# Modelling tree diameter from airborne laser scanning derived variables: a comparison of spatial statistical models

Christian Salas[1,a,c], Liviu Ene[b], Timothy G. Gregoire[a], Erik Næsset[b], and Terje Gobakken[b]

[a] School of Forestry and Environmental Studies, Yale University, New Haven, Connecticut, USA
[b] Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway
[c] Departamento de Ciencias Forestales, Universidad de La Frontera, Temuco, Chile

## Summary

In the past tree height has been more difficult to measure accurately in the field than tree diameter at breast height. As a consequence, models to predict height from diameter measurements have been widely developed in the forestry literature. Through the use of airborne laser scanning technology (e.g., LiDAR), tree variables such as height and crown diameter can be measured accurately, a development which has spawned the need for models to predict diameter from airborne laser-derived measurements. Although some work has been done for fitting such models, none have incorporated spatial information to improve the accuracy of the predicted diameters. Using a simple linear model for predicting tree diameter from laser-derived tree height and crown diameter measurements, we compared the performance of ordinary least squares (OLS), generalized least squares with a non-null correlation structure (GLS), linear mixed-effects model (LME), and geographically weighted regression (GWR). Our data were obtained from 36 sample plots established in Norway. This is the first study to examine the use of spatial statistical models for tree-level LiDAR data. Prediction errors in tree diameter with LME are of 3.5%, with GWR are 10%, and with OLS are of 17%. LME also exhibited low variability in predicting performance across all the validation classes. Giving the difficulties of using parametric statistical inference (such as maximum likelihood based indices) for GWR, we used permutation tests and bootstrapping as a way for detecting statistical differences. LME was significantly better than the other models, as well as GWR was to OLS and GLS. Our results indicate that the LME model produced the best predictions of tree diameter from LiDAR-based variables to a degree that has not been possible earlier.

## Introduction

Airborne laser scanning is a powerful remote sensing system for collecting topographic data having applications in a variety of disciplines. Airborne laser scanning systems (ALS) have lead a revolution in remote sensing technology during the last 10–14 years (Popescu y Wynne 2004). As pointed out by Reutebuch *et al*. (2005), the basis of this revolution is the ability to measure directly the three-dimensional structure (i.e., terrain, vegetation, and infrastructure) of imaged areas and to separate biospatial data (measurements of aboveground vegetation) from geospatial data (measurements of the terrain surface) using active remote sensing technologies. In ALS, a scanner which distributes the transmitted pulses across the flight direction of the platform is attached to the laser, and such systems can measure the 3D position of points on the ground and in vegetation canopies with an accuracy of a few decimetres (Næsset *et al*. 2004). Among the laser scanner systems currently available, LiDAR (light detection and ranging) sensors offer impressive performance that challenge physical barriers in the optical and electronic domain by offering a high density of points at scanning frequencies of 50,000 pulses/second, multiple echoes per laser pulse, intensity measurements for the returning signal, and centimeter accuracy for horizontal and vertical positioning (Popescu y Wynne 2004). These features, make LiDAR useful for directly assessing vegetation characteristics, and overall providing new tools for measuring and monitoring biospatial data across the landscape (Popescu y Wynne 2004, Reutebuch *et al*. 2005). Further reviews on ALS and LiDAR studies in general and in forestry, can be found in Næsset *et al*. (2004) and references therein.

[1]Corresponding author. School of Forestry and Environmental Studies, Yale University, 360 Prospect Street, New Haven, CT 06511-2104, USA. Tel. +1(203)432-9398. Fax. +1(203)432-3809. E-mail: christian.salas@yale.edu

Height-diameter models are key in conventional forest inventories. In conventional forest sampling, i.e., ground-based measurements of tree variables within sample plots, diameter at breast height (*dbh*) is measured for all the trees within plots, and height (*h*) is measured only for a sample, or also referred as subsample, of trees within plots, because *h* is more difficult and expensive to measure than *dbh*. In this setting, models that predict *h* as a function of *dbh* (hereafter referred to as height-diameter models, or *h-dbh* models[2]) are fitted in order to later predict *h* of those trees based on tree *dbh* alone. Several studies have been conducted on *h-dbh* models, e.g. Thorey (1932), Meyer (1940), Curtis (1967), Yuancai y Parresol (2001), including model comparisons (e.g., Zhang 1997, Peng *et al*. 2001), up to models incorporating stand-level variables (e.g., Staudhammer y LeMay 2000, Sánchez *et al*. 2003) and random-effects (e.g., Lappi y Bailey 1988, Robinson y Wykoff 2004, Lynch *et al*. 2005). Several model forms have been proposed and fitted, van Laar y Akça (p. 122, 2007) list a few.

Modelling efforts using data from laser scanning are mainly at the plot level. Forest canopy height model (or CHM) is a digital elevation map of the top-of-forest canopy (Nelson *et al*. 1998). The CHM is a plot-level (or stand-level) height model. A CHM is relatively easy to obtain through airborne laser scanning techniques, e.g., LiDAR, and methods for calibrating it have been advised. Several efforts for estimating plot-level tree heights derived from LiDAR have been conducted (e.g., Popescu *et al*. 2002, Popescu y Wynne 2004). As pointed out by Næsset *et al*. (2004), who reviewed the Nordic experience in laser scanning, the main objective especially in Norway has been to develop methods that are directly suited for practical forest inventory at the stand level. In this setting, regression models are fitted using laser-derived variables and field plot variables, in order to predict stand variables (plot-wise). The stand variables of chief interest in this regard, according to Næsset *et al*. (2004) have been: mean tree height, basal area, and stand volume. Other variables have been also predicted such as biomass and carbon (e.g., Nelson *et al*. 2004), diameter and basal area distribution (e.g., Næsset y Gobakken 2004), and number of trees (e.g., Hudak *et al*. 2005, 2008).

In a laser-scanning-based forest inventory framework there is a shift from *h-dbh* to *dbh-h* models. When trees are measured on the ground, their positions can be linked with laser measurements (Persson *et al*. 2004), through a process known as segmentation, and the trees identified from the segmentation, are known as linked trees. Individual heights of the linked trees can be roughly estimated from laser scanning data by using the approximated ground level and the highest hit of the treetop (Rönnholm *et al*. 2004). Nevertheless, tree diameter is not available from LiDAR data, but it is a crucial variable in order to analyze the forest structure. Analogous to the need in ground-based forest inventory to develop *h-dbh* models to predict *h* of non-measured trees, in a laser-based forest inventory we need *dbh-h* models to predict *dbh*. More generally LiDAR-derived variables, not just *h*, can be used as predictor variables.

Spatial correlation especially affects statistical inference of fitted models. When data are spatially collected, it is likely that points close together have more similar values (or different) of variables than points that are farther apart (Schabenberger y Gotway 2005, Ives y Zhu 2006). Due to this, residuals of models being fit using OLS would likely be autocorrelated. As pointed out by Kissling y Carl (2008), the presence of spatial autocorrelation is problematic for classical statistical test because these methods assume independently distributed errors. The effects of ignoring spatial correlation are summarized by Hoeting (2009), as follows. Effective sample size decreases as the correlation between observations increases (Schabenberger y Gotway 2005). Ignoring spatial correlation when selecting covariates for inclusion in models can lead to the exclusion of relevant covariates in the model (Hoeting *et al*. 2006). Models which incorporate spatial correlation allow for direct incorporation of the sampling design, using model-based inference, in the modelling by accounting for the hierarchical dependences between members of a population (Cressie *et al*. 2009). Further references on applications mixing hierarchical modelling and sampling design can be found in Hoeting (2009).

Models that explicitly use spatial information are not common in forest inventories. When a traditional regression model is fitted with ordinary least squares (OLS), beyond the implicit spatial arrangement of the data, spatial correlation plays no role in the modelling process (Brunsdon *et al*. 1998). The advantages of taking spatial information into account using ground-based measurements for modelling *h-dbh* (Zhang y Shi 2004), tree basal area growth-*dbh* (Zhang *et al*. 2004), tree crown area-*dbh* (Zhang y Gove 2005, Zhang *et al*. 2005) have been examined. In a laser scanning context, the works on modelling *dbh* from LiDAR variables have not taken spatial information into consideration. For example, Hyyppä *et al*. (2001) built a model, using OLS that predict tree diameter *dbh* as a function of

[2] Height-diameter models are sometimes referred to as diameter-height models (e.g., Schröder y Álvarez 2001, Zhang y Shi 2004). We prefer to be consistent with mathematics, and reserve the left word in the couplet to identify the response variable, as in Henry y Aarssen (1999).

height (*lh*) and tree crown diameter derived from LiDAR (*lcd*). They only compared the forest inventory estimates of the mean height obtained using that function versus the same parameter estimates using ground-data, however, they did not report any statistics or comparison of the fitted model. The same model was later fit by Schardt *et al*. (2004) and García *et al*. (2007) with good results, but again not offering greater details on the performance of the model. Both Persson *et al*. (2004) and Heurich *et al*. (2004) proposed slightly different models using the same predictor variables. Persson *et al*. (2004) did report the fit statistics of the model, and obtained errors of around 10% in predicting *dbh*. In recent studies, Heurich y Thoma (2008) and Heurich (2008), both height and density-related laser variables (e.g., height percentiles and total penetration rate. Næsset 2002), have been incorporated in models for predicting *dbh*.

Testing the specification of spatial regression models, has not been a straightforward endeavor. Brunsdon *et al*. (1999b) and Leung *et al*. (2000) compared the use of OLS and geographically weighted regression (GWR) using a parametric test, a $F$-test based on the residual sum of squares of each model, and maximum likelihood-based tests have been later proposed (e.g., Fotheringham *et al*. 2002) too. However, as we will shall later argue, those tests are not valid because they are suitable for parametric models, but not for non-parametric ones such as GWR. Zhang y Shi (2004) and Zhang *et al*. (2004) compared the use of OLS and GWR using the $F$-test, as described in Leung *et al*. (2000). Later, comparisons of GWR were extended not only to OLS but also to other models (e.g., Zhang y Gove 2005, Zhang *et al*. 2005, 2008) such as linear mixed-effects models (LME). Nevertheless, in the straight statistical sense of the model, their LME models are not actually mixed-effects models, because no random-effects coefficients were included in their models. The statistical model that they were fitting was a generalized linear model with a non-null correlation structure. Brunsdon *et al*. (1999a) compared GWR and a model with random coefficients (in GWR, parameter coefficients are not assumed to be random, Brunsdon *et al*. 1999a), which they fitted in a Bayesian framework. A Bayesian approach proposed by Gelfand *et al*. (2003), termed spatially varying coefficient process (SVCP), has been applied by Waller *et al*. (2007), Wheeler y Waller (2009), and Finley *et al*. (Submitted) with success. This approach involves, in general terms, fitting a LME using Markov chain Monte Carlo (MCMC) algorithms assigning prior distributions to fixed and random effects (Waller *et al*. 2007). We prefer the frequentist statistics framework. Further discussion about frequentist and Bayesian statistics can be found elsewhere (e.g., Efron 1986, Dennis 1996, Ellison 2004).

Reasoning that building statistical models that (may) use spatial information offer both a better prediction performance and a sounder statistical framework than using statistical models that do not use spatial information, we compare the performance of alternative models fitted using ordinary least squares (OLS), generalized least squares with spatial correlation structure (GLS), linear mixed-effects model (LME), and geographically weighted regression (GWR), for modelling the spatial variation in tree diameter at breast height as a function of laser-derived height relationship in forest stands.

# Methods

## Data

**Field data** The study area is located in the municipality of Aurskog, situated in the south-eastern Norway ($59°80'$ N, $11°55'$ E, 172-388 m a.s.l.). The sample plots were located in heterogeneous forest conditions regarding the tree species composition, site qualities, and development classes. The main three species were Norway spruce [*Picea abies* (L.) Karst.], Scots pine (*Pinus silvestris* L.), and birch (*Betula* sp.). At the plot level, the species composition by volume ranged from 0-98% for spruce, 0-100% for pine, and 0-34% for birch. The terrain across the study area is gentle compared to average terrain conditions for productive forests in Norway. Nonetheless the local topography varied significantly among plots. The plot center coordinates (i.e., the $x$ and $y$ position) were determined by differential dual-frequency Global Positioning System (GPS) and Global Navigation Satellite System (GLONASS) measurements, using two Topcon dual-frequency receivers. On each plot, all trees with *dbh* $\geq$ 5 cm were callipered, and the stem locations were mapped using a Sokkia SET5F total station. The tree species and vegetation status were recorded for each measured tree. Total tree height ($h$) was measured using a Suunto hypsometer, on a total of 618 reference trees selected from across all plots.

**LiDAR data** The LiDAR data was obtained under leaf-on canopy conditions on June, 2006. Five non-overlapping strips were flown from an average flying altitude of 800 m, using a PA31 Piper Navajo (LN-LAS) airplane. The strips were E-W oriented, with an N-S spacing of ca. 8.7 km. Data were acquired with an Optech ALTM 3100 system, operating with a pulse repetition frequency of 100 kHz, a scanning frequency of 70 kHz, and using a half-angle of $5^{\circ}$. Laser data were processed by the contractor Blom Geomatics A.S. The GPS/INS processing was performed using Applanix POSPac software, and then planimetric coordinates ($x$ and $y$) and ellipsoidal heights of the first and last return echoes were obtained from the navigation and laser data using Optech REALM tools. The 3D adjustment and classification was performed using TerraScan and TerraModel software from TerraSolid Ltd. The TIN model representing the terrain surface was subtracted from the height values of all the echoes recorded, to obtain the heights above the surface of the echoes, with an accuracy of $\pm$ 10 cm.

**Individual tree detection** Only the first returns located inside buffers extending 10 m around the plot borders were retained for canopy height models (CHM) creation. The average density of the first echoes within the buffers after data processing was 7.3 pulses/m$^2$. For each plot, a CHM was obtained by griding the irregularly spaced laser returns into a 0.35 m regular grid. For single tree delineation, we followed the procedure described by Hyyppä *et al*. (2001). The CHM was smoothed using the low-pass Gaussian filter described by Hyyppä *et al*. (2001). Regional maxima were identified in the smoothed CHMs by searching in 8-connected neighborhoods for finding the pixels having higher values than their external boundary neighbors. A marker-based watershed algorithm was employed for individual tree crown delineation. Individual tree variables (tree height, crown width and stem location) were obtained from the point clouds located inside the delineated segments. The tree height and stem location were obtained directly from the echo with the highest $z$-value within each segment, which was considered to represent the top of the delineated tree. The polygons representing the crown projections of the delineated trees on the ground were approximated using the convex hull of the ($x$, $y$)-coordinates of the returns located within each segment. Thus, the crown widths were estimated as the diameter of a circle with the same area as the crown polygons. In total, we have 1,438 trees with both field-based and lidar-derived variables (Table 1). The minimum number of sample trees per plot was 22, which is a reasonable number of trees for fitting *dbh*-LiDAR derived variables models on a per plot basis.

Table 1: Descriptive statistics for 1,438 trees. *dbh* is diameter at breast height, *h* is total height, *lh* and *lcd* are lid derived height and crown diameter, respectively.

| Statistic | *dbh* (cm) | *h* (m) | *lh* (m) | *lcd* (m) |
|---|---|---|---|---|
| Minimum | 5.0 | 5.0 | 3.4 | 0.3 |
| Maximum | 49.5 | 35.8 | 29.6 | 4.2 |
| Mean | 19.2 | 15.2 | 15.1 | 1.5 |
| Std.Dev | 7.8 | 4.7 | 4.8 | 0.7 |

## Statistical models

Slight variations have been proposed for *dbh*-LiDAR derived variables models. Hyyppä *et al*. (2001), Schardt *et al*. (2004), and García *et al*. (2007) used the following mean function

$$\mathrm{E}[dbh] = \beta_0 + \beta_1 lh + \beta_2 lcd, \tag{1}$$

where $\beta_0, \beta_1$, and $\beta_2$ are coefficients to be estimated, *lh* is the laser-derived height, and *lcd* is the laser-derived tree crown diameter. In the sequel, we shall use (1) as the mean function in various statistical models. Our aim is to compare a variety of ways to incorporate spatial information in statistical models for *dbh*. Rather than pool data from all plots together, we fit the various models to data from each plot separately, thereby avoiding having to account for variations in site quality, aspect, slope, and other topographical features among plots. We largely adopt the statistical notation of Schabenberger y Gotway (2005) and Pinheiro y Bates (2000), where a response and predictor variables $y$

and $x$s have been recorded at spatial locations $\mathbf{s}_i, ..., \mathbf{s}_n$, i.e., specific geographic locations for each $i$th observation with spatial coordinates $\mathbf{s}_i = [x_i, y_i]'$, represented by $\mathbf{y}(\mathbf{s})$ and $\mathbf{X}(\mathbf{s})$, respectively. The additive error term is represented by $\mathbf{e}(\mathbf{s})$.

**Model 1** The first statistical model is as follows

$$\begin{aligned} dbh(\mathbf{s}_i) &= \beta_0 + \beta_1 lh(\mathbf{s}_i) + \beta_2 lcd(\mathbf{s}_i) + e(\mathbf{s}_i), \\ e(\mathbf{s}_i) &\sim \mathcal{N}\left(0, \sigma^2\right), \\ \mathbf{\Sigma}\left[e(\mathbf{s}_i), e(\mathbf{s}_j)\right] &= \sigma^2 \mathbf{I_n}, \end{aligned} \tag{2}$$

where *dbh*$(\mathbf{s}_i)$, *lh*$(\mathbf{s}_i)$, and *lcd*$(\mathbf{s}_i)$ are *dbh*, *lh*, and *lcd*, recorded at spatial location $\mathbf{s}_i$, respectively. Residual of the model at location $\mathbf{s}_i$ are represented by $\mathbf{e}(\mathbf{s}_i)$, and they are assumed Normally distributed with mean 0 and variance $\sigma^2$. $\beta_0$, $\beta_1$, and $\beta_2$ are coefficient parameters of the model. $\mathbf{\Sigma}\left[e(\mathbf{s}_i), e(\mathbf{s}_j)\right]$ is the covariance matrix of the residuals, and $\mathbf{I_n}$ is an identity matrix of size $n$ (the sample size).

Eq. (2) can be more generally represented in matrix form as follows,

$$\mathbf{y}(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + \mathbf{e}(\mathbf{s}) \tag{3}$$

where $\mathbf{y}(\mathbf{s})$ is the vector with the response variable recorded at location $\mathbf{s}$, $\mathbf{X}(\mathbf{s})$ is the matrix of predictor variables spatially recorded, $\boldsymbol{\beta}$ is a vector with the coefficient parameters of the model, and $\mathbf{e}(\mathbf{s})$ is the vector of random errors.

With ordinary least squares (OLS), the coefficient vector is estimated by

$$\widehat{\boldsymbol{\beta}}_{OLS} = \left\{\mathbf{X}(\mathbf{s})'\mathbf{X}(\mathbf{s})\right\}^{-1}\mathbf{X}(\mathbf{s})'\mathbf{y}(\mathbf{s}). \tag{4}$$

Model 1 does not incorporate any kind of spatial information. As pointed out by Schabenberger y Gotway (2005, p. 316), model 1 is a spatial regression model since the response variable $\mathbf{y}(\mathbf{s})$, and the predictor variables comprising $\mathbf{X}(\mathbf{s})$, are recorded at spatial locations. However, the spatial information serves only to link $\mathbf{y}(\mathbf{s})$ and $\mathbf{X}(\mathbf{s})$, yet, there is nothing in the model that explicitly considers spatial pattern or spatial relationships.

**Model 2** The second model differs from (2) only in its specification of the error covariance. For this model

$$\mathbf{\Sigma}\left[e(\mathbf{s}_i), e(\mathbf{s}_j)\right] = \sigma^2 \mathbf{\Omega}, \tag{5}$$

where $\mathbf{\Omega}$ is the correlation matrix, the elements of which specify the spatial correlation between all pairs of trees on a plot. If the correlation parameters in $\mathbf{\Omega}$ were known, then the generalized least squares (GLS) estimator

$$\widehat{\boldsymbol{\beta}}_{GLS} = \left\{\mathbf{X}(\mathbf{s})'\mathbf{\Omega}^{-1}\mathbf{X}(\mathbf{s})\right\}^{-1}\mathbf{X}(\mathbf{s})'\mathbf{\Omega}^{-1}\mathbf{y}(\mathbf{s}) \tag{6}$$

would be feasible. Usually the parameters in $\mathbf{\Omega}$ must be estimated from data or auxiliary information. Using $\widehat{\mathbf{\Omega}}$ as the estimated correlation matrix (see bellow), the estimated GLS estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}_{EGLS} = \left\{\mathbf{X}(\mathbf{s})'\widehat{\mathbf{\Omega}}^{-1}\mathbf{X}(\mathbf{s})\right\}^{-1}\mathbf{X}(\mathbf{s})'\widehat{\mathbf{\Omega}}^{-1}\mathbf{y}(\mathbf{s}). \tag{7}$$

As pointed out by Cressie (1993) and exemplified by Pinheiro y Bates (2000), spatial correlation structures are generally represented by their semivariogram, instead of their correlation function. We derived $\widehat{\mathbf{\Omega}}$ for each plot after

deciding the following: (a) the functional form of the semivariogram, (b) whether the semivariogram would include a nugget effect, and (c) the distance metric to be used. Regarding (a), we assessed the following semivariogram functions; exponential, Gaussian, linear, rational quadratic, and spherical. All of them are fully described in Pinheiro y Bates (2000, p. 230–233), as well as further details on them can be found in spatial statistics books (e.g., Waller y Gotway 2004, Schabenberger y Gotway 2005, Diggle *et al*. 2007). Regarding (b), we assessed the following distance metrics between the $x$ and $y$ coordinates among trees: Euclidean, maximum, and Manhattan. Finally, regarding (c), we assessed the effect of adding a nugget to the semivariogram function. We fit model 2 and compared all the possible combinations of semivariogram function-distance metrics-nugget effect, and selected the best model using the Akaike Information Criterion (AIC, Akaike 1973). A referee suggested that a Matern type of correlation function would better capture a nugget effect, however, we did not use this function in our study, but did add nugget effect.

**Model 3** The third model adds random effects to the fixed effect coefficient parameters:

$$\begin{aligned} dbh(\mathbf{s}_i) &= (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})lh(\mathbf{s}_i) + (\beta_2 + b_{2i})lcd(\mathbf{s}_i) + e(\mathbf{s}_i), \\ \mathbf{b_i} &= (b_{0i}, b_{1i}, b_{2i})' \sim \mathcal{N}\left(0, \mathbf{D}\left(\boldsymbol{\phi}\right)\right), \\ e(\mathbf{s}_i) &\sim \mathcal{N}\left(0, \sigma^2\right), \\ \boldsymbol{\Sigma}\left[e(\mathbf{s}_i), e(\mathbf{s}_j)\right] &= \sigma^2\boldsymbol{\Omega}, \end{aligned} \tag{8}$$

where $\mathbf{D}\left(\boldsymbol{\phi}\right)$ is a $3 \times 3$ random-effects covariance matrix parameterized by $\boldsymbol{\phi}$. The inclusion of a vector of random effects permits the model to be individualized to each tree, while also permitting a pooling of all the data when fitting the model. Importantly, the random individual effects induces an intra-individual correlation structure that accounts for the lack of independence among trees on the same plot. It is this feature that has great appeal for modelers of correlated data, even though the induced correlation function may not be easily discernible (Schabenberger y Gregoire 1996). Further details on mixed-effects models can be found in Pinheiro y Bates (2000) and Schabenberger y Pierce (2002), as well as applications in a spatial context in Schabenberger y Gotway (2005) and in Bivand *et al*. (2008).

In matrix notation, (8) may be written as

$$\mathbf{y}(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + \mathbf{Z}(\mathbf{s})\mathbf{b} + \mathbf{e}(\mathbf{s}). \tag{9}$$

The flexibility permitted by a mixed model of this sort comes at the price of added complexity, both statistical and computational. In practice it is often found that data do not support a random effect for each fixed effect parameter, and also that when the correlation structure is modelled by $\mathbf{D}\left(\boldsymbol{\phi}\right)$, it is often sufficient to specify $\boldsymbol{\Omega} = \mathbf{I}_n$.

Customarily $\boldsymbol{\beta}$ in (9) is estimated by maximum likelihood or restricted maximum likelihood. followed by the best linear unbiased predictors ("BLUP", Robinson 1991) of the random effects. We used the `nlme` package (Pinheiro *et al*. 2008) implemented in R for fitting LME models.

Having fitted a LME model, the mean response at $\boldsymbol{x}$ can be estimated by $\boldsymbol{x}\widehat{\boldsymbol{\beta}}$, where $\boldsymbol{x}$ is the vector of covariate values. An individual observation can be predicted by adding the corresponding BLUP of the random effect. Prediction of a new observation outside the sample data is impossible owing to the impossibility of computing the BLUP of the random effect for that observation. Nevertheless, given a rather new (i.e., not used before) small sample of response-predictor variables pair of data (e.g., 2-4) and a fitted LME, it is possible to predict the random-effects as well, that is calibrate the mean response (see examples in Lappi y Bailey 1988, Lappi 1991, Lynch *et al*. 2005). A new observation can be think in our context as new trees that appear after running a different segmentation algorithm or trees that reach the $dbh$ minimum in a second measurement. In any case they were not part of the sample used for fitting the model.

In model 3 we are not directly adding spatial information. The random effects in Eq. 8, show that we are estimating an overall (i.e., fixed) mean intercept, and modelling local deviations from this mean. These deviations are driven by

a distribution, in our case $b_{0i} \sim \mathcal{N}\left(0, \sigma^2_{b_0}\right)$, that determines how much variation would be assigned to each tree from the mean intercept.

**Model 4** The fourth model allows the coefficients to vary by spatial location

$$\begin{aligned} \mathit{dbh}(\mathbf{s}_i) &= \beta_0(\mathbf{s}_i) + \beta_1(\mathbf{s}_i)\mathit{lh}(\mathbf{s}_i) + \beta_2(\mathbf{s}_i)\mathit{lcd}(\mathbf{s}_i) + e(\mathbf{s}_i) \\ e(\mathbf{s}_i) &\sim \mathcal{N}\left(0, \sigma^2\right), \\ \mathbf{\Sigma}\left[e(\mathbf{s}_i), e(\mathbf{s}_j)\right] &= \sigma^2 \mathbf{I_n}, \end{aligned} \tag{10}$$

where $\beta_{0(\mathbf{s}_i)}$, $\beta_{1(\mathbf{s}_i)}$, and $\beta_{2(\mathbf{s}_i)}$ represent the coefficient parameters for the $i$th observation, the other notation was previously defined. Model 4 can be fitted using geographically weighted regression (GWR), a non-parametric regression method that has been portrayed by Fotheringham *et al*. (2002) and previous work of the same authors (e.g., Brunsdon *et al*. 1996, 1998, Fotheringham *et al*. 1998).

In GWR a vector of coefficient parameters is required for each observation, that is to say, each observation has different coefficient parameters, therefore, having a local feature. Sometimes, only a portion of the coefficients of a GWR model can be local and the other ones can remain global, which has been called "mixed-GWR" (Fotheringham *et al*. 2002, p. 65–73). The estimated parameter vector for each observation in GWR (Brunsdon *et al*. 1998) is obtained using weighted least squares regression, as follows

$$\widehat{\boldsymbol{\beta}}_{iGWR} = \left\{\mathbf{X}(\mathbf{s})'\mathbf{W}(\mathbf{s}_i)^{-1}\mathbf{X}(\mathbf{s})\right\}^{-1}\mathbf{X}(\mathbf{s})'\mathbf{W}(\mathbf{s}_i)^{-1}\mathbf{y}(\mathbf{s}), \tag{11}$$

where $\mathbf{W}(\mathbf{s}_i)$ is the diagonal weight matrix for the $i$th observation (i.e., tree). This matrix in general is defined as $\mathbf{W}(\mathbf{s}_i) = \text{diag}\{W(s_i, s_j)\}$, where the scalar weight $W(s_i, s_j)$ for the $j$th observation depends on its proximity to the $i$th observation. A simple scheme of weighting would be to exclude from the model fitting observations beyond a distance $b$ from the $i$th observation, which is termed the moving window approach (Fotheringham *et al*. 2002). With this approach the spatial weighting has the problem of discontinuity in the parameter estimates (Brunsdon *et al*. 1998, Schabenberger y Gotway 2005). An alternative approach is to use a continuous spatial weighting function, as in general represented by

$$W(\mathbf{s}_i, \mathbf{s}_j) = f(\text{dist}_{(\mathbf{s}_i, \mathbf{s}_j)}, b), \tag{12}$$

where $f$ is a kernel function with bandwidth $b$. With (12) both the kernel and the bandwidth need to be stipulated. We assessed a Gaussian and a near-Gaussian kernels as being spatial weighting functions and selected the one that achieve a lower AIC[3] of the GWR model. The bandwidth is the searching radius around each observation and is key in GWR. The bandwidth can greatly affect the properties of the $\widehat{\boldsymbol{\beta}}_{iGWR}$ (Brunsdon *et al*. 1998). It governs the extent to which the resulting local calibration results are smoothed (Fotheringham *et al*. 2002). We used a fixed bandwidth of 6 m for all the plots, determined based on the range of the empirical semivariogram (i.e., distance in which the semivariogram reach the sill) of *dbh* per plot, and is consistent with the studies by Zhang *et al*. (2004) and Pukkala (1989). We used the `spgwr` package (Bivand y Yu 2009) implemented in R for fitting GWR models.

## Comparing models

We first computed the studentized model residuals and plotted them in a spatial fashion. Zhang y Gove (2005) called these type of residuals "local Z-values". These values would allow us for comparing the magnitudes of the residuals in space. Notice that we used the studentized residuals definition given by Schabenberger y Pierce (2002, Eq. 4.23),

[3] We used the AIC of GWR as defined in Fotheringham *et al*. (2002, eq. 2.33, p. 61).

and in order to compute this type of residual for GWR, we used the $\mathbf{L}$ matrix of Leung *et al*. (2000, Eq. 16) as the hat matrix.

**Goodness-of-fit** For each model (OLS, GLS, LME, and GWR), we computed the following statistics in order to assess the goodness-of-fit of the tested models: (a) The aggregated difference (AD) or mean residual, computed as follows

$$AD = \frac{1}{n}\sum_{i}^{n} r_i, \tag{13}$$

where $r_i = dbh(\mathbf{s}_i) - \widehat{dbh}(\mathbf{s}_i)$, $dbh(\mathbf{s}_i)$ is the observed tree diameter at the $i$th location, $\widehat{dbh}(\mathbf{s}_i)$ is the predicted tree diameter at the $\mathbf{s}_i$ location, and $n$ the sample size in the field plot being analyzed.

(b) the root mean square differences (*RMSD*) as follows,

$$RMSD = \sqrt{\frac{1}{n}\sum_{i}^{n} r_i^2}, \tag{14}$$

and (c) the aggregated of the absolute value of the differences (*AAD*), as follows

$$AAD_j = \frac{1}{n}\sum_{i}^{n} |r_i| \,. \tag{15}$$

We fitted and validated the models using the same observations (on a per plot basis). We deliberately avoided the use of validation or cross-validation, because withholding observations alters the pattern of spatial correlation of the observations. We preferred to use the term *RMSD* rather than *RMSE* (root mean squared error) following Stage y Crookston (2007) and Hudak *et al*. (2008).

At this stage, we assess the models exclusively according to their goodness-of-fit. We follow the methodology for selecting models proposed by Salas (2002). The methodology is as follows. First, partition the data into validation classes. We use the percentiles of the variable *lh* in order to define five validation classes, with similar number of observations in each class. Second, compute the statistics detailed above (*AD*, *RMSD*, *AAD*) within validation classes. Third, compute the mean and standard deviation of among-class values of these statistics, therefore having an among-class mean and an among-class standard deviation for *AD*, and so on). Fourth, assign ranking to the among-class mean and among-class standard deviation of each statistic. Ranking score 1 is assigned to the model with lowest among-class mean (or closer to 0 for *AD*) and lowest among-class standard deviation (which indicates more stable results across validation classes). Fifth, add all these ranking scores (six in total, two scores by three statistics) and sort the models according to this sum (no weighting of statistics was considered), the model with the lowest sum would be rank as the best. The methodology aim to rank best not only the models with overall better precision and accuracy but also with a consistent behavior across validation classes.

**Statistical inference** Statistical inference for GWR offers several challenges. Once we have built and maximized the log-likelihood function of a parametric model, it is straightforward to compare models using indices such as; Akaike Information Criterion (AIC, Akaike 1973) or the correction of AIC of Hurvich *et al*. (1998), among others. However, this is possible only for parametric models (in our study; OLS, GLS, and LME). On the other hand, and although we could think in building a likelihood function for GWR (e.g., Eq. 4.2 in Fotheringham *et al*. 2002), the actual way in which the parameters are estimated does not involve any optimal solution for this likelihood function, because all the coefficients are fitted independently from one observation to another. That is to say, the log-likelihood in Fotheringham *et al*. (2002) for GWR permits an infinity of maxima. Non-parametric statistical tests provide a more suitable framework for assessment of GWR. Permutation tests, also called randomization (Manly 2006) or re-randomization

(Welch 1987) tests, offer a non-parametric simple and straightforward technique for conducting hypothesis testing among statistics of two groups. In order to conduct the permutation tests, we first defined the statistics that would allow us to compare the models, which we will denote as $\widehat{\theta}$, as follows

$$\widehat{\theta} = \frac{1}{m}\sum_{j=1}^{m}\left(RMSD_{aj} - RMSD_{bj}\right) \tag{16}$$

where $m$ is the total number of sample plots, $RMSD_{aj}$ and $RMSD_{bj}$ are the root mean square differences (14) in % of models $a$ and $b$, respectively, for the $j$-th plot. We set model $a$ to be the base one and model $b$ the new model, or the one with lower ***RMSD***. Our hypothesis under test is $H_0 : \mu_{RMSD_a} - \mu_{RMSD_b} = 0$, vs. $H_a : \mu_{RMSD_a} > \mu_{RMSD_b}$. The following step was to randomly assign (without replacement) $m$ values of ***RMSD***s to model a and the the rest to model b, and based on this new sample (called "permutation sample" by Hesterberg *et al*. 2003), we computed $\widehat{\theta}$ (Eq. 16) for the permutation sample. The next step is to repeat this resampling many times, we used 10,000 permutations. The distribution of the statistic from these resamples forms the sampling distribution under the condition that $H_0$ is true. Finally, we computed the proportion of the distribution which exceeded the difference observed between the two models (i.e., $\widehat{\theta}$ using the original sample), which is the estimate of the $P$-value for the one-sided test, given the null hypothesis is true. We conducted this simple two-sample permutation test procedure between every pair of models. We also used other variants of non-parametric tests, but with all of them we achieved the same conclusions.

## Results

Most of our plots shown a random pattern in the spatial distribution of their trees, and closer toward a regular distribution. We fit a smooth curve to the empirical semivariogram in order to detect a general trend of the autocorrelation of *dbh*, which was used later for fitting GWR models. The semivariogram shown spatial correlation but only at small distances. Overall, our type of forest studied here do not offer us a great variability of spatial structures.

A rather simplified version of model 3 was sufficient for our data. Although our base formulation of model 3 (Eq. 8) allowed both the intercept and the slopes coefficient parameters to have a random-effect, we found better fit, as measured by AIC, when having only the intercept random-effects. That is, the final model 3 used was $dbh(\mathbf{s}_i) = (\beta_0 + b_{0i}) + \beta_1 lh(\mathbf{s}_i) + \beta_2 lcd(\mathbf{s}_i) + e(\mathbf{s}_i)$. Furthermore, we did not find significant improvement by imposing a marginal correlation, therefore, we set $\mathbf{\Omega} = \mathbf{I}_n$ in (8).

Studentized residuals changed both in magnitude and spatial distribution. Change, regards to OLS, in the spatial distribution pattern (i.e., distribution of positive and negative values) of the studentized residuals were only possible to observed for GWR (Figure 1d). The other statistical models do not change the spatial pattern of the residuals, which is in agreement with the findings of Zhang y Shi (2004), Zhang *et al*. (2004). However, LME produces much smaller studentized residuals than the other models (Figure 1c).

The ***RMSD*** of the models, that is computed based on all the trees of each plot (Figure 2), shows a clear difference between the LME and the rest of models, being not only the more precise but also by far the most accurate. ***AD*** and ***AAD*** by validation classes point again to LME as the model with a more consistent predictive perform across all validation classes. In the same manner, the proportion of field plots (i.e. cases) where each model was assigned to be the best, according to the proposed methodology, portrait to LME as the best in almost all (95%) the field plots used in our study (Table 2).

Statistical inference assessment of the ***RMSD*** between pairs of models showed significant differences. Statistical significance comparisons of the models did favor the best predictive models. There is indeed an ordering of the models, not only considering the pure ***RMSD*** but also its statistical significance, which is LME preferred to GWR preferred to OLS (Figure 3). Nevertheless, we did not find strong evidence against the null hypothesis of no differences between the ***RMSD***s of GLS and OLS. This is expected because GLS really affects (or improve) the estimates of the standard errors of the coefficient estimates, over the ones obtained using OLS.

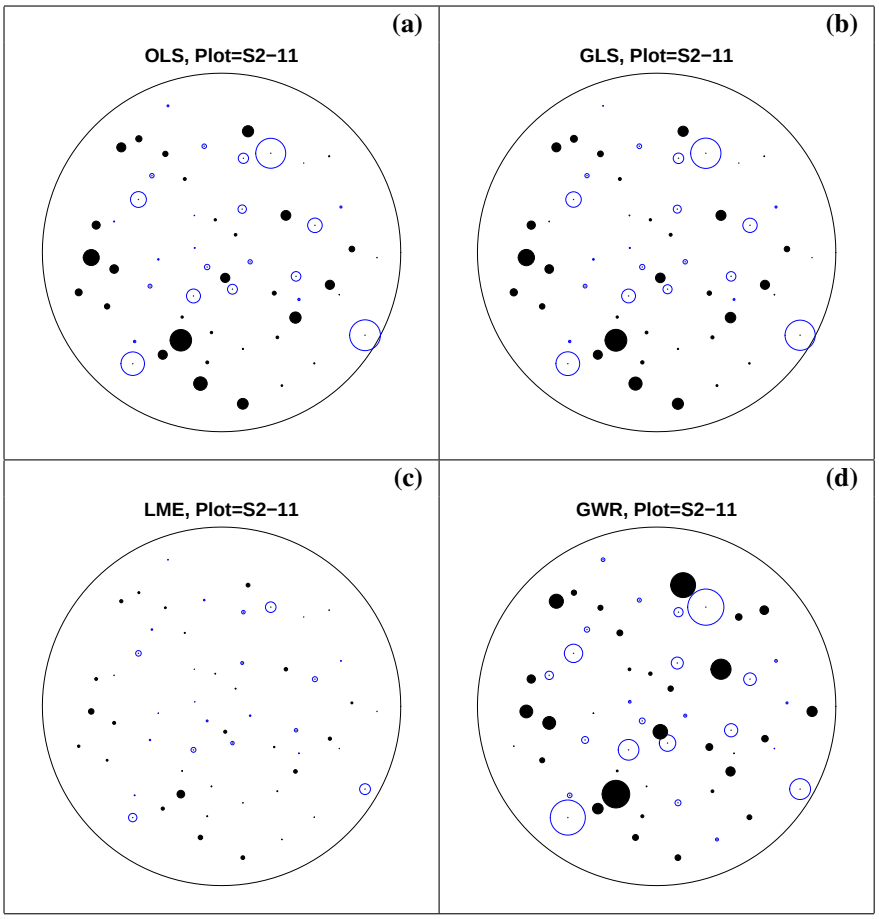


Figure 1: Graph of studentized model residuals (local $Z$-values) from the four modelling techniques [OLS(a), GLS (b), LME (c), and GWR (d)] when predicting diameter in a sample plot. The black dots represent positive local $Z$-value, and the circles represent negative local $Z$-value.

Table 2: Percentage of the plots (i.e., cases) where each model was chosen as the best according to its prediction performance.

| Model | Position | | | |
|---|---|---|---|---|
| | 1st | 2nd | 3rd | 4rd |
| GLS | 0.0 | 2.8 | 44.4 | 52.8 |
| GWR | 5.6 | 86.1 | 5.6 | 2.8 |
| LME | 94.4 | 5.6 | 0.0 | 0.0 |
| OLS | 0.0 | 5.6 | 50.0 | 44.4 |

## Discussions

The linear mean function (Eq. 1) seems to be appropriate for these data. We obtained a $RMSD$ of 18% for our model fit by OLS. Among other studies with models having the same predictor variables: Persson *et al*. (2004) obtained a 10% error, computed over all their sample trees, but with a much smaller sample (562 trees) and a smaller diameter range too; and unfortunately neither Heurich *et al*. (2004) nor Schardt *et al*. (2004), García *et al*. (2007) reported their *dbh* model errors. Heurich (2008), using models with more predictor variables, obtained a $RMSD$ of 19.9%[4] for his *dbh* models in 936 trees. Therefore we think our mean function performs reasonable well, and is appropriate to be used as reference for adding extra statistical features.

[4] This value was derived and computed proportional to the sample size of his forest types.

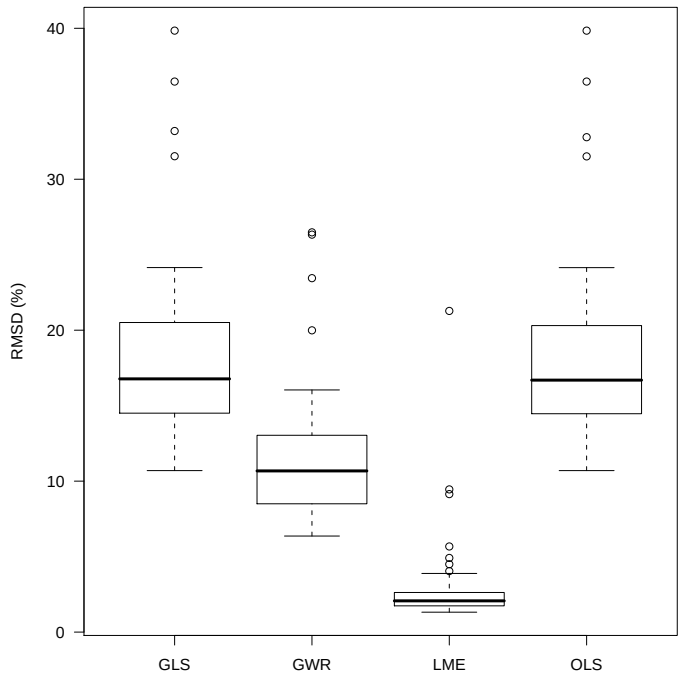


Figure 2: Box plot of the root mean square differences (*RMSD*) in percentage for predicting tree diameter in a field plot (i.e., each data point represents a *RMSD* for a plot) from the assessed statistical models. GLS is generalized least squares with correlation structure, GWR is geographically weigthed regression, LME is linear mixed-effects, and OLS is ordinary least squares.

LME offered a series of key features in comparison to the other statistical models. We obtained the best precision and accuracy in predictions using LME. Prediction errors with LME are much lower (4%) than what can be achieved having a rather overparametrized model such as GWR (10%), where a full model is fitted for each observation. Not only were the *RMSD*s from LME statistically significant different from all the other models, but also is a more appropriate model for spatially correlated data. GLS models the spatial correlation of the residuals, its prediction errors are not significantly different from the one obtained from a simpler OLS model. However, standard error estimates under OLS are biased, which raises doubts about the correctness of statistical inference following OLS estimation. Finally, LME also offers us a statistical pathway to further explore the relationship being modelled. We could extract the random-effects and relate them to other potential variables that are needed and are not being currently considered in our skeleton model.

The use of GWR in practical applications is unclear. Although several studies have portrayed GWR as performing well and having a series of advantages over other model fitting methods (e.g., Zhang y Shi 2004, Zhang *et al*. 2004, Zhang y Gove 2005, Zhang *et al*. 2005), few studies have focused on how a GWR model could be used to predict $Y$ for a new observation which was not used in fitted the model (i.e., impute). One way for doing so is having the original data set used for fitting the model, and that is done in a type of cross-validation manner (e.g., see the explanation of GWR prediction in Leung *et al*. 2000, , p.27). Indeed, alternatives such as interpolating in a map given the GWR predictions for known points or interpolating the parameter estimates (which is the final aim of GWR), is possible, but the statistical properties of interpolated predictions may be difficult to discern. On the other hand, LME use in an imputation setting is straightforward. For LME we could only use the fixed-effect parameter estimates to generate population-average prediction. It is also possible to calibrate the imputation using the BLUP of the random-effects using only a few new observations (Lappi y Bailey 1988). Notice that those new few observations needed for calibrating a LME would not be enough for fitting a new GWR model, therefore being much more practical than GWR.

Parametric statistical inference is questionable for GWR models. We proposed the use of permutation tests, as well as bootstrapping (which was not shown here) in order to test the null hypothesis of no difference in the population means of the prediction errors (*RMSD*) between models. The results at least for our data were still pretty clear: there

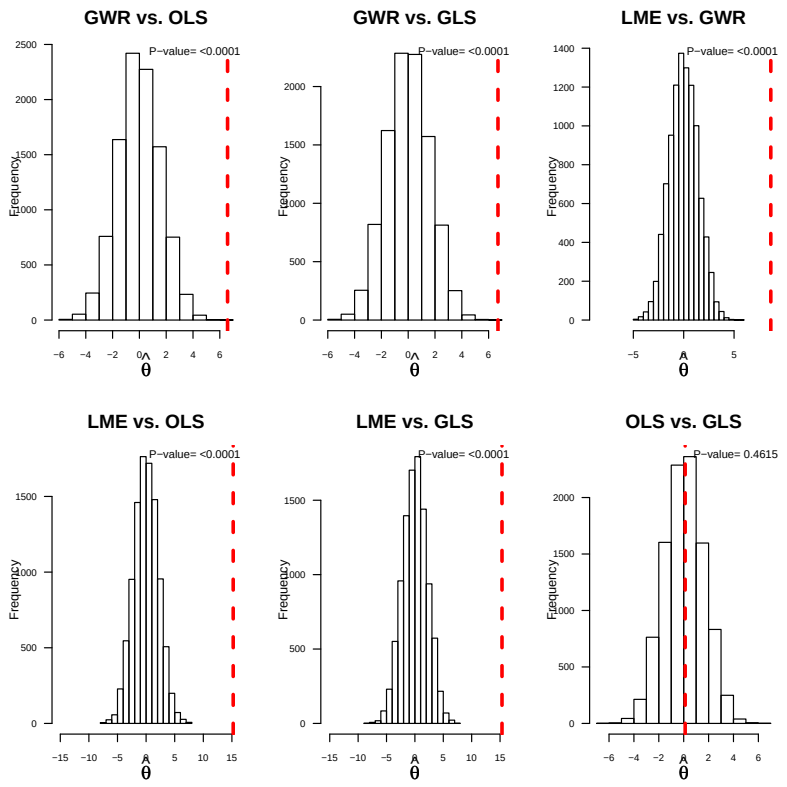


Figure 3: Permutation distribution of $\widehat{\theta}$ (i.e., the mean difference between *RMSD*s on predicting tree height from OLS, GLS, LME, and GWR models) using a simple two-sample permutation test. The vertical line show the statistics for our dataset. 10,000 permutations were conducted. The empirical one-sided $P$-value for the significance of the statistic is also printed.

are significant difference for LME over GWR, being also more parsimonious and holding a stronger statistical theory. Further research into the use of permutation and resampling methods to compare the performance of GWR models to parametric models is needed.

## Concluding remarks

Incorporating the tree spatial information available from LiDAR data into modelling tree diameter, allows for taking into account the spatial correlation of the data, contributing not only to improving prediction performance but also more appropriate statistical inference. Past modelling efforts from LiDAR data only used OLS, therefore our study has shown new alternatives for same efforts, as well as, reduce significatively prediction errors. Given the same type of LiDAR data, we can achieve a better modelling of tree diameter, a key variable for forest characterization. The different pitfalls pointed out for GWR avoid a sound statistical use of it, however, nevertheless, we recognized it can be a useful exploratory spatial data technique, which it seemed to be its original aim. To our knowledge, this is the first study analyzing the use of spatial statistical models for LiDAR data. Our results show that a mixed-effects model offers the best predictions of *dbh* from LiDAR -based variables.

## References

Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. *In* Petrov BN, G Czaki eds. Second International Symposium on Information Theory. Akademiai Kiadó, Budapest, Hungary, p. 267–281.

Bivand R, D Yu. 2009. `spgwr`: Geographically weighted regression. R package version 0.5-7.

Bivand RS, EJ Pebesma, V Gómez. 2008. Applied Spatial Data Analysis with R. Springer, New York, NY, USA. 374 p.

Brunsdon C, M Aitkin, S Fotheringham, M Charlton. 1999a. A comparison of random coefficient modelling and geographically weighted regression for spatially non-stationary regression problems. *Geographical & Environmental Modelling* 3(1):47–62.

Brunsdon C, AS Fotheringham, M Charlton. 1999b. Some notes on parametric signficance tests for geographically weighted regression. *Journal of Regional Science* 39(3):497–524.

Brunsdon C, AS Fotheringham, ME Charlton. 1996. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis* 28(4):281–298.

Brunsdon C, S Fotheringham, M Charlton. 1998. Geographically weighted regression–modelling spatial non-stationary. *The Statistician* 47(3):431–443.

Cressie N. 1993. Statistics for Spatial Data. Wiley Interscience, New York, USA. 928 p.

Cressie N, CA Calder, JS Clark, JMV Hoef, CK Wikle. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications* 19(3):553–570.

Curtis RO. 1967. Height-diameter and height-diameter-age equations for second-growth Douglas-fir. *For. Sci.* 13(4):365–375.

Dennis B. 1996. Should ecologists become Bayesians? *Ecological Applications* 6(4):1095–1103.

Diggle PJ, , PJ Ribeiro. Jr. 2007. Model-based Geostatistics. Springer, New York, USA. 228 p.

Efron B. 1986. Why isn't everyone a Bayesian? (with discussion). *The American Statistician* 40(1):1–11.

Ellison AM. 2004. Bayesian inference in ecology. *Ecology Letters* 7:509–520.

Finley AO, S Banerjee, RE McRoberts. Submitted. Hierarchical spatial models for predicting tree species assemblages across large domains. *(Manuscript submitted to the Annals of Applied Statistics)* .

Fotheringham AS, ME Charlton, C Brunsdon. 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A* 30:1905–1927.

Fotheringham S, C Brunsdon, M Charlton. 2002. Geographically Weighted Regression: the analysis of spatially varying relationships. Wiley, New York, USA. 269 p.

García R, JC Súarez, G Patenaude. 2007. Delineation of individual tree crowns for LiDAR tree and stand parameter estimation in scotish woodlands. *In* Fabrikant SI, M Wachowicz eds. The European Information Society: Leading the Way with Geo-information. Springer. New York, USA, p. 55–87.

Gelfand AE, HJ Kim, CF Sirmans, S Banerjee. 2003. Spatial modeling with spatially varying coefficient processes. *J. Am. Stat. Assoc.* 98(462):387–396.

Henry HA, LW Aarssen. 1999. The interpretation of stem diameter-height allometry in trees: biomechanical constrainsts, neighbour effects, or biased regressions? *Ecology Letters* 2:89–97.

Hesterberg T, S Monaghan, DS Moore, A Clipson, R Epstein. 2003. Bootstrap Methods and Permutation Tests. W. H. Freeman and Company, New York, USA. 74 p.

Heurich M. 2008. Automatic recognition and measurement of single trees based on data from airborne laser scanning over the richly structured natural forests of the bavarian forest national park. *For. Ecol. Manage.* 255:2416–2433.

Heurich M, A Persson, J Holmgren, E Kennel. 2004. Detecting and measuring individual trees with laser scanning in mixed mountain forest of central Europe using an algorithm developed for Swedish boreal forest conditions. *In* Thiers M, B Kock, H Spiecker, H Weinacker eds. Proceedings of the ISPRS (International Society of Photogrammetry and Remote Sensing) working group part 8/2. Int. Arch. Photogramm. Remote Sensing, Vol. 36. Freiburg, Germany, p. 307–312.

Heurich M, F Thoma. 2008. Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural European beech (*Fagus sylvatica*) and Norway spruce (*Picea abies*) forests. *Forestry* 81(5):645–661.

Hoeting JA. 2009. The importance of accounting for spatial and temporal correlation in analyses of ecological data. *Ecological Applications* 19(3):574–577.

Hoeting JA, RA Davis, AA Merton, SE Thompson. 2006. Model selection for geostatistical models. *Ecological Applications* 16(1):87–98.

Hudak A, N Crookston, J Evans, D Hall, M Falkowski. 2008. Nearest neighbor imputation of specie-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment* 112:2232–2245 (and corrigendum in 113:289–290).

Hudak AT, JS Evans, MJ Falkowski, NL Crookston, PE Gessler, P Morgan, A Smith. 2005. Predicting plot basal area and tree density in mixed-conifer forest from lidar and advanced land imager (ALI) data. *In* The 26th Canadian Symposium on Remote Sensing. Wolfville, Nova Scotia, p. CD–ROM. 7 p.

Hurvich CM, JS Simonoff, CL Tsai. 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. R. Stat. Soc. B.* 60(2):271–293.

Hyyppä J, O Kelle, M Lehikoinen, M Inkinen. 2001. A segmentation-based method to retrieve stem volume estimates from 3-d tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing* 39(5):969–975.

Ives AR, J Zhu. 2006. Statistics for correlated data: Phylogenies, space, and time. *Ecological Applications* 16(1):20–32.

Kissling WD, G Carl. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography* 17(1):59–71.

Lappi J. 1991. Calibration of height and volume equations with random parameters. *For. Sci.* 37:781–801.

Lappi J, RL Bailey. 1988. A height prediction model with random stand and tree parameters: An alternative to traditional site index methods. *For. Sci.* 34:907–927.

Leung Y, CL Mei, WX Zhang. 2000. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A* 32:9–32.

Lynch TB, AG Holley, DJ Stevenson. 2005. A random-parameter height-dbh model for cherrybark oak. *South. J. Appl. For.* 29(1):22–26.

Manly BFJ. 2006. Randomization, Bootstrap and Monte Carlo Methods in Biology. 3rd edition. Chapman & Hall/CRC, London, England. 455 p.

Meyer HA. 1940. A mathematical expression for height curves. *J. Forestry* 38(5):415–420.

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

Næsset E, T Gobakken. 2004. Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scandinavian J. For. Res.* 19:529–542.

Næsset E, T Gobakken, J Holmgren, H Hyyppä, J Hyyppä, M Maltamo, M Nilsson, H Olsson, A Persson, U Söderman. 2004. Laser scanning of forest resources: The Nordic experience. *Scandinavian J. For. Res.* 19:482–499.

Nelson RE, A Short, M Valenti. 2004. Measuring biomass and carbon in Delaware using airborne profiling LiDAR. *Scandinavian J. For. Res.* 19:500–511.

Nelson RF, TG Gregoire, , RG Oderwald. 1998. The effects of fixed-area plot width on forest canopy height simulation. *For. Sci.* 44(3):438–444.

Peng C, L Zhang, J Liu. 2001. Developing and validating nonlinear height-diameter models for major tree species of Ontario's boreal forests. *North. J. Appl. For.* 18(3):87–94.

Persson Å, J Holmgren, U Söderman. 2004. Detecting and measuring individual trees using an airbone laser scanner. *Photogramm. Eng. Rem. Sens.* 68(9):925–932.

Pinheiro J, D Bates, S DebRoy, D Sarkar, the R Core Team. 2008. `nlme`: Linear and Nonlinear Mixed Effects Models. R package version 3.1-89.

Pinheiro JC, DM Bates. 2000. Mixed-effects Models in S and Splus. Springer-Verlag, New York, USA. 528 p.

Popescu SC, RH Wynne. 2004. Seeing the trees in the forest: using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. *Photogramm. Eng. Rem. Sens.* 70(5):589–604.

Popescu SC, RH Wynne, RE Nelson. 2002. Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window. *Computers and Electronics in Agriculture* 37(1-3):71–95.

Pukkala T. 1989. Predicting diameter growth in even-aged scots pine stands with a spatial and non-spatial model. *Silva Fennica* 23(2):101–116.

Reutebuch SE, HE Andersen, RJ McGaughey. 2005. Light detection and ranging (LIDAR): An emerging tool for multiple resource inventory. *J. Forestry* 103(6):286–292.

Robinson AP, WR Wykoff. 2004. Imputing missing height measures using a mixed-effects modeling strategy. *Can. J. For. Res.* 34:2492–2500.

Robinson GK. 1991. That BLUP is a good thing: The estimation of random effects. *Stat. Sci.* 6(6):15–51.

Rönnholm P, J Hyyppä, H Hyyppä, H Haggrén, X Yu, H Kaartinen. 2004. Calibration of laser-derived tree height estimates by means of photogrammetric techniques. *Scandinavian J. For. Res.* 19(6):524–528.

Salas C. 2002. Ajuste y validación de ecuaciones de volumen para un relicto del bosque de roble-laurel-lingue. *Bosque* 23(2):81–92.

Sánchez CAL, JG Varela, FC Dorado, AR Alboreca, RR Soalleiro, JGA González, FS Rodríguez. 2003. A height-diameter model for *Pinus radiata* D. Don in Galicia (Northwest Spain). *Ann. For. Sci.* 60(3):237–245.

Schabenberger O, CA Gotway. 2005. Statistical Methods for Spatial Data Analysis. Chapman & Hall/CRC, Boca Raton, FL, USA. 512 p.

Schabenberger O, TG Gregoire. 1996. Population-avaeraged and subject-specific approaches for clustered categorical data. *Journal of Statistical Computation and Simulation* 54:231–253.

Schabenberger O, FJ Pierce. 2002. Contemporary Statistical Models for the Plant and Soil Sciences. CRC Press, Boca Raton, FL, USA. 738 p.

Schardt M, M Ziegler, A Wimmer, R Wack, J Hyyppä. 2004. Assessment of forest parameters by means of laser scanning. *In* Thiers M, B Kock, H Spiecker, H Weinacker eds. Proceedings of the ISPRS (International Society of Photogrammetry and Remote Sensing) working group part 8/2. Int. Arch. Photogramm. Remote Sensing, Vol. 36. Freiburg, Germany, p. 272–276.

Schröder J, JG Álvarez. 2001. Comparing the performance of generalized diameter-height equations for maritime pine in Northwestern Spain. *Forstw. Cbl.* 120:18–23.

Stage AR, NL Crookston. 2007. Partitioning error components for accuracy-assessment of near-neighbor methods of imputation. *For. Sci.* 53:62–72.

Staudhammer C, VM LeMay. 2000. Height prediction equations using diameter and stand density measures. *For. Chron.* 76(2):303–309.

Thorey LG. 1932. A mathematical method for the construction of diameter height curves based on site. *For. Chron.* 8(2):121–132.

van Laar A, A Akça. 2007. Forest Mensuration. Springer, Dordrecht, The Netherlands. 383 p.

Waller LA, CA Gotway. 2004. Applied spatial statistics for public health data. John Wiley & Sons, Hoboken, N.J, USA. 494 p.

Waller LA, L Zhu, CA Gotway, DM Gorman, PJ Gruenewald. 2007. Quantifying geographic variations in associations between alcohol distribution and violence: a comparison of geographically weighted regression and spatially varying coefficient models. *Stochastic Environmental Research and Risk Assessment* 21(5):573–588.

Welch WJ. 1987. Rerandomizing the median in matched-pairs designs. *Biometrika* 74(3):609–614.

Wheeler DC, LA Waller. 2009. Comparing spatially varying coefficient models: a case study examining violent crime rates and their relationships to alcohol outlets and ilegal drug arrests. *Journal of Geographical Systems* 11:1–22.

Yuancai L, BR Parresol. 2001. Remarks on height-diameter modeling. USDA For. Serv. Res. Not. SE-10, USA. 8 p.

Zhang L. 1997. Cross-validation of non-linear growth functions for modelling tree height-diameter relationships. *Annals of Botany* 79(3):251–257.

Zhang L, H Bi, P Cheng, CJ Davis. 2004. Modeling spatial variation in tree diameter-height relationships. *For. Ecol. Manage.* 189(2):317–329.

Zhang L, JH Gove. 2005. Spatial assessment of model errors from four regression techniques. *For. Sci.* 51(4):334–346.

Zhang L, JH Gove, L Heath. 2005. Spatial residual analysis of six modeling techniques. *Ecol. Modelling* 186:154–177.

Zhang L, Z Ma, L Guo. 2008. Spatially assessing model errors of four regression techniques for three types of forest stands. *Forestry* 81(2):209–225.

Zhang L, H Shi. 2004. Local modeling of tree growth by geographically weighted regression. *For. Sci.* 50(2):225–244.