UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Tecnura

https://revistas.udistrital.edu.co/index.php/Tecnura/issue/view/1277

DOI: https://doi.org/10.14483/22487638.22579

INVESTIGACIÓN

Aboveground Biomass Estimation of Colombian Cocoa Agroforestry Systems using Multispectral Images and Regression Methods

Estimación de la biomasa aérea de sistemas agroforestales cacaoteros colombianos mediante imágenes multiespectrales y métodos de regresión

Claudia V. Correa[1], Tatiana Gélvez-Barrera[2], Laura V Galvis G[3], Edwin Vargas[4], Jonathan Monsalve[5], Ariolfo Camacho[6], Iván Ramírez[7], Hoover Rueda-Chacón[8]

Fecha de Recepción: 14 de agosto de 2024 **Fecha de Aceptación:** 2 de diciembre de 2025

Cómo citar: Correa., C. V., Gélvez-Barrera., T., Galvis., L. V., Vargas., E., Monsalve., J., Camacho., A., Ramírez., I, and Rueda-Chacón., H. (2025). Aboveground Biomass Estimation of Colombian Cocoa Agroforestry Systems using Multispectral Images and Regression Methods. *Tecnura*, 29(86), 3–16. https://doi.org/10.14483/22487638.22579

Abstract

Context: Aboveground biomass (AGB) estimation for agricultural and environmental applications traditionally relies on time-consuming manual methods at ground level. Emerging approaches use remote sensing data evaluating spectral responses and vegetation indices in order to estimate the AGB more efficiently. Nonetheless, such techniques face various challenges in accounting for complex spatial distributions, especially when dealing with agroforestry systems (AFS).

Objective: This work aimed to estimate AGB in a Colombian cocoa AFS from multispectral images acquired with an un-manned aerial vehicle (UAV).

Methodology: In this work, the AGB was estimated by computing different vegetation indices from the measured spectral reflectances and evaluating two linear regression models, i.e., principal component regression (PCR) and partial least squares regression (PLSR), as well as a nonlinear regression model, a neural network composed by a perceptron with a single layer. Control points were obtained via on-ground biomass manual acquisition.

Results: Numerical experiments resulted in a coefficient of determination of R2 = 0.58 for the best linear model, while the nonlinear model reached R2 = 0.86.

[1] PhD. in Electrical and Computer Engineering. Deep Science SAS, Colombia. Email: clavicop@deepscience.com.co
[2] PhD. in Engineering. Deep Science SAS, Colombia. Email: tatiana.gelvez@deepscience.com.co
[3] PhD. in Electrical and Computer Engineering. Deep Science SAS, Colombia. Email: laura.galvis@deepscience.com.co
[4] PhD. in Engineering. Deep Science SAS, Colombia. Email: edwin.vargas@deepscience.com.co
[5] PhD. in Engineering. Deep Science SAS, Colombia. Email: jonathan.monsalve@deepscience.com.co
[6] PhD. in Computer Science. Deep Science SAS, Colombia. Email: ariolfo@deepscience.com.co
[7] Conexalab SAS. Email: iramirez@conexalab.com
[8] PhD. in Electrical and Computer Engineering. Assistant Professor at the Department of Computer Science, Universidad Industrial de Santander, Bucaramanga, Colombia. Email: hfarueda@uis.edu.co

Conclusions: AGB estimation in a cocoa AFS can be effectively performed using multispectral data acquired with a UAV and a simple nonlinear regression model.
***Keywords*:** aboveground biomass estimation, agroforestry systems, cocoa, multispectral images, regression methods

Resumen

Contexto: La estimación de la biomasa aérea (AGB) para aplicaciones agrícolas y ambientales tradicionalmente se basa en métodos manuales a nivel del suelo que requieren mucho tiempo. Los enfoques emergentes utilizan datos de teledetección que evalúan respuestas espectrales e índices de vegetación para estimar la AGB de una manera más eficiente. No obstante, estas técnicas enfrentan varios desafíos a la hora de tener en cuenta distribuciones espaciales complejas, especialmente en sistemas agroforestales (SAF).
Objetivo: Este trabajo buscó estimar la AGB en un SAF de cacao colombiano a partir de imágenes multiespectrales adquiridas con un vehículo aéreo no tripulado (UAV).
Métodología: En este trabajo se estimó la AGB calculando diferentes índices de vegetación a partir de reflectancias espectrales medidas y evaluando dos modelos de regresión lineal, i.e., regresión de componentes principales (PCR) y regresión de mínimos cuadrados parciales (PLSR), y un modelo de regresión no lineal, una red neuronal de un solo perceptrón. Los puntos de control se obtuvieron mediante una adquisición manual de biomasa en tierra.
Resultados: Los experimentos numéricos dieron como resultado un coeficiente de determinación de R2 = 0.58 para el mejor modelo lineal, mientras que el modelo no lineal alcanzó un R2 = 0.86.
Conclusiones: Es posible realizar una estimación eficaz de la AGB en un sistema agroforestal de cacao a partir de datos multiespectrales adquiridos con un UAV y un modelo de regresión no lineal simple.
***Palabras clave*:** estimación biomasa aérea, sistemas agroforestales, cacao, imágenes multiespectrales, métodos de regresión

Introduction

Biomass is the largest and most important renewable source that is currently available and can be used to produce different forms of energy. It is currently the fourth largest energy source after coal, oil, and natural gas [1]. Given that forests have an inherent ability to retain carbon dioxide and can regulate the cycling of carbon, interest in estimating forest biomass has increased.

The aboveground biomass (AGB) measures the amount of organic matter in living and dead plant materials [2, 3], which is a critical indicator of plant health. It is expressed as oven-dry tons per unit, including leaves, twigs, branches, the main bole, and the bark [4]. Therefore, it is widely used in smart and precision agriculture for growth crop monitoring [5] and environmental applications such as carbon sequestration monitoring [6].

Traditional AGB estimation methods are based on manual sampling and dry weight calculations, which destroy the samples [7, 8]. Alternative methods employ tree level measures to generate allometric functions aimed at evaluating the ground-level biomass, and extrapolation methods are later used

to assess the whole area [7, 9–12]. These methods, however, are time-consuming and expensive considering the extensive areas that must be covered [6].

Recently, remotely sensed data from satellites and unmanned aerial vehicles (UAVs) have been used as a more practical method for biomass mapping and estimation in larger study areas [13–15]. Remote sensing data contain information on the spectral response, density, shade, and texture of the vegetation cover, which is related to various indices [16] and can therefore be used to estimate the AGB [8]. To name a few, the authors of [7] use spaceborne optical, synthetic aperture radar (SAR), and light detection and ranging (LiDAR) data to estimate AGB in agroforestry systems (AFS) while focusing on varying climate conditions. In [18], UAV LiDAR data and multispectral satellite images are used along with topographic correction methods for AGB estimation, providing a new indicator in mountain grassland applications. In [19], the AGB in sorghum is estimated via machine learning techniques, exploiting the relationship between vegetation indices and sorghum AGB for different spatial resolutions. Further methods have estimated biomass using multispectral UAV imagery from crops such as dry strawberry [20] and potato [21].

AGB estimation through remotely sensed data has become relevant in AFS, where vegetation, crop density, and sometimes livestock are combined. Specifically, AFS are a diverse and sustainable farming approach that combines agriculture and forestry, optimizing land use and contributing to environmental conservation, sustainable development and ensuring long-term productivity [22]. In Colombia, cocoa AFS have become popular due to their potential for supporting sustainable development and providing alternative livelihoods in areas previously affected by conflict [23, 24]. In addition, cocoa is one of the most representative products for the Colombian economy, not only for internal consumption but also for international trading [25, 26]. Previous works on cocoa AFS in Colombia have investigated the carbon storage potential and productivity in the southwest of the country [23, 27] by employing different allometric models.

The accurate estimation of AGB from remote sensing data still faces limitations, including the difficulty of considering the variability in the spatial distribution, structure, and composition of AFS

[28]. Furthermore, particularly for Colombia, there is still a strong dependence on traditional on-site crop measurements as well as on destructive AGB methods. In light of the above, this work aims to explore the potential of remotely acquired spectral imagery and machine learning to estimate the AGB in a Colombian cocoa AFS.

Specifically, this work employed a UAV to collect multispectral images over an AFS located in a northeastern region of Colombia, in addition to a set of on-site biomass control points manually calculated through allometric equations. The proposed pipeline to analyze and calculate the AGB included

(i) the calculation of vegetation indices (VIs) from the spectral bands, (ii) a correlation analysis between the predictors (spectral bands and VIs) and the observed biomass, and (iii) the performance analysis of two multivariate linear regression models, *i.e.*, principal component regression (PCR) [29, 30] and partial least squares regression (PLSR) [31], and a nonlinear regression model, a neural network (NN) [32]. The results demonstrate that the near-infrared (NIR) and green spectral bands, along with the normalized difference vegetation index (NDVI), provide the highest correlation with the AGB of the cocoa AFS. The best model achieves a coefficient of determination as high as $R^2 = 0.86$.

Methodology

This section describes the study area and summarizes the methods used to estimate the AGB from multispectral imagery.

Study area

The study area is an 896 × 726 m^2 region in the northeastern part of Colombia, as highlighted in Figure 1a. Particularly, this area is located at 6°51'N, 73°24'W, in the municipality of San Vicente de Chucurí, Department of Santander. In Figure **1**b, the red line traces the entire municipality, and the small blue box limits the observed region: a mountainous agricultural area with a cocoa AFS. As shown in Figures 1c-d, the yellow box highlights the cocoa trees, which were manually located and segmented.

At-ground allometric biomass measurements

40 cocoa trees (*Theobroma cacao*) were manually measured at ground level, including height and diameter, and later analyzed. Following the work of [33], we used an allometric equation to calculate the cacao species biomass in individual trees:

$$\text{Biomass} = 10^{-1,625 + 2,63 \times \log10(d30)} \tag{1}$$

where d_{30} refers to the tree trunk diameter at a height of 30 cm. Based on Equation (1), the observed biomass for the 40 cocoa trees ranges from 3 to 25 kilograms per tree.

Aboveground multispectral measurements

Multispectral images were acquired using a Parrot Sequoia multispectral camera mounted on a UAV at a height of 100 m, at 10:15 am, with image overlap of 75%. This camera has a synchronized array of four single-band cameras, each with a 1.2 MP sensor featuring a 3.75 µm pixel pitch and a 4 mm lens, but with a different interference filter in front. The interference filters employed by this camera span along the green (G: 530 to 570 nm), red (R: 640 to 680 nm), red-edge (RE: 730 to 740 nm), and near-infrared (NIR: 770 to 810 nm) spectral bands. The acquired images have a resolution of 7474 × 6053 pixels, with

a ground sample distance (GSD) of 12 cm. Agisoft Metashape and its automatic reflectance calibration procedures were used to generate the orthomosaic.

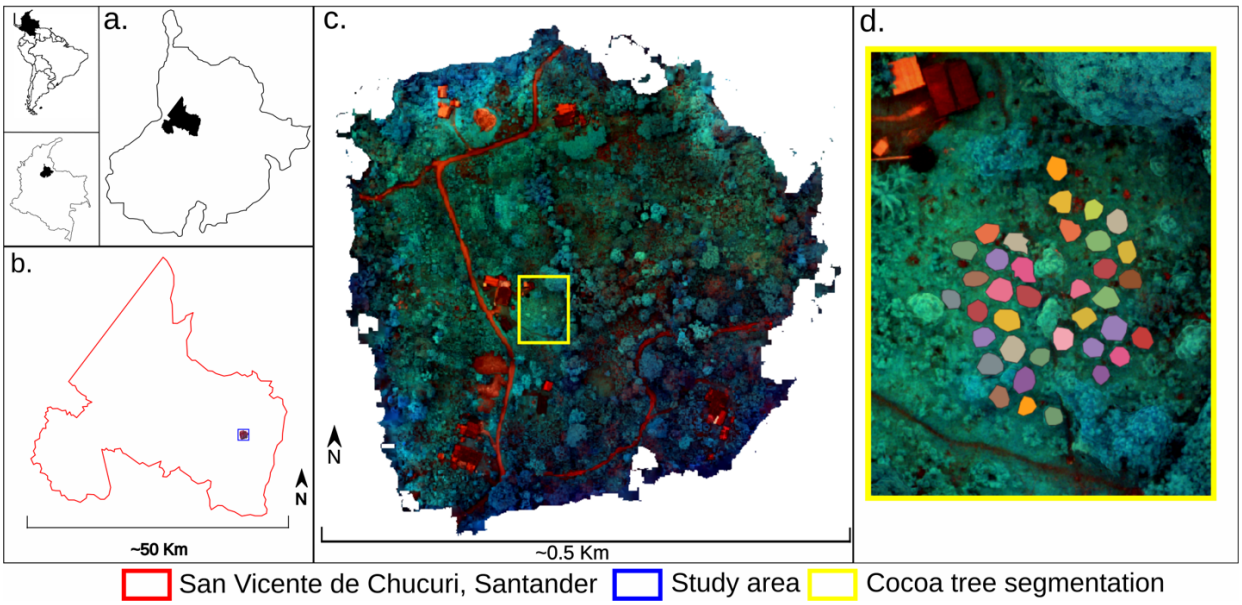


Figure 1. a) Study area located at 6°51′N, 73°24′W in San Vicente de Chucuri, Santander, Colombia. b) The red line traces the entire municipality, while the small blue box limits the observed region. c) Mountainous area with the AFS under analysis. d)Yellow box showing the cocoa tree segmentation.

Calculating VIs

VIs are calculated based on the ratio between two or more bands to contrast the high absorption by leaf pigments (chlorophylls, carotenoids, and xanthophylls) in the visible spectral region (400–700 nm) and the high reflectance by leaves in the NIR region (700–1300 nm). For this study, we chose six of the most representative VIs that have been associated with AGB in the state of the art [16]: the NDVI, GNDVI, NDRE, RDVI, CVI, and OSAVI. Table 1 shows the formulas used to calculate each index.

Regression models for AGB estimation

We considered two regression approaches to estimate AGB from multispectral information and calculated VIs. The first uses linear regressions such as PCR and PLSR. The second uses a learning-based perceptron that captures nonlinear relationships between the predictors.

PCR and PLS regression

For the linear regression method, let $\mathbf{y} \in R^m$ denote a column vector with the m observed bio-mass values, and $\mathbf{X} \in R^{m \times n}$ a matrix with the n predictors (four spectral bands and six VIs), for each m-th observation. Thus, a linear model of the form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$ can be established, where $\boldsymbol{\beta} \in R^n$ is the vector of coefficients to be estimated and $\epsilon \in R^m$ accounts for the approximation error. These coefficients are commonly

estimated via ordinary least squares. However, due to the high correlations within the spectral variables, PCR and PLSR are preferred; they break such correlations by projecting the original data onto a lower-dimensional subspace.

Mathematically, the projection of the matrix $\mathbf{X}$ to a new subspace can be represented as $\mathbf{C} = \mathbf{X}\mathbf{W}^T$, where $\mathbf{W} \in \mathrm{R}^{r \times n}$ is known as the projection matrix and $\mathbf{C} \in \mathrm{R}^{m \times r}$ contains r linearly independent variables, *i.e.*, the principal components or scores. When $\mathbf{W}$ is calculated using the SIMPLS algorithm [34], $\mathbf{C}$ contains the PLS coefficients. Furthermore, when $\mathbf{C}$ is employed instead of $\mathbf{X}$, the regression model becomes, $\mathbf{y}' = \mathbf{X}\boldsymbol{\beta}' + \epsilon'$, with $\boldsymbol{\beta}' = \boldsymbol{\beta}\mathbf{Q}^T$, where $\mathbf{Q} \in \mathrm{R}^{n \times r}$ represents the PLS regression over the observation vector $\mathbf{y}$. PLS uses the observed samples $\mathbf{y}$ to construct $\mathbf{W}$ and $\mathbf{Q}$, to obtain linearly independent variables from $\mathbf{X}$, which are maximally correlated with the known observations.

Nonlinear regression model

To consider nonlinear interactions between the n predictors, we explored the use of NNs, introducing nonlinearity through activation functions [35]. In mathematical terms, let $\{\mathbf{x}_i, \mathbf{y}_i\}$ be the paired training dataset for $i = 1, \ldots, m$, where $\mathbf{y}_i$ denotes the i-th observed biomass and $\mathbf{x}_i \in \mathrm{R}^n$ contains the corresponding predictors (rows of $\mathbf{X}$). A neural network G_θ, with trainable weights can learn the coefficients to predict the AGB from the multispectral predictors, obtaining the optimal weights by solving the following optimization problem:

$$\theta^* \varepsilon \arg\min_{\theta} \left\{ \frac{1}{m} \sum_{i=1}^{m} L(x_i, G_\theta(y_i)) \right\} \tag{2}$$

where $L(\cdot)$ denotes the loss function. The model architecture includes an input layer that receives the predictors as the input, a flatten layer, and a dense layer that outputs predictions with the specified number of classes. We selected the mean-squared error (MSE) as the loss function and used the adaptive moment estimation (ADAM) optimizer, with the activation function set to be the parametric rectified linear unit (PReLU) [32].

Results

This section reports on the results obtained to estimate the AGB from remote sensing data via the linear and nonlinear regression models. Firstly, a correlation analysis is conducted to study how the predictors are correlated between them and to the observed biomass measured at ground level. Then, we present the results obtained with the two regression methodologies. We used the coefficient of determination (R^2) and the root-mean-squared error (RMSE) to evaluate their performance. In this section, note that B1 refers to the green (B) band, B2 to the red (R) band, B3 to the red-edge (RE) band, and B4 to the near-infrared (NIR) band of the multispectral images.

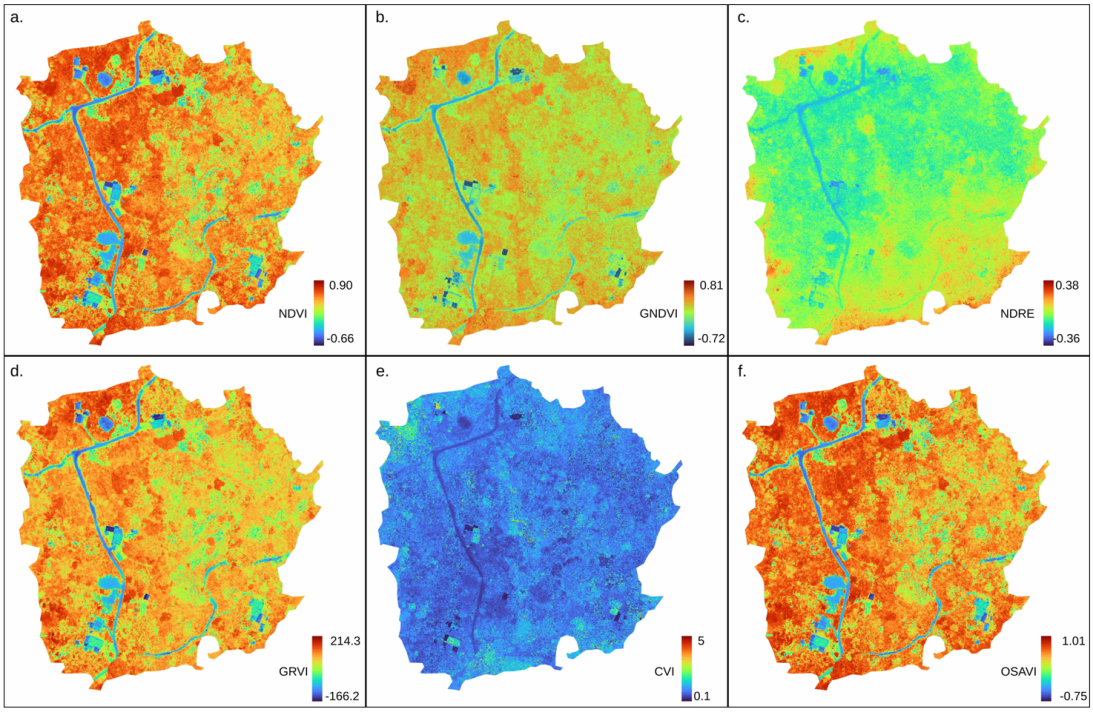


Figure 2. Maps of the cocoa AFS for the six vegetation indices under analysis: a) NDVI, b) GNDVI, c) NDRE, d) GRVI, e) CVI, and f) OSAVI.

VI maps

40 biomass data points were measured at ground level using the allometric Equation (1), and they were assumed to be the ground truth control points. The multispectral dataset was acquired in the same region where the cocoa trees were present, as detailed in Figure 1, and the 40 cocoa trees were manually segmented to calculate the averaged reflectance, with which we calculated the VIs. For reference, the maps of the calculated vegetation indices for the full cocoa AFS are shown in Figure 2.

Correlation analysis

The predictors were correlated between them and against the biomass. Figure 3 shows that the greatest correlation with biomass comes from the red-edge band (B3=0.24), followed by green (B1=0.19) and red (B2=0.19). As expected, higher correlation values were obtained between the spectral bands and the VIs. For instance, B1 and GNDVI have a correlation of -0.94, B2 and NDVI exhibit -0.96, and B2 and RDVI report -0.90.

Furthermore, an exploratory analysis was performed to determine the data variance explained by the principal components. Given the bias-variance trade-off, selecting many components may lead to overfitting, and too few may cause underfitting. Thus, keeping the minimum number of components before explained data variance plateaus yields minimal information loss and greater prediction power.

Figure 4a shows that two components explain more than 95% of the variance in **X**, while at least six components are required to explain 40% of the biomass (**y**). Figure 4b shows the score plot where the dominant predictors for the first two principal components (PC1 and PC2) explain 95% of the variance. In particular, CVI and GNDVI are in the PC1-positive and PC2-positive quadrant; NDRE, B4, RDVI, NDVI, and OSAVI are in the PC1-positive and PC2-negative quadrant; B2 is in the PC1-negative and PC2-positive quadrant; and B1 is in the PC1-negative and PC2-negative quadrant.

AGB estimation using PCR and PLSR

The PCR and PLSR algorithms were run following a cross-validation strategy with 70-30% partitions. The spectral variables were mean-centered and normalized. Furthermore, singular value decomposition (SVD) and the SIMPLS algorithm were used to estimate the PCA and PLS components, respectively. We evaluated the R^2 and the RMSE to address the variable selection step. Fig. 5a summarizes the statistics of different models, varying the number of components from 1 to 10. The results suggest that the best prediction model (high R^2 and low RMSE) is attained with seven components, with an averaged $R^2 = 0.58$ (Fig. 5b), and an RMSE of prediction of 3.44 for both PCR and PLSR, as shown in Fig. 5c. The latter suggests and error of around (3.4499/25) * 100% = 13.80% in the maximum biomass value reported (25 kg/tree).

AGB estimation using nonlinear regression

To evaluate the performance of the regression NN, and given the limited number of samples (m = 40), we employed a k-fold cross-validation strategy to avoid optimistic bias and obtain a more reliable estimate of generalization performance. We carried out a k-fold cross-validation for k = [2, 4, 6, 8], *i.e.*, we used k samples for validation and m - k samples for training. The learning rate was fixed to $8e^{-3}$ using 600 epochs. The dataset was normalized with respect to the maximum. For each experiment, Fig. 6a reports the mean and standard deviation, where we obtained up to $R^2 = 0.8589$ for the prediction task (Fig. 6b) and a prediction RMSE of 2.090 (Fig. 6c). The latter suggests and error of around (2.090/25) * 100% = 8.36% in the maximum biomass value reported (25 kg/tree). The high variance in the models is due to the few available training points. Such behavior can be improved when more data become available.

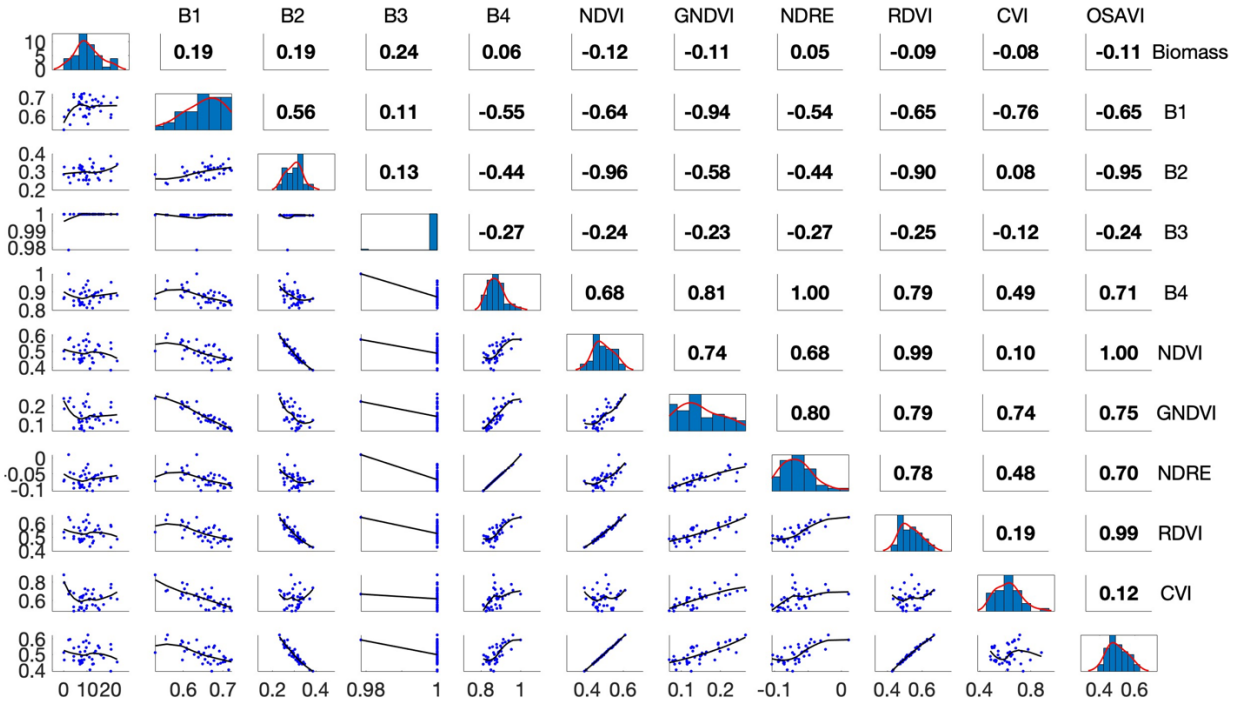
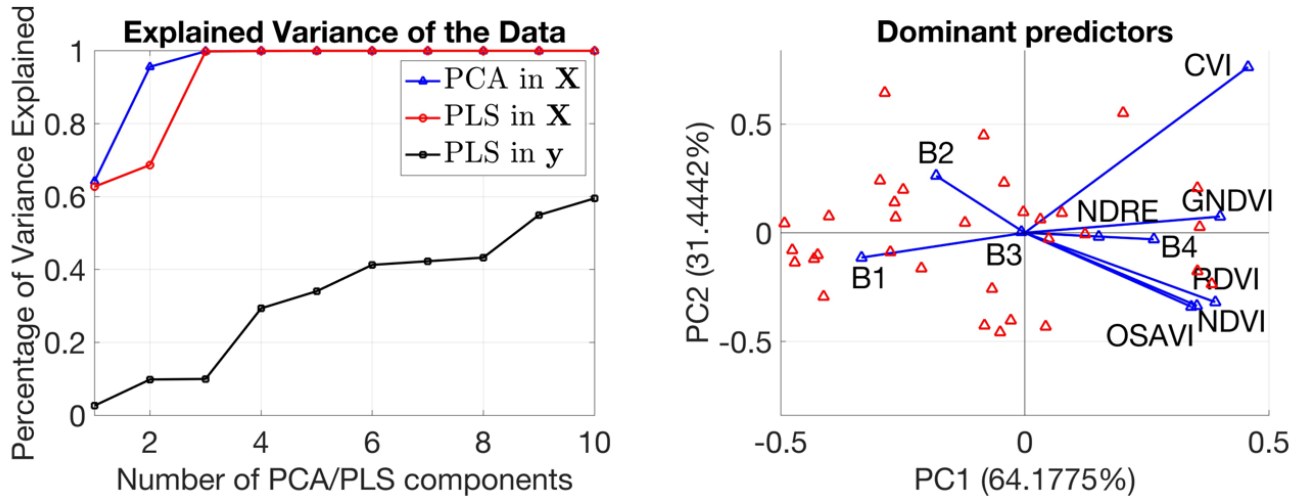
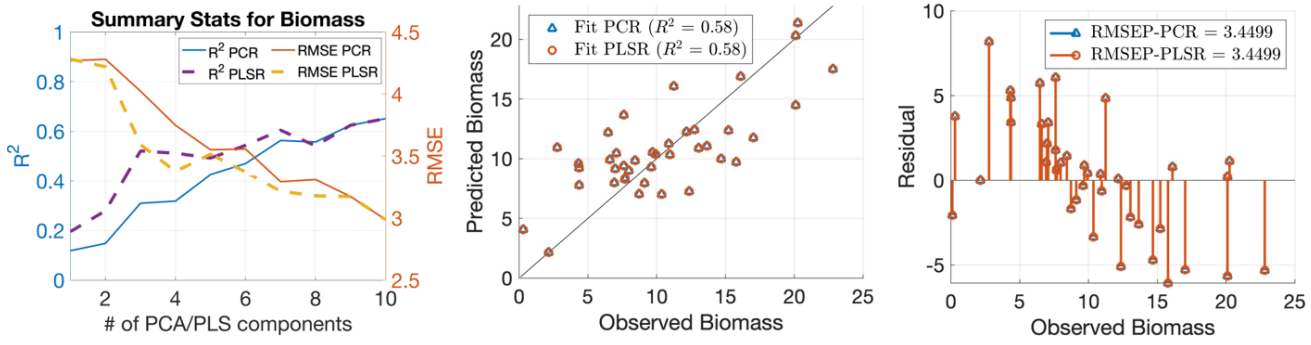


Figure 3. Correlation plot between biomass and predictors. Greater correlation comes from B3 (0.24), B1, and B2 (0.19). Higher correlation values are observed between spectral bands and vegetation indices, as expected.



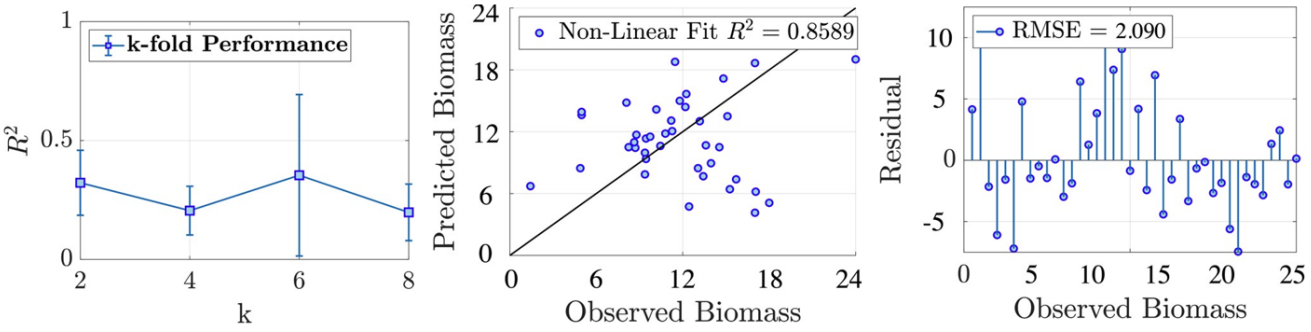
(a) (b)

Figure 4. a) Explained variance *vs.* number of components (two components explain over 95% of the variance in **X**, while at least six are required to explain 40% of the biomass). b) Score plot showing the dominant predictors for the first two principal components (PC1 on the x-axis and PC2 on the y-axis).



(a) (b) (c)

Figure 5. PCR/PLSR results. a) Summary of different regression models with varying number of components. b) The best fitting result with seven components. c) Residuals of best fit.



(a) (b) (c)

Figure 6. a) Average performance of nonlinear regression for different k-folds. b) Best fitting result. c) Best fit residuals.

Discussion

Overall, the results demonstrate the potential of multispectral UAV imagery combined with regression techniques to estimate AGB in Colombian cocoa AFS. The correlation analysis showed that, although individual predictors such as the red-edge band have moderate correlations with biomass (B3=0.24), the predictors are highly intercorrelated. This supports the selection of few predictors that ensure a good model fit while preventing overfitting, with seven components providing the best trade-off. The best linear models reached R^2 = 0.58, suggesting that they cannot fully represent complex spectral–biomass relationships. In contrast, the nonlinear NN improved performance, achieving R^2 = 0.86 and reducing the RMSE to 2.090 kg/tree (8.36% of the maximum biomass), highlighting the importance of nonlinear interactions in heterogeneous AFS. These results indicate that comparable AGB estimation performance can be achieved with smaller, lower-resolution datasets, in contrast to recent studies on strawberry [20] and potato [21] crops in different countries, which used large, very high-resolution (2 cm) multitemporal datasets and reported R^2 values around 0.8.

From a practical perspective, this methodology could be used by institutions such as FEDECA-CAO as a scalable tool for monitoring cocoa AFS. Specifically, periodic UAV-based multispectral surveys combined with nonlinear models would generate spatially explicit AGB maps at farm or regional scales, thereby supporting carbon accounting and certification processes, as described in [36]. Additionally, these biomass maps could serve as indicators of crop health, facilitating the early identification of low-productivity areas and enabling targeted agronomic interventions.

Among the identified limitations, the small sample size (40 trees) likely contributed to variability in model performance and may restrict generalization to other cocoa farms. Second, the entirety of the UAV imagery was acquired on a single date. Therefore, models may be sensitive to phenological stage and seasonal variations. Finally, the atmospheric conditions present at the time of acquisition, such as cloud cover or illumination differences, could have influenced spectral reflectance, potentially affecting predictions.

Conclusions

This work estimated the AGB of a cocoa AFS in Colombia while employing a dataset of multispectral images acquired with an UAV, using linear regression models (PCR and PLSR) and a learning-based nonlinear regression model (NN). These models exploit at-ground biomass control points. Numerical experiments resulted in an $R^2 = 0.58$ for the best linear model, with an RMSE of 3.4499, whereas the non-linear model attained an $R^2 = 0.86$ and an RMSE of 2.090. In terms of percentage error, the best linear model yielded 13.80%, and the non-linear model reported only 8.36%. This suggests that a simple nonlinear regression model can better predict the biomass content of a cocoa AFS.

Future work should expand the analysis to include multiple acquisition dates across phenological stages and environmental conditions for temporal assessment. Additionally, integrating complementary data such as LiDAR-derived structural metrics or textural information from high-resolution imagery could enhance AGB estimation accuracy. Exploring advanced nonlinear modeling and ensemble methods may further improve prediction while maintaining operational scalability, strengthening UAV based biomass estimation in cocoa AFS.

Acknowledgments

The data used in this work were provided by the National Federation of Cacao Growers of Colombia (FEDECACAO).

References

[1] L. Torre, "Biomass estimation using lidar data," Int. J. Sust. Energy Manage. Plan., vol. 17, p. 79–90, Jun. 2018. https://journals.aau. dk/index.php/sepm/article/view/2037

[2] H. Tao et al., "Estimation of crop growth parameters using UAV-based hyperspectral remote sensing data," Sensors, vol. 20, no. 5, art. 1296, 2020. https://doi.org/10.3390/s20051296

[3] P. Lourenço, S. Godinho, A. Sousa, and A. C. Gonçalves, "Estimating tree aboveground biomass using multispectral satellite-based data in mediterranean agroforestry system using random forest algorithm," Remote Sens. App. Soc. Environ., vol. 23, art. 100560, 2021. https://www.sciencedirect.com/science/article/pii/S2352938521000963

[4] S. Brown, Estimating biomass and biomass change of tropical forests: A primer. Rome, Italy: FAO, 1997.

[5] T. Wang et al., "Applications of UAS in crop biomass monitoring: A review," Front. Plant Sci., vol. 12, art. 616689, 2021. https://doi.org/10.3389/fpls.2021.616689

[6] S. Sankar, M. Lewis, and P. Hosein, "Above ground biomass estimation of a cocoa planta-tion using machine learning," in 2022 IEEE Int. Conf. Data Min. Work. (ICDMW), 2022, pp. 1–8.

[7] Z. Hosseini, H. Naghavi, H. Latifi, and S. Bakhtiari Bakhtiarvand, "Estimating biomass and carbon sequestration of plantations around industrial areas using very high resolution stereo satellite imagery," iForest Biogeosci. Forestry, vol. 12, no. 6, art. 533, 2019. https://doi.org/10.3832/ifor3155-012

[8] Y. Li, M. Li, C. Li, and Z. Liu, "Forest aboveground biomass estimation using landsat 8 and sentinel-1a data with machine learning algorithms," Sci. Rep., vol. 10, no. 1, art. 9952, 2020. https://doi.org/10.1038/s41598-020-67024-3

[9] A. Nyamukuru, C. Whitney, J. R. Tabuti, J. Esaete, and M. Low, "Allometric models for aboveground biomass estimation of small trees and shrubs in african savanna ecosystems," Trees Forest People, vol. 11, art. 100377, 2023. https://www.sciencedirect.com/science/article/pii/S2666719323000092

[10] R. Aabeyir, S. Adu-Bredu, W. A. Agyare, and M. J. C. Weir, "Allometric models for estimating aboveground biomass in the tropical woodlands of Ghana, West Africa," Forest Ecosyst., vol. 7, art. 41, 2020. https://doi.org/10.1186/s40663-020-00250-3

[11] A. B. Bazrgar, N. Thevathasan, A. Gordon, and J. Simpson, "Allometric equations for estimating aboveground biomass carbon in five tree species grown in an intercropping agroforestry system in southern ontario, canada," Agroforestry Syst., vol. 98, no. 3, pp. 739–749, 2024. https://doi.org/10.1007/s10457-023-00942-z

[12] V. L. Morán-Villa et al., "Above-ground biomass estimation by developing allometric equations for theobroma cacao in tabasco, mexico," Agroforestry Systems, vol. 98, no. 3, pp. 537–549, 2024. [Online]. Available: https://doi.org/10.1007/s10457-023-00928-x

[13] P. Sharma et al., "Geospatial technology in agroforestry: Status, prospects, and constraints,"Environ. Sci. Poll. Res., vol. 30, 30, pp. 116459–116487, 2023. https://doi.org/10.1007/s11356-022-20305-y

[14] M. Zhang et al., "A spatio-temporal fusion strategy for improving the estimation accuracy of the aboveground biomass in grassland based on gf-1 and modis," Ecol. Ind., vol. 157, art. 111276, 2023. https://www.sciencedirect.com/science/article/pii/S1470160X23014188

[15] A. Hojo, R. Avtar, T. Nakaji, T. Tadono, and K. Takagi, "Modeling forest above-ground biomass using freely available satellite and multisource datasets," Ecol. Infor., vol. 74, art. 101973, 2023. https://www.sciencedirect.com/science/article/pii/S157495412300002X

[16] P. Lourenço, "Biomass estimation using satellite-based data," in Forest Biomass, A. C. Gonçalves, A. Sousa, and I. Malico, Eds. Rijeka: IntechOpen, 2021, ch. 3. [Online]. Available: https://doi.org/10.5772/intechopen.93603

[17] D. Kanmegne Tamga, H. Latifi, T. Ullmann, R. Baumhauer, J. Bayala, and M. Thiel, "Estimation of aboveground biomass in agroforestry systems over three climatic regions in West Africa using Sentinel-1, Sentinel-2, ALOS, and GEDI data," Sensors, vol. 23, no. 1, art. 349, 2022. https://doi.org/10.3390/s23010349

[18] A. Chen et al., "Fusion of lidar and multispectral data for aboveground biomass estimation in mountain grassland," Remote Sens., vol. 15, no. 2, art. 405, 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/2/405

[19] Q. Liu et al., "Smartagb: Aboveground biomass estimation of sorghum based on spatial resolution, machine learning and vegetation index," EAI End. Trans. Internet of Things, vol. 9, no. 1, art. e1, Mar. 2023. https://publications.eai.eu/index.php/IoT/article/view/2904

[20] C. Zheng, A. Abd-Elrahman, V. Whitaker, and C. Dalid, "Prediction of strawberry dry biomass from uav multispectral imagery using multiple machine learning methods," Remote Sens., vol. 14, no. 18, art. 4511, 2022. https://www.mdpi.com/2072-4292/14/18/4511

[21] S. Luo et al., "Multi-dimensional variables and feature parameter selection for aboveground biomass estimation of potato based on uav multispectral imagery," Front. Plant Sci., vol. 13, art. 948249, 2022. https://www.frontiersin.org/articles/10.3389/fpls.2022.948249

[22] P. K. R. Nair, "Carbon sequestration studies in agroforestry systems: a reality-check,"Agroforestry Syst., vol. 86, no. 2, pp. 243–253, 2012. https://doi.org/10.1007/s10457-011-9434-z

[23] W. Ballesteros-Possú, J. C. Valencia, and J. F. Navia-Estrada, "Assessment of a cocoa-based agroforestry system in the southwest of colombia," Sustainability, vol. 14, no. 15, art. 9447, 2022. https://www.mdpi.com/2071-1050/14/15/9447

[24] P. C. Abbott et al., "An analysis of the supply chain of cacao in Colombia," Purdue University and International Center for Tropical Agriculture, Technical Report, 2018. [Online]. Available: https://www.purdue.edu/colombia/partnerships/cacaoforpeace/docs/Cacao%20for%20Peace%20Final%20English.pdf

[25] UPRA, "Plan de ordenamiento productivo: Análisis situacional de la cadena productiva del cacao y su agroindustria en colombia," Unidad de Planeación Rural Agropecuaria, Dirección del Uso Eficiente del Suelo y Adecuación de Tierras, Technical Report, 2023. [Online]. Available: https://upra.gov.co/es-co/Paginas/pop-cacao.aspx

[26] UPRA, "Resultados preliminares: evaluaciones agropecuarias," Unidad de Planeación Rural Agropecuaria, Technical Report, 2023. [Online]. Available: https://upra.gov.co/es-co/Evas_Documentos/Resultados_EVA_2023.pdf

[27]V. E. Mena-Mosquera and H. J. A. C, "Potencial de reducción de emisiones y captura de carbono en bosques y sistemas agroforestales con cacao en el pacífico colombiano," Rev. Bio. Trop., vol. 69, no. 4, pp. 1252–1263, 2021.

[28] R. Singh et al., "Optimising carbon fixation through agroforestry: Estimation of aboveground biomass using multi-sensor data synergy and machine learning," Ecol. Infor., vol. 79, art. 102408, 2024. https://www.sciencedirect.com/science/article/pii/S1574954123004375

[29] I. T. Jolliffe, "A note on the use of principal components in regression," J. Royal Stat. Soc. Ser. C App. Stat., vol. 31, no. 3, pp. 300–303, 1982. http://www.jstor.org/stable/2348005

[30] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning: with Applications in R. London, UK: Springer, 2013. [Online]. Available: https://faculty.marshall.usc.edu/gareth-james/ISL/3

[31] S. Wold, M. Sjostrom, and L. Eriksson, "Pls-regression: A basic tool of chemometrics," Chemome-trics Intell. Lab. Syst., vol. 58, pp. 109–130, 2001

[32] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA:MIT Press, 2016, http://www. deeplearningbook.org

[33] H. Andrade, M. Segura, E. Somarriba, and M. Villalobos, "Valoración biofísica y financiera de la fijación de carbono por uso del suelo en fincas cacaoteras indígenas de Talamanca, Costa Rica," in Agroforres. Amér., vol. 46, pp. 45-50, 2008. https://repositorio.catie.ac.cr/bitstream/handle/11554/10054/A11506e.pdf?sequence=1&isAllowed=y

[34] S. de Jong, "Simpls: An alternative approach to partial least squares regression," Chemometrics Intell. Lab. Syst., vol. 18, no. 3, pp. 251–263, 1993. https://www.sciencedirect.com/science/article/pii/016974399385002X

[35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015. https://doi.org/10.1038/nature14539

[36]R. Budiarto and B. G. Dewanto, "Unmanned aerial vehicles for assessing biomass and carbon stocks in mangrove forests: A systematic review," Sust. Fut., vol. 10, art. 101425, 2025. https://doi.org/10.1016/j.sftr.2025.101425