Modeling mangrove aboveground biomass using LiDAR derived metrics

Jonathan V. Solórzano, Candelario Peralta Carreta, J. Alberto Gallardo Cruz

2024-07-23

# Introduction

Mangroves are tropical forests that occur under brackish waterlogged regimes that represent some of the world’s largest Carbon sinks (Donato et al. 2011; Alongi 2014). Frequently, around Mangroves, Tropical Swamp Forests can be found under waterlogged conditions with reduced salinity, that contain equally important Carbon sinks (Lugo, Brinson, and Brown 1990). Although these forests usually contain large carbon stocks in their below-ground component, their above-ground is usually easier to study with field sampling techniques and remote sensing (Kauffman et al. 2016). These forests usually contain a limited number of species; however, they provide other important ecosystem services such as providing the habitat for fish nurseries and mitigate storms intensity (Hutchison et al. 2014).

Mangroves’ and Tropical Swamp Forests’ above-ground biomass (AGB) monitoring has frequently been accompanied by the use of remote sensing technologies (Kuenzer et al. 2011). This has been due to the advantages of remote sensing of being able to extrapolate data obtained by non-destructive approaches (structure metrics obtained form field sampling) to larger regions and being able to develop spatially explicit AGB models. Thus, the above-ground biomass of these forests is usually obtained by obtaining structural metrics (e.g., DAP, height and specific wood gravity), extrapolating it to an ha and relating it to a particular remote sensed metric.

# Methods

## Data

The model with the lowest average RMSE obtained from the cross validation procedure was regarded as the best model and selected to be fitted using all the training data and evaluated using the test set.

## Algorithms

Three types of algorithms were used to train the models: a linear regression and two machine learning algorithms, random forest and XGBoost. Due to their assumptions and characteristics, linear regression was considered the simplest model and less flexible (considering only linear relations), while XGBoost was considered the most complex and flexible one (using boosting to fit the random trees).

All models were fitted in two different stages: 1) cross-validation models (CV-models) to identify the best combination of predictive variables with the cross validation dataset and 2) final models (F-models) trained with the complete training dataset and evaluated on the test set. The importance of each variable incorporated in the models that achieved the lowest RMSE was calculated in order to identify the most important predictive variables. This importance was calculated on the Fmodels. All the models training and evaluation was performed in R.4.4.1 (R Core Team 2024) using tidymodels (Kuhn and Wickham 2020), randomForest (Liaw and Wiener 2002), xgboost (Chen et al. 2024) and yardstick (Kuhn, Vaughan, and Hvitfeldt 2024) packages.

### Linear model

Linear models is a simple model that considers only linear relations between predictive and response variables that has been used to model AGB using remote sensing metrics.

### Random forest

Random forest is an algorithm frequently used in remote sensing applications to perform both classification and regression tasks. This algorithm consists of an ensemble of random decision trees to obtain a prediction from the majority of trees (average prediction).

The trained random forest models included three predictive variables, 500 random trees and 1 variable tried at each split (Breiman 2001).

### XGBoost

XGBoost is an algorithm based also an ensemble of decision trees; however instead of building random trees, each successive tree is built on the residual errors of the previous one (i.e., boosting). Thus, it is usually regarded as an algorithm that can surpass random forest performance.

### Confidence intervals

Once the Fmodels were fitted, the confidence intervals for all the predicted area were calculated using a bootstraped dataset and evaluated on the complete LiDAR data.

## Correlation analyses

Finally, to better understand the relationship between the variables included in the model selected as the best overall (i.e., lowest RMSE on the test set), a Pearson’s correlation analysis was used to characterize the relation between each predictive variable and AGB.

# Results

## Field data

The AGB data of the Mangrove and TSF plots showed a mean AGB value of 140.32 Mg/ha with a standard deviation = 52.74 Mg/ha. In turn, the range of AGB values was min = 42.03 Mg/ha and max = 262.17 Mg/ha.

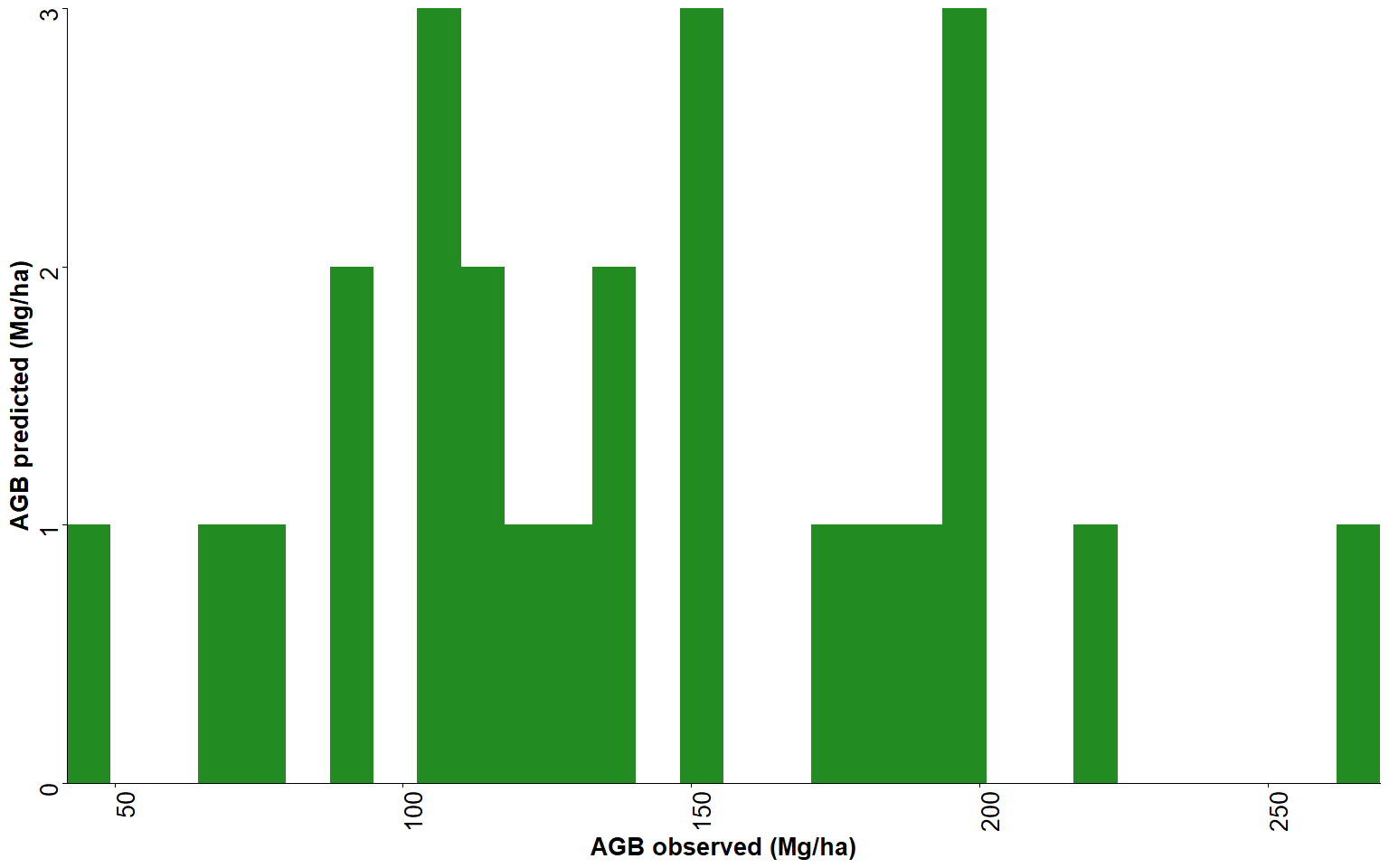


Figure 1. Histogram of the AGB values in the sampling plots.

## Model

The model that achieved the lowest error on the test set was the random forest one ( = 21.43; = 0.69), followed by the linear regression ( = 33.21; = 0.26) and XGBoost ( = 36.71; = 0.09).

Table 1. Models that achieved the lowest RMSE on the test data for Random forest, XGBoos and Linear

| Model | Var 1 | Var 2 | Var 3 | RMSE train | RMSE test | *R2* train | *R2* test |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Random forest | zq35 | zmean | p4th | 14.50 | 21.43 | 0.93 | 0.69 |
| XGBoost | zq55 | p4th | zq95 | 5.88 | 36.71 | 0.99 | 0.09 |
| Linear | zmean | p5th | p2th | 19.74 | 33.21 | 0.86 | 0.26 |

# References

Alongi, Daniel M. 2014. “Carbon Cycling and Storage in Mangrove Forests.” Annu. Rev. Mar. Sci 6: 195–219. <https://doi.org/10.1146/annurev-marine-010213-135020>.

Breiman, L. 2001. “Random Forests.” Journal Article. Machine Learning 45 (5): 5–32. <https://doi.org/10.1023/A:1010933404324>.

Chave, Jérôme, Maxime Réjou-Méchain, Alberto Búrquez, Emmanuel Chidumayo, Matthew S. Colgan, Welington B. C. Delitti, Alvaro Duque, et al. 2014. “Improved Allometric Models to Estimate the Aboveground Biomass of Tropical Trees.” Journal Article. Global Change Biology 20: 3177–90. <https://doi.org/10.1111/gcb.12629>.

Chen, Tianqi, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, et al. 2024. Xgboost: Extreme Gradient Boosting. <https://CRAN.R-project.org/package=xgboost>.

Donato, Daniel, J. Boone Kauffman, Daniel Murdiyarso, Sofyan Kurnianto, Melanie Stidham, and Markku Kanninen. 2011. “Mangroves Among the Most Carbon-Rich Forests in the Tropics.” Nature Geoscience 4: 293–97. <https://doi.org/10.1038/ngeo1123>.

Hutchison, James, Andrea Manica, Ruth Swetnam, Andrew Balmford, and Mark Spalding. 2014. “Predicting Global Patterns in Mangrove Forest Biomass.” Conservation Letters 7: 233–40. <https://doi.org/10.1111/conl.12060>.

Kauffman, J. Boone, Humberto Hernandez Trejo, Maria del Carmen Jesus Garcia, Chris Heider, and Wilfrido M. Contreras. 2016. “Carbon Stocks of Mangroves and Losses Arising from Their Conversion to Cattle Pastures in the Pantanos de Centla, Mexico.” Wetlands Ecology and Management 24: 203–16. <https://doi.org/10.1007/s11273-015-9453-z>.

Kuenzer, Claudia, Andrea Bluemel, Steffen Gebhardt, Tuan Vo Quoc, and Stefan Dech. 2011. Remote Sensing of Mangrove Ecosystems: A Review. Remote Sensing. Vol. 3. <https://doi.org/10.3390/rs3050878>.

Kuhn, Max, Davis Vaughan, and Emil Hvitfeldt. 2024. Yardstick: Tidy Characterizations of Model Performance. <https://CRAN.R-project.org/package=yardstick>.

Kuhn, Max, and Hadley Wickham. 2020. Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles. <https://www.tidymodels.org>.

Liaw, Andy, and Matthew Wiener. 2002. “Classification and Regression by randomForest.” R News 2 (3): 18–22. <https://CRAN.R-project.org/doc/Rnews/>.

Lugo, A. E., M. Brinson, and S. Brown. 1990. Forested Wetlands. Forested Wetlands. Vol. 2. <https://doi.org/10.2307/2260741>.

R Core Team. 2024. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Réjou-Méchain, Maxime, Ariane Tanguy, Camille Piponiot, Jérôme Chave, and Bruno Hérault. 2017. “Biomass: An r Package for Estimating Above-Ground Biomass and Its Uncertainty in Tropical Forests.” Methods in Ecology and Evolution 8 (September): 1163–67. <https://doi.org/10.1111/2041-210X.12753>.

Roussel, Jean-Romain, David Auty, Nicholas C. Coops, Piotr Tompalski, Tristan R. H. Goodbody, Andrew Sánchez Meador, Jean-François Bourdon, Florian de Boissieu, and Alexis Achim. 2020. “lidR: An r Package for Analysis of Airborne Laser Scanning (ALS) Data.” Remote Sensing of Environment 251: 112061. <https://doi.org/10.1016/j.rse.2020.112061>.

Salum, Rafaela B., Sharon A. Robinson, and Kerrylee Rogers. 2021. “A Validated and Accurate Method for Quantifying and Extrapolating Mangrove Above‐ground Biomass Using Lidar Data.” Remote Sensing 13 (July). <https://doi.org/10.3390/rs13142763>.

Solórzano, Jonathan V., J. Alberto Gallardo-Cruz, Candelario Peralta-Carreta, Rubén Martínez-Camilo, and Ana Fernández-Montes de Oca. 2020. “Plant Community Composition Patterns in Relation to Microtopography and Distance to Water Bodies in a Tropical Forested Wetland.” Aquatic Botany 167. <https://doi.org/10.1016/j.aquabot.2020.103295>.

Solórzano, Jonathan V., José Alberto Gallardo-Cruz, Edgar Javier González, Candelario Peralta-Carreta, Matías Hernández-Gómez, Ana Fernández-Montes de Oca, Luis Gerardo Cervantes-Jiménez, et al. 2018. “Contrasting the Potential of Fourier Transformed Ordination and Gray Level Co-Occurrence Matrix Textures to Model a Tropical Swamp Forest’s Structural and Diversity Attributes.” Journal of Applied Remote Sensing 12: 036006. <https://doi.org/10.1117/1.JRS.12.036006>.

Woods, M., K. Lim, and P. Treitz. 2008. “Predicting Forest Stand Variables from LiDAR Data in the Great Lakes  St. Lawrence Forest of Ontario.” The Forestry Chronicle 84 (6): 827–39. <https://doi.org/10.5558/tfc84827-6>.

Zanne, Amy E., G. Lopez-Gonzalez, David A. Coomes, Jugo Ilic, Steven Jansen, Simon L. Lewis, Regis B. Miller, Nathan G. Swenson, Michael C. Wiemann, and Jerome Chave. 2009. “Towards a Worldwide Wood Economics Spectrum.” Dataset. https://doi.org/<https://doi.org/10.5061/dryad.234>.