International Journal of Applied Earth Observation and Geoinformation 49 (2016) 107–116

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Modelling the standing timber volume of Baden-Württemberg—A large-scale approach using a fusion of Landsat, airborne LiDAR and National Forest Inventory data

Joachim Maack[a,*], Marcus Lingenfelder[b], Holger Weinacker[a], Barbara Koch[a]

[a] University of Freiburg, Chair of Remote Sensing and Landscape Information Systems (FeLis), Germany
[b] University of Freiburg, Chair of Forest Operations, Germany

ARTICLE INFO

Article history:
Received 20 November 2015
Received in revised form 9 February 2016
Accepted 10 February 2016
Available online 18 February 2016

Keywords:
Timber volume modelling
Large-scale interpolation
Generalized additive models
Airborne LiDAR
Landsat 7

ABSTRACT

Remote sensing-based timber volume estimation is key for modelling the regional potential, accessibility and price of lignocellulosic raw material for an emerging bioeconomy. We used a unique wall-to-wall airborne LiDAR dataset and Landsat 7 satellite images in combination with terrestrial inventory data derived from the National Forest Inventory (NFI), and applied generalized additive models (GAM) to estimate spatially explicit timber distribution and volume in forested areas. Since the NFI data showed an underlying structure regarding size and ownership, we additionally constructed a socio-economic predictor to enhance the accuracy of the analysis. Furthermore, we balanced the training dataset with a bootstrap method to achieve unbiased regression weights for interpolating timber volume. Finally, we compared and discussed the model performance of the original approach ($r^2 = 0.56$, NRMSE = 9.65%), the approach with balanced training data ($r^2 = 0.69$, NRMSE = 12.43%) and the final approach with balanced training data and the additional socio-economic predictor ($r^2 = 0.72$, NRMSE = 12.17%). The results demonstrate the usefulness of remote sensing techniques for mapping timber volume for a future lignocellulose-based bioeconomy.

1. Introduction

In 2012 the European Commission adopted a strategy to support a bioeconomy based on renewable resources (European Commission, 2012). Following this strategy, one essential goal for Germany is to replace fossil fuels and materials with renewable products derived from algae, crop and wood (details see BMBF, 2010). As sources for sustainable wood production are limited, the demand for accurate, large-scale assessments of forest resources is growing (Treuhaft et al., 2003; Rudel et al., 2005). There is strong interest in estimating biomass or timber volume for different utilizations such as fuel or new products like bio-plastics. Before a bioeconomy can emerge it is important to calculate the potential supply of future resources. With this study we would like to present one example for the federal state of Baden-Württemberg (Germany). We estimated the standing timber volume using wall-to-wall remote sensing data in combination with ground truth data derived from the German National Forest Inventory (NFI). We estimated the timber volume instead of biomass since it is required for further market and price calculations.

New remote sensing platforms and sensors with higher spatial, spectral and temporal resolutions are continuously being developed. A variety of studies have estimated forest biomass with various data and methodologies (Lu, 2006; Hartig et al., 2012; Fassnacht et al., 2014). A sound procedure is combining ground truth reference data from field surveys with statistical predictors derived from remotely sensed information (Woodhouse et al., 2012). The crucial parameter for accurate estimation is tree height as it is strongly linked to timber volume (Koch, 2010). Light detection and ranging (LiDAR) systems represent one of the most accurate methods for tree height measurements on a local to regional scale (Lefsky et al., 2002; Clark et al., 2011; Næsset et al., 2011). While airborne missions are often too expensive for federal state-wide investigations, we fortunately got access to a unique wall-to-wall LiDAR dataset that covers the whole federal state of Baden-Württemberg (approximately 35742 km^2 with a resolution of 0.8 pulses per m^2). This data was originally collected by the State Agency for Spatial Information and Rural Development Baden-Württemberg for calculating a statewide Digital Terrain Model

* Corresponding author.
E-mail address: joachim.maack@felis.uni-freiburg.de (J. Maack).

http://dx.doi.org/10.1016/j.jag.2016.02.004
0303-2434/

Fig. 1. Study area Baden-Württemberg (black) located in southwest Germany.

(DTM). The raw data allowed us to derive a DTM, a Digital Surface Model (DSM) and a Canopy Height Model (CHM).

Recent studies have shown that combining spectral or texture information with LiDAR data enhances the accuracy of biomass estimations (Fassnacht et al., 2014; Maack et al., 2015; Kattenborn et al., 2015). We combined LiDAR data with Landsat 7 images to get additional information about vegetation productivity, vitality, and species composition. Furthermore, we embrace the idea that abiotic factors like terrain characteristics that influence tree growth should be taken into consideration for estimating forest properties (Bonan et al., 1992; Sveinbjörnsson et al., 2002; Loranty and Goetz, 2012; Maack et al., 2015).

Several regression models were utilized for estimating forest characteristics such as biomass or biodiversity (Fassnacht et al., 2014). Machine learning algorithms like Classification and Regression Trees (CART) or generalized additive models (GAM) have already shown promising performance in several studies (Fassnacht et al., 2014; Kattenborn et al., 2015; Maack et al., 2015). They feature clear advantages such as reducing effects of co-linearity between predictor variables (Bright et al., 2012) and handling non-linear relationships between predictor and response variables (Guisan et al., 2002) when compared to classical linear models. We therefore used randomForest (RF), which is a very frequently utilized CART-based model (see Fassnacht et al., 2014), and generalized additive models (GAM), and checked their performance for large-scale timber volume estimations. The best performing algorithm was then used for the final modelling approach. Accordingly we evaluated the potential of satellite-based images (Landsat 7) and airborne laser scanning-derived (ALS) Canopy Height Models (CHMs) to estimate forest timber volume accurately on an extensive spatial scale in southwest Germany.

2. Study site and material

2.1. Study site Baden-Württemberg, Germany

Baden-Württemberg (center at 48° 32′ 16″N, 9° 2′ 28″E) is the third largest federal state of Germany located in the southwest corner of the country. It extends over 35,751 km^2 whereby about 14000 km^2 (39%) of the area is forested. The dominant tree species

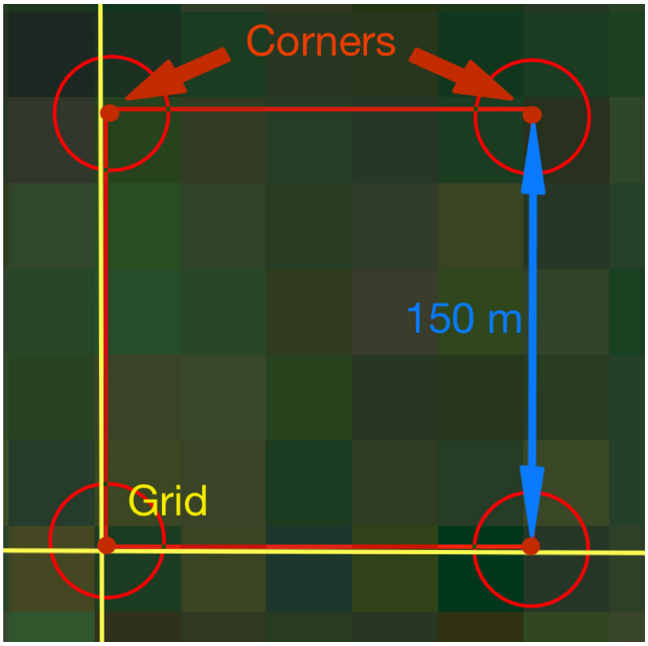

Fig. 2. The sampling scheme of the National Forest Inventory system in Baden-Württemberg. The sampling is performed along a 2 × 2 km grid (yellow). The measurements are performed at the corners of a 150 m square. The lower left point of the square is the intersection point of the grid, which is called a tract. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are 38% spruce (*Picea abies* (L.) H. Karst) and 21% beech (*Fagus sylvatica* L.) followed by 8% fir (*Abies alba* Mill.), 7% oak (*Quercus robur* L., *Quercus petraea* (Matt.) Liebl., *Q. rubra* L.), 7% pine (mainly *Pinus sylvestris* L.), 3% douglas fir (*Pseudotsuga menziesii* (Mirb.) Franco) and 2% larch (*Larix decidua* Mill.). All other species are summarized as other broadleaf-trees with long (10%) or short (4%) lifespans (details in Kändler et al., 2005). The study area includes all types of forest ownership ranging from state forest (24%) to corporate forest (40%) and private forests (36%) (for more details, see Kändler et al., 2005; Kändler and Cullmann 2014). The area is diverse in climate since it is located in a transitional zone between maritime climate to the west and continental climate to the east (Kullen, 1989). Mean annual temperatures range from 4 °C in montane regions to 9 °C in the lowlands. Overall, temperature is highly altitude dependent. The terrain ranges from 85 to 1493 m above sea level, and features a variety of different landscapes from lowlands and flat hills to low mountain ranges.

2.2. Field sampling

The ground truth data originate from the NFI in Baden-Württemberg, a standardized sampling procedure of all forested areas on a 2 × 2 km grid ($n = 4339$) with up to 4 field sampling points (corners) centered around on each tract of the grid (Fig. 2). The terrestrial field measurements are collected periodically in a 10-year cycle at permanently marked positions. In addition to data derived from the second sampling period (2001, 2002), the repeated measurements from the third period (2011–2012) were included for detecting change and enhancing GPS-accuracy. This field sampling recorded information about tree species, age, height, diameter at breast height (1.3 m; DBH) and forest ownership. Tree data were collected using an angle-count sampling methodology. Accordingly, timber volume in $m^3\,ha^{-1}$ was calculated from height and DBH using allometric equations (details see Polley, 2001).

2.3. Remote sensing data acquisition and preprocessing

The ALS data was collected between 2000 and 2005, and covered the whole area of Baden-Württemberg with a point density of 0.8 per m^2 and a spatial overlap of about 80–110% depending on the relief energy. The relief energy describes the height differences of the terrain within the study area. Data collection was performed during wintertime (detailed information about the ALS field survey is given in Schleyer, 2001). We calculated a DTM with a resolution of 1 m × 1 m. A Canopy Height Model (CHM) was subsequently produced by subtracting the DTM from the first LiDAR pulses. LiDAR data preprocessing was done using TreesVis 0.86 (Weinacker et al., 2004).

We enriched our dataset with additional spectral data using five, cloud free Landsat 7 images from a two year period (2001–2003, June–August) that covered the whole study area. Landsat 7 images have shown their potential for estimating biomass in several studies (Zheng et al., 2004; Hall et al., 2006; Powell et al., 2010). Pre-processing was performed in ENVI 5.1 (Exelis Visual Information Solutions, Boulder, Colorado, USA) using Landsat 7 settings; we applied radiometric correction and FLAASH atmospheric correction. Statistical analysis was carried out using R statistical sofware (R Core Team, 2015) version 3.1.2. GAM (Wood, 2006; Chambers and Hastie, 1992) and RF (Breiman, 2001) were implemented using "mgcv" (Wood, 2001) and "randomForest" packages. Spatial interpolation was performend using "raster" and "rgdal" packages. Since the remote sensing data and terrestrial field measurements covered the entire federal state, an extrapolation beyond the spatial range of observations was not neccesary.

2.4. Outlier reduction and signal smoothing

The NFI of Baden-Württemberg provided federal state-wide terrestrial reference data on forest timber volume. The data was not per se usable for remote sensing approaches because of some special characteristics. The most problematic issue was the combination of an angle-count sampling scheme with GPS inaccuracies. While the sampling points are permanently marked in the field, the GPS reference data is biased by this inaccuracy. The mean GPS difference between the meansurement of 2001/02 and 2011/12 was about 10 m. The only accuracy reference was the difference between these two measurements. Furthermore the remote sensing data featured a resolution of 30 × 30 m and was collected between 2000 and 2005. This meant a timelag between ground truth sampling and acquiring remote sensing data influenced the approach. Inaccuracies in spatial and temporal dimensions thus had to be addressed. A temporal problem occured if plots were harvested after the terrestrial measurement and before the LiDAR flight. To counter this issue we used the additional data of the third NFI (2011/2012). Field samples with a timber volume change of more than 15% between the second and third NFI were excluded

Table 1
Range of timber volume values at tree and plot level.

Ownership	Number of treesMean[N]	Number of treesMax[N]	Mean plotVolume[$m^3\,ha^{-1}$]
Private	10.8	25	453
Corporate	9.2	23	381
National	8.9	18	329

because it is unclear if the change happened before or after collecting the LiDAR data. The 15% boundary was a compromise between our mean timber volume estimation error and the time lag between terrestrial field measurement and remote sensing data acquisition. This procedure excluded 215 points (Fig. 1).

To handle the spatial relationship between field measurements and remote sensing data we first visually checked the reference data and the sampling radii in areas where high-resolution (~0.5 m) aerial photographs were available. It became clear that some field plots where located near forest edges, roads, lakes or glades. It also became obvious that the sampling radii differed greatly from 30 m to just a few centimeters. To get a sound approach we had to minimize the impact of big objects that were obviously not trees, and we had to harmonize the remote sensing data resolution with the varying sampling radii that were additionally affected by GPS inaccuracies. To handle the first issue we resampled the images at 5 × 5 m before extracting the training data for the interpolation, and we visually checked different sampling radii between 15 and 30 m. Finally, we applied circular buffers with a 20 m radius as a compromise between average ground truth sampling radius and sensor resolution. This way we still had the signal of the roads or lakes in the pixels, but their weighting was considerably lower.

To handle the second issue we used the mean of all corners of a tract if they belonged to the same owner (see Fig. 2). We thus treated each tract as a cluster with repetead measurements similar to the approach of (Kattenborn et al., 2015). This also countered GPS-inaccuracies and "edge effects" like overlapping crowns or other objects (e.g., pylons and high-voltage lines). This procedure potentially shrank the value range of the training data and thus the explanatory power of the model for small areas. The spatial extended to reference data changed from about 0.125 to 0.5 ha. In fact it shrank the value range slightly by 3%. For a statewide approach it made no difference since we were aiming for positioning of biomass fractioning plants with catchment areas of more that 10 km.

3. Influencing factors

Terrain dependent factors like climate, soil, and nutrient conditions have an effect on tree growth, vitality and species composition (Bonan et al., 1992; Sveinbjörnsson et al., 2002; Loranty and Goetz, 2012). Further studies have also pointed out characteristics like terrain altitude, slope and exposition influence statistical biomass modeling (Bonan et al., 1992; Sveinbjörnsson et al., 2002; Loranty and Goetz 2012; Maack et al., 2015).

A saturation effect of tree height occurs since older trees grow slower in height than in diameter compared to younger trees (Ryan and Yoder, 1997). Since tree height is one of the most important predictors of timber volume, this effect influences timber volume modeling. If LiDAR resolution is sufficient, this effect can be taken into account by individual crone diameter measurements since they feature a strong link to stem diameter (e.g., Popescu et al., 2003). Unfortunately, the LiDAR resolution of our approach was insufficient for single tree detection and crown diameter measurements.

Management alters the structure of forests and therefore influences forest productivity (Andersson and Birot, 2004). Depending on the resolution of the height measurement techniques (e.g., LiDAR or RADAR) understory vegetation layers influence plot-based mean height and reduces timber volume estimation accuracy. We assumed management activities like periodic thinning strongly reduced understory vegetation and thus enhance the accuracy of LiDAR height measurements with relatively low point density. Furthermore, we expected the mean age of trees was lower in highly managed forests.

The NFI-data allowed us to identify three ownership classes: highly managed state forest, highly but heterogeneously managed corporate forest and diverse private forest with varying management impacts (Schaffner 2001; Hausler and Scherer-Lorenzen, 2001). To examine this issue we split the data into different classes moderated by the ownership and size of the stand.

3.1. Socio-economic predictor

Mean timber volume from NFI data varied for forests under different ownerships (Tables 1 and 2). Due to differences in forest size and ownership from ground truth sampling, we created an additional socio-economic predictor. This predictor was constructed using 24 classes with a combination of ownership type and forest size where "P1" is private forest between 1–5 ha and "N8" is national forest owned by the state larger than 1000 ha (see Table 2). We used the given size classes of the NFI. We assumed that this predictor indicated forest density and thus would enhance the accuracy of spatial timber volume estimations; the final set of predictors is listed in Table 3.

3.2. Feature space set-up

We selected predictors based on the availability of wall-to-wall remote sensing products and experience from other studies performed in Baden-Württemberg. The initial set of predictors included Landsat 7 Bands 1–5 and 7, and the corresponding NDVI. The LiDAR-derived predictors were calculated from the raw data using TreesVis 0.86 (Weinacker et al., 2004) with a resolution of 1 m. Subsequently the LiDAR products were adjusted to Landsat 7's resolution of 30 m to calculate the mean, standard deviation, maximum and 10th to 90th height quantiles of the canopy height as well as terrain height above sea level, slope and exposition. To select the final set of predictors we checked pairwise correlations (Pearsoní's r) between all predictors and excluded one of the intercorrelated predictors ($r > 0.7$). To decide which one of an intercorrelated pair had to be removed we considered the level of significance. If both predictors featured the same level of significance we checked the change in RMSE for the final timber volume estimation model. RMSE was calculated using 10 repetitions of a 10-fold cross validation ($n = 100$ runs). All variables including the socio-economic predictor went through this procedure.

3.3. Regression models and training data

In order to identify the best performing algorithm we compared RF and GAM using NRMSE and r^2. The RF is a classification and regression model that builds a number of uncorrelated random decision trees. For classification and regression, each tree of the random forest votes for one decision and the path with most votes will be finally applied for the regression (for details, see Breiman, 2001). This structure makes the RF very efficient in terms of parallelization and handling large datasets. GAM is a semi- or non-parametric regression technique. It creates an additive, non-linear smooth function for each predictor-response relationship (for details, see Wood, 2001). As expected, both algorithms worked well and resulted in comparable results. Compared to the RF, GAM provided better results in terms of both explained variance (r^2 0.53–0.56) and lower uncertainties (NRMSE 11.25% to 9.64%). The NRMSE values were calculated using the full dataset and 100 repetitions of a 10-fold cross validation ($n = 1000$ runs). With regard to these findings, we performed further investigations using GAM. The results of both models indicated a biased regression slope that led to overestimation and underestimation of small and large timber volumes, respectively. We thus balanced the regression using more evenly distributed training data. We divided

Table 2
Socio-economic predictor classes combining ownership and size.

Size [ha]	Class (private forest)	Mean timber volume (private forest) [m^3 ha^{-1}]	Class (corporate forest)	Mean timber volume (corporate forest) [m^3 ha^{-1}]	Class (national forest)	Mean timber volume (national forest) [m^3 ha^{-1}]
1–5	P1	466.7	C1	509.8	N1[a]	–
5–20	P2	490.9	C2	362.2	N2[a]	–
20–50	P3	488.6	C3	349.4	N3[a]	–
50–100	P4	486.6	C4	357.8	N4[a]	–
100–200	P5	499.3	C5	365.2	N5[a]	–
200–500	P6	422.5	C6	340.3	N6[a]	–
500–1000	P7	404.4	C7	396.6	N7[a]	–
>1000	P8	377.5	C8	398.9	N8	329.2

[a] No data.

Table 3
Final set of predictors.

Predictor	Source	Description	Level of significance (*p*-value)
CHM mean height	LiDAR	Mean of surface minus terrain in meters	<0.001
CHM standard deviation	LiDAR	Standard deviation of surface minus terrain in %	<0.001
DTM mean height	LiDAR	Terrain height in meters above sea level	<0.01
Slope	LiDAR	Terrain slope in %	<0.001
Green band	Landsat 7	Reflectance 520–600 nm	<0.001
Red band	Landsat 7	Reflectance 630–690 nm	<0.001
NIR band	Landsat 7	Reflectance 770–900 nm	<0.001
Ownership and Size	NFI	24 classes	<0.001

the dataset into five classes of timber volume (0–200, 200–400, 600–800, 800–1200 m^3 ha^{-1}), from which the training data was randomly drawn with the same number of samples for each class.

4. Results

4.1. Model performance and uncertainties

The GAM showed high flexibility and therefore was valuable for large-scale assessments of extensive and diverse areas. The results of the original approach indicated a higher model performance for state-(NRMSE 8.9%) and publicly-owned (NRMSE 8.9%) forests in comparison to private (NRMSE 11.8%) forests. Furthermore, the difference between r^2 and adjusted r^2 indicated a slight overfit for the national forest dataset. The predictive r^2, RMSE and NRMSE displayed the performance for interpolating timber volume for the whole study site. An estimation error between 110.2 m^3 ha^{-1} (8.9%) and 147.1 m^3 ha^{-1} (11.8%) was calculated through 100 model runs using a 10-fold cross-validation (see Table 4). The explained variance was 53–63% (predictive r^2). The goodness of fit plots (Fig. 3) indicated an overestimation of low and an underestimation of high timber volume values. We hence trained the model with a more evenly distributed dataset. This approach resulted in a better regression fit (see Fig. 4) with higher r^2 values, but also with higher RMSE values (Table 5). The resulting r^2 values varied between 0.64 (private forest) and 0.80 (state forest) for different ownership classes. Also, NRMSE values showed a varying mean error between 14.87% for private and 11.01% for national forest (see Table 5). The divergent estimation performance between the different types of ownership showed there was some underlaying structure that influenced the estimation which could help further explain the spatial timber volume distribution in Baden-Württemberg. Private forests have on average 20% more trees per hectare and 37% larger timber volumes than national forests at the plot level (Table 1).

The new predictor enhanced the measured performance in terms of lower RMSE and higher r^2 values (Tables 4 and 5). Furthermore, the goodness of fit plots (Figs. 3–5) revealed much better regression slopes with consequently more accurate estimation of low and high timber volume values and a more reasonable distribution of the fitted values. A Shapiro–Wilk-Test validated the normal distribution ($W = 0.98$) of the residuals (Fig. 5(right)).

Fig. 6(a) shows the map of the final approach using the mean of 50 bootstrap runs. It represents the spatial distribution of timber volume for the whole state with a resolution of 30 × 30 m. The close-ups (b and c) show the clear separation of forested and unforested areas as well as subtle differences in terms of standing timber volume within the forests.

5. Discussion

These results show potential timber volume interpolation across a large area. We used data from terrestrial forest inventory in combination with generalized additive models (GAM) and a CART-based (RF) regression technique for estimating timber volume. Overall the GAM performed very well, and the results are in line with other studies using related data and methodology (an overview is given in Fassnacht et al., 2014). While there are many approaches for estimating timber volume or biomass, in southwest Germany a map of this spatial extent and resolution is, so far, unique and enables us to do further investigations, calculations and simulations for a future, lignocellulosebased bioeconomy.

5.1. Model performance

First of all, the GAM outperformed randomForest in terms of r^2 and RMSE. That contradicts findings of Kattenborn et al., (2015) where RF clearly showed better performance for estimating biomass with a fusion of hyperspectral and a TandemX derived CHM (Kattenborn et al., 2015). However, that study was carried out on a homogenous and relatively small study area. In addition, the authors stated the GAM performed well for separating urban from forested areas, and might work well for a more diverse study area with lager sample sizes. This possibility is supported by a study of Moisen and Frescino, (2002) that revealed higher accuracies for biomass estimation from GAM—compared to CART-based models on five test sites using 531–1393 forest plots. RF often shows excellent performance for handling big and complex datasets like hyperspectral images. The GAM outperformed the RF

Table 4
Model results using original training data.

Ownership	Share [%]	Count	R^2	Adj. R^2	Pred. R^2	SD (Pred. R^2)	RMSE [m^3 ha^{-1}]	NRMSE [%]
State	24	837	0.63	0.59	0.55	0.08	110.2	8.88
Publ. corporation	42	1465	0.57	0.56	0.54	0.05	110.5	8.91
Private	34	1353	0.53	0.52	0.50	0.06	147.1	11.8
All	100	3655	0.56	0.55	0.54	0.04	120.1	9.65
All + ownership and size	100	3655	0.57	0.56	0.55	0.03	119.6	9.64

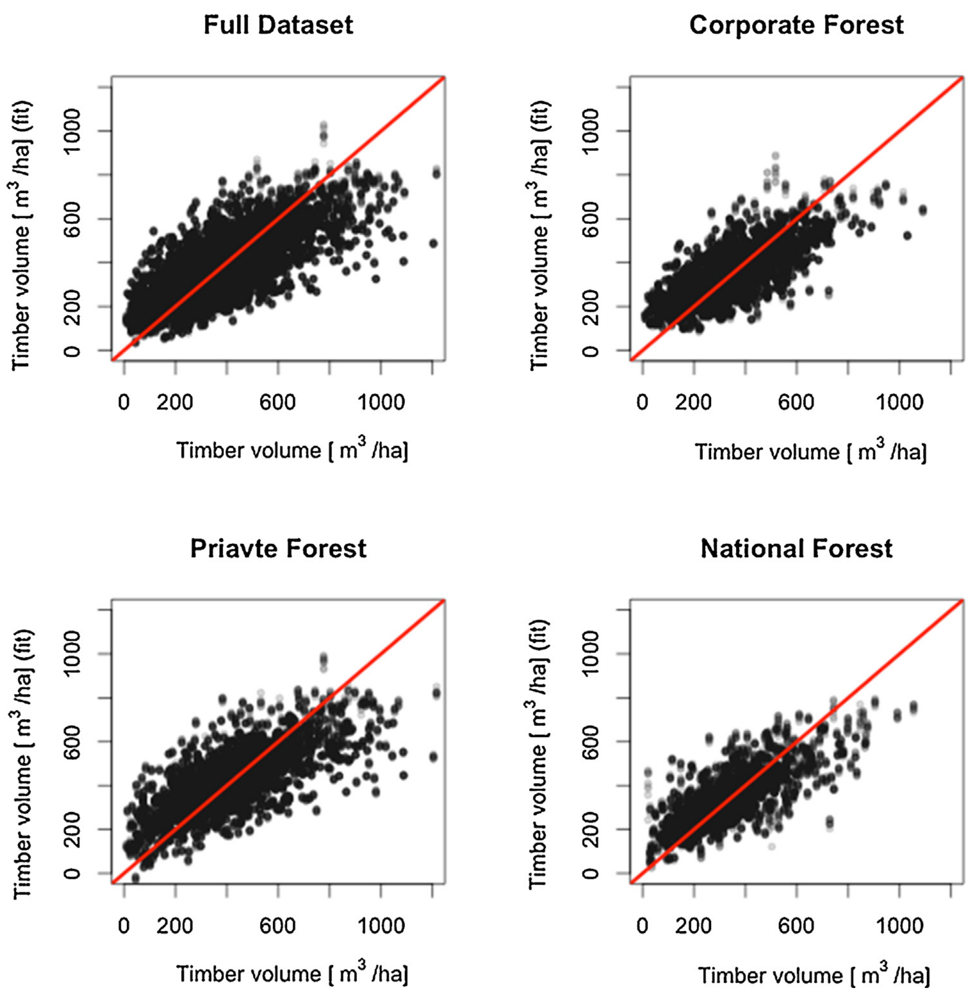

Fig. 3. Goodness of fit plots for the different ownership classes. Response (*x*-axis) and predicted (*y*-axis) timber volume values based on the GAM using varying combinations of the full dataset (upper left, n = 3655) corporate forest (upper right, n = 1465), private forest (lower left, n = 1353) and national forest (lower right, n = 837); solid red lines idicating optimal regression fits (n = 100 10-fold-CV). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Model results using the stratified sampling method for selecting training data (mean of 1000 10-fold-CV runs).

Ownership	Share [%]	Count	R^2	Adj. R^2	Pred. R^2	SD (Pred. R^2)	RMSE [m^3 ha^{-1}]	NRMSE [%]
State	24	500	0.78	0.78	0.73	0.12	141.8	11.01
Publ. corporation	42	500	0.80	0.78	0.76	0.09	136.5	11.43
Private	34	500	0.64	0.62	0.58	0.11	184.4	14.87
All	100	1000	0.69	0.68	0.68	0.06	154.1	12.43
All + ownership and size	100	1000	0.72	0.71	0.70	0.04	151.0	12.17

most likely due to the small number and simple structure of the predictors used. It was plausible that the used predictors showed a non-linear but additive relationship to timber volume, which could be described by a continuous function.

The first approach using a 10-fold crossvalidation (see Table 4) resulted in relatively good RMSE and NRMSE values. However, the corresponding scatterplots revealed biased regression weights through too shallow regression slopes. We assumed that this was caused by the high number of samples in the mid-range growing stock between 200 and 400 m^3/ha. We countered this issue by building five timber volume classes and drawing samples randomly from those classes until we got the same amount of data for each class. Thus, the regression weights were more evenly distributed and resulted in clearly steeper regression slopes (0.72–0.46)

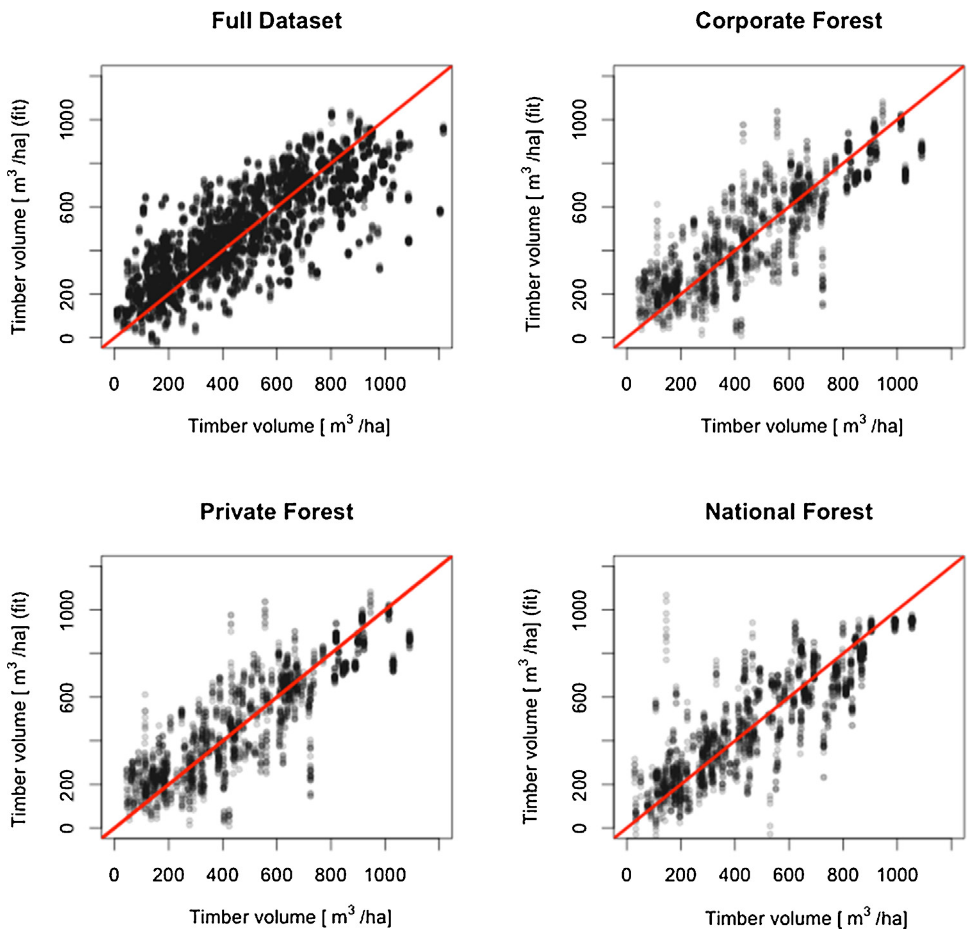

Fig. 4. Goodness of fit plots for the different ownership classes using weighted training data. Response (*x*-axis) and predicted (*y*-axis) timber volume values based on the GAM using varying combinations of the full dataset (upper left, $n = 1000$), corporate forest (upper right, $n = 500$), private forest (lower left, $n = 500$) and national forest (lower right, $n = 837$); solid red lines indicate optimal regression fits ($n = 100$ 10-fold-CV). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

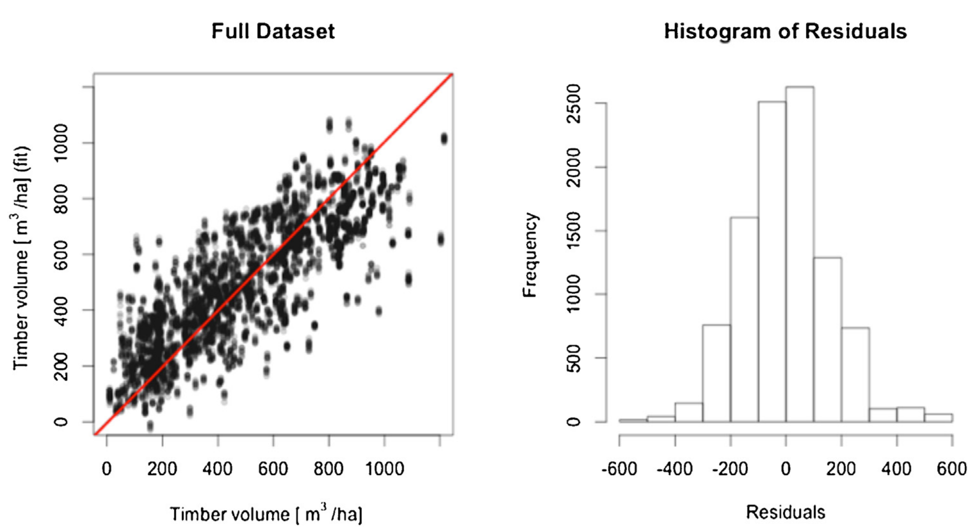

Fig. 5. Goodness of fit plot (left) for the final regression using weighted training data and additional socio-economic predictor. Response (*x*-axis) and predicted (*y*-axis) timber volume values based on the GAM using varying combinations of the full dataset (left, $n = 1000$). The solid red line indicates optimal regression fit ($n = 10$ 10-fold-CV). Residuals of the final regression were normally distributed (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(compare 3 and 4) and higher r^2 values (compare Tables 4 and 5). Furthermore, RMSE and NRMSE values (see Table 5 and Fig. 5) shows higher though more reasonable distributed error rates. The assumption that ownership and size of the forest might provide additional explanation was substantiated by the increased performance when using information about forest ownership and size

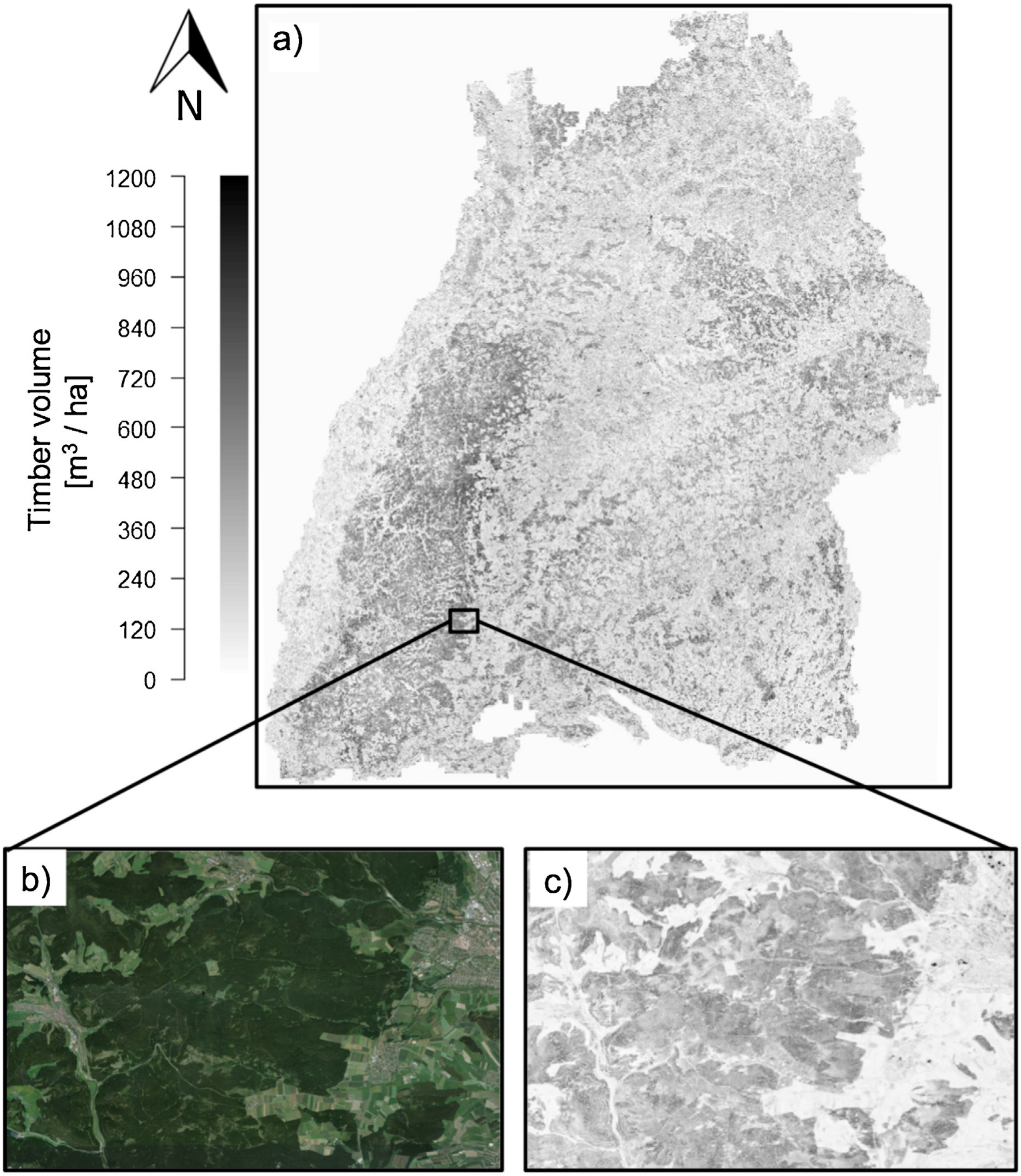

Fig. 6. The spatial distribution of timber volume for the whole federal state (a) with selected close-ups of the eastern border region of the Black Forest comparing a high-resolution spectral image (b) and the timber volume map (c).

(see Tables 4 and 5). The weaker performance for private forests was likely a result of the very heterogenous range of management approaches and intensities in comparison to the other ownership classes. For large-scale timber volume assessments we therefore recommend human impact on forests in terms of management be considered. However, the particular underlying struture has to be analyzed to construct possible new predictors and thus enhance estimation performance.

6. Limitations

Several sources of uncertainties affect indirect timber volume estimations. First, field samples inherit GPS positioning inaccuracies, and they rely on strongly simplified conversion factors to calculate the per-hectare volume. Furthermore, vegetation density including regeneration layer and shrubs is often not measured by field campaigns yet influence remote sensing-based height measurements. Third, there is no linear relationship between tree height and timber volume since younger trees grow faster in height than in diameter than older trees. Finally, tree species and soil dependent characteristics further influence tree growth.

The regression slope was affected by errors using the full, ground truth dataset. We assumed that understory vegetation influenced LiDAR measurements and in turn the estimation of stand height. In addition, very old trees might cause the underestimation of high timber volume values since they grow more in diameter than in height. A density parameter such as the number of trees or mean crown diameter per area should be used as an additional predictor to counter this issue. For two reasons we were not able to provide such a predictor. Firstly, the resolution of our LiDAR data (0.8 pulses/m^2) is not sufficient to detect single trees in dense stands. Furthermore, studies reported good accuracies for single tree delineation using airborne LiDAR with higher resolution (e.g., Reitberger et al., 2009; Vauhkonen et al., 2011). Secondly, the terrestrial inventory was based on angle-count samples with dynamic radii that did not allow for accurate, plot-based single tree extraction since there might have been trees within the sampling radii that were not sampled due to inherent measurement specifications (see Polley, 2001). However, if no remote sensing signal is available, other data may provide additional information about the forest structure or density. We were able to derive a density parameter for the forests of Baden-Württemberg from information on forest ownership and size, which allowed us to clearly improve our interpolation. We assume that the presented methodologies and results are transferable to countries with similar management strategies. This, though, has to be further investigated.

7. Outlook

The estimated timber volume map is a basis for calculating the economic and sustainable potential for a lignocellulose-based bioeconomy in Baden-Württemberg. We still need to convert the standing volume to regrowth and build a framework for calculating harvesting and transportation costs. Besides timber from forests, there is additional potential from trees outside forests (TOF), short rotation coppice and waste that ought to be integrated in further calculations. The final objective is a location analysis for possible lignocellulosic biomass fractioning or conversion plants to extract chemical intermediates that can replace petrochemicals for multiple applications.

Acknowledgments

This work was supported by a grant from the Ministry of Science, Research and the Arts of Baden-Württemberg (funding code: 7533-10-5-79). The study was developed in cooperation with the Forest Research Institute (FVA) in Baden-Württemberg (www.fva-bw.de). We would like to thank Dr. Gerad Kändler of the Forest Research Institute (FVA) in Freiburg for providing NFI and LiDAR data. The LiDAR data was originally collected by the State Agency for Spatial Information and Rural Development Baden-Württemberg (www.lgl-bw.de) Az.:2851.9-1/3. Felix Storch from the department of Silvicultures of the University of Freiburg shared his deep understanding of the National Forest Inventory data with us. ForstBW provided the information about forest ownership and size. NASA and USGS made the Landsat imaging available.

References

Andersson, F., Birot, Y., 2004. Towards the Sustainable Use of Europe's Forests–Forest Ecosystem and Landscape Research: Scientific Challenges and Opportunities. European Forest Institute.
Bonan, G.B., Pollard, D., Thompson, S.L., 1992. Effects of boreal forest vegetation on global climate. Nature 359 (6397), 716–718.
Bundesministerium for Building und Forschung (BMBF), 2010. Nationale Forschungsstrategie BioÖkonomie 2030—Unser Weg zu einer bio-basierten Wirtschaft. URL: http://www.bmbf.de/pub/biooekonimie.pdf.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Bright, B.C., Hicke, J.A., Hudak, A.T., 2012. Estimating aboveground carbon stocks of a forest affected by mountain pine beetle in Idaho using lidar and multispectral imagery. Remote Sens. Environ. 124, 270–281.
Clark, M.L., Roberts, D.A., Ewel, J.J., Clark, D.B., 2011. Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. Remote Sens. Environ. 115 (11), 2931–2942.
European Commission, 2012. Innovating for Sustainable Growth: A Bioeconomy for Europe. http://ec.europa.eu/research/bioeconom/pdf/201202_innovating_sustainable_growth_en.pdf.
Fassnacht, F.E., Hartig, F., Latifi, H., Berger, C., Hernández, J., Corvalán, P., Koch, B., 2014. Importance of sample size: data type and prediction method for remote sensing-based estimations of aboveground forest biomass. Remote Sens. Environ. 154, 102–114.
Guisan, A., Edwards, T.C., Hastie, T., et al., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Modell. 157 (2), 89–100.
Hall, R.J., Skakun, R.S., Arsenault, E.J., Case, B.S., 2006. Modeling forest stand structure attributes using Landsat ETM+ data: application to mapping of aboveground biomass and stand volume. For. Ecol. Manage. 225 (1), 378–390.
Hausler, A., Scherer-Lorenzen, M., 2001. Sustainable Forest Management in Germany: the Ecosystem Approach of the Biodiversity Convention Reconsidered. Federal Ministry of Environment, Bonn, Germany.
Hartig, F., Dyke, J., Hickler, T., Higgins, S.I., O'Hara, R.B., Scheiter, S., Huth, A., 2012. Connecting dynamic vegetation models to data—an inverse perspective. J. Biogeogr. 39 (12), 2240–2252.
Chambers, J.M., Hastie, T.J., 1992. Statistical models in S. Scientific Modeling and Simulation SMNS, 1.
Kändler, G., Cullmann, D., 2014. Der Wald in Baden-Württemberg. Ausgewählte Ergebnisse der dritten Bundeswaldinventur. Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg (FVA).
Kändler, G., Schmidt, M., Breidenbach, J., 2005. Der Wald in Baden-Württemberg im Jahr 2002 und seine Entwicklung seit 1987-Die wichtigsten Ergebnisse der zweiten Bundeswaldinventur. FVA Forstliche Versuchs-und Forschungsanstalt Baden-Württemberg, Artikel vom, 1, 2005.
Kattenborn, T., Maack, J., Faßnacht, F., Enßle, F., Ermert, J., Koch, B., 2015. Mapping forest biomass from space—fusion of hyperspectral EO1-hyperion data and Tandem-X and WorldView-2 canopy height models. Int. J. Appl. Earth Obs. Geoinf. 35, 359–367.
Koch, B., 2010. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. ISPRS J. Photogramm. Remote Sens. 65 (6), 581–590.
Kullen, S., 1989. Baden-Württemberg, 3rd. ed. Klett, Stuttgart.
Lefsky, M.A., Cohen, W.B., Harding, D.J., Parker, G.G., Acker, S.A., Gower, S.T., 2002. Lidar remote sensing of above-ground biomass in three biomes. Global Ecol. Biogeogr. 11 (5), 393–399.
Loranty, M.M., Goetz, S.J., 2012. Shrub expansion and climate feedbacks in Arctic tundra. Environ. Res. Lett. 7 (1), 011005.
Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. Int. J. Remote Sens. 27 (7), 1297–1328.
Maack, J., Kattenborn, T., Enßle, F., Hernández, J., Corvalán, P., Koch, B., 2015. Modeling forest biomass using very-high-resolution data—combining textural, spectral and photogrammetric predictors derived from spaceborne stereo images. Int. J. Remote Sens.
Moisen, G.G., Frescino, T.S., 2002. Comparing five modelling techniques for predicting forest characteristics. Ecol. Modell. 157 (2), 209–225.
Næsset, E., Gobakken, T., Solberg, S., Gregoire, T.G., Nelson, R., Ståhl, G., Weydahl, D., 2011. Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: a case study from a boreal forest area. Remote Sens. Environ. 115 (12), 3599–3614.
Polley, H., 2001. Aufnahmeanweisung für die Bundeswaldinventur II: (2001–2002). Bundesministerium für Verbraucherschutz, Ernährung und Landwirtschaft.
Popescu, S.C., Wynne, R.H., Nelson, R.F., 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Can. J. Remote Sens. 29 (5), 564–577.
Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B., Ohmann, J.L., 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: a comparison of empirical modeling approaches. Remote Sens. Environ. 114 (5), 1053–1068.
R Core Team, 2015. R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org/.
Reitberger, J., Schnörr, C., Krzystek, P., Stilla, U., 2009. 3D segmentation of single trees exploiting full waveform LIDAR data. ISPRS J. Photogramm. Remote Sens. 64 (6), 561–574.
Rudel, T.K., Coomes, O.T., Moran, E., Achard, F., Angelsen, A., Xu, J., Lambin, E., 2005. Forest transitions: towards a global understanding of land use change. Global Environ. Change 15 (1), 23–31.
Ryan, M.G., Yoder, B.J., 1997. Hydraulic limits to tree height and tree growth. Bioscience, 235–242.
Schaffner, S.J., 2001. Realisierung von Holzvorräten im Kleinprivatwald (Doctoral dissertation, Technische Universität München).
Schleyer, A., 2001. Das Laserscan-DGM von Baden-Württemberg. Wichmann Verlag, Heidelberg.
Sveinbjörnsson, B., Hofgaard, A., Lloyd, A., 2002. Natural causes of the tundra-taiga boundary. Ambio, 23–29.
Treuhaft, R.N., Asner, G.P., Law, B.E., 2003. Structure-based forest biomass from fusion of radar and hyperspectral observations. Geophys. Res. Lett. 30 (9).
Vauhkonen, J., Ene, L., Gupta, S., Heinzel, J., Holmgren, J., Pitkänen, J., Solberg, S., Wang, Y., Weinacker, H., Hauglin, K.M., Lien, V., Packalén, P., Gobakken, T., Koch, B., Næsset, E., Tokola, T., Maltamo, M., 2011. Comparative testing of single-tree detection algorithms under different types of forest. Forestry.
Weinacker, H., Koch, B., Weinacker, R., 2004. TREESVIS: a software system for simultaneous ED-real-time visualisation of DTM, DSM, laser raw data, multispectral data, simple tree and building models. International Archives of Photogrammetry. Remote Sens. Spatial Inf. Sci. 36, 90–95.

Wood, S.N., 2001. mgcv:GAMs and generalized ridge regression for R. R News 1 (2), 20–25.
Wood, S.N., 2006. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC Press.
Woodhouse, I.H., Mitchard, E.T., Brolly, M., Maniatis, D., Ryan, C.M., 2012. Radar backscatter is not a direct measure of forest biomass. Nat. Clim. Change 2 (8), 556–557.
Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J., Ryu, S.R., 2004. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. Remote Sen. Environ. 93 (3), 402–411.