

## Stratified aboveground forest biomass estimation by remote sensing data



Hooman Latifi <sup>a,\*</sup>, Fabian E. Fassnacht <sup>b,c</sup>, Florian Hartig <sup>d</sup>, Christian Berger <sup>e</sup>, Jaime Hernández <sup>f</sup>, Patricio Corvalán <sup>f</sup>, Barbara Koch <sup>c</sup>

<sup>a</sup> University of Wuerzburg, Department of Remote Sensing in Cooperation with German Aerospace Center, Oswald-Kuelpe-Weg 86, Wuerzburg D-97074 Germany

<sup>b</sup> Institute for Geography and Geoecology, Karlsruhe Institute of Technology, Kaiserstraße 12, D-76131 Karlsruhe, Germany

<sup>c</sup> University of Freiburg, Chair of Remote Sensing and Landscape Information Systems, Tennenbacherstrasse 4, Freiburg D-79106 Germany

<sup>d</sup> University of Freiburg, Department of Biometry and Environmental System Analysis, Tennenbacherstrasse 4, Freiburg D-79106 Germany

<sup>e</sup> University of Jena, Department of Earth Observation, Loebdergraben 32, Jena D-07743, Germany

<sup>f</sup> Laboratorio de Geomática y Ecología del Paisaje, Universidad de Chile, Av. Santa Rosa 11315, Santiago de Chile, Chile

### ARTICLE INFO

#### Article history:

Received 1 November 2014

Received in revised form 23 January 2015

Accepted 27 January 2015

Available online 5 February 2015

#### Keywords:

LiDAR and hyperspectral remote sensing

Aboveground biomass

Statistical prediction

Post-stratification

Model performance

Factorial design

### ABSTRACT

Remote sensing-assisted estimates of aboveground forest biomass are essential for modeling carbon budgets. It has been suggested that estimates can be improved by building species- or strata-specific biomass models. However, few studies have attempted a systematic analysis of the benefits of such stratification, especially in combination with other factors such as sensor type, statistical prediction method and sampling design of the reference inventory data. We addressed this topic by analyzing the impact of stratifying forest data into three classes (broadleaved, coniferous and mixed forest). We compare predictive accuracy (a) between the strata (b) to a case without stratification for a set of pre-selected predictors from airborne LiDAR and hyperspectral data obtained in a managed mixed forest site in southwestern Germany. We used 5 commonly applied algorithms for biomass predictions on bootstrapped subsamples of the data to obtain cross validated RMSE and  $r^2$  diagnostics. Those values were analyzed in a factorial design by an analysis of variance (ANOVA) to rank the relative importance of each factor. Selected models were used for wall-to-wall mapping of biomass estimates and their associated uncertainty. The results revealed marginal advantages for the strata-specific prediction models over the unstratified ones, which were more obvious on the wall-to-wall mapped area-based predictions. Yet further tests are necessary to establish the generality of these results. Input data type and statistical prediction method are concluded to remain the two most crucial factors for the quality of remote sensing-assisted biomass models.

© 2015 Elsevier B.V. All rights reserved.

## Introduction

The estimation of aboveground forest biomass from remotely-sensed data is currently of great interest, due to important applications ranging from forest management to environmental and climate policy. Forest biomass is directly linked to carbon stocks, which are crucial for establishing future mitigation scenarios under climate change. The importance of forest biomass in

the context of such mitigation strategies is demonstrated by international initiatives such as reducing emissions from deforestation and forest degradation (REDD and REDD+) (e.g., Hill et al., 2013). Furthermore, biomass estimates can support surveys assessing the bioenergy potential of certain landscapes and help to monitor the sustainability of forest resources (e.g., Rosillo Calle et al., 2008).

Metrics from light detection and ranging (LiDAR) data have been frequently reported to provide good estimates of aboveground biomass across different geographical units (e.g., Hall et al., 2005; Næsset and Gobakken, 2008; Bright et al., 2012). A possibility to improve predictive accuracy could be including additional information, for example on species composition, in the estimation process. This could be achieved by various techniques. One is combining LiDAR information with optical data, but results have been mixed. Whereas some improvements could be obtained (e.g., Popescu

\* Corresponding author. Tel.: +49 931 3189638; fax: +49 931 31896380.

E-mail addresses: [hooman.latifi@uni-wuerzburg.de](mailto:hooman.latifi@uni-wuerzburg.de) (H. Latifi), [fabian.fassnacht@kit.edu](mailto:fabian.fassnacht@kit.edu) (F.E. Fassnacht), [florian.hartig@biom.uni-freiburg.de](mailto:florian.hartig@biom.uni-freiburg.de) (F. Hartig), [christian.berger@uni-jena.de](mailto:christian.berger@uni-jena.de) (C. Berger), [jhernand@uchile.cl](mailto:jhernand@uchile.cl) (J. Hernández), [pcovala@uchile.cl](mailto:pcorvala@uchile.cl) (P. Corvalán), [barbara.koch@felis.uni-freiburg.de](mailto:barbara.koch@felis.uni-freiburg.de) (B. Koch).

**Table 1**

Summary statistics of reference aboveground biomass values within the study site.

Min. value	Max. value	First quantile	Median	Mean	Third quantile	No. of samples
9.02	372.9	114	165.7	167.8	216.4	297

et al., 2004), these were occasionally reported to be only marginal (Kulawardhana et al., 2014), particularly in case of pure deciduous stands (Tonolli et al., 2011). Previous studies using predictors from LiDAR-based biomass models (Packalén and Maltamo, 2006, 2007; Breidenbach et al., 2010a,b) show promising results for predicting biomass on species level. Further refinements have been reported by incorporating hyperspectral metrics (e.g., Sarrazin et al., 2011). However, in many cases (e.g., in highly mixed stands) a realistic biomass prediction at tree species level will be severely restricted by factors such as spectral mixture due to tree crown overlaps. In such cases, a coarser division (i.e., post-stratification) into species groups (or communities) or into major strata of coniferous, deciduous and mixed stands is a compromise to retrieve strata-specific estimates (e.g., Eckert, 2012; Latifi et al., 2012). A practical example under which a similar stratification approach is applied is the Forest Inventory and Analysis program of the US, where remote sensing data are used to stratify sample plots from a nation-wide regular grid to subpopulations. The proportionally-allocated samples of each subpopulation are eventually inventoried in the field (e.g., Reams et al., 2005).

A superiority of species (or strata) – specific biomass models to those predicting the entire units at once has been found in a number of previous reports (Breidenbach et al., 2010a,b; Latifi et al., 2012). In case of LiDAR data, this may be related to the differing interactions of the laser pulse signals with the architecture of broadleaved and coniferous trees, as stated by Heurich and Thoma, (2008) who suggested the stratification into deciduous, coniferous and mixed strata for LiDAR-assisted forest parameter estimation.

There are several examples on comparisons between modeling approaches while predicting area-based biomass (e.g., Breidenbach et al., 2010a; Latifi et al., 2010; Powell et al., 2010; Main-Knorn et al., 2015; Gagliasso et al., 2014). However, studies addressing the general issue of post-stratification of the input data for remote sensing-based estimates are still scarce (see Heurich and Thoma, 2008; Dahlke et al., 2013). It has been suggested that classifying inventory plots information to forest types or districts may improve the precision of forest attribute estimation (Reams et al., 2005; Nelson, 2010; Latifi and Koch, 2012), particularly when the aim is to design a multi-level forest inventory for large area estimations (Katila and Tomppo, 2002; Andersen et al., 2011). However, recent reports also state an existing shortage of statistical analysis on post-stratified estimation of forest attributes to be a function of restriction in the sample size in small scale domains (McRoberts et al., 2012), who also provided examples on regional inferences of standing timber volume (McRoberts et al., 2013). Yet in order to draw reliable conclusions on the effect of stratification on forest biomass estimates, stratification approaches are needed to be examined in interaction with several other parameters which are known to influence remote sensing-based biomass estimates (e.g., sensor type, prediction method, sample size).

Here, we explore the question of whether stratification of sampling units into major forest types can influence the predictive quality of area-based forest biomass modeling. We based the models on a number of pre-selected predictors from sets of LiDAR and hyperspectral data. We did not consider building models based on combined LiDAR and hyperspectral predictors due to the previously-available reports on the fairly similar performance of LiDAR and combined

LiDAR + Hyperspectral data for the examined dataset (e.g., Latifi et al., 2012, Fassnacht et al., 2014).

Commonly applied parametric and non-parametric prediction methods were used on bootstrapped subsamples of the data to obtain a relative accuracy measure (RMSE) as well as the degree of variance explained by the models ( $r^2$ ) under cross-validation. Two subsequent analyses of Variance (ANOVA) were used to compare the differences in RMSE and  $r^2$  (a) between the strata (b) between the stratified and the non-stratified case with differences in predictive accuracy from other factors (prediction method, input data type and sample size). This allows us to systematically assess the importance of factors which typically occur when modeling stratified forest biomass by means of remote sensing data.

## Materials and methods

### Study site

The study site consists of nearly 900 ha of managed pure and mixed stands located in the vicinity of the southwestern German city of Karlsruhe ( $8^{\circ}24'09''E$ ,  $49^{\circ}03'37''N$  to  $8^{\circ}25'49''E$ ,  $49^{\circ}01'15''N$ ). The dominant tree species is scots pine (*Pinus sylvestris* L., with 56.3% of the total timber volume), occurring with other species such as European Beech (*Fagus sylvatica* L., with 17.8% of the total volume), Sessile Oak (*Quercus petraea* Liebl.) and Pedunculate Oak (*Quercus robur* L.) (jointly 14.9% of the total volume) and other deciduous trees (5.8% of the total volume). Further tree species including *Pseudotsuga menziesii*, *Picea* sp., *Abies* sp. and *Larix* sp. (jointly 5.2% of the total volume) are also sporadically present within the stands. The age of the stands ranges between 30 and 130 years. The stands were either comprised of dense, young stands (mainly pure Scots pine or pure oak trees) or of older stands (with Scots pine as the dominating species) with varying densities. The stands were mostly two-story with a second tree story consisting of broadleaves i.e., Beech and Hornbeam (*Carpinus betulus* L.).

### Field and remote sensing datasets

The reference biomass values were calculated from 297 plots inventoried in 2006. The systematically-gridded plot design was comprised of concentric circles of 2, 3, 6 and 12 m radii in a  $200 \times 100$  m grid. In each plot, trees with (DBH)  $<10$  cm,  $<15$  cm,  $<30$  cm, and  $\geq 30$  cm were measured if their distance to the plot center was 2, 3, 6 and 12 m, respectively (State Forest Service of Baden-Württemberg, 2009). The aboveground biomass of each tree was then calculated by applying species-specific allometric functions (Zell, 2008). The yielded biomass values were summed up to derive total biomass in tons per hectare. The descriptive statistics for the reference biomass values is summarized in Table 1.

**Table 2**

Number of samples for the three sample size classes which were applied in the individual model runs of the two experiments.

Experiment	Samples class 1	Sample class 2	Sample class 3
Broadleaved	34	48	73
Coniferous	38	48	76
Mixed	34	48	73
Reference samples	42	49	72

The LiDAR dataset included metrics extracted from pre-processed, full waveform data collected in August 2007 (leaf-on condition) by Toposys GmbH by means of a Harrier56 system and a Riegl LMS-Q560 laser scanner. Further technical specifications of the applied LiDAR data can be found in Latifi et al., (2012). Pulse-form data were derived from the original waveform dataset, from which solely the first pulse data was used. Following a digital terrain model (DTM) generation by means of the TreesVis software package (Weinacker et al., 2004), the available point cloud was normalized to derive the actual tree heights, where a height of 2 m was used as a threshold for deriving tree metrics, with a rationale of these being mainly representative for canopy hits (Packalén and Maltamo, 2006).

In addition to the LiDAR data, spectral metrics were extracted from an airborne hyperspectral scene (HyMap) acquired in August 2009 by the HyEurope campaign of the German Aerospace Center (DLR). The atmospherically/geometrically-corrected scene had a pixel size of 4 m. The data consisted of 125 spectral bands featuring a spectral resolution of 13–17 nm covering a range between 0.45 and 2.48 μm of the electromagnetic spectrum (Cocks et al., 1998). See Latifi et al., (2012) for further technical details on the applied LiDAR and hyperspectral data. A total number of 13 metrics were derived for the subsequent modeling. The LiDAR-extracted metrics included the mean height, the maximum height, 10th, 70th, and 90th height quantiles of first-pulse height as predictors. The rationale for this selection stems from prior experiences regarding the relevance of mean- and top-heights extracted features for biomass studies (e.g., Tsui et al., 2012).

From the HyMap data, eight predictors were extracted which mostly correlate with species information, vegetation density and leaf water content. The predictors consisted of normalized difference vegetation index (NDVI), the normalized difference water index (NDWI), and Chlorophyll-VI as suggested by Gitelson et al., (2003). In addition, five pre-selected original hyperspectral bands (closest bands corresponding to 518 nm, 681 nm, 1235 nm, 1477 nm and 2032 nm of the electromagnetic spectrum) were derived due to their previously-reported relevance to vegetation biomass (Thenkabail et al., 2004a; Latifi et al., 2012) and species information (Fassnacht et al., 2014; Thenkabail et al., 2004b).

#### Description of the experiments

Two experiments were conducted based on the stratification of samples prior to modeling. The first experiment explored the prediction of aboveground biomass for each of the three broadleaved, coniferous and mixed forest strata. The second test compared the stratified predictions to the predictions of total (i.e., unstratified) samples.

The assignment of the sample plots to the coniferous, deciduous and mixed strata was based on the information recorded for the inventoried field data. Thus, the percentage of coniferous and deciduous timber volume within each inventory plot was initially derived. Each plot was then assigned to the coniferous or deciduous stratum if this portion was 70% for either of the strata. Otherwise the given plot was assigned to the mixed stratum (see Latifi et al., 2012).

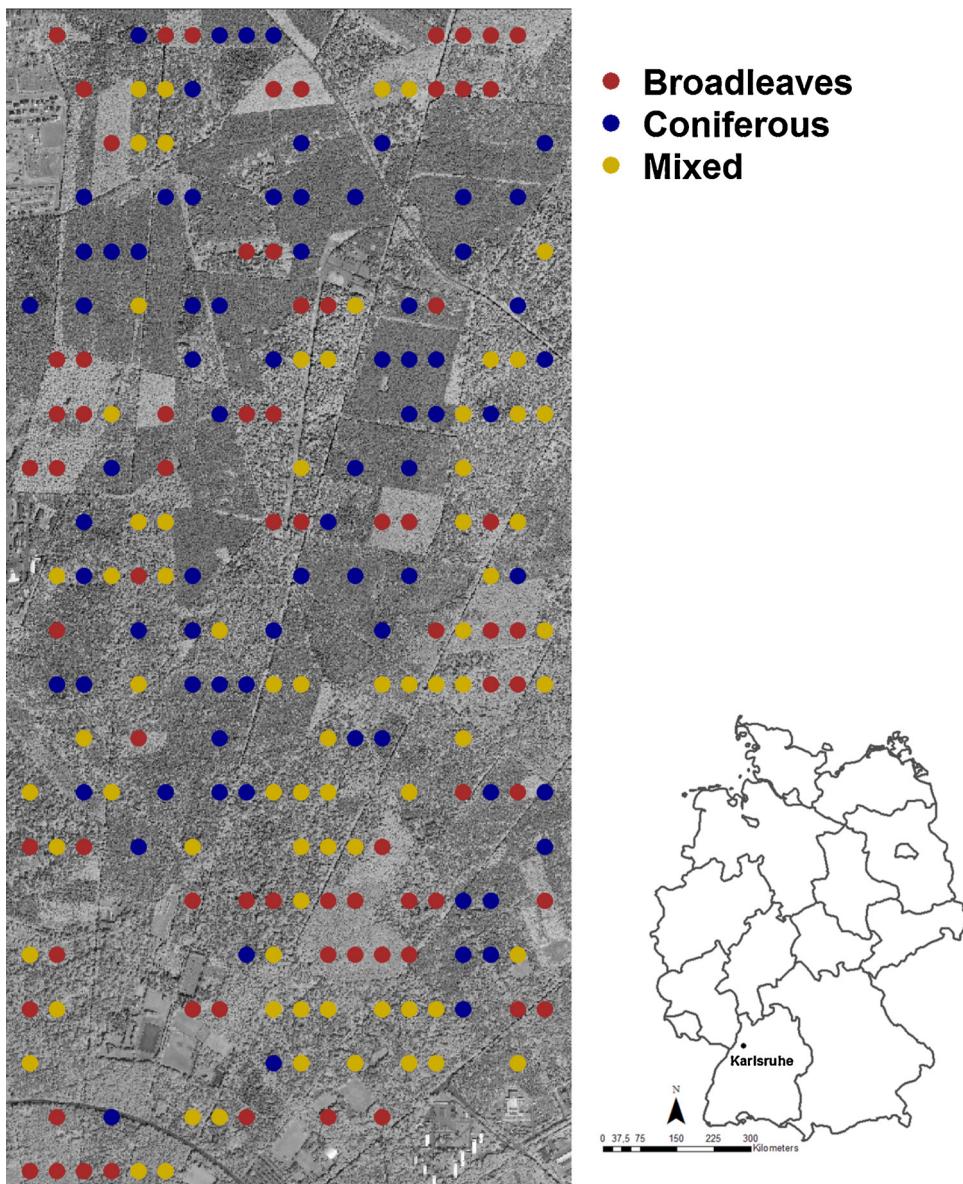
The application of the sample allocation methods such as proportional or Neyman (e.g., Peltoniemi et al., 2007) in stratification would inherently yield in differing stratum-specific sample sizes (see Westfall et al., 2011). This has been reported to pose as a highly-affecting factor on model performance in case of the current dataset in Karlsruhe test site, in which the proportion of plots covering the coniferous stratum is notably larger than those of the other two strata (Latifi et al., 2012). Since the aim here was to keep the explored factors independent from stratum size, we minimized this effect by deliberately forming three strata with nearly the same

number of sample units. Thus, the initially-higher number of 151 coniferous sample units was pruned by a systematic selection of every second plot from the set of samples sorted based on their total biomass. This finally led to 75 sample units for the coniferous stratum, as well as 73 sample units for both of the broadleaved and mixed strata. An overview of the location of the sample plots within the study area is provided in Fig. 1.

#### Biomass modeling

The modeling chain of the two experiments consisted of the following steps (Fig. 2):

- I. Separate matrices of response and predictor variables were formed for each of the two sources of remote sensing metrics. The  $n$  plots were then sorted according to their reference biomass values. From the sorted dataset, five groups of equal sample size (i.e.,  $ns = (n/5)$ ) were derived which is analogous to splitting the data into 20% percentiles. This split was integrated to reduce potential effects originating from differing data ranges and distributions of the response values (see step II). For the first experiment, this process was repeated for each forest stratum. The splitting process was independently conducted for each source of predictors (i.e., LiDAR and hyperspectral features).
- II. We performed a 500-times permuted stratified bootstrap on the groups built in step I through which samples from each of the five sample groups were drawn with replacements. Then, a single input dataset per bootstrap was formed by combining the drawn subsets from the five sample groups. As a consequence, sample units from the full range of available biomass values were available in each bootstrapped stratum-specific input dataset. The within-group drawn sample size  $x$  was varied three times ( $x = ns/2, x = ns/1.5, x = ns$ ) which led to forming three sample size classes with differing number of sample units for each class. Input datasets of class 1 ( $x = ns/2$ ) embraced the least sample units, whereas those of class 3 ( $x = ns$ ) contained the most.
- III. Subsequently, each of the created 500 input datasets of each input sample size class were fit by 5 selected prediction methods, including k-nearest neighbor (KNN), support vector machines (SVMs), gaussian processes (GPs), random forest (RF) and step-wise linear regression (LMSTEP). A summary of the applied prediction methods is included in Appendix 1. Cross-validations were conducted for each case, where the model outputs included the best model parameters (selected based on lowest RMSE values) as well as the RMSE and  $r^2$  diagnostics. 3-fold cross-validation with 5 repetitions was performed. A higher number of folds was not possible because of too few sample units in the hold-out samples.
- IV. To quantify the influence of each tested factor on the predictive error, analysis of variance (ANOVA) was applied (see Chambers et al., (1992) for details on ANOVA). Independent variables in the ANOVA were the sample size (as applied by sample size class), the prediction method, the sensor type (LiDAR/hyperspectral), the either (a) stratification into broadleaved, coniferous, mixed, or (b) and presence/absence of stratification, as well as their corresponding interactions. For the presence/absence of stratification, the sample sizes were kept approximately equal. Table 2 provides an overview of the sample sizes for the entire experiment.
- V. The selected best models of each prediction method were used for wall-to-wall mapping of the biomass estimates (and the corresponding coefficients of variation (CV)). To this aim, the mean values of the entire 500 model runs were calculated for both predicted biomass values and the standard error of the estimates.



**Fig. 1.** Geographical Location of the study site in Baden–Württemberg state- Germany shown on the NIR band of an aerial orthophoto and overlaid by the stratified sample plots (the coniferous stratum has been already pruned to yield similar number of samples as the broadleaves and mixed strata).

The model training, prediction and mapping processes were entirely implemented in an open source domain in R (Development Core Team, 2014) by integrating a number of libraries including kernlab (Karatzoglou et al., 2004), RandomForest (Liaw and Wiener, 2012) and caret (Kuhn and Johnson, 2013).

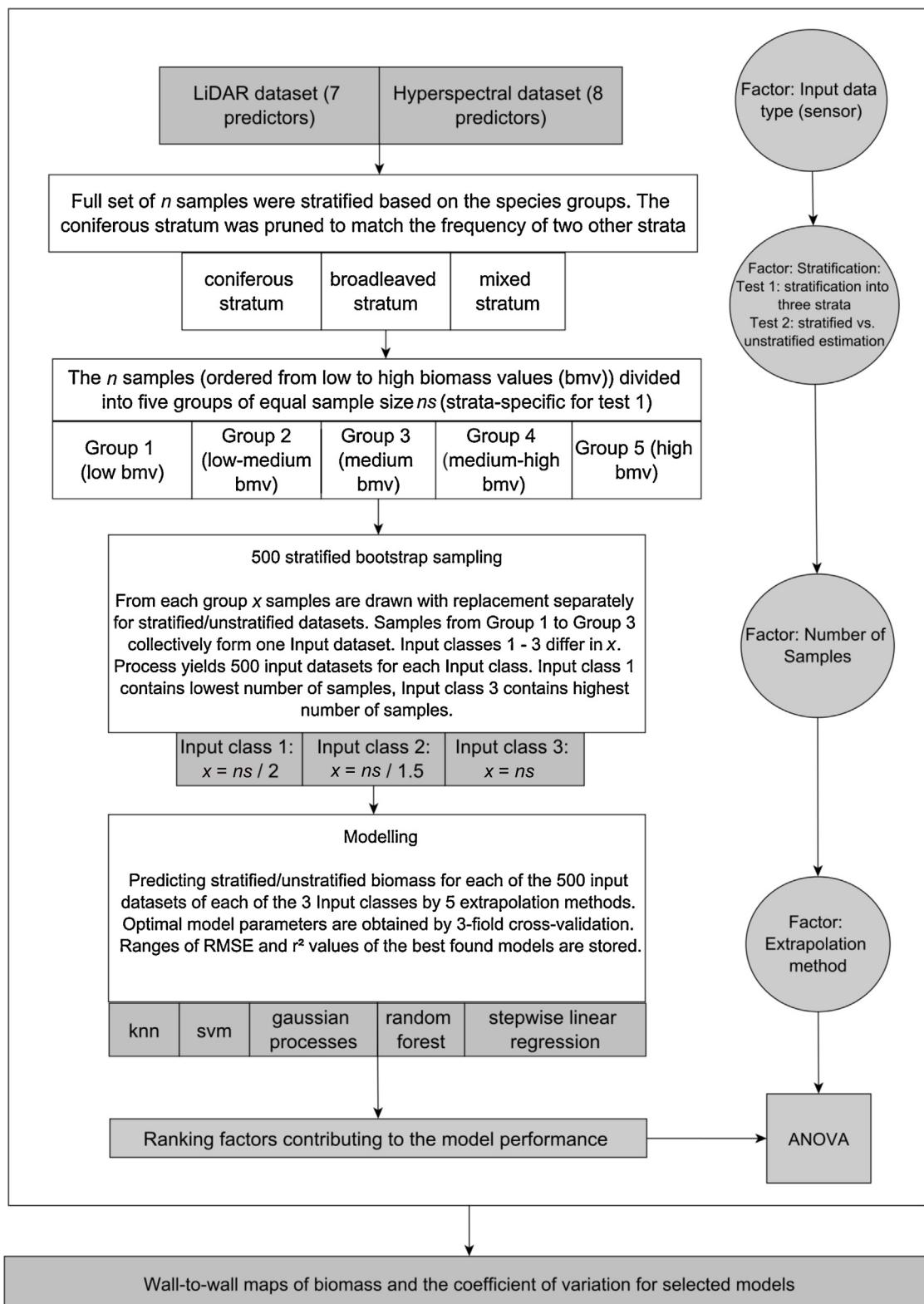
### 3. Results

#### Model performances

Figs. 3–6 summarize the results of model performances. For the strata-specific models, the median  $r^2$  values varied between 0.17 (kNN for broadleaves) and 0.48 (RF for coniferous) for the hyperspectral predictors. For the LiDAR predictors they ranged between 0.42 (SVM for coniferous) and 0.6 (RF for coniferous). No explicit trend was observed among the applied modeling approaches. When applying hyperspectral metrics, lower  $r^2$  rates were returned by the broadleaved stratum than by both coniferous and mixed strata (lowest  $r^2 \sim 0.17$  for broadleaved compared

to lowest  $r^2 \sim 0.22$  and  $r^2 \sim 0.20$  for mixed and coniferous strata). When applying LiDAR-based predictors, a lower amount of variance (lowest median  $r^2 \sim 0.42$ ) was explained for coniferous stratum than for mixed (lowest  $r^2 \sim 0.47$ ) and broadleaved (lowest  $r^2 \sim 0.46$ ) strata (Fig. 3).

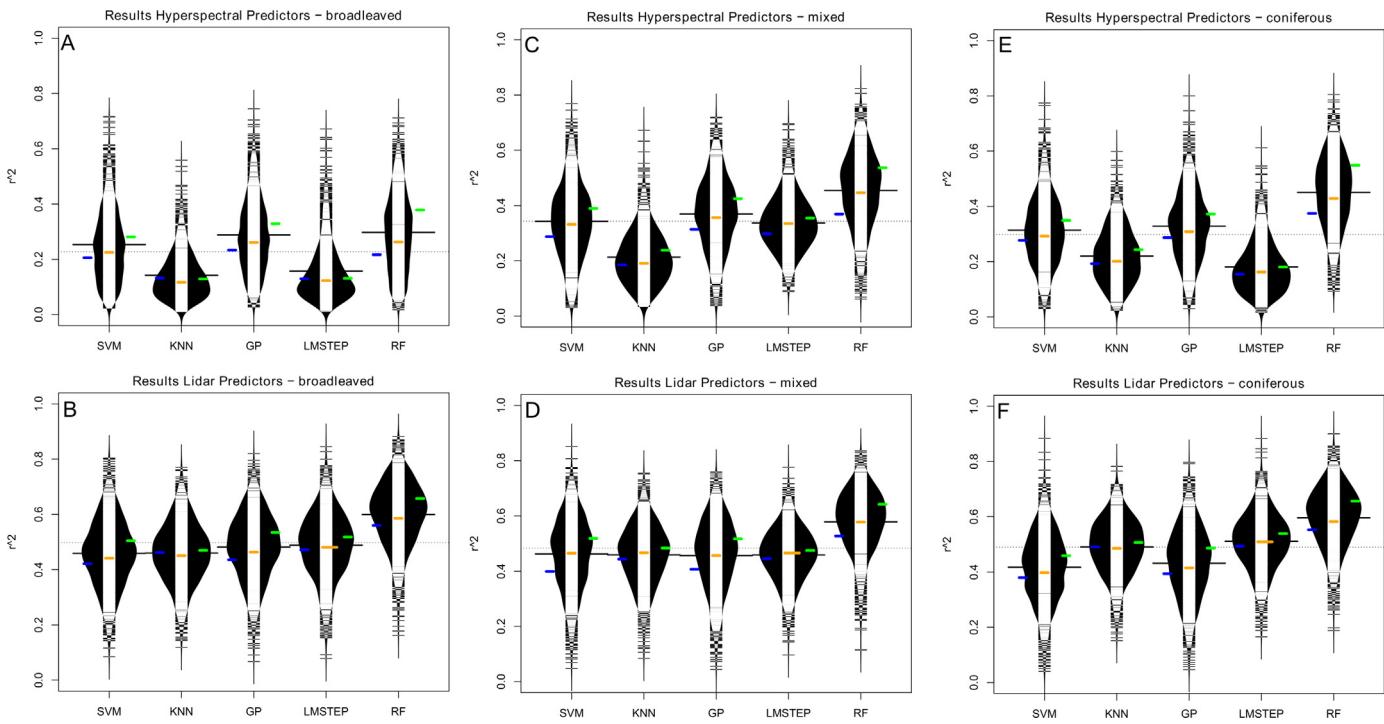
When applying hyperspectral predictors, the RMSE rates showed comparable patterns with lowest performances for the broadleaved stratum (highest median RMSE  $\sim 77$  t/ha for LMSTEP), followed by coniferous (highest median RMSE  $\sim 72$  t/ha for LMSTEP) and mixed (highest median RMSE  $\sim 64$  t/ha for kNN) strata (Fig. 4). Applying LiDAR predictors turned out fairly better performances by returning lowest RMSE for the coniferous stratum (mean RMSE  $\sim 44$  t/ha by RF), followed by the broadleaves (lowest mean RMSE  $\sim 45$  t/ha by RF) and the mixed (mean RMSE  $\sim 46$  t/ha) strata (Fig. 4). A general tendency toward improved model performance was observed along with increasing the number of input samples, as shown by the colored horizontal stripes in the Fig. 4. Furthermore, the RF showed to generally return higher performances (i.e., lower RMSEs) compared to other applied extrapolation methods.



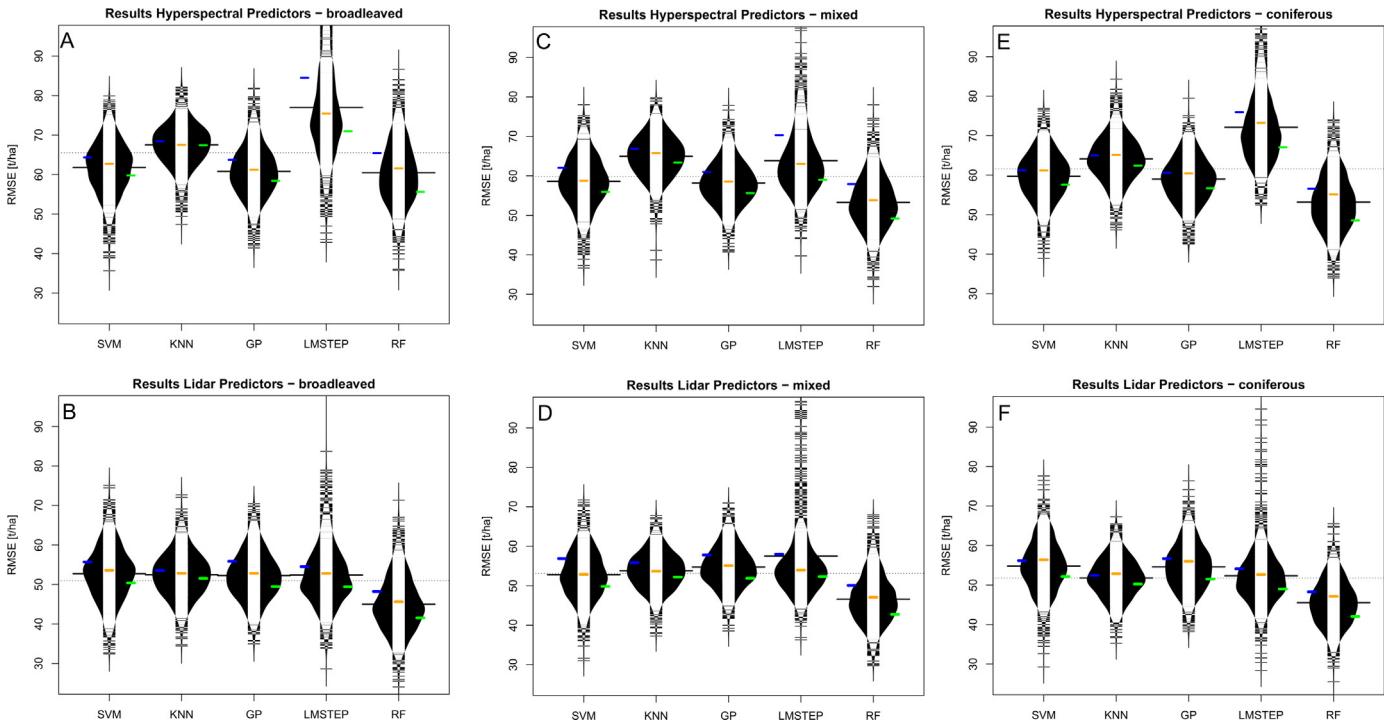
**Fig. 2.** The experimental set-up of the biomass modeling approach.

Figs. 5 and 6 summarize the model performances for the stratified (panels a and b) and the unstratified (panels c and d) model runs. One may particularly note the generally improved rates for both RMSE and  $r^2$  diagnostics for the stratified runs (highest median  $r^2 \sim 0.59$ , lowest median RMSE  $\sim 45$  t/ha by LiDAR data and RF

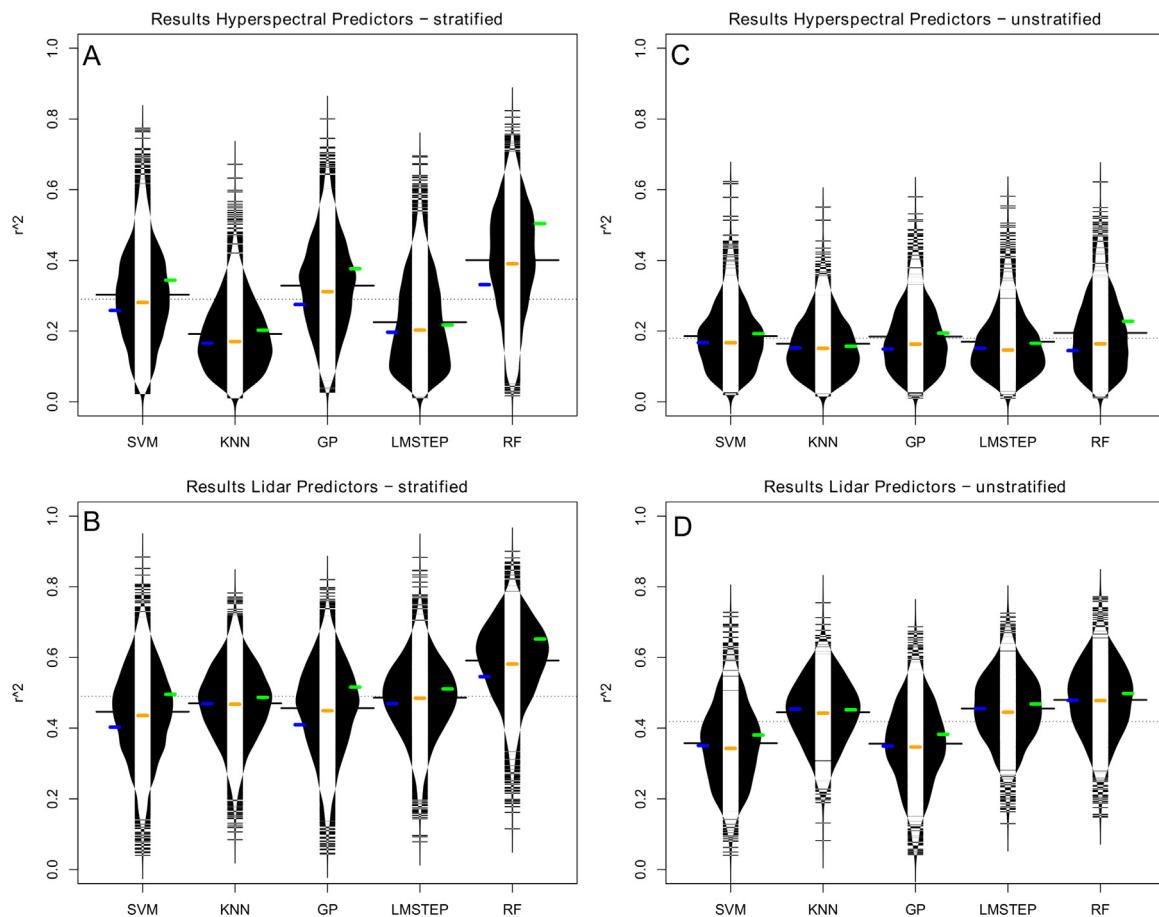
model) as compared to those built by unstratified samples (highest median  $r^2 \sim 0.48$ , lowest median RMSE  $\sim 50$  t/ha by LiDAR data and RF model). In addition to the overall median rates, improvements were also observed for the median values of each individual prediction method across all the applied sample sizes.



**Fig. 3.** The variance explained by the stratified models by applying two separate sets of hyperspectral (top) and LiDAR (bottom) predictors. The beanplots illustrates the distribution of the  $r^2$  values from the 500 bootstrapped models as obtained by the 3-fold-cross validation for each extrapolation method (LMSTEP = stepwise linear models, SVM = support vector machines, KNN = k-nearest neighbor, RF = random forest, GP = Gaussian processes) and sample size (class 1–3) for hyperspectral-based models of broadleaved (A), mixed (C) and coniferous (E) strata, as well as for LiDAR-based models of broadleaves (B), mixed (D) and coniferous (F) strata. The horizontal black line in each beanplot shows the mean value. The median  $r^2$  for each of the three sample size classes is given with the blue, yellow and green horizontal stripes for class 1–3, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** The RMSE [t/ha] returned by the stratified models by applying two separate sets of hyperspectral (top) and LiDAR (bottom) predictors. The beanplots illustrates the distribution of the RMSE values from the 500 bootstrapped models as obtained by the 3-fold-cross validation for each extrapolation method (LMSTEP = stepwise linear models, SVM = support vector machines, KNN = k-nearest neighbor, RF = random forest, GP = Gaussian processes) and sample size (class 1–3) for hyperspectral-based models of broadleaved (A), mixed (C) and coniferous (E) strata, as well as for LiDAR-based models of broadleaves (B), mixed (D) and coniferous (F) strata. The horizontal black line in each beanplot shows the mean value. The median RMSE for each of the three sample size classes is given with the blue, yellow and green horizontal stripes for class 1–3, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 5.** The stratified vs. unstratified models as compared by the variance explained by the models applying two separate sets of hyperspectral (top) and LiDAR (bottom) predictors. The beanplots illustrates the distribution of the  $r^2$  values from the 500 bootstrapped models as obtained by the 3-fold-cross validation for each extrapolation method (LMSTEP = stepwise linear models, SVM = support vector Machines, KNN = k-nearest neighbor, RF = random forest, GP = Gaussian processes) and sample size (class 1–3) for hyperspectral-based stratified (A) and unstratified (C) models, as well as for LiDAR-based stratified (B) and unstratified (D) models. The horizontal black line in each beanplot shows the mean value. The median  $r^2$  for each of the three sample size classes is given with the blue, yellow and green horizontal stripes for class 1–3, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

#### ANOVA results

Tables 3 and 4 summarize the outcomes of the ANOVAs by quantifying the influence of prediction methods (SVM, KNN, GP, LMSTEP, and RF), sample size (class 1–3), sensor type (LiDAR, hyperspectral) and modification (three individual strata as well as stratified vs. unstratified prediction) on the obtained performance diagnostics. The factors have been ranked based on their calculated sum of squares (ssq).

All in all, the highest ranked factors within the first ANOVA based on the explained variance were sensor type ( $ssq = 450.7$ ), prediction method ( $ssq = 145.7$ ) and the sample size ( $ssq = 42$ ). The stratification seemed to be only of minor importance ( $ssq = 19.6$ ). When exploring RMSE rates, the same ranking was observed. The second ANOVA (stratified vs. unstratified models) showed fairly differing rankings based on either  $r^2$  or RMSE rates: the highest  $r^2$ -ranked factors included sensor type ( $ssq = 660.7$ ), prediction method ( $ssq = 128$ ) and stratification ( $ssq = 92.8$ ), while the highest ranks based on RMSE included sensor type ( $ssq = 1,685,775$ ), prediction method ( $ssq = 618,125$ ) and the number of samples ( $ssq = 249,304$ ).

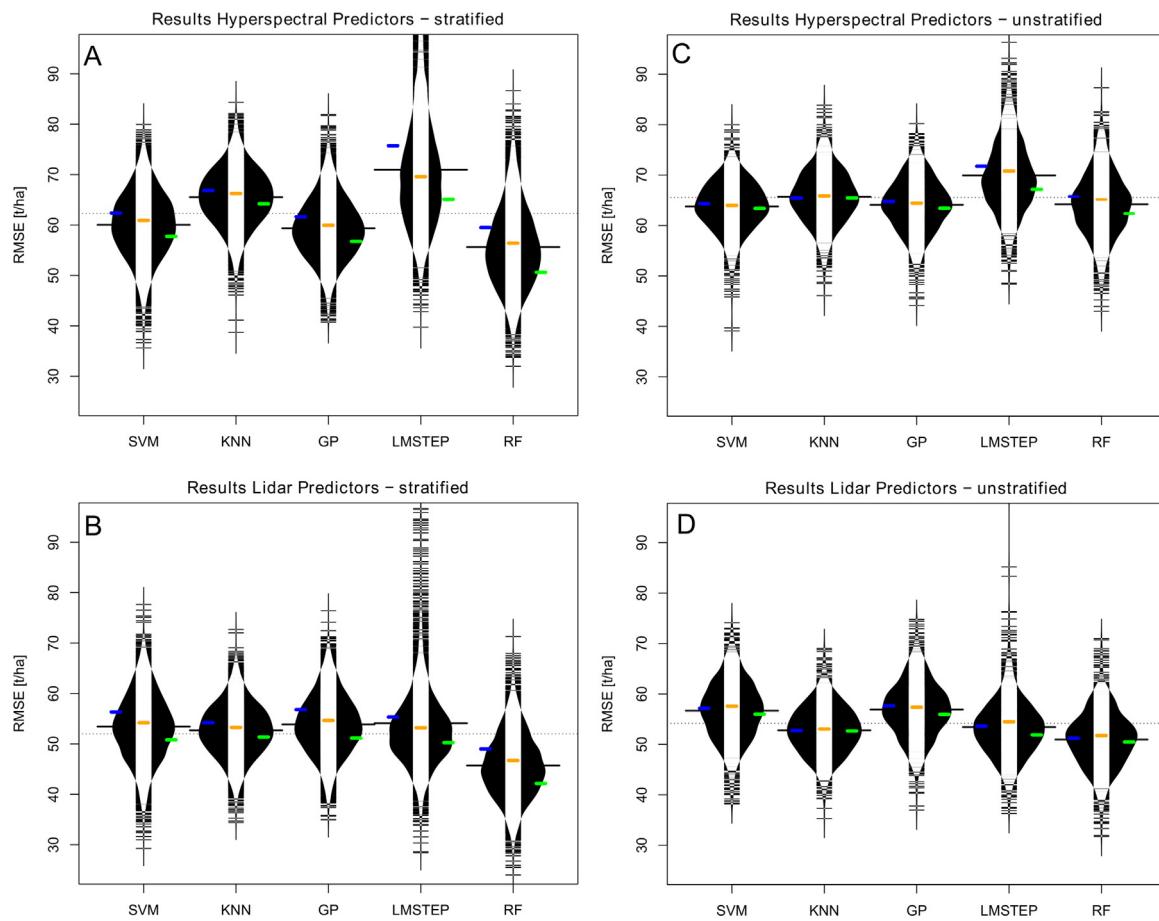
#### Area-based predictions

Figs. 7 and 8 contain examples of wall-to-wall predictions of biomass (left side) and their corresponding CVs (right side). The

mapped CVs provide information on the extent of variability relative to the mean which is analogous to a measure of relative standard uncertainty of the predicted biomass. The maps show the strata-specific (Fig. 7) and unstratified (Fig. 8) estimates predicted with the highest within-group drawn sample size (class 3) by means of the best performing approach (RF) on LiDAR predictors. Slight differences in the spatial distribution of predicted biomass across the different strata as well as differences in the value range can be observed. Additionally, one may note the somewhat higher CVs (most obviously located within the non-forested areas in e.g., upper left part of the maps) in both coniferous and broadleaved strata (panels A and C in Fig. 7) as compared to that of mixed stratum (panel B in Fig. 7). Despite that observation, fairly similar patterns were observed throughout the mapped biomass for all the strata, which reflect the previously-mentioned subtle effect of the sample stratification. The Fig. 8 depicts an exemplified prediction map of an unstratified model, in which the predicted pixels represent the biomass values with a degree of smoothness as opposed to the heterogeneity unveiled when mapping based on the strata-specific estimates.

#### Discussion

Regional forest inventories are often based on systematic sampling grids, and thus are constrained to previously-defined sampling designs and intensities (McRoberts et al., 2012). Since a



**Fig. 6.** The stratified vs. unstratified models as compared by the RMSE [t/ha] by the models applying two separate sets of hyperspectral (top) and LiDAR (bottom) predictors. The beanplots illustrates the distribution of the RMSE values from the 500 bootstrapped models as obtained by the 3-fold-cross validation for each extrapolation method (LMSTEP = stepwise linear models, SVM = support vector machines, KNN = k-nearest neighbor, RF = random forest, GP = Gaussian processes) and sample size (class 1–3) for hyperspectral-based stratified (A) and unstratified (C) models, as well as for LiDAR-based stratified (B) and unstratified (D) models. The horizontal black line in each beanplot shows the mean value. The median RMSE for each of the three sample size classes is given with the blue, yellow and green horizontal stripes for class 1–3, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 3**

Results of ANOVA conducted to explain variance of  $r^2$  and RMSE as obtained for the additional experiments on test site Karlsruhe.

	Df	$r^2$ SumSq	RMSE SumSq
PredMeth	4	145.7	677886
SensType	1	450.7	1201864
SampSi	3	42.0	271392
Strat	2	19.6	27712
PredMeth:SensType	4	41.1	195998
PredMeth:SampSi	8	13.7	52194
PredMeth:Strat	8	9.6	27507
SensType:SampSi	2	0.0	1148
SensType:Strat	2	32.4	117983
SampSi:Strat	4	1.4	9853
PredMeth:SensType:SampSi	8	1.6	6943
PredMeth:SensType:Strat	8	15.0	76194
PredMeth:SampSi:Strat	16	0.4	2227
SensType:SampSi:Strat	4	0.4	725
PredMeth:SensType:SampSi:Strat	15	0.3	4928
Residuals	44910	543.1	2049713

PredMeth = Prediction method, SampSi = number of input samples, SensType = input datatype / sensor, Strat = stratification, here: stratification into broadleaved, coniferous, mixed stands. The colons (:) shows the interaction between two given factors.

**Table 4**

Results of ANOVA conducted to explain variance of  $r^2$  as obtained for the additional experiments on test site Karlsruhe.

	Df	$r^2$ SumSq	RMSE SumSq
PredMeth	4	128.0	618125
SensType	1	660.7	1685775
SampSi	3	38.6	249304
Strat	1	92.8	83277
PredMeth:SensType	4	49.8	249304
PredMeth:SampSi	8	12.8	50628
PredMeth:Strat	4	27.5	87885
SensType:SampSi	2	0.0	1468
SensType:Strat	1	4.2	3165
SampSi:Strat	3	5.0	32537
PredMeth:SensType:SampSi	8	1.7	7475
PredMeth:SensType:Strat	4	2.7	5003
PredMeth:SampSi:Strat	8	1.5	5479
SensType:SampSi:Strat	2	0.0	7
PredMeth:SensType:SampSi:Strat	7	0.1	428
Residuals	59940	769.6	2807363

PredMeth = Prediction method, SampSi = Number of input samples, SensType = Input datatype / sensor, Strat = Stratification, here: stratified models vs. non stratified samples. The colons (:) shows the interaction between two given factors.

stratified sampling can often not be integrated into forest inventory plans prior to the inventory, a post-stratification of the previously-recorded plots into major forest types have been stated to partially enable more precise estimations of forest attributes (Westfall et al., 2011; McRoberts et al., 2012).

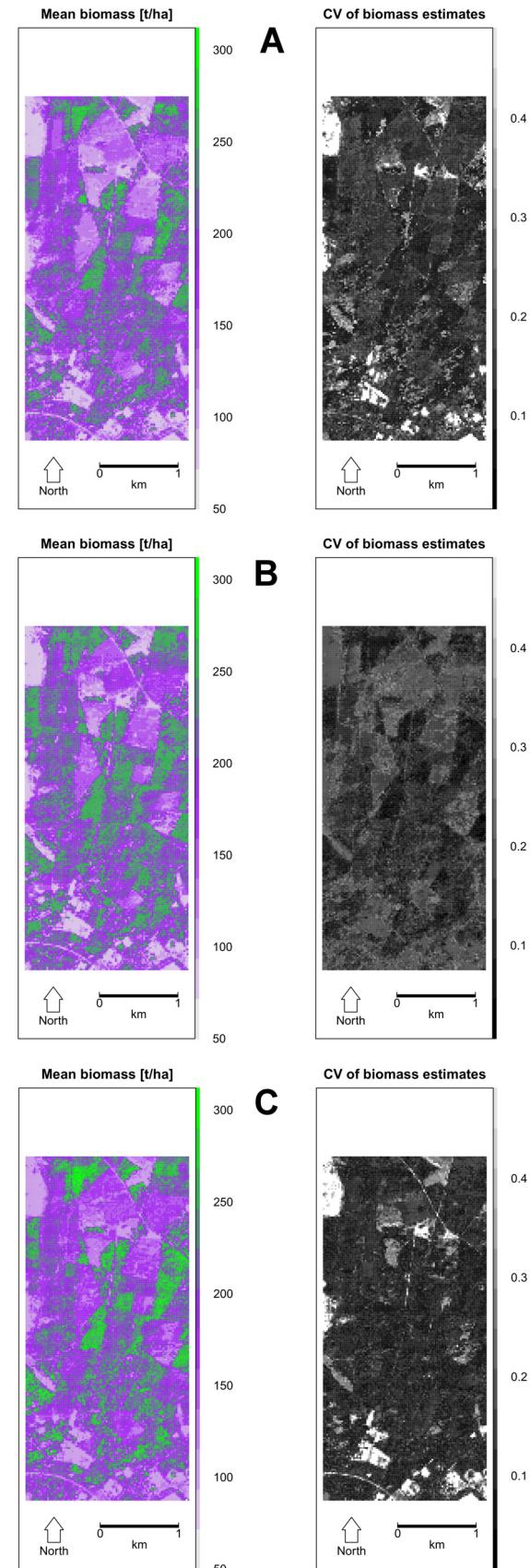
Here, we investigated the effect of stratification on remote sensing-based aboveground biomass predictions in combination with a set of crucial factors that have been shown to affect the quality and robustness of small-scale forest biomass estimation by remote sensing data. The considered factors include sample size, prediction method and sensor type. For each combination of those factors and the strata, we calculated cross-validated  $r^2$  and RMSE diagnostics with replicated datasets created by bootstrapping. A subsequent ANOVA allowed us to separate the relative influence of the above-mentioned factors on the predictive error.

The results of our experiment suggest that the importance of a post stratification of the reference samples is of lower importance compared to factors such as the sensor type (i.e., the input predictor data) and the applied prediction method. Here, the strata-specific models were only slightly better than the total ones, as revealed by the ANOVA analysis in which the comparison between stratified and total models was included as one of the examined factors. A previous study across another German test site stated the stratification of sample units to bear different levels of effects on LiDAR-based estimation of stand parameters (Heurich and Thoma, 2008). They reported the variance explanation of models for stand height and DBH to be only marginally enhanced by stratification, whereas the standing volume and density were significantly enhanced.

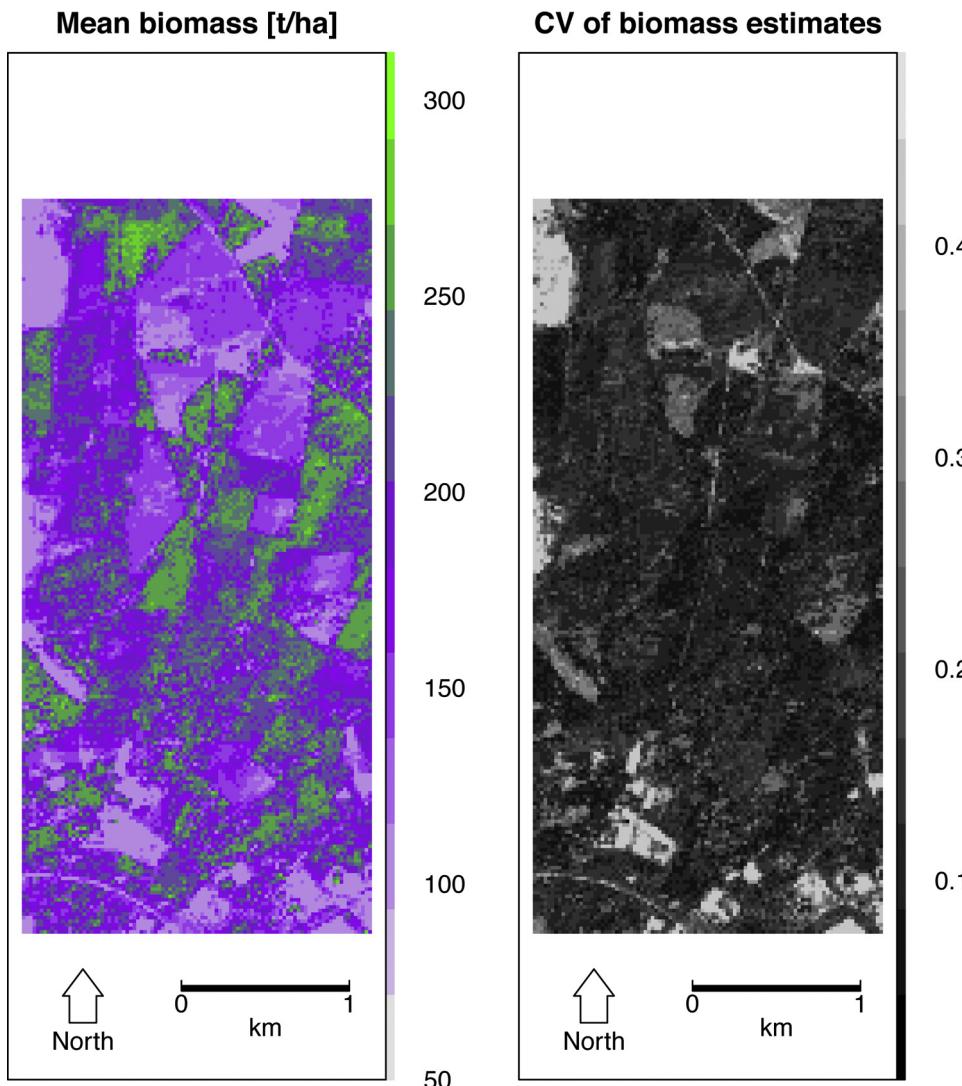
In a typical stratification case, the strata-specific sample sizes are defined based on the size of each subpopulation. Though several numbers have been suggested as minimum number of samples allowed within each forest stratum (Särndal et al., 1992; Ott, 1999), this minimum depends chiefly on the population structure (Cochran, 1977; Westfall et al., 2011). In a general sense, the less-populated strata are presumed to return higher prediction errors which in turn affect the total prediction error (see e.g., the RMSE values of up to 261% for Douglas fir (*Pseudotsuga menziesii* reported by Breidenbach et al., 2010a)). This was also mentioned by e.g., Westfall et al., (2011) who warn about biased estimates of standard errors at small within-strata sample sizes.

In our test site, an asset in applying forest strata-specific models over the total ones has previously been confirmed to be marginal and restricted to the more homogenous coniferous stratum (mainly consisting of Scots Pine in Karlsruhe test site) (Latifi et al., 2012). In the latter study a proportional allocation of the sample units resulted in a notably higher number of sample units for the coniferous stratum. Here we minimized the effect of stratum size on the model outputs by deliberately keeping the sample sizes of the three strata on a comparable level. However, the observed results concerning this were similar to the latter study, where only a minimal improvement was observed for the stratified models over the unstratified ones. This confirms that the predictive power of a stratified model is not only a function of the homogeneity (and spatial dominance) of the higher populated strata, but also a function of more crucial factors such as sensor type and prediction method. Further tests have to be done to confirm whether this observation is specific to our test site, or whether it applies in general.

By solely applying multispectral predictors, Labrecque et al., (2006) reported slightly higher total explained variance (i.e.,  $r^2$  values) for the stratified models compared to the unstratified ones. They also confirmed that a post-stratification was most effective when predicting the strata which are formed based on species groups (i.e., coniferous, broadleaves and mixed) rather than those based on the species level. In addition, our results are in line with those from Labrecque et al., (2006), in which the stratification



**Fig. 7.** Wall-to-wall predicted maps (left) and the CV of estimates (right) of mean stratified biomass for broadleaves (A), mixed (B) and coniferous (C) strata. The predictions are resulted from RF models incorporating the class 3 sample size and LiDAR predictors.



**Fig. 8.** Wall-to-wall predicted maps (left) and the CV of estimates (right) of mean unstratified biomass. The predictions are resulted from RF models incorporating the class 3 sample size and LiDAR predictors.

caused almost no advantage in terms of reported RMSE and estimation bias. In another case, a post-stratification to landcover types did only marginally reduce the standard error of LiDAR-assisted biomass estimates (Andersen et al., 2011).

The RF models were shown here to generally exceed other tested prediction methods, particularly in presence of larger sample sizes. The higher performance of RF compared to the other tested methods confirms several previous reports on the ability of RF to flexibly deal with multiple modeling contexts in inventory of forest biomass (e.g., Breidenbach et al., 2010a; Yu et al., 2011; Latifi and Koch, 2012), although some other studies did not find significant improvements compared to parametric regressions (Penner et al., 2013). For RF, one may note that it is affiliated with an internal subsampling, which presumably causes the single estimates to be highly variant when the number of input subsamples is low (Wagner et al., 2014). Here, we also observed the higher performance of models with highest number of samples per stratum compared to those containing the lower number of samples. Nevertheless, the sample size in our study was not notably affecting the achieved  $r^2$  and RMSE rates when focusing on the ANOVA results.

Considering the rank of the other factors in ANOVA, the high importance of the sensor type was expectedly observed. The significant dominance of 3D metrics from LiDAR data has been repeatedly confirmed compared to other sources of optical data for modeling forest structural attributes (Sexton et al., 2009; Koch, 2010; Clark et al., 2011).

The examples of wall-to-wall predictions on pixel level mirrored the general patterns within the study area well. In the stratified predictions, the mixed stratum proved to be somewhat less prone to high prediction uncertainties as shown by its comparatively less CV values. This was most presumably due to the absence of extreme biomass values compared to both broadleaves and coniferous strata in combination with the applied prediction method. For the mixed strata, there was only one reference sample plot with forest biomass of >300 t/ha. Since RF (and in general nearest neighbor imputation methods) can only predict within the range of values offered by the reference sample units, the smaller range of values may have contributed to the decreased CV. In addition, the presence of both coniferous and broadleaved trees in the mixed strata may have caused a smoothing effects on the wall-to-wall biomass estimates.

Although being moderate, differences existed in the wall-to-wall estimates of the three broadleaved, mixed and coniferous strata. One should keep in mind that this poses new challenges concerning the use of final maps or the calculated statistics when modeling spatially-explicit biomass estimates for larger areas. The main idea behind applying remote sensing data for biomass estimations is to extrapolate the plot-based information to larger geographic domains which provides the required area-based information that would be costly to collect in the field surveys. Applying stratification will lead to as many wall-to-wall prediction maps as the number of models being built for the individual strata. If the objective is to transport the advantages in the model performances of the applied stratification (which were yet observed to be small in the current study) to the wall-to-wall maps, approaches have to be developed to merge the various mapping results into a single product. If spatially explicit, strata-specific information concerning each location in the study area are available, a straightforward approach would possibly be to apply the best strata-specific model to each location which belongs to the corresponding stratum. The final product would then be a mosaic of the predictions of all locations, each predicted with the corresponding strata-specific model. If this information is not available, one alternative would be to use a remote sensing-based classification of the study site into the defined forest strata and then follow the procedure as outlined above. This could be a motivating subject for future research.

## Conclusion

In the context of remote sensing-assisted estimation of forest biomass, we examined the importance of sensor type, statistical prediction method, sample size as well as the influence of stratification of the sample units in two experiments. The results lead us to the conclusion that the sensor type (hyperspectral/LiDAR) showed to be the essential source of impact on the yielded predictive performance of the models. This was followed by the effect of prediction method, while sample size turned out to be only of a relatively low importance. The modeling based on forest strata did show slight improvements over the use of unstratified models. Thus, this study mainly bears the message that model diagnostics of biomass estimations can be more drastically influenced by data- and method-driven factors compared to those factors directly related to the stratification process. Nevertheless, a certain improvement in model diagnostics was achieved with the stratification procedure. Further investigations with further datasets are required to draw final conclusions on the question of how far the stratification approaches can contribute to improve remote sensing-based biomass estimations. We therefore encourage further tests across other test sites with the algorithms presented here to allow an exact comparison to our results.

We believe that the ranking of the importance of the most relevant factors affecting the modeling of the target variable, as shown in the present study, can help to objectively assess the performance of the various modeling approaches in the literature. This may be a key step toward a quantitative ranking of different modeling strategies for predicting biomass values of forest ecosystems on larger geographical domains, but also beyond that for other remote sensing applications.

## Acknowledgements

This study was partly funded by the German Aerospace Center (DLR) and the German Federal Ministry of Economy and Technology based on the Bundestag resolution 50EE1025 and 50EE1265–66. The authors would like to acknowledge the valuable suggestions of Prof.

Dr. Carsten Dormann concerning the applied methodology and the visualization of the results.

## Appendix 1.

### *Summary of the applied modeling approaches*

#### *Approach 1: Gaussian processes*

Gaussian processes provide a probabilistic approach for learning generic regression problems with kernels (([Rasmussen and Williams, 2006](#)). The model relates the predictor and the response variables (canopy parameter)  $X \in \mathbb{R}^B$  of the form:

$$\hat{y} = f(X) = \sum_{i=1}^N \alpha_i K(X_i, X), \quad (\text{S4})$$

Where  $X_{i=1}^N$  are the predictor variables applied in the training phase,  $\alpha_i \in \mathbb{R}$  is the weight assigned to them and  $K$  is a function evaluating the similarity between the predictor dataset from the test data and the entire  $N$  training predictor set,  $X_i, i=1, \dots, N$ .

Two main advantages of GPR are 1) estimation of the prediction variance in addition to the mean value and 2) the ability to apply sophisticated kernel functions, since all the hyperparameters can be learned efficiently by maximizing the marginal likelihood in the training set ([Verrelst et al., 2012](#)).

The method is considered as being advantageous in multiple aspects. It uses a weighting strategy for the optimization which is relevant to the predictor variables. Furthermore, the inverse of the specific parameter controlling the spread of the relations for each particular predictor represents the relevance of that predictor variable. That is, the higher this parameter is, the more extended are the relations along that predictor (i.e., containing less informative content).

#### *Approach 2: Random forest*

The regression tree method of RF ([Breiman, 2001](#)) has been reported to be an efficient prediction approach, especially when the number of descriptors is very large ([Svetnik et al., 2003](#)). The algorithm works as follows ([Liaw and Wiener, 2002; Latifi and Koch, 2012](#)):

1. Bootstrap samples are drawn from the original data.
2. For each bootstrap sample, an unpruned regression tree is grown. The best splits are chosen from the randomly-sampled variables at each node.
3. New predictions are made by aggregating the predictions of the total number of trees. In this way, the mode votes from the total trees will be the final predicted value of the respective variable.

As an k-NN imputation method, the distance between target and reference units is calculated as “one minus the proportion of terminal nodes from all regression trees where the target observation is in the same terminal node as the specific reference unit” ([Crookston and Finley, 2008](#)).

#### *Approach 3: k-NN*

Generally speaking, in a k-NN method the weighted mean of response value of the most similar neighbor(s) is assigned to a target unit of interest, where the similarity is defined in a feature space consisted of candidate predictor variables ([Latifi and Koch, 2012](#)). The k-NN estimator can be defined as a weighted average of neighboring response values according to ([Härdle et al., 2004](#))

$$\hat{y}_k(x) = \frac{1}{n} \sum_{i=1}^n \omega_{ki}(x) y_i, \quad (\text{S5})$$

where the weights  $\omega_{ki(x)}$  are defined as

$$\omega_{ki}(x) = \begin{cases} \frac{n}{k} & \text{if } i \in J_x, \\ 0 & \text{otherwise.} \end{cases} \quad (S6)$$

With the set of indices  $J_x = \{i : X_i \text{ is one of the } k \text{ nearest observations to } x\}$  and  $n$  the total number of reference sample plots. Thus, the NN estimator has the character of a kernel estimator in the form of the Nadaraya–Watson estimator with uniform kernel  $K_{d(k)}(u) = 1/2I(|u| \leq d^{(k)})$  and a variable bandwidth  $d^{(k)}=d^{(k)}(x)$ , with  $d^{(k)}(x)$  being the distance between the vector of variables for the target unit and its  $k^{\text{th}}$  neighbor

$$\hat{y}_k(x) = \frac{\sum_{i=1}^n K_{d(k)}(x)(d_i)y_i}{\sum_{i=1}^n K_{d(k)}(x)d_i} \quad (S7)$$

In general, the distance between the target unit with a vector  $x$  of  $p$  variables to any neighbor  $i$  having the  $p$ -dimensional vector of variables  $x_i$  can be measured by

$$d_i = \sqrt{(x - x_i)^T \Omega (x - x_i)} \quad (S8)$$

In the original form of k-NN, the weighting matrix  $\Omega$  is a  $p$ -dimensional identity matrix. That is,  $d_i$  becomes the Euclidian distance. The original k-NN approach is modified when using a different  $\Omega$  which turns out to form e.g., Mahalanobis distance (if  $\Omega$  is set as inverse of the covariance matrix of the predictor variables  $\Omega = (\text{Cov}[X])^{-1}$ ; with  $X$  being the  $(n \times p)$  matrix of auxiliary variables  $i$ ).

#### Approach 4: SVM

The SVM algorithm (Vapnik, 2000) is a well-known machine learning approach which has been widely used for classification/regression based on remote sensing data (e.g., Mountrakis et al., 2011; Pal and Mather 2005). The SVM algorithm implemented in “caret” uses “kernlab” R package (Karatzoglou et al., 2004) which is developed by Cortes and Vapnik (1995) and Vapnik (2000). The idea behind SVM is based on structural risk minimization (Vapnik, 2000). For a linearly separable binary classification ( $y_i \in +1, -1$ ) the SVM classifier is constructed by the hyper planes given by:

$$W^T x + W_0 = \pm 1 \text{ where } x \in R^d \text{ and } W \in R^{d+1} \quad (S9)$$

With the following conditions for all  $i \in [1, d]$

$$\text{minimize}_J(w, \xi) = \frac{1}{2} \|W\|^2 + c \sum_{i=1}^d \xi_i \quad (S10)$$

$$\text{subject to } W^T x + W_0 \geq 1 - \xi_i, \quad (S11)$$

$$W^T x + W_0 \leq -1 + \xi_i, \quad (S12)$$

$$W^T x + W_0 \leq -1 + \xi_i \quad (S13)$$

The non-negative margin errors  $\xi_i$  are known as the slack variables. For a nonlinear classification, the input data is mapped onto a higher dimensional feature space using a non-linear mapping function:

$x \rightarrow \phi(x) \in R^k$  where  $k \gg d$ . (S14) The classification is then performed in the mapped space instead of the input feature space. Mapping of the entire samples to higher dimension is computationally expensive and this is avoided by using a positive definite kernel function.

## References

- Andersen, H.-E., Strunk, J., Temesgen, H., Atwood, D., Winterberger, K., 2011. Using multilevel remote sensing and ground data to estimate forest biomass resources in remote regions: a case study in the boreal forests of interior Alaska. *Can. J. Remote Sens.* 37 (6), 1–16.
- Breidenbach, J., Nothdurft, A., Kandler, G., 2010a. Comparison of nearest neighbor approaches for small area estimation of tree species-specific forest inventory attributes in central europe using airborne laser scanner data. *Eur. J. For. Res.* 129 (5), 833–846.
- Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., Solberg, S., 2010b. Prediction of species specific forest inventory attributes using a nonparametric semi-individual treecrown approach based on fused airborne laser scanning and multispectral data. *Remote Sens. Environ.* 114, 911–924.
- Bright, B.C., Hudak, A.T., McGaughey, R., Andersen, H.-E., Negrión, J., 2012. Predicting live and dead tree basal area of bark beetle affected forests from discrete-return lidar. *Can. J. Remote Sens.* 39 (s1), 99–111.
- Chambers, J.M., Freeny, A., Heiberger, R.M., 1992. In: Chambers, S.J.M., Hastie, T.J. (Eds.), *Analysis of Variance: Designed Experiments*. Wadsworth & Brooks/Cole, Chapter 5 of Statistical Models.
- Clark, M.L., Roberts, D.A., Ewel, J.J., Clark, D., 2011. Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. *Remote Sens. Environ.* 115, 2931–2942.
- Cochran, W.G., 1977. *Sampling Techniques*, 3rd ed. John Wiley & Sons, New York.
- Cocks, T., Jenssen, R., Stewart, A., Wilson, I., Shields, T., 1998. The HyMap airbornehyperspectral sensor: the system, calibration and performance. In: Presented at 1st EARSEL Workshop on Imaging Spectroscopy, Zurich, October 1998.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20 (3), 273–297.
- Dahlke, M., Breidt, F.J., Opsomer, J.D., Van Keilegom, I., 2013. Nonparametric endogenous post-stratification estimation. *Stat. Sinica* 23, 189–211.
- Eckert, S., 2012. Improved forest biomass and carbon estimations using texture measures from Worldview-2 satellite data. *Remote Sens.* 4, 810–829.
- Fassnacht, F., Nueman, C., Förster, M., Buddenbaum, H., Ghosh, A., Clasen, A., Joshi, P.K., Koch, B., 2014. Comparison of feature reduction algorithms for classifying tree-species with hyperspectral data on three central-european test sites. *IEEE J. Select. Top. Appl. Earth Obs. Remote Sens.* 7 (6), <http://dx.doi.org/10.1109/JSTARS.2014.2329390>.
- Gagliasso, D., Hummel, S., Temesgen, H., 2014. A comparison of selected parametric and non-parametric imputation methods for estimating forest biomass and basal area. *Open J. For.* 4 (1), 42–48.
- 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. *J. Plant Physiol.* 160, 271–282.
- Hall, S.A., Burke, I.C., Boox, D.O., Kaufmann, M.R., Stocker, J.M., 2005. Estimating stand structure using discrete-return lidar: an example from low density, fire-prone ponderosa pine forests. *For. Ecol. Manage.* 208, 189–209.
- Heurich, M., Thoma, F., 2008. Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural European beech (*Fagus sylvatica*) and Norway spruce (*Picea abies*) forests. *Forestry* 81 (5), 645–661.
- Hill, T.C., Williams, M., Bloom, A.A., Mitchard, E.T.A., Ryan, C.M., 2013. Are inventory based and remotely sensed above-ground biomass estimates consistent? *PLoS ONE* 8 (9), e74170, <http://dx.doi.org/10.1371/journal.pone.0074170>.
- Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. Kernlab- An S4 package or kernel methods in R. *J. Stat. Softw.* 11 (9), 1–20.
- Katila, M., Tomppo, E., 2002. Stratification by ancillary data in multisource forest inventories employing k-nearest neighbor estimation. *Can. J. For. Res.* 32, 1548–1561.
- 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, 581–590.
- Kuhn, M., Johnson, K., 2013. *Applied Predictive Modeling*. Springer Science+Business Media, New York, pp. 995.
- Kulawardhana, R.W., Popescu, S.C., Feagin, R.A., 2014. Fusion of lidar and multispectral data to quantify salt marsh carbon stocks. *Remote Sens. Environ.*, <http://dx.doi.org/10.1016/j.rse.2013.10.036>.
- Labrecque, S., Fournier, R.A., Luther, J.E., Piercy, D., 2006. A comparison of four methods to map biomass from Landsat-TM and inventory data in western Newfoundland. *For. Ecol. Manage.* 226, 129–144.
- 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.
- 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. *Int. J. Remote Sens.* 33 (21), 6668–6694.
- Latifi, H., Fassnacht, F., Koch, B., 2012. Forest structure modelling with combined airborne hyperspectral and LiDAR data. *Remote Sens. Environ.* 121, 10–25.
- Liaw, A., Wiener, M., 2012. R-package random forest userguide. R Development Core Team. Available at: <http://cran.r-project.org/web/packages/randomForest/randomForest.pdf> (accessed on: 22.04.14.).
- Main-Knorn, M., Moisen, G.G., Healey, S.P., Keeton, W.S., Freeman, E.A., Hostert, P., 2015. Evaluating the remote sensing and inventory-based estimation of biomass in the western carpathians. *Remote Sens.* 3, 1427–1446.
- McRoberts, R.E., Gobakken, T., Næsset, E., 2012. Post-stratified estimation of forest area and growing stock volume using lidar-based stratifications. *Remote Sens. Environ.* 125, 157–166.
- McRoberts, R.E., Næsset, E., Gobakken, T., 2013. Inference for lidar-assisted estimation of forest growing stock volume. *Remote Sens. Environ.* 128, 268–275.

- Næsset, E., Gobakken, T., 2008. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sens. Environ.* 112, 3079–3090.
- Nelson, R., 2010. Model effects on GLAS-based regional estimates of forest biomass and carbon. *Int. J. Remote Sens.* 31 (5), 1359–1372.
- Ott, P., 1999. A Comparison of two Sampling Designs for use in the B.C. Forest Inventory: Stratified Probability Proportional to Size with Replacement (PPSWR) and Multiple Pass Ordered Systematic (MPOS) Sampling. Technical Report. British Columbia Ministry of Forests, Land and Natural Resources Operations, Available Online at: <http://www.for.gov.bc.ca/hts/vri/technical/technical/comparison.pdf>
- Packalén, P., Maltamo, M., 2006. Predicting the plot volume by tree species using airborne laser scanning and aerial photographs. *For. Sci.* 52 (6), 611–622.
- Packalén, P., Maltamo, M., 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sens. Environ.* 109, 328–341.
- Peltoniemi, M., Heikkilä, J., Mäkipää, R., 2007. Stratification of regional soil sampling by model-predicted change in soil carbon in forested mineral soils. *Silva Fenn.* 41, 527–539.
- Penner, M., Pitt, D.G., Woods, M.E., 2013. Parametric vs. nonparametric LiDAR models for operational forest inventory in boreal Ontario. *Can. J. Remote Sens.* 39 (5), 426–443.
- Popescu, S.C., Wynne, R.H., Scrivani, J.A., 2004. Fusion of small-footprint LiDAR and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, USA. *For. Sci.* 50 (4), 551–565.
- 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, 1053–1068.
- R Development Core Team., 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <<http://www.R-project.org>> (accessed on: 21.04.14.).
- Reams, G.A., Smith, W.D., Hansen, M.H., Bechtold, W.A., Roesch, F.A., Moisen, G.G., 2005. The forest inventory and analysis sampling frame. In: Bechtold, W.A., Patterson, P.L. (Eds.), In The enhanced Forest Inventory and Analysis program – national sampling design and estimation procedures. U.S. For. Serv. Gen. Tech. Rep., pp. 11–26, SRS-80f.
- Rosillo Calle, F., De Groot, P., Hemstock, S.L., Wooda, J. (Eds.), 2008. The Biomass Assessment Handbook: Bioenergy for a Sustainable Environment. Earthscan Publishing, London, UK, ISBN 1-84407-285-1.
- Särndal, C., Swensson, B., Wretman, J., 1992. Model Assisted Survey Sampling. Springer-Verlag, New York.
- Sarrazin, M.J.D., van Aardt, J.A.N., Asner, G.P., McGlinchy, J., Messinger, D.W., Wu, J., 2011. using small-footprint waveform LiDAR and hyperspectral data for canopy-level species classification and herbaceous biomass modeling in savanna ecosystems. *Can. J. Remote Sens.* 37, 653–665.
- Sexton, J.O., Bax, T., Siquiera, P., Swenson, J.J., Hensley, S., 2009. Comparison of lidar, radar, and field measurements of canopy height in pine and hardwood forests of southeastern north America. *For. Ecol. Manage.* 257, 1136–1147.
- State Forest Service of Baden-Württemberg, 2009. User manual of the forest inventory program Baden-Württemberg BI2005. [in German].
- Thenkabail, P.S., Enclona, E.A., Ashton, M.S., Legg, C., Jean De Dieu, M., 2004a. Hyperion, IKONOS, ALI and ETM+ sensors in the study of African rainforests. *Remote Sens. Environ.* 90, 23–43.
- Thenkabail, P.S., Enclona, E.A., Ashton, M.S., Van Der Meer, B., 2004b. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sens. Environ.* 91, 354–376.
- Tonolli, S., Dalmonte, 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 Sens. Environ.* 115, 2486–2498.
- 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 J. Photogramm. Remote Sens.* 69, 121–133.
- Wagner, S., Hastie, T., Efron, B., 2014. Confidence intervals for random forests: the jackknife and the infinitesimal jackknife. *J. Mach. Learn. Res.* 15 (1), 1625–1651.
- 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. In: Proceedings of the ISPRS working group VIII/2, Freiburg, October 3–6 2004, pp. 90–95, ISSN 1682–1750.
- Westfall, J.A., Patterson, P.L., Coulston, J.W., 2011. Post-stratified estimation: within-strata and total sample size recommendations. *Can. J. For. Res.* 41, 1130–1139.
- Yu, X., Hyppä, J., Vastaranta, M., Holopainen, M., Viitala, R., 2011. Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. *ISPRS J. Photogramm. Remote Sens.* 66, 28–37.
- Zell, J., 2008. Methoden für die Ermittlung, Modellierung und Prognose der Kohlenstoffspeicherung in Wäldern auf Grundlage permanenter Großrauminventuren. Thesis (PhD). Faculty of Forest and Environmental Studies, University of Freiburg, pp. 162p (in German with English summary).
- Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. The MIT Press, New York.
- Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G., Moreno, J., 2012. Machine learning regression algorithms for biophysical parameter retrieval: opportunities for Sentinel-2 and -3. *Remote Sens. Environ.* 118, 127–139.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Crookston, N.L., Finley, A.O., 2008. YalImpute: an R package for KNN imputation. *J. Stat. Software* 23 (10), 16p.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R-News* 2, 18–22.
- Härdle, W., Müller, M., Sperlich, S., Werwatz, A., 2004. Nonparametric and Semiparametric Models. Springer Series in Statistics, Springer, Berlin.
- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. *ISPRS J. Photogramm. Remote Sens.* 66 (3), 247–259.
- Pal, M., Mather, M., 2005. Support vector machines for classification in remote sensing. *Int. J. Remote Sens.* 26 (5), 1007–1011.
- Vapnik, V.N., 2000. The Nature of Statistical Learning Theory. Springer, New York.