

## Mapping the spatial distribution of Colombia's forest aboveground biomass using SAR and optical data

P. Rodríguez-Veiga<sup>1, 2,\*</sup>, A.P. Barbosa-Herrera<sup>3</sup>, J.S. Barreto-Silva<sup>3</sup>, P. C. Bispo<sup>1</sup>, E. Cabrera<sup>3</sup>, C. Capachero<sup>3</sup>, G. Galindo<sup>3</sup>, Y. Gou<sup>1</sup>, L.M. Moreno<sup>3</sup>, V. Louis<sup>1</sup>, P. Lozano<sup>3</sup>, A. M. Pacheco-Pascagaza<sup>1</sup>, I. P. Pachon-Cendales<sup>3</sup>, J.F. Phillips-Bernal<sup>3</sup>, J. Roberts<sup>1</sup>, N. R. Salinas<sup>3</sup>, L. Vergara<sup>3</sup>, A.C. Zuluaga<sup>3</sup>, H. Balzter<sup>1,2</sup>

<sup>1</sup> Centre for Landscape and Climate Research, University of Leicester, Leicester, UK - prv4@leicester.ac.uk

<sup>2</sup> NERC National Centre for Earth Observation (NCEO), Leicester, UK

<sup>3</sup> Institute of Hydrology, Meteorology and Environmental Studies (IDEAM), Bogota, Colombia - pbarbosa@ideam.gov.co

### Commission III

**KEY WORDS:** biomass, carbon, SAR, multispectral, machine-learning, forest inventory

### ABSTRACT:

An assessment on the amount and spatial distribution of forest aboveground biomass (AGB) for the forests in Colombia was generated using in-situ national forest inventory data (IDEAM, 2018), in combination with multispectral optical data and synthetic aperture radar (SAR) satellite imagery. ALOS-2 PALSAR-2 gamma-0 backscatter annual mosaics (2015-2017) provided by JAXA were normalised and corrected using previous ALOS PALSAR annual mosaics (2007-2010) as reference. A multi-temporal Landsat 7 & 8 composite over the whole of Colombia was used for the year  $2016 \pm 1$ . The national forest inventory in-situ plots used to train our model consisted of 5-subplots each and were collected during the period 2015-2017 in the main biomes of the country. A sample of permanent 1ha plots (PPMs) were also measured. Nationally developed allometries (Alvarez et al, 2012) were used to estimate AGB. A non-parametric random forests (RF) algorithm was used within a k-fold framework to retrieve AGB at 30m spatial resolution for the whole of Colombia. The algorithm was trained using forest inventory plots and validated at plot (0.35 ha) and PPM level (1 ha). The accuracy assessment found coefficients of determination ( $R^2$ ) of 0.68 and 0.61, and relative root mean square errors (Rel. RMSE) of 49% and 34% at plot and at PPM level, respectively. The results showed that the average AGB for the country was  $118.1 \text{ t ha}^{-1}$  ( $45.6 \text{ t ha}^{-1}$  for Caribe,  $75.4 \text{ t ha}^{-1}$  Andes,  $122.5 \text{ t ha}^{-1}$  Pacifico,  $32.7 \text{ t ha}^{-1}$  Orinoquia, and  $200.5 \text{ t ha}^{-1}$  for the Amazonia, regionally), and that the total carbon stocks for the country were  $6.7 \text{ Pg C}$  for the period 2015-2017.

### 1. INTRODUCTION

Forest cover area in Colombia amounts to 52% of the national territory. Forests are one of the most important resources in terms of goods and ecosystem services, which are vital for the sustainable development of the country. Thus, forest monitoring is essential for decision-making processes related to forest management and reduction of deforestation.

The government of Colombia signed the Joint Declaration of Intent in 2015 with the governments of Norway, Germany and the United Kingdom to cooperate in reducing emissions from deforestation and forest degradation (i.e. REDD+) and promoting sustainable development in Colombia. In the last years, Colombia has strengthened its monitoring capabilities by developing the National Forest Information System, the Forest and Carbon Monitoring System, and the National Forest Inventory (NFI). Those have allowed to improve the monitoring of forest dynamics associated to disturbances leading to greenhouse gas emissions, and to promote the conservation and sustainable management of forests.

Aboveground biomass (AGB) is an Essential Climate Variable (ECV) required by the International Panel on Climate Change (IPCC) to carry out greenhouse gas inventories of forest ecosystems. AGB is the basic unit to account for carbon stored in vegetation. Activities aiming at the reduction of greenhouse

gas emissions from deforestation and forest degradation require information on the spatial distribution of the carbon stocks. This information can be used to evaluate the carbon gains due to new forest or growth, as well as the loss due to forest disturbances such as deforestation or degradation.

The National Forest Inventory of Colombia started in 2015 and it is an initiative aiming to monitor forest resources nationwide (IDEAM, 2018). The NFI collects information on vegetation structure, composition, richness, biodiversity, biomass, growing stock, soil carbon, and forest dynamics. This information allows to monitor the status of forests, the carbon stocks, and the composition of forest resources through time. However, the number of sampling units to date is quite limited due to difficulty to access the most remote forest areas (i.e. Amazon region).

The aim of this study is to estimate the amount and spatial distribution of Colombia's AGB using the current NFI sample and Earth Observation (EO) data. Then, the results will be compared with the current NFI estimates at national and regional level to understand where the main discrepancies are.

\* Corresponding author

## 2. MATERIALS

### 2.1 National Forest Inventory (NFI)

The NFI uses two types of sampling units: plots and 1-ha square plots (PPMs). Approximately 300 plots and 26 PPMs have been measured since 2015 (IDEAM, 2018). However, the Amazon region still presents very sparse sampling in the core areas (Figure 1).

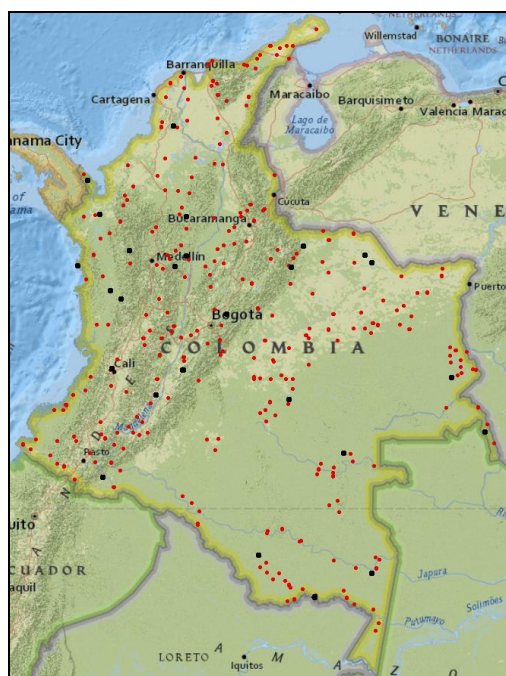


Figure 1. Location of plots (red dots) and PPMs (black squares)

A NFI plot is composed of five 30 m diameter circular subplots (SPFs) (Figure 2). Thus, each plot consists of 0.35 ha of actual measurements (0.07 ha per SPF). The dimensions of the PPMs are 100m x 100m and usually share the central coordinates with the centre of the central SPF (SPF-1) within an existing plot (Figure 2). Additionally, the SPF follow a concentric nested sampling for different diameter classes (Figure 2).

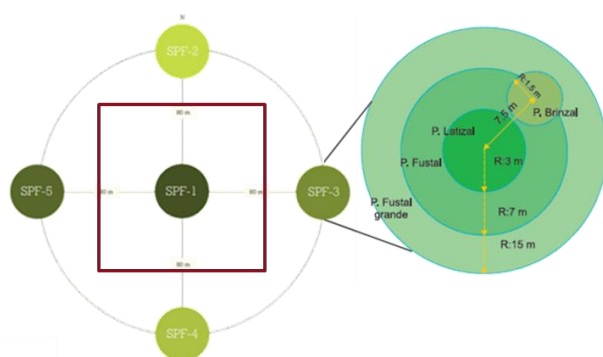


Figure 2. Structure of plots and 1-ha square plots

The NFI plots are used as the main training/validation sampling unit in this study, while PPMs provided an additional validation scenario. The size and design of the PPMs would be more optimal than the plots for using with EO data (Réjou-Méchain et al., 2014), but the reduced number of those limits its current

use. A 10 m buffer was applied in this study around each 30 m SPF in order to reduce location errors. This resulted in actual 50 m diameter subplots (~0.2 ha per subplot x 5 = 1 ha per plot) used to sample the overlapping pixels from the EO datasets for the training of the algorithm.

The NFI calculates carbon stocks nationally and regionally using the ratio estimator frequently used for the Global Forest Resources Assessment reporting (Marklund et al., 2005).

### 2.2 Earth Observation Datasets

A combination of multispectral and SAR datasets was used in this study. The JAXA's ALOS-2 PALSAR-2 freely available 25m resolution mosaics with standard SAR pre-processing (Shimada et al., 2014) were used. This processing consists of calibration, multi-looking (output of 16 looks), projection, ortho-rectification, slope correction using SRTM DEM and an additional destriping process (Shimada and Isoguchi, 2002). We additionally process these mosaics with the following steps: 1) removal of problematic pixels based on quality layers, 2) multi-temporal filtering (5x5 window), 3) normalization and correction based on temporal statistics of ALOS PALSAR previous mosaics, and 4) gap-filling. Landsat 7 and 8 multispectral 30m spatial resolution normalized top of atmosphere (TOA) annual composites (Hansen et al., 2013) were also used. The multispectral dataset Proba V 100m (Dierckx et al., 2014), and the 30m Digital Elevation Model from the Shuttle Radar Topography Mission (SRTM) were also explored during the variable selection stage.

## 3. METHODS

### 3.1 Allometric models

Non-destructive sampling of tree level variables such as diameter at breast height ( $D$ ), canopy height ( $H$ ), and species identification were collected from trees, palms and shrubs. AGB was calculated for trees using the pantropical allometric models of Chave et al. (2014) and the national models of Alvarez et al. (2012), and then compared. In both cases the allometric models with  $D$  as sole predictor variable were used. In the case of palms and shrubs, the Tiepolo et al. (2002) allometric models were used.

### 3.2 Algorithm

The implementation of the non-parametric machine learning algorithm Random Forest (Breiman, 2001) in the Google Earth Engine platform was used. The regression version of the algorithm was implemented within a k-fold framework to maximize the number of plots available for training and validation of the algorithm (i.e. 300 plots). Therefore, all the data was used for training and validation purposes. The dataset was stratified by three AGB levels, and randomly sampled into the folds to ensure that all folds have similar probability distribution functions of AGB.

The EO input variables for the algorithm were selected based on quantitative analyses (i.e. jackknife analyses) as well as qualitative observations (minimum presence of artefacts in the resulting map).

#### 4. RESULTS

The allometric models from Chave et al. (2014) derived in a 11% higher AGB estimate in comparison to Alvarez et al. (2012) models (average AGB of 125 t ha<sup>-1</sup> vs. 113 t ha<sup>-1</sup>) (Figure 3). The national models from Alvarez et al. (2012) were finally used to develop the AGB map.

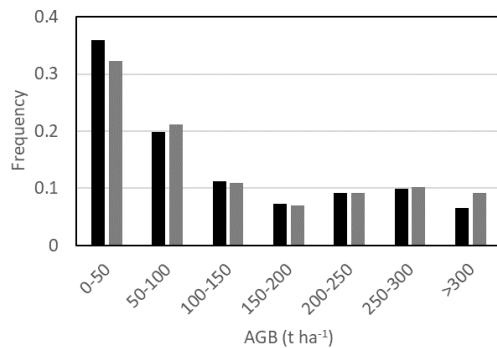


Figure 3. Histograms of AGB estimates from NFI plots using Chave et al. (2014) and Alvarez et al. (2012) models

The results of the Jackknife analyses based on root mean square error (RMSE), bias, and coefficient of determination ( $R^2$ ) indicated that the combination of ALOS-2 PALSAR-2 and Landsat variables showed the best balance between the lowest possible bias and RMSE, and the highest possible  $R^2$  (Figure 4)

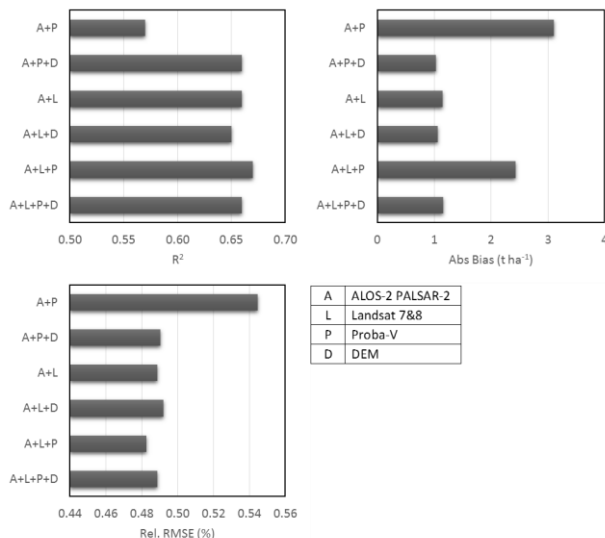


Figure 4. Jackknife analyses based on  $R^2$ , bias and RMSE for the potential input variables

The AGB map was therefore generated using ALOS-2 PALSAR-2 and Landsat 7 & 8 composite variables at 30 m spatial resolution for the whole of Colombia (Figure 5).

The average AGB density estimated for the country using the AGB map was 118.1 t ha<sup>-1</sup>. By regions, the average AGB density values were 45.6 t ha<sup>-1</sup> for Caribe, 75.4 t ha<sup>-1</sup> for Andes, 122.5 t ha<sup>-1</sup> for Pacifico, 32.7 t ha<sup>-1</sup> for Orinoquia, and 200.5 t ha<sup>-1</sup> for the Amazonia. The total forest carbon stock for the country was estimated in 6.7 Pg C for the period 2015-2017. The Amazon region has more than half of the AGB stocks in the country.

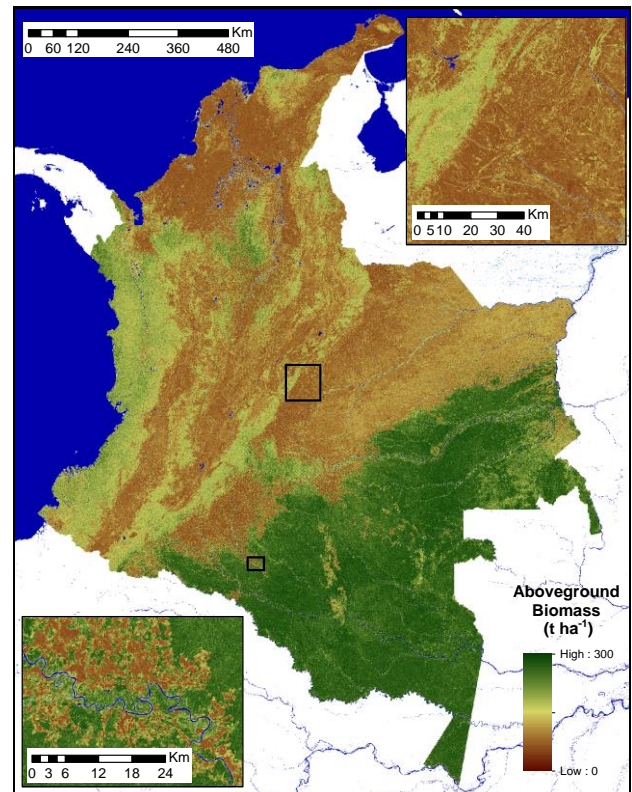


Figure 5. Colombia AGB map for year 2016±1

The  $R^2$  of the final map was 0.68, the Rel. RMSE 49%, and the bias -2.6 t ha<sup>-1</sup> (Figure 6). The additional validation using the PPMs showed similar results with  $R^2=0.61$ , Rel. RMSE=34% and Bias=-7.1 t ha<sup>-1</sup>.

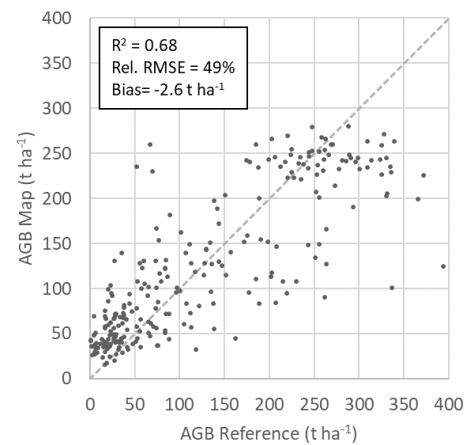


Figure 6. Scatterplot AGB map vs. AGB reference

The AGB stocks calculated using the AGB map were comparable to current estimates from the NFI. Table 1 shows that although the AGB estimates derived from the map are lower than the NFI estimates, the differences are not significant. The total AGB estimate for the whole country was only a 2.93% lower than the NFI national estimate, while regionally the differences were between 0.87% and 28.19% (Table 1).

Table 1. Comparison between AGB map estimates and NFI estimates by region

Region	Total AGB (Pg)	Difference to NFI (Pg)	Difference to NFI (%)
Amazonía	9.19	1.453	13.65
Andes	2.20	0.252	10.28
Caribe	0.70	0.006	0.87
Orinoquía	0.55	0.216	28.19
Pacífico	0.82	0.147	15.14
<b>Total</b>	<b>13.47</b>	<b>0.406</b>	<b>2.93</b>

## 5. DISCUSSIONS

The AGB plot data calculated using Alvarez et al. (2012) allometric models showed 11% lower values than AGB data calculated by Chave et al. (2014) models. These differences could be related to the different sampling datasets used to develop the models per se (national vs. pantropical). However, it can also be related to the use of the simple models using  $D$  as predictor variable instead of the models using  $D$ ,  $H$ , and specific wood gravity ( $\rho$ ). The AGB map was finally derived by training the algorithm with AGB plot data derived from Alvarez et al. (2012) allometric models.

The largest difference between the AGB map estimates and the NFI estimates occurs in the Orinoquia and Pacifico regions. Those are the regions with the lowest number of plots available from the NFI and the differences might indicate the need for more sample plots. Additionally, even though Orinoquia region is mostly covered by savannah vegetation (forest represents only 13% of the area), more than half of the plots sampled in that region were located in forests. This might lead to an overestimation by the NFI in that region using the ratio estimator method and explain the large difference found with the AGB map in that region (28.19%). Caribe was the region with the highest agreement to the AGB map estimates with just a 0.87% difference. This region had the highest sampling per unit of area, which

## 6. CONCLUSION

This is the most detailed and accurate map to date produced for Colombia using national data. Initial results using a combination of SAR ALOS-2 PALSAR-2 and multispectral Landsat 7 & 8 annual composites show promising results. Future work however should explore the use of allometric models with  $D$ ,  $H$ , and  $\rho$  as predictors to generate more accurate AGB plot data to train the algorithm, as well as larger plots more suited to use in combination with EO data.

## ACKNOWLEDGEMENTS

The authors would like to thank JAXA, NASA, and USGS for freely providing the EO data. Thanks also to Google for providing the processing platform (Google Earth Engine). This research was supported by the Forests 2020 project. The Forests 2020 project is funded by the UK Space Agency International Partnership Program. P. Rodríguez Veiga and H. Balzter were supported by the UK's National Centre for Earth Observation (NCEO).

## REFERENCES

- Alvarez E, Duque A, Saldarriaga J, Cabrera K, de las Salas G, del Valle I, et al. Tree above-ground biomass allometries for carbon stocks estimation in the natural forests of Colombia. *Forest Ecology and Management*. 2012;267(0):297-308.
- Breiman, L. 2001. Random forests. *Machine Learning*, 45, 5-32.
- Chave, J., Réjou- Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., Duque, A., Eid, T., Fearnside, P. M. & Goodman, R. C. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global change biology*, 20, 3177-3190.
- Dierckx, W., Sterckx, S., Benhadj, I., Livens, S., Duhoux, G., Van Achteren, T., Francois, M., Mellab, K. & Saint, G. 2014. PROBA-V mission for global vegetation monitoring: standard products and image quality. *International journal of remote sensing*, 35, 2589-2614.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O. & Townshend, J. R. G. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342, 850-853.
- IDEAM (Institute of Hydrology, Meteorology and Environmental Studies). National Forest Inventory. Url: <http://www.ideam.gov.co/web/ecosistemas/inventario-forestal-nacional>. 2018.
- Marklund, L.G. and Schoene, D.I.E.T.E.R., 2006. Global assessment of growing stock, biomass and carbon stock. *Forest Resources Assessment Programme Working paper*, 106.
- Réjou-Méchain, M., Muller-Landau, H. C., Detto, M., Thomas, S. C., Toan, T. & Saatchi, S. S. 2014. Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks. *Biogeosciences*, 11.
- Shimada, M. & Isoguchi, O. 2002. JERS-1 SAR mosaics of Southeast Asia using calibrated path images. *International Journal of Remote Sensing*, 23, 1507-1526.
- Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R. & Lucas, R. 2014. New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment*, 155, 13-31.
- Tiepolo, G., Calmon, M., & Feretti, A. R. 2002. Measuring and monitoring carbon stocks at the Guaraqueçaba climate action project, Paraná, Brazil. In *Proceedings of the International Symposium on Forest Carbon Sequestration and Monitoring* (pp. 11-15). Taipei: Taiwan Forestry Research Institute.

Revised January 2019