European Journal of Forest Research manuscript No.
(will be inserted by the editor)

Combining canopy height and tree species map information for large scale timber volume estimations under strong heterogeneity of auxiliary data and variable sample plot sizes

Andreas Hill · **Henning Buddenbaum** · **Daniel Mandallaz**

Received: date / Accepted: date

Abstract A timber-volume regression model applicable to the entire forest area of the federal German state of Rhineland-Palatinate is identified using a combination of airborne laser scanning (ALS)-derived metrics and information from a satellite-based tree species classification map available on the federal state level. As is common in many forest inventory datasets, strong heterogeneity in the LiDAR data due to different acquisition dates and misclassifications in the tree species classification map had noticeable effects on the regression model's performance. This article specifically addresses techniques that improve the performance of ordinary least square regression models under such restricting conditions. We introduce a calibration technique to neutralize the effect of misclassifications in the tree species variable that originally caused a residual inflation of 5%. Incorporating the calibrated tree species information improved the model accuracy by 5% in adjusted R^2 and suggests the use of such information in forthcoming inventories. We also found that including ALS quality information as categorical variables within the regression model considerably mitigates issues with time lags between the ALS and terrestrial data acquisition and ALS quality variations (9% increase in adjusted R^2). The model achieved an adjusted R^2 of 0.48 (RMSE$_{cv}$ of 137 m^3/ha) under incorporation of the tree species and ALS quality information, and was thus improved by 13% (16 m^3/ha) compared to the simple model only containing ALS height metrics (adjusted R^2=0.35, RMSE$_{cv}$=153 m^3/ha).

Keywords OLS Regression · standing timber volume · ALS canopy height model · satellite-based tree species classification · calibration · forest inventory · angle count sampling

Andreas Hill
Department of Environmental Systems Science, ETH Zurich, Universitaetstrasse 22, 8092 Zurich, Switzerland
Tel.: +41 44 632 32 36
E-mail: andreas.hill@usys.ethz.ch

Henning Buddenbaum
Environmental Remote Sensing and Geoinformatics Department, Trier University, 54286 Trier, Germany
Tel.: +49 651 201 4729
E-mail: buddenbaum@uni-trier.de

Daniel Mandallaz
Department of Environmental Systems Science, ETH Zurich, Universitaetstrasse 22, 8092 Zurich, Switzerland
Tel.: +41 44 632 88 35
E-mail: daniel.mandallaz@usys.ethz.ch

1 Introduction

Forest inventory methods are the primary tools used to assess the current state and development of forests over time. They provide reliable evidence-based information that is used to define and identify management actions as well as to adapt forest management strategies to both national and international guidelines. Two methods that have become particularly attractive are so-called *double-sampling* (Mandallaz, 2008) and *mapping* (Brosofske et al, 2014) procedures. The core concept of these methods is to use predictions of the terrestrial target variable at additional sample locations where the terrestrial information has not been gathered. These predictions are produced by models that use explanatory variables derived from *auxiliary data*, commonly in the form of spatially exhaustive remote sensing data in the inventory area. Especially models to predict timber volume based on airborne laser scanning (ALS) have been extensively investigated for a long time (Næsset, 1997) . The specific scope of double-sampling is to enlarge the terrestrial sample size by a much larger sample of predictions of the target variable in order to gain higher estimation precision without performing additional expensive terrestrial measurements. Model-based and model-assisted regression estimators are used in a broad range of double sampling concepts and methods (Gregoire and Valentine, 2007; Köhl et al, 2006; Mandallaz, 2013a,b; Saborowski et al, 2010; Schreuder et al, 1993) and have been applied to existing inventory systems (Breidenbach and Astrup, 2012; von Lüpke and Saborowski, 2014; Magnussen et al, 2014; Mandallaz et al, 2013; Massey et al, 2014). While double-sampling methods provide reliable estimates for a given spatial unit, e.g. a forest district, they do not provide information about the spatial distribution of the estimated quantity within this area. For this reason, the same modeling technique used in double-sampling procedures has also been intensively used to produce exhaustive prediction maps that provide pixel-wise estimations of a target variable in high spatial resolution (Bohlin et al, 2017; Hill et al, 2014; Latifi et al, 2010; Nink et al, 2015; Tonolli et al, 2011).

To allow for an area-wide application of the prediction model, both double sampling and mapping methods require that the remote sensing data are available over the entire inventory area. This is usually not a limiting factor in *small-scale* applications. In the optimal case, the remote sensing data are in principle collected in accordance to the specific study objective. Quality standards that have often been addressed are that *a)* the remote sensing data should be acquired close to or even at the time of the terrestrial inventory in order to ensure best possible comparability between the target variable on the ground and the remote sensing derived variables (McRoberts et al, 2015); *b)* the remote sensing technology and its spectral and spatial resolution should be chosen according to the modelling purpose (Köhl et al, 2006); and *c)* the variation in quality of the remote sensing data over the inventory area should be minimized in order to avoid artificial noise in the data (Naesset, 2014). Despite the increasing availability and decreasing costs of remote sensing data (White et al, 2016), these quality standards of the remote sensing data can often not be guaranteed for *large-scale* applications (Maack et al, 2016), and trade-offs must be accepted (Jakubowski et al, 2013). The prime objective is then to produce the best possible prediction model given the restrictions imposed by the available remote sensing information. The exploration of scarcely used remote sensing products and the optimization of prediction models under severe quality restrictions in the remote sensing data are thus one of the challenges in large-scale model-supported inventory applications.

Among the still rarely used remote sensing data in large scale applications, the integration of tree species information in prediction models - especially for timber volume estimation - has been stated as some of the most promising but often missing information (Koch, 2010; White et al, 2016). As timber volume estimations on the single tree level in forest inventories are often based on species-specific biomass and volume equations (Husmann et al, 2017; Zianis et al, 2005), the application of species-specific models is expected to be a key factor for improving estimation precision (White et al, 2016). Breidenbach et al (2008) found that their timber volume prediction model based on ALS canopy height metrics could be significantly improved by including a variable estimating the deciduous proportion derived from leaf-off ALS data. Similar gains in model performance were also reported by Straub et al (2009) and Latifi et al (2012) who used broadleaf and coniferous information based on color infrared orthophotos as a categorical explanatory variable. However, studies that explore the use of more species-specific information (i.e. a further discrimination of tree species) as explanatory variables in prediction models have been rare. Further investigations are thus necessary especially in countries whose forests are characterized by a larger variety of tree species that may also occur in mixed and uneven-aged stands (McRoberts et al, 2010). The area-wide tree species information in most studies was obtained from satellite and airborne remote sensing sensors based on automatic classification methods. Whereas the presence of misclassifications has already been addressed (Latifi et al, 2012), an issue that has so far been neglected is how misclassifications actually affect the prediction model (Gustafson, 2003).

A frequently encountered problem in large scale forest inventories is the lack of temporal synchronicity between the remote sensing acquisition and the terrestrial survey. As a result, the available remote sensing data often exhibit notable time-lags with respect to the date of the terrestrial inventory. This has often been addressed as a major drawback, especially for the application of model-assisted change estimation (Massey and Mandallaz, 2015).

Our study is embedded in the current implementation of model-assisted regression estimators (Mandallaz, 2013a,b; Mandallaz et al, 2013) for estimating the standing timber volume within the state and communal forest management units over the entire state of Rhineland-Palatinate (RLP, Germany). With respect to this overall objective, the aim of this study was to derive an ordinary least square (OLS) regression model to generate predictions of the standing timber volume associated with a sample location of the Third German National Forest Inventory (BWI3) over the entire federal state forest area (6155 km^2). A merged ALS dataset from different acquisition years and a satellite-based tree species classification map for the five main tree species in RLP was available for the entire inventory area and consequently used to derive predictor variables. The major limiting factors for using these data in a regression analysis are **(i)** variation in the ALS data quality as well as time-lags of up to 10 years between the ALS acquisitions and the terrestrial survey, **(ii)** misclassifications in the tree species classification map and **(iii)** the ambiguous choice of a suitable extraction area (*support*) for all remote sensing information under angle count sampling in the terrestrial survey (variable sample plot sizes). For this reason, we address the following specific research questions:

1. How can tree species map information be optimally used within a regression model that predicts timber volume? What effects do misclassifications have on the predictions and how can these effects be minimized?
2. What are the effects of quality restrictions and substantial time lags between the ALS- and terrestrial data acquisition on the regression model and how can these effects be mitigated?
3. Does support size influence model accuracy? What is the optimal support size and what are the determining factors?

2 Materials and Methods

2.1 Study Area

The German federal state Rhineland-Palatinate (RLP) is located in the western part of Germany and borders Luxembourg, France and Belgium (figure 1). With 42.3% (appr. 8400 km^2) of the entire state area (19850 km^2) covered by forest, RLP is one of the two states with the highest forest coverage among all federal states of Germany (von Thünen-Institut, 2014). The most frequent tree species in RLP are European beech (*fagus sylvatica*, 21.8%), oak (*quercus petrea* and *quercus robur*, 20.2%), Norway spruce (*picea abies*, 19.5%), Scots pine (*pinus sylvestris*, 9.9%), Douglas fir (*pseudozuga menziesii*, 6.4%), European larch (*larix decidua*, 2.4%) and Silver fir (*abies alba*, 0.7%). The share of broadleaf tree species is 58.7%. The forests of RLP further exhibit heterogeneous structures (von Thünen-Institut, 2014): around 82% of the forest area in RLP are mixed forest stands (i.e. at least two different tree species occur in the same stand) and 69% of the forest area exhibit a multi-layered vertical structure. While the average tree age is around 80 years, most of the forest area (20%) is occupied by trees between 40 and 60 years of age, whereas 27% of the trees are older than 100 years. Spatially variable climate conditions have a strong influence on the local growth dynamics as well as tree species composition and create a large variety of forest structures, ranging from characteristic oak coppices (Moselle valley), pure spruce, beech and scots pine forests (e.g. Hunsrück and Palatinate forest) to mixed forests comprising variable proportions of oak, larch, spruce, Scots pine and beech. Accordingly, RLP has been divided into 16 bioclimatic growing regions that form homogeneous areas with respect to the afore mentioned characteristics (Gauer and Aldinger, 2005).

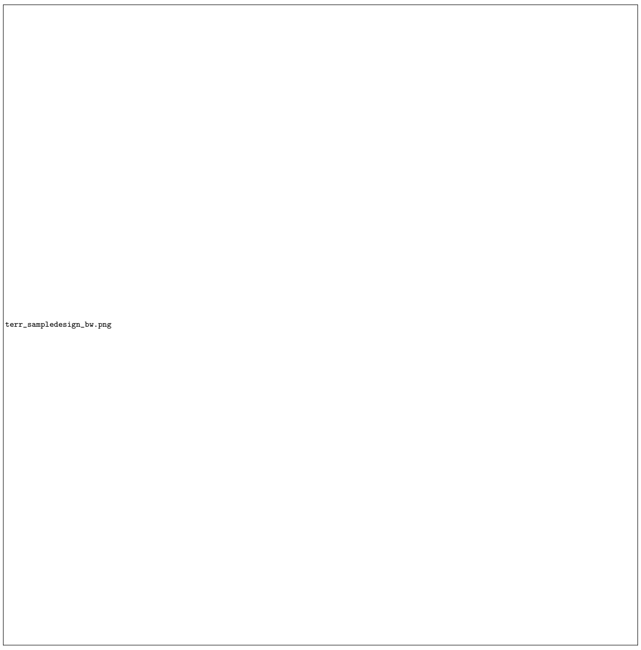


Fig. 1: Spatial distribution of the BWI3 cluster samples over Rhineland-Palatinate

2.2 Terrestrial Inventory Data

The German National Forest Inventory is carried out over the entire forest area of Germany in reoccurring time periods of 10 years. The most recent inventory (BWI3) has been conducted in the years 2011 and 2012. In this framework, Rhineland-Palatinate is covered by a 2x2 km grid that defines the sample locations for the terrestrial survey. A sample unit consists of four sample locations (also referred to as *sample plots*) that are arranged in squares (so called *clusters*) with a side length of 150 metres (figure 1). The number of plots per cluster can however vary between 1 and 4 depending on forest/non-forest decisions on the plot level (Bundesministerium für Ernährung, 2011). In the field survey of the BWI3, sample trees for timber volume estimations are selected according to the angle count sampling technique (Bitterlich, 1984), using a basal area factor (*BAF*) of 4 that is respectively adjusted for boundary effects at the forest border (Bundesministerium für Ernährung, 2011). A further selection criterion for a tree to be recorded is a diameter at breast height (*dbh*) of at least 7 cm. This sampling technique was applied to 8092 sample plots (2810 clusters) in RLP, resulting in the collection of 56561 sample trees for which the dbh, the tree diameter at 7 m (*D*7) and the tree species were recorded for all trees. Tree height measurements were conducted only for a subset of all sample trees and used to predict the height for the remaining sample. During the last inventory, all plot center positions were remeasured with a differential GPS technique. Knowledge about the exact plot positions were considered crucial to provide optimal comparability between the terrestrially observations and the information derived from the auxiliary information. A detailed analysis of horizontal DGPS errors in RLP by Lamprecht et al (2017) indicated that horizontal DGPS errors do not exceed a range of 8 meters for 80% of all plots. For 162 plots, the DGPS coordinates were replaced by their former target coordinates due to missing or implausible values. In order to derive a volume estimation for each sample tree, the BWI3 estimates a taper curve for each sample tree by calibrating the random effects term of linear mixed-effects taper models with the set of diameters and corresponding height measurements taken from the respective sample tree (Kublin et al, 2013). The integration of the derived taper curves consequently lead to a volume prediction for each sample tree. Since the overall objective of the study was to subsequently use the identified regression model for design-based timber volume estimations within the state and communal forest management units, we already restricted the sample plots used for modeling to the state and communal forest area (73% of the entire forest area of RLP). This provides the advantage that when used as an internal model in design-based estimators, the regression model predictions already hold the assumption on the residuals to be zero on average for state and communal forest by construction of OLS technique (Mandallaz, 2013a,b; Mandallaz et al, 2013). The dataset of this study hence comprised 5791 plots (2055 clusters). For this sample, the timber volume density per hectare on plot level, $Y(x)$, was calculated according to the formula of one-phase one-stage sampling (Mandallaz, 2008). The timber volume density per hectare on plot level was used as the response variable in the regression analysis.

2.3 Auxiliary Information

2.3.1 LiDAR Canopy Height Model

Between 2003 and 2013, the topographic survey institution of RLP acquired airborne laser scanning (LiDAR) data over the entire state of RLP at leaf-off condition (Figure 2). The objective of this campaign was to derive a countrywide digital terrain and surface model based on the acquired LiDAR point clouds. During the extended acquisition period, airborne laser scanning technology and data quality evolved significantly. The tiles recorded in 2002 and 2003 have a rather poor quality with about only 1 point per 5x5 m^2, while more recently acquired datasets contain more than 125 points per 5x5 m^2 raster cell. The data was delivered as two separate point clouds: one cloud contained filtered ground returns, whereas the other cloud contained first pulses from non-ground objects. All point clouds were delivered as three-column (easting, northing, and height above sea level) ASCII files in tiles of 1 km^2. Before interpolating the point clouds to regular rasters, the clouds were thinned. For the ground data, the mean value of each raster cell in the final resolution of 5x5 m^2 was calculated. For the surface model, both ground and vegetation point clouds were first united, and the maximum value for each raster cell was determined respectively. The combination of both point clouds was necessary in order to avoid large spaces without laser points between vegetated areas that would otherwise have been filled with unrealistic values in the interpolation step. The thinned point clouds were aggregated to larger tiles in order to decrease the number of seamlines in the final mosaic. The aggregated tiles were then interpolated to raster images using a Delauney interpolation in the Matlab software (Mathworks, 2017). The resulting two elevation models were then used to calculate a canopy height model (*CHM*) in raster format, providing discrete information about the canopy surface height of the forest area in a spatial resolution of 5 meters.

As explanatory variables, the mean canopy height (*meanheight*) and the standard deviation (*stddev*) were calculated as the mean and standard deviation of all raster values within a predefined square around each sample plot center. The square (i.e. *support* of the explanatory variable, see section 2.4) was previously intersected with the state and communal forest area defined by a polygon mask and thereby corrected for edge effects at the forest border. The tree height is one prominent predictor variable in the taper functions of the BWI3 that are used to calculate a timber volume value for each sample tree (Kublin, 2003; Kublin et al, 2013). A visual inspection of the tree volumes of all sample trees collected in the BWI3 within RLP against their tree heights also revealed the characteristic shape of an allometric relationship between these variables (Online Resource 1). It was hypothesized that this relationship on single-tree level is also apparent on the aggregated level of a sample plot and cluster, and can be used within the frame of regression modeling.

The strength of correlation between *meanheight* and timber volume on plot level was expected to show high variation according to the mentioned time-lag up to 10 years between LiDAR acquisition and terrestrial survey. The quality of the height information was also expected to vary according to changing sensor technologies and different point densities used over the years. For these reasons, the LiDAR acquisition year (*lidaryear*) for each sample plot was considered as a potential categorical explanatory variable to explain the variation in the data introduced by these factors. For this purpose, the acquisition year *2008* was further divided into *2008* and *2008_1*. In the latter, the data quality turned out to be very poor due to sensor failures during the acquisition. Additionally, the years *2006* and *2007* as well as *2012* and *2013* were pooled in order to increase the number of observations per factor level for modelling reasons. As a result, the *lidaryear* variable comprised nine categories (*2002*, *2003*, *2007*, *2008*, *2008_1*, *2009*, *2010*, *2011* and *2012*).

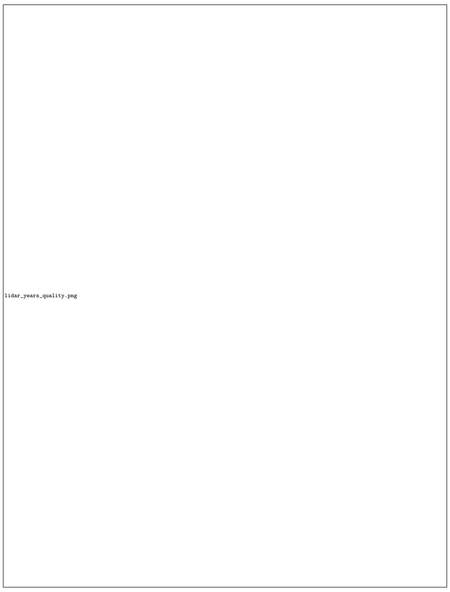


Fig. 2: Separate LiDAR acquisitions in Rhineland-Palatinate over the years. The colors also indicate the quality of the data: *light*: low point densities (1/5x5m^2), *dark*: high point densities (>100/5x5m^2)

2.3.2 Tree Species Classification Map

A countrywide satellite-based classification map of the five main tree species (European beech, Sessile and Pedunculate oak, Norway spruce, Douglas fir, Scots pine) described in Stoffels et al (2015) was used to derive tree species information on sample plot level. The classified tree species map has a grid size of 5 meters and predicts five of the seven tree species that are used in the BWI3 taper functions (Kublin et al, 2013) to calculate the timber volume of a sample tree. Due to unavailable satellite data for the classification, the tree species map excluded one patch with an area of 415 km^2 in the south-west part of RLP, and two further patches with an area of 76 km^2 and 100 km^2 in the northern part (Stoffels et al, 2015). The tree species information was consequently missing for 407 (7%) of the 5791 sample locations.

Prediction of main plot tree species

A visual inspection of all BWI3 sample trees of RLP suggested that a stratification of the relation between tree height and timber volume according to these seven tree species may provide a considerable reduction in variation within the tree species groups (Online Resource 1). This led to the hypothesis that this tree species specific signal might also be apparent on sample plot and cluster level and can consequently be used to increase the accuracy of the prediction model. Based on the tree species classification map, the main tree species of each sample plot was calculated as an additional categorical explanatory variable (*treespecies*) with six categories following a similar approach as Latifi et al (2012): one of the five tree species was assigned as the main plot tree species if its proportion within the edge-corrected support around the sample location exceeded a predefined threshold. If this threshold was not reached by any of the five tree species, the respective sample plot was assigned the category 'Mixed'.

Calibration

Our analyses revealed that the prediction of the main tree species for a sample plot can be subject to misclassifications (section 3.1). Errors in the explanatory variables of linear regression models can however lead to a bias of the regression coefficients in the direction of zero due to an artificial introduction of noise (Carroll et al, 2006). This can cause an inflation of the residual variance and a consequent decrease of the model accuracy (Magnussen et al, 2010). In case of classification the impacts of misclassifications on the model properties are even harder to predict (Gustafson, 2003). While errors in the explanatory variables do not affect the unbiasedness of the estimators in the model-assisted framework, a reduction or elimination of the classification errors could provide an improvement of the regression model accuracy and thereby potentially lead to smaller prediction and estimation errors. We therefore addressed the effect of misclassifications in the *treespecies* variable by the following analysis:

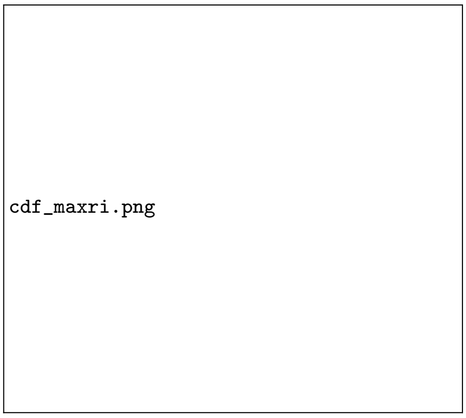
a) we investigated the effect on the regression model performance (regression coefficients, model accuracy) when substituting the *predicted* by the *actual* main plot tree species derived from the sampled trees of the respective sample plot under identical threshold settings
b) we used the random forest algorithm (Breiman, 2001; Liaw and Wiener, 2002) in the statistical software *R* (R Core Team, 2016) to define a *calibration model* in order to improve the classification accuracy of the initially predicted main plot tree species, correct for potential systematic misclassifications and thus minimize the effect of misclassifications on the regression model. The random forest algorithm is a machine learning algorithm that grows a large number of decorrelated classification trees by considering only a subset of all provided predictor variables for each split. In the case of classification, new data are thus predicted by aggregating the predictions of all trees using a majority vote. For our purpose of predicting the actual main tree species of a sample plot (target variable), we provided the random forest algorithm with a full set of *p* predictor variables that comprised the initial prediction of the main plot tree species

(*treespecies*), the mean canopy height (*meanheight*) and standard deviation (*stddev*) derived from the CHM, the proportion of coniferous trees estimated from the tree species classification map (*prop.conif*) and the bioclimatic growing region (*wgb*) at the sample location. The algorithm was grown with 2000 trees, considering $\sqrt{p} \approx 3$ of the predictors for each split.

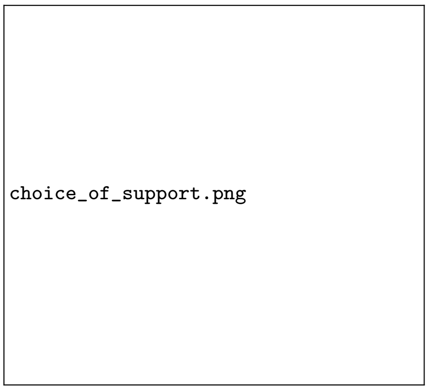
2.4 Choice of Support under Angle Count Sampling

One characteristic of angle count sampling applied in the BWI3 is that a sample plot does not have a fixed radius in which trees are selected (*fixed-radius plot*), but each tree generates an individual radius from the plot center depending on its diameter at breast height (*variable-radius plot*). This tree-individual radius is known as the *limiting distance* from the plot center where the tree would still be included in the sample. A consequence of the absence of a fixed plot radius is the question about the optimal support (Hollaus et al, 2007), i.e. the spatial extent around the plot center in which the auxiliary information is evaluated and transformed into an explanatory variable. It has widely been hypothesized that the best relationship between the target variable on the ground and any explanatory variable derived from the auxiliary information is obtained if the support is spatially identical to the sample plot extent. In case of angle count sampling, an individual extent for each sample plot can be approximated by regarding the maximum limiting distances of its sample trees as the outer plot radius. However, many model-assisted applications under double-sampling do not allow for a between-plot change of the support for a specific explanatory variable (Mandallaz, 2013a,b).

For this reason, the task is to find a unique support for each auxiliary information that leads to the best overall model accuracy. Deo et al (2016) conducted extensive analysis to identify optimal supports for modelling standing timber volume for *variable-radius plot* designs in conifer forests. They analysed 24 different radii (i.e. circular supports) in which they extracted 57 metrics from a LiDAR derived point cloud with an average point density of 18 pulses per square meter. They successively evaluated the prediction performance of each support size by using the LiDAR metrics in a random forest algorithm and comparing the resulting model accuracies. In order to identify the best-performing supports for our explanatory variables, we followed a similar approach. The explanatory variables were calculated using *individual* (i.e. plot-varying) supports (*ind*), i.e. an individual support extent was used for each plot according to the maximum limiting distance of all sample trees associated to the respective sample plot. We then compared the model accuracies achieved by the individual supports against the model accuracies from a set of *fixed* (i.e. non plot-varying) supports. The extents of the fixed supports were chosen from the cumulative distribution function (ECDF) of the maximum limiting distances of all 5791 sample plots of the analysed forest area (Fig. 3a). We considered the 25^{th} ($q25$, 9 meters), 50^{th} ($q50$, 12 meters), 80^{th} ($q80$, 15 meters) and the 100^{th} ($q100$, 38 meters) percentiles, resulting in support side lengths of 18, 24, 30 and 76 meters (Fig. 3). While in this study we used squares to extract the auxiliary information, also other support-shapes are possible (e.g. circles, hexagons). We also want to emphasize that the use of different support sizes for each explanatory variable is perfectly valid in the infinite population framework of model-assisted estimators (Mandallaz, 2013a,b).



(a) ECDF of maximum limiting distances of all BWI3 sample locations in RLP



(b) Rectangular supports used to extract explanatory variables around sample locations. *Dash dot dot line*: q100, *dash dot line*: q80, *dot dot line*: q50, *dot line*: q25, *solid line*: individual support, *triangles*: sample trees

Fig. 3: Identification (*a)*) and visualization (*b)*) of potential supports used for calculating the predictor variables on plot level

2.5 Model Validation

In order to judge the quality of the *treespecies* variable, the user's accuracy for each classified species and the overall accuracy of the classification scheme was calculated based on the confusion matrix (Congalton and Green, 2008), using the main plot tree species calculated from the sample trees as reference data. The classification accuracy was performed for all support sizes for both the calibrated and the uncalibrated *treespecies* variables. The measures of the regression model accuracy using both CHM- and *treespecies* variables were defined as the 10-fold cross-validated root mean square error (RMSE_{cv}) and the adjusted coefficient of determination (adjusted R^2) of the multiple linear regression model defined in equation 1. Additionally, we considered the interaction terms *meanheight*:*treespecies*, *meanheight*^{2}:*treespecies*, *meanheight*:*lidaryear*, *stddev*:*lidaryear* and *meanheight*:*stddev* and performed a variable selection based on the Akaike Information Criterion (AIC) (Akaike, 2011) in order to minimize the number of variables in the model. Due to a pronounced unbalanced design in the *treespecies*-*lidaryear* strata (Online Resource 2), no interaction between *treespecies* and *lidaryear* was possible. We evaluated the model for all support combinations, considering the use of individual support sizes for each auxiliary information, using both the calibrated and the uncalibrated *treespecies* variable. The calibration model (section 2.3.2) for the *treespecies* variable was recalculated for each respective support-threshold setting.

$$
\begin{aligned}
Y(x) = & \beta_0 + \beta_1 * meanheight + \beta_2 * meanheight^2 + \\
& \beta_3 * stddev + \\
& \beta_4 * lidaryear_1 + \ldots + \beta_{12} * lidaryear_9 + \\
& \beta_{13} * treespecies_1 + \ldots + \beta_{18} * treespecies_6 + e(x)
\end{aligned} \tag{1}
$$

206 sample plots included no sample trees and the timber volume density $Y(x)$ was thus set to *zero*. These *zero-plots* were removed from the modeling dataset since they acted as leverage points in cases where the LiDAR height metrics were recorded long before the terrestrial survey. Together with the missing tree species information (section 2.3.2), the modeling dataset was limited to 5206 observations.

3 Results

3.1 Classification Accuracies

Effect of Support Size and Threshold

Before calibration, the lowest user's accuracies (*UA*) for most tree species were realized using high thresholds of 80% and 100% for deciding the main tree species on the plot level (figure 4a). A plausible reason for this is that raising the threshold to higher values (e.g. 80%, 100%) distinctively increases the probability of the reference class (based on the sample trees of the sample location) to be assigned as class 'Mixed', while the much coarser spatial resolution of the tree species map causes the *predicted* class to remain classified as one of the five tree species. However, as the support size is increased, so does the number of tree species raster cells to be evaluated at the sample location, thereby increasing the probability that the predicted class will be 'Mixed'. For this reason, most tree species exhibit an increase in user's accuracy under higher thresholds with higher support sizes. This scale-threshold dependency of the user's accuracy particularly affects tree species that most commonly occur in mixed forest stands in Rhineland-Palatinate (*Scots pine*, *oak* and *beech*), whereas the user's accuracies for tree species that are mostly prominent in pure forest stands (*spruce*, *Douglas fir*) logically turned out to be much more robust to changes in the thresholds and support sizes.

Among the uncalibrated tree species predictions, *beech* and *spruce* produced the best predictions achieving UAs of up to 70% and 80%. Although the predictions for *Douglas fir* and *Scots pine* generally performed less well than *beech* and *spruce*, similar UAs can be produced by adjusting the threshold and support choices. UAs for *oak* never performed better than 50%. A detailed table of the user's and overall accuracies is provided in Online Resource 3.

cos_tspec_mixed_cal_nocal.png

(a)

cos_tspec_oaa_mixed.png

(b)

Fig. 4: Classification accuracy for the main tree species of a sample location *before* and *after* calibration: *a)* user's accuracies. *b)* overall accuracies. *n*: number of validation data per class

Effect of Calibration

Calibration substantially diminished the effect of the scale-threshold dependency for the five tree species and also increased the UAs for *Scots pine* and *oak*. Whereas the UA level of *spruce* remained unchanged, the UAs for *beech* were found to be slightly lower after calibration. The overall accuracy under each support choice was always considerably increased by calibrating the tree species prediction (figure 4b). With respect to the calculated random forest models, the initial tree species prediction (*treespecies*) and the information about the growing region (*wgb*) turned out to be the most valuable information, followed by the estimated proportion of coniferous trees (*prop.conif*) and the mean canopy height (*meanheight*).

3.2 Regression Model Accuracies

Effect of Support Size and Threshold

Figure 5 shows the accuracies of the regression model (equation 1) achieved under all possible combinations of support sizes for the auxiliary data. The stepwise selection procedure always included all considered single and interaction terms. In terms of adjusted R^2 and RMSE_{cv}, the analysis revealed that the choice of the CHM support size controls the overall level of the model's accuracy. The information about the main plot tree species can then be used to further improve the model fit under suitable *treespecies* support and threshold settings. When using the uncalibrated *treespecies* variable, an increase of the *treespecies* support size causes an increase in the model performance if low thresholds are used, whereas high thresholds (80%, 100%) cause a decrease in the model performance. This threshold-dependency could be removed by calibrating the *treespecies* variable. The highest adjusted R^2 and the lowest RMSE_{cv} were realized using the $q50$ support for the CHM variables in combination with the $q100$ support and a threshold of 100% for the calibrated *treespecies* variable (adjusted R^2=0.48 and RMSE_{cv}=137 m^3/ha). However, various support and threshold combinations for the CHM and *treespecies* variables can be used to yield almost identical RMSE_{cv} and adjusted R^2 values. A detailed table of the model accuracies is given in Online Resource 4.

Effect of Misclassifications

We can assess magnitude of the misclassification effect by comparing the adjusted R^2's of models that use the predicted tree species (calibrated and uncalibrated) as an explanatory variable to models that use the error-free tree species variables acquired from the terrestrial survey. Note that only the model with the predicted tree species variables can be applied to additional sample locations where no terrestrial survey has been carried out. Figure 6 provides a comparison of the adjusted R^2 achieved under the use of the error-free tree species predictor variable against the adjusted R^2 realized cos_cal_nocal_bw.png

Fig. 5: 10-fold RMSE$_{cv}$ and adjusted R^2 realized under various support choices for the CHM and *treespecies* explanatory variables

under the use of the tree species variable containing missclassifications. This analysis was carried out for all models that were analysed in section 3.2, i.e. for all possible support and threshold combinations for the CHM and *treespecies* predictor variables.

As expected, the highest adjusted R^2 for every evaluated model was always achieved using the error-free tree species variable, whereas the missclassifications in the tree species variable led to a systematic decrease of the model accuracy. This is in agreement with the potential effects of erroneous explanatory variables discussed in Carroll et al (2006) and Gustafson (2003), i.e. an increase of variability (noise) in the data that can increase the amount of unexplainable variance and thereby reduce the model accuracy.

The calibration of the initially predicted main plot tree species using the random forest classification algorithm (section 2.3.2) turned out to not only improve the classification accuracies (section 3.1), but also to considerably decrease the effect of the missclassifications on the regression model predictions and accuracy. Figure 6 (*right*) shows that the adjusted R^2 under the actual and the calibrated predicted tree species variable are in general much closer to, and in many cases even on the identity line. Whereas the misclassifications in the uncalibrated *treespecies* variable led to a residual inflation of 1% - 5%, it was only between 0% and 1% after calibration. Further analysis revealed that when using the calibrated *treespecies* variables, the regression coefficients were almost identical to the ones received using the actual main plot tree species.

3.3 Final Regression Model

In order to address research questions 1 and 2 (i.e. the gain in model accuracy by tree species information and effect of heterogeneity in the LiDAR data), we investigated the model properties in more detail. For this purpose, we decided to cos_r2_cal_vs_nocal.png

Fig. 6: Effect on the adjusted R^2 when substituting the actual main tree species with the predicted main tree species of a sample plot. The differentiation into two distinct point clouds results from the poor model performance under support size $q100$ for the CHM variables (i.e. the *lower* point cloud). The *dotted* line tracks the the model with the highest adjusted R^2 under the use of the error-free *treespecies* variable

use the support settings of $q50$ for both auxiliary data with a threshold of 100% for the tree species variable as the regression model of choice. The reason for this choice was that *a)* the model provided almost the highest adjusted R^2 among all validated models while reducing the data handling complexity for upcoming applications (i.e. identical support sizes for all remote sensing data) and *b)* the calibration neutralized the effects of misclassifications on the model predictions. The interaction term between *meanheight*^{2} and *treespecies* (i.e. considering separate curvatures for each tree species) turned out not to have any influence on the model accuracy and was thus dropped, resulting in an adjusted R^2 of 0.49 and RMSE$_{cv}$ of 132.12 m^2/ha.

Interpretation of Final Regression Model

Figure 7 provides a visualisation of the tree species prediction functions separated by the LiDAR acquisition years. Sample plots classified as *oak* and *Scots pine* revealed to have an almost identical relationship (nearly identical slopes) for the mean canopy height - timber volume relationship. They only differ by a marginally higher intercept for *Scots pine* plots, meaning that given the same mean canopy height a sample plot dominated by *Scots pine* yields a marginally higher timber volume on the plot level than a plot dominated by *oak*. *Beech*-dominated sample plots tend to achieve a higher timber volume than *oak* and *Scots pine* for canopy heights below 20 meters, but realize the lowest timber volumes for canopy heights above 20 metres. Sample plots dominated by any of the remaining coniferous tree species (*Douglas fir*, *spruce*) revealed to have higher slopes than broadleaf classified plots. This indicates that given the same mean canopy height, sample plots dominated by *Douglas fir* and *spruce* yield higher timber volume values than broadleaf- or *Scots pine* dominated sample plots, and this difference becomes more pronounced with increasing mean canopy heights. Within the group of coniferous-dominated sample plots, *spruce* turned out to have the highest slope, thereby yielding the highest timber volume values for mean canopy heights above 15 meters. An undesired characteristic of the model is that the predicted timber volume can in some cases ($< 1\%$) take negative values for low canopy heights (e.g. for *spruce*-dominated plots with *meanheight* below 5 meters and *stddev* of 4 meters). However, we chose not to use a log-transformation of the response variable. Doing so would have prevented the subsequent calculation of the g-weight variance of the model assisted estimators (Mandallaz, 2013a; Mandallaz et al, 2013), which is only possible for response variables on the original scale.

Effect of Time-Lags and Heterogeneity in LiDAR Data

Incorporating the LiDAR acquisition year as a categorical variable (*lidaryear*) in the regression model substantially accounted for the variability in the data introduced by *a)* the time-lags between LiDAR acquisition and terrestrial survey, and *b)* variation in LiDAR data quality which are due to sensor- and post processing techniques (table 2). Whereas the adjusted R^2 for the regression model without considering the LiDAR acquisition year as additional predictor variable was 0.35 (0.40 including the tree species variable), the stratification by the LiDAR acquisition year led to adjusted R^2 of 0.44 (0.48), thereby increasing the proportion of explained variance by up to 8%.

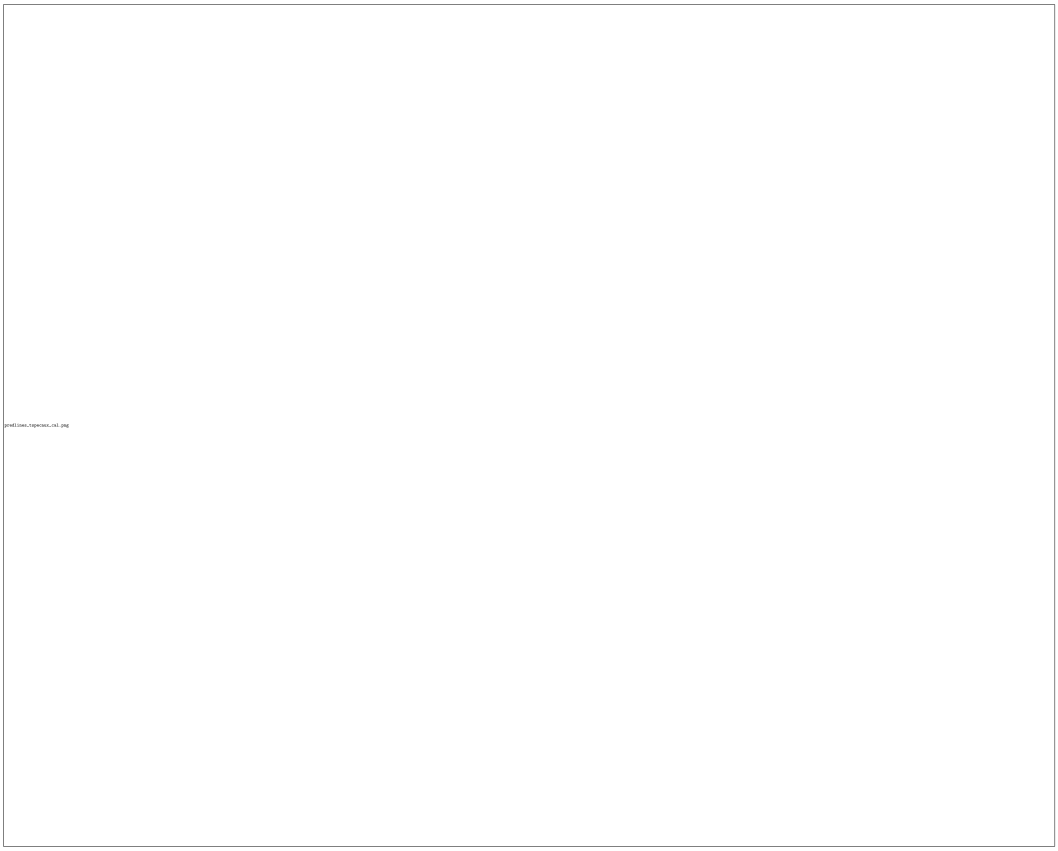


Fig. 7: Visualization of the timber volume prediction function (*final regression model*) on sample plot level for each main plot tree species and LiDAR acquisition year. For visualization purposes, the predictor variable *stddev* was set to its average value within the respective *treespecies* and *lidaryear* group
. The terrestrially observed timber volume values are plotted in the background.

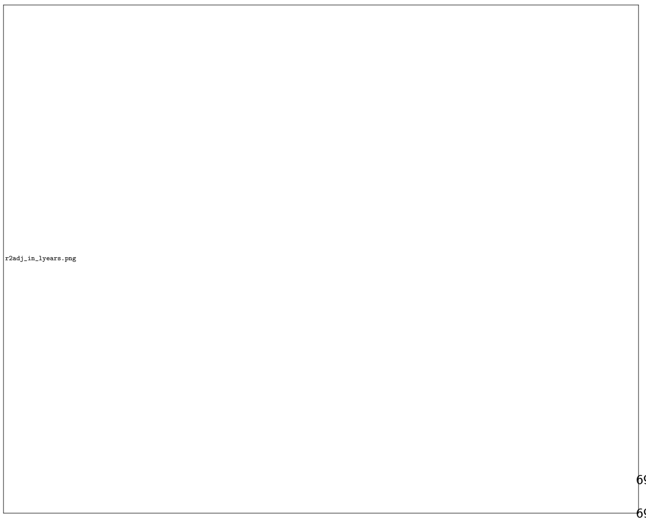


Fig. 8: R^2 of the final regression model achieved *within* the LiDAR acquisition year strata. *Grey points*: R^2 of submodel 1 (no stratification according to LiDAR acquisition year or tree species). *Crosses*: R^2 of submodel 2 (*without* tree species stratification). *Filled triangles*: R^2 of final model using the *uncalibrated* tree species variable. *Empty triangles*: R^2 of final model using the *calibrated* tree species variable. *Diamonds*: R^2 achieved using the error-free tree species variable (derived from sample trees). *Dotted line*: Overall adjusted R^2 of submodel 2. *Solid line*: Overall adjusted R^2 of final model using the *calibrated* tree species variable. *Two-dashed line*: Overall adjusted R^2 of final model using the *error-free* tree species variable

Table 1: R^2, RMSE and Residual Square Sum (SSE) of final regression model within LiDAR acquisition year strata (*lidaryear*). *n*: number of validation data

LiDAR acquisition year	R^2	rmse	SSE	n
2012	0.55	139.54	7535278	387
2011	0.55	145.21	17880553	848
2010	0.48	122.16	16907662	1133
2009	0.43	127.17	8652419	535
2008_1	0.34	170.49	11161424	384
2008	0.50	124.39	10475374	677
2007	0.49	129.72	6950192	413
2003	0.32	146.37	11139814	520
2002	0.43	139.79	6038601	309

We further analysed the model residuals within each LiDAR acquisition year (within-group variation) for the final model and nested submodels. It turned out that the R^2 vary distinctly between the LiDAR acquisition year strata (figure 8). More precisely, the within-group R^2 can be higher and lower than the overall R^2 of the respective model. Figure 8 shows that a stratification according to the LiDAR acquisition years (submodel 2) can already increase the R^2 in most acquisition year strata, compared to the basic model using only the LiDAR height metrics as predictor variables (submodel 1). In some LiDAR acquisition year strata (i.e. 2007, 2008), this increase in R^2 even reached 7% - 11%. The accuracies for the final model are also given in table 1.

Added Value of Tree Species Map Information

Introducing the predicted main tree species of a sample plot as an additional categorical variable to submodel 2 yielded a further 4% increase in the adjusted R^2 (table 2). This improvement was even more pronounced in LiDAR acquisition years close or identical to the year of the terrestrial inventory (up to 7% increase in R^2, figure 8. The analysis illustrated once more that misclassifications in the tree species variable generally reduce model accuracy compared to using error-free tree species information. The residual inflations caused by the misclassifications in the uncalibrated *treespecies* variable within the *lidaryear* strata were up to 5%. However, the calibration was able to substantially decrease or even remove the effects of misclassifications on the model accuracy in all LiDAR acquisition year strata.

4 Discussion

4.1 Stratification according to Tree Species and LiDAR Acquisitions

Incorporating the main tree species of a sample location in the timber volume regression model significantly increased the model accuracy and revealed strong evidence for the existence of a tree species specific behaviour concerning timber volume on the plot level. This result seems reasonable regarding the species specific taper functions on single-tree level applied in the BWI3 (Kublin, 2003; Kublin et al, 2013). Further evidence and specification of the tree species effects on sample plot level - up to modeling individual tree species - would be desirable. However, this was not possible in our study because the stratification according to the LiDAR acquisition years severely limited the flexibility of species-specific prediction functions and model interpretability. In particular, using the LiDAR acquisition years as categorical variables led to highly unbalanced datasets when stratifying according to the main plot tree species, and prevented the use of further stratification variables such as bioclimatic growing regions due to confounding effects and consequent singularities in the design matrices. A stratification to the LiDAR acquisition years however proved to be a means in accounting for the artificially introduced noise in the data caused by quality variations and the large time-lags between the remote sensing and terrestrial data. Incorporating the calibrated tree species information further improved the model accuracy by 4% in adjusted R^2. Compared to the simple model only containing LiDAR height metrics, including the LiDAR quality and calibrated tree species information increased the adjusted R^2 by 13% in total. A differentiated evaluation of the final regression model revealed that the highest R^2 where achieved within LiDAR acquisitions year strata identical with the year of the terrestrial survey. Also the gain in the R^2 by including the tree species information was largest (i.e. 7%) in combination with LiDAR information acquired in the year of the terrestrial inventory. These insights were particularly interesting with respect to the further use of the regression model for small area estimations. Small area estimators generally gain modeling strength by defining the prediction model *globally* (i.e. using all data in the inventory area), and then applying the so-derived prediction model to a subset of observations located within the area of interest (Mandallaz et al, 2016). Consequently, the proposed stratification technique in the prediction model could be expected to yield a gain in model accuracy and a reduction of the small area estimation errors if the small area domain mostly includes data from strata that have high within-strata model accuracies. This hypothesis is subject to ongoing analysis.

4.2 Calibration of Tree Species Map Information

The accuracy assessment of the initially derived main plot species from the classification map revealed the presence of misclassifications that led to a decrease in model accuracy. One reason for the misclassifications were that the classification algorithm of Stoffels et al (2015) was exclusively trained in pure stands with the objective to predict the *dominant tree species* of a forest stand. Thus, our requirements on the classification map differed considerably from the ones imposed by Stoffels et al (2015) and have to be considered as far more difficult to meet. Firstly, the reference data used in the accuracy assessment also included understory trees that were recorded in the BWI3 sample. Secondly, determining an exact spatial validation unit for a sample location (support) is not possible due to the properties of angle count sampling (section 2.4). Thirdly, distinct discrepancies in the spatial scale between the reference data and the classification map severely hamper exact predictions of the main plot tree species especially in mixed forest stands. The latter issue caused a pronounced dependency of the user's accuracy on the support and threshold choice, particularly for tree species that most commonly occur in mixed forest structures, i.e. *Scots pine* (91%), *oak* (90%) and *beech* (85%) (von Thünen-Institut, 2014). With respect to this set-up, the application of our calibration method proved to be of high value. It led to an increase in the classification accuracies, particularly for those tree species that performed worse in the uncalibrated setup, and thereby successfully minimized and even removed the deleterious effect of misclassifications on model accuracy and regression coefficients. We consider

Table 2: Accuracy metrics for submodels of final OLS regression model

model terms	model	parameters	R^2_{adj}	RMSE_{cv}
meanheight + stddev + meanheight^2 + treespecies + lidaryear + meanheight:treespecies + meanheight:lidaryear + meanheight:stddev + stddev:lidaryear	final model	39	0.48	137.49
meanheight + stddev + meanheight^2 + meanheight:stddev	submodel 1	5	0.35	153.02
meanheight + stddev + meanheight^2 + lidaryear + meanheight:lidaryear + meanheight:stddev + stddev:lidaryear	submodel 2	29	0.44	142.82
meanheight + stddev + meanheight^2 + treespecies + meanheight:treespecies + meanheight:stddev	submodel 3	15	0.40	142.82

this *a posteriori calibration* a valuable method for future studies where an external tree species map (i.e. the map was not created for the specific study objective) is used in prediction models. Whereas the extensive analysis in our study deepened the understanding of the afore mentioned scale-effects, an alternative method for future applications could be to use map-derived percentages of each tree species as predictor variables in the random forest algorithm in order to directly predict the terrestrially observed main plot tree species.

4.3 Choice of Support under Angle Count Sampling

The validation of different support sizes underlined that the support choice can impact prediction accuracy. In the present study, differences in the model accuracies turned out to be small for most support choices. An exception was the choice of the $q100$ support for the CHM derived variables (76 meter side length), where the model accuracy was considerably worse than what was achieved under optimal settings. With the exception of the latter, the accuracy differences according to adjusted R^2 and RMSE_{cv} were very similar to those found by Deo et al (2016) when evaluating the model performance of optimal support sizes for a range of various basal area factors. An analysis to find the best support settings therefore seems to be advisable prior to further applications of model-assisted or model-dependent inventory methods so as not to lose model accuracy by unsuitable support choices. The concept of the demonstrated analysis method for identifying suitable supports can be transferred to any kind of auxiliary information, predictor variable and prediction model.

Contrary to our hypothesis, the use of plot-individual supports did not yield the best prediction performances. A plausible reason for this is that determining an exact plot radius under angle count sampling is technically infeasible, and thus, angle count sampling does not seem to be adequate when linking inventory information with remote sensing data. However, the extensive analysis carried out in our study indicated that the optimal support size depends on the spatial resolution of the remote sensing data as well as the context in which the derived information is used in the prediction model. In the case of transforming the tree species information map into a suitable categorical predictor variable, the use of a large support size of 76 meter side length turned out to yield the best model accuracy. However, only few sample locations in the study area were actually characterized by limiting circles of that particular size.

5 Conclusion

The objective of this study was to identify a suitable ordinary least square regression model that can be applied over the entire forest area of Rhineland-Palatinate using model-assisted estimators. The large amount of data that was gathered in the frame of this study allowed for extensive modeling possibilities, but had the side effect of contributing to high heterogeneity in the response and explanatory variables. Whereas the variability of the response variable (timber volume on plot level) is due to the very heterogeneous forest structures and bioclimatic growing regions in RLP, a considerable amount of heterogeneity in the explanatory variables was introduced by quality restrictions in the remote sensing data. This was particularly true for the LiDAR derived canopy height information that was gathered in a time span of ten years around the date of the terrestrial inventory and revealed pronounced quality variations. With an adjusted R^2 of 0.48 and a RMSE_{cv} of 137 m^3/ha, the model accuracy was still very close to those found in similar studies (Maack et al, 2016). Our analyses strongly indicate that the acquisition of the auxiliary information close to the date of the terrestrial survey is a key factor in order to increase the model accuracy. We also expect the tree species information in the timber volume model to become even more relevant if the temporal synchronocity and the quality of the canopy height information is improved. An up-to-date canopy height model would also circumvent a stratification according to different LiDAR acquisition characteristics, lead to a more balanced dataset when stratifying for the main plot tree species and allow for incorporating information that can further explain the variation within each tree species group. With respect to the latter, information about the bioclimatic growing conditions, soil properties and the stand density on plot level are expected to further improve the model's predictive performance. Promising steps with respect to more up-to-date auxiliary information have already been made, as the topographic survey institution of RLP is currently processing a canopy height model from aerial imagery acquisitions for 2011 and 2012 covering the entire federal state. These aerial photography acquisitions will in the future be conducted in a two-year period, allowing to derive up-to-date canopy height information in the framework of future forest inventories. As the availability of countrywide imagery-based surface models has been increasing (Ginzler and Hobi, 2015), investigating the performance between areal and LiDAR derived canopy height models and their consequent predictive power in the frame of timber volume estimations (Ullah et al, 2017) are tasks for subsequent analysis. Additionally, availability of satellite data for tree species classification map production with respect to up-to-dateness and coverage has recently been increasing in the frame of the Sentinel-2 mission (ESA, 2017).

Acknowledgements We want to express our gratitude to Prof. H. Heinimann (Chair of Land Use Engineering, ETH Zurich) for supporting this study. We want to explicitly thank Dr. Johannes Stoffels from the Environmental Sensing and Geoinformatics Group of Trier University for providing the tree species classification map as well as for constructive discussions when it came to interpreting the results. Special gratitude is owed to the State Forest Service of Rhineland-Palatinate, in particular Dr. Joachim Langshausen, Jürgen Dietz and Claus-Andreas Lessander, for collaboration and providing the forest inventory and geodata. We also want to thank Kai Husmann and Christoph Fischer from the Northwest German Forest Research Institution Göttingen for their advice in processing the terrestrial inventory data, and Alexander Massey and Michael Hill for proofreading. We also want to thank two anonymous reviewers for their valuable support to improve the initial manuscript.

Conflict of Interest The authors declare that they have no conflict of interest.

References

Akaike H (2011) Akaike's Information Criterion, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 25–25. DOI 10.1007/978-3-642-04898-2_110, URL http://dx.doi.org/10.1007/978-3-642-04898-2_110

Bitterlich W (1984) The relascope idea. Relative measurements in forestry. Commonwealth Agricultural Bureaux

Bohlin J, Bohlin I, Jonzén J, Nilsson M (2017) Mapping forest attributes using data from stereophotogrammetry of aerial images and field data from the national forest inventory. SILVA FENNICA 51(2)

Breidenbach J, Astrup R (2012) Small area estimation of forest attributes in the norwegian national forest inventory. European Journal of Forest Research 131(4):1255–1267, DOI 10.1007/s10342-012-0596-7

Breidenbach J, Kublin E, McGaughey R, Andersen HE, Reutebuch SE (2008) Mixed-effects models for estimating stand volume by means of small footprint airborne laser scanner data. Photogrammetric Journal of Finland 21(1):4–15

Breiman L (2001) Random forests. Machine learning 45(1):5–32, DOI 10.1023/A:1010933404324

Brosofske KD, Froese RE, Falkowski MJ, Banskota A (2014) A review of methods for mapping and prediction of inventory attributes for operational forest management. Forest Science 60(4):733–756

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006) Measurement error in nonlinear models: a modern perspective. CRC press, DOI 10.1201/9781420010138, URL https://doi.org/10.1201/9781420010138

Congalton RG, Green K (2008) Assessing the accuracy of remotely sensed data: principles and practices. CRC press, DOI 10.1201/9781420055139, URL https://doi.org/10.1201/9781420055139

Deo RK, Froese RE, Falkowski MJ, Hudak AT (2016) Optimizing variable radius plot size and lidar resolution to model standing volume in conifer forests. Canadian Journal of Remote Sensing 42(5):428–442, DOI 10.1080/07038992.2016.1220826, URL http://dx.doi.org/10.1080/07038992.2016.1220826, http://dx.doi.org/10.1080/07038992.2016.1220826

Bundesministerium für Ernährung LuV (2011) Aufnahmeanweisung für die dritte Bundeswaldinventur BWI3 (2011 - 2012). URL https://www.bundeswaldinventur.de/index.php?id=421

ESA (2017) Sentinel-2 earth observation mission. URL http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2, accessed: 2017-03-29

Gauer J, Aldinger E (2005) Waldökologische Naturräume Deutschlands-Wuchsgebiete. Mitteilungen des Vereins für Forstliche Standortskunde und Forstpflanzenzüchtung 43:281–288

Ginzler C, Hobi ML (2015) Countrywide stereo-image matching for updating digital surface models in the framework of the swiss national forest inventory. Remote Sensing 7(4):4343–4370, DOI 10.3390/rs70404343, URL http://www.mdpi.com/2072-4292/7/4/4343

Gregoire TG, Valentine HT (2007) Sampling strategies for natural resources and the environment. CRC Press

Gustafson P (2003) Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments. CRC Press, DOI 10.1201/9780203502761, URL https://doi.org/10.1201/9780203502761

Hill A, Breschan J, Mandallaz D (2014) Accuracy assessment of timber volume maps using forest inventory data and lidar canopy height models. Forests 5(9):2253–2275, DOI 10.3390/f5092253, URL http://www.mdpi.com/1999-4907/5/9/2253

Hollaus M, Wagner W, Maier B, Schadauer K (2007) Airborne laser scanning of forest stem volume in a mountainous environment. Sensors 7(8):1559–1577, DOI 10.3390/s7081559, URL http://www.mdpi.com/1424-8220/7/8/1559

Husmann K, Rumpf S, Nagel J (2017) Biomass functions and nutrient contents of european beech, oak, sycamore maple and ash and their meaning for the biomass supply chain. Journal of Cleaner Production DOI 10.1016/j.jclepro.2017.03.019, URL https://doi.org/10.1016/j.jclepro.2017.03.019

Jakubowski MK, Guo Q, Kelly M (2013) Tradeoffs between lidar pulse density and forest measurement accuracy. Remote Sensing of Environment 130:245–253, DOI 10.1016/j.rse.2012.11.024, URL https://doi.org/10.1016/j.rse.2012.11.024

Koch B (2010) Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. ISPRS Journal of Photogrammetry and Remote Sensing 65(6):581–590, DOI 10.1016/j.isprsjprs.2010.09.001, URL https://doi.org/10.1016/j.isprsjprs.2010.09.001

Köhl M, Magnussen SS, Marchetti M (2006) Sampling methods, remote sensing and GIS multiresource forest inventory. Springer Science & Business Media

Kublin E (2003) Einheitliche beschreibung der schaftform–methoden und programme–bdatpro. Forstwissenschaftliches Centralblatt 122(3):183–200

Kublin E, Breidenbach J, Kändler G (2013) A flexible stem taper and volume prediction method based on mixed-effects b-spline regression. European journal of forest research 132(5-6):983–997, DOI 10.1007/s10342-013-0715-0

Lamprecht S, Hill A, Stoffels J, Udelhoven T (2017) A machine learning method for co-registration and individual tree matching of forest inventory and airborne laser scanning data. Remote Sensing 9(5), DOI 10.3390/rs9050505, URL http://www.mdpi.com/2072-4292/9/5/505

Latifi H, Nothdurft A, Koch B (2010) Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/lidar-derived predictors. Forestry 83(4):395–407, DOI 10.1093/forestry/cpq022, URL https://doi.org/10.1093/forestry/cpq022

Latifi H, Nothdurft A, Straub C, Koch B (2012) Modelling stratified forest attributes using optical/lidar features in a central european landscape. International Journal of Digital Earth 5(2):106–132, DOI 10.1080/17538947.2011.583992, URL http://dx.doi.org/10.1080/17538947.2011.583992

Liaw A, Wiener M (2002) Classification and regression by randomforest. R News 2(3):18–22, URL http://CRAN.R-project.org/doc/Rnews/

von Lüpke N, Saborowski J (2014) Combining double sampling for stratification and cluster sampling to a three-level sampling design for continuous forest inventories. European journal of forest research 133(1):89–100, DOI 10.1007/s10342-013-0743-9, URL http://dx.doi.org/10.1007/s10342-013-0743-9

Maack J, Lingenfelder M, Weinacker H, Koch B (2016) Modelling the standing timber volume of baden-württemberg—a large-scale approach using a fusion of landsat, airborne lidar and national forest inventory data. International Journal of Applied Earth Observation and Geoinformation 49:107–116, DOI 10.1016/j.jag.2016.02.004, URL https://doi.org/10.1016/j.jag.2016.02.004

Magnussen S, Næsset E, Gobakken T (2010) Reliability of lidar derived predictors of forest inventory attributes: A case study with norway spruce. Remote Sensing of Environment 114(4):700–712, DOI 10.1016/j.rse.2009.11.007, URL https://doi.org/10.1016/j.rse.2009.

11.007

Magnussen S, Mandallaz D, Breidenbach J, Lanz A, Ginzler C (2014) National forest inventories in the service of small area estimation of stem volume. Canadian Journal of Forest Research 44(9):1079–1090, DOI 10.1139/cjfr-2013-0448, URL https://doi.org/10.1139/cjfr-2013-0448

Mandallaz D (2008) Sampling techniques for forest inventories. CRC Press, DOI 10.1201/9781584889779, URL https://doi.org/10.1201/9781584889779

Mandallaz D (2013a) Design-based properties of some small-area estimators in forest inventory with two-phase sampling. Canadian Journal of Forest Research 43(5):441–449, DOI 10.1139/cjfr-2012-0381, URL https://doi.org/10.1139/cjfr-2012-0381

Mandallaz D (2013b) A three-phase sampling extension of the generalized regression estimator with partially exhaustive information. Canadian Journal of Forest Research 44(4):383–388, DOI 10.1139/cjfr-2013-0449, URL https://doi.org/10.1139/cjfr-2013-0449

Mandallaz D, Breschan J, Hill A (2013) New regression estimators in forest inventories with two-phase sampling and partially exhaustive information: a design-based monte carlo approach with applications to small-area estimation. Canadian Journal of Forest Research 43(11):1023–1031, DOI 10.1139/cjfr-2013-0181, URL https://doi.org/10.1139/cjfr-2013-0181

Mandallaz D, Hill A, Massey A (2016) Design-based properties of some small-area estimators in forest inventory with two-phase sampling - revised version. Tech. rep., Department of Environmental Systems Science, ETH Zurich, DOI 10.3929/ethz-a-010579388, URL https://doi.org/10.3929/ethz-a-010579388

Massey A, Mandallaz D (2015) Design-based regression estimation of net change for forest inventories. Canadian Journal of Forest Research 45(12):1775–1784, DOI 10.1139/cjfr-2015-0266, URL https://doi.org/10.1139/cjfr-2015-0266, https://doi.org/10.1139/cjfr-2015-0266

Massey A, Mandallaz D, Lanz A (2014) Integrating remote sensing and past inventory data under the new annual design of the swiss national forest inventory using three-phase design-based regression estimation. Canadian Journal of Forest Research 44(10):1177–1186, DOI 10.1139/cjfr-2014-0152, URL https://doi.org/10.1139/cjfr-2014-0152

Mathworks (2017) Matlab version 9.2.0.538062 (r2017a)

McRoberts RE, Tomppo EO, Næsset E (2010) Advances and emerging issues in national forest inventories. Scandinavian Journal of Forest Research 25(4):368–381, DOI 10.1080/02827581.2010.496739, URL http://dx.doi.org/10.1080/02827581.2010.496739

McRoberts RE, Næsset E, Gobakken T, Bollandsås OM (2015) Indirect and direct estimation of forest biomass change using forest inventory and airborne laser scanning data. Remote Sensing of Environment 164:36–42, DOI 10.1016/j.rse.2015.02.018, URL https://doi.org/10.1016/j.rse.2015.02.018

Naesset E (2014) Area-based inventory in norway – from innovation to an operational reality. In: Forest Applications of Airborne Laser Scanning - Concepts and Case Studies, Springer, chap 11, pp 216–240, DOI 10.1007/978-94-017-8663-8

Nink S, Hill J, Buddenbaum H, Stoffels J, Sachtleber T, Langshausen J (2015) Assessing the suitability of future multi-and hyperspectral satellite systems for mapping the spatial distribution of norway spruce timber volume. Remote Sensing 7(9):12,009–12,040, DOI 10.3390/rs70912009, URL http://www.mdpi.com/2072-4292/7/9/12009

Næsset E (1997) Estimating timber volume of forest stands using airborne laser scanner data. Remote Sensing of Environment 61(2):246 – 253, DOI https://doi.org/10.1016/S0034-4257(97)00041-2, URL http://www.sciencedirect.com/science/article/pii/S0034425797000412

R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/

Saborowski J, Marx A, Nagel J, Böckmann T (2010) Double sampling for stratification in periodic inventories—infinite population approach. Forest ecology and management 260(10):1886–1895, DOI 10.1016/j.foreco.2010.08.035, URL https://doi.org/10.1016/j.foreco.2010.08.035

Schreuder HT, Gregoire TG, Wood GB (1993) Sampling methods for multiresource forest inventory. John Wiley & Sons

Stoffels J, Hill J, Sachtleber T, Mader S, Buddenbaum H, Stern O, Langshausen J, Dietz J, Ontrup G (2015) Satellite-based derivation of high-resolution forest information layers for operational forest management. Forests 6(6):1982–2013, DOI 10.3390/f6061982, URL http://www.mdpi.com/1999-4907/6/6/1982

Straub C, Dees M, Weinacker H, Koch B (2009) Using airborne laser scanner data and cir orthophotos to estimate the stem volume of forest stands. Photogrammetrie-Fernerkundung-Geoinformation 2009(3):277–287, DOI 10.1127/0935-1221/2009/0022, URL https://doi.org/10.1127/0935-1221/2009/0022

von Thünen-Institut (2014) Dritte Bundeswaldinventur 2012. URL https://bwi.info, accessed: 2017-02-03

Tonolli S, Dalponte M, Vescovo L, Rodeghiero M, Bruzzone L, Gianelle D (2011) Mapping and modeling forest tree volume using forest inventory and airborne laser scanning. European Journal of Forest Research 130(4):569–577, DOI 10.1007/s10342-010-0445-5, URL http://dx.doi.org/10.1007/s10342-010-0445-5

Ullah S, Dees M, Datta P, Adler P, Koch B (2017) Comparing airborne laser scanning, and image-based point clouds by semi-global matching and enhanced automatic terrain extraction to estimate forest timber volume. Forests 8(6):215, DOI 10.3390/f8060215, URL http://www.mdpi.com/1999-4907/8/6/215

White JC, Coops NC, Wulder MA, Vastaranta M, Hilker T, Tompalski P (2016) Remote sensing technologies for enhancing forest inventories: A review. Canadian Journal of Remote Sensing 42(5):619–641, DOI 10.1080/07038992.2016.1207484, URL http://dx.doi.org/10.1080/07038992.2016.1207484

Zianis D, Muukkonen P, Mäkipää R, Mencuccini M, et al (2005) Biomass and stem volume equations for tree species in Europe. Silva Fennica Monographs 4