Research Article

Estimating the Aboveground Biomass of an Evergreen Broadleaf Forest in Xuan Lien Nature Reserve, Thanh Hoa, Vietnam, Using SPOT-6 Data and the Random Forest Algorithm

The Dung Nguyen 1,2 and Martin Kappas 1

1Department of Cartography, GIS and Remote Sensing, Georg-August-Universität Göttingen, Göttingen 37077, Germany
2Department of Forest Inventory and Planning, Vietnam National University of Forestry, Xuan Mai, Ha Noi 156200, Vietnam

Correspondence should be addressed to The Dung Nguyen; tnguyen4@gwdg.de

Received 28 February 2020; Revised 7 August 2020; Accepted 12 August 2020; Published 27 August 2020

Academic Editor: Thomas Campagnaro

Copyright © 2020 The Dung Nguyen and Martin Kappas. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Forest biomass is an important ecological indicator for the sustainable management of forests. The aim of this study was to estimate forest aboveground biomass (AGB) by integrating SPOT-6 data with field-based measurements using the random forest (RF) algorithm. In total, 52 remote sensing variables, including spectral bands, vegetation indices, topography data, and textures, were extracted from SPOT-6 images to predict the forest AGB of Xuan Lien Nature Reserve, Vietnam. To determine the optimal predictor variables for AGB estimation, 10 different RF models were built. To evaluate these models, 10-fold cross-validation was applied. We found that a combination of spectral and vegetation indices and topography variables offer the highest prediction results ($R^2 = 0.74$ and RMSE $= 61.24$ Mg ha$^{-1}$). Adding texture features into the predictor variables did not improve the model performance. In addition, the SPOT-6 sensor has the potential to predict forest AGB using the RF algorithm.

1. Introduction

The forest ecosystem is one of the primary sources of carbon storage in the terrestrial ecosystem and constitutes approximately 80% of all living terrestrial biomass [1]. With a massive carbon pool, the forest ecosystem plays an important role in reducing global warming [2, 3]. Most activities related to forest biomass assessments focus on the aboveground biomass (AGB) of living trees because AGB represents the largest amount of total biomass in forests. The accurate assessment and evaluation of forest AGB stores and their spatiotemporal patterns are important for the sustainable management of forests [4, 5]. Estimating AGB is one of the most important steps in measuring and evaluating the carbon stocks and carbon sequestration of forests [6].

In general, field measurements (including destructive sampling or using allometric equations/conversion factors) and remote sensing (RS) are the main methods used to estimate forest AGB [7]. The traditional method based on field measurements is the most accurate but is difficult to cover large areas due to it being expensive, labor-intensive, impractical, harmful to nature, and time-consuming at a large scale [8–10]. Compared to the traditional approach, the RS technique has advantages in its ability to obtain effective and repeatable vegetation information in large areas, especially for remote regions [11]. The forest AGB can be estimated from different RS sensor types, including synthetic aperture radars [11–13], light detection and ranging (LiDAR), and optical sensors. The radar and LiDAR datasets have the advantage of penetrability through the forest canopy to obtain more information in the following, such as trunks and branches, which contain more than 60% of the AGB [11]; this information will help achieve a higher accuracy. The limitations of these types of datasets are their high costs and large data volume requirements to capture the information in large-scale areas [14–17]. Besides radar and LiDAR, very high-resolution (VHR) optical images, such as IKONOS, WorldView-2, GeoEye-1, and SPOT-6 or SPOT-7,
also allow one to estimate AGB by using the empirical relationships between the AGB and RS spectral bands, vegetation indices (VIs), and texture and topographic information with acceptable accuracy. For example, Motlagh et al. [18], Hirata et al. [19], Hussin et al. [20], Karna et al. [21], Li et al. [22], and Gara et al. [23] successfully predicted forest biomass based on those LiDAR and VHR sensors combined with sufficient field data.

Regardless of the RS data sources, there are no RS techniques that are capable of providing a direct measurement of biomass. As a result, biomass prediction accuracy increases when combined with field-sampled data, especially when using machine learning approaches to build biomass models [4, 24–26]. Machine learning algorithms allow one to analyze a large number of predictor variables from remote sensing data, thereby filling in the missing data and reducing the error of the prediction models [27–29]. A wide variety of machine learning algorithms have been employed to estimate AGB, including an artificial neural network (ANN), K-nearest neighbor (KNN), support vector machine (SVM), and random forest (RF). In recent years, RF has been widely used to develop predictive models for AGB at the local, regional, and global areas because it can run efficiently on large datasets with a high accuracy. Furthermore, RF has the ability to determine the importance of variables [30, 31].

Evergreen broadleaf (EB) forests are estimated to cover more than 57% of national forests [32–35] and harbour approximately 44% of the total forest carbon stock in Vietnam [35]. This forest type plays an important key role in ecosystem carbon sequestration in Vietnam. However, in Vietnam, there have only been a few studies on forest carbon estimation with the integration of RS techniques, especially using VHR sensors. For instance, Dang et al. [36] used Sentinel-2 satellite images, combined with field-measured data, to estimate the AGB in Yok Don National Park. The research of Pham and Brabyn [8] successfully proved the accuracy of predicting the AGB of mangrove forests in Can Gio (73%) by integrating spectral information, vegetation type, texture features, and vegetation indices from SPOT-4 and SPOT-5 images.

The main objective of this study is (i) to test the ability of the spectral and vegetation indices and topographic and texture features derived from SPOT-6 images to predict AGB in combination with field data using the RF algorithm and (ii) to identify the most desirable predictors for AGB estimation.

2. Materials and Methods

2.1. Study Site. This study was conducted at Xuan Lien Nature Reserve, Thanh Hoa, Vietnam, located at 19°52’-20°02’N, 104°58’-105°15’E, which covers 23,404 ha of two forest types in the southwest of Thanh Hoa province. This reserve is bordered by Cao River in the north, the Nghe An province in the south and west, and the Ta Leo and Bu Khong mountains and the confluence of Cao and Chu Rivers to the east (Figure 1). The study area is situated in a belt of mountains from Sam Neua in Laos to the Thuong Xuan and Nhu Xuan districts in Thanh Hoa province, which contain some high peaks (e.g., Ta Leo (1400 m), Bu Cho (1563 m), Bu Hon Han (1208 m), and an unnamed 1605 m peak). The mean temperature is about 23–24°C, and the mean annual rainfall is approximately 1700–1900 mm, which occurs mainly from May to October and accounts for 90% of the total annual rainfall [37]. The main soil in the nature reserve is feralite soil: feralite humus soil in the medium-high mountains (FH), feralite soil in the lowlands (F), and alluvial soil (P) associated with streams or rivers and the valley bottom [37].

The vegetation in the study area was mainly closed evergreen broadleaf forest, which was classified into three forest types [38] based on the classification of Thai Van Trung [39]. The first forest type is distributed from medium to high montane, consisting of mixed coniferous and broadleaf evergreen forests (MCBEV) between 800 m and 1605 m (a.s.l). This forest type is generally undisturbed and dominated by upper storeyed broadleaf tree species from the families of Fagaceae, Lauraceae, Euphorbiaceae, Fabaceae, Magnoliaceae, Dipterocarpaceae, and Sapotaceae [37, 38]. The second forest type is located in the low montane broadleaf evergreen forests (BEV), which are distributed under 800 m a.s.l and have been weakly impacted by human activities. Common species include Leguminosae, Euphorbiaceae, Lauraceae, Rutaceae, Rosaceae, and Meliaceae [37, 38]. The final forest type is secondary forests (SF), which are mainly a mix of Neohouzeaua dullooa, Dendrocalamus patellaris, Bambusa sp., and broadleaf evergreen forest [37, 38].

2.2. Field Data Collection. Field surveys were conducted in February 2015. In total, 180 plots (20 m × 25 m) were sampled and inventoried. These plots were randomly generated in ArcGIS 10.4 and then located in the field using a GPS device with errors up to 5 m. Within the plots, the diameter at breast height (dbh) and the total height (h) of each living tree with dbh greater than 5 cm were measured using a diameter tape and a Vertex Hypsometer, respectively. Tree species were also recorded for each measured tree.

2.3. Aboveground Biomass Estimation. We considered only the aboveground living tree biomass for carbon estimations. Aboveground biomass (AGB) was estimated as the sum of the individual components (stumps, stems, bark, branches, seeds, and foliage) of the individual living trees that were predicted using appropriate allometric equations [6]. These allometric equations were carefully chosen depending on the forest types and the tree or bamboo species available in the input dataset. For the evergreen broadleaf forests, we used the biomass equation developed by Huy et al. [40], which was specifically developed for evergreen broadleaf forests in the North Central region of Vietnam (1). For bamboo forests, we opted for the equation from Vu et al. [41], which was developed for bamboo forests at a national scale (2). For mixed forests of bamboo and evergreen broadleaf forests, both (1) and (2) were used to estimate the total biomass. All of the selected equations above are based on the tree/bamboo diameters (dbh) and total heights (h).
Finally, to synchronize the estimated AGB for each sample plot to the remotely sensed data, the AGB values were prorated and scaled to obtain the per-hectare values.

2.4. Remotely Sensed (RS) Data. Due to the availability, SPOT-6 dataset was opted as RS data in this paper. SPOT-6 is an optical satellite that was developed by Astrium with the capacity to obtain panchromatic and multispectral imagery at spectral resolutions of 1.5 m and 6 m, respectively [42]. Two orthorectified scenes of SPOT-6 images taken on 20 May and 05 December 2013 were obtained for this research. Both image scenes consist of four multispectral bands (blue: 450–520 nm, green: 530–590 nm, red: 625–695 nm, and near-infrared (NIR): 760–890 nm), each with a 6 m spatial resolution and one panchromatic band (450–745 nm) with a 1.5 m spatial resolution [42]. The digital number (DN) of the SPOT-6 images was first used to calculate the radiance data and then convert those data to the reflectance value using atmospheric correction in ENVI 5.4. We applied the FLAASH (Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercube) radiative transfer model to correct the atmospheric interference in each image [43].

The 6 m spatial resolution digital elevation model (DEM) was first created from a topographic map with 5 m contour lines [44] using the “Topo to Raster” interpolation method in ArcGIS 10.4. The topographic data (elevation, slope, and aspect) were then generated from a 6 m DEM.

2.5. Variables for AGB Prediction. To explore the effectiveness of the SPOT-6 sensor for estimating forest AGB, different types of RS features were considered. These features included raw spectral bands, topographic data, vegetation indices (VIs), and texture (Table 1). Based on the coordinates, size, and shape of each sample plot, we created a polygon shapefile using the “rectangles, ovals, and diamonds” plugin in QGIS 1.8.0 [52], which we then overlaid onto the RS data. The values of all pixels within each polygon plot were derived for the four different spectral bands and then averaged for each plot. The extracted values were then
used to calculate the 9 VIs. We used the following vegetation indices, most often used in remote sensing-based studies on forest biomass and its properties [4, 45, 53, 54]: NDVI (normalized difference vegetation index), RVI (renormalized difference vegetation index), DVI (difference vegetation index), MSR (modified simple ratio), SAVI (soil-adjusted vegetation index), OSAVI (optimized soil-adjusted vegetation index), GEMI (global environmental monitoring index), EVI (enhanced vegetation index), GLCM (grey-level co-occurrence matrix), and aspect, were also considered as factors affecting the forest’s structure, composition, and distribution [55–57]. The texture feature calculations were carried out using PCI Geomatica 2013. These calculations were performed on all images using a 5 × 5 (900 m²) 6-m-pixel window [50]. For each spectral band, eight texture parameters, as per Haralick et al. [51], were calculated. In total, 52 independent variables were used.

### 2.6. Correlation between the AGB and RS Data

The analysis of the relationship between the AGB and RS data was carried out using the RF algorithm that was integrated into the randomForest package in R software [58]. RF is an ensemble machine learning algorithm that has been widely used in biomass modeling, with the advantages of being able to handle a large number of input variables and identify the most significant variables, as well as to reduce or even overcome the overfitting problem and thereby improve model accuracy [8, 59, 60]. The RF algorithm (RF) was first developed by Breiman [30]. This ensemble learning method generates many decision trees from a randomly selected sample via bootstrapping, known as a training dataset. The features for modeling at each node of the decision trees are also randomly selected. The results are then obtained by averaging the predictions from all decision trees. To estimate the model errors, a subset of samples, comprising the remaining data from the original dataset (called out-of-bag data or OOB data), is used as validation samples. These OOB data are not only used to calculate prediction errors by comparing the predictions from the training dataset with

### Table 1: Variables used in this study for estimating biomass.

| Categories         | Variables       | Algorithm       | References |
|--------------------|-----------------|-----------------|------------|
| Raw spectral features | Blue            | B1 (mean)       |            |
|                    | Green           | B2 (mean)       |            |
|                    | Red             | B3 (mean)       |            |
|                    | NIR             | B4 (mean)       |            |
| Topography         | DEM             |                 |            |
|                    | Slope           |                 |            |
|                    | Aspect          |                 |            |
| Vegetation indices | NDVI            | NDVI = B4 − B5/5 + B3 | [45] |
|                    | RVI             | RVI = B5/B1     | [45] |
|                    | DVI             | DVI = B5 − B1   | [45] |
|                    | RDVI            | RDVI = (B7 − B8)/(B7 + B8) | [46] |
|                    | MSR             | MSR = (B8/B7 − 1)/(√(B7/B8) + 1) | [14] |
|                    | SAVI            | SAVI = (B5 − B3/B5 + B3 + 0.5) × (1 + 0.5) | [47] |
|                    | OSAVI           | OSAVI = (1 + 1.6)(B7 − B5/B7 + B5 + 0.16) | [48] |
|                    | GEMI            | GEMI = n(1 − 0.25n) − B5 − 0.125/1 − B5 | [49] |
|                    | EVI             | EVI = 2.5(B3 − B5/B7 + 6.0B8 − 7.5B1 + 1) | [47] |
| Texture (derived from each spectral band) | GLCM mean (mean) | Mean_j = \sum_{i=0}^{N-1} P_{ij} | [50, 51] |
|                    | GLCM variance (Var) | Var = \sum_{i=0}^{N-1} P_{ij} \times (i − Mean)^2 | [50, 51] |
|                    | Homogeneity (Hom) | Hom = \sum_{i=0}^{N-1} (P_{ij}/1 + (i − j)^2) | [50, 51] |
|                    | Contrast (Con)  | Con = \sum_{i=0}^{N-1} P_{ij} \times (i − j)^2 | [50, 51] |
|                    | Dissimilarity (Dis) | Dis = \sum_{i=0}^{N-1} | P_{ij} \times |i − j| | [50, 51] |
|                    | Entropy (Ent)   | Ent = \sum_{i=0}^{N-1} (−P_{ij} \times Ln(P_{ij})) | [50, 51] |
|                    | Angular second moment (ASM) | ASM = \sum_{i=0}^{N-1} (P_{ij}^2) | [50, 51] |
|                    | Correlation (Cor) | Cor = \sum_{i=0}^{N-1} P_{ij} |(i − Mean)\times (j − Mean)/\sqrt{Var_i \times Var_j} | [50, 51] |
|                    | Inverse difference (InvD) | InvD = \sum_{i=0}^{N-1} P_{ij} for i = 1 | [50, 51] |

Note: NIR: near infrared; DEM: digital elevation model; NDVI: normalized difference vegetation index; RVI: ratio vegetation index; DVI: difference vegetation index; RDVI: renormalized difference vegetation index; MSR: modified simple ratio; SAVI: soil-adjusted vegetation index; OSAVI: optimized soil-adjusted vegetation index; GEMI: global environmental monitoring index; EVI: enhanced vegetation index; GLCM: grey-level co-occurrence matrix; P_{ij} is the probability of values i and j occurring in adjacent pixels in the original image within the window defining the neighborhood; i refers to the digital number (DN) value of a target pixel; j is the DN value of its immediate neighbor; and N is the number of grey levels.
In RF modeling, there are two important training parameters that need specification: \( ntree \) is the number of trees to grow in the forest, and \( mtry \) is the number of randomly selected variables used in each node of the tree. A good RF model, which is built from the desirable values of \( ntree \) and \( mtry \), will have a low root mean square error (RMSE). To find the \( ntree \) value that corresponds to a desirable predictor, different \( ntree \) values varying from 50 to 1000 with an interval of 50 were tested. The final \( ntree \) value was selected based on the stability of the RMSE (see Figure 2). To identify the optimal \( mtry \) values, we used the `tuneRF` function in the `randomForest` package.

To evaluate the importance of each variable, RF defines two measures, which are computed from the OOB data. The first measure is the percent increase in the mean square error (%IncMSE) that was calculated for the prediction of each tree [31]. Higher %IncMSE values indicate a more important predictor. The second measure is the total decrease in node impurities (IncNodePurity), which is the average of the residual sum of squares over all trees when splitting the variables at each node [31]. Higher IncNodePurity values indicate a more important variable. According to Strobl et al. [61], the IncNodePurity method is biased and not recommended for use. Therefore, in this study, we only use the % IncMSE measure to identify the importance of variables.

Overall, 10 RF models were built to determine the most desirable predictor for forest AGB estimation (Table 2).

2.7. Model Validation. For validation, the original data were randomly divided into two separate parts: a training dataset (70%) and a testing dataset (30%). Each RF model’s performance was validated through a 10-fold cross-validation. The validation measures include the adjusted coefficient of determination \( (R^2_{adj}) \) and the root mean square error (RMSE).

3. Results

3.1. Tree AGB Estimation from Field Data. Table 3 and Figure 3 show the results of tree AGB calculations for each forest type at the plot level from field data measurements. The results show that the forest AGB ranges from 18.32 Mha\(^{-1}\) to 543.86 Mha\(^{-1}\). The average AGB estimated for Xuan Lien Nature Reserve was 158.23 Mha\(^{-1}\) for the four forest types. The MCBEV forests had the highest AGB followed by the BEV and SF forests. Secondary forests had the lowest AGB and were mostly mixed bamboo and evergreen forests or developed on abandoned agriculture land.

In total, 189 species from 55 families were recorded in the field. The five most dominant species were *Castanopsis indica*, *Engelhardia roxburghiana*, *Ormosia sp.*, *Fokienia hodginsii*, and *Archidendron balansae*.

3.2. Variable Importance and Variable Selection for the Final RF Models. Because models RF2, RF3, RF4, RF5, and RF6 are a combination of spectral features, vegetation indices, topographic data, and texture features, only models RF1 (all variables), RF7 (spectral variables), RF8 (vegetation indices’ variables), RF9 (topographic variables), and RF10 (texture variables) were used to investigate the importance of the predictor variables. Each RF model was run 100 times to determine the variation of each variable’s importance.
Table 2: Different RF models and their settings.

| Model | Variable combination | Number of variables | ntree | mtry |
|-------|----------------------|---------------------|-------|------|
| RF1   | Spectral, topography, vegetation indices, and texture | 52 | 300 | 26 |
| RF2   | Spectral, vegetation indices, and texture | 49 | 300 | 6 |
| RF3   | Spectral, topography, and vegetation indices | 16 | 300 | 12 |
| RF4   | Spectral and vegetation indices | 13 | 300 | 9 |
| RF5   | Spectral and topography | 7 | 300 | 3 |
| RF6   | Spectral and texture | 40 | 300 | 29 |
| RF7   | Spectral | 4 | 300 | 3 |
| RF8   | Vegetation indices | 9 | 300 | 7 |
| RF9   | Topography | 3 | 300 | 2 |
| RF10  | Texture | 36 | 300 | 12 |

Table 3: Summary statistics for the forest aboveground biomass (AGB) at the plot level.

| No. | Forest type | No. of plots | Min AGB (Mg ha\(^{-1}\)) | Max AGB (Mg ha\(^{-1}\)) | Mean AGB (Mg ha\(^{-1}\)) | Standard deviation (Mg ha\(^{-1}\)) |
|-----|-------------|--------------|--------------------------|--------------------------|-----------------------------|----------------------------------------|
| 1   | MCBEV       | 64           | 40.88                    | 543.88                   | 251.81                      | 125.43                                 |
| 2   | BEV         | 29           | 64.56                    | 303.76                   | 153.36                      | 61.61                                  |
| 3   | SF          | 87           | 18.32                    | 252.01                   | 91.01                       | 53.78                                  |
| Total/average | 180 | 18.32 | 543.88 | 158.23 | 113.36 |

Figure 3: Distribution of the AGB of the field data by dbh in three forest types: medium to high montane mixed coniferous and broadleaf evergreen forests (MCBEVs), low montane broadleaf evergreen forests (BEVs), and secondary forests (SFs).
Figure 4: Ranking of each variable’s importance measured in %IncMSE by running the RF algorithm 100 times for different types of predictors: all variables (model RF1—top 30 variables), spectral reflectance (model RF7), vegetation indices (model RF8), topography (model RF9), and texture features (model RF10).
Among all the variables (model RF1), the most important variables are elevation, spectral band 3 (red), the GLCM mean of the green and blue band, RVI, MSR, NDVI, OSAVI, RDVI, and SAVI (Figure 4, RF1). When using spectral features as predictors, the most important band is band 3 (red) following by band 4 (NIR), band 2 (green), and band 1 (blue) (Figure 4, RF7). Among the nine VIs used for AGB estimation, there is no large difference in the %IncMSE values (Figure 4, RF8), for which DVI and EVI have highest values. If we use only topographic data to predict AGB, elevation has the largest influence followed by aspect and slope (Figure 4, RF9). For texture features, the result from model RF10 reveals that the texture means of the green and blue bands are the two most important variables for AGB estimation (Figure 4, RF10).

**Figure 5:** The number of variables (x-axis) used vs. the RMSE (y-axis) of 10 RF models based on 100-times 10-fold cross-validation. The red line indicates the mean RMSE; the green and blue lines indicate the mean RMSE + SD (standard deviation) or the mean RMSE–SD, respectively.
**Figure 6: Continued.**

(a) $y = 0.12 + 0.84x$

- $R^2_{\text{adj}} = 0.69$
- RMSE = 52.53 (Mg ha$^{-1}$)

(b) $y = -66 + 1.4x$

- $R^2_{\text{adj}} = 0.5$
- RMSE = 80.27 (Mg ha$^{-1}$)

(c) $y = -66 + 1.4x$

- $R^2_{\text{adj}} = 0.56$
- RMSE = 71.24 (Mg ha$^{-1}$)

(d) $y = -66 + 1.4x$

- $R^2_{\text{adj}} = 0.46$
- RMSE = 87.07 (Mg ha$^{-1}$)

(e) $y = 15 + 0.9x$

- $R^2_{\text{adj}} = 0.57$
- RMSE = 67.92 (Mg ha$^{-1}$)

(f) $y = 38 + 0.63x$

- $R^2_{\text{adj}} = 0.2$
- RMSE = 83.93 (Mg ha$^{-1}$)
To select variables for the final models, knowing the importance of each variable is not enough. Each model must have an optimal number of variables, which will improve the model’s accuracy. To obtain the optimal variables, we used a 10-fold cross-validation method (running 100 times). The optimal variables for the 10 models are shown in Figure 5.

3.3. Performance of the Ten RF Models. Figure 6 shows the results of the ten models, which are consistent with the processing done in Section 3.2. Model RF6 shows the lowest result followed by models RF10, RF8, and RF7. Models RF4 and RF9 have similar accuracy and present a slightly lower result than model RF2. Model RF1 presents the highest result but requires 52 variables and resulted in overfitting. Although models RF3 and RF5 have a result lower than model RF1, they require only 7 and 16 variables, respectively, and do not show over- or underfitting in their results.

4. Discussion

The main objective of this study was to test the possibility of using SPOT-6 images for estimating the AGB of evergreen broadleaf forests in Xuan Lien Nature Reserve using the random forest algorithm. Figure 4 (RF1) shows that elevation was the most important variable for predicting AGB. This is mainly because vegetation types strongly vary along the altitudinal gradient within the study area [38]. Similarly, other studies have proven that forest biomass has a significant relationship with vegetation types and elevation [3, 8, 62–64]. The next three most important variables were
red reflectance and the two textures derived from the green and blue bands.

Among the four spectral bands, the red band has the strongest correlation with forest AGB (Figure 4, RF7) following by NIR and the green and blue bands. The possible reason for this result is that red reflectance and NIR reflectance are more sensitive to vegetation characteristics (e.g., tree species or stem volume) than other visible types of reflectance [65]. The performance of the regression was improved when we combined vegetation indices and/or topographical features with spectral band reflectance. Some similar findings were presented by Pandit et al. [4] and Adam et al. [66], who stated that using VIs improves results because VIs diminish the influence of environmental conditions and shadow effects on reflectance.

In this study, we have shown that the AGB and different VIs have a significant correlation with each other. The most useful VI for predicting forest AGB was DVI followed by EVI, RV1, OSAVI, SAVI, MSR, GEMI, NDVI, and RDVI; however, the difference between these VIs was not very high.

Figure 5 clearly shows that the texture features (models RF10, RF6, and RF2—that included texture as a predictor) are less important than other features in AGB estimation. In other words, the accuracy of the model was not improved when using texture as an additional predictor for AGB estimation. This result is similar to that of Pham and Brabyn [8].

5. Conclusions

This study used the RF algorithm for modeling and predicting the forest AGB in Xuan Lien Nature Reserve, Vietnam, using VHR SPOT-6 data combined with field-based data. The results showed a significant statistical relationship between the AGB and the SPOT-6 data. The SPOT-6 data effectively predicted the AGB of the EB forest with $R^2_{adj} = 0.74$ and RMSE = 61.24 Mg ha$^{-1}$. The accuracy of AGB estimation was affected by many factors, among which elevation was indicated to be the most important for AGB models. The random forest model selection of important variables showed that using elevation and vegetation indices and spectral reflectance could significantly improve biomass estimations in evergreen broadleaf forests. The RF algorithm is also suitable for estimating the AGB of evergreen broadleaf forests.

Based on the results of the study above, some future work should be considered. For example, when applying the model from this study to different types of forests in different ecoregions, the topography, spectral reflectance, and texture should be taken into consideration. Although the method introduced in this study is applicable to other forest ecosystems in Vietnam, evaluating forest types is one of the necessary impact predictors.

It is also possible to use other RS data sources or machine learning algorithms, which may have better fit estimations. Therefore, the next step of this study is to compare different machine learning techniques to predict forest AGB using two optical sensor types (SPOT-6 and Sentinel-2 MSI).

In this study, we found that elevation is one of the most important predictors for forest AGB estimation. Therefore, when designing a field survey for forest biomass estimation, the elevation should always be recorded.

Data Availability

The data that support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors are thankful to the support of various foundations. The authors are also grateful to all the forest rangers and local members for their kind support and assistance during the fieldwork. The authors would like to thank the Vietnamese Government for their financial support and Xuan Lien Nature Reserve for providing permission to collect their field data. The publication of this paper was supported financially by the Open Access Grant Program of the German Research Foundation (DFG) and the Open Access Publication Fund of the Georg-August University of Göttingen.

References

[1] F. S. Chapin, P. A. Matson, and H. A. Mooney, *Principles of Terrestrial Ecosystem Ecology*, Springer, New York, NY, USA, 2002.
[2] R. A. Houghton, “Aboveground forest biomass and the global carbon balance,” *Global Change Biology*, vol. 11, no. 6, pp. 945–958, 2005.
[3] P. Vicharnakorn, R. Shrestha, M. Nagai, A. Salam, and S. Kiritiprayoon, “Carbon stock assessment using remote sensing and forest inventory data in savannahket, Lao PDR,” *Remote Sensing*, vol. 6, no. 6, pp. 5452–5479, 2014.
[4] S. Pandit, S. Tsuyuki, and T. Dube, “Estimating above-ground biomass in sub-tropical buffer zone community forests, Nepal, using Sentinel 2 data,” *Remote Sensing*, vol. 10, no. 4, p. 601, 2018.
[5] M. Zhang, H. Du, G. Zhou et al., “Estimating forest above-ground carbon storage in hang-jia-hu using landsat TM/OLI data and random forest model,” *Forests*, vol. 10, no. 11, p. 1004, 2019.
[6] S. Eggleston, L. Buendia, and K. Miwa, *IPCC Guidelines for National Greenhouse Gas Inventories. Intergovernmental Panel on Climate Change*, The Institute for Global Environmental Strategies, Kanagawa, Japan, 2006.
[7] D. Lu, “The potential and challenge of remote sensing-based biomass estimation,” *International Journal of Remote Sensing*, vol. 27, no. 7, pp. 1297–1328, 2007.
[8] L. T. H. Pham and L. Brabyn, “Monitoring mangrove biomass change in Vietnam using SPOT images and an object-based approach combined with machine learning algorithms,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 128, pp. 86–97, 2017.
[9] L. Du, T. Zhou, Z. Zou, X. Zhao, K. Huang, and H. Wu, “Mapping forest biomass using remote sensing and national
forest inventory in China,” *Forests*, vol. 5, no. 6, pp. 1267–1283, 2014.

[10] D. Deb, J. P. Singh, S. Deb, D. Datta, A. Ghosh, and R. S. Chaurasia, “An alternative approach for estimating above ground biomass using resources-2 satellite data and artificial neural network in Bundelkhand region of India,” *Environmental Monitoring and Assessment*, vol. 189, no. 11, p. 576, 2017.

[11] K. M. Bergen and M. C. Dobson, “Integration of remotely sensed radar imagery in modeling and mapping of forest biomass and net primary production,” *Ecological Modelling*, vol. 122, no. 3, pp. 257–274, 1999.

[12] O. Hamdan, I. M. Hasmadi, and H. K. Aziz, “Combination of SPOT-5 and ALOS PALSAR images in estimating above-ground biomass of lowland Dipterocarp forest,” *IOP Conference Series: Earth and Environmental Science*, vol. 18, p. 012016, 2014.

[13] H. Huang, C. Liu, X. Wang, X. Zhou, and P. Gong, “Integration of multi-resource remotely sensed data and allometric models for forest aboveground biomass estimation in China,” *Remote Sensing of Environment*, vol. 221, pp. 225–234, 2019.

[14] J. M. Chen, “Evaluation of vegetation indices and a modified Simple Ratio for boreal applications,” *Canadian Journal of Remote Sensing*, vol. 22, no. 3, pp. 229–242, 2014.

[15] H.-E. Andersen, J. Strunk, and H. Temesgen, “Using airborne light detection and ranging as a sampling tool for estimating forest biomass resources in the upper banana valley of interior Alaska,” *Western Journal of Applied Forestry*, vol. 26, no. 4, pp. 157–164, 2011.

[16] E. Nesset, “Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data,” *Remote Sensing of Environment*, vol. 80, pp. 88–99, 2002.

[17] L. I. Duncanson, K. O. Niemann, and M. A. Wulder, “Estimating forest canopy height and terrain relief from GLAS waveform metrics,” *Remote Sensing of Environment*, vol. 114, no. 1, pp. 138–154, 2010.

[18] M. G. Motlagh, S. B. Kafaky, A. Mataji, and R. Akhavan, “Estimating and mapping forest biomass using regression models and Spot-6 images (case study: hycyanian forests of north of Iran),” *Environmental Monitoring and Assessment*, vol. 190, no. 6, p. 352, 2018.

[19] Y. Hirata, N. Furuya, H. Saito et al., “Object-based mapping of aboveground biomass in tropical forests using LiDAR and very-high-spatial-resolution satellite data,” *Remote Sensing*, vol. 10, no. 3, p. 438, 2018.

[20] Y. A. Hussin, H. Gilani, L. van Leeuwen et al., “Evaluation of object-based image analysis techniques on very high-resolution satellite image for biomass estimation in a watershed of hilly forest of Nepal,” *Applied Geomatics*, vol. 6, no. 1, pp. 59–68, 2014.

[21] Y. K. Karna, Y. A. Hussin, H. Gilani et al., “Integration of worldview-2 and airborne LiDAR data for tree species level carbon stock mapping in Kayar Khola watershed, Nepal,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 38, pp. 280–291, 2015.

[22] W. Li, Z. Niu, X. Liang et al., “Geostatistical modeling using LiDAR-derived prior knowledge with SPOT-6 data to estimate temperate forest canopy cover and above-ground biomass via stratified random sampling,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 41, pp. 88–98, 2015.

[23] T. W. Gara, A. Murwira, and H. Ndaimani, “Predicting forest carbon stocks from high resolution satellite data in dry forests of Zimbabwe: exploring the effect of the red-edge band in forest carbon stocks estimation,” *Geocarto International*, vol. 31, no. 2, pp. 176–192, 2015.

[24] A. Rosenqvist, A. Milne, R. Lucas, M. Imhoff, and C. Dobson, “A review of remote sensing technology in support of the kyoto protocol,” 2019, https://www.sciencedirect.com/science/article/abs/pii/S1462901103000704.

[25] B. Kong, H. Yu, R. Du, and Q. Wang, “Quantitative estimation of biomass of alpine grasslands using hyperspectral remote sensing,” *Rangeland Ecology & Management*, vol. 72, no. 2, pp. 336–346, 2019.

[26] Y. Patil, A. Singh, N. Naik, and S. Unnikrishnan, “Estimation of mangrove carbon stocks by applying remote sensing and GIS techniques,” *Wetlands*, vol. 35, no. 4, pp. 695–707, 2015.

[27] X. Dou, Y. Yang, and J. Luo, “Estimating forest carbon fluxes using machine learning techniques based on eddy covariance measurements,” *Sustainability*, vol. 10, no. 1, p. 203, 2018.

[28] P. M. López-Serrano, C. A. López-Sánchez, J. G. Álvarez-González, and J. García-Gutiérrez, “A comparison of machine learning techniques applied to landsat-5 TM spectral data for biomass estimation,” *Canadian Journal of Remote Sensing*, vol. 42, no. 6, pp. 690–705, 2016.

[29] I. Ali, F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola, “Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data,” *Remote Sensing*, vol. 7, no. 12, pp. 16398–16421, 2015.

[30] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[31] D. R. Cutler, T. C. Edwards, K. H. Beard et al., “Random forests for classification in ecology,” *Ecology*, vol. 88, no. 11, pp. 2783–2792, 2007.

[32] V. T. Phuong, “Forest environment of Vietnam: features of forest vegetation and soils,” in *Forest Environments In the Mekong River Basins*, 1, H. Sawada, N. Chappell, J. V. LaFrankie, A. Shimizu, and M. Araki, Eds., pp. 189–200, Springer, Tokyo, Japan, 2007.

[33] Vietnam Administration of Forestry, “Vietnam forest data sharing system,” 2020, http://maps.vnforest.gov.vn/.

[34] C. T. Ha, *Vietnam National Forest Status of 2012*, Annual Report of Ministry of Agriculture and Rural Development, Hanoi, Vietnam, 2013.

[35] Ministry of Agriculture and Rural Development, “Vietnam’s modified submission on reference levels for REDD+ results based payments under UNFCCC,” 2016, https://redd.unfccc.int/files/vietnam_frl_modified_submission_final_for_posting.pdf.

[36] A. T. N. Dang, S. Nandy, R. Srinet, N. V. Luong, S. Ghosh, and A. Senthil Kumar, “Forest aboveground biomass estimation using machine learning regression algorithm in Yok Don National Park, Vietnam,” *Ecological Informatics*, vol. 50, pp. 24–32, 2019.

[37] T. T. Le, V. C. Le, D. T. Bui et al., *A Feasibility Study for the Establishment of Xuan Lien Nature Reserve, Thanh Hoa Province, Vietnam*, BirdLife International Vietnam Programme, Hanoi, Vietnam, 1999.

[38] L. T. Trai, L. V. Cham, B. D. Tuyen et al., *A Feasibility Study for the Establishment of Xuan Lien Nature Reserve, Thanh Hoa Province, Vietnam*, BirdLife International Vietnam Programme, Hanoi, Vietnam, 1999.

[39] V. T. Thãi, “Thấm thức và vườn Việt Nam (trên quan điểm hệ sinh thái),” in *Lâm Thự 2. Có Sựa ChưaKhoa học và kỹ thuật*, Hanoi, Vietnam, 1978.

[40] B. Huy, K. Kralicek, K. P. Poudel et al., “Allometric equations for estimating tree aboveground biomass in evergreen
broadleaf forests of Viet Nam,” *Forest Ecology and Management*, vol. 382, pp. 193–205, 2016.

[41] T. P. Vu, V. X. Nguyen, T. M. L. Nguyen, and D. T. Phung, *Allometric Equations at National Scale for Tree Biomass Assessment in Vietnam*. Part B—Allometric Equations for Bamboo Forests, Vietnam Academy of Forest Sciences and UN-REDD Vietnam, Hanoi, Vietnam, 2014.

[42] Astrium Services. SPOT 6/7 Imagery-User Guide.

[43] M. L. Clark, D. A. Roberts, J. J. Ewel, and D. B. Clark, “Ecosystems for monitoring vegetation,” *Remote Sensing of Environment*, vol. 55, no. 2, pp. 95–107, 1996.

[44] G. Rondeaux, M. Steven, and F. Baret, “Optimization of soil-adjusted vegetation indices,” *Remote Sensing of Environment*, vol. 51, no. 3, pp. 375–384, 1995.

[45] A. R. Huete, “A soil-adjusted vegetation index (SAVI),” *Remote Sensing of Environment*, vol. 25, no. 3, pp. 295–309, 1988.

[46] G. Rondeaux, M. Steven, and F. Baret, “Optimization of soil-adjusted vegetation indices,” *Remote Sensing of Environment*, vol. 55, no. 2, pp. 95–107, 1996.

[47] B. Pinty and M. M. Verstraete, “GEMI: a non-linear index to monitor global vegetation from satellites,” *Vegetatio*, vol. 101, no. 1, pp. 15–20, 1992.

[48] M. Hall-Beyer, “Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales,” *International Journal of Remote Sensing*, vol. 38, no. 5, pp. 1312–1338, 2017.

[49] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, “Textural features for image classification,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6, pp. 610–621, 1973.

[50] P. Kapusta, B. R. Bohard, and D. Ovals: Plugin for QGIS (ver. 1.1.2), 2012, https://plugins.qgis.org/plugins/rectovalDigit/version/1.1.2/.

[51] L. Cortes, “Estimation of above-ground forest biomass using landsat ETM+, aster GDEM and lidar,” *Forest Research: Open Access*, vol. 03, no. 2, 2014.

[52] M. L. Clark, D. A. Roberts, J. J. Ewel, and D. B. Clark, “Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors,” *Remote Sensing of Environment*, vol. 115, no. 11, pp. 2931–2942, 2011.

[53] R. Muscarella, S. Kolyaie, D. C. Morton, J. K. Zimmerman, M. Uriarte, and T. Jucker, “Effects of topography on tropical forest structure depend on climate context,” *Journal of Ecology*, vol. 108, no. 1, p. 145, 2019.

[54] Q. Wang, R. Punchi-Manage, Z. Lu et al., “Effects of topography on structuring species assemblages in a subtropical forest,” *Journal of Plant Ecology*, vol. 280, p. 47, 2016.

[55] S. Ediriweera, T. Danaher, and S. Pathirana, “The influence of topographic variation on forest structure in two woody plant communities: a remote sensing approach,” *Forest Systems*, vol. 25, no. 1, p. 049, 2016.

[56] R Core Team, *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013.