BIOMASS ESTIMATION BASED ON MULTILINEAR REGRESSION AND MACHINE LEARNING ALGORITHMS IN THE MAYOMBE TROPICAL FOREST, IN THE DEMOCRATIC REPUBLIC OF CONGO

OPELELE, O. M.[1,2,3] – YU, Y.[1,2*] – FAN, W.[1,2*] – CHEN, C.[1,2] – KACHAKA, S. K.[3]

[1]*School of Forestry, Northeast Forestry University, Harbin 150040, Heilongjiang, PR China*

[2]*Key Laboratory of Sustainable Forest Ecosystem Management – Ministry of Education, School of Forestry, Northeast Forestry University, Harbin 150040, Heilongjiang, PR China*

[3]*Department of Natural Resources Management, Faculty of Agricultural Sciences, University of Kinshasa, 117 Kinshasa XI, Mont-Amba/Lemba, Democratic Republic of Congo*

***Corresponding authors**
e-mail: fanwy@163.com, yuying4458@163.com; phone: +86-139-4605-5384

(Received 28th Aug 2020; accepted 19th Nov 2020)

Abstract. Accurate forest aboveground biomass estimations have always been of crucial importance for sustainable forest management. However, a choice of the suitable statistical modelling method and predictor variables from remotely sensed data remains the keystone for providing accurate aboveground biomass estimates. The present study intended to compare the potential of four modelling techniques, including RandomForest, Support vector machine, multilinear regression, and K-nearest neighbour for estimating aboveground biomass using vegetation indices, spectral information, and both vegetation indices and spectral bands. The results have revealed that machine learning algorithms provide better results than the multilinear regression method. Indeed, the multilinear regression method produced the lowest R^2 and the greatest RMSE. Besides, the RandomForest performed better by providing accurate results compared to other machine learning algorithms. However, comparing the three sets of predictors, the vegetation indices have yielded accurate results of aboveground biomass and the strongest modelling power. Our results have also revealed that the RF is the best choice for predicting aboveground biomass for the purpose of reducing over- or under-estimation problems. This study has demonstrated the potential of the machine learning algorithms in predicting aboveground biomass in the tropical forest, using freely remotely sensed data derived from sensors with medium spatial resolution.
Keywords: *aboveground biomass, Landsat 8 Operational Land Imager, Mayombe, machine learning*

Introduction

Forest ecosystems store up to 80% and 40% of aboveground and underground carbon, respectively (Mohammadi et al., 2017) and can strongly contribute to mitigating effects of climate change (Moroni, 2013; Caputo, 2009; Brown et at., 1996; Zhang et al., 2014). Numerous studies have shown that biomass is an essential parameter for the carbon sequestration description. However, forest programs based on reducing carbon emissions require the accurate estimation of forest biomass.

In general, the biomass comprises of aboveground biomass (AGB) and belowground biomass (BGB) (Lu, 2006). Therefore, due to the difficult works related to the acquisition and calculation of the BGB, several studies are focused principally on AGB.

Forest biomass has already been measured using several allometric methods based on numerous tree measurements (Manyanda et al., 2019; Mohammadi et al., 2017; Vashum and Jayakumar, 2012). Therefore, forest biomass estimation relying on field inventory

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

is more expensive, arduous, and unrealizable in unreachable areas, thereby making it practicable only in relatively small and accessible areas. Currently, it is possible to optimize this inventory work and reduce the field measurement costs using techniques combining remote sensing technology and field inventories. Indeed, the remote sensing technology, with its capabilities of providing updated information on large areas, has been successfully used for spatial distribution and temporal variation of forest biomass. Several studies have reported that remote sensing variables are useful predictors of biomass because of a strong correlation between biomass and reflectance at different wavelengths (Phua and Saito, 2003; Lu et al., 2004; Zheng et al., 2004). Also, McRoberts et al. (2013) have stated that biomass models using remotely sensed data produce more accurate results than other traditional models. Previously, much research has been completed to estimate AGB in forest ecosystems (Zhang et al., 2014; Dixon et al., 1994; Saatchi et al., 2009; Pflugmacher et al., 2014; López-Serrano et al., 2016; Zhu and Liu, 2015; Glenn et al., 2016). In this research, different remote sensing data (spectral bands and variables derived from spectral bands) and various modelling techniques have been used. However, the remote sensing technology for modelling aboveground biomass implies several main issues. Some of them are related to the remotely sensed derived spectral information that is used as predictor variables (Wang et al., 2013; Lu et al., 2016; Frazier et al., 2014), while others are related to the suitable statistical modelling approach (Shao et al., 2016; Alrababah et al., 2011). According to Lu et al. (2006), two categories of techniques have been used for modelling forest biomass, including parametric and nonparametric methods.

The parametric methods are related to statistical regression, such as linear regression (Lu et al., 2016). In fact, multilinear models are frequently applied to estimate aboveground biomass. However, the relationships between aboveground biomass and predictor variables derived from remotely sensed data might not be linear; consequently, it can lead to overestimation or underestimation problems for small or high aboveground biomass values (Zhao et al., 2016). Therefore, much research has been conducted to examine the use or the potential of nonparametric algorithms, including support vector regression, K-nearest neighbour, and random forest (Li et al., 2014; Vauhkonen et al., 2010; Lu et al., 2016; Gleason and Im, 2012). However, Kumar and Mutanga (2017) have mentioned that among various AGB modelling methods based on different remotely sensed and field data, it is difficult to state or declare that there is one more suitable model than others without assessing their performance separately. In fact, in the field of aboveground biomass estimation using remote sensing technology, the result accuracy depends on different factors, including forest types, remote sensing sensors, and topographical features. Similarly, Feng et al. (2017) have reported that no single modelling method has been determined to be the best for predicting aboveground biomass. Also, Fassnacht et al. (2014) have demonstrated that in the frame of modelling forest biomass based on remote sensing technique, the modelling approach is as important as the data type in deriving accurate AGB estimates.

However, among all previous studies that have tackled the main methods for modelling aboveground biomass, it is unclearly known how data types, forest types, and modelling methods affect aboveground biomass prediction results, especially in the Mayombe tropical forest of the Democratic Republic of Congo, where less research has been carried out for aboveground biomass estimation.

In this research, the main objective was to compare different statistical modelling methods for generating estimates of AGB in the Mayombe forest. The machine learning

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

algorithms and multilinear regression method were then used to estimate the aboveground forest biomass based on remotely sensed data and field biomass measurement. After comparing the performance of the four different models, the suitable model was used to produce a forest biomass map of our study area.

Materials and methods

Study area

The study area is located in the central zone of the Biosphere Reserve of Luki (*Fig. 1*), in the Southwestern part of the Democratic Republic of Congo. It is located between 5.5-5.6°S in latitude and 13.08-13.24°E in longitude. Its total land area is estimated at 8347 ha, entirely located in the Congolese Mayombe tropical forest. The region is dominated by tropical forest and humid tropical climate (Aw5, according to Köppen's classification). This climate is characterized by two seasons notably a rainy season of seven months (mid-October to mid-May) and a dry season of five months (mid-May to mid-October). The annual average temperature and the annual average precipitation are 28.8 °C and 1032.72 mm, respectively. The vegetation is dominated by the primary forest.

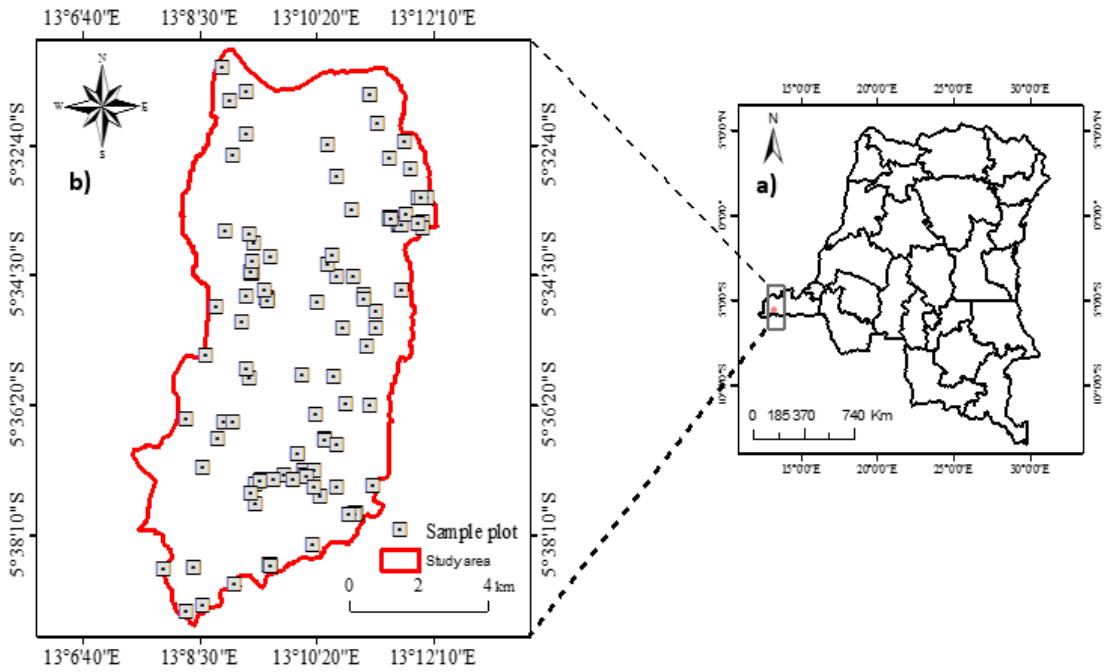

Figure 1. *Study area location and forest sample plot. (a) The Democratic Republic of Congo, (b) the study area*

Data collection from the field and aboveground biomass estimation

One hundred fourteen square plots with size 30 × 30 m (corresponding to a Landsat pixel) were installed in the forest from June to July of the year 2019. We recorded the plot centre geographic coordinates using GPS (global positioning system) receivers. For the purpose of reducing the GPS horizontal locational error (~5 to 10 m), we considered

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

final plot positions based on the criterion that the forest structure and composition in a 10-m buffer around the plot are the same as within the plot. The tree height (m) and tree diameter at breast height (cm) were measured on each tree inside of each sample plot. Also, the name of each tree with diameter at breast height (DBH) greater than 10 cm was reported. To compute the aboveground biomass of individual trees using diameter and tree height, allometric models were used. Several allometric models have been developed for the tropical forest. However, due to the literature review, no allometric model for estimating aboveground biomass is available locally for the Mayombe tropical forest of the Democratic Republic of Congo. Therefore, the allometric equation of Fayolle et al. (2013) was adapted to convert field data to AGB per tree. The predicted tree aboveground forest biomass within each plot was summed to represent plot biomass (expressed in Mg/900 m^2). After that, the expansion factor was used to calculate the AGB per hectare for each plot.

Remote sensing dataset and preprocessing

Surface reflectance Landsat image of 2019 (the 18th of June), with 30 m resolution (path 183 and row 64), was acquired through the USGS Earth Explorer (earthexplorer.usgs.gov). The image was cloud-free and was corrected for atmospheric and topography conditions by the provider. Using the spectral bands (red, blue, and near-infrared bands), we calculated various vegetation indices (*Table 1*), which were used to estimate the aboveground biomass of the area under study. In this paper, based on the correlation between different spectral variables and biomass, numerous machine learning algorithms and a multilinear regression model were performed to predict aboveground biomass. Based on the geographic coordinates of the centre of each sample plot, the spectral variable values, were calculated within the area of each plot using R software; then, we established a database of predictors (vegetation indices and spectral bands) versus biomass values.

Table 1. *Information on remote sensing variables*

Image spectral information	**Band 2 (Blue), Band 3 (Green), Band 4 (Red), Band 5 (NIR), Band 6 (SWIR1), Band 7 (SWIR2)**
Vegetation indices	**Equation**
Soil adjusted vegetation index (SAVI)	((NIR – R) / (NIR + R + L)) × (1 + L) (Eq.1)
Normalized difference vegetation index (NDVI)	NIR – R / NIR + R (Eq.2)
Ratio vegetation index (RDVI)	$(NDVI \times DVI)^{0.5}$ (Eq.3)
Optimized soil-adjusted vegetation index (OSAVI)	(NIR – R / NIR + R + L) × (1 + L) (Eq.4) where *L* is the soil brightness correction factor; *L* = 0.5 works well in most situations and is the default value
Simple ratio (SR)	NIR / R (Eq.5)
Modified soil adjusted vegetation index (MSAVI)	(2 * NIR + 1 – sqrt ((2 * NIR + 1)2 – 8 * (NIR – R))) / 2 (Eq.6)
Difference vegetation index (DVI)	NIR – R (Eq.7)
Enhanced vegetation index (EVI)	5 × (NIR – R) / (NIR + 6 R – 7.5 Blue + 1) (Eq.8)

NIR: near-infrared; R: red band

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

Among the four models under study, we selected the best model based on the performance of each of them in terms of predictive power and the accuracy in AGB estimates. Then, we performed the selection of the most important variables using the best model. Finally, the best model was used to produce the map of the AGB spatial distribution in the study area.

Modelling methods

In this study, four different modelling methods were tested to estimate forest AGB including RandomForest (RF), Support vector regression (SVR), multilinear regression (MLR), and K-nearest neighbour (KNN).

MLR is one among parametric prediction methods commonly used to predict forest AGB using remotely sensed data (Zhu and Liu, 2015; Fassnacht et al., 2014; Lu et al., 2016; Zhao et al., 2016). For this study, we used the ordinary least squares regression method to predict forest aboveground biomass values. However, compared to the machine learning algorithm, the linear regression approach depends on certain assumptions, including the linearity in the relationship between explained and explanatory variables, independence, and normal distribution of errors with a mean value of zero and constant variance. Therefore, the non-respect of these assumptions leads to their violation. Thus, the method is less flexible when facing nonlinear problems (Li et al., 2017), and cannot adequately handle the multicollinearity problem (Ju et al., 2008).

The RF algorithm has been widely used to predict aboveground biomass (Avitabile et al., 2012; Chen, 2015; Pflugmacher et al., 2014; Vauhkonen et al., 2010; Hudak et al., 2012; Tanase et al., 2014). A random forest (RF) algorithm is a tree-based modelling method using a set of rule-based decisions to assess the relationships between a response variable and its predictor variables (Gleason and Im, 2012). This method can generate a large number of small trees built through a different randomly permuted sample from the input dataset (Breiman, 2001). The target data are categorized through two offspring at each node split to maximize homogeneity, and the best split is selected. Finally, the target data for each tree are achieved using bootstrap resampling (Were et al., 2015). Applying unique tree bagging and selection of a random subset of covariates results in minimization of within-group variance and overcoming the over-fitting problem (Park et al., 2016).

The nearest neighbour approach is one of the nonparametric methods used in remote sensing technology (Shataee, 2013), to predict the values of variables using the information of its neighbours (Cover and Hart, 1967). With these techniques, predictions are computed as linear combinations of observations for population units in a sample that are similar or nearest in the space of auxiliary variables to population units requiring predictions (Chirici et al., 2016). The performance of this algorithm depends on the number of neighbours retained by the model (López-Serrano et al., 2016).

The support vector machine (SVM) algorithm states that each ensemble of predictor variables has a unique relationship to the response variable, and sets of explanatory variables can be used to identify the rules to predict a response variable from a set of predictor variables (Mountrakis et al., 2011). It changes the input dataset into a multidimensional hyperplane space by using a kernel function to separate groups of input data with similar response variables to predict a response variable (Were et al., 2015). Indeed, hyperplanes are multidimensional space. In consequence, each explanatory variable is represented by axes from which hyperplanes are built. In that space, the explained variables are placed by projecting it following its explanatory

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

variable values. The SVM applies support vectors to assign each target to a well-fragmented space (Görgens et al., 2015). The main idea behind SVM is to minimize structural risk and moderate the overfitting problem (Latifi et al., 2015).

Model development and verification of estimation models

In the field of biomass estimation, the assessment of the model's performance and accuracy are of crucial importance in terms of selecting a suitable model (Mayer and Butler, 1993). Thus, the k-fold cross-validation approach was considered to examine the performance of the different models. In this research, the 10-fold cross-validation approach was mainly applied as it involves the random partitioning of the original dataset into k subsets (10) with equal size. Among k subsets, every single subset should be held out and used as testing data, while the others k subsets are using as training data. The procedure has a k-times number of repetitions. Also, every k subset is used one time for testing the model. Then, the results are averaged depending on the k number, to provide an overall accuracy. This technique provides many advantages, of which, all instances can be used for validating and training the model. In addition to the cross-validation method, number of validation measures (*Eqs. 9-11*), including the root mean square percentage error (RMSPE), root mean squared error (RMSE), R2, were calculated to evaluate the model's performance and assess the accuracy of the model using the testing dataset (25%). All statistical analyses were carried out using R software.

Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{\mathrm{n}}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \quad \text{(Eq.9)}$$

Root mean square percentage error (RMSPE):

$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}}{y_i}\right)^2} \times 100 \quad \text{(Eq.10)}$$

Coefficient of determination:

$$\mathrm{R}^2 = 1 - \sqrt{\sum_{i=1}^{n}\frac{(yi-\hat{y}i)\text{^}2}{(yi-\bar{y})\text{^}2}} \quad \text{(Eq.11)}$$

where yi is the observed AGB, ŷ is the predicted AGB, n is the number of plots, and ȳ is the mean observed AGB.

Results

Analysis of correlation between AGB and predictors subsection

Table 2 presents the Pearson's correlation coefficient between all the explanatory variables and the forest AGB. Among all the vegetation indices, the forest AGB was significantly correlated with the NDVI, SR, and RDVI, respectively. For the spectral information, the forest AGB was significantly correlated with the Band4, Band3,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

Band2, Band7, and Band6, respectively. However, it has been noted that the NDVI, SR, and RDVI had the highest positive correlation coefficient values among all the predictor variables retained for the present study. At the same time, all spectral bands were negatively correlated with AGB.

Table 2. *Pearson's correlation coefficients between independent variables and AGB*

Statistical parameters							
Variables	**r**	**Variables**	**r**	**Variables**	**r**	**Variables**	**r**
DVI	-0.069	MSAVI	0.0005	Band2	-0.42***	Band6	-0.29**
SAVI	0.021	OSAVI	0.13	Band3	-0.47***	Band7	-0.35**
EVI	-0.020	SR	0.53***	Band4	-0.52***		
NDVI	0.81***	RDVI	0.42***	Band5	-0.12		

Level of significance, *** < 0.001; ** < 0.01; r, correlation coefficient

Field-based AGB estimates and predicted AGB estimates

Table 3 presents the descriptive statistics of the field-level forest aboveground biomass and predicted aboveground biomass. *Table 3* indicates that the observed AGB biomass ranged between 192.22 to 301.11 t ha^{-1}, with a mean value of 250.34 t ha^{-1}, and a standard deviation of 29.39 t ha^{-1}. For the predicted biomass, it has been found some slight differences among different models and data sources. Also, the predicted biomass was slightly greater or smaller than observed biomass, implying the overall underestimation or overestimation problems.

Table 3. *Plot-level aboveground and predicted biomass statistics of 114 plots*

Data	Model	Min (t/ha)	Max (t/ha)	Mean (t/ha)	SD (t/ha)
Vegetation indices	MLR	187.77	286.91	251.26	26.29
	SVM	204.10	290.53	251.41	25.61
	RF	199.32	290.92	250.39	28.15
	KNN	206.00	287.11	250.23	26.21
Spectral information	MLR	193.98	279.44	250.03	18.79
	SVM	204.90	284.33	248.61	21.53
	RF	209.90	282.19	251.03	20.92
	KNN	210.89	278.00	250.42	20.27
Combination VI and SI	MLR	32.19	291.76	242.23	54.12
	SVM	204.11	289.59	251.44	25.04
	RF	200.74	291.75	250.62	27.91
	KNN	212.04	279.01	251.08	22.80
Field data		192.22	301.11	250.34	29.39

SD, Min, and Max represent standard deviation, minimum, and maximum

AGB estimates based on the machine learning and multilinear regression models

The performance assessment results of the different machine learning algorithms and multilinear regression based on different datasets, including vegetation indices, spectral

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

information, and combination of both vegetation indices and spectral information, are presented in *Table 4*. Overall, the machine learning algorithms provide the best prediction performance of aboveground biomass compared to the multilinear regression method, but each algorithm has its performance in predicting aboveground biomass using different data sources. For example, for the support vector machine method, the vegetation indices-based predictors and the combination of both vegetation indices and spectral information produced the highest R^2 values and smallest RMSE and RMSEPE value; spectral information-based variables. For the K- nearest neighbours and RF algorithm, the conclusion is similar to the support vector machine, but for the RF, the R^2 value between vegetation indices-based variables and combination of both vegetation indices and spectral information-based variables were similar. For multilinear regression method, the highest R^2 was reached with vegetation indices-based variables. However, spectral information-based variables provide the greatest RMSE and RMSPE values compared with vegetation indices-based variables and a combination of both vegetation indices and spectral information, except for multilinear regression that has the greatest value with the combination of all predictors. Thus, in general, the vegetation indices-based predictors produced the most accurate results of AGB estimation regardless of which models were used, while the spectral information-based predictors yielded poor performance. The combination of both vegetation indices and spectral information could not improve the modelling performance. According to these results, the RandomForest algorithm provides the most accurate results in predicting biomass, but the accuracy decreases when using spectral information-based predictors. Besides, the Support vector machine yielded the best estimation of the aboveground biomass with spectral information-based predictors. Using the RF algorithm, the most important variables selected were NDVI, Band4, RDVI, SR, Band2, Band6, Band7, Band3, and SAVI, in decreasing order of importance (*Fig. 2*).

Underestimation and overestimation measure based on the machine learning and multilinear regression models

The goodness of fit can be visualized with the scatterplots showing the linear relationships between the predicted AGB and observed AGB from the sample plots (*Fig. 3*).

After computing the overestimation and underestimation mean values with regard to the mean observed values of aboveground biomass (*Table 5*), it was found that spectral data have much more significant overestimation and underestimation mean values than vegetation indices data and combination of both spectral information and vegetation indices data. The vegetation indices data have the smallest overestimation and underestimation mean values. The fact of combining both spectral information data and vegetation indices data improves the overestimation and underestimation problems regardless of which modelling approaches were used, compared to when the models use only spectral information. In view of the modelling algorithms based on the three datasets, in general, the machine learning algorithms have much smaller overestimation and underestimation problem than multilinear regression regardless of which data sources were used. Comparing the performance of machine learning algorithm in dealing with this issue, it was found that the RF and SVM have much smaller overestimation and underestimation problems than the KNN algorithm. However, the RF algorithm has produced the best performance with the lowest overestimation and underestimation problems.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

Table 4. *Comparison of AGB prediction performance*

Data	Model	RMSE (t/ha)	RMSE-CV (t/ha)	RMSPE (t/ha)	R^2
Vegetation indices	MLR	17.48	2.12	0.06	0.60
	SVM	16.38	1.32	0.06	0.65
	RF	10.22	1.08	0.04	0.86
	KNN	14.75	2.05	0.06	0.72
Spectral information	MLR	27.89	2.25	0.07	0.15
	SVM	24.72	2.04	0.10	0.20
	RF	26.55	2.13	0.11	0.08
	KNN	27.44	2.13	0.11	0.01
Vegetation indices and spectral information	MLR	89.13	1.51	1.58	0.25
	SVM	16.95	1.42	0.07	0.62
	RF	10.44	1.06	0.04	0.86
	KNN	18.47	1.46	0.07	0.55

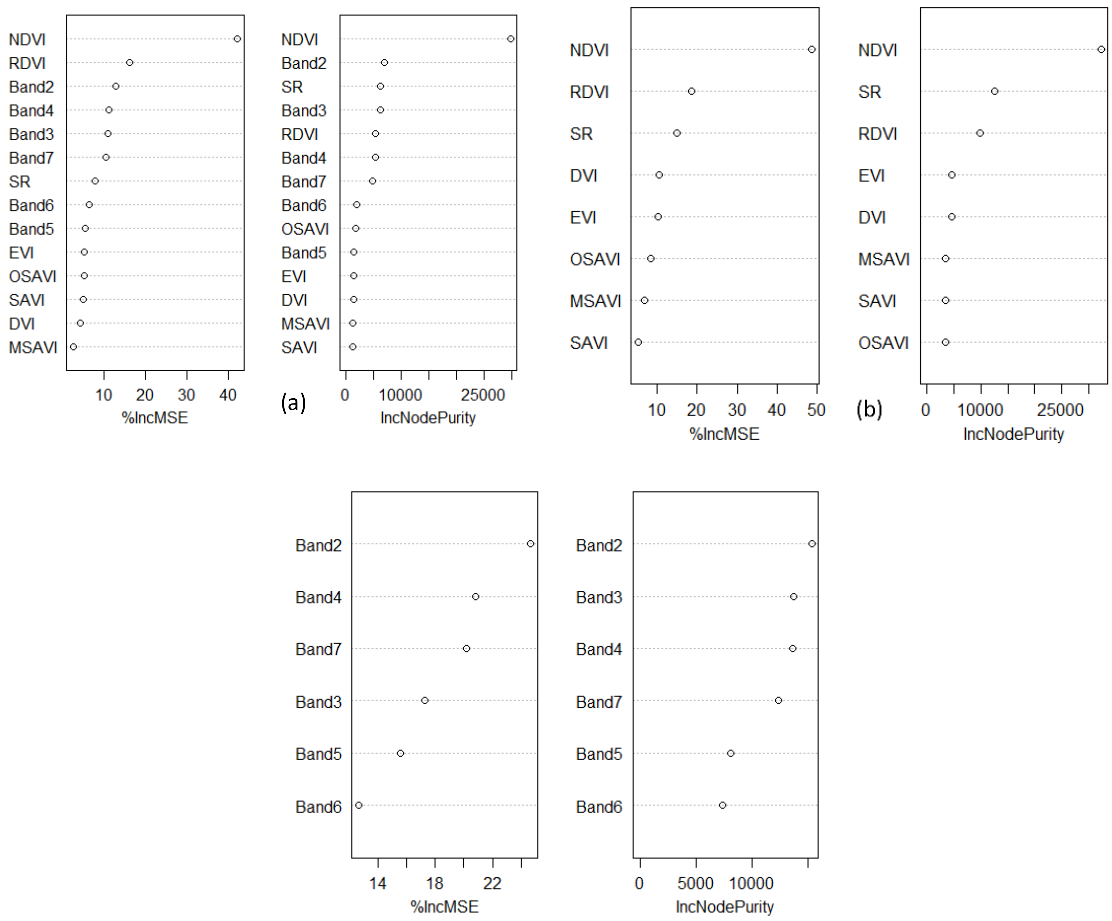

Figure 2. *Important variable ranking by Random forest algorithm. (a) RandomForest with a combination of both vegetation indices and spectral information; (b) RandomForest with vegetation indices; (c) RandomForest with spectral information*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

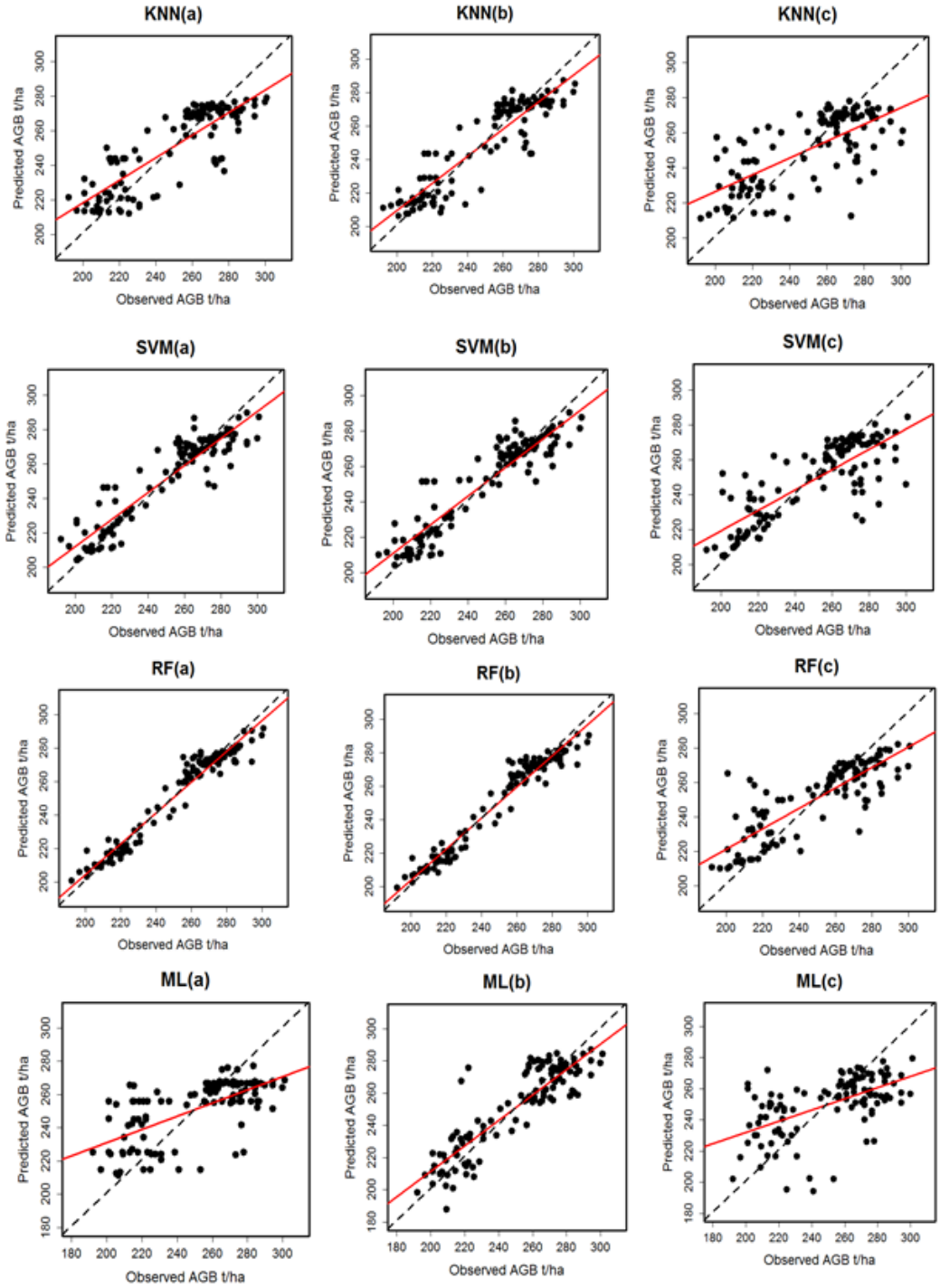

Figure 3. *Observed AGB versus predicted AGB by KNN, SVM, RF and ML. The solid line indicates the optimal regression of observed versus predicted AGB and the dashed line indicates the 1:1 line of perfect agreement. (Note: (a) represents VI and SI, (b) represents VI and (c) represent SI*

AGB mapping using RF algorithm

Based on the RF algorithm, the suitable algorithm in this study concerning its best performance, as shown in *Table 4* (R^2 = 0.86 and RMSE = 10.22 t ha^{-1}), the map of the

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

forest AGB was generated. *Figure 4* presents the spatial distribution of the forest AGB within our study area. The prediction results produced different statistics, notably 190.32 t ha^{-1}, 274.02 t ha^{-1}, 302.8 t ha^{-1}, respectively minimum, mean, and maximum predicted biomass value.

Table 5. *A comparison of mean values of overestimation or underestimation from different data and algorithms*

Data	Model	Overestimation	Underestimation
Vegetation indices	MLR	11.88	9.65
	SVM	8.74	7.45
	RF	5.17	5.62
	KNN	8.67	11.35
Spectral information	MLR	18.96	18.92
	SVM	10.62	14.08
	RF	11.79	12.55
	KNN	15.15	17.88
Combination VI and SI	MLR	9.81	23.13
	SVM	8.52	8.05
	RF	5.15	5.34
	KNN	11.94	14.11

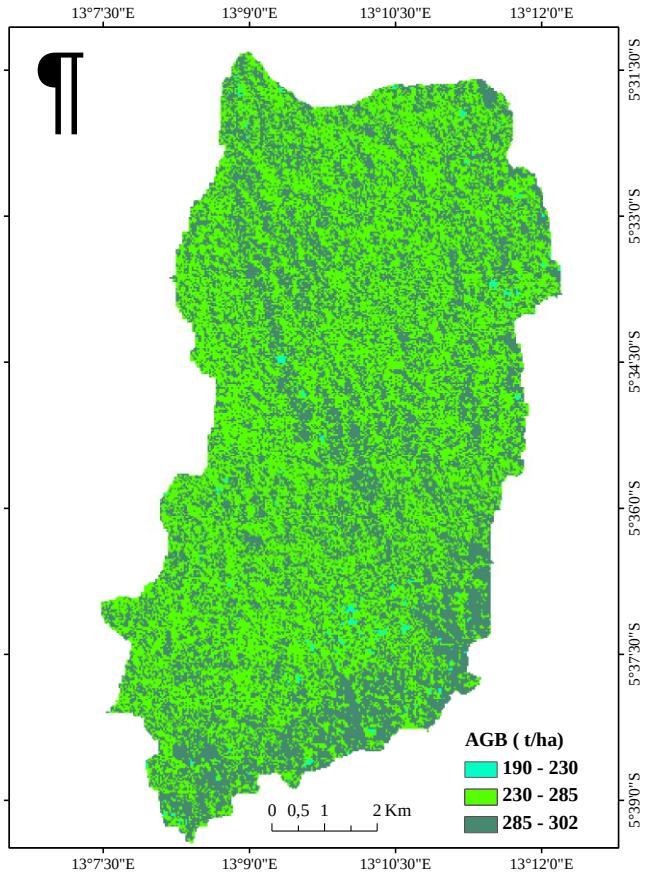

Figure 4. *Aboveground biomass of the study area using the RF algorithm*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

Discussion

Satellite images with medium resolution represent an alternative method to estimate aboveground biomass, especially where the acquisition of hyperspectral images still has serious problems. In fact, up to now, the acquisition of hyperspectral images constitutes a severe problem due to their high costs, limited accessibility or availability, and their manoeuvrability. However, Forest aboveground biomass estimation using remote sensing approaches remains a significant challenge requiring numerous studies to found related solutions concerning different modelling approaches and remote sensing data. According to GAO et al. (2018), the AGB model performance depends on remote sensing data, modelling algorithms, and forest type. Lu (2006) has stated that in the field of the aboveground biomass estimation based remote sensing technique, researchers should take great attention regarding many parameters, among others, the remote sensing variables and the modelling technique or algorithms. In this study, we compared the potential of four methods, including RandomForest, K-nearest neighbour, support vector machine, and multilinear regression, to predict the aboveground biomass of tropical forest using the Landsat 8 multispectral OLI.

In general, the machine learning algorithms have demonstrated the higher ability to predict aboveground biomass in comparison with the multilinear regression method, regardless of data sources that were used (vegetation indices, spectral bands, and combination of both). Indeed, the multilinear regression method yielded the lowest R^2 and the greatest RMSE compared to all other models. Our results are in line with those found by Sadeghi et al. (2018) and Pandit and Dube (2018). Indeed, comparing the performance of the RF machine learning algorithm and the multilinear regression, the authors have reported that the RF algorithm provides high accuracy than multiple linear regression for mapping aboveground biomass. Another study conducted by Feng et al. (2017) has demonstrated the strong ability of RF and SVR to yield better aboveground biomass estimation than multilinear regression. Thus, many reasons were developed to explain the low accuracy of multilinear regression compared to the machine learning algorithm. Numerous studies have pointed out that the weaker performance of the linear regression method is based on the fact of the complexity and non-linearity between remote sensing-based variables and aboveground biomass (Baccini et al., 2004; Foody et al., 2003; Muukkonen and Heiskanen, 2005). Gao et al. (2018) stated that the multiple linear regression method was an essential tool of modelling aboveground biomass, especially for the biomass range of 40–120 Mg/ha. Another reason is that the linear regression method is less flexible when facing nonlinear problems (Li et al., 2017), and cannot adequately handle the multicollinearity problem (Ju et al., 2005). Still, according to Safari et al. (2018), when the biomass value is lower than saturation values, such as 150-ton⁄ha, as stated by Feng et al. (2017) and Gizachew et al. (2016), and when the relationship between Landsat-derived predictors and observed aboveground biomass is expected linear, the linear regression approach can be more efficient. In our study, the biomass range is 192.22–301.11 t/ha, greater than 40–120 Mg/ha. This can explain why the linear regression performed poorly than the machine learning algorithm.

However, it is essential to point out that the RF algorithm improves the forest aboveground biomass estimation. Indeed, the RF algorithm provided better results in AGB estimation regardless of the data source that was used. Indeed, recently, numerous studies have successfully revealed the effectiveness of the RandomForest algorithm to estimate aboveground biomass, and their results are similar to ours. Supporting our

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

results, Liu et al. (2017) reported a significant improvement of the RandomForest algorithm compared with the multilinear regression and support vector machine. In this study, the RandomForest algorithm has produced better performance in terms of accuracy and predictive power of the model (R^2 = 0.95 RMSE = 17.73 Mg/ha). Another study carried out by Latifi et al. (2010), comparing the effectiveness of the K- nearest neighbour methods and the RandomForest algorithm to predict volume and biomass, have found accurate biomass estimation with the RandomForest algorithm, compared to all other nonparametric methods. Similarly, Sadeghi et al. (2018) also revealed that the RF algorithm could produce the accurate results in mapping aboveground biomass at the level of boreal forest stands (R^2 = 0.62, RMSE = 26 Mg ha^{-1}).

In this research, we also analysed the power of the RandomForest algorithm using the important variable selection method, completed using the parameter tuning procedure. The present study has demonstrated the advantage of parameter tuning in predicting AGB employing the RF model. For instance, using the vegetation indices-based variables, the R^2 value augmented from 0.86 to 0.91, while the RMSE value declined from 10.22 to 9.12 t ha^{-1}. According to Kuhn et al. (2008), removing the less important variables from the model has no impact on its performance. However, the results obtained in this study proved the opposite. Hence, Biau and Scornet (2016) have mentioned that due to its complexity, it is difficult to understand the computation of the Random Forest's parameters.

Nevertheless, the RandomForest model provides the minimum number of subgroup input predictors that will improve the effectiveness of the model using the OOB method. Based on OOB data, the RandomForest algorithm produces an estimation error without bias for the testing data. Moreover, several parameters such as tree numbers (Oshiro et al., 2012), splitting at each node of each tree (Grömping, 2009), determine the performance of the algorithm. The sample numbers in each cell, below which the cell is not divided (Kuhn, 2008), but equals the default value of the node size (Tyralis and Papacharalampous, 2017). However, the present study used the default value, as suggested by the literature. The RandomForest algorithm variables have many limit factors, one of which is that the ideal number of predictors with the smallest error is not selected automatically (Adam et al., 2012). According to Grömping (2009), numerous parameters can affect variable importance in RandomForest, one of which is based on the choice of mtry.

Additionally, the author confirmed that it should be better to avoid redundancy to achieve a suitable model in finding a minimum number of variables for a good prediction. Thus, the model does not necessarily need to contain all the appropriate predictors, as long as the results are accurate. In line with Oshiro (2012), in the present study, the RandomForest algorithm identified 15 input predictors as the minimum subgroup with acceptable prediction power.

The percentIncMSE and IncNodePurity were used in ranking predictors concerning their performance to estimate forest AGB provided an improvement in the accuracy of the model. Though the present study accomplished a satisfactory result, it is essential to point out that several parameters influence the effectiveness of the model, including the number of trees (Genuer et al., 2010) and the split numbers. Kuhn and Johnson (2013) recommended considering a minimum number of trees at 1000 for optimizing the parameterization of the model. Likewise, as recommended by Verikas et al. (2011), optimizing variable numbers to divide a node, rather than default values, leads to numerous predictor rankings.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

However, the performance of each machine learning algorithm varied depending on data sources. Indeed, vegetation indices were better related to aboveground biomass than spectral information-based variables. The use of spectral information-based predictors did severely increase the performance of the model. Our findings support those discovered by Pandit et al. (2018) and Lu (2006). In their studies, the authors found that the vegetation indices-based variables were better correlated to AGB than spectral bands. Indeed, some researchers have revealed the strong relationship between vegetation indices-based predictors and biomass (Pearson et al., 1976; Bedard and LaPointe, 1987; Hardisky et al., 1984; Deering and Haas, 1980). For example, Piao et al. (2007) have reported that vegetation indices computed from spectral information reflect the photosynthetic activity of the vegetation and are consequently used for biomass monitoring.

Nevertheless, the fact of combining both vegetation indices-based predictors and spectral bands-based predictors could not improve the modelling performance. Hence, using RandomForest and, according to the increasing order of importance as shown in *Figure 2*, the most crucial vegetation indices-based variables are NDVI, RDVI, SR, DVI, EVI, OSAVI, MSAVI, and SAVI. Our results are in line with those found by Shao and Zhang (2016) and Pandit et al. (2018). The authors have found similar variables for estimating forest biomass by using Landsat 8 OLI sensor.

Finally, the present study found that the biomass was more linked to the vegetation indices than the spectral bands. Indeed, the best models have been found through the use of vegetation indices-based predictors. However, the results based on the linear model were less precise compared to those found using machine learning algorithms. This may be due to the complex structure of tropical forests that sometimes makes linear models ineffective in estimating biomass under these conditions. It should be emphasized that, of all the different existing AGB modelling methods, using various sensors and field data, the choice of the best model should be preceded by a parsimonious evaluation considering the number of parameters of which the sensor type, the forest types, and other environmental conditions. However, as Kumar and Mutanga (2017) reported, several factors influencing uncertainties in the aboveground biomass estimation methods, including vegetation types, landscape types, seasons, and data availability.

Seeing that our study area is located in relatively steep terrain with almost homogeneity in the structure of the canopy, and containing high aboveground biomass, we suggest for further research to investigate the potential of these four methods for predicting forest biomass in various forest conditions, to fill the gap existing in the lack of aboveground forest biomass data.

Conclusions

The present study compared the potential of different machine learning algorithms (RF, KNN, and SVM) and the linear regression method for AGB estimation in the tropical forest using Landsat images and field data. Our study demonstrated that the machine learning algorithm has the potential to estimate AGB with high precision compared to the linear regression method. Comparing the three sets of predictors, the vegetation indices have yielded accurate results and the strongest modelling power in the present study. Our findings show that the RF algorithm is the best alternative for biomass estimation given its best accuracy and high modelling power, regardless of the

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

data source that was used. This study revealed as others the potential of machine learning algorithms based on Landsat images in predicting AGB in the tropical forest, using freely remotely sensed data, allied to field measurement data. This research confirms that machine learning algorithms, especially the RF and SVM, are powerful tools for aboveground biomass using variables derived from sensors with medium spatial resolution. However, we suggest for further research to examine the potential of these machine learning algorithms for predicting forest aboveground biomass in various forest conditions.

Funding. This study was supported by The Fundamental Research Funds for the Central Universities (2572019CP12).

REFERENCES

[1] Adam, E. M., Mutanga, O., Rugege, D., Ismail, R. (2012): Discriminating the papyrus vegetation (Cyperus papyrus L.) and its co-existent species using random forest and hyperspectral data resampled to HYMAP. – Int. J. Remote Sens. 33: 552-569.
[2] Alrababah, M. A., Alhamad, M. N., Bataineh, A. L., Bataineh, M. M., Suwaileh, A. F. (2011): Estimating east Mediterranean forest parameters using Landsat ETM. – International Journal of Remote Sensing 32(6): 1561-1574.
[3] Avitabile, V., Baccini, A., Friedl, M. A., Schmullius, C. (2012): Capabilities and limitations of landsat and land cover data for aboveground woody biomass estimation of Uganda. – Remote Sensing of Environment 117: 366-380.
[4] Baccini, A., Friedl, M. A., Woodcock, C. E., Warbington, R. (2004): Forest biomass estimation over regional scales using multisource data. – Geophysical Research Letters 31: 1-4.
[5] Bedard, J., LaPointe, G. (1987): The estimation of dry green biomass in hayfields from canopy spectroreflectance measurements. – Grass Forage Sci. 42: 73-78.
[6] Biau, G., Scornet, E. (2016): A random forest guided tour. – Test 25: 197-227.
[7] Breiman, L. (2001): Random forests. – Mach. Learn. 45(1): 5-32.
[8] Brown, S., Sathaye, J., Cannell, M., Cannell, M., Kauppi, P. E. (1996): Mitigation of carbon emissions to the atmosphere by forest management. – Commonw. For. Rev. 75: 80-91.
[9] Caputo, J. (2009): Sustainable Forest Biomass: Promoting Renewable Energy and Forest Stewardship. – Policy paper, Environmental and Energy Study Institute, Washington, DC.
[10] Chen, Q. (2015): Modeling aboveground tree woody biomass using national-scale allometric methods and airborne LiDAR. – ISPRS Journal of Photogrammetry and Remote Sensing 106: 95-106.
[11] Chirici, G., Mura, M., McInerney, D., Py, N., Tomppo, E. O., Waser, L. T., Travaglini, D., McRoberts, R. E. (2016): A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. – Remote Sensing of Environment 176: 282-294.
[12] Cover, T., Hart, P. (1967): Nearest neighbor pattern classification. – IEEE Transactions on Information Theory 13(1): 21-27.
[13] Deering, D. W., Haas, R. H. (1980): Using Landsat digital data for estimating green biomass. – NASA Technical Memorandum #80727, Greenbelt, MD.
[14] Dixon, R. K., Brown, S., Houghton, R. E. A., Solomon, A. M., Trexler, M. C., Wisniewski, J. (1994): Carbon pools and flux of global forest ecosystems. – Science 263: 185-190.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

[15] Fassnacht, F. E., Hartig, F., Latifi, H., Berger, C., Hernández, J., Corvalán, P., Koch, B. (2014): Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. – Remote Sens. Environ. 154: 102-114.
[16] Fayolle, A., Doucet, J-L., Gillet, J-F., Lejeune, P. (2013): Tree allometry in Central Africa: testing the validity of pantropical multi-species allometric equations for estimating biomass and carbon stocks. – Forest Ecology and Management 305: 29-37.
[17] Feng, Y., Lu, D., Chen, Q., Keller, M., Moran, E., dos-Santos, M. N., Bolfe, E. L., Batistella, M. (2017): Examining effective use of data sources and modeling algorithms for improving biomass estimation in a moist tropical forest of the Brazilian Amazon. – International Journal of Digital Earth 10(10): 996-1016.
[18] Foody, G. M., Boyd, D. S., Cutler, M. E. (2003): Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. – Remote Sensing of Environment 85: 463-474.
[19] Frazier, R. J., Coops, N. C., Wulder, M. A., Kennedy, R. (2014): Characterization of aboveground biomass in an unmanaged boreal forest using Landsat temporal segmentation metrics. – ISPRS Journal of Photogrammetry and Remote Sensing 92: 137-146.
[20] Gao, Y., Lu, D., Li, G., Wang, G., Chen, Q., Liu, L., Li, D. (2018): Comparative analysis of modeling algorithms for forest aboveground biomass estimation in a subtropical region. – Remote Sensing 10(4).
[21] Genuer, R., Poggi, J. M., Tuleau-Malot, C. (2010): Variable selection using random forests. Pattern Recognit. – Lett. 31: 2225-2236.
[22] Gizachew, B., Solberg, S., Næsset, E., Gobakken, T., Bollandsås, O. M., Breidenbach, J., Zhabu, E., Mauya, E. W. (2016): Mapping and estimating the total living biomass and carbon in low-biomass woodlands using Landsat 8 CDR data. – Carbon Balance and Management 11(1).
[23] Gleason, C. J., Im, J. (2012): Forest biomass estimation from airborne LiDAR data using machine learning approaches. – Remote Sens. Environ 125: 80-91.
[24] Glenn, N. F., Neuenschwander, A., Vierling, L. A., Spaete, L., Li, A., Shinneman, D. J., Philiod, D., Arkle, R., McIlroy, S. K. (2016): Landsat 8 and ICESat-2: Performance and potential synergies for quantifying dryland ecosystem vegetation cover and biomass. – Remote Sensing of Environment 185: 233-242.
[25] Görgens, E. B., Montaghi, A., Rodriguez, L. C. E. (2015): A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. – Comput. Electron. Agric. 116: 221-227.
[26] Grömping, U. (2009): Variable importance assessment in regression: linear regression versus random forest. – Am. Stat. 63: 308-319.
[27] Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., Dextraze, L. (2002): Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. – Remote Sensing of Environment 81: 416-426.
[28] Hardisky, M. A., Daiber, F. C., Roman, C. T., Klemas, V. (1984): Remote sensing of biomass and annual net aerial primary productivity of a salt marsh. – Remote Sens. Environ. 16: 91-106.
[29] Hudak, A. T., Strand, E. K., Vierling, L. A., Byrne, J. C., Eitel, J. U. H., Martinuzzi, S., Falkowski, M. J. (2012): Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. – Remote Sensing of Environment 123: 25-40.
[30] Ju, C., Cai, T., Yang, X. (2008): Topography-based modeling to estimate percent vegetation cover in semi-arid Mu Us sandy land, China. – Computers and Electronics in Agriculture 64: 133-139.
[31] Kuhn, M. (2008): Building Predictive Models in R Using the caret Package. – J. Stat. Softw. 28: 1 26.
[32] Kuhn, M., Johnson, K. (2013): Applied Predictive Modeling. – Springer, New York.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

[33] Kumar, L., Mutanga, O. (2017): Remote sensing of above-ground biomass. – Remote Sens. 9: 935.
[34] 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.
[35] Latifi, H., Fassnacht, F. E., Hartig, F., Berger, C., Hernández, J., Corvalán, P., Koch, B. (2015): Stratified aboveground forest biomass estimation by remote sensing data. – International Journal of Applied Earth Observation and Geoinformation 38: 229-241.
[36] Li, M., Im, J., Quackenbush, L. J., Liu, T. (2014): Forest biomass and carbon stock quantification using airborne LiDAR data: a case study over huntington wildlife forest in the Adirondack park. – IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(7): 3143-3156.
[37] Li, Y., Zhou, L. W., Wang, R. Z. (2017): Urban biomass and methods of estimating municipal biomass resources. – Renewable and Sustainable Energy Reviews 80: 1017-1030.
[38] Liu, K., Wang, J., Zeng, W., Song, J. (2017): Comparison and evaluation of three methods for estimating forest above ground biomass using TM and GLAS data. – Remote Sensing 9(4).
[39] López-Serrano, P. M., López-Sánchez, C. A., Álvarez-González, J. G., García-Gutiérrez, J. A. (2016): Comparison of machine learning techniques applied to Landsat-5 TM spectral data for biomass estimation. – Canadian Journal of Remote Sensing 42(6): 690-705.
[40] Lu, D. (2006): The potential and challenge of remote sensing-based biomass estimation. – International Journal of Remote Sensing 27: 1297-1328.
[41] Lu, D., Mausel, P., Brondızio, E., Moran, E. (2004): Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. For. – Ecol. Manag. 198: 149-167.
[42] Lu, D., Chen, Q., Wang, G., Liu, L., Li, G., Moran, E. (2016): A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. – Int. J. Digital Earth 9(1): 63-105.
[43] Manyanda, B. J., Mugasha, W. A., Nzunda, E. F., Malimbwi, R. E. (2019): Biomass and volume models based on stump diameter for assessing degradation of miombo woodlands in Tanzania. – International Journal of Forestry Research 2019: 1-15.
[44] Mayer, D. G., Butler, D. G. (1993): Statistical validation. – Ecological Modelling 68: 21-32.
[45] McRoberts, R. E., Næsset, E., Gobakken, T. (2013): Inference for LiDAR-assisted estimation of forest growing stock volume. – Remote Sens. Environ. 128: 268-275.
[46] Mohammadi, Z., Limaei, S. M., Lohmander, P., Olsson, L. (2017): Estimating the aboveground carbon sequestration and its economic value (case study: Iranian Caspian forests). – Journal of Forest Science 63(11): 511-518.
[47] Moroni, M. T. (2013): Simple models of the role of forests and wood products in greenhouse gas mitigation. – Australian Forestry 76: 50-57.
[48] Mountrakis, G., Im, J., Ogole, C. (2011): Support vector machines in remote sensing: a review. – ISPRS J. Photogramm. Remote Sens. 66(3): 247-259.
[49] Muukkonen, P., Heiskanen, J. (2005): Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data. – Remote Sensing of Environment 99: 434-447.
[50] Oshiro, T. M., Perez, P. S., Baranauskas, J. A. (2012): How Many Trees in a Random Forest. – In: Perner, P. (ed.) International Workshop on Machine Learning and Data Mining in Pattern Recognition. Springer, Berlin.
[51] Pandit, S., Tsuyuki, S., Dube, T. (2018): Landscape-scale aboveground biomass estimation in buffer zone community forests of Central Nepal: coupling in situ measurements with Landsat 8 Satellite Data. – Remote Sensing 10(11).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

[52] Park, S., Im, J., Jang, E., Rhee, J. (2016): Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions. – Agricultural and Forest Meteorology 216: 157-169.
[53] Pearson, R. L., Miller, L. D., Tucker, C. J. (1976): Hand-held spectral radiometer to estimate gramineous biomass. – Appl. Opt. 15(2): 416-418.
[54] Pflugmacher, D., Cohen, W. B., Kennedy, R. E., Yang, Z. (2014): Using Landsat-derived disturbance and recovery history and LiDAR to map forest biomass dynamics. – Remote Sensing of Environment 151: 124-137.
[55] Phua, M.-H., Saito, H. (2003): Estimation of biomass of a mountainous tropical forest using Landsat TM data. – Canadian Journal of Remote Sensing 29: 429-440.
[56] Piao, S. L., Fang, J. Y., Zhou, L. M., Tan, K., Tao, S. (2007): Changes in biomass carbon stocks in China's grasslands between 1982 and 1999. – Glob. Biogeochem. Cycles 21. DOI: 10.1029/2005GB002634.
[57] R Core Team (2019): R: A Language and Environment for Statistical Computing. – R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.
[58] Saatchi, S., Malhi, Y., Zutta, B., Buermann, W., Anderson, L. O., Araujo, A. M., Phillips, O. L., Peacock, J., Steege, H. T., Gonzalez, G. L. (2009): Mapping landscape scale variations of forest structure, biomass, and productivity in Amazonia. – Biogeosci. Discuss. 6(3): 5461-5505.
[59] Sadeghi, Y., St-Onge, B., Leblon, B., Prieur, J. F., Simard, M. (2018): Mapping boreal forest biomass from a SRTM and TanDEM-X based on canopy height model and Landsat spectral indices. – Int. J. Appl. Earth Obs. Geoinf. 68: 202-213.
[60] Safari, A., Sohrabi, H., Powell, S. (2018): Comparison of satellite-based estimates of aboveground biomass in coppice oak forests using parametric, semiparametric, and nonparametric modeling methods. – Journal of Applied Remote Sensing 12(04).
[61] Shao, Z., Zhang, L. (2016): Estimating forest aboveground biomass by combining optical and SAR data: a case study in Genhe, Inner Mongolia, China. – Sensors 16: 834.
[62] Shataee, S. (2013): Forest attributes estimation using aerial laser scanner and TM data. – Forest Systems 22(3): 484-496.
[63] 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.
[64] Tyralis, H., Papacharalampous, G. (2017): Variable selection in time series forecasting using random forests. – Algorithms 10: 114.
[65] Vashum, K. T., Jayakumar, S. (2012): Methods to estimate above-ground biomass and carbon stock in natural forests - a review. – J Ecosyst Ecogr 2: 116.
[66] Vauhkonen, J., Korpela, I., Maltamo, M., Tokola, T. (2010): Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. – Remote Sensing of Environment 114: 1263-1276.
[67] Verikas, A., Gelzinis, A., Bacauskiene, M. (2011): Mining data with random forests: a survey and results of new tests. – Pattern Recognit. 44: 330-349.
[68] Wang, X., Shao, G., Chen, H., Lewis, B. J., Qi, G., Yu, D., Zhou, L., Dai, L., Dai, L. (2013): An application of remote sensing data in mapping landscape-level forest biomass for monitoring the effectiveness of forest policies in northeastern china. – Environmental Management 52(3): 612-620.
[69] Were, K., Bui, D. T., Dick, Ø. B., Singh, B. R. (2015): A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. – Ecological Indicators 52: 394-403.
[70] Zhang, Y., Liang, S., Sun, G. (2014): Forest biomass mapping of northeastern China using GLAS and MODIS data. – IEEE J. Sel. Top. In Appl. Earth Obs. And Remote Sens. 7(1): 140-152.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377

[71] Zhao, P., Lu, D., Wang, G., Liu, L., Li, D., Zhu, J., Yu, S. (2016): Forest aboveground biomass estimation in Zhejiang Province using the integration of Landsat TM and ALOS PALSAR data. – International Journal of Applied Earth Observation and Geoinformation 53: 1-15.

[72] Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J., Ryu, S. R. (2004): Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. – Remote Sensing of Environment 93: 402-411.

[73] Zhu, X., Liu, D. (2015): Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. – ISPRS J. Photogramm. Remote Sens. 102: 222-231.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_359377