Remote Sensing of Environment 154 (2014) 102–114

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass

F.E. Fassnacht [a,f,*], F. Hartig [b], H. Latifi [c], C. Berger [d], J. Hernández [e], P. Corvalán [e], B. Koch [f]

[a] Karlsruhe Insitute of Technology, Institute of Geography and Geoecology, Kaiserstraße 12, 76131 Karlsruhe, Germany
[b] University of Freiburg, Department of Biometry and Environmental System Analysis, Tennenbacherstrasse 4, 79106 Freiburg, Germany
[c] University of Wuerzburg, Department of Remote Sensing, Oswald-Kuelpe-Weg 86, 97074 Wuerzburg, Germany
[d] University of Jena, Department of Earth Observation, Loebdergraben 32, 07743 Jena, Germany
[e] Laboratorio de Geomática y Ecología del Paisaje, Universidad de Chile, Av. Santa Rosa 11315, Santiago de Chile, Chile
[f] University of Freiburg, Chair of Remote Sensing and Landscape Information Systems, Tennenbacherstrasse 4, 79106 Freiburg, Germany

ARTICLE INFO

Article history:
Received 9 January 2014
Received in revised form 25 July 2014
Accepted 31 July 2014
Available online 15 September 2014

Keywords:
Biomass estimation
ANOVA
Sample size
Prediction method
LiDAR
EO1-Hyperion
HyMap
Hyperspectral

ABSTRACT

Estimates of forest biomass are needed for various technical and scientific applications, ranging from carbon and bioenergy policies to sustainable forest management. As local measurements are costly, there is a great interest in obtaining reliable estimates over large areas from remote sensing data. Currently, such estimates are obtained with a variety of data sources, statistical methods and prediction standards, and there is no agreement on what are best practices for this task.

To improve our understanding of how these different methods affect prediction quality, we first conducted a systematic review of the available literature to identify the most common sensor types and prediction methods. Based on the review, we identified sample size of the reference points on the ground, prediction method (stepwise linear regression, support vector machines, random forest, Gaussian processes and k-nearest neighbor), and sensor type as the main differences that could potentially affect predictive quality. We then compared those factors in two case study areas in Germany and Chile, for which airborne discrete return Light Detection And Ranging (LiDAR) and airborne hyperspectral as well as airborne discrete return LiDAR and spaceborne hyperspectral data were available. For each factor combination, we calculated Pearson's coefficient of correlation between observations and predictions (r^2) and root mean squared error (RMSE) for bootstrapped estimates using k-fold cross-validation with a varying number of folds. Finally, Analysis of Variance (ANOVA) was used to quantify the influence of the factors on the predictive error of the biomass models.

Our results confirm previous findings that predictor data (sensor) type is the most important factor for the accuracy of biomass estimates, with LiDAR being preferable to hyperspectral data. In contrast to some previous studies, complementing LiDAR with hyperspectral data did not improve predictive accuracy. Also the prediction method had a substantial effect on accuracy and was generally more important than the sample size. In most cases, random forest performed best and stepwise linear models worst, judging from r^2 and RMSE under cross-validation. Additional results suggested that r^2 may deliver unrealistically large values when the hold-out sample during the cross-validation is too small.

In conclusion, our literature review revealed that different methods for biomass estimation are currently used, with no general agreement on best practices. In our case studies, we found substantial accuracy differences between those methods, with LiDAR data, in combination with a random forest algorithm and a large number of reference sample units on the ground yielding the lowest error for biomass predictions. The comparatively high importance of the statistical prediction method seems particularly relevant, as they suggest that choosing the appropriate statistical method may be more effective than obtaining additional field data for obtaining good biomass estimates. Considering the costs of improving accuracy of global and regional biomass estimates by ground measurements, it seems sensible to invest in further comparative studies, preferably with a wider range of sites and including also RADAR sensors, to establish robust best-practice recommendations for obtaining regional and global biomass estimates from remote-sensing data.

© 2014 Elsevier Inc. All rights reserved.

* Corresponding author at: Karlsruhe Insitute of Technology, Institute of Geography and Geoecology, Kaiserstraße 12, 76131 Karlsruhe, Germany.
E-mail addresses: fabian.fassnacht@kit.edu (F.E. Fassnacht), florian.hartig@biom.uni-freiburg.de (F. Hartig), hooman.latifi@uni-wuerzburg.de (H. Latifi), christian.berger@uni-jena.de (C. Berger), jhernand@uchile.cl (J. Hernández), pcorvala@uchile.cl (P. Corvalán), barbara.koch@felis.uni-freiburg.de (B. Koch).

http://dx.doi.org/10.1016/j.rse.2014.07.028
0034-4257/© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Forest ecosystems account for the dominant share of terrestrial biomass stocks (Houghton, Hall, & Goetz, 2009). There is a strong interest in estimating these stocks with sufficient resolution across larger spatial scales, for example for bioenergy production, sustainable forest management, detection of land-use change and the assessment of carbon stocks for initiatives such as REDD and REDD+ (Hartig et al., 2012; Koch, 2010; Treuhaft, Asner, & Law, 2003). The wall-to-wall estimation of forest biomass over large areas by ground-based measurements requires a dense network of inventory plots to reach good accuracies. In many regions, this is infeasible due to high costs and required man power. This limitation is particularly evident in many sparsely-populated areas with notable portion of natural forest ecosystems that are considered crucial for climate and biodiversity. Using remote sensing data is therefore the only practical option to predict biomass on these scales with affordable effort.

Many previous studies have evaluated the possibility to estimate biomass by means of remote sensing information. Practically all major data sources have been used, including RAdio Detection And Ranging (RADAR) (e.g., Li et al., 2007; Saatchi, Marlier, Chazdon, Clark, & Russell, 2011; Sun et al., 2011; Tanase et al., 2014; Tsui, Coops, Wulder, & Marshall, 2013), Light Detection And Ranging (LiDAR) (e.g. Clark, Roberts, Ewel, & Clark, 2011; Dubayah et al., 2010; Hudak et al., 2012; Laurin et al., 2014; Næsset et al., 2013) and optical multi and hyperspectral data (e.g., Laurin et al., 2014; Morel, Fisher, & Malhi, 2012). Study areas range across all major forest ecosystems, including the boreal (Hyyppä et al., 2008); temperate (Latifi, Nothdurft, & Koch, 2010; Tsui, Coops, Wulder, Marshall, & McCardle, 2012) and tropical (Drake, Dubayah, Knox, Clark, & Blair, 2002; Treuhaft et al., 2010) zones as well as particular ecosystems in the sub-tropical zones and the mangroves (Li et al., 2007; Proisy, Couteron, & Fromard, 2007). Most studies follow an approach in which field-measured biomass values are used to train statistical or machine-learning methods in predicting biomass by remote sensing predictors, and the majority report favorably on the accuracy of their biomass predictions. Unfortunately, the diversity of data sources and study locations is matched by an equal diversity of statistical methods and modeling standards. Therefore, it is difficult to compare studies, and there is still no agreement on best practices to estimate biomass from remote sensing data.

In this study, we compare the performance of statistical methods for estimating biomass while varying the remote sensing sensor as well as the reference data size. We first conduct a literature review to identify the five most widely-used statistical prediction methods, and then apply the identified methods to airborne LiDAR and hyperspectral datasets from two study areas in southern Germany and central Chile. Using Analysis of Variance (ANOVA), we rank the impact of data type, statistical prediction method, and the size of the reference data according to their influence on model performance, and give recommendations for the most suitable algorithms for biomass estimation from remote sensing data.

2. Literature review

We explored Thomson Reuter's ISI web of knowledge database to access the relevant studies of the last 13 years (2000–2013) by searching for keywords “REMOTE SENSING” AND “BIOMASS” AND “RADAR” OR “LIDAR” OR “ALS” OR “SAR”. From the total 474 returned results, we first eliminated all studies that were either not directly dealing with biomass, or were not conducted in forest biomes. From the remaining 213 studies, 113 were finally identified as relevant since they were either using forest biomass as the target variable in a case study, or focusing on the estimation of forest biomass from remote sensing data in the form of a review. The 113 studies were reviewed with regard to reference measurements (type and number), predictor data (sensor) type and prediction method. To keep the systematically reviewed literature to a manageable number of studies, we decided to exclude studies focusing on other response variables such as growing stock volume, as for example applied by McRoberts, Gobakken, and Næsset (2012), McRoberts, Næsset, and Gobakken (2013), Steinmann, Mandallaz, Ginzler, and Lanz (2013) and Üreyen et al. (2014).

2.1. Reference measurements

Most of the reviewed studies used local reference measurements to establish a statistical relationship between biomass and remote-sensing predictors. For understanding the statistical problems that may arise in this setting, three considerations are important. Firstly, local biomass measurements are relatively labor-intensive. Therefore, the sample sizes of the reference sets that were used in the reviewed studies were often small compared to the total number of predictor variables available from remote sensing data. Typical were a few tens to a few hundreds of reference plots (Fig. 1). It should be noted though that plot sizes differed among different studies, so that the number of plots is only a proxy for the information available for model calibration. Secondly, biomass was usually not measured directly, but predicted with allometric models based on other variables such as tree diameter and tree height (e.g. Zianis, Muukkonen, Mäkipää, & Mencuccini, 2005). In some of the reviewed studies site-specific allometries were developed from individual tree measurements, other applied allometries were from the literature (e.g., Carreiras, Vasconcelos, & Lucas, 2012; Latifi, Fassnacht, & Koch, 2012; Morel et al., 2012). In both cases, significant deviations from the true biomass values can occur due to individual differences between trees as well as different site conditions. Finally, when correlating the resulting biomass values with remote sensing data, we should keep in mind that the remote sensing data provide variables such as height that are correlated with biomass, but there is no sensor that is able to directly measure biomass (Woodhouse, Mitchard, Brolly, Maniatis, & Ryan, 2012). Care should therefore be taken when extrapolating the identified correlations between local reference values and the remote sensing signal to different conditions (see Dormann et al., 2012, e.g., for a general discussion of extrapolation problems).

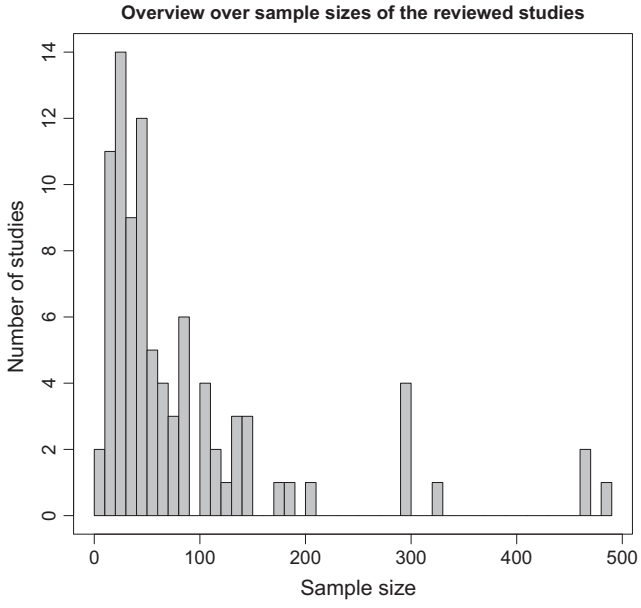

Fig. 1. Frequency distribution of the sample sizes of the reference measurements in the reviewed studies.

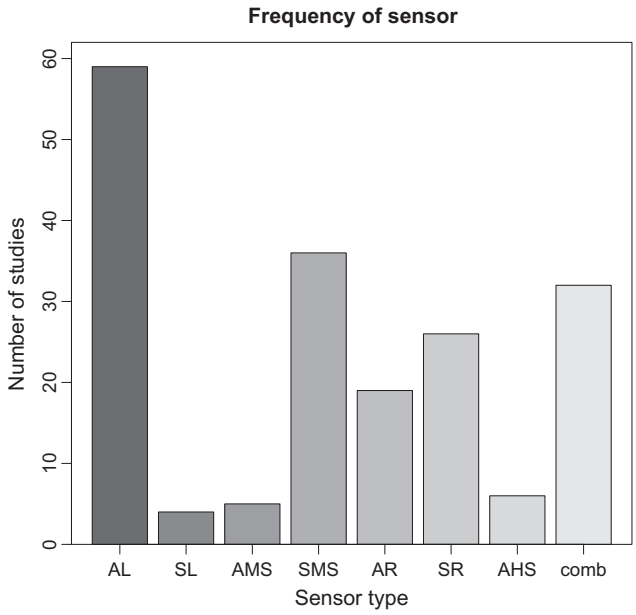

Fig. 2. Frequency distribution of the data sources (sensors) employed in the reviewed studies. (AL = airborne LiDAR, SL = spaceborne LiDAR, AMS = airborne multispectral, SMS = spaceborne multispectral, AR = airborne RADAR, SR = spaceborne RADAR, AHS = airborne hyperspectral, comb = studies using data from at least two of the aforementioned data sources.)

2.2. Data types (sensors)

The reviewed literature used a variety of remote sensing data types including optical, LiDAR, and RADAR (mostly Synthetic Aperture RADAR (SAR)) to estimate biomass. The most frequently applied sensors were discrete return airborne LiDAR, spaceborne multispectral, and airborne or spaceborne RADAR-systems (Fig. 2).

RADAR data can be obtained with a variety of wavelengths and technologies. One common option is correlating SAR backscatter intensity with biomass (Balzter et al., 2003; Santos et al., 2003). However, the signal usually saturates at low biomass values (ca. 50–100 t/ha) (Imhoff, 1995; Koch, 2010). As an alternative to SAR backscatter intensity, the interferometric and polarimetric coherence have been often employed to predict biomass (Askne & Santoro, 2005; Eriksson, Santoro, Wiesmann, & Schmullius, 2003; Tansey et al., 2004; Wagner et al., 2003). Coherence saturation levels are generally higher than those reported for backscatter intensity. Under favorable conditions, correlations exist for values of up to 250–300 t/ha (Koskinen, Pulliainen, Hyyppä, Engdahl, & Hallikainen, 2001; Santoro, Shvidenko, McCallum, Askne, & Schmullius, 2007). Provided that problems concerning saturation can be solved, RADAR data would be a very interesting source of information for biomass estimates at global scales due to its independence from clouds and therefore the possibility to obtain continuous global coverage (Kurvonen, Pulliainen, & Hallikainen, 1999; Rauste, 2005; Santoro, Beer, et al., 2007).

LiDAR systems actively emit high-frequency pulses of (laser-) light and the corresponding echoes are received by the sensor to scan the terrain for height information (Goodwin, Coops, & Culvenor, 2006; Nelson, Krabill, & Tonelli, 1988). When using pulse-form laser scanning systems in forested environments, the LiDAR signal is partially reflected by the top of the canopy (first returns), while other parts are reflected from the intermediate- and understories as well as the ground vegetation (last returns). Hence, the time interval that the LiDAR pulse needs to return to the sensor (Zimble et al., 2003) and the intensity of the reflected energy (Blair, Rabine, & Hofton, 1999; Bortolot & Wynne, 2005) are appropriate to infer forest height as well as horizontal and vertical forest structure. Height and structural information from LiDAR data is typically summarized in form of descriptive statistical attributes such as mean height, height percentiles and comparable derivatives (Tsui et al., 2012) which have been successfully applied to estimate forest biomass (e.g., Clark et al., 2011; Latifi et al., 2010; Næsset et al., 2011; Tian et al., 2012).

Reflectance of forests with dense canopy cover as measured by passive multi- and hyperspectral optical sensors mostly originates from reflections of the sunlight in the topmost part of the canopy. It has therefore been suggested that single-look optical data do not deliver detailed information on the vertical stand structure and are barely sensitive to forest biomass (Steininger, 2000). In general, one would expect passive optical data to be less sensitive to biomass than active sensors (Koch, 2010; Rahman, Csaplovics, & Koch, 2007), since the latter source is normally available in 3D form such as LiDAR-based or interferometric height models or LiDAR point clouds. Nevertheless, some studies have been using multi-spectral information to estimate forest biomass (e.g., Chopping et al., 2011; Zheng et al., 2004) with mixed results. It is assumed that the wide availability of multispectral data in combination with low or even no acquisition cost increased the number of studies using these data (exclusively or in combination with data from an active sensor). While the direct estimation of biomass with multi- and hyperspectral data is hampered by its low potential to gather structural information, optical data can unfold its full potential by estimating other forest attributes which are potentially related to forest biomass. For example accurate tree species information (e.g., Carleer & Wolff, 2004; Ghosh, Fassnacht, Joshi, & Koch, 2014) may serve as a complement to structural information estimated from an active sensor system, since tree biomass is related to the height and diameter of a tree, combined with the species-dependent wood density.

In summary, model performances and errors reported in the reviewed studies suggest a superiority of the active over the passive sensor systems when modeling forest biomass from remote sensing.

2.3. Prediction methods

2.3.1. Common modeling approaches

The reviewed studies showed that different prediction methods are applied for the estimation of forest biomass (Fig. 3). By far the largest number of studies applied different types of linear models (e.g., Morel

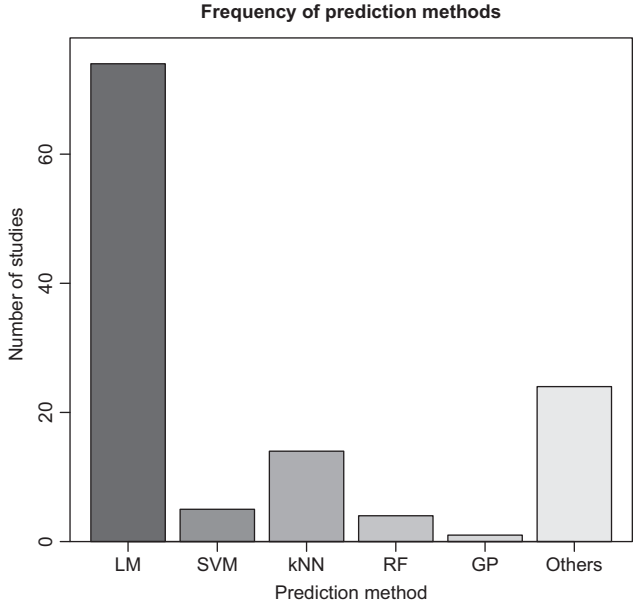

Fig. 3. Frequency distribution of the prediction methods of the reviewed studies. (lin = linear models, SVM = support vector machines, NN = nearest neighbor-based methods, RF = random forest, GP = Gaussian processes.)

et al., 2012; Straub & Koch, 2011). A second group of frequently-used models was based on nonparametric nearest neighbor approaches (e.g., Karjalainen, Kankare, Vastaranta, Holopainen, & Hyyppa, 2012; Nothdurft, Soborowski, & Breidenbach, 2009; Straub, Weinacker, & Koch, 2010). Some studies used other machine learning methods such as support vector machines (SVM) (e.g., Chen, Hay, & Zhou, 2010; Zhao, Popescu, Meng, Pang, & Agca, 2011) and random forest (RF) (e.g., Latifi & Koch, 2012; Yu, Hyyppa, Vastaranta, Holopainen, & Viitala, 2011). Finally, there was a sizable group of approaches that were often only implemented in single studies. In studies working with RADAR data, also physically-based (e.g., Wang & Qi, 2008) and semi-empirical models (e.g., Carreiras et al., 2012) have been applied.

The use of methods is related to the date of publication. In recent years, more flexible, often nonparametric methods from fields of geostatistics and machine learning have become more prevalent. Methods such as geostatistical smoothing, SVM, nearest neighbor as well as classification-and regression tree (CART)-based methods perform often better than standard linear regression models at identifying local relationships between remote sensing predictors and reference measurements based on a limited number of sample units (Gibbons & Chakraborti, 2003). This is relevant, since field measurements are often limited in studies focusing on biomass estimations.

The observed trend in modeling towards more flexible spatial models can be interpreted as an evidence of a genuine methodical advancement as well as simply a result of ongoing technical improvements in performing expensive statistical computations. In addition, one may note the shortage of prior knowledge about correlation structures for complex data (Bright, Hicke, & Hudak, 2012; Carreiras et al., 2012), which contributes to an increased application of machine learning methods. This lack of knowledge is particularly evident when combining numerous remote sensing predictors originating from differing sensors (i.e. ranging from spectral indices to predictors related to LiDAR height and intensity). However, it may also be that existing prior knowledge is available (e.g. for fully managed boreal/temperate stands) (see e.g. Breidenbach, Næsset, Lien, Gobakken, & Solberg, 2010; Karjalainen et al., 2012; Packalen & Maltamo, 2006; Packalen & Maltamo, 2007; Straub et al., 2010), but the inclusion of this prior knowledge requires new statistical approaches, for example Bayesian statistics (e.g., Hartig et al., 2012).

2.3.2. Model diagnostics

A general insight from statistics and machine learning is that more complex models will usually fit better to any given dataset. However, a more complex model is neither necessarily more likely to identify the essential mechanisms, nor does it necessarily lead to improved predictions when applied to new datasets. The reason is that going towards more complex models increases the risk of overfitting, i.e. explaining variance through a large number of redundant variables, which may lead to undesirably high variance on parameter estimates and unreliable predictions (see e.g. Hawkins, 2004). Finding the appropriate model complexity across this so-called bias-variance trade-off is a crucial topic in spatial statistics (Burnham & Anderson, 2002; Johnson & Omland, 2004).

There are multiple mechanisms to avoid overfitting. Some are based on fixed penalties for model complexity, such as the Akaike Information Criterion (AIC). Another common method is to split the data in two parts, one for model calibration, and one for evaluating the predictive error. Methods that use this approach include leave-one-out cross-validation, k-fold cross-validation, as well as repeated splits in training and validation samples (Kuhn & Johnson, 2013). Moreover, resampling methods (e.g. via jackknife or bootstrapping) are helpful by providing explicit estimates on mean and standard deviation of non-parametric estimates. Some methods mentioned in Section 2.3 already include one of these safeguards. In particular, machine learning algorithms such as RF or SVM typically split the data in training and validation sets to avoid overfitting.

To assess model performance, the most common measures in the reviewed studies by far were r^2 (correlation between observations and predictions), root mean squared error (RMSE) and mean residual deviation (many authors called this "bias"). However, one should note that not all the reviewed literature included cross-validations, or similar analyses of predictive error and mean residual deviation. Thus, some studies report model performance measures for the data that were used to fit the model, while others report it for independent validation data, for which generally a larger error is to be expected.

3. Case study methods

Our literature review revealed a high diversity of sensor types and methods used for predicting forest biomass. So far, however, only few studies have conducted a comparative analysis of those methodological differences (e.g., Garcia-Gutierrez, Gonzalez-Ferreiro, Mateos-Garcia, Riquelme-Santos, & Miranda, 2011; Gleason & Im, 2012; McRoberts et al., 2013; Tian et al., 2012). A possible effect of sample size or the applied validation procedure on the model performance has to the best of our knowledge not yet been addressed. To address those issues, we conducted two case studies, in which we examine the effect of reference data (number of reference sample units), the remote sensing datasets (predictor data (sensor) type), the prediction method and the validation procedure. We varied all these factors in a factorial experiment to assess their impact on the selected model diagnostics. The different options included for each factor were based on our literature review (prediction methods), with some constraints imposed by the available datasets (predictor data types and sample sizes).

3.1. Study areas

3.1.1. Karlsruhe, Germany

The study site in Karlsruhe (8° 25′ 00″ E and 49° 02′ 20″ N) covers an area of about 900 ha managed forest in the southwestern German federal state of Baden-Württemberg. The site consists of stands of native coniferous and deciduous species spreading along a topographically-gentle terrain. The dominating tree species include Scots Pine (*Pinus sylvestris*, L), European Beech (*Fagus sylvatica* L.), Sessile Oak (*Quercus petraea* Liebl.) and Pedunculate Oak (*Quercus robur* L.). Other species (*Picea* sp., *Abies* sp., and *Pseudotsuga menziesii*) occasionally occur in the stands. Stand ages range from approximately 30 to 130 years. Young stands are often composed of pure pine or pure oak trees and tend to be dense. In the old stands, pine is normally the dominating species. The density of the old stands varies from stand to stand but is often rather low, particularly in the very old stands. In many cases, there is a second tree layer consisting of Beech and Hornbeam below the old Pines.

3.1.2. Monte Oscuro, Chile

The second test site is located in the Maule region of central Chile (35° 7′ 0″ S, 70° 55′ 26″ E) and covers an area of approximately 1260 ha. Forest vegetation is dominated by roble beech (*Nothofagus oblique* (Mirb.) Oerst.). Other species include Mañío de hojas largas (*Podocarpus salignus* D. Don), Naranjillo (*Citronella mucronata* (Ruiz & Pav.) D. Don), Piñol (*Lomatia dentata* (Ruiz et 176 Pavón) R. Br.), Peumo (*Cryptocarya alba* (Mol.) Looser) and Olivillo (*Aextoxicon punctatum* Ruiz et Pavón.) which occur in smaller numbers on the site. The area is in a quasi-natural state and can be considered as a second growth native forest with a very low management impact. The terrain is rough and ranges from altitudes of 700 m to 1400 m above sea level. Most parts of the area consist of naturally regrowing secondary forest that reestablished itself after harvesting activities for construction wood in the 1950s (large impact) and 1990s (only 30 ha was affected). The intensity of the harvesting activities depended on the location within the site and resulted in an unsystematic pattern of stands with differing ages. Furthermore, individual trees in inaccessible areas might reach

ages of up to several hundred years. The vegetation density in undisturbed regrowing areas is generally very high, due to the lack of management activities and shows very pronounced horizontal and vertical structural diversity.

3.2. Field measurements and remote sensing datasets

3.2.1. Karlsruhe, Germany

Aboveground biomass values for 297 inventory plots were collected in 2006. The inventory consisted of concentric circular plots in a 200 m × 100 m sampling grid. Trees with DBH >7 cm, 10 cm, 15 cm and 30 cm were measured in rings with 2 m, 3 m, 6 m and the (full) 12 m radii of each plot, respectively. Individual tree biomass values were obtained by applying species-specific allometric models (Zell, 2008). Subsequently, expansion factors (e.g., expansion factor ring 1 = 1 ha/area of ring 1) were applied to the individual tree biomass values depending on the DBH. The expanded biomass values were finally summed up to obtain total biomass values in tons per hectare (details see Latifi et al., 2012).

The remote sensing dataset used here is described in detail in Latifi et al. (2012). The dataset includes metrics extracted from pre-processed, full waveform LiDAR data collected by Toposys GmbH with a Harrier56 system and a Riegl LMS-Q560 laser scanner launched on a helicopter. The fullwave data were transformed into a multi pulse-form point cloud, from which only the first pulse data were utilized in the study. The point cloud with an average point density of 16 points/m^2 was normalized using a Digital Terrain Model (DTM) as calculated in the TreesVis software (Weinacker, Koch, & Weinacker, 2004).

As a second data source, a hyperspectral scene from the airborne HyMap sensor was used. The scene was acquired during the HyEurope campaign of the German Aerospace Center (DLR) in August 2009 under nearly perfect weather conditions. The image was processed in a standard processing chain of the DLR to correct for atmospheric and terrain effects. The resulting image had a pixel size of 4 m. The HyMap sensor produces images with 125 spectral channels featuring a spectral resolution of 13 to 17 nm between 0.45 and 2.48 μm of the wavelength range (Cocks, Jenssen, Stewart, Wilson, & Shields, 1998). For further technical details of the LiDAR and hyperspectral data we refer to Latifi et al. (2012).

3.2.2. Monte Oscuro, Chile

Data on the Chilean study site were collected in early 2011. The sampling design used a systematic sampling grid with 200 m × 200 m grid size containing 150 clusters, consisting of five sub-plots with 8 m radius each (one central plot and four surrounding plots located 30 m apart from the central plot in the four cardinal directions). Trees with DBH >5 cm, 10 cm and 20 cm were measured in rings with 2 m, 4 m and the (full) 8 m radii of the sub-plot, respectively. Tree height was measured for selected trees within each cluster. Missing heights were estimated from an allometric model relating DBH to tree height (Eq. (S1) in Supplementary data 1). Uncertainties in these model based height predictions were ignored and were assumed to be observations without errors.

Single tree biomass values were obtained by a second allometric model (Eq. (S2) in Supplementary data 1) developed for roble beech by Gayoso, Guerra, and Alarcón (2002). This model was applied to all tree species. A correction factor for the tree density was applied for tree species other than roble (Eq. (S3) in Supplementary data 1).

Following the estimation of the single tree biomass values, plot-specific values in tons per hectare were calculated using expansion factors for the three defined rings with differing areas and then summing up the biomass values of the individual trees. The biomass values of the 5 sub-plots of each cluster were aggregated into a single mean reference value, which was used to calibrate the remote-sensing models. The variability of biomass among the individual sub-plots was not considered explicitly. The rationale was that the mean value of a cluster is expected to correspond well to the available 30 m pixel size of Hyperion data, since it reduces the effects of positioning errors raised from the rough terrain. Furthermore, using the clusters instead of the individual sub-plots was assumed to marginally smooth the substantial structural diversity within the study area.

Table 1
Summary of the biomass values of test sites Karlsruhe and Monte Oscuro. All values are in tons per hectare (t/ha).

Test site	Minimum	1st Quant.	Median	Mean	3rd Quant.	Maximum	N samples
Karlsruhe	9.02	114.00	165.70	167.80	216.40	372.90	297
Monte Oscuro	11.61	86.72	115.40	123.20	151.90	296.50	150

A summary of the biomass values of test sites in Karlsruhe and Monte Oscuro is provided in Table 1. The distributions of biomass values for Karlsruhe and Monte Oscuro are illustrated in Figs. S1 and S2 of Supplementary data 2.

In Monte Oscuro, a laser scanner identical to the system used in Karlsruhe was applied to collect LiDAR data from a Piper PA-24 Comanche in February 2011. The LiDAR data had an average point density of 4.6 points/m^2. LiDAR data processing was identical to the procedure described for test site Karlsruhe.

The hyperspectral data originate from the Hyperion sensor on board of the EO-1 platform. The image was acquired during the peak of the vegetation period on February 25, 2011 under cloud-free conditions. Noisy bands were manually removed (see Datt, McVicar, Van Niel, Jupp, & Pearlman, 2003). The Hyperion tools 2.0 plug-in for the ENVI software package (White, 2011) was used to prepare the dataset for subsequent atmospheric correction which was conducted with the "Fast Line of sight Atmospheric Analysis of Spectral Hypercubes" (FLAASH) implementation in ENVI (ITT, Visual Information Solutions, 2009). As a final processing step, the Hyperion scene was georectified to an ortho-photograph which had been acquired simultaneous to the recorded LiDAR data. The RMSE obtained by the georectification using evenly distributed pass-points and a second-order polynomial was 0.9 pixels (equivalent to 27 m).

3.3. Feature space set-up

We selected predictors based on previous experiences across these and other test sites as well as based on the conducted literature review. An alternative would have been to select predictors with an automated feature selection algorithm (e.g. Latifi et al., 2012; Tomppo & Halme, 2004), which was not performed to avoid the possible confounding of the analysis with potential uncertainties of an additional statistical algorithm.

For LiDAR data, we selected mean height, maximum height, 10th, 70th, and 90th height quantiles of the first-pulse points as predictor variables. These predictors were selected based on the literature review (e.g., Tonolli et al., 2011; Tsui et al., 2012) as well as earlier experiences, partly with a similar dataset (e.g., Latifi et al., 2012).

For the hyperspectral predictors, we hypothesized that the association between optical data and biomass would originate mostly from species information, vegetation density and leaf water content. Therefore, we selected 8 hyperspectral predictors that have been linked to those three factors in earlier studies. The 8 predictors consisted of three vegetation indices (VI) including normalized difference vegetation index (NDVI, as a measure of vegetation density), the normalized difference water index (NDWI, as a measure of leaf water content) and an additional Chlorophyll-VI proposed by Gitelson, Gritz, and Merzlyak (2003) (as a measure of vegetation density), as well as of five heuristically-selected original hyperspectral bands, located at 518 nm, 681 nm, 1235 nm, 1477 nm and 2032 nm (in the case of the Hyperion dataset). These bands have been reported to carry important information on forest biomass (Latifi et al., 2012; Thenkabail, Enclona,

Ashton, Legg, & Jean De Dieu, 2004) and species information (Fassnacht et al., 2014; Thenkabail, Enclona, Ashton, & Van Der Meer, 2004). For the HyMap sensor, the corresponding closest bands were selected

3.4. Biomass modeling

The following gives an overview over the applied modeling set-up. The set-up consisted of 5 main processing steps:

I) After preparing an initial matrix of response and predictor variables, the dataset of size n was ordered according to ascending biomass values and was subdivided in five equal-size subgroups of size ns = n/5 (analogous to splitting the dataset into 20% percentiles).
II) From the resulting stratified dataset, we created subsamples for each stratum via bootstrapping (Efron & Tibshirani, 1994). That is we drew 500 datasets with replacement for each of the 4 desired sample sizes per each of the 3 sensor types. The bootstrapping procedure incorporates effects of sampling variability which is a requirement for the ANOVA analysis described in step V. The stratification ensures that samples from the full range of available biomass values were included in each bootstrapped input dataset. To generate the 4 desired sample size classes the number of sample units x drawn from each of the five subgroups was varied four times (x = ns/4, x = ns/3, x = ns/2, x = ns). Input datasets of class 1 (x = ns/4) contained the fewest number of sample units, while those of class 4 (x = ns) contained the most.
III) For each combination of sample size and sensor, the 500 input datasets were fit by 5 prediction methods including k-nearest neighbor (KNN), SVM, Gaussian processes (GP), RF and stepwise linear regression (LMSTEP). More details on the prediction methods are provided in Supplementary data 3.
IV) For each prediction method as well as each input dataset, a cross-validation was applied which resulted in 1) obtaining the best model parameters (indicated by the model with lowest RMSE) and 2) retrieving diagnostics of RMSE and r^2. We applied the cross-validation with three different settings: 10-fold, 5-fold and 3-fold, each with 5 repetitions.
V) Subsequently, we used an ANOVA to estimate the contribution of each of the varied factors (sensor, sample size, method and cross-validation setting) and their corresponding interactions to describe the variance reflected in the computed model diagnostics (r^2 and RMSE) of the model runs.

The results from the runs that applied the 5-fold cross-validation with 5 repetitions will be presented in Section 4.2, while the effects of a varying number of folds is included in the ANOVA results of Section 4.1. A link to the R-script containing our source code is enclosed in Supplementary data 4. In addition to the model diagnostics, we show spatial biomass predictions and their coefficient of variation (CV) for the entire study sites. For the maps, we calculated the predictors described in Section 3.3 on a grid over the study area with cell sizes corresponding to the field plots described in Section 3.2.

4. Results

4.1. ANOVA

In Karlsruhe, the highest ANOVA sum of squares values (ssv) (which indicate an important contribution to the explained variance of r^2) were reported for predictor data type (968.6) and prediction method (473.0) as well as for their interaction (247.6) (Table 2). The number of sample units seems to be less important as indicated by the smaller ssv (116.1). ANOVA results for RMSE as dependent variable show the same pattern, with predictor data type and prediction method as well as their interaction having the highest ssv (2,182,360, 1,203,330 and 566,076, respectively, see Table 2). For Karlsruhe, the number of folds in the k-fold cross-validation settings did not have a notable effect on the model diagnostics (ssv of 71.8 and 82,518 for r^2 and RMSE, respectively).

Table 2
Results of ANOVA conducted to explain the variance of r^2 and RMSE as obtained for the different experiments on test site Karlsruhe (KA) and Monte Oscuro (MO).

Test site		KA		MO	
Response variable		RMSE	r^2	RMSE	r^2
	Df	SumSq	SumSq	SumSq	SumSq
PredMeth	4	1,203,330	473.0	977,606	409.7
NumSamp	3	465,024	116.1	348,625	62.1
PredData	2	2,182,360	968.6	1,150,474	648.7
Folds	2	82,518	71.8	190,116	445.9
PredMeth:NumSamp	12	155,898	64.3	228,013	117.9
PredMeth:PredData	8	566,076	247.6	294,817	136.9
PredMeth:folds	8	7341	1.9	26,868	1.6
NumSamp:PredData	6	7590	2.6	34,072	30.2
NumSamp:folds	6	652	5.8	22,539	108.7
InData:folds	4	506	2.7	5978	9.9
PredMeth:NumSamp: PredData	24	19,691	12.9	93,467	16.5
PredMeth:NumSamp:folds	24	4006	0.8	44,785	2.8
PredMeth:PredData:folds	16	1144	1.2	21,931	1.8
NumSamp:PredData:folds	12	119	0.3	7575	1.5
PredMeth:NumSamp: PredData:folds	48	173	0.1	31,675	0.7
Residuals	89,820	1,237,038	408.5	17,211,248	829.9

PredMeth = prediction method, NumSamp = number of input sample units, PredData = predictor data (sensor) type/sensor, folds = number of folds in the k-fold cross-validation.

Regarding variance in r^2, Monte Oscuro showed a largely identical trend as shown in Karlsruhe (Table 2). Predictor data type and prediction method showed high ssv (648.7 and 409.7, respectively), while the number of sample units reached only low ssv (ssv = 62.1). A striking difference was the large effect of the number of folds (ssv = 445.9) being the second most important factor for explaining the variance in r^2 on the test site in Monte Oscuro.

The results for RMSE for Monte Oscuro (Table 2) follow the patterns of the abovementioned results for Karlsruhe. Largest contribution to the explained variance was returned for the predictor data type (ssv = 1,150,474) followed by the prediction method (ssv = 977,606) and the number of sample units (ssv = 348,625). The number of folds were not of great importance to explain variance in RMSE (ssv = 190,116).

4.2. Model performances

Figs. 4 and 5 summarize the model performances obtained for all model runs conducted on the study areas Karlsruhe (results for Monte Oscuro are provided in Supplementary data 5). In Karlsruhe, the presence of LiDAR metrics generally led to smaller RMSE and higher r^2 values compared to the hyperspectral-only results. An exception was the kNN model applied on the combined LiDAR and hyperspectral dataset, where RMSE values increased and r^2 values decreased compared to the sole LiDAR models. For all other prediction methods, only marginal differences in RMSE and r^2 were observed when running models on either the sole LiDAR or the combined LiDAR and hyperspectral datasets. Concerning the number of sample units a relatively stable increase in accuracy and decrease in RMSE were observed along with increasing sample size (colored horizontal lines from left to right). Exceptions were the r^2 values of SVM and KNN applied on hyperspectral metrics, where a slight decrease in r^2 was observed when switching from class 1 with smallest number of sample units to class 2.

Comparisons across the five prediction methods indicated that RF outperformed all other tested methods, especially when many sample

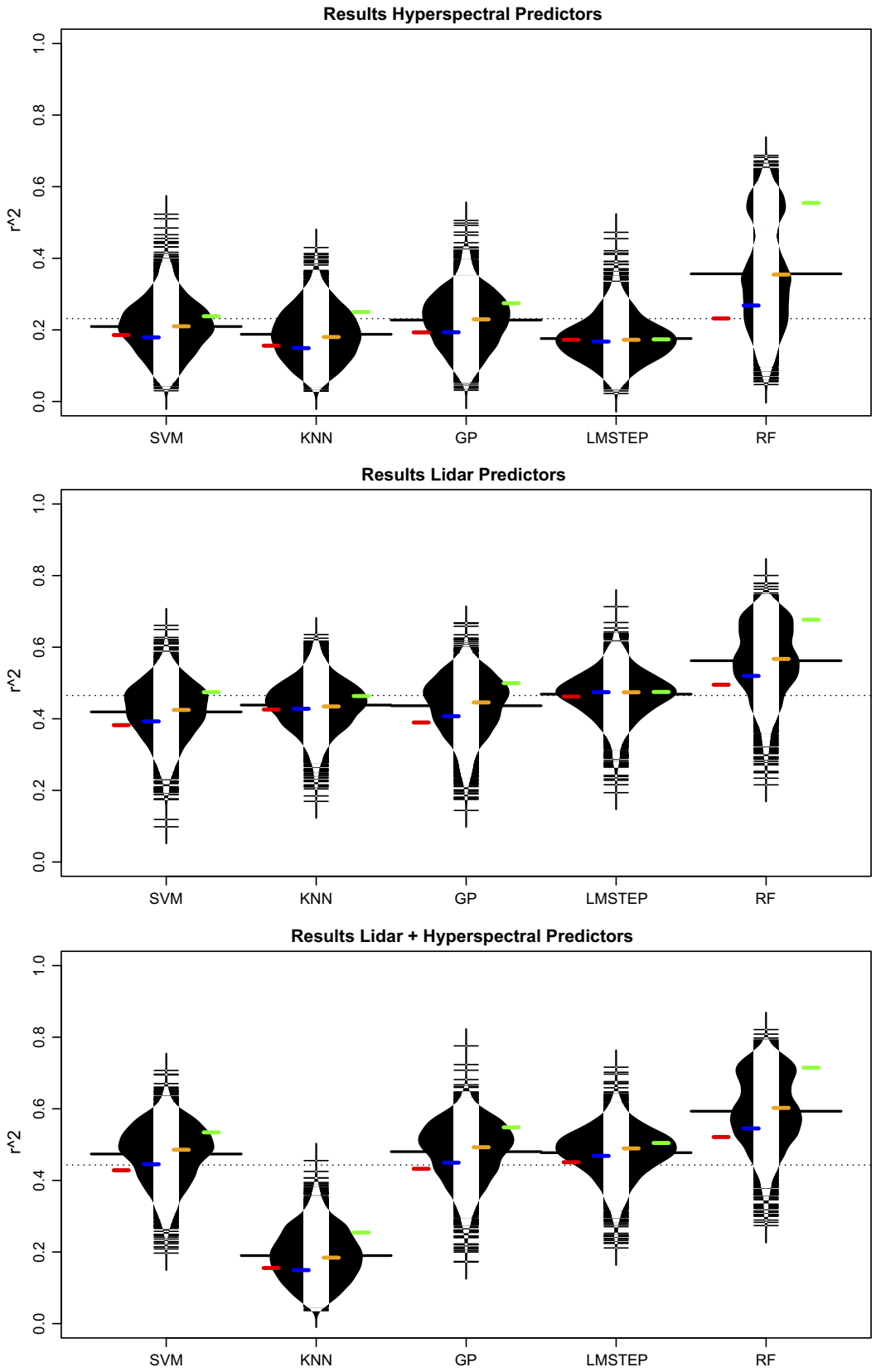

Fig. 4. Results for test site Karlsruhe. The beanplots illustrate the distribution of the mean r^2 values from the 500 bootstrapped models as obtained by the 5-fold-cross validation for each prediction method (LMSTEP = stepwise linear models, SVM = support vector machines, KNN = k-nearest neighbor, RF = random forest, GP = Gaussian processes) and predictor data (sensor) type. Furthermore, the median values of the corresponding accuracy measures for each of the four sample size classes are given with the colored horizontal stripes (class 1 to class 4 from left to right in colors red, blue, yellow and green).

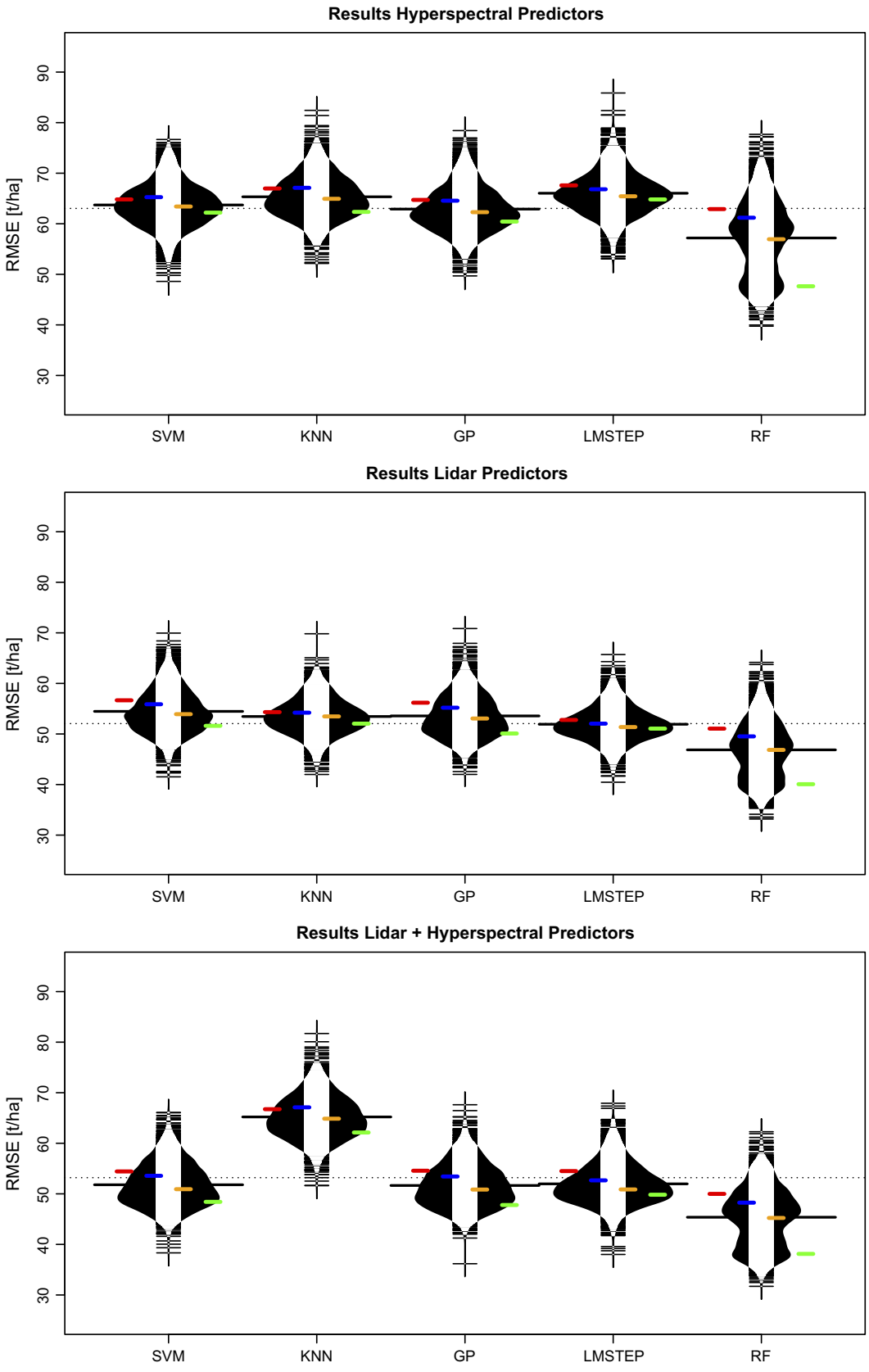

Fig. 5. Results for study area Karlsruhe. The beanplots illustrate the distribution of the mean RMSE values from the 500 bootstrapped models. Explanations follow those from Fig. 4.

units were available. The absolute differences in lowest RMSE and highest r^2 between RF and the next best model for the LiDAR models reached from approximately 2 t/ha and 0.03 (class 1, next best model LMSTEP) to about 10 t/ha and 0.175 (class 4, next best model GP). The absolute differences in RMSE and r^2 between RF and the worst models (LMSTEP and KNN) reached values of 17 t/ha and 0.21. A downside is

that RF showed highest variances in the obtained accuracy measures, which were particularly large when solely applying the hyperspectral data. LMSTEP, in most cases, turned out slightly higher RMSE and lower r^2 values as compared to all other methods. Results indicated symmetric distributions for all accuracy diagnostics for the entire methods for the 500 bootstrap runs. However, RF was an exception, in which more or less bi-modal distributions were observed in all presented cases (Figs. 4, 5).

The results for Monte Oscuro showed similar qualitative patterns as Karlsruhe. Models incorporating LiDAR data produced greater accuracies than those built solely based on hyperspectral data. Similarly, the differences between model diagnostics of LiDAR-only and combined LiDAR/hyperspectral data were marginal. For models based on the combined LiDAR and hyperspectral data, the notable low performance of KNN was again observed. The beanplots (provided in Supplementary data 5) show symmetric distributions of the RMSE values for all prediction methods, while depicting positively skewed distributions of r^2 values for RF.

Concerning the varying number of sample units, notable dissimilarities occurred in Monte Oscuro compared to Karlsruhe. The r^2 values were mostly high when the predictor dataset consisted of fewer sample units. In the case of the hyperspectral dataset, the values then decreased along with increasing number of sample units. When LiDAR data were available, the r^2 values in most cases decreased when going from class 1 to class 2 and started to increase again. For RMSE this trend was virtually not existent.

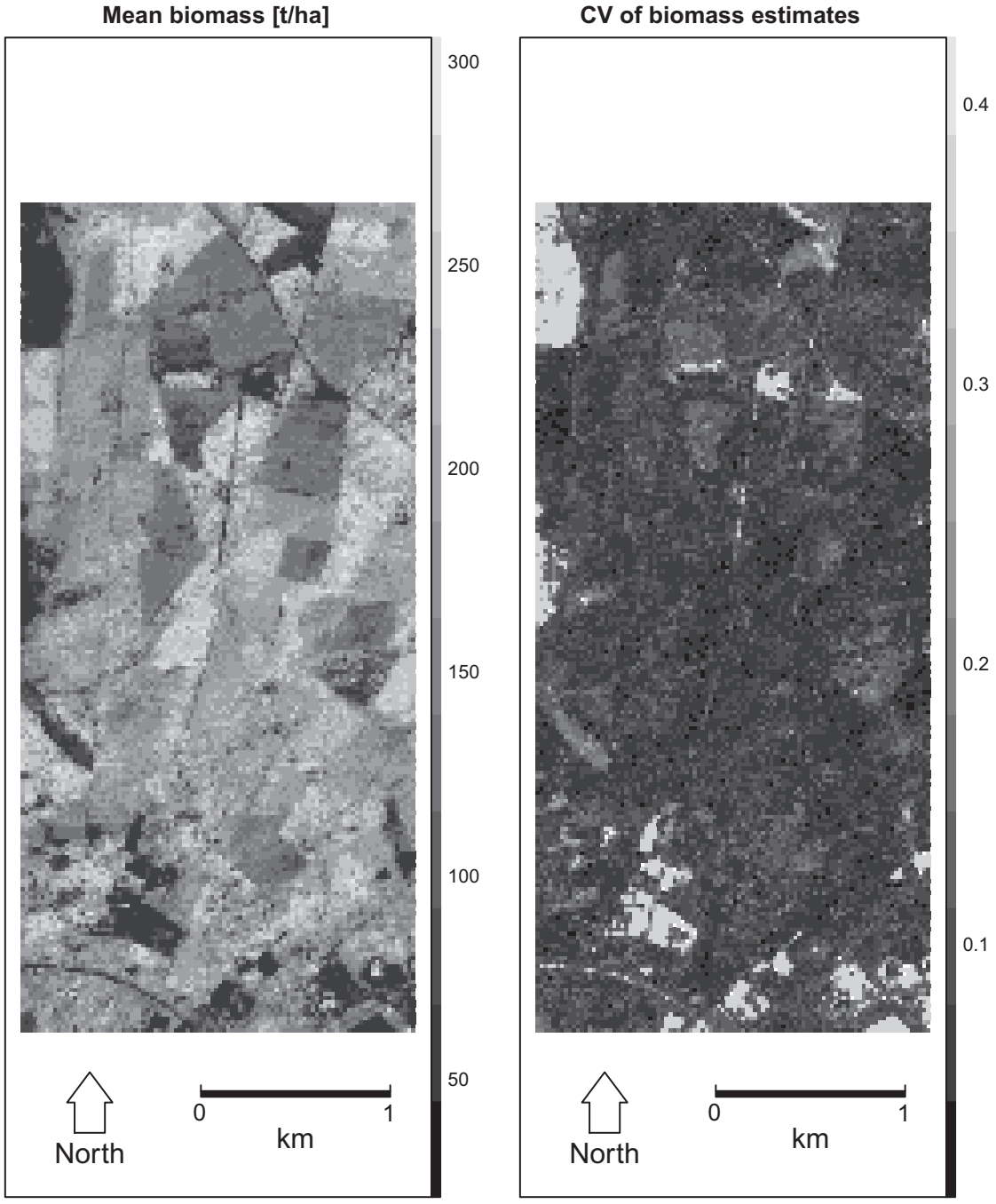

Fig. 6. Wall-to-wall map of mean biomass estimates for test site Karlsruhe. Left map shows mean biomass predictions as obtained from the 500 bootstrapped model runs, using LiDAR data, random forest and largest sample size (class 4). Additionally, the coefficient of variation (CV) of the biomass estimates is given on the right.

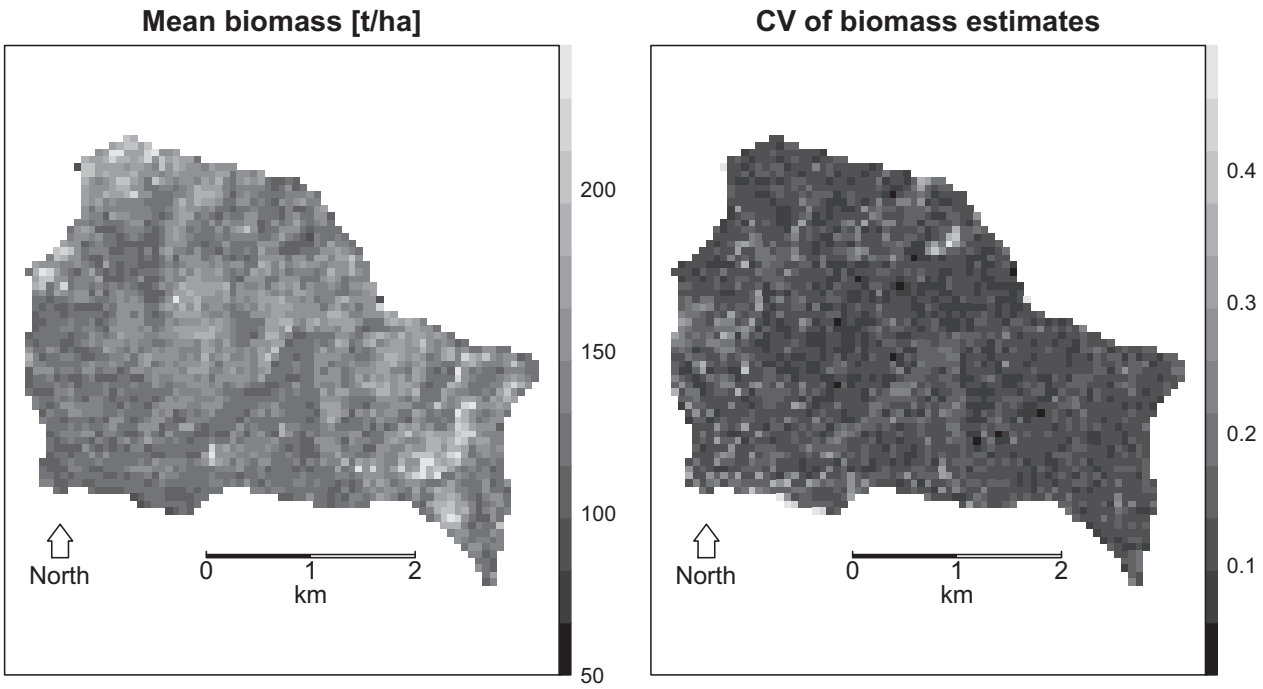

Fig. 7. Wall-to-wall map of mean biomass estimates for test site Monte Oscuro. Explanations follow those from Fig. 6.

All in all, the observed variances of the model diagnostics were higher and the RMSE values smaller in Monte Oscuro when compared to Karlsruhe. Selected residual plots for the study area Karlsruhe are provided in Supplementary data 6.

4.3. Wall-to-wall predictions

Fig. 6 and 7 illustrate the exemplified wall-to-wall predictions for test sites Karlsruhe and Monte Oscuro as obtained from the LiDAR runs with the highest number of sample units and RF model.

For test site Karlsruhe, the mean biomass estimates often show homogenous patterns within stand borders, which is plausible due to the typically identical management within those stands. The estimates are reasonably well spread over the original range of reference biomass values (Table 1), although extreme biomass values are less frequently observed. This is in agreement with the distribution of the residuals displayed and discussed in the Supplementary data 6. Nonforested areas can be identified as homogeneous areas of low predicted biomass estimates. In the forested areas CV of the biomass estimates indicates reasonably low values.

For Monte Oscuro, large areas show medium predicted biomass values, with only two regions in the Northwest and Southeast reaching higher biomass estimates. This agrees fairly well with the known situation on the site. The two mentioned areas showing high biomass estimates are located on relatively hard accessible high altitudes and have therefore been less affected by wood exploitation. As in Karlsruhe, extreme values are less frequent.

Further maps of biomass estimates as obtained from other prediction methods, sample sizes and predictor data type are provided in the Supplementary data 8.

5. Discussion

Our literature review revealed a large diversity in the methods that are used to estimate forest biomass from remote sensing data, with no clear agreement on what methods perform best. We therefore set up two case studies, in which we analyzed the effect of sample size (4 cases), predictor data (sensor) type (3 cases), prediction method (5 cases) and cross-validation (3 cases) on the accuracy of forest biomass estimates from LiDAR and hyperspectral data. We generated resampled datasets by stratified bootstrapping, and used ANOVA to rank the importance of the four factors on the accuracy of the biomass predictions, measured by cross-validated RMSE and r^2.

Biomass estimates of Monte Oscuro had generally lower RMSE values than Karlsruhe but showed a slightly increased variance. We conjecture that the greater structural complexity of the forest in the unmanaged Monte Oscuro stands hampers finding reliable remote sensing predictors and therefore increased the variance of the RMSE. The generally lower RMSE values are most probably due to the smaller range of reference biomass values in Monte Oscuro compared to Karlsruhe (Table 1). As RMSE is an absolute measure of the deviation between model and data, larger ranges of the reference values create the potential for larger deviances when the model fit is unsatisfactory.

Our ANOVA results indicate that the predictor data type is the most important factor for the accuracy of biomass predictions. This is in line with other studies (e.g., Clark et al., 2011; Latifi et al., 2012) as well as a recent meta-analysis (Zolkos, Goetz, & Dubayah, 2013), which reported that LiDAR has a notably higher information content than optical data for modeling forest biomass. The sole use of hyperspectral data resulted in relatively low model performances at both test sites. This is in accordance with earlier findings (e.g., Cháidez, 2009; Clark et al., 2011; Latifi et al., 2012; Laurin et al., 2014). The combination of hyperspectral and LiDAR information did not improve results compared to the LiDAR-only models. This is in line with earlier findings (e.g., Clark et al., 2011; Hyde, Nelson, Kimes, & Levine, 2007; Nelson et al., 2007), although other studies reported more positive results (e.g., Anderson et al., 2008; Tsui et al., 2012). We stress that our results are based on forest data. Optical data may be more suitable to predict vegetation biomass in other biomes such as grasslands.

The prediction method turned out to be the second most important factor for the accuracy of biomass prediction in most cases. Although also earlier studies reported notable differences in the performance of prediction methods (Garcia-Gutierrez et al., 2011; Gleason & Im,

2012; Latifi et al., 2010), we were somewhat surprised that the effect was so pronounced.

Among the considered methods, random forest (RF) outperformed the other tested approaches, especially when many sample units were used. We attribute the good performance, that was also found in other studies (e.g., Garcia-Gutierrez et al., 2011; Latifi et al., 2010), to the flexibility of the RF approach, which notably differs from all other tested methods due to its conceptual design. A problem with RF may be that the applied subsampling in the algorithm may result in considerable variance of the estimates when applied to a small number of sample units (compare results regarding sample size; see also Latifi et al., 2012).

Stepwise linear regression (LMSTEP) in most cases performed worse than other tested models, particularly when using hyperspectral data. We explain this by the fact that relationships between hyperspectral predictors and observed biomass are likely nonlinear and therefore not well modeled by LMSTEP. In contrast, LiDAR metrics (e.g. mean tree height) can be expected to show a more linear association to biomass due to the relationship between height and volume. This agrees with our results in which LMSTEP performed similar to most other tested prediction methods when LiDAR data were applied.

A striking observation concerning the prediction methods was the substantially worse performances of KNN for the combined LiDAR and hyperspectral datasets. We assume that this is related to the lack of an effective procedure to weight the prediction strength of the predictors in our implementation of KNN.

Somewhat surprisingly, the number of sample units was of lower importance than the prediction method for explaining the variance in r^2 and RMSE on both test sites. Nevertheless, Figs. 4 and 5 show that a practically relevant effect on r^2 and RMSE can be attributed to sample size (compare colored stripes), especially for RF. Furthermore, the importance was higher for Monte Oscuro compared to Karlsruhe. We explain this by the generally-smaller available sample size for the Chilean test site, and the observation that model performance reacts particularly sensitive when the sample size falls below a minimum number of samples.

Small sample size is also related to a further issue. For Monte Oscuro, the number of folds was found to have a notable effect on the explained variance of r^2. For example, for model runs with a 10-fold cross-validation with 5 repetitions (results in Supplementary data 7), we observed that the models with smallest sample size (class 1) produced very high r^2 values on test site Monte Oscuro, which obviously did not match the predictive power of the models. Further analysis revealed that the very small sample size of the hold-out sample (3–4 samples) during the cross-validation caused this result. Reducing the number of folds in the cross-validation settings mitigated this effect, as it is also apparent in the eventually-presented results. We conclude that r^2 values (calculated as Pearson's correlation coefficient between observed and predicted samples) can be strongly influenced by the applied number of folds within the cross-validation when the sample size is small. In Karlsruhe, where the samples size was generally bigger, the influence of the cross-validation settings was marginal. RMSE values were not notably influenced by the validation settings on both test sites, which suggest that RMSE is more robust to this problem. The discussed issue is interesting because a substantial number of the reviewed studies used a very limited sample size (Fig. 1) and therefore, might have been affected by similar problems. Zolkos et al. (2013) conducted a meta-analysis on LiDAR-based estimation of forest biomass and found a negative relationship between model errors and the size of inventory plots. Although other explanations are possible, one reason may simply be that larger inventory plots are often correlated with smaller sample sizes due to the increased workload. Studies with larger plots may therefore face an increased risk to suffer from statistical artifacts due to small sample size.

Regarding the generality of our results, it should be noted that our ranking of the five prediction methods was based on their r^2 and RMSE values. It is known that methods that perform well considering those criteria may have weaknesses when considering other aspects of the model-data fit. Powell et al. (2010), for example, compared RF and two other prediction methods (reduced major axis regression, gradient nearest neighbor imputation) for predicting field biomass values from Landsat time series. Similar as in the present study, RF outperformed the other tested methods in terms of lowest RMSE values; however, the results of the other methods outperformed RF in maintaining the observed variance of the reference biomass values which, depending on the application, can be a major criteria for model selection.

Another consideration regarding the generality of the ranking is that we selected relatively simple algorithms as representations of regression (LMSTEP) and nearest neighbor approaches (KNN with an unweighted Euclidean distance as distance metric). The use of non-linear regression (as e.g. applied in Næsset et al. 2013; McRoberts et al., 2013), as well as the further optimization of the KNN approach by applying efficient weighting procedures of the predictor variables (e.g., with a genetic algorithm or an optimized distance metric) could change the ranking of these methods compared to RF.

Concerning the ANOVA results, it has to be considered that the amount of residual variance in the ANOVA originates from the variance in r^2 and RMSE that is created by the bootstrapping. Because of potential larger-scale environmental effects, and also because of potential observation uncertainties, we would expect that the within-sample variance used by the bootstrap tends to be lower than the out-of-sample variance that would originate from repeating this analysis over multiple study areas. It can therefore be expected that the reported ANOVA uncertainties are slightly too optimistic, while the estimated effects of the different factors should be unbiased by this consideration.

We also concede that our reports have to be interpreted cautiously, as they are limited to two sites as well as to two remote sensing sources of information and a limited number of prediction methods. We think that, among those three, the choice of sites is least likely to substantially change our conclusions: although the two test sites differ notably in terms of tree species, topography and forest structure, the observed trends were still relatively stable. Extending our research to more diverse datasets, potentially also integrating RADAR information, would be of high interest. Further studies could also investigate whether other prediction methods, such as nonlinear regression or optimized nearest neighbor approaches alter our findings. In addition, the methodology used here could be improved in the future by attempting to correct prediction bias, as well as by exploring the uncertainty on larger area estimates (see hints given by Næsset et al. (2011) and McRoberts et al. (2013)).

Finally, there is the possibility of improving model estimates in other ways than improving the statistical algorithm. Examples of this would be methods to select the predictor variables; incorporating additional environmental covariates such as topography, soil or climate information; or creating separate estimates for different forest types.

6. Conclusions

Our study showed that the prediction method had a considerable impact on the accuracy of the biomass estimates, nearly equally important as data type, and more important than sample size. This indicates that collecting more reference data is not necessarily the most effective option for improving the accuracy of biomass estimates. Yet, our results also indicate the need for a minimum number of reference samples (per remote sensing dataset) and a sound selection of the validation methods to avoid overly optimistic r^2 values. In addition, the overall best algorithm, RF, benefited strongly from a larger sample size.

Within the limitations of the applied datasets and tested prediction methods, our recommendation for minimizing predictive error of forest biomass estimates is therefore to use LiDAR data with a preferably large number of reference samples in combination with RF. Moreover, when reporting model performance, we do not recommend using the

correlation between predictions and observations as the sole indicator of performance, as we saw considerable dependence of Pearson's r^2 on the selection of sample units and the sample size. RMSE was more stable, but does not consider the variability within the data that the model has to predict. Therefore, multiple performance measures, at least RMSE and one measure of correlation, should be ideally reported.

For the future, expanding our results to a larger number of datasets (more test sites, more sensor systems and especially the inclusion of Radar information) would further increase our understanding of the role of the statistical model set-up in the estimation of forest biomass from remote sensing. We invite other researchers to repeat our analysis on new datasets, using the code we provide. The integration of additional predictors (e.g., topographic information or the pre-stratification into forest types) would be a further possible extension of our work.

Acknowledgment

This study was partly funded by the German National Space Agency DLR (Deutsches Zentrum für Luft- und Raumfahrt e.V.) on behalf of the German Federal Ministry of Economy and Technology (grant numbers: 50EE1025 and 50EE1265). Further financial support was received from BioComsa, Consorcio Tecnológico de Biocombustibles S.A., Chile. The authors would like to acknowledge the valuable suggestions of Prof. Dr. Carsten Dormann concerning the applied methodology and the visualization of the results. Furthermore, we would like to thank Stuart Frye from NASA for the friendly support in obtaining the EO-1 Hyperion data. Last but not least, we acknowledge three anonymous reviewers, who made a large effort to improve earlier versions of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.rse.2014.07.028. The R-codes used in this study can be accessed at: https://github.com/fabianfassnacht/biomass.

References

Anderson, J. E., Plourde, L. C., Martin, M. E., Braswell, B. H., Smith, M. L., Dubayah, R. O., et al. (2008). Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment*, *112*(4), 1856–1870.

Askne, J., & Santoro, M. (2005). Multitemporal repeat pass SAR interferometry of boreal forests. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(6), 1219–1228.

Balzter, H., Rowland, C., Milne, R., Stebler, O., Patenaude, G., Dawson, T., & Saich, P. (2003). Potential of polarimetric SAR interferometry for forest carbon accounting. *Proceedings of Geoscience and Remote Sensing Symposium, IGARSS '03, 21–25 July, Toulouse, France. 3.* (pp. 1945–1947).

Blair, J. B., Rabine, D. L., & Hofton, M. A. (1999). The Laser Vegetation Imaging Sensor (LVIS): a medium-altitude, digitation-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing*, *54*, 115–122.

Bortolot, Z. J., & Wynne, R. H. (2005). Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *59*, 342–360.

Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., & Solberg, S. (2010). Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sensing of Environment*, *114*, 911–924.

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 Sensing of Environment*, *124*, 270–281.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach.* Springer (488 pp.).

Carleer, A., & Wolff, E. (2004). Exploitation of very high resolution satellite data for tree species identification. *Photogrammetric Engineering & Remote Sensing*, *70*(1), 135–140.

Carreiras, J. M. B., Vasconcelos, M. J., & Lucas, R. M. (2012). Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa). *Remote Sensing of Environment*, *121*, 426–442.

Cháidez, J. J. N. (2009). Allometric equations and expansion factors for tropical try forest trees of eastern Sinaloa, Mexico. *Tropical and Subtropical Agroecosystems*, *10*(1), 45–52.

Chen, G., Hay, G. J., & Zhou, Y. (2010). Estimation of forest height, biomass and volume using support vector regression and segmentation from lidar transects and Quickbird imagery. *Proceedings of 18th International Conference On Geoinformatics, 18–20 June, Beijing, China.* IEEE GRSS, Geographical Society of China (4 pp.).

Chopping, M., Schaaf, C. B., Zhao, F., Wang, Z., Nolin, A. W., Moisen, G. G., et al. (2011). Forest structure and aboveground biomass in the southwestern United States from MODIS and MISR. *Remote Sensing of Environment*, *115*, 2943–2953.

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 Sensing of Environment*, *115*, 2931–2942.

Cocks, T., Jenssen, R., Stewart, A., Wilson, I., & Shields, T. (1998). The HyMap airborne hyperspectral sensor: The system, calibration and performance. *Proceedings of 1st EARSEL Workshop on Imaging Spectroscopy, Zurich, Switzerland, October 1998.*

Datt, B., McVicar, T. R., Van Niel, T. G., Jupp, D. L. B., & Pearlman, J. S. (2003). Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(6), 1246–1259.

Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., et al. (2012). Correlation and process in species distribution models: Bridging a dichotomy. *Journal of Biogeography*, *39*, 2119–2131.

Drake, J. B., Dubayah, R. O., Knox, R. G., Clark, D. B., & Blair, J. B. (2002). Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*, *81*(2–3), 378–392.

Dubayah, R. O., Sheldon, S. L., Clark, D. B., Hofton, M. A., Blair, J. B., Hurtt, G. C., et al. (2010). Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *Journal of Geophysical Research – Biogeosciences*, *115*(G2) (17 pp.).

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap.* CRC Press.

Eriksson, L. E. B., Santoro, M., Wiesmann, A., & Schmullius, C. C. (2003). Multitemporal JERS repeat-pass coherence for growing-stock volume estimation of Siberian forest. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(7), 1561–1570.

Fassnacht, F. E., Neumann, C., Förster, M., Buddenbaum, H., Ghosh, A., Clasen, A., et al. (2014). Comparison of feature reduction algorithms for classifying tree-species with hyperspectral data on three central-european test sites. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, http://dx.doi.org/10.1109/JSTARS.2014.2329390 (online available).

Garcia-Gutierrez, J., Gonzalez-Ferreiro, E., Mateos-Garcia, D., Riquelme-Santos, J. C., & Miranda, D. (2011). A comparative study between two regression methods on LiDAR data: A case study. *Proceedings of 6th International Conference on Hybrid Artificial Intelligent Systems, May 23–25, Wroclaw, Poland, Part II* (pp. 311–318).

Gayoso, J., Guerra, J., & Alarcón, D. (2002). *Contenido de carbono y funciones de biomasa en especies nativas y exóticas.* Document, Nr. 1, Project: FONDEF Medición de la capacidad de captura de carbono en bosques de Chile y promoción en el mercado mundial. Universidad Austral de Chile, Instituto Forestal (53 pp., 8 Appendices).

Ghosh, A., Fassnacht, F. E., Joshi, P. K., & Koch, B. (2014). A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observations and Geoinformation*, *26*, 49–63.

Gibbons, J. D., & Chakraborti, S. (2003). *Nonparametric statistical interference.* New York: Marcel Denker, Inc (645 pp.).

Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, *160*, 271–282.

Gleason, C. J., & Im, J. (2012). Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sensing of Environment*, *125*, 80–91.

Goodwin, N. R., Coops, N. C., & Culvenor, D. S. (2006). Assessment of forest structure with airborne LiDAR and the effects of platform altitude. *Remote Sensing of Environment*, *103*(2), 140–152.

Hartig, F., Dyke, J., Hickler, T., Higgins, S. I., O'Hara, R. B., Scheiter, S., et al. (2012). Connecting dynamic vegetation models to data — an inverse perspective. *Journal of Biogeography*, *39*, 2240–2252.

Hawkins, D. M. (2004). The problem of overfittting. *Journal of Chemical Information and Computer Science*, *44*, 1–12.

Houghton, R., Hall, F., & Goetz, S. (2009). Importance of biomass in the global carbon cycle. *Journal of Geophysical Research*, *114* (13 pp.).

Hudak, A. T., Strand, E. K., Vierling, L. A., Byrne, J. C., Eitel, J. U. H., Martinuzzi, S., et al. (2012). Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. *Remote Sensing of Environment*, *123*, 25–40.

Hyde, P., Nelson, R., Kimes, D., & Levine, E. (2007). Exploring LiDAR–RaDAR synergy — Predicting aboveground biomass in a southwestern ponderosa pine forest using LiDAR, SAR and InSAR. *Remote Sensing of Environment*, *106*(1), 23–38.

Hyyppä, J., Hyyppä, H., Leckie, D., Gougon, F., Yu, X., & Maltamo, M. (2008). Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing*, *29*(5), 1339–1366.

Imhoff, M. L. (1995). Radar backscatter and biomass saturation: Ramifications for global biomass inventory. *IEEE Transactions on Geoscience and Remote Sensing*, *33*(2), 511–518.

ITT Visual Information Solutions (2009). *ENVI 4.7 SP2.*

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology and Evolution*, *19*(2), 101–108.

Karjalainen, M., Kankare, V., Vastaranta, M., Holopainen, M., & Hyyppa, J. (2012). Prediction of plot-level forest variables using TerraSAR-X stereo SAR data. *Remote Sensing of Environment*, *117*, 338–347.

Koch, B. (2010). Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*, 581–590.

Koskinen, J. T., Pulliainen, J. T., Hyyppä, J. M., Engdahl, M. E., & Hallikainen, M. T. (2001). The seasonal behavior of interferometric coherence in boreal forest. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(4), 820–829.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York: Springer Science + Business Media.
Kurvonen, L., Pulliainen, J., & Hallikainen, M. (1999). Retrieval of biomass in boreal forests from multitempotal ERS-1 and JERS-1 SAR images. *IEEE Transactions on Geoscience and Remote Sensing, 37*(1), 198–205.
Latifi, H., Fassnacht, F., & Koch, B. (2012). Forest structure modeling with combined airborne hyperspectral and LiDAR data. *Remote Sensing of Environment, 121*, 10–25.
Latifi, H., & Koch, B. (2012). Evaluation of most similar neighbour and random forest methods for imputing forest inventory variables using data from target and auxiliary stands. *International Journal of Remote Sensing, 33*(21), 6668–6694.
Latifi, H., Nothdurft, A., & Koch, B. (2010). Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: Application of multiple optical/LiDAR-derived predictors. *Forestry, 83*, 395–407.
Laurin, G. V., Chen, Q., Lindsell, J. A., Coomes, D. A., Del Frate, F., Guerriero, L., et al. (2014). Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data. *ISPRS Journal of Photogrammetry and Remote Sensing, 89*, 49–58.
Li, X., Yeh, A. G. -O., Wang, S., Liu, K., Liu, X., Qian, J., et al. (2007). Regression and analytical models for estimating mangrove wetland biomass in South China using Radarsat images. *International Journal of Remote Sensing, 28*, 5567–5582.
McRoberts, R. E., Gobakken, T., & Næsset, E. (2012). Post-stratified estimation of forest area and growing stock volume using lidar-based stratifications. *Remote Sensing of Environment, 125*, 157–166.
McRoberts, R. E., Næsset, E., & Gobakken, T. (2013). Inference for lidar-assisted estimation of forest growing stock volume. *Remote Sensing of Environment, 128*, 268–275.
Morel, A. C., Fisher, J. B., & Malhi, Y. (2012). Evaluating the potential to monitor aboveground biomass in forest and oil palm in Sabah, Malaysia, for 2000–2008 with Landsat ETM plus and ALOS-PALSAR. *International Journal of Remote Sensing, 33*, 3614–3639.
Næsset, E., Gobakken, T., Bollandsås, O. M., Gregoire, T. G., Nelson, R., & Ståhl, G. (2013). Comparison of precision of biomass estimates in regional field sample surveys and airborne LiDAR-assisted surveys in Hedmark County, Norway. *Remote Sensing of Environment, 130*(15), 108–120.
Næsset, E., Gobakken, T., Solberg, S., Gregoire, T. G., Nelson, R., Stahl, G., et al. (2011). Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: A case study from a boreal forest area. *Remote Sensing of Environment, 115*, 3599–3614.
Nelson, R. F., Hyde, P., Johnson, P., Emessiene, B., Imhoff, M. L., Campbell, R., et al. (2007). Investigating RaDAR–LiDAR synergy in a North Carolina pine forest. *Remote Sensing of Environment, 110*(1), 98–108.
Nelson, R., Krabill, W., & Tonelli, J. (1988). Estimating forest biomass and volume using airborne laser scanner data. *Remote Sensing of Environment, 24*(2), 247–267.
Nothdurft, A., Soborowski, J., & Breidenbach, J. (2009). Spatial prediction of forest stand variables. *European Journal of Forest Research, 128*(3), 241–251.
Packalen, P., & Maltamo, M. (2006). Predicting the plot volume by tree species using airborne laser scanning and aerial photographs. *Forest Science, 52*(6), 611–622.
Packalen, P., & Maltamo, M. (2007). The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment, 109*, 328–341.
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 Sensing of Environment, 114*, 1053–1068.
Proisy, C., Couteron, P., & Fromard, F. (2007). Predicting and mapping mangrove biomass from canopy grain analysis using Fourier-based textural ordination of IKONOS images. *Remote Sensing of Environment, 109*, 379–392.
Rahman, M., Csaplovics, E., & Koch, B. (2007). An efficient regression strategy for extracting forest biomass information from satellite sensor data. *International Journal of Remote Sensing, 26*(7), 1511–1519.
Rauste, Y. (2005). Multi-temporal JERS SAR data in boreal forest biomass mapping. *Remote Sensing of Environment, 97*, 263–275.
Saatchi, S., Marlier, M., Chazdon, R. L., Clark, D. B., & Russell, A. E. (2011). Impact of spatial variability of tropical structure on radar estimation of aboveground biomass. *Remote Sensing of Environment, 115*(11), 2836–2849.
Santoro, M., Beer, C., Shvidenko, A., McCallum, I., Wegmüller, U., Wiesmann, A., et al. (2007). Comparison of forest biomass estimates in Siberia using spaceborne SAR, inventory based information and the LPJ dynamic global vegetation model. *Proceedings of ENVISAT Symposium, 23–27 April, Montreux, Switzerland*.
Santoro, M., Shvidenko, A., McCallum, I., Askne, J., & Schmullius, C. (2007). Properties of ERS-1/2 coherence in the Siberian boreal forest and implications for stem volume retrieval. *Remote Sensing of Environment, 106*, 154–172.
Santos, J. R., Freitas, C. C., Araujo, L. S., Dutra, L. V., Mura, J. C., Gama, F. F., et al. (2003). Airborne P-band SAR applied to the aboveground biomass studies in the Brazilian tropical rainforest. *Remote Sensing of Environment, 87*, 482–493.
Steininger, M. (2000). Satellite estimation of tropical secondary forest above-ground biomass: Data from Brazil and Bolivia. *International Journal of Remote Sensing, 21*, 1139–1157.
Steinmann, K., Mandallaz, D., Ginzler, C., & Lanz, A. (2013). Small area estimations of proportion of forest and timber volume combining Lidar data and stereo aerial images with terrestrial data. *Scandinavian Journal of Forest Research, 28*(4), 373–385.
Straub, C., & Koch, B. (2011). Estimating single tree stem volume of *Pinus sylvestris* using airborne laser scanner and multispectral line scanner data. *Remote Sensing, 3*(5), 929–944.
Straub, C., Weinacker, H., & Koch, B. (2010). A comparison of different methods for forest resource estimation using information from airborne laser scanning and CIR orthophotos. *European Journal of Forest Research, 129*, 1069–1080.
Sun, G., Ranson, J. K., Guo, Z., Zhang, Z., Montesano, P., & Kimes, D. (2011). Forest biomass mapping from lidar and radar synergies. *Remote Sensing, 115*, 2906–2916.
Tanase, M. A., Panciera, R., Lowell, K., Tian, S., Hacker, J. M., & Walker, J. P. (2014). Airborne multi-temporal L-band polarimetric SAR data for biomass estimation in semi-arid forests. *Remote Sensing of Environment, 145*, 93–104.
Tansey, K. J., Luckman, A. J., Skinner, L., Balzter, H., Strozzi, T., & Wagner, W. (2004). Classification of forest volume resources using ERS tandem coherence and JERS backscatter data. *International Journal of Remote Sensing, 25*, 751–768.
Thenkabail, P. S., Enclona, E. A., Ashton, M. S., Legg, C., & Jean De Dieu, M. (2004). Hyperion, IKONOS, ALI and ETM+ sensors in the study of African reinforests. *Remote Sensing of Environment, 90*, 23–43.
Thenkabail, P. S., Enclona, E. A., Ashton, M. S., & Van Der Meer, B. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment, 91*, 354–376.
Tian, X., Su, Z., Chen, E., Li, Z., van der Tol, C., Guo, J., et al. (2012). Estimation of forest above-ground biomass using multi-parameter remote sensing data over a cold and arid area. *International Journal of Applied Earth Observation and Geoinformation, 14*, 160–168.
Tomppo, E., & Halme, M. (2004). Using coarse scale forest variables as ancillary information and weighting of variables in k-nn estimation: A genetic algorithm approach. *Remote Sensing of Environment, 92*, 1–20.
Tonolli, S., Dalponte, M., Neteler, M., Rodeghiero, M., Vescovo, L., & Gianelle, D. (2011). Fusion of airborne LiDAR and satellite multispectral data for the estimation of timber volume in the Southern Alps. *Remote Sensing of Environment, 115*, 2486–2498.
Treuhaft, R. N., Asner, G. P., & Law, B. E. (2003). Structure-based forest biomass from fusion of radar and hyperspectral observations. *Geophysical Research Letters, 30*, 1472.
Treuhaft, R. N., Goncalves, F. G., Drake, J. B., Chapman, B. D., dos Santos, J. R., Dutra, L. V., et al. (2010). Biomass estimation in a tropical wet forest using Fourier transforms of profiles from lidar or interferometric SAR. *Geophysical Research Letters, 37*, L23403.
Tsui, O. W., Coops, N. C., Wulder, M. A., & Marshall, P. L. (2013). Integrating airborne LiDAR and space-borne radar via multivariate kriging to estimate above-ground biomass. *Remote Sensing of Environment, 139*, 340–352.
Tsui, O. W., Coops, N. C., Wulder, M. A., Marshall, P. L., & McCardle, A. (2012). Using multi-frequency radar and discrete-return LiDAR measurements to estimate above-ground biomass and biomass components in a coastal temperate forest. *ISPRS Journal of Photogrammetry and Remote Sensing, 69*, 121–133.
Üreyen, S., Hüttich, C., & Schmullius, C. (2014). Modeling growing stock volume using SAR data and OBIA: Effects of scale parameter and textural and geometrical features. *Journal of Remote Sensing Technology, 2*(1), 108–117.
Wagner, W., Luckman, A., Vietmeier, J., Tansey, K., Balzter, H., Schmullius, C., et al. (2003). Large-scale mapping of boreal forest in SIBERIA using ERS tandem coherence and JERS backscatter data. *Remote Sensing of Environment, 85*, 125–144.
Wang, C., & Qi, J. (2008). Biophysical estimation in tropical forests using JERS-1 SAR and VNIR imagery. II. Aboveground woody biomass. *International Journal of Remote Sensing, 29*, 6827–6949.
Weinacker, H., Koch, B., & Weinacker, R. (2004). TREESVIS — A software system for simultaneous 3D-realtime visualization of DTM, DSM, laser raw data, multispectral data, simple tree and building models. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Freiburg, Germany. Vol. XXXVI*. (pp. 90–95) (Part 8/W2).
White, D. (2011). Hyperion tools. (Available at:). http://www.exelisvis.com/Learn/CodeDetail/TabId/220/ArtMID/904/ArticleID/9669/Hyperion-Tools.aspx
Woodhouse, I. H., Mitchard, E. T. A., Brolly, M., Maniatis, D., & Ryan, C. M. (2012). Radar backscatter is not a 'direct measure' of forest biomass. *Nature Climate Change, 2*, 556–557.
Yu, X., Hyyppa, J., Vastaranta, M., Holopainen, M., & Viitala, R. (2011). Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*, 28–37.
Zell, J. (2008). *Methoden für die Ermittlung, Modellierung und Prognose der Kohlenstoffspeicherung in Wäldern auf Grundlage permanenter Großrauminventuren*. (Ph.D Dissertation). Faculty of Forest and Environmental Studies, University of Freiburg (162 pp., in German with English summary).
Zhao, K., Popescu, S., Meng, X., Pang, Y., & Agca, M. (2011). Characterizing forest canopy structure with lidar composite metrics and machine learning. *Remote Sensing of Environment, 115*, 1978–1996.
Zheng, D. L., Rademacher, J., Chen, J. Q., Crow, T., Bresee, M., le Moine, J., et al. (2004). Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sensing of Environment, 93*, 402–411.
Zianis, D., Muukkonen, P., Mäkipää, R., & Mencuccini, M. (2005). *Biomass and stem volume equations for tree species in Europe. SILVA FENNICA, Monographs, 4*, .
Zimble, D. A., Evans, D. L., Carlson, G. C., Parker, R. C., Grado, S. C., & Gerard, P. D. (2003). Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote Sensing of Environment, 87*, 171–182.
Zolkos, S. G., Goetz, S. J., & Dubayah, R. (2013). A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment, 128*, 289–298.