

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

¹ **Abstract** A timber-volume regression model applicable ⁶ available on the federal state level. As is common in many
² to the entire forest area of the federal German state of ⁷ forest inventory datasets, strong heterogeneity in the LiDAR
³ Rhineland-Palatinate is identified using a combination of ⁸ data due to different acquisition dates and misclassifications
⁴ airborne laser scanning (LiDAR)-derived metrics and infor- ⁹ in the tree species classification map had noticeable effects
⁵ mation from a satellite-based tree species classification map ¹⁰ on the regression model's performance. This article specif-
¹¹

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

¹² ¹³ ¹⁴ ¹⁵ ¹⁶ ¹⁷ ¹⁸ ¹⁹ ²⁰ ²¹ ²² ²³ 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 6% in adjusted R^2 and suggests the use of such information in forthcoming inventories. We also found that including LiDAR quality information as categorical variables within the regression model considerably mitigates issues with time lags between the LiDAR and terrestrial data acquisition and LiDAR quality variations (9% increase in adjusted R^2). The model achieved an adjusted R^2

of 0.49 (RMSE_{cv} of 132 m³/ha) under incorporation of the tree species and LiDAR quality information, and was thus improved by 14% (16 m³/ha) compared to the simple model only containing LiDAR height metrics (adjusted $R^2=0.35$, RMSE_{cv}=148 m³/ha).

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

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 recently become particularly attractive are so-called *double-sampling*

(Mandallaz, 2008) and *mapping* (Beaudoin et al, 2014) procedures. The core concept of these methods is to use predictions of the terrestrial target variable at additional sam-

ple 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. 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 As-trup, 2012; von Lüpke and Saborowski, 2014; Magnusson 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 pixelwise estimations of a target variable in high spatial resolution (Hill et al, 2014; Latifi et al, 2010; Nink et al, 2015; Tonolli et al, 2011; Van Aardt et al, 2008).

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 ad-

dressed 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 use 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). Straub et al (2009) and Latifi et al (2012) reported a notable gain in model performances when using stratification according to broadleaf and coniferous tree species on sample plot level in conjugation to canopy height metrics for timber volume estimations. One of the rare examples of using more species-specific information is Packalén and Maltamo (2006), who applied a separate prediction of the sample plot timber volume for Scots pine, Norway spruce and a deciduous-species group. However, further investigation is 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

130 a result, the available remote sensing data often exhibit no₁₅₇
 131 table time-lags with respect to the date of the terrestrial in₁₅₈
 132 vventory. This has often been addressed as a major drawback,₁₅₉
 133 especially for the application of model-assisted change esti₁₆₀
 134 mation ([Massey and Mandallaz, 2015](#)).₁₆₁

135 Our study is embedded in the current implementation of₁₆₂
 136 model-assisted regression estimators ([Mandallaz, 2013a,b](#),₁₆₃
 137 [Mandallaz et al, 2013](#)) for estimating the standing timber₁₆₄
 138 volume within the state and communal forest management₁₆₅
 139 units over the entire state of Rhineland-Palatinate (RLP, Ger₁₆₆
 140 many). With respect to this overall objective, the aim of this₁₆₇
 141 study was to derive an ordinary least square (OLS) regres₁₆₈
 142 sion model to generate predictions of the standing timber₁₆₉
 143 volume associated with a sample location of the Third Ger-

144 man National Forest Inventory (BWI3) over the entire fed₁₇₀
 145 eral state forest area (6155 km²). A merged LiDAR dataset
 146 from different acquisition years and a satellite-based tree

147 species classification map for the five main tree species in₁₇₂
 148 RLP was available for the entire inventory area and conse₁₇₃
 149 quently used to derive predictor variables. The major limit₁₇₄
 150 ing factors for using these data in a regression analysis are₁₇₅
 151 (i) variation in the LiDAR data quality as well as time-lags₁₇₆
 152 of up to 10 years between the LiDAR acquisitions and the₁₇₇
 153 terrestrial survey, (ii) misclassifications in the tree species₁₇₈
 154 classification map and (iii) the ambiguous choice of a suit₁₇₉
 155 able extraction area (*support*) for all remote sensing infor₁₈₀
 156 mation under angle count sampling in the terrestrial survey₁₈₁

(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 LiDAR- 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 km²) of the entire state area (19850 km²) 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%),

182 Douglas fir (*pseudotsuga menziesii*, 6.4%), European larch
 183 (*larch decidua*, 2.4%) and Silver fir (*abies alba*, 0.7%). The
 184 share of broadleaf tree species is 58.7%. The forests of
 185 RLP further exhibit heterogeneous structures (von Thünen-
 186 Institut, 2014): around 82% of the forest area in RLP are
 187 mixed forest stands (i.e. at least two different tree species
 188 occur in the same stand) and 69% of the forest area exhibit a
 189 multi-layered vertical structure. While the average tree age
 190 is around 80 years, most of the forest area (20%) is occupied
 191 by trees between 40 and 60 years of age, whereas 27% of
 192 the trees are older than 100 years. Spatially variable climate
 193 conditions have a strong influence on the local growth dy-
 194 namics as well as tree species composition and create a large
 195 variety of forest structures, ranging from characteristic oak
 208 coppices (Moselle valley), pure spruce, beech and scots pine
 196 forests (e.g. Hunsrück and Palatinate forest) to mixed forests
 209 comprising variable proportions of oak, larch, spruce, Scots
 197 pine and beech. Accordingly, RLP has been divided into 16
 210 bioclimatic growing regions that form homogeneous areas
 200 with respect to the afore mentioned characteristics (Gauer
 214 and Aldinger, 2005).

203 2.2 Terrestrial Inventory Data

204 The German National Forest Inventory is carried out over
 218 the entire forest area of Germany in reoccurring time peri-
 205 ods of 10 years. The most recent inventory (BWI3) has been
 206 conducted in the years 2011 and 2012. In this framework,
 207

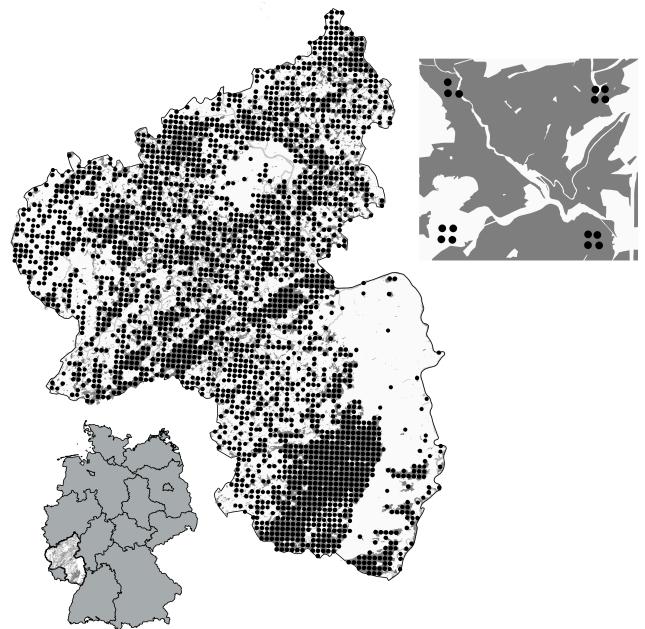


Fig. 1: Spatial distribution of the BWI3 cluster samples over Rhineland-Palatinate

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 technique²⁴⁵ was applied to 8092 sample plots (2810 clusters) in RLP, resulting in the collection of 56561 sample trees for which the²⁴⁶ *dbh*, the absolute tree height, the tree diameter at 7 m (*D7*)²⁴⁷ and the tree species were recorded. All plot center positions²⁴⁸ were determined with a differential GPS technique. In order²⁴⁹ to derive a volume estimation for each sample tree, the²⁵⁰ BWI3 estimates a taper curve for each sample tree by calibrating²⁵¹ the random effects term of linear mixed-effects taper²⁵² models with the set of diameters and corresponding height²⁵³ measurements taken from the respective sample tree (Kublin²⁵⁴ et al., 2013). The integration of the derived taper curves²⁵⁵ subsequently lead to a volume prediction for each sample tree.²⁵⁶ We restricted the sampling frame exclusively to the area of²⁵⁷ state and communal forest management units, which²⁵⁸ constitute 73% of the entire forest area of RLP (von Thünen-Institut,²⁵⁹ 2014). The dataset of this study hence comprised²⁶⁰ 5791 plots (2055 clusters). For this sample, the timber volume²⁶¹ density per hectare on plot level, $Y(x)$, was calculated²⁶² according to the formula of one-phase one-stage sampling²⁶³ (Mandallaz, 2008). The timber volume density per hectare²⁶⁴ on plot level was used as the response variable in the regression²⁶⁵ 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 $5 \times 5 \text{ m}^2$, while more recently acquired datasets contain more than 125 points per $5 \times 5 \text{ m}^2$ 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 $5 \times 5 \text{ m}^2$ 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

271 otherwise have been filled with unrealistic values in the inter-²⁹⁸
272 polation step. The thinned point clouds were aggregated to²⁹⁹
273 larger tiles in order to decrease the number of seamlines
274 in the final mosaic. The aggregated tiles were then interpo-
275 lated to raster images using a Delauney interpolation in the
276 Matlab software ([Mathworks, 2017](#)). The resulting two ele-
277 vation models were then used to calculate a canopy height
278 model (*CHM*) in raster format, providing discrete informa-
279 tion about the canopy surface height of the forest area in a³⁰⁰
280 spatial resolution of 5 meters.³⁰¹

281 As explanatory variables, the mean canopy height³⁰²
282 (*meanheight*) and the standard deviation (*stddev*) were cal-³⁰³
283 culated as the mean and standard deviation of all raster val-³⁰⁴
284 ues within a predefined square around each sample plot cen-³⁰⁵
285 ter. The square (i.e. *support* of the explanatory variable,³⁰⁶
286 see section [2.4](#)) was previously intersected with the state³⁰⁷
287 and communal forest area defined by a polygon mask and³⁰⁸
288 thereby corrected for edge effects at the forest border. The³⁰⁹
289 tree height is one prominent predictor variable in the taper³¹⁰
290 functions of the BWI3 that are used to calculate a timber vol-³¹¹
291 ume value for each sample tree ([Kublin, 2003; Kublin et al.,³¹²](#)
292 [2013](#)). A visual inspection of the tree volumes of all sample³¹³
293 trees collected in the BWI3 within RLP against their tree³¹⁴
294 heights also revealed the characteristic shape of an allomet-³¹⁵
295 ric relationship between these variables (Online Resource³¹⁶
296 1). It was hypothesized that this relationship on single-tree³¹⁷
297 level is also apparent on the aggregated level of a sample plot³¹⁸

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*).

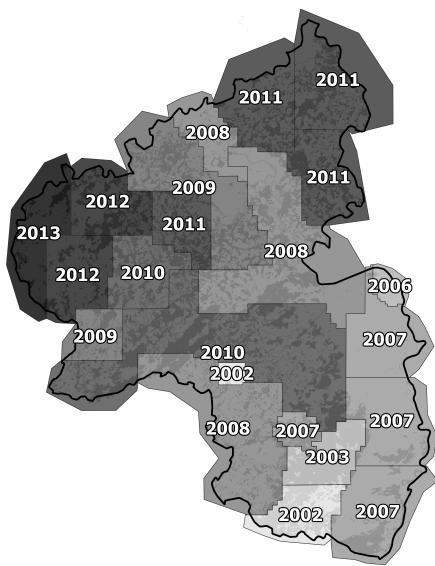


Fig. 2: Separate LiDAR acquisitions in Rhineland-Palatinate over the years. The colors also indicate the quality of the data: *light*: low point densities ($1/5 \times 5 m^2$), *dark*: high point densities ($>100/5 \times 5 m^2$)

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 regres-

357 sion coefficients in the direction of zero due to an artificial³⁸⁴
 358 introduction of noise (Carroll et al, 2006). This can cause an³⁸⁵
 359 inflation of the residual variance and a consequent decrease³⁸⁶
 360 of the model accuracy (Magnussen et al, 2010). In case of³⁸⁷
 361 classification the impacts of misclassifications on the model³⁸⁸
 362 properties are even harder to predict (Gustafson, 2003)³⁸⁹
 363 While errors in the explanatory variables do not affect the³⁹⁰
 364 unbiasedness of the estimators in the model-assisted frame³⁹¹
 365 work, a reduction or elimination of the classification errors³⁹²
 366 could provide an improvement of the regression model ac-³⁹³
 367 curacy and thereby potentially lead to smaller prediction and³⁹⁴
 368 estimation errors. We therefore addressed the effect of mis-³⁹⁵
 369 classifications in the *treespecies* variable by the following³⁹⁶
 370 analysis:

- 371 a) we investigated the effect on the regression model³⁹⁸
 372 performance (regression coefficients, model accuracy)³⁹⁹
 373 when substituting the *predicted* by the *actual* main plot
 374 tree species derived from the sampled trees of the re-⁴⁰⁰
 375 spective sample plot under identical threshold settings⁴⁰¹
- 376 b) we used the random forest algorithm (Breiman, 2001)⁴⁰²
 377 Liaw and Wiener, 2002) in the statistical software R (R⁴⁰³
 378 Core Team, 2016) to define a *calibration model* in order⁴⁰⁴
 379 to improve the classification accuracy of the initially pre-⁴⁰⁵
 380 dicted main plot tree species, correct for potential sys-⁴⁰⁶
 381 tematic misclassifications and thus minimize the effect⁴⁰⁷
 382 of misclassifications on the regression model. The ran-⁴⁰⁸
 383 dom forest algorithm is a machine learning algorithm⁴⁰⁹

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,

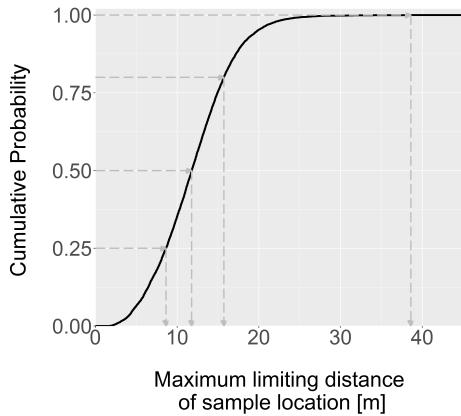
⁴¹⁰ 2007), i.e. the spatial extent around the plot center in which
⁴¹¹ the auxiliary information is evaluated and transformed into
⁴¹² an explanatory variable. It has widely been hypothesized
⁴¹³ that the best relationship between the target variable on the
⁴¹⁴ ground and any explanatory variable derived from the auxil-
⁴¹⁵ iary information is obtained if the support is spatially iden-
⁴¹⁶ tical to the sample plot extent. In case of angle count sam-
⁴¹⁷ pling, an individual extent for each sample plot can be ap-
⁴¹⁸ proximated by regarding the maximum limiting distance
⁴¹⁹ of its sample trees as the outer plot radius. However, many
⁴²⁰ model-assisted applications under double-sampling do not
⁴²¹ allow for a between-plot change of the support for a specific
⁴²² explanatory variable (Mandallaz, 2013a,b).

⁴²³ For this reason, the task is to find a unique support
⁴²⁴ for each auxiliary information that leads to the best over-
⁴²⁵ all model accuracy. Deo et al (2016) conducted extensive
⁴²⁶ analysis to identify optimal supports for modelling standing
⁴²⁷ timber volume for *variable-radius plot* designs in conifer
⁴²⁸ forests. They analysed 24 different radii (i.e. circular sup-
⁴²⁹ ports) in which they extracted 57 metrics from a LiDAR⁴⁵⁶
⁴³⁰ derived point cloud with an average point density of 18
⁴³¹ pulses per square meter. They successively evaluated the
⁴³² prediction performance of each support size by using the⁴⁵⁸
⁴³³ LiDAR metrics in a random forest algorithm and com-⁴⁵⁹
⁴³⁴ paring the resulting model accuracies. In order to identify⁴⁶⁰
⁴³⁵ the best-performing supports for our explanatory variables,⁴⁶¹
⁴³⁶ we followed a similar approach. The explanatory variables⁴⁶²

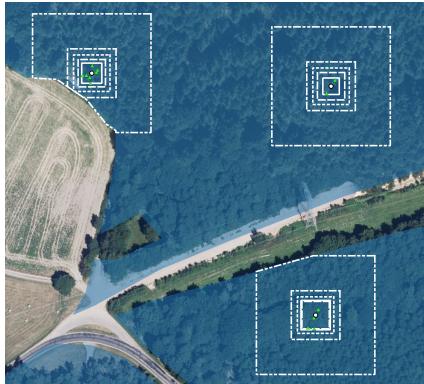
were calculated using *individual* (i.e. plot-varying) supports
⁴³⁷ (*ind*), i.e. an individual support extent was used for each
⁴³⁸ plot according to the maximum limiting distance of all sam-
⁴³⁹ ple trees associated to the respective sample plot. We then
⁴⁴⁰ compared the model accuracies achieved by the individual
⁴⁴¹ supports against the model accuracies from a set of *fixed*
⁴⁴² (i.e. non plot-varying) supports. The extents of the fixed sup-
⁴⁴³ ports were chosen from the cumulative distribution function
⁴⁴⁴ (ECDF) of the maximum limiting distances of all 5791 sam-
⁴⁴⁵ ple plots of the analysed forest area (Fig. 3a). We considered
⁴⁴⁶ the 25th (q_{25} , 9 meters), 50th (q_{50} , 12 meters), 80th (q_{80} , 15
⁴⁴⁷ meters) and the 100th (q_{100} , 38 meters) percentiles, result-
⁴⁴⁸ ing in support side lengths of 18, 24, 30 and 76 meters (Fig.
⁴⁴⁹ 3). In contrast to the approach of Deo et al (2016), the the-
⁴⁵⁰ory of the model-assisted regression estimators by Mandal-
⁴⁵¹laz (2013a,b) required the explanatory variables to be calcu-
⁴⁵²lated using supports that theoretically allow for a tessellation
⁴⁵³of the entire forest area (i.e. rectangular supports). Based on
⁴⁵⁴the underlying theory the use of different support sizes for
⁴⁵⁵each explanatory variable is however perfectly valid.

2.5 Model Validation

In order to judge the quality of the *treespecies* variable,
⁴⁵⁶ the user's accuracy for each classified species and the
⁴⁵⁷ overall accuracy of the classification scheme was calcu-
⁴⁵⁸lated based on the confusion matrix (Congalton and Green,
⁴⁵⁹ 2008), using the main plot tree species calculated from



(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

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 $\text{meanheight}:\text{treespecies}$, $\text{meanheight}^2:\text{treespecies}$, $\text{meanheight}:\text{lidaryear}$, $\text{stddev}:\text{lidaryear}$ and $\text{meanheight}:\text{stddev}$ 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.

$$\begin{aligned}
 Y(x) = & \beta_0 + \beta_1 * \text{meanheight} + \beta_2 * \text{meanheight}^2 + \\
 & \beta_3 * \text{stddev} + \\
 & \beta_4 * \text{lidaryear}_1 + \dots + \beta_{12} * \text{lidaryear}_9 + \\
 & \beta_{13} * \text{treespecies}_1 + \dots + \beta_{18} * \text{treespecies}_6 + e(x)
 \end{aligned} \tag{1}$$

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

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.

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.

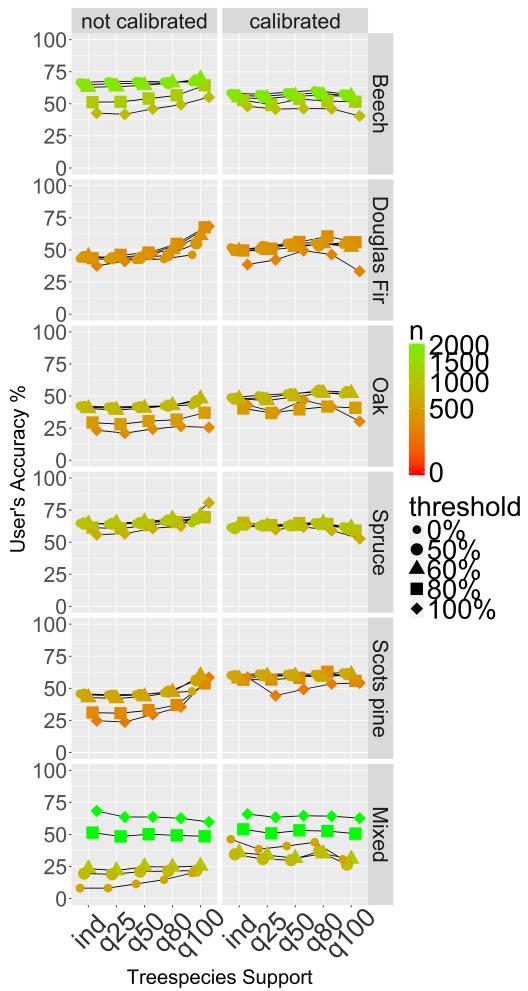
3 Results

3.1 Classification Accuracies

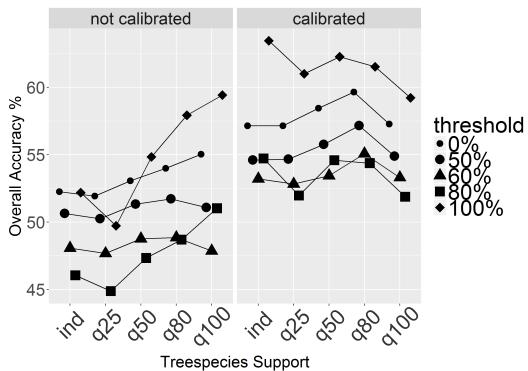
Effect of Support Size and Threshold

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

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.



(a)



(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 thresh-

old 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 $R^2=0.49$ and RMSE_{cv}=132 m³/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

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 coef-

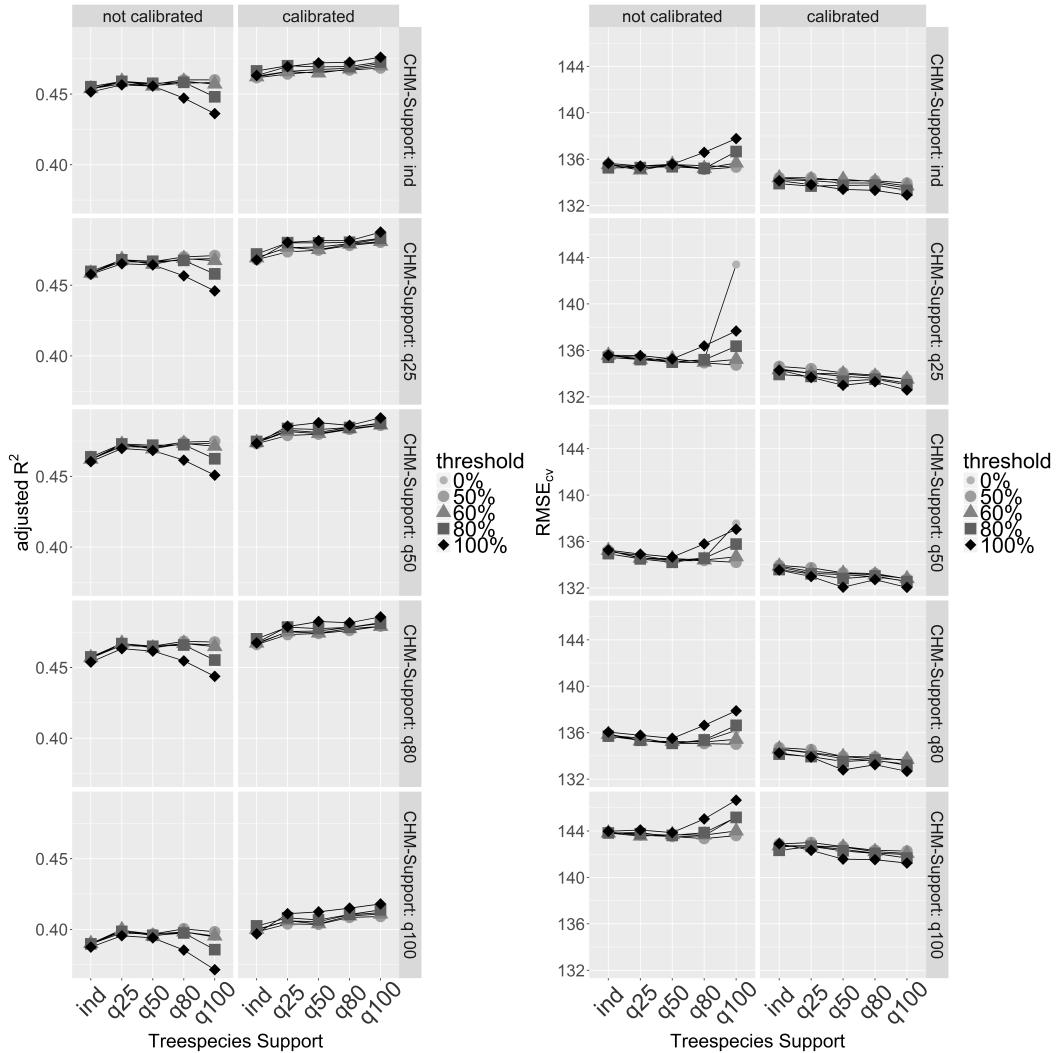


Fig. 5: 10-fold RMSE_{cv} and adjusted R^2 realized under various support choices for the CHM and *treespecies* explanatory variables

ficients 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 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)

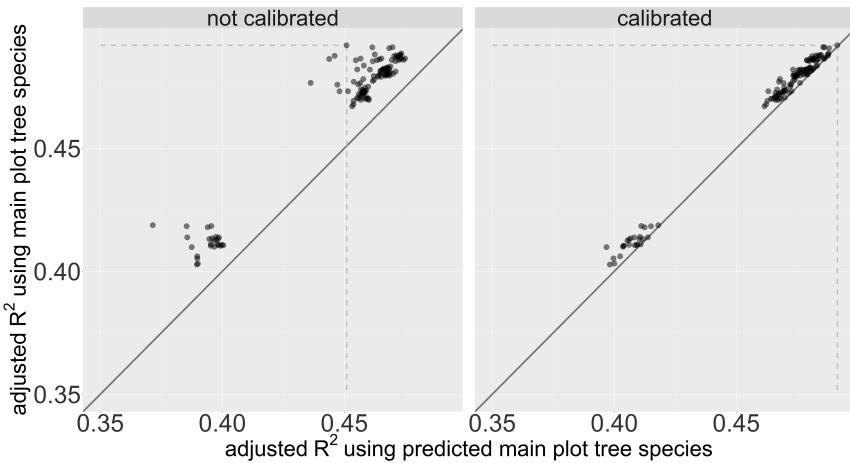


Fig. 6: Effect on the adjusted R^2 when substituting the actual main tree species with the predicted main tree species of a sample plot. The differentiation into two distinct point clouds results from the poor model performance under support size q_{100} for the CHM variables (i.e. the *lower* point cloud). The *dotted* line tracks the the model with the highest adjusted R^2 under the use of the error-free *treespecies* variable

turned out not to have any influence on the model accuracy⁶³⁸ and was thus dropped, resulting in an adjusted R^2 of 0.49⁶³⁹ and RMSE_{cv} of 132.12 m²/ha.

626 Interpretation of Final Regression Model

Figure 7 provides a visualisation of the tree species prediction functions separated by the LiDAR acquisition years.⁶⁴⁵ 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

654 meters and $stddev$ of 4 meters). However, we chose not to
 655 use a log-transformation of the response variable. Doing so
 656 would have prevented the subsequent calculation of the g-
 657 weight variance of the model assisted estimators (Mandal-
 658 laz, 2013a; Mandallaz et al, 2013), which is only possible
 659 for response variables on the original scale.

660 Effect of Time-Lags and Heterogeneity in LiDAR Data

661 Incorporating the LiDAR acquisition year as a categorical
 662 variable (*lidaryear*) in the regression model substantially
 663 accounted for the variability in the data introduced by *a)*
 664 the time-lags between LiDAR acquisition and terrestrial sur-
 665 vey, and *b)* variation in LiDAR data quality which are due
 666 to sensor- and post processing techniques (table 2). Whereas
 667 the adjusted R^2 for the regression model without considering
 668 the LiDAR acquisition year as additional predictor variable
 669 was 0.35 (0.41 including the tree species variable), the strat-
 670 ification by the LiDAR acquisition year led to adjusted R^2 of
 671 0.44 (0.48), thereby increasing the proportion of explained
 672 variance by up to 9%.

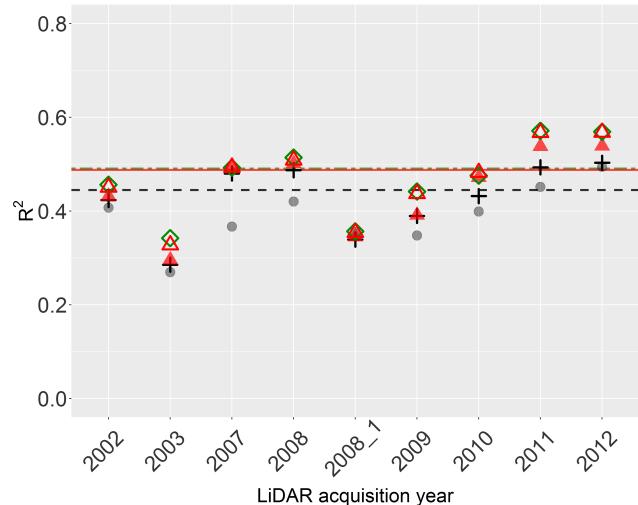


Fig. 8: R^2 of the final regression model achieved *within* the LiDAR acquisition year strata. Grey points: R^2 of submodel 1 (no stratification according to LiDAR acquisition year or tree species). Crosses: R^2 of submodel 2 (*without* tree species stratification). Filled triangles: R^2 of final model 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

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

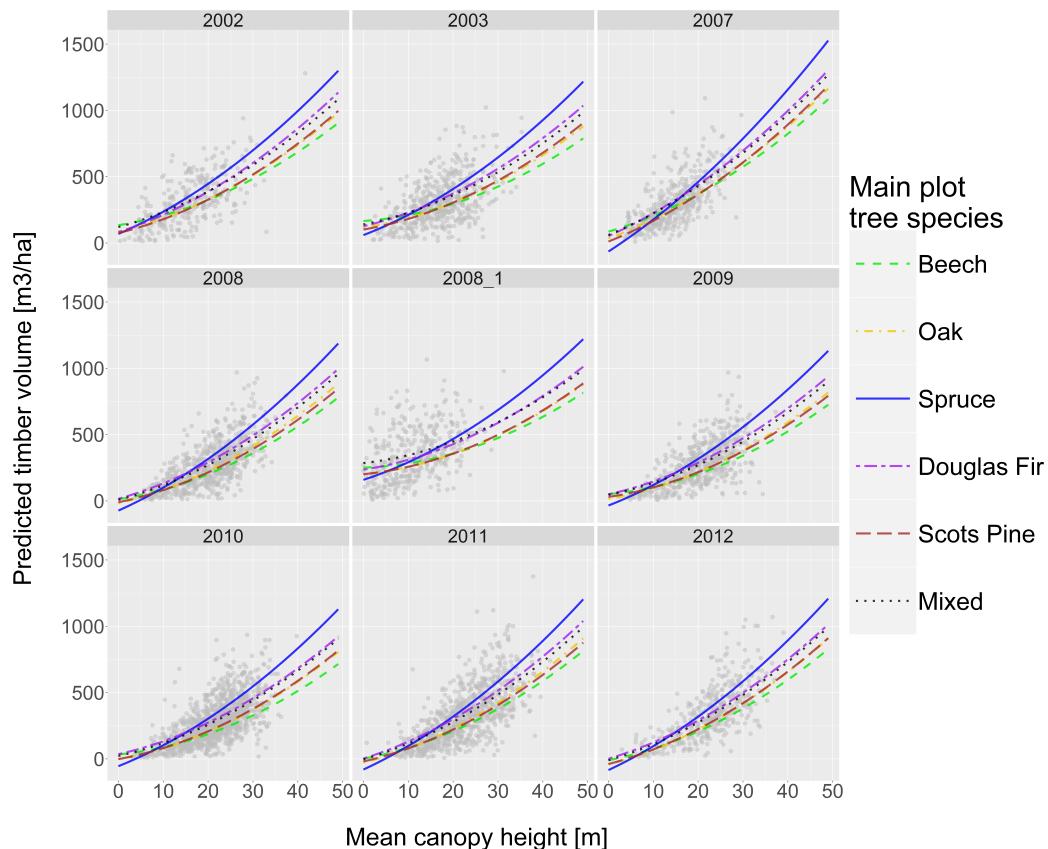


Fig. 7: Visualization of the timber volume prediction function (*final regression model*) on sample plot level for each main plot tree species and LiDAR acquisition year. For visualization purposes, the predictor variable *stddev* was set to its average value within the respective *treespecies* and *lidaryear*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

only the LiDAR height metrics as predictor variables (sub-⁶⁹² trial inventory (R^2 of 0.57 in 2011 and 2012, figure 8). The model 1). In some LiDAR acquisition year strata (i.e. 2007,₆₉₃ analysis illustrated once more that misclassifications in the 2008), this increase in R^2 even reached 8% - 13%. The ac-₆₉₄ tree species variable generally reduce model accuracy com-
685 curacies for the final model are also given in table 1.₆₉₅ pared to using error-free tree species information. The resid-

Added Value of Tree Species Map Information
686 Using the predicted main tree species of a sample plot as₆₉₆ ual inflations caused by the misclassifications in the uncali-
687 an additional categorical variable yielded a further increase₆₉₇ brated *treespecies* variable within the *lidaryear* strata were
688 of the model accuracy of 6% in the R^2 (table 2). This im-₆₉₈ up to 5%. However, the calibration was able to substantially
689 provement was particularly pronounced in LiDAR acquisi-₆₉₉ decrease or even remove the effects of misclassifications on
690 tion years that are close or identical to the year of the terres-₇₀₀ the model accuracy in all LiDAR acquisition year strata.

Table 1: R^2 , RMSE and Residual Square Sum (SSE) of final regression⁷¹⁵ model within LiDAR acquisition year strata (*lidaryear*). n : number of⁷¹⁶ validation data⁷¹⁷

LiDAR acquisition year	R^2	rmse	SSE	n
2012	0.57	135.20	7073596	387
2011	0.57	136.41	15779849	848
2010	0.48	119.57	16199324	1133
2009	0.44	122.87	8077013	535
2008	0.51	121.15	9936203	677
2008_1	0.35	158.76	9678912	384
2007	0.49	127.71	6736041	413
2003	0.33	142.24	10521254	520
2002	0.45	135.08	5638379	309

quisition 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 6% 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 14%. A differentiated evaluation revealed that the R^2 within LiDAR acquisitions year strata identical with the year of the terrestrial survey were much higher than those of previous acquisition years, showing differences of up to 22% between the R^2 's (0.35 compared to 0.57, table 2). The gain in the R^2 and the prediction performance when including the tree species information was also largest in combination with LiDAR information acquired in the year of the terrestrial inventory. These insights were also particularly interesting with respect to the further use of the regression model for small area estimations. Small area es-

701 4 Discussion

702 4.1 Stratification according to Tree Species and LiDAR

703 Acquisitions

704 Incorporating the main tree species of a sample location in⁷³¹
 705 the timber volume regression model significantly increased⁷³²
 706 the model accuracy and revealed strong evidence for the ex-⁷³³
 707 istence of a tree species specific behaviour concerning tim-⁷³⁴
 708 ber volume on the plot level. This result seems reasonable⁷³⁵
 709 regarding the species specific taper functions on single-tree⁷³⁶
 710 level applied in the BWI3 (Kublin, 2003; Kublin et al, 2013).⁷³⁷

711 Further evidence and specification of the tree species effects⁷³⁸
 712 on sample plot level - up to modeling individual tree species⁷³⁹
 713 - would be desirable. However, this was not possible in our⁷⁴⁰
 714 study because the stratification according to the LiDAR ac-

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	0.49	132.12
meanheight + stddev + meanheight ² + meanheight:stddev	submodel 1	5	0.35	148.03
meanheight + stddev + meanheight ² + lidaryear + meanheight:lidaryear + meanheight:stddev + stddev:lidaryear	submodel 2	29	0.44	137.52
meanheight + stddev + meanheight ² + treespecies + meanheight:treespecies + meanheight:stddev	submodel 3	15	0.41	137.52

742 estimators generally gain modeling strength by defining the
 743 prediction model *globally* (i.e. using all data in the inventory
 744 area), and then applying the so-derived prediction model to
 745 a subset of observations located within the area of interest
 746 ([Mandallaz et al, 2016](#)). Consequently, the proposed stratifi-
 747 cation technique in the prediction model could be expected
 748 to yield a gain in model accuracy and a reduction of the
 749 small area estimation errors if the small area domain mostly
 750 includes data from strata that have high within-strata model
 751 accuracies. This hypothesis is subject to ongoing analysis.
 752

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 trees⁷⁹⁰ 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 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. It has already been indicated that using predictor variables derived from auxiliary data of much higher spatial resolution than those used in this study (e.g. based on individual tree detections) will require supports that correspond not only to the actual spatial extent, but also to the exact position of the sample locations (Lamprecht et al., 2017).

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 vari-

ables. 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 0.49 and a RMSE_{cv} of $132 \text{ m}^3/\text{ha}$, 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 synchronicity 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 information⁸⁸⁶ have already been made, as the topographic survey institu⁸⁸⁷ tion of RLP is currently processing a canopy height model⁸⁸⁸ from aerial imagery acquisitions for 2011 and 2012 cover⁸⁸⁹ ing the entire federal state. These aerial photography acqui⁸⁹⁰ sitions will in the future be conducted in a two-year period,⁸⁹¹ allowing to derive up-to-date canopy height information in⁸⁹² the framework of future forest inventories. As the availabil⁸⁹³ ity of countrywide imagery-based surface models has been⁸⁹⁴ increasing (Ginzler and Hobi, 2015), investigating the per⁸⁹⁵ formance between areal and LiDAR derived canopy height⁸⁹⁶ models and their consequent predictive power in the frame⁸⁹⁷ of timber volume estimations (Ullah et al, 2017) are tasks⁸⁹⁸ for subsequent analysis. Additionally, availability of satel⁸⁹⁹

lite data for tree species classification map production with⁹⁰⁰ respect to up-to-dateness and coverage has recently been in⁹⁰¹ creasing in the frame of the Sentinel-2 mission (ESA, 2017).⁹⁰²

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.

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
- Beaudoin A, Bernier P, Guindon L, Villemaire P, Guo X, Stinson G, Bergeron T, Magnussen S, Hall R (2014) Mapping attributes of Canada's forests at moderate resolution through k nn and modis imagery. Canadian Journal of Forest Research 44(5):521–532, DOI 10.1139/cjfr-2013-0401, URL <https://doi.org/10.1139/cjfr-2013-0401>

- 915 Bitterlich W (1984) The relascope idea. Relative measure-⁹⁴² bundeswaldinventur.de/index.php?id=421
- 916 ments in forestry. Commonwealth Agricultural Bureaux ⁹⁴³
- 917 Breidenbach J, Astrup R (2012) Small area estimation of ⁹⁴⁴
- 918 forest attributes in the norwegian national forest inven-⁹⁴⁵
- 919 tory. European Journal of Forest Research 131(4):1255–⁹⁴⁶
- 920 1267, DOI 10.1007/s10342-012-0596-7 ⁹⁴⁷
- 921 Breiman L (2001) Random forests. Machine learning ⁹⁴⁸
- 922 45(1):5–32, DOI 10.1023/A:1010933404324 ⁹⁴⁹
- 923 Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM⁹⁵⁰
- 924 (2006) Measurement error in nonlinear models: a modern ⁹⁵¹
- 925 perspective. CRC press, DOI 10.1201/9781420010138,⁹⁵²
- 926 URL <https://doi.org/10.1201/9781420010138> ⁹⁵³
- 927 Congalton RG, Green K (2008) Assessing the accuracy ⁹⁵⁴
- 928 of remotely sensed data: principles and practices. CRC⁹⁵⁵
- 929 press, DOI 10.1201/9781420055139, URL <https://doi.org/10.1201/9781420055139> ⁹⁵⁶
- 930 Deo RK, Froese RE, Falkowski MJ, Hudak AT (2016)⁹⁵⁸
- 932 Optimizing variable radius plot size and lidar reso-⁹⁵⁹
- 933 lution to model standing volume in conifer forests.⁹⁶⁰
- 934 Canadian Journal of Remote Sensing 42(5):428–442,⁹⁶¹
- 935 DOI 10.1080/07038992.2016.1220826, URL <http://dx.doi.org/10.1080/07038992.2016.1220826>, ⁹⁶²
- 936 <http://dx.doi.org/10.1080/07038992.2016.1220826>. ⁹⁶³
- 938 Bundesministerium für Ernährung LuV (2011) Auf-⁹⁶⁶
- 940 nahmeanweisung für die dritte Bundeswaldinven-⁹⁶⁷
- 941 tur BWI3 (2011 - 2012). URL <https://www.bundeswaldinventur.de/index.php?id=421> ⁹⁶⁸
- 943 ESA (2017) Sentinel-2 earth observation mission. URL http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2, accessed: 2017-03-29
- 947 Gauer J, Aldinger E (2005) Waldökologische Naturräume Deutschlands-Wuchsgebiete. Mitteilungen des Vereins für Forstliche Standortskunde und Forstpflanzenzüchtung 43:281–288
- 951 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>
- 955 Gregoire TG, Valentine HT (2007) Sampling strategies for natural resources and the environment. CRC Press
- 959 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>
- 963 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>
- 967 Hollaus M, Wagner W, Maier B, Schadauer K (2007) Airborne laser scanning of forest stem volume in a mountain-

- ous environment. Sensors 7(8):1559–1577, DOI 10.3390/995
s7081559, URL <http://www.mdpi.com/1424-8220/996>
- 7/8/1559 997
- Husmann K, Rumpf S, Nagel J (2017) Biomass functions⁹⁹⁸
and nutrient contents of european beech, oak, sycamore⁹⁹⁹
maple and ash and their meaning for the biomass¹⁰⁰⁰
supply chain. Journal of Cleaner Production DOI 10.1007/
1016/j.jclepro.2017.03.019, URL <https://doi.org/1002>
[10.1016/j.jclepro.2017.03.019](https://doi.org/10.1016/j.jclepro.2017.03.019) 1003
- Jakubowski MK, Guo Q, Kelly M (2013) Tradeoffs be¹⁰⁰⁴
tween lidar pulse density and forest measurement ac¹⁰⁰⁵
curacy. Remote Sensing of Environment 130:245–253¹⁰⁰⁶
DOI 10.1016/j.rse.2012.11.024, URL [https://doi.org/10.1016/j.rse.2012.11.024](https://doi.org/1007) 1008
- Koch B (2010) Status and future of laser scanning, synthetic¹⁰⁰⁹
aperture radar and hyperspectral remote sensing data for¹⁰¹⁰
forest biomass assessment. ISPRS Journal of Photogramm¹⁰¹¹
etry and Remote Sensing 65(6):581–590, DOI 10.1012/
1016/j.isprsjprs.2010.09.001, URL <https://doi.org/1013>
[10.1016/j.isprsjprs.2010.09.001](https://doi.org/10.1016/j.isprsjprs.2010.09.001) 1014
- Köhl M, Magnussen SS, Marchetti M (2006) Sampling¹⁰¹⁵
methods, remote sensing and GIS multiresource forest in¹⁰¹⁶
ventory. Springer Science & Business Media 1017
- Kublin E (2003) Einheitliche beschreibung der¹⁰¹⁸
schaftform–methoden und programme–bdatpro¹⁰¹⁹
Forstwissenschaftliches Centralblatt 122(3):183–200 1020
1021
- Kublin E, Breidenbach J, Kändler G (2013) A flexi-
ble stem taper and volume prediction method based
on mixed-effects b-spline regression. European jour-
nal 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 ma-
chine learning method for co-registration and individual
tree matching of forest inventory and airborne laser scan-
ning 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 pre-
diction 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) Mod-
elling stratified forest attributes using optical/lidar fea-
tures 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

- 1022 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>
- 1023
- 1024
- 1025
- 1026 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>
- 1027
- 1028
- 1029
- 1030
- 1031
- 1032
- 1033
- 1034 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>
- 1035
- 1036
- 1037
- 1038
- 1039
- 1040 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>
- 1041
- 1042
- 1043
- 1044
- 1045
- 1046 Mandallaz D (2008) Sampling techniques for forest inventories. CRC Press, DOI 10.1201/9781584889779, URL <https://doi.org/10.1201/9781584889779>
- 1047
- 1048
- 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>

- 1076 doi.org/10.1139/cjfr-2015-0266, <https://doi.org/10.1139/cjfr-2015-0266> 1102
1077 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>
- 1078 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. *Candinian Journal of Forest Research* 44(10):1177–1186, DOI 10.1139/cjfr-2014-0152, URL <https://doi.org/10.1139/cjfr-2014-0152> 1109
1083 Packalén P, Maltamo M (2006) Predicting the plot volume by tree species using airborne laser scanning and aerial photographs. *Forest Science* 52(6):611–622
- 1084 Mathworks (2017) Matlab version 9.2.0.538062 (r2017a) 1111
1085 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> 1115
1090 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/>
- 1091 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> 1122
1097 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 1123
1100 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* 1127
1101 1128

- 1129 6(6):1982–2013, DOI 10.3390/f6061982, URL [http://¹¹⁵⁶org/10.14358/PERS.74.8.1033](http://org/10.14358/PERS.74.8.1033)
- 1130 www.mdpi.com/1999-4907/6/6/1982 ¹¹⁵⁷
- 1131 Straub C, Dees M, Weinacker H, Koch B (2009) Using air¹¹⁵⁸
1132 borne laser scanner data and cir orthophotos to estimate¹¹⁵⁹
1133 the stem volume of forest stands. Photogrammetrie¹¹⁶⁰
1134 Fernerkundung-Geoinformation 2009(3):277¹¹⁶¹
1135 287, DOI 10.1127/0935-1221/2009/0022, URL¹¹⁶²
1136 <https://doi.org/10.1127/0935-1221/2009/0022>¹¹⁶³
- 1137 von Thünen-Institut (2014) Dritte Bundeswaldinventur¹¹⁶⁴
1138 2012. URL <https://bwi.info>, accessed: 2017-02-03¹¹⁶⁵
- 1139 Tonolli S, Dalponte M, Vescovo L, Rodeghiero M, Bruzzone
1140 L, Gianelle D (2011) Mapping and modeling forest tree
1141 volume using forest inventory and airborne laser scan-
1142 ning. European Journal of Forest Research 130(4):569–
1143 577, DOI 10.1007/s10342-010-0445-5, URL [http://¹¹⁶⁶
1144 dx.doi.org/10.1007/s10342-010-0445-5](http://dx.doi.org/10.1007/s10342-010-0445-5)
- 1145 Ullah S, Dees M, Datta P, Adler P, Koch B (2017) Compar-
1146 ing airborne laser scanning, and image-based point clouds
1147 by semi-global matching and enhanced automatic ter-
1148 rain extraction to estimate forest timber volume. Forests
1149 8(6):215, DOI 10.3390/f8060215, URL [http://¹¹⁶⁷
1150 mdpi.com/1999-4907/8/6/215](http://www.)
- 1151 Van Aardt J, Wynne R, Scrivani J (2008) Lidar-based map-
1152 ping of forest volume and biomass by taxonomic group
1153 using structurally homogenous segments. Photogrammet-
1154 ric Engineering & Remote Sensing 74(8):1033–1044,
1155 DOI 10.14358/PERS.74.8.1033, URL [https://doi.¹¹⁶⁸](https://doi.)
- 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/¹¹⁶⁹10.1080/07038992.2016.1207484](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