Using canopy heights from digital aerial photogrammetry to enable spatial transfer of forest attribute models: a case study in central Europe

Christoph Stepper[1,2 *], Christoph Straub[1], Markus Immitzer[3] and Hans Pretzsch[2]

[1] *Bavarian State Institute of Forestry (LWF), Department of Information Technology, Research Group: Remote Sensing, Hans-Carl-von-Carlowitz-Platz 1, D-85354 Freising; E-Mail: Christoph.Straub@lwf.bayern.de (C.Str.)*

[2] *Chair for Forest Growth and Yield Science, Faculty of Forest Science and Resource Management, Technische Universität München, Hans-Carl-von-Carlowitz-Platz 2, D-85354 Freising, E-Mail: Hans.Pretzsch@lrz.tu-muenchen.de (H.P.)*

[3] *Institute of Surveying, Remote Sensing and Land Information (IVFL), University of Natural Resources and Life Sciences, Vienna (BOKU), Peter-Jordan-Straße 82, A-1190 Vienna, Austria; E-Mail: markus.immitzer@boku.ac.at (M. I.)*

***** Author to whom correspondence should be addressed; E-Mail: c.stepper@mytum.com (C.Ste.)

Using canopy heights from digital aerial photogrammetry to enable spatial transfer of forest attribute models: a case study in central Europe

This paper describes a workflow utilizing detailed canopy height information derived from digital airphotos combined with ground inventory information gathered in state-owned forests and regression modelling techniques to quantify forest growing stocks in private woodlands, for which little information is generally available. Random forest models were trained to predict three different variables at the plot level: quadratic mean diameter of the 100 largest trees (d_{100}), basal area weighted mean height of the 100 largest trees (h_{100}), and gross volume (V). Two separate models were created – one for a spruce- and one for a beech-dominated test site. We examined the spatial portability of the models by using them to predict the aforementioned variables at actual inventory plots in nearby forests, in which simultaneous ground sampling took place. When data from the full set of available plots were used for training, the predictions for d_{100}, h_{100}, and V achieved out-of-bag model accuracies (scaled RMSEs) of 15.1%, 10.1% and 35.3% for the spruce- and 15.9%, 9.7%, and 32.1% for the beech-dominated forest, respectively. The corresponding independent RMSEs for the nearby forests were 15.2%, 10.5%, and 33.6% for the spruce- and 15.5%, 8.9%, and 33.7% for the beech-dominated test site, respectively.

Running Head: Aerial photogrammetry for attribute model transfer

Keywords: remote sensing; digital aerial photogrammetry; semi-global matching; forest inventory; area-based approach; random forest; private forests

Introduction

Private forest owners play a key role in sustainable forest management in Europe, especially with respect to developing stable climate-adapted forests and supplying timber to forest-based industries (MCPFE 2003). When discussing strategies for mitigating climate change, ensuring biodiversity protection, or enhancing rural development, policy makers must be aware of the importance of privately owned forests. According to the latest report on the state of Europe's forests, private forest owners hold 51% of the total forest area in Europe (FOREST EUROPE 2015). However, these percentages vary across countries. Particularly in northern Europe (Finland, Sweden, Norway), but also in western and central European countries (France, Austria, Slovenia), private forest properties sum up to more than two thirds of the total forest area (Schmithüsen & Hirsch 2010). In Germany, 48% of the forest area is privately owned, with significant regional variation due to historic developments (Thünen-Institut 2016). Especially in those regions with large amounts of private forest in the northern, western and central European countries, the vast majority of properties are small (often < 10 ha; FOREST EUROPE 2015) and managed by family members, most of whom are not forestry professionals.

In the aforementioned countries, the most common instrument used by the state authorities to influence private forest management is subsidies which are payed, e.g., for conservation of forest biodiversity, silvicultural activities, forest inventory and planning (Rametsteiner & Sotirov 2015). The need to support management activities in the privately owned forestry sector is greater than ever due to the ever-increasing demand for roundwood. Also important is the fact that owners of small forest properties today are less motivated than owners of large forest properties to sell timber due to both the low personal economic significance of such income and the high transaction costs

involved (Verkerk et al. 2011). This discrepancy creates a gap between theoretically available wood in European forests and the actual amount of timber supplied to domestic markets. Felling rates (i.e., proportions of increments utilized by fellings) in the northern, western and central European countries range between 62% and 79%. Köhl et al. (2015) argue that it would be possible to mobilize a substantial amount of timber by increased fellings without violating sustainability constraints in those countries. However, in order to promote the mobilisation of wood and at the same time to support small private forest owners in sustainable forest management, detailed information about the current stocking within these forest properties is needed.

Traditionally, ground-based forest inventories are conducted to acquire measurements describing the current state of forests. These data serve as the basis for establishing management plans, albeit most often only for public forest holdings or large private properties. For example, in Bavaria, Germany, a recurring terrestrial sample-based forest inventory system has been in place since the 1980s (A. Schnell, personal communication, 18 April 2016) in the state forests (which account for 30% of the total forest area in Bavaria; Thünen-Institut 2016). No comparable sampling is done in privately-owned forest parcels (56% of the total forest area), leading to a lack of up-to-date and reliable data for these forests.

In this paper, we propose a remote sensing-based approach in order to bridge this data gap at relatively low cost, thus providing information to both the state authorities for fulfilling their advisory mandate and the private forest owners to help with sustainable management. By utilizing existing ground plot data from the state forests in combination with readily available remote sensing data which covers large areas, prediction models for forest attributes such as mean stem diameter, mean and top heights, or gross volumes can be created. Subsequently these models can be applied to

estimate these attributes in private forests in the same geographic area, i.e., private forests, which lack inventory information. Despite the availability of data and the practical necessity for spatially explicit forest inventory information, to our knowledge, the spatial portability of prediction models has received little attention in the recently published literature on remote sensing for forestry applications.

Some research has been conducted to examine the efficacy of repeat airborne laser scanning (ALS) acquisitions for quantifying dynamics in forests, and to monitor this change in spatially explicit ways by comparing wall-to-wall predictions from different points in time. For example, Hudak et al. (2012) and Fekety et al. (2015) utilized repeat ALS measurements collected over a six-year period for a study area in northern Idaho, USA, to quantify fluxes in aboveground biomass pools. In another case study for a small boreal forest area in south-eastern Norway, Næsset et al. (2013) used multi-temporal ALS data acquired at 11-year intervals to demonstrate how model-assisted approaches can support field sample surveys to reduce uncertainty in biomass change estimates.

With regard to prediction across larger spatial extents, interesting research was published by Zald et al. (2016). In their study for Saskatchewan, Canada, they used spectral indices and change metrics derived from Landsat TM/ETM+ pixel composites as predictors and ALS-based response variables describing forest structure from LiDAR transects to establish nearest-neighbour prediction models. Spatially complete information over 37 million ha of forestland was generated via model imputation. Achieved accuracies in terms of RMSE were e.g., 2.11 m for Lorey's height or 24.99 m^3ha^{-1} for stem volume. This approach, however, was designed to facilitate management of vast forestlands, as are common in the boreal forest region of North America. Considering the great difference in forest conditions in Europe, the smaller

areas of interest, and a highly fragmented landscape, other remote sensing data seem more appropriate for gaining insight into forests for which there are no available field data.

In recent years, aerial photographs have become particularly interesting for practical forestry applications, as such images are now recorded digitally with high radiometric resolution and high geometric accuracy (White, Wulder & Vastaranta et al. 2013). Additionally, improvements in computing power and the development of enhanced algorithms for image matching – for instance, semi-global matching (Hirschmüller 2008; Haala 2013) – have generated new interest in the use of photogrammetrically derived height data for describing forest characteristics. Most surveying authorities in central Europe have established routine observation programs. For example, in Germany, digital airphotos covering the entire country are now acquired every 2 to 3 years. Thus, the basis is already laid for the development of new approaches for integrating digital aerial photogrammetry (DAP) and products derived from airphotos such as point clouds and raster surface models into forest planning and consulting (Straub et al. 2015).

DAP products can serve as surrogates for costlier ALS-based equivalents for various forest inventory applications. Several studies have been published following the conceptual frameworks established in ALS work (e.g., Nilsson 1996; Næsset 2002b, 2004; Packalén & Maltamo 2007; White, Wulder & Varhola et al. 2013) and early photogrammetric studies (e.g., Carson et al. 1996; St-Onge & Achaichia 2001; Næsset 2002a; Korpela & Anttila 2004). Recent studies have examined the use of DAP products to quantify canopy height changes (Stepper et al. 2015a; Vastaranta et al. 2015), and more frequently, for building prediction models for forest attributes via the area-based approach (Bohlin et al. 2012; Straub et al. 2013; Vastaranta et al. 2013; Pitt

et al. 2014; Rahlf et al. 2014; Gobakken et al. 2015; Stepper et al. 2015b; White et al. 2015). While different parametric and non-parametric regression methods have been used, and forest environments ranging from relatively homogeneous to complex have been assessed, most studies have reported modelling accuracies similar to or only slightly worse than the predictive models created using ALS data.

To evaluate model performances at the plot level, in most cases either built-in validation (e.g., out-of-bag) or cross-validation (cv) techniques, such as leave-one-out or 10-fold cv have been applied. Pitt et al. (2014) used independent validation plots and Gobakken et al. (2015) performed tests on independent stands to quantify model accuracy and bias in addition to using cross-validation. Independent model validation was also conducted by Immitzer et al. (2016), who broadened the scope of digital photogrammetry and forest attribute modelling by utilizing stereo WorldView-2 satellite data in combination with angle-count sampled plots from the German national forest inventory to approximate growing stock volumes. They validated their prediction models both at the plot- and stand-levels using additional management forest inventory data as an independent validation dataset. With achieved accuracies comparable to area-based predictions using fixed-area plots for training, they demonstrated the practicability of their proposed approach.

The studies mentioned above show promise for using DAP in area-based predictions of forest attributes in different forest ecosystems. However, in our judgement, they still have not made use of the full potential of DAP-based height information that covers large spatial extents. Little work has evaluated the efficacy of prediction models based on the combination of DAP data and concurrently collected field measurements for spatial extrapolation, i.e., for model transfers to forest regions that do not have associated field data. These forest areas are of particular interest for the

165 forest authorities throughout Europe, as they are often responsible for assisting the
166 owners of small private forests with silvicultural planning and management. Thus, these
167 officials have called for spatially explicit information on these forests.

168 The primary objective of this research is to examine the potential for using aerial
169 photography covering a large spatial extent, ground plot measurements from state
170 forests, and statistical regression modelling to predict a set of forest attributes in forest
171 areas, which generally are not assessed via traditional field-based inventories.
172 Specifically, we train random forest prediction models for three forest attributes –
173 quadratic mean diameter of the 100 largest trees (d_{100}), basal area weighted mean height
174 of the 100 largest trees (h_{100}), and gross volume (V) – in both a spruce- and a beech-
175 dominated forest area in northern Bavaria, Germany for which field data exists. These
176 test sites were selected as we assume these areas to be representative of the predominant
177 forest conditions in central Europe. In a second step, the spatial portability of our
178 models to nearby forests was tested using independent reference data to create
179 performance estimates for model accuracy and bias. As a secondary objective, the
180 impact of varying amounts of training data on the robustness of the models was
181 evaluated, once again focusing on model transferability to other forest areas located in
182 the vicinity of the training areas.

183 **Materials and Methods**

184 ***Study areas***

185 We selected two distinct study areas, both of which are located in the northern part of
186 Bavaria, Germany (Figure 1), to investigate the transferability of forest attribute models
187 for commonly occurring forest types in Germany. The study areas *Frankenwald*
188 (FRAN) and *Steigerwald* (STEI) differ considerably from one another with respect to

189 forest tree species composition: conifers dominate in the FRAN study area, whereas the
190 stands in the STEI test site are primarily stocked with broadleaved trees.

191 The FRAN study area is centred at 50°20’ N, 11°35’ E. The topography is
192 characterized by steeply cut valleys, mostly running north-south, and the elevation
193 ranges from 360 m to 760 m above sea level (a.s.l) with an average elevation of 575 m
194 a.s.l. To distinguish training areas from test areas, the FRAN site was split into two
195 subareas (FRAN_train, FRAN_test; cf. Figure 1b and Table 1), covering a total of
196 15,651 ha, of which 10,353 ha are state-owned land which is mostly managed forest.
197 The main tree species in this study area are Norway spruce (*Picea abies* [L.] Karst.,
198 74.0%) and European beech (*Fagus sylvatica* L., 17.3%).

199 The STEI test site is centred at 49°55’ N, 10°35’ E, and comprises a total area of
200 16,554 ha (STEI_train, STEI_test; cf. Figure 1c and Table 1). An eastward sloping high
201 plain, interspersed with steep valley cuts, dominates the terrain, and the elevation ranges
202 from 220 m to 490 m a.s.l. (mean elevation: 385 m a.s.l.). The state-owned land within
203 this study area (11,207 ha) is primarily comprised of broadleaf-dominated forest stands.
204 The dominant tree species are European beech (41.5%), sessile oak (*Quercus petraea*
205 (Mattuschka) Liebl., 18.7%), and Scots pine (*Pinus sylvestris* L., 14.4%).

206

207 **Table 1:** Description of the two study areas – FRAN and STEI. The number of ground
208 plots lists first the total number of sampling plots used for model training and testing,
209 and in parentheses the various decreasing densities of training data used to test the
210 importance of the number of data points on overall model performance.

Test site	Subarea	Total area (ha)	State-owned forest area (ha)	No. of ground plots (-)
FRAN	FRAN_train	9,347	6,541	1,477 (751; 358; 181; 89; 42)
	FRAN_test	6,304	3,812	790
STEI	STEI_train	12,005	7,739	1,848 (924; 453; 223; 112; 60)
	STEI_test	4,549	3,468	819

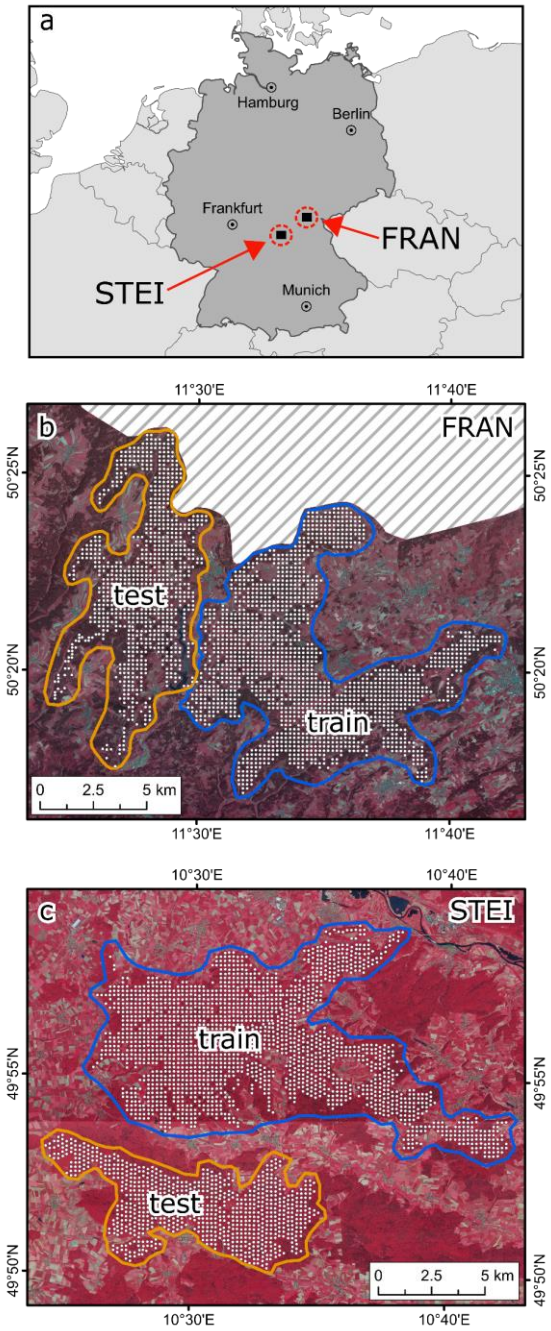


Figure 1: Map of the study areas *Frankenwald* (FRAN) and *Steigerwald* (STEI). (**a**) Geographic locations of the two study areas in Germany. (**b**) FRAN test site: spatial distribution of the ground plots, separated into a training set (train) in the eastern part and a test set (test) in the western part, (**c**) STEI test site: spatial distribution of the ground plots, training set (train) in the northern part and test set (test) in the southern part. Background image of subfigures (b) and (c): colour-infrared orthophoto mosaics from the aerial photos used in the study.

Data sets

Ground plot data

The ground plot data used in this study were collected in 2014 and 2010 for the test sites FRAN and STEI, respectively. Based on a regular grid of 200 m x 200 m, permanently marked sample plots are systematically distributed across the state forests in both study areas, resulting in a total of 2,267 plots located in FRAN and 2,667 located in STEI (see Figure 1). The ground plots are circular in shape (radius 12.62 m), with each covering an area of 500 m^2.

All field measurements were acquired in accordance with the field inventory guidelines of the Bavarian State Forest Enterprise (BaySF, Neufanger 2011). The centres of the field plots were georeferenced using a Trimble GeoExplorer XT GPS device (maximum deviations of ±3-5 m; H. Grünvogel, personal communication, 4 August 2014). Within each plot, tree characteristics such as species, diameter at breast height (DBH; stem diameter at 1.3 m above ground), and tree height were recorded for individual trees. Using these data, individual tree-based stem volumes were computed. For details regarding field data acquisition, refer to Stepper et al. (2015b).

We used only the data collected for trees with DBH ≥ 7 cm (the minimum threshold for merchantable timber) to compute the estimates for the following forest-related attributes at the plot level:

(1) Quadratic mean diameter of the 100 largest trees (d_{100}): Diameter of the mean basal area tree of the 100 stems per hectare with the largest DBHs. Only those trees which were assigned to either the overstory or the emergent tree layer by the field workers were selected to compute the plot-level d_{100}. To generate the

list of the 100 trees per hectare with the largest DBHs, we followed the algorithm described in Stepper et al. (2015a).

(2) Basal area-weighted mean height of the 100 trees per hectare with the largest diameters (h_{100}): The basal area-weighted mean height was computed for the same tree collective used to derive d_{100}. This plot-level h_{100} was calculated to be used as a proxy for stand top height.

(3) Gross volume (V): Sum of the single tree stem volumes per hectare which refers here only to the total merchantable timber (m^3ha^{-1}), i.e., the aboveground inside-bark stem volume up to a minimum diameter limit of 7 cm.

These three forest attributes were selected for assessment due to their high relevance for forest management planning. The d_{100} value indicates the stand development phase and the trees' fitness for felling (i.e., having passed specified diameter thresholds for harvesting). The h_{100} hints at site productivity and by this, allows for the estimation of growth rates and the determination of a reasonable annual cut. The ratio h_{100}/d_{100}, often referred to as the slenderness of the dominant trees, is commonly used as an indicator for mechanical tree stability; values of $h_{100}/d_{100} < 0.8$ [m/cm] indicate sufficient stability against snow- and wind-breakage. Gross standing volume V is essential for silvicultural planning, especially with respect to determining the amount of harvestable wood. In mature stands, V is used to determine the optimal time to take measures to initiate natural regeneration. In order to ensure a successor generation of mixed-species stands, V needs to be monitored and systematically reduced once regeneration has been established.

Descriptive statistics for the inventory data from the test sites FRAN and STEI are listed in Table 2. Overall, the mean growing stock volumes of about 300 m^3ha^{-1} are somewhat above the national average (264 m^3ha^{-1}) as reported by the most recent

German national forest inventory (Thünen-Institut 2016). The gross volumes in the two study areas are very similar, with slightly higher mean values in the beech-dominated STEI area. On a related note, the mean values of d_{100} and h_{100} are marginally lower in the FRAN area. The rather high mean tree dimensions and standing volumes, as calculated from the ground plot measurements (Table 2), reflect the fact that in both test sites the majority of stands are in a mature development phase. According to Borchert (2007), such mature forest stands dominate the forest landscape in Bavaria. As discussed by Verkerk et al. (2011), the situation is similar across central Europe and these mature forests are of particular relevance for efforts to support wood mobilization. Volume needs to be continuously reduced, especially in these stands with advanced tree age, in order to establish and foster multi-species regeneration, which is the management objective in both of the study sites (Müller 2005).

Table 2: Summary of ground plot characteristics in the training and test subareas within the two test sites (n = the number of ground plots available for examination). Here, d_{100} = quadratic mean diameter of the 100 largest trees per ha, h_{100} = basal area weighted mean height of the 100 largest trees per ha, V = gross volume per ha.

	Forest attribute	**Min.**	**1st Qu.**	**Median**	**3rd Qu.**	**Max.**	**Mean**	**SD**
FRAN_train (n = 1477)	d_{100} [cm]	7.5	34.5	42.3	48.1	71.4	40.4	11.5
	h_{100} [m]	4.0	24.2	28.8	32.1	44.7	27.3	7.0
	V [m³ha^{-1}]	2.2	172.5	288.0	399.0	1051.4	292.0	160.2
FRAN_test (n = 790)	d_{100} [cm]	7.5	33.1	41.0	47.2	67.5	39.1	12.0
	h_{100} [m]	3.0	23.6	28.4	31.3	42.5	26.3	7.5
	V [m³ha^{-1}]	1.4	151.2	296.6	410.3	846.2	291.2	173.0
STEI_train (n = 1848)	d_{100} [cm]	8.5	36.9	44.0	50.0	84.5	42.9	11.1
	h_{100} [m]	8.4	24.9	28.3	31.7	44.5	28.0	5.6
	V [m³ha^{-1}]	3.2	212.3	297.6	384.6	1075.2	301.1	135.1
STEI_test (n = 819)	d_{100} [cm]	13.5	37.8	44.8	51.2	83.5	44.4	10.4
	h_{100} [m]	12.5	25.9	29.0	31.6	43.2	28.6	4.6
	V [m³ha^{-1}]	21.7	204.4	288.2	368.2	986.7	296.1	127.3

Digital aerial imagery

Digital imagery covering the study areas was acquired as part of the regularly scheduled aerial survey of Bavaria. An UltraCam-Xp camera was used for image acquisition in both areas. For each of the two test sites complete coverage was achieved through two separate flights. The images for the FRAN test site were acquired on 21 May and 8 June 2014, and for the STEI test site, in two separate flights on 7 May 2011. The imagery is characterized by a ground sampling distance of 0.20 m and minimum along-track / across-track overlaps of 75% / 40% for FRAN and 75% / 30% for STEI. This resulted in a total of 450 and 372 stereo images, respectively.

For further photogrammetric processing, we used the panchromatic data, which was recorded with a radiometric resolution of 12 bit. Image matching was done using RSG software (v. 7.46.11, Joanneum Research 2015) by means of the semi-global matching (SGM, Hirschmüller 2008) algorithm, as per Stepper et al. (2015b). All stereo pairs from the along-track overlap were matched to generate dense, image-based point clouds. The measurements from the matched stereo pairs were merged for each of the two test sites and ALS-based terrain heights (from 1 m spatial resolution topographic mapping surveys conducted in 2009) were subtracted from the image-based surface heights to derive vegetation heights above ground. The point clouds were then clipped to the respective extents of the 500 m^2 ground plots, resulting in final point clouds with average densities of 168 points$\cdot m^{-2}$ and 117 points$\cdot m^{-2}$ for FRAN and STEI, respectively. (See Figure S1 in the supplementary material for an illustration of a height-normalized point cloud clipped to the extent of a 500 m^2 ground plot).

Area-based modelling of forest attributes

Figure 2 illustrates the workflow for model training and validation used in our study. As

shown, the spatial portability of the trained models was evaluated by predicting forest attribute values at independent 'test' sample plots from nearby forests. The modelling procedure was developed in the open-source statistical software R (v. 3.2.2; R Core Team 2015) using the R packages *randomForest* (Liaw & Wiener 2002) and *caret* (Kuhn 2015).

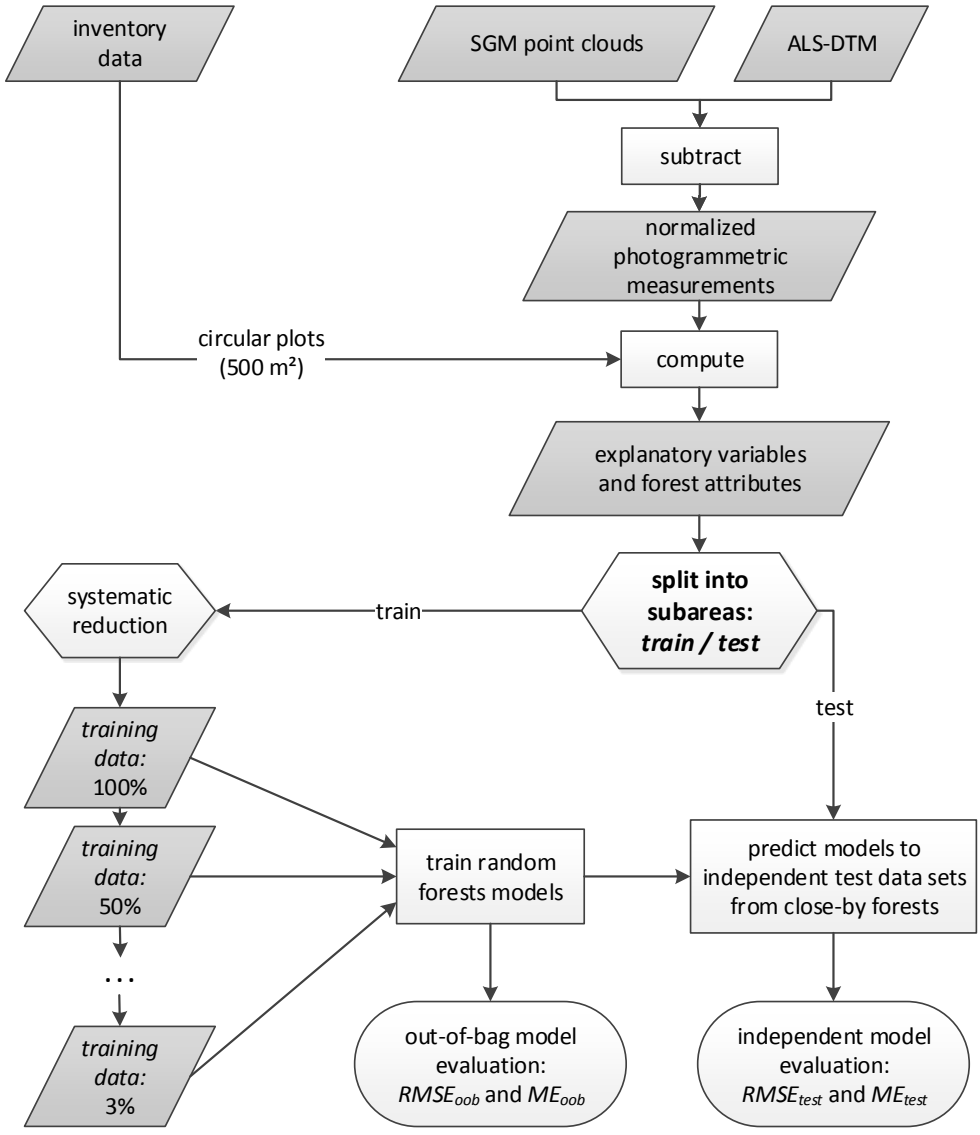


Figure 2: Procedure for area-based prediction of different forest attributes with models created from image-based point clouds and ground plot inventory data. To assess the importance of the density of ground sample points, different subsets of the available training data were used to create additional separate models: 50%, 25%, 12%, 6%, and 3% of the original number of ground plots. The spatial portability of the trained models was assessed by predicting independent samples from nearby forests.

For each ground plot (size: 500 m^2), candidate explanatory variables were derived from the image-based point clouds. These metrics were computed by means of FUSION (v. 3.42, McGaughey 2014) functionalities and scripts coded in the R environment. Informed by, e.g., White et al. (2015), who demonstrated that image-based canopy metrics are highly prone to multicollinearity, and in the interest of parsimony, we chose a reduced set of variables for modelling – assigned to the categories *height*, *variation of height* and *canopy cover* (Table 3). The selection of variables was supported by a previous analysis of principal components with corresponding component loadings, as well as correlation analysis of the potential explanatory variables with the three target forest attributes.

Table 3: Metrics derived from the image-based point clouds used as explanatory variables in predictive random forest models of the target variables – d_{100}, h_{100} and V.

Category	Metric	Description	Unit
Height	*H.mean*	Average of point heights	m
	H.P90	90th percentile of point heights	m
Variation of height	*H.CoV*	Coefficient of variation of point heights	-
	H.SD	Standard deviation of point heights	-
	H.skew	Skewness of point heights	-
	H.kurt	Kurtosis of point heights	-
	Rumple	Ratio of canopy outer surface area to ground surface area (Kane et al. 2010)	-
	VCI	Vertical complexity index (modified after van Ewijk et al. 2011; based on the Shannon-index measuring entropy (Shannon 1948), which is commonly used in ecology to describe diversity): $VCI = -\sum_i^{HB} p_i \cdot \ln(p_i)$, where *HB* is the total number of height bins (here: 0-1m, 1-2m, …, 49-50m), and p_i is the proportion of points in the respective height bin *i* $\left(\sum_i^{HB} p_i = 1\right)$.	-
Canopy closure	*CC.mean*	Proportion of points above mean canopy height	%

As mentioned above, the machine learning algorithm random forest (RF, Breiman 2001) was applied. In RF, uncorrelated decision trees are trained based on bootstrap-sampled versions of the training data. During the tree building process, a random subset of predictor variables is selected as candidates for splitting at each node. The trees are then assembled to form the final regression model by averaging these individual decisions. Following the recommendations of Kuhn & Johnson (2013), 1000-tree ensembles were run to set up the RF models (n_{tree} = 1000), with the number of randomly selected variables at each node equal to one-third of the total number of predictors P (m_{try} = P/3). The suitability of RF in regression mode for the prediction of forest attributes based on canopy height metrics has been demonstrated in several recent studies, e.g., Pitt et al. (2014), Immitzer et al. (2016) or Straub & Stepper (in press).

RF models for $d_{100,}$ h_{100} and V were trained separately for each of the two study areas using the full set of available inventory plots in the respective FRAN_train and STEI_train data sets. Additionally, to investigate the impact that varying amounts of training data had on model performance, the original sample grid (200 m x 200 m) was systematically reduced as per Stepper et al. (2015b). We sequentially eliminated every second inventory plot from the previous grid and eventually, six different training data sets (i.e., inventory grid densities) comprising 100%, 50%, 25%, 12%, 6%, and 3% of the full set of inventory plots, were created (see Table 1). The regression models were again trained using each of the reduced sets of input data in both the FRAN_train and STEI_train sites.

To examine thoroughly the spatial portability of all of the established models, each was then used to approximate the variables $d_{100,}$ h_{100} and V at the inventory plots in the nearby forests (FRAN_test and STEI_test). Thus, two methods for assessing the performance of the models were used:

(1) Using the built-in out-of-bag (oob) performance measures computed along the way during training of the RF models.

(2) Quantifying the performance after predicting the 'test' data sets – these are the independent inventory plots from the FRAN_test and STEI_test areas for which the plot data was not used to build the models. This evaluation provides an unbiased sense of model effectiveness when predicting values at adjacent, similarly structured forests.

To evaluate the performance of the models, we chose the root mean squared error (RMSE, eq. 1) as goodness-of-fit statistic, i.e., to quantify model accuracies. Additionally, the mean error (ME, eq. 2) was calculated as a measure for systematic errors (i.e., to assess model bias).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\mathrm{n}}} \quad (1)$$

$$\mathrm{ME} = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{\mathrm{n}} \quad (2)$$

where y_i refers to the observed value and $\hat{y}_i$ to the predicted value of the i^{th} of n sampled observations. To make the results comparable, the absolute RMSE and ME values were scaled by dividing through the observations' means ($\bar{y}$) and given in percentage terms (sRMSE% and sME%).

Results

A comparative analysis of RF prediction models using the area-based approach in both a spruce- and a beech-dominated forest was carried out for each model. Test site-specific models were built using the training data from each of the two forest areas separately – and applied to predict the independent validation data from the respective test areas.

Feature importance of predictor variables

The RF variable importance values were utilized to explore the relationship between the predictor and the response variables, i.e., the DAP-based metrics and the selected forest attributes. For the models of each of the three response variables created using the full training data sets, the resulting values of predictor variable importance for each of the two study sites are presented in Figure 3.

For the models of d_{100} and h_{100} for both sites, *H.P90* and *H.mean* appear at the top of the importance metric, with a clearer distance between top first and second in the spruce-dominated site than in the beech site. A closer look at the importance values for the d_{100} models reveals that the variables *Rumple*, *H.CoV*, *H.SD*, and *VCI* were of intermediate importance, whereas *H.skew*, *H.kurt*, and *CC.mean* had almost no importance. Similar results were observed for the two models of h_{100}, with *Rumple* being the third most important predictor variable, and importance values tapering off to the least important variable *CC.mean*. As is visible in the bottom line of Figure 3, *H.mean* was the top predictor for volume *V*, in both of the test sites, with *H.P90* taking second place.

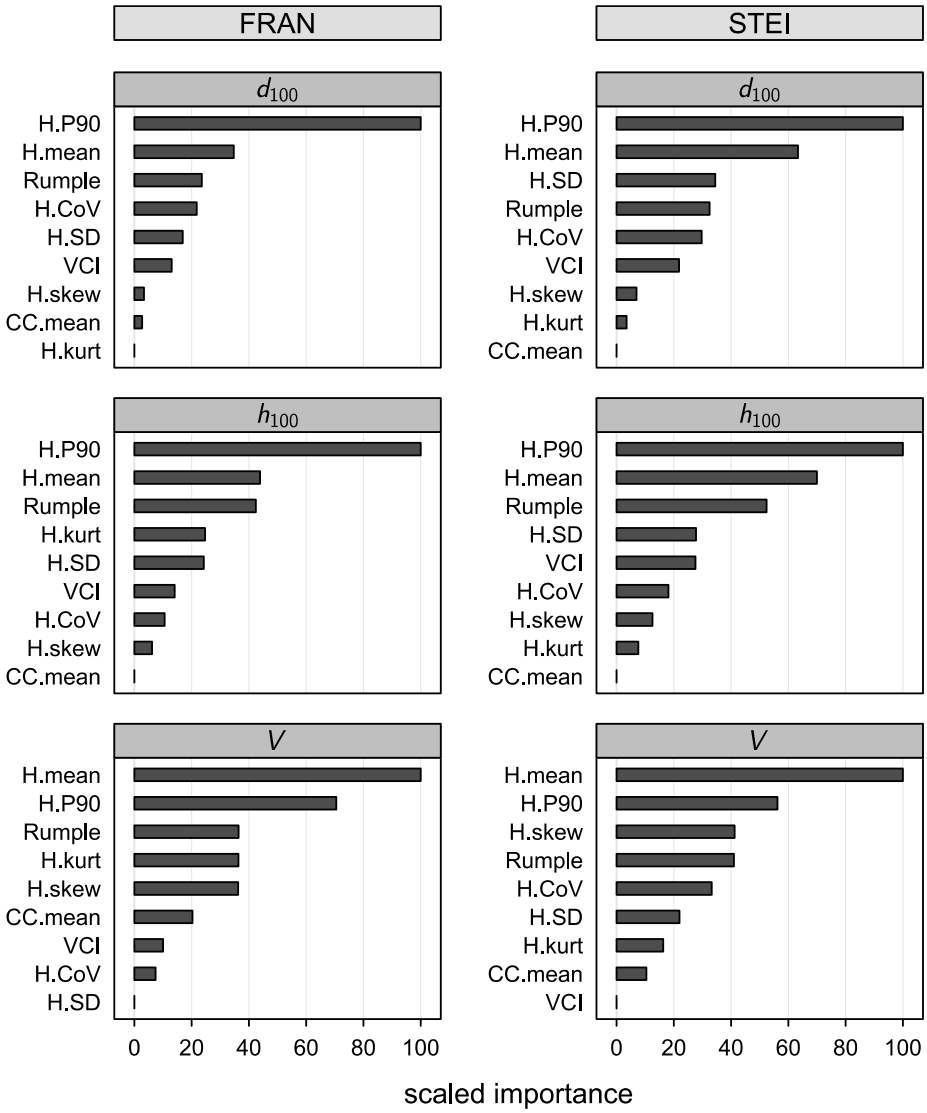
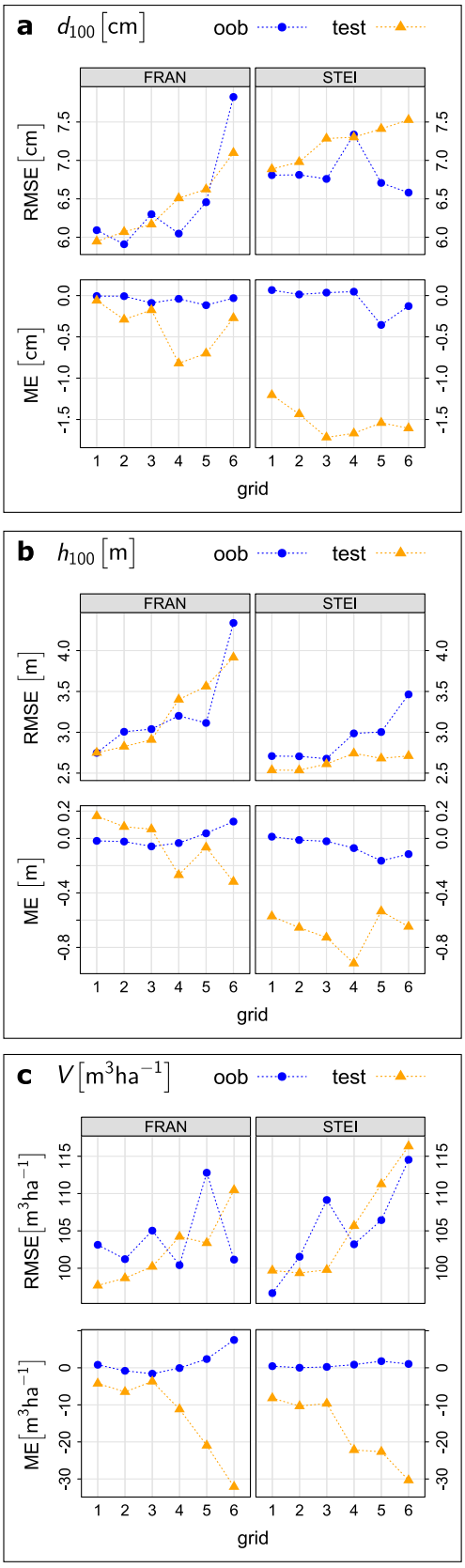


Figure 3: Variable importance measures for the models of d_{100}, h_{100}, and V generated using the full set of training data for each site (i.e., grid 1). The importance measures were computed as: mean decrease in accuracy (%incMSE) normalized by the "standard errors" and scaled to have a maximum value of 100 (cf. website of the *caret* package, Kuhn 2016).

Model performance

The performance measures of the models for d_{100}, h_{100}, and V are summarized in Table 4 and Table 5 for the FRAN and the STEI study areas, respectively. These tables report on both the 'oob' evaluation and the transfer to the independent test data. Figure 4 displays the results at a glance, condensing the effects of (i) forest attribute considered, (ii) study site, i.e., forest type, and (iii) varying amounts of training data. It highlights the differences between the 'oob' and 'test' evaluation results, both for RMSE and ME.



415

416 **Figure 4:** Influence of number of inventory plots used for training on the RMSE and
417 mean error (ME) of the random forest models for the three different attributes (**a**: d_{100};
418 **b**: h_{100}, **c**: V). Shown separately for the two test sites – FRAN and STEI.

Table 4: Summary of the performance measures for the random forest models predicting d_{100}, h_{100}, and V in the FRAN study area. Results are shown for both the out-of-bag (oob) model evaluation and the independent assessment of predictive performance using samples from the nearby forests. Results listed separately for each of the six different inventory plot densities used for model training (grids 1–6: 1477, 751, 358, 181, 89, and 42 sample plots) and independent validation (test: 790 sample plots).

Forest attribute	Mode	Grid	Mean Observed	RMSE	sRMSE%	ME	sME%
d_{100} [cm]	oob	1	40.4	6.1	15.1	0.0	0.0
		2	40.3	5.9	14.7	0.0	0.0
		3	40.0	6.3	15.7	-0.1	-0.2
		4	39.3	6.0	15.4	0.0	-0.1
		5	39.9	6.5	16.2	-0.1	-0.3
		6	37.9	7.8	20.6	0.0	-0.1
	test	1	39.1	5.9	15.2	-0.1	-0.2
		2		6.1	15.5	-0.3	-0.7
		3		6.2	15.8	-0.2	-0.4
		4		6.5	16.6	-0.8	-2.1
		5		6.6	16.9	-0.7	-1.8
		6		7.1	18.1	-0.3	-0.7
h_{100} [m]	oob	1	27.3	2.7	10.1	0.0	-0.1
		2	27.4	3.0	11.0	0.0	-0.1
		3	27.2	3.0	11.2	-0.1	-0.2
		4	26.7	3.2	12.0	0.0	-0.1
		5	27.0	3.1	11.6	0.0	0.1
		6	25.2	4.3	17.2	0.1	0.5
	test	1	26.3	2.7	10.5	0.2	0.6
		2		2.8	10.8	0.1	0.3
		3		2.9	11.1	0.1	0.3
		4		3.4	12.9	-0.3	-1.0
		5		3.6	13.6	-0.1	-0.3
		6		3.9	14.9	-0.3	-1.2
V [m^3ha^{-1}]	oob	1	292.0	103.1	35.3	0.8	0.3
		2	287.8	101.2	35.2	-0.8	-0.3
		3	286.7	105.0	36.6	-1.6	-0.6
		4	285.3	100.4	35.2	-0.1	0.0
		5	285.9	112.8	39.5	2.4	0.8
		6	241.3	101.1	41.9	7.5	3.1
	test	1	291.2	97.7	33.6	-4.2	-1.5
		2		98.7	33.9	-6.5	-2.2
		3		100.2	34.4	-3.7	-1.3
		4		104.2	35.8	-11.2	-3.8
		5		103.4	35.5	-21.0	-7.2
		6		110.5	37.9	-32.1	-11.0

Table 5: Summary of the performance measures for the random forest models predicting d_{100}, h_{100}, and V in the STEI study area. Results are shown for both the out-of-bag (oob) model evaluation and the independent assessment of predictive performance using samples from the nearby forests. Results listed separately for each of the six different inventory plot densities used for model training (grids 1–6: 1848, 924, 453, 223, 112, and 60 sample plots) and independent validation (test: 819 sample plots).

Forest attribute	Mode	Grid	Mean Observed	RMSE	sRMSE%	ME	sME%
d_{100} [cm]	oob	1	42.9	6.8	15.9	0.1	0.2
		2	42.6	6.8	16.0	0.0	0.0
		3	41.9	6.8	16.1	0.0	0.1
		4	41.1	7.3	17.8	0.0	0.1
		5	41.7	6.7	16.1	-0.4	-0.9
		6	41.5	6.6	15.9	-0.1	-0.3
	test	1	44.4	6.9	15.5	-1.2	-2.7
		2		7.0	15.7	-1.4	-3.2
		3		7.3	16.4	-1.7	-3.9
		4		7.3	16.4	-1.7	-3.8
		5		7.4	16.7	-1.5	-3.5
		6		7.5	17.0	-1.6	-3.6
h_{100} [m]	oob	1	28.0	2.7	9.7	0.0	0.0
		2	27.9	2.7	9.7	0.0	0.0
		3	27.6	2.7	9.7	0.0	-0.1
		4	27.1	3.0	11.0	-0.1	-0.3
		5	27.2	3.0	11.0	-0.2	-0.6
		6	27.0	3.5	12.8	-0.1	-0.4
	test	1	28.6	2.5	8.9	-0.6	-2.0
		2		2.5	8.9	-0.7	-2.3
		3		2.6	9.1	-0.7	-2.5
		4		2.7	9.6	-0.9	-3.2
		5		2.7	9.4	-0.5	-1.9
		6		2.7	9.5	-0.6	-2.3
V [m³ ha^{-1}]	oob	1	301.1	96.7	32.1	0.4	0.1
		2	299.8	101.5	33.9	0.0	0.0
		3	299.3	109.1	36.5	0.2	0.1
		4	283.4	103.2	36.4	0.8	0.3
		5	284.4	106.4	37.4	1.8	0.6
		6	279.6	114.5	41.0	1.0	0.4
	test	1	296.1	99.7	33.7	-8.2	-2.8
		2		99.4	33.6	-10.3	-3.5
		3		99.8	33.7	-9.7	-3.3
		4		105.7	35.7	-22.2	-7.5
		5		111.3	37.6	-22.7	-7.7
		6		116.3	39.3	-30.3	-10.3

Out-of-bag model performance

Focusing on the 'oob' results, the MEs suggest no bias when modelling the different attributes, irrespective of the number of training samples used in either study site. Only in the model for *V*, did ME increase marginally when grid 6 was used, and here only for the conifer-dominated (FRAN) site.

The accuracies for predicting d_{100} in the conifer-dominated FRAN area were approximately the same for grids 1 through 5 ($RMSE_{oob}$ ranging from 5.9 cm to 6.5 cm). The increased error for grid 6 ($RMSE_{oob}$ = 7.8 cm) may be explained by the obviously reduced range of d_{100} values mapped by the limited amount of field data in the most reduced sample size and the related leverage of over-predicted values at the lower end (cf. Figure S2 and Figure S3 in supplementary material). The RMSEs for the h_{100} models in the FRAN area followed a similar course, slightly increasing from grid 1 to grid 5 ($RMSE_{oob}$ between 2.7 m and 3.2 m), and showing a more pronounced increase when grid 6 was used ($RMSE_{oob}$ = 4.3 m). We attribute this to the same reasons described above for d_{100}. The highest $RMSE_{oob}$-values were recorded for the prediction of *V* (the most 'integrative' variable considered), with results ranging between 100.4 m^3ha^{-1} and 112.8 m^3ha^{-1}, but with no distinct systematic increase pattern for the progressively smaller sets of training data.

In general, the 'oob' estimates for the models from the broadleaf-dominated STEI test site were comparable to those from the FRAN area. $RMSE_{oob}$-values for d_{100} were between 6.6 cm and 7.3 cm for STEI, thus slightly higher than those for FRAN (and in line with the somewhat higher mean observed values). However, there was no decline in accuracy when moving from grid 1 to 6 in the STEI area (see Figure 4). When predicting h_{100}, however, the $RMSE_{oob}$ increased from 2.7 m to 3.5 m from grid 1 to grid 6. The most significant decrease in accuracy was observed for the *V* models in

458 the STEI test site, with $RMSE_{oob}$-values changing from 96.7 m^3ha^{-1} (grid 1) to

459 114.5 m^3ha^{-1} (grid 6).

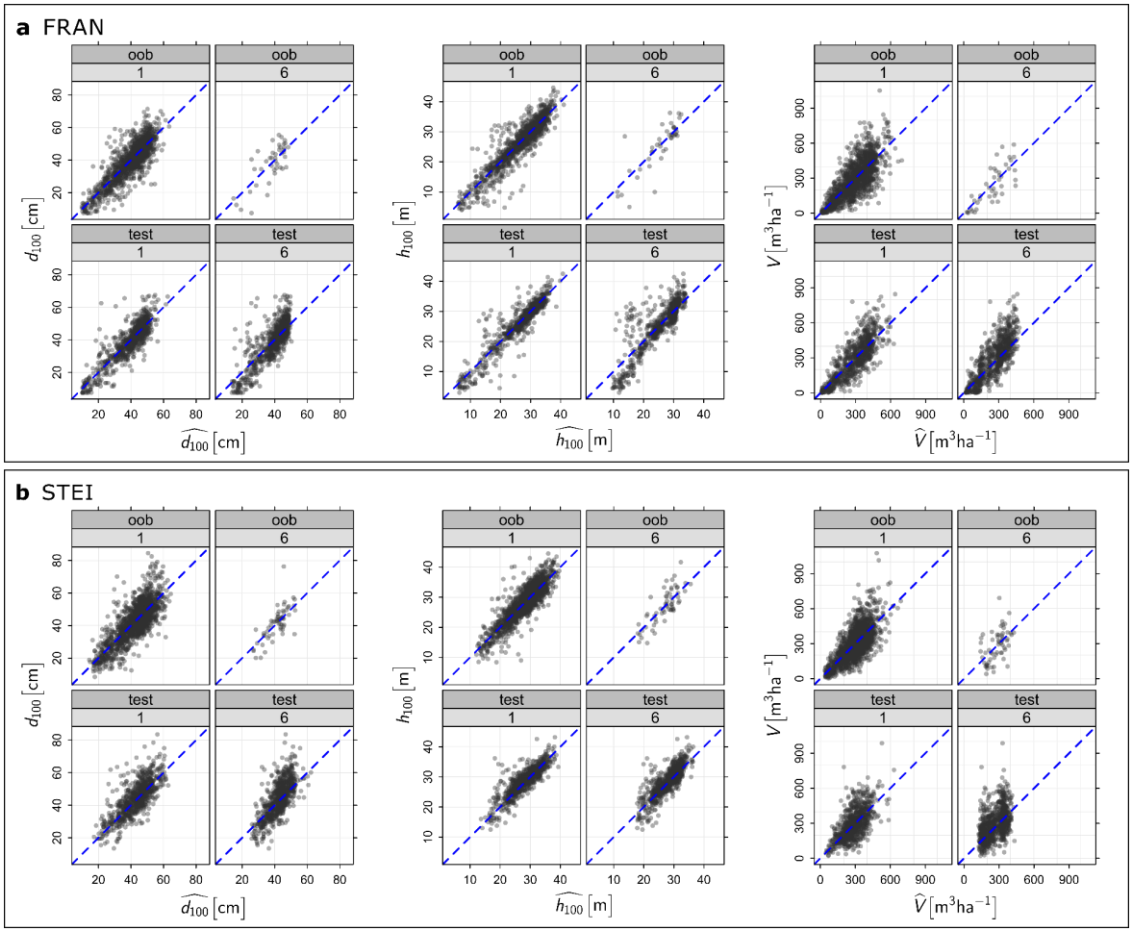
460 *Model transfer to nearby forests*

461 As can be seen from the scatter plots in Figure 5 (and Figure S3 in the supplementary

462 material), the picture changes when examining the predictions at inventory plots of the

463 neighbouring forest areas, i.e., the independent validation data.

464



465

466 **Figure 5:** Observed vs. predicted values of d_{100}, h_{100}, and V, as obtained from the out-
467 of-bag (oob) evaluations and the model predictions at the independent validation data
468 (test) in the study areas FRAN (**a**) and STEI (**b**). Selected scatter plots are presented
469 for the models using the full set of inventory plots (grid 1) and the most heavily reduced
470 set (grid 6). The scatter plots for all intermediate reduction steps are provided in
471 supplementary Figure S3.

472

In the FRAN study area, the $RMSE_{test}$-values for predicting d_{100} were comparable to the 'oob' measures and again, increased slightly from grid 1 to grid 6 (5.9 cm to 7.1 cm). The same was observed for h_{100}, where the range of $RMSE_{test}$-values from grid 1 to grid 6 was 2.7 m to 3.9 m. Regarding *V*, the RMSEs increased from 97.7 m^3ha^{-1} (grid 1) to 110.5 m^3ha^{-1} (grid 6). For both d_{100} and h_{100}, only small negative ME_{test}-values were introduced to the models from the reduced sets 'grid 4' and onwards. However, this trend towards heavier systematic errors was more pronounced for the prediction of *V*, where the minimum ME was -32.1 m^3ha^{-1} (-11.0%) for grid 6. The negative MEs hint at the phenomenon frequently observed when using an inadequate number of training samples – namely that small values are over-predicted and large ones under-predicted. This numerical result is clearly illustrated by the scatter plots of the observed versus predicted values (Figure 5 and Figure S3).

Looking at the independent validation results of the models from the STEI study area, the $RMSE_{test}$-values for d_{100} and *V* also increased from grid 1 to grid 6 (6.9 cm to 7.5 cm and 99.7 m^3ha^{-1} to 116.3 m^3ha^{-1}). In the case of h_{100}, however, the $RMSE_{test}$-values stayed rather constant across the different grids, and were even lower than the corresponding 'oob' estimates. Still, the ME_{test}-values for the STEI area reveal some distinct systematic errors – this being valid for each of the three attributes. For d_{100} and h_{100}, the negative MEs were approximately at the same level for each of the different grids, whereas for *V*, the ME decreased from -8.2 m^3ha^{-1} (-2.8%) to -30.3 m^3ha^{-1} (-10.3%). Analogous to the outcomes from the FRAN area, these results support the assumption of biased predictions for models built with training data that is not sufficient to reproduce the entire range of observed data.

Discussion

The main focus of our study was to examine and evaluate the usefulness of DAP-based canopy heights for producing predictive models for selected forest attributes that are transferrable to areas which are usually not assessed via systematic ground-based inventories. In an effort to meet this goal, we chose to follow the established area-based approach and RF regression for model building.

Across all of the three forest attributes considered, the random forest importance measures confirmed the relevance of metrics directly describing canopy height (*H.mean*, *H.P90*) for prediction. Further, as can be clearly seen in Figure 3, the *Rumple* metric appeared to contribute significantly to the models' performance. No obvious pattern was found for the remaining metrics describing the variation in height (*H.CoV*, *H.SD*, *H.skew*, *H.kurt*, *VCI*). *CC.mean*, the selected metric for describing canopy closure, made no significant contribution to the models' predictive ability. While these findings are in line with those reported by White et al. (2015), who emphasized the differences in the predictive contribution of ALS and DAP-based metrics, we cannot conclusively assess whether the importance values found here also hold true in other forest environments. Thus, especially with regard to "non-standard" metrics like *Rumple* or *VCI* derived from DAP-data, we would like to encourage further research in forests with different characteristics.

The built-in 'oob' predictions of the RF models were used for internal model validation. With regard to the models that were trained with the largest available set of inventory plots (grid 1), the differences in the accuracies found (as represented by the scaled RMSEs) for d_{100}, h_{100}, and V were minimal between the spruce- and the beech dominated sites. Thus, at least for the attributes considered here, our results suggest that RF modelling can be applicable to different forest types. However, as most of the stands

in both test sites are made up of either pure conifer or pure broadleaf groups, further research is necessary to examine model performance in forests with more heterogeneous tree mixtures (i.e., with mainly interspecific tree neighbourhoods).

In the context of previous published work from other geographic regions, we can evaluate our results for the different forest attributes as follows:

- With respect to d_{100}, we achieved 'oob' sRMSEs of 15.1% and 15.9% for grid 1 models in FRAN and STEI, respectively. We are not aware of another study that has reported on this particular metric. However, assuming that we can expect similar model behaviour for d_g (quadratic mean diameter computed for the entire tree collective), we refer to the sRMSEs reported in the following studies for comparison: Vastaranta et al. (2013): 21.7%; Gobakken et al. (2015): 18.7%; Pitt et al. (2014): 15.8%. However, in order to evaluate the figures properly, one should be aware of the following aspects: (i) all of the studies cited were conducted for boreal, conifer-dominated forests with stands stocked with thinner trees with less variation in diameter, (ii) d_{100} is likely to be more closely correlated to canopy height than d_g. Since the effects of (i) and (ii) act contrarily with respect to prediction from DAP-based canopy metrics, general conclusions about the achieved accuracies must be made with some caution. Nonetheless, we can state, that for the two test sites used in our study – both of which have rather high average d_{100} values – reasonable predictions were possible.
- The 'oob' sRMSEs obtained for h_{100} were 10.1% for FRAN and 9.7% for STEI. These values are very much in line with reported accuracies for top and mean heights from other studies (*top height*: Gobakken et al. (2015): 9.2%; Pitt et al. (2014): 10.8%; *mean height*: Vastaranta et al. (2013): 11.2%; Gobakken et al. (2015): 10.2%; White et al. (2015): 14.0%). Our results, hereby, confirm the

commonly reported finding that highly accurate approximations of canopy heights can be achieved by means of DAP.

- Regarding *V*, the 'oob' sRMSEs of 35.3% (FRAN) and 32.1% (STEI) are within the same magnitude as those reported by, e.g., Straub et al. (2013) for a temperate, mixed-species forest (sRMSE = 37.9%) and White et al. (2015) for a complex coastal forest (36.9%). However, better accuracies have often been reported for boreal test sites (Vastaranta et al. (2013): 24.5%; Pitt et al. (2014): 26.8%; Gobakken et al. (2015): 21.7%; Rahlf et al. (2015): 22%). As *V* integrates over the entire canopy, i.e., the vertical growing space, these differences might be related to the respective forest structures. Complexity in multi-layered stands is likely to add more scatter to these models than occurs in models of stands with mostly single-layered stand structures.

This comparison to related studies thus, confirms that our established prediction models were generally sound. We further infer that model accuracies for predicting forest attribute models by means of DAP-based data are rather insensitive to forest type.

Given this background on internal model accuracies, we discuss the novelty of our research: the spatial portability of the models to nearby forests. One underlying assumption of using models created with assistance from remotely sensed auxiliary data to predict conditions in other areas where no ground sampling has taken place is that the forest structure conditions in the target areas are within the range of sampled observations in the training regions. More specifically, it is essential that the relationship between the remotely sensed measurements and the forest attributes of interest is comparable between the training and the extrapolation areas. The result from our experiment supports this assumption. The portability of the models was successful for any of the attributes in either of the two areas, when the entire training dataset

571 (grid 1) was used in model generation (i.e., the ‘oob’ and the ‘test’ predictions achieved
572 comparable accuracies and there was no indication of bias).

573 However, systematic errors were introduced to the models for both sites when
574 the amount of training data was reduced. This effect was most prominent in the beech-
575 dominated STEI area. For both test sites, absolute deviations of sMEs > 10% were
576 observed for *V* when the least dense inventory grid (with a minimum of 3% of the
577 original number of ground plots) was used. These deviations, mostly related to over-
578 predictions of small values and under-predictions of high values, highlight the common
579 – and well-known – disadvantage of all decision tree-based models: an extrapolation
580 beyond the training data range is not possible (Baccini et al. 2004).

581 In addition to the minimum requirements for the field data, a second issue arises
582 when using public forest data to predict conditions in privately owned forests, at least in
583 the central European context. Often, forest management in privately owned areas is very
584 unlike that done in state-managed forests. This is especially true for small parcels of
585 forest land. This, in turn, can create other differences in, e.g., tree species composition,
586 tree shape (slenderness), stem densities or gross volumes, to name a few. Due to that, a
587 model transferred across “management realms” might result in biased predictions to
588 some extent. This aspect was not covered by our study, as the “test” areas are managed
589 similarly to the “train” areas in the two test sites considered. However, in order to deal
590 with expected differences in the forest structure due to different management strategies,
591 some sort of bias correction could be possible. For this purpose, existing NFI data could
592 be used to derive a correction factor (valid for larger areas) by relating measurements
593 from NFI field plots located in state forests to those measurements from privately
594 owned forests. Still, the applicability of this idea must be tested in a future research
595 project.

So far, most published applications of DAP have focused on area-based modelling. However, recently published studies on DAP data have begun to illuminate the wide field of possible future applications for these kind of data in forestry. Rahlf et al. (2015) segmented an image-based CHM to tree-crown objects and adopted the semi-ITC (individual tree crown) approach to predict forest attributes. St-Onge et al. (2015) similarly delineated individual tree crowns and assessed tree heights and species (using 3d shape and spectral properties). Penner et al. (2015) compared ALS and DAP point clouds to reconstruct tree size distributions and Bohlin et al. (2015) combined image-based CHMs from leaf-on and leaf-off surveys to quantify the broadleaf fraction of total volumes. As reported from Switzerland (Ginzler & Hobi 2015; Waser et al. 2015), nationwide applications of DAP surface models are already in operational use within the national forest inventory.

Conclusion

In this study, we propose a workflow utilizing detailed canopy height information derived from digital aerial images combined with ground inventory information gathered in state-owned forests and regression modelling techniques to quantify forest growing stocks in private woodlands, for which little information is generally available. In both a conifer- and a broadleaf-dominated test site, predictions at nearby forests resulted in reasonable estimates for quadratic mean diameter of the 100 largest trees (d_{100}), basal area weighted mean height of the 100 largest trees (h_{100}), and gross volume (V). However, as our examination of the use of different sample sizes of training data brought to light, considerable biases can be introduced to such predictions when the amount of training data is limited.

The forests selected for our study represent common forest environments in central Europe and thus, the results are of high relevance for achieving sustainable management of these forests. They can be particularly effective tools for helping owners of small private forest estates understand their current resources and options for future management. In view of present policies for converting labile monocultural stands with high stocking levels into more stable and climate-adapted forests, we expect the approach outlined in this paper can help with prioritizing regions where state forest authorities should concentrate their efforts. However, we see the study presented here as merely proof of concept. The proposed approach must undergo further examination before recommending it for broad scale implementation in practice.

Acknowledgements

The research project E49 SAPEX-DLB was funded through the Bavarian State Ministry of Food, Agriculture and Forestry. We thank the Bavarian Administration for Surveying for providing the remote sensing data, the Bavarian State Forest Enterprise for providing the field inventory data, and Laura Carlson for language editing.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Baccini A, Friedl MA, Woodcock CE, Warbington R. 2004. Forest biomass estimation over regional scales using multisource data. Geophys Res Lett. 31. doi:10.1029/2004GL019782.

Bohlin J, Wallerman J, Fransson JE. 2012. Forest variable estimation using photogrammetric matching of digital aerial images in combination with a high-resolution DEM. Scand J For Res. 27:692–699. doi:10.1080/02827581.2012.686625.

644 Bohlin J, Wallerman J, Fransson JE. 2015. Deciduous forest mapping using change
645 detection of multi-temporal canopy height models from aerial images acquired
646 at leaf-on and leaf-off conditions. Scand J For Res:1–27.
647 doi:10.1080/02827581.2015.1130850.
648 Borchert H. 2007. Veränderungen des Waldes in Bayern in den letzen 100 Jahren
649 [Changes of the Bavarian forests in the last 100 years]. LWF Wissen. [Internet].
650 [cited 2016 Mar 25]:42–49. Available from: http://www.lwf.bayern.de/mam/
651 cms04/service/dateien/w58-veraenderung-des-waldes-in-bayern-in-den-letzten-
652 100-jahren.pdf.
653 Breiman L. 2001. Random forests. Mach Learn. 45:5–32.
654 Carson WW, Miller SB, Walker AS. 1996. Automated forest inventory using a digital
655 photogrammetric workstation. Proceedings of the second international airborne
656 remote sensing conference and exhibition:251–257.
657 Fekety PA, Falkowski MJ, Hudak AT. 2015. Temporal transferability of LiDAR-based
658 imputation of forest inventory attributes. Can J For Res. 45:422–435.
659 doi:10.1139/cjfr-2014-0405.
660 FOREST EUROPE. 2015. State of Europe's Forests 2015 [Internet]. Madrid: FOREST
661 EUROPE Liaison Unit Madrid; [cited 2016 Mar 30]. Available from: http://
662 www.foresteurope.org/docs/fullsoef2015.pdf.
663 Ginzler C, Hobi M. 2015. Countrywide Stereo-Image Matching for Updating Digital
664 Surface Models in the Framework of the Swiss National Forest Inventory.
665 Remote Sens. 7:4343–4370. doi:10.3390/rs70404343.
666 Gobakken T, Bollandsås OM, Næsset E. 2015. Comparing biophysical forest
667 characteristics estimated from photogrammetric matching of aerial images and
668 airborne laser scanning data. Scand J For Res. 30:73–86.
669 doi:10.1080/02827581.2014.961954.
670 Haala N. 2013. The Landscape of Dense Image Matching Algorithms. In: Fritsch D,
671 Fritsch D, editors. Photogrammetric Week ,13 (2013); Stuttgart. Berlin,
672 Offenbach: Wichmann.
673 Hirschmüller H. 2008. Stereo Processing by Semiglobal Matching and Mutual
674 Information. IEEE Trans Pattern Anal Mach Intell. 30:328–341.
675 doi:10.1109/TPAMI.2007.1166.

676 Hudak AT, Strand EK, Vierling LA, Byrne JC, Eitel JU, Martinuzzi S, Falkowski MJ.
677 2012. Quantifying aboveground forest carbon pools and fluxes from repeat
678 LiDAR surveys. Remote Sens Environ. 123:25–40.
679 doi:10.1016/j.rse.2012.02.023.
680 Immitzer M, Stepper C, Böck S, Straub C, Atzberger C. 2016. Use of WorldView-2
681 stereo imagery and National Forest Inventory data for wall-to-wall mapping of
682 growing stock. For Ecol Manage. 359:232–246.
683 doi:10.1016/j.foreco.2015.10.018.
684 Joanneum Research. 2015. Remote Sensing Software: All-in-one-solution for all data
685 types and sources [Internet]. Graz: DIGITAL- Institute for Information and
686 Communication Technologies; [cited 2015 Dec 6]. Available from: http://
687 www.remotesensing.at/en/remote-sensing-software.html.
688 Kane VR, McGaughey RJ, Bakker JD, Gersonde RF, Lutz JA, Franklin JF. 2010.
689 Comparisons between field- and LiDAR-based measures of stand structural
690 complexity. Can J For Res. 40:761–773. doi:10.1139/X10-024.
691 Köhl M, Domínguez Torres G, Marchetti M, Tomé M, Martínez de Arano I, Corona P,
692 Thorsen BH, Lasserre B, Pettenella D, Zasada M. 2015. Part II: European
693 Forests: Status, Trends and Policy Responses: Criterion 3: Maintenance and
694 Encouragement of Productive Functions of Forests (Wood and Non-woods)
695 [Internet]. In: Ministerial Conference on the Protection of Forests in Europe,
696 editor. State of Europe's Forests 2015. Madrid: [publisher unknown]; p. 111–
697 132 ; [cited 2016 Mar 30]. Available from: http://www.foresteurope.org/docs/
698 fullsoef2015.pdf.
699 Korpela I, Anttila P. 2004. Appraisal of the mean height of trees by means of image
700 matching of digitised aerial photographs. Photogramm J Finland. 19:23–36.
701 Kuhn M. 2015. caret: Classification and Regression Training [Internet]. [place
702 unknown]: [publisher unknown]; [cited 2015 Dec 6]. Available from: http://
703 cran.r-project.org/package=caret.
704 Kuhn M. 2016. The caret package: Variable Importance [Internet]. [place unknown]:
705 [publisher unknown]; [updated 2016 Jan 19; cited 2016 Mar 17]. Available
706 from: http://topepo.github.io/caret/varimp.html.
707 Kuhn M, Johnson K. 2013. Applied Predictive Modeling. New York: Springer. ISBN:
708 978-1-4614-6848-6.

709 Liaw A, Wiener M. 2002. Classification and Regression by randomForest. R news.
710 2:18–22.
711 McGaughey RJ. 2014. FUSION/LDV: Software for LIDAR Data Analysis and
712 Visualization [Version 3.42]. Seattle, WA: U.S. Department of Agriculture,
713 Forest Service, Pacific Northwest Research Station, University of Washington.
714 [MCPFE] Fourth Ministerial Conference on the Protection of Forests in Europe, editor.
715 2003. Vienna Resolution 2: Enhancing Economic Viability of Sustainable Forest
716 Management in Europe [Internet]. Vienna, Austria: [publisher unknown]; [cited
717 2016 Mar 29]. Available from: http://www.foresteurope.org/docs/MC/MC_
718 vienna_resolution_v2.pdf.
719 Müller J. 2005. Waldstrukturen als Steuergröße für Artengemeinschaften in kollinen bis
720 submontanen Buchenwäldern [Forest Structures as Control Quantity for Species
721 Communities in Colline to Submontane Beech Forests] [Dissertation]. Freising:
722 Technische Universität München.
723 Næsset E. 2002a. Determination of mean tree height of forest stands by digital
724 photogrammetry. Scand J For Res. 17:446–459.
725 Næsset E. 2002b. Predicting forest stand characteristics with airborne scanning laser
726 using a practical two-stage procedure and field data. Remote Sens Environ.
727 80:88–99.
728 Næsset E. 2004. Practical large-scale forest stand inventory using a small-footprint
729 airborne scanning laser. Scand J For Res. 19:164–179.
730 doi:10.1080/02827580310019257.
731 Næsset E, Bollandsås OM, Gobakken T, Gregoire TG, Ståhl G. 2013. Model-assisted
732 estimation of change in forest biomass over an 11year period in a sample survey
733 supported by airborne LiDAR: A case study with post-stratification to provide
734 “activity data”. Remote Sens Environ. 128:299–314.
735 doi:10.1016/j.rse.2012.10.008.
736 Neufanger M. 2011. Richtlinie für die mittel- und langfristige Forstbetriebsplanung in
737 den Bayerischen Staatsforsten: Forsteinrichtungsrichtlinie [Guideline for the
738 intermediate and long-term forest management planning in the Bavarian State
739 Forests]. Regensburg: [publisher unknown].
740 Nilsson M. 1996. Estimation of tree heights and stand volume using an airborne lidar
741 system. Remote Sens Environ. 56:1–7. doi:10.1016/0034-4257(95)00224-3.

742 Packalén P, Maltamo M. 2007. The k-MSN method for the prediction of species-
743 specific stand attributes using airborne laser scanning and aerial photographs.
744 Remote Sens Environ. 109:328–341. doi:10.1016/j.rse.2007.01.005.
745 Penner M, Woods M, Pitt D. 2015. A Comparison of Airborne Laser Scanning and
746 Image Point Cloud Derived Tree Size Class Distribution Models in Boreal
747 Ontario. Forests. 6:4034–4054. doi:10.3390/f6114034.
748 Pitt DG, Woods M, Penner M. 2014. A Comparison of Point Clouds Derived from
749 Stereo Imagery and Airborne Laser Scanning for the Area-Based Estimation of
750 Forest Inventory Attributes in Boreal Ontario. Can J Remote Sens. 40:214–232.
751 doi:10.1080/07038992.2014.958420.
752 R Core Team. 2015. R: A Language and Environment for Statistical Computing
753 [Internet]. Vienna, Austria: R Foundation for Statistical Computing. Available
754 from: http://www.r-project.org/.
755 Rahlf J, Breidenbach J, Solberg S, Astrup R. 2015. Forest Parameter Prediction Using
756 an Image-Based Point Cloud: A Comparison of Semi-ITC with ABA. Forests.
757 6:4059–4071. doi:10.3390/f6114059.
758 Rahlf J, Breidenbach J, Solberg S, Næsset E, Astrup R. 2014. Comparison of four types
759 of 3D data for timber volume estimation. Remote Sens Environ. 155:325–333.
760 doi:10.1016/j.rse.2014.08.036.
761 Rametsteiner E, Sotirov M. 2015. Overall Policies, Institutions and Instruments for
762 Sustainable Forest Management: Part I. In: Ministerial Conference on the
763 Protection of Forests in Europe, editor. State of Europe's Forests 2015. Madrid:
764 [publisher unknown]; p. 39–63.
765 Schmithüsen F, Hirsch F. 2010. Private forest ownership in Europe: Geneva Timber and
766 Forest Study Paper 26 [ECE/TIM/SP/26] [Internet]. Geneva: [publisher
767 unknown]. (no. 26); [cited 2016 Mar 25]. Available from: http://www.unece.org/
768 fileadmin/DAM/timber/publications/SP-26.pdf.
769 Shannon CE. 1948. A Mathematical Theory of Communication. The Bell System
770 Technical Journal. 27:379–423. English.
771 Stepper C, Straub C, Pretzsch H. 2015a. Assessing height changes in a highly structured
772 forest using regularly acquired aerial image data. Forestry. 88:304–316.
773 doi:10.1093/forestry/cpu050.

774 Stepper C, Straub C, Pretzsch H. 2015b. Using semi-global matching point clouds to
775 estimate growing stock at the plot and stand levels: application for a broadleaf-
776 dominated forest in central Europe. Can J For Res. 45:111–123.
777 doi:10.1139/cjfr-2014-0297.
778 St-Onge B, Achaichia N. 2001. Measuring forest canopy height using a combination of
779 LiDAR and aerial photography data. Int Arch Photogramm Remote Sens Spatial
780 Inf Sci. 34:131–137.
781 St-Onge B, Audet F, Bégin J. 2015. Characterizing the Height Structure and
782 Composition of a Boreal Forest Using an Individual Tree Crown Approach
783 Applied to Photogrammetric Point Clouds. Forests. 6:3899–3922.
784 doi:10.3390/f6113899.
785 Straub C, Stepper C. in press. Using Digital Aerial Photogrammetry and the Random
786 Forest Approach to Model Forest Inventory Attributes in Beech- and Spruce-
787 dominated Central European Forests. Photogramm Fernerkund Geoinf.
788 Straub C, Stepper C, Seitz R, Waser LT. 2013. Potential of UltraCamX stereo images
789 for estimating timber volume and basal area at the plot level in mixed European
790 forests. Can J For Res. 43:731–741. doi:10.1139/cjfr-2013-0125.
791 Straub C, Stepper C, Seitz R, Waser LT. 2015. Amtliche Fernerkundungsdaten in der
792 Forstwirtschaft - Anwendungspotential in Bayern [Official Remote Sensing
793 Data in Forestry - Potentials for Application in Bavaria]. In: Wallner A, Seitz R,
794 editors. Der gepixelte Wald -Reloaded. Der gepixelte Wald -Reloaded; 17.-18.
795 März 2015; Freising. 250th ed. Freising: Zentrum Wald-Forst-Holz
796 Weihenstephan; p. 7–17 (Forstliche Forschungsberichte München; vol. 214).
797 Thünen-Institut, editor. 2016. Dritte Bundeswaldinventur - Ergebnisdatenbank [Third
798 National Forest Inventory - Result Database] [Internet]. [place unknown]:
799 [publisher unknown]; [cited 2016 Mar 16]. Available from: https://bwi.info/.
800 van Ewijk K, Treitz PM, Scott NA. 2011. Characterizing Forest Succession in Central
801 Ontario using Lidar-derived Indices. Photogramm Eng Remote Sens. 77:261–
802 269. doi:10.14358/PERS.77.3.261.
803 Vastaranta M, Niemi M, Wulder MA, White JC, Nurminen K, Litkey P, Honkavaara E,
804 Holopainen M, Hyyppä J. 2015. Forest stand age classification using time series
805 of photogrammetrically derived digital surface models. Scand J For Res:1–37.
806 doi:10.1080/02827581.2015.1060256.

807 Vastaranta M, Wulder MA, White JC, Pekkarinen A, Tuominen S, Ginzler C, Kankare
808 V, Holopainen M, Hyyppä J, Hyyppä H. 2013. Airborne laser scanning and
809 digital stereo imagery measures of forest structure: comparative results and
810 implications to forest mapping and inventory update. Can J Remote Sens.
811 39:382–395. doi:10.5589/m13-046.
812 Verkerk PJ, Anttila P, Eggers J, Lindner M, Asikainen A. 2011. The realisable potential
813 supply of woody biomass from forests in the European Union. For Ecol Manage.
814 261:2007–2015. doi:10.1016/j.foreco.2011.02.027.
815 Waser L, Fischer C, Wang Z, Ginzler C. 2015. Wall-to-Wall Forest Mapping Based on
816 Digital Surface Models from Image-Based Point Clouds and a NFI Forest
817 Definition. Forests. 6:4510–4528. doi:10.3390/f6124386.
818 White JC, Stepper C, Tompalski P, Coops NC, Wulder MA. 2015. Comparing ALS and
819 Image-Based Point Cloud Metrics and Modelled Forest Inventory Attributes in a
820 Complex Coastal Forest Environment. Forests. 6:3704–3732.
821 doi:10.3390/f6103704.
822 White JC, Wulder MA, Varhola A, Vastaranta M, Coops NC, Cook BD, Pitt D, Woods
823 M. 2013. A best practices guide for generating forest inventory attributes from
824 airborne laser scanning data using the area-based approach. Victoria, BC,
825 Canada: [publisher unknown]. 50 p. (Information report; FI-X-010). ISBN: 978-
826 1-100-22385-8.
827 White JC, Wulder MA, Vastaranta M, Coops NC, Pitt D, Woods M. 2013. The Utility
828 of Image-Based Point Clouds for Forest Inventory: A Comparison with Airborne
829 Laser Scanning. Forests. 4:518–536. doi:10.3390/f4030518.
830 Zald HS, Wulder MA, White JC, Hilker T, Hermosilla T, Hobart GW, Coops NC. 2016.
831 Integrating Landsat pixel composites and change metrics with lidar plots to
832 predictively map forest structure and aboveground biomass in Saskatchewan,
833 Canada. Remote Sens Environ. 176:188–201. doi:10.1016/j.rse.2016.01.015.
834

Supplementary material

Figure S1: *Image-based point cloud covering the area of one ground plot (500 m^2), normalized to heights above ground using ALS-based terrain height measurements. The canopy is described by the upper part of the point cloud which clearly portrays the shape of two trees/tree groups higher than the closed canopy surface. The blue points represent the ground.*

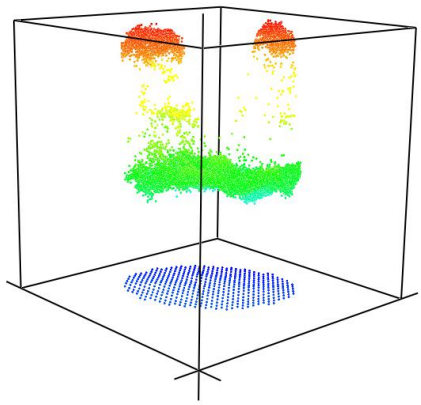


Figure S2: *Boxplots displaying the sample distributions of d_{100}, h_{100} and V, calculated using data from each of the six different inventory grids and the data from the independent test subareas in the two study sites – FRAN and STEI. The widths of the boxes indicate the relative amount of data within each sample. This analysis was conducted to determine whether the reduced sets of training data encompassed the complete range of structural variability existing in the respective test areas.*

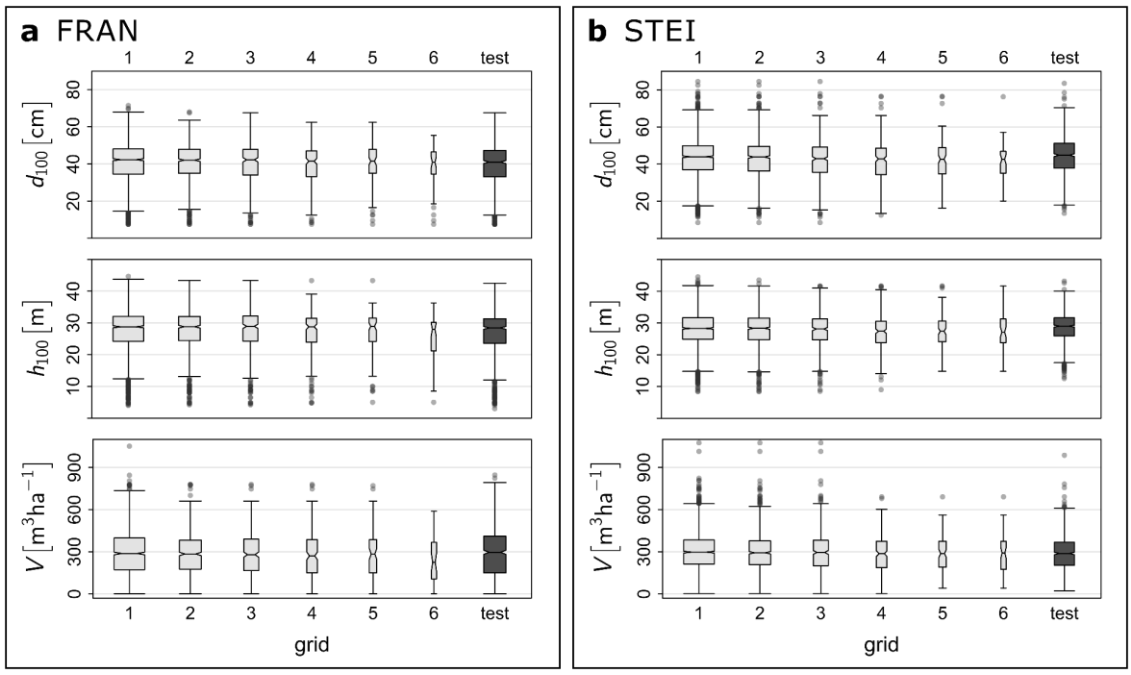
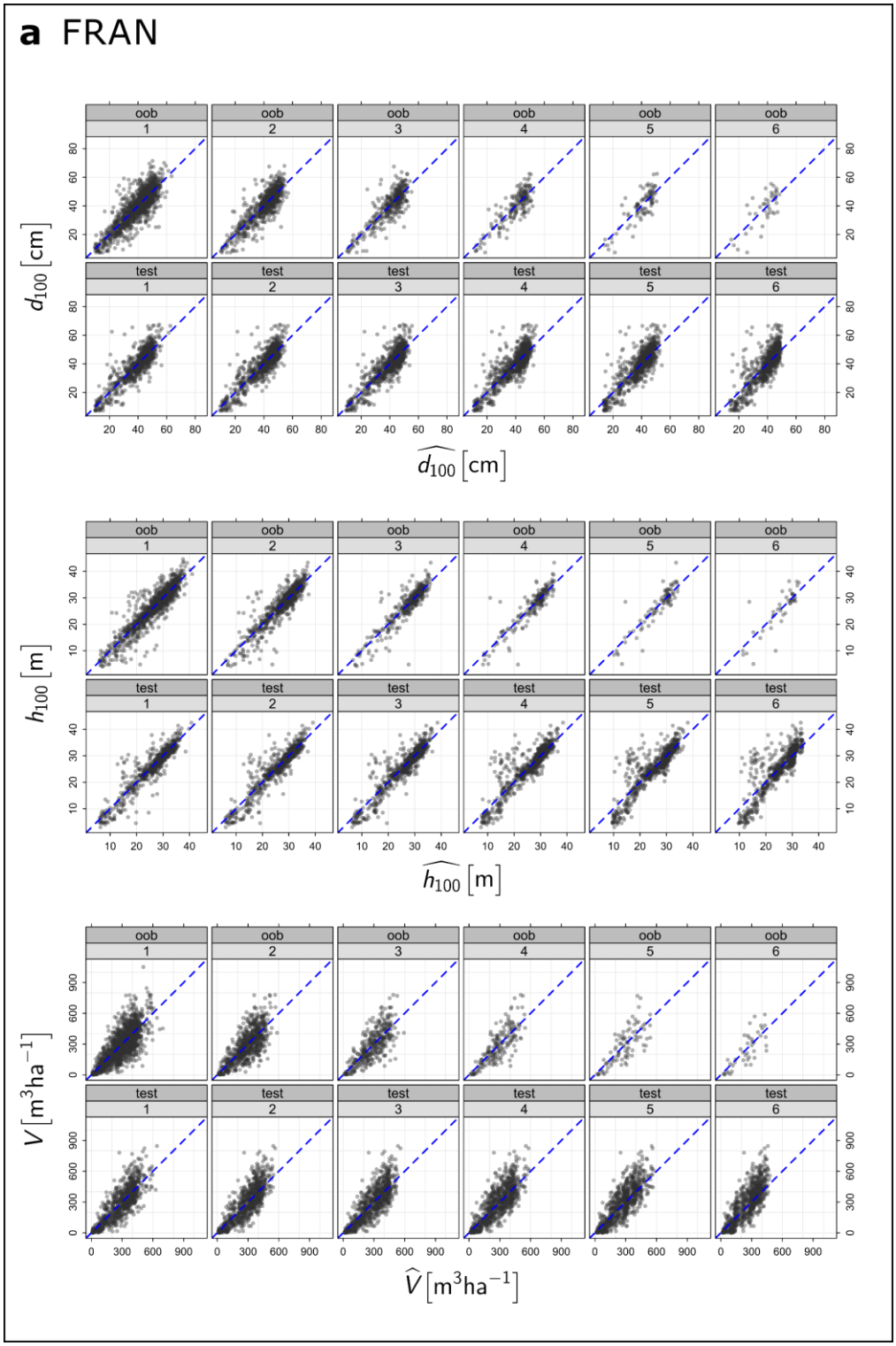
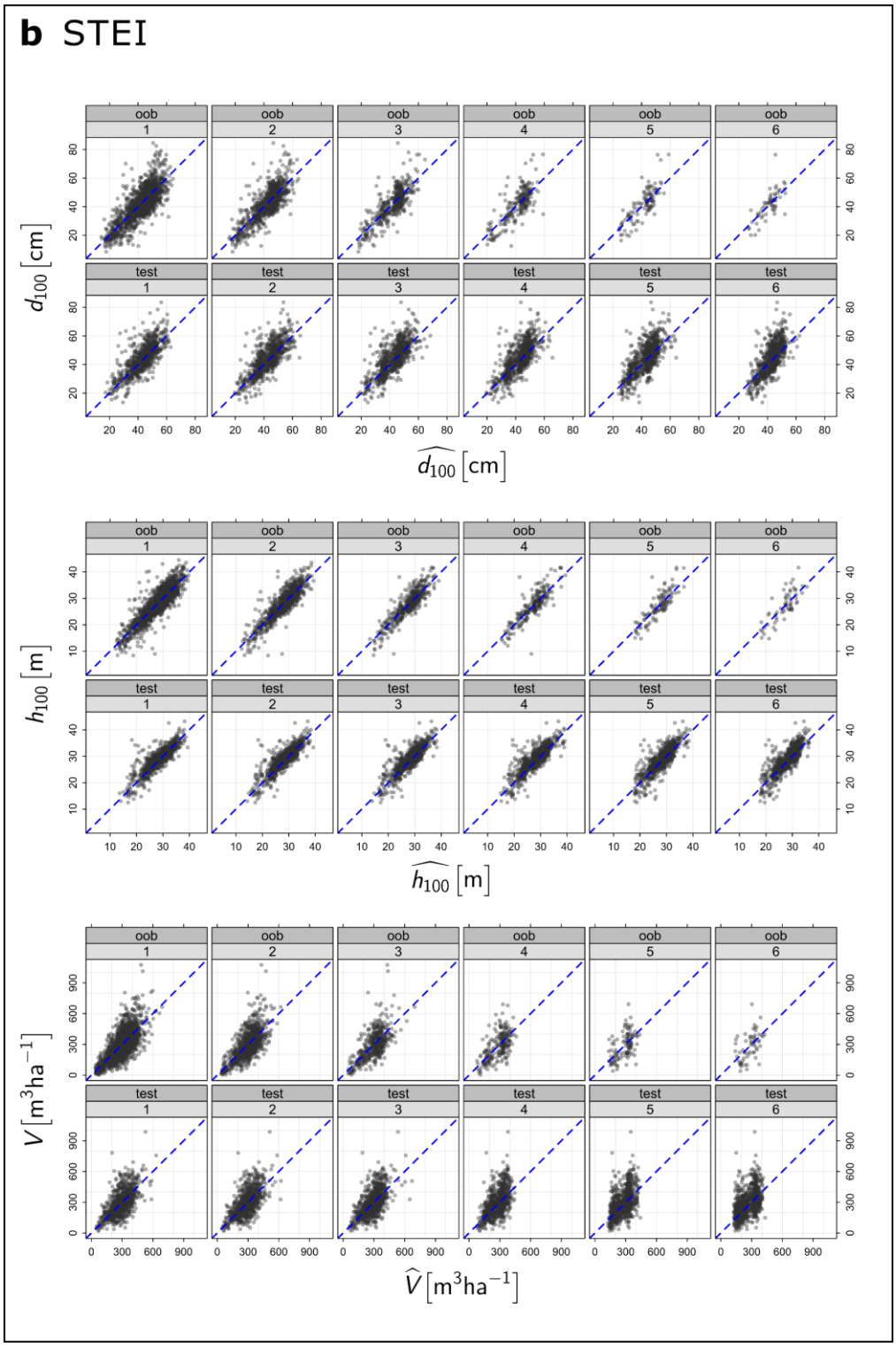


Figure S3: *Scatter plots of predicted vs. observed values for the three selected forest attributes in each of the two study areas (FRAN and STEI). The plots are displayed in panels separating the out-of-bag predictions (oob) and the predictions at the independent test data (test), in each case moving through the different plot densities used for model building (grids 1 - 6).*




b STEI
oob
test
1
2
3
4
5
6
d_{100} [cm]
$\widehat{d_{100}}$ [cm]
h_{100} [m]
$\widehat{h_{100}}$ [m]
V [m^{3}ha^{-1}]
$\widehat{V}$ [m^{3}ha^{-1}]

857