Article

Not peer-reviewed version

Comparison Between Traditional Forest Inventory and Remote Sensing with Random Forest for Estimating the Periodic Annual Increment in a Dry Tropical Forest

Anelisa Pedroso Finger *, Rinaldo Luiz Caraciolo Ferreira, Mayara Dalla Lana, José Antônio Aleixo da Silva, Emanuel Araújo Silva *, Fábio Marcelo Breuning, Polyanna Da Conceicao Bispo, Veraldo Liesenberg, Sara Sebastiana Nogueira

Posted Date: 21 May 2025

doi: 10.20944/preprints202505.1655.v1

Keywords: Vegetation indices; Spectral reflectance; Google Earth Engine; Machine learning; Forest productivity; Climate variability; Basal area; Dryland monitoring; Landsat imagery; Semi-arid ecosystems

 doi:10.20944/preprints202505.1655.v1

Article

Comparison Between Traditional Forest Inventory and Remote Sensing with Random Forest for Estimating the Periodic Annual Increment in a Dry Tropical Forest

Anelisa Pedroso Finger [1,*], Rinaldo Luiz Caraciolo Ferreira[1], Mayara Dalla Lana[2], José Antônio Aleixo da Silva[1], Emanuel Araújo Silva [1,4], Fábio Marcelo Breunig[3], Polyanna Conceição Bispo[5], Veraldo Liesenberg[6] and Sara Sebastiana Nogueira[1]

[1] Department of and nd rest Sciences, Federal Rural University of Pernambuco, Recife-PE, CEP 52171-900, Brazil
[2] Federal Institute of Pernambuco, Federal Institute of Pernambuco - Garanhuns Campus, Garanhuns-PE, CEP 55299390, Brazil
[3] Department of Geography, Earth Science Sector, Federal University of Paraná (DGEOG/SCT/UFPR). Curitiba-PR, CEP 81531-970, Brasil
[4] Federal University of Santa Maria, Department of Forest Engineering - Frederico Westphalen Campus, Frederico Westphalen-RS, CEP 98400000, Brazil
[5] Department of Geography, School of Environment Education and Development, University of Manchester, Manchester M13 9PL, UK
[6] State University of Santa Catarina, Agro-Veterinary Center, Lages-SC, CEP 88520000, Brazil
* Correspondence: anefinger@yahoo.com.br

Abstract: This study evaluates the effectiveness of combining remote sensing techniques with the Random Forest algorithm for estimating the Periodic Annual Increment (PAI) in a dry tropical forest lo-cated within the Caatinga biome in northeastern Brazil. The analysis integrates forest inventory data collected from permanent plots monitored between 2011 and 2019 with Landsat satellite imagery processed through the Google Earth Engine platform. By incorporating surface reflec-tance and vegetation indices, the approach significantly improved the accuracy of productivity estimates while reducing the costs and efforts associated with traditional field-based methods. The Random Forest model achieved a strong performance (R^2 = 0.8867; RMSE = 0.87), and its predictions were further refined using post-processing correction factors. These results demon-strate the potential of data-driven modeling to support forest monitoring and sustainable man-agement practices, especially in ecosystems vulnerable to the impacts of climate change.

Keywords: vegetation indices; spectral reflectance; google earth engine; machine learning; forest productivity; climate variability; basal area; dryland monitoring; landsat imagery; semi-arid ecosystems

1. Introduction

Native forests play a central role in the provision of timber and non-timber products, such as wood, fibers, and fuels. Beyond their industrial utility, these resources are renewable and exhibit high energy efficiency, making them environmentally responsible alternatives to fossil fuel-derived products, plastics, and metals. When sustainable management techniques are adopted, these forests contribute to soil and water conservation [1], the recovery of degraded areas [2], and offer social benefits, including job creation and regional development [3].

Research on dry tropical forests in Brazil has investigated aspects related to forest recovery [4], structure and dynamics [5,6], carbon stock [7–9], and, more recently, the use of remote sensing for

 doi:10.20944/preprints202505.1655.v1

vegetation monitoring [10,11]. These efforts aim to bridge knowledge gaps regarding the ecology of these forests, particularly in the context of climate change [12], which exacerbates their vulnerability due to the increasing frequency of extreme climatic events [13,14].

Although tropical forests are recognized for their high biodiversity, they may be dominated by a few species at local scales. This phenomenon is also observed in seasonally dry biomes, such as the Caatinga, where sparse vegetation can exhibit high diversity due to environmental heterogeneity and species' adaptive strategies to extreme climatic conditions. Studies indicate that despite being predominantly xerophytic and of low density, the Caatinga harbors significant biodiversity, with structural variations influenced by climatic and edaphic factors [15,16]. Such dominance patterns often reflect successional and demographic variations over time [17]. Climatic factors, such as precipitation and temperature [18,19], along with soil fertility and other edaphic conditions [20,21], shape the structure and composition of these plant communities. Furthermore, species resilience to environmental disturbances is crucial for their survival and dispersal [22].

To better understand forest dynamics, growth assessment through the Periodic Annual Increment (PAI) has emerged as a critical metric. This measure allows for the evaluation of forest growth in response to environmental conditions. In dry tropical forests, such as the Caatinga, seasonal water availability plays a central role in influencing growth processes and natural regeneration. Extreme climatic events, such as severe droughts, can substantially alter forest dynamics, highlighting the need for studies correlating PAI with these environmental variables [23,24]. It is also crucial to further investigate the Cerrado and Caatinga biomes, considering anthropogenic impacts [25].

Traditionally, PAI monitoring has been conducted through permanent plots, where periodic measurements provide data on vegetation development. However, this method presents significant challenges, including the need for extensive fieldwork, high costs, and logistical difficulties in remote areas. These limitations can restrict the spatial and temporal scope of studies, compromising the effectiveness of continuous monitoring [26]. There is evidence of an encroachment process occurring in the Cerrado biome [27,28], which may also be taking place in the Caatinga. To understand such phenomena, methods with broad spatial and temporal coverage are essential.

With the advancement of remote sensing technologies, it has become possible to overcome some of these barriers. The use of satellite imagery and unmanned aerial vehicles (UAVs) allows for large-scale data collection with higher frequency and lower cost. One of the main applications of these technologies is the use of vegetation indices (VIs) and spectral reflectance, which are metrics derived from vegetation reflectance that enable the monitoring of vegetation cover, plant vigor, and ecosystem processes over time [29–33]. For example, [34] developed a low-cost UAV-based forest restoration monitoring protocol, demonstrating its effectiveness in Brazil's Atlantic Forest. Similarly, [35] implemented a national-scale forest monitoring system in Tanzania using remote sensing data to support management and conservation decisions. These examples illustrate how remote sensing can complement traditional methods, providing a more comprehensive and updated view of forest dynamics.

Landsat satellite imagery has proven to be a valuable tool for multi-temporal analysis, enabling high-precision vegetation cover change detection over more than four decades [36–39], including harmonized data products [40,41]. The availability of orbital data and the demands associated with dry tropical forest monitoring require robust information extraction methods.

Machine learning algorithms, such as Random Forest, Gradient Boosting, Support Vector Machines (SVM), and Artificial Neural Networks (ANN), can enhance predictive capabilities and the analysis of complex ecosystems [42–46]. These algorithms, by handling large volumes of multivariate data, have demonstrated efficiency in ecological modeling, enabling the identification of relevant patterns and key variables influencing forest growth and species resilience [47–49].

Among these, Random Forest stands out for its robustness in analyzing large-scale multivariate datasets. This algorithm employs multiple decision trees to construct highly accurate predictive models, facilitating the identification of key factors affecting forest growth [42,50,51]. Its application

 doi:10.20944/preprints202505.1655.v1

in ecological studies provides valuable insights into the factors influencing forest dynamics and species resilience [50,52,53].

To assess the potential of machine learning techniques and satellite imagery, this study aims to compare the Periodic Annual Increment (PAI) estimates obtained through traditional field-based forest inventory methods with those derived from remote sensing techniques, employing machine learning algorithms such as Random Forest. The Caatinga biome was selected as the study area.

2. Materials and Methods

To provide a clear overview of the research workflow, Figure 1 presents a schematic representation of the methodological steps adopted in this study, highlighting the integration of field and orbital data, variable processing, statistical analysis, and the use of the Random Forest algorithm for the estimation and spatialization of the Periodic Annual Increment (PAI).

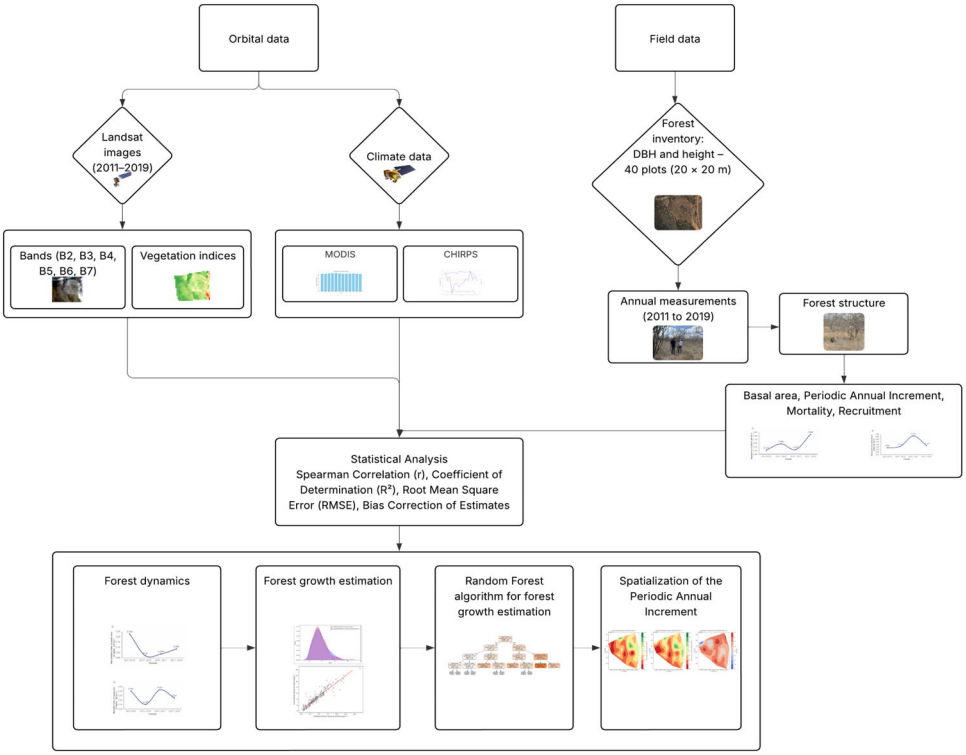

Figure 1. Methodological flowchart summarizing the integration of field and satellite data, extraction of spectral and structural variables, statistical analysis and Random Forest modeling for the estimation and spatialization of the Periodic Annual Increment (PAI) in a dry tropical forest area.

2.1. Study Area Location

The data were collected from a 1.6-hectare sampling area, located within Fazenda Itapemirim (Agrimex S.A, Floresta, Pernambuco, Brazil). The farm covers 5,695.65 hectares, where native vegetation was cleared in 1987 using tractors and chains [54]. At the time, the purpose of removing the native vegetation was to establish exotic species. However, the project failed, leading to the abandonment of the area, which subsequently entered a natural regeneration process.

The study area has been monitored since 2008 through 40 permanent plots, each measuring 20 m × 20 m (400 m²), totaling a sampled area of 1.6 ha. The plots are evenly spaced 80 m apart, measured from the center, and the buffer zone between plots is 60 m (Figure 2).

 doi:10.20944/preprints202505.1655.v1

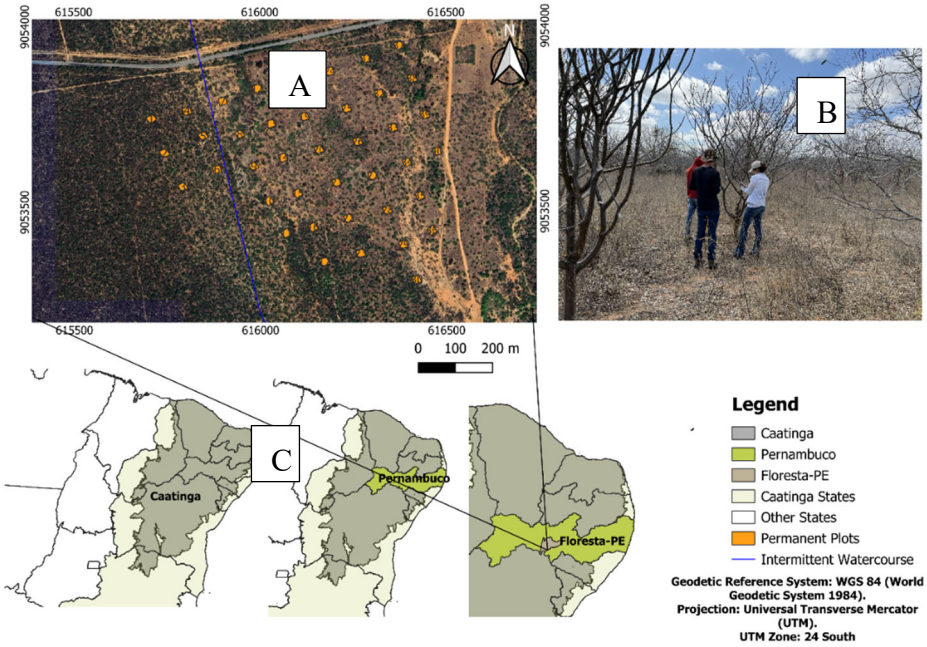

Figure 2. (A) Distribution of the 40 permanent plots in the study area, shown over high-resolution satellite imagery. Orange points indicate plot locations, and the blue line represents an intermittent wa-tercourse. (B) Fieldwork activities conducted in the Caatinga vegetation within one of the per-manent plots, illustrating the dry forest structure. (C) Geographic location of the study area: the Caatinga biome in northeastern Brazil, highlighting the state of Pernambuco and the municipality of Floresta, where the field experiment was conducted.

2.2. Field Data Collection

For standardization purposes, plants were identified and tagged at a height of 1.30 m above ground. Annually, starting in 2011, all shrub and tree species were identified and measured in each sampling unit, recording the circumference at 1.30 m above ground level ($C \geq 6$ cm) and total height (H). Additionally, newly recruited individuals (plants and/or stems reaching the minimum C criterion), as well as dead and fallen individuals, were recorded. The study followed the nomenclature and measurement standards for forest inventory [55].

A plant was considered any part belonging to a single root system, based on the observation of its above ground structure. In cases where bifurcations occurred below 1.30 m in the same plant and met the minimum inclusion criteria, they were recorded as stems.

Data were collected each September to coincide with the dry season. Circumference measurements were taken using a measuring tape, while height was measured with a graduated ruler, both with centimeter precision. Circumferences were converted to diameter at 1.30 m above ground level (D) by dividing C by π.

2.3. Climatic Data

Temperature and precipitation data for the period from 2009 to 2023 were obtained from the LP DAAC database of the USGS and the Climate Hazards Center – UC Santa Barbara, University of California. The dataset was sourced from the MODIS/006/MOD11A1 and CHIRPS/DAILY products, based on the geographic coordinates of the study area[56,57].

2.4. Remote Sensing Data

Landsat 5 and Landsat 8 satellite images were selected from the complete collection available on the Google Earth Engine platform, using the LANDSAT/LT05/C02/T1_L2 and

 doi:10.20944/preprints202505.1655.v1

LANDSAT/LC08/C02/T1_L2 libraries, respectively. Landsat 5 was used for the year 2011, while Landsat 8 data were employed for the years 2013, 2015, 2017, and 2019.

The images were preprocessed by applying cloud and radiometric saturation masks, limiting cloud coverage to less than 60% to ensure data quality. After preprocessing, the spectral surface reflectance bands were corrected for each sensor, and vegetation indices were calculated for each analyzed year. These data were extracted specifically for the study area, ensuring temporal and spatial consistency and comparability (Table 1).

Table 1. Vegetation indices used in the study area.

Vegetation Index	Formula	Reference Source
Normalized Difference Vegetation Index (NDVI)	(NIR - RED) / (NIR + RED)	ROUSE et al., 1973
Enhanced Vegetation Index (EVI)	2.5 * (NIR - RED) / (NIR + 6 * RED - 7.5 * BLUE + 1)	HUETE et al., 2002
Soil-Adjusted Vegetation Index (SAVI)	(1 + L) * (NIR - RED) / (NIR + RED + L)	HUETE,1988
Green Normalized Difference Vegetation Index (GNDVI)	(NIR - GREEN) / (NIR + GREEN)	GITELSON; MERZLYAK, 1994
Normalized Difference Water Index (NDWI)	(NIR - SWIR) / (NIR + SWIR) or (GREEN - NIR) / (GREEN + NIR)	GAO, 1996
Normalized Difference Built-up Index (NDBI)	(SWIR - NIR) / (SWIR + NIR)	ZHA et al., 2003
Moisture Stress Index (MSI)	SWIR / NIR	ROCK et al., 1986
Normalized Burn Ratio (NBR)	(NIR - SWIR) / (NIR + SWIR)	KEY; BENSON, 2001
Normalized Difference Snow Index (NDSI)	(GREEN - SWIR) / (GREEN + SWIR)	RIGGS et al., 1994
Leaf Area Index (LAI)	Ln((0.69 – SAVI) / 0,59) / 0,91	TUCKER et al., 1981

In which: NIR: Near-Infrared; RED: Red; GREEN: Green; BLUE: Blue; SWIR: Short-Wave Infrared; L: Soil adjustment factor for SAVI, which typically ranges from 0 (for dense vegetation cover) to 1 (for sparse vegetation cover); in this study, a value of 0.5 was used.

To ensure compatibility between the two sensors, Band 6 of the Landsat 5/TM satellite and Bands 1, 8, 9, 10, and 11 of the Landsat 8/OLI satellite were excluded from the analysis.

2.5. Recruitment and Mortality Dynamics

In this study, a two-year monitoring period was considered between surveys conducted from 2011 to 2019. Thus, an individual (stem) was classified as a recruit if it reached C ≥ 6 cm only in the second measurement. This means that individuals (stems) that emerged in 2013 were counted for the 2011–2013 interval, those that emerged in 2015 for the 2013–2015 interval, those that appeared in 2017 for the 2015–2017 interval, and those that appeared in 2019 for the 2017–2019 interval.

The recruitment rate was calculated using Equation 1.

 doi:10.20944/preprints202505.1655.v1

$$TR = \left[1 - \left(1 - \frac{N_r}{N_f}\right)^{\frac{1}{t}}\right] x\ 100 \quad (1)$$

Where: TR = Recruitment rate (%); t = Time elapsed between inventories (years); Nf = Final count of living trees (stems) per hectare; Nr = Number of recruited individuals (stems) per hectare

For mortality estimation, an individual (stem) was considered dead if it was an adult shrub or tree with C ≥ 6 cm in the 2011 survey but was found dead in the subsequent 2013 survey (first interval: 2011–2013). For the second interval (2013–2015), trees recorded as alive in 2013 but dead in 2015 were counted. Similarly, for the third interval (2015–2017), trees that were alive in 2015 but dead in 2017 were considered. Finally, for the fourth interval (2017–2019), trees recorded as alive in 2017 but dead in 2019 were included in the mortality count.

The mortality rate was calculated using Equation 2.

$$TM = \left[1 - \left(1 - \frac{N_m}{N_0}\right)^{\frac{1}{t}}\right] x\ 100 \quad (2)$$

Where: TM = Mortality rate (%); t = Time elapsed between inventories (years); N_0 = Initial number of trees (stems) per hectare; Nm = Number of dead individuals (stems) per hectare.

Growth estimates were obtained using the methodology proposed by [52], and the Periodic Annual Increments (PAI) for field data were calculated based on the ratio between gross and net growth (including and excluding recruits, respectively) and the interval between measurement periods, according to Equations 3, 4, 5, and 6.

$$Cb_i = G_2 + M + MF + C + CF - G_1 \quad (3)$$

$$Cb = G_2 + M + MF + C + CF - GI - GF - G_1 \quad (4)$$

$$Cl_I = Cb_i - M + MF - G_1 \quad (5)$$

$$Cl = G_2 - GI + GF - G_1 \quad (6)$$

Where: Cbi represents the gross growth including recruits; Cb is the gross growth excluding recruits; ClI corresponds to the net growth including recruits; Cl is the net growth excluding recruits. G_2 is the basal area ($m^2\ ha^{-1}$) in the final inventory, at the end of the growth period, while G_1 is the basal area ($m^2\ ha^{-1}$) in the initial inventory, at the beginning of the growth period. M represents the basal area of individual mortality ($m^2\ ha^{-1}$), and MF is the basal area of stem mortality in living individuals ($m^2\ ha^{-1}$). C refers to the basal area of harvested individuals (recruits) in $m^2\ ha^{-1}$, and CF is the basal area of harvested stems (recruits within individuals) in $m^2\ ha^{-1}$. GI denotes the basal area of recruited individuals ($m^2\ ha^{-1}$), and GF represents the basal area of recruited stems within individuals ($m^2\ ha^{-1}$).

2.6. Periodic Annual Increments

The periodic annual increments in basal area (PAI, $mm.ha^{-1}$) were obtained by calculating the ratio between Cbi, Cb, ClI, and Cl, respectively, and the interval between measurement periods, considering both individuals and stems.

Based on the estimation of Cb, the mean annual periodic increment in diameter was calculated using Equation 7.

$$\overline{PAI}_d = \frac{1}{n.a}\sqrt{\frac{40000.C_b}{\pi}} \quad (7)$$

Where: IPA represents the mean annual periodic increment in diameter ($mm.ha^{-1}$); Cb is the gross growth excluding recruits; n is the number of individuals; and a is the interval between measurement occasions (2 years).

 doi:10.20944/preprints202505.1655.v1

2.7. Modeling with Random Forest

Digital image processing was performed using QGIS (version 3.10.4). The extrac-tion of pixel reflectance values for the analyzed periods was carried out using the Point Sampling Tool plugin, which allowed capturing reflectance values associated with the study area. The extracted data, including reflectance and vegetation indices, were or-ganized into an Excel spreadsheet, containing detailed information such as plot identi-fication, metric coordinates, basal areas, periodic annual increments, reflectance values from Landsat 5 and Landsat 8 bands, and the vegetation indices calculated for each plot and analyzed period. These variables were later used as predictors in machine learning modeling.

To estimate the Periodic Annual Increment (PAI) using the remote sensing varia-bles of the study area, the Random Forest technique was employed, implemented in the Scikit-learn library in Python [58,59]. Random Forest is an ensemble learning algorithm based on the aggregation of multiple decision trees. Each tree is built from a bootstrap sample of the dataset, and the final prediction is obtained by averaging the outputs of individual trees. This approach reduces overfitting and improves generalization, mak-ing it particularly robust for modeling complex and nonlinear relationships, as often encountered in ecological and remote sensing data [42,50,51]. In addition to this library, Pandas [60] was used for data analysis, Scipy.stats [61] for computing Spearman's cor-relation, Matplotlib.pyplot [62] and Seaborn [63] for generating graphs, and Numpy [64] for matrix calculations.

Each model was trained with 400 estimators (trees), balancing prediction accuracy and computational efficiency. This number was chosen based on preliminary testing with different values (ranging from 100 to 500), which showed that performance gains stabilized around 400 estimators. Thus, this value was selected as an optimal point to avoid unnecessary computational cost while ensuring model robustness. The data were split into 80% for training and 20% for validation, following common predictive mod-eling practices [65].

The selection of variables was guided by the significance of Spearman's correlation [66] (Equation 8).

$$r_s = 1 - \frac{6\sum d_i^2}{n\,(n^2-1)} \quad (8)$$

Where d_i represents the difference between the ranks of each pair of observations, and n is the number of observations.

The model performance metrics included the Coefficient of Determination (R^2) and the Root Mean Square Error (RMSE) (Equations 9 and 10, respectively), to assess prediction accuracy [67].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \quad (9)$$

Where n is the number of observations, y_i represents the observed values, $\hat{y}_i$ represents the predicted values from the model, and $\bar{y}$ is the mean of the observed values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}} \quad (10)$$

Where y_i is the observed value, $\hat{y}_i$ is the predicted value from the model, and n is the number of observations.

2.8. Bias Correction of Estimates

Due to the observed bias in the Random Forest models, considering the complexity of the forest structure in a Dry Tropical Forest, a correction factor was applied based on the Periodic Annual Increment (PAI).

The correction factor was obtained as follows:

 doi:10.20944/preprints202505.1655.v1

a) Estimation Difference Calculation: First, the absolute difference between the actual values and the predicted values was calculated and stored in a new column in the spreadsheet. This was done to quantify the estimation error for each observation (Equation 11).

$$\Delta_i = |y_i - \hat{y}_i| \quad (11)$$

Where y_i represents the actual observed value of observation i, and $\hat{y}_i$ is the value predicted by the Random Forest model.

b) Data Classification: Next, the actual data were classified into ranges using the Qcut function, which divides the data into quantiles. The data were segmented into six equally sized classes, with an additional unit added to each class to avoid zero-based indexing, resulting in classes numbered 1 to 6. These classes were stored in a new column (Equation 12).

$$C_i = Classe(yi), C_i \quad \epsilon \{1, 2, \ldots, k\} \quad (12)$$

Where k = 6 is the number of classes.

c) Correction Factor Calculation: For each previously defined class, a mean correction factor was calculated based on the estimation difference. This was done by grouping the data by class and obtaining the mean difference for each group. These factors represent the average adjustment required for each value class, considering both basal area and periodic annual increment (Equation 13).

$$F_c = \frac{1}{nC} \sum_{i \epsilon C} \Delta_i \quad (13)$$

Where n_C is the number of observations in class C, and Δ_i represents the esti-mation difference for observation i in class C.

d) Application of the Correction Factor: The correction factor was then applied to adjust the original predictions. For each observation, the respective correction factor was added to the estimated value, resulting in a new column of corrected predictions (Equation 14).

$$\hat{y}_i^{corr} = \hat{y}_i + FC_i \quad (14)$$

e) Reevaluation of Model Metrics: After applying the correction factor, the model evaluation metrics were recalculated, including the coefficient of determination (R^2) and the root mean square error (RMSE) (Equations 15 and 16).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i^{corr})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i^{corr})^2} \quad (16)$$

2.9. Generation of Spatial Maps

After constructing the Random Forest models for estimating the Periodic Annual Increment (PAI), spatial distribution maps were generated. The Inverse Distance Weighting (IDW) interpolation technique was used, implemented through the Scipy.interpolate library. This method assigns greater weight to values closer to the predicted location and is well established for the spatial interpolation of environmental variables[68].

The maps were created using both estimated and observed values from the 40 plots. For each plot, the PAI value was interpolated across a grid covering the entire study area using IDW, allowing for the visualization of a spatial continuum of the variables of interest.

IDW was selected due to its simplicity, low computational cost, and suitability for datasets with a limited number of spatial points, as in this case (n = 40 plots). Although geostatistical methods such as Kriging provide estimates of spatial uncertainty, they require assumptions about stationarity and spatial autocorrelation that are difficult to satisfy with small sample sizes.

 doi:10.20944/preprints202505.1655.v1

3. Results and Discussion

3.1. Growth Dynamics

Figure 3 illustrates the annual variation in precipitation and the minimum, mean, and maximum temperatures in the study area over the years. Based on the graphical analysis, important trends can be observed, helping to contextualize the climatic factors that influence forest growth and the Periodic Annual Increment (PAI) in the Caatinga biome.

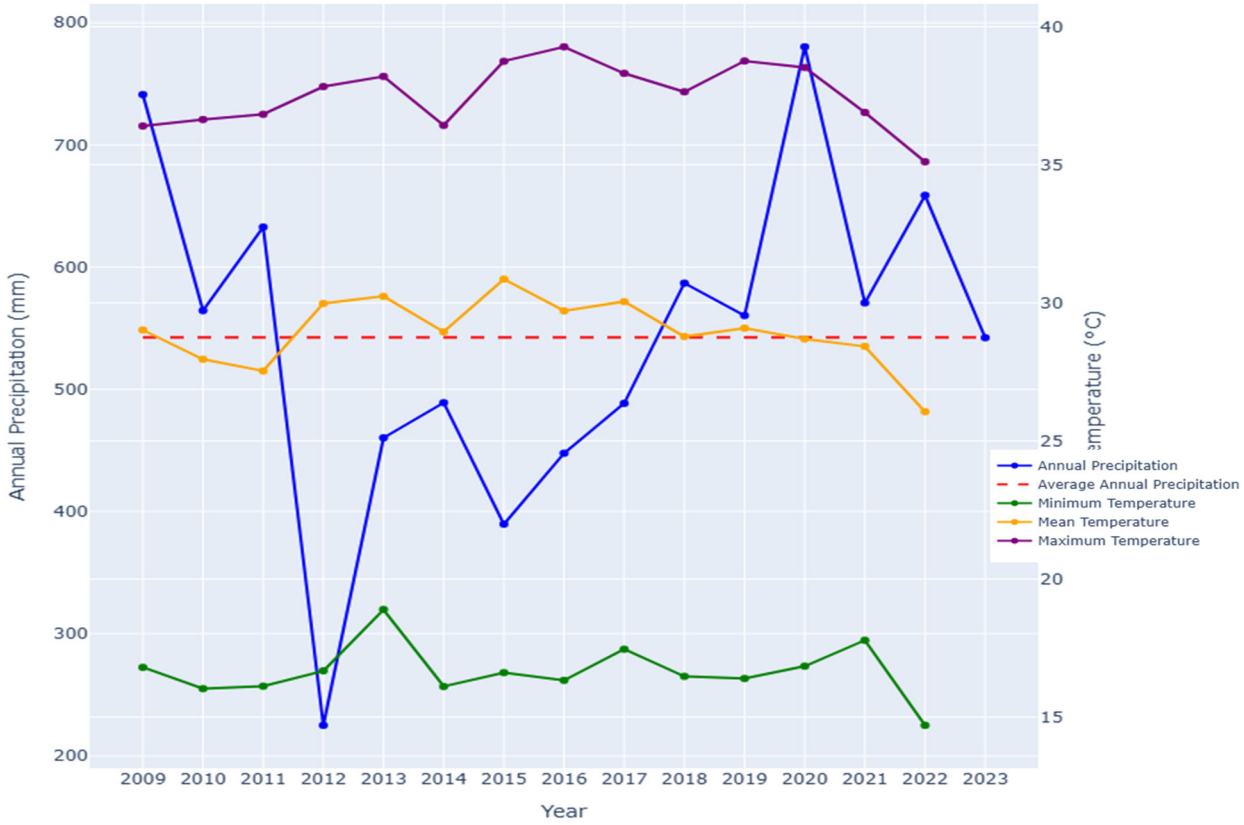

Figure 3. Annual variability of mean temperature and precipitation over 14 years of study, highlighting seasonal patterns and climatic fluctuations during the analyzed period.

Precipitation variability is a defining characteristic of the Caatinga biome, known for its cycles of intense and intermittent droughts. This variability has a direct impact on the Periodic Annual Increment (PAI), as tree growth in the Caatinga is highly dependent on water availability.

Between 2011 and 2019, total mortality in the study area exceeded the recruitment of both stems and individuals (Figures 4a, 4b and 5). This imbalance resulted in a negative net growth rate, highlighting the struggle for survival faced by various species and individual trees during this period. This timeframe coincided with a severe drought recorded in Northeastern Brazil, which lasted from 2010 to 2016 [5,69]. The recorded water deficit likely had a negative influence on vital plant processes, leading to a reduced growth rate and an increase in vegetation mortality.

 doi:10.20944/preprints202505.1655.v1

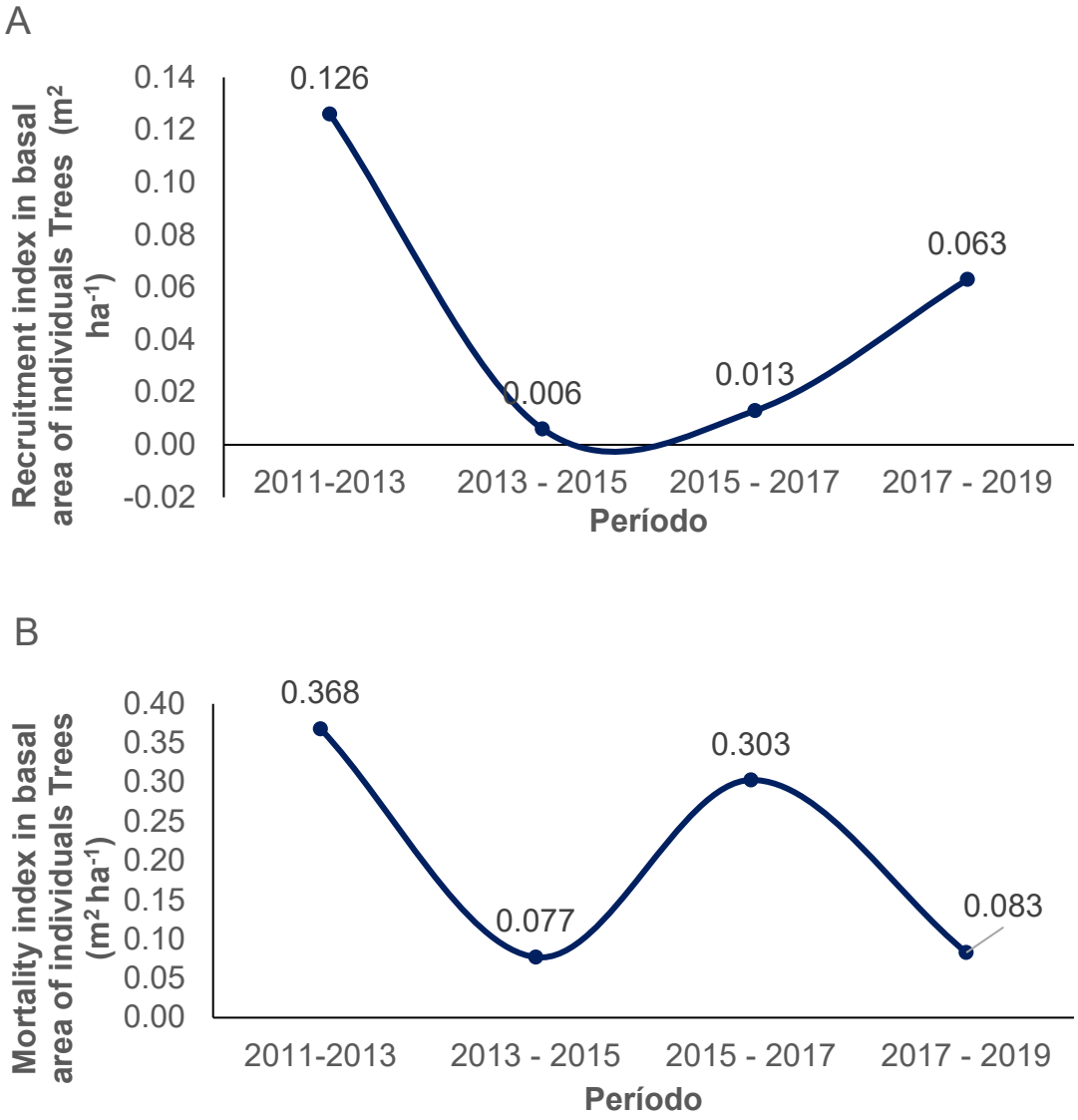

Figure 4. a. Recruitment and mortality index of individuals in basal area over the analyzed periods in Floresta-PE. (A) Recruitment index in basal area ($m^2\ ha^{-1}$) of individuals over the analyzed periods; (B) Mortality index in basal area ($m^2\ ha^{-1}$) of individuals over the analyzed periods.

 doi:10.20944/preprints202505.1655.v1

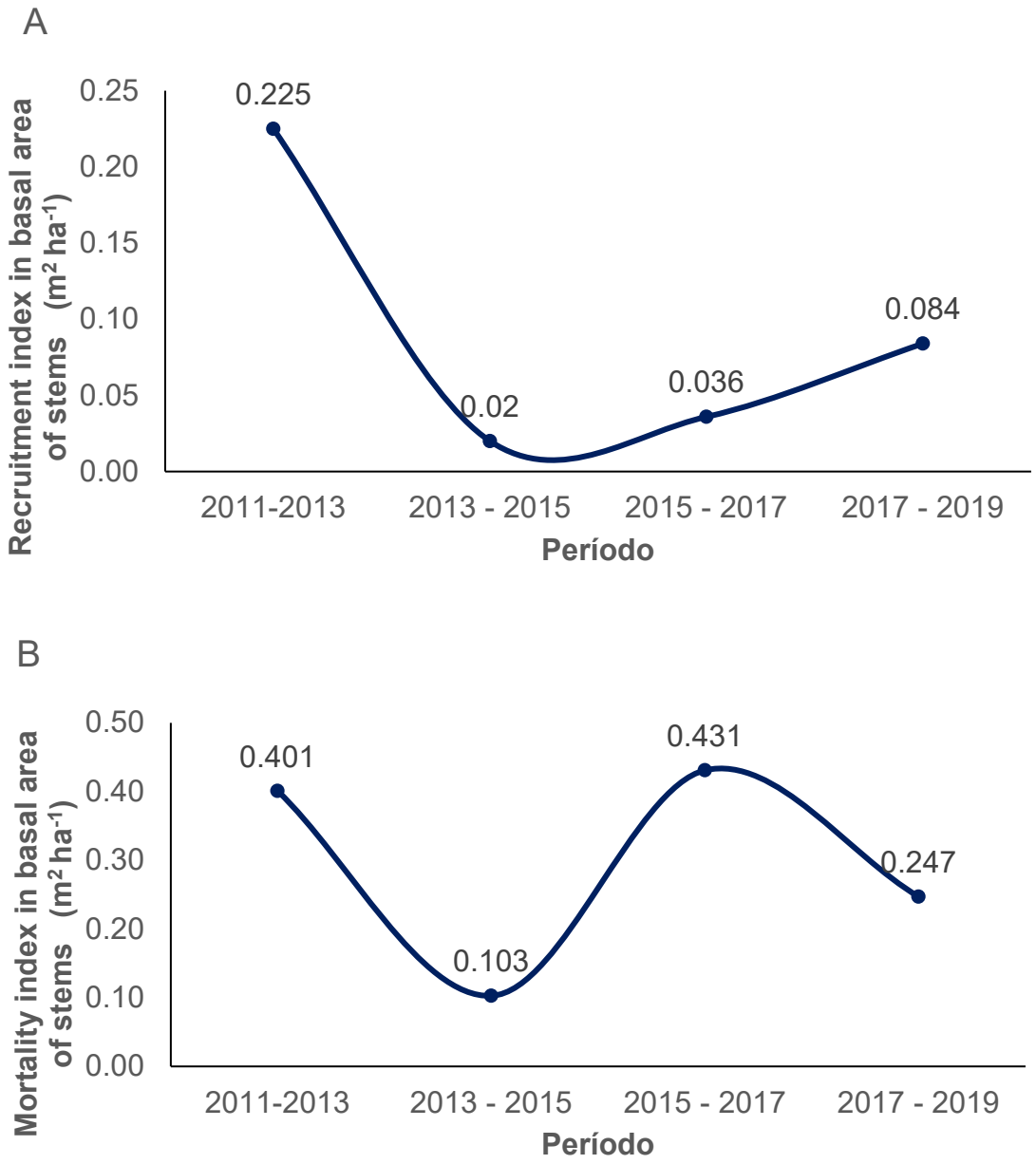

Figure 4. b. Recruitment and mortality index of stems in basal area over the analyzed periods in Floresta-PE. (A) Recruitment index in basal area ($m^2\ ha^{-1}$) of stems over the analyzed periods; (B) Mortality index in basal area ($m^2\ ha^{-1}$) of stems over the analyzed periods.

The analysis of biennial periods revealed that the maximum basal area increment was recorded in the last analyzed interval, reaching 2.844 $m^2\ ha^{-1}$ (Figure 5). This growth can be attributed to a reduction in individual mortality and a simultaneous increase in the mortality of smaller stems.

This pattern suggests an adaptive strategy of Caatinga species, in which there is a selective allocation of resources, such as nutrients and water, favoring stems with larger diameters. This natural selection mechanism, known as self-thinning, contributes to the optimization of resource use, promoting the survival and growth of more robust stems.

 doi:10.20944/preprints202505.1655.v1

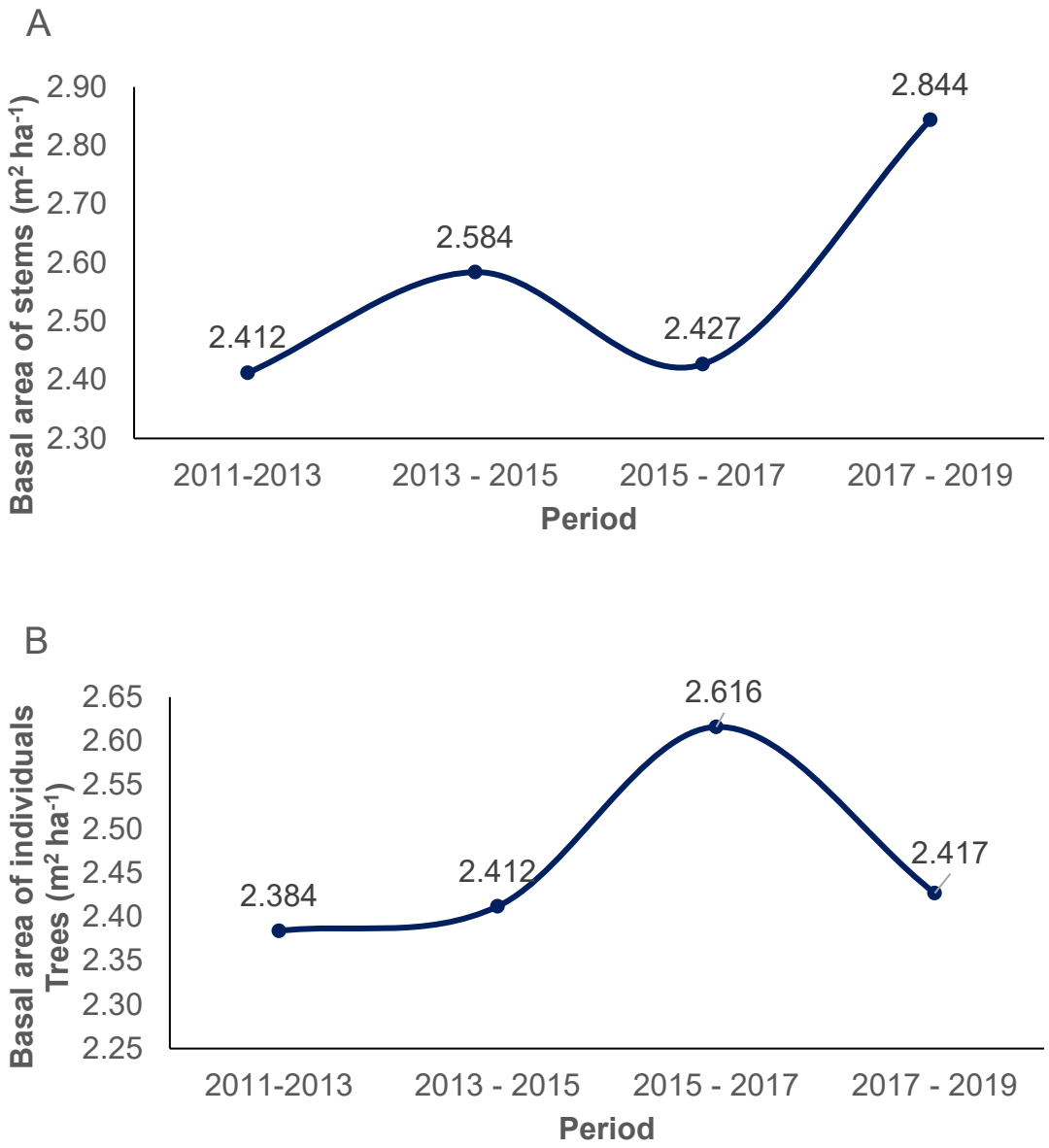

Figure 5. Variation in basal area (m² ha⁻¹) over the biennial periods. (A) Basal area of stems per hectare in different biennial periods in Floresta-PE; (B) Basal area of individuals per hectare over the analyzed periods in Floresta-PE.

The higher mortality rate compared to the recruitment of new individuals resulted in a negative net growth for several species between 2011 and 2019. This phenomenon reflects the challenges these species face in developing in this region, as the low recruitment rate significantly impacts the dynamics of the local community (Figure 4). [54] and [70] emphasize that population dynamics vary significantly among different species; while some populations respond positively to environmental conditions, others exhibit negative responses. This pattern highlights the complexity and variability inherent in ecological dynamics across different contexts.

Given the semi-arid context, a region frequently affected by intense drought periods, the local vegetation may experience a decline in the number of individuals, leading to an increase in mortality rates. This scenario can result in a negative net growth of populations, as indicated by [71]. The vulnerability of semi-arid flora to such extreme climatic conditions underscores the challenges these ecosystems face in maintaining their integrity and biodiversity.

 doi:10.20944/preprints202505.1655.v1

The predominance of stems over the total number of individuals can be attributed to an adaptive survival mechanism of these species, where bifurcation is a common characteristic among specimens (Figure 6).

Figure 6. Distribution of the number of stems per individual in the shrub-tree community, highlighting the structural pattern of the vegetation.

A study conducted in the Brazilian Cerrado biome by [72] demonstrated that wood density significantly influences water translocation within plants. Young plants, characterized by lower wood density, facilitate water transport, contributing to the development of new stems. However, as wood density increases, particularly under high-temperature conditions, water transport efficiency decreases, limiting water availability to certain branches. This process leads to natural pruning, where the less-supplied branches are eliminated, thereby optimizing the distribution of water resources within the plant. Such a process may contribute to woody encroachment, a phenomenon previously recorded in the Cerrado biome [27].

The results of the present study emphasize the complexity and dynamics of the Caatinga biome, addressing key aspects related to recruitment, mortality, basal area, and Periodic Annual Increment (PAI). The analysis of recruitment highlighted that climatic events, such as precipitation, play a determinant role in vegetation regeneration. This pattern aligns with the findings of [54], who identified a direct relationship between rainfall occurrence and the recruitment of woody species. In the context of forest management, these data reinforce the importance of considering climatic variability in the development of conservation strategies.

The observed mortality was particularly high during severe drought periods, such as in 2012, confirming previous studies by [73] and [69], which highlighted the vulnerability of Caatinga species to prolonged water deficits. Mortality, combined with recruitment, reflects the high turnover rate of dry tropical forests, showcasing their resilience but also their sensitivity to climate change. Additionally, mortality and recruitment values indicate the crucial role of interannual precipitation variability, this factor plays a key role in controlling forest dynamics, as discussed by [74].

 doi:10.20944/preprints202505.1655.v1

The observed forest dynamics patterns were largely influenced by water availability and the edaphic characteristics of the region. Areas with greater hydrological stability exhibited higher recruitment rates, supporting the observations of [10] regarding the impact of soil resources on the regeneration and growth of woody species. These dynamics are particularly significant in semi-arid regions, where climatic fluctuations play a central role in shaping the structure and composition of forests.

Between 2015 and 2021, *Cenostigma bracteosum* exhibited the highest basal area increment, with an average value of 0.210 m^2 ha^{-1} $year^{-1}$ over the analyzed period [3]. Other species with notable growth included Cnidoscolus quercifolius and Jatropha mollissima, whose PAI in basal area ranged between 0.165 and 0.198 m^2 ha^{-1} $year^{-1}$ from 2018 to 2021, coinciding with slightly higher precipitation levels.

Conversely, some species exhibited negative growth performance throughout the study. Bauhinia cheilantha and Combretum glaucocarpum recorded negative PAI values in basal area in some periods, with declines of -0.0200 m^2 ha^{-1} $year^{-1}$ and -0.0010 m^2 ha^{-1} $year^{-1}$, respectively.

A study by [5], conducted between 2011 and 2015, reported lower diameter PAI values on average compared to the findings of [3], likely due to the prolonged drought period during that interval. The species Cenostigma bracteosum exhibited the highest basal area increment values, with 0.1217 m^2 ha^{-1} $year^{-1}$ in the 2011–2013 period and 0.0937 m^2 ha^{-1} $year^{-1}$ in 2013–2015. Other species with significant growth included Mimosa ophthalmocentra and Cnidoscolus quercifolius, which had basal area PAI values of 0.0871 m^2 ha^{-1} $year^{-1}$ and 0.0766 m^2 ha^{-1} $year^{-1}$, respectively.

Some species faced significant regeneration challenges. Combretum glaucocarpum, for instance, recorded a negative basal area PAI of -0.0010 m^2 ha^{-1} $year^{-1}$ between 2013 and 2015. Additionally, between 2011 and 2013, 318 dead individuals and 611 lost stems were recorded, resulting in a mortality rate of 13.51% for individuals and 10.26% for stems. Between 2013 and 2015, mortality rates decreased to 222 individuals and 501 stems, with mortality rates of 10.71% and 8.46%, respectively.

These results align with the present study, indicating that during periods of greater water scarcity, species exhibited reduced growth, whereas higher precipitation periods led to increased growth rates.

3.2. Random Forest Model Performance

The statistical analysis of the variables (Table 2) considered in this study included mean, median, minimum and maximum values, standard deviation (SD), and coefficient of variation (CV%). These metrics were calculated for the Periodic Annual Increment (PAI), vegetation indices, and spectral bands obtained from Landsat 5 and Landsat 8 satellite images.

Table 2. Descriptive statistics of the analyzed variables for the study area.

Variable	Mean	Median	Min	Max	SD	CV (%)
PAI	4.30	3.94	0.58	16.57	2.59	60.27
Blue	0.06	0.05	0.02	0.09	0.02	32.94
Green	0.09	0.09	0.04	0.14	0.02	25.30
Red	0.12	0.13	0.03	0.20	0.04	36.11
NIR	0.27	0.26	0.19	0.38	0.03	12.89
SWIR1	0.33	0.34	0.18	0.45	0.07	21.48
SWIR2	0.23	0.24	0.08	0.36	0.07	31.81

 doi:10.20944/preprints202505.1655.v1

NDVI	0.38	0.29	0.17	0.82	0.19	50.36
EVI	0.24	0.17	0.10	0.56	0.12	51.61
SAVI	0.25	0.19	0.11	0.54	0.12	46.55
GNDVI	0.49	0.43	0.31	0.76	0.12	24.51
NDWI	-0.10	-0.16	-0.29	0.31	0.15	-160.12
NDBI	0.10	0.16	-0.31	0.29	0.15	160.12
MSI	1.27	1.38	0.53	1.83	0.35	27.33
NBR	0.10	0.03	-0.19	0.62	0.21	214.83
NDSI	-0.57	-0.56	-0.64	-0.51	0.02	-4.36
LAI	0.95	0.93	0.67	1.35	0.13	13.31

The Periodic Annual Increment (PAI) had an average of 4.30 $mm.ha^{-1}$ and a median of 3.94 $mm.ha^{-1}$, with significant variation between the minimum (0.58 $mm.ha^{-1}$) and maximum (16.57 $mm.ha^{-1}$) values, reflected in the high coefficient of variation (CV) of 60.27%.

Among the spectral bands, the NIR (related to cell wall structure) and SWIR1 (associated with leaf water content) exhibited lower coefficients of variation, 12.89% and 21.48%, respectively, suggesting greater stability in reflectance values. On the other hand, indices such as NDBI and NBR displayed extremely high coefficients of variation, 160.12% and 214.83%, respectively, indicating greater variability, possibly due to the structural heterogeneity of the study area.

The EVI and SAVI indices, known for their sensitivity to vegetation, showed relatively high coefficients of variation, 51.61% and 46.55%, respectively, highlighting the spatial and temporal variability of vegetation in the analyzed area. Additionally, SAVI has been recognized as suitable for areas with low vegetation cover, as described in the study by [75]. This index proved useful for biomass modeling in semi-arid ecosystems. Furthermore, the NDWI water index exhibited a contrasting behavior, with a negative mean value (-0.10) and a high coefficient of variation (-160.12%), possibly linked to the restricted water conditions of the semi-arid region, consistent with the findings of [76]. These authors explored the seasonal dynamics of water resources in dry tropical forests, emphasizing the vegetation response to limited water availability and the challenges in detecting critical areas within semi-arid ecosystems.

The variability observed in spectral indices, such as NDVI and EVI, reinforces the importance of these indices in analyzing the structural heterogeneity and vegetation dynamics in dry ecosystems. The high coefficients of variation (CV = 50.36% for NDVI) found in this study reflect the complexity and structural heterogeneity of the Caatinga, characteristics that have also been described in remote sensing analyses of seasonal tropical forests.

NDVI is widely used to represent seasonal differences in vegetation, whereas EVI is effective in correcting soil influences, especially in heterogeneous areas [75]. This index was useful for biomass modeling in semi-arid ecosystems, which is consistent with our results, where the Caatinga, characterized by sparse vegetation and seasonal variability, exhibited high variability in these indices.

The lower coefficients of variation observed for the NIR (12.89%) and SWIR1 (21.48%) spectral bands further emphasize the importance of these variables in capturing information related to biomass and vegetation health. This behavior reflects the direct interaction of these bands with leaf structure and vegetation density, as highlighted by [43], who demonstrated the critical role of NIR,

 doi:10.20944/preprints202505.1655.v1

along with the red-edge band, in distinguishing vegetation types and understanding their interactions with soil attributes in the Brazilian semi-arid region.

Overall, the variability expressed by the coefficients of variation reflects the intrinsic heterogeneity of vegetation, environmental conditions, and spectral indices in the study area, reinforcing the complexity of monitoring dry tropical forests. These results highlight the importance of robust methods, such as the use of machine learning algorithms, for data modeling and analysis.

The selection of variables based on Spearman's correlation allowed for a significant reduction in the number of spectral variables and vegetation indices used in the predictive model for Periodic Annual Increment (PAI). Initially, 16 variables were evaluated, of which only four showed statistically significant correlations ($p < 0.05$) with PAI (Table 3).

Table 3. Significant correlations between variables and Periodic Annual Increment (PAI) for model construction.

Variable	Correlation	p-value
NIR	0.254451	0.000004
LAI	0.244708	0.000010
NDSI	0.201745	0.000281
NDVI	0.118065	0.034761

Among the selected variables, NIR (Near Infrared) exhibited the highest correlation ($r = 0.254$, $p = 0.000004$), standing out as the most strongly related variable to Periodic Annual Increment (PAI). This result aligns with the role of NIR in capturing information on biomass and vegetation health, reflecting its interaction with leaf structure and vegetation density [43]. These findings are consistent with [77], who demonstrated that Landsat 8 spectral bands, such as NIR, possess high predictive capacity in forest ecosystems due to their sensitivity to structural characteristics and vegetation health.

Additionally, [78] emphasized the importance of NIR in capturing leaf structure-related information, such as chlorophyll content and physical leaf characteristics, which are essential for distinguishing between different vegetation types. These findings support the strong correlation observed between NIR and PAI ($r = 0.254$) in the present study, reinforcing its role as a key predictor variable in modeling vegetation productivity.

The Leaf Area Index (LAI) also showed a significant correlation ($r = 0.245$, $p = 0.000010$), further reinforcing its importance in PAI modeling, as LAI is directly related to primary vegetation productivity and growth dynamics.

The Normalized Difference Snow Index (NDSI) presented a moderate correlation ($r = 0.202$, $p = 0.000281$), suggesting that the model is sensitive to soil and vegetation properties in arid environments, where reflectance conditions may be influenced by leafless vegetation. Etymologically, the word "Caatinga" (Brazilian Dry Tropical Forest) originates from the Tupi-Guarani language and means "white forest", composed of caa (forest) and tinga (white).

The NDVI (Normalized Difference Vegetation Index), widely used as an indicator of vegetation cover, exhibited a relatively low but significant correlation ($r = 0.118$, $p = 0.034761$), indicating its contribution to the model despite its weaker association with PAI. This behavior aligns with the findings of [74] who demonstrated that NDVI is effective in differentiating vegetation classes. It is also consistent with the results of [79], who highlighted NDVI's sensitivity to climatic variations in semi-arid regions of Northeastern Brazil over a 20-year period. Their study identified that NDVI captures both seasonal and interannual vegetation patterns, reflecting changes in vegetation cover in response to climatic and environmental fluctuations.

The reduction in the number of variables based on correlation criteria not only simplifies the model but also reduces potential multicollinearity issues, ensuring greater efficiency and interpretability in the predictive modeling process. These results highlight the importance of

 doi:10.20944/preprints202505.1655.v1

selecting key variables that capture the most relevant variations in forest growth dynamics, particularly in complex ecosystems such as dry tropical forests.

The vegetation indices and the NIR band used in this study demonstrated a positive correlation with PAI, confirming their utility as proxies for estimating vegetation productivity. [74] previously highlighted the robustness of using remote sensing data to capture variations in vegetation cover and vigor. The Spearman correlation analysis revealed that remote sensing-derived variables had significant relationships with PAI, supporting the findings of [10], who emphasized the role of these indicators in modeling forest processes.

The reduction in the number of variables based on correlation criteria not only simplifies the model but also reduces potential multicollinearity issues, ensuring greater efficiency and interpretability in the predictive modeling process. As emphasized by [80], Spearman's correlation coefficient is a robust tool for identifying monotonic associations between variables, especially when data do not meet normality or linearity assumptions. By replacing original values with ranks, Spearman's method allows for reliable analysis, even in cases of asymmetrical distributions, which are frequently observed in environmental data.

The initial evaluation of the Random Forest model for estimating PAI showed promising results, but also highlighted challenges related to the intrinsic variability of field data. Before applying adjustments, the model presented a coefficient of determination (R^2) of 0.5055, indicating that approximately 50.55% of the variation in actual PAI values was explained by the model, while the Root Mean Square Error (RMSE) was 1.83 $mmha^{-1}$. These results are consistent with the performance reported by [81] in arid regions, where Random Forest was identified as the most accurate algorithm for predicting environmental and soil properties in challenging scenarios with high dimensionality and correlated variables.

The study by [77] further supports the potential of Random Forest as a powerful tool for environmental modeling, even when compared to more recent algorithms such as XGBoost. The ability of Random Forest to randomly sample variables and observations enhances its resistance to multicollinearity, enabling more effective modeling in complex ecosystems.

An analysis of the estimated values revealed that the model tended to underestimate higher PAI values and overestimate lower ones, as evidenced by an average percentage difference of 0.29% and a total absolute difference of 4.00 $mmha^{-1}$. This bias may be attributed to the structural complexity of the dry tropical forest and the model's limitations in fully capturing the variability of spectral and vegetation index data.

The initial challenges of the model included the underestimation of high PAI values and overestimation of low PAI values, a behavior also observed by [77] in forest biomass estimations using Random Forest. This limitation is attributed to the algorithm's difficulty in handling extreme data points without additional adjustments. Nonetheless, the high Spearman correlation (0.846) demonstrated that the model effectively captured general data trends, even though it lacked precision in extreme values.

The application of Random Forest for drought monitoring in grasslands has highlighted that adjustments based on error distribution can minimize discrepancies and improve overall model performance, a strategy that has also been implemented in the present study [82].

Given this limitation, a correction factor was introduced, based on the distribution of estimation errors. The adjustment involved dividing the data into six classes, using PAI value intervals, and computing average correction factors for each class. These factors were then applied to adjust the model's estimated values, resulting in a substantial improvement in prediction accuracy (Table 4).

Table 4. Correction factor for model bias in estimating Periodic Annual Increment (PAI) ($mmha^{-1}$).

Class	Intervals ($mmha^{-1}$)	Correction Factor (FC)
1	0.5835 – 1.8916	-0.8685
2	1.8916 – 2.9572	-0.6340

 doi:10.20944/preprints202505.1655.v1

3	2.9572 – 3.9390	-0.4750
4	3.9390 – 4.8921	-0.1901
5	4.8921 – 6.4431	0.2911
6	6.4431 – 16.5665	1.5334

After applying the correction factor, the model's results were significantly improved. The R^2 increased to 0.8867, indicating that 88.67% of the variation in the actual PAI values is now explained by the corrected model. This substantial increase in R^2 reflects a better fit between the estimated and real values. Simultaneously, the RMSE was reduced to 0.87 $mm.ha^{-1}$, demonstrating a marked reduction in the average prediction error (Figure 7).

This outcome aligns with the findings of [83] and [84], who underscored the importance of post-modeling adjustments in reducing discrepancies in heterogeneous systems. Moreover, the use of correction factors reinforces the applicability of Random Forest in complex scenarios, such as dry tropical forests. [75] highlighted the algorithm's capacity to capture complex patterns and handle nonlinear data, demonstrating its relevance for robust estimates in heterogeneous ecosystems.

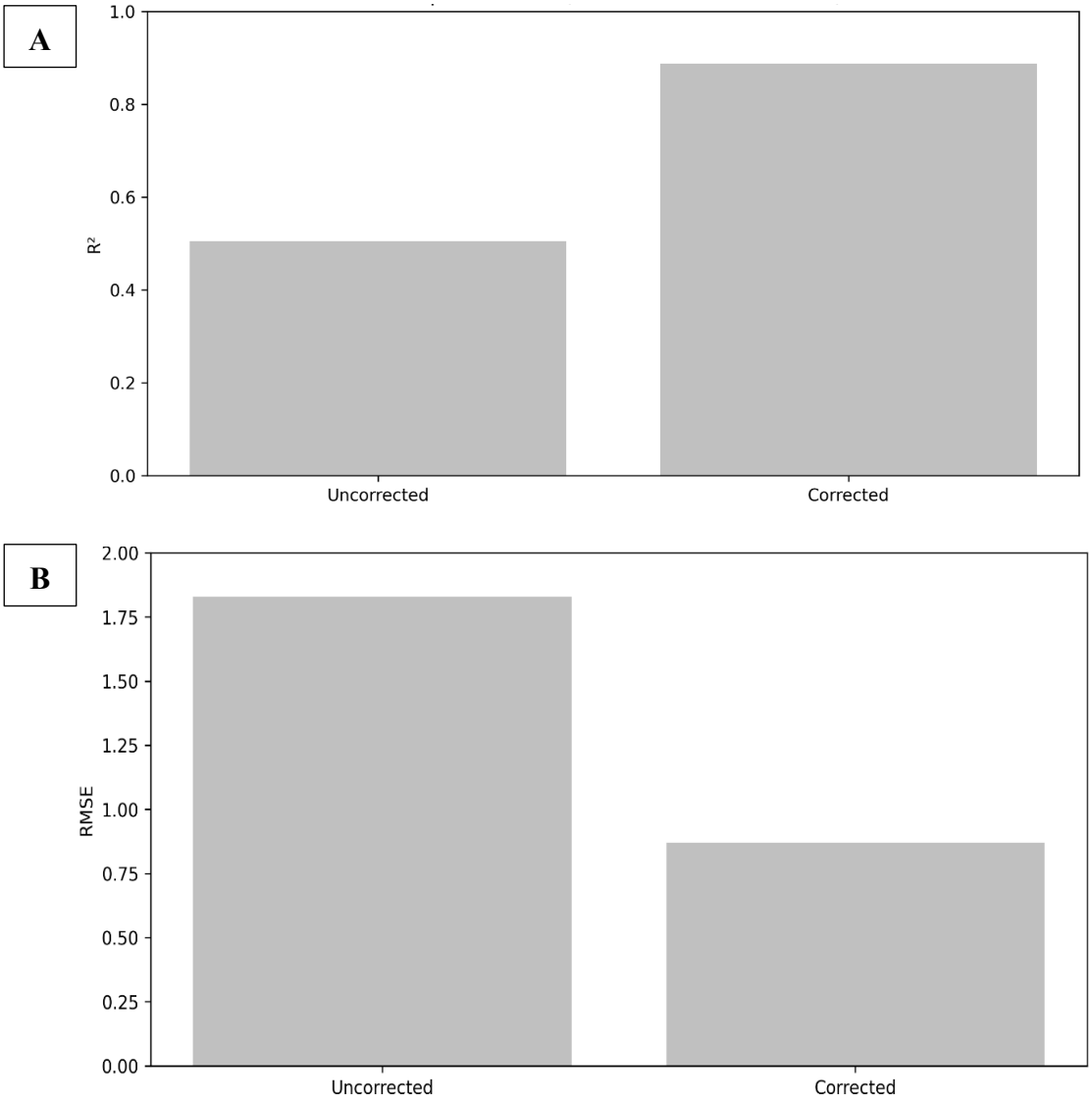

Figure 7. Comparison of the Random Forest model's performance before and after the application of the correction factor. (A) Coefficient of determination (R^2) for uncorrected and corrected models. (B) Root Mean Square Error (RMSE) for uncorrected and corrected models.

 doi:10.20944/preprints202505.1655.v1

Furthermore, the Spearman correlation increased to 0.946 after applying the correction factor, indicating a more robust relationship between the actual and estimated values. This improvement not only reflects enhanced metrics such as R^2 and RMSE but also demonstrates greater consistency in the ranking trends of the corrected data, which is crucial for monitoring and management studies. The high post-correction correlation shows that the adjusted model is capable of capturing the patterns of forest growth with higher fidelity, even in an environment characterized by high variability.

Figure 8 presents detailed comparisons between the actual and estimated values of the Periodic Annual Increment (PAI) before and after the Random Forest model correction. Density analyses reveal improved alignment between the estimated and actual values after applying the correction factor. The distribution peaks become closer, indicating a significant reduction in the discrepancy between the estimated maximum and minimum values.

Scatter plots further illustrate the relationship between the actual and estimated values before and after correction. It is evident that, following the application of the correction factor, the values align more consistently, signifying a substantial improvement in the model's predictive accuracy.

These results underscore the effectiveness of the correction factor in adjusting discrepancies—especially in extreme values—thereby enhancing the model's ability to capture the variations in forest growth within the study area.

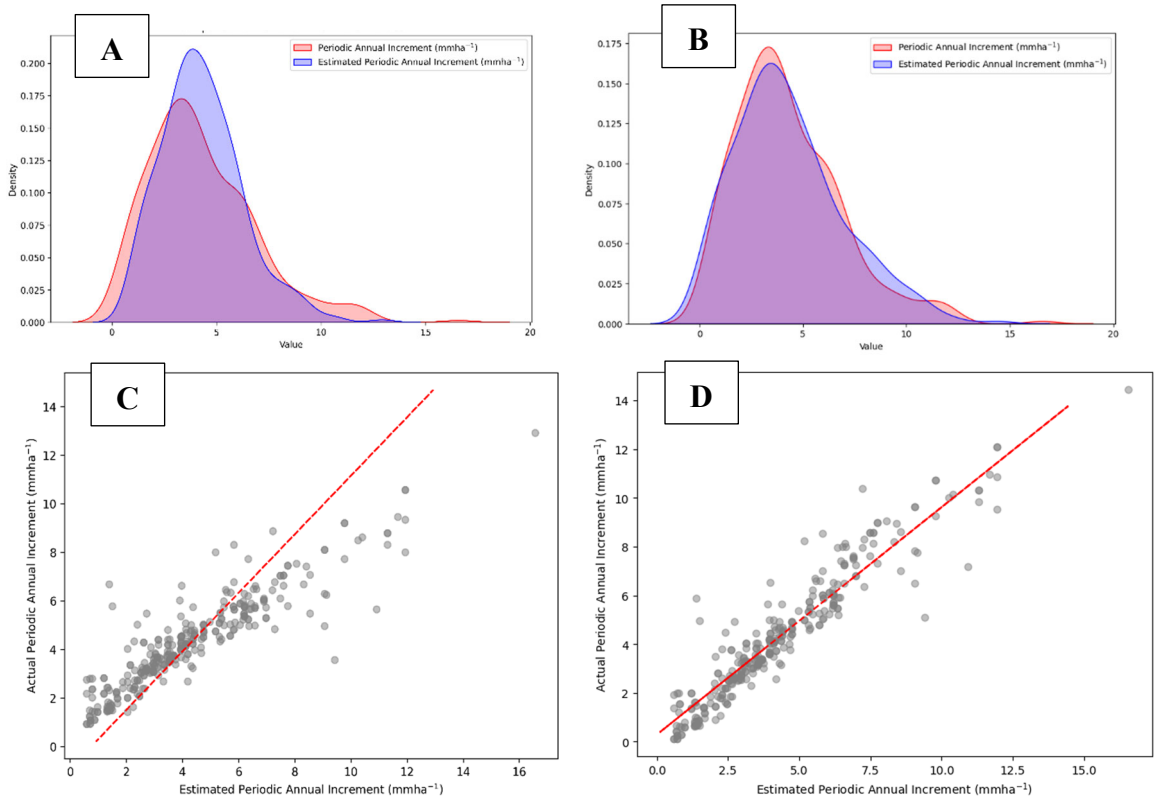

Figure 8. Comparison of the Random Forest model's performance before and after the application of the correction factor. (A) Density plot of observed vs. estimated PAI values before correction. (B) Density plot of observed vs. estimated PAI values after correction. (C) Scatter plot comparing observed and estimated PAI before correction. (D) Scatter plot comparing observed and estimated PAI after correction.

Post-modeling adjustments are essential for addressing discrepancies in predictions for complex systems such as dry tropical forests. The application of correction factors allows estimates to be aligned with observed data, particularly in cases where data variability challenges traditional modeling approaches [85]. The corrected model demonstrated significantly improved performance,

 doi:10.20944/preprints202505.1655.v1

underscoring its potential as a valuable tool for the management and monitoring of dry tropical forests. Furthermore, the strong correlation between precipitation and PAI (r = 0.946) suggests that environmental factors, such as water availability, play a crucial role in forest growth and should be incorporated into future studies to refine predictive estimates.

Table 5 shows that there is a positive relationship between precipitation and PAI, with periods of higher precipitation being associated with higher mean PAI values. This indicates that precipitation may be an important factor in tree growth in the Caatinga, a biome sensitive to moisture variations. Similar results were found by [86], who demonstrated the significant influence of precipitation on the seasonality of vegetation in the Caatinga biome, reinforcing the biome's sensitivity to water availability [87].

Table 5. Precipitation and PAI values for the studied periods.

Period	Mean Precipitation (mm)	Mean PAI (mm)
2011-2013	286.2	4.04
2013-2015	245.9	3.73
2015-2017	372.5	4.23
2017-2019	461.7	6.14

The values between the mean precipitation and the mean Periodic Annual Increment (PAI) exhibited a correlation of approximately 0.92. This indicates a strong positive correlation, suggesting that in this dataset, periods with higher precipitation tend to be associated with higher tree growth rates. According to recent studies, precipitation plays a crucial role in the population dynamics of Caatinga vegetation. [88] found that the growth of woody species in the Brazilian semi-arid region is directly influenced by water availability, especially in years with above-average precipitation patterns.

Furthermore, research conducted by [89] highlighted that the Leaf Area Index and biomass of the Caatinga are strongly dependent on precipitation patterns, demonstrating that water availability is a limiting factor for vegetation development. These findings align with the observations in Table 5, which reinforce the importance of precipitation as one of the main determining factors for the Periodic Annual Increment of trees in the Caatinga.

Periods with below-average precipitation, such as those observed in 2012 and 2014, tend to result in lower mean PAI values, as previously shown. Conversely, years with greater water availability, such as 2019, coincide with higher forest growth rates, reinforcing the strong positive correlation between precipitation and PAI ($\varrho = 0.92$).

The combination of highly variable precipitation and elevated temperatures creates a challenging environment for forest growth in the Caatinga biome. These factors underscore the need to consider not only water availability but also the role of extreme temperatures in adjusting and refining predictive models of PAI. Additionally, these climatic data provide a basis for exploring adaptive management strategies that can mitigate the effects of climate change on the growth of tree species in this ecosystem.

The Periodic Annual Increment (PAI), a widely used metric for evaluating forest growth, revealed significant variations over the analyzed years. PAI values were lower during dry periods and higher in years with above-average precipitation, as observed in 2019. This pattern is consistent with the physiological response mechanisms of species to water deficits, as discussed by [7] and [73]. Mapping the PAI revealed marked spatial variations, reflecting the interactions between climatic and edaphic factors, as also observed by [90] in a similar study.

 doi:10.20944/preprints202505.1655.v1

The positive relationship between precipitation and PAI, with a correlation of 0.92, reinforces the importance of water availability for tree growth in the Caatinga, as evidenced by studies from [91] and [92]. This result highlights the need for continuous monitoring of climatic conditions and their incorporation into forest management models to mitigate the impacts of climate change.

The analysis of these climatic factors demonstrates the relevance of using robust models—such as the corrected Random Forest—to estimate PAI, incorporating environmental variables and their interactions. The variability in precipitation and the persistence of high temperatures underscore the importance of continuously monitoring climatic conditions in the sustainable management of the Caatinga.

The spatial distribution maps of the actual and estimated PAI, along with the uncertainty map, provided valuable insights into the forest dynamics of the region (Figure 9). It was observed that areas with higher species diversity and closer proximity to recharge sources exhibited the highest increments, while plots located in more arid areas showed the largest estimation errors. This pattern highlights the importance of integrating climatic and edaphic data to refine predictions, as suggested by [93]. In the error map, the greatest discrepancies were associated with specific plots, which may be related to species diversity and local growth dynamics.

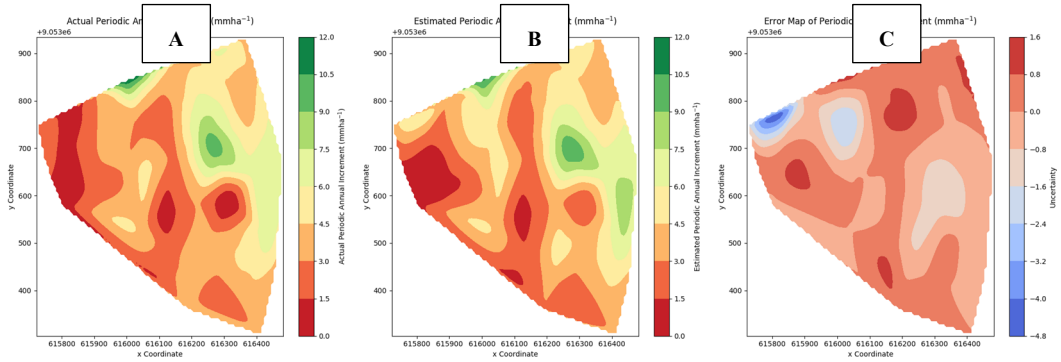

Figure 9. (A) Spatial distribution map for the actual PAI (mm ha^{-1}); (B) Spatial distribution map for the estimated PAI (mm ha^{-1}); (C) Error map of the PAI distribution (mm ha^{-1}) for the study area.

4. Conclusions

The estimates generated by the Random Forest model showed high agreement with field data, demonstrating its efficacy in predicting the Periodic Annual Increment (PAI) in dry tropical forests. The application of a correction factor to the model was decisive in improving the precision of the estimates, resulting in a significant increase in the coefficient of determination (R^2) and a reduction in the root mean square error (RMSE). This adjustment underscores the importance of integrating post-processing methods in forest modeling studies, particularly in ecosystems with high spatial and temporal variability.

This work demonstrates that the integration of remote sensing, machine learning, and field data is an effective approach for understanding the forest dynamics of the Caatinga. In addition to contributing to scientific advancement in the field, the results presented provide important support for the planning and sustainable management of this biome in the context of climate change.

Despite the promising results, the study presents limitations that should be acknowledged. The relatively small number of sampling plots (n = 40) may limit the generalization of the model to broader spatial scales. In addition, the model was calibrated using spectral and climatic conditions specific to the study area, which may differ across other regions or under varying seasonal contexts. Future research should consider integrating additional environmental variables (e.g., soil texture, land use history) and expanding the spatial and temporal extent of data collection to improve model accuracy and applicability in other ecosystems.

Author Contributions: Conceptualization, A.P.F. and R.L.C.F.; methodology, A.P.F., E.A.S., V.L. and F.M.B.; software, A.P.F.; validation, A.P.F., R.L.C.F., and M.D.L.; formal analysis, A.P.F. and E.A.S.; investigation, A.P.F.;

 doi:10.20944/preprints202505.1655.v1

resources, P.C.B.; data curation, A.P.F.; writing—original draft preparation, A.P.F.; writing—review and editing, A.P.F., E.A.S., V.L., P.C.B., and S.S.N.; visualization, A.P.F.; supervision, R.L.C.F., M.D.L., and J.A.A.S.; project administration, S.S.N.; funding acquisition, P.C.B. All authors have read and agreed to the published version of the manuscript.

Funding: Funding: This research was funded by University Open Access Fund (University of Manchester). The present work was carried out with support from Fundação de Amparo à Ciência e Tecnologia de Pernambuco (FACEPE)—Brazil - Pernambuco—Financial code IBPG-0318-5.02/19, and National Council for Scientific and Technological Development (CNPq) (grant numbers 305452/2023-1 and 317538/2021-7).

Data Availability Statement: The data presented in this study are not publicly available at this time but will be made available upon reasonable request after the publication of the manuscript.

Acknowledgments: The authors would like to thank the Universidade Federal Rural de Pernambuco (UFRPE) for supporting this research. We are also grateful to the Pipipã Indigenous Community for their valuable assistance during fieldwork. This work was supported by FACEPE through a research scholarship. We also acknowledge the University of Manchester Open Access Fund for supporting the publication of this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Perrando, E.R.; Breunig, F.M.; Galvão, L.S.; Bostelmann, S.L.; Martarello, V.; Straube, J.; Conte, B.; Sestari, G.; Burgin, M.R.B.; De Marco, R. Evaluation of the Effects of Pine Management on the Water Yield and Quality in Southern Brazil. *Journal of Sustainable Forestry* **2021**, *40*, 217–233, doi:10.1080/10549811.2020.1746916.
2. Kumar Mishra, R.; Agarwal, R. Sustainable Forest Land Management to Restore Degraded Lands. **2024**, doi:10.5772/INTECHOPEN.1004793.
3. DA SILVA, A.F. ESTRUTURA E DINÂMICA RELACIONADAS A FATORES AMBIENTAIS EM FLORESTA SECA SUBMETIDA AO MANEJO FLORESTAL , Universidade Federal Rural de Pernambuco: Recife, 2023.
4. Salazar, A.A.; Arellano, E.C.; Muñoz-sáez, A.; Miranda, M.D.; da Silva, F.O.; Zielonka, N.B.; Crowther, L.P.; Silva-ferreira, V.; Oliveira-reboucas, P.; Dicks, L. V. Restoration and Conservation of Priority Areas of Caatinga's Semi-Arid Forest Remnants Can Support Connectivity within an Agricultural Landscape. *Land 2021, Vol. 10, Page 550* **2021**, *10*, 550, doi:10.3390/LAND10060550.
5. Melo, C.L.S.M.S.; F.R.L.C.; S.J.A.A.; M.M.Á.H.; C.G.H.G. Dynamics of Dry Tropical Forest after Three Decades of Vegetation Suppression. *Floresta e Ambiente* **2019**, *26*.
6. DOS SANTOS, N.Á.T. DINÂMICA DA FLORESTA SECA SOB DIFERENTES HISTÓRICOS DE USO: DISTRIBUIÇÃO DIAMÉTRICA DE INDIVÍDUOS E FUSTES, Universidade Federal Rural de Pernambuco: Recife, 2021.
7. Dalla Lana, M.; Ferreira, R.L.C.; da Silva, J.A.A.; Duda, G.P.; Cespedes, G.H.G. Carbon Content in Shrub-Tree Species of the Caatinga. *Floresta e Ambiente* **2019**, *26*, e20170617, doi:10.1590/2179-8087.061717.
8. Abreu, J.C.; S.J.A.A.; F.R.L.C.; R.S.J.S.S.; T.J.I.S.; F.A.A.; V.P.H.; V.A.B.T.; S.B.L.S.; T.L.A.A.; S.A.A. Mixed Models for Nutrients Prediction in Species of the Brazilian Caatinga Biome. *Árvore* **2023**, *47*.
9. Salami, G.; Ferreira, R.L.C.; Silva, J.A.A. da; Freire, F.J.; Silva, E.A.; Diodato, M.A.; Pessi, D.D. Distribuição Espaço-Temporal Da Biomassa e Carbono Florestal Acima Do Solo Em Floresta Tropical Seca Sob Pressão Antrópica, Brasil. *Terr@ Plural* **2024**, *18*, 1–27, doi:10.5212/TERRAPLURAL.V.18.2423225.014.
10. MOREIRA, G.L. Modelagem de Variáveis Biofísicas Em Floresta Tropical Seca Por Meio de Geotecnologias. **2021**.
11. de Oliveira, C.P.; Ferreira, R.L.C.; da Silva, J.A.A.; de Lima, R.B.; Silva, E.A.; da Silva, A.F.; de Lucena, J.D.S.; Dos Santos, N.A.T.; Lopes, I.J.C.; Pessoa, M.M. de L.; et al. Modeling and Spatialization of Biomass and Carbon Stock Using LiDAR Metrics in Tropical Dry Forest, Brazil. *Forests 2021, Vol. 12, Page 473* **2021**, *12*, 473, doi:10.3390/F12040473.
12. Beck, H.E.; McVicar, T.R.; Vergopolan, N.; Berg, A.; Lutsko, N.J.; Dufour, A.; Zeng, Z.; Jiang, X.; van Dijk, A.I.J.M.; Miralles, D.G. High-Resolution (1 Km) Köppen-Geiger Maps for 1901–2099 Based on Constrained CMIP6 Projections. *Scientific Data 2023 10:1* **2023**, *10*, 1–16, doi:10.1038/s41597-023-02549-6.

 doi:10.20944/preprints202505.1655.v1

13. Ottoni, F.P.; Filgueira, C.T.S.; Lima, B.N.; Vieira, L.O.; Rangel-Pereira, F.; Oliveira, R.F. Extreme Drought Threatens the Amazon. *Science (1979)* **2023**, *382*, 1253–1255, doi:10.1126/SCIENCE.ADM8147/ASSET/DBE5B50D-810B-46C6-BFD4-8A79932D9820/ASSETS/SCIENCE.ADM8147.FP.PNG.
14. AR6 Synthesis Report: Climate Change 2023. Available online: https://www.ipcc.ch/report/ar6/syr/ (accessed on 20 March 2025).
15. Santos, J.C.; Leal, I.R.; Almeida-Cortez, J.S.; Fernandes, G.W.; Tabarelli, M. Caatinga: The Scientific Negligence Experienced by a Dry Tropical Forest. *Trop Conserv Sci* **2011**, *4*, 276–286, doi:10.1177/194008291100400306/ASSET/383E751D-291C-4B8F-882B-1BB60E311D7C/ASSETS/IMAGES/LARGE/10.1177_194008291100400306-FIG4.JPG.
16. Leal, I.R.; Da Silva, J.M.C.; Tabarelli, M.; Lacher, T.E. Changing the Course of Biodiversity Conservation in the Caatinga of Northeastern BrazilCambiando El Curso de La Conservación de Biodiversidad En La Caatinga Del Noreste de Brasil. *Conservation Biology* **2005**, *19*, 701–706, doi:10.1111/J.1523-1739.2005.00703.X.
17. Pitman, N.C.A.; Silman, M.R.; Terborgh, J.W. Oligarchies in Amazonian Tree Communities: A Ten-Year Review. *Ecography* **2013**, *36*, 114–123, doi:10.1111/J.1600-0587.2012.00083.X.
18. Ter Steege, H.; Pitman, N.C.A.; Sabatier, D.; Baraloto, C.; Salomão, R.P.; Guevara, J.E.; Phillips, O.L.; Castilho, C. V.; Magnusson, W.E.; Molino, J.F.; et al. Hyperdominance in the Amazonian Tree Flora. *Science (1979)* **2013**, *342*, doi:10.1126/SCIENCE.1243092/SUPPL_FILE/TERSTEEGE_APPENDIX.XLSX.
19. Bonal, D.; Burban, B.; Stahl, C.; Wagner, F.; Hérault, B. The Response of Tropical Rainforests to Drought—Lessons from Recent Research and Future Prospects. *Ann For Sci* **2016**, *73*, 27–44, doi:10.1007/S13595-015-0522-5/FIGURES/4.
20. Elias, F.; Marimon, B.S.; Marimon-Junior, B.H.; Budke, J.C.; Esquivel-Muelbert, A.; Morandi, P.S.; Reis, S.M.; Phillips, O.L. Idiosyncratic Soil-Tree Species Associations and Their Relationships with Drought in a Monodominant Amazon Forest. *Acta Oecologica* **2018**, *91*, 127–136, doi:10.1016/J.ACTAO.2018.07.004.
21. Lloyd, J.; Domingues, T.F.; Schrodt, F.; Ishida, F.Y.; Feldpausch, T.R.; Saiz, G.; Quesada, C.A.; Schwarz, M.; Torello-Raventos, M.; Gilpin, M.; et al. Edaphic, Structural and Physiological Contrasts across Amazon Basin Forest-Savanna Ecotones Suggest a Role for Potassium as a Key Modulator of Tropical Woody Vegetation Structure and Function. *Biogeosciences* **2015**, *12*, 6529–6571, doi:10.5194/BG-12-6529-2015.
22. Powers, J.S.; F.X.; S.-A.A.; M.D. Ter Steege, H.; Pitman, N. C. A.; Sabatier, D.; Baraloto, C.; Salomão, R. P. Et Al. Hyperdominance in the Amazonian Tree Flora. Science, 2013, v. 342, n. 6156, p. 325-+. . *Environmental Research Letters* **2018**, *13*.
23. Dickinson, M.B.; Hermann, S.M.; Whigham, D.F. Low Rates of Background Canopy-Gap Disturbance in a Seasonally Dry Forest in the Yucatan Peninsula with a History of Fires and Hurricanes. *J Trop Ecol* **2001**, *17*, 895–902, doi:10.1017/S0266467401001663.
24. Santos, J.R. dos; Narvaes, I.S.; Graca, P.M.L.A.; Goncalves, F.G.; Santos, J.R. dos; Narvaes, I.S.; Graca, P.M.L.A.; Goncalves, F.G. Polarimetric Responses and Scattering Mechanisms of Tropical Forests in the Brazilian Amazon. *Advances in Geoscience and Remote Sensing* **2009**, doi:10.5772/8340.
25. da Conceição Bispo, P.; Picoli, M.C.A.; Marimon, B.S.; Marimon Junior, B.H.; Peres, C.A.; Menor, I.O.; Silva, D.E.; de Figueiredo Machado, F.; Alencar, A.A.C.; de Almeida, C.A.; et al. Overlooking Vegetation Loss Outside Forests Imperils the Brazilian Cerrado and Other Non-Forest Biomes. *Nature Ecology & Evolution 2023 8:1* **2023**, *8*, 12–13, doi:10.1038/s41559-023-02256-w.
26. Chave, J.; Réjou-Méchain, M.; Búrquez, A.; Chidumayo, E.; Colgan, M.S.; Delitti, W.B.C.; Duque, A.; Eid, T.; Fearnside, P.M.; Goodman, R.C.; et al. Improved Allometric Models to Estimate the Aboveground Biomass of Tropical Trees. *Glob Chang Biol* **2014**, *20*, 3177–3190, doi:10.1111/GCB.12629.
27. Rosan, T.M.; Aragão, L.E.O.C.; Oliveras, I.; Phillips, O.L.; Malhi, Y.; Gloor, E.; Wagner, F.H. Extensive 21st-Century Woody Encroachment in South America's Savanna. *Geophys Res Lett* **2019**, *46*, 6594–6603, doi:10.1029/2019GL082327.
28. Raymundo, D.; Oliveira-Neto, N.E.; Martini, V.; Araújo, T.N.; Calaça, D.; de Oliveira, D.C. Assessing Woody Plant Encroachment by Comparing Adult and Juvenile Tree Components in a Brazilian Savanna. *Flora* **2022**, *291*, 152060, doi:10.1016/j.flora.2022.152060.

 doi:10.20944/preprints202505.1655.v1

29. Bilar, A.B.C.; Pimentel, R.M. de M.; Cerqueira, M.A. Monitoramento Da Cobertura Vegetal Através de Índices Biofísicos e Gestão de Áreas Protegidas. *Geosul* **2018**, *33*, 236–259, doi:10.5007/2177-5230.2018V33N68P236.
30. García-Haro, F.J.; Campos-Taberner, M.; Muñoz-Marí, J.; Laparra, V.; Camacho, F.; Sánchez-Zapero, J.; Camps-Valls, G. Derivation of Global Vegetation Biophysical Parameters from EUMETSAT Polar System. *ISPRS Journal of Photogrammetry and Remote Sensing* **2018**, *139*, 57–74, doi:10.1016/J.ISPRSJPRS.2018.03.005.
31. Piles, M.; Camps-Valls, G.; Chaparro, D.; Entekhabi, D.; Konings, A.G.; Jagdhuber, T. Remote Sensing of Vegetation Dynamics in Agro-Ecosystems Using Smap Vegetation Optical Depth and Optical Vegetation Indices. *International Geoscience and Remote Sensing Symposium (IGARSS)* **2017**, *2017-July*, 4346–4349, doi:10.1109/IGARSS.2017.8127964.
32. Verrelst, J.; Malenovský, Z.; Van der Tol, C.; Camps-Valls, G.; Gastellu-Etchegorry, J.P.; Lewis, P.; North, P.; Moreno, J. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. *Surv Geophys* **2019**, *40*, 589–629, doi:10.1007/S10712-018-9478-Y/FIGURES/9.
33. Gutiérrez-Castorena, E.V.; Silva-Núñez, J.A.; Gaytán-Martínez, F.D.; Encinia-Uribe, V.V.; Ramírez-Gómez, G.A.; Olivares-Sáenz, E. Evaluation of Statistical Models of NDVI and Agronomic Variables in a Protected Agriculture System. *Horticulturae 2025, Vol. 11, Page 131* **2025**, *11*, 131, doi:10.3390/HORTICULTURAE11020131.
34. Albuquerque, R.W.; Ferreira, M.E.; Olsen, S.I.; Tymus, J.R.C.; Balieiro, C.P.; Mansur, H.; Moura, C.J.R.; Costa, J.V.S.; Branco, M.R.C.; Grohmann, C.H. Forest Restoration Monitoring Protocol with a Low-Cost Remotely Piloted Aircraft: Lessons Learned from a Case Study in the Brazilian Atlantic Forest. *Remote Sensing 2021, Vol. 13, Page 2401* **2021**, *13*, 2401, doi:10.3390/RS13122401.
35. John, E.; Bunting, P.; Hardy, A.; Silayo, D.S.; Masunga, E. A Forest Monitoring System for Tanzania. *Remote Sensing 2021, Vol. 13, Page 3081* **2021**, *13*, 3081, doi:10.3390/RS13163081.
36. Wulder, M.A.; White, J.C.; Masek, J.G.; Dwyer, J.; Roy, D.P. Continuity of Landsat Observations: Short Term Considerations. *Remote Sens Environ* **2011**, *115*, 747–751, doi:10.1016/J.RSE.2010.11.002.
37. Asner, G.P.; Powell, G.V.N.; Mascaro, J.; Knapp, D.E.; Clark, J.K.; Jacobson, J.; Kennedy-Bowdoin, T.; Balaji, A.; Paez-Acosta, G.; Victoria, E.; et al. High-Resolution Forest Carbon Stocks and Emissions in the Amazon. *Proc Natl Acad Sci U S A* **2010**, *107*, 16738–16742, doi:10.1073/PNAS.1004875107/SUPPL_FILE/PNAS.201004875SI.PDF.
38. Saatchi, S.S.; Harris, N.L.; Brown, S.; Lefsky, M.; Mitchard, E.T.A.; Salas, W.; Zutta, B.R.; Buermann, W.; Lewis, S.L.; Hagen, S.; et al. Benchmark Map of Forest Carbon Stocks in Tropical Regions across Three Continents. *Proc Natl Acad Sci U S A* **2011**, *108*, 9899–9904, doi:10.1073/PNAS.1019576108/SUPPL_FILE/ST03.DOC.
39. Mitchard, E.T.A.; Flintrop, C.M. Woody Encroachment and Forest Degradation in Sub-Saharan Africa's Woodlands and Savannas 1982–2006. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2013**, *368*, doi:10.1098/RSTB.2012.0406.
40. Berra, E.F.; Fontana, D.C.; Yin, F.; Breunig, F.M. Harmonized Landsat and Sentinel-2 Data with Google Earth Engine. *Remote Sensing 2024, Vol. 16, Page 2695* **2024**, *16*, 2695, doi:10.3390/RS16152695.
41. Claverie, M.; Ju, J.; Masek, J.G.; Dungan, J.L.; Vermote, E.F.; Roger, J.C.; Skakun, S. V.; Justice, C. The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Set. *Remote Sens Environ* **2018**, *219*, 145–161, doi:10.1016/J.RSE.2018.09.002.
42. Breiman, L. Random Forests. *Mach Learn* **2001**, *45*, 5–32, doi:10.1023/A:1010933404324/METRICS.
43. da Silveira, H.L.F.; Galvão, L.S.; Sanches, I.D.A.; de Sá, I.B.; Taura, T.A. Use of MSI/Sentinel-2 and Airborne LiDAR Data for Mapping Vegetation and Studying the Relationships with Soil Attributes in the Brazilian Semi-Arid Region. *International Journal of Applied Earth Observation and Geoinformation* **2018**, *73*, 179–190, doi:10.1016/J.JAG.2018.06.016.
44. Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning. **2009**, doi:10.1007/978-0-387-84858-7.
45. Goodfellow, I.; B.Y.; C.A. *Deep Learning*; 2016;
46. Zhang, J.; Yin, K. Application of Gradient Boosting Model to Forecast Corporate Green Innovation Performance. *Front Environ Sci* **2023**, *11*, 1252271, doi:10.3389/FENVS.2023.1252271/BIBTEX.

 doi:10.20944/preprints202505.1655.v1

47. Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover Classification. *ISPRS Journal of Photogrammetry and Remote Sensing* **2012**, *67*, 93–104, doi:10.1016/J.ISPRSJPRS.2011.11.002.
48. Powell, S.L.; Cohen, W.B.; Healey, S.P.; Kennedy, R.E.; Moisen, G.G.; Pierce, K.B.; Ohmann, J.L. Quantification of Live Aboveground Forest Biomass Dynamics with Landsat Time-Series and Field Inventory Data: A Comparison of Empirical Modeling Approaches. *Remote Sens Environ* **2010**, *114*, 1053–1068, doi:10.1016/J.RSE.2009.12.018.
49. Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. RANDOM FORESTS FOR CLASSIFICATION IN ECOLOGY. *Ecology* **2007**, *88*, 2783–2792, doi:10.1890/07-0539.1.
50. Liaw, A.; Wiener, M. Classification and Regression by RandomForest. **2002**, *2*.
51. Machado, G.; Mendoza, M.R.; Corbellini, L.G. What Variables Are Important in Predicting Bovine Viral Diarrhea Virus? A Random Forest Approach. *Vet Res* **2015**, *46*, 1–15, doi:10.1186/S13567-015-0219-7/FIGURES/7.
52. Ferreira, R.L.C.D.S.J.A.A.; A.J.F.T.; D.L.R.B.; D.L.M. Components of Growth for Tropical Dry Deciduous Forest, Brazil. In Proceedings of the In: ASA, CSSA & SSSA INTERNATIONAL ANNUAL MEETING, Long Beach.; Long Beach: , 2014.
53. Christina, R.; Silva, S.; Luiz, R.; Ferreira, C.; Antônio, J.; Da Silva, A.; Maria, I.; Meunier, J.; Berger, R. Aspectos Fitossociológicos e de Crescimento de Commiphora Leptophloeos No Semiárido Brasileiro. *Pesqui Florest Bras* **2017**, *37*, 11–18, doi:10.4336/2017.pfb.37.89.1224.
54. PIMENTEL, D.J.O. Dinâmica Da Vegetação Lenhosa Em Área de Caatinga, Floresta – PE. **2012**.
55. Silva, G.F.; Mendonça, A.R.; Dias, A.N.; Nogueira, G.S.; Silva, J.A.A.; Oliveira, M.L.R.; Ferreira, R.L.C. *Padronização Da Simbologia Em Mensuração e Manejo Florestal.*; Alegre - ES, 2022;
56. Zhengming Wan MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1 Km SIN Grid Available online: https://lpdaac.usgs.gov/products/mod11a1v061/ (accessed on 21 March 2025).
57. Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. The Climate Hazards Infrared Precipitation with Stations—a New Environmental Record for Monitoring Extremes. *Scientific Data 2015 2:1* **2015**, *2*, 1–21, doi:10.1038/sdata.2015.66.
58. Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Müller, A.C.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; et al. API Design for Machine Learning Software: Experiences from the Scikit-Learn Project. **2013**.
59. Pedregosa FABIANPEDREGOSA, F.; Michel, V.; Grisel OLIVIERGRISEL, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Vanderplas, J.; Cournapeau, D.; Pedregosa, F.; Varoquaux, G.; et al. Scikit-Learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
60. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the scipy; SciPy, 2010; pp. 56–61.
61. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods 2020 17:3* **2020**, *17*, 261–272, doi:10.1038/s41592-019-0686-2.
62. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng* **2007**, *9*, 90–95, doi:10.1109/MCSE.2007.55.
63. Waskom, M.L. Seaborn: Statistical Data Visualization. *J Open Source Softw* **2021**, *6*, 3021, doi:10.21105/JOSS.03021.
64. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362, doi:10.1038/s41586-020-2649-2.
65. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. Bias-Variance Trade-Off for k-Fold Cross-Validation. *An Introduction to Statistical Learning - with Applications in R* **2013**, *7*, 24.
66. Spearman, C. The Proof and Measurement of Association between Two Things. *Am J Psychol* **1904**, *15*, 101, doi:10.2307/1412159.

 doi:10.20944/preprints202505.1655.v1

67. Cort J. Willmott; Kenji Matsuura Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. . *Clim Res* **2005**, *30*, 79–82.
68. Shepard, D. A Two-Dimensional Interpolation Function for Irregularly-Spaced Data. *Proceedings of the 1968 23rd ACM National Conference, ACM 1968* **1968**, 517–524, doi:10.1145/800186.810616.
69. Jose A. Marengo; Ana P. Cunha; Lincoln M. Alves A Seca de 2012-15 No Semiárido Do Nordeste Do Brasil No Contexto Histórico. *Revista Climanálise* **2016**, *4*, 49–54.
70. da Costa Júnior, D.S.; Ferreira, R.L.C.; da Silva, J.A.A.; da Silva, A.F.; Pessoa, M.M. de L. Dinâmica de Crescimento de Uma Floresta Tropical Sazonalmente Seca No Semiárido Brasileiro. *Ciência Florestal* **2022**, *32*, 1594–1616, doi:10.5902/1980509867697.
71. Worbes, M. Annual Growth Rings, Rainfall-Dependent Growth and Long-Term Growth Patterns of Tropical Trees from the Caparo Forest Reserve in Venezuela. *Journal of Ecology* **1999**, *87*, 391–403, doi:10.1046/J.1365-2745.1999.00361.X.
72. Bucci, S.J.; Goldstein, G.; Meinzer, F.C.; Scholz, F.G.; Franco, A.C.; Bustamante, M. Functional Convergence in Hydraulic Architecture and Water Relations of Tropical Savanna Trees: From Leaf to Whole Plant. *Tree Physiol* **2004**, *24*, 891–899, doi:10.1093/TREEPHYS/24.8.891.
73. DE LA CRUZ, D.B.C. IDADE RELATIVA E TEMPO DE PASSAGEM PARA Cenostigma Bracteosum (TUL.) E. Gagnon & G. P. Lewis NO SEMIÁRIDO DE PERNAMBUCO, Universidade Federal Rural de Pernambuco: Recife, 2024.
74. Da Silva, J.L.B.; De Albuquerque Moura, G.B.; De França E Silva, Ê.F.; Lopes, P.M.O.; Da Silva, T.T.F.; Lins, F.A.C.; De Oliveira Silva, D.A.; Ortiz, P.F.S. Dinâmica Espaço-Temporal Da Cobertura Vegetal de Caatinga Por Sensoriamento Remoto Em Município Do Semiárido Bras. *Revista Brasileira de Ciências Agrárias* **2019**, *14*, 1–10, doi:10.5039/AGRARIA.V14I4A7128.
75. Estrada Zúñiga, A.C.; Cárdenas Rodriguez, J.; Víctor, J.; Saya, B.; Ñaupari Vásquez, J.; Zúñiga, E.; Rodriguez, C.; Ñaupari Vásquez, J. V Estimación de La Biomasa de Una Comunidad Vegetal Altoandina Utilizando Imágenes Multiespectrales Adquiridas Con Sensores Remotos UAV y Modelos de Regresión Lineal Múltiple, Máquina de Vectores Soporte y Bosques Aleatorios. *Scientia Agropecuaria* **2022**, *13*, 301–310, doi:10.17268/sci.agropecu.2022.027.
76. Ferreira, M.B.; Ferreira, R.L.C.; Da Silva, J.A.A.; De Lima, R.B.; Silva, E.A.; De Sousa, A.N.; De, L.; Cruz, D.B.C.; Da Silva, M. V; Araújo, G.; et al. Spatial-Temporal Dynamics of Water Resources in Seasonally Dry Tropical Forest: Causes and Vegetation Response. *AgriEngineering 2024, Vol. 6, Pages 2526-2552* **2024**, *6*, 2526–2552, doi:10.3390/AGRIENGINEERING6030148.
77. Li, Y.; Li, M.; Li, C.; Liu, Z. Forest Aboveground Biomass Estimation Using Landsat 8 and Sentinel-1A Data with Machine Learning Algorithms. *Scientific Reports 2020 10:1* **2020**, *10*, 1–12, doi:10.1038/s41598-020-67024-3.
78. Guzmán-Quesada, J.A. High-Resolution Remote Sensing for Monitoring Lianas and Trees in Tropical Dry Forests, University of Alberta: Edmonton, 2021.
79. Barbosa, H.A.; Huete, A.R.; Baethgen, W.E. A 20-Year Study of NDVI Variability over the Northeast Region of Brazil. *J Arid Environ* **2006**, *67*, 288–307, doi:10.1016/J.JARIDENV.2006.02.022.
80. Miot, H.A. Análise de Correlação Em Estudos Clínicos e Experimentais. *J Vasc Bras* **2018**, *17*, 275–279, doi:10.1590/1677-5449.174118.
81. Suleymanov, A.; Gabbasova, I.; Komissarov, M.; Suleymanov, R.; Garipov, T.; Tuktarova, I.; Belan, L. Random Forest Modeling of Soil Properties in Saline Semi-Arid Areas. *Agriculture 2023, Vol. 13, Page 976* **2023**, *13*, 976, doi:10.3390/AGRICULTURE13050976.
82. Wang, Q.; Zhao, L.; Wang, M.; Wu, J.; Zhou, W.; Zhang, Q.; Deng, M. A Random Forest Model for Drought: Monitoring and Validation for Grassland Drought Based on Multi-Source Remote Sensing Data. *Remote Sensing 2022, Vol. 14, Page 4981* **2022**, *14*, 4981, doi:10.3390/RS14194981.
83. Fremout, T.; Cobián-De Vinatea, J.; Thomas, E.; Huaman-Zambrano, W.; Salazar-Villegas, M.; Limache-de la Fuente, D.; Bernardino, P.N.; Atkinson, R.; Csaplovics, E.; Muys, B. Site-Specific Scaling of Remote Sensing-Based Estimates of Woody Cover and Aboveground Biomass for Mapping Long-Term Tropical Dry Forest Degradation Status. *Remote Sens Environ* **2022**, *276*, 113040, doi:10.1016/J.RSE.2022.113040.

 doi:10.20944/preprints202505.1655.v1

84. Breidenbach, J.; McRoberts, R.E.; Astrup, R. Empirical Coverage of Model-Based Variance Estimators for Remote Sensing Assisted Estimation of Stand-Level Timber Volume. *Remote Sens Environ* **2016**, *173*, 274–281, doi:10.1016/J.RSE.2015.07.026.
85. Stojanova, D.; Panov, P.; Kobler, A.; Džeroski, S.; Taškova, K. LEARNING TO PREDICT FOREST FIRES WITH DIFFERENT DATA MINING TECHNIQUES. In Proceedings of the Conference on data mining and data warehouses; 2006; pp. 255–258.
86. Alberto, J.; Becerra, B.; De, S.; Jean, C.; Henry, P.; Ometto, B. Relação Das Sazonalidades Da Precipitação e Da Vegetação No Bioma Caatinga: Abordagem Multitemporal. In Proceedings of the XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR; RECIFE, April 25 2015; pp. 6668–6674.
87. da Costa Júnior, D.S.; Ferreira, R.L.C.; da Silva, J.A.A.; Silva, J.W.L.; Pessoa, M.M. de L. Influências Ambientais No Incremento Anual Da Área Basal, Em Ambiente de Caatinga. *Ciência Florestal* **2024**, *34*, e67699–e67699, doi:10.5902/1980509867699.
88. ANDRADE, W.M.; LIMA, E. de A.; RODAL, M. de J.N.; ENCARNAÇÃO, C.R.F.; PIMENTEL, R.M. de M. *Revista de Geografia*. RECIFE August 2009, pp. 161–184.
89. Galvíncio, J.D.; Mendes, S.M.; Morais, Y.C.B.; Souza, W.M.; De Moura, M.S.B.; Santos, W. Correlação Linear Entre a Precipitação e o Índice de Área Foliar Do Bioma Caatinga. *Revista Brasileira de Geografia Física* **2020**, *13*, 3304–3314, doi:10.26848/RBGF.V13.07.P3304-3314.
90. Jardim, A.M. da R.F.; Araújo Júnior, G.D.N.; da Silva, M.V.; Dos Santos, A.; da Silva, J.L.B.; Pandorfi, H.; de Oliveira-Júnior, J.F.; Teixeira, A.H. de C.; Teodoro, P.E.; de Lima, J.L.M.P.; et al. Using Remote Sensing to Quantify the Joint Effects of Climate and Land Use/Land Cover Changes on the Caatinga Biome of Northeast Brazilian. *Remote Sens (Basel)* **2022**, *14*, 1911, doi:10.3390/RS14081911/S1.
91. Ferreira, J.M.S.; Ferreira, H. dos S.; Silva, H.A. da; Santos, A.M. dos; Galvíncio, J.D. Análise Espaço-Temporal Da Dinâmica Da Vegetação de Caatinga No Município de Petrolina - PE (Analysis Space-Time from Dynamics of Caatinga Vegetation in the Municipality of Petrolina – PE). *Revista Brasileira de Geografia Física* **2012**, *5*, 904–922, doi:10.26848/RBGF.V5I4.232892.
92. Liu, Z.; Huang, T.; Zhang, X.; Wu, Y.; Xu, X.; Wang, Z.; Zou, F.; Zhang, C.; Xu, C.; Ou, G. Interacting Sentinel-2A, Sentinel 1A, and GF-2 Imagery to Improve the Accuracy of Forest Aboveground Biomass Estimation in a Dry-Hot Valley. *Forests 2024, Vol. 15, Page 731* **2024**, *15*, 731, doi:10.3390/F15040731.
93. Oliveira, S.N.; de Carvalho Júnior, O.A.; Gomes, R.A.T.; Guimarães, R.F.; McManus, C.M. Deforestation Analysis in Protected Areas and Scenario Simulation for Structural Corridors in the Agricultural Frontier of Western Bahia, Brazil. *Land use policy* **2017**, *61*, 40–52, doi:10.1016/J.LANDUSEPOL.2016.10.046.