Leveraging SAR and Optical Remote Sensing for Enhanced Biomass Estimation in the Amazon with Random Forest and XGBoost Models

Rodrigo Antunes[1], Luiz Junior[1], Gilson Costa[2], Raul Feitosa[3], Edilson de Souza Bias[4], Abimael Cereda Junior[7], Catherine Almeida[5], Laura E. Cué La Rosa[6], Patrick Happ[3], Leonardo Chiamulera[1]

[1]Conecthus Research Institute, (rodrigo.antunes, luiz.junior, leonardo.chiamulera)@conecthus.org.br
[2]Rio de Janeiro State University (UERJ), gilson.costa@ime.uerj.br
[3]Pontifical Catholic University of Rio de Janeiro (PUC-Rio), raul@ele.puc-rio.br ; patrick@exacta.inf.puc-rio.br
[4]University of Brasilia (UnB), edbias@unb.br
[5]São Paulo State University (UNESP), catherine.almeida@unesp.br
[6]Wageningen University & Research (WUR), laura.cuelarosa@wur.nl
[7]Geografia das Coisas, ceredajunior@geografiadascoisas.com.br

Keywords: Remote Sensing, Aboveground Biomass, Machine Learning, Forest Inventory, Geodatabase.

Abstract

This study addresses the challenge of estimating above-ground biomass (AGB) in the Amazon rainforest by developing a reference geographical database, which provides the ground truth, and comparing the relative importance of using Synthetic Aperture Radar (SAR) and optical remote sensing data to automatically infer AGB. In the experiments reported in this article, we assessed how those two remote sensing data sources impact the accuracy of AGB estimates produced by regression models built with Random Forest (RF) and Extreme Gradient Boosting (XGBoost). The research involved compiling a comprehensive database from many available forest inventories, integrating parcel- and tree-level data to enable precise biomass estimation. The methodology included setting up a spatial data analysis environment, standardizing data, and implementing an experimental protocol with feature selection and leave-one-out cross-validation. The results demonstrate that both kinds of data, i.e., SAR and optical, and their combination can be used for estimating AGB, providing valuable insights for forest management and climate change mitigation efforts. The reference database is available upon request to the corresponding authors.

1. Introduction

Global warming has been a central issue in environmental discussions for decades, drawing attention and action proposals from environmental agencies and uniting nations worldwide around a common goal. Importantly, these discussions extend beyond governmental organizations to include academia, non-governmental organizations, and the private sector, all working towards the same objective.

The rise in global, regional, and local temperatures can be attributed to several factors. Rapid urbanization worldwide, reliance on non-renewable energy for transportation, unplanned agricultural expansion, and deforestation are major contributors to this trend.

The critical role of intact vegetation in mitigating the greenhouse effect is widely recognized. The excessive release of carbon dioxide can be significantly offset through carbon sequestration, a natural process facilitated by oceans, soil, and forests.

Focusing on Brazil, the Amazon rainforest stands out for its vast carbon sequestration potential. Spanning 6.7 million km² across several South American countries, including Bolivia, Peru, Colombia, Venezuela, Guyana, Suriname, and French Guiana, the portion within Brazil, known as the Brazilian Legal Amazon, covers 5 million km². The Amazon biome accounts for 49% of Brazil's territory, covering 8,510,000 km².

Forests are pivotal in carbon sequestration, storing approximately 80% of the terrestrial biomass (Gardon et al., 2020). Beyond their role in carbon storage, forests' biomass is a critical indicator of their health and an essential metric for quantifying their ecosystem services (Herold et al., 2019; Reichstein and Carvalhais, 2019). Accurate biomass estimation supports environmental preservation and sustainable management, indicating the nature and extent of environmental degradation (Ghasemi et al., 2011; Saatchi et al., 2011).

Amidst escalating global warming, monitoring forest biomass and carbon stocks has gained urgency, leading to initiatives like REDD+ (Reducing Emissions from Deforestation and Forest Degradation). Developed by the United Nations Framework Convention on Climate Change (UNFCCC), REDD+ emphasizes the economic value of carbon sequestration in forests, necessitating reliable carbon stock estimates (Pati et al., 2022; Pothong et al., 2022).

However, tropical forests pose significant challenges for modeling their properties due to their complex dynamics, resulting in considerable uncertainties in carbon stock estimates (Mitchard et al., 2014, 2013; Sinha et al., 2015). Field quantification of forest carbon typically involves collecting data on tree diameter, height, and species in sample plots, using these variables to estimate above-ground biomass (AGB) through allometric equations (e.g., Chave et al., 2014). Despite its high accuracy, such an in situ method is labor-intensive and less feasible in vast or inaccessible areas.

Remote sensing offers a viable alternative for biomass estimation, leveraging various sensors (optical, SAR, LiDAR) to calibrate biomass models. Integrating field data with remote sensing through statistical and machine learning models (e.g., linear regression, random forests, artificial neural networks) can enhance AGB estimates, as demonstrated in Almeida et al. (2019). Nonetheless, effectively combining different sensor data and modeling algorithms remains a complex challenge.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-2024-21-2024

The present study aims to improve the estimation of above-ground biomass (AGB) in the Amazon region. By assembling a comprehensive reference geographical database and employing remote sensing data from Synthetic Aperture Radar (SAR) and optical sensors, we compare the efficacy of Random Forest (RF) and Extreme Gradient Boosting (XGBoost) algorithms in biomass estimation. The challenges of modeling in tropical forests, such as the curse of dimensionality due to the large number of available feature types and small number of target samples (parcels) within the selected study area, are addressed, highlighting the significance of accurate carbon stock estimates for environmental preservation and sustainable management.

Our contributions are twofold: first, we provide a structured reference database for model training and evaluation, enhancing data accessibility for biomass quantification in the Amazon. Second, we conduct a comparative analysis of SAR and optical remote sensing data to ascertain their relative importance and combined efficacy in AGB estimation.

The subsequent sections of this paper are organized as follows: "Materials and Methods" outlines the creation of the reference database and the technological platforms utilized for spatial data analysis and biomass calculation. "Results and Discussion" presents the findings from our regression analysis, comparing the performance of RF and XGBoost models across different remote sensing data sets. Finally, the "Conclusion" section summarizes the study's implications for forest management and climate change mitigation, proposing directions for future research to further improve the accuracy and applicability of biomass estimation models.

2. Materials and Methods

The first goal of this research is to assemble a reference database for training and evaluating biomass estimation methods. This database covers regions in the Amazon, including parts of the Brazilian states of Amazonas, Pará, Rondônia, Acre, and Mato Grosso (Figure 1).

In the pursuit of state-of-the-art practices not only in cartographic representation but also in the preparation and use of spatial data analysis environments under the premise of Geographic Intelligence, the following technological platforms were employed to prepare the reference geographical database:

- ArcGIS Pro, a comprehensive Geographic Information System (GIS) for input, storage, spatial data analysis, and output, whether as cartographic products or any other informational output.
- ArcGIS Online WebGIS, for publishing, manipulating, geodata visualization, and analyzing geographic data in a cloud computing environment.
- The R programming language and the BIOMASS library (Roéjou-Méchain *et al.*, 2016), for conducting the necessary calculations for above-ground biomass estimation.
- The Python Programming language for geodata extraction, standardization, and transformation.

The database used in this research was constructed from data sourced from the study "Forest Inventory and Biophysical Measurements, Brazilian Amazon, 2009-2018" (dos Santos *et al.*, 2022). From the forest inventories provided in the aforementioned sources, approximately fifty by fifty meters (2500 m2) parcels were selected, amounting to 19 inventories and 385 parcels.

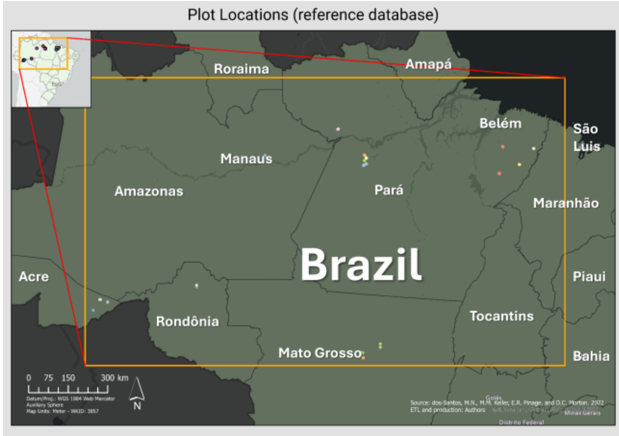

Figure 1. Extents of the forest inventories (reference database)

A comprehensive effort to standardize and correct errors in the data from the selected inventories was undertaken, including

- Importing parcel polygons and tree points in shapefile format into spatial databases (GDB and PostGIS);
- Resolving geometric-topological issues, such as negative areas;
- Standardizing numerical data to ensure compatibility of data types and numerical magnitudes;
- Standardizing attributes and modeling of the Geographic Database.

The geographic database's objects were thus integrated and made compatible (Figure 1) in a Geographic Coordinate System, with WGS 84 as the Reference System.

To preserve cartographic integrity, especially for potential metric calculations, spatial data were also organized in a plane coordinate system (Universal Transverse Mercator), in five groups by UTM Zone, with SIRGAS 2000 datum, as shown in Table 1. This meticulous data preparation process ensures the database's reliability for further analysis, facilitating accurate biomass estimation and contributing to the integrity of cartographic and metric calculations within the research.

The database objects represent trees (points) and parcels (polygons). The trees occur within the parcels, and each corresponding point has the attributes listed in Table 2. The polygons associated with the parcels of the forest inventories have the attributes listed in Table 3.

Once the Geographic Database was systematically organized, standardized, and normalized, we could calculate the aboveground biomass (AGB) utilizing the BIOMASS library. By developing a straightforward R script, drawing on the approach outlined by Réjou-Méchain *et al.* (2016), we could calculate the AGB for each tree and extend these calculations to encompass the entire parcels.

https://doi.org/10.5194/isprs-annals-X-3-2024-21-2024 | © Author(s) 2024. CC BY 4.0 License.

Study Area Name	UTM Zone
BON_A01_2014, HUM_A01_2014, TAL_A01_2014,	19S
DUC_A01_2016, JAM_A02_2013, JAM_A03_2013	20S
ANA_A01_2015_2018, FN_A01_2015, FNA_A01_2013, FST_A01_2013, SAN_A01_2014_2016, SAN_A01b_2016_2018, SAN_A02_2014, TAP_A01_2015_2016_2018, TAP_A03_2015_2018	21S
CAU_A01_2014_2018, TAC_A01_2015,	22S
AND_A01_2013_2018, PAR_A01_2018	23S

Table 1. Study Areas and respective UTM Zones.

Attribute	Description/Unit
PlotID	Parcel identification code
area	Code name to the area
tree	Tree identification code
common.name	Tree common name
scientific.name	Tree scientific name
family.name	Tree family name
DBH	Diameter at breast height (1.3 meters above the ground), in centimeters
type	Tree class: Liana, Palms, Trunked Palms, or Others
Dead	Standing dead
D.Class	Decomposition Class (from Keller *et al.*, 2004)
H	Tree total height, in meters
WD	Wood density, in g/cm^3
date	Date, in ISO 8601 format
UTM.easting	Abscissa of UTM Coordinate (x), in meters
UTM.northing	Ordinate of UTM Coordinate (y) UTM, in meters

Table 2. Attributes associated with the points (trees) of the reference database.

Attribute	Description/Unit
PlotID	Parcel identification code
area	Name to the area (ref. Table 1)
Shape_Lenght	Perimeter, in meters
Shape_Area	Area, in square meters

Table 3. Attributes associated with the polygons (parcels) of the reference database.

The allometric equation used to calculate the AGB per tree was the same one proposed by Chave et al. (2014) and implemented in the BIOMASS library, namely:

$$AGB\_tree = 0.0673 \times (WD \times H \times DBH^2)^{0.976}$$

Following the AGB calculation processes, the outputs were seamlessly integrated into the geographic reference database, introducing the attributes specified in Tables 4 and 5 and enriching the database with detailed biomass information.

Attribute	Description/Unit
AGB_tree	AGB, in Mg

Table 4. Attribute appended to the reference database's points (trees) objects after biomass calculation.

Attribute	Description/Unit
AGB_ha	AGB, in Mg ha^{-1}

Table 5. Attribute appended to the reference database's polygons (parcels) objects after biomass calculation.

2.1 Study Area

To investigate the relative importance of different types of remote sensing data (i.e., SAR and optical) in biomass estimation, we selected a particular area (and respective set of parcels) within the reference database.

The particular study area is within the Adolpho Ducke Forest Reserve, northeast of Manaus, managed by the National Institute of Amazonian Research (INPA). This area covers 100 km² of primary forest (Hopkins, 2005), and the selected study area covers 21 km², as illustrated in Figure 2 (orange box).

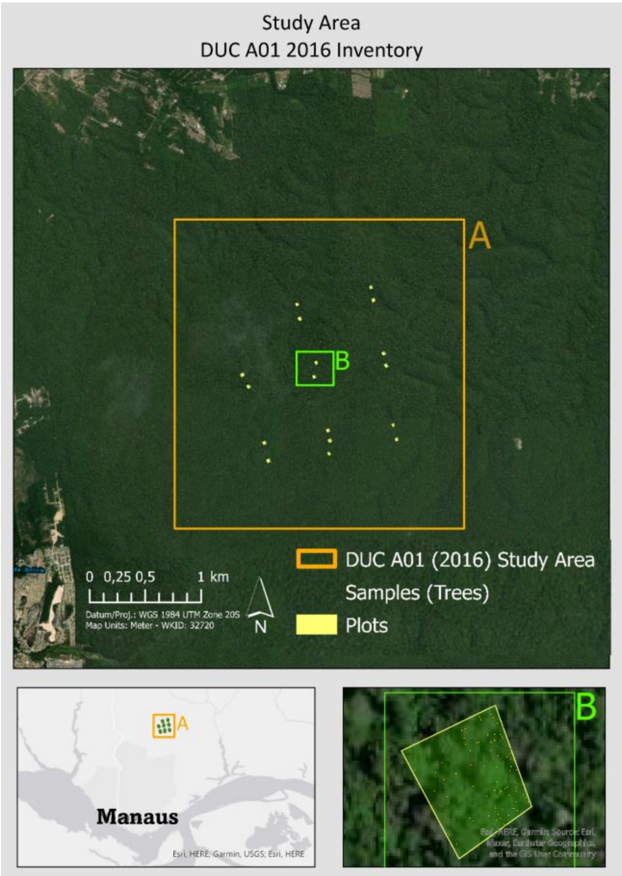

Figure 2. Location of the selected Study Area

Hopkins (2005) identifies the main vegetation type in the study area as *terra firme* forest, noting the absence of floodplain and igapó forests attributed to the local rivers' lack of regular flooding. The reserve's habitats are differentiated into plateau forests, with nutrient-poor, clayey, and well-drained soil, where

the canopy ranges from 30-40 meters, occasionally reaching 50-60 meters through emergent trees.
Furthermore, Hopkins (2005) describes lowland forests as having sandy, very moist soil and a canopy of 25-30 meters. Slope forests on inclines feature sandy soil at lower elevations and vegetation that serves as a transitional zone between plateau and lowland forests. *Campinarana* forests on plains near creeks are characterized by sandy soil and abundant leaf litter, with canopy heights between 15-25 meters.

The respective forest inventory was carried out in August 2016, and comprises a total of 17 parcels.

2.2 Remote Sensing Data

For the biomass estimation regression study, we employed images from optical orbital sensors and synthetic aperture radar (SAR) systems. The selected imagery included:

- A Sentinel-2 (optical) image from August 16, 2016, processed to Level 1C. It was orthorectified and adjusted for Top-of-Atmosphere reflectance, with its bands resampled to achieve a 10-meter resolution per pixel.

- A Sentinel-1 (SAR) image from September 25, 2016, presented in Ground Range Detected (GRD) format, with a 10-meter resolution per pixel, which underwent calibration and orthorectification using the Sentinel-1 Toolbox.

- An Alos Palsar-2 ScanSAR image from August 29, 2016, calibrated and orthorectified using the ALOS World 3D (AW3D30) Digital Surface Model to achieve a 25-meter resolution per pixel.

In the experiments, we used all available Sentinel-2 bands except bands 1, 9, and 10. For the Sentinel-1 (C-band) image, we used both VV and VH polarizations. For the Palsar-2 (L-band) image, we used HH and HV polarizations.

2.3 Biomass estimation method

We employed two machine learning algorithms in the experiments: Random Forest (RF) and Extreme Gradient Boosting (XGBoost).

Random Forest (Breiman, 2001) constructs an ensemble of decision trees through a process known as bootstrap aggregating or bagging (Breiman, 1996). This method involves creating multiple subsets of the original dataset through random sampling with replacement, ensuring each subset is slightly different. For each of these subsets, a decision tree is grown. At inference time, Random Forest computes the outcome by averaging the predictions made by all the individual trees.

XGBoost (Friedman, 2001) also builds an ensemble of decision trees, sharing similarities with Random Forests in that it aims to minimize a specific loss function during training. This loss function accounts for the discrepancies between predicted and actual outcomes and incorporates a regularization term to manage the model's complexity, mitigating the risk of overfitting. However, unlike Random Forests, which constructs its trees in parallel, XGBoost builds its model sequentially. Each tree is added to the ensemble to correct the residuals or errors left by the previously trained trees. The gradient of the loss function guides this correction. Furthermore, the training involves assigning weights to each tree based on their contribution to reducing the overall prediction error.
At inference, each new input data passes through the ensemble of trees. To obtain the final prediction, XGBoost aggregates the scores from all trees. The weights computed during the training phase are considered in the aggregation step, with more accurate trees having a greater influence on the final prediction.

Our choice of Random Forest and XGBoost was motivated by their success in many studies on biomass estimation, e.g. (Li et al., 2020; Torre-Tojal et al., 2022). We favored them over deep learning approaches due to the limited availability of labeled training data. The selected methods require fewer samples to achieve high accuracy, avoiding the overfitting common in deep learning models trained with scarce data.

Furthermore, such methods are more easily interpretable than deep learning algorithms. This characteristic facilitates the scientific and decision-making communities' validation and adoption of solutions based on them.

2.4 Experimental Protocol

We extracted several features from the satellite imagery bands (Sentinel-1, Sentinel-2, and Palsar-2) and the vegetation indices EVI and NDVI derived from the Sentinel-2 bands. The total set of features used in the experiments comprise the mean and standard deviations computed for each Sentinel-2 band (except bands 1, 9, and 10), for the EVI and NDVI indices, and for each polarization of the Sentinel-1 and Palsar-2 image data, amounting to 32 individual features, for each of the 17 parcels of the DUC_A01_2016 inventory. Our experimental protocol implemented an exploratory feature selection process, systematically investigating the predictive power of various combinations of these features on the model's performance.

We adopted the leave-one-out (LOO) cross-validation technique to evaluate the model and mitigate overfitting. In this validation method, each instance in the dataset is sequentially used as a single data point for the test set, while the remainder of the data serves as the training set. This approach is particularly beneficial in scenarios where the dataset size is limited, as it maximizes the training data usage while ensuring a thorough assessment of the model on every data point.

In tuning the XGBoost regressor, we adjusted several hyperparameters to optimize performance. We set the number of trees in the model (estimators) to 50, limited the depth of each tree to 2 layers to prevent overly complex models, and specified a learning rate of 0.1 to control how quickly the model adapts to the problem. To introduce randomness and thus enhance the model's generalization capability, we applied two types of subsampling: a rate of 0.6 for selecting samples (subsample), and a feature sampling rate (Feature Sampling Rate for Tree Construction) of 0.7. We also employed regularization techniques, setting alpha and lambda to 10. These regularization parameters add penalties on the model's complexity, with alpha for L1 regularization and lambda for L2 regularization, further aiding in the prevention of overfitting by discouraging overly complex models.

A similar procedure was conducted for Random Forest. The final configuration consists of 50 trees, a minimum requirement of 2 samples for splitting, and at least six samples in each leaf. The decision on which features to consider during splits is restricted to the square root of the total features available. Additionally, the

https://doi.org/10.5194/isprs-annals-X-3-2024-21-2024 | © Author(s) 2024. CC BY 4.0 License.

trees are capped at a maximum depth of 3 to prevent the model from becoming excessively complex. The model employs absolute error as its splitting criterion, aiming to minimize prediction errors in absolute terms.

We used the Root Mean Squared Error (RMSE) metric to evaluate the models’ accuracy. To address potential variability in results due to stochastic processes inherent in algorithms like XGBoost, we ran each experiment 20 times with 20 different seeds. We reported the RMSE mean and standard deviation.

3. Results and Discussion

3.1 Reference Database

With the results obtained from data processing with the BIOMASS package and integrated into the database (utilizing Python and ArcGIS Pro), a geovisualization web application was created in ArcGIS Online, the so-called dashboard, as shown in Figure 3. This dashboard allows the user to browse the obtained data, select topics of interest, and have an interactive view, including filters by scale and attributes. This dataviz is possible because the geographical database has been modeled and normalized using tools such as Model Builder (ArcGIS Pro) and the support of Extract, transform, load (ETL) processes with Python codes.

Figure 3 shows an example of the study area (DUC_A01_2016), containing 17 parcels and an average AGB of 281.5 Mg ha^{-1} from 1,169 inventoried trees, with an average AGB per tree of 0.9 Mg. In the same view, the graphs with the tree's diameter DBH and WD average wood density, respectively, can be read and interpreted.

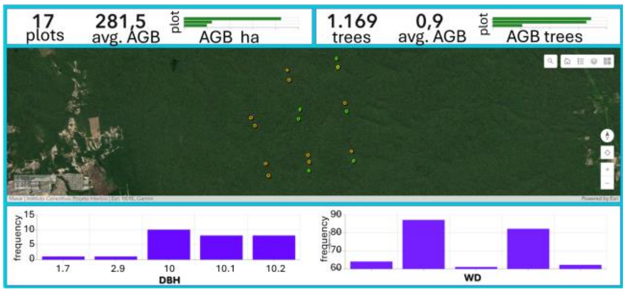

Figure 3. Exploratory Panel (geodataviz Dashboard).

In addition to the species name, the dashboard allows the analysis of the same parameters (parcel number, AGB, DBH, and WD) in individual trees, as shown in Figure 4.

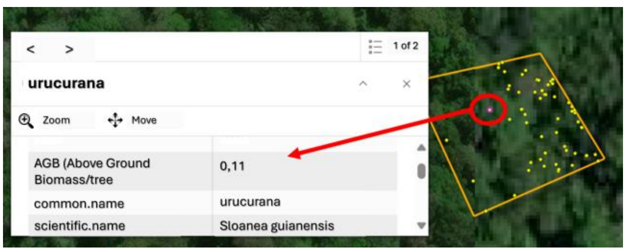

Figure 4. Querying a specific sample (tree) attributes from Table 2.

3.2 AGB Regression Analysis

Figures 5 and 6 show the distribution of RMSE values obtained from the prediction of the XGBoost and RF regression models, respectively, each configured with a different combination of features. In the figures, 'S1' denotes features from Sentinel-1, 'S2' represents features from Sentinel-2, 'P2' corresponds to PALSAR-2 features, and 'EVI' and 'NDVI' are vegetation indices derived from Sentinel-2 data.

By inspecting the figures, one can observe that the errors in the AGB are, on average, in the range of 54 to 58 Mg/ha RMSE, which is compatible with the state-of-the-art. In (Arévalo et al., 2023), which used Landsat data and XGBoost for estimating biomass in the Amazon, the error ranged from 64 to 92 Mg/ha RMSE.

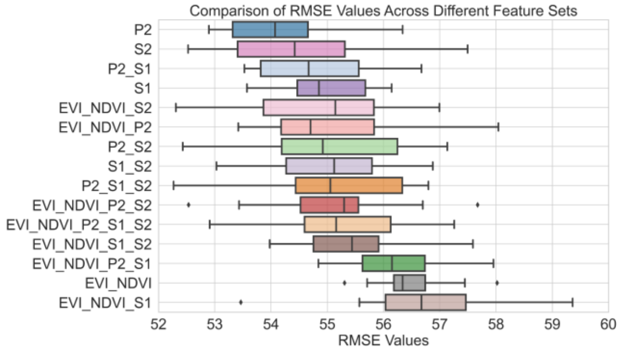

Figure 5. Boxplot of RMSE Values Across Different Feature Set Combinations for the XGBoost model.

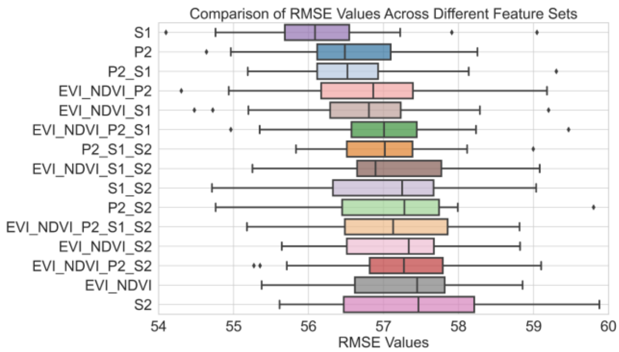

Figure 6. Boxplot of RMSE Values Across Different Feature Set Combinations for the RF model.

It is also clear that the results with the XGBoost regression models were superior to those obtained with RF in most cases (i.e., considering the different input feature sets).

Considering the XGBoost results, the best feature sets are not identifiable, considering the overlapping ranges of the different boxplots. Interestingly, however, the three best results are associated with features coming either from the SAR data (PALSAR-2 and the combination of PALSAR-2 and Sentinel-1 bands) or the optical data (Sentinel-2 bands). Considering both regression methods, the features from the SAR sensors alone or combined provided better regression models.

Interestingly, considering the sole use of Sentinel-2 features, the XGBoost model was the second best, while the RF counterpart was the worst. This indicates that the first method can better handle more features when creating regression models.

Focusing again on the XGBoost results, combining all available features (i.e., PALSAR-2, Sentinel-1, Sentinel-2, and the vegetation indices) delivered relatively poorer results. This may indicate that, after a certain amount, adding more features makes the respective regression models more complex (introducing the commonly known curse of dimensionality problem). If that is the case, including more training samples, i.e., data from different sites in the reference database, would be beneficial for generating better regression models.

4. Conclusion

This research investigates the estimation of above-ground biomass (AGB) in the Amazon rainforest, focusing on the enhanced accuracy achieved by combining Synthetic Aperture Radar (SAR) and optical remote sensing data. Creating a reference database and evaluating Random Forest (RF) and Extreme Gradient Boosting (XGBoost) algorithms for biomass estimation are central to this study.

The findings confirmed that integrating SAR and optical data may improve the accuracy of AGB estimation by exploiting the complementarity of information contained in each source to capture the complex dynamics of tropical forest ecosystems.

In future works, we are committed to expanding our reference database with additional data from diverse geographical regions and temporal spans. We also intend to explore advanced feature selection techniques and dimensionality reduction methods to manage model complexity and improve interpretability.

We also aim to integrate spaceborne LiDAR data with SAR and optical data. Additionally, we will investigate the regression models' sensitivity to variations in the volume of training samples by analyzing different subsets of data extracted from the reference database. This will help us understand the impact of training data volume on model performance.

Acknowledgments

We acknowledge the support of the Conecthus Research Institute and Callidus in developing this research project.

References

Almeida, C., Galvão, L., Sato, Y., Lopes, A., De Alencastro Graça, P.M., Ferreira-Ferreira, J., Aragão, L., Ometto, J., Jacon, A., Pereira, F., Silva, C., Longo, M., 2019. Combining LiDAR and hyperspectral data for aboveground biomass modeling in the Brazilian Amazon using different regression algorithms. *Remote Sensing of Environment* 232, 111323 https://doi.org/10.1016/j.rse.2019.111323

Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., 1996. Bagging Predictors. *Machine Learning* 24, 123–140. https://doi.org/10.1007/BF00058655

Chave, J., Réjou-Méchain, M., Burquez, A., Chidumayo, E., Colgan, M., Delitti, W., Duque, A., Eid, T., Fearnside, P., Goodman, R., Henry, M., Martinez-Yrizar, A., Mugasha, W., Muller-Landau, H., Mencuccini, M., Nelson, B., Ngomanda, A., Nogueira, E., Ortiz, E., Vieilledent, G., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology* 20, 3177–3190. https://doi.org/10.1111/gcb.12629

Dos-Santos, M.N., Keller, M.M., Pinage, E.R., Morton, D.C., 2022. Forest Inventory and Biophysical Measurements, Brazilian Amazon, 2009-2018. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/2007

Friedman, J., 2000. Greedy Function Approximation: A Gradient Boosting Machine. Ann. Statist. 29 (5) 1189 - 1232, October 2001. https://doi.org/10.1214/aos/1013203451

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

Ghasemi, N., Sahebi, M.R., Mohammadzadeh, A., 2011. A review on biomass estimation methods using synthetic aperture radar data 1. *International Journal of Geomatics and Geosciences*, 1 (4), 776-788.

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. Surv. Geophys. 40, 757–778. https://doi.org/10.1007/s10712-019-09510-6

Hopkins, M.J.G., 2005. Flora da Reserva Ducke, Amazonas, Brasil. *Rodriguésia* 56, 09–25. https://doi.org/10.1590/2175-78602005568602

Li, Y., Li, M., Li, C., Liu, Z., 2020. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. *Scientific Reports* 10, 9952. https://doi.org/10.1038/s41598-020-67024-3

Mitchard, E., Feldpausch, T., Brienen, R., Lopez-Gonzalez, G., Monteagudo, A., Baker, T., Lewis, S., Lloyd, J., Quesada, C., Gloor, M., ter Steege, H., Meir, P., Alvarez-Davila, E., Araujo-Murakami, A., Aragão, L., Arroyo Padilla, L., Aymard, G., Bánki, O., Bonal, D., Phillips, O., 2014. Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. *Global Ecology Biogeography* 23. https://doi.org/10.1111/geb.12168

Mitchard, E., Saatchi, S., Baccini, A., Asner, G., Goetz, S., Harris, N., Brown, S., 2013. Uncertainty in the Spatial Distribution of Tropical Forest Biomass: A Comparison of Pan-tropical Maps. *Carbon Balance and Management* 8, 10. https://doi.org/10.1186/1750-0680-8-10

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

Pothong, T., Elliott, S., Chairuangsri, S., Chanthorn, W., Shannon, D.P., Wangpakapattanawong, P., 2022. New allometric equations for quantifying tree biomass and carbon sequestration in seasonally dry secondary forest in northern Thailand. *New Forests* 53, 17–36. https://doi.org/10.1007/s11056-021-09844-3

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

Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J., Hérault, B., 2017. Biomass: an r package for estimating above-ground biomass and its uncertainty in tropical forests. *Methods in Ecology and Evolution* 8, 1163–1167. https://doi.org/10.1111/2041-210X.12753

Saatchi, S., Marlier, M., Chazdon, R.L., Clark, D.B., Russell, A.E., 2011. Impact of spatial variability of tropical forest structure on radar estimation of aboveground biomass. *Remote Sensing of Environment* 115, 2836–2849. https://doi.org/10.1016/j.rse.2010.07.015

Sinha, S., Jeganathan, C., Sharma, L.K., Nathawat, M.S., 2015. A review of radar remote sensing for biomass estimation. *International Journal of Environmental Science and Technology* 12, 1779–1792. https://doi.org/10.1007/s13762-015-0750-0

Torre-Tojal, L., Bastarrika, A., Boyano, A., Lopez-Guede, J.M., Graña, M., 2022. Above-ground biomass estimation from LiDAR data using random forest algorithms. *Journal of Computational Science* 58, 101517. https://doi.org/10.1016/j.jocs.2021.101517