

Article

Comparing Empirical and Semi-Empirical Approaches to Forest Biomass Modelling in Different Biomes Using Airborne Laser Scanner Data

Endre H. Hansen ^{1,*}, Liviu T. Ene ¹, Ernest W. Mauya ^{1,2}, Zdeněk Patočka ³, Tomáš Mikita ³,
Terje Gobakken ¹ and Erik Næsset ¹

¹ Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, NO-1432 Ås, Norway; liviu.ene@nmbu.no (L.T.E.); ernestmauya@gmail.com (E.W.M.); terje.gobakken@nmbu.no (T.G.); erik.naeset@nmbu.no (E.N.)

² Department of Forest Mensuration and Management, Sokoine University of Agriculture, P.O. Box 3013, Chuo Kikuu, Morogoro, Tanzania

³ Department of Forest Management and Applied Geoinformatics, Mendel University, Zemědělská 3, 613 00 Brno, Czech Republic; zdenek.patočka@mendelu.cz (Z.P.); tomas.mikita@mendelu.cz (T.M.)

* Correspondence: endre.hansen@nmbu.no; Tel.: +47-9544-3705

Academic Editors: Christian Ginzler and Lars T. Waser

Received: 10 March 2017; Accepted: 12 May 2017; Published: 16 May 2017

Abstract: Airborne laser scanner (ALS) data are used operationally to support field inventories and enhance the accuracy of forest biomass estimates. Modelling the relationship between ALS and field data is a fundamental step of such applications and the quality of the model is essential for the final accuracy of the estimates. Different modelling approaches and variable transformations have been advocated in the existing literature, but comparisons are few or non-existent. In the present study, two main approaches to modelling were compared: the empirical and semi-empirical approaches. Evaluation of model performance was conducted using a conventional evaluation criterion, i.e., the mean square deviation (MSD). In addition, a novel evaluation criterion, the model error (ME), was proposed. The ME was constructed by combining a MSD expression and a model-based variance estimate. For the empirical approach, multiple regression models were developed with two alternative transformation strategies: square root transformation of the response, and natural logarithmic transformation of both response and predictors. For the semi-empirical approach, a nonlinear regression of a power model form was chosen. Two alternative predictor variables, mean canopy height and top canopy height, were used separately. Results showed that the semi-empirical approach resulted in the smallest MSD in three of five study sites. The empirical approach resulted in smaller ME in the temperate and boreal biomes, while the semi-empirical approach resulted in smaller ME in the tropical biomes.

Keywords: airborne laser scanning; biomass modelling; light detection and ranging (LiDAR); model-based variance; model error

1. Introduction

Airborne laser scanning (ALS) has become an important source of auxiliary data for estimating tree heights, diameter distribution, timber volume, and forest biomass [1,2]. For forest inventory purposes, the ALS data are often applied within an estimation framework known as the area-based approach. In this method, introduced in Næsset [3] and further described in Næsset [4], models are constructed using spatially coincident observations of ground-based response values (e.g., biomass) measured on field plots and variables derived from ALS data. The models are subsequently used

to predict forest attributes for individual cells which tessellate the area of interest (AOI). The cell predictions are finally aggregated to estimates for the entire AOI or regions within the AOI.

Hollaus et al. [5] and Hollaus et al. [6] described three main approaches to modelling the relationship between forest attributes and the remotely sensed data: physical, empirical, and semi-empirical. While physical modelling approaches have had limited practical applications due to difficulties with inverting the models [5], both the empirical and semi-empirical approaches are commonly used. Following the empirical approach, a variety of features are extracted from the ALS echoes and related to field-based observations using statistical methods such as regression analysis, nearest neighbors, neural networks, or ensemble learning to model the relationship (e.g., [7–10]). Universal for these methods when used in the empirical approach is that the model selection is performed objectively. One of the most common methods, at least the most frequently implemented method in commercial operations in Norway, is ordinary least square regression (OLS). To improve the linear relationship between the response and predictors, transformations of the response or of both the response and predictors are often used. Two common transformation approaches are the square-root transformation (SQRT) of the response, and the logarithmic transformation (LOG) of both response and predictors. Comparisons of SQRT and LOG models generally show small differences in terms of model performance. Næsset [11], for instance, reported only small differences in model performance when modelling aboveground biomass (AGB) in young forests in Norway using SQRT, LOG, and linear model forms. Saarela et al. [12] found that SQRT models performed slightly better compared to linear and LOG models when using ALS data to model timber volumes in Finland. The use of SQRT and LOG transformations requires that model predictions are back-transformed to original scale. Although different methods of back-transformation of LOG models have been suggested (e.g., [13–15]), achieving unbiasedness is challenging [16]. For back-transformation of SQRT models, the commonly applied correction factor presented by Gregoire et al. [17] results in “negligibly small” biased predictions. Regression methods that render the need for transformation of the response obsolete have also been used e.g., OLS without transformation of the response [11], quadratic polynomials [18], nonlinear least square (NLS) [19], and generalized linear modelling (GLM) [20]. These methods have however not reached the same level of use in operational settings as the LOG and SQRT. This could be because of reported advantages with the LOG and SQRT methods, or with difficulties in applying the alternative methods.

Although less used in operational settings, the semi-empirical approach is common in research of the relationship between forest attributes and ALS-derived predictors (e.g., [5,21–23]). Following this approach, the model form and predictors are selected a priori, often based on the theoretical relationship between the response and predictor or predictors. Allometry of tree height and girth is often described by a power model [24,25] and is therefore commonly applied to model AGB, substituting tree height with ALS-derived variables and girth with AGB. This approach was taken by Asner, et al. [23] to model forest AGB on the island of Hawaii. The authors regressed AGB against mean canopy height derived from the ALS data using a NLS modelling procedure. Asner et al. [26] advocated the same approach following the argument of a theoretical relationship between mean canopy height and carbon stored in the AGB. Having only one predictor variable, this approach makes it easy to fit the NLS model using statistical software such as e.g., R [27]. Increasing the number of predictors makes it increasingly difficult to fit the model. Although Magnussen et al. [21] used two predictors, other authors have used OLS and LOG transformations to achieve the desired model form (e.g., [22,28]). The variables were however selected subjectively making the approach semi-empirical. Bouvier et al. [22] selected four predictor variables based on their theoretical complimentary properties describing the canopy height, heterogeneity of canopy height, canopy cover, and variation in leaf area density. Tompalski et al. [28] took a similar approach, however selecting the most correlated ALS-derived variable to each of the four model properties. Chen [29] modelled AGB in three different study sites in northwestern USA. He compared two power models, each with one predictor variable, one multiple OLS model with a logarithmic transformed response and three pre-selected predictors,

and two nonparametric methods. Results showed that the power model with the mean of all ALS echoes as predictor had the smallest estimated prediction error in two of the study sites, whilst the multiple OLS model form with three variables had the smallest estimated prediction errors in one study site.

Although the empirical approach has been successful and is currently used by commercial companies in many operational applications [30], the method has its limitations. In the presence of a large number of candidate predictors, with many of them being strongly correlated, there is a potential risk of multicollinearity problems that complicates the selection of model predictors [31] (Chapter 7). Furthermore, the empirical models are to some degree often affected by local effects such as geographical region, forest type, ALS acquisition parameters, and forest inventory design [5]. In this context, the semi-empirical approach has the advantage of simplifying the modelling step and enabling the re-use and re-calibration of the models with new ALS data [21].

The use of predictive models for supporting forest inventories is traditionally done within either a design-based or a model-based inferential framework. Even though the design-based framework has been predominant, the case for the model-based framework is now being made [32]. One of the main advantages of the model-based framework is that it does not, as opposed to design-based, rely on a probabilistic field sample. Instead, the inference is based on the model as a valid model for the distribution of Y_i random variables. The sampled values y_i are considered as a realization of the model [33] (p. 40). The model cannot be observed, but the parameters of the model can be estimated from the sample. Furthermore, the model-based framework provides more flexibility in terms of model transferability and small area estimation [32]. The variability of the predictive model can be evaluated by the mean square error (MSE). However, when used within the model-based inferential framework, the error of estimates for predictive models with small MSE can be larger compared to other candidate models that are nearly as good as the best one according to the MSE criterion (e.g., [12]). Evaluation of predictive models within the model-based inferential framework can instead be assessed in terms of (1) model specification to ensure that the model is correctly specified and (2) a model-based variance estimate (e.g., [34]). In order to evaluate the errors of predictive models, we studied a criterion that incorporates both a cross-validated MSE (i.e., mean square difference (MSD)) and a model-based variance estimate. The proposed evaluation criterion is referred to as the model error (ME).

While both the empirical and semi-empirical modelling approaches are used for AGB estimation, to the best of our knowledge, only the study by Hollaus et al. [5] has attempted to compare the two alternatives. The study by Hollaus et al. [5] found that the two approaches resulted in similar accuracy, and advocated the use of the semi-empirical approach based on its simplicity. Because the two approaches are commonly used, and with little scientific basis for choosing one over the other, it is of interest to assess possible differences in performance. It is also of interest to generalize the choice of approach, and the diverse data available for this study provided an opportunity for such generalizations.

The main objective of the present study was therefore to compare the two approaches in a variety of biomes. We used both MSD and ME as model evaluation criteria, and compared the performance of two empirical and two semi-empirical modelling approaches. The analyses were performed on five datasets representing four different forest biomes: tropical moist and -dry, temperate, and boreal forests.

2. Materials and Methods

2.1. Field Data

Data from five study sites from four different forest biomes were used in the present study (Table 1). These datasets containing 1182 field plots in total allowed for an extensive comparison of both the empirical and semi-empirical approaches.

The first dataset (S1) was collected in Amani Nature Reserve, northeastern Tanzania (5°08' S, 38°37' E). It represents a tropical moist forest biome. The area receives around 2000 mm rainfall per

year, and most of the rain falls in the two wet seasons, April–May and October–November. Daily mean temperatures vary from about 16 to 25 °C. The area is covered by both old-growth natural tropical forest with a multi-layered canopy and previously harvested areas with single layered, light demanding, pioneer species. Forest inventory data such as diameter at breast height (DBH), tree height (H), and tree species were collected on 153 rectangular plots of about 50 m × 20 m in size over a period of four years, 2008–2012 [35]. The horizontal area of the plots varies from 639 to 1239 m², with a mean of 914 m² (Table 1). The variation in plot size is due to the procedure of plot establishment where the plots were laid out along the terrain slope, without slope correction. A threshold of ≥ 10 cm was used for recording the DBH. The corner coordinates of each plot were established by means of differential global navigation satellite system (GNSS), using survey-grade receivers. Errors in the x–y coordinates of the plot corners were estimated to an average of 0.57 m based on random errors reported from the post-processing software [36] and empirical experience of the relationship between reported error and the true error documented by Næsset [37]. H was measured using Vertex IV hypsometer and a DBH–H model [35] (Equation (1)) was developed from the sample trees with height measurements collected on each plot and used to predict the H of all trees. AGB predicted for each individual tree in the sample was obtained using a local AGB model with DBH and H as independent variables [38] (p. 43). Further details about the field data can be found in Hansen et al. [35].

The second dataset (S2) was collected in a tropical dry forest during 2011 and the first quarter of 2012 [39]. The site is located in Liwale district (9°54' S, 37°38' E), southeastern Tanzania, and it is characterized by a climate with two rain seasons during November–January and March–May. The mean annual rainfall ranges between 600 and 1000 mm, and daily mean temperatures range between 20 and 30 °C. The forest structure has high diversity both in terms of structure because of varying soil conditions as well as a high species diversity. Dominant species included *Brachystegia* sp., *Julbernardia* sp., and *Pterocarpus angolensis*. Field plots were laid out in a clustered design as part of Tanzania's National Forestry Resources Monitoring and Assessment (NAFORMA) program [40]. To avoid problems with spatial auto-correlation, we selected the two plots farthest away from each other from each cluster for the present study. The minimum distance between plots was 905 m, and the average distance within a cluster was 1438 m. The semi-variance of predicted wood volume studied during the design of the NAFORMA sample survey is indicated to level off at a distance of around 250 m for most of the forest types in Tanzania [41]. Wood volume and AGB are strongly correlated variables. To test for an effect of the clustering, a log-likelihood ratio test between GLM models with and without a compound symmetry covariance structure that should account for heteroscedasticity was performed. The models were fit using maximum likelihood using the “nlme” [42] package in R [27]. The test showed no significant cluster effect. Collection of forest inventory data was performed on concentric circular plots of 1, 5, 10, and 15 m in radius, following the field protocol of NAFORMA [43]. Minimum thresholds for DBH for registration on the concentric plots were ≥ 1 , ≥ 5 , ≥ 10 , and ≥ 20 cm respectively. The center coordinate of each plot was established with the same equipment and procedure as in the first dataset. Errors in the x–y coordinates of the plot corners were estimated to an average of 0.19 m [39]. The height of every fifth tree was measured using a Suunto hypsometer and DBH–H models were developed for each stratum. These DBH–H models were used to predict the H of all trees. Tree AGB was predicted using allometric models developed for Tanzanian woodlands by Mugasha et al. [44]. Further details about the field data can be found in Mauya et al. [39].

The third dataset (S3) was collected in a temperate forest in the Czech Republic during January, July, and August of 2015 [45]. The study site was located in Mendel University Training Forest, Křtiny (49°17' N, 16°44' E). Mean annual rainfall is 600 mm and mean annual temperature is 7.5 °C. Tree species composition can be described as coniferous dominant with *Picea abies*, *Larix decidua*, and *Pinus sylvestris* making up 76%, 4%, and 2% of the trees, respectively. The rest of the trees are *Fagus sylvatica* (17%) and other species (2%). The canopy is generally closed and with only one layer. Forest inventory data were collected on 50 circular plots with radius 12.62 m. The plots were selected by stratified sampling using an existing forest management plan. Plot centers were determined by

GNSS using Topcon Hiper Pro with applied RTK corrections from the CZEPOS (Czech Positioning System) reference network. GNSS measurements were performed at five-second intervals for at least 20 minutes. On each plot, DBH and H of all trees above a minimum DBH ≥ 7 cm were measured. H was measured using a Vertex laser hypsometer. The age of forest stands was in the range of 60–130 years. Tree specific AGB was calculated using allometric models (*Alnus glutinosa*, *Fagus sylvatica*, *Picea abies*, and *Pseudotsuga menziesii* [46]; *Pinus sylvestris* and *Abies alba* [47]; *Larix decidua* [48]).

The fourth dataset (S4) was collected in a boreal forest located in Aurskog-Høland municipality (59°50' N, 11°30' E), southeastern Norway, during the fall of 2006 [49]. Conifers (*Picea abies* and *Pinus sylvestris*) dominate the forest. Some deciduous species are also present, mainly *Betula pubescens*, especially in younger stands. A total of 201 circular field plots of 200 m² in size were systematically distributed. Plot centers were determined by GNSS using Topcon LegacyE receivers observing pseudo-range and carrier phase of the global positioning system and global navigation satellite system. All trees on the field plots with DBH ≥ 4 cm were callipered and DBH was recorded in two cm diameter classes. An average of nine trees per plot were sampled for height measurement with a probability proportional to stem basal area. H was measured with a Vertex hypsometer. Volume of each sample tree was estimated by means of species-specific volume models of individual trees dependent on DBH and H [50–52]. A so-called tariff height for each tree on the plot using tariff height curves [53]. For trees with a field measured H and an H estimated from the tariff height curves, separate volumes were estimated using both heights. The ratio between the mean volume predicted from field measurements and the volume for the sample trees was then used to adjust the volumes of all trees. Using the ratio estimated volumes, H were predicted for trees without a field measured H using single tree volume models [50–52]. Trees with DBH <5 cm and H >1.3 m were counted and their H estimated by means of models presented by Tomter [54]. AGB of each component; stump, stem, bark, dead and living branches, and foliage of each tree was calculated using species-specific allometric models developed by Marklund [55] with DBH and field measured or predicted H as independent variables.

The fifth dataset (S5) was collected in Hedmark county (61°40' N, 11°40' E) during 2005–2007 [56] as part of the Norwegian national forest inventory. A total of 648 circular field plots of 250 m² in size were systematically distributed. Plot centers were determined by GNSS using Topcon LegacyE receivers observing pseudo-range and carrier phase of the global positioning system and global navigation satellite system for a minimum of 15 min. Forest conditions and dominant species in S5 are relatively similar to the forest in S4. The AGB density, however, is considerably lower (Table 1). All trees on the field plots with DBH ≥ 5 cm were callipered. An average of ten trees per plot were sampled for height measurement with a probability proportional to stem basal area. H was measured with a Vertex hypsometer. AGB for trees with DBH ≥ 5 cm was estimated using the same procedure as in S4. Trees with DBH <5 cm and H >1.3 m were counted and their H estimated by means of models presented by Tomter [54]. The original dataset covered both forest and non-forest areas. In the present study, only plots with registered AGB were used.

For all datasets, the AGB predictions of individual trees were summed for each plot and expanded to per hectare unit (Table 1).

2.2. ALS Data Acquisition and Initial Processing

The ALS datasets were acquired at different times and with different sensors and acquisition parameters. The acquisition parameters are summarized in Table 2. Post flight processing of the ALS data was performed using the TerraScan software (Terrasolid Ltd., Helsinki, Finland) [57]. A triangulated irregular network (TIN) surface was constructed using the algorithm presented by Axelsson [58]. Following the construction of the TIN, the elevation of each ALS echo relative to the TIN was computed, resulting in an elevation above the ground surface for each echo.

Table 1. Summary statistics for the five study sites.

| Site | Biome | Location | Position | N | Plot Size (m ²) | AGB (Mg·ha ^{−1}) | | | |
|------|----------------|----------------|--------------------|-----|-----------------------------|----------------------------|-----|-----|------|
| | | | | | | Mean | SD | Min | Max |
| S1 | Tropical moist | Tanzania | 5°08' S, 38°37' E | 153 | 914 | 462 | 207 | 43 | 1147 |
| S2 | Tropical dry | Tanzania | 9°54' S, 37°38' E | 130 | 707 | 67 | 54 | 0 | 350 |
| S3 | Temperate | Czech Republic | 49°17' N, 16°44' E | 50 | 500 | 323 | 79 | 84 | 493 |
| S4 | Boreal | Norway | 59°50' N, 11°30' E | 201 | 200 | 128 | 78 | 20 | 407 |
| S5 | Boreal | Norway | 61°40' N, 11°40' E | 648 | 250 | 64 | 66 | 0 | 405 |

AGB, aboveground biomass; SD, standard deviation.

Table 2. Airborne laser scanner (ALS) acquisition parameters.

| | Study Site | | | | |
|-----------------------------------|-----------------------|---------------------|-------------------------|---------------------|----------------------|
| | S1 | S2 | S3 | S4 | S5 |
| Sensor | Leica ALS70 | Leica ALS70 | Leica ALS70 | Optech ALTM 3100 | Optech ALTM 3100 |
| Date | January–February 2012 | February–March 2012 | September 2014 | June–September 2005 | July–September 2006 |
| Flight speed (m·s ^{−1}) | 70 | 77 | 70 | 75 | 75 |
| Flying altitude (m a.g.l.) | 800 | 1320 | 700 | 1850 | 800 |
| Pulse repetition frequency (kHz) | 339 | 193 | 302 | 50 | 55 |
| Reference | Hansen et al. [35] | Mauya et al. [39] | Patočka and Mikita [45] | Næsset et al. [49] | Gobakken et al. [56] |

2.3. ALS-Derived Predictor Variables

The Leica ALS70 sensors provided records of up to five echoes per pulse and the Optech ALTM 3100 sensor four echoes. In the present study, we categorized each echo as “single”, “first of many”, “intermediate”, or “last of many” based on the echo sequence. The “single” and “first of many” echoes were combined into one dataset and denoted as “first” while “single” and “last of many” were combined into another dataset and denoted as “last”. The separation of “first” and “last” echoes were based on the assumption that the echoes from a pulse that has already been intercepted by the canopy (“last”) provide divergent information from the echoes not previously intercepted (“first”), and is a common procedure for processing ALS data [59]. Predictor variables were derived from the vertical distribution of echoes for each plot. The variables used in the empirical approach included percentiles of echo elevation, relative canopy density, and mean (h_{mean}) and maximum (h_{max}) elevation. Height percentiles were calculated at 10% intervals ($h_{10}, h_{20}, \dots, h_{90}$). These variables were calculated using a canopy threshold of 1.3 m, frequently used to eliminate echoes from undergrowth and rocks. Canopy density was computed by dividing the height from the canopy threshold to the 95th percentile height into ten equal intervals as recommended by Næsset and Gobakken [60]. Cumulative proportions of echoes in the ten intervals to total number of echoes were calculated (d_0, d_1, \dots, d_9). Height percentiles and canopy density variables were calculated from the “first” and “last” data separately.

Studies following a semi-empirical approach have often used only one [5,26] or two [3,21] ALS-derived variables to model AGB or timber volumes. A variable capturing both the height and density of the canopy is often used. By calculating the height of ALS echoes without a canopy threshold, the height variables describe both the height and density of the canopy. Hollaus et al. [5] used a variable derived from “first” echoes only for modelling timber volumes in Austria, whereas others have used both a variable from all echoes [26] and a variable from first echoes [61] to model AGB in tropical biomes. Næsset [3] used the mean height of all echoes together with a canopy density variable to model timber volumes in eastern Norway. Furthermore, Magnussen et al. [21] found that complementing the canopy height of “first” echoes with a measure for the variance in canopy height reduced the RMSE with 11% when modelling timber volumes in eastern Norway. For simplistic reasons, and the ability to compute the heteroscedasticity consistent covariance matrix estimator described in Section 2.4.4, we chose to explore the approach used by e.g., Asner et al. [26] and Asner and Mascaro [61]. We therefore derived two variables from the ALS data to be used in the semi-empirical approach. Although the calculations are not identical, we use the terms of Asner et al. [26] and derived the mean canopy height (MCH) from all ALS echoes and the top canopy height (TCH) from ALS echoes in the “first” dataset.

2.4. Statistical Modelling

As described in the introduction, there exists a number of statistical methods which have been used in both empirical and semi-empirical approaches. We have chosen two common approaches in operational settings and research i.e., OLS used in the empirical approach (described in Section 2.4.1), and NLS used in the semi-empirical approach (described in Section 2.4.2).

2.4.1. Ordinary Least Square Modelling

Multiple regression models expressed as LOG functions are frequently used for estimating AGB with ALS data in the empirical approach (e.g., [60,62]). Another common transformation used for predictive AGB models is the SQRT (e.g., [11,12]). Both transformations are applied to improve the linear relationship between the response and the predictors and to mitigate heteroscedasticity in the models. Thus, the models used for the empirical approach were formulated as:

$$\ln(y_i) = \beta_0 + \ln(X_i) \times \beta + e_i \quad (1)$$

and

$$\text{sqrt}(y_i) = \beta_0 + X_i \times \beta + e_i \quad (2)$$

where y_i is field values of AGB in plot i , β_0 is the model intercept coefficient, e_i is the model residual for plot i , and β is the vector of model parameters associated with the X matrix of ALS predictors. Residuals were assumed to be normally distributed with mean zero and a constant variance ($e_i \sim N(0, \sigma_i^2)$). Selection of predictor variables was performed using a best subset regression procedure implemented in the “leaps” package [63] in R, constrained to include a maximum of three predictors in the models. To avoid overfitting and multicollinearity, the models were selected using the Bayesian information criterion and variance inflation factors were kept below 10. Models without cross validation (see Section 2.4.3) were reported for assessing selected predictors.

Transformation of the response introduce a bias when the predicted biomass is back-transformed to arithmetic scale. To correct for bias in the LOG models, the correction factor for the uniform minimum variance unbiased estimator [14] was applied. Initial bias correction showed that the correction factor presented by Bradu and Mundlak [14] gave significant bias ($p > 0.05$) at two study sites (S2 and S5). Thus, a correction factor presented by Snowdon [13] was applied for predictions in S2 and S5. SQRT models were back-transformed according to Gregoire et al. [17]. In the present study, one LOG (Equation (1)) and one SQRT (Equation (2)) model was fit for each study site. These are referred to as OLS_{LOG} and OLS_{SQRT} , respectively.

2.4.2. Nonlinear Regression

Instead of relying on transformations and multiple variables, the choice of modelling technique in the semi-empirical approach is often to use NLS. The use of the NLS technique enables the model to be fitted non-linearly through successive approximations by which only initial starting values for the approximation have to be stated [64] (Chapter 8). With several parameters to be estimated, the selection of starting values becomes difficult. Taking the semi-empirical approach with only one predictor, selection of starting values is simplified. We fit two nonlinear power models using the standard “nls” procedure in R with models formulated as:

$$y_i = a \times H_i^b + e_i \quad (3)$$

where a and b are the model parameters to be estimated, and H is the MCH and TCH used in separate models denoted NLS_{MCH} and NLS_{TCH} , respectively.

2.4.3. Model Evaluation Criteria

We first assessed the models based on the mean square deviation (MSD, Equation (4)) produced by a 10-fold cross validation procedure with:

$$MSD = \sum_{k=1}^K \frac{n_k}{n} MSE_k \quad (4)$$

where K is the total number of folds (10), n is the total number of observations, n_k is the number of observations in the k -th fold ($k = 1, 2, \dots, K$), MSE_k is the mean square error of AGB predictions in the k -th fold, MSD is the mean square deviation. Successful use of the MSD as a model criterion assumes a design-based estimation strategy as the criterion is based on the observations and predictions on the sample units. We therefore sought to construct a model evaluation criterion that could be more informative than the MSE in cases where the field sample is not a probability sample. This was done by incorporating a model-based variance estimate based on a 10-fold cross validation. This criterion is referred to as the model error (ME, Equation (5)). The ME criterion combines both an expression of the MSD (Equation (5), first term) and a model-based variance estimate (Equation (5), second term):

$$ME = \sqrt{\left(\frac{MSD}{(n-1)DF}\right)^2 + \hat{var}(\hat{\mu})^2} \quad (5)$$

where DF is the model degrees of freedom and $\hat{\text{var}}(\hat{\mu})$ is a model-based variance estimate calculated from the cross validation as:

$$\text{var}(\hat{\mu}) = \sum_{k=1}^K \bar{\mathbf{X}}_k' \text{var}(\hat{\beta}_k) \bar{\mathbf{X}}_k = \sum_{k=1}^K (\mathbf{X}_k' \mathbf{X}_k)^{-1} \mathbf{X}_k' \mathbf{V}_k (\mathbf{X}_k' \mathbf{X}_k)^{-1} \quad (6)$$

where \mathbf{X}_k is a matrix of variables derived from the ALS data in the k -th fold, $\bar{\mathbf{X}}_k \approx \partial f(\bar{\mathbf{X}}_k; \beta_k) / \partial \beta_k$ is a matrix of the approximated mean variables of \mathbf{X}_k in the k -th fold, and \mathbf{V}_k is a matrix containing the weights of the observations in the diagonal cells and the error correlations in the off diagonal cells in the k -th fold.

2.4.4. Covariance Matrix Estimators

A central part of the variance estimator in Equation 6 is the matrix of estimated covariances between the parameter estimates $\hat{\beta}$ of the regression model. If the model errors are homoscedastic ($\mathbf{V} = \sigma^2 \mathbf{I}$), Equation (6) simplifies to:

$$\hat{\text{var}}(\hat{\beta}) = \hat{\sigma}^2 (\mathbf{X}' \mathbf{X})^{-1} \quad (7)$$

In the case of AGB modelling however, the variance of the response is likely to be related to the mean of the response, and thus we often have heteroscedasticity in our regression models. In this case, Long and Ervin [65] recommended using a heteroscedasticity consistent covariance matrix. OLS models with significant ($p < 0.05$) heteroscedasticity were identified using the Breusch–Pagan test [66]. Residual plots were used to visually assess heteroscedasticity in the NLS models. In the presence of heteroscedasticity, heteroscedasticity consistent covariance matrix estimators were used as estimators of $\text{var}(\hat{\beta})$. For the OLS models, heteroscedasticity consistent covariance matrix estimators of type HC₃, presented by MacKinnon and White [67], were computed using the “sandwich” package [68] in R. The HC₃ (Equation (8)) estimator is frequently used and recommended for small samples [65]. Computation of the HC₃ estimator requires the model projection matrix. Since this is not available for NLS models, a nonlinear heteroscedasticity consistent covariance matrix estimator described by White [69] (p. 821) was therefore adopted for a nonlinear heteroscedasticity consistent covariance matrix estimator (NHC), Equation (9). The nonlinear NHC estimator was implemented in R and used to compute the final variance estimate in the presence of heteroscedasticity:

$$\text{HC}_3 = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \text{diag} \left(\frac{\hat{e}_i^2}{(1 - p_{ii})^2} \right) \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \quad (8)$$

$$\text{NHC} \cong \sum \hat{e}_i^2 (\mathbf{Z}' \mathbf{Z})^{-1} \quad (9)$$

where $\mathbf{Z} \approx \partial f(\mathbf{X}; \beta) / \partial \beta$ is a matrix of the approximated partial derivatives of the NLS model, \hat{e}_i^2 is the estimated residuals squared, and p_{ii} is the diagonal elements of the projection matrix.

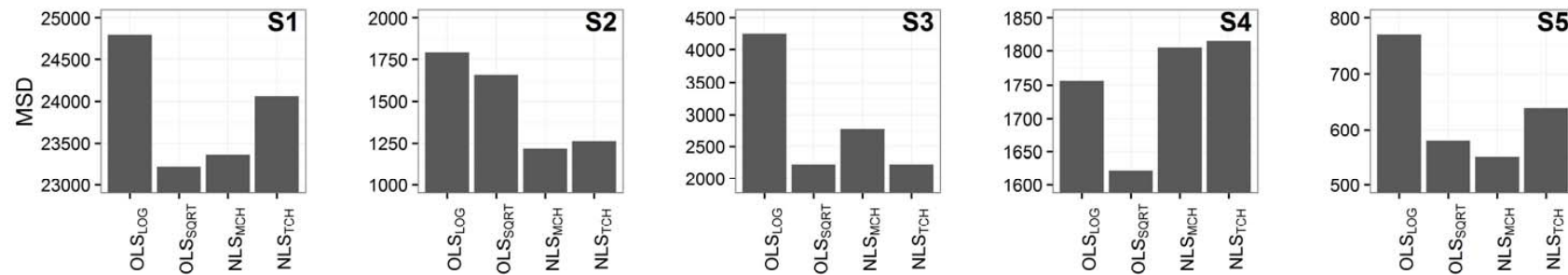


Figure 1. Calculated mean square deviation (MSD) for S1, S2, S3, S4, and S5. OLS_{LOG} and OLS_{SQRT} represent the empirical approach to modelling. NLS_{MCH} and NLS_{TCH} represent the semi-empirical approach to modelling.

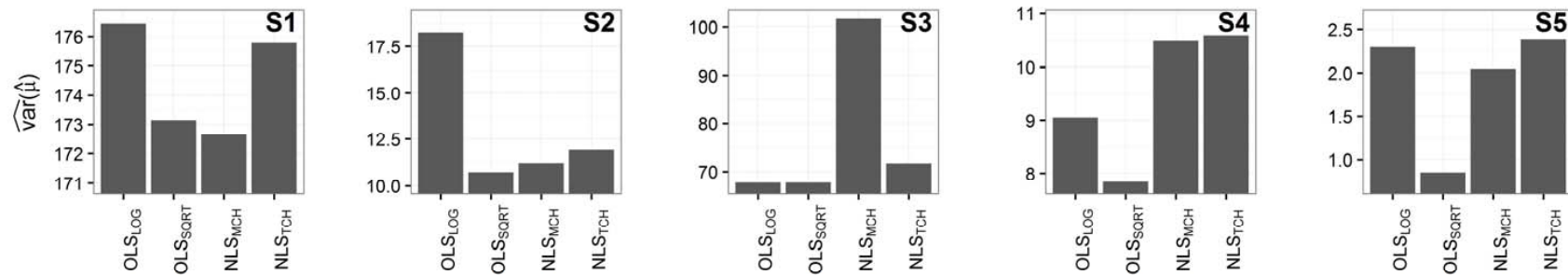


Figure 2. Calculated model-based variance estimates ($\widehat{\text{var}}(\hat{\mu})$) for S1, S2, S3, S4, and S5. OLS_{LOG} and OLS_{SQRT} represent the empirical approach to modelling. NLS_{MCH} and NLS_{TCH} represent the semi-empirical approach to modelling.

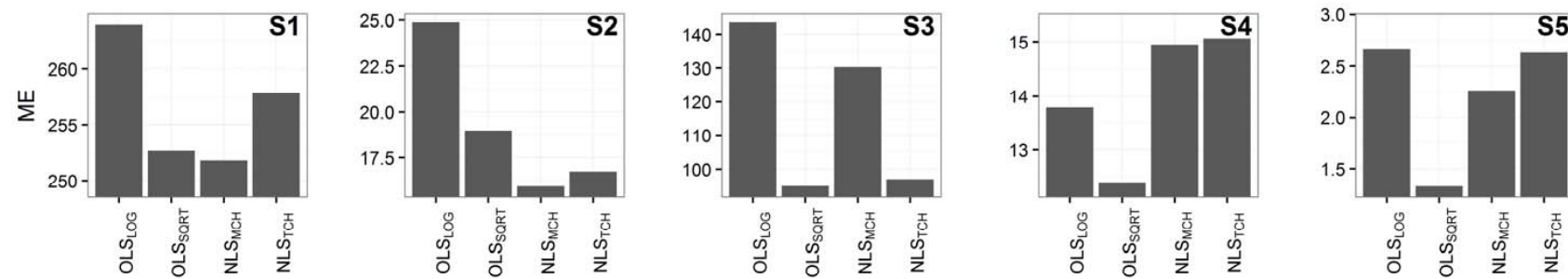


Figure 3. Calculated model error (ME) for S1, S2, S3, S4, and S5. OLS_{LOG} and OLS_{SORT} represent the empirical approach to modelling. NLS_{MCH} and NLS_{TCH} represent the semi-empirical approach to modelling.

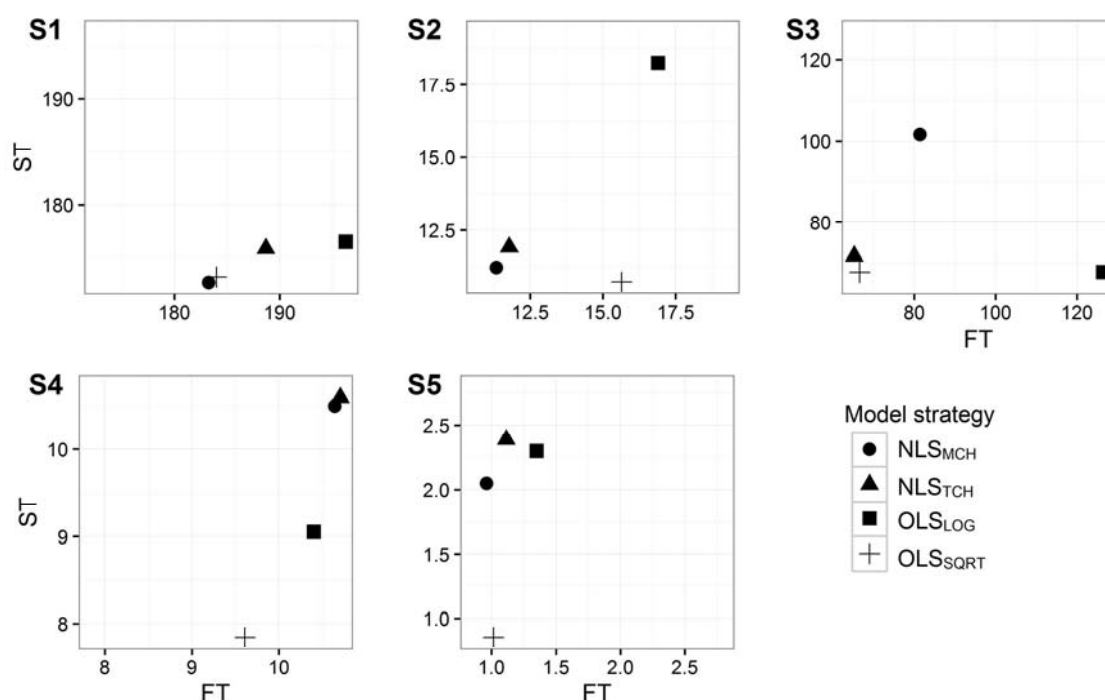


Figure 4. Scatter plot of the two components of the ME for S1, S2, S3, S4, and S5. FT and ST are the first and second terms of Equation (5), respectively.

3. Results

Two OLS models and two NLS models were fit separately for the five study sites. The empirical approach was assessed by the two OLS models formulated as one LOG and one SQRT model, referred to as OLS_{LOG} and OLS_{SQRT} respectively. Models for the five study sites, fit using all observations, were reported for reference (see Supplementary Table S1). The semi-empirical approach was assessed in terms of the two NLS models with MCH and TCH as predictor variables, referred to as NLS_{MCH} and NLS_{TCH} respectively. Comparison of the model performances showed that the empirical approach resulted in the smallest MSD in three of the five study sites (Figure 1). In S1, S3 and S5, the empirical approach with a SQRT model resulted in the smallest MSD. In S3, the semi-empirical (NLS_{TCH}) and empirical (OLS_{SQRT}) approaches were almost equal in terms of MSD, with the OLS_{SQRT} resulting in slightly smaller MSD. In terms of model-based estimated variance (Figure 2), the empirical approach resulted in the smallest variance estimates in four of the five study sites. In S1, the NLS_{MCH} resulted in the smallest variance estimate. Considering the ME, the empirical approach had smaller ME in three of five study sites (Figure 3). In the two tropical study sites (S1 and S2), the NLS_{MCH} model produced a smaller ME estimate compared to the other models. The scatterplot in Figure 4 provides additional information on the ME criterion. For sites S1, S2, S3, and S5, the first term of Equation (4) (i.e., an expression of MSD) was smallest for the NLS models. The second term of Equation (4) (i.e., estimated variance) however was larger, and the OLS models resulted in smaller ME for S3, S4, and S5.

When comparing the two different transformation strategies in the empirical approach, the OLS_{SQRT} models resulted in smaller MSD in all study sites (Figure 1). It also resulted in smaller ME in all study sites, compared to LOG models (Figure 3). For the two NLS modelling approaches, NLS_{MCH} resulted in smaller MSD and ME values in all study sites except S3 where the NLS_{TCH} resulted in the smallest values of MSD and ME.

4. Discussion

Both empirical and semi-empirical approaches to modelling are common practice. Even so, to our knowledge, only one study has been published that compares the two approaches for biomass estimation purposes using ALS data. The study by Hollaus et al. [5] found that the performances of an empirical and a semi-empirical approach were similar in terms of the coefficient of determination. Hollaus et al. [5], somewhat unexpectedly, found that the MSE and standard deviation favored the semi-empirical approach. When assessing only MSD in the present study, we made similar observations for two of the five study sites (S2 and S5). However, a model-based variance estimate can be used in addition to the MSD to select a model that not only has a small estimated prediction error (MSD), but also a high estimated stability in the model parameters (slope and intercept), i.e., that model parameters are less influenced by new observations as indicated by the model-based variance estimate. The ME proposed in the present study is a novel combination of the MSD and a model-based variance estimator. Scatter plots of the two components of the ME (Figure 4) can be used to visualize the two components and their respective magnitudes. This can aid in deciding which estimators have the desired properties in terms of small MSD, small model-based variance, or a combination of the two. Using the proposed ME criterion, the empirical approach was found to result in smaller errors in the boreal and temperate biomes. In the tropical biome, the semi-empirical approach resulted in smaller ME.

Comparisons of OLS_{LOG} and OLS_{SQRT} models show that the OLS_{SQRT} models resulted in smaller MSD and ME. In three of the study sites, the difference in MSD was quite small (<8%). In S3 and S5 however, the OLS_{SQRT} models resulted in MSD values that were 48% and 25% smaller than in the OLS_{LOG} models. This is in contrast to Hansen et al. [35] who reported approximately 7% smaller MSD using logarithmic transformation compared to SQRT. However, only the response was transformed in Hansen et al. [35]. Other comparative studies have however also found advantages of using SQRT models compared to LOG [12,70].

Comparisons of NLS_{MCH} and NLS_{TCH} using MCH and TCH respectively as predictor variables showed an advantage of using MCH in all study sites except for S3 where NLS_{TCH} resulted in 20% smaller MSD compared to NLS_{MCH} . Although Asner and Mascaro [61] argue that TCH is a more consistent predictor, less affected by sensor effects compared to MCH, our results suggest that using TCH instead of MCH comes at a cost of loss in prediction accuracy. A possible explanation could be that the MCH captures variations in the forest canopy structure better, compared to TCH. This is supported by the result for S3 where NLS_{TCH} resulted in smaller MSD and ME compared to NLS_{MCH} . S3 is characterized by low variation in age and AGB (Table 1).

The types of statistical methods used to model the relationship between forest attributes and ALS data have possibly affected the results in the present study. Several other methods for modelling are available, most notably GLM, nearest neighbors techniques, neural networks, and the increasingly popular Random Forest algorithm. To assess these alternative methods is a great undertaking that we considered to be outside the scope of the present study. Instead, we chose to focus on two commonly used modelling methods: OLS and NLS. Furthermore, several of the non-parametric methods mentioned do not have available model-based variance estimators, and would require a different approach to produce a criterion similar to the ME.

Even though the empirical approach resulted in smaller ME in boreal and temperate study sites, the semi-empirical approach could be a viable option based on advantageous properties such as re-use or re-calibration of models in new study sites [21]. Models that are relatively unaffected by noise, and that have a robust relationship between the response and predictor variable facilitate re-use or re-calibration. The model-based variance estimate, incorporated in the ME, could be calculated for new study sites, provided available ALS data. Thus, the proposed ME criterion could aid in the decision of modelling approach by providing a means of comparing the performance in terms of estimated variance to the estimated mean square error.

5. Conclusions

The results of the analysis showed that both approaches, empirical and semi-empirical, could be used to model the relationship between ALS data and AGB. The two approaches showed however differences in performance in terms of the two evaluation criteria. The semi-empirical approach resulted in the smallest MSD in three of five study sites, and in all three biomes represented in the present study. The ME criterion showed that the empirical approach resulted in smaller errors in the temperate and boreal biomes, while the semi-empirical approach resulted in smaller ME in the tropical biomes.

Additionally, the results showed that OLS_{SQRT} resulted in the smallest MSD and ME for all study sites taking the empirical approach. For the semi-empirical approach, the NLS_{MCH} resulted in the smallest MSD and ME in four of the five study sites.

Supplementary Materials: The following are available online at www.mdpi.com/1999-4907/8/5/170/s1, Table S1: OLS and SQRT models, fit using all observations.

Author Contributions: E.H. Hansen, E. Næsset, and L.T. Ene conceived and designed the experiments; E.H. Hansen performed the experiments; E.W. Mauya, Z. Patočka, T. Mikita, T. Gobakken and E. Næsset contributed materials; E.H. Hansen wrote the paper with contributions from all co-authors through the editorial process.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nelson, R. How did we get here? An early history of forestry lidar. *Can. J. Remote Sens.* **2013**, *39*, S6–S17. [CrossRef]
2. Vauhkonen, J.; Maltamo, M.; McRoberts, R.; Næsset, E. Introduction to forestry applications of airborne laser scanning. In *Forestry Applications of Airborne Laser Scanning*; Maltamo, M., Næsset, E., Vauhkonen, J., Eds.; Springer: Dordrecht, The Netherlands, 2014; Volume 27, pp. 1–16.
3. Næsset, E. Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sens. Environ.* **1997**, *61*, 246–253. [CrossRef]
4. Næsset, E. Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS J. Photogramm. Remote Sens.* **1997**, *52*, 49–56. [CrossRef]
5. Hollaus, M.; Wagner, W.; Schadauer, K.; Maier, B.; Gabler, K. Growing stock estimation for alpine forests in Austria: A robust lidar-based approach. *Can. J. For. Res.* **2009**, *39*, 1387–1400. [CrossRef]
6. Hollaus, M.; Mücke, W.; Roncat, A.; Pfeifer, N.; Briese, C. Full-waveform airborne laser scanning systems and their possibilities in forest applications. In *Forestry Applications of Airborne Laser Scanning*; Maltamo, M., Næsset, E., Vauhkonen, J., Eds.; Springer: Dordrecht, The Netherlands, 2014; pp. 43–61. [CrossRef]
7. Fassnacht, F.E.; Hartig, F.; Latifi, H.; Berger, C.; Hernández, J.; Corvalán, P.; Koch, B. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. *Remote Sens. Environ.* **2014**, *154*, 102–114. [CrossRef]
8. Næsset, E.; Bollandsås, O.M.; Gobakken, T. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. *Remote Sens. Environ.* **2005**, *94*, 541–553. [CrossRef]
9. Li, Y.Z.; Andersen, H.E.; McGaughey, R. A comparison of statistical methods for estimating forest biomass from light detection and ranging data. *West. J. Appl. For.* **2008**, *23*, 223–231.
10. Gleason, C.J.; Im, J. Forest biomass estimation from airborne lidar data using machine learning approaches. *Remote Sens. Environ.* **2012**, *125*, 80–91. [CrossRef]
11. Næsset, E. Estimating above-ground biomass in young forests with airborne laser scanning. *Int. J. Remote Sens.* **2011**, *32*, 473–501. [CrossRef]
12. Saarela, S.; Schnell, S.; Grafström, A.; Tuominen, S.; Nordkvist, K.; Hyyppä, J.; Kangas, A.; Ståhl, G. Effects of sample size and model form on the accuracy of model-based estimators of growing stock volume. *Can. J. For. Res.* **2015**, 1524–1534. [CrossRef]
13. Snowdon, P. A ratio estimator for bias correction in logarithmic regressions. *Can. J. For. Res.* **1991**, *21*, 720–724. [CrossRef]

14. Bradu, D.; Mundlak, Y. Estimation in lognormal linear models. *J. Am. Stat. Assoc.* **1970**, *65*, 198–211. [[CrossRef](#)]
15. Baskerville, G.L. Use of logarithmic regression in the estimation of plant biomass. *Can. J. For. Res.* **1972**, *2*, 49–53. [[CrossRef](#)]
16. Clifford, D.; Cressie, N.; England, J.R.; Roxburgh, S.H.; Paul, K.I. Correction factors for unbiased, efficient estimation and prediction of biomass from log–log allometric models. *For. Ecol. Manag.* **2013**, *310*, 375–381. [[CrossRef](#)]
17. Gregoire, T.G.; Lin, Q.F.; Boudreau, J.; Nelson, R. Regression estimation following the square-root transformation of the response. *For. Sci.* **2008**, *54*, 597–606.
18. Næsset, E.; Ørka, H.O.; Solberg, S.; Bollandsås, O.M.; Hansen, E.H.; Mauya, E.; Zahabu, E.; Malimbwi, R.; Chamuya, N.; Olsson, H.; et al. Mapping and estimating forest area and aboveground biomass in miombo woodlands in Tanzania using data from airborne laser scanning, tandem-x, rapideye, and global forest maps: A comparison of estimated precision. *Remote Sens. Environ.* **2016**, *175*, 282–300. [[CrossRef](#)]
19. McRoberts, R.E.; Næsset, E.; Gobakken, T. Accuracy and precision for remote sensing applications of nonlinear model-based inference. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 27–34. [[CrossRef](#)]
20. Ene, L.T.; Næsset, E.; Gobakken, T. Simulation-based assessment of sampling strategies for large-area biomass estimation using wall-to-wall and partial coverage airborne laser scanning surveys. *Remote Sens. Environ.* **2016**, *176*, 328–340. [[CrossRef](#)]
21. Magnussen, S.; Næsset, E.; Gobakken, T.; Frazer, G. A fine-scale model for area-based predictions of tree-size-related attributes derived from lidar canopy heights. *Scand. J. For. Res.* **2012**, *27*, 312–322. [[CrossRef](#)]
22. Bouvier, M.; Durrieu, S.; Fournier, R.A.; Renaud, J.-P. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne lidar data. *Remote Sens. Environ.* **2015**, *156*, 322–334. [[CrossRef](#)]
23. Asner, G.P.; Flint Hughes, R.; Varga, T.A.; Knapp, D.E.; Kennedy-Bowdoin, T. Environmental and biotic controls over aboveground biomass throughout a tropical rain forest. *Ecosystems* **2009**, *12*, 261–278. [[CrossRef](#)]
24. Chave, J.; Andalo, C.; Brown, S.; Cairns, M.A.; Chambers, J.Q.; Eamus, D.; Folster, H.; Fromard, F.; Higuchi, N.; Kira, T.; et al. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* **2005**, *145*, 87–99. [[CrossRef](#)] [[PubMed](#)]
25. Hulshof, C.M.; Swenson, N.G.; Weiser, M.D. Tree height–diameter allometry across the United States. *Ecol. Evol.* **2015**, *5*, 1193–1204. [[CrossRef](#)] [[PubMed](#)]
26. Asner, G.P.; Mascaro, J.; Muller-Landau, H.C.; Vieilledent, G.; Vaudry, R.; Rasamoelina, M.; Hall, J.S.; van Breugel, M. A universal airborne lidar approach for tropical forest carbon mapping. *Oecologia* **2012**, *168*, 1147–1160. [[CrossRef](#)] [[PubMed](#)]
27. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016.
28. Tompalski, P.; Coops, N.C.; White, J.C.; Wulder, M.A. Enriching ALS-derived area-based estimates of volume through tree-level downscaling. *Forests* **2015**, *6*, 2608–2630. [[CrossRef](#)]
29. Chen, Q. Modeling aboveground tree woody biomass using national-scale allometric methods and airborne lidar. *ISPRS J. Photogramm. Remote Sens.* **2015**, *106*, 95–106. [[CrossRef](#)]
30. Næsset, E. Area-based inventory in Norway—from innovation to an operational reality. In *Forestry Applications of Airborne Laser Scanning*; Maltamo, M., Næsset, E., Vauhkonen, J., Eds.; Springer: Dordrecht, The Netherlands, 2014; Volume 27, pp. 215–240. [[CrossRef](#)]
31. Hastie, T.; Tibshirani, R.; Friedman, J.; Franklin, J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd ed.; Springer Science Business Media: New York, NY, USA, 2009; p. 745. ISBN: 978-0-387-84857.
32. Magnussen, S. Arguments for a model-dependent inference? *Forestry* **2015**, *88*, 317–325. [[CrossRef](#)]
33. Kangas, A. Model-based inference. In *Forest Inventory: Methodology and Applications*; Kangas, A., Maltamo, M., Eds.; Springer: Dordrecht, The Netherlands, 2006; pp. 39–52. ISBN: 978-1-4020-4381-9.
34. McRoberts, R.E.; Næsset, E.; Gobakken, T. Inference for lidar-assisted estimation of forest growing stock volume. *Remote Sens. Environ.* **2013**, *128*, 268–275. [[CrossRef](#)]
35. Hansen, E.H.; Gobakken, T.; Bollandsås, O.; Zahabu, E.; Næsset, E. Modeling aboveground biomass in dense tropical submontane rainforest using airborne laser scanner data. *Remote Sens.* **2015**, *7*, 788–807. [[CrossRef](#)]
36. Anonymous. *Pinnacle User's Manual*; Javad Positioning Systems: San Jose, CA, USA, 1999.

37. Næsset, E. Effects of differential single- and dual-frequency gps and glonass observations on point accuracy under forest canopies. *Photogramm. Eng. Remote Sens.* **2001**, *67*, 1021–1026.
38. Masota, A.M.; Bollandas, O.M.; Zahabu, E.; Eid, T. Allometric biomass and volume models for lowland and humid montane forests. In *Allometric Tree Biomass and Volume Models in Tanzania*; Malimbwi, R.E., Eid, T., Chamshama, S.A.O., Eds.; Department of Forest Mensuration and Management, Faculty of Forestry and Nature Conservation, Sokoine University of Agriculture: Morogoro, Tanzania, 2016; pp. 35–46. ISBN: 978-9976-9930-1-1.
39. Mauya, E.; Ene, L.; Bollandas, O.; Gobakken, T.; Næsset, E.; Malimbwi, R.; Zahabu, E. Modelling aboveground forest biomass using airborne laser scanner data in the miombo woodlands of Tanzania. *Carbon Balance Manag.* **2015**, *10*, 28. [[CrossRef](#)] [[PubMed](#)]
40. Ministry of natural resources and tourism (MNRT). *National Forest Resources Monitoring and Assessment of Tanzania Mainland (NAFORMA). Main Results*; Ministry of Natural Resources and Tourism (MNRT): Dar Es Salaam, Tanzania, 2015; p. 106.
41. Tomppo, E.; Malimbwi, R.; Katila, M.; Mäkisara, K.; Henttonen, H.M.; Chamuya, N.; Zahabu, E.; Otieno, J. A sampling design for a large area forest inventory: Case Tanzania. *Can. J. For. Res.* **2014**, *44*, 931–948. [[CrossRef](#)]
42. Pinheiro, J.; Bates, D.; DebRoy, S.; Sarkar, D.; R Development Core Team. *Nlme: Linear and Nonlinear Mixed Effects Models*; R Package Version 3.1-125; R Foundation for Statistical Computing: Vienna, Austria, 2016.
43. Ministry of Natural Resources and Tourism (MNRT). *NAFORMA Field Manual—Biophysical*; Ministry of Natural Resources and Tourism (MNRT): Dar Es Salaam, Tanzania, 2011; p. 96.
44. Mugasha, W.A.; Eid, T.; Bollandas, O.M.; Malimbwi, R.E.; Chamshama, S.A.O.; Zahabu, E.; Katani, J.Z. Allometric models for prediction of above- and belowground biomass of trees in the miombo woodlands of Tanzania. *For. Ecol. Manag.* **2013**, *310*, 87–101. [[CrossRef](#)]
45. Patočka, Z.; Mikita, T. Využití plošného přístupu ke zpracování dat leteckého laserového skenování v inventarizaci lesa (Use of area-based approach to process the airborne laser scanning data in forest inventory). *Zprávy Lesn. Výzkumu* **2016**, *62*, 115–124. (In Czech)
46. Zianis, D.; Muukkonen, P.; Mäkipää, R.; Mencuccini, M. *Biomass and Stem Volume Equations for Tree Species in Europe*; Finnish Society of Forest Science, Finnish Forest Research Institute: Vantaa, Finland, 2005; Volume 4, p. 63. ISBN: 951-40-1983-0.
47. Muukkonen, P.; Mäkipää, R. Biomass equations for European trees: Addendum. *Silva Fenn.* **2006**, *40*, 763–773. [[CrossRef](#)]
48. Widlowski, J.L.; Verstraete, M.; Pinty, B.; Gobron, N. *Allometric Relationships of Selected European Tree Species*; Rep. EUR 20855 EN; European Commission Joint Research Centre: Ispra, Italy, 2003; p. 61.
49. Næsset, E.; Gobakken, T.; Solberg, S.; Gregoire, T.G.; Nelson, R.; Stahl, G.; Weydahl, D. Model-assisted regional forest biomass estimation using lidar and insar as auxiliary data: A case study from a boreal forest area. *Remote Sens. Environ.* **2011**, *115*, 3599–3614. [[CrossRef](#)]
50. Braastad, H. Volume tables for birch. *Medd. Nor. Skogforsøksves.* **1966**, *21*, 265–365.
51. Brantseg, A. Furu sønnafjells. Kubering av stående skog. Funksjoner og tabeller. *Medd. Nor. Skogforsøksves.* **1967**, *22*, 689–739.
52. Vestjordet, E. Functions and tables for volume of standing trees, Norway spruce. *Medd. Nor. Skogforsøksves.* **1967**, *22*, 543–574.
53. Fitje, A.; Vestjordet, E. Bestandshøydekurver og nye høydeklasser for gran. *Medd. Nor. Inst. Skogforsk.* **1977**, *34*, 23–68.
54. Tomter, S.M. *Beregning av Volum de Første år Etter Bestandsetablering (Volume Estimation during the First Years after Stand Establishment)*; Unpublished note; Norwegian Institute of Land Inventory: Ås, Norway, 1998. (In Norwegian)
55. Marklund, L.G. *Biomassafunktioner för Tall, Gran och björk i Sverige [Biomass Functions for Pine, Spruce and Birch in SWEDEN]*; Report 45; Department of Forest Survey, Swedish University of Agricultural Sciences: Umeå, Sweden, 1988; (In Swedish). ISBN: 9157635242.
56. Gobakken, T.; Næsset, E.; Nelson, R.; Bollandas, O.M.; Gregoire, T.G.; Ståhl, G.; Holm, S.; Ørka, H.O.; Astrup, R. Estimating biomass in Hedmark county, Norway using national forest inventory field plots and airborne laser scanning. *Remote Sens. Environ.* **2012**, *123*, 443–456. [[CrossRef](#)]
57. Terrasolid. *Terrascan User's Guide*; Terrasolid Oy: Jyväskylä, Finland, 2012; p. 311.

58. Axelsson, P. DEM generation from laser scanner data using adaptive TIN models. *Int. Arch. Photogramm. Remote Sens.* **2000**, *33*, 110–117.
59. Ene, L.T.; Næsset, E.; Gobakken, T.; Bollandsås, O.M.; Mauya, E.W.; Zahabu, E. Large-scale estimation of change in aboveground biomass in miombo woodlands using airborne laser scanning and national forest inventory data. *Remote Sens. Environ.* **2017**, *188*, 106–117. [[CrossRef](#)]
60. Næsset, E.; Gobakken, T. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sens. Environ.* **2008**, *112*, 3079–3090. [[CrossRef](#)]
61. Asner, G.P.; Mascaro, J. Mapping tropical forest carbon: Calibrating plot estimates to a simple lidar metric. *Remote Sens. Environ.* **2014**, *140*, 614–624. [[CrossRef](#)]
62. Lim, K.S.; Treitz, P.M. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. *Scand. J. For. Res.* **2004**, *19*, 558–570. [[CrossRef](#)]
63. Lumley, T.; Miller, A. Leaps: Regression Subset Selection. R Package Version 2.9. Available online: <http://cran.r-project.org/src/contrib/Archive/leaps> (accessed on 15 May 2017).
64. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002; p. 495. ISBN: 978-0-387-21706-2.
65. Long, J.S.; Ervin, L.H. Using heteroscedasticity consistent standard errors in the linear regression model. *Am. Stat.* **2000**, *54*, 217–224. [[CrossRef](#)]
66. Breusch, T.S.; Pagan, A.R. A simple test for heteroscedasticity and random coefficient variation. *Econometrica* **1979**, *47*, 1287–1294. [[CrossRef](#)]
67. MacKinnon, J.G.; White, H. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *J. Econom.* **1985**, *29*, 305–325. [[CrossRef](#)]
68. Zeileis, A. Econometric computing with HC and HAC covariance matrix estimators. *J. Stat. Softw.* **2004**, *11*. [[CrossRef](#)]
69. White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **1980**, *48*, 817–838. [[CrossRef](#)]
70. Ørka, H.O.; Gobakken, T.; Næsset, E. Predicting attributes of regeneration forests using airborne laser scanning. *Can. J. Remote Sens.* **2016**, *42*, 541–553. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).