}{\hat{\mu}}\right) \tag{13}$$

where the variance $V\left(\frac{1}{\hat{\mu}}\right)$ depends on the number of detected trees M and is given by Quinn & Gallucci (1980, Eq. 13)

$$V\left(\frac{1}{\hat{\mu}}\right) = \left(\frac{2}{\pi}\right) B^2(M) V\left(\frac{1}{\hat{\sigma}}\right) \tag{14}$$

where $B(M) = \sqrt{\frac{2}{M}}\Gamma\left(\frac{M}{2}\right)/\Gamma\left[\frac{(M-1)}{2}\right]$ is a bias correction factor for $\frac{1}{\hat{\sigma}}$, and $\Gamma(\cdot)$ is the gamma function. The bias factor $B(M)$ is close to 1 for $M > 50$ (Quinn & Gallucci, 1980); hence, in our study we used the approximation $B(M) = 1$. The variance of $\frac{1}{\hat{\sigma}}$ is given by Quinn & Gallucci (1980, Eq. 11)

$$V\left(\frac{1}{\hat{\sigma}}\right) = \frac{\left[\frac{M}{(M-2)} - \frac{1}{B^2(M)}\right]}{\sigma^2} \tag{15}$$

By replacing σ with the estimator $\hat{\sigma}$ [Eq. (7)] we obtain an estimator for $V\left(\frac{1}{\hat{\sigma}}\right)$; employing Eqs. (13) and (14) we obtain an approximation for $\hat{V}(\hat{Q})$:

$$\hat{V}(\hat{Q}) \cong w^2 \left(\frac{2}{\pi}\right) \frac{\left[\frac{M}{(M-2)} - 1\right]}{\hat{\sigma}^2} \tag{16}$$

Finally, an approximately unbiased variance estimator $\hat{V}(\hat{\tau}_{MLS})$ is obtained as:

$$\hat{V}(\hat{\tau}_{MLS}) \cong A^2 \left[\bar{\hat{y}}^2 w^2 \left(\frac{2}{\pi}\right) \frac{\left[\frac{M}{(M-2)} - 1\right]}{\hat{\sigma}^2} + \hat{Q}^2 \frac{1}{N^2} \sum_{k=1}^{N} \sum_{l=1}^{N} \sum_{i=1}^{n_{MLS}} \sum_{j=1}^{n_{MLS}} a_{ki} a_{lj} \widehat{Cov}(y_i, y_j) \right.$$
$$\left. + w^2 \left(\frac{2}{\pi}\right) \frac{\left[\frac{M}{(M-2)} - 1\right]}{\hat{\sigma}^2} \frac{1}{N^2} \sum_{k=1}^{N} \sum_{l=1}^{N} \sum_{i=1}^{n_{MLS}} \sum_{j=1}^{n_{MLS}} a_{ki} a_{lj} \widehat{Cov}(y_i, y_j) \right] \quad (17)$$

2.2.6. Comparing with estimates from fixed area plots

The circular plots were used to make an independent estimate of stem number for purposes of comparison. We applied the fixed area sampling approach (*e.g.,* Gregoire & Valentine, 2008, p. 215). The population total of the forest attribute parameter was estimated as (Gregoire & Valentine, 2008, Eq. 7.3a, p.216):

$$\hat{\tau}_{FD} = \frac{1}{n_{FD}} \sum_{i=1}^{n_{FD}} \hat{\tau}_{FD_i} \quad (18)$$

Where $n_{FD} = 54$ field plots and $\hat{\tau}_{FD_i} = \sum_{k \in s} \frac{y_k}{\pi_k}$, s denotes the sample of n_{FD} field plots, y_k is the attribute for the k^{th} tree, $\pi_k = \frac{a_k}{A}$ is the probability of including the k^{th} tree in the sample, and a_k is the inclusion zone of the k^{th} tree.

The variance was estimated as (Gregoire & Valentine, 2008, Eq. 7.6a, p. 216):

$$\hat{V}(\hat{\tau}_{FD}) = \frac{1}{n_{FD}(n_{FD}-1)} \sum_{i=1}^{n_{FD}} \left(\hat{\tau}_{FD_i} - \hat{\tau}_{FD}\right)^2 \quad (19)$$

2.2.7. Relative standard error and confidence interval

Relative standard errors (RSEs) were computed as

$$RSE = 100\% \frac{\sqrt{\hat{V}(\hat{\tau})}}{\hat{\tau}} \quad (20)$$

The prediction interval (PI) in the case of stem number prediction based on MLS data was estimated as

$$\widehat{PI} = \hat{\tau} \pm z\alpha_{/2}\sqrt{\hat{V}(\hat{\tau}_{MLS})} \quad (21)$$

and the confidence interval (CI) in the case of stem number estimation based on field plot measurements only was estimated as

$$\widehat{CI} = \hat{\tau} \pm z\alpha_{/2}\sqrt{\hat{V}(\hat{\tau}_{FD})} \quad (21)$$

where $z\alpha_{/2}$ is the standard normal distribution's critical value, which equals 1.96 for $\alpha = 0.05$, i.e. for 95% confidence.

3. Results

The predicted number of stems (with DBH>10 cm) based on MLS data and the estimated number of stems based of field data, their corresponding estimated RSEs, and the PI and the CI, are presented in Table 2. It can be seen that the PI (based on MLS data) covers the stem number estimate using field data and that the CI (based on field data) covers the stem number prediction using MLS data.

Table 2 about here

In Table 3 the estimated detection model parameter $\hat{\sigma}$, the estimated correction factor $\hat{Q}$, its variance $\hat{V}(\hat{Q})$, the average predicted stem density $\bar{\hat{y}}$, and its variance $\hat{V}(\bar{\hat{y}})$ are presented. It can be noted that the correction for non-detection is substantial, i.e. the stem density based on trees detected with MLS data is increased with 54%.

Table 3 about here

Figure 7 shows a map of predicted stem densities for each 5×5 m grid-cell across the stand.

Figure 7 about here

In Figure 7 it can be observed how patches of denser and sparser stem densities were allocated along the MLS track. Further, areas far from the MLS track obtained predicted values equal to the mean stand density, since the spatial autocorrelation was zero at distances greater than about 60 m.

4. Discussion

Airborne laser scanning has emerged as a very useful technique for forest inventories (*e.g.,* Gregoire *et al.,* 2016) and there is much promise in using TLS for the same purpose (*e.g.,* Raumonen *et al.,* 2013; Astrup *et al.,* 2014). However, several issues related to identifying and measuring trees and setting up adequate sampling protocols remain to be solved before the latter technique will be operational. In this study, we propose an approach for making use of MLS data obtained from subjectively selected MLS tracks in a forest stand. We applied Kriging (*e.g.,* Thompson 1992) to handle the issue that a non-probability sample was selected. With MLS it is possible to select large samples from forest stands and provided that there is significant spatial autocorrelation not all parts of a stand need to be visited. In this study we found that the covariance between plot level stem densities remained non-zero up to about 60 m. Thus, in case measurements are obtained through traversing a stand in a manner that leaves few parts further than 60 m away from the survey line, Kriging can compensate for the subjective sample selection and make the stand level predictions model-unbiased (*e.g.,* Thompson 1992). A second issue addressed in this study was to compensate for non-detected trees. To achieve this we followed an approach suggested by Ducey & Astrup (2013), i.e. we used distance sampling theory (*e.g.,* Buckland *et al.,* 2005) to adjust for non-detection. By multiplying the stand level average stem density estimated from MLS data with a correction factor and multiplying it with the study area size we obtained a predicted value of stem number at stand level and its corresponding variance. Assessing the error components contributing to the total variance in the case study, we found that the variance due to Kriging was the major component. We also made comparisons with an independent field sample plot survey within the same stand and found the results to be quite similar. Studying the map of predicted stem densities (Figure 7), patches of denser and sparser densities along the MLS track tend to have smooth shapes due to the spatial interpolation. While sharp boundaries may not be captured, we argue that this is less important when the main objective is to make predictions at stand level. Further, the variance of the grid-cell level stem density predictions was found to increase rather rapidly with distance from the MLS track.

The approach we suggest has many shortcomings, and it is easy to make a long list of topics that need further attention. In the bullet list below we suggest some important topics to be addressed in future research studies for improving the methodology.

- The probability of detecting trees most likely depends on the trees' DBH, and thus it may be advisable to fit detection functions by DBH size groups, or to use DBH as a covariate (*e.g.,* Ducey & Astrup 2013). However, in this study we focused on mid- and overstory trees and used a common detection function for all trees larger than 10 cm.

- How large should the MLS plots be? Perhaps 10×50 m is too large, considering that stands are often relatively small and that large plots will often cross stand boundaries. Within the design-based inventory framework for conventional field plots, procedures for handling plots near the boundary (and trees whose inclusion zones cross the boundary) are well-developed (*e.g.,* Gregoire 1982; Ducey *et al.*, 2004). In general, more work is needed for determining how boundary issues should be treated in MLS surveys. In this study the width (10 m) was chosen to be rather narrow because this made it possible to approximate the ATV track with a straight line.

- What detection function should ideally be used in MLS-based surveys? In this study, we used the half-normal model, but we did not account for the right-truncation of the detection function in the parameter estimation. Although the literature suggests that this type of truncation has minimal effect on the estimates (*e.g.,* Feller 1971) it would be advisable to study this issue further in case the half-normal model is used. Also, since there cannot be any trees in the ATV track it might be appropriate to let the zero distance in estimating the detection function be located some short distance away from the ATV path. No such correction was employed in the current study, as can be observed in Figure 6.

- More studies are needed for optimizing the MLS path, considering what spatial autocorrelation can be expected. In our study we found that the distance beyond which the covariance was close to zero was about 60 metres. Thus, the ATV should probably have aimed at traversing the stand in a manner so that only few parts of the stand were located more than 60 metres away from the track.

- The variance estimation procedures we propose are not entirely straightforward to apply. An alternative could be to apply jackknifing or spatial bootstrapping (*e.g.,* Anselin, 1990; Tomczak, 1998; McRoberts *et al.,* 2011). During the study, we made a simple initial test of jackknifing for the variance estimation and obtained a smaller value compared to the variance estimate reported in Table 2. Also, methods to account for the uncertainty due to the estimation of the covariogram model should be developed.

- Additional auxiliary data, such as remotely sensed data from digital air photos (*e.g.,* Bohlin *et al.,* 2012; Breidenbach & Astrup 2012; Puliti *et al.,* 2017) or satellites (*e.g.,* Tomppo *et al.,* 2002; McRoberts *et al.,* 2011) could be incorporated into the prediction procedures, in which case cokriging (*e.g.,* Myers 1983) would be applied.

Whereas studies such as those by Lin *et al.* (2012) and Forsman *et al.* (2016) focus on obtaining accurate tree-level measurements from the MLS point cloud, the present study is a first step towards utilising MLS data for obtaining stand level estimates once the problems related to tree level measurements are solved. Such applications would potentially be very valuable for obtaining accurate stand level data for forest management purposes. We focused on stem number prediction; however, other tree-based attributes such as basal area, growing stock volume and mean DBH would also be possible to predict from reliable tree level data along the ATV track, perhaps building on the work by Astrup *et al.* (2014) for conventional point-based TLS. Given the fast development of laser technology, we think that procedures like the one we suggest in this study has great potential for providing accurate stand level information at low cost in the future.

5. Conclusions

MLS coupled with adequate sampling protocols and computational algorithms has a potential to become an efficient tool for stand level forest inventories. This would not require any traditional field measurements at all, or MLS might be operated in tandem with field measurements when those are sparse. However, further studies addressing many remaining uncertainties are needed before MLS-based stand level surveys could become operational.

Acknowledgment

This study was financially supported by the Norwegian Institute of Bioeconomy Research (NIBIO) and the Distributed, Integrated And Harmonized Forest Information For Bioeconomy Outlooks (DIABOLO) project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 633464. The authors are grateful to Dr Ronald E. McRoberts and an anonymous reviewer for their valuable comments, which helped to improve the article.

References

Andersen, H.-E., Strunk, J., & Temesgen, H. (2011). Using airborne light detection and ranging as a sampling tool for estimating forest biomass resources in the Upper Tanana Valley of Interior Alaska. *Western Journal of Applied Forestry, 26*, 157–164.

Anselin, L. (1990). Some robust approaches to testing and estimation in spatial econometrics. *Regional Science and Urban Economics, 20(2),* 141-163.

Astrup, R., Ducey, M. J., Granhus, A., Ritter, T., & von Lüpke, N. (2014). Approaches for estimating stand-level volume using terrestrial laser scanning in a single-scan mode. *Canadian Journal of Forest Research, 44(6)*, 666-676.

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

Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L., & Kimes, D. S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. *Remote Sensing of Environment, 112*, 3876–3890.

Breidenbach, J. & Astrup, R. (2012). Small area estimation of forest attributes in the Norwegian National Forest Inventory. *European Journal of Forest Research 131(4),* 1255-1267.

Buckland, S. T., Anderson, D. R., Burnham, K. P., & Laake, J. L. (2005). *Distance sampling*. Wiley Online Library.

Cressie, N. (1990). The origins of kriging. *Mathematical geology, 22(3)*, 239–252.

Ducey, M.J., Gove, J.H. & Valentine, H.T. (2004). A walkthrough solution to the boundary overlap problem. *Forest Science, 50(4)*, 427–435.

Ducey, M. J., & Astrup, R. (2013). Adjusting for nondetection in forest inventories derived from terrestrial laser scanning. *Canadian Journal of Remote Sensing, 39*, 410–425.

Ducey, M. J., Astrup, R., Seifert, S., Pretzsch, H., Larson, B. C., & Coates, K. D. (2013). Comparison of forest attributes derived from two terrestrial Lidar systems. *Photogrammetric Engineering & Remote Sensing, 79*, 245–257.

Ene, L.T., Næsset, E., Gobakken, T., Bollandsås, O.M., Mauya, E.W. & Zahabu, E. (2017). Large-scale estimation of change in aboveground biomass in miombo woodlands using airborne laser scanning and national forest inventory data. *Remote Sensing of Environment, 188*, 106–117.

Feller, W. (1971). *An introduction to probability theory and its applications: volume II.* 2nd edition, London-New York-Sydney-Toronto: John Wiley & Sons.

Forsman, M., Holmgren, J. & Olofsson, K. (2016). Tree Stem Diameter Estimation from Mobile Laser Scanning Using Line-Wise Intensity-Based Clustering. *Forests, 7(9)*, p.206.

Glennie, C., Brooks, B., Ericksen, T., Hauser, D., Hudnut, K., Foster, J., & Avery, J. (2013). Compact multipurpose mobile laser scanning system: Initial tests and results. *Remote Sensing, 5*, 521–538.

Gobakken, T., Næsset, E., Nelson, R., Bollandsås, O.M., Gregoire, T.G., Ståhl, G., Holm, S., Ørka, H.O. &Astrup, R. (2012). Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. *Remote Sensing of Environment, 123*, 443–456.

Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association, 55(292),* 708-713.

Gregoire, T.G. (1982). The unbiasedness of the mirage correction procedure for boundary overlap. *Forest Science. 28(3)*; 504–508.

Gregoire, T. G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research, 28(10),* 1429–1447.

Gregoire, T. G., & Valentine, H. T. (2008). *Sampling strategies for natural resources and the environment volume 1*. CRC Press.

Gregoire, T.G., Næsset, E., McRoberts, R.E., Ståhl, G., Andersen, H.E., Gobakken, T., Ene, L. & Nelson, R. (2016). Statistical rigor in LiDAR-assisted estimation of aboveground forest biomass. *Remote Sensing of Environment, 173*, 98–108.

Hyyppä, J., & Hallikainen, M. (1996). Applicability of airborne profiling radar to forest inventory. *Remote Sensing of Environment, 57*, 39–57.

Hyyppä, J., & Inkinen, M. (1999). Detecting and estimating attributes for single trees using laser scanner. *The photogrammetric journal of Finland, 16*, 27–42.

Jaakkola, A., Hyyppä, J., Kukko, A., Yu, X., Kaartinen, H., Lehtomäki, M., & Lin, Y. (2010). A low-cost multi-sensoral mobile mapping system and its feasibility for tree measurements. *ISPRS journal of Photogrammetry and Remote Sensing, 65*, 514–522.

Kangas, A., & Maltamo, M. (Eds.). (2006). *Forest inventory: methodology and applications* (Vol. 10). Springer Science & Business Media.

Landsskogtakseringen (2008). Landsskogtakseringens feltinstruks 2008, Hndbok fra Skog og landskap 05/08. Skog og landskap s, Norway. 152 p.

Lehtola, V., Virtanen, J.-P., Rönnholm, P., & Nüchter, A. (2016). Localization corrections for mobile laser scanner using local support-based outlier filtering. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 3*.

Liang, X., Kankare, V., Hyyppä, J., Wang, Y., Kukko, A., Haggrén, H., Yu, X., Kaartinen, H., Jaakkola, A., Guan, F. et al. (2016). Terrestrial laser scanning in forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing, 115*, 63–77.

Liang, X., Kukko, A., Kaartinen, H., Hyyppä, J., Yu, X., Jaakkola, A., & Wang, Y. (2014). Possibilities of a personal laser scanning system for forest mapping and ecosystem services. *Sensors, 14*, 1228–1248.

Lin, Y., Hyyppä, J., Jaakkola, A. & Yu, X. (2012). Three-level frame and RD-schematic algorithm for automatic detection of individual trees from MLS point clouds. *International journal of remote sensing, 33(6),* 1701–1716.

Magnussen, S. (2015). Arguments for a model-dependent inference? *Forestry*, *88*, 317 –325.

McRoberts, R. E. (2010). Probability-and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. *Remote Sensing of Environment, 114(5),* 1017-1025.

McRoberts, R. E., Magnussen, S., Tomppo, E. O., & Chirici, G. (2011). Parametric, bootstrap, and jackknife variance estimators for the k-Nearest Neighbors technique with illustrations using forest inventory and satellite image data. *Remote Sensing of Environment, 115(12),* 3165-3174.

Myers, D. E. (1983). Estimation of linear combinations and co-kriging. *Mathematical Geology, 15(5),* 633-637.

Neigh, C. S., Nelson, R. F., Ranson, K. J., Margolis, H. A., Montesano, P. M., Sun, G., Kharuk, V., Næsset, E., Wulder, M. A., & Andersen, H.-E. (2013). Taking stock of circumboreal forest carbon with ground measurements, airborne and spaceborne lidar. *Remote Sensing of Environment*, 137, 274–279.

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment, 80(1),* 88–99.

Olofsson, K., Holmgren, J., & Olsson, H. (2014). Tree stem and height measurements using terrestrial laser scanning and the ransac algorithm. *Remote Sensing, 6(5),* 4323–4344.

Olofsson, K., & Holmgren, J. (2016). Single Tree Stem Profile Detection Using Terrestrial Laser Scanner Data, Flatness Saliency Features and Curvature Properties. *Forests, 7(9)*, 207.

Puliti, S., Gobakken, T., Ørka, H. O., & Næsset, E. (2017). Assessing 3D point clouds from aerial photographs for species-specific forest inventories. *Scandinavian Journal of Forest Research, 32(1),* 68-79.

Ørka, H. O., Gobakken, T., & Næsset, E. (2016). Predicting Attributes of Regeneration Forests Using Airborne Laser Scanning. *Canadian Journal of Remote Sensing 42(5)*, 541-553.

Quinn, T.J., & Galucci, V.F. (1980). Parametric Models for Line-Transect Estimators of Abundance, *Ecology*, 61(2), 293–302. *Ecol. Soc. of Amer.,* URL: http://www.esajournals.org.

Raumonen, P., Kaasalainen, M., Åkerblom, M., Kaasalainen, S., Kaartinen, H., Vastaranta, M., Holopainen, M., Disney, M., & Lewis, P. (2013). Fast automatic precision tree models from terrestrial laser scanner data. *Remote Sensing, 5*, 491–520.

Raumonen, P., Casella, E., Calders, K., Murphy, S., Åkerblom, M., & Kaasalainen, M. (2015). Massive-scale tree modelling from TLS data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2.*

Rönnholm, P., Liang, X., Kukko, A., Jaakkola, A., & Hyyppä, J. (2016). Quality analysis and correction of mobile backpack laser scanning data. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 3.*

Saarela, S., Grafström, A., Ståhl, G., Kangas, A., Holopainen, M., Tuominen, S., Nordkvist, K., & Hyyppä, J. (2015). Model-assisted estimation of growing stock volume using different combinations of LiDAR and Landsat data as auxiliary information. *Remote Sensing of Environment, 158*, 431–440.

Ståhl, G., Saarela, S., Schnell, S., Holm, S., Breidenbach, J., Healey, S.P., Patterson, P.L., Magnussen, S., Næsset, E., McRoberts, R.E. & Gregoire, T.G. (2016). Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. *Forest Ecosystems 3(5),* 1-11

Thompson, S. K. (1992). *Sampling*. Wiley.

Tomczak, M. (1998). Spatial interpolation and its uncertainty using automated anisotropic inverse distance weighting (IDW)-cross-validation/jackknife approach. *Journal of Geographic Information and Decision Analysis, 2(2),* 18-30.

Tomppo, E., Nilsson, M., Rosengren, M., Aalto, P., & Kennedy, P. (2002). Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass. *Remote Sensing of Environment, 82(1),* 156-171.

Vaaja, M., Virtanen, J.-P., Kurkela, M., Lehtola, V., Hyyppä, J., & Hyyppä, H. (2016). The effect of wind on tree stem parameter estimation using terrestrial laser scanning. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 117–122.

Watt, P., & Donoghue, D. (2005). Measuring forest structure with terrestrial laser scanning. *International Journal of Remote Sensing, 26*, 1437–1446.

Wulder, M. A., White, J. C., Nelson, R. F., Næsset, E., Ørka, H. O., Coops, N. C., Hilker, T., Bater, C. W., & Gobakken, T. (2012). Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment, 121*, 196–209.

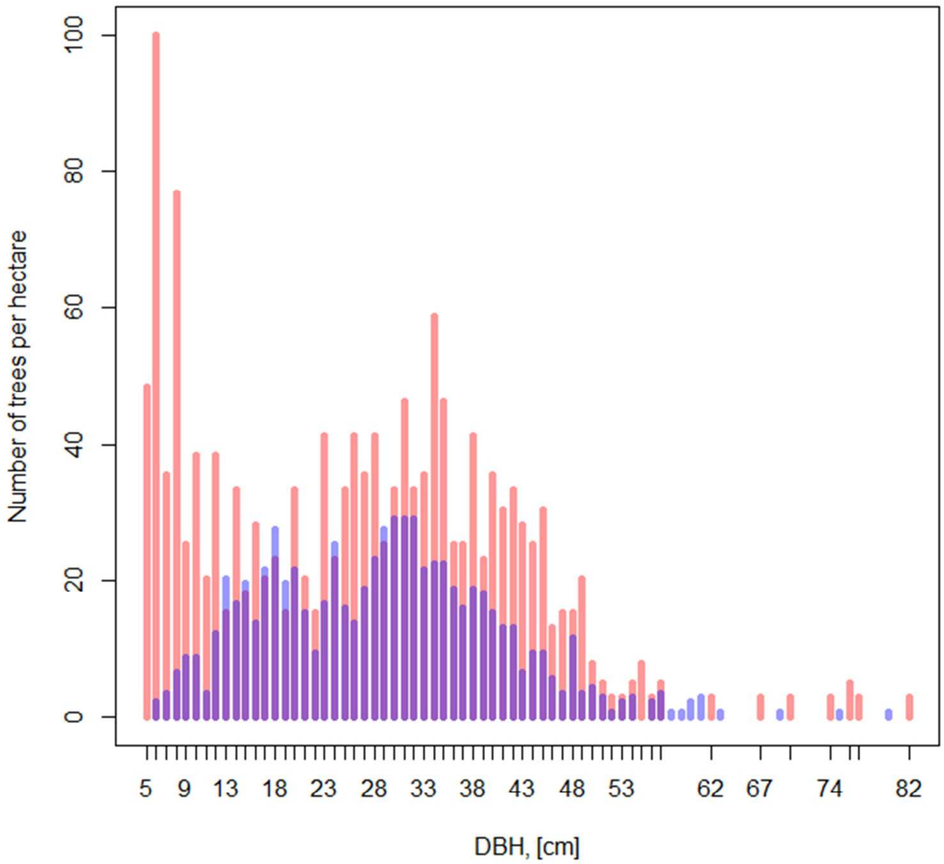

Figure 1: The DBH frequency for trees measured in the field sample plots is indicated with red colour, the DBH frequency for trees identified from the MLS point cloud using tree extraction is indicated with blue colour, the violet colour shows the intersection of these frequencies. The number of trees are shown per hectare.

237x237mm (72 x 72 DPI)

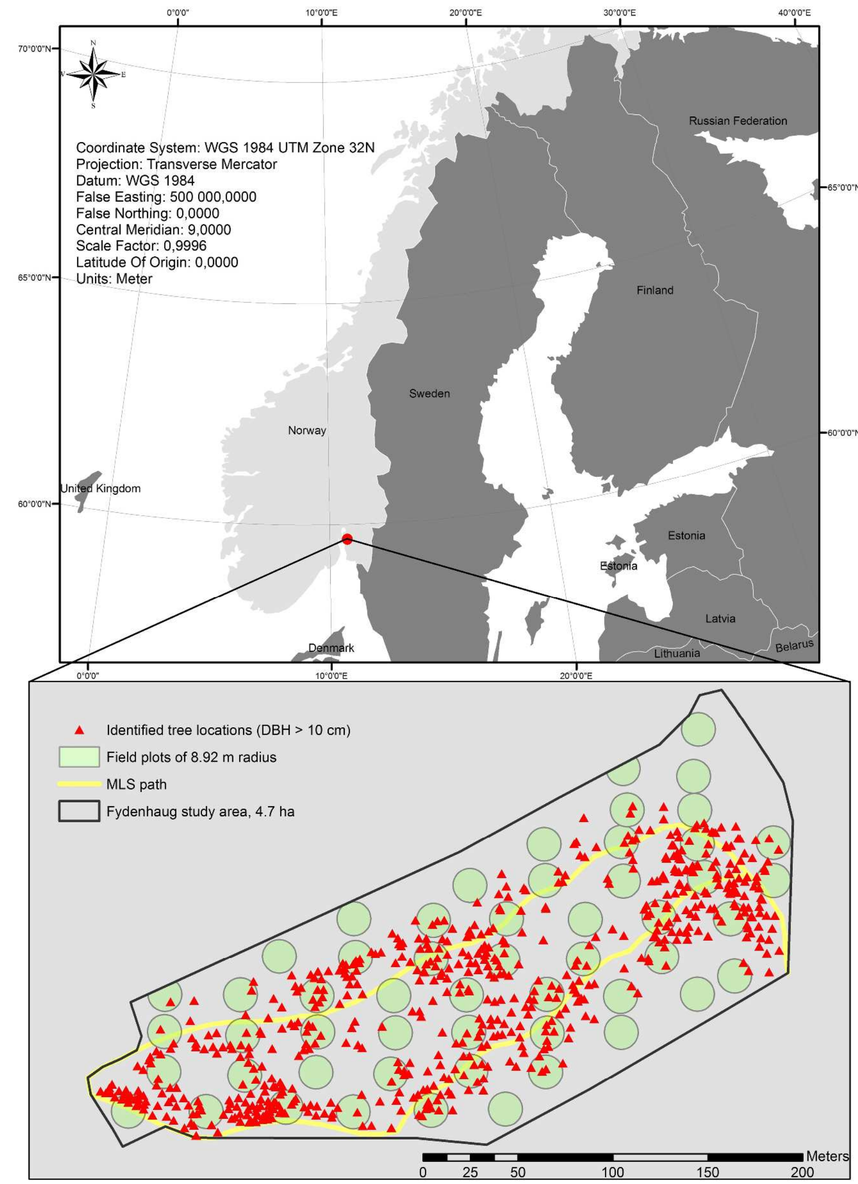

Figure 2: Frydenhaug study area and its location within Norway. Note: the tree locations are identified by the tree detection algorithm using MLS survey data.

296x419mm (300 x 300 DPI)

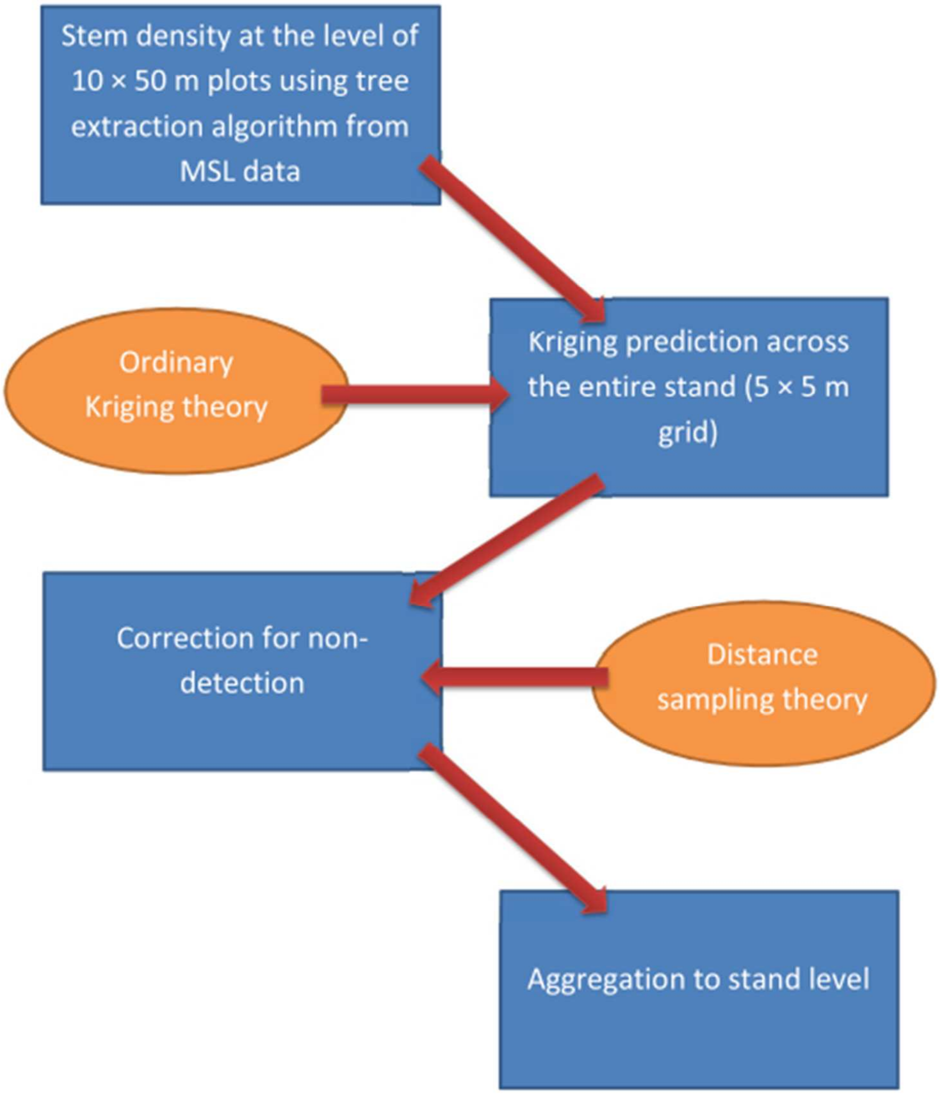

Figure 3: An overview of the procedure to predict stem number at stand level.

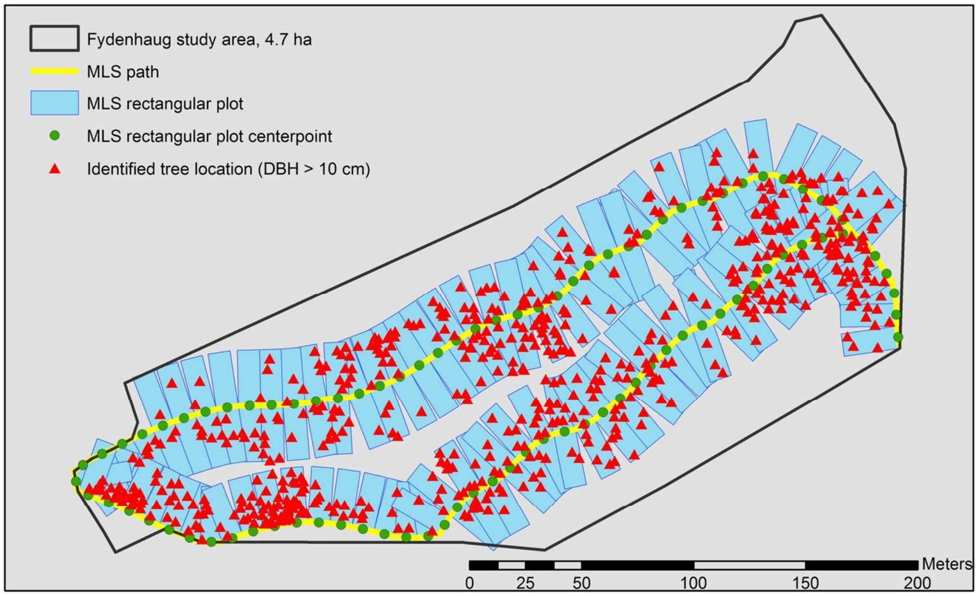

Figure 4: Spatial distribution of the rectangular plots every 10 m along the MLS track. Note: the tree locations are identified by the tree detection algorithm using MLS survey data.

133x84mm (300 x 300 DPI)

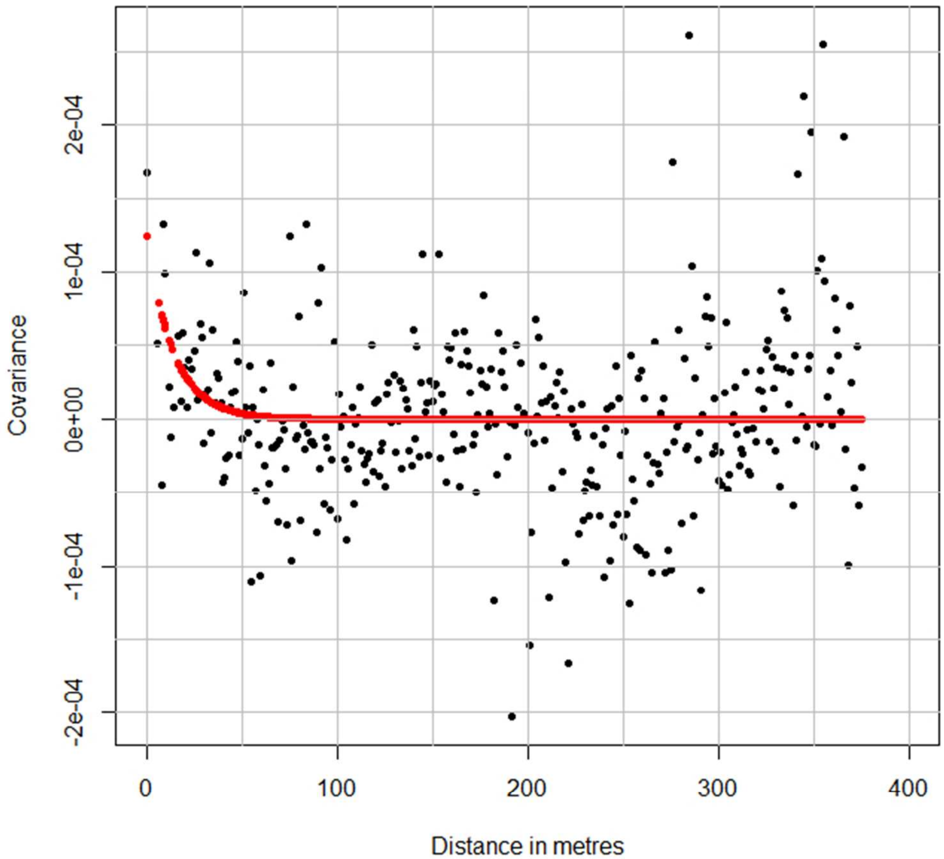

Figure 5: Empirical covariance values (black dots) and the fitted negative exponential covariogram model (red curve) as a function of distance for trees of DBH > 10 cm.

237x237mm (72 x 72 DPI)

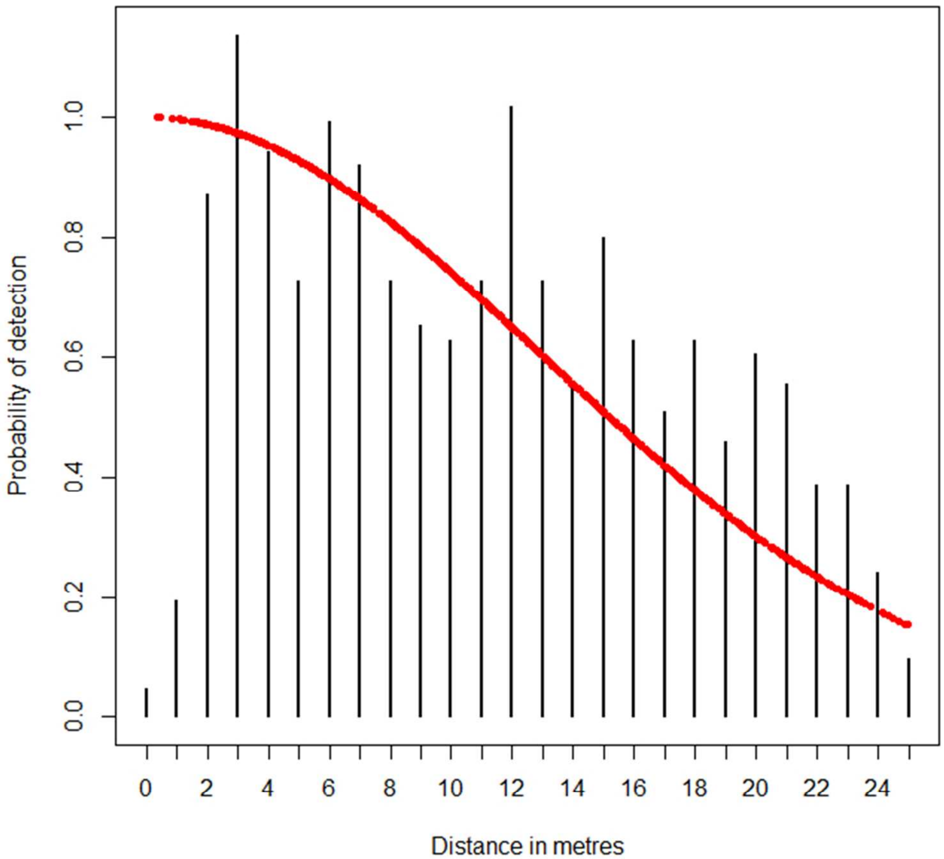

Figure 6: Empirical detection probability values (black bars) and the fitted detection function (red curve).

237x237mm (72 x 72 DPI)

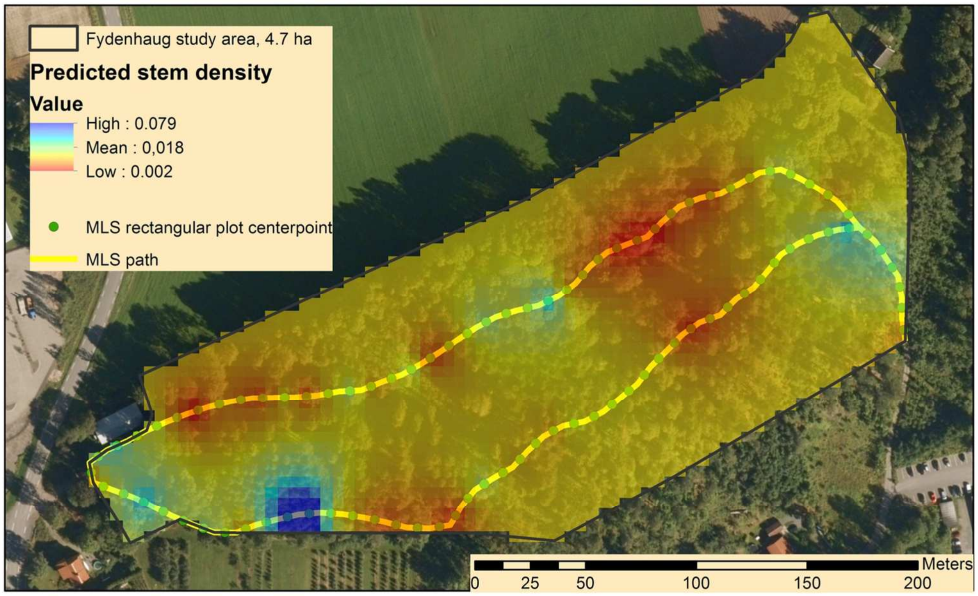

Figure 7: Stem density predictions across Frydenhaug study area.

133x84mm (300 x 300 DPI)

Table 1: The fitted negative exponential covariogram model's characteristics for trees with DBH > 10 cm.

Model parameter	Estimate	Standard Error	t-value	p-value
α	14.22	5.68	2.57	0.012
β	1.24×10^{-4}	4.49×10^{-5}	2.50	0.0060
				Residual standard error: 5.72×10^{-5}

Table 2: The estimated total number of trees, and the corresponding estimated RSE, PI values based on MLS and CI values based on field data.

Based on	$\hat{\tau}$ [trees]	RSE [%]	PI/CI_{lower} [trees]	PI/CI_{upper} [trees]
MLS data only	1366	12.86	1020	1712
Field data only	1677	7.46	1431	1923

Table 3: The estimated detection function parameter, the correction factor and its variance, and the estimated average of predicted stem density across $N = 1875$ points and its variance.

Kriging prediction of stem densities				Adjusting for non-detection
$\bar{\hat{y}}$	$\hat{V}(\bar{\hat{y}})$	$\hat{\sigma}$	$\hat{Q}$	$\hat{V}(\hat{Q})$
1.88×10^{-2}	4.77×10^{-6}	12.90	1.54	7.18×10^{-3}