Estimation of Above Ground Biomass Using GEDI Data and Remote Sensing in La Joya - La Barreta Ecological Park, Querétaro

[1]Luis Felipe Castelblanco Rivera, [2]Nelly Lucero Ramírez-Serrato, [3]Samuel Villarreal-Rodríguez, [1]Mario Guevara.

[1]Instituto de Geociencias, Universidad Nacional Autónoma de México, campus Juriquilla, Querétaro, México
[2]Departamento de Recursos Naturales, Instituto de Geofísica, Universidad Nacional Autónoma de México, Coyoacán, Mexico
[3]Departamento de Ingeniería Sustentable, Centro de Investigación en Materiales Avanzados, S.C, Subsede Durango, Durango, Mexico

Keywords: Above Ground Biomass (AGB), GEDI, Remote Sensing, Biomass Density, Ecological Park.

Abstract

Accurate estimation of aboveground biomass (AGB) is crucial to understanding the carbon cycle in conservation areas. This study developed a predictive model for AGB density in La Joya - La Barreta Ecological Park, Querétaro, by integrating LiDAR data from the GEDI mission with passive sensor information (Sentinel-1 and Sentinel-2) and topographic variables. This park is a vital space for recreation, environmental education, and carbon credit generation. We used Global Ecosystem Dynamics Investigation (GEDI) AGB density data from April 2024, with values ranging from 2.5 to 368.2 Mg/ha. For modelling, we processed satellite imagery and a digital terrain model, generating a comprehensive set of 178 spectral indices and topographic variables. Through Recursive Feature Elimination (RFE), five optimal covariates were selected: the LS-Factor, Analytical Hillshading, and Channel Network Base Level (topographic variables), along with Sentinel-2's Triangle Water Index (TWI) and Enhanced Modified Bare Soil Index (EMBI) spectral indices. The Quantile Regression Forest (QRF) model predicted AGB density with an RMSE of 1.8 Mg/ha and an R^2 of 0.59, indicating a robust predictive capability for local applications. AGB density predictions ranged from 5.2 to 113.4 Mg/ha, with a mean of 32.9 Mg/ha. A total biomass of 8171.9 Mg was estimated for the park, containing 3652.8 Mg of organic carbon. These results provide a cost-effective basis for monitoring and verifying the park's conservation projects, highlighting the importance of topography and spectral indices in biomass distribution.

1. Introduction

Forests store approximately 80% of the terrestrial biomass (Gardon et al., 2020). Accurately estimating of Above Ground Biomass (AGB) is crucial for sustainable forest management, carbon cycle monitoring, and quantify ecosystem health and services (Herold et al., 2019; Reichstein & Carvalhais, 2019) and to improve our understanding of terrestrial ecosystems (Stoffel et al., 2008).

With the intensification of global warming, monitoring forest biomass and carbon reserves has become increasingly critical. This need has prompted initiatives such as REDD+ (Reducing Emissions from Deforestation and Forest Degradation). Established by the United Nations Framework Convention on Climate Change, REDD+ highlights the economic importance of forest carbon sequestration, making accurate and reliable assessments of carbon stock essential (Pati et al., 2022).

Field-based AGB measurements rely on destructive sampling methods, which are used to develop of allometric equations for plot sampling. Despite their high accuracy, these methods are limited in spatial coverage (Lefsky et al., 2002). Remote sensing offers an efficient alternative for mapping AGB at broad scales. For example, building on such advancements, Reichstein & Carvalhais (2019) developed a reference database by integrating Synthetic Aperture Radar (SAR) and optical remote sensing data to estimate AGB in the Amazon rainforest. They assessed the performance of machine learning algorithms, specifically Random Forest (RF) and Extreme Gradient Boosting (XGBoost), both of which yielded strong results.

The Global Ecosystem Dynamics Investigation (GEDI) LiDAR instrument, launched in late 2018 and mounted on the International Space Station (ISS), plays a key role in monitoring forest ecosystems by providing detailed information on canopy structure and biomass (Dubayah et al., 2020). The GEDI collects waveform LiDAR allows the characterization of the structure of forest canopies, delivering key metrics such as surface elevation, topography, canopy height, relative height metrics, plant area index (PAI), and gridded above-ground biomass (Potapov et al., 2021). Although GEDI data consists of sparsely distributed footprints, machine learning (ML) models are used to achieve continuous and comprehensive estimates of canopy height, integrating GEDI measurements with optical and SAR remote sensing data (Bhuyan et al., 2024; Jiang et al., 2021).

Recent advances have expanded AGB modeling by integrating GEDI with complementary sensors such as ICESat-2 and airborne LiDAR, which enhance canopy structure characterization (Guo et al., 2023; Jiang et al., 2021). Other studies have explored deep learning methods (e.g., convolutional neural networks) for mapping biomass at regional scales (Bhuyan et al., 2024). However, such approaches often require extensive computational resources or dense field calibration data. In contrast, our workflow prioritizes a cost-effective and replicable methodology using open-access datasets (GEDI, Sentinel-1/2, and DEM), tailored for local-scale conservation and carbon monitoring applications. This framework contributes to bridging the gap between global-scale GEDI applications and local conservation management, particularly in semi-arid environments where data availability is limited.

This research explored the potential of integrating the GEDI LiDAR mission with the passive sensor data (Sentinel-1 and Sentinel-2) and topographic variables to estimate AGB density in La Joya - La Barreta ecological park in Querétaro. Besides being a recreational and environmental education space, the park is a nature reserve prioritized by the municipal

https://doi.org/10.5194/isprs-annals-X-3-W3-2025-29-2026 |

administration for generating carbon credits through carbon sequestration in both soil and vegetation.

The main objective was to develop a robust AGB predictive model using AGB density data derived from the GEDI program and covariates obtained from satellite images and digital elevation models. This aimed to establish a cost-effective methodology for monitoring and verifying conservation projects in La Joya - La Barreta ecological park in Querétaro.

2. Methodology

The methodological workflow consisted of four main steps: (1) acquisition and preprocessing of GEDI above-ground biomass density data; (2) generation of environmental covariates from Sentinel-1, Sentinel-2, and topographic information; (3) selection of the most relevant predictors using Recursive Feature Elimination (RFE); and (4) spatial prediction of AGB density using a Quantile Regression Forest (QRF) model, along with uncertainty estimation. The following subsections describe each step in detail.

2.1 Study Area

The park is located in the northwest portion of the municipality of Querétaro, in the state of Querétaro, Mexico, in the Santa Rosa de Jauregui municipal delegation, lying between 20° 49′ 20″ N to 20° 48′ 24″ N latitude and 100° 32′ 17″ W to 100° 30′ 39″ W longitude, spanning an area of 247.91 ha. The park covered by submontane scrubland, pine-oak forest, grassland, and scrubland (Figure 1).

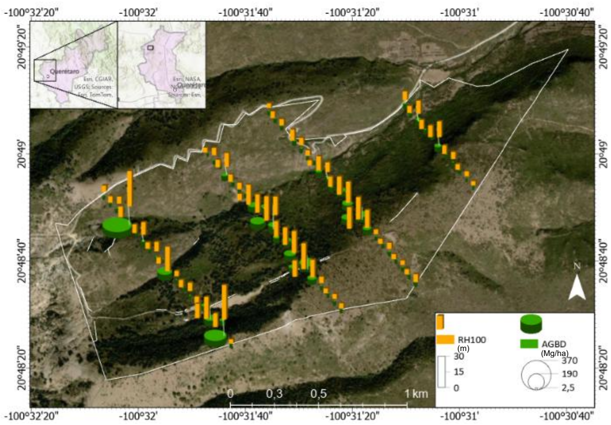

Figure 1. Location of La Joya–La Barreta Ecological Park, Querétaro, Mexico, and visualization of the GEDI AGB density, (Mg/ha) product for April 2024. Insets show the vegetation height (RH100: Relative Height at which 100% of the waveform energy is accumulated).

2.2 Above Ground Biomass

For this study, we used AGB density data derived from the GEDI LiDAR mission product (Dubayah et al., 2020), recorded during April 2024 in La Joya - La Barreta ecological park (Figure 1). were used. AGB density values are expressed in megagrams per hectare (Mg/ha), recorded values range from a minimum of 2.5 Mg/ha, a maximum of 368.2 Mg/ha, a mean of 20.6 Mg/ha, and a standard deviation of 66.3 Mg/ha. This product is derived from vegetation structure metrics obtained through the analysis of return intensity and waveform. One of the key variables for estimating AGB density is vegetation canopy height, which is obtained from the RH100 waveform response, representing the relative height above ground where 100% of the LIDAR waveform return energy has accumulated. This height exhibits a high correlation (correlation coefficient of 0.95) with the estimated biomass (Figure 2).

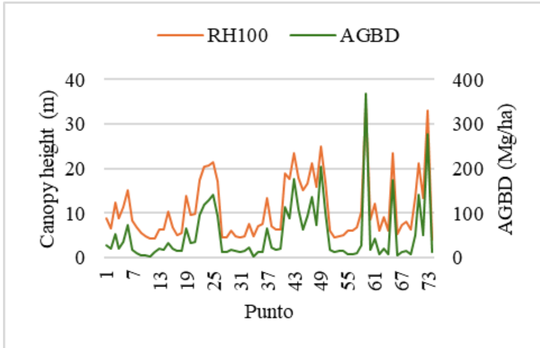

Figure 2. Canopy height (RH100) and AGB density values from the GEDI mission in La Joya – La Barreta park.

The AGB density data obtained from the GEDI mission were first pre-processed by removing outliers with values exceeding the mean plus 1.5 times the standard deviation (151.64 Mg/ha). In a second step, they were transformed using a logarithm to reduce their asymmetry and improve the model's ability to represent their variability.

2.3 Environmental Covariates

The Digital Elevation Model (DEM) used corresponds to a bare-earth Digital Terrain Model (DTM) obtained through the SRTM-derived Copernicus DEM (GLO-30) dataset, accessed via the elevatr package (Hollister et al., 2023) with a spatial resolution of 8.92 m. Topographic covariates were generated from terrain analysis using the basic terrain analysis tools of SAGA GIS (Conrad et al., 2015). Additionally, Sentinel-1 and Sentinel-2 satellite imagery for the study area was acquired using the rsi package in R (Mahoney et al., 2025) for April 2024. From the Sentinel-2 images, a total of 178 spectral indices were calculated, covering a wide range of vegetation, bare soil, and moisture indices available in the Awesome Spectral Indices repository (Montero et al., 2023). A subset of these variables is shown in Figure 3.

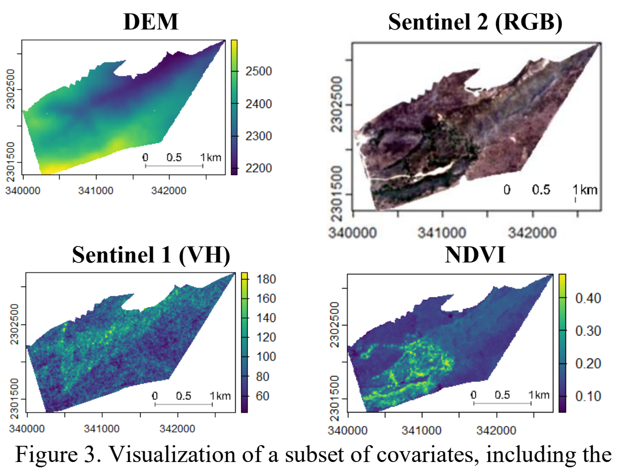

Figure 3. Visualization of a subset of covariates, including the Digital Elevation Model (DEM), a true-color visualization of the Sentinel-2 image, Sentinel-1 cross-polarization radar response (VH), and the Normalized Difference Vegetation Index (NDVI).

https://doi.org/10.5194/isprs-annals-X-3-W3-2025-29-2026

In total, 206 environmental covariates were generated to ensure comprehensive coverage of potential predictors influencing above-ground biomass. This large set was designed to capture diverse biophysical processes related to vegetation structure, moisture, and topography. The inclusion of numerous indices follows the exploratory approach commonly used in biomass modeling studies (e.g., Guo et al., 2023; Potapov et al., 2021) where an extensive pool of spectral and topographic variables allows data-driven feature selection. Although only five covariates were retained by the RFE process, starting with a broader set ensured that potentially important predictors were not excluded a priori, as their relevance can vary depending on vegetation type, scale, and sensor characteristics.

2.4 Modeling

Recursive Feature Elimination (RFE) (Kuhn, 2008) was used to select the covariates that contributed most significantly to the prediction of the biomass. This method allowed the identification of an optimal subset of covariates that best explain biomass variability while minimizing the Root Mean Square Error (RMSE).

To model the relationship between the selected covariates and the AGB density, the Quantile Regression Forest (QRF) algorithm (Meinshausen, 2006) was implemented. This approach provided both the value prediction and the standard deviation as a measure of prediction uncertainty. The workflow is schematized in Figure 4.

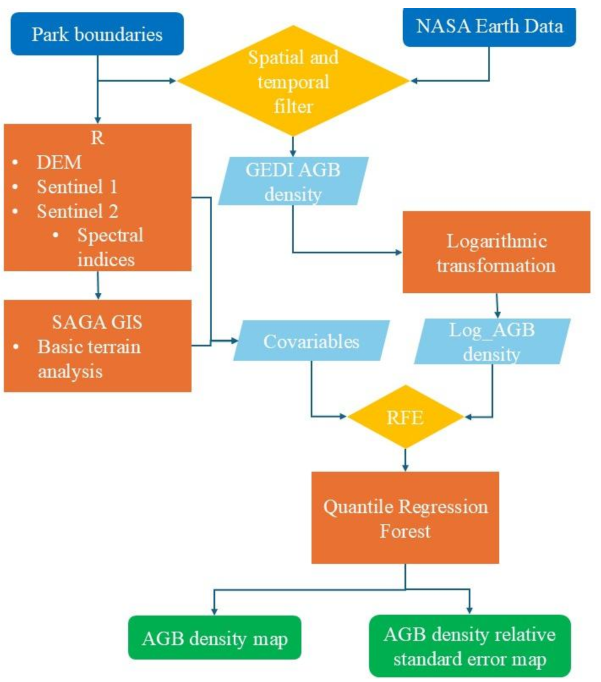

Figure 4. Workflow diagram for obtaining the variable of interest, environmental covariates, and modeling.

The feature selection process using RFE evaluated 206 covariates and identified the optimal set of five covariates to minimize RMSE in AGB density prediction (Figure 5). This subset includes three topographic variables: the LS-Factor (Slope Length), Analytical Hillshading, and Channel Network Base Level; and two spectral indices derived from Sentinel-2: TWI (Niu et al., 2022) which captures moisture variation based on infrared spectral response, and EMBI (Zhao & Zhu, 2022), an enhanced vegetation greenness index with a shortwave infrared correction to reduce its sensitivity to soil background distortions (Figure 6).

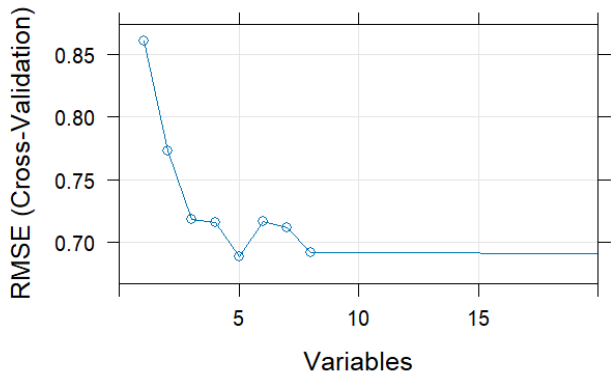

Figure 5. Variable selection plot with the RFE method using minimum RMSE.

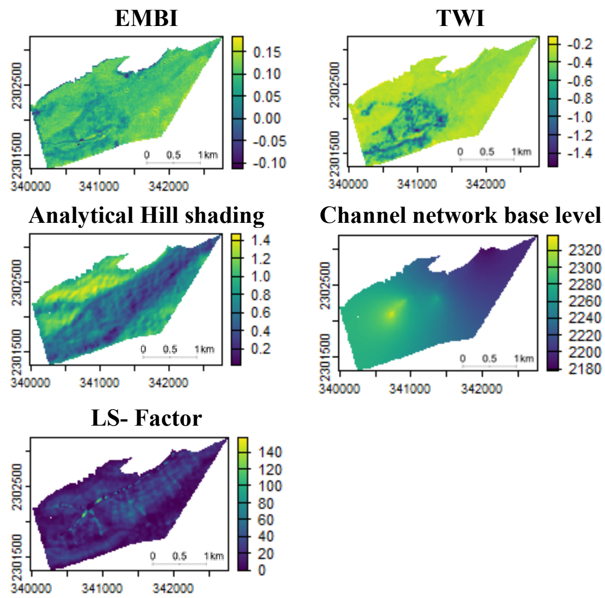

Figure 6. Visualization of the covariates selected by the RFE method.

Finally, using the trained QRF model and the rasters of the selected covariates, the spatial predictions of the most probable AGB density value and their associated standard deviation were performed. The standard deviation was used to calculate the standard relative error, which served as a measure of the uncertainty of the predictions using the following equation:

$$RSE = \frac{sd}{mpv} * 100 \tag{1}$$

where RSE = relative standard error
sd = standard deviation at each pixel
mpv = most probable value at each pixel

The Quantile Regression Forest (QRF) algorithm (Meinshausen, 2006) was selected because it extends the Random Forest approach by estimating conditional distributions rather than only mean predictions, thus allowing direct quantile-based uncertainty assessment. Compared to other ensemble methods such as XGBoost or Support Vector Machines, QRF provides more stable predictions in small-to-moderate sample sizes and is less sensitive to hyperparameter tuning, which is advantageous for spatial modeling. Although we report mean predictions and standard deviations for clarity, the model inherently captures full quantile distributions, which could be further explored in future analyses to refine uncertainty interpretation.

2.5 Model Performance Evaluation

The QRF model, based on decision trees, performs bootstrapping (Bagging) of the data to create new training sets for each tree. Using this feature, data not used for training a tree are used to predict its value as an individual external validation called Out-of-bag. With this method, model performance metrics such as the Root Mean Square Error (RMSE) and coefficient of determination (R^2) are obtained and reported as model performance metrics.

This methodological workflow provides a replicable and transparent framework for integrating GEDI LiDAR data with multisensor covariates in biomass modeling, aligning with best practices in recent remote sensing applications

3. Results

The estimation of AGB through the integration of GEDI data, satellite imagery, and topographic variables in La Joya - La Barreta ecological park allowed the generation of high-spatial resolution AGB density maps. The results show the spatial distribution of the AGB density variable in its original and transformed form (Figure 7), which helped to identify the variability patterns associated with the topographic and structural characteristics of the ecosystem. The validation of the predictive model, based on cross-resampling techniques and training/test data partitioning, demonstrated a satisfactory fit, with relatively low error metrics and adequate predictive capacity.

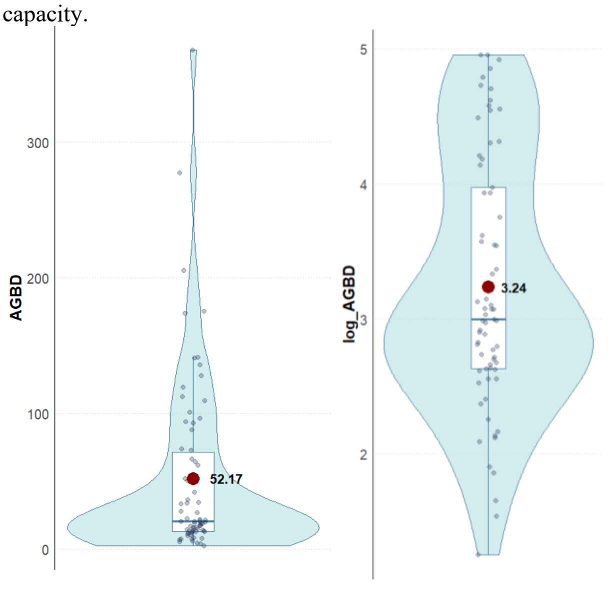

Figure 7. Frequency distribution, quartiles, and mean biomass density values (AGB density) and their logarithmic transformation (log_ AGB density).

The maps generated include both the most probable AGB density value and its relative standard error, providing a comprehensive representation of the park's above-ground biomass (Figure 8). This information is essential for evaluating the area's carbon sequestration potential, supporting the implementation of conservation projects, and establishing a baseline for future monitoring campaigns.

The predicted AGB density values range from 5.2 to 113.4 Mg/ha, with a mean of 32.9 Mg/ha and a standard deviation of 24.8 Mg/ha. The uncertainty of these predictions ranges from 1.3 to 3.3 Mg/ha, with a mean of 2.1 Mg/ha and a standard deviation of 0.36 Mg/ha. The modeled values showed a total accumulated AGB in the park of 8171.9 Mg, which is estimated to contain 3652.8 Mg of organic carbon.

The Out-of-Bag model performance evaluation demonstrated a strong capacity to estimate AGB density, yielding a RMSE of 1.8 Mg/ha and a coefficient of determination (R^2) of 0.59. The R^2 value indicates that 59% of the observed variance in above-ground biomass can be explained by the selected covariates.

4. Discussion

The importance of the LS-Factor (Slope Length), Analytical Hillshading, and the Channel Network Base Level as key covariates highlights how topography strongly influences critical ecological processes for vegetation growth in the La Joya - La Barreta Ecological Park. The LS-Factor affects not just water availability and soil stability, but also erosion and where sediment gets deposited. These processes have a direct impact on soil depth and nutrient levels, which are essential for plant growth. Analytical Hillshading isn’t just about showing sunlight exposure—it also affects soil temperature and evapotranspiration, creating microclimates that can either help or limit the growth of certain plants, and thus influence how much biomass builds up. The Channel Network Base Level gives a sense of how close an area is to drainage systems and how much water tends to collect there. This is closely tied to soil moisture and groundwater, which are key for biomass growth, especially in semi-arid regions like Querétaro.

Adding to this topographic perspective, the use of spectral indices like TWI (Topographic Wetness Index) and EMBI (Enhanced Modified Biomass Index) demonstrate the value of optical data for capturing variations in biomass structure and health. TWI is good at identifying areas with higher soil moisture, which often support more biomass because there’s more water available. EMBI, which is designed to boost the signal from vegetation and reduce background noise from the soil using shortwave infrared correction, gives a clearer and more reliable picture of plant health. It works well as an indicator of how much photosynthetically active biomass is present.

The RMSE of 1.8 achieved in this study indicates a good predictive capacity of the model for regional-scale applications. When compared to large-scale biomass prediction efforts, such as that by Shendryk (2022) who using GEDI data, incorporating elevation models, land cover classification, and Sentinel 1/2 images as covariates, reported R^2 values ranging from 0,66–0,74 and RMSE values of 55–81 Mg/ha a crucial distinction emerges. Shendryk’s results were based on much coarser spatial resolutions (100 m and 200 m, respectively). In comparison, our model uses a much finer resolution of 10 m, which offers a big advantage for managing the park at a more local scale. Even though our R^2 is slightly lower at 0.59, our much lower RMSE of 1.8 Mg/ha shows that our estimates of biomass are more accurate at this finer scale—an important benefit when it comes to monitoring conservation efforts within the park.

https://doi.org/10.5194/isprs-annals-X-3-W3-2025-29-2026

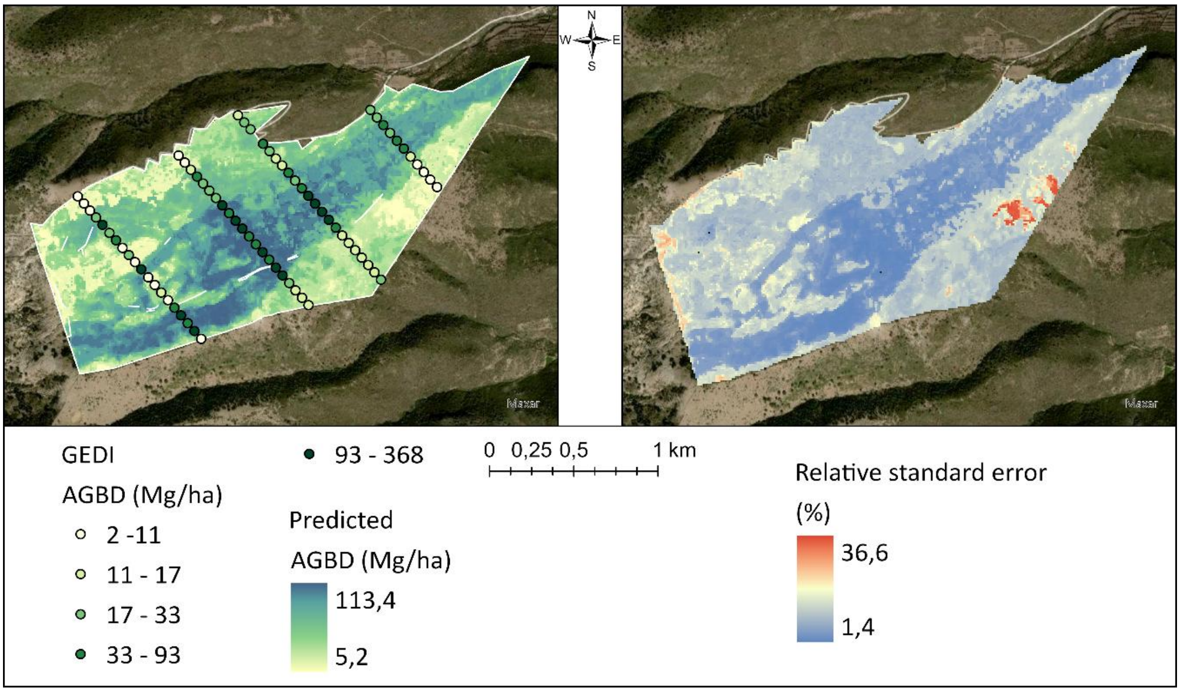

Figure 8. Maps of the most probable AGB density prediction and its relative standard error.

Similarly, while Saarela et al. (2018) reported relative standard errors between 8.0% and 25.4% using Landsat and LiDAR data, our study showed a wider range—from 1.4% to 36.6%. Still, most of our values were under 20%, showing that our model delivers similar or even better precision, especially considering the natural variability found in park ecosystems.

Areas exhibiting a relative standard error above 30% (Figure 8) are mainly located in steep terrain and transition zones between vegetation types. These regions often show high spectral and structural heterogeneity, where small changes in illumination, shadow, or canopy cover strongly affect the reflectance and radar response. In addition, GEDI footprints in these zones are relatively sparse, which limits the representativeness of training data and contributes to greater local uncertainty. Similar patterns of elevated error in complex topography have been reported in other GEDI-based AGB studies (e.g., Bhuyan et al., 2024; Shendryk, 2022).

The R^2 value of 0.59 indicates that while the model captures a significant part of the AGB variability, a considerable percentage remains unexplained. Bhuyan et al. (2024) obtained similar results when modelling canopy height using GEDI data, implementing RF, Extreme Gradient Boosting, Support Vector Machine, and kNN algorithms, with R^2 of 0.553, 0.557, 0.559, and 0.515, respectively.

This lack of explainability could be due to the inherent complexity of biomass distribution, the spatial resolution of the covariates, or the presence of other environmental variables not included in the model but also by uncaptured biological factors such as stand age, tree density, species composition, or the presence of diseases and invasive species. It has been found that AGB estimation can be improved by including field measurements of allometric variables and the use of GEDI's RH95 index along with information derived from Sentinel-1 and -2 bands, obtaining R^2 values from 0.66 to 0.91 (Guo et al., 2023), However, the inclusion of the RH95 index would prevent wall-to-wall prediction.

While our remote sensing-based model demonstrates strong predictive capabilities, it is important to acknowledge that on-going ground-truth biomass measurements are being conducted by the municipality within La Joya - La Barreta Ecological Park. These independent field data, once publicly available, will provide a crucial opportunity for further validation and potential refinement of our model's accuracy. Direct comparisons between our remote sensing estimates and these in-situ measurements will be a key next step to bolster the robustness and practical applicability of our biomass assessment for carbon accounting and conservation projects.

Although independent field validation is not yet available. The predicted mean AGB values (32.9 Mg/ha) are consistent with field-based estimates reported for dry forest and scrubland ecosystems in central Mexico, which typically range between 25 and 45 Mg/ha (Pati et al., 2022). This agreement supports the plausibility of the current predictions despite the lack of direct field calibration.

Despite satisfactory model performance ($R^2 = 0.59$), the absence of biophysical variables such as stand age, tree density, or species composition likely contributed to the unexplained variance. Additionally, climatic factors (e.g., mean annual precipitation or temperature) are known to influence biomass accumulation but were beyond the scope of this study. Upcoming research should integrate such variables to better capture ecological drivers and reduce residual uncertainty.

Future research could explore the inclusion of additional environmental variables, such as annual precipitation or mean temperature, which are often crucial drivers of biomass production, especially in semi-arid regions. Investigating more advanced modeling methods, including Deep Learning approaches that can capture more complex non-linear relationships, or geo-statistical models that explicitly account for the spatial autocorrelation of biomass, could also lead to improved accuracy. Moreover, integrating data from other sources like drone-based photogrammetry or low-altitude

https://doi.org/10.5194/isprs-annals-X-3-W3-2025-29-2026

LiDAR could provide even more detailed structural information.

These findings have significant practical implications for the management of La Joya - La Barreta Ecological Park. The high-resolution AGB density maps generated can effectively inform specific management decisions, such as identifying areas with high biomass for priority protection or areas with lower biomass that present strong potential for reforestation or restoration efforts. The methodology employed is highly cost-effective and scalable, making it a viable approach for ongoing monitoring of the park's carbon sequestration capacity. This robust baseline information is crucial for integrating the park's conservation efforts into REDD+ initiatives, providing a verifiable foundation for tracking and reporting carbon stock changes and thus securing future carbon credits. Beyond its local implications, this approach demonstrates the feasibility of using open-access GEDI and Sentinel data for cost-effective biomass monitoring in other semi-arid regions of Mexico and Latin America, contributing to scalable frameworks for REDD+ and ecosystem restoration initiatives.

5. Conclusions

This study demonstrates the feasibility of integrating GEDI mission LiDAR data with topographic and spectral covariates derived from Sentinel-2 for above-ground biomass estimation in La Barreta Ecological Park. Topographic covariates (LS-Factor, Analytical Hillshading and Channel network base level) and the spectral indices TWI and EMBI emerged as the most influential predictors. Although the model shows acceptable performance, these findings provide a valuable basis for future studies, suggesting the need to explore other data sources, such as drones (photogrammetry or low-altitude LiDAR), or more advanced modeling methods to improve the accuracy of biomass estimation. These results highlight the operational value of GEDI-based biomass modeling for conservation management and carbon accounting, supporting science-based decision-making for protected areas.

6. Data availability

The supplementary material is available on https://doi.org/10.5281/zenodo.15802798. This material contains the R script and sources such as shapefiles, raster files, and databases to replicate the results obtained in the study. The data is open access according to the Creative Commons Attribution Share Alike 4.0 International Licence.

7. Acknowledgments

Luis Castelblanco acknowledges support from scholarship #4047895 and complementary support 67f5645e3c05a02f021d48b3 of the National Council of Humanities, Sciences, and Technologies; abbreviated SECIHTI. Mario Guevara acknowledges support from Conahcyt , (Grant / Award Number: 'grant no. CF-2023-I-1846') UNAM PAPIIT , (Grant / Award Number: 'IV200223') UNESCO-IGCP , (Grant / Award Number: 'grant no. 765'). This work was supported by Luis Aguilar, Alejandro Ávalos, Alejandro De León, and Jair García from the National Laboratory for Advanced Scientific Visualization.

8. References

Bhuyan, M., Jeganathan, C., & Pujar, G. S. (2024). Enhancing Forest Canopy Height Mapping in Kaziranga National Park, Assam, by Integrating LISS IV and SAR data with GEDI LiDAR data Using Machine Learning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *X-3–2024*, 39–44. https://doi.org/10.5194/isprs-annals-X-3-2024-39-2024

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). *System for Automated Geoscientific Analyses (SAGA)* (v. 2.1.4, Geosci. Model Dev., 8, 1991-2007). https://doi.org/doi:10.5194/gmd-8-1991-2015

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P. L., Qi, W., & Silva, C. (2020). The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*, *1*, 100002. https://doi.org/10.1016/j.srs.2020.100002

Gardon, F. R., Santos, R. F. dos, & Rodrigues, R. R. (2020). Brazil's forest restoration, biomass and carbon stocks: A critical review of the knowledge gaps. *Forest Ecology and Management*, *462*, 117972. https://doi.org/10.1016/j.foreco.2020.117972

Guo, Q., Du, S., Jiang, J., Guo, W., Zhao, H., Yan, X., Zhao, Y., & Xiao, W. (2023). Combining GEDI and sentinel data to estimate forest canopy mean height and aboveground biomass. *Ecological Informatics*, *78*, 102348. https://doi.org/10.1016/j.ecoinf.2023.102348

Herold, M., Carter, S., Avitabile, V., Espejo, A. B., Jonckheere, I., Lucas, R., McRoberts, R. E., Næsset, E., Nightingale, J., Petersen, R., Reiche, J., Romijn, E., Rosenqvist, A., Rozendaal, D. M. A., Seifert, F. M., Sanz, M. J., & De Sy, V. (2019). The Role and Need for Space-Based Forest Biomass-Related Measurements in Environmental Management and Policy. *Surveys in Geophysics*, *40*(4), 757–778. https://doi.org/10.1007/s10712-019-09510-6

Hollister, J., Shah, T., Nowosad, J., Robitaille, A. L., Beck, M. W., & Johnson, M. (2023). *elevatr: Access Elevation Data from Various APIs* (R package version 0.99.0). https://doi.org/https://doi.org/10.5281/zenodo.8335450

Jiang, F., Zhao, F., Ma, K., Li, D., & Sun, H. (2021). Mapping the Forest Canopy Height in Northern China by Synergizing ICESat-2 with Sentinel-2 Using a Stacking Algorithm. *Remote Sensing*, *13*(8), 1535. https://doi.org/10.3390/rs13081535

Kuhn, M. (2008). Building Predictive Models in *R* Using the **caret** Package. *Journal of Statistical Software*, *28*(5). https://doi.org/10.18637/jss.v028.i05

Lefsky, M. A., Cohen, W. B., Harding, D. J., Parker, G. G., Acker, S. A., & Gower, S. T. (2002). Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology and Biogeography*, *11*(5), 393–399. https://doi.org/10.1046/j.1466-822x.2002.00303.x

Mahoney, M., Carvalho, F., & Sumner, M. (2025). *Efficiently Retrieve and Process Satellite Imagery* (0.2.0.9000). https://doi.org/10.5281/zenodo.10926857

Meinshausen, N. (2006). Quantile Regression Forests. *Journal of Machine Learning Research*. https://people.math.ethz.ch/~nicolai/quantregforests.pdf

Montero, D., Aybar, C., Mahecha, M. D., Martinuzzi, F., Söchting, M., & Wieneke, S. (2023). A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research. *Scientific Data*, *10*(1), 197. https://doi.org/10.1038/s41597-023-02096-0

Niu, L., Kaufmann, H., Xu, G., Zhang, G., Ji, C., He, Y., & Sun, M. (2022). Triangle Water Index (TWI): An Advanced Approach for More Accurate Detection and Delineation of Water Surfaces in Sentinel-2 Data. *Remote Sensing*, *14*(21), 5289. https://doi.org/10.3390/rs14215289

Pati, P. K., Kaushik, P., Khan, M. L., & Khare, P. K. (2022). Allometric equations for biomass and carbon stock estimation of small diameter woody species from tropical dry deciduous forests: Support to REDD+. *Trees, Forests and People*, *9*, 100289. https://doi.org/10.1016/j.tfp.2022.100289

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C. E., Armston, J., Dubayah, R., Blair, J. B., & Hofton, M. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*, 112165. https://doi.org/10.1016/j.rse.2020.112165

Reichstein, M., & Carvalhais, N. (2019). Aspects of Forest Biomass in the Earth System: Its Role and Major Unknowns. *Surveys in Geophysics*, *40*(4), 693–707. https://doi.org/10.1007/s10712-019-09551-x

Shendryk, Y. (2022). Fusing GEDI with earth observation data for large area aboveground biomass mapping. *International Journal of Applied Earth Observation and Geoinformation*, *115*, 103108. https://doi.org/10.1016/j.jag.2022.103108

Stoffel, M., Cramer, W., Luterbacher, U., & Toth, F. (2008). Methods for Estimating Above-Ground Biomass. In *Carbon Inventory Methods Handbook for Greenhouse Gas Inventory, Carbon Mitigation and Roundwood Production Projects* (pp. 113–147). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6547-7_10

Zhao, Y., & Zhu, Z. (2022). ASI: An artificial surface Index for Landsat 8 imagery. *International Journal of Applied Earth Observation and Geoinformation*, *107*, 102703. https://doi.org/10.1016/j.jag.2022.102703

Appendix

The explanatory covariates used in this study are listed below:

Source	**Indices**
Sentinel-1	VH , VV
Sentinel-2	A, B, G, R, RE1, RE2, RE3, N, N2, WV, S1, S2
Spectral Indices	AFRI1600, AFRI2100, ANDWI, ARI, ARI2, AVI, AWEInsh, AWEIsh, BAI, BAIM, BAIS2, BCC, BI, BITM, BIXS, BLFEI, BNDVI, BRBA, BaI, CIG, CIRE, CRI550, CRI700, CSI, CVI, DBSI, DSI, DSWI1, DSWI2, DSWI3, DSWI4, DSWI5, DVI, EMBI, ENDVI, EVIv, ExG, ExGR, ExR, FCVI, GARI, GBNDVI, GCC, GEMI, GLI, GM1, GM2, GNDVI, GOSAVI, GRNDVI, GRVI, GVMI, IKAW, IPVI, IRECI, LSWI, MBI, MCARI, MCARI1, MCARI2, MCARI705, MCARIOSAVI, MCARIOSAVI705, MGRVI, MIRBI, MLSWI26, MLSWI27, MNDVI, MNDWI, MRBVI, MSAVI, MSI, MSR, MSR705, MTCI, MTVI1, MTVI2, MuWIR, NBAI, NBR, NBR2, NBRSWIR, NBRplus, NBSIMS, ND705, NDBI, NDCI, NDDI, NDGlaI, NDII, NDMI, NDPonI, NDREI, NDSI, NDSII, NDSWIR, NDSaII, NDSoI, NDTI, NDVI, NDVI705, NDVIMNDWI, NDWI, NDYI, NGRDI, NHFD, NIRv, NLI, NMDI, NRFIg, NRFIr, NSDS, NSDSI1, NSDSI2, NSDSI3, NWI, NormG, NormNIR, NormR, OSAVI, OSI, PI, PISI, PSRI, RCC, RDVI, REDSI, RENDVI, RGBVI, RGRI, RI, RI4XS, RNDVI, RVI, S2REP, S2WI, S3, SI, SIPI, SLAVI, SR, SR2, SR3, SR555, SR705, SWI, SWM, SeLI, TCARI, TCARIOSAVI, TCARIOSAVI705, TCI, TDVI, TGI, TRRVI, TTVI, TVI, TWI, TriVI, UI, VARI, VARI700, VI700, VIBI, VIG, VgNIRBI, VrNIRBI, WI1, WI2, WI2015, WRI, bNIRv, mND705, mSR705, sNIRvLSWI, sNIRvNDVILSWIP, sNIRvNDVILSWIS, sNIRvSWIR Indices available at https://github.com/awesome-spectral-indices/awesome-spectral-indices
Terrain analysis	elevation, Analytical Hillshading, Aspect, Channel Network Base Level, Channel Network Distance, Closed Depressions, Convergence Index, Geomorphons, LS-Factor, Plan Curvature, Profile Curvature, Relative Slope Position, Slope, Topographic Wetness Index, Total Catchment Area, Valley Depth. Terrain analysis documentation available in: https://saga-gis.sourceforge.io/saga_tool_doc/9.8.1/ta_compound_0.html