International Journal of Applied Earth Observation and Geoinformation 38 (2015) 229–241

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

Stratified aboveground forest biomass estimation by remote sensing data

Hooman Latifi[a,*], Fabian E. Fassnacht[b,c], Florian Hartig[d], Christian Berger[e], Jaime Hernández[f], Patricio Corvalán[f], Barbara Koch[c]

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

ARTICLE INFO

Article history:
Received 1 November 2014
Received in revised form 23 January 2015
Accepted 27 January 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.

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.

* Corresponding author. Tel.: +49 931 3189638; fax: +49 931 31896380.
E-mail addresses: hooman.latifi@uni-wuerzburg.de (H. Latifi), fabian.fassnacht@kit.edu (F.E. Fassnacht), florian.hartig@biom.uni-freiburg.de (F. Hartig), 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.jag.2015.01.016
0303-2434/

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

Whereas some improvements could be obtained (e.g., Popescu 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 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 systematical 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°24′09″E, 49°03′37″N to 8°25′49″E, 49°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 × 100 m grid. In each plot, trees with (DBH) <10 cm, <15 cm, <30 cm, and ≥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