Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Environmental gradients shape leaf phenology strategies and biomass dominance in a tropical forests

Alice Cristina Rodrigues
Universidade Federal de Viçosa
Pedro Manuel Villa
Universidade Federal de Juiz de Fora
Écio Souza Diniz
Centre for Research on Ecology and Forestry Applications
Alex Pires Coelho
Universidade Federal de Viçosa
Lhoraynne Pereira Gomes
Universidade Federal de Viçosa
Fernando Silla
Universidad de Salamanca
Andreza Viana Neri
andreza.neri@ufv.br
Universidade Federal de Viçosa

Research Article

Keywords: Aboveground carbon stock, Random Forest, Classification and Regression Tree, Support Vector Machines, Deciduousness

Posted Date: October 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11078079/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Ecological systems are shaped by complex, nonlinear interactions among environmental drivers, often leading to emergent functional patterns and trade-offs that cannot be inferred from single-factor analyses. In tropical forests, fine-scale environmental heterogeneity can generate contrasting functional strategies with distinct consequences for biodiversity and ecosystem functioning. Here, we investigate how interacting abiotic filters structure tree leaf phenology groups evergreen, semideciduous, and deciduous and how these emergent functional states relate to species richness and aboveground carbon storage in a late-secondary Atlantic forest in southeastern Brazil. Using machine-learning approaches capable of capturing nonlinear relationships (Random Forest, Support Vector Machine, and Classification and Regression Trees), we evaluated the relative importance of topographic, edaphic, climatic, and successional drivers on phenological group organization. Random Forest achieved the highest predictive performance, revealing elevation as the dominant environmental filter, followed by soil fertility and acidity, whereas climatic water deficit and forest stand age played secondary roles at this local scale. The resulting functional organization of the forest exhibited a clear emergent trade-off: evergreen species maximized local species richness, while deciduous species dominated aboveground carbon storage. No single phenological strategy simultaneously optimized both properties, indicating that biodiversity and carbon accumulation arise from different, interacting components of the system. Our findings demonstrate that environmental heterogeneity can produce disproportionate functional outcomes through nonlinear filtering, generating emergent trade-offs between diversity and biomass in tropical forests. By integrating empirical data with machine-learning methods as tools for revealing system structure rather than prediction alone, this study highlights the value of complexity-based approaches for understanding and managing multifunctional forest ecosystems under environmental change.

1. Introduction

The distribution, composition, and functioning of tropical forests are largely shaped by environmental filters such as soil properties, topography, and climate. These abiotic drivers influence plant growth, species distributions, community assembly, and ecosystem processes across multiple spatial scales (Toledo et al., 2012; Moeslund et al., 2013; Brown et al., 2013; Poorter et al., 2017; Rozendaal et al., 2019). Tropical forests, in turn, harbor more than half of global biodiversity and play a central role in climate regulation and the provision of ecosystem services, including carbon storage and hydrological regulation (Matos et al., 2020; Rodrigues et al., 2023). Understanding how abiotic gradients structure forest communities is therefore essential for predicting vegetation responses to global change and for informing conservation and management in biodiversity hotspots such as the Brazilian Atlantic Forest.

At local and mesoscale, topographic and edaphic heterogeneity generate strong environmental gradients that modulate resource availability (e.g. light, water, and nutrients) and shape tree community assembly (Brown et al., 2013; Ali et al., 2018; Jucker et al., 2018; Rodrigues et al., 2019, 2020, 2021). Elevation, slope, and convexity affect solar radiation, evapotranspiration, air temperature, and soil moisture, thereby

mediating the fine-scale climatic water balance experienced by plants (Hoylman et al., 2018, 2019, 2021). Soil properties, including fertility and acidity, further interact with topography to influence functional trait distributions and aboveground biomass (Ali et al., 2018, 2019; Rodrigues et al., 2021). While forest stand age reflects successional trajectories and can act as an indirect proxy for changes in structure and composition (Mora et al., 2018; Ali et al., 2019), its relative importance compared to topographic–edaphic filters at local scales remains less clear.

Climatic drivers also play a key role in controlling forest structure, diversity, and function (Jucker et al., 2018; Ali et al., 2019; Bennett et al., 2020, 2023; Doughty et al., 2023). In particular, climatic water deficit (CWD) integrates the balance between precipitation and atmospheric demand, strongly influencing local moisture regimes, plant water stress, and the distribution of plant functional types (Stephenson, 1998; Holden et al., 2019; Hoylman et al., 2019). Under climate change, increasing temperature, altered precipitation regimes, and more frequent extreme events are expected to drive shifts in functional composition, including changes in leaf phenology strategies (Aguirre-Gutiérrez et al., 2019; Beigaitė et al., 2022). For example, drier conditions can favor more drought-avoiding, deciduous species over evergreen species, with consequences for biodiversity and carbon cycling in tropical forests (Aguirre-Gutiérrez et al., 2019; Bennett et al., 2023).

In the Brazilian Atlantic Forest, tree species can be broadly grouped into three leaf phenology categories: evergreen, semideciduous, and deciduous (Singh and Kushwaha, 2016; Rodrigues et al., 2023). These functional groups differ in leaf lifespan, drought response, and carbon allocation strategies, which in turn affect forest productivity, resilience, and carbon storage. Recent studies have shown that variation in the relative abundance of these phenological groups is associated with fine-scale topographic and edaphic gradients and has strong implications for aboveground carbon stocks and biodiversity co-benefits during secondary succession (Rodrigues et al., 2019, 2020, 2022, 2023). However, the extent to which multiple abiotic drivers jointly predict the spatial distribution of leaf phenology functional groups, and how these groups contribute to species richness and carbon storage, remains insufficiently quantified at local scales.

Machine learning methods have become powerful tools in ecology for modeling complex, nonlinear relationships and integrating multiple predictors of vegetation patterns and ecosystem services (Cutler et al., 2007; Reichstein et al., 2019; Scowen et al., 2021; Beigaitė et al., 2022; Cetin et al., 2022). Algorithms such as Random Forest (RF), Support Vector Machines (SVM), and Classification and Regression Trees (CART) have been widely used to classify forest types, map tree species or functional types, and identify environmental thresholds controlling vegetation distributions (De'ath and Fabricius, 2000; Breiman, 2001; Onishi and Ise, 2021; Beigaitė et al., 2022). Yet, most applications rely on remote-sensing products at regional to global scales, whereas fewer studies use detailed field data from permanent plots to explore how local abiotic gradients drive functional group composition and associated ecosystem functions in tropical forests.

Here, we use long-term data from permanent plots in a late-secondary Atlantic forest to investigate how abiotic drivers shape the distribution of tree leaf phenology groups and their contributions to species richness and aboveground carbon stocks. Specifically, we apply three supervised machine learning algorithms (RF, SVM, and CART) to classify trees into evergreen, semideciduous, and deciduous functional groups based on climate (CWD), topography (elevation, slope, convexity), soil properties (sum of basic exchangeable cations, SB; exchangeable acidity, H + Al), and stand age. We hypothesize that (i) topographic and soil gradients are the main abiotic drivers of leaf phenology functional composition at this local scale, with climate playing a secondary but detectable role, and (ii) machine learning models can accurately predict the spatial distribution of leaf phenology groups from these abiotic predictors. Additionally, we evaluate how these functional groups differ in their contributions to species richness and aboveground carbon stocks, providing insights into the carbon–biodiversity co-benefits associated with different phenological strategies. This framework is particularly relevant for guiding conservation and restoration planning in tropical secondary forests under ongoing climate and land-use change.

2. Material and methods

2.1. Study site and vegetation sampling

We studied a Semideciduous Seasonal Atlantic Forest fragment in Minas Gerais state, Southeastern Brazil (20°45'14"S, 42°45'53"W) of approximately 75 ha that was used for coffee cultivation until 1926. Since then, it has been in natural regeneration and forest succesion. According to the Köppen-Geiger classification, the study area climate is tropical of altitude (Cw_b), with a dry season between May and September, and wet season occurring between December and March (Alvares et al., 2013). The mean annual temperature is 21°C and the mean annual precipitation is 1,270 mm, with the highest volumes of rain concentrated from October to March (140 mm to 200 mm per month) and often low values of accumulated precipitation from April to August (50 mm per month) (Avila-Diaz et al., 2020; UFV, 2020). The study area is located between 620 and 820 m elevation, and the relief varies from strongly undulating to mountainous and has marked differences in the spatial distribution of topographical variables, mainly elevation and convexity (Rodrigues et al., 2020, 2019). The site is characterized by the presence of two dominant soil classes: a red-yellow alsicose latosol covers hilltops and mountainsides; while a cambic yellow-red podzolic dominates the upper fluvial terraces (Ferreira-Júnior et al., 2007).

Our research group has been monitoring Atlantic forest fragments through the methodology of permanent plots for almost four decades (1984–2023). Thus, we use a valuable and extensive set of data of the dynamics of this forest ecosystem that can provide us evidence of the effects of climate change and other environmental factors on the structure of these forests. We use plots established in 1984 and 1993, with 1984, 1993, 1998, 2003, 2004, 2011, and 2017 measurements. The forest sampling areas are located in two contrasting topographic conditions, each comprising a continuous 1-ha plot subdivided into 100 contiguous subplots of 10 × 10 m (totaling 200 subplots and 2 ha of sampled area). Within each plot, all trees having diameter at breast height (DBH) 10 cm were identified to the species level and tagged for the measurement of tree height. All individuals were identified using specialized

literature, samples from the Herbarium VIC, or by taxonomists. The Angiosperm Phylogeny Group IV (APG IV 2016) was used for táxon classification. In addition to the species identification, we classified the tree species into three functional groups concerning the leaf phenology (deciduous, semideciduous, and evergreen species). The species were classified according to field observations and using literature.

2.2. Topographic variables

We measured three topographic variables (i.e., slope, elevation, and convexity) within each plot. Topographic variables were measured using a Ttotal Station, which measures vertical and horizontal angles and linear distances and is positioned at an obstacle-free location and aimed at a prism. The prism sits on a metal stick and is placed over the point to be measured. The Total Station then emits a laser beam that reflects in the prism and returns to the equipment. Using the response time of the laser beam to the equipment and the angle of rotation of the station's bezel, the internal computer calculates the angles and distances and stores the data in its internal memory (Kahmen and Faig 1988). The data was then transferred to a computer and analyzed with AutoCAD software (Autodesk Inc., San Rafael, CA, USA). Elevation was calculated using the mean elevation at each of the four corners of the plot. The slope (measured in degrees) was the mean angular deviation of the horizontal of each of the four triangular planes formed by the connection of three of its edges (Guo et al. 2017). Convexity was determined by subtracting the elevation at the center of the quadrat from the mean elevation of the eight surrounding plots. On edge plots, convexity was calculated as the altitude of the plot of interest minus the mean altitude of the surrounding plots (Lan et al. 2011).

2.3. Soil properties

Within each plot, a composite sample of the surface soil (0–10 cm depth) was collected. Soil properties of the samples were measured in the Soil Laboratory at the Federal University of Viçosa, following standard protocols (EMBRAPA 1997). The following soil properties were assessed: soil organic carbon (C); total N; available P, K, Ca, Mg, Fe, Zn; effective cation exchange capacity (CEC); exchangeable acidity potential (H + Al); sum of basic exchangeable cations (SB); base saturation (V); aluminum saturation (m); pH and organic matter (OM).

2.4. Climate data

We calculated the climate water deficit (CWD) for the study areas, which is a measure used to verify the intensity of drought during the year (Aguirre-Gutiérrez et al., 2019). CWD represents the most negative value of the water deficit, that is, the difference between precipitation and evapotranspiration within each year. To calculate CWD for a representative vegetation sampling period, we characterize each year's climate conditions with vegetation data, obtaining the average CWD across sampling years. Absolute (ΔCWDAbs) and relative (ΔCWDRel) changes in CWD were then calculated by subtracting ΔCWDAbs and ΔCWDR between vegetation sampling years for the study area. We used daily, monthly and annual precipitation and evapotranspiration datasets (recorded from 1980 to 2017) to estimate water balance. The potential evapotranspiration (PET) and actual evapotranspiration (AET) output from these water balance models was used to calculate the climatic water deficit (CWD = PET − AET; mm). We then

calculated the total annual CWD for each year, which represents the unmet atmospheric demand for moisture, an ecologically relevant metric of accumulated drought stress. The dataset was collected in the “Instituto Nacional de Meteorologia” database – INMET (INMET, 2020). The weather station where data were collected is based in Viçosa, Minas Gerais State, at the same Campus of the Federal University of Viçosa. Breaking point detection techniques were used to evaluate the temporal evolution of water deficit and investigate the incidence of discontinuity in the time series.

2.5. Estimation of aboveground carbon (AGC) and species richness

Estimating aboveground carbon (AGC) stocks in forests is commonly derived from allometric equations predicting forest biomass. Typically, the carbon content of various tree organs is assumed to be approximately 50% of their dry biomass (Chave et al., 2009). In this study, aboveground biomass (AGB) for each sampled tree stem was calculated using the general allometric equation (Eq. 1) developed by Chave et al. (2014), which incorporates tree diameter at breast height (DBH, cm), total height (H, m), and species-specific wood density (ρ, g cm⁻³). To obtain wood density values, we used data from the Global Wood Density Database (Chave et al., 2009; Zanne et al., 2009), applying genus- or family-level averages when species-specific information was unavailable. Tree diameter was measured in the field, while tree height was determined using a laser rangefinder.

$$AGB = 0.0673\,(\rho \times DBH^2 \times H)^{0.976}$$ Eq. 1

The total AGB per forest patch was calculated as the sum of the AGB of all trees with DBH ≥ 10 cm and subsequently converted to megagrams per hectare (Mg ha⁻¹) following Ali et al. (2017). Species-level biomass was estimated by summing the biomass of all stems belonging to a given species. All AGB estimations were conducted using the ‘BIOMASS’ package in R (Réjou-Méchain et al., 2017). Finally, we estimated the species richness and carbon distribution into the following groups in each plot: evergreen, semideciduous and deciduous.

2.6. Data analysis

2.6.1. Evaluation of trend and detection of breaking points

Different breaking point detection methods were used to evaluate the temporal CWD variation based on the incidence of discontinuity in the time series. Thus, the following breaking point detection methods were used: Mann-Kendall (M-K), Pettitt test, Buishad U test, Standard Normal Homogeinity test, Lanzante's test procedures (Yu and Ruggieri, 2019; Pohlert, 2020). These tests were used to identify trends and deviations in the temporal series of CWD data on seasonal scales, which show a significant breaking point in the time series of 30 years of data in the studied area region (Figure. 2). For example, all these tests (at 95% significance level) allows if the null hypothesis (H0) lies in the lack of no trend in climate extremes over time and that the alternative hypothesis (H1) lies in the evidence of a monotonic trend in them (increasing or decreasing). These analyses were calculated in the “trend” package (Pohlert,

2020). This test was calculated using seasonal scales, the premise of which is that significant changes based on the period of higher water deficit may indicate climate changes. These methods are widely used in climate studies to determine trends in time series data, indicating that the trend is greater when there is a larger deviation from zero (Yu and Ruggieri, 2019).

2.6.2. Principal Analysis component

We used a principal component analysis (PCA) on the correlation matrix to describe the topographical and soil gradients between leaf phenology groups, reducing the number of redundant variables on the PCA axes. This analysis was preceded by variable standardization to equalize their contributions on the PCA ordination axes using the 'FactoMineR' package (Husson et al. 2018). Thus, we constructed two PCA analyses, The first PCA represents the fertility and texture gradient due to the high correlation of the variables with the PCA axes. Subsequently, we analyzed the Spearman correlation between properties related to fertility and texture with the PCA axes to evaluate the contribution of variables (Figure S1 from Supplementary Material, SM).

2.6.3 Machine learning algorithms

We used three main machine learning algorithms for classification to compare the effectiveness and performance in predicting the leaf functional groups of the tree community (evergreen, deciduous, and semideciduous) using multiple predictive variables, such as CWD, elevation, slope, convexity, SB, HAL, and stand age (Fig. 1). Thus, we tested the Random Forest (RF), Support Vector Machine (SVM) and classification and regression tree (CART) algorithms to predict the importance relative of climate, soil properties, topography and stand age for the leaf phenological groups classification (i.e., Onishi et al. 2021; Beigaitė et al. 2022; Cetin et al. 2021). We tested the Random Forest (RF) using the randomForest (Liaw and Wiener, 2002) package, which had the same predictions and results in this study. Meanwhile, the support vector Machines (SVM) classifier used the caret package. Finally, we used the rpart (Therneau and Atkinson, 2021) and the caret (Kuhn et al., 2016) packages for fitting the decision trees – CART (i.e., Beigaitė et al. 2022).

2.6.3.1. Algorithms description

The Random Forest model is a supervised machine learning algorithm based on decision trees using 'bootstrap' aggregation (bagging), which is robust to data reduction, generating an internal unbiased estimate of the generalization error as the forest building progresses and having higher accuracy than individual decision trees and low sensitive to parameter adjustment than other machine learning models (Breiman 2001, Khaledian and Miller, 2020). In Random Forest, each tree is built using a deterministic algorithm by selecting a random set of covariates and a random sample from the training dataset (Khaledian and Miller, 2020).

The Support Vector Machine (SVM) is a technique used for classification tasks, which operates by identifying the optimal decision boundary that separates data points from different groups (or classes),

and then predicts the class of new observations based on this separation boundary (Khaledian and Miller, 2020; Diniz et al. 2021). The different groups classification might be separable by a linear straight line or by a non-linear boundary line. Thus, SVM can handle both linear and non-linear class boundaries and can be used for both two-class and multi-class classification problems. However, the separation boundary is generally nonlinear, and the SVM algorithm performs a non-linear classification using the kernel strick (Khaledian and Miller, 2020), mainly to predict association of abiotic factors and phenological functional groups (i.e., Beigaitė et al. 2022). Therefore, in this study the *radial kernel* transformation was used (i.e. Diniz et al. 2021).

The classification and regression tree (CART) algorithm is a non-parametric supervised learning algorithm, which is used for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes (Breiman 1984). This algorithm reachs reasonably high accuracy and extract the abiotic thresholds responsible for the separation of different leaf functional groups, which employ ensembles of decision trees (Breiman 1984; Beigaitė et al. 2022). However, as we focused on the extraction of the threshold values, we used a single tree model, which is more transparent for interpretation and has a lower risk of overfitting the data (Beigaitė et al. 2022). Then, a decision tree model was built interatively by first splitting the training data set based on the abiotic variable that is the most informative regarding leaf functional groups classes. Therefore, the most informative variable is often selected until the observations at the final nodes of the tree are classified. This algorithm typically utilizes Gini impurity to identify the ideal attribute to split on (James et al., 2013). To avoid possible overfitting, we regulated the complexity parameter, which indirectly controls the number of splits by imposing a relative cost for each split (Maindonald and Braun, 2013).

2.6.3.2. Data processing for modeling

In supervised machine learning the dataset is split into two subsets. One subset, the training data, is used to ‘train’ the algorithm how to carry out the task e.g., how to classify (i.e., Breiman 2001, Diniz et al. 2021; Beigaitė et al. 2022). This training data contains the target output and the user indicates what it is. The second subset, the test data, is reserved to ‘test’ the algorithm's performance in carrying out its task. In this phase, the target is not supplied to the algorithm so that the output produced by the algorithm can be compared to the target output data (Breiman, 2001). Thus, in this study, each time the algorithm was run using the createDataPartition function of the caret package, it will be trained on 70% of the data and tested on 30% (i.e., Breiman 2001, Diniz et al. 2021, Fig. 1).

2.6.3.3. Accuracy assessment methods

We have used the machine learning algorithm's Kappa index and three main accuracy methods (i.e., k-fold cross-validation, Repeated k-fold cross-validation, and Bootstrap) before estimating the relative importance of all predictors. The best classification found was chosen based on the accuracy rate (> 70%) and Kappa index (~ 50 and 60%). The k-fold cross-validation method involves the training set being split into k smaller sets (i.e., Breiman 2001; Diniz et al. 2021; Beigaitė et al. 2022); for example, we have used k = 10 in this study. This will split the dataset into 10 parts (10 folds) and the algorithm will be run

10 times (i.e., i.e., Breiman 2001, Diniz et al. 2021; Beigaitė et al. 2022). This process is completed until accuracy is determined for each instance in the dataset, and an overall accuracy estimate is provided. The Repeated k-fold Cross Validation is the process of splitting the data into k-folds that can be repeated several times, and the final model accuracy is taken as the mean from the number of repeats (i.e., Breiman 2001, Diniz et al. 2021; Beigaitė et al. 2022). Thus, we used 10 resamples iterations based on the repeatedcv argument of the “Caret” package. Finally, the bootstrap reesapling involves taking random samples from the dataset (with re-selection) against which to evaluate the model, and typically a large number of resampling iterations are performed (thousands or tens of thousands). Thus, the results provide an indication of the variance of the model’s performance (i.e., Breiman 2001, Diniz et al. 2021; Beigaitė et al. 2022). In the three methods randomly determined the calculation of the hyperparameters using “random” argument (Bergstra and Bengio, 2012).

2.6.3.4. Variables importance

Variable importance refers to how much a given model "uses" that variable to make accurate predictions (i.e. i.e. Breiman 2001, Diniz et al. 2021; Beigaitė et al. 2022). The more a model relies on a variable to make predictions, the more important it is to the model. It can apply to many different models, each using different methods described in the previous section. However, in this study, we consider the relative importance of predicted variables with RF that showed higher accuracy (i.e., Breiman 2001, Diniz et al. 2021).

3. Results

3.1 Performance of machine-learning algorithms

The classification of leaf functional groups (deciduous, semideciduous, and evergreen) based on multiple abiotic predictors (climate, topography, soil properties, and stand age; Fig. 1) showed the highest predictive performance when using the Random Forest (RF) algorithm compared to Classification and Regression Trees (CART) and Support Vector Machines (SVM) (Table 1). Among the validation approaches, the k-fold cross-validation applied to RF produced the best results, reaching an accuracy of 0.753 and a Kappa coefficient of 0.629 (Table 1). CART showed moderate predictive performance, especially under the Repeated k-fold validation, although consistently inferior to RF. Conversely, SVM exhibited the lowest efficiency across all validation methods, with accuracy values ranging from 0.60 to 0.64 and Kappa coefficients between 0.40 and 0.47 (Table 1).

Table 1
Performance of the learning algorithms in the caret package to classification of leaf phenology groups. RF = Random Forest; SVM = Support Vector Machines and; CART = Classification and Regression Tree Algorithm. Methods 1, 2 and 3 are repeated cross-validation; cross-validation and bootstrap respectively, for k-fold cross validation that the three models were tested.

Method 1	**RF**	**SVM**	**CART**
Kappa	0.622	0.473	0.674
Accuracy (%)	0.749	0.649	0.510
Method 2			
Kappa	0.629	0.460	0.470
Accuracy (%)	0.753	0.640	0.648
Method 3			
Kappa	0.569	0.410	0.455
Accuracy (%)	0.713	0.607	0.637

3.2 Relative importance of predictors (Random Forest)

Based on the best-performing RF model (k-fold cross-validation), topographic and soil variables emerged as the most influential predictors of functional group classification. Elevation showed the highest explanatory contribution across all three phenological groups, with importance values between 92 and 100% (Table 2). Soil properties were the next most relevant predictors, particularly base saturation (SB) and aluminum saturation (HAL) (Fig. 2). These variables were especially influential for deciduous species, whereas elevation alone had the strongest effect on the classification of evergreen species (Table 2). Variables related to terrain configuration (convexity and slope), climate (CWD), and stand age displayed substantially lower importance values in the RF models (Table 2).

Table 2
Relative importance of explanatory variables predicting leaf functional groups classifications in each learning algorithm in the caret package. RF = Random Forest; CART = Classification and Regression Tree Algorithm. Methods 1, 2 and 3 are repeated cross-validation; cross-validation and bootstrap respectively, for k-fold cross validation that the three models were tested. SB: sum of basic exchangeable cations; HAL: changeable acidity potential; CWD: climatological water deficit.

Random Forest			
Predictor	Importance		
Method 1	Deciduous	Evergreen	Semi deciduous
Elevation	**94.25**	**100.00**	**92.56**
SB	**91.91**	**64.51**	**52.94**
HAL	**66.69**	**53.27**	**54.39**
Convexity	39.49	38.14	27.77
Slope	33.93	31.79	16.83
CWD	23.44	14.47	23.80
Stand age	0.00	0.81	7.56
Method 2	Deciduous	Evergreen	Semi deciduous
Elevation	98.34	91.29	100.00
SB	80.00	55.42	36.46
HAL	44.66	43.45	39.19
Convexity	25.23	23.33	18.66
Slope	17.56	15.25	12.00
CWD	14.57	5.27	16.61
Stand age	1.84	0.00	6.95
Method 3	Deciduous	Evergreen	Semi deciduous
Elevation	100.00	97.32	90.80
SB	89.09	60.18	54.74
HAL	57.80	48.46	46.79
Convexity	39.78	39.61	23.35
Slope	32.84	26.97	16.64

Random Forest			
CWD	22.88	11.82	20.45
Stand age	0.10	0.00	5.51

3.3 CART decision tree analysis

Results obtained from the CART model using the Repeated k-fold Cross Validation method corroborated the patterns observed with RF. Elevation, SB, and HAL remained the most important predictors, maintaining the same relative order of importance (Fig. 3). The decision tree initiated its splitting based on elevation (< 703 m). At lower elevations, the subsequent key separation was driven by aluminum saturation (HAL ≥ 8.5), which contributed mainly to the classification of evergreen species. Therefore, functional group discrimination at these sites depended on the combined effects of elevation and soil properties, particularly SB and HAL (Fig. 3).

3.4 Climate variability and CWD

The temporal analysis of Climatic Water Deficit (CWD) detected an abrupt changepoint between 1980 and 2017, indicating shifts in climatic conditions during the study period (Fig. 4). Despite this variation, CWD did not rank among the main predictors within the RF models. However, CWD contributed to the separation between deciduous and evergreen groups within the CART decision tree, evidencing a secondary but ecologically relevant influence (Fig. 4).

3.5 Functional groups, species richness, and carbon stock

The evergreen functional group accounted for the highest proportion of species richness, comprising 40.7% of all recorded species in the study area, followed by the semideciduous (33.0%) and deciduous groups (26.3%) (Fig. 5). In contrast, aboveground carbon stock exhibited an inverse pattern among functional groups. Deciduous species contributed the largest share of carbon storage, representing 56.15% of the total aboveground biomass, whereas evergreen and semideciduous species accounted for 22.0% and 21.08%, respectively (Fig. 5). Together, these results reinforce the existence of a functional trade-off between species diversity and carbon storage among leaf phenology groups, with evergreen species maximizing richness while deciduous species dominate carbon accumulation.

4. Discussion

Our study in the Brazilian Atlantic Forest investigated how abiotic factors influence the distribution of tree leaf phenology groups (evergreen, deciduous, and semideciduous) using machine-learning approaches. We hypothesized that the occurrence of these functional groups would be primarily shaped by environmental gradients and could be reliably predicted by statistical learning models. Our results strongly supported this hypothesis, demonstrating that topography and soil properties — particularly

elevation, base saturation (SB), and aluminum saturation (HAL) — are the main drivers of leaf phenology group classification across the landscape.

Among the evaluated algorithms, the Random Forest model showed the highest predictive performance, highlighting its robustness for capturing complex, non-linear relationships between vegetation patterns and environmental variables. In contrast, forest stand age played a minor role in the classification process, presenting very low relative importance compared with the dominant abiotic predictors. This finding indicates that phenological functional composition is not driven directly by successional age per se, but rather by the environmental filters that operate consistently across successional stages.

These results have important implications for understanding forest responses to climate change and land-use dynamics. As shifts in climatic conditions and topographic-edaphic gradients reshape species distributions, the ability to predict changes in functional phenology groups becomes essential for anticipating impacts on species richness and aboveground carbon stocks. Consequently, integrating machine-learning predictions with functional trait approaches provides a valuable framework for guiding conservation planning, restoration strategies, and carbon-biodiversity co-benefits assessments in tropical forest landscapes under global change scenarios.

4.1. Abiotic factors strongly influence the distribution of functional groups based on leaf phenology

Our results showed that topography, mainly elevation, followed soil properties SB and HAL, which were the main factors that were the most important predictive variables for leaf phenology group classification at the local scale in this tropical forest. Despite the climatic water deficit breaking point based on climatic water deficit, indicating variation in the climate during the period analyzed, CWD was not the main predictor for leaf phenology group classification. For example, Hoylman et al., (2021), found that the climatic water balance imparts a strong control over the spatial distribution of plant functional types at large spatial scales. Still, this research explains that CWD provides an integrated metric on the effects of regional-scale climate patterns on water and energy available to plants, but these resources are mediated by local topography on a fine scale. These findings, along with our results, would explain the greater importance of topographic and soil predictors in relation to CWD in our study area. However, a larger time series of climatic data and a larger spatial or latitudinal scale would probably be needed to detect a greater influence of CWD as a predictive variable for classification. For example, the thresholds of climate extremes (e.g. extreme cold or drought) have been found to be essential for the dominance of certain vegetation types such as evergreen needleleaf forest and deciduous needleleaf forest classification at global scale using machine learning (Beigaitė et al. 2022).

Although breaking points in CWD related to regional climate change have been reported (Ávila-Díaz et al., 2020), CWD showed low importance in our models for classifying leaf phenology groups, except for its contribution to separating deciduous species in the CART analysis. This weak signal may reflect high climate variability and the growing frequency of extreme events, which limit the ability of relatively short

climate time series to capture consistent phenological responses (Díaz et al., 2021). Future studies using long-term vegetation monitoring from permanent plots and scenario-based modeling approaches may better resolve climate-driven shifts in functional group composition at local scales.

Conversely, variables that do not change temporally in the study, could have a spatial influence on the classification of functional groups as detected by RF and CART algorithms. The decision trees for separating different leaf phenology groups start the splitting based on the elevation variable and then simultaneously on the SB and HAL, followed by CWD. These results are consistent with our ecological understanding on the environmental filtering processes that promote vegetation classification. The observed leaf phenology group classification due to topography is probably linked to the underlying spatial variation of light, soil nutrients and climatic water availability, which are strongly influenced by topographic factors (Moeslund et al. 2013; Ali et al. 2019). Previously have been shown that evergreen species are associated with low elevations and high topographic heterogeneity, while deciduous species dominate higher elevation habitats and low heterogeneity (i.e., low fertility, convexity, and slope) in this study area (Rodrigues et al. 2022). Topography heterogeneity may affect resources such as light, water and constraints local soil nutrient and contribute to variance in microclimates and the climatic water balance (Hoylman et al 2019; Moeslund et al., 2013), within which trees grow, that in turn strongly affect forest species structure, composition and function (Jucker et al., 2018).

Previous studies (Rodrigues et al. 2021, 2022), have shown a strong correlation of evergreen species with area with higher topographic heterogeneity, and deciduous species with area of less topographic heterogeneity. Furthermore, the highest SB values are allocated to the lowest elevations, showing a relationship between topography and soil fertility distribution (Rodrigues et al., 2022). The CWD split detected by CART may reflect the ecophysiological responses of leaf phenology groups to water stress (Geange et al., 2021). Deciduous species, which typically exhibit lower tolerance to prolonged water deficit and rely on leaf shedding as a drought-avoidance strategy, tend to experience higher mortality under severe climatic water stress. This differential sensitivity may explain their separation from evergreen species along the CWD gradient, contributing to the observed split in the decision tree.

4.2. Machine learning algorithms and the effectiveness and performance in predicting the leaf functional groups

RF demonstrated the best performance in our analysis, achieving the highest accuracy in classifying leaf phenology functional groups. This effectiveness of RF can be attributed to its ability to handle large amounts of data and complexity, as well as its ability to reduce overfitting through the use of multiple decision trees (Cutler et al., 2007). The success of the RF algorithm against the other can be explained by the fact that the RF is suitable for handling unbalanced samples and adds additional randomness to the classification model during the as well as searching for only the best features among a random subset of features in the splitting process of each node (Indira et al., 2020).

Most of the tree species classification studies in the literature have used data from LiDAR or other remote sensing tools (Cetin et al., 2022; Shi et al., 2018). In this study, we used a set of field data at a

local scale to test the effects of environmental drivers on the classification of tree functional groups. In our study, as well as in those using data at regional scales, machine learning proved to be a robust tool to predict these patterns. Thus, the obtained accuracies of the machine learning-based classifications in our study are comparable with similar studies in regional scales.

The performance of SVM in our analysis was lower than that of RF and CART. Although SVM is known for its ability to handle complex, nonlinear data, it can be sensitive to tuning parameters and less effective on large, complex data sets, such as those found in ecological studies (Hsu et al., 2003). This may explain the lower accuracy observed with SVM in our analysis. CART showed moderate accuracy in classifying leaf phenology functional groups. Although it has a simpler and more intuitive interpretation compared to RF and SVM, CART may be less efficient in dealing with complexity and overfitting (De'ath and Fabricius, 2000). However, its ability to identify clear patterns in data sets can be valuable in exploratory analysis.

In short, each algorithm presented distinct advantages and limitations in classifying functional groups of leaf phenology in the Atlantic Forest. While RF stands out for its high accuracy and ability to deal with complexity, SVM can be sensitive to tuning parameters, and CART offers a simpler interpretation, but with less efficiency in dealing with data complexity.

5. Conclusions

Our study demonstrates that the spatial distribution of evergreen, semideciduous, and deciduous tree species in this Atlantic forest fragment is primarily structured by fine-scale abiotic filters, particularly topography and soil properties. Elevation emerged consistently as the strongest predictor of leaf phenology group classification, followed by indicators of soil fertility and acidity (SB and HAL). In contrast, forest stand age and climatic water deficit (CWD) showed low relative importance in the models, indicating that, at this local scale and successional stage, environmental filtering driven by topographic–edaphic gradients outweigh the effects of forest age and interannual climate variability on phenological functional composition. Among the evaluated machine-learning algorithms, Random Forest provided the highest predictive performance, highlighting the effectiveness of ensemble-based methods for capturing the complex and non-linear relationships between environmental drivers and functional vegetation patterns using field plot data. Our results also reveal a clear functional trade-off between species richness and aboveground carbon storage among leaf phenology groups. Evergreen species contributed most to local species richness, whereas deciduous species dominated carbon accumulation, emphasizing that no single functional group simultaneously maximizes both biodiversity and carbon storage. Maintaining a heterogeneous mixture of phenological strategies is therefore essential for balancing these complementary ecosystem services. Overall, our findings reinforce the importance of considering fine-scale abiotic heterogeneity and functional traits when predicting forest dynamics, carbon cycling, and biodiversity patterns in tropical secondary forests. This approach can support more effective conservation and restoration planning by aligning functional group composition with site-specific environmental conditions. Future studies integrating longer climate series, broader

spatial gradients, and biotic interactions may further refine predictions of functional responses to climate change and enhance strategies for managing the Atlantic Forest under ongoing global change.

Declarations

Conflict of interest statement

The authors declare that they have no conflicts of interest.

Acknowledgement

This study was financed in part by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) through grant BPD-00878-22 and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We are deeply grateful to all researchers, and specialists who contributed to data collection, species identification, and logistical support throughout the long-term monitoring of the study area. We also thank the team of the Laboratory of Plant Ecology and Evolution (LEEP/UFV) for technical assistance and valuable discussions that greatly improved this work.

Data Availability

The data and scripts supporting the findings of this study are publicly available in the GitHub repository: https://github.com/alicecristina261-cmyk/Script-Machine-learning.git

Author Contributions

Alice C. Rodrigues, Andreza V. Neri and Pedro M. Villa conceived the study ideas and designed the methodological framework. Alice C. Rodrigues, and Lhoraynne P. Gomes collected, organized, and processed the field, environmental, and climatic data.

Alice C. Rodrigues, Écio S. Diniz, and Pedro M. Villa performed the statistical analyses and machine-learning modeling. Alice C. Rodrigues and Andreza V. Neri led the writing of the manuscript. Fernando Silla and Alex Coelho contributed substantially to the critical review and improvement of the manuscript. All authors critically reviewed the manuscript drafts and approved the final version for publication.

References

1. Ali A, Lin SL, He JK, Kong FM, Yu JH, Jiang HS (2018a) Climatic water availability is the main limiting factor of biotic attributes across large-scale elevational gradients in tropical forests. Sci Total Environ 647:1211–1221
2. Ali A, Lin SL, He JK, Kong FM, Yu JH, Jiang HS (2019) Elucidating space, climate, edaphic and biodiversity effects on aboveground biomass in tropical forests. Land Degrad Dev. https://doi.org/10.1002/ldr.3278
3. Alvares CA, Stape JL, Sentelhas PC, de Moraes Gonçalves JL, Sparovek G (2013) Köppen's climate classification map for Brazil. Meteorol Z 22:711–728. https://doi.org/10.1127/0941-2948/2013/0507
4. Aguirre-Gutiérrez J, Oliveras I, Rifai S, Fauset S, Adu-Bredu S, Affum-Baffoe K, Baker TR, Feldpausch TR, Gvozdevaite A, Hubau W, Kraft NJB, Lewis SL, Moore S, Niinemets Ü, Peprah T, Phillips OL, Ziemińska K, Enquist B, Malhi Y (2019) Drier tropical forests are susceptible to functional changes in response to a long-term drought. Ecol Lett 22:855–865. https://doi.org/10.1111/ele.13243
5. APG IV (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc 141:399–436
6. Avila-Diaz A, Justino F, Lindemann DS, Rodrigues JM, Ferreira GR (2020) Climatological aspects and changes in temperature and precipitation extremes in Viçosa, Minas Gerais. An Acad Bras Cienc 92:e20190388. https://doi.org/10.1590/0001-3765202020190388
7. Bennett AC (2020) Climate more important than soils for predicting forest biomass at the continental scale. Ecography 43:1692–1705
8. Bennett AC, Rodrigues de Sousa T, Monteagudo-Mendoza A (2023) Sensitivity of South American tropical forests to an extreme climate anomaly. Nat Clim Change 13:967–974. https://doi.org/10.1038/s41558-023-01776-4
9. Beigaitė R, Tang H, Bryn A, Skarpaas O, Stordal F, Bjerke JW, Žliobaitė I (2022) Identifying climate thresholds for dominant natural vegetation types at the global scale using machine learning: Average climate versus extremes. Glob Change Biol 28:3557–3579. https://doi.org/10.1111/gcb.16110
10. Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. J Mach Learn Res 13:281–305
11. Brown C, Burslem DFRP, Illian JB, Bao L, Brockelman W, Cao M, Chang LW, Dattaraja HS, Davies S, Gunatilleke CVS, Gunatilleke IAUN, Huang J, Kassim AR, LaFrankie JV, Lian J, Lin L, Ma K, Mi X, Nathalang A, Noor S, Ong P, Sukumar R, Su SH, Sun IF, Suresh HS, Tan S, Thompson J, Uriarte M, Valencia R, Yap SL, Ye W, Law R (2013) Multispecies coexistence of trees in tropical forests: spatial signals of topographic niche differentiation increase with environmental heterogeneity. *Proceedings of the Royal Society B*, 280, 1764
12. Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees. CRC, Boca Raton. https://doi.org/10.1201/9781315139470
13. Breiman L (2001) Random forests. Mach Learn 45:5–32

14. Cetin Z, Yastikli N (2022) The use of machine learning algorithms in urban tree species classification. ISPRS Int J Geo-Information 11:226. https://doi.org/10.3390/ijgi11040226
15. Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. Ecology 88:2783–2792. https://doi.org/10.1890/07-0539.1
16. De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178–3192
17. Díaz S (2021) Summary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services. IPBES Secretariat, Bonn
18. Doughty CE, Keany JM, Wiebe BC (2023) Tropical forests are approaching critical temperature thresholds. Nature 621:105–111. https://doi.org/10.1038/s41586-023-06391-z
19. EMBRAPA (1997) Manual de métodos de análises de solo, 2nd edn. Empresa Brasileira de Pesquisa Agropecuária, Rio de Janeiro
20. Ferreira-Júnior WG, Silva AF, Schaefer CEGR, Meira Neto JAA, Dias AS, Ignácio M, Medeiros MCMP (2007) Influence of soils and topographic gradients on tree species distribution in a Brazilian Atlantic tropical semideciduous forest. Edinb J Bot 64:137–157. https://doi.org/10.1017/S0960428607000832
21. Guo Y, Wang B, Mallik AU, Huang F, Xiang W, Ding T, Wen S, Lu S, Li D, He Y, Li X (2017) Topographic species–habitat associations of tree species in a heterogeneous tropical forest. J Plant Ecol 57:1–10
22. Hadley W (2009) ggplot2: Elegant graphics for data analysis. Springer, New York
23. Holden ZA, Jolly WM, Swanson A, Warren DA, Jencso K, Maneta M, Landguth EL (2019) TOPOFIRE: a topographically resolved wildfire danger and drought monitoring system for the conterminous United States. Bull Am Meteorol Soc 100:1607–1613
24. Hoylman ZH, Jencso KG, Hu J, Holden ZA, Martin JT, Gardner WP (2019) The climatic water balance and topography control spatial patterns of atmospheric demand, soil moisture, and shallow subsurface flow. Water Resour Res 55:2370–2389
25. Hoylman ZH, Jencso KG, Hu J, Martin JT, Holden ZA, Seielstad CA, Rowell EM (2018) Hillslope topography mediates spatial patterns of ecosystem sensitivity to climate. J Geophys Research: Biogeosciences 123:353–371
26. Hoylman ZH, Jencso KG, Hu J, Martin JT, Holden ZA, Seielstad CA, Rowell EM (2021) The influence of hydroclimate and management on forest regrowth across the western U.S. Environ Res Lett 16:064057. https://doi.org/10.1088/1748-9326/abec03
27. Hsu CW, Chang CC, Lin CJ (2003) A practical guide to support vector classification. National Taiwan University, Taipei
28. Husson F, Josse J, Le S, Mazet J (2018) FactoMineR package: multivariate exploratory data analysis and data mining. R package version 1.0.14

29. Indira B, Valarmathi K (2020) A perspective of the machine learning approach for packet classification in software defined networks. Intell Autom Soft Comput 26:795–805
30. Instituto Nacional de Meteorologia (INMET) (2020) Dados meteorológicos. Available at: http://www.inmet.gov.br/portal/
31. James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. Springer, New York
32. Jucker T, Bongalov B, Burslem DFRP, Nilus R, Dalponte M, Lewis SL, Phillips OL, Qie L, Coomes DA (2018) Topography shapes the structure, composition and function of tropical forest landscapes. Ecol Lett 21:989–1000
33. Kahmen H, Faig W (1988) Surveying. Walter de Gruyter, Berlin
34. Khaledian Y, Miller BA (2020) Selecting appropriate machine learning methods for digital soil mapping. Appl Math Model 81:401–418. https://doi.org/10.1016/j.apm.2019.12.016
35. Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, Benesty M, Lescarbeau R, Ziem A, Scrucca L, Tang Y (2016) caret: Classification and regression training. R package
36. Lan GY, Hu YH, Cao M, Zhu H (2011) Topography-related spatial distribution of dominant tree species in a tropical seasonal rain forest in China. For Ecol Manag 262:1507–1513
37. Liaw A, Wiener M (2002) Classification and regression by randomForest. R News 2:18–22
38. Maindonald J, Braun J (2013) Data analysis and graphics using R: An example-based approach. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139194648
39. Matos FAR, Magnago LFS, Miranda AC, de Menezes LFT, Gastauer M, Safar NVH, Schaefer CERG, da Silva MP, Simonelli M, Edwards FA, Silva M (2020) Secondary forest fragments offer important carbon and biodiversity cobenefits. Glob Change Biol 26:509–522. https://doi.org/10.1111/gcb.14824
40. Moeslund JE, Arge L, Bøcher PK, Dalgaard T, Svenning JC (2013) Topography as a driver of local terrestrial vascular plant diversity patterns. Nord J Bot 31:129–144
41. Onishi M, Ise T (2021) Explainable identification and mapping of trees using UAV RGB image and deep learning. Sci Rep 11:903. https://doi.org/10.1038/s41598-020-79653-9
42. Pohlert T (2020) *trend: Non-parametric trend tests and change-point detection*. R package version 1.1.1
43. Poorter L, van der Sande M, Arets AJM, Ascarrunz N, Enquist B, Finegan B, Licona JC, Martínez-Ramos M, Mazzei L, Meave JA, Muñoz R, Nytch CJ, Ortiz-Malavassi E, Pérez-García EA, Rodríguez-Velázquez J, Russo SE, Saenz GP, Schall P, Schnitzer SA, Souza FM, Stein A, Zavala MA, Peña-Claros M (2017) Biodiversity and climate determine the functioning of Neotropical forests. Glob Ecol Biogeogr 26:1423–1434
44. R Development Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

45. Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N and Prabhat (2019) Deep learning and process understanding for data-driven Earth system science. Nature 566:195–204. https://doi.org/10.1038/s41586-019-0912-1
46. Rodrigues AC, Villa PM, Neri AV (2019) Fine-scale topography shapes richness, community composition, stem and biomass hyperdominant species in Brazilian Atlantic Forest. Ecol Ind 102:208–217. https://doi.org/10.1016/j.ecolind.2019.02.033
47. Rodrigues AC, Villa PM, Ali A, Ferreira-Júnior WG, Neri AV (2020) Fine-scale habitat differentiation shapes the composition, structure and aboveground biomass but not species richness of a tropical Atlantic Forest. J Forestry Res 31:1599–1611. https://doi.org/10.1007/s11676-019-00994-x
48. Rodrigues AC, Villa PM, Ferreira-Júnior WG, Schaefer CERG, Neri AV (2021) Effects of topographic variability and forest attributes on fine-scale soil fertility in late-secondary succession of Atlantic Forest. Ecol Processes 10:62. https://doi.org/10.1186/s13717-021-00333-1
49. Rodrigues AC, Villa PM, Silla F, Gomes LP, Meira-Neto JAA, Torres CMME, Neri AV (2022) Functional composition enhances aboveground carbon stock during tropical late-secondary forest succession. Plant Biosystems 156. https://doi.org/10.1080/11263504.2022.2073394
50. Rodrigues AC, Silla F, Meira-Neto JAA, Gomes LP, Villa PM, Neri AV (2023) Carbon and biodiversity cobenefits of second-growth tropical forest: The role of leaf phenology. For Ecol Manag 546:121377. https://doi.org/10.1016/j.foreco.2023.121377
51. Rozendaal DMA, Bongers F, Aide TM, Poorter L (2019) Biodiversity recovery of Neotropical secondary forests. Sci Adv 5:eaau3114. https://doi.org/10.1126/sciadv.aau3114
52. Scowen M, Athanasiadis JN, Bullock JM, Eigenbrod F, Willcock S (2021) The current and future uses of machine learning in ecosystem service research. Sci Total Environ 149263. https://doi.org/10.1016/j.scitotenv.2021.149263
53. Shi Y, Skidmore A, Heurich M (2018) Important LiDAR metrics for discriminating forest tree species in Central Europe. ISPRS J Photogrammetry Remote Sens 137:163–174
54. Shi Y, Skidmore A, Holzwarth S, Heiden U, Pinnel N, Zhu X, Heurich M (2018) Tree species classification using plant functional traits from LiDAR and hyperspectral data. Int J Appl Earth Obs Geoinf 73:207–219
55. Singh KP, Kushwaha CP (2016) Deciduousness in tropical trees and its potential as indicator of climate change: A review. Ecol Ind 69:699–706. https://doi.org/10.1016/j.ecolind.2016.04.011
56. Stephenson NL (1998) Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. J Biogeogr 25:855–870
57. Therneau T, Atkinson B, Ripley B (2021) rpart: Recursive partitioning. R package version 4:1–15
58. Universidade Federal de Viçosa – UFV (2020) Boletim meteorológico 2020. Departamento de Engenharia Agrícola, Estação Climatológica Principal, Viçosa
59. Villa PM, Martins SV, Oliveira Neto SN, Villa-Lobos J, Rodrigues AC, Teshome M, Carvalho FA, Gastauer M (2018) Woody species diversity as an indicator of the forest recovery after shifting

cultivation disturbance in the northern Amazon. Ecol Ind 95:687–694. https://doi.org/10.1016/j.ecolind.2018.08.005

60. Villa PM, Ali A, Martins SV, Oliveira SN, Rodrigues AC, Teshome M, Carvalho FA, Heringer G, Gastauer M (2020) Stand structural attributes and functional trait composition overrule the effects of functional divergence on aboveground biomass during Amazon forest succession. For Ecol Manag 477:118481. https://doi.org/10.1016/j.foreco.2020.118481
61. Wang Q, Punchi-Manage R, Lu Z, Franklin SB, Wang Z, Li Y, Chi X, Bao D, Guo Y, Lu J (2016) Effects of topography on structuring species assemblages in a subtropical forest. J Plant Ecol 10:440–449. https://doi.org/10.1093/jpe/rtw047

Figures

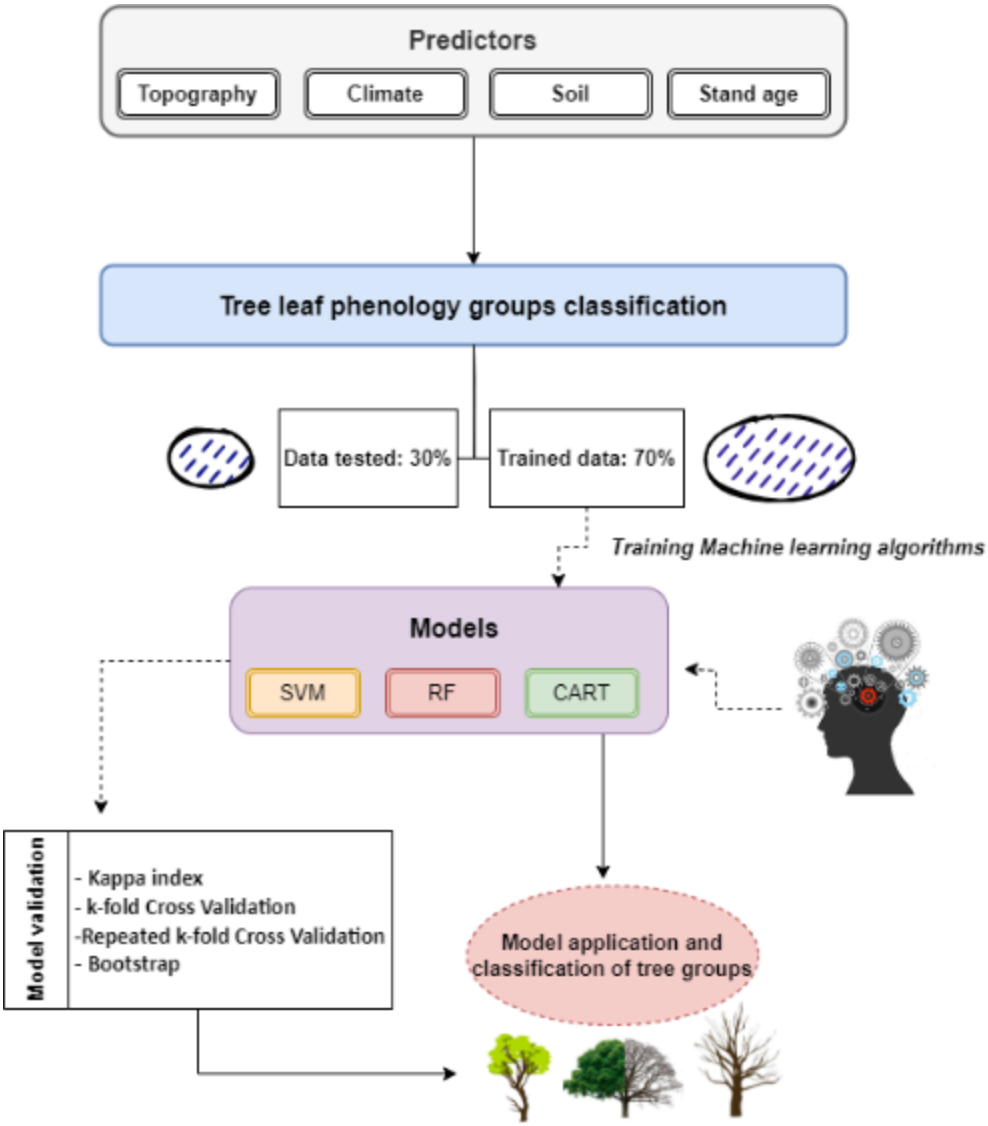


Figure 1

Modeling flowchart to predict the classification of tree functional groups under the effect of environmental variables. Environmental variables used as predictor variables include topography (elevation, convexity, and slope), soil (SB, HAL), climate (CWD), and stand age. The following models were trained and tested: Support Vector Machine (SVM), Random Forest (RF) and Classification and Regression Tree (CART). Kappa Index, k-fold Cross Validation, Repeated k-fold Cross Validation and Bootstrap were used as accuracy assessment methods.

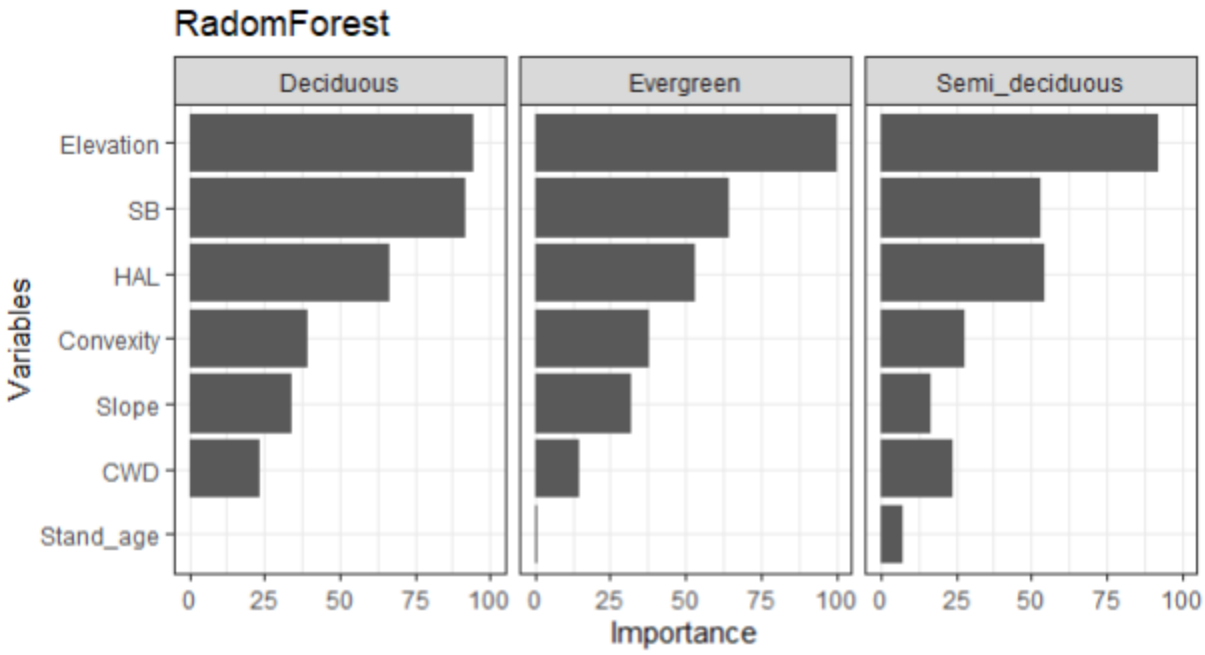


Figure 2

Relative importance of explanatory variables predicting leaf functional groups classification using Radom Forest Algorithm based on k-fold Cross Validation method. The predictors are ranked in order of importance starting with elevation, SB (sum of basic exchangeable cations), HAL (changeable acidity potential), Convexity, Slope, CWD (Climatological Water Deficit) and Stand age (i.e., successional stage of the late-secondary forest).

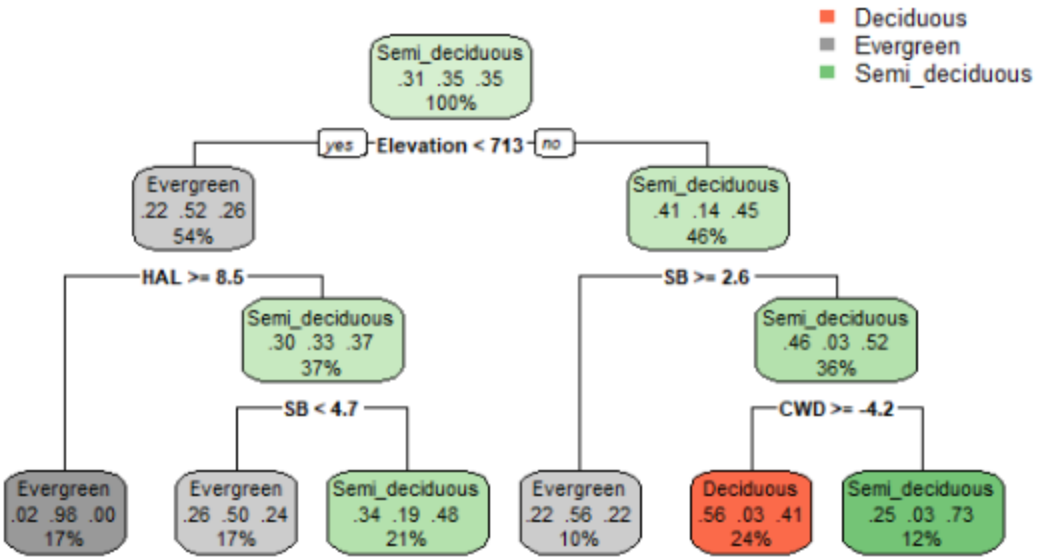


Figure 3

Decision tree using the multiple abiotic predictors. elevation, SB (sum of basic exchangeable cations), HAL (changeable acidity potential), CWD (Climatological Water Deficit) are indicated in the best CART selected.

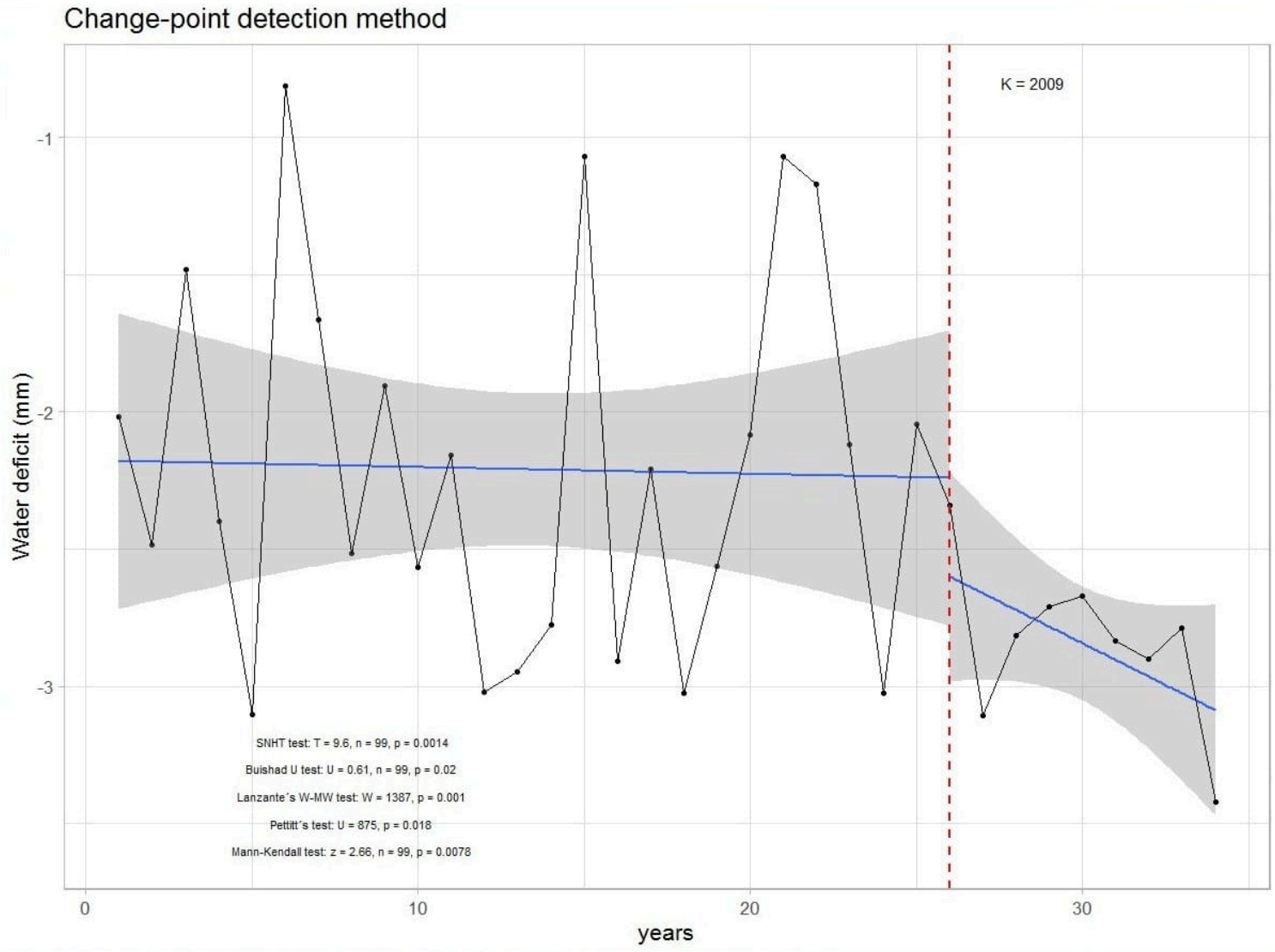


Figure 4

The temporal distribution of the water deficit presented an abrupt change point from 1980–to 2017. The dashed red line indicates a non-homogeneous series and the respective year of rupture.

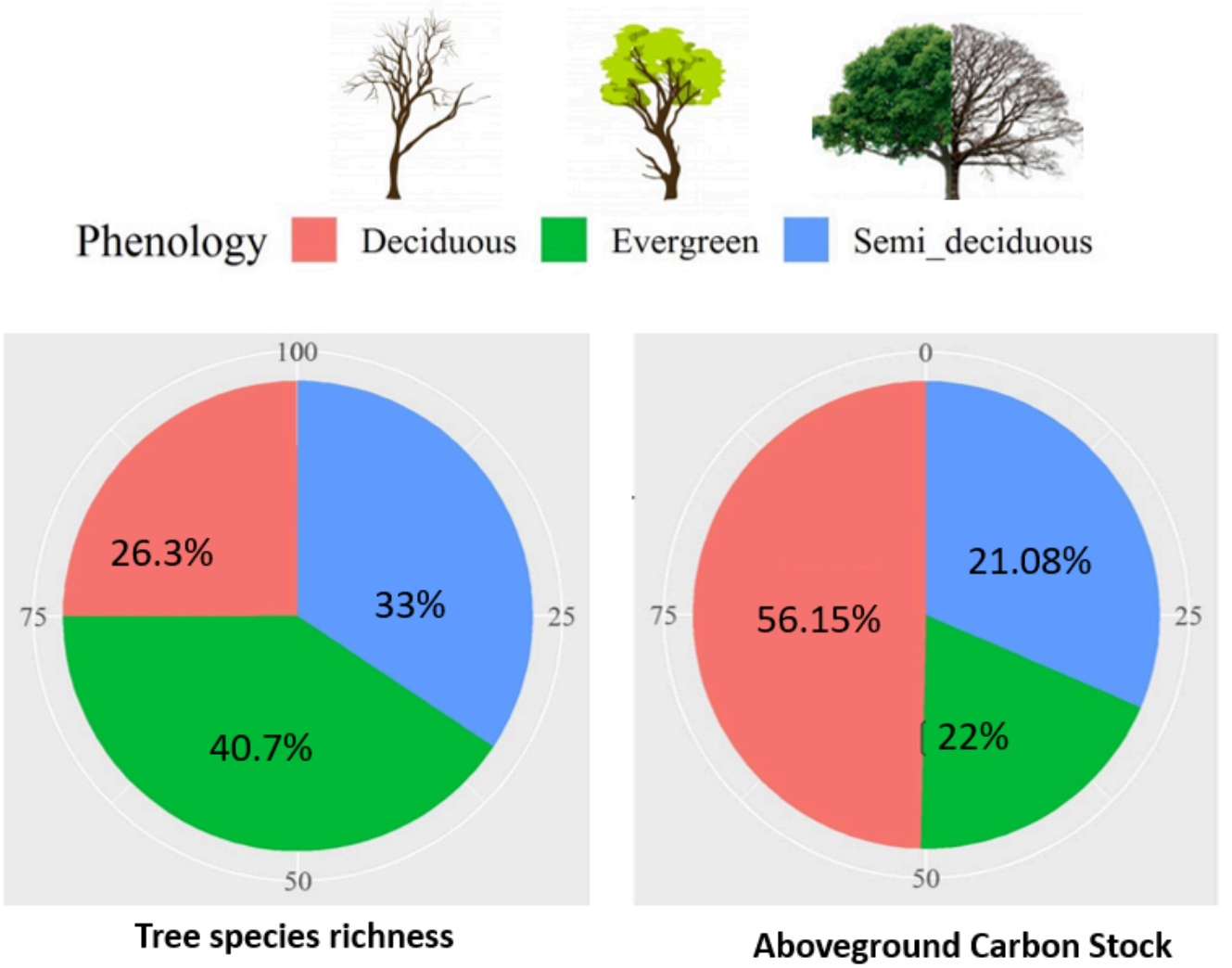


Figure 5

The relative proportion of aboveground carbon stock (AGC) and species richness in each functional group is based on the total number of tree species in the community per area.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx