Canadian Journal of Forest Research
Revue canadienne de recherche forestière

Analysis of spatial correlation in linear models to predict forest variables from LiDAR auxiliary information

Journal:	*Canadian Journal of Forest Research*
Manuscript ID	cjfr-2016-0296
Manuscript Type:	Article
Date Submitted by the Author:	10-Jul-2016
Complete List of Authors:	Mauro Gutiérrez, Francisco; Oregon State University, College of foretsry. Forest Engineering Resources and Management. Monleon, Vicente; USDA Forest Service Temesgen, Hailemariam; Oregon State University, Forestry; Ruíz Fernández, Luis Ángel; Geo-Environmental Cartography and Remote Sensing Group, Department of Cartographic Engineering, Geodesy and Photogrammetry, Universitat Politècnica de València., Cartographic Engineering, Geodesy and Photogrammetry
Keyword:	Spatial correlation, LiDAR, forest inventory, linear models, spatial models

Analysis of spatial correlation in linear models to predict forest variables from LiDAR auxiliary information.

F. Mauro

Oregon State University, College of Forestry, Forest Engineering Resources and Management Department. 053 Peavy Hall, Corvallis, OR 97331, USA E-mail: francisco.mauro@oregonstate.edu

V.J. Monleon

US Forest Service, Pacific Northwest Research Station, Corvallis Forestry Sciences Laboratory 3200 SW Jefferson Way, Corvallis, OR 97331, USA E-mail: vjmonleon@fs.fed.us

H. Temesgen

Oregon State University. College of Forestry, Forest Engineering Resources and Management Department. 233 Peavy Hall, Corvallis, OR 97331, USA E-mail: temesgen.hailemariam@oregonstate.edu

L.A. Ruiz

Geo-Environmental Cartography and Remote Sensing Group, Department of Cartographic Engineering, Geodesy and Photogrammetry, Universitat Politècnica de València, ESP E-mail: laruiz@cgf.upv.es

Abstract

Accounting for spatial correlation of LiDAR model residuals can improve the precision of model based estimators. To model such spatial correlation, sample designs providing close enough observation are needed but they are difficult to implement. Aiming to provide references about the gains that can be obtained by accounting for the spatial correlation of model residuals, we analyzed: 1) The spatial correlation patterns of residuals from LiDAR linear models developed to predict volume, total and stem biomass per hectare, quadratic mean diameter (QMD), basal area, mean and dominant height, and stand density; 2) How the plot size changes the spatial correlation patterns.

For all variables the correlation range of model residuals consistently increased with the plot radius and was always below 60 m except for stand density, where it reached 85 m. Excluding QMD, depending on the radius and variable of interest, correlation ranges of model residuals were from 1.06 to 8.16 times shorter than those observed for the raw variables. Based on the sort correlation ranges observed, the assumption of independent residuals accepted in numerous studies without enough empirical evidence, seems to be reasonable appropriate which raises questions about the practical need of accounting for spatial correlation in LiDAR inventories.

***Keywords**:* Spatial correlation, LiDAR, forest inventory, linear models, spatial models.

1 Introduction

Remotely sensed auxiliary information from airborne laser scanners (ALS), in combination with the area based approach (ABA), have been extensively used to assist forest inventories during the last two decades (Næsset 1997a, Magnussen et al. 1999, Næsset and Bjerknes 2001, Andersen et al. 2005, González-Ferreiro et al. 2012). Under the ABA, a study area is covered by a grid (i.e. a compact tessellation with non-overlapping units) containing auxiliary information for each pixel. Pixels are considered population elements or units, so that the grid implicitly defines a pseudo sampling frame. Then, the variables of interest are measured in a sample of field plots of size similar to that of the pixels. These field plots are regarded as elements of population for which both auxiliary information and the variables of interest are known.

The ABA in combination with model-based estimation methods have had a prominent role in forest inventories assisted with LiDAR and we will focus our study on it. The impact of the potential spatial correlation of the model residuals, and how this spatial correlation can be accounted for, have received significant attention and different authors have provided estimators to account for it in forests inventories assisted with spatially explicit auxiliary information (McRoberts, 2006; McRoberts et al., 2007; Breidenbach. et al., 2008; Magnussen et al., 2009; Ver Hoef and Temesgen, 2013; Temesgen and Ver Hoef, 2014; Finley et al., 2014; Magnussen et al., 2016a, Magnussen et al., 2016b). A related issue that has not been studied in the literature is how field plot size affects the potential spatial correlation. This issue has important practical consequences as plot size has a large impact on both fieldwork costs and the ABA work-flow.

1.1 Spatial correlation in forest management inventories

Spatial correlation of model residuals has been frequently ignored in operational forest inventories assisted with remotely sensed auxiliary information, thus assuming that model residuals are independent with little empirical evidence. Sampling designs typically use grids of plots where the field observations are too far to detect the spatial correlation. In those cases (Woods et al. 2011, Mauro et al. 2016), the assumed independence for the residuals is more a consequence of the inability to observe spatial correlation patterns than an empirically tested result of the model. Assuming independence because of a lack of field measurements may result in loss of predictive power as knowledge about the residual spatial correlation can be used to improve predictions. In addition, omitting accounting for the spatial correlation can result in unrealistic measures of uncertainty (Breidenbach et al. 2016). For these reasons, there has been an increasing interest on analyzing the spatial correlation of model residuals in LiDAR based forest inventories.

For a linear spatial model, best linear unbiased prediction (BLUP, or kriging in the geostatistical literature) incorporates the spatial correlation to improve predictions. This technique rely on a model obtained for the residual spatial correlation that is used in a further stage where the variable of interest in unsampled locations is estimated, both to create maps of its distribution and to estimate block averages The improvement of the prediction, however, is greatest for the pixels closest to the observed plots and is negligible for pixels located beyond the range of the semivariogram, when the spatial correlation is close to 0. Thus, the shape of the semivariogram closest to the origin is of the greatest importance for spatial prediction (Cressie 1993). In addition, in some cases, and especially so in LiDAR based forest inventories, spatially explicit auxiliary information variables with high explanatory power may account for an important part of the spatial correlation of the variable of interest so the correlation that is left among model residuals can show short spatial ranges (Breidenbach. et al. 2008, Finley et al. 2014, Breidenbach et al. 2016). The need of observations of the spatial correlation close to the origin and the potentially short range of such correlation raise important questions about the survey designs needed to account for this factor.

To detect and model spatial correlation of model residuals, fieldwork design needs to ensure the existence of pairs of observations at distances where spatial correlation is still present. For certain variables of interest and auxiliary information sources, the observed spatial correlation range of the residuals was larger than the minimum distance between plots, reaching distances of about 1 km to 3 km (Magnussen et al. 2009, Ver Hoef and Temesgen 2013, Finley et al. 2014). However, in forest inventory applications, especially when using LiDAR or photogrammetric point clouds, it seems to be more common to observe residual spatial correlation that vanishes at distances from 10 m to 200 m (Breidenbach. et al. 2008, Finley et al. 2014, Breidenbach et al. 2016). These distances are significantly shorter than, or at most close to, the minimum separation between field observations. This problem could be even more significant in large scale forest inventories where, for example, (McRoberts et al. 2007) reported spatial correlation ranges of presence/absence of forestland that did not even reach 200 m. These issues raise important questions regarding fieldwork protocols. Zimmerman, (2006) studied this problem and concluded that designs with uniformly spread plot locations are optimal for prediction in unsampled locations when the spatial correlation parameters are known, but clustered sample plots are optimal for estimation of the spatial correlation parameters. In actuality, spatial correlation parameters are not known and prediction at unsampled locations is also needed, so best designs are those in which clusters of nearby sample plots are uniformly spread throughout the study area (Zimmerman, (2006).

The desirable option to study spatial correlation is to rely on data to directly analyze it, which may require special designs. Recent studies have proposed ways to obtain estimates of the spatial correlation when it cannot be directly estimated due to a lack field observations close enough to observe the correlation. These methods rely on estimating the residual spatial correlation from model predictions of the variable of interest (Magnussen et al., 2016a, Magnussen et al., 2016b). Unfortunately, these approaches try to overcome the problem of lack of field information by relying on strong assumptions relating spatial correlation of predictions and spatial correlation of model residuals, that cannot be empirically confirmed in the context of those studies. Insights about the potential importance of the spatial correlation of model residuals could be obtained from previous studies such as (Gunnarsson et al. 1998) where no auxiliary information was used to predict different forest attributes of interest. Based on the spatial nature of LiDAR predictors one could expect smaller correlation ranges for the residuals than those observed for the attributes of interest, so information from previous studies might be considered as an upper bound for the spatial correlation of LiDAR model residuals.

Due to difficulties to obtain data appropriate for studying spatial correlation of model residuals models using LiDAR as auxiliary information, studies on this topic are not very numerous (Breidenbach. et al. 2008, Finley et al. 2014). This number increases by one if one considers photogrammetric point clouds. The number of variables analyzed is very small, limited to volumes and stand table data, and no previous study has considered variables such as mean height, dominant height, above ground biomass, quadratic mean diameter or basal area.

1.2 Support region overlap

Field plots and pixels are not points, but fixed areas. When using either LiDAR or photogrammetric point clouds, spatial correlation has been always studied assuming that the distance between units is the distance between their centroids (Breidenbach. et al., 2008; Magnussen et al., 2009; Finley et al., 2014; S. Magnussen et al., 2016 a, Magnussen et al., 2016 b). An inherent consequence of having a support area is that, as the distance between units decreases, support regions can overlap. While this may not be an issue if the population is portioned into a grid of non-overlapping units (pixels) and a sample taken from those units, the reality of LiDAR supported ABA inventory is that the population grid and field plots are misaligned, therefore field plots and pixels can overlap (Figure 1). For example, field plot locations are often determined before knowing the LiDAR grid (e.g. Finley et al., 2014; Mauro et al., 2016), which would likely result in misalignment. The same would happen for systematic sampling designs, if the separation between plots is not a multiple of the pixel size (i.e. when national forest inventory field plots are used e.g. Breidenbach and Astrup, 2012). Even if the locations of the field plots were planned to coincide with a pixel center, field crews typically navigate to pre-selected plot locations using coarse acquisition (C\A) code based Global Navigation Satellite Systems (GNSS) techniques (Breidenbach et al. 2016), resulting in positioning errors and misalignment between field plots and pixels that cannot be corrected in a subsequent post-processing of GNSS phase observations taken at the plot center. Studies on GNSS positioning using real time C\A code in forested environments reported errors that reached 21.60 m when no external corrections were used and 14.01 m when using code corrections (Andersen et al. 2009). Coordinates of plot centers can be differentially corrected later using relatively long phase observations, which would allow computing more accurate coordinates and ensure a correct matching of LiDAR point cloud and plot measurements. However, the computation of refined coordinates does not solve the initial misalignment of field plots and grid units.

Plot overlap induces a correlation, because the overlapping area is measured by both plots. However, spatial statistics methods make use of spatial correlation but do not try to explain the reasons that generate it (McRoberts 2006, Breidenbach. et al. 2008, Magnussen et al. 2009, Ver Hoef and Temesgen 2013, Finley et al. 2014, Temesgen and Ver Hoef 2014, Breidenbach et al. 2016). Even when induced by overlap, this correlation is a form of spatial correlation, inasmuch as it is a function of the distance between plot centers: it is 1 when the distance between the plot centers is 0, and decreases as the distance increases. There are neither practical nor theoretical reasons to distinguish between the correlation induced by overlap and a hypothetical, non-overlapping spatial correlation at distances where overlap does exist. Because of misalignment, partial overlap between the grid that partitions the population and the field plots is a reality, so knowing the correlation at those distances where the overlap is still present would help improve the prediction for grid cells that share area with the field plots. Estimation of the empirical autocorrelation function is the same whether there is overlap or not, for it is simply the correlation between two plots at a given distance between plot centers: it does distinguish whether there is overlap or not. Once this function is estimated, fitted models may account for differences between overlapping and non-overlapping portions of the autocorrelation function if warranted by the data.

1.3 Field plot size

Field plot size has important consequences for both fieldwork cost and the ABA workflow, and may also influence the spatial correlation. Larger plots contain more trees, thus requiring more measurements and being more expensive. However, larger plots may be less sensitive to positioning errors (Gobakken and Næsset 2009) and would result in better estimator quality (Ruiz et al. 2014), as the variance of the estimators is smaller in larger plots. In addition, at close distances, correlation of residuals from larger plots is expected to be greater than correlation of residuals from smaller plots, as more plot overlap can be expected for larger plots. Therefore, the use of smaller plots is desirable to avoid expensive fieldwork, and could reduce the need to account for spatial correlation. It is therefore necessary to find compromise solutions for the plot size, and information on how plot size and spatial correlation interact is relevant to that end. However, to the best of our knowledge, no study has analyzed such interaction in a LiDAR assisted inventory context.

2 Objectives

The objectives of the study are:

1. Study the spatial correlation of residuals from models to predict forest attributes using LiDAR and relate spatial correlation ranges of the residuals to spatial correlation ranges for the raw attributes of interest. We focus on analyzing the spatial correlation at short distances, which are the most relevant for spatial prediction.
2. Examine the interaction between plot size and spatial correlation.

We studied a group of variables that can be considered as representative sample of the type of variables estimated in forest management inventories. Variables of interest were volume (V (m^3/ha)), total biomass per hectare (B_{tot} (kg/ha)), stem biomass per hectare (B_{stem} (kg/ha)), quadratic mean diameter (QMD (cm)), basal area (G (m^2/ha)), mean tree height (H_m (m)), dominant height (H_o (m)) and stand density (N (stems/ha)). Plot radii ranged from 7.5 m to 12.5 m.

3 Material and methods

3.1 Study area and AOI hierarchy

The study area is a 4000 ha forest located in "La Serranía de Cuenca", central Spain, described in Ruiz et al. (2014). Approximately 5% of the area is non-forested (considering as forest those areas with at least 10% canopy cover (FAO 2012)). European black pine (*Pinus nigra* Arn.) and scots pine (*Pinus sylvestris* L.) are the main species and appear mixed in different proportions. Black pine, however, dominates the forest in approximately 80% of the study area. In addition, other conifers such as Spanish juniper (*Juniperus thurifera* L.), maritime pine (*Pinus pinaster* Ait.) and hardwoods (e.g. holm oak (*Quercus ilex* L.) and Portuguese oak (*Quercus faginea*. Lam)) appear scattered over the study area. Slopes are very steep and the configuration of the hydrological network, with a main river crossing the study area from north to south and several seasonal tributaries running in east or west direction to join the main stream, result in a patch of areas with clearly differentiated slopes and orientations. The study area contains a total of 55 delineated stands ranging in area from 28.34 ha to 75.92 ha. These stands were merged into 13 management units (MU) containing one or more stands. Stands grouped to form each MU are subject to similar treatments so the composition and structure of the MU is homogeneous. The area of the MUs ranges from 30.64 ha to 392.34 ha.

3.2 LiDAR data

LiDAR data were collected in November 2008 using an Optech ALTM-1225 operating at 25 kHz and a maximum scanning angle of ± 18°. The minimum nominal LiDAR density was 4 points/m^2. The resulting average point density was 11.4 points/m^2, however, the point density was not homogeneous, due to irregular overlap of scanning stripes. The LiDAR point cloud was thinned using the software lastools (Isenburg 2013) to obtain a homogeneous density of 4 points/m^2. Ground points were filtered from the LiDAR point cloud and used to obtain a digital terrain model of 0.5 m pixel, that was employed to normalize the LiDAR point cloud. A visual inspection of the DTM and of the normalize point cloud were performed to ensure that these products were in fact free of spikes and outliers. All these processes were performed using FUSION (Mc Gaughey 2014).

3.3 Field data collection and locating of field plots

A total of 85, 25 m radius field plots (1963.5 m^2) were measured in December 2008. The radius of field plots typically used in LiDAR based forest inventories range from 9 m to 12.5 m (Ruiz et al. 2014), so the plots in this study are 4 to 7.72 times larger than commonly used field plots, which allowed studying the spatial correlation at relatively short distances.

Plots were located on the nodes of a 500 m regular grid. Field crews navigated to the preselected plot centers using a navigation grade Global Positioning System (GPS) using C\A code. Coordinates, relative to the plot center, of each tree with diameter at breast height (DBH) larger than 7 cm were obtained using a measuring tape and a compass. The expected accuracy of the relative positioning based on previous experience was around 0.5 m. Each tree was measured for DBH and height using a caliper and a Hagölf Vertex III hypsometer. Volume of each tree was computed using species specific regional equations developed by the Spanish National Forest Inventory (NFI) using *DBH* and *H* as predictors. Tree level total and stem biomass were computed for each tree using species specific models developed by (Montero et al. 2005) using DBH as the only predictor.

Positioning errors of navigation grade GPS devices can frequently exceed 5 m and should be corrected to ensure a precise co-registration with the LiDAR data. For each plot, trees were first positioned using their coordinates relative to the plot center and overlaid on the orthophoto and on the LiDAR point cloud resulting from the filtering of the ground points. Then a manual correction was performed by a photo interpreter, and relied on the identification of at least seven different trees in both the digital canopy height model (DCMH), the orthophoto of the study area and the ground point cloud. Tree stem locations were identified as maxima in the DCHM and gaps in in the ground point cloud derived from LiDAR. All trees in a plot were manually translated and rotated as a block until most isolated and easy to identify trees overlapped with the stem locations identified from the LiDAR image (Figure 2). Certain trees were moved independently in each plot when their position was identified on the ground point cloud and on the orthophoto. These trees were less than a 0.5% of the total. The average displacement of the plot center was 1.13 m, the standard deviation of the displacement was 1.72 m and the maximum displacement was 9.14 m.

3.4 Model fitting and spatial correlation assessment

For each variable of interest, we estimated linear spatial models where the mean of the distribution was a function of typical LiDAR covariates (i.e. percentiles, moments, means, minimums and maximums of the LiDAR elevations as well as cover parameters such as percentages of returns above different height thresholds, Mc Gaughey, 2014) and correlation between residuals for two locations a function of the distance separating them. Suitable model were selected as follows: first, selected the LiDAR predictors, then we considered a weighting schema to account for heteroscedasticity. Then, we added a random effect for the management unit. Finally, we modeled the spatial autocorrelation as a function of subplot distance.

3.4.1 Computation of sub-plot level values and auxiliary information

For each 25 m radius plot, groups of subplots of radii 7.5, 8, 8.5,... 11.5, 12 and 12.5 m were created. Each group of subplots was obtained by first defining a subplot, concentric to the 25 m radius plot. New subplots were defined by moving outwards the central subplot in steps of 0.5 m following E-W, SE-NW, S-N and SW-NE directions, until the edge of the subplots were tangent to the 25 m radius plot. The number of subplots in each 25 m plot, the total number of subplots and the maximum distances between subplots is indicated in Table 1. Note that the number of steps and the maximum distance between subplots ($max\_distance(radius)$) is different for each radius (Figure 3) and equals 50 m minus two times the subplot radius. For example, if we consider 10 m radius subplots, $max\_distance$ equals 30 m and it is possible to allocate 30/0.5 +1=61 subplots in each moving direction within a large plot. For each subplot, the variables of interest were calculated and expanded to a per hectare basis when appropriate. Similarly, a set of 30 LiDAR predictors were computed for each subplot using FUSION (Mc Gaughey 2014).

3.4.2 Non-spatial models

Linear fixed effects models were fit to the variables of interest, as a function of the LiDAR variables. Because of the very large number of potential predictor variables, a parsimonious model was selected as follows: First, fixed effects linear models were selected using the R package leaps (Lumley 2009) based on the 12.5 m radius subplot only. The maximum number of predictors was set to 3 independent variables per model. We obtained the best 5 models in terms of adjusted coefficient of determination when considering 1, 2 and 3 auxiliary variables which makes a total of 15 models. These models were denoted as $m_{0,vrbl,l}$ where $vrbl$ is a sub-index to denote the variable of interest, and subscript l = 1, 2,..., 15 indicates the candidates.

Typically, the variance of the model residuals was not constant, so a new set of 15 models, $m_{1,vrbl,l}$ accounting for heteroscedasticity, was fit using the R package nlme (Pinheiro et al. 2015). For each of the 15 models selected previously, the standard deviation of the residuals σ_e was assumed to be proportional to a power of the predictor most correlated with the variable of interest, mcp^{η}, so that $\sigma_e = \sigma_{e0} mcp_l^{\eta}$, where η was a parameter. Models $m_{0,vrbl,l}$and $m_{1,vrbl,l}$ can be respectively expressed for a given unit j in the i^{th}management unit as

$$y_{ij} = \boldsymbol{\beta}\boldsymbol{x}_{ij} + e_{ij} \quad [1]$$

Where the variance of e_{ij}, $V(e_{ij}) = \sigma_{0e}^2$ for $m_{0,vrbl,l}$, and $V(e_{ij}) = \sigma_{0e}^2 mcp_{l,i,j}^{2\eta}$. Note that $m_{0,vrbl,l}$ is a particular (i.e. nested) case of $m_{1,vrbl,l}$, where $\eta = 0$. Then, $m_{0,vrbl,l}$ and $m_{1,vrbl,l}$ were compared using a likelihood-ratio and $m_{0,vrbl,l}$ was selected when including the heteroscedasticity did not improved the model fit significantly, and $m_{1,vrbl,l}$ otherwise (Pinheiro and Bates 2000).

Finally, a management unit random effect was added to the models selected in the previous step. Resulting models can be expressed as

$$y_{ij} = \boldsymbol{\beta}\boldsymbol{x}_{ij} + v_i + e_{ij} \quad [2]$$

Here v_i are, the Management unit random effects. These effects are assumed to be independent and identically distributed variables with variance $V(v_i) = \sigma_v^2$. and $V(e_{ij})$ is the one determined in the previous step. These models were denoted as $m_{2,vrbl,l}$ and both $m_{0,vrbl,l}$ and $m_{1,vrbl,l}$are specific cases of $m_{2,vrbl,l}$. The significance of the MU random effect was tested using a likelihood-ratio test (Pinheiro and Bates 2000). Selected models were denoted by $m^{*}_{vrbl,l}$ and Pearson´s standardized residuals and normality of management unit random effects were graphically assessed and one of the fifteen candidates was selected and called m^{**}_{vrbl}.

Finally, for other subplot radii, we kept constant the fixed effects selected for the 12.5 m radius plot model, m^{**}_{vrbl}. Then heteroscedastic variance patterns and MU random effects were included and their significance was tested as we did with the models for the 12.5 m radius subplots. The resulting models were denoted as $m^{**}_{vrbl,rad}$, where sub-index rad was included as it becomes necessary hereafter to index the subplot radius.

5.1.1 Spatial correlation assessment

To analyze the importance of the spatial correlation, Pearson's standardized residuals from $m^{**}_{vrbl,rad}$ were obtained for each subplot, radius and variable. Pearson correlations were computed for all pairs of subplots separated distances d of (0.5 m, 1 m, 1.5 m,…$max\_distance(radius)$) in each moving direction by using only pairs of observations on the same moving line (dashed-lines Figure 3). The result of this step can be regarded as a directional empirical correlation function. Then, all the pairs were pooled together to compute an isotropic empirical correlation function. Empirical correlation at distance d are denoted hereafter as $\omega_{vrbl,rad}(d)$ where subscripts meanings are the ones indicated in the previous section. Patterns of correlation were examined in order to select a suitable spatial correlation model for each variable of interest.

In a last step, for each variable and radius, the spatial correlation of the residuals of $m^{**}_{vrbl,rad}$ was modeled. Covariance of model residuals associated to location b and location c were expressed as $Cov(\epsilon_b, \epsilon_c) = \sigma_b\sigma_c G\left(d_{b,c}, \rho, \theta\right)$ where $G(d_{b,c}, \rho, \theta)$ is the correlation function, σ_b and σ_c, are the standard deviation of the unit level random effects of the b^{th} and c^{th} subplot, $d_{b,c}$ is the Euclidean distance between those elements. The model shape for $G(d_{b,c}, \rho, \theta)$ was chosen after observing the empirical correlation function and it is a mixture of two components:

$$G(d_{b,c}, \rho, \theta) = \theta I(d_{b,c} < 2rad)\left\{\left[acos\left(\frac{d_{b,c}}{2rad}\right)\right]\frac{2}{\pi} - \left(\frac{d_{b,c}}{\pi rad^2}\sqrt{rad^2 - \frac{{d_{b,c}}^2}{4}}\right)\right\} + (1-\theta)e^{-(\frac{d_{b,c}}{\rho})} \quad [3]$$

The first component is the proportion of overlapping area; the second component is a pure exponential model without nugget effect. $I(d_{b,c} < 2rad)$ is an indicator function and θ the weight for the first component. The effective range, denoted as φ hereafter is defined as the distance for which the correlation descends to 0.05, and it is a function of both ρ, θ and the plot radius.

5.1.2 Spatial correlation of raw variables and comparison to residual correlation

The spatial analysis described above was conducted on the residuals of LiDAR models to predict different attributes of interest. Due to the spatial nature of LiDAR predictors, one can expect a reduction of the spatial correlation of model residuals when compared to the spatial correlation of raw variables of interest. Once the auxiliary information is taken into account through the model the spatial correlation that is left in the residuals can be substantially smaller than that present for the raw variables. To assess the reduction of the spatial correlation once the auxiliary information was considered, we examined the spatial correlation of the raw variables of interest. As with the residuals, we modeled the spatial correlation patterns observed for the response variables using the correlation function in [4]. For each variable of interest and subplot radius empirical correlations, the covariance function, the correlation function, its parameters and the effective range

($w(d), Cov(d_{b,c}, \rho, \theta), G(d_{b,c}, \rho, \theta), \rho, \theta, \varphi$) were indexed using sub-indexes $vrbl$ and rad to denote the variable of interest and the subplot radius. A super-script res or raw was added to indicate model residuals or raw variables respectively. We computed the ratios of the effective empirical correlation ranges of the residuals and raw variables $\gamma_{vrbl,rad}^{dep,res} = \frac{\varphi_{vrbl,rad}^{raw}}{\varphi_{vrbl,rad}^{res}}$. These ratios summarize the reduction of the spatial correlation range, when comparing the raw variables with the residuals (i.e. random part that is left once the auxiliary information is used to predict a response).

5.1.3 Influence of plot size in the spatial correlation of residuals

To analyze how plot size interacted with the spatial correlation of the residuals two different comparisons were performed. First, for each variable we plotted the ranges of the spatial correlation models for the residuals against the plot radius and computed the correlation coefficients between these two variables. Same analysis was performed for the spatial correlation ranges of the raw variables. Second, a more detailed analysis directly using the empirical correlations observed rather than model parameters was performed. In this analysis subplot radius for which no overlap occurred at a given distance were considered. For each variable of interest and distances from 20 m to 30 m, all pairs $[\omega(d), rad]$ (computed Pearson's correlation at distance d and plot radius) were gathered and the effect of increasing the plot radius in the empirical correlation $\omega^{res}(d)$ of model residuals was tested by means of a Kendal's τ test.

6 Results

For all variables and subplot radii the empirical correlation at the maximum possible distance was always below 0.26, and in most cases it did not exceed 0.1. For 38.6% of the analyzed combinations variable of interest-radius (34 out of 88), the empirical correlation at the maximum distance was below 0.05.Thus, the sample always covered more than 74% of the range of possible values for the spatial correlation, in most cases the coverage was larger than 90% (Table 2) and in more than a third of the cases the empirical correlation apparently reached the range.

The estimated parameters of the selected models, $m^{**}_{vrbl,rad}$, for each variable of interest and subplot radius as well as the spatial correlation parameters for raw variables and residuals are shown in Appendix A, Table A1. The exploratory analysis revealed that the correlation of residuals as a function of distance showed decreasing pattern without marked differences between directions (Appendix A, Figure A1 shows H_m as an example) and an isotropic model without nugget effect (Eq. 1) was appropriate to capture the variability of the empirical correlation function (Figure 4). For G, with subplots radii 7.5,8.5,9 and 9.5 m and for N with subplots radius of 9.5, 10 and 10.5 m the right tail of the empirical correlation function for the residuals was specially flat. That resulted in models with a very large ρ parameter for the exponential part resulting in unrealistically large values when computing the effective ranges. This seven cases were removed in the remaining analysis but their parameters are reported in Apendix A Table A1. This effect was especially prominent for G, where computed ranges for these four subplot radii were orders of magnitude longer than those observed for other subplots radius with the same variables or for other variables.

For all raw variables, the effective spatial correlation range calculated from the model was always less than 200 m (Figure 5). For the residuals, except for N and the seven cases commented before, the effective range was below 60 m. The raw variable that shows the shortest spatial correlation range is QMD. In this study, the LiDAR variables do not explain much of the variability of QMD, and when included, the spatial correlation range of the residuals for this variable increased. The prediction of this variable is very poor and it seems that the LiDAR information introduces noise as $\gamma^{dep,res}_{QMD,rad}$ is close but smaller than one. Among the variables included in this study, N is the one that typically shows weakest correlation with LiDAR auxiliary information (Næsset 2002). In this case, the raw variable N exhibits the largest correlation range. The reduction in the correlation range after including the LiDAR is low and $\gamma^{dep,res}_{QMD,rad}$, ranges from 1.06 to 3.45. After QMD, the variables with the smalest correlation ranges are V, H_o and H_m being those ranges slightly lower for the first two. Early studies on prediction of forest variables have shown that LiDAR information is highly correlated to these structural variables, specially to H_o (Næsset 1997b, Magnussen et al. 1999, Næsset and Økland 2002, Næsset and Bjerknes 2001). The high predictive power of LiDAR data for this variable explains why the spatial correlation range of the residuals for H_o decreases to distances 3.14 to 5.14 times shorter than that observed for the raw variable ($\gamma_{H_o,rad}^{dep,res}$) and become even smaller than the spatial correlation ranges of the residuals for $H_{m,}$ Figure 5). For V, the values of $\gamma_{V,rad}^{dep,res}$ are similar or larger than those obtained for $\gamma_{H_o,rad}^{dep,res}$ and considerable larger than those observed for B_{tot} and B_{stem} (Figure 5). This can be explained by the high correlation of LiDAR predictors with tree height, a variable that was included in the tree volume equations, but not in the biomass equations. The remaining variables (B_{tot}, B_{stem}, G) show correlation ranges for the residuals larger than those observed for QMD,H_m, H_o and V and smaller than those observed for N and similarly, the reduction of the spatial correlation when the LiDAR auxiliary information is included is larger than that observed for QMD and smaller than that observed for H_o and V. Both B_{tot}, B_{stem}, and G, are related in different ways to N and tree height, which may explain this average behavior.

The spatial correlation range consistently increased with the plot radius for both the model residuals and the raw responses. For all variables, except for G and N, where four and three radius were excluded, this positive correlation was significant for the residual part (Figure 5). The correlation between plot radius and empirical correlation for non-overlapping plots was studied at 19 different distances for each variable (152 pairs variable-distance). For 136 cases (approximately 90% of the cases), the Kendall's τ coefficient was positive indicating that typically larger radii result in larger empirical correlation (Figure 6). This result suggests that assuming uncorrelated residuals might have a larger impact in subsequent estimates when using larger plots which raises questions for further research about the use of different plot sizes with different spatial correlation ranges.

7 Discussion

The spatial correlation of model residuals have received important attention recently as they might be required:1) to improve predictions 2) for computation of uncertainty measures for pixel and\or stand level estimates (Magnussen et al. 2016a) and 3) to up-scale LiDAR predictions from different inventories made with different plot\pixel sizes to a common area (Magnussen et al. 2016b). Both studies recognized that directly studying spatial correlation of model residuals requires field observations that are difficult to obtain and proposed methods to anticipate the spatial correlation of model residuals. Unfortunately, the proposed methods try to overcome this problem by relying on strong assumptions, such as proportionality of the spatial correlation range of predictions and model residuals that, in the context of those studies, cannot be empirically confirmed. The spatial correlation models for the residuals obtained here are empirical results that can be directly used in the three stages mentioned above. The ratios $\gamma_{vrbl,rad}^{raw,res}$ could be used to anticipate the spatial correlation of model residuals if previous information about the raw forest attributes of interest, such as that provided by Gunnarsson et al., (1998), were available in a similar study area. Although they are not direct estimates of the spatial correlation needed by the methods described in Magnussen et al. (2016a) and Magnussen et al.,(2016b) to obtain mean square error estimators and to scale up model predictions, the ratios $\gamma_{vrbl,rad}^{raw,res}$ were empirically obtained and provide an alternative way to anticipate spatial correlation of model residuals.

The demanding fieldwork needed is the main reason why studies analyzing correlation of residuals from models specifically based on LiDAR auxiliary information are not very numerous, (Breidenbach. et al., 2008; Finley et al., 2014). This number increases by one when photogrammetric point clouds are considered (Breidenbach et al., 2016). Breidenbach. et al., (2008) and Breidenbach et al., (2016) only consider volume as dependent variable and Finley et al., (2014) analyzed stand tables (i.e. number of stems in predefined diameter classes). For volume there exist two references in the literature, however, the one that uses LiDAR (Breidenbach. et al., 2008) reported spatial correlations that descend below 0.05 for distances of 202 m of larger. The design used in that study, provide pairs of observations with a separation between plots of 100, 200 m 223 m and larger, in conclusion, only two of those distances were within the spatial correlation range, and reported results lack robustness. The second study on volume is more consistent but it does not consider LiDAR auxiliary information. While some relations exist between some of the variables considered here (i.e. QMD or G), and those considered in (Finley et al., 2014) they are not completely the same. In our study we include eight different variables and for seven of them (B_{tot}, B_{stem}, QMD, G, H_m, H_o, and N), no study to the date has reported correlation ranges of residuals from LiDAR models. Another novelty of the present study is the analysis of the effect of the plot size ion the spatial correlation. A consistent increase of the spatial correlation range was observed when increasing the plot radius. This results raises questions about the implications of using different plot sizes to derive estimates and measures of uncertainty for different subpopulations.

Excluding the 4 and 3 subplot radius discarded for G and N respectively, where the empirical correlation was flat at the end, for all the variables analyzed here the spatial correlation range of the residuals never surpassed 100 m and for most variables they were always below 60 meters (Figure 5 and Table A1). This makes clear the need of sampling designs with very close observations if the modeler aims at analyzing\using the spatial correlation of model residuals. A controversial point of our field sampling design is that it provides overlapping subplots. As mentioned in the introduction, overlap occurs in real applications and it should not be disregarded. In addition, except for 12.5 m radius subplots, the design used here provides distances without overlap. Even if the overlap was considered an issue, if the spatial correlation models fit well in the section without overlap, the overlap causes no harm in subsequent estimates.

Limitations from our sampling design did not allowed us to study spatial dependences for distances lager than 25 to 35 meters depending on the subplot radius. However, this constraint can be regarded as minor. Extrapolations would be needed for distances lager than the maximum distance studied here, however, the correlation of the residuals decreases when the distance increases, and we empirically confirmed that for the maximum distances studied here the empirical spatial correlation have almost disappeared (Table 2), being for most variables and subplot radius smaller than 0.1 at the maximum distance. Therefore, extrapolation errors for distances within the range will be bounded by a small quantity.

For most variables, the spatial correlation ranges of the residuals were so sort that the assumption of independent residuals seem to be reasonably accurate, at least for prediction. The effect of omitting the spatial correlation analysis in the prediction stage in a spatial correlation scenario like the one observed here can be illustrated with the following example. If we considered a systematic design with plots on the nodes of a rectangular grid, a hypothetical plot density of 0.1 plot/ha, a management unit of 50 ha and a pixel size of 15 m, the MU would contain approximately 2222 pixels and 5 plots. If the spatial correlation of the residuals vanishes at 40 m, plots could be assumed to be independent for model fitting purposes. Incorporating the spatial structure to improve the predictive performance of the models would have very little impact, compared to a model that assumes independence: it would only affect the prediction for about 35 pixels per measured plot (i.e., pixels that surround a plot and are closer than 40 m). In total, only predictions for around 175 pixels out of 2222 (which approximately represents 8% of the total number of pixels in the MU) would be different from those obtained omitting the spatial correlation.

However, the spatial correlation should be accounted for not only on the prediction stage, but also when computing uncertainty measures for predictions. The effect of omitting the spatial correlation on the computation of variances and mean squared errors of predictions would result in overoptimistic uncertainty measures. Breidenbach et al., (2016) analyzed the effect of such model misspecification in variance estimators for V using 9 m radius plots and found that the omission of the spatial correlation resulted in variance estimates that were 15% smaller than those obtained when accounting for the spatial correlation. Only volume and one plot radius were studied but a

15% underestimation for the variance is an important quantity and further studies should analyze the same effect for different plot sizes and for other forest attributes like the ones studied here.

Using our results as a reference for further forest inventories in similar areas, we can say that commonly used systematic sampling designs are inappropriate for modeling spatial correlation. Even with plots overlaid on the nodes of an extremely dense grid, where distance between nodes is 100 m (1 plot/ha), estimating the spatial correlation would be very difficult, or even impossible for all variables. Sample designs where plot locations are random could provide pairs of observations within the correlation ranges, but still, sampling efforts have to be very high due to the short correlation ranges observed for residuals of most variables. Based on our experience, the best alternatives for operational forest inventories aiming at analyzing this factor are the use of large plots with georeferenced tree positions, like the ones used in this study, or clusters of subplots as those used in some national inventories (e.g., the U.S. Forest Inventory and Analysis (FIA), (Bechtold and Patterson 2005)). In the latter case, it may be necessary to incorporate more subplots at a greater range of distances, as the actual FIA design, for example, only allow to consider two distances (36.58 m and 63.35 m) between subplots, which seems insufficient to model spatial correlation patterns. In any case, both designs would allow obtaining clusters of observations, which is the option recommended by Zimmerman, (2006) to optimize sample designs to account for residual spatial correlation when both fixed effects and spatial correlation parameters are unknown.

An advantage of the design employed here is that it provides a relatively operational way of obtaining data that can be used to directly model spatial correlation of residuals from LiDAR models. To obtain large plots with georeferenced trees it is necessary to obtain tree coordinates relative to the plot center as well as absolute coordinates of the plot center. Specially promising for the first task are photogrammetric point clouds obtained using inexpensive cameras providing centimetric accuracies for relative coordinates (Gatziolis et al. 2015). Unfortunately, these newer technologies or even highly accurate devices such as terrestrial laser scanners or total stations do not solve the problem of the absolute positioning of the field plot center.

8 Conclusions

In situations where highly correlated auxiliary information is available, the assumption of uncorrelated residuals that has been implicitly accepted in large number LiDAR assisted forest inventory applications seems to be reasonable accurate, and misspecification by omitting accounting for spatial correlations may not have a significant effect on model predictions. However, the effect of such misspecification on uncertainty measures needs to be studied.

Sampling designs able to provide clusters of plots separated by small distances are needed to study spatial correlation, as it tends to vanish at distances shorter than the minimum separation between plots employed in most LiDAR assisted inventories.

Spatial correlation ranges increased with the plot size.

Except for QMD, once the LiDAR information was included, spatial correlation ranges of the residual were smaller than the spatial correlation ranges for the raw variables. The reduction was greatest for variables highly correlated with LiDAR.

Acknowledgements

The authors wish to thank Dr. Jay Ver Hoef for his valuable comments on earlier versions of the manuscript. The US Bureau of Land Management and the Spanish Ministries of Industry, Tourism and Trade and Science and Innovation for the financial support provided in the framework of the projects: *Use of LIDAR and other remote sensing data with FIA plots for mapping forest inventory in Southwest Oregon, InForest II*, and *CGL2010-19591/BTE*.

9 References

Andersen, H.-E., Clarkin, T., Winterberger, K., and Strunk, J. 2009. An Accuracy Assessment of

Positions Obtained Using Survey- and Recreational-Grade Global Positioning System Receivers across a Range of Forest Conditions within the Tanana Valley of Interior Alaska. West. J. Appl. For. **24**(9): 128–136.

Andersen, H.-E., McGaughey, R.J., and Reutebuch, S.E. 2005. Estimating forest canopy fuel parameters using LIDAR data. Remote Sens. Environ. **94**(4): 441–449.

Bechtold, W.A., and Patterson, P.L. 2005. The enhanced forest inventory and analysis program- national sampling design and estimation procedures.

Breidenbach., J., Kublin, E., McGaughey, R., Andersen, H.E., and Reutebuch, S. 2008. Mixed-effects models for estimating stand volume by means of small footprint airborne laser scanner data. Photogramm. J. Finl. **21**(1): 4–15.

Breidenbach, J., McRoberts, R.E., and Astrup, R. 2016. Empirical coverage of model-based variance estimators for remote sensing assisted estimation of stand-level timber volume. Remote Sens. Environ. **173**: 274–281. doi:10.1016/j.rse.2015.07.026.

FAO. 2012. Forest Resources Assessment Working Paper 180. FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS., Rome.

Finley, A.O., Banerjee, S., Weiskittel, A.R., Babcock, C., and Cook, B.D. 2014. Dynamic spatial regression models for space-varying forest stand tables. Environmetrics **25**(8): 596–609. doi:10.1002/env.2322.

Gatziolis, D., Lienard, J.F., Vogs, A., and Strigul, N.S. 2015. 3D Tree Dimensionality Assessment Using Photogrammetry and Small Unmanned Aerial Vehicles. PLoS ONE **10**(9): e0137765. doi:10.1371/journal.pone.0137765.

Gobakken, T., and Næsset, E. 2009. Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. Can. J. For. Res. **39**(5): 1036–1052.

González-Ferreiro, E., Diéguez-Aranda, U., and Miranda, D. 2012. Estimation of stand variables in Pinus radiata D. Don plantations using different LiDAR pulse densities. Forestry.

Gunnarsson, F., Holm, S., Holmgren, P., and Thuresson, T. 1998. On the potential of Kriging for forest management planning. Scand. J. For. Res. **13**(1–4): 237–245. doi:10.1080/02827589809382981.

Isenburg, M. 2013. LAStools - efficient tools for LiDAR processing. http://lastools.org. Available from http://rapidlasso.com.

Lumley, T. 2009. leaps: regression subset selection. Available from http://CRAN.R-project.org/package=leaps.

Magnussen, S., Eggermont, P., and LaRiccia, V.N. 1999. Recovering Tree Heights from Airborne Laser Scanner Data. For. Sci. **45**: 407–422(16).

Magnussen, S., Frazer, G., and Penner, M. 2016a. Alternative mean-squared error estimators for synthetic estimators of domain means. J. Appl. Stat.: 1–24. doi:10.1080/02664763.2016.1142942.

Magnussen, S., Mandallaz, D., Lanz, A., Ginzler, C., Næsset, E., and Gobakken, T. 2016b. Scale effects in survey estimates of proportions and quantiles of per unit area attributes. For. Ecol. Manag. **364**: 122–129. doi:10.1016/j.foreco.2016.01.013.

Magnussen, S., McRoberts, R.E., and Tomppo, E.O. 2009. Model-based mean square error estimators for k-nearest neighbour predictions and applications using remotely sensed data for forest inventories. Remote Sens. Environ. **113**(3): 476–488.

Mauro, F., Molina, I., García-Abril, A., Valbuena, R., and Ayuga-Téllez, E. 2016. Remote sensing estimates and measures of uncertainty for forest variables at different aggregation levels. Environmetrics **27**(4): 225–238. doi:10.1002/env.2387.

Mc Gaughey, R.J. 2014. FUSION/LDV: Software for LIDAR Data Analysis and Visualization.

McRoberts, R.E. 2006. A model-based approach to estimating forest area. Remote Sens. Environ. **103**(1): 56–66. doi:10.1016/j.rse.2006.03.005.

McRoberts, R.E., Tomppo, E.O., Finley, A.O., and Heikkinen, J. 2007. Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. Remote Sens. Environ. **111**(4): 466–480. doi:10.1016/j.rse.2007.04.002.

Montero, G., Ruiz-Peinado, R., Muñoz, M., España. Ministerio de Educación y Ciencia, and Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria. 2005. Producción de biomasa y fijación de CO2 por los bosques españoles. Ministerio de Educación y Ciencia. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria. Available from https://books.google.com/books?id=j4o41tVIV40C.

Næsset, E. 1997a. Estimating timber volume of forest stands using airborne laser scanner data. Remote Sens. Environ. **61**(2): 246–253.

Næsset, E. 1997b. Determination of mean tree height of forest stands using airborne laser scanner data. ISPRS J. Photogramm. Remote Sens. **52**(2): 49–56.

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

Næsset, E., and Bjerknes, K.-O. 2001. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. Remote Sens. Environ. **78**(3): 328–340.

Næsset, E., and Økland, T. 2002. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. Remote Sens. Environ. **79**(1): 105–115.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and Team, R.C. 2015. nlme: Linear and Nonlinear Mixed Effects Models.

Pinheiro, J.C., and Bates, D.M. 2000. Mixed-Effects Models in S and S-Plus. Springer. Berlin.

Ruiz, L.A., Hermosilla, T., Mauro, F., and Godino, M. 2014. Analysis of the Influence of Plot Size and LiDAR Density on Forest Structure Attribute Estimates. Forests **5**(5): 936–951. doi:10.3390/f5050936.

Temesgen, H., and Ver Hoef, J.M. 2014. Evaluation of the spatial linear model, random forest and gradient nearest-neighbour methods for imputing potential productivity and biomass of the Pacific Northwest forests. Forestry. doi:10.1093/forestry/cpu036.

Ver Hoef, J.M., and Temesgen, H. 2013. A Comparison of the Spatial Linear Model to Nearest Neighbor (k-NN) Methods for Forestry Applications. PLoS ONE **8**(3): e59129. doi:10.1371/journal.pone.0059129.

Woods, M., Pitt, D., Penner, M., Lim, K., Nesbitt, D., Etheridge, D., and Treitz, P. 2011. Operational implementation of a LiDAR inventory in Boreal Ontario. For. Chron. **87**(4): 512–528. doi:10.5558/tfc2011-050.

Zimmerman, D.L. 2006. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics **17**(6): 635–652. doi:10.1002/env.769.

Tables

Table 1Number of subplots, maximum distances, and number of pairs of observations at 0.5m and at max_distance used to compute spatial correlation parameters.

Subplot radius	Subplots/25 m plot	Total subplots	Max distance (m)	Pairs at 0.5 m	Pairs at max distance
7.5	284	24140	35	23800	
8	276	23460	34	23120	
8.5	268	22780	33	22440	
9	260	22100	32	21760	
9.5	252	21420	31	21080	
10	244	20740	30	20400	340
10.5	236	20060	29	19720	
11	228	19380	28	19040	
11.5	220	18700	27	18360	
12	212	18020	26	17680	
12.5	204	17340	25	17000	

Table 2. Empirical correlations observed for each variable and subplot radius at the máximum distance. $.max\_distance(radius) = 50 - 2radius$

Radius(m)	V	B_{tot}	B_{stem}	G	H_m	H_o	QMD	N
	Variable							
7.5	0.04	0.07	0.07	0.15	-0.03	0.02	0.12	0.26
8	-0.02	0.03	0.04	0.03	-0.03	0.02	0.10	0.26
8.5	0.03	0.10	0.08	0.19	-0.03	0.02	0.11	0.26
9	0.02	0.11	0.08	0.19	0.00	0.00	0.12	0.26
9.5	0.02	0.11	0.08	0.19	0.03	0.02	0.13	0.26
10	0.01	0.11	0.08	0.19	0.06	0.02	0.13	0.25
10.5	0.00	0.12	0.08	0.20	0.09	0.04	0.12	0.24
11	0.01	0.12	0.09	0.21	0.10	0.03	0.11	0.25
11.5	0.02	0.12	0.09	0.06	0.13	0.04	0.10	0.24
12	0.03	0.05	0.07	0.06	0.14	0.05	0.10	0.24
12.5	0.06	0.03	0.04	0.07	0.16	0.05	0.10	0.24

Figure captions

Figure 1. Field plot and grid of pixels. Note the overlap between plot and the four pixels surrounding it.

Figure 2. Manual correction of field plot positions. Tree locations were corrected by translating and rotating around the plot center, all trees as a block. For certain trees easy to identify, coordinates were changed to match the DCHM and the stem location identified on the ground point cloud after the first correction (rotation and translation).

Figure 3. Example of 25 m radius plot and subplots of radius 7.5 m, 10 m and 12.5 m (for clarity, other subplots radii are omitted) moving in an East to West direction. South East to North West, South to North and South West to North East directions in which field plots were moved in 0.5 m steps are marked with dashed lines. Trees are plotted according to their crown radius

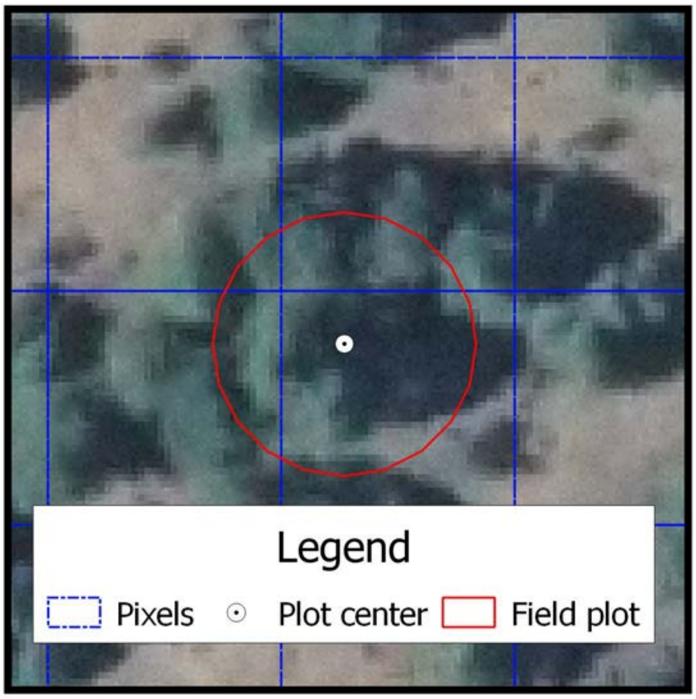
Figure 4. Spatial correlation models for the residuals from $m^{**}_{vrbl,rad}$ and for the raw variables.

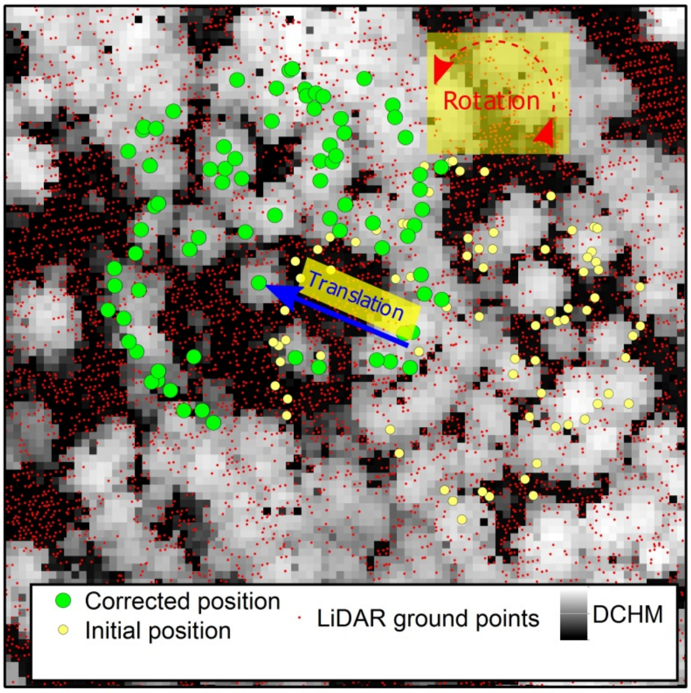
Figure 5. Distances (φ) for which correlation between pairs of observations decreases to 0.05 and parameters $\gamma_{vrbl,rad}^{raw,res}$.

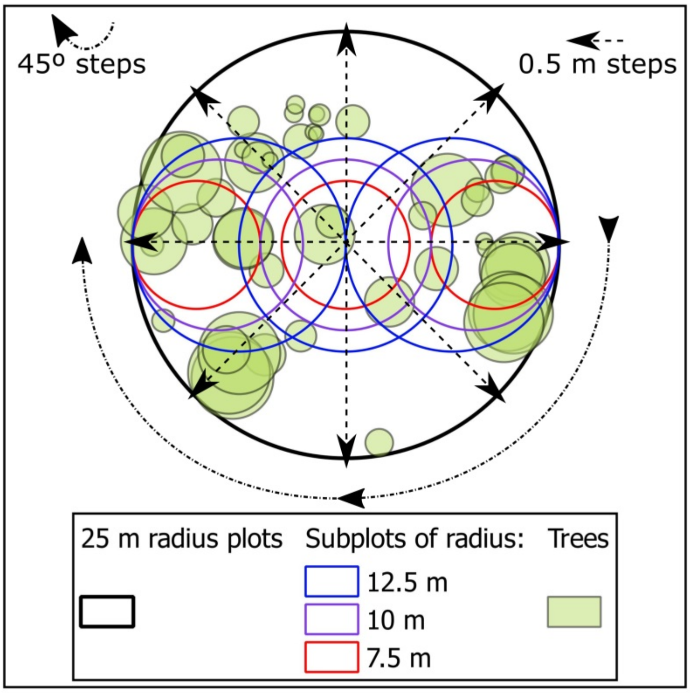
Figure 6. Results for the Kendall's τ significance test for each variable and subplot radius. Only non-overlapping plots are considered. Subfigures a,b,c and d are examples included as a graphical legend for the figure in the upper panel.

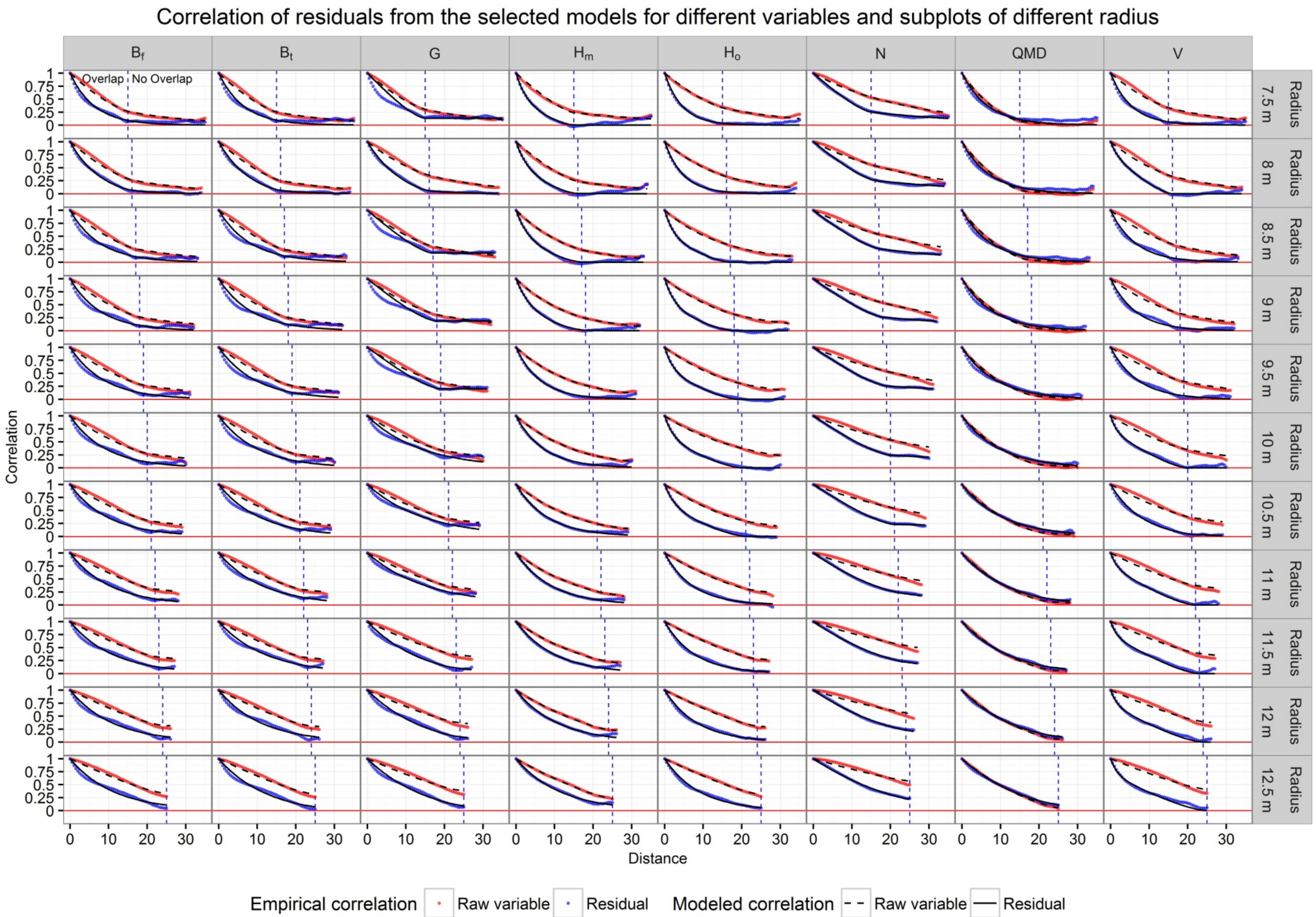
Figure A1. Empirical correlation functions observed for H_m and for its residuals in different directions and for different subplot radii. Fifth column shows the empirical correlation computed assuming isotropy.

Figures



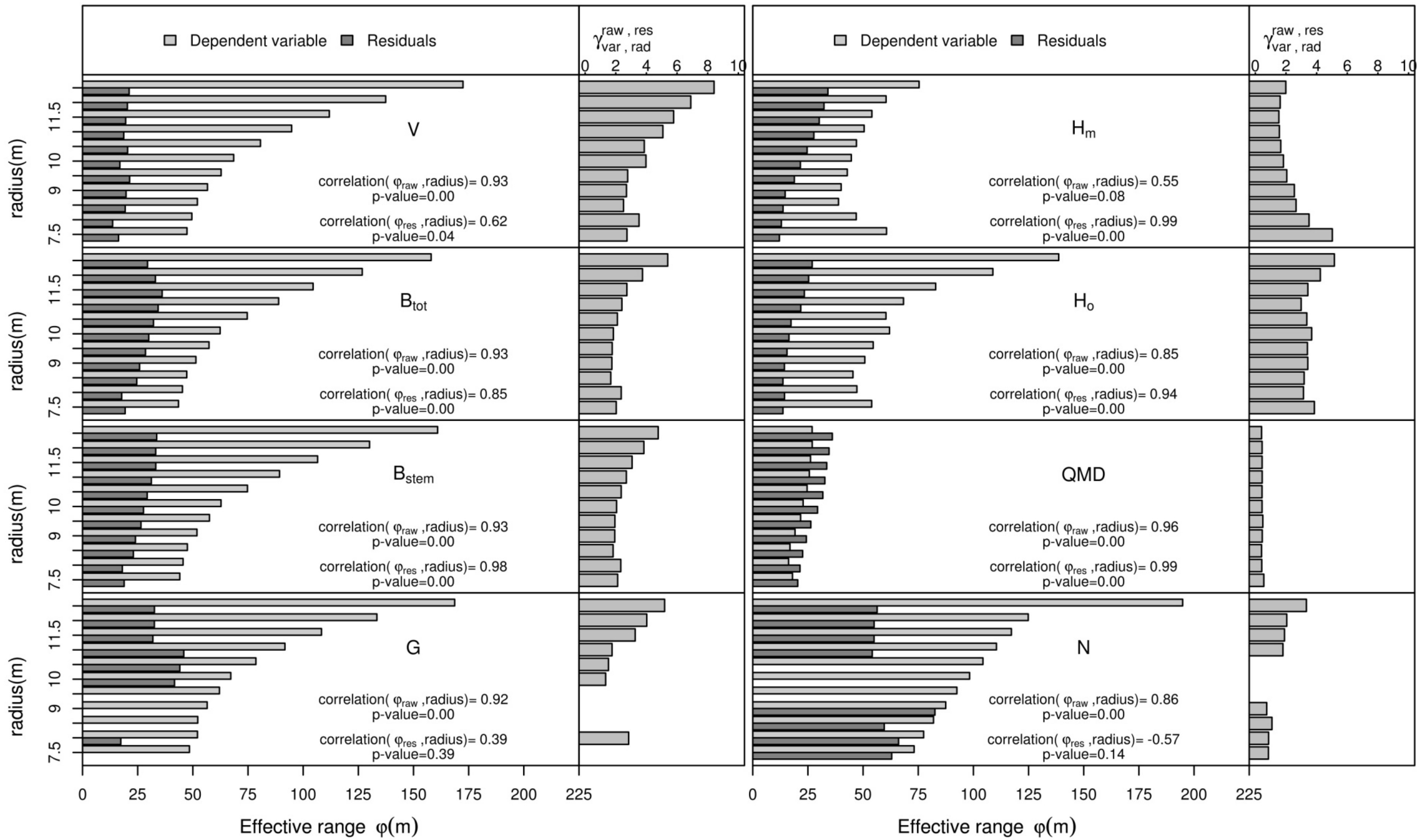

Rotation
Translation
Corrected position
Initial position
LiDAR ground points
DCHM


45º steps
0.5 m steps
25 m radius plots
Subplots of radius:
Trees
12.5 m
10 m
7.5 m


Correlation of residuals from the selected models for different variables and subplots of different radius
Bf
Bt
G
Hm
Ho
N
QMD
V
Overlap
No Overlap
Radius 7.5 m
Radius 8 m
Radius 8.5 m
Radius 9 m
Radius 9.5 m
Radius 10 m
Radius 10.5 m
Radius 11 m
Radius 11.5 m
Radius 12 m
Radius 12.5 m
Correlation
Distance
Empirical correlation
Raw variable
Residual
Modeled correlation
Raw variable
Residual

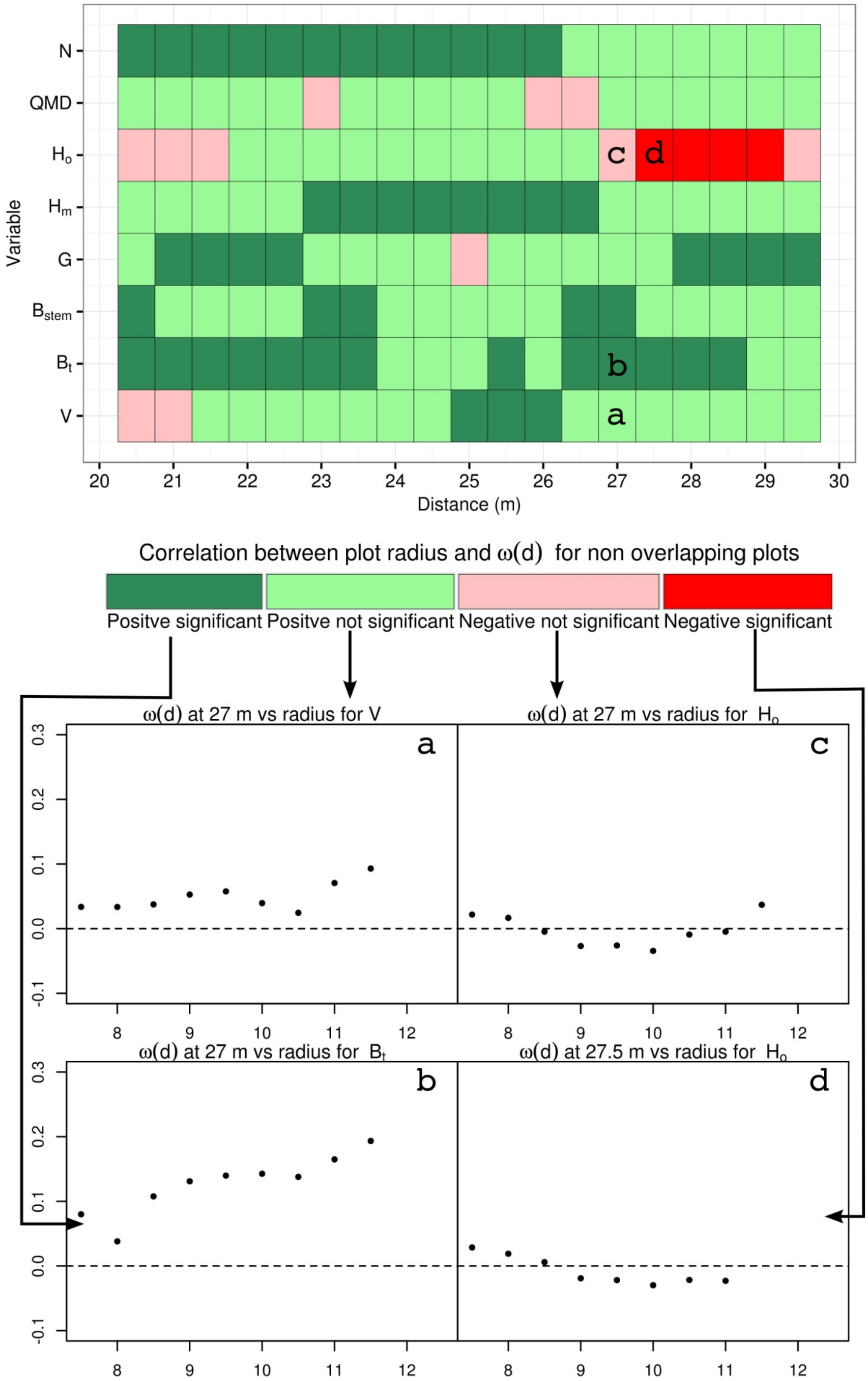
803

37



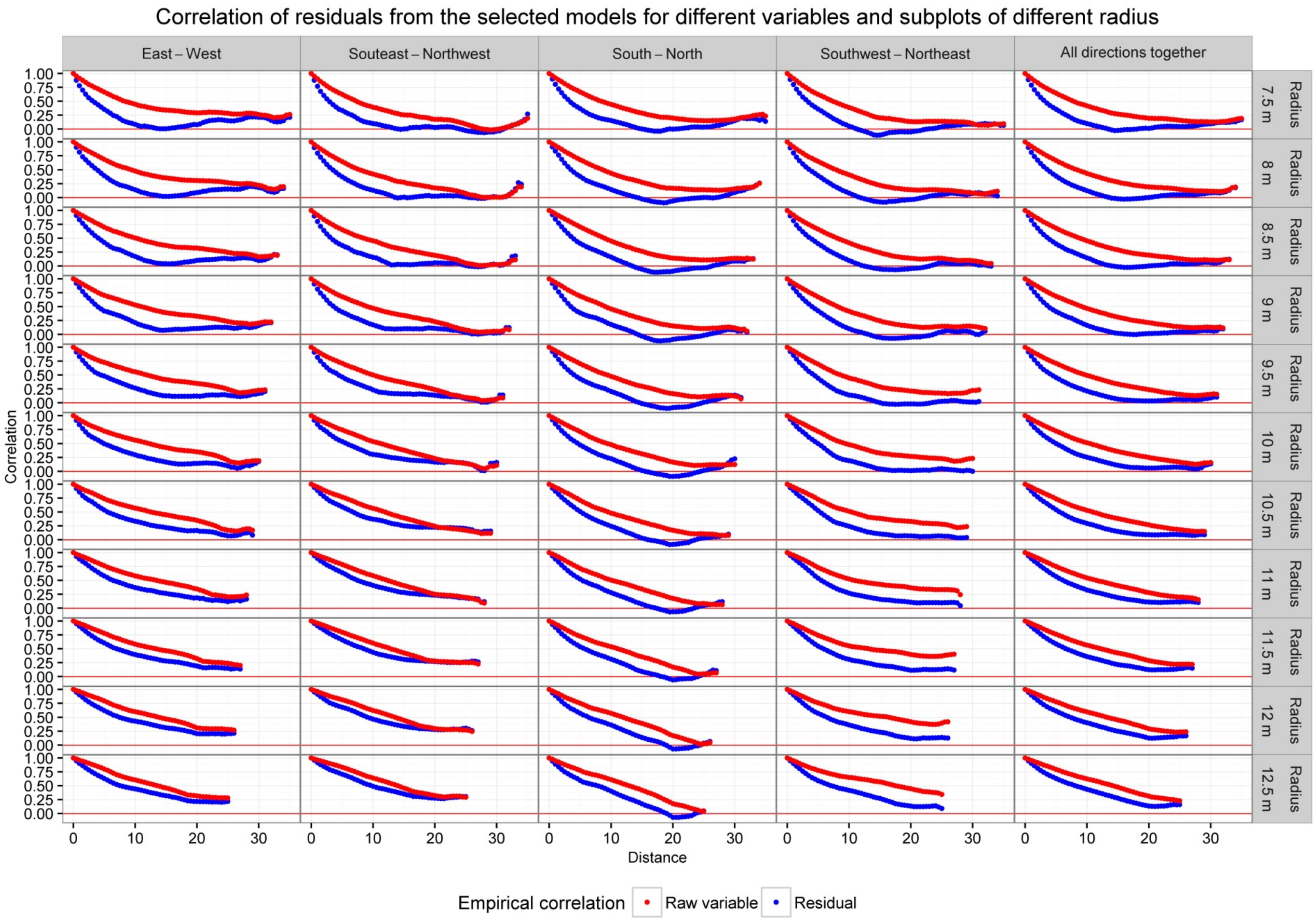
Dependent variable
Residuals
$\gamma^{raw,\,res}_{var,\,rad}$
V
correlation(φraw ,radius)= 0.93
p-value=0.00
correlation(φres ,radius)= 0.62
p-value=0.04
Btot
correlation(φraw ,radius)= 0.93
p-value=0.00
correlation(φres ,radius)= 0.85
p-value=0.00
Bstem
correlation(φraw ,radius)= 0.93
p-value=0.00
correlation(φres ,radius)= 0.98
p-value=0.00
G
correlation(φraw ,radius)= 0.92
p-value=0.00
correlation(φres ,radius)= 0.39
p-value=0.39
Hm
correlation(φraw ,radius)= 0.55
p-value=0.08
correlation(φres ,radius)= 0.99
p-value=0.00
Ho
correlation(φraw ,radius)= 0.85
p-value=0.00
correlation(φres ,radius)= 0.94
p-value=0.00
QMD
correlation(φraw ,radius)= 0.96
p-value=0.00
correlation(φres ,radius)= 0.99
p-value=0.00
N
correlation(φraw ,radius)= 0.86
p-value=0.00
correlation(φres ,radius)= -0.57
p-value=0.14
radius(m)
Effective range φ(m)

804



N
QMD
Ho
Hm
G
Bstem
Bt
V
Variable
c
d
b
a
20
21
22
23
24
25
26
27
28
29
30
Distance (m)
Correlation between plot radius and ω(d) for non overlapping plots
Positve significant
Positve not significant
Negative not significant
Negative significant
ω(d) at 27 m vs radius for V
ω(d) at 27 m vs radius for Ho
ω(d) at 27 m vs radius for Bt
ω(d) at 27.5 m vs radius for Ho

805


Correlation of residuals from the selected models for different variables and subplots of different radius
East – West
Souteast – Northwest
South – North
Southwest – Northeast
All directions together
Radius 7.5 m
Radius 8 m
Radius 8.5 m
Radius 9 m
Radius 9.5 m
Radius 10 m
Radius 10.5 m
Radius 11 m
Radius 11.5 m
Radius 12 m
Radius 12.5 m
Correlation
1.00
0.75
0.50
0.25
0.00
Distance
0
10
20
30
Empirical correlation
Raw variable
Residual

806

40

Appendix

Table A1. Estimated model parameters for each variable of interest and sub-plots radius. β_0 denotes the intercept and β_1, β_2 and β_3 are the regression parameters associated to x_1, x_2 and x_3 in respectively. The most correlated predictor (mcp) is. x_1. The modeled standard deviation of residual was equal to $\sigma_e mcp^{\eta}$. The parameter σ_v is the standard deviation of the management unit random effects. $\varphi^{res}_{vrbl,rad}$ and $\varphi^{dep}_{vrbl,rad}$ represent the effective spatial correlation range of residuals and raw variables and $\gamma^{raw,res}_{vrbl,rad}$ denotes the ratio $\frac{\varphi^{raw}_{vrbl,rad}}{\varphi^{res}_{vrbl,rad}}$..

Variable	rad	β_0	β_1	β_2	β_3	σ_e	η	σ_v	Residuals			Raw Variables			$\gamma^{raw,res}_{vrbl,rad}$
Aux variables									ρ^{res}	θ^{res}	φ^{res}	ρ^{raw}	θ^{raw}	φ^{raw}	
V(m^3/ha) x1=Elv_mean x2=Elv_P01 x3=Elv_P40	7.5	0.30	24.61	0.59	6.50	14.68	0.78	9.01E-02	5.44	0.00	16.32	16.93	0.18	47.34	2.80
	8.0	1.49	28.77	46.33	7.13	33.51	--	1.19E-01	1.84	0.75	13.65	17.60	0.16	49.55	2.82
	8.5	1.34	24.31	0.18	8.57	12.68	0.82	3.40E-01	6.45	0.00	19.34	18.60	0.18	52.05	2.80
	9.0	1.21	24.66	0.52	8.74	12.05	0.83	3.72E-01	6.58	0.00	19.74	20.29	0.18	56.63	2.79
	9.5	0.99	25.09	0.83	8.34	11.59	0.81	3.57E-01	7.12	0.00	21.35	22.62	0.20	62.81	2.78
	10.0	0.76	25.49	2.00	8.05	11.12	0.81	3.73E-01	1.47	0.71	16.93	24.86	0.21	68.51	2.76
	10.5	0.55	25.84	2.90	7.81	10.71	0.81	4.42E-01	7.22	0.17	20.44	29.90	0.26	80.64	2.70
	11.0	0.82	26.05	4.85	7.68	10.22	0.81	6.08E-01	1.69	0.73	18.69	36.00	0.30	94.77	2.63
	11.5	0.88	26.30	5.85	7.59	9.77	0.81	7.67E-01	1.50	0.73	19.55	43.40	0.34	111.85	2.58
	12.0	1.46	26.51	10.42	7.44	9.26	0.83	8.67E-01	1.67	0.72	20.34	54.81	0.39	137.40	2.51
	12.5	1.94	26.97	17.11	7.20	8.98	0.80	9.18E-01	1.68	0.70	21.13	70.25	0.42	172.44	2.45
B_{tot} (kg/ha) x1=Elv_mean x2=CRR	7.5	-482.86	20498.27	292.61		12217.26	0.78	4.56E-02	6.44	0.00	19.32	17.47	0.40	43.49	2.49
	8.0	-7556.84	16271.57	102258.52		24586.30	--	1.94E-01	6.51	0.24	17.75	17.98	0.38	45.29	2.52
	8.5	-392.22	20748.02	677.19		11224.13	0.76	9.41E-02	8.19	0.00	24.58	18.89	0.39	47.24	2.50
	9.0	-602.99	20933.98	708.84		10830.21	0.74	1.08E-01	8.62	0.00	25.86	20.51	0.39	51.39	2.51
	9.5	-661.42	21019.10	666.94		10439.33	0.73	1.39E-01	9.51	0.00	28.54	22.99	0.39	57.33	2.49

Variable Aux variables	rad	β_0	β_1	β_2	β_3	σ_e	η	σ_v	Residuals			Raw Variables			$\gamma_{vrbl,rad}^{raw,res}$
									ρ^{res}	θ^{res}	φ^{res}	ρ^{raw}	θ^{raw}	φ^{raw}	
	10.0	-789.75	21073.42	1046.41		10064.25	0.73	1.87E-01	10.01	0.00	30.02	25.11	0.40	62.37	2.48
	10.5	-740.51	21074.72	1409.97		9746.68	0.72	2.52E-01	10.72	0.00	32.16	30.72	0.43	74.65	2.43
	11.0	-511.05	21011.64	1929.48		9433.28	0.71	3.53E-01	11.42	0.00	34.27	37.62	0.47	88.88	2.36
	11.5	-246.73	20868.41	3249.89		9154.14	0.70	4.57E-01	12.03	0.00	36.09	45.02	0.49	104.49	2.32
	12.0	-1888.65	18434.29	47567.08		9322.38	0.60	5.95E-01	11.01	0.00	33.03	56.31	0.52	126.78	2.25
	12.5	-2477.69	17912.21	59069.51		9168.77	0.56	7.31E-01	10.56	0.19	29.46	72.00	0.55	158.00	2.19
	7.5	-183.97	13322.18	181.63		8124.30	0.79	5.64E-02	6.29	0.00	18.87	17.32	0.36	44.08	2.55
	8.0	-6244.16	11504.19	67144.69		16910.59	--	1.99E-01	6.58	0.23	18.04	17.67	0.34	45.58	2.58
	8.5	76.33	13492.76	440.80		7419.93	0.77	1.50E-01	7.68	0.00	23.05	18.54	0.35	47.52	2.56
	9.0	-69.16	13671.49	450.32		7144.41	0.76	1.62E-01	8.00	0.00	23.99	20.23	0.35	51.87	2.56
B_{stem} (kg/ha)	9.5	-145.27	13779.77	420.10		6868.77	0.75	1.80E-01	8.83	0.00	26.50	22.48	0.35	57.52	2.56
x1=Elv_mean	10.0	-275.36	13869.16	610.48		6590.74	0.75	2.18E-01	9.21	0.00	27.63	24.73	0.37	62.77	2.54
x2=CRR	10.5	-288.86	13926.97	699.01		6370.98	0.74	2.72E-01	9.77	0.00	29.30	30.20	0.41	74.75	2.48
	11.0	-181.39	13943.14	840.66		6142.53	0.74	3.80E-01	10.40	0.00	31.19	37.15	0.45	89.29	2.40
	11.5	-109.95	13952.33	1203.12		5912.13	0.74	4.95E-01	11.07	0.00	33.20	45.23	0.47	106.54	2.36
	12.0	-816.20	13209.10	16580.32		5827.23	0.70	6.16E-01	11.06	0.00	33.17	56.83	0.51	130.08	2.29
	12.5	-1258.10	12903.82	24075.74		5756.19	0.66	7.33E-01	11.21	0.00	33.62	72.04	0.53	160.93	2.23
	7.5	11.36	-0.10	2.46	-0.95	7.60	0.04	5.93E-02	6.80	0.00	20.39	7.53	0.45	18.01	2.39
	8.0	14.45	-0.13	2.70	-1.30	8.03	-0.01	6.47E-02	7.14	0.00	21.41	7.37	0.55	16.24	2.20
	8.5	16.76	-0.14	2.79	-1.48	8.39	-0.05	6.68E-02	7.54	0.00	22.62	7.61	0.54	16.87	2.22
QMD (cm)	9.0	17.97	-0.15	2.69	-1.45	8.53	-0.08	6.55E-02	8.10	0.00	24.29	8.08	0.47	19.15	2.37
x1= Elv_P95	9.5	18.56	-0.16	2.58	-1.36	8.38	-0.08	6.68E-02	8.77	0.00	26.32	8.86	0.42	21.69	2.45
x2= Rt_Abvmean	10.0	19.54	-0.17	2.47	-1.30	8.48	-0.11	6.60E-02	9.79	0.00	29.37	9.64	0.47	22.82	2.37
x3=Elv_P99	10.5	20.25	-0.18	2.34	-1.22	8.66	-0.13	6.34E-02	10.59	0.00	31.77	10.36	0.46	24.63	2.38
	11.0	21.04	-0.18	2.28	-1.22	8.79	-0.16	6.24E-02	11.31	0.10	32.69	10.99	0.48	25.64	2.33
	11.5	21.48	-0.18	2.18	-1.16	8.86	-0.18	6.31E-02	11.59	0.10	33.54	11.02	0.46	26.18	2.38
	12.0	21.75	-0.18	2.07	-1.09	8.70	-0.18	6.62E-02	12.26	0.16	34.58	11.66	0.50	26.96	2.31

Canadian Journal of Forest Research

Variable Aux variables	rad	β_0	β_1	β_2	β_3	σ_e	η	σ_v	Residuals ρ^{res}	Residuals θ^{res}	Residuals φ^{res}	Raw Variables ρ^{raw}	Raw Variables θ^{raw}	Raw Variables φ^{raw}	$\gamma_{vrbl,rad}^{raw,res}$
	12.5	22.04	-0.19	2.01	-1.05	8.62	-0.20	6.69E-02	13.38	0.26	36.06	12.91	0.60	26.92	2.09
G (m2/ha) x1=Elv_mean x2=CRR	7.5	-0.15	4.84	0.08		2.82	0.74	5.65E-06	1E05	0.87	1E05	18.46	0.31	48.45	2.62
	8.0	-2.03	3.11	32.49		5.16	--	1.46E-05	6.73	0.34	17.35	18.39	0.15	52.09	2.83
	8.5	-0.15	4.86	0.15		2.55	0.73	1.09E-05	1E05	0.82	1E05	19.36	0.26	52.23	2.70
	9.0	-0.18	4.86	0.17		2.45	0.72	1.31E-05	1E05	0.81	1E05	20.90	0.25	56.49	2.70
	9.5	-0.19	4.85	0.18		2.36	0.71	1.64E-05	1E05	0.80	1E05	22.98	0.26	62.05	2.70
	10.0	-0.22	4.84	0.31		2.29	0.70	2.07E-05	13.90	0.00	41.69	24.95	0.26	67.20	2.69
	10.5	-0.22	4.81	0.51		2.24	0.69	2.55E-05	14.70	0.00	44.11	29.75	0.30	78.58	2.64
	11.0	-0.18	4.78	0.80		2.18	0.67	3.17E-05	15.30	0.00	45.89	35.63	0.34	91.73	2.57
	11.5	-1.03	3.45	24.06		2.59	0.34	3.77E-05	11.64	0.23	31.89	42.95	0.38	108.30	2.52
	12.0	-1.21	3.37	26.54		2.23	0.45	4.93E-05	12.11	0.26	32.56	54.59	0.42	133.41	2.44
	12.5	-1.26	3.35	27.36		2.13	0.44	5.86E-05	12.24	0.28	32.59	70.87	0.46	168.67	2.38
H_m (m) x1=Elv_AAD x2=Elv_P75	7.5	2.78	2.33	-0.26		2.71	-0.04	1.54E-01	2.87	0.41	12.07	29.05	0.60	60.63	2.09
	8.0	3.11	2.27	-0.27		2.64	-0.07	1.54E-01	3.22	0.41	12.97	19.90	0.47	46.91	2.36
	8.5	3.42	2.19	-0.26		2.58	-0.10	1.53E-01	3.37	0.41	13.73	13.90	0.18	38.86	2.80
	9.0	3.69	2.09	-0.25		2.53	-0.13	1.59E-01	4.27	0.30	14.68	13.96	0.12	40.06	2.87
	9.5	3.88	2.03	-0.24		2.48	-0.16	1.63E-01	6.29	0.00	18.86	14.80	0.10	42.84	2.89
	10.0	4.05	1.99	-0.24		2.43	-0.17	1.63E-01	7.20	0.00	21.59	15.57	0.12	44.66	2.87
	10.5	4.13	1.96	-0.24		2.37	-0.19	1.66E-01	8.21	0.00	24.64	16.43	0.13	47.00	2.86
	11.0	4.19	1.94	-0.24		2.31	-0.19	1.76E-01	9.23	0.00	27.70	17.94	0.17	50.48	2.81
	11.5	4.23	1.94	-0.25		2.24	-0.20	1.89E-01	10.04	0.00	30.12	19.48	0.20	53.99	2.77
	12.0	4.29	1.91	-0.24		2.16	-0.20	2.07E-01	10.76	0.00	32.29	23.26	0.33	60.43	2.60
	12.5	4.32	1.90	-0.24		2.08	-0.19	2.25E-01	11.36	0.00	34.07	32.80	0.50	75.39	2.30
H_o (m) x1=Elv_AAD	7.5	3.41	2.28			2.83	-0.04	1.87E-01	4.56	0.00	13.69	20.46	0.30	53.90	2.63
	8.0	3.95	2.21			2.77	-0.08	2.02E-01	4.80	0.00	14.39	16.54	0.13	47.20	2.85
	8.5	4.40	2.16			2.72	-0.11	2.19E-01	2.38	0.48	13.68	15.14	0.00	45.42	3.00
	9.0	4.34	2.09			2.56	-0.14	2.09E-01	2.77	0.44	14.35	17.73	0.12	50.79	2.86

Variable Aux variables	rad	β_0	β_1	β_2	β_3	σ_e	η	σ_v	Residuals ρ^{res}	Residuals θ^{res}	Residuals φ^{res}	Raw Variables ρ^{raw}	Raw Variables θ^{raw}	Raw Variables φ^{raw}	$\gamma_{vrbl,rad}^{raw,res}$
	9.5	4.68	2.07			2.53	-0.16	2.36E-01	2.24	0.54	15.52	19.25	0.15	54.59	2.84
	10.0	5.00	2.04			2.48	-0.18	2.54E-01	2.34	0.56	16.42	20.66	0.00	61.99	3.00
	10.5	5.27	2.02			2.44	-0.18	2.80E-01	2.59	0.58	17.34	22.22	0.24	60.38	2.72
	11.0	5.16	1.99			2.33	-0.20	2.81E-01	7.26	0.00	21.77	25.92	0.30	68.32	2.64
	11.5	5.38	1.98			2.28	-0.20	3.35E-01	7.80	0.00	23.39	33.52	0.41	82.91	2.47
	12.0	5.25	1.97			2.17	-0.21	3.41E-01	8.44	0.00	25.32	46.48	0.48	108.83	2.34
	12.5	5.45	1.96			2.11	-0.21	3.98E-01	8.98	0.00	26.94	61.90	0.53	138.67	2.24
N (stems/ha) x1=PercR1_Abvmea x2=PercRt_Abvmea x3=Elv_AAD	7.5	-1.14	34.10	-26.24	-10.82	7.13	0.94	1.61E+01	29.60	0.58	62.98	24.38	0.00	73.13	3.00
	8.0	-3.84	33.72	-25.44	-12.35	6.11	0.97	2.22E+01	30.73	0.57	66.09	25.82	0.00	77.45	3.00
	8.5	-6.50	33.46	-24.89	-12.92	5.34	1.00	3.18E+01	25.70	0.49	59.60	27.31	0.00	81.93	3.00
	9.0	-9.85	33.29	-24.55	-12.31	4.88	1.02	3.98E+01	38.90	0.58	82.57	29.15	0.00	87.46	3.00
	9.5	-13.13	33.02	-24.03	-12.64	4.70	1.02	4.43E+01	53.15	0.63	105.99	30.84	0.00	92.52	3.00
	10.0	-15.02	32.72	-23.42	-14.01	4.42	1.03	5.25E+01	63.42	0.67	120.40	32.76	0.00	98.28	3.00
	10.5	-15.88	32.51	-22.94	-15.12	4.16	1.04	6.12E+01	145.03	0.73	244.45	34.78	0.00	104.33	3.00
	11.0	-15.54	32.48	-22.70	-16.44	3.90	1.05	7.10E+01	20.50	0.30	54.16	36.82	0.00	110.45	3.00
	11.5	-15.89	32.45	-22.54	-16.96	3.63	1.06	8.17E+01	20.80	0.30	54.97	39.07	0.00	117.21	3.00
	12.0	-16.03	32.33	-22.25	-17.87	3.40	1.08	9.03E+01	20.25	0.24	55.00	41.62	0.00	124.86	3.00
	12.5	-16.40	32.17	-21.89	-18.88	3.17	1.09	1.01E+02	21.25	0.29	56.38	69.89	0.19	194.84	2.79

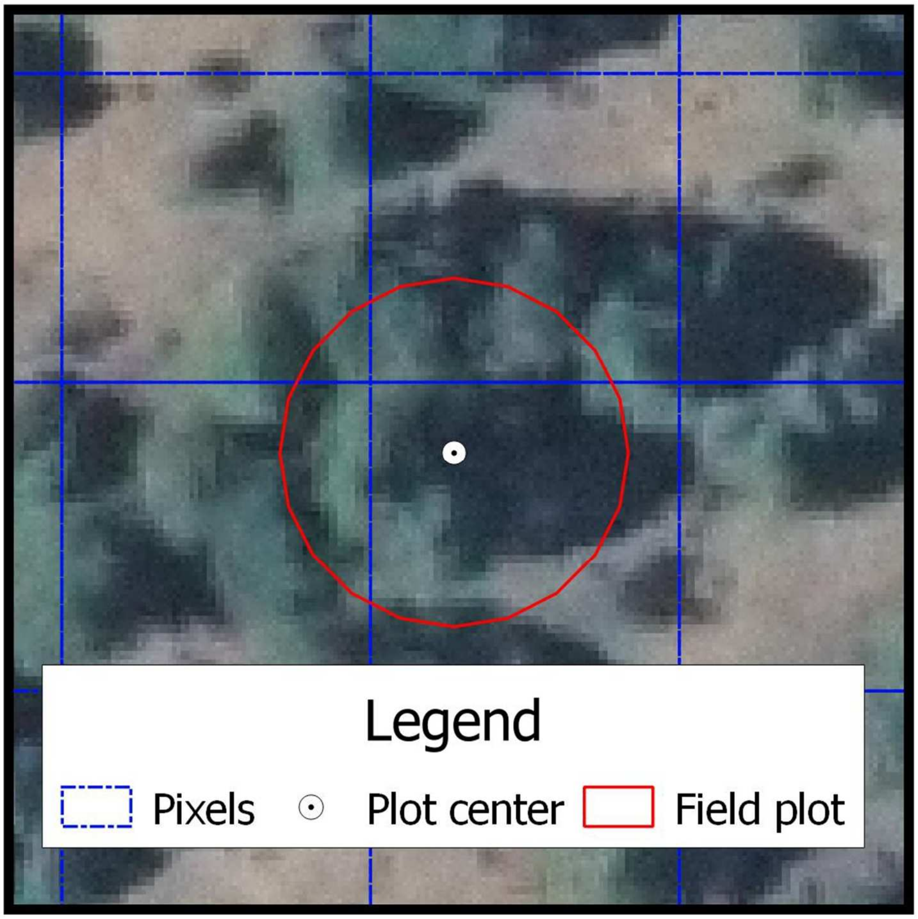


Figure 1. Field plot and grid of pixels. Note the overlap between plot and the four pixels surrounding it.
Figure 1.
99x99mm (300 x 300 DPI)

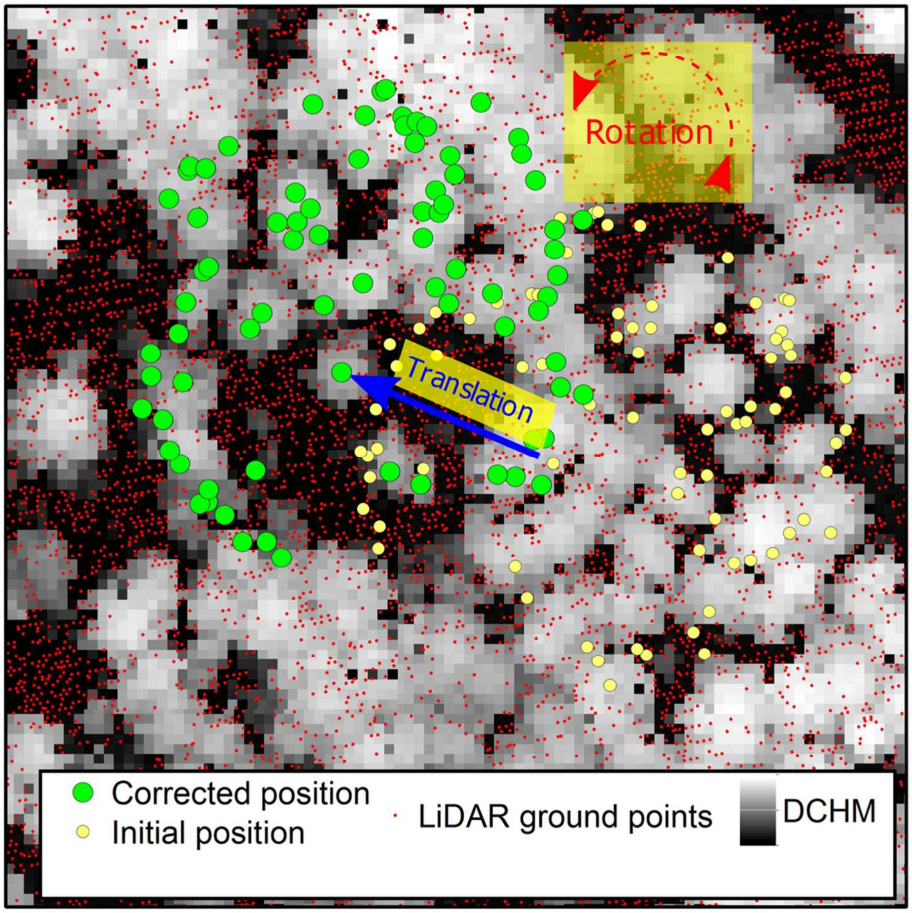


Figure 2. Manual correction of field plot positions. Tree locations were corrected by translating and rotating around the plot center, all trees as a block. For certain trees easy to identify, coordinates were changed to match the DCHM and the stem location identified on the ground point cloud after the first correction (rotation and translation).
Figure 2.
99x99mm (300 x 300 DPI)

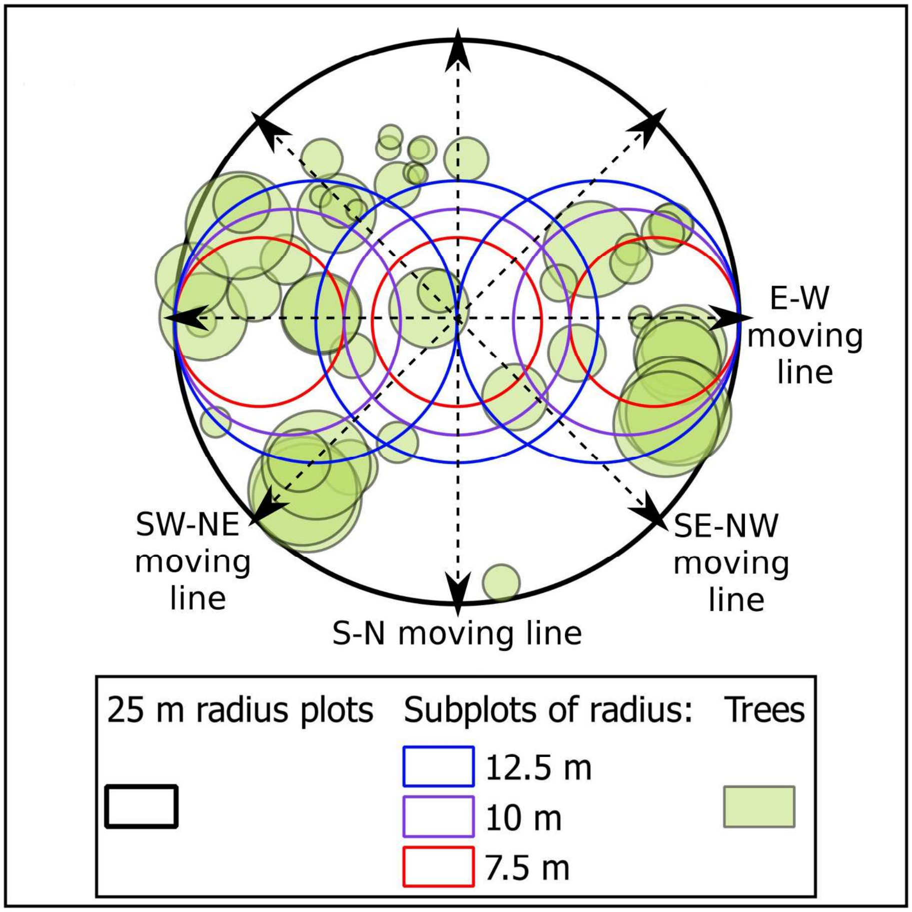


Figure 3. Example of 25 m radius plot and subplots of radius 7.5 m, 10 m and 12.5 m (for clarity, other subplots radii are omitted) moving in an East to West direction. South East to North West, South to North and South West to North East directions in which field plots were moved in 0.5 m steps are marked with dashed lines. Trees are plotted according to their crown radius
Figure 3.
100x100mm (600 x 600 DPI)

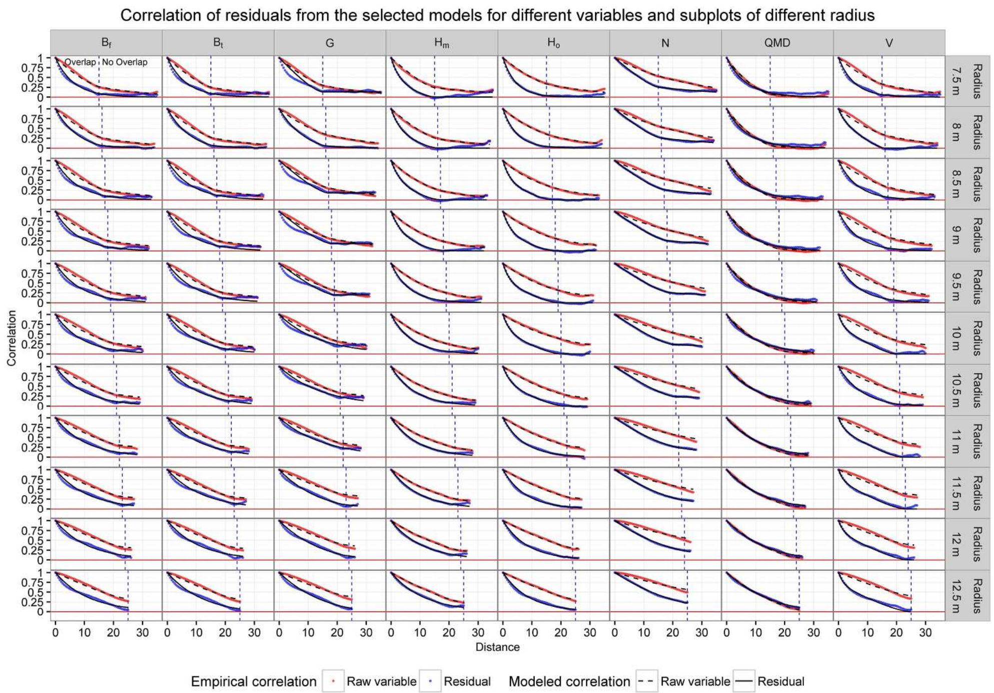


Figure 4. Spatial correlation models for the residuals from $m^{**}_{varbl,rad}$ and for the raw variables.
Figure 4.
199x142mm (300 x 300 DPI)

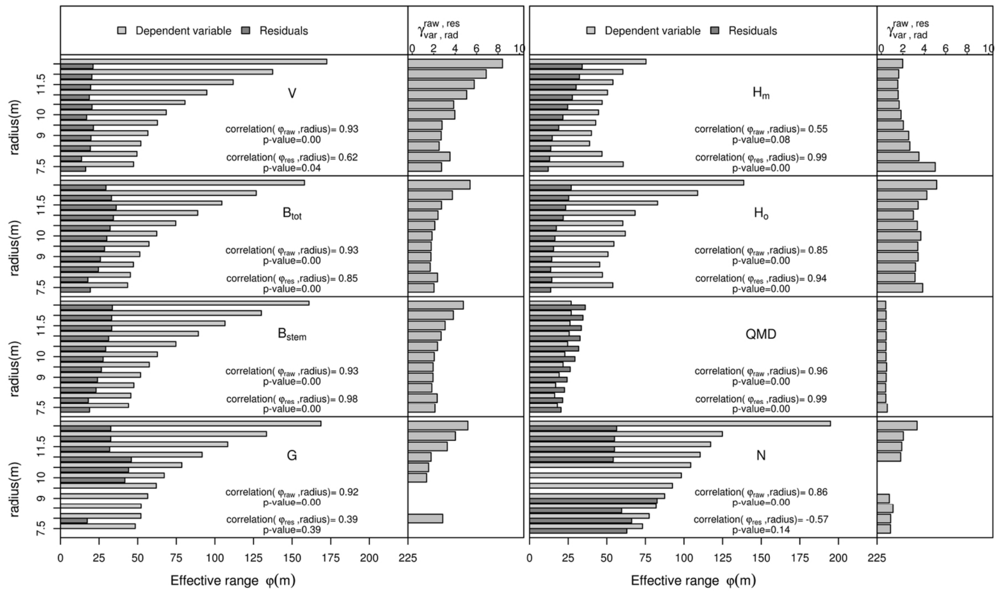


Figure 5. Distances (φ) for which correlation between pairs of observations decreases to 0.05 and parameters $\gamma\_^{raw,res}_{vrbl,rad}$
Figure 5.
152x91mm (300 x 300 DPI)

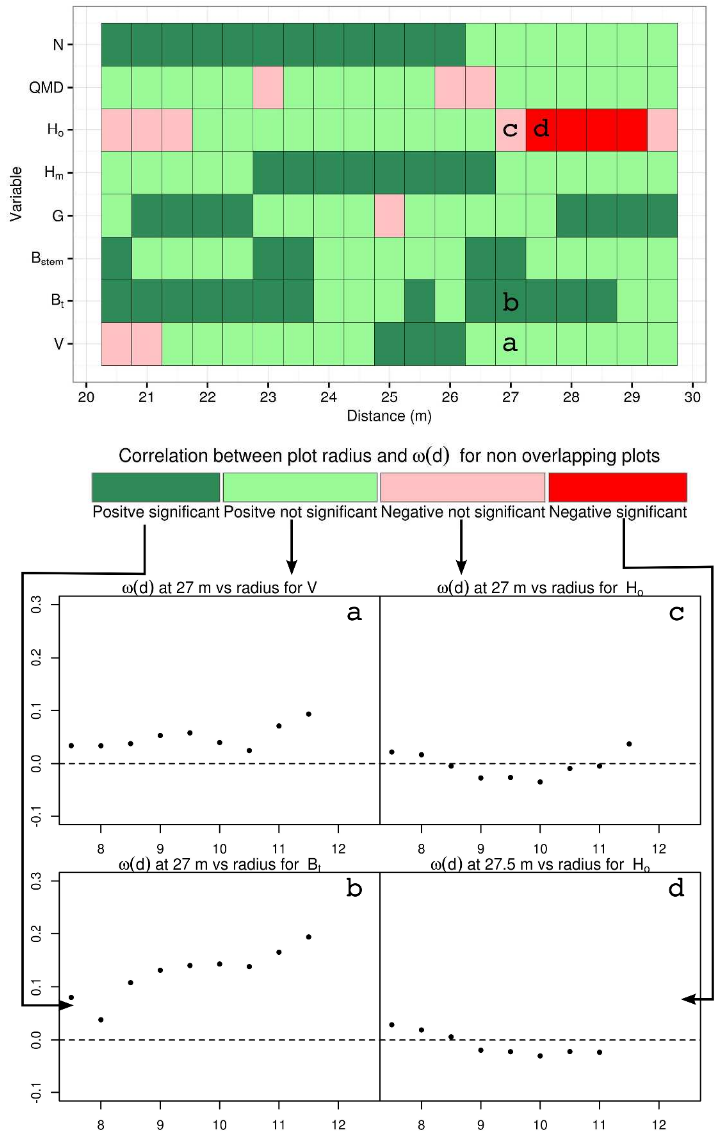


Figure 6. Results for the Kendall's τ significance test for each variable and subplot radius. Only non-overlapping plots are considered. Subfigures a,b,c and d are examples included as a graphical legend for the figure in the upper panel.
Figure 6.
1027x1712mm (89 x 89 DPI)

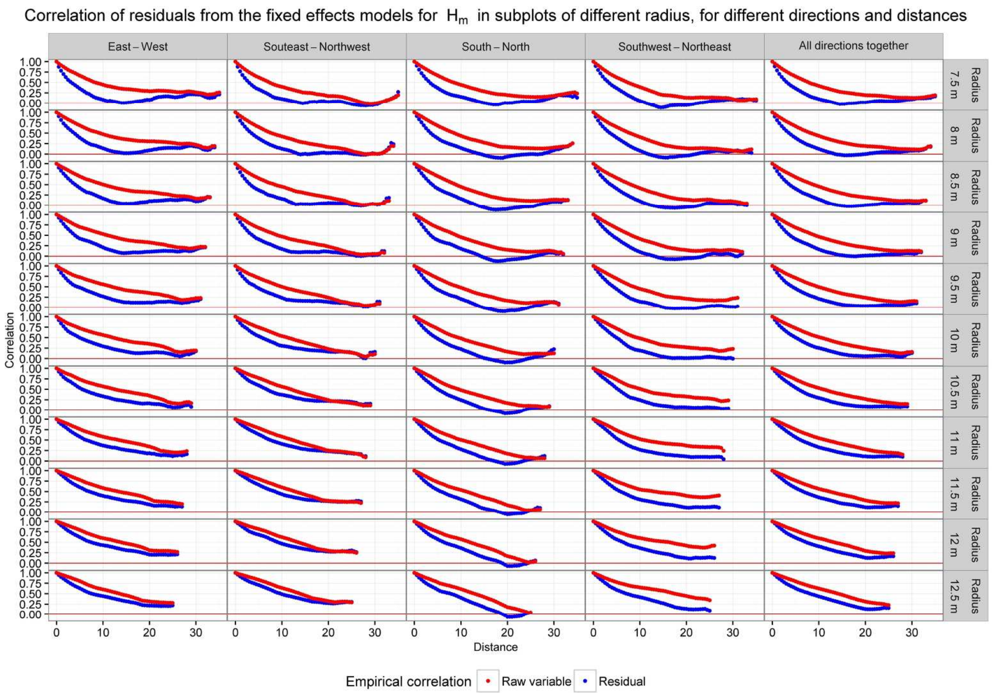


Figure A1. Empirical correlation functions observed for Hm and for its residuals in different directions and for different subplot radii. Fifth column shows the empirical correlation computed assuming isotropy.

Figure A1.

199x142mm (300 x 300 DPI)