

Using semi-global matching point clouds to estimate growing stock at the plot and stand levels: application for a broadleaf-dominated forest in central Europe

Christoph Stepper, Christoph Straub, and Hans Pretzsch

Abstract: Dense image-based point clouds have great potential to accurately assess forest attributes such as growing stock. The objective of this study was to combine height and spectral information obtained from UltraCamXp stereo images to model the growing stock in a highly structured broadleaf-dominated forest (77.5 km^2) in southern Germany. We used semi-global matching (SGM) to generate a dense point cloud and subtracted elevation values obtained from airborne laser scanner (ALS) data to compute canopy height. Sixty-seven explanatory variables were derived from the point cloud and an orthoimage for use in the model. Two different approaches — the linear regression model (lm) and the random forests model (rf) — were tested. We investigated the impact that varying amounts of training data had on model performance. Plot data from a previously acquired set of 1875 inventory plots was systematically eliminated to form three progressively less dense subsets of 937, 461, and 226 inventory plots. Model evaluation at the plot level (size: 500 m^2) yielded relative root mean squared errors (RMSEs) ranging from 31.27% to 35.61% for lm and from 30.92% to 36.02% for rf. At the stand level (mean stand size: 32 ha), RMSEs from 14.76% to 15.73% for lm and from 13.87% to 14.99% for rf were achieved. Therefore, similar results were obtained from both modeling approaches. The reduction in the number of inventory plots did not considerably affect the precision. Our findings underline the potential for aerial stereo imagery in combination with ALS-based terrain heights to support forest inventory and management.

Key words: digital aerial images, dense image matching, forest attribute estimation, random forests, linear regression, feature selection, wall-to-wall mapping, inventory design, sampling density.

Résumé : Les nuages de points basés sur des images denses présentent un grand potentiel pour évaluer avec précision les attributs forestiers tels que le volume sur pied. L'objectif de cette étude était de combiner la hauteur et l'information spectrale obtenues à partir d'images stéréo captées avec l'UltraCamXp pour modéliser le volume sur pied dans une forêt très structurée dominée par des feuillus ($77,5 \text{ km}^2$) dans le sud de l'Allemagne. Nous avons utilisé la méthode d'appariement semi-globale afin de générer un nuage de points dense, puis nous avons soustrait les valeurs d'élévation obtenues à partir d'un scanner laser aéroporté (SLA) afin de calculer la hauteur du couvert forestier. Soixante-sept variables explicatives ont été obtenues à partir du nuage de points et d'une ortho-image pour être utilisées dans le modèle. Deux approches différentes ont été testées : un modèle de régression linéaire (ml) et un modèle de forêts aléatoires (fa). Nous avons étudié l'impact de différentes quantités de données d'entraînement sur la performance des modèles. Les données des placettes provenant d'un inventaire réalisé précédemment, comptant un ensemble de 1875 placettes, ont été systématiquement éliminées pour former progressivement trois sous-ensembles de moins en moins denses comptant respectivement 937, 461 et 226 placettes d'inventaire. L'évaluation des modèles à l'échelle de la placette (taille : 500 m^2) a donné des erreurs quadratiques moyennes relatives allant de 31,27 % à 35,61 % pour ml et de 30,92 % à 36,02 % pour fa. À l'échelle du peuplement (taille moyenne du peuplement : 32 ha), des erreurs quadratiques moyennes relatives de 14,76 % à 15,73 % pour ml et de 13,87 % à 14,99 % pour fa ont été observées. Ainsi, des résultats similaires ont été obtenus à partir des deux approches de modélisation. La réduction du nombre de placettes d'inventaire n'a pas eu d'incidence considérable sur la précision. Nos résultats soulignent le potentiel de l'imagerie aérienne stéréo combinée à des hauteurs de terrain obtenues au moyen d'un SLA pour appuyer l'inventaire et la gestion des forêts. [Traduit par la Rédaction]

Mots-clés : images aériennes digitales, appariement d'images denses, estimation d'attributs forestiers, forêts aléatoires, régression linéaire, sélection de caractéristiques, cartographie en couverture complète, plan d'inventaire, densité d'échantillonnage.

1. Introduction

Sustainable management of the multiple functions and services of highly structured temperate forests requires spatially explicit information about their state and development. In recent decades, a lot of effort has been put into studies in which airborne laser scanning (ALS) was investigated for characterizing both the vertical and horizontal structures of different forest types. Further-

more, many such studies have focused on the integration of ALS into forest inventories. Comprehensive review articles covering this field have been published by Lim et al. (2003), Næsset et al. (2004), Hyppä et al. (2008), and Wulder et al. (2008). Detailed insights into the concepts of ALS and various forestry case studies are described in Maltamo et al. (2014).

A great number of studies have demonstrated the applicability of ALS for estimating various forest attributes using the area-

Received 24 June 2014. Accepted 13 October 2014.

C. Stepper and C. Straub. Bavarian State Institute of Forestry (LWF), Department of Information Technology, Research Group: Remote Sensing, Hans-Carl-von-Carlowitz-Platz 1, D-85354 Freising, Bavaria, Germany.

H. Pretzsch. 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, Bavaria, Germany.

Corresponding author: Christoph Stepper (e-mail: Christoph.Stepper@lwf.bayern.de).

based approach (ABA). [White et al. \(2013a\)](#) presented a synoptic view of the ABA and the processing steps that it requires. Within the ABA, several variables related to the statistical distribution of canopy heights, derived either from point clouds or from surface models, are used as predictors for the estimation of forest attributes (e.g., growing stock (GS)). Basically, two different types of statistical methods can be distinguished: (i) parametric methods, most often linear regression (lm; used by, e.g., [Nilsson \(1996\)](#), [Means et al. \(2000\)](#), [Næsset et al. \(2005\)](#), [Maltamo et al. \(2006\)](#), and [Lindberg and Hollaus \(2012\)](#)) and (ii) nonparametric methods, where various k -nearest neighbor (kNN) approaches or random forests (rf) were mostly implemented (used by, e.g., [Packalen and Maltamo \(2006\)](#), [Hudak et al. \(2008\)](#), [Breidenbach et al. \(2010\)](#), [Latifi et al. \(2010\)](#), [Hudak et al. \(2012\)](#), and [Penner et al. \(2013\)](#)).

A sufficient amount of field data is necessary for successful model calibration, i.e., to link remotely sensed data to the forest attribute(s) of interest. Field sampling strategies must, therefore, strive to capture the entire range of variability over the total area being considered. However, as fieldwork is costly and time consuming, one aims to reduce the number of inventory plots gathered to the number that is actually essential for sound estimations. Up to now, just a few studies have focused on this task, and with regard to airborne remote-sensing systems, ALS was employed most often. [Gobakken and Næsset \(2008\)](#) assessed the effect of reducing the number of field plots used to make stand-level predictions and reported only minor accuracy losses when the number of plots was reduced by up to 50%. [Hawbaker et al. \(2009\)](#) and [Maltamo et al. \(2011\)](#) emphasized the benefits of a stratified sampling design, using the ALS data as auxiliary information for stratification. In the latter study, a substantial reduction in the number of plots was possible without adverse effects on model performance. [Strunk et al. \(2012\)](#) evaluated the effect of different sample sizes on the precision of a regression estimator for selected forest variables, e.g., volume. The authors applied a simulation approach and found that reducing the sample size induces noticeable losses of model precision.

Recently, several research studies have emphasized the potential of dense point clouds derived from aerial stereo images as a possible alternative to ALS (e.g., [Leberl et al. 2010](#)). This development has been instigated by a number of different innovations ([Haala and Rothermel 2012](#)): (i) the introduction of digital cameras in aerial surveys allows for the acquisition of images with a large degree of overlap and a good signal-to-noise ratio, (ii) recent improvements in matching algorithms such as pixelwise semi-global matching (SGM; [Hirschmüller 2008](#)) enable the generation of dense photogrammetric point clouds with high levels of accuracy, and (iii) rapid progress in hardware capacity enables the processing of large image blocks and the derived 2.5D products.

A summary of recently published studies dealing with image matching and forest attribute estimation is provided in [Table 1](#). These studies, mostly conducted in boreal forests, have verified the potential of image-based 2.5D information as an alternative to ALS for use in forest attribute estimation. [Bohlin et al. \(2012\)](#) followed the ABA suggested by [Næsset \(2002\)](#) to estimate tree heights, stem volume, and basal area in a coniferous forest in southern Sweden. They derived metrics related to canopy height, density, and texture from the image-based height data and used them as explanatory variables in regression analysis. For two other studies, both conducted on the same test site in southern Finland, an image-based canopy height model (CHM) served as the input for the extraction of explanatory variables at the plot level. In both studies, different kNN approaches were applied as statistical methods for imputation ([Järnstedt et al. 2012](#); [Vastaranta et al. 2013](#)). Working at a test site with mixed forests in southern

Germany, [Straub et al. \(2013a\)](#) used the spectral information of airborne imagery as auxiliary data for a stratified linear regression. Distinct models for conifer- and broadleaf-dominated sample plots were generated to estimate the GS and basal area by means of height and canopy cover metrics together with a surface-roughness metric, derived from image-based height measurements. [Nurminen et al. \(2013\)](#) used explanatory variables from image-based point clouds and examined the performance of the rf algorithm for the prediction of forest attributes at the plot level.

[White et al. \(2013b\)](#) outlined both the benefits (e.g., additional spectral information, lower costs, high return frequency) and the limitations (e.g., necessity for an ALS-based digital terrain model (DTM) for height normalization and restriction to 2.5D information of the outer canopy surface) of image-based point clouds for forest inventory applications. They emphasized the applicability of these data in the ABA for forest attribute estimation and highlighted some key future research questions. According to the authors, more research is needed in a variety of forest types and stand conditions to prove the capability of image-based point clouds in the ABA. Moreover, [Straub et al. \(2013a\)](#) pointed out that image-based height information gained from state-of-the-art matching algorithms such as SGM might potentially describe the canopy surface in more detail, which is desirable when using these data to estimate GS. Furthermore, few studies have used both height- and structure-related information in combination with spectral data from stereo imagery as auxiliary information.

Therefore, we chose to estimate the GS at the plot and stand levels, using UltraCamXp stereo images for a large and highly structured broadleaf-dominated forest area (77.5 km^2) in southern Germany. GS here refers to the total merchantable timber volume in cubic metres per hectare ($\text{m}^3 \cdot \text{ha}^{-1}$), defined as the inside-bark volume of all stems and branches with a top diameter equal to or greater than 7 cm. The main objectives of the present study were as follows:

- application of SGM to generate a very dense 2.5D point cloud that would provide a detailed characterization of the surface structure of the forest canopy;
- examination of several spectral features computed from the orthoimagery as additional explanatory variables for estimating GS, in combination with height and structural variables derived from the image-based height data;
- comparative testing of rf vs. ordinary least squares lm to estimate GS at the plot level, as well as at the stand level;
- investigation of the effect of different sampling densities, i.e., systematically reduced numbers of terrestrial inventory plots, on the precision and accuracy of the estimation of GS.

As GS is one of the most important forest attributes, we focused on this characteristic throughout the study. All remote-sensing data (i.e., the aerial stereo images and the ALS-based DTM) were acquired as part of the official surveys conducted by the Bavarian State Office for Surveying and Geoinformation. The field measurements employed in this study were obtained from the regularly updated inventory data gathered by the Bavarian State Forest Enterprise. Thus, the findings of this study are relevant both for forest research and for practical applications.

2. Materials

2.1. Study area

This study was conducted within the Steigerwald test site ($49^\circ 55' \text{N}, 10^\circ 35' \text{E}$) in the northwestern part of Bavaria, Germany ([Fig. 1](#)). The test site comprises a total area of 120 km^2 , and the terrain is best described as a high plain (consisting mainly of Upper Triassic sandstone layers) gently sloping eastward and interspersed with steep valley cuts (see [Supplementary Fig. S1a¹](#)).

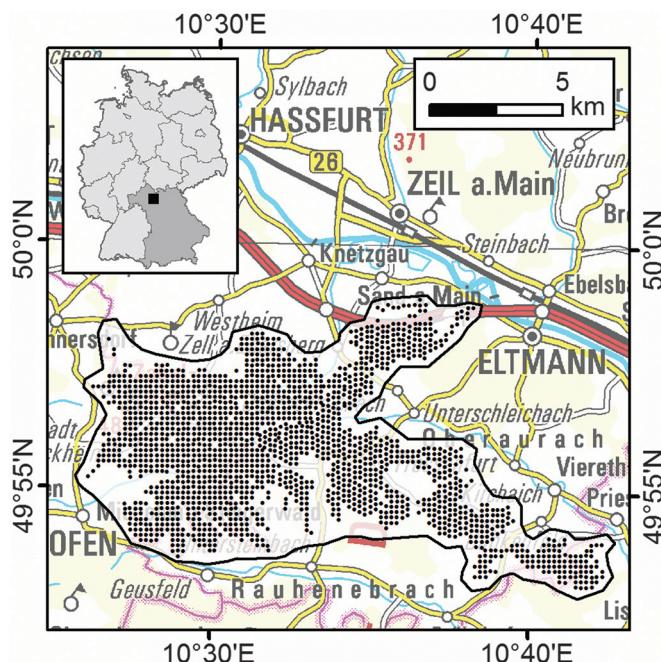
¹Supplementary data are available with the article through the journal Web site at <http://nrcresearchpress.com/doi/suppl/10.1139/cjfr-2014-0297>.

Table 1. Recently published studies where digital aerial imagery was used to estimate growing stock via the ABA.

Study	Bohlin et al. (2012)	Järnstedt et al. (2012)	Nurminen et al. (2013)	Straub et al. (2013a)	Vastaranta et al. (2013)
Test site	Southern Sweden	Southern Finland	Central Finland	Southern Germany	Southern Finland
Forest type	Coniferous hemi-boreal	Conifer dominated	Conifer dominated	Mixed central-European	Conifer dominated
Camera system	DMC	UltracamXp	UltracamD	UltracamX	UltracamXp
Matching software	Match-T	SOCET SET (NGATE)	SOCET SET (NGATE)	LPS (eATE)	SOCET SET (NGATE)
No. of plots	344	402	89	225	500
Plot radius (m)	10	9.77	Variable (5.64–20)	12.62	9.77
Observed growing stock ($\text{m}^3 \cdot \text{ha}^{-1}$)					
Minimum	0.5	0.4	44.7	0	43.6
Maximum	804	615.4	380.7	968.4	586.2
Mean	262	160.1	179	321.99	187.7
SD	—	—	92.5	225.6	94.3
Regression type	Multiple linear regression	kNN (euclidean distance)	Random forests	Stratified multiple linear regression	kNN (random forests)
Validation method	Leave-one-stand-out cross-validation	Leave-one-out cross-validation	Out-of-bag cross-validation	Leave-one-out cross-validation	Out-of-bag cross-validation
RMSE ($\text{m}^3 \cdot \text{h}^{-1}$)	32.8	60.06	41.03	121.99	46.1
RMSE (%)	13.1	40.39	22.62	37.89	24.5

Note: The RMSE values from the Bohlin et al. (2012) study were calculated at the stand level.

Fig. 1. Geographic location of the Steigerwald test site in northwestern Bavaria, Germany. The inset map displays the location of the test site in Germany (light grey) and in Bavaria (dark grey). The detailed map depicts the outer boundary of the test site and the spatial distribution of the full set of forest inventory plots collected in the test site by the Bavarian State Forest Enterprise. Note that this figure is provided in color online.



Elevations range from 220 m to 500 m above sea level (a.s.l.) (mean elevation: 336 m a.s.l.). The land cover within the area is largely forest, which comprises a total of 77%. The remainder is a mixture of agricultural land (21%) and settlement, which includes infrastructure (2%). With regard to the forested area, about 77.5 km² are contiguously state-owned land and can be characterized as managed forest with a great variety of stand development stages (see descriptive statistics of the mean age of overstory trees in Table 3). Interpretation of aerial photographs of the test site revealed various crown closures of the forest stands. Some stands

are of great ecological relevance with minimized harvesting (Mergner 2013). Broadleaf tree species are dominant within the test site, whereas coniferous tree species sum to a total portion of only 20%. The most common tree species are European beech (*Fagus sylvatica* L., 42%), sessile oak (*Quercus petraea* (Mattuschka) Liebl., 22%), and Scots pine (*Pinus sylvestris* L., 10%).

2.2. Field data

The field data used in this study were recorded in 2010 as part of the regular forest inventory conducted by the Bavarian State Forest Enterprise (Neufanger 2011). The permanent field plots where the data were collected are laid out in a regular grid pattern of 200 m × 200 m (Fig. 1), with each field plot representing roughly 4 ha of forested land. The plot centers are permanently marked, and during the inventory, each particular plot location was georeferenced using a Trimble GeoExplorer XT GPS device. Generally, the GPS measurements achieved accuracies with maximum deviations of ±3–5 m (H. Grünvogel, personal communication, 4 August 2014).

For each sample plot, all relevant tree attributes were recorded within one of three concentric circles according to certain thresholds of diameter at breast height (DBH; 1.3 m). The thresholds employed are listed in Table 2. Tree species, DBH, age class, and vertical-layer affiliation (i.e., regeneration, understory, and over-story) were recorded for all sample trees taller than 1.3 m. Tree heights were measured for at least two trees from every possible combination of species, age class, and vertical layer occurring at that sample plot. Using the measured tree heights and DBHs, generalized height-diameter models (established in Bavaria based on the equation by Petterson (1955)) were adjusted at each plot. By this method, tree heights could be modeled for all remaining trees in the plots for which no terrestrial height measurement was recorded. Single-tree volumes were estimated using allometric models, which were previously established in Bavaria. These models require the DBH and tree height as independent variables, as well as tree species specific form factors (for a detailed explanation of the form factor, please refer to van Laar and Akça (2007)).

The finally available individual-tree data from each plot were used to compile per-hectare values for various forest-related attributes, e.g., dominant tree height (H_{dom}), basal area (BA), and growing stock (GS). In total, data were recorded from 1878 circular 500 m² field inventory plots. At the outset of our investigation, we eliminated three of these original plots from all further investigations, as a visual examination of the orthoimages showed that the

Table 2. Radii and DBH thresholds for the concentric circles assembling a circular inventory sample plot.

Circle No.	Radius (m)	Area (m^2)	DBH (cm)
1	2.82	25	<12
2	6.31	125	12–29.9
3	12.62	500	≥30

Note: DBH, diameter at breast height.

plots were clear-cut between when the field measurements were taken in 2010 and when the aerial survey was conducted in 2011. Descriptive statistics from the remaining 1875 inventory plots are shown in Table 3.

The state-owned forested land within the test site was subdivided into relatively homogeneous units, i.e., forest stands, for forest management and silvicultural planning purposes. In accordance with Neufanger (2011), these stands were delineated manually by forest-management professionals of the Bavarian State Forest Enterprise, based on aerial-image interpretation and field surveys in 2012, considering characteristics such as tree species composition, developmental stage, and vertical, as well as horizontal, stand structures. An intersection of the stand geometries with the sample-plot locations of the regular forest inventory was conducted to obtain a reasonable selection of stands that were later used for the assessment at the stand level. We tested thresholds of at least four, five, and six complete sample plots per stand for the selection of reference stands. To arrive at a good balance for the trade-off between covering the range of observed GSs for the different stands and a minimum amount of sample plots per stand for solid stand calculations, we chose the five sample plot threshold for selecting the stand reference data in our study. In total, 94 stands fulfilled this requirement and were further used in the stand-level assessment.

2.3. Remote-sensing data

We used the remote-sensing data provided by the Bavarian Office for Surveying and Geoinformation for our study. The aerial imagery was acquired during leaf-on conditions in May 2011 as part of the regularly scheduled aerial survey, which is updated every 3 years (Stössel 2009). To cover the entire test site, images from two different flights both conducted on 7 May 2011 using the same Vexcel UltraCamXp camera (Leberl et al. 2012) were processed, resulting in a ground sampling distance of 0.20 m. For both flights, imagery with along- and across-track overlaps of 75% and 30%, respectively, was recorded. This resulted in a total set of 163 stereo images from six different flight lines (see Supplementary Fig. S1b¹). The data provided included both panchromatic images (PAN) and PAN-sharpened multispectral images (blue, green, red, and near infrared) of the same radiometric resolution (12 bit).

To achieve canopy heights from the image-based 2.5D point cloud, additional bare-earth height information was required. An ALS-based DTM (1 m resolution; see Supplementary Fig. S1a¹) derived from a topographic mapping survey in 2009 with a last return density of $1.37 \cdot m^{-2}$ was available for the test site. According to the data provider, 95% of the last returns have a vertical accuracy of ≤ 0.16 m, and the horizontal accuracy is given as ± 0.5 m.

3. Methods

In the following sections, we will describe the different processing steps in detail: first, derivation of the remote-sensing products, i.e., a dense point cloud, a CHM, and the orthoimagery; second, calculation of explanatory variables for the inventory plots; third, model development for the different inventory grid densities and evaluation at the plot level; and finally, wall-to-wall application of the models and evaluation at the stand level.

3.1. Derivation of remote-sensing products

We applied the SGM approach to compute a dense image-based point cloud. This algorithm, developed by Hirschmüller (2008), approximates a global cost minimization between corresponding pixels in base and match images by minimizing matching costs, which are aggregated along a certain number of 1D path directions through the image. Matching costs measure the dissimilarity between corresponding pixels, and disparity jumps are penalized by additional costs. This induces smooth disparities along the search path (Hirschmüller 2011). By these means, the pixelwise SGM approach provides computation of dense point clouds within reasonable run times on large images (Haala 2011). SGM, as implemented in the Remote Sensing Software Package Graz (RSG, version 7.41; see Joanneum Research (2014)) was used in this study. The matching procedure in RSG is hierarchically structured as coarse to fine matching. An initial coarse, top-level disparity is calculated based on the additional information from the available ALS-based DTM, with consequent disparity calculations on four image pyramid levels. The cost function implemented in RSG and used in the current study compares two image patches and is defined as the Hamming distance of the two census transforms within 9×9 pixel windows on the epipolar-rectified images.

We used the panchromatic band of the digital stereo images for the matching procedure. Due to the low across-track overlap of the aerial images, we used only stereo-image pairs from the along-track overlap for computation. In total, 308 stereo-image pairs were selected for processing. From the dense matching with SGM, we achieved a mean point density for the stereo-image pairs of $21.35 \text{ points} \cdot m^{-2}$. Aggregating the data from the overlapping stereo-image pairs resulted in a final point cloud with a mean density of $103.65 \text{ points} \cdot m^{-2}$.

Using the image-based point heights, we computed a gridded digital surface model (DSM) of 1 m spatial resolution, in accordance with the DTM. The algorithm employed, implemented in the software FUSION (McGaughey 2013), assigns the elevation of the highest point within each $1 \text{ m} \times 1 \text{ m}$ grid cell to the grid-cell center. All points located more than 50 m above the DTM were considered to be outliers and were eliminated. We applied a 3×3 median filter to the DSM, as a smoothed surface was essential for the orthorectification of the multispectral aerial images to prevent blurred orthoimages. We combined all single orthoimages into one mosaic using most-nadir seamlines and resampled the image to 1 m resolution to coincide with the resolution of the DSM (see Supplementary Fig. S1d¹). To ensure that no systematic vertical deviation existed between the image-based heights and the ALS-based DTM, we computed the mean height difference between DSM and DTM heights for nine street polygons in the test site. The mean deviation for these control polygons was 0.17 m (standard deviation (SD), 0.29 m), which we consider to be acceptable for the purpose of our study.

Additionally, the original image-based point cloud was normalized by subtracting the ALS-based DTM heights; again, we eliminated all points located more than 50 m above the DTM as they were considered outliers. Using the normalized point cloud, we computed a gridded CHM (see Supplementary Fig. S1c¹) of 1 m spatial resolution. For consistency, we applied the same method as for the DSM computation.

3.2. Calculation of explanatory variables for the inventory plots

Based on these derived remote-sensing products, a total of 67 explanatory variables were computed for each of the 500 m^2 inventory plots (see, for example, extracted 2.5D point cloud for one inventory plot in Supplementary Fig. S2¹). A list of all variables including a concise explanation of the computation is provided in Supplementary Table S1¹.

The FUSION cloudmetrics command was used for the calculation of the point-cloud variables. For a detailed description of the

Table 3. Descriptive statistics calculated from the 1875 field inventory plots gathered by the Bavarian State Forest Enterprise in the Steigerwald test site and used to train the various models investigated in this study: mean age of overstory trees (age), dominant tree height (H_{dom}), basal area (BA), and growing stock (GS).

Forest attribute	Minimum	1st quartile	Median	3rd quartile	Maximum	Mean	SD
Age (years)	3.0	62.0	104.0	138.0	230.0	100.6	44.9
H_{dom} (m)	0.0	24.3	27.8	31.2	43.0	27.2	6.3
BA ($\text{m}^2 \cdot \text{ha}^{-1}$)	0.0	22.3	28.6	35.7	92.0	29.2	10.9
GS ($\text{m}^2 \cdot \text{ha}^{-1}$)	0.0	208.2	294.7	382.6	1075.5	297.9	137.5

Note: The minimum values of 0.0 were recorded in areas with stocking consisting only of trees with height <1.3 m (DBH), which existed at some inventory plots. SD, standard deviation.

equations implemented, please refer to McGaughey (2013). In accordance with Nilsson (1996) and Næsset (2002), all height measurements below 2 m were not considered in the calculation to exclude noncanopy measurements (e.g., points representing the ground, stones, or shrubs). For the extraction of variables from the gridded CHM and the orthoimage, we developed our own software routines. Prior to conducting any further calculations, the 67 explanatory variables were centered and scaled so that all predictors had a zero mean and a common standard deviation of one (Kuhn and Johnson 2013).

According to White et al. (2013a), multicollinearity among explanatory variables can result in unstable predictions. Therefore, we employed correlation analysis to identify potential linear relationships between the extracted variables. Pairwise correlations were computed for all variables, and the resulting correlation matrix was visually examined using the corrplot R package (Wei 2013). Here, each pairwise correlation is colored according to its magnitude, and the predictors are grouped into “clusters” of collinearity using the hierarchical clustering approach described in Everitt et al. (2011).

Consequently, the explanatory variables were subdivided into the following groups:

- (1) 26 height (H) metrics calculated from the point clouds: minimum, maximum, mean, mode, percentile values (1st, 5th, 10th, 20th, 25th, 30th, 40th, 50th, 60th, 70th, 75th, 80th, 90th, 95th, and 99th), and proportion values for various height strata (0–5 m, 5–10 m, 10–15 m, 15–20 m, 20–30 m, 30–40 m, and 40–50 m);
- (2) five variability (V) metrics based on the point clouds: canopy relief ratio, absolute average deviation, L-moment coefficient of variation, L-moment skewness, and L-moment kurtosis;
- (3) 10 canopy cover (CC) metrics derived from the point clouds and the CHM: ratios of either points or pixels above certain height thresholds (points: 2 m, mean, and mode; CHM: 2 m, 5 m, 10 m, 15 m, 20 m, 30 m, and 40 m);
- (4) 26 spectral (SP) metrics derived from the orthoimage: mean and standard deviation for all bands (blue, green, red, and near infrared) and for the artificial normalized-difference vegetation index (NDVI) layer and texture metrics (contrast, correlation, energy, and homogeneity) extracted from the grey-level co-occurrence matrices (GLCM; averaged for the four directions 0°, 45°, 90°, and 135°) for all bands, in accordance with Haralick et al. (1973).

3.3. Model development for the different inventory grid densities and evaluation at the plot level

The explanatory variables derived from the aerial stereo images are highly prone to multicollinearity. We tried to overcome the modeling problems related to that issue by implementing two different approaches for the estimation of the GS, namely, the (parametric) ordinary least squares linear regression model (lm), with an associated feature selection, and the (nonparametric) random forests (rf) approach. We developed our modeling procedure in R (version 3.0.1, 64 bit), an open-source statistical software (R Core Team 2013), and

used the packages caret (Kuhn 2013) and bootStepAIC (Rizopoulos 2009) to implement the different modeling steps.

Resampling methods such as cross-validation can be used to produce appropriate estimates of model performance. As suggested by Kuhn and Johnson (2013), we applied 10-fold cross-validation repeated five times so that 50 different held-out sets were used to assess the model performance. Kuhn and Johnson (2013) stated that repeating cross-validation can be favorable to effectively increase the precision of model estimates. To determine the estimation precision and accuracy of the different models examined in this study, the root mean squared error (RMSE) and bias were calculated for k sets of held-out samples, referred to as folds f (eqs. 1 and 2; in accordance with Straub et al. 2010). The resulting figures, i.e., the RMSE_f and bias_f values for the 50 folds, were then aggregated to mean values, i.e., the RMSE and bias (absolute and relative), which correspond to the repeated cross-validation estimates of model performance (eqs. 3–6).

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

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

$$(3) \quad \text{RMSE} = \frac{\sum_{j=1}^k \text{RMSE}_{f,j}}{k}$$

$$(4) \quad \text{RMSE (\%)} = \frac{\text{RMSE}}{\text{mean } (\bar{y}_f)} \times 100$$

$$(5) \quad \text{bias} = \frac{\sum_{j=1}^k \text{bias}_{f,j}}{k}$$

$$(6) \quad \text{bias (\%)} = \frac{\text{bias}}{\text{mean } (\bar{y}_f)} \times 100$$

Here, y_i is the observed value and \hat{y}_i is the predicted value for the i th of n sample plots in one of the k held-out sets and \bar{y}_f is the mean of n observed values in that held-out set.

3.3.1. Ordinary least squares linear regression model

Considering lm, linear relationships were used to predict the GS. In addition, we tested exponential relationships for modeling. However, using these for modeling did not perform better with respect to the cross-validated RMSEs; therefore, linear additive models were ultimately selected for use in this study (eq. 7).

$$(7) \quad \text{GS} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

where β_i are the unknown regression coefficients, X_i are the independent variables derived from the remote-sensing data, and ε is the error term.

Feature selection was constructed as a two-phase procedure, and the variable groups described in section 3.2 were used. The H metrics were tested first because previous correlation analysis indicated that these variables had the highest predictive potential.

Therefore, in the first phase, the initial model was built by evaluating each H metric separately as the sole explanatory variable for the response variable GS. The model with the smallest RMSE was chosen as the best of these models, calculated as shown in eq. 3. Next, each possible combination of a V metric and a CC metric, together with the previously selected H metric, was used to create a model. Here again, the best-performing model with respect to the RMSE was selected. Finally, this variable set, i.e., the “best” H metric and the “best” combination of structural metrics (one V metric and one CC metric), was integrated with each of the SP metrics separately to find the best model. Therefore, the final selection is parsimonious as it consists of exactly one explanatory variable from each of the four groups.

In the second phase, the statistical significance of each of the selected predictor variables was evaluated via a bootstrapped application of the stepAIC algorithm (R package MASS; Venables and Ripley 2007). Using bootStepAIC (Rizopoulos 2009), the final model from phase one was refitted to 100 bootstraps from all sample plots, and a subsequent stepwise regression (direction: both; penalty criterion: BIC) was performed on each bootstrap. In accordance with Austin and Tu (2004), the results from the bootstraps were summarized, and only explanatory variables with statistical significance ($\alpha = 0.01$) were chosen for inclusion in the final linear model.

3.3.2 Random forests

The rf approach, a popular ensemble regression tree algorithm first published by Breiman (2001), was applied as the nonparametric modeling approach. This technique overcomes the problems of multidimensionality and multicollinearity. In rf modeling, different regression trees are built based on random subsets of the available data, and a subsequent majority vote of these decision trees is used to form the final regression model. Detailed descriptions of rf modeling can be found in Genuer et al. (2008) and Hastie et al. (2009).

The modeling procedure was implemented in caret, which makes use of the package functionalities of randomForest (Liaw and Wiener 2002). We followed the recommendation of Kuhn and Johnson (2013) with regard to the estimation procedure and fixed the number of regression trees for the forest (n_{tree}) at 1000, which gives a good balance between computational expense and model performance. Previous testing using a larger number of decision trees did not gain higher model accuracies.

The number of variables randomly selected from the total set P (here 67 variables) at each node (commonly referred to as m_{try}) is the tuning parameter used to optimize the performance of the rf model. The rf models built with various values for m_{try} ranging from 2 to P did not show substantial differences in terms of the RMSEs achieved. Due to this and the fact that optimization runs for m_{try} are computationally intensive, we used a setting of $m_{try} = P/3$ for rf in regression mode, as recommended by Breiman (2001).

3.3.3. Systematic reduction of the number of inventory plots used to train the model

To investigate the impact that varying amounts of training data had on model performance, the original sample grid (200 m × 200 m) was systematically reduced in three iterative steps. In each of the iterative steps, every second inventory plot was eliminated from the previous grid. Thus, four different training data sets, i.e., inventory grid densities, were created. Examples of each of these

grids for a small part of the test site are shown in Fig. 2. Henceforth, these grids will be referred to as the following:

- grid 1: 200 m × 200 m, original sample grid comprising all inventory plots within the test site (1875 plots);
- grid 2: 282.8 m × 282.8 m, remaining inventory plots after the first reduction step, i.e., every second inventory plot from grid 1 was removed (50% of the original sample, 937 plots);
- grid 3: 400 m × 400 m, remaining inventory plots after the second reduction step, i.e., every second inventory plot from grid 2 was removed (25% of the original sample, 461 plots);
- grid 4: 565.7 m × 565.7 m, remaining inventory plots after the final reduction step, i.e., every second inventory plot from grid 3 was removed (12% of the original sample, 226 plots).

3.3.4. Comparison of different modeling approaches and inventory grid densities

In total, the following seven different models were constructed using the two modeling approaches (lm and rf) described above:

- lm_best_H: only the H metric with the best predictive performance was used;
- lm_best_H_V_CC: the selected H metric and the “best” combination of one V metric and one CC metric were used;
- lm_best_H_V_CC_SP: the “best” metric from each of the four different groups was used;
- rf_best_H_V_CC_SP: only the “best” predictors (those used in the lm_best_H_V_CC_SP model) were included in the rf modeling process;
- rf_all_H: all H metrics were used in the rf modeling process;
- rf_all_H_V_CC: all H metrics together with all V and CC metrics were used;
- rf_all_H_V_CC_SP: all explanatory variables (H, V, CC, and SP) were used in the rf model-building process.

Finally, each model was trained using the field data from each of the four inventory grids described above, resulting in 28 different models. For comparison purposes, the RMSE and bias values from the resampling procedure (both the results for the 50 different held-out sets and their arithmetic means) were considered.

3.4. Wall-to-wall mapping and stand-level evaluation

As recommended by, e.g., Næsset (2002), Breidenbach et al. (2010), Woods et al. (2011), and White et al. (2013a), the cell size used for predictive modeling in a wall-to-wall manner should coincide with the ground inventory plot size. To meet this requirement, the entire test site was subdivided into contiguous 500 m² hexagons (see Fig. 6a), which were used as the basic prediction units. The same metrics calculated for the inventory plots were generated for each prediction unit.

To predict the GS of the entire forest area, each of the different regression models was applied to each 500 m² prediction unit in the test site. Subsequently, the hexagons were intersected with the stand geometries. Stand-level GS was determined by taking the arithmetic mean of the predicted GSs from all of the prediction units located completely within the corresponding stand boundary.

For the accuracy assessment at the forest stand level, all stands containing at least five inventory plots (94 in total) were taken into consideration. These stands had a mean size of 32.0 ha (SD: 11.7 ha), and the observed GS, derived as the mean from the terrestrial inventory plots, ranged from 129.6 to 499.0 m³·ha⁻¹, with an arithmetic mean of 316.4 m³·ha⁻¹ (SD: 70.0 m³·ha⁻¹). As in the plot-level evaluation, the RMSE and bias were calculated at the stand level for each of the different combinations of model and inventory grids, and the results were compared.

Fig. 2. Systematic reduction in inventory plot density used in this study to test the effect of different sampling densities. Grid density depicted here for a small section of the test site. (a) Full set of actual inventory plots, located based on a 200 m × 200 m grid, each plot represents 4 ha (grid 1). (b) First reduction step, each inventory plot represents 8 ha (grid 2). (c) Second reduction step, each inventory plot represents 16 ha (grid 3). (d) Final reduction step, plots are arranged on a 565.7 m × 565.7 m grid, each represents 32 ha (grid 4). The background image (CIR-orthoimage) illustrates the heterogeneity of the forest conditions within the Steigerwald test site. Note that this figure is provided in color online.

4. Results

4.1. Comparison of the different inventory grids

Figure 3 depicts the GS distribution of each of the four training data sets (grids 1–4). The distributions of GSs calculated from the four different sets of inventory plots had mean (SD) of 297.9 (137.5), 296.6 (141.1), 294.9 (146.8), and 280.3 (131.9) m³·ha⁻¹ for grids 1, 2, 3, and 4, respectively. Several outliers with GS values above 650 m³·ha⁻¹ can be observed in Fig. 3. Two inventory plots even had GSs of more than 1000 m³·ha⁻¹. The outliers themselves are distributed similarly across the first three inventory plot layouts. In the least dense inventory grid (12% of the original number of plots), only two outliers remain, thus accounting, at least in part, for the lower mean GS. Nonetheless, the boxplots clearly show that, at least in our test site, even a sample of inventory plots in which each plot represents an area of 32 ha (grid 4) can provide a reasonable characterization of the GS. To prove this statistically, a Kruskal-Wallis rank sum test was applied to compare the distributions from grids 1, 2, 3, and 4, as the normal distribution prerequisite for ordinary ANOVA was not fulfilled for grids 1, 2, and 3 (test results are given in the supplementary material¹). The results of this test show that the four different grids do not differ significantly.

4.2. Model building and plot-level evaluation

Table 4 shows the absolute and relative RMSE and bias as obtained from the 10-fold cross-validation repeated five times at the plot level for each of the different modeling approaches combined with each of the different inventory grids. When all models and all grids were considered, the estimations were unbiased, and the relative cross-validated RMSE at the plot level ranged from 30.92% to 36.02% (Table 4).

In the lm modeling procedure, the explanatory variable E.MEAN (mean height calculated from the image-based point cloud) resulted in the lowest RMSE of all H metrics tested when using grid 1 as training data. Therefore, E.MEAN was selected for inclusion in the final lm_best_H model (RMSE: 32.50%). When evaluating this selected H metric together with the structural metrics, E.LSKEW (L-moment skewness based on the point cloud) as the “best” V metric and E.CC_MEAN (ratio of all points above the mean height) as the “best” CC metric were added to the model lm_best_H_V_CC (RMSE: 31.86%). In the final selection, MeanNDVI (mean value of the NDVI) as the “best” SP metric was included to the model lm_best_H_V_CC_SP (RMSE: 31.27%). The selected predictors from the four groups of explanatory variables proved statistically significant in the bootStepAIC algorithm; therefore, all of them remained in the final model.

For the rf models, the performance of the rf_best_H_V_CC_SP model (relative RMSE: 32.55%) was inferior to that of the corresponding lm_best_H_V_CC_SP model (relative RMSE: 31.27%). The models rf_all_H and rf_all_H_V_CC yielded relative RMSEs of 32.42% and 32.31%, thus not outperforming the linear models with feature selection. However, the rf_all_H_V_CC_SP model showed the best performance when applied to grid 1, where it achieved a relative RMSE of 30.92%. The explanatory variables MeanNDVI, E.MEAN, and cc_15m (ratio of all pixels above 15 m calculated from

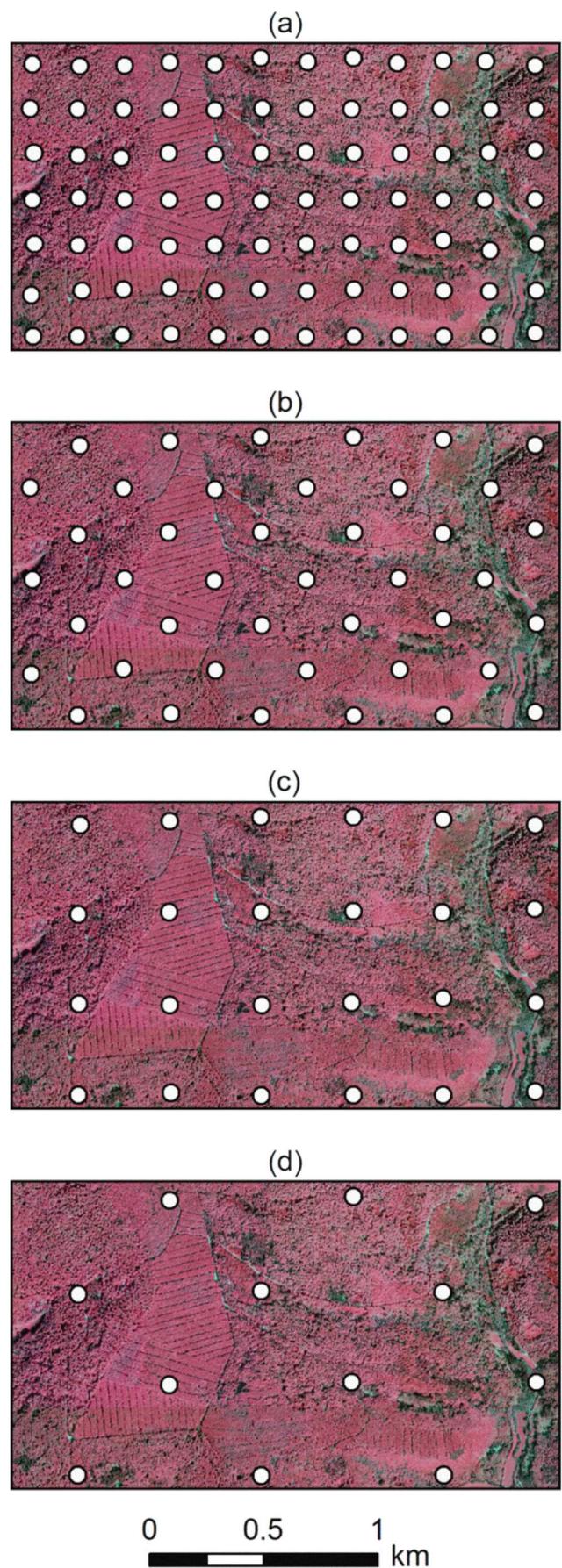


Fig. 3. Boxplots displaying the sample distribution of the growing stock (GS; $\text{m}^3 \cdot \text{ha}^{-1}$) calculated using data from the four different inventory grids. The width of the boxes indicates the number of inventory plots within each sample. Two inventory plots had GS values of more than $1000 \text{ m}^3 \cdot \text{ha}^{-1}$; one of these sites is located within a long-term forest growth test site dominated by beech, whereas the other one is part of an old-growth forest stand where very little harvesting takes place.

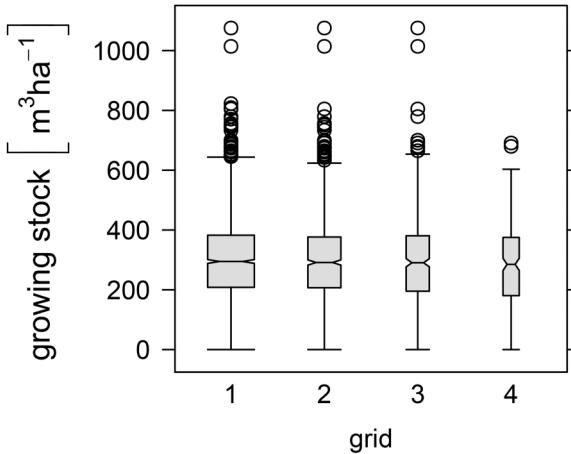


Table 4. Absolute and relative RMSE and bias as obtained from the 10-fold cross-validation (repeated five times) for each combination of the various modeling approaches and inventory grid densities investigated in this study.

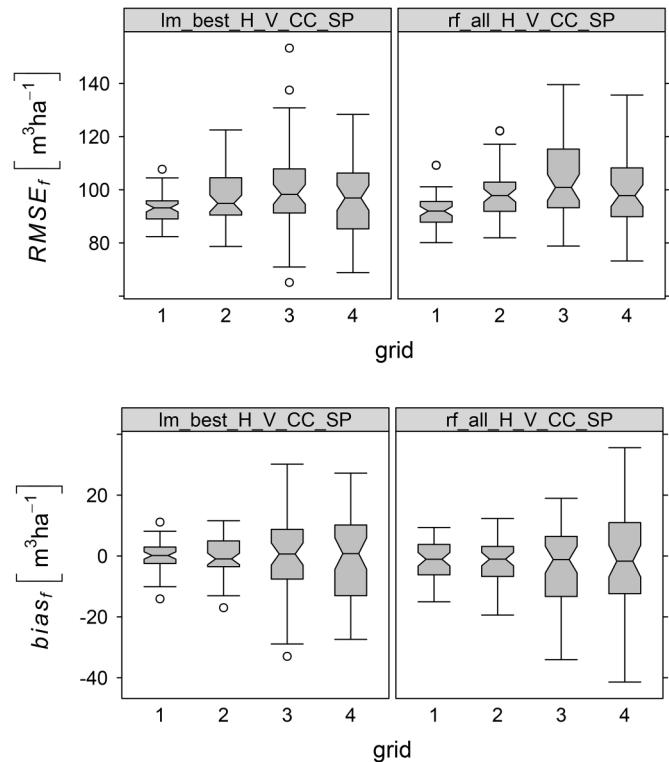
Model	Grid	RMSE ($\text{m}^3 \cdot \text{ha}^{-1}$)	RMSE (%)	Bias ($\text{m}^3 \cdot \text{ha}^{-1}$)	Bias (%)
lm_best_H	1	96.83	32.50	0.01	0.00
	2	100.50	33.88	0.02	0.01
	3	105.00	35.61	0.04	0.01
	4	99.22	35.37	0.13	0.05
lm_best_H_V_CC	1	94.90	31.86	-0.01	0.00
	2	98.47	33.20	-0.06	-0.02
	3	103.20	34.99	0.06	0.02
	4	98.03	34.95	-0.06	-0.02
lm_best_H_V_CC_SP	1	93.15	31.27	0.00	0.00
	2	97.10	32.74	-0.09	-0.03
	3	101.80	34.52	0.08	0.03
	4	96.78	34.50	-0.01	0.00
rf_best_H_V_CC_SP	1	96.96	32.55	-0.32	-0.11
	2	100.30	33.82	-0.53	-0.18
	3	105.70	35.84	-0.04	-0.01
	4	98.12	34.98	0.33	0.12
rf_all_H	1	96.57	32.42	-1.79	-0.60
	2	101.80	34.32	-2.64	-0.89
	3	106.20	36.02	-1.82	-0.62
	4	100.20	35.75	-0.83	-0.30
rf_all_H_V_CC	1	96.26	32.31	-1.96	-0.66
	2	101.40	34.19	-2.82	-0.95
	3	105.40	35.75	-2.41	-0.82
	4	99.87	35.63	-1.32	-0.47
rf_all_H_V_CC_SP	1	92.10	30.92	-1.39	-0.46
	2	98.11	33.08	-1.72	-0.58
	3	104.90	35.58	-2.62	-0.89
	4	99.12	35.36	-0.93	-0.33

Note: All data presented were calculated at the inventory plot level.

the CHM with 1 m spatial resolution) yielded the highest importance scores in that particular rf model.

For all models, the RMSE values, in general, increased slightly as the training sample grid density decreased (i.e., from grid 1 to grid 3). However, for grid 4, the absolute RMSE also decreased

Fig. 4. Distribution of the absolute RMSE_f and bias_f values, as obtained from the 50 held-out samples of the 10-fold cross-validation (repeated five times). Boxplots are given for the modeling approaches lm_best_H_V_CC_SP and rf_all_H_V_CC_SP for each of the four inventory grids tested. In Supplementary Fig. S3¹, the boxplots for all different modeling approaches are depicted.



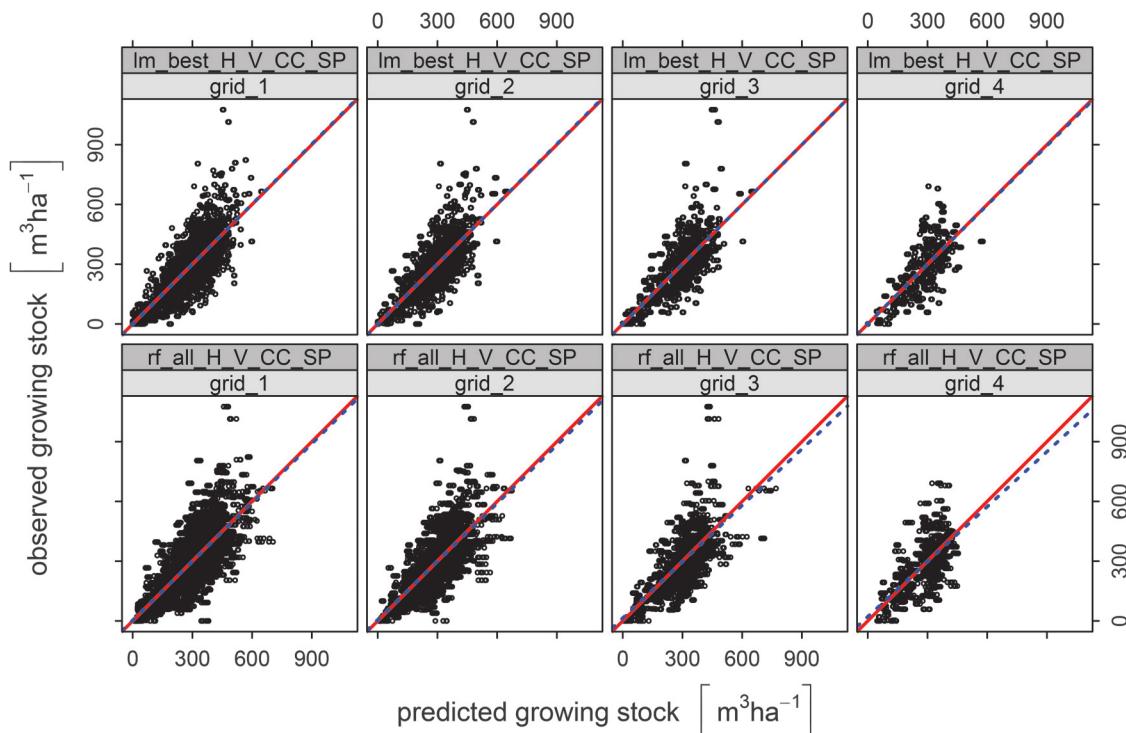
somewhat, which could be due to the fact that the training data in grid 4 were less influenced by outliers, as explained in section 4.1 and illustrated in Fig. 3. Due to the lower mean GS ($280.3 \text{ m}^3 \cdot \text{ha}^{-1}$) obtained using the inventory plots in grid 4, this effect is not reflected in the relative RMSE.

Figure 4 displays the results from the resampling process, i.e., the calculated RMSE_f and bias_f values from all 50 sets of held-out samples, for the modeling approaches lm_best_H_V_CC_SP and rf_all_H_V_CC_SP (the plots for all different modeling approaches are depicted in Supplementary Fig. S3¹). The boxplots clearly show that the distribution of the 50 RMSE_f and bias_f values widened when moving from grid 1 to grid 4. Nonetheless, the medians are similar. Scatterplots of the predicted vs. the observed GS at the plot level are shown in Fig. 5 for the models lm_best_H_V_CC_SP and rf_all_H_V_CC_SP (see Supplementary Fig. S4¹ for all models). The scatterplots illustrate that, in all cases, the deviations from the 1:1 line increased with higher GS values. Furthermore, concerning the inventory plots with high GS values (previously described as outliers with more than $650 \text{ m}^3 \cdot \text{ha}^{-1}$; see Fig. 3), a systematic underestimation can be observed irrespective of the model applied. Nevertheless, the 1:1 line and the fitted linear trend line matched well for all models. Thus, a reasonable relationship between the predicted and observed values was obtained with all models for the entire range of presented GSs.

4.3. Wall-to-wall mapping and stand-level evaluation

Following model building at the plot level, the GS was predicted for each of the hexagonal prediction units using each combination of the different models and inventory grids. The results of this wall-to-wall mapping process is exemplarily shown in Fig. 6 for the rf_all_H_V_CC_SP model applied to grid 1 for a subset of

Fig. 5. Predicted vs. observed growing stock (GS; $\text{m}^3 \cdot \text{ha}^{-1}$) at the plot level, as obtained from the 50 held-outs of the 10-fold cross-validation (repeated five times). Linear trend lines are given as dashed (blue online) lines, and the 1:1 lines are given as solid (red online). Scatterplots are presented for the modeling approaches lm_best_H_V_CC_SP and rf_all_H_V_CC_SP for each of the four inventory grids tested. In Supplementary Fig. S4¹, the scatterplots for all different combinations of modeling approach and inventory grid are depicted. Note that this figure is provided in color online.



the study area (the same map extent as depicted in Fig. 2). Superimposed over this result in black lines are the manually delineated forest stand boundaries used for forest management.

The color-coded map (Fig. 6a) illustrates the magnitude of predicted GSs for the prediction units and shows the variability even within the different forest stands. Yet, in general, a causal connection between the predicted GSs for the prediction units and the manually delineated forest stands is obvious. Figure 6b depicts the aggregated GS estimates for the different stands obtained by averaging the predictions from the basic subunits within each stand region.

The accuracy assessment at the stand level resulted in absolute RMSE values ranging from $43.88 \text{ m}^3 \cdot \text{ha}^{-1}$ to $49.76 \text{ m}^3 \cdot \text{ha}^{-1}$ and relative RMSEs ranging from 13.87% to 15.73% (Table 5). Comparing the different models, no major deviations were detected. As further shown in Table 5, in some cases, the aggregated estimations at the stand level show a minor prediction bias ranging from -3.74% to 2.71%. For all modeling approaches, the bias was negative for grids 1, 2, and 3. However, using the models based on the grid 4 data only, positive biased predictions at the stand level were observed. This indicates an overestimation when the first three grids were used and an underestimation using grid 4 at the stand level.

The scatterplots in Fig. 7 show the predicted vs. the observed GS values at the stand level for the modeling approaches lm_best_H_V_CC_SP and rf_all_H_V_CC_SP (see Supplementary Fig. S5¹ for all models). In comparison with the plot-level estimation (Fig. 5; Supplementary Fig. S4¹), the amount of scatter reduced considerably for all model-grid combinations. In general, the scatter was similar for all models and all grids. Only for the rf models applied to grid 4, some saturation effects were visible in stands with more than $400 \text{ m}^3 \cdot \text{ha}^{-1}$ of observed GS. To illustrate possible deviations of the predicted vs. the observed GSs at the

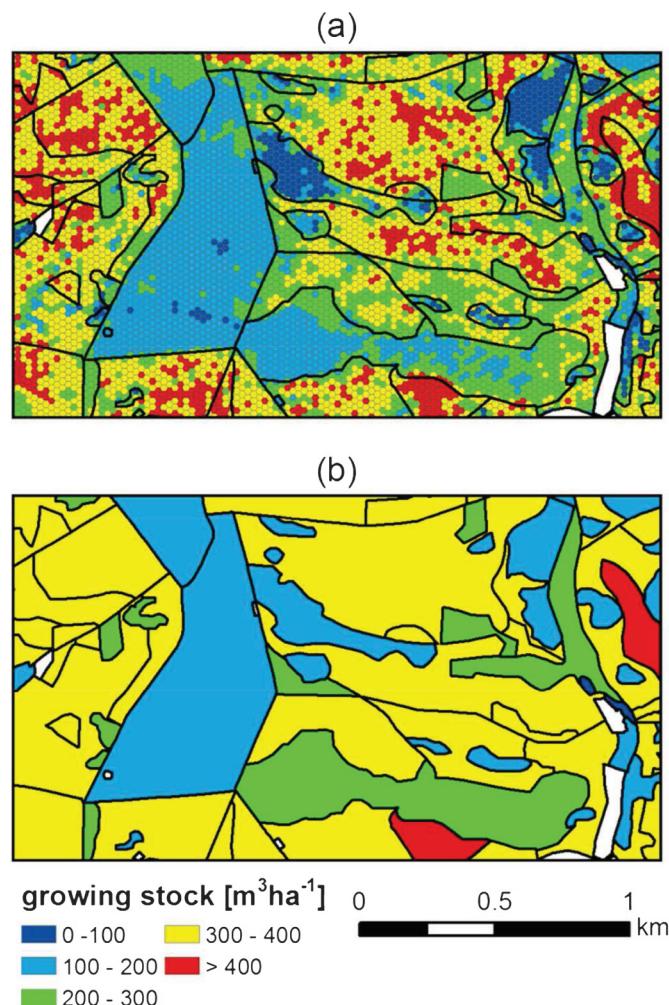
stand level from the 1:1 line, we added linear trend lines to the scatterplots. Overall, the two lines matched well for all models, indicating no systematic shift in predicted GSs.

5. Discussion

5.1. Image-based height data to describe the canopy surface

Very dense point clouds with more than $100 \text{ measurements} \cdot \text{m}^{-2}$ were obtained from the stereo images using the SGM approach, thus providing a detailed characterization of the forest canopy surface (see Supplementary Fig. S2¹). Visual examination of the 2.5D data revealed that the pixel-based SGM even produced such dense measurements in regions with low image texture, e.g., shaded areas at forest stand boundaries. We regard this as an advantage over conventional feature- and area-based matching algorithms, e.g., normalized cross-correlation (NCC). However, we agree with White et al. (2013b) that further research is still necessary, including investigations in a variety of forest types with a diversity of stand conditions, to give a reliable recommendation concerning the performance of different matching algorithms. Up to now, this has largely been evaluated only for either urban or agricultural landscapes, for example, in the benchmark study initiated by the EuroSDR (Haala 2013). That study used aerial imagery acquired especially for research purposes, i.e., with high across-track overlaps of more than 60%. Studies using such specialized and highly targeted datasets are not necessarily comparable to studies such as ours, which make use of "off-the-shelf" imagery available from the national surveying administrations in many European countries. To fully and effectively investigate the feasibility of these datasets for further operational usage, we believe that a comprehensive assessment of matching algorithms in various forest areas with stereo imagery acquired by official and regular aerial surveys is necessary.

Fig. 6. Illustration of the wall-to-wall mapping result (for the same geographic extent presented in Fig. 2). (a) Predicted growing stock (GS; $\text{m}^3 \cdot \text{ha}^{-1}$) estimates for the 500 m 2 hexagons based on the model rf_all_H_V_CC_SP, with all inventory plots (grid 1) in the test site. The boundaries of the manually delineated forest stands (black lines) are superimposed on the illustration. (b) Averaged GS estimates for the stands based on the values from all hexagons that are located completely within the stand polygons. Note that this figure is provided in color online.



5.2. Growing stock estimation at the plot and stand levels

In accordance with other recently published articles (Table 1), our study confirms the great potential of stereo images and dense image matching for estimating GS, at both the plot and stand levels.

The feature selection procedure in the lm approach revealed that the mean canopy height, computed from the image-based point cloud, was the most important predictor of GS. This is in line with the results obtained in various ALS studies (e.g., Straub et al. (2010), Woods et al. (2011), and Hudak et al. (2012)), as well as with the findings from recent studies dealing with image-based point clouds (as listed in Table 1), where H metrics were also found to be superior to other predictors. In the rf approach using all predictor variables, the mean height yielded the second highest importance score (after the mean NDVI, which was ranked as the most important variable), which further underlines the importance of this variable in estimation of GS. Concerning the derivation of the H metrics from the image-based point cloud, we applied the 2 m height cutoff, in accordance with many previous studies, e.g., Nilsson (1996), Næsset

Table 5. Absolute and relative RMSE and bias for each combination of the various modeling approaches and inventory grid densities investigated in this study.

Model	Grid	RMSE ($\text{m}^3 \cdot \text{ha}^{-1}$)	RMSE (%)	Bias ($\text{m}^3 \cdot \text{ha}^{-1}$)	Bias (%)
lm_best_H	1	49.76	15.73	-8.00	-2.53
	2	49.33	15.59	-4.85	-1.53
	3	49.75	15.72	-8.62	-2.72
	4	48.64	15.37	1.14	0.36
lm_best_H_V_CC	1	48.45	15.31	-7.63	-2.41
	2	47.88	15.13	-3.83	-1.21
	3	48.15	15.22	-7.28	-2.30
	4	47.39	14.98	1.70	0.54
lm_best_H_V_CC_SP	1	47.45	14.99	-5.75	-1.82
	2	46.96	14.84	-2.10	-0.66
	3	47.28	14.94	-6.34	-2.00
	4	46.70	14.76	2.85	0.90
rf_best_H_V_CC_SP	1	45.75	14.46	-5.72	-1.81
	2	45.70	14.44	-1.50	-0.47
	3	44.99	14.22	-4.15	-1.31
	4	47.43	14.99	8.57	2.71
rf_all_H	1	45.94	14.52	-10.19	-3.22
	2	45.26	14.30	-6.01	-1.90
	3	44.74	14.14	-8.12	-2.56
	4	46.90	14.82	6.62	2.09
rf_all_H_V_CC	1	46.33	14.64	-10.81	-3.41
	2	45.61	14.41	-6.63	-2.09
	3	45.62	14.41	-9.67	-3.06
	4	46.68	14.75	5.66	1.79
rf_all_H_V_CC_SP	1	44.00	13.90	-10.30	-3.26
	2	43.88	13.87	-8.15	-2.58
	3	44.20	13.97	-11.83	-3.74
	4	45.56	14.40	3.97	1.25

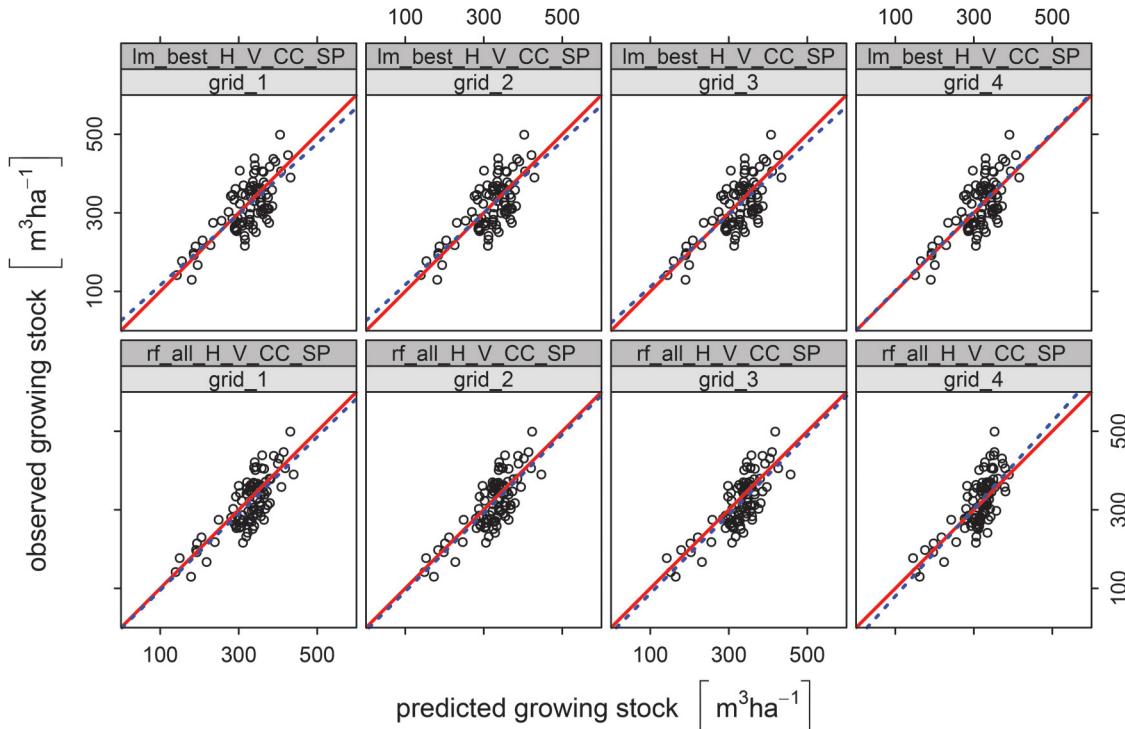
Note: Data presented here were calculated at the forest stand level for 94 validation stands with an average stand size of 32 ha, each containing a minimum of five terrestrial inventory plots.

(2002), Nurminen et al. (2013), and Vastaranta et al. (2013). However, further research should reveal the most appropriate value for this height threshold in more detail, especially concerning image-based point clouds and different forest types.

We evaluated the additional gain in modeling performance when including structural variables, i.e., variability and canopy cover metrics, as explanatory variables to the models. In both the lm and rf approaches, the RMSE could be reduced only marginally by adding these variables. This suggests that for image-based data, at least in our study, these metrics are not very contributive for estimating GS, although the structural variables were tested statistically significant in the bootStepAIC algorithm within the lm modeling procedure.

With regard to SP metrics, the NDVI appeared to be the most important contributing predictor. It was chosen for inclusion in the lm_best_H_V_CC_SP model and was also ranked as the most important variable in the rf_all_H_V_CC_SP model. Moreover, in both models, the NDVI was more important than any of the various spectral texture metrics calculated based on the GLCM. By adding metrics describing the spectral information from the imagery to the models, some measurable improvement in terms of RMSE could be obtained, for both lm and rf. Therefore, based on our investigation and considering the results from Straub et al. (2013a), we conclude that the inclusion of spectral information can slightly improve model performance when estimating GS. However, it has to be kept in mind that differences in the spectral information, e.g., the NDVI, can occur if datasets are acquired during different phenological stages. In the case that only imagery from different seasons is available for a specific test site, tests are required to determine whether the resulting spectral data can be combined or if it is advantageous to treat them separately.

Fig. 7. Predicted vs. observed growing stock (GS; $\text{m}^3 \cdot \text{ha}^{-1}$) for 94 validation stands. All stands that fully contained a minimum of five terrestrial inventory plots were used for the comparison. The predicted GS was derived by averaging the predictions from the hexagonal prediction units within each stand unit. The observed GS was calculated as the arithmetic mean of the terrestrial inventory plot measurements within each stand. Linear trend lines are given as dashed (blue online) lines, and the 1:1 lines are given as solid (red online). Scatterplots are given for the modeling approaches lm_best_H_V_CC_SP and rf_all_H_V_CC_SP for each of the four inventory grids tested. In Supplementary Fig. S5¹, the scatterplots for all different combinations of modeling approach and inventory grid are depicted. Note that this figure is provided in color online.



Concerning the two different modeling approaches investigated here — the lm and the rf models — our results do not clearly favor one over the other. This is in line with the findings of Penner et al. (2013), who compared parametric and nonparametric methods for forest-attribute estimation and stated that none of the methods consistently yielded better results than the others. As an example, the relative RMSEs obtained using grid 1 with both models were similar to one another (lm_best_H_V_CC_SP: 31.27%; rf_all_H_V_CC_SP: 30.92%) and substantially smaller than the standard deviation for the GS (46.16%) of the observed values. In addition, the distribution of the RMSE_f values obtained was similar for both approaches, indicating a comparable level of estimation uncertainty. In the recently published studies listed in Table 1, both parametric and nonparametric regression approaches were implemented. With respect to our findings, we assume that the differences in the relative RMSEs presented in the last row of Table 1 were less influenced by the regression type applied than by the forest structure and the distribution of the observed GS among the different test sites (compare the mean and SD values in Table 1). Väistönen et al. (2013) and Nurminen et al. (2013) reported RMSEs at the plot level of 24.5% and 22.62%, respectively, for single-layered, even-aged, and conifer-dominated forests in Finland, with similar means and SDs of the observed GS. Straub et al. (2013a) achieved a relative RMSE of 37.89% at the plot level for a multi-aged, multi-layered, and mixed-species forest in southern Germany. Thus, although our results are more precise with regard to the RMSE than those obtained by Straub et al. (2013a), they are not as precise as those found in studies conducted in forests under Nordic conditions. The test site used in our study is more homogeneous with respect to the species mixture and GS distribution than the forest investigated in Straub et al. (2013a). This might

explain the lower RMSEs obtained in our study, the smallest of which was 30.92% (obtained using the rf_all_H_V_CC_SP model applied to grid 1).

Considering stand-level estimations, the RMSEs were reduced by 50% compared with the plot-level estimations. To the best of our knowledge, only Bohlin et al. (2012) have previously assessed stand-level predictions of GS based on airborne stereo imagery. In their study, conducted for a boreal test site with a mean stand size of 2.8 ha, they achieved a stand-level RMSE of 13.1%, which is in a similar range to the lowest RMSE of 13.87% obtained in our study using the rf_all_H_V_CC_SP model. Moreover, Holmgren and Jonsson (2004) and Straub et al. (2013b) evaluated models for GS estimation at the stand level based on ALS data. In their stand-level analysis in a conifer-dominated test site in Sweden (median stand size: 1.3 ha), Holmgren and Jonsson (2004) achieved a RMSE of 14.1%, and for a mixed, highly structured forest in Germany (mean stand size: 10 ha), Straub et al. (2013b) obtained a RMSE of 17.14% in their stand-level analysis. However, as these studies were conducted in different forest types and the stand sizes vary noticeably from those of the Steigerwald test site, direct comparisons of the resulting estimation errors are not possible. Nonetheless, the reported RMSEs all ranged in a similar magnitude. These numbers indicate that image-based estimations are competitive with ALS-based estimations when used for stand-level prediction.

5.3. Number of inventory plots used as training data

With regard to the number of inventory plots used in the ABA, Holopainen et al. (2013) stated that training data should represent the entire population and cover the entire spectrum of variation in it as much as possible. Considering the information presented in the boxplots in Fig. 3 and the results from the Kruskal-Wallis

test, we conclude that even the minimum number of plots in grid 4 fulfilled the requirements for a sound sampling of the GS in the test site. This is in line with the model precision achieved at the plot level, where the RMSEs obtained using each of the various grids were in a similar range. However, as the boxplots shown in Fig. 4 reveal, the uncertainty of the estimation increased when the number of inventory plots used for training was reduced. This finding was observed for both the lm and the rf models. Moreover, our results from the wall-to-wall mapping and stand-level prediction showed similar precision and accuracy for all grids. However, the scatterplots in Fig. 7 indicate a saturation effect for stands with more than $400 \text{ m}^3 \cdot \text{ha}^{-1}$ of observed GS, but this effect was found only for the rf models applied to grid 4. Accordingly, with regard to the rf models, the slight increase in the RMSE achieved using grid 4 might be related to this effect. A possible explanation for this phenomenon could be found in the disadvantages inherent to all imputation methods, i.e., no extrapolation is possible beyond the range of the training data; therefore, the predicted values tend towards the center of the distribution (Penner et al. 2013).

The inventory grid 1 ($200 \text{ m} \times 200 \text{ m}$) provided more than sufficient information for modeling the GS in our test site. When aiming at cost savings, thinning the inventory grid and, therefore, reducing the amount of fieldwork necessary appears to be possible. Nevertheless, the results from the rf models based on grid 4 and used for stand-level prediction revealed that a sample of training data that is too small might possibly lead to prediction errors.

5.4. Operational implementation for forest management

The successful implementation of ALS data into operational forest inventories and forest management activities has already been described in several publications (e.g., Næsset 2007 and Woods et al. 2011). However, it is commonly agreed that ALS acquisitions are more expensive than aerial-image surveys (White et al. 2013b). Therefore, exploring aerial imagery as an alternative to ALS is a current and relevant research topic, particularly with respect to the need for up-to-date and detailed spatially explicit information for practicing sustainable forest management. This demand is already high at present but will further increase with advancing global change (Pretzsch et al. 2014). In our study, we were able to demonstrate the feasibility of using standard aerial images from regularly updated surveys in combination with an existing ALS-based DTM to estimate the GS for a large forested area. The resulting map depicts the GS for the entire forest based on 500 m^2 prediction units and provides valuable insights into the distribution of GS across the forest. Furthermore, aggregation to the stand level is feasible with a reasonable level of estimation accuracy. As the estimation was successfully applied to a large test site (77.5 km^2), the presented method using regularly updated aerial images might be a valuable contribution to meeting the growing needs in forest inventory and planning.

6. Conclusions and outlook

This study underlined the high potential of existing aerial stereo imagery for estimating growing stock. We successfully applied SGM to compute a dense point cloud characterizing the canopy surface of a 77.5 km^2 broadleaf-dominated forest in central Europe. Our assessment of ordinary least squares linear regression and random forests models revealed that both approaches are suitable for modeling, with slightly better precision achieved using the rf model. In addition to explanatory variables obtained from the point cloud and the CHM, spectral metrics derived from orthoimagery contributed to the GS estimation. At both the plot and the stand levels, very satisfying precisions and accuracies were obtained. Moreover, the results of our study demonstrated that a systematic reduction in the number of plots in the field inventory grid used for training purposes did not substantially affect the model performance. Our findings underline the potential of aerial stereo imagery

in combination with an existing ALS-based DTM as appropriate data to support forest inventory and management.

In future research, we intend to assess the effectiveness of modeling using regularly updated aerial image data to detect changes in height and volume increment. Moreover, we will attempt to estimate other forest attributes (e.g., H_{dom} , BA, stand-density index, etc.) using aerial imagery. We will further test and improve the procedures presented here under various forest conditions in central Europe, e.g., in areas with different topographic conditions and species compositions. Finally, the transferability of the model approaches presented here to adjacent forested areas with similar conditions will be investigated, as this can support forest management in areas where terrestrial forest inventory data are lacking.

Acknowledgements

The authors would like to acknowledge the funding of this project (E49 SAPEX-DLB) by the Bavarian State Ministry of Food, Agriculture, and Forestry (in German: Bayerisches Staatsministerium für Ernährung, Landwirtschaft und Forsten). We thank the Bavarian Office for Surveying and Geoinformation for providing the remote-sensing data, as well as the Bavarian State Forest Enterprise for providing the field inventory data. We also thank Karl-Heinz Gutjahr from Joanneum Research, Graz, for his software support, Laura Carlson for her language editing, Rudolf Seitz for his helpful comments on earlier drafts of this paper, and three anonymous reviewers for their suggestions, which helped to improve the manuscript.

References

- Austin, P.C., and Tu, J.V. 2004. Bootstrap methods for developing predictive models. *Am. Stat.* **58**(2): 131–137. doi:[10.1198/0003130043277](https://doi.org/10.1198/0003130043277).
- Bohlin, J., Wallerman, J., and Fransson, J.E.S. 2012. Forest variable estimation using photogrammetric matching of digital aerial images in combination with a high-resolution DEM. *Scand. J. For. Res.* **27**(7): 692–699. doi:[10.1080/02827581.2012.686625](https://doi.org/10.1080/02827581.2012.686625).
- Breidenbach, J., Nothdurft, A., and Kändler, G. 2010. Comparison of nearest neighbour approaches for small area estimation of tree species-specific forest inventory attributes in central Europe using airborne laser scanner data. *Eur. J. For. Res.* **129**(5): 833–846. doi:[10.1007/s10342-010-0384-1](https://doi.org/10.1007/s10342-010-0384-1).
- Breiman, L. 2001. Random forests. *Mach. Learn.* **45**(1): 5–32. doi:[10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- Everitt, B.S., Landau, S., Leese, M., and Stahl, D. 2011. *Cluster analysis*. John Wiley & Sons, Ltd., Chichester, UK.
- Genuer, R., Poggi, J.-M., and Tuleau, C. 2008. Random forests: some methodological insights. *Institut national de recherche en informatique et en automatique*, Technical Report 6729.
- Gobakken, T., and Næsset, E. 2008. Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. *Can. J. For. Res.* **38**(5): 1095–1109. doi:[10.1139/X07-219](https://doi.org/10.1139/X07-219).
- Haala, N. 2011. Multiray photogrammetry and dense image matching. *In Proceedings of the 53rd Photogrammetric Week*. Edited by D. Fritsch. Institute for Photogrammetry, University of Stuttgart, Germany. pp. 185–195.
- Haala, N. 2013. The landscape of dense image matching algorithms. *In Proceedings of the 54th Photogrammetric Week*. Edited by D. Fritsch. Institute for Photogrammetry, University of Stuttgart, Germany. pp. 271–284.
- Haala, N., and Rothermel, M. 2012. Dense multi-stereo matching for high quality digital elevation models. *Photogramm. Fernerkun.* **2012**(4): 331–343. doi:[10.1127/1432-8364/2012/0121](https://doi.org/10.1127/1432-8364/2012/0121).
- Haralick, R.M., Shanmugam, K., and Dinstein, I. 1973. Textural features for image classification. *IEEE Trans. Syst., Man, Cybern.* **3**(6): 610–621. doi:[10.1109/TSMC.1973.4309314](https://doi.org/10.1109/TSMC.1973.4309314).
- Hastie, T., Tibshirani, R., and Friedman, J.H. 2009. *The elements of statistical learning: data mining, inference, and prediction*. 2nd edition. Springer, New York.
- Hawbaker, T.J., Keuler, N.S., Lesak, A.A., Gobakken, T., Contrucci, K., and Radeloff, V.C. 2009. Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. *J. Geophys. Res.* **114**. doi:[10.1029/2008JG000870](https://doi.org/10.1029/2008JG000870).
- Hirschmüller, H. 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(2): 328–341. doi:[10.1109/TPAMI.2007.1166](https://doi.org/10.1109/TPAMI.2007.1166).
- Hirschmüller, H. 2011. Semi-global matching — motivation, developments, and applications. *In Proceedings of the 53rd Photogrammetric Week*. Edited by

- D. Fritsch. Institute for Photogrammetry, University of Stuttgart, Germany. pp. 5–11.
- Holmgren, J., and Jonsson, T. 2004. Large scale airborne laser scanning of forest resources in Sweden. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XXXVI**(Part 8/W2): 157–160.
- Holopainen, M., Vastaranta, M., Liang, X., Hyypä, J., Jaakkola, A., and Kankare, V. 2013. Estimation of forest stock and yield using LiDAR data. In *Remote Sensing of Natural Resources*. CRC Press. pp. 259–290.
- Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E., and Falkowski, M.J. 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sens. Environ.* **112**(5): 2232–2245. doi:[10.1016/j.rse.2007.10.009](https://doi.org/10.1016/j.rse.2007.10.009).
- Hudak, A.T., Strand, E.K., Vierling, L.A., Byrne, J.C., Eitel, J.U., Martinuzzi, S., and Falkowski, M.J. 2012. Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. *Remote Sens. Environ.* **123**: 25–40. doi:[10.1016/j.rse.2012.02.023](https://doi.org/10.1016/j.rse.2012.02.023).
- Hyypä, J., Hyppä, H., Leckie, D., Gougeon, F., Yu, X., and Maltamo, M. 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *Int. J. Remote Sens.* **29**(5): 1339–1366. doi:[10.1080/01431160701736489](https://doi.org/10.1080/01431160701736489).
- Järnstedt, J., Pekkarinen, A., Tuominen, S., Ginzler, C., Holopainen, M., and Viitala, R. 2012. Forest variable estimation using a high-resolution digital surface model. *ISPRS J. Photogramm. Remote Sens.* **74**: 78–84. doi:[10.1016/j.isprsjprs.2012.08.006](https://doi.org/10.1016/j.isprsjprs.2012.08.006).
- Joanneum Research. 2014. Remote sensing software package graz [online]. Available from dig.joanneum.at/rsg/ [accessed 14 February 2014].
- Kuhn, M. 2013. caret: classification and regression training [online]. Available from caret.r-forge.r-project.org/ [accessed 18 June 2014].
- Kuhn, M., and Johnson, K. 2013. Applied predictive modeling. Springer, New York.
- Latifi, H., Nothdurft, A., and Koch, B. 2010. Nonparametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. *Forestry*, **83**(4): 395–407. doi:[10.1093/forestry/cpq022](https://doi.org/10.1093/forestry/cpq022).
- Leberl, F., Irschara, A., Pock, T., Meixner, P., Gruber, M., Scholz, S., and Wiechert, A. 2010. Point clouds: lidar versus 3D vision. *Photogramm. Eng. Remote Sens.* **76**(10): 1123–1134. doi:[10.14358/PERS.76.10.1123](https://doi.org/10.14358/PERS.76.10.1123).
- Leberl, F., Gruber, M., Ponticelli, M., and Wiechert, A. 2012. The UltraCam story. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XXXIX**(B1): 39–44. doi:[10.5194/isprarchives-XXXIX-B1-39-2012](https://doi.org/10.5194/isprarchives-XXXIX-B1-39-2012).
- Liaw, A., and Wiener, M. 2002. Classification and regression by randomForest. *R News*, **2**(3): 18–22.
- Lim, K., Treitz, P., Wulder, M., St-Onge, B., and Flood, M. 2003. LiDAR remote sensing of forest structure. *Prog. Phys. Geogr.* **27**(1): 88–106. doi:[10.1191/0309133303pp360r](https://doi.org/10.1191/0309133303pp360r).
- Lindberg, E., and Hollaus, M. 2012. Comparison of methods for estimation of stem volume, stem number, and basal area from airborne laser scanning data in a hemi-boreal forest. *Remote Sens.* **4**(12): 1004–1023. doi:[10.3390/rs4041004](https://doi.org/10.3390/rs4041004).
- Maltamo, M., Eerikäinen, K., Packalen, P., and Hyypä, J. 2006. Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry*, **79**(2): 217–229. doi:[10.1093/forestry/cpl007](https://doi.org/10.1093/forestry/cpl007).
- Maltamo, M., Bollandsas, O.M., Næsset, E., Gobakken, T., and Packalen, P. 2011. Different plot selection strategies for field training data in ALS-assisted forest inventory. *Forestry*, **84**(1): 23–31. doi:[10.1093/forestry/cpq039](https://doi.org/10.1093/forestry/cpq039).
- Maltamo, M., Næsset, E., and Vauhkonen, J. 2014. Forestry applications of airborne laser scanning. Springer Netherlands, Dordrecht.
- McGaughey, R.J. 2013. FUSION/LDV: software for LiDAR data analysis and visualization. Version 3.30. USDA Forest Service, Pacific Northwest Research Station, University of Washington, Seattle, Washington.
- Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L., and Hendrix, C.J. 2000. Predicting forest stand characteristics with airborne scanning lidar. *Photogramm. Eng. Remote Sens.* **66**(11): 1367–1372.
- Mergner, U. 2013. Exkursionsführer Forstbetrieb Ebrach.
- Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* **80**(1): 88–99. doi:[10.1016/S0034-4257\(01\)00290-5](https://doi.org/10.1016/S0034-4257(01)00290-5).
- Næsset, E. 2007. Airborne laser scanning as a method in operational forest inventory: status of accuracy assessments accomplished in Scandinavia. *Scand. J. For. Res.* **22**(5): 433–442. doi:[10.1080/02827580701672147](https://doi.org/10.1080/02827580701672147).
- Næsset, E., Gobakken, T., Holmgren, J., Hyypä, H., Hyypä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å., and Söderman, U. 2004. Laser scanning of forest resources: the nordic experience. *Scand. J. For. Res.* **19**(6): 482–499. doi:[10.1080/02827580410019553](https://doi.org/10.1080/02827580410019553).
- Næsset, E., Bollandsås, O.M., and Gobakken, T. 2005. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. *Remote Sens. Environ.* **94**(4): 541–553. doi:[10.1016/j.rse.2004.11.010](https://doi.org/10.1016/j.rse.2004.11.010).
- Neufanger, M. 2011. Richtlinie für die mittel — und langfristige Forstbetriebsplanung in den Bayerischen Staatsforsten. Forsteinrichtungsrichtlinie — FER 2011.
- Nilsson, M. 1996. Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sens. Environ.* **56**(1): 1–7. doi:[10.1016/0034-4257\(95\)00224-3](https://doi.org/10.1016/0034-4257(95)00224-3).
- Nurminen, K., Karjalainen, M., Yu, X., Hyypä, J., and Honkavaara, E. 2013. Performance of dense digital surface models based on image matching in the estimation of plot-level forest variables. *ISPRS J. Photogramm. Remote Sens.* **83**: 104–115. doi:[10.1016/j.isprsjprs.2013.06.005](https://doi.org/10.1016/j.isprsjprs.2013.06.005).
- Packalen, P., and Maltamo, M. 2006. Predicting the plot volume by tree species using airborne laser scanning and aerial photographs [online]. *For. Sci.* **52**(6): 611–622.
- Penner, M., Pitt, D.G., and Woods, M.E. 2013. Parametric vs. nonparametric LiDAR models for operational forest inventory in boreal Ontario. *Can. J. Remote Sens.* **39**(05): 426–443. doi:[10.5589/m13-049](https://doi.org/10.5589/m13-049).
- Pettersson, H. 1955. Die Massenproduktion des Nadelwaldes. *Mitt. Forstl. Forsch. Anst. Schweden*, Stockholm.
- Pretzsch, H., Biber, P., Schütze, G., Uhl, E., and Rötzer, T. 2014. Forest stand growth dynamics in Central Europe have accelerated since 1870. *Nat. Commun.* **5**: 4967. doi:[10.1038/ncomms5967](https://doi.org/10.1038/ncomms5967).
- R Core Team. 2013. R: a language and environment for statistical computing [online]. Available from www.R-project.org/.
- Rizopoulos, D. 2009. bootStepAIC: bootstrap stepAIC [online]. Available from CRAN.R-project.org/package=bootStepAIC.
- Stössel, W. 2009. Die Bayernbefliegung wird digital [online]. Mitteilungen des DVW e.V. Gesellschaft für Geodäsie, Geoinformation, und Landmanagement (1): 69–84. Available from www.dvw.de/sites/default/files/landesverband/bayern/anhang/beitragskontext/2014/stoessel_0.pdf [8 December 2014].
- Straub, C., Weinacker, H., and Koch, B. 2010. A comparison of different methods for forest resource estimation using information from airborne laser scanning and CIR orthophotos. *Eur. J. For. Res.* **129**(6): 1069–1080. doi:[10.1007/s10342-010-0391-2](https://doi.org/10.1007/s10342-010-0391-2).
- Straub, C., Stepper, C., Seitz, R., and Waser, L.T. 2013a. Potential of UltraCamX stereo images for estimating timber volume and basal area at the plot level in mixed European forests. *Can. J. For. Res.* **43**(8): 731–741. doi:[10.1139/cjfr-2013-0125](https://doi.org/10.1139/cjfr-2013-0125).
- Straub, C., Tian, J., Seitz, R., and Reinartz, P. 2013b. Assessment of Cartosat-1 and WorldView-2 stereo imagery in combination with a LiDAR-DTM for timber volume estimation in a highly structured forest in Germany. *Forestry*, **86**(4): 463–473. doi:[10.1093/forestry/cpt017](https://doi.org/10.1093/forestry/cpt017).
- Strunk, J., Temesgen, H., Andersen, H.-E., Flewelling, J.P., and Madsen, L. 2012. Effects of lidar pulse density and sample size on a model-assisted approach to estimate forest inventory variables. *Can. J. Remote Sens.* **38**(5): 644–654. doi:[10.5589/m12-052](https://doi.org/10.5589/m12-052).
- van Laar, A., and Akçay, A. 2007. Forest mensuration. 2nd edition. Springer, Dordrecht.
- Vastaranta, M., Wulder, M.A., White, J.C., Pekkarinen, A., Tuominen, S., Ginzler, C., Kankare, V., Holopainen, M., Hyppä, J., and Hyppä, H. 2013. Airborne laser scanning and digital stereo imagery measures of forest structure: comparative results and implications to forest mapping and inventory update. *Can. J. Remote Sens.* **39**(05): 382–395. doi:[10.5589/m13-046](https://doi.org/10.5589/m13-046).
- Venables, W.N., and Ripley, B.D. 2007. Modern applied statistics with S. 4th edition, corr. print. Springer, New York.
- Wei, T. 2013. corrrplot: visualization of a correlation matrix. Available from cran.r-project.org/web/packages/corrrplot/index.html.
- White, J.C., Wulder, M.A., Varhola, A., Vastaranta, M., Coops, N.C., Cook, B.D., Pitt, D., and Woods, M. 2013a. A best practices guide for generating forest inventory attributes from airborne laser scanning data using the area-based approach. Natural Resources Canada, Canadian Forest Service, Canadian Wood Fibre Centre, Victoria, British Columbia, Canada, Information Report FI-X-010.
- White, J., Wulder, M., Vastaranta, M., Coops, N., Pitt, D., and Woods, M. 2013b. The utility of image-based point clouds for forest inventory: a comparison with airborne laser scanning. *Forests*, **4**(3): 518–536. doi:[10.3390/f4030518](https://doi.org/10.3390/f4030518).
- Woods, M., Pitt, D., Lim, K., Nesbitt, D., Etheridge, D., and Treitz, P. 2011. Operational implementation of a LiDAR inventory in boreal Ontario [online]. *For. Chron.* **87**(4): 512–528. doi:[10.5558/tfc2011-050](https://doi.org/10.5558/tfc2011-050).
- Wulder, M.A., Bater, C.W., Coops, N.C., Hilker, T., and White, J.C. 2008. The role of LiDAR in sustainable forest management. *For. Chron.* **84**(6): 807–826. doi:[10.5558/tfc84807-6](https://doi.org/10.5558/tfc84807-6).

Using Semi-Global Matching point clouds to estimate growing stock at the plot and stand levels: Application for a broadleaf-dominated forest in central Europe

Christoph Stepper^{1,*}, Christoph Straub¹, Hans Pretzsch²

¹Bavarian State Institute of Forestry (LWF), Department of Information Technology, Research Group: Remote Sensing, Hans-Carl-von-Carlowitz-Platz 1, D-85354 Freising

²Chair 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

*Corresponding author: Christoph.Stepper@lwf.bayern.de

Supplementary material

Supplementary table: Explanatory variables used to model growing stock

Table S1: List of explanatory variables derived from the remote sensing products.

No.	Name	Explanation	Unit	Data Source
height metrics	1 <i>E.MIN</i>	Minimum of point heights > 2 m	m	Image-based point cloud
	2 <i>E.MAX</i>	Maximum of point heights > 2 m	m	Image-based point cloud
	3 <i>E.MEAN</i>	Average of point heights > 2m	m	Image-based point cloud
	4 <i>E.MODE</i>	Mode of point heights > 2m	m	Image-based point cloud
	5 <i>E.P01</i>	1 st percentile of point heights > 2m	m	Image-based point cloud
	6 <i>E.P05</i>	5 th percentile of point heights > 2m	m	Image-based point cloud
	7 <i>E.P10</i>	10 th percentile of point heights > 2m	m	Image-based point cloud
	8 <i>E.P20</i>	20 th percentile of point heights > 2m	m	Image-based point cloud
	9 <i>E.P25</i>	25 th percentile of point heights > 2m	m	Image-based point cloud
	10 <i>E.P30</i>	30 th percentile of point heights > 2m	m	Image-based point cloud
	11 <i>E.P40</i>	40 th percentile of point heights > 2m	m	Image-based point cloud
	12 <i>E.P50</i>	50 th percentile of point heights > 2m	m	Image-based point cloud
	13 <i>E.P60</i>	60 th percentile of point heights > 2m	m	Image-based point cloud
	14 <i>E.P70</i>	70 th percentile of point heights > 2m	m	Image-based point cloud
	15 <i>E.P75</i>	75 th percentile of point heights > 2m	m	Image-based point cloud
	16 <i>E.P80</i>	80 th percentile of point heights > 2m	m	Image-based point cloud
	17 <i>E.P90</i>	90 th percentile of point heights > 2m	m	Image-based point cloud
	18 <i>E.P95</i>	95 th percentile of point heights > 2m	m	Image-based point cloud
	19 <i>E.P99</i>	99 th percentile of point heights > 2m	m	Image-based point cloud
	20 <i>E.PROP_0_5</i>	Proportion of point measurements in stratum $0m \leq x < 5m$	%	Image-based point cloud
	21 <i>E.PROP_5_10</i>	Proportion of point measurements in stratum $5m \leq x < 10m$	%	Image-based point cloud
	22 <i>E.PROP_10_15</i>	Proportion of point measurements in stratum $10m \leq x < 15m$	%	Image-based point cloud
	23 <i>E.PROP_15_20</i>	Proportion of point measurements in stratum $15m \leq x < 20m$	%	Image-based point cloud
	24 <i>E.PROP_20_30</i>	Proportion of point measurements in stratum $20m \leq x < 30m$	%	Image-based point cloud
	25 <i>E.PROP_30_40</i>	Proportion of point measurements in stratum $30m \leq x < 40m$	%	Image-based point cloud
	26 <i>E.PROP_40_50</i>	Proportion of point measurements in stratum $40m \leq x < 50m$	%	Image-based point cloud
variability metrics	27 <i>CanRelRat</i>	Canopy relief ratio ((mean-min)/(max-min))	-	Image-based point cloud
	28 <i>E.AAD</i>	Absolute average deviation of point heights > 2m	m	Image-based point cloud
	29 <i>E.LCV</i>	L-moment coefficient of variation (second L-moment ratio)	-	Image-based point cloud
	30 <i>E.LSKEW</i>	L-moment skewness (third L-moment ratio)	-	Image-based point cloud
	31 <i>E.LKURT</i>	L-moment kurtosis (fourth L-moment ratio)	-	Image-based point cloud
canopy cover metrics	32 <i>E.CC_2.00</i>	Canopy cover at 2m (proportion of points above 2m)	%	Image-based point cloud
	33 <i>E.CC_MEAN</i>	Canopy cover at mean canopy height (proportion of points above mean canopy height)	%	Image-based point cloud
	34 <i>E.CC_MODE</i>	Canopy cover at modal canopy height (proportion of points above modal canopy height)	%	Image-based point cloud
	35 <i>cc_2m</i>	Canopy cover at 2m (proportion of pixels above 2m)	%	Canopy Height Model
	36 <i>cc_5m</i>	Canopy cover at 5m (proportion of pixels above 5m)	%	Canopy Height Model
	37 <i>cc_10m</i>	Canopy cover at 10m (proportion of pixels above 10m)	%	Canopy Height Model
	38 <i>cc_15m</i>	Canopy cover at 15m (proportion of pixels above 15m)	%	Canopy Height Model
	39 <i>cc_20m</i>	Canopy cover at 20m (proportion of pixels above 20m)	%	Canopy Height Model
	40 <i>cc_30m</i>	Canopy cover at 30m (proportion of pixels above 30m)	%	Canopy Height Model
	41 <i>cc_40m</i>	Canopy cover at 40m (proportion of pixels above 40m)	%	Canopy Height Model

No.	Name	Explanation	Unit	Data Source
spectral metrics	42 <i>MeanB</i>	Mean value for the blue band	-	Orthoimagery
	43 <i>MeanG</i>	Mean value for the green band	-	Orthoimagery
	44 <i>MeanR</i>	Mean value for the red band	-	Orthoimagery
	45 <i>MeanNIR</i>	Mean value for the near infrared band	-	Orthoimagery
	46 <i>MeanNDVI</i>	Mean value for the calculated NDVI [*] layer	-	Orthoimagery
	47 <i>DevB</i>	Standard deviation for the blue band	-	Orthoimagery
	48 <i>DevG</i>	Standard deviation for the green band	-	Orthoimagery
	49 <i>DevR</i>	Standard deviation for the red band	-	Orthoimagery
	50 <i>DevNIR</i>	Standard deviation for the near-infrared band	-	Orthoimagery
	51 <i>DevNDVI</i>	Standard deviation for the calculated NDVI [*] layer	-	Orthoimagery
	52 <i>ConB</i>	Mean contrast, based on the GLCM [†] of the blue band	-	Orthoimagery
	53 <i>ConG</i>	Mean contrast, based on the GLCM of the green band	-	Orthoimagery
	54 <i>ConR</i>	Mean contrast, based on the GLCM of the red band	-	Orthoimagery
	55 <i>ConnIR</i>	Mean contrast, based on the GLCM of the near infrared band	-	Orthoimagery
	56 <i>CorrB</i>	Mean correlation, based on the GLCM of the blue band	-	Orthoimagery
	57 <i>CorrG</i>	Mean correlation, based on the GLCM of the green band	-	Orthoimagery
	58 <i>CorrR</i>	Mean correlation, based on the GLCM of the red band	-	Orthoimagery
	59 <i>CorrNIR</i>	Mean correlation, based on the GLCM of the near-infrared band	-	Orthoimagery
	60 <i>EnergyB</i>	Mean energy, based on the GLCM of the blue band	-	Orthoimagery
	61 <i>EnergyG</i>	Mean energy, based on the GLCM of the green band	-	Orthoimagery
	62 <i>EnergyR</i>	Mean energy, based on the GLCM of the red band	-	Orthoimagery
	63 <i>EnergyNIR</i>	Mean energy, based on the GLCM of the near-infrared band	-	Orthoimagery
	64 <i>HomB</i>	Mean homogeneity, based on the GLCM of the blue band	-	Orthoimagery
	65 <i>HomG</i>	Mean homogeneity, based on the GLCM of the green band	-	Orthoimagery
	66 <i>HomR</i>	Mean homogeneity, based on the GLCM of the red band	-	Orthoimagery
	67 <i>HomNIR</i>	Mean homogeneity, based on the GLCM of the near-infrared band	-	Orthoimagery

* Normalized difference vegetation index

† Grey-level co-occurrence matrix

Supplementary calculations: Test results for the comparison of the distributions of growing stock values obtained from each of the four different inventory grid densities (*grids 1- 4*)

(all computations done in R)

1. Test for homogeneity of variances

```
> GrowingStock.bartlett <- bartlett.test(GrowingStock ~ grid, data = trainingData_allGrids)
> GrowingStock.bartlett

Bartlett test of homogeneity of variances

data: GrowingStock by grid
Bartlett's K-squared = 4.8272, df = 3, p-value = 0.1849
```

Based on the *p*-value of 0.1849 obtained, the Null-hypothesis could not be rejected. Therefore, we could assume homogeneity of variance.

2. Test for normality

```
> GrowingStock.shapiro <- tapply(trainingData_allGrids$GrowingStock,
  trainingData_allGrids$grid, shapiro.test)
> grid <- levels(trainingData_allGrids$grid)
> for (i in grid)
+ cat("grid ", i, ":", " ", GrowingStock.shapiro[[i]]$p.value, sep = "", "\n")

grid 1: 3.011689e-13
grid 2: 2.015005e-11
grid 3: 2.539675e-09
grid 4: 0.1824137
```

The *p*-values obtained from the Shapiro-Wilk-test indicate that the samples in *grids 1 – 3* do not fulfil the criteria necessary to assume that the data are normally distributed. This assumption held only for the *grid 4* sample, with a *p*-value of 0.1824 (significance level α : 0.05).

3. Kruskal–Wallis rank sum test , used to determine whether the samples can be assumed to come from the same distribution

```
> GrowingStock.kruskal <- kruskal.test(GrowingStock ~ grid, data = trainingData_allGrids)
> GrowingStock.kruskal

Kruskal-Wallis rank sum test

data: GrowingStock by grid
Kruskal-Wallis chi-squared = 2.9252, df = 3, p-value = 0.4033
```

Based on the results of the Kruskal-Wallis rank sum test presented above, we can assume all samples originated from the same distribution.

Supplementary figures

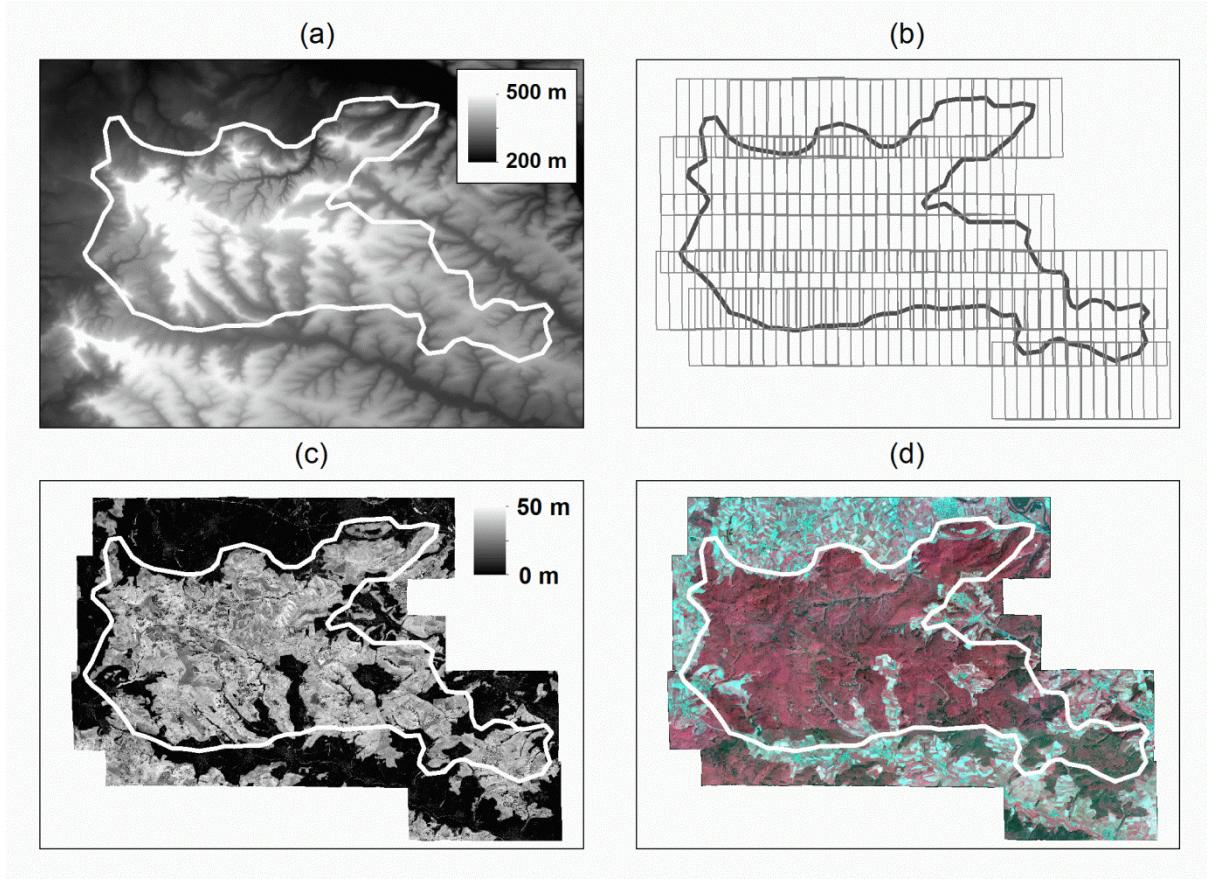


Fig. S1: Overview of the remote sensing data used in the study. (a) ALS-based DTM (1 m resolution) of the test site showing the elevation above sea level in m. (b) Image-block of 163 UltraCamXp stereo images (along-track overlap 75 %, across-track overlap 30 %, with a higher degree of overlap in the southern part of the study area due to the fact that two different flights were conducted). (c) CHM (1 m resolution) derived from the image-based point cloud and normalized by means of the ALS-based DTM. Heights are given in m above ground. (d) DSM-rectified orthoimage-mosaic, i.e. orthorectified using an image-based digital surface model and resampled to 1 m resolution; displayed here as a color-infrared image.

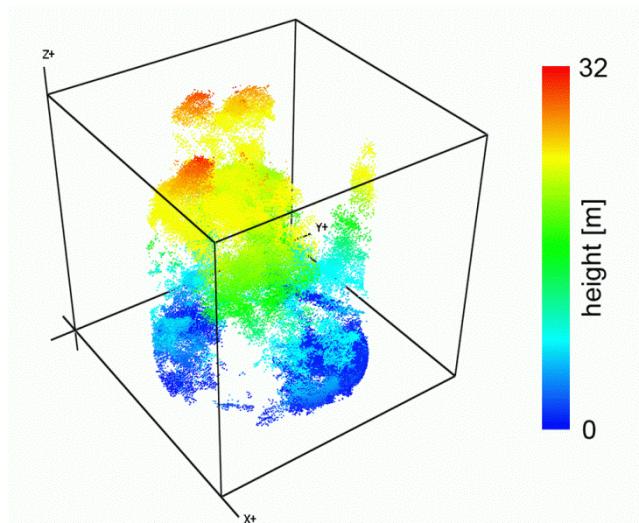


Fig. S2: Image-based point cloud normalized to heights above ground using an ALS-based DTM and clipped to the extent of one 500 m² inventory plot.

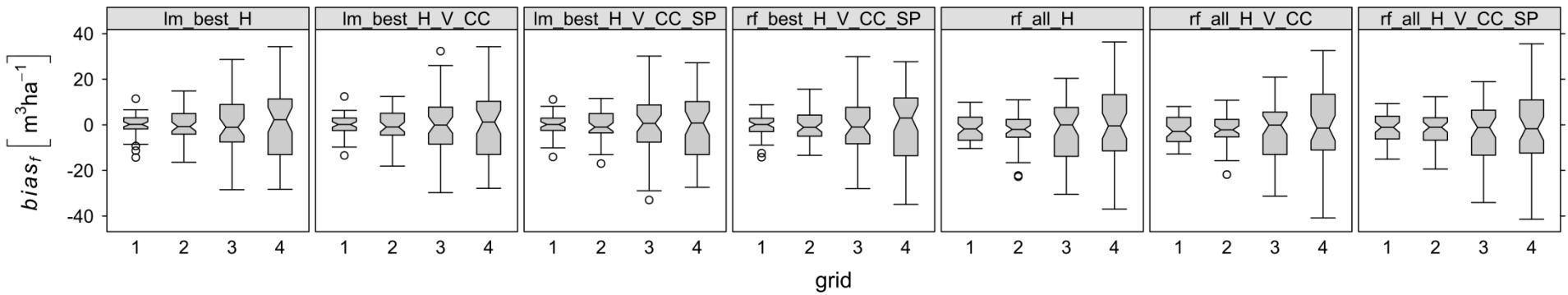
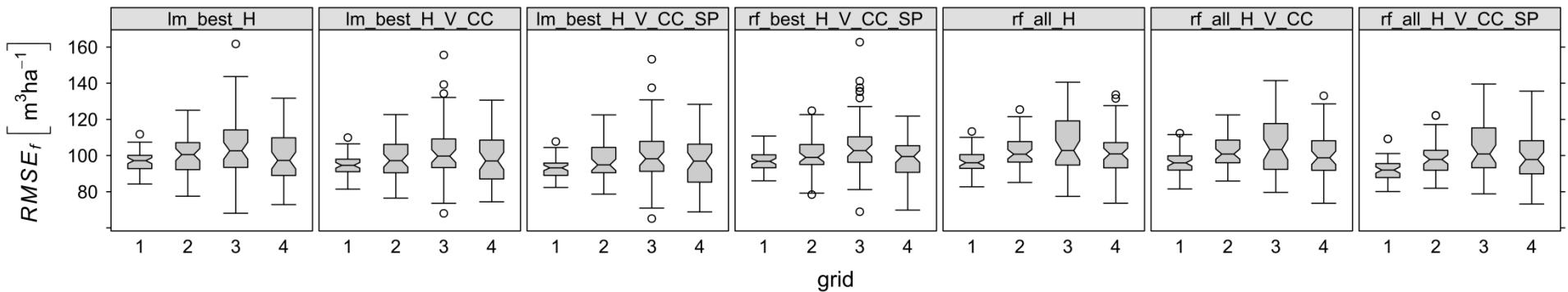


Fig. S3: Distribution of the absolute RMSE_f - and bias_f -values as obtained from the 50 held-out samples of the 10-fold cross-validation repeated five times for each of the different modeling approaches and inventory grids tested.

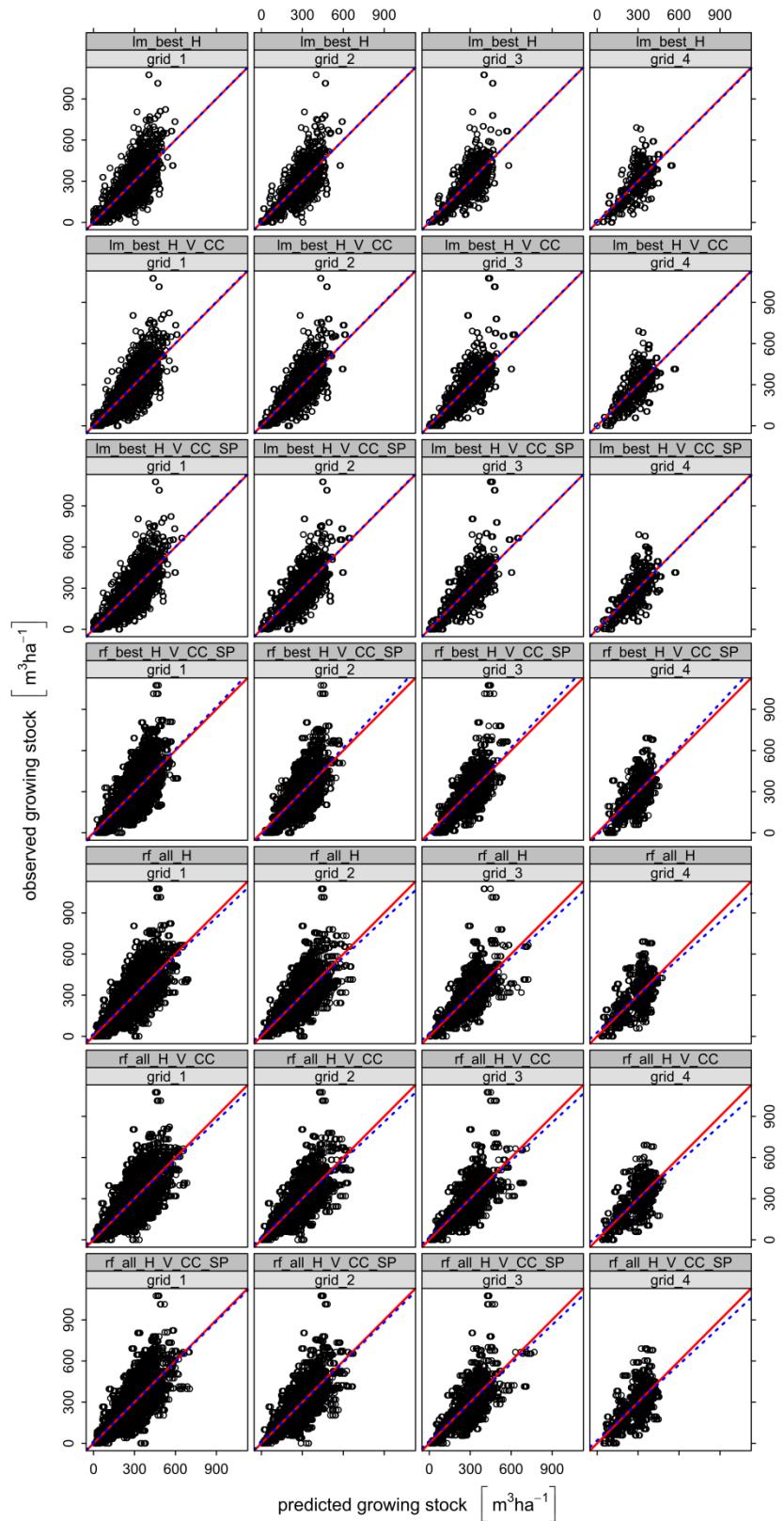


Fig. S4: Predicted vs. observed growing stock ($\text{m}^3 \text{ha}^{-1}$) at the plot level as obtained from the 50 held-outs of the 10-fold cross-validation (repeated five times). Linear trend lines are given as dashed blue lines and the 1:1 lines are given as solid red. Separate plots are shown for each of the different combinations of modeling approach and inventory grid.

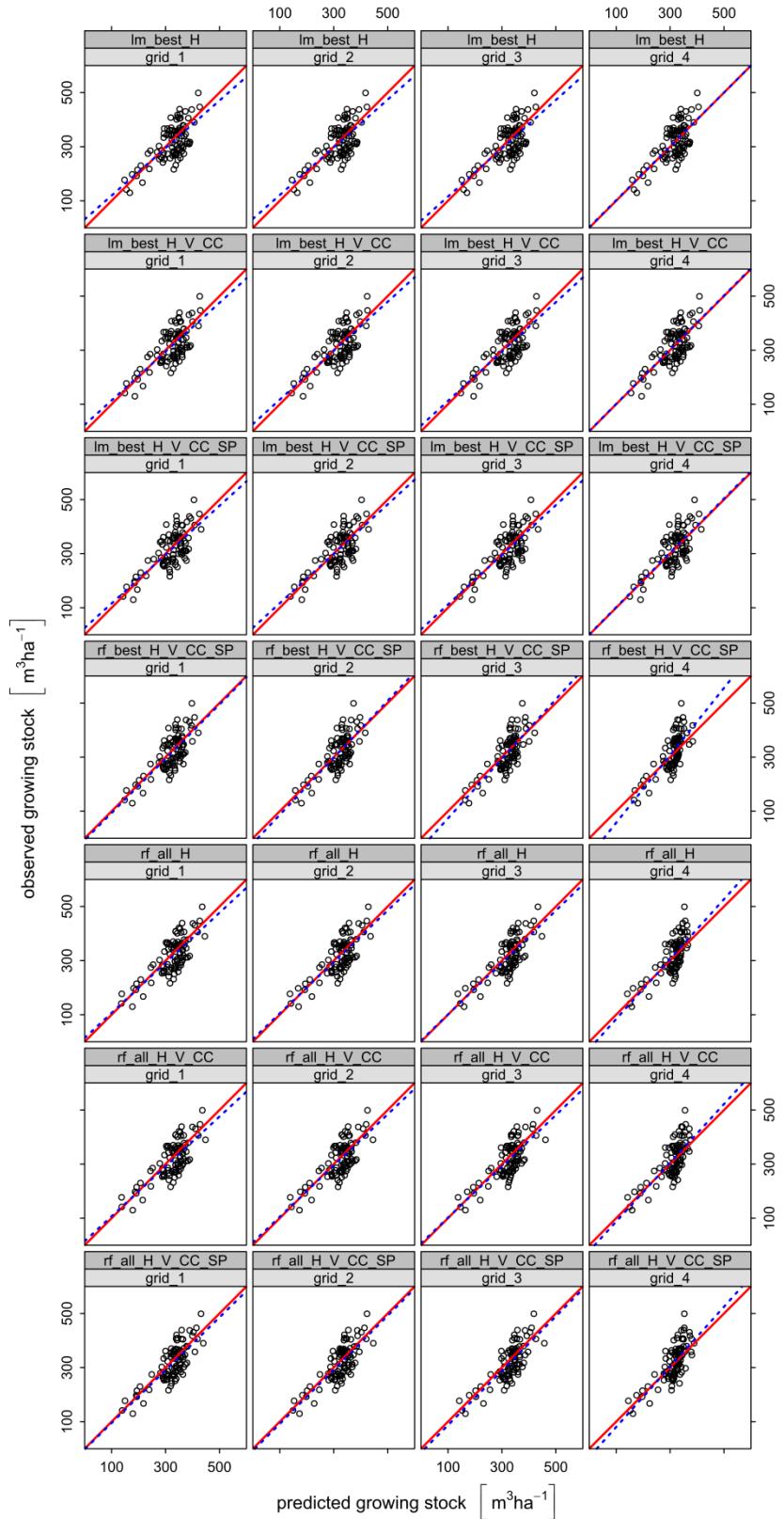


Fig. S5: Predicted vs. observed growing stock ($\text{m}^3 \text{ha}^{-1}$) for 94 validation stands. All stands which fully contained a minimum of five terrestrial inventory plots were used for the comparison. The predicted growing stock was derived by averaging the predictions from the hexagonal prediction units within each stand unit. The observed growing stock volume was calculated as the arithmetic mean of the terrestrial inventory plot measurements within each stand. Linear trend lines are given as dashed blue lines and the 1:1 lines are given as solid red. Separate plots are shown for each of the different combinations of modeling approach and inventory grid.