Remote Sensing of Environment 114 (2010) 700–712

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Reliability of LiDAR derived predictors of forest inventory attributes: A case study with Norway spruce

S. Magnussen [a,*], E. Næsset [b], T. Gobakken [b]

[a] Canadian Forest Service, Pacific Forestry Center, 506 West Burnside Road, Victoria, BC Canada V8Z 1M5
[b] Department of Ecology and Natural Resource Management, Norwegian University of Life Science, P.O. Box 5003, N-1432 Ås, Norway

ARTICLE INFO

Article history:
Received 28 July 2009
Received in revised form 12 November 2009
Accepted 16 November 2009

Keywords:
Airborne laser scanner
Replication variance
Errors-in-variables
Reliability ratio
Attenuation
Calibration
Generalized least squares
Change-point

ABSTRACT

To increase the application domain (re-use) of LiDAR-based models the random replication effects in the predictor(s) must be considered. We quantify these effects in a linear predictor (X) of four forest inventory attributes (Lorey's height HT, basal area BA, volume VOL, and stem density TPH) with LiDAR data acquired over 40 spruce-dominated large plots in southeastern Norway. A grid-based random thinning of the raw multi-echo LiDAR data, to five target densities between 0.25 m^{-2} and 2.0 m^{-2}, generated 100 replications with each density. A DTM was estimated for each replicate and target pulse density. The four linear predictors were constructed from two indicators of canopy density and a posited average effect of a power-transform of echoes classified as canopy returns. Replication variance varied significantly among plots but the reliability ratio of X was high ($\geq$0.92) for HT, BA and VOL but lower for TPH, especially at low pulse densities. Reliability ratios increased with pulse density. Replication variance attenuated the linear regression coefficients by about 10% and inflated the residual variance by 3–6%. A proposed calibration was effective in reducing the impact of replication effects. A proposed bootstrap procedure can be used in practice to obtain good approximations of the replication variance. With echo-densities of approximately 1 m^{-2} or higher the replication effects do not warrant the effort of a calibration.

1. Introduction

Airborne laser scanners (ALS) are increasingly used in forest inventories to provide ancillary metrics (**X**) with good, or at least adequate, predictive power of desired forest resource attributes (**Y**) (Næsset, 2004a). The list of successful applications is long and growing (for examples, Andersen et al., 2006; Holmgren, 2004; Maltamo et al., 2007; Nilsson, 1996; Sexton et al., 2009). In parallel with technical improvements of sensors and data processing the number and types of both X and Y have increased steadily from an initial focus on stand (plot) height (Magnussen et al., 1999; Næsset, 1997; Nilsson, 1996) to a much broader suite including forest structural parameters (Hall et al., 2005), diameter distributions (Gobakken and Næsset, 2005), biomass (Lefsky et al., 1999), leaf area index (Jensen et al., 2008), forest succession (Falkowski et al., 2009), and avian diversity (Clawges et al., 2008). Powerful combinations of point-based ALS metrics and pixel-level multispectral image data have consolidated ALS as essential for boosting accuracy of image-based predictions (Anderson et al., 2008; Coops et al., 2004; Hudak et al., 2006; Leckie et al., 2003; St-Onge et al., 2008). High-density and multi-date ALS data make it possible to pursue 2D and 3D-renditions of forest cover and an object-oriented information extraction (Hopkinson et al., 2008; Korpela, 2006; Pascual et al., 2008; Popescu and Zhao, 2008; Solberg et al., 2006; Vepakomma et al., 2008). Hesitation to use ALS in the context of forest inventory is generally argued on non-technical grounds.

Investments in developing good ALS-based predictive models are considerable, from the cost of data and data processing to the cost of extracting and validating useful predictors (Hyyppä et al., 2008). Ideally the models would not need to be updated, calibrated, or refitted every time they are used with data from a source other than the original, or when they are applied at a spatial resolution different from the one used for model fitting, as, for example, a stand or a plot may have. Næsset et al. (2005) found a significant difference between model predictions of Lorey's height derived from two inventories of otherwise similar populations and ancillary variables. Added costs of frequent model calibration and validation could become an issue for wider ALS applications in forest inventories and provide an incentive for making models more robust (portable).

Several studies have confirmed that changing sensors or one or more of the parameters controlling the ALS data acquisition process can significantly affect the temporal profile of pulse returns, and the spatial and vertical distributions of laser return points (Goodwin et al., 2006; Hopkinson, 2007; Hopkinson et al., 2008; Næsset, 2004a, 2005, 2009). The impact on extracted predictors and the precision (and bias) of predicted inventory attributes vary according to attribute and to specific changes in the process leading to the extracted predictor. Large changes in flying altitude can have a significant effect on the size

* Corresponding author. Tel.: +1 250 363 0712; fax: +1 250 363 0775.
E-mail address: steen.magnussen@nrcan.gc.ca (S. Magnussen).

0034-4257/$ – see front matter.
doi:10.1016/j.rse.2009.11.007

of the effective footprint and thus canopy penetration. It is therefore normal to see both upward and downward shifts in extracted height metrics and canopy density indicators as flying altitude increases (Næsset, 2004a, 2005, 2009). From these studies it seems safe to conclude that a variation between campaigns in flying altitude of approximately 10% or less should have little impact on the inventory results.

Variation in the density of reflected laser pulses has an immediate effect on both bias and accuracy of predicted forest inventory attributes. The latter can be deduced from sampling theory but is often confounded by a possible spatial autocorrelation in estimates of canopy surface elevation. Density-induced changes in bias stem from shifts in extracted distribution of ground and canopy surface heights (Gobakken and Næsset, 2008; Hopkinson, 2007; Magnusson et al., 2007; Thomas et al., 2006). It seems possible to predict the general direction and approximate magnitude of these changes through simulation or geometric analysis of a virtual forest canopy (Chasmer et al., 2006; Goodwin et al., 2007; Lovell et al., 2005). Seasonal variation and other biotic changes in the optical properties of a canopy superimpose variations in these general trends (Hopkinson, 2007; Hopkinson et al., 2005; Næsset, 2005). Prediction models that use one or more quantiles from an extracted canopy height distribution as predictor(s) are especially sensitive to density effects because quantiles are order statistics and as such are functions of sample size viz. spatial scale (Harter, 1970; Magnussen, 1999). It appears that a density of 1 returned pulse per square meter of ground surface affords a fairly accurate digital terrain model (DTM) with a root mean squared error (RMSE) of 30 cm or less. In terrains amenable to conventional forestry operations the same density also vouches for a reliable (relative error of 15% or less) extraction of canopy surface heights (Gobakken and Næsset, 2008; Hodgson and Bresnahan, 2004; Hopkinson et al., 2008; Magnusson et al., 2007). Larger field plots can partially compensate a lower pulse density in terms of precision of estimated inventory attributes (Gobakken and Næsset, 2008) because of an increased number of pulses lead to lower sampling errors of extracted predictors.

Scan angle, terrain form, crown shapes and stem density all affect the extracted canopy height distribution based on geometric principles and knowledge about the geometric properties of the terrain and canopy (Breidenbach et al., 2008; Goodwin et al., 2007; Holmgren et al., 2003; Lovell et al., 2005; Magnussen et al., 1999). The just cited studies suggest that scan angle and terrain form may not have a large effect when half-scan angles are less than 7° and terrain slopes are less than 20°.

Factors affecting extracted ALS-based predictors of forest inventory attributes also include instrument errors, positional errors and post-capture data-processing steps. Taken together, they suggest that extracted predictors should be considered as random variables and not as fixed entities. Random is used as a generic term and it includes all factors that for different reasons preclude a perfect replication of the predictors under normal operating conditions (Fuller, 1987). We refer collectively to these factors as replication effects. In routine forest inventory applications with LiDAR extracted predictors it seems reasonable to assume that sampling effects dominate the replication effecs given that sampling of the entire vegetation surface is typically sparse (Chasmer et al., 2006; Jensen et al., 2006; Lovell et al., 2005; Popescu and Zhao, 2008).

Replication effects in LiDAR-derived predictors are typically ignored. To our knowledge the sparse Bayesian model fitting approach proposed by Junttila et al. (2008) is the first to consider model parameters as random which implicitly acknowledge the random nature of the predictors. Ignoring the random nature of a predictor means that model parameters and their precision, as well as confidence intervals for predictions, will differ from those obtained in absence of random variation in the predictor(s) (Carroll et al., 1995; Fuller, 1987). Conversely, apparent differences between 'inventory models' (Hopkinson et al., 2008; Næsset et al., 2005; Vepakomma et al., 2008) may partly be due to random effects in the predictors. Depending on the reliability of the predictor, the model form, and the nature of the random effects the overall impact will range from trivial to serious (Carroll et al., 1995).

To lessen the sensitivity of a model to random effects in the predictors, the analyst can attempt a calibration if either replicated data or an estimate of the assumed replication effect is available (Brown, 1982; Carroll et al., 1995). A calibrated model is — at least in theory — more portable and should be expected to perform better than a non-calibrated model (Carroll et al., 1995). Tangible, practical benefits of a calibration may not accrue in all cases. For example, if a calibrated model is used with predictors that have different replication effects than in the data used for model fitting. The success of a calibration depends critically on model choice, quality of the estimated replication variance, and the structure of these random effects (Carroll et al., 1995). Models used only with predictors extracted from a single data set with a unique set of prescribed flight and data acquisition parameters do not need calibration.

In this study we quantify the assumed replication effects in LiDAR-derived predictors of four forest inventory attributes (Lorey's height, basal area, volume, and stem density) at five density levels of pulse-returns. Because replication effects are model-dependent and tend to become very complex in case of multiple correlated predictors and mixed effects models (Fuller, 1987, ch. 4), we chose to limit our exposé to linear models with a single (trait-specific) predictor considered as the sum of a fixed part and a random replication effect. To be relevant our simple model must be comparable in terms of prediction accuracy to more complex models favoured in practice. This challenge triggered a suite of purpose-specific data-processing step detailed in an Appendix.

To reduce the impact of replication effects we propose a calibrated generalized least squares (CGLS) procedure for the estimation of (linear) model parameters. A practical implementation of the CGLS procedure requires estimates of the replication effects either from replicated data or a bootstrapping of a single data set. We propose a modified bootstrap procedure for the latter case. Measurement (equation) errors (Carroll et al., 1995, p. 229) in the field data have been ignored throughout.

2. Materials and methods

2.1. Study area and field data

A sample of 40 field plots in the municipality of Aurskog–Høland (59°50′N, 11°34′E, 120–390 m a.s.l.) in southeastern Norway furnished the field data. The plots are circular with an area of 1000 m^2 (36 plots) and 500 m^2 (4 plots) located within a 960 km^2 forested area. The field data were collected during the fall of 2007 and winter of 2008. On each plot, diameter at breast height (1.3 m) was callipered on all trees. For each major species two trees in each of five diameter classes — covering the entire range of diameters — were measured for height. Up to five tree heights were measured in each secondary species. Secondary species trees were selected with probability proportional to their stem basal area. The total number of height measurements per plot ranged from 11 to 20 with an average of 17.

Based on the sample trees six different height-diameter regression models were developed. These models were specific for three species (Norway spruce, Scots pine, and deciduous spp.) and two site productivity classes (high, low). Models were applied to predict the height of trees without height measurements. Lorey's basal area weighted mean height (HT_{Lor} m) was calculated from the predicted and measured heights. Volume of each tree was computed by means of volume equations of individual trees (Braastad, 1966; Brantseg, 1967; Vestjordet, 1967), which are based on height and diameter as predictor variables. Total plot volume (VOL $m^3\ ha^{-1}$ over bark) was computed as the sum of the individual tree volumes.

Plot basal area (BA $m^2\ ha^{-1}$) was computed as basal area per hectare from the stem breast height diameter measurements. Stem

number was computed as number of trees per hectare (TPH ha^{-1}). The species distribution by volume was spruce (27–100%, mean 89%), pine (0–59%, mean 4%), deciduous spp. (0–33%, mean 7%). A summary of the field data is in Table 1.

The plot center coordinates (x and y) were determined by differential dual-frequency Global Positioning System (GPS) and Global Navigation Satellite System (GLONASS) measurements. The Pinnacle version 1.00 (Anon., 1999) post-processing software package was used. Based on the positional standard errors reported by Pinnacle, the estimated accuracy of the planimetric plot coordinates (x and y) ranged from <0.1 to 0.35 m, with an average of 0.12 m.

2.2. Laser scanner data

Laser data for this study were acquired under leaf-on conditions on 16 June 2006 using and Optech ALTM 3100 laser scanner operated from an altitude of approximately 800 m a.g.l. from a fixed-wing PA31 Piper Navajo aircraft. The flight speed was 75 ms^{-1} and the pulse repetition frequency was 100 kHz. The scan frequency was 70 Hz, resulting in a point density of emitted pulses of approximately 7.4 m^{-2}. The maximum half scan-angle was 5°.

The ALTM 3100 sensor is capable of recording up to four echoes per pulse. The raw data set for this study had 1,329,488 registered echoes of which 1242 (0.09%) had extreme z-values (approximately 100 times the average) and were deleted from further processing. To prepare for the DTM construction the ALS data were clipped to squares of 2116 m^2 (46 m × 46 m) centered on each 1000 m^2 plot and for the 500 m^2 plots to squares of 1521 m^2 (39 m × 39 m). After this clipping the data comprised 606,136 emitted pulses with a total of 849,302 registered echoes (average 1.4 echoes per pulse). Table 2 gives the distribution of echo types.

2.3. Overview of data-processing and modelling steps

To emulate replication effects in ALS data we need to work with the data as provided by a sensor. Since we have no actual replications we resort to Monte Carlo simulation. Replications must conform to the spatial realities on the ground which precludes resampling at the original density. Instead we perform a simulated sampling of pulses within small spatial units to densities well below the original 7.4 pulses m^{-2} (see Section 2.3.1).

Our models for predicting $Y = HT_{Lor}$, BA, VOL, and TPH were, for reasons given above, limited to linear models with a single predictor. As in Zhao et al. (2009) we believe that the expected value of canopy heights raised to some power is a useful predictor of the four inventory attributes. However, canopy properties like openness and density are bound to modify the relationship. We posit that there is a function of these three predictors that combines them into a single powerful predictor. To arrive at this goal we must be able to extract the canopy heights from the raw ALS data which means that we must not only generate a DTM for each simulated replication (cf. Section 2.3.2) but also somehow manage to separate 'canopy' echoes from 'below-canopy' echoes (cf. Section 2.3.3). The former are returned from the surfaces in the live canopy, the latter from the surfaces below the live canopy. Once we have accomplished the separation, we define and estimate the integral of canopy heights (in Section 2.3.3) and canopy properties (in Section 2.3.4). The function that combines the three predictors into a single composite predictor (X) is detailed in Section 2.3.5. Replication effects in (X) are quantified by their relative magnitude viz. reliability coefficient (cf. Section 2.3.6).

After extracting the trait-specific predictors we estimate — for each replicate and target pulse density — the parameters of a generalized least squares (GLS) regression of the four inventory attributes on the four LiDAR dervied predictors (cf. Section 2.3.7).When the analyst has available estimates of the reliability coefficient of X a set of calibrated generalized regression coefficients (CGLS) can be obtained. We demonstrate this case with reliability coefficients obtained from either replicated data or a bootstrap resampling of a single data set. Goodness of fit is assessed by the usual summary statistics (R^2, residual error, error of parameters) and coverage rates of 95% confidence intervals of a prediction of a plot-level inventory attribute (cf. Section 2.3.5).

2.3.1. Target pulse densities and simulation of ALS data replications

To simulate a repeated data capture with all flight parameters except the pulse density at their original values, we generated 100 random data replications for each plot and target densities of $\lambda = 0.25$, 0.50, 0.75, 1.0, and 2.0 pulse m^{-2}. Each cell (j,k) in the grid of the ith plot with n_{ijk} observed pulses was thinned by selecting at random n^{*}_{ijk} pulses with n^{*}_{ijk} determined by a random draw from a binomial distribution $BIN(n_{ijk}, \lambda \times \hat{\lambda}_i^{-1})$ where $\hat{\lambda}_i$ is the observed pulse density in the ith plot and λ is the target density. Our thinning algorithm preserves, at a resolution of 1 m^2, the relative spatial distribution of pulses in the original data.

2.3.2. Digital terrain models (DTM)

A DTM was generated for every combination of plot (40) × replicate (100) × target density (5). The last echo from each pulse was used for this purpose (Table 1) as they have *a priori* the highest likelihood of having hit a surface near the ground. The ground elevation (z) at each 1 m × 1 m grid cell center was estimated in an iterative algorithm adapted from Axelsson (2000). For an arbitrary grid-location the ground elevation was estimated via cubic spline interpolation (De Boor et al., 1978).

2.3.3. Canopy heights

Our modelling paradigm stated that an X equal to the expected value of canopy heights raised to some power (γ) is a good predictor of Y. We therefore need to obtain an estimate of the distribution of canopy heights — in order to compute the expected value — and find the value of γ that maximizes the correlation with Y.

In pursuit of these objectives we first computed a vegetation (surface) height $\Delta\hat{z}(x,y)$ for all echoes as the difference between the observed z-value and the estimated ground elevation at the location of an echo. A few (~0.2%) negative vegetation heights (range −0.6 to −0.02, std. dev. 0.14 m) were set to zero. The next step is to separate canopy echoes from below-canopy echoes. We separated the echoes based on finding a statistically significant change-point in the distribution of vegetation heights. Fig. 1 gives a (randomly chosen) example of the outcome of a change-point analysis. Details of the change-point analysis are in the Appendix A. We always found a statistically significant change-point located between 1.7 m and 11.0 m above the estimated ground-surface. The median change-point was at a height of 5.5 m.

For a given plot, target density, and replication the expected value of canopy heights raised to a power of γ was approximated by a weighted mean of echoes classified as 'canopy', i.e. $\overline{\Delta z}^{\gamma} = \sum_{i\in\text{canopy}} \Delta\hat{z}_i^{\gamma} \hat{f}(\Delta\hat{z}_i) \times (\sum_{i\in\text{canopy}} \hat{f}(\Delta\hat{z}_i))^{-1}$ where $\hat{f}(\Delta\hat{z}_i)$ is a nonparametric Epanechnikov kernel estimate of the density $f(\Delta\hat{z})$. Kernel bandwidths were determined from data as recommended by Silverman (1986, p 42). In theory the expected value should be computed as an integral, but considering computing time and the finite nature of the target population (trees) we opted for the above approximation. Compared to an arithmetic mean the above weighting scheme reduced the among-

Table 1
Field data summary. Mean, minimum, maximum, and standard deviation (sdev.) of four inventory attributes: Lorey's height (HT_{Lor}, Loetsch et al., 1973, p. 131), basal area (BA), over bark total stem volume (VOL), and stem density (TPH).

	HT_{Lor}	BA	VOL	TPH
	m	$m^2\ ha^{-1}$	$m^3\ ha^{-1}$	ha^{-1}
Mean	16.3	24.3	202.4	1093
Minimum	10.3	13.3	78.9	440
Maximum	25.4	41.3	451.0	2360
Sdev.	3.3	8.4	102.0	485

Table 2
Frequency distribution of LiDAR echoes. Proportions of echoes by type are in parentheses.

	Number of received echoes per pulse			
	1	2	3	>3
First returned echo	387944 (0.46)	190322 (0.22)	25910 (0.03)	830 (0.0)
Second returned echo		191272 (0.23)	25344 (0.03)	820 (0.0)
Third or later returned echoes			26023 (0.03)	837 (0.0)

replicate variance in $\overline{\Delta}\, z^{\gamma}$ by approximately 45% but also introduced a small amount of bias (<5%).

The value of $\overline{\Delta}\, z^{\gamma}$ was determined for a range of γ-values, from −1 to 6 in steps of 0.2. The value that produced the strongest correlation with Y was chosen as 'best'. Changes in γ smaller than 0.2 have virtually no impact on the correlation coefficient.

2.3.4. Canopy properties

We argued that the relationship between Y and $\overline{\Delta}\, z^{\gamma}$ is modified by canopy properties such as, for example, openness and density. We define openness (P_{ground}) as the proportion of all echoes with Δz<1.3 m and canopy density (P_{canopy}) as the proportion of all echoes classified as 'canopy' in the change-point analysis. Summary statistics of these two proportions are in the Appendix.

2.3.5. Predictors of inventory attributes

To achieve our goal of finding a single LiDAR-derived predictor of Y with a predictive power comparable to what could have been achieved with multiple predictors we must find a suitable transformation of $\overline{\Delta}\, z^{\gamma}$, P_{ground} and P_{canopy} into a single powerful predictor X that maximizes the correlation between X and Y. After an extensive search we opted for the following transformations:

$$
\begin{aligned}
X_{\text{HT}} &= \overline{\Delta}\, z^{3.2} \times 10^{-3} \\
X_{\text{BA}} &= \sqrt[0.25]{\overline{\Delta} z^{3.2}(1-\hat{P}_{\text{ground}})(1-\hat{P}_{\text{canopy}})^{-0.5}} \\
X_{\text{VOL}} &= \sqrt{\overline{\Delta} z^{3.2}(1-\hat{P}_{\text{ground}})^{1.5}(1-\hat{P}_{\text{canopy}})^{-0.5}} \\
X_{\text{TPH}} &= 100-100(\hat{P}_{\text{ground}})^{3}\hat{P}_{\text{canopy}}
\end{aligned}
\tag{1}
$$

The predictors in Eq. (1) have some intuitive appeal: only $\overline{\Delta}\, z$ enters as a predictor for HT_{Lor}, and only the two proportions characterizing the canopy appear as predictors for TPH. In a mature spruce/pine stratum with distinct site-differences, we should not expect $\overline{\Delta} z^{\gamma}$ to contain much information about stem density. The scaling constants in Eq. (1) are arbitrary and are only used in order to avoid very large and very small regression coefficients.

2.3.6. Reliability of X

The above derived predictors are not fixed effects. For a given plot and target density their values will vary from one replication to another. To account for this fact in a linear regression of Y on X we need an estimate of the replication variance σ_{xx} of X (Fuller, 1987, p 3). A large replication variance makes X less reliable as a predictor and vice versa. The reliability of X is quantified by a reliability ratio κ_x which takes values between 0 (no reliability) and 1 (complete reliability). In settings with replicated data the estimation of σ_{xx} and κ_x is straightforward. In most cases, however, there are no actual replications and the two quantities must be estimated by some form of resampling of the available data. Estimation details for the two scenarios are in the Appendix.

2.3.7. Model fitting

In the regressions of Y on X we treat the four inventory attributes as correlated multivariate observations on a plot. We therefore use generalized least squares regression (Fomby et al., 1980) which treats the four traits in the ith plot i.e. $\mathbf{Y}_i=\{\text{HT}_{\text{Lor}}, \text{BA}, \text{VOL}, \text{TPH}\}_i$ as observations from a single subject (plot). We propose the following linear relationship between $\mathbf{Y}_i$ and the extracted LiDAR predictors $\mathbf{X}_{ijk}=\{X_{\text{HT}}, X_{\text{BA}}, X_{\text{VOL}}, X_{\text{TPH}}\}_{ijk}$ in the ith plot, the kth target pulse density and jth replicate:

$$
\mathbf{Y}_i = \boldsymbol{\beta}_{0,kj} + \boldsymbol{\beta}_{x,kj} Diag(\mathbf{X}_{ikj}) + \boldsymbol{\varepsilon}_{ikj},\ i = 1,\ldots,40 \tag{2}
$$

where $Diag(\mathbf{X}_{ijk})$ is a 4×4 diagonal matrix with $\mathbf{X}_{ijk}$ along the diagonal and zero elsewhere.

Generalized least squares (GLS) estimates of $\boldsymbol{\beta}_{0,kj}$ and $\boldsymbol{\beta}_{x,kj}$ were obtained for each target pulse density ($k=1,\ldots,5$) and replicate ($j=1,\ldots,100$) by standard methods (Fomby et al., 1980). A GLS-prediction and accuracy statistics of a $\mathbf{Y}_i$ not used in model-fitting were obtained for all combinations of plots, target pulse densities, and replications in a leave-one-out cross-validation scheme (Draper and Smith, 1981, p 30). GLS estimates were also obtained with the mean of X in 100 replications ($\overline{X}_{ik}$). These estimates are virtually free of any attenuation due to random replication effects. They serve as benchmarks.

When reliable estimates of the replication variances in X are available, either from replicated data or bootstrapping of a single data set, the impact of replication effects on the GLS regression coefficients in Eq. (2) can be reduced through a calibration step. We demonstrate calibration of GLS regression coefficients for both scenarios. The abbreviation CGLS is used for calibrated GLS results. Details are in the Appendix.

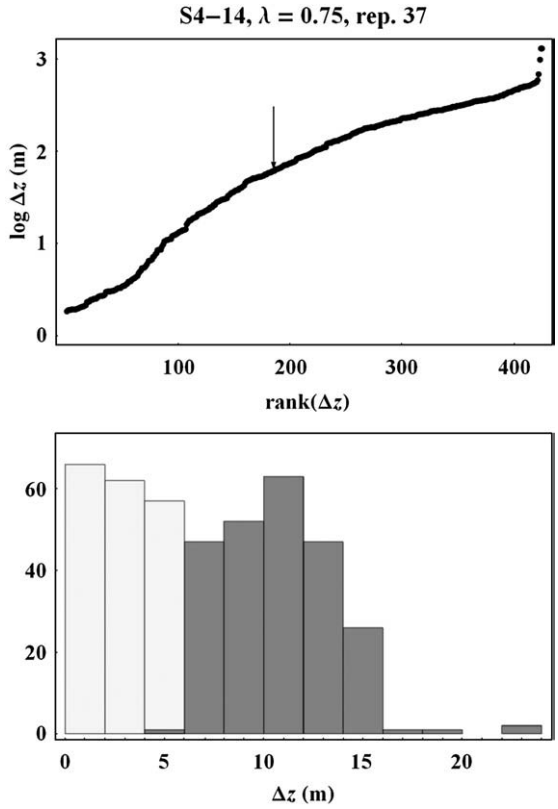

Fig. 1. Classifying echoes to 'canopy' and 'below-canopy' (an example). The arrow points to a statistically significant change-point in the linear regression of $\log(\Delta)$ on $\text{rank}(\Delta\hat{z})$. Echoes to the left and right of the change-point are classifed as 'below-canopy' and 'canopy' returns, respectively. The vegetation (surface) heights of the two classes are displayed in the histogram.

Standard confidence intervals (95%) for GLS and CGLS predictions of a 'new' Y were estimated. The rate (across replications) at which these intervals covered the actual field value of Y is reported as coverage rates.

3. Results

3.1. Replication variance

The among-replication variance of the linear predictors X decreased, as expected, with increasing target density. An increase of 1.0 m^{-2} in λ decreased the variance at a rate of λ^{-d} with d close to1 for X_{HT} and X_{TPH} ($\hat{d} \approx 0.96$) but significantly below 1.0 for X_{BA} and X_{VOL} ($\hat{d}_{BA} = 0.48$, $\hat{d}VOL = 0.41$) which suggests a fairly strong spatial covariance in the underlying data for these two predictors. P_{GROUND}, like X_{HT} and X_{TPH}, also declined at an approximate rate of 1 commensurate with the rate expected under spatial independence (Lichstein et al., 2002).

Fig. 2 illustrates the trend in replication variance across λ and predictors. To remove trait dependent scale effects, the replication variance is expressed in relative terms as the coefficient of variation CV% (the standard deviation divided by the mean × 100). Replication variances differed significantly ($P<0.001$) among plots (Chi-squared test, Shoemaker, 2003). The ratios of the largest to the smallest replication variance were, in all cases, in excess of 100. Eight plots had replication variances more than 10 times as large as the mean of the

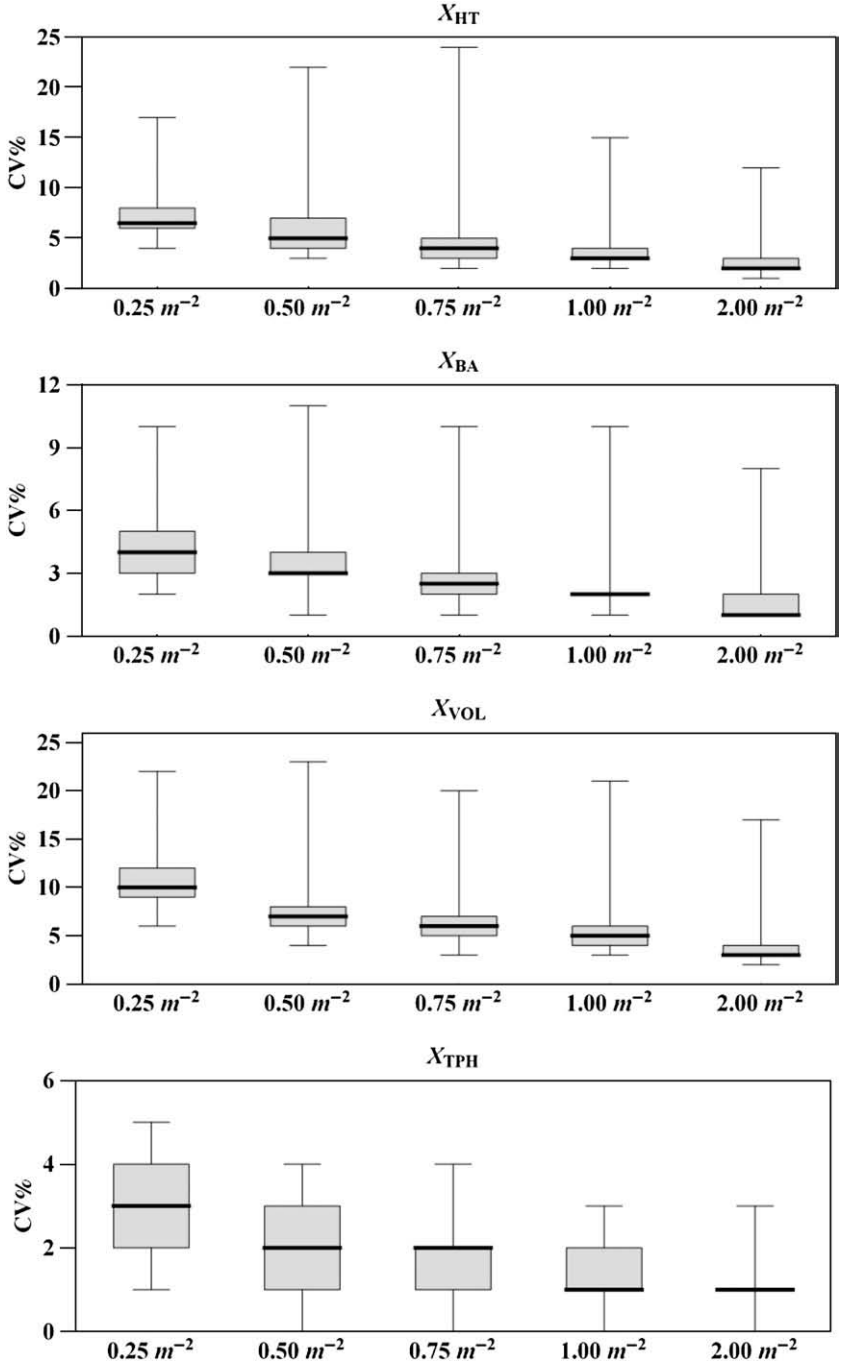

Fig. 2. Replication coefficient of variation in five target pulse densities and four explanatory variables (X_{HT}, X_{BA}, X_{VOL}, X_{TPH}). Plot averages are indicated by horizontal black lines while the 25% and 75% quantiles are displayed as the lower and upper edge of the whisker-boxes. Whiskers indicate plot extremes.

remaining 32. Replication variances of X tend to increase with the plot mean of X.

For X_{HT} the reliability ratio κ_{HT} (the ratio of the among-plot variance to the sum of the among-plot and replicate variance, Carroll et al., 1995, p 22) was very high ($\geq$0.99) in all target densities (Fig. 3). Measurement errors in X_{HT} are thus unlikely to be of concern for predictions of HT_{Lor}. Reliability ratios of X_{BA} and X_{VOL} were somewhat lower. With a target density of $\lambda = 0.25 m^{-2}$ they were 0.92 and 0.95, respectively. A doubling of the target density from 0.25 m^{-2} to 2.0 m^{-2} raised their reliability ratios by about 0.05 (Fig. 3). The lowest reliability ratios belong to X_{TPH} with κ_{TPH} estimated at 0.81 with $\lambda = 0.25$ m^{-2} and increasing to 0.97 with $\lambda = 2.0$ m^{-2}.

The generally high reliability ratios for the predictors in Eq. (1) were not materially different from the ratios of otherwise popular canopy-height (Δz) deciles (not shown) which were in the range 0.84–0.99. Again, ratios increased with density up to the seventh decile and then stabilized.

3.2. Effects of replication variance on prediction of inventory attributes

When the average of X in 100 random replications was used as a predictor the models for HT_{Lor}, BA, and VOL (cf. Eq. (2)) achieved a good fit to the data (Fig. 4). The fit for TPH was poor (R^2<0.4) and we henceforth drop reference to TPH results, but those interested can obtain them upon request. Data from a single replicate will, of course, produce a noisier relationship. Each replication had, on average, 13 plots with X-values more than one standard deviation away from their mean in 100 replications, which is close to the 12 expected for normally distributed data.

Compared to a linear regressor with zero replication variance, a regressor with a random replication effect will lower the coefficient of determination (R^2), increase the residual variance ($\sigma^2_{y|x}$), and attenuate the associated regression coefficient (β_x) towards zero (Carroll et al., 1995, p 42). Given the above effects of target density on reliability ratios, the magnitude of changes in the three regression statistics also depends on the target density. This is illustrated for X_{VOL} and X_{BA} in Figs. 5–7. Changes imparted by replication effects in X_{HT} were minor and not detailed. Results are averages of 100 leave-one-out regressions as per the linear model in (2). As the reliability ratio increases with λ so does R^2 and β_x, while $\sigma^2_{y|x}$ decreases. Changes are again particularly manifest between a target density of 0.25 m^{-2} and 0.5 m^{-2}. The value of the regression statistics with $\overline{X}$ is indicated by a dashed line. The reliability ratio for $\overline{X}$ is above 0.99 in all target densities and traits HT_{Lor}, BA, and VOL with no effect of target density on the regression statistics.

For X_{VOL} the R^2 is 0.86 with $\lambda = 2.0 m^{-2}$, or about 0.07 higher than at the lowest target density. For X_{BA} the corresponding figures are 0.78 and 0.06, respectively. All R^2 values for X_{HT} were between 0.83 and 0.86. A perfect calibration of a linear model with normal distributed residuals and a constant replication variance in X should,

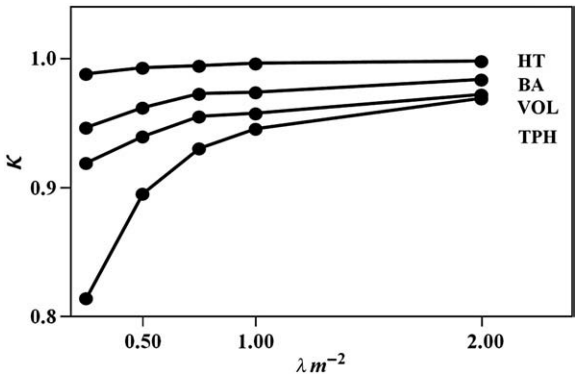

Fig. 3. Reliability ratios (κ) of predictors of HT, BA, VOL, and TPH in five target pulse densities (λ).

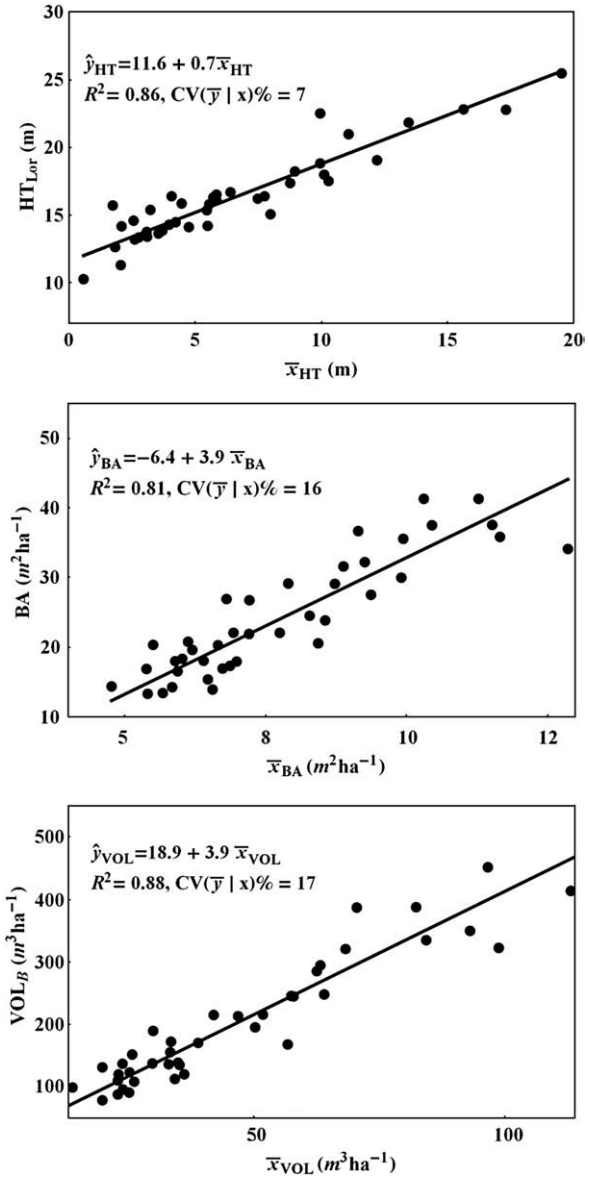

Fig. 4. Regression models (GLS) for HT_{Lor}, BA, and VOL, with the average of X_{HT}, X_{BA}, and X_{VOL} in 100 replications used as a predictor.

theoretically, boost the R^2 by a factor of κ^{-1} (Fuller, 1987, p 4). In practice this is only partially possible when the replication variance is heterogeneous, possibly increasing in X, and estimated from the data (Fuller, 1987, p 271). As a result, the R^2 achieved with the proposed CGLS is only approximating the R^2 obtained with $\overline{X}$ (Fig. 5).

In linear regressions, an improvement in the reliability ratio of a predictor lowers the residual variance (Fig. 6). The contribution of a random replication effect in X to the residual variance is $\hat{\beta}^2 \times \hat{\kappa}^{-2} \times \hat{\sigma}^2_u$ when the effects are non-differential with an average of zero, and a variance independent of X and Y (Carroll et al., 1995, p 22). As suggested by the trends in the reliability ratios, the largest drop in the residual variance occurs between a target density of 0.25 m^{-2} and 0.5 m^{-2}. Further increases in target density barely change the residual variance. In HT-regressions, the coefficient of residual variation dropped by only 1% between a target density of 0.25 m^{-2} and 2.0 m^{-2}.

Attenuation of regression slopes due to replication effects is expected to be proportional to κ^{-1} for normal distributed effects with a constant variance (Fuller, 1987, p 3). Accordingly, $\hat{\beta}_{HT}$ is barely attenuated (<1%). For BA and VOL, the slopes in regressions with X from a single replicate were significantly lower at all densities (t-test,

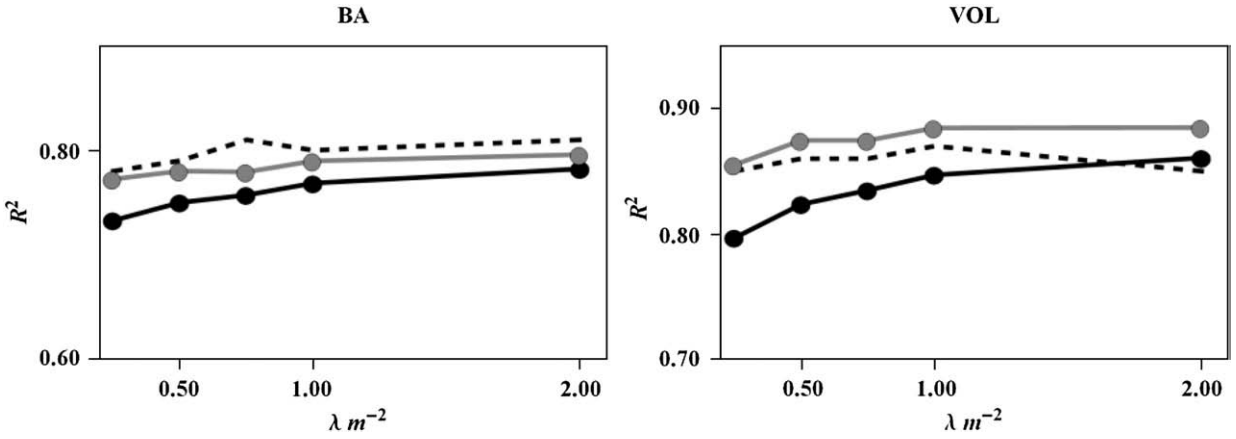

Fig. 5. Average coefficient of determination (R^2) in 100 leave-one-out linear regressions of Y_{BA} against $X_{BA,}$ and Y_{VOL} against X_{VOL} in five target pulse densities (λ). Black symbols: generalized least squares (GLS). Gray symbols: calibrated GLS (CGLS). The dotted lines indicate a 95% jackknife confidence interval for $\hat{R}^2_{GLS}$. The dashed line is the R^2 achieved with the mean of X in 100 replications as a predictor.

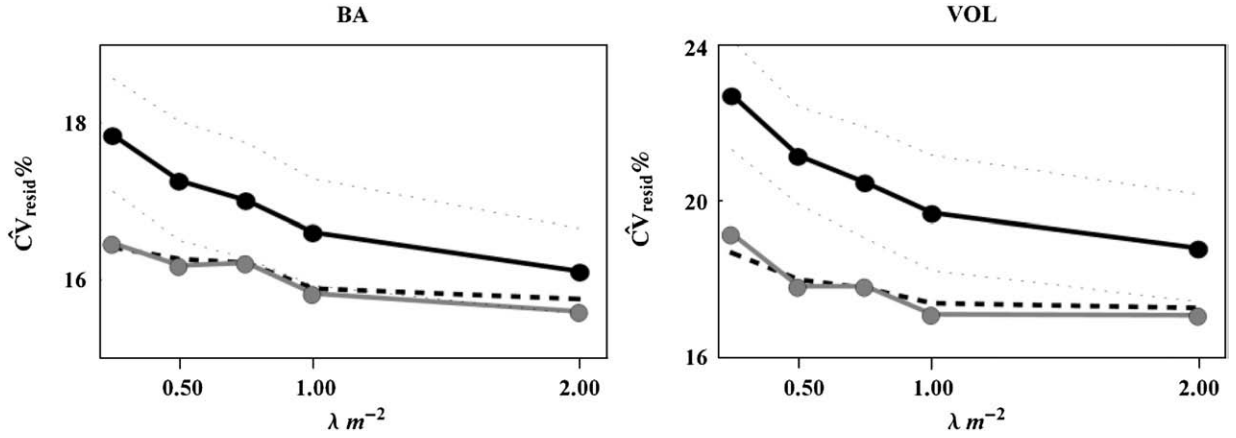

Fig. 6. Average coefficient of residual variation ($CV_{resid}\%$) in 100 leave-one-out regressions of Y_{BA} against X_{BA} and Y_{VOL} against X_{VOL} in five target pulse densities (λ) (black: GLS, gray: CGLS). The dotted lines indicate a 95% jackknife confidence interval for $CV^{GLS}_{resid}\%$. The dashed line is the $CV_{resid}\%$ achieved with the mean of X in 100 replications as a predictor.

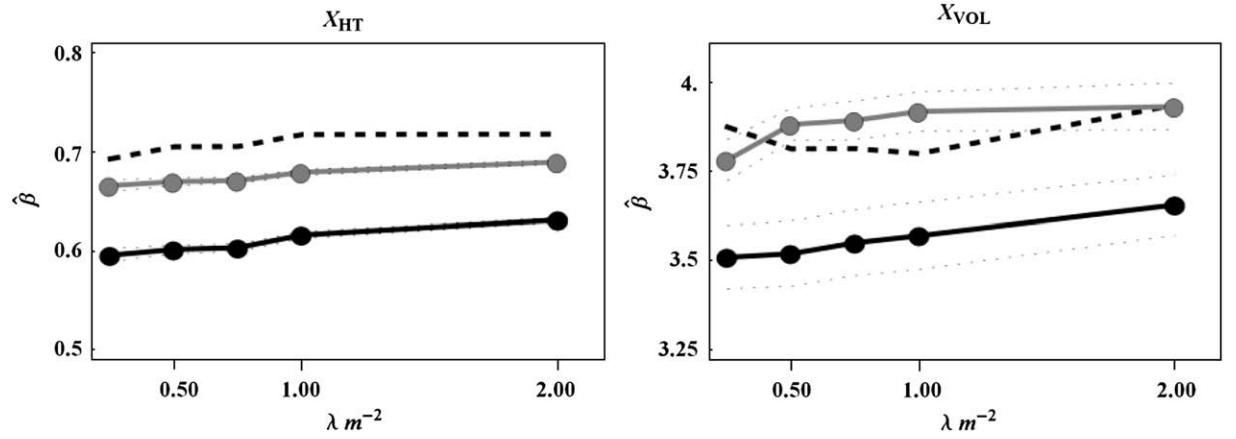

Fig. 7. Trends in regression slopes for HT_{Lor} and VOL across five target pulse densities (λ). Plotted values (circles) are averages of 100 leave-one-out regressions (black: GLS, gray: CGLS). Dotted lines indicate 95% confidence intervals of the estimates. The dashed line is the GLS slope coefficient when the mean of X in 100 replications is used as a predictor.

$P<0.01$) than the slopes obtained with $(\overline{X}$ (Fig. 7). A change in target density from 0.25 m^{-2} to 2.0 m^{-2} increased the slopes $\hat{\beta}_{BA}$ and $\hat{\beta}_{VOL}$ by just 4% ($P>0.28$).

An important consequence of replication errors in X is their additive effects on the variance of predicted Y-values. With plot-dependent replication variances and a tendency for the replication variance to depend on X, the variance of predicted Y-values is further confounded by a covariance between X and the variance of the regression parameters.

3.3. Calibration of predictive models

Calibrated slopes (CGLS) were, as expected, greater than their GLS counterparts and almost reach the values attainable with $(\overline{X}$ Fig. 6). A CGLS estimation of the regression parameters offers potential advantages of: i) less biased predictions for plots in the first and fourth quartile of X, ii) a reduction in the variance of predictions, iii) improved predictions when X is outside the sample range, and iv) accommodation of predictors with a different error variance than in the sample data (Fuller, 1987, p 75). The advantage of less bias is demonstrated for VOL in Fig. 8 where the average of GLS and CGLS leave-one-out predicted Y-values are plotted against their field values. The CGLS averages track closer to the one-to-one line than do GLS averages in all five densities, albeit with the anticipated shrinking of discrepancies as density increases. For BA, the bias advantage was slightly larger, but for HT_{Lor} it was negligible. CGLS leave-one-out predictions were also less variable with a reduction of about 15% in standard deviation at $\lambda=0.25\ m^{-2}$ and about 6% at $\lambda=2.0\ m^{-2}$. CGLS 95% confidence intervals for leave-one-out predictions of VOL were slightly shorter than corresponding GLS intervals, yet achieved better coverage (0.94 versus 0.91) for target pulse densities 0.25 m^{-2} and 0.50 m^{-2}. At higher densities the coverage rates were similar.

3.4. Bootstrap results

The average replication variance estimated from the bootstrap data in replications one to six matched the actual replication variance well. Results for the first six replicates are shown for X_{HT} and X_{VOL} in Fig. 9. Bootstrap estimates of the replication variance of X_{HT} were both larger ($\lambda=0.25 m^{-2}$) and slightly smaller ($\lambda\geq 0.5 m^{-2}$) than variances seen in the simulated replications of data. The discrepancies will be of little concern due to the high reliability ratio of X_{HT}. For VOL, bootstrap estimates of replication variance were mostly within approximately 10% of their counterparts estimated from 100 replications. Only one bootstrap variance estimate (density 2.0 m^{-2}) was significantly different from the variance estimated across 100 replications (F-ratio = 1.89, $P=0.03$). Bootstrap estimates were always strongly and significantly correlated with the estimates based on 100 replications ($0.75\leq\hat{\rho}\leq 0.96$), and rankings of plot-specific estimates were generally preserved ($0.81\leq\hat{\rho}_{Spearman}\leq 0.98$). Bootstrap estimates of the reliability ratios were always within 0.05 from the estimates derived from 100 replications.

Regression coefficients (CGLS) obtained with bootstrap estimates of replication variance(s) were, on average, within 1% (HT), 5% (BA) and 3% (VOL) from the CGLS coefficients determined with the 100-replicate estimates of replication variances (Fig. 10). Results from individual replications (1 through 6) were closer to CGLS than to GLS estimates. Bootstrap estimates of the proportion of the residual variance due to errors in X were within 0.07 of the proportion determined using the 100-replicate estimates of replication variances.

4. Discussion

Random variations in the parameters controlling the capture and processing of LiDAR data also generate random effects in statistics and extracted ancillary variables used as predictors in models of attributes of interest (Breidenbach et al., 2008; Gobakken and Næsset, 2008; Holmgren et al., 2003; Morsdorf et al., 2008; Næsset, 2009; Wulder et al., 2008). We have used the term replication effect to capture non-intended random deviations from specified parameters and data capture conditions. At low to medium-high echo densities, we expect sampling variance to dominate the random effects. When sampling variance is the dominating random effect, various resampling methods can be used to get a good first-order approximation of their magnitude (Bannerjee et al., 2004, p 366; Lahiri et al., 1999; Nordman and Lahiri, 2004). Our proposed modified bootstrap method seems adequate. Whenever the resampling effects appear to be important a calibration of models should be considered.

The term replication effect has been used as if it can be understood *prima facie*. Reality is, of course, much more complex. In the context of LiDAR extracted predictors of forest inventory attributes, the problem

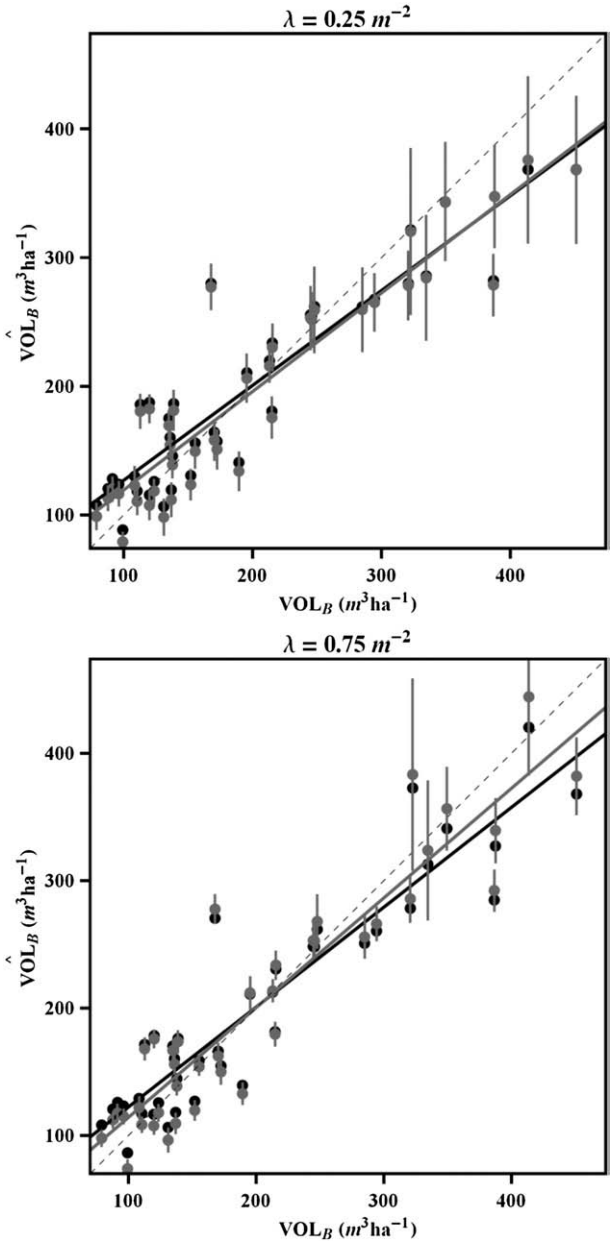

Fig. 8. Average of predicted values of VOL in 100 leave-out-out GLS (black) and CGLS (gray) regressions plotted against their field values in two pulse densities. Ordinary least squares trend lines for GLS (black) and CGLS (gray) are provided for reference. A 1:1 line is indicated by a dashed line. Vertical gray lines are 95% intervals for CGLS predictions.

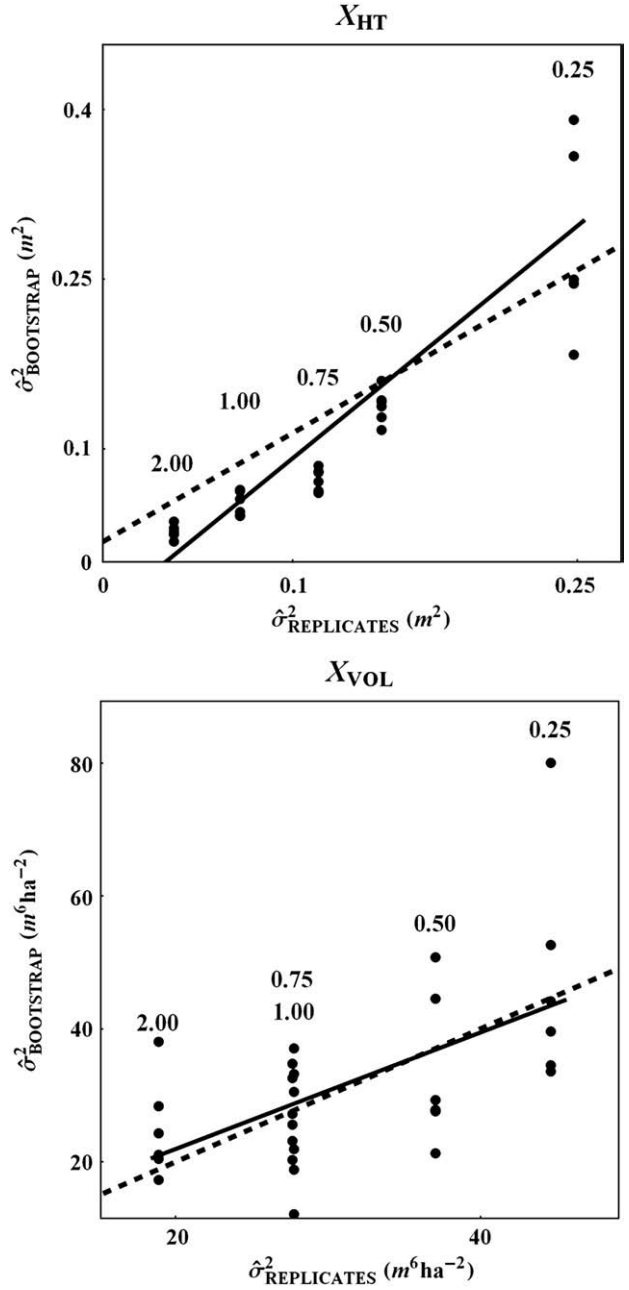

Fig. 9. Bootstrap estimates of the replication variance of X_{HT} and X_{VOL} in five target pulse densities (0.25, 0.5, 0.75, 1.0, and 2.0 m^{-2}) and replications one to six (black dots) plotted against the variance in100 simulated replications of data. A least-squares trend line is shown in full. A one-to-one line is provided for reference (dashed).

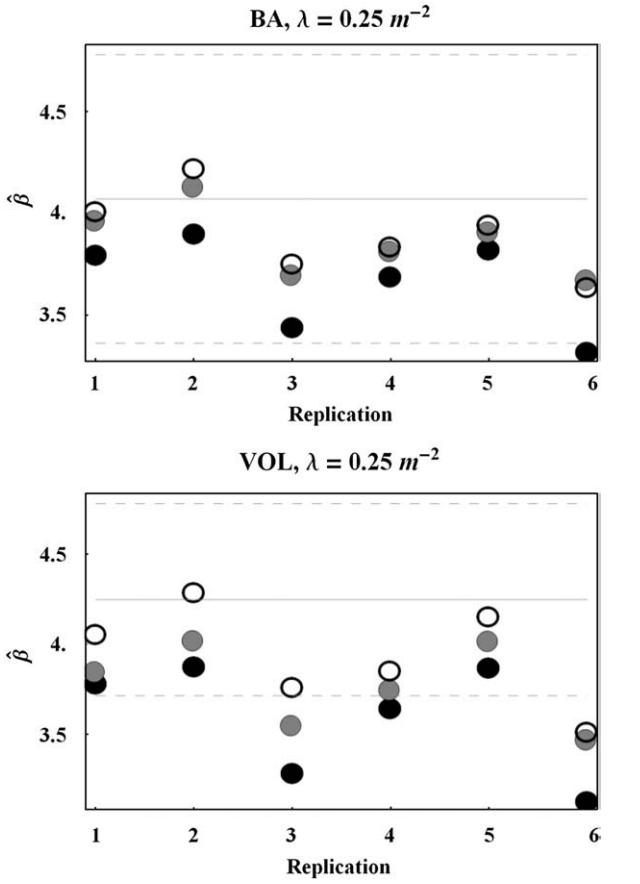

Fig. 10. Examples of GLS and CGLS regression coefficients for BA and VOL in replications one to six with $\lambda = 0.25$ m^{-2}. Black dots: GLS estimates, gray dots: CGLS based on estimates of replication variance in100 replications. Open circles: CGLS based on bootstrap estimates of replication variance. Solid line: GLS regression coefficients obtained with X as the mean of 100 replications. A 95% confidence interval for the latter is indicated by dashed lines.

begins with defining the observation as it would be in absence of a replication effect. Without fixed protocols and standards of measurement, and with ill-defined observational units, we recommend avoiding any simplistic interpretation of 'random replication effects' (Lesser and Kalsbeek, 1999). LiDAR data present an inseparable mixture of random sampling and non-sampling effects. Our terminology has been one of convenience yet consonant in the current problem setting.

Our simulation of replications by random thinning of a single LiDAR data set has undoubtedly underestimated the true replication variance as it would materialize in practical settings with random perturbations of the data by unintended but unavoidable variations in data capture parameters, flight conditions, flight direction, and temporal variation in vegetation surface properties. From published results (Breidenbach et al., 2008; Gobakken and Næsset, 2008; Hopkinson et al., 2005; Magnusson et al., 2007; Næsset, 2004a), we surmise that the underestimation is probably modest, perhaps no more than 5% for the entertained set of five target pulse densities and chosen predictors. It would be larger if flight parameters are not kept reasonably constant (±10%) and half-scan angles are not limited to a narrow range. Larger variations in flight altitude, scan angles, and flight patterns, will have significant effects on the reliability ratio of LiDAR derived predictors. Carefully planned experiments with actual and simulated replications are needed to resolve these issues beyond conjecture.

Our 1 m × 1 m grid-based spatial random thinning of pulses to a given target pulse density emulates ALS data with different levels of pulse density. In settings with a single flight line and all flight and sensor parameters except pulse density are kept constant, a lowering of the pulse density acts as a spatial uniform thinning of pulses emitted at a higher rate. Our thinning, however, does not emulate a change in the pulse repetition frequency (PRF) which would also affect the properties of the outgoing signal and hence the returns (Hopkinson, 2007). Our thinning preserves, on a relative scale, the original spatial variation in density of returned pulses. Alternative thinning procedures tend to generate a more uniform density distribution (Gobakken and Næsset, 2008; Holmgren, 2004; Magnusson et al., 2007; Næsset, 2009).

A linear regression model with a single predictor is ideally suited for demonstrating the impact of random replication effects in a predictor on regression coefficients and descriptive statistics summarizing the regression model. We intentionally departed from a more traditional path of non-linear regression with multiple predictors (for examples, Hyyppä et al., 2008; Næsset, 2004b) because the demonstration would

have succumbed to the complexity of dealing with multivariate non-linear error structures (Fuller, 1987, ch 4). Simplicity was achieved at the cost of additional data processing steps aimed at extracting the single most powerful linear predictor. Inspiration for the average effect of a power transform of vegetation height came from Zhao et al (2009) who used a related device in their scale-invariant prediction model. Many practitioners will undoubtedly continue to prefer non-linear models with multiple predictors in order to forgo the search for a suitable univariate (composite) predictor. While they can assess the magnitude of replication effects in the predictors via the proposed modified bootstrap-procedure the impact on model-prediction will be more difficult to assess. Cook and Stefanski's (1994) calibration procedure can be used, but it is much more involved than the herein detailed CGLS. A successful calibration will improve the portability (re-use) of a model, but only within the domain covered by the validation (field) data.

Most of the studies cited so far have quantified the effects of changes in the parameters of ALS data on various canopy height statistics and predictions of inventory attributes. The effects range from trivial to important, depending on the inventory attribute and the change in the data capture processes (Gobakken and Næsset, 2008; Holmgren et al., 2003; Magnusson et al., 2007; Morsdorf et al., 2008; Næsset, 2009; Næsset et al., 2005). However, LiDAR-derived predictors have been treated as fixed and not as random variables subject to uncontrolled variation. When a predictor is random, so is its associated model parameter and model fitting should be adapted accordingly (Longford, 1993, p 46). With random predictors the error variance of a prediction is inflated relative to the error variance based on fixed predictor values and regression coefficients are usually attenuated (Carroll et al., 1995, p 206). For models with several random predictors the combined effect of the random components is cumulative (Fuller, 1987, p 292).

Random effects in predictors complicates the statistical inference because determination of optimal plot size, test power (Tosteson et al., 2003), model and variable selection, and hypothesis testing (Carroll et al., 1995, ch 11; Laster et al., 1993) become non-standard (Lehmann and Casella, 1998).

Unless one has the assurance that random effects in the predictors are unimportant,we caution against the use of non-calibrated regression models outside the domain defined by the original data used for model fitting (Carroll et al., 1995, p 19; Zhao et al., 2009). Otherwise the risk is campaign or scale effects in predicted inventory attribute values (Gobakken and Næsset, 2008; Hopkinson et al., 2008; Magnusson et al., 2007; Næsset et al., 2005; Rombouts et al., 2008). Popular regression and classification tree techniques (CART) such as 'random forest' (Falkowski et al., 2009; Lin and Jeon, 2006) are especially sensitive to random effects. In this study a density of at least 1 multi-echo laser pulse per square meter ensures predictors with a very high reliability ratio and no need for a calibration. Yet even at a pulse density as low as 0.5 per square meter the analyst should still be able to extract reliable predictors but a calibration is now more likely to improve the estimated regressions.

A calibrated model can be expected to be more robust against random data perturbations (Carroll et al., 1995, p 32), which can translate into non-trivial savings by extending the temporal domain of a model. Calibration does not extend the application domain of a model nor does it correct for errors in inventory attributes. Only additional validation (field) can extend the application domain. Errors in the dependent variable make an efficient calibration more difficult (Carroll et al., 1995; Fuller, 1987).

As demonstrated, calibration can be straightforward and efficient when the model is linear with a single predictor (Kulathinal et al., 2002; McCullagh and Nelder, 1989). More specialized methods will be needed for non-linear, semi-parametric, and non-parametric models (Carroll et al., 1995, ch 6–9; Lin and Carroll, 2000; Schafer and Purdy, 1996; Suh and Schafer, 2002). We also tested the generic simulation-extrapolation procedure (SIMEX, Cook and Stefanski, 1994; Stefanski and Cook, 1995) and found it to work well, but long computation times make it less attractive than CGLS for linear models. It holds for all calibrations that their efficiency deteriorates as the model becomes more complex and the replication effects are poorly quantified (Carroll et al., 1995, p 61).

We demonstrated a replication variance that varied with the value of a predictor — a phenomenon expected to be the rule rather than the exception in forest inventories. A mean–variance relationship may compromise the efficiency of a calibration unless the error structure is somehow known (Carroll et al., 1995, p 50; Fuller, 1987, ch 3; Kulathinal et al., 2002).

Actual replications of ALS data will be rare in operational settings. At best, some of the field plots may be over-flown twice which does allow an estimation of replication effects and possibly a variance function to capture signs of heteroscedasticity (Carroll et al., 1995, p 17). With no actual replications, a bootstrap procedure adapted to the estimation problem seems to capture the portion of the replication effects associated with sampling variance reasonably well. With pre-processed ALS data it becomes important to add the replication effects to the processed data — for example, by adding the random effects to DTM predictions (Andersen et al., 2006; Hodgson and Bresnahan, 2004; Kobler et al., 2007; Reutebuch et al., 2003).

An alternative to our extraction and calibration of predictors is a calibration of the distribution of canopy heights (Carroll et al., 1995, ch 12; Staudenmayer et al., 2008). The calibration is then followed by a conventional search for suitable predictors. However, quantifying random replication effects on a distribution requires actual replications of data.

5. Conclusions

Replication effects in LiDAR extracted predictors of forest inventory attributes are clearly a nuisance issue. A conservative interpretation of our results suggests that replication effects are unimportant as long as the pulse density is one per square meter or higher and when replication effects are dominated by sampling variance. Since these assumptions are not always met in practice the analyst must address the potential of attenuating and variance inflating replication effects in LiDAR extracted predictors. A data resampling scheme, akin to the proposed modified bootstrap procedure, can quantity sampling effects but not the impact of variations in flight parameters and sensor parameters. A calibration of models with LiDAR extracted predictors is recommended when the reliability of one or more of the predictors is below 0.9.

Acknowledgements

The authors would like to thank Mr. Marius Hauglin, Mr. Vegard Lien and Mr. John Gunnar Dokk for a well conducted field work and Dr. Ole Martin Bollandsås for computing the field values of the sample plots used in this study. Three reviewers made numerous helpful suggestions of improvements.

Appendix A

Simulated replications

To initiate simulated replications with preset fixed target densities all echoes from a single pulse were assigned to a 1 m^2 grid cell based on the average of their (x and y) coordinates. Plot specific averages of the number of pulses per square meter varied from 6.3 to 8.3. Nine plots had one or two empty cells. Eight pulses per cell was by far the most common number and 95% of all cells had between 4 and 11 pulses. A maximum of 15 was observed in six cells. The achieved echo density in a replicate was always close to its nominal target.The mean absolute deviation was 0.3% with $\lambda = 0.25\ m^{-2}$and decreased to 0.1% with $\lambda = 2.0\ m^{-2}$. The among-replication coefficient of variation varied from 1.3% ($\lambda = 2.0 m^{-2}$) to 4.3% ($\lambda = 0.25\ m^{-2}$).

DTM algorithm and assessment of errors

The construction principle for the DTM is that the terrain surfaces in the 40 plots are relatively smooth with a maximum slope of 12.5°. In the first iteration ($iter=1$) the terrain z-value ($z^{(iter)}$) assigned to a 1 m×1 m grid cell center was the lowest value among 900 nearest neighbours (search radius≅16.92 $\lambda^{-0.5}$ m). In the second and following iterations, the current z-value of a cell was replaced by the minimum z-value among $NN^{(iter)}=900\times1.25^{-(iter-1)}$ nearest neighbours if $\mathrm{Abs}(z^{(iter)}-z^{(iter-1)})<\bar{r}_w^{(iter)}Tan[12.5°/180°\ \pi]$, and otherwise retained ($\bar{r}_w^{(iter)}$ is the harmonic mean distance to the $NN^{(iter)}$ nearest neighbours from a grid cell center). Iterations were continued until no z-value was replaced or until $NN^{(iter)}<1$, whichever came first. Five to seven iterations sufficed in most cases to complete a DTM.

To check our DTM algorithm we also computed a DTM from the original unthinnned data and compared it to the DTM produced by the data-provider with the proprietary TerraScan software (Anon., 2005). Estimates of terrain elevation (z) at the grid cell centers were compared whereby the z-values from the TerraScan DTM were obtained via bilinear interpolation from the provided DTM given as a triangulated irregular network (Okabe et al., 2000).The mean difference in elevation across 40 plots was 0.05 m with a standard error of 0.09 m. The average within-plot correlation coefficient between the two sets of gridded z-values was 0.96.

In the DTM generated for the thinned data the among-replication variation in z-values for a specific grid point declined, as expected with increasing target density. With a target density of $\lambda=0.25\ \mathrm{m}^{-2}$ the standard deviation was, on average across grid centers and plots, 0.38 m. With $\lambda=2.0\ \mathrm{m}^{-2}$ the standard deviation was 0.25 m. In the lowest target density the replicate standard deviation in elevation of a single grid point could reach 0.94 m but only 0.7 m in the highest target density. For a large majority of grid points the replicate standard deviation was well below 0.3 m regardless of target density. Reliability ratios of the produced DTMs varied from an average of 0.95 with $\lambda=0.25\ \mathrm{m}^{-2}$ to 0.98 with $\lambda=2.0\ \mathrm{m}^{-2}$ where averages are across plots with the same target density.

Change-point analysis

The classification of echoes to 'canopy' and 'below-canopy' was based on a test of the null hypothesis (H_0) of one type of echoes versus the alternative hypothesis (H_A) of two types. The most powerful test for this purpose — and the data at hand — is a test of no change in the expected linear trend between the logarithm of Δz and the rank of Δz (Csörgő, 1997, p 199). A rejection leads to acceptance of H_A with echoes separated at a value of Δz that maximizes the test statistic (Kim and Siegmund, 1989). Our test statistic was the maximum change in Akaike's corrected information criterion ($\mathrm{AIC_c}(H_0)-\mathrm{AIC_c}(H_A)$); when it was larger than 50 we accepted the proposed separation of echoes (Vaida and Blanchard, 2005) because it strongly suggests the presence of a change-point (McQuarrie and Tsai, 1998, Section 2.3.1). Cluster-analysis (Everitt et al., 2001) and the EM (Magnussen and Boudewyn, 1998) separation methods were also tried. They were uniformly inferior because they frequently identified a separation close to the ground or at the top of the canopy, values that for the studied plots seemed implausible.

Summary statistics for P_{ground} and P_{canopy}

Target pulse density (λ) had little effect on canopy openness viz. P_{ground}. The maximum difference between any two densities was always less than 0.03 when averaged over replications. The average for all densities was 0.71. Plots, however, differed significantly in their canopy openness (Chi-square test, $P<0.001$) with P_{ground} ranging from 0.53 to 0.86.

Canopy density viz. P_{canopy} was, on average over plots and replications, 0.73 in the lowest target pulse density and 0.76 in the highest. An analysis of variance with arcsine transformed proportions failed to reject the null hypothesis of no density effect ($\hat{F}=0.31$, $P=0.87$). Plot averages (over 100 replications), on the other hand, showed large and significant differences (Chi-square test, $P<0.001$) from a low of 0.31 to a high of 0.95. Low averages were generally associated with low BA and TPH, and high averages with high BA and TPH. It was noted that the replication variance of P_{canopy} declined with increasing target pulse density at a rate inversely proportional to the number of 'canopy' echoes.

Estimating the reliability of X

In our study with 100 replications of X for every plot (and target pulse density) the among-plot variance σ_{xx} and reliability ratio κ_x of X is estimated from an analysis of variance as the ratio of the among-plot varianceof X to the total variance of X. Specifically $\hat{\kappa}_x=\hat{\sigma}_{x'x'}(\hat{\sigma}_{x'x'}+\hat{\sigma}_{uu})^{-1}$ where $\hat{\sigma}_{uu}$ is the estimate of the average within-plot replicate variance.

During these analyses we discovered a few (1.2%) aberrant X-values in 27 plots with a disproportionate large influence on the above variance components and reliability ratios. An X-value in the ith plot and jth replicate was deemed an outlier if $(x_{ij}-\tilde{x}_i)/Mean|x_{ij}-\tilde{x}_i|>t_{99,0.99}=3.86$ where $\tilde{x}_i$ is the 5% trimmed mean of x_{ij} (Atkinson, 1994). Outliers were replaced by imputations obtained via linear extrapolation of a plot-specific rank-based interpolation-function fitted to non-outlying values.

Without actual data replications estimation of σ_{xx} and κ_x proceeds by some form of resampling of the available data (Efron and Tibshirani, 1993, ch 4). Providers of ALS data have usually separated echoes returning from an estimated ground-surface from those returning from non-ground surfaces. We therefore illustrate how to apply a modified bootstrap procedure to estimate σ_{xx} and κ_x from non-replicated observations of Δz.

In each bootstrap replication we repeat the separation of echoes to 'canopy' and 'below-canopy' by the above detailed change-point analysis and then obtained $\overline{\Delta z}^{\gamma}$, P_{ground}, P_{canopy}, X_{HT}, X_{BA},X_{VOL}, X_{TPH} as before. For the sake of demonstration we completed a total of 100 bootstrap samples for each plot, target density, and the first six replications generated by the random thinning of the original ALS data.

In a first attempt we used a conventional bootstrap procedure. To mimic errors in the underlying DTM model we added to each drawn Δz-value a random Gaussian residual ($\varepsilon_{\mathrm{DTM}}$)with a mean of zero and a target-density specific variance (e.g. 0.2 m^2 for $\lambda=0.25\ \mathrm{m}^{-2}$ and 0.1 m^2 for $\lambda=2.0\ \mathrm{m}^{-2}$) obtained from our analysis of the reliability of the estimated DTMs.

The unconstrained bootstrap procedure generated estimates of the replication variance that were systematically too large compared to the replication variances estimated from the replicated data. The overestimation reached 40% for $\mathrm{HT_{Lor}}$ and BA, and 25% for VOL and TPH. A modified resampling procedure resolved the bias problem. In the modified procedure 95% of the pseudo−observations in a bootstrap sample came from a constrained resampling procedure and 5% from an unconstrained procedure. To generate the pseudo data from the constrained procedure we estimated for each bootstrap replicate: *i*) a linear interpolation function for the empirical cumulative distribution function (*ecdf*) of $\Delta z+\varepsilon_{\mathrm{DTM}}$ and *ii*) a simultaneous 95% confidence intervals (SCI95) around the *ecdf* as outlined in Owen (1995). This gives a (simultaneous) 95% percentile confidence interval for each value of $\Delta z+\varepsilon_{\mathrm{DTM}}$. Random draws from the percentile intervals provided random percentiles that – in turn – were converted to quantiles of vegetation height via a cubic-spline approximation to the inverse of the *ecdf* interpolation function (De Boor et al., 1978). The unconstrained part of each bootstrap sample (i.e. 5%) was generated by random sampling via the inverse to the ecdf interpolation function. In this modified bootstrap procedure each value $\Delta z+\varepsilon_{\mathrm{DTM}}$ is regarded as an expected value (over an infinite number of replications). From that follows that the variation in the

drawn quantiles only represents the expected variance; the variance of the expectations is ignored. Because each observation is a sample of size one and sampling is directed at percentiles with a binomial distribution (Serfling, 1980) the variance of the expectations is equal to the expected variance. Consequently the modified bootstrap estimates of the replication variance should be expected to be downward biased by approximately 50%. Accordingly we doubled all bootstrap estimates of variance.

Calibrated generalized least squares (CGLS)

With estimates of the reliability of X available to the analyst it becomes possible to obtain calibrated estimates of the regression coefficients in Eq. (2). Let $\hat{\sigma}_{uu}(x)$ and $\hat{\sigma}_{x'x'}(x)$ denote trait-specific estimates of the replication variance and the among-plot variance of X. The calibrated GLS (CGLS) estimates are the values of $\beta_{0,kj}$ and $\beta_{x,kj}$ that minimize the following weighted sum of squared residuals (Fuller, 1987, p 187)

$$SSQ_{CGLS}(\beta_{0,j},\beta_{x,j}|\hat{\theta}) = (\mathbf{Y}_i-\beta_{0,j}+\beta_{x,j}\hat{\mathbf{X}}_{ij})'Diag\,(\hat{\theta}_j)^{-1}(\mathbf{Y}_i-\beta_{0,j}+\beta_{x,j}\hat{\mathbf{X}}_{ij})$$

$$\hat{\theta}_{x_t,j} = \hat{\sigma}_{\varepsilon\varepsilon,j}(x_t)-\hat{\beta}^2_{x,j}(\hat{\sigma}_{uu}(x_t)+\hat{\sigma}_{x'x'}(x_t))(\hat{\sigma}_{uu}(x_t)\hat{\sigma}^{-1}_{x'x'}(x_t)+1),$$

$$t=\{HT,VOL,BA,TPH\} \quad (3)$$

where $Diag(\cdot)$ in our case is a 4×4 diagonal matrix with non-zero elements on its diagonal and 0 elsewhere. The idea behind the calibration is to remove from the residual variance the effect of the replication variance on the regression slope (attenuation) and the residual variance (confounding) effects. As for GLS, the iteration begins with ordinary least squares estimates of the regression coefficients and the residual variances $\sigma_{\varepsilon\varepsilon}(x)$ of the four inventory attributes. We only use precision weighting in (4) since attempts to use a full covariance matrix $\hat{\theta}_{x_t,x'_t,t}=\{HT,VOL,BA,TPH\}$ frequently resulted in unstable estimates or non-positive definite covariance matrices. The variance of a CGLS-prediction of a new trait-specific Y is

$$\widehat{var}_{CGLS}(Y_{new}|x_{new}) = \hat{\sigma}_{\varepsilon\varepsilon}(CGLS_x)\left[1+\frac{1}{n}+\frac{(x_{new}-\bar{x})^2}{\hat{\sigma}_{xx}(n-1)}\right] + \hat{\beta}^2_{CGLS_x}\hat{\sigma}_{uu}(x_{new}) + 2\frac{\hat{\beta}_{CGLS_x}cov(x_i,\hat{\sigma}^{0.5}_{uu}(x_i))}{n} \quad (4)$$

where $\hat{\sigma}_{uu}(x_{new})$ is an estimate of the replication variance associated with x_{new}. In case of homogenous replication variances the last term vanishes. In the current study it does not vanish because we noticed a positive covariance between X and its estimated replication variance. Note, if a calibrated x_{new} is used as a predictor the second term in Eq. (4) drops out (Carroll et al., 1995, p 50). A CGLS prediction of Y was made for each of the four inventory attributes, and for each plot, replicate, and target pulse density in a leave-one-out cross-validation scheme. Predictor values were not calibrated.

References

Andersen, H. E., Reutebuch, S. E., & McGaughey, R. J. (2006). A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. *Canadian Journal of Remote Sensing*, *32*, 355−366.

Anderson, J. E., Plourde, L. C., Martin, M. E., Braswell, B. H., Smith, M. L., Dubayah, R. O., Hofton, M. A., & Blair, J. B. (2008). Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment*, *112*, 1856−1870.

Anon. (1999). Pinnacle user's manual. *Javad Positioning Systems. San Jose, CA* 123 pp.

Anon. (2005). TerraScan user's guide Jyvaskyla, Finland: Terrasolid Ltd 169 pp.

Atkinson, A. C. (1994). Fast very robust methods for the detection of multiple outliers. *Journal of the American Statistical Association*, *89*, 1329−1339.

Axelsson, P. (2000). DEM generation from laser scanner data using adaptive TIN models. *Photogrammetric and remote sensing archives*, *33 Part B4/1 Part 4*, 110−117.

Bannerjee, S., Carlin, B. P., Gelfand, A. E., Isham, V., Keiding, N., Louis, T., Reid, N., Tibshirani, R., & Tong, H. (2004). Hierarchical modelling and analysis for spatial data Boca Raton: Chapman & Hall/CRC.

Braastad, H. (1966). Volume tables for birch. (In Norwegian with English summary). *Meddelser Norske Skogforsøgsvesen*, *Vol. 21*, 265−365.

Brantseg, A. (1967). Volume functions and tables for Scots pine in southern Norway. (In Norwegian with English summary). *Meddelser Norske Skogforsøgsvesen*, *Vol. 22*, 695−739.

Breidenbach, J., Koch, B., Kindler, G., & Kleusberg, A. (2008). Quantifying the influence of slope, aspect, crown shape and stem density on the estimation of tree height at plot level using lidar and InSAR data. *International Journal of Remote Sensing*, *29*, 1511−1536.

Brown, P. J. (1982). Multivariate calibration. *Journal Royal Statistical Society, Series B*, *44*, 287−321.

Carroll, R. J., Ruppert, D., & Stefanski, L. A. (1995). *Measurement error in nonlinear models.* London: Chapman & Hall.

Chasmer, L., Hopkinson, C., & Treitz, P. (2006). Investigating laser pulse penetration through a conifer canopy by integrating airborne and terrestrial lidar. *Canadian Journal of Remote Sensing*, *32*, 116−125.

Clawges, R., Vierling, K., Vierling, L., & Rowell, E. (2008). The use of airborne lidar to assess avian species diversity, density, and occurrence in a pine/aspen forest. *Remote Sensing of Environment*, *112*, 2064−2073.

Cook, J. R., & Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, *89*, 1314.

Coops, N., Wulder, M. A., Culvenor, D. S., & St Onge, B. (2004). Comparison of forest attributes extracted from fine spatial resolution multispectral and lidar data. *Canadian Journal of Remote Sensing*, *30*, 855−866.

Csörgő, M. H. L. (1997). Limit theorems in change-point analysis Chicester: Wiley.

De Boor, C., John, F., LaSalle, J. P., Sirovich, L., & Whitham, G. B. (1978). A practical guide to splines New York: Springer.

Draper, N. R., & Smith, H. (1981). Applied regression analysis New York: Wiley.

Efron, B., & Tibshirani, R. J. (1993). An introduction to the bootstrap Boca Raton: Chapman & Hall.

Everitt, B. S., Landau, S., & Leese, M. (2001). Cluster analysis London: Arnold.

Falkowski, M. J., Evans, J. S., Martinuzzi, S., Gessler, P. E., & Hudak, A. T. (2009). Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA. *Remote Sensing of Environment*, *113*, 946−956.

Fomby, T. B., Hill, R. C., & Johnson, S. R. (1980). *Advanced econometric methods.* New York: Springer-Verlag.

Fuller, W. A. (1987). *Measurement error models.* New York: Wiley.

Gobakken, T., & Næsset, E. (2005). Weibull and percentile models for lidar-based estimation of basal area distribution. *Scandinavian Journal of Forest Research*, *20*, 490−502.

Gobakken, T., & Næsset, E. (2008). Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research*, *38*, 1095−1109.

Goodwin, N. R., Coops, N. C., & Culvenor, D. S. (2006). Assessment of forest structure with airborne LiDAR and the effects of platform altitude. *Remote Sensing of Environment*, *103*, 140−152.

Goodwin, N. R., Coops, N. C., & Culvenor, D. S. (2007). Development of a simulation model to predict LiDAR interception in forested environments. *Remote Sensing of Environment*, *111*, 481−492.

Hall, S. A., Burke, I. C., Box, D. O., Kaufmann, M. R., & Stoker, J. M. (2005). Estimating stand structure using discrete-return lidar: an example from low density, fire prone ponderosa pine forests. *Forest Ecology and Management*, *208*, 189−209.

Harter, H. L. (1970). Order statistics and their use in testing and estimation. Vol. II Washington D.C.: US Government Printing Office 172 pp.

Hodgson, M. E., & Bresnahan, P. (2004). Accuracy of airborne lidar-derived elevation: Empirical assessment and error budget. *Photogrammetric Engineering and Remote Sensing*, *70*, 331−340.

Holmgren, J. (2004). Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scandinavian Journal of Forest Research*, *19*, 543−553.

Holmgren, J., Nilsson, M., & Olsson, H. (2003). Simulating the effects of lidar scanning angle for estimation of mean tree height and canopy closure. *Canadian Journal of Remote Sensing*, *29*, 623−632.

Hopkinson, C. (2007). The influence of flying altitude, beam divergence, and pulse repetition frequency on laser pulse return intensity and canopy frequency distribution. *Canadian Journal of Remote Sensing*, *33*, 312−324.

Hopkinson, C., Chasmer, L., & Hall, R. J. (2008). The uncertainty in conifer plantation growth prediction from multi-temporal lidar datasets. *Remote Sensing of Environment*, *112*, 1168−1180.

Hopkinson, C., Chasmer, L. E., Sass, G., Creed, I. F., Sitar, M., Kalbfleisch, W., & Treitz, P. (2005). Vegetation class dependent errors in lidar ground elevation and canopy height estimates in a boreal wetland environment. *Canadian Journal of Remote Sensing*, *31*, 191−206.

Hudak, A. T., Crookston, N. L., Evans, J. S., Falkowski, M. J., Smith, A. F. M., Gessler, P. E., & Morgan, P. (2006). Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. *Canadian Journal of Remote Sensing*, *32*, 126−138.

Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X., & Maltamo, M. (2008). Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, *29*, 1339−1366.

Jensen, J. L. R., Humes, K. S., Conner, T., Williams, C. J., & DeGroot, J. (2006). Estimation of biophysical characteristics for highly variable mixed-conifer stands using small-footprint lidar. *Canadian Journal of Forest Research*, *36*, 1129−1138.

Jensen, J. L. R., Humes, K. S., Vierling, L. A., & Hudak, A. T. (2008). Discrete return lidar-based prediction of leaf area index in two conifer forests. *Remote Sensing of Environment*, *112*, 3947−3957.
Junttila, V., Maltamo, M., & Kauranne, T. (2008). Sparse Bayesian estimation of forest stand characteristics from airborne laser scanning. *Forest Science*, *54*, 543−552.
Kim, H. -J., & Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression. *Biometrika*, *76*, 409−423.
Kobler, A., Pfeifer, N., Ogrinc, P., Todorovski, L., Ostir, K., & Dzeroski, S. (2007). Repetitive interpolation: A robust algorithm for DTM generation from aerial laser scanner data in forested terrain. *Remote Sensing of Environment*, *108*, 9−23.
Korpela, I. (2006). Geometrically accurate time series of archived aerial images and airborne lidar data in a forest environment. *Silva Fennica*, *40*, 109−126.
Kulathinal, S. B., Kuulasmaa, K., & Gasbarra, D. (2002). Estimation of an errors-in-variables regression model when the variances of the measurement errors vary between the observations. *Statistics in Medicine*, *21*, 1089−1101.
Lahiri, S. N., Kaiser, M. S., Cressie, N. A. C., & Hsu, N. J. (1999). Prediction of spatial cumulative distribution functions using subsampling. *Journal of the American Statistical Association*, *94*, 86−110.
Laster, L. L., Pickands, J., & Li, I. (1993). Repeated measurement design with random selections. *Biometrics*, *49*, 75−83.
Leckie, D., Gougeon, F., Hill, D., Quinn, R., Armstrong, L., & Shreenan, R. (2003). Combined high-density lidar and multispectral imagery for individual tree crown analysis. *Canadian Journal of Remote Sensing*, *29*, 633−649.
Lefsky, M. A., Harding, D., Cohen, W. B., Parker, G., & Shugart, H. H. (1999). Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of Environment*, *67*, 83−98.
Lehmann, E. L., & Casella, G. (1998). Theory of point estimation New York: Springer.
Lesser, V. M., & Kalsbeek, W. D. (1999). Nonsampling errors in environmental surveys. *Journal of Agricultural Biological and Environmental Statistics*, *4*, 473−488.
Lichstein, J. W., Simons, T. R., Shriner, S. A., & Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, *72*, 445−463.
Lin, X., & Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, *95*, 520−534.
Lin, Y., & Jeon, Y. (2006). Random forests and adaptive nearest neighbour methods. *Journal of the American Statistical Association*, *101*, 578−590.
Loetsch, F., Zöhrer, F., & Haller, K. E. (1973). Forest inventory Munich: BLV-Verlagsgesellschaft.
Longford, N. T. (1993). Random coefficients models New York: Clarendon.
Lovell, J. L., Jupp, D. L. B., Newnham, G. J., Coops, N. C., & Culvenor, D. S. (2005). Simulation study for finding optimal lidar acquisition parameters for forest height retrieval. *Forest Ecology and Management*, *214*, 398−412.
Magnussen, S. (1999). Effect of plot size on estimates of top height in Douglas-fir. *Western Journal of Applied Forestry*, *14*, 17−27.
Magnussen, S., & Boudewyn, P. (1998). Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research*, *28*, 1016−1031.
Magnussen, S., Eggermont, P., & LaRiccia, V. N. (1999). Recovering tree heights from airborne laser scanner data. *Forest Science*, *45*, 407−422.
Magnusson, M., Fransson, J. E. S., & Holmgren, J. (2007). Effects on estimation accuracy of forest variables using different pulse density of laser data. *Forest Science*, *53*, 619−626.
Maltamo, M., Suvanto, A., & Packalén, P. (2007). Comparison of basal area and stem frequency diameter distribution modelling using airborne laser scanner data and calibration estimation. *Forest Ecology and Management*, *247*, 26−34.
McCullagh, P., & Nelder, J. A. (1989). Generalized linear models London: Chapman & Hall.
McQuarrie, A. D. R., & Tsai, C. L. (1998). Regression and time series model selection Singapore: World Scientific.
Morsdorf, F., Frey, O., Meier, E., Itten, K. I., & Allgower, B. (2008). Assessment of the influence of flying altitude and scan angle on biophysical vegetation products derived from airborne laser scanning. *International Journal of Remote Sensing*, *29*, 1387−1406.
Næsset, E. (1997). Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *52*, 49−56.
Næsset, E. (2004). Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. *Remote Sensing of Environment*, *91*, 243−255.
Næsset, E. (2004). Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. *Scandinavian Journal of Forest Research*, *19*, 164−179.
Næsset, E. (2005). Assessing sensor effects and effects of leaf-off and leaf-on canopy conditions on biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, *98*, 356−370.
Næsset, E. (2009). Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, *113*, 148−159.
Næsset, E., Bollandsås, O. M., & Gobakken, T. (2005). Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. *Remote Sensing of Environment*, *94*, 541−553.
Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sensing of Environment*, *56*, 1−7.
Nordman, D. J., & Lahiri, S. N. (2004). On optimal spatial subsample size for variance estimation. *Annals of Statistics*, *32*, 1981−2027.
Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. (2000). Spatial tessellations. Concepts and applications of Voronoï diagrams Chicester: Wiley.
Owen, A. B. (1995). Nonparametric likelihood confidence bands for a distribution function. *Journal of the American Statistical Association*, *90*, 516−521.
Pascual, C., García-Abril, A., García-Montero, L. G., Martín-Fernández, S., & Cohen, W. B. (2008). Object-based semi-automatic approach for forest structure characterization using lidar data in heterogeneous *Pinus sylvestris* stands. *Forest Ecology and Management*, *255*, 3677−3685.
Popescu, S. C., & Zhao, K. (2008). A voxel-based lidar method for estimating crown base height for deciduous and pine trees. *Remote Sensing of Environment*, *112*, 767−781.
Reutebuch, S. E., McGaughey, R. J., Andersen, H. E., & Carson, W. W. (2003). Accuracy of a high-resolution lidar terrain model under a conifer forest canopy. *Canadian Journal of Remote Sensing*, *29*, 527−535.
Rombouts, J., Ferguson, T. S., & Leech, J. (2008). Campaign and site effects in LiDAR prediction models for site quality assessment of radiata pine plantations in south Australia. In R. A. Hill, J. Rosette, & J. Suárez (Eds.), *Proceedings of SilviLaser 2008: 8th international conference on LiDAR applications in forest assessment and inventory*. Edinburgh, UK. pp 39–48.
Schafer, D. W., & Purdy, K. G. (1996). Likelihood analysis for errors-in-variables regression with replicate measurements. *Biometrika*, *83*, 813−824.
Serfling, R. J. (1980). Approximation theorems of mathematical statistics New York: J. Wiley Sons.
Sexton, J. O., Bax, T., Siqueira, P., Swenson, J. J., & Hensley, S. (2009). A comparison of lidar, radar, and field measurements of canopy height in pine and hardwood forests of southeastern North America. *Forest Ecology and Management*, *257*, 1136−1147.
Shoemaker, L. H. (2003). Fixing the *F* test for equal variances. *The American Statistician*, *57*, 105−114.
Silverman, B. W. (1986). Density estimation for statistics and data analysis London: Chapman & Hall.
Solberg, S., Næsset, E., & Bollandsås, O. M. (2006). Single-tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. *Photogrammetric Engineering & Remote Sensing*, *72*, 1369−1378.
St-Onge, B., Hu, Y., & Vega, C. (2008). Mapping the height and above-ground biomass of a mixed forest using lidar and stereo Ikonos images. *International Journal of Remote Sensing*, *29*, 1277−1294.
Staudenmayer, J., Ruppert, D., & Buonaccorsi, J. R. (2008). Density estimation in the presence of heteroscedastic measurement error. *Journal of the American Statistical Association*, *103*, 726−736.
Stefanski, L. A., & Cook, J. R. (1995). Simulation-extrapolation: The measurement error Jackknife. *Journal of the American Statistical Association*, *90*, 1247−1256.
Suh, E. Y., & Schafer, D. W. (2002). Semiparametric maximum likelihood for nonlinear regression with measurement errors. *Biometrics*, *58*, 448−453.
Thomas, V., Treitz, P., McCaughey, J. H., & Morrison, I. (2006). Mapping stand-level forest biophysical variables for a mixedwood boreal forest using lidar: An examination of scanning density. *Canadian Journal of Forest Research*, *36*, 34−47.
Tosteson, T. D., Buzas, J. S., Demidenko, E., & Karagas, M. (2003). Power and sample size calculations for generalized regression models with covariate measurement error. *Statistics in Medicine*, *22*, 1069−1082.
Vaida, F., & Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, *92*, 351−370.
Vepakomma, U., St-Onge, B., & Kneeshaw, D. (2008). Spatially explicit characterization of boreal forest gap dynamics using multi-temporal lidar data. *Remote Sensing of Environment*, *112*, 2326−2340.
Vestjordet, E. (1967). Functions and tables for volume of standing trees. Norway spruce. (In Norwegian with English summary). *Meddelser Norske Skogforsøgsvesen* 543–574 pp.
Wulder, M. A., Magnussen, S., Harding, D., Coops, N. C., Boudewyn, D., & Seemann, D. (2008). Stability of surface LiDAR height estimates on a point and polygon basis. *Journal of Forest Planning*, *13*, 279−286.
Zhao, K., Popescu, S., & Nelson, R. (2009). Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment*, *113*, 182−196.