Assessing the performance of random forest regression for estimating canopy height in tropical dry forests

Christian Javier Pinza-Jiménez, Yeison Alberto Garcés-Gómez
Faculty of Engineering and Architecture, Universidad Católica de Manizales, Manizales, Colombia

Article Info	ABSTRACT
	Accurate estimation of forest canopy height is essential for monitoring forest ecosystems and assessing their carbon storage potential. This study evaluates the effectiveness of different remote sensing techniques for estimating forest canopy height in tropical dry forests. Using field data and remote sensing data from airborne lidar and polarimetric synthetic aperture radar (SAR), a random forest (RF) model was developed to estimate canopy height based on different indices. Results show that the normalize difference build-up index (NDBI) has the highest correlation with canopy height, outperforming other indices such as relative vigor index (RVI) and polarimetric vertical and horizontal variables. The RF model with NDBI as input showed a good fit and predictive ability, with low concentration of errors around 0. These findings suggest that NDBI can be a useful tool for accurately estimating forest canopy height in tropical dry forests using remote sensing techniques, providing valuable information for forest management and conservation efforts. *This is an open access article under the <u>CC BY-SA</u> license.*
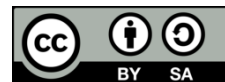
Corresponding Author:

Yeison Alberto Garcés-Gómez
Faculty of Engineering and Architecture, Universidad Católica de Manizales
Cra 23 No 60-63, Manizales, Colombia
Email: ygarces@ucm.edu.co

1. INTRODUCTION

The evaluation of the structure and dynamics of a forest ecosystem is crucial to understand its functioning and response to different disturbances and climate changes [1]. In this context, remote sensing technologies have become fundamental tools to obtain accurate and continuous information on the structure and dynamics of forests [2], [3]. Light detection and ranging (LIDAR) and synthetic aperture radar (SAR) have proven to be useful for obtaining precise measurements of the height and structure of forests in different types of forest ecosystems [4], [5].

This article presents an evaluation of the relationship between height metrics obtained through the global ecosystem dynamics investigation (GEDI), LIDAR sensor and SAR images from the L-band advanced land observing satellite-2 phased array type L-band synthetic aperture radar (ALOS PALSAR-2) sensor in the tropical dry forest of Caldas Department, Colombia. To analyze this relationship, the random forest algorithm is used, known for its ability to analyze large datasets and generate accurate predictive models [6]. The capacity of this algorithm to estimate canopy height from these different data sources is evaluated [7].

The estimation of canopy height is fundamental for sustainable forest management and understanding forest dynamics [8]. However, direct measurements of canopy height can be costly, labor-intensive, and even impossible in remote or extensive areas [9]. For this reason, satellite-based approaches have been developed to estimate canopy height at global and regional scales. The use of satellite data can be considered an indirect method for estimating forest canopy heights, which has become increasingly accurate with the evolution of satellite technologies [10].

In this context, the GEDI by National Aeronautics and Space Administration (NASA) [11] has stood out for providing valuable information on forest height through medium-resolution time series [12]. Although field validation is important to correlate two types of satellite data, the lack of it does not invalidate the usefulness of GEDI as an input for canopy height estimation [13]. GEDI datasets are improving height models and offering an important tool for global and regional forest management [4].

In the search to improve the accuracy of canopy height models, a combination of LIDAR and SAR data [14] has been used, which offer detailed information on the structure and dynamics of tropical forests. These data allow identifying patterns in the vertical distribution of forests and improving the accuracy of canopy height models [15]. Despite the advances achieved with this methodology, there are still challenges in terms of model accuracy. In particular, the lack of structural information on the forest has been a significant obstacle to improving the accuracy of such models [16]. The combination of LIDAR and SAR data offers a unique synergy for the evaluation of canopy height and texture in any type of forest [12], allowing for detailed information on structure and its vertical distribution. SAR data from the ALOS PALSAR-2 sensor [17], through the generation of various polarimetric indices, are suitable for texture detection, improving the accuracy of canopy height models [13].

The estimation of canopy height in tropical forests is valuable for the management and conservation of natural resources, including the identification and monitoring of biodiversity, assessment of forest productivity, and carbon monitoring. GEDI and SAR data provide information at global and regional scales, but machine learning algorithms are needed to fully evaluate the information. The evaluation of the relationship between GEDI height metrics and SAR polarimetric indices using machine learning techniques is crucial for improving the understanding of tropical forests and promoting their sustainable conservation and management.

The article is organized as follows: section 2 presents the materials and methods used for the development of the research. In section 3, the results of the study are shown and discussed. Section 4 presents the conclusions.

2. MATERIALS AND METHODS

2.1. Study area

The study area corresponds to the tropical dry forest ecosystem in the eastern part of the Caldas Department in Colombia. This is an area of great biological and environmental importance, which makes it suitable for investigating its dynamics, functioning, and response of this type of ecosystem to environmental changes. It is characterized by a prolonged dry season, high temperatures, and limited availability of water [18]. It is in the valley of the Magdalena River, one of the most important rivers in the country, at an elevation ranging from 150 to 400 meters above sea level. The spatialization of the study area was elaborated by the [18], as part of the project called "Conservation status of the floristic component in the tropical dry forest of Caldas," as shown in Figure 1.
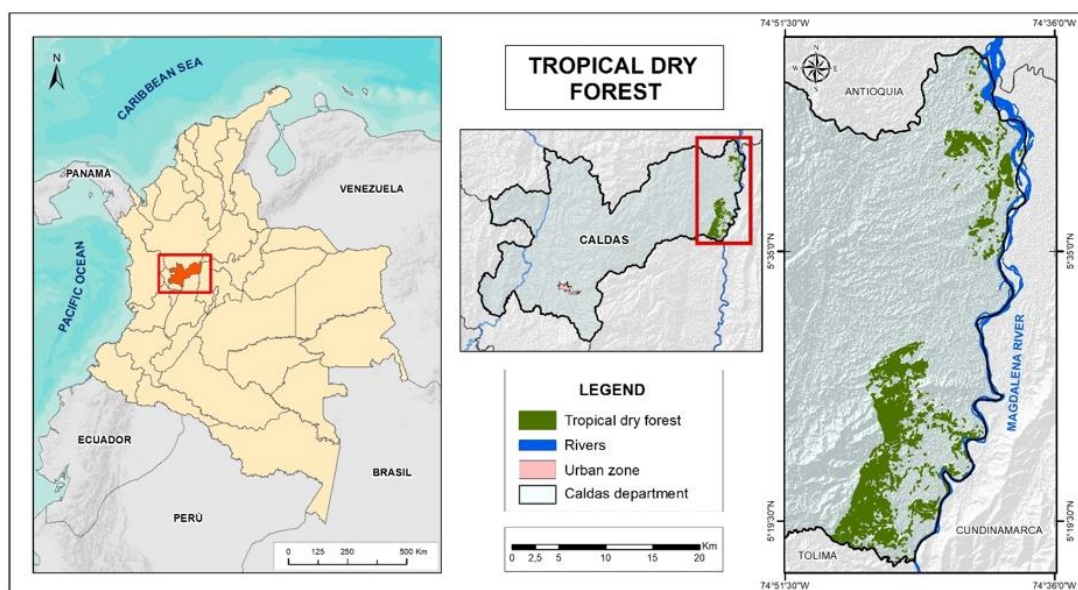

Figure 1. Location of the study area, tropical dry forest, Caldas, Colombia

The tropical dry forest is characterized as an ecosystem with a closed tree canopy and vegetation that varies between semi-deciduous and fully deciduous foliage. Despite its ecological importance, the tropical dry forest has been one of the least studied ecosystems, even though it has been highly intervened by human activity. In this sense, it is crucial to investigate and better understand the dynamics and functioning of this ecosystem in the selected study area in order to generate scientific knowledge that contributes to its conservation and sustainable management. The dry forest area is highly fragmented due to human activity, which has led to the creation of multiple polygons with isolated extensions and distributions [19]. To facilitate image processing and data extraction, a quadrant was delimited that covers the multiple polygons. However, it is emphasized that the analyses and results obtained correspond exclusively to the dry forest area.

2.2. Data acquisition

In the present research, two satellite images were used: the GEDI and the ALOS PALSAR-2. It should be noted that the image processing and data extraction were performed through Google earth engine (GEE), which allowed for greater efficiency and speed in the image analysis [20]. "GEDI is a data collection containing measurements of surface height acquired through the instrument aboard the NASA ICESat-2 satellite [21]. This instrument uses a laser system to measure the distance between the satellite and the earth's surface, and from these measurements, vegetation height can be determined [22]. The selected image consists of a collection of images obtained through GEE with a spatial resolution of 25 meters and a temporal range from 2019 to 2020. The rh95 metric was used because, according to previous studies, it has shown better results compared to other metrics [13].

The ALOS PALSAR-2 image is acquired by a satellite from the Japan Aerospace Exploration Agency-JAXA [23], which uses a SAR technology in L-band to capture images of the earth's surface [24]. The image is a collection obtained through GEE, with a spatial resolution of 25 meters and standard pre-processing provided by JAXA, which includes geometric and terrain slope corrections [25]. The image collection is available annually and has been constructed using a mean filter that combines images acquired from 2015 to 2020.

2.3. Data processing

The methodological scheme in Figure 2 shows the development of data processing. To process the ALOS PALSAR-2 image, several activities were carried out, including structuring the image collection in GEE. First, the collection was filtered by dates and region of interest to select the most suitable images for analysis and their corresponding composition during the aforementioned analysis period. Subsequently, the Pauli polarimetric decomposition technique was used to obtain a new polarization (VV) from the original polarizations (horizontal receive (HH) and vertical receive (HV)).
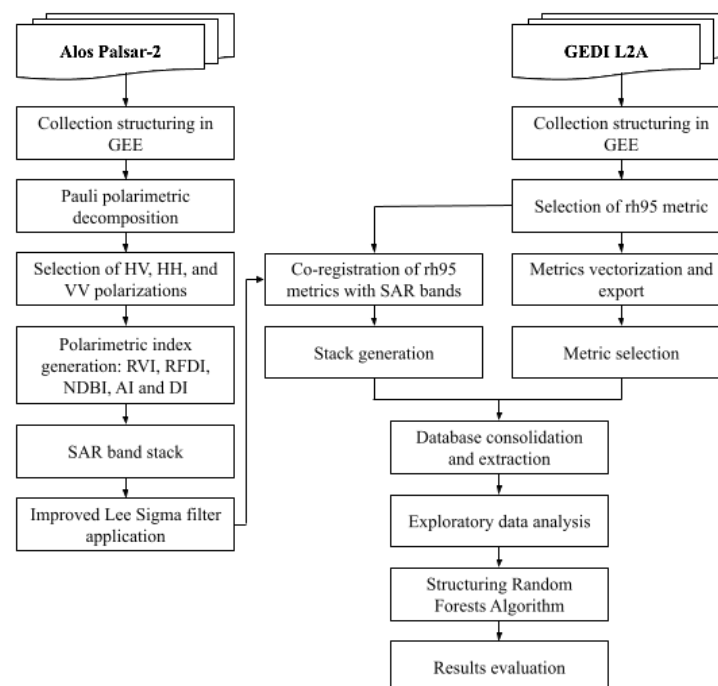

Figure 2. General workflow used in the research

This technique allows for the decomposition of the polarimetric scattering matrix into three backscattering matrices, which in turn facilitates the analysis of the polarimetric data of the image and highlights important features of the land surface, such as the forest structure of a forest [26]. Various polarimetric indices were generated from the HH, HV, and VV polarizations, providing complementary information about the forest composition [10]. Polarimetric indices provide a quantitative measure of the polarimetric properties of the land surface and are used to characterize the physical properties of objects in the image [27]. Additionally, they can be used to analyze specific features such as the asymmetry of surface backscattering, object shape, and orientation [28]. This can be directly associated with the structure and composition of the tropical dry forest. Table 1 presents the polarizations from ALOS PALSAR platform and Table 2 the polarimetric indices employed in the research.

Figure 3 shows a view of the quadrant that includes the study area. On the right, a sentinel-2 image is observed in red, green, and blue (RGB) combination, while on the left, a section of the study area is presented in different frames. In this section, the generated polarimetric indices are visualized in a color scale, and the tropical dry forest area is delimited with a black contour.

Table 1. ALOS PALSAR polarizations

Polarizations	Description
HH	Horizontal transmission and horizontal reception
HV	Horizontal transmission and vertical reception
VV	Vertical transmission and vertical reception

Table 2. Generated polarimetric indexes

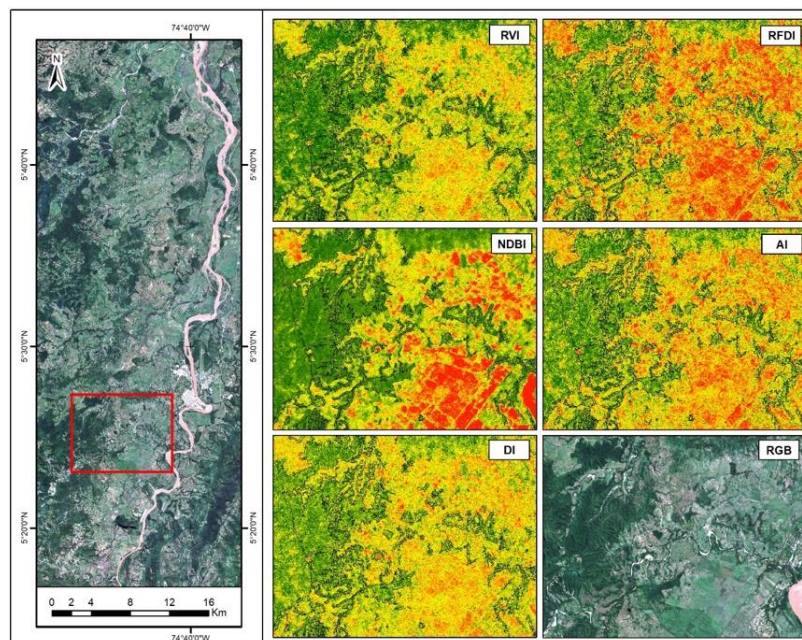| Name	Formula
Radar vegetation index (RVI)	$(4*HV)/(HH+HV)$
Radar forest degradation index (RFDI)	$(HH-HV)/(HH+HV)$
Normalized differential backscatter index (NDBI)	$(HH-HV)/(HH*HV)$
Anisotropy index (AI)	$(HH-VV)/(HH+VV)$
Double bounce index (DI)	$(2*HV)/(HH+VV)$

Figure 3. Visualization of polarimetric indexes accompanied by a true color sentinel-2 scene

From the polarizations and the generated polarimetric indices, they were compiled into an image stack that allows all the obtained information to be visualized together. Additionally, it is important to highlight that in order to improve the image quality and reduce the noise produced by the speckle phenomenon, a filtering technique based on the improved sigma lee filter was applied. This technique allowed obtaining sharper and

clearer images, facilitating the identification and analysis of objects and features present in the image [29]. Finally, two image conversions were performed to improve interpretation Figure 4: from linear units to decibel units to expand the dynamic range and highlight areas of higher or lower signal intensity, and from decibels to power backscatter to show the amount of energy backscattered by objects in the scene, following the methodology of study [10].

A two-step process was carried out to process the GEDI image collection and ensure accurate data integration. First, the collection was structured using GEE and the rh95 metric was selected as the variable of interest. Subsequently, geospatial processing procedures were carried out for vectorization of the selected metrics, followed by their export as a shapefile. To ensure spatial consistency and accurate alignment, co-registration of GEDI and ALOS PALSAR imagery was performed prior to integration into a multiband raster file. This process ensured that each pixel of both images corresponded to exactly the same geographic location within the dry forest study area [30]. The final stack included the height metrics (rh95) and polarimetric indices mentioned previously. Once the shapefile was imported as an asset to GEE, the corresponding data from the image stack was extracted. This allowed for the creation of a CSV database with the GEDI and ALOS PALSAR-2 data, which was used to structure the regression algorithm in Python in Google Colaboratory, allowing for detailed and precise analysis of the information.
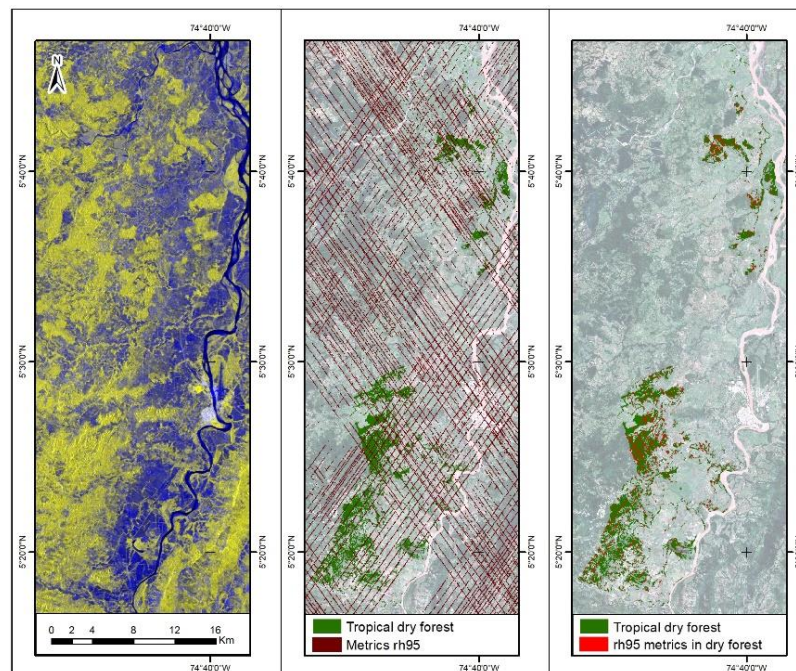

Figure 4. On the left is a visualization of the ALOS PALSAR-2 image in polarization combination. In the center, all metrics within the quadrant are presented. On the right, the metrics within the specific dry forest area are shown

2.4. Structuring of the random forest algorithm

After importing the shapefile as an asset in GEE, an exploratory analysis of the dataset was conducted, which included obtaining descriptive statistics as well as creating and analyzing histograms and boxplots for each variable. The goal was to evaluate the presence of missing and outlier values. Values that could affect the quality of the results were removed. After the exploratory analysis, various activities were carried out to further analyze the obtained data. Firstly, the Shapiro-Wilk test was used to evaluate the normality of the data [31]. Secondly, Spearman correlation was calculated to evaluate the nonlinear relationship between two variables, and scatter plots were generated to visualize the relationship between canopy height and SAR bands [32]. These activities allowed identifying the key relationships between the variables, which contributed to the selection of the most relevant variables for the structure of the random forest algorithm.

Once the variables were selected, we proceeded to implement the random forest model to predict canopy height. Using the Scikit-learn library in Python, the independent variables ('HH', 'HV', 'VV', 'RVI', 'RFDI', 'NDBI', 'AI', and 'DI') were assigned to the X variable, while the dependent variable 'GEDI rh95' was assigned to the Y variable. The model implementation was based on the random forest regressor class of

Scikit-learn. To improve the results, a function was created that allowed iterating between different values of hyperparameters such as number of trees, maximum depth and random seed, allowing to select the model that best fit the data. This strategy helped to select the model that best fit the data and optimally captured the relationships between the independent variables and the dependent variable. In addition, a 5-iteration cross-validation was applied to assess the accuracy of the model. This technique allowed comparing the results obtained with the values of the rh95 metric and evaluating the predictive capacity of the model using a specific metric on different data sets to reduce the risk of overfitting and increase the reliability of the results. Subsequently, the model was fitted using all training data and predictions were made. Evaluation metrics, such as the coefficient of determination (R2), root mean squared error (MSE) and root mean squared error (RMSE), were calculated to assess model performance. In addition, graphs comparing actual values with predicted values were generated, allowing visualization of the quality of the predictions and the relationship between observed and estimated values.

3. RESULTS AND DISCUSSION

After carrying out the descriptive data analysis that included several activities, such as obtaining descriptive statistics, generating and analyzing histograms and boxplots, and removing outliers for each variable, the Shapiro-Wilk test was performed to assess the normality of the data. The initial number of records in the dataset was a total of 1,028. After removing outliers, a total of 957 records remained in the dataset Figure 5. For this analysis, a significance level of 0.05 was used to evaluate the normality of the data using the Shapiro-Wilk, Anderson-Darling, Cramér-von Mises, and Kolmogorov-Smirnov tests. This test generated a test statistic and a p-value, where if the p-value is less than 0.05, the null hypothesis that the data come from a normal distribution can be rejected. After applying the test, the variables HH, NDBI, AI, DI, and GEDI rh95 result in not normality. Upon finding that more than 50% of the data did not meet the normality assumption, it was decided to use non-parametric statistics. Therefore, the Spearman correlation test [32] was performed to evaluate the relationship between the variables of interest, which is suitable for non-normally distributed variables. However, it was also decided to calculate the Spearman correlation matrix to observe the positive and negative correlations between the variables.
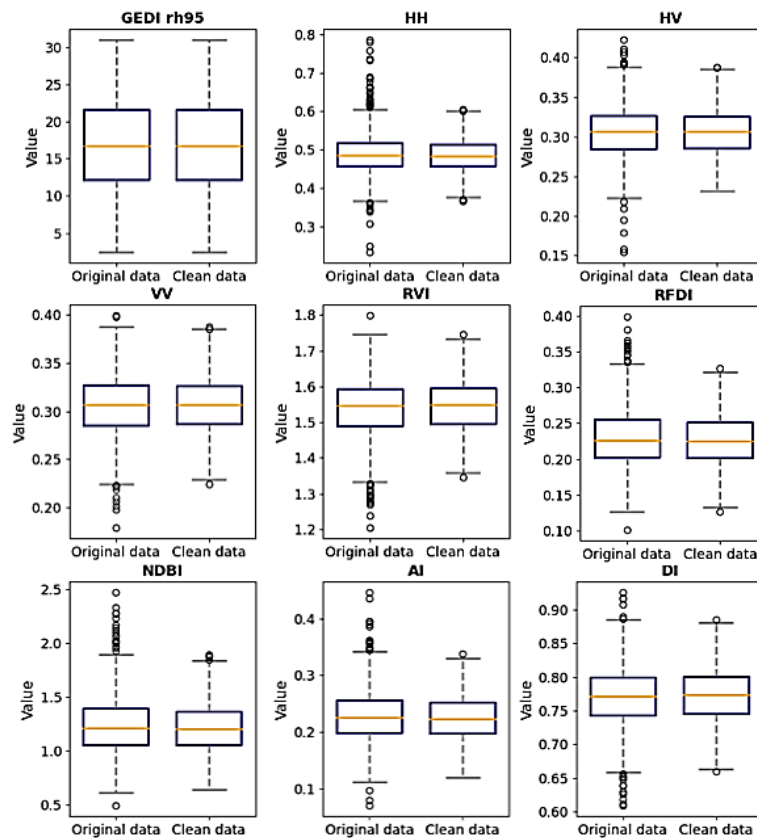

Figure 5. Box plots for each variable before and after outlier cleanup

After applying the Spearman correlation test to the research data, a correlation matrix was generated that allowed the identification of positive and negative correlations between variables. This is important for selecting the variables to be included in the model for estimating heights with GEDI and ALOS PALSAR-2, avoiding the inclusion of variables with a high degree of correlation among them, which can affect the accuracy of the model and generate multicollinearity. The Spearman correlation matrix showed that the dependent variable "GEDI rh95" positively correlates with the variable RVI and negatively correlates with the variables RFDI and NDBI. The p-values of these correlations were significant, suggesting that these variables are relevant in estimating forest height. These correlations are also useful in identifying variables that have high correlation with each other and reducing the complexity of the model.

3.1. Random forest model application

The random forest machine learning algorithm was used, which combines the prediction of multiple decision trees to improve accuracy and reduce overfitting [6]. The scikit-learn library of Python was used to build the model, and the k-fold cross-validation technique was applied to evaluate its predictive ability. For each independent variable (HH, HV, VV, RVI, RFDI, NDBI, AI, and DI), a for loop was executed to build a model and obtain specific metrics and scatter plots. A random forest model with 100 trees was constructed, and 5-fold cross-validation was used for each independent variable. This process involves dividing the data into k subsets and training the model on k-1 subsets, using the other subset as a validation set. This was repeated k times to obtain a more accurate evaluation of the model's generalization ability.

Scatter plots were generated for each of the eight independent variables obtained from the SAR bands and vegetation height obtained by GEDI (rh95) to analyze their relationship. Each plot allows visualizing the relationship between an independent variable and vegetation height in the study area, and a total of eight plots were obtained, one for each independent variable. Subsequently, the model was fit with all training data, and predictions were made with all data. To evaluate the predictive ability of the model, evaluation metrics (R^2, MSE, and RMSE) were calculated for each independent variable. The results were plotted, showing the actual vs. predicted values for each independent variable in a scatter plot Figure 5. Overall, a positive relationship was observed between the actual and predicted values for all independent variables. Figure 6 shows the eight scatter plots, where the corresponding independent variable for each SAR band is on the x-axis, and the vegetation height (rh95) obtained by GEDI is on the y-axis. Each plot shows the scatter plot of data points and the fitted regression line, which represents the relationship between the independent variable and canopy height.

Based on the results obtained, it can be observed that the variables with the highest relationship to GEDI rh95 are NDBI and RVI, with R^2 values of 0.76 and 0.72, respectively. These polarimetric indices are related to the response of the tropical dry forest because NDBI is based on the difference in backscattering between horizontal and vertical waves, allowing the detection of degraded areas [33]. In the case of RVI, this index is related to the density and vertical structure of vegetation, making it a good measure for evaluating forest biomass [34].

On the other hand, the variables with a lower relationship to GEDI rh95 are VV and HV, with R^2 values of 0.53 and 0.57, respectively. These polarizations are related to the response of the dry forest because SAR polarimetry is sensitive to the structure and geometry of vegetation, allowing the detection of changes in forest structure [27]. However, in this case, it can be observed that the relationship with GEDI rh95 is lower, which may be due to the low density of vegetation in the tropical dry forest ecosystem. Regarding the other variables, it can be observed that they have a moderate relationship with GEDI rh95, with R^2 values ranging between 0.61 and 0.64. These polarimetric indices are related to the response of the tropical dry forest because SAR polarimetry allows for the characterization of vegetation structure and geometry, which can detect changes in forest structure. However, it is observed that the relationship with GEDI rh95 is lower than that of NDBI and RVI, suggesting that these indices may not be as sensitive for evaluating forest biomass in the tropical dry forest ecosystem. Taking this into account, NDBI is the variable with the best relationship for predicting canopy height measured from rh95. Performance plots of the model are presented in Figure 7. In the real value distribution plot, it can be observed that most values are between 2.5 and 30, with a higher concentration around 17.5. On the other hand, the predicted value distribution plot shows a similar distribution, which suggests that the model can predict rh95 values with some precision. In the error distribution plot, it can be observed that most errors are around 0, indicating that the model does not have a systematic tendency to overestimate or underestimate rh95 values.

In summary, it can be concluded that the NDBI variable is the one that best explains the relationship with rh95, as evidenced by the high R^2 value obtained. The evaluation of the performance graphs supports this conclusion by showing a similar distribution between the actual and predicted values, and a low concentration of errors around 0. The RF model with this input variable shows a good fit and predictive capacity and can be used to estimate forest height with reasonable accuracy.
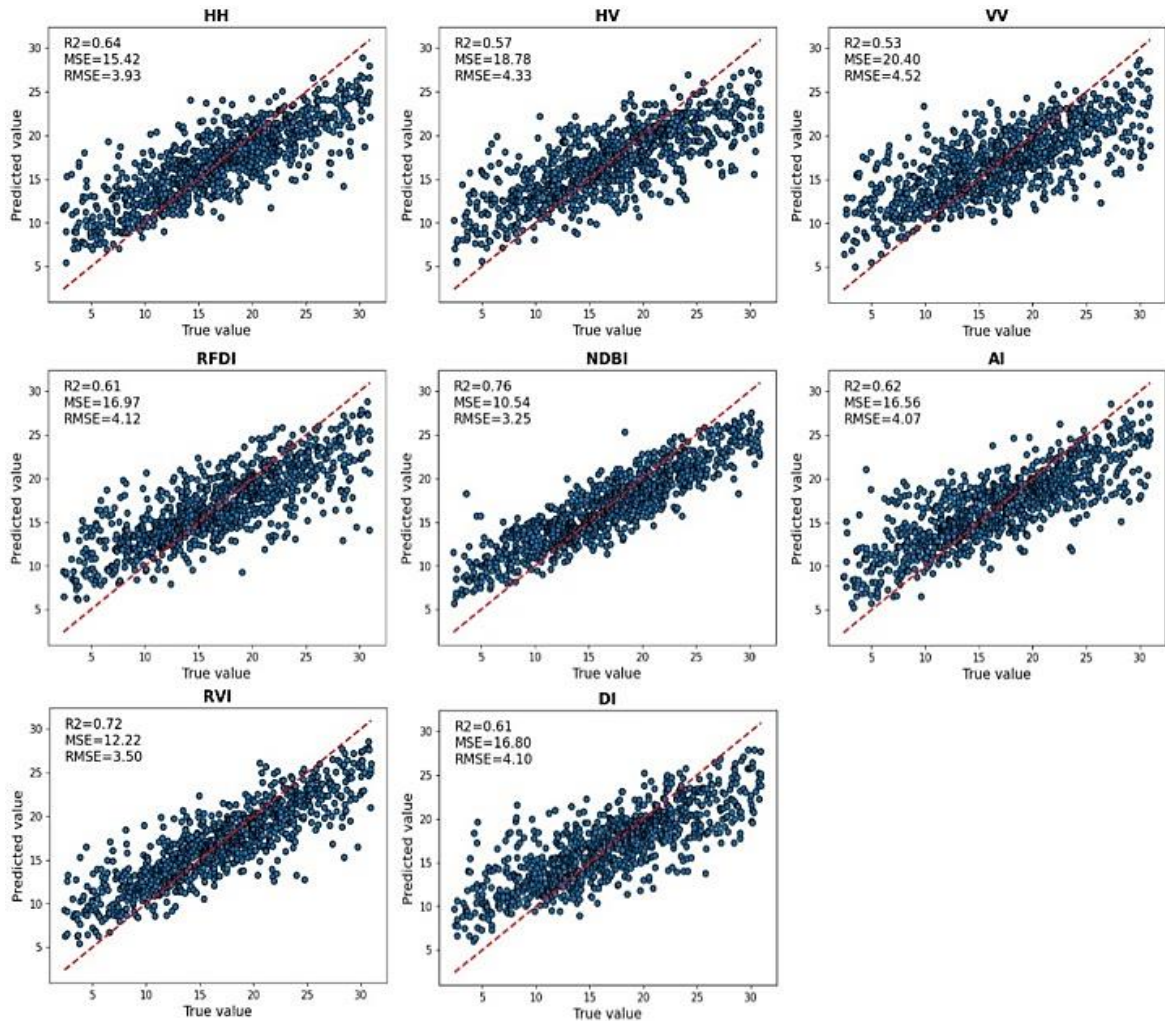
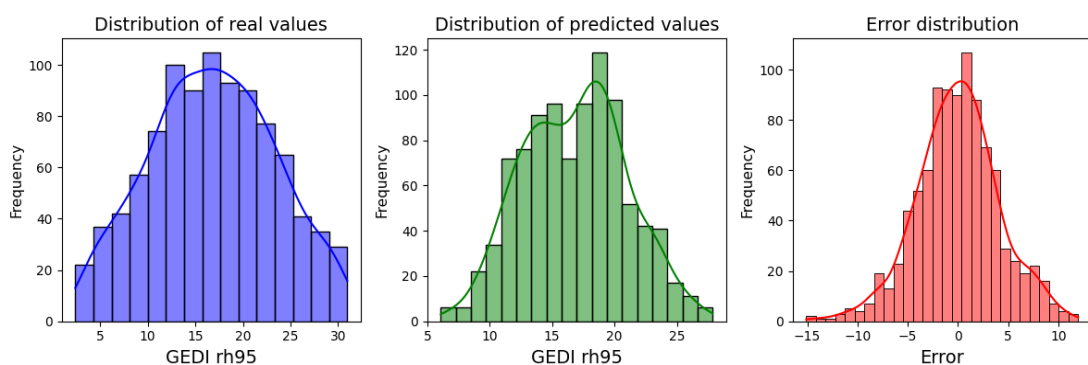Figure 6. Scatter plots for each independent variable and its relationship to rh95

Figure 7. Performance graphs of the NDBI variable with the rh95 metric

4. CONCLUSION

After carefully examining the study's findings, it can be said that using polarimetric indices to estimate vegetation height is a useful and promising method. Estimates of the height of the forest canopy can be more accurately and thoroughly determined thanks to the extra information about the structure and composition of vegetation that is provided by polarimetric analysis of SAR data. In particular, the regression between polarimetric indices and the dependent variable GEDI rh95 has found use of the RF algorithm to be a valuable

tool. The method is a good choice for this kind of study since it can manage a large number of independent variables and their intricate non-linear relationships with the dependent variable. It is crucial to remember that the accuracy and dependability of vegetation height estimations in tropical forests can be increased by combining data from several sources, such as that collected from SAR satellites and GEDI mission data. The management and maintenance of these ecosystems can benefit greatly from the information that these analysis approaches can offer. Additionally, the significance of using the Google Earth Engine platform for effectively processing and scaling up enormous volumes of data is emphasized. This platform allowed for the rapid and accurate completion of this research since it allowed for the processing and analysis of a sizable volume of SAR pictures and GEDI data in a reasonable length of time. In conclusion, this study highlights the significance of RF algorithm and SAR polarimetry for increasing the precision of vegetation height estimation in tropical forests. These findings highlight the necessity of combining various data sources and analysis methods in order to better comprehend and monitor forest ecosystems.

REFERENCES

[1]   E. U. Botero, "Climate change and its effects on biodiversity in Latin America (in Spanyol)," *Comisión Económica para América Latina y el Caribe (CEPAL)*, 2016. Accessed Apr. 27, 2023. [Online], Available: https://repositorio.cepal.org/bitstream/handle/11362/39855/S1501295_en.pdf?sequence=1

[2]   S. Abbas, M. S. Wong, J. Wu, N. Shahzad, and S. M. Irteza, "Approaches of satellite remote sensing for the assessment of above-ground biomass across tropical forests: Pan-tropical to national scales," *Remote Sensing*, vol. 12, no. 20, Oct. 2020, doi: 10.3390/rs12203351.

[3]   A. Karandikar and A. Agrawal, "Performance analysis of change detection techniques for land use land cover," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 4, pp. 4339–4346, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4339-4346.

[4]   J. C. Fagua, P. Jantz, S. Rodriguez-Buritica, L. Duncanson, and S. J. Goetz, "Integrating LiDAR, multispectral and SAR data to estimate and map canopy height in tropical forests," *Remote Sensing*, vol. 11, no. 22, Nov. 2019, doi: 10.3390/rs11222697.

[5]   S. Swetanisha, A. R. Panda, and D. K. Behera, "Land use/land cover classification using machine learning models," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 2, pp. 2040–2046, Apr. 2022, doi: 10.11591/ijece.v12i2.pp2040-2046.

[6]   L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[7]   R. Gupta and L. K. Sharma, "Mixed tropical forests canopy height mapping from spaceborne LiDAR GEDI and multisensor imagery using machine learning models," *Remote Sensing Applications: Society and Environment*, vol. 27, Aug. 2022, doi: 10.1016/j.rsase.2022.100817.

[8]   M. R. Pourrahmati *et al.*, "Capability of GLAS/ICESat data to estimate forest canopy height and volume in mountainous forests of Iran," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 11, pp. 5246–5261, Nov. 2015, doi: 10.1109/JSTARS.2015.2478478.

[9]   K. D. Fieber *et al.*, "Validation of canopy height profile methodology for small-footprint full-waveform airborne LiDAR data in a discontinuous canopy environment," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 104, pp. 144–157, Jun. 2015, doi: 10.1016/j.isprsjprs.2015.03.001.

[10] S. Saatchi, "SAR Methods for Mapping and Monitoring Forest Biomass," in *The Synthetic Aperture Radar (SAR) Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation, A. Flores, K. Herndon, R. Bahadur, W. L. Ellenburg, S. Thapa and E. Cherrington, Eds.*, California: SAR, 2019, pp. 115–145. Accessed: Aug 25, 2023. [Online]. Available: https://servirglobal.net/Global/Articles/Article/2674/sar-handbook-comprehensive-methodologies-for-forest-monitoring-and-biomass-estimation

[11] L. Duncanson *et al.*, "Aboveground biomass density models for NASA's global ecosystem dynamics investigation (GEDI) lidar mission," *Remote Sensing of Environment*, vol. 270, Mar. 2022, doi: 10.1016/j.rse.2021.112845.

[12] U. Khati, M. Lavalle, and G. Singh, "The role of time-series L-band SAR and GEDI in mapping sub-tropical above-ground biomass," *Frontiers in Earth Science*, vol. 9, Oct. 2021, doi: 10.3389/feart.2021.752254.

[13] X. Zhou *et al.*, "Improving GEDI forest canopy height products by considering the stand age factor derived from time-series remote sensing images: A case study in Fujian, China," *Remote Sensing*, vol. 15, no. 2, Jan. 2023, doi: 10.3390/rs15020467.

[14] P. Scarth, J. Armston, R. Lucas, and P. Bunting, "A structural classification of Australian vegetation using ICESat/GLAS, ALOS PALSAR, and landsat sensor data," *Remote Sensing*, vol. 11, no. 2, Jan. 2019, doi: 10.3390/rs11020147.

[15] P. Lourenço, "Biomass estimation using satellite-based data," in *Forest Biomass - From Trees to Energy*, IntechOpen, 2021. doi: 10.5772/intechopen.93603.

[16] D. Morin *et al.*, "Improving heterogeneous forest height maps by integrating GEDI-based forest height information in a multi-sensor mapping process," *Remote Sensing*, vol. 14, no. 9, Apr. 2022, doi: 10.3390/rs14092079.

[17] M. Urbazaev, F. Cremer, M. Migliavacca, M. Reichstein, C. Schmullius, and C. Thiel, "Potential of multi-temporal ALOS-2 PALSAR-2 ScanSAR data for vegetation height estimation in Tropical Forests of Mexico," *Remote Sensing*, vol. 10, no. 8, Aug. 2018, doi: 10.3390/rs10081277.

[18] A. Benitez *et al.*, *The tropical dry forest in Colombia* (in Spanyol). Alexander von Humboldt Biological Resources Research Institute, 2014. Accessed: Aug 25, 2023. [Online]. Available: http://repository.humboldt.org.co/handle/20.500.11761/9333

[19] L. Miles *et al.*, "A global overview of the conservation status of tropical dry forests," *Journal of Biogeography*, vol. 33, no. 3, pp. 491–505, Mar. 2006, doi: 10.1111/j.1365-2699.2005.01424.x.

[20] A. Tassi and M. Vizzari, "Object-oriented LULC classification in Google earth engine combining SNIC, GLCM, and machine learning algorithms," *Remote Sensing*, vol. 12, no. 22, Nov. 2020, doi: 10.3390/rs12223776.

[21] R. Dubayah and J. B. Blair, "Global ecosystem dynamics investigation (GEDI) level 2 user guide," *United States Geological Survey*, 2021.

[22] C. A. Silva *et al.*, "Fusing simulated GEDI, ICESat-2 and NISAR data for regional aboveground biomass mapping," *Remote Sensing of Environment*, vol. 253, Feb. 2021, doi: 10.1016/j.rse.2020.112234.

[23] JAXA, "PALSAR phased array type L-band synthetic aperture radar," *ALOS PALSAR*, 2020. https://www.eorc.jaxa.jp/ALOS/en/alos/sensor/palsar_e.htm (accessed: Aug 25, 2023).

[24] S. Mermoz, M. Réjou-Méchain, L. Villard, T. Le Toan, V. Rossi, and S. Gourlet-Fleury, "Decrease of L-band SAR backscatter with biomass of dense forests," *Remote Sensing of Environment*, vol. 159, pp. 307–317, Mar. 2015, doi: 10.1016/j.rse.2014.12.019.

[25] M. Shimada *et al.*, "New global forest/non-forest maps from ALOS PALSAR data (2007–2010)," *Remote Sensing of Environment*, vol. 155, pp. 13–31, Dec. 2014, doi: 10.1016/j.rse.2014.04.014.

[26] S. Sinha, "H/A/α polarimetric decomposition of dual polarized Alos Palsar for efficient land feature detection and biomass estimation over tropical deciduous forest," *Geography, Environment, Sustainability*, vol. 15, no. 3, pp. 37–46, Oct. 2022, doi: 10.24057/2071-9388-2021-095.

[27] R. Nasirzadehdizaji, F. B. Sanli, S. Abdikan, Z. Cakir, A. Sekertekin, and M. Ustuner, "Sensitivity analysis of multi-temporal sentinel-1 SAR parameters to crop height and canopy coverage," *Applied Sciences*, vol. 9, no. 4, Feb. 2019, doi: 10.3390/app9040655.

[28] P. Zeng, W. Zhang, Y. Li, J. Shi, and Z. Wang, "Forest total and component above-ground biomass (AGB) estimation through C- and L-band polarimetric SAR Data," *Forests*, vol. 13, no. 3, Mar. 2022, doi: 10.3390/f13030442.

[29] D. Closson, F. Holecz, P. Pasquali, and N. Milisavljevic, *Land applications of radar remote sensing*, InTech, 2014, doi: 10.5772/55833.

[30] T.-A. Teo and S.-H. Huang, "Automatic co-registration of optical satellite images and airborne Lidar data using relative and absolute orientations," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 5, pp. 2229–2237, Oct. 2013, doi: 10.1109/JSTARS.2012.2237543.

[31] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, Dec. 1965, doi: 10.2307/2333709.

[32] J. C. F. de Winter, S. D. Gosling, and J. Potter, "Comparing the Pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data.," *Psychological Methods*, vol. 21, no. 3, pp. 273–290, Sep. 2016, doi: 10.1037/met0000079.

[33] D. Dobrinić, M. Gašparović, and D. Medak, "Sentinel-1 and 2 time-series for vegetation mapping using random forest classification: A case study of Northern Croatia," *Remote Sensing*, vol. 13, no. 12, Jun. 2021, doi: 10.3390/rs13122321.

[34] C. Szigarski *et al.*, "Analysis of the radar vegetation index and potential improvements," *Remote Sensing*, vol. 10, no. 11, Nov. 2018, doi: 10.3390/rs10111776.

BIOGRAPHIES OF AUTHORS

Christian Javier Pinza-Jiménez 🆔 📧 SC ▸ is geographer graduated from the Universidad de Nariño and holds a master's degree in remote sensing from the Universidad Católica de Manizales, Colombia. He has extensive experience in geomatics and remote sensing, particularly in applications related to land-use planning, disaster risk management, and natural resource conservation. Additionally, he has a strong interest in applying cloud computing techniques for processing optical and SAR images to investigate and characterize forest cover, study changes in land cover, and analyze various environmental issues. He is also interested in applying machine learning algorithms to predict different parameters related to forest monitoring. He can be contacted at email: geo.christianpinza@gmail.com.

Yeison Alberto Garces-Gómez 🆔 📧 SC ▸ Received bachelor's degree in electronic engineering, and master's degrees and Ph.D. in Engineering from Electrical, Electronic and Computer Engineering Department, Universidad Nacional de Colombia, Manizales, Colombia, in 2009, 2011 and 2015, respectively. He is full professor at the academic unit for training in natural sciences and mathematics, Universidad Católica de Manizales, and teaches several courses such as experimental design, statistics, physics. His main research focus is on applied technologies, embedded system, power electronics, power quality, but also many other areas of electronics, signal processing and didactics. He published more than 30 scientific and research publications, among them more than 10 journal papers. He worked as principal researcher on commercial projects and projects by the Ministry of Science, Tech and Innovation, Republic of Colombia. He can be contacted at email: ygarces@ucm.edu.co.