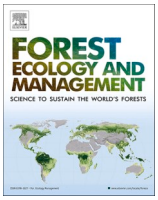
Forest Ecology and Management 580 (2025) 122520

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco



Predicting species-specific diameter growth rate for Caribbean trees using mixed-effects extreme gradient boosting

Sheng-I Yang [a,*], Thomas J. Brandeis [b], Eileen H. Helmer [c], Humfredo Marcano-Vega [d]

[a] Warnell School of Forestry and Natural Resources, University of Georgia, 180 E Green Street, Athens, GA 30602, USA
[b] USDA Forest Service, Southern Research Station, 4700 Old Kingston Pike, Knoxville, TN 37919, USA
[c] USDA Forest Service, International Institute of Tropical Forestry, Jardín Botánico Sur, 1201 Calle Ceiba, San Juan, PR 00926, USA
[d] USDA Forest Service, Southern Research Station, Jardín Botánico Sur, 1201 Calle Ceiba, San Juan, PR 00926, USA

ARTICLE INFO

Keywords:
Puerto Rico
U.S. Virgin Islands
Tree growth
Environmental variables
Machine learning

ABSTRACT

Caribbean islands encompass diverse forest ecosystems which provide valuable ecosystem services to their inhabitants. Currently, researchers rely on global generic or broad regional models or default values to predict tree species growth rates, which, in some cases, have been developed in temperate forests. The ability to understand the factors influencing the growth rates of Caribbean forest ecosystems, including both native and non-native species, has implications for projecting changes in and managing these forests under climate change scenarios. The objective of this study was to predict species-specific diameter growth rates for Caribbean trees with mixed-effects extreme gradient boosting (XGBoost), which is a hybrid approach combining a mixed model and a machine learning algorithm. The predictability of the models with and without the inclusion of climatic and environmental variables as predictors was examined. Long-term data collected by the US Department of Agriculture (USDA), Forest Service, Forest Inventory and Analysis (FIA) program in Puerto Rico and the U.S. Virgin Islands were used in analyses.

Results show that mixed-effects XGBoost is advantageous for providing species-specific predictions from both fixed and random effects, as well as capturing the remaining variability from XGBoost. Models produced more accurate predictions of growth rate for forests in the U.S. Virgin Islands than Puerto Rican forests. Among 30 variables examined, average diameter at breast height, average total tree height, average height-diameter ratio and average competition index play the most important roles for both islands. Adding topography- and climate-related variables can improve the prediction accuracy of annual diameter increment. This work will serve as a working example to demonstrate the application of the methodology to continuously monitor forest resources for the vulnerable ecosystems in Caribbean as well as other mixed-species forests.

1. Introduction

Caribbean islands encompass diverse forest ecosystems which provide valuable ecosystem services to their inhabitants (Brandeis and Turner, 2013a, 2013b; Helmer et al., 2002; López-Marrero and Heartsill-Scalley, 2012, Heartsill Scalley and González, 2016), and the degree to which they can be negatively impacted has been highlighted by recent hurricanes and droughts (Hu and Smith, 2018, López-Marrero et al., 2019, Uriarte et al., 2019, Hall et al., 2020, Zimmerman et al., 2021, Ramos-Scharrón et al., 2022). Our ability to estimate the impacts of these increasingly frequent and severe disturbances and predict forest recovery has been hindered by the lack of local or regional allometric equations and growth models. Currently, researchers rely on global generic or broad regional models or default values, which, in some cases, have been developed in temperate forests (see Brandeis et al., 2007; Brandeis and Turner, 2013a, 2013b for examples). Additionally, many tree species from outside the Caribbean have been naturalized there, resulting in "novel" species assemblages (Chinea and Helmer, 2003; Brandeis et al., 2009a; Lugo et al., 2012) for which we have lacked information on how both native and non-native species interactions might affect ecosystem functions and tree growth. In some instances, these naturalized species have become dominant components of Caribbean forest ecosystems (Brandeis et al., 2007, Brandeis and Turner, 2013a, 2013b; Marcano-Vega et al., 2015, Marcano-Vega and Williamson,

* Corresponding author.
E-mail addresses: syang23@uga.edu (S.-I. Yang), thomas.j.brandeis@usda.gov (T.J. Brandeis), eileen.helmer@usda.gov (E.H. Helmer), humfredo.marcano@usda.gov (H. Marcano-Vega).

https://doi.org/10.1016/j.foreco.2025.122520
Received 9 August 2024; Received in revised form 16 January 2025; Accepted 17 January 2025
Available online 30 January 2025

2017, Marcano-Vega, 2019, 2023). The biomass of some naturalized species can spike after extreme climate events, specifically severe storms, even when they die at higher rates during such events (Helmer et al., 2023). The ability to understand all of the factors influencing the growth rates of these and native species has implications for projecting changes in and managing these forests under climate change scenarios.

Thus, there is a need to develop Caribbean regional models for the tree species currently found there. These models would contribute to developing reliable statistical procedures that are essential to providing quantitative information for prudent resource management decision making, forest resilience monitoring, and meeting national and international greenhouse gas emissions and sustainability reporting requirements. There have been previous efforts devoted to developing quantitative methods for Caribbean trees (e.g., Yang et al., 2022; Brandeis et al., 2005; Brandeis et al., 2009b, Brandeis and Randolph, 2010), but the availability of increasing amounts of data from remeasured permanent plots provides opportunities for applying new modeling techniques to these larger datasets.

Over the past decades, various models were developed to predict individual tree growth ranging from single-species, even-aged to mixed-species, uneven-aged populations (Burkhart and Tomé, 2012; Petter et al., 2021; Vanclay, 1995). Compared to monocultures, predicting tree growth for uneven-aged, mixed-species forests is challenging due to (1) lack of tree age information, (2) small sample size for uncommon species, and (3) large variation among individuals. Stand age has proven a critical predictor to model tree growth, but it has nebulous meaning in the context of uneven-aged or mixed-species forests (Burkhart and Tomé 2012) and is largely not feasible in tropical forests where trees typically lack annual growth rings. Modeling efforts have focused on including other variables (e.g., average canopy height, functional traits, wood density, environmental factors) as surrogates for stand age in tree growth models for mixed-species forests (e.g., Adame et al., 2014; Li and Prentice, 2024). To address the issues of small sample size and high species diversity, grouping individuals based on similar tree characteristics (e.g., taxonomic rank, tree size) has been commonly applied in the past (e.g., Vanclay, 1991; Phillips et al., 2002; Zhang et al., 2022). Recently, mixed-effects modeling approaches provide another quantitative solution to produce reliable predictions of tree growth without data grouping (e.g., Russell et al., 2014; Yang et al., 2022; Kuehne et al., 2020). With the inclusion of both fixed and random components in a mixed model, variables of interest can be predicted at both regional scales and for local, plot-level predictions.

In the recent decades, a variety of machine learning algorithms were developed and integrated into mixed models. The algorithms were used to estimate the fixed effects of the model as they can capture the complex relationships between covariates. For example, Yang et al. (2022) applied mixed-effects random forests (MERF) with species as a random effect to characterize height-diameter relationships for Caribbean trees. They reported that MERF provides more accurate predictions than the original random forests and parametric models (Yang et al., 2022). Another application of machine learning in mixed modeling is to capture the remaining variability after a mixed model is fitted, also known as a hybrid model or residual learning method (e.g., Cornelio et al., 2022; de Mattos Neto et al., 2022). In this approach, a mixed model is first constructed, followed by training a machine learning algorithm with the residuals from the fitted mixed model as the response variable. Final predictions are made by combining the outputs of the mixed model and the predicted residuals from the machine learning algorithm. This method provides more interpretable fixed effects from the mixed model while leveraging the ability of machine learning to capture nonlinear and complex patterns in the data.

Gradient boosting techniques, which involve training a sequence of simple models where each one corrects the errors of the previous model, are widely used in machine learning. For example, Moisen et al. (2006) applied stochastic gradient boosting for mapping presence and basal area of 13 species in Utah. Extreme gradient boosting (XGBoost) proposed by Chen and Guestrin (2016) is another popular tree-based approach where the prediction is made with a sequence of decision trees. Specifically, XGBoost operates through an iterative process where each new decision tree is trained on the residuals (or errors) of the predictions made by the previous tree. This sequential approach allows XGBoost to progressively improve predictions by focusing on the mistakes of the prior models. XGBoost has been studied and applied in various fields, including forestry. For example, Watt et al. (2021) reported that XGBoost improves the prediction precision of site index and productivity measure for radiata pine (*Pinus radiata*) in New Zealand. MacMillan et al. (2022) developed quantitative models with XGBoost for predicting individual fire expenditures in British Columbia, Canada. Jha et al. (2023) used XGBoost to characterize height-diameter relationships for temperate and pantropical forests. Although XGBoost has proven useful, the application of a hybrid method that combines linear mixed-effects models with XGBoost in predicting tree growth for diverse species has not been explicitly investigated in the literature.

Therefore, the objective of this study was to predict species-specific diameter growth rates for Caribbean trees using a hybrid model. Specifically, we examined the predictability of the models with and without the inclusion of climatic and environmental variables as predictors, as well as compared the prediction accuracy between linear mixed and hybrid models. Long-term data collected by the US Department of Agriculture (USDA), Forest Service, Forest Inventory and Analysis (FIA) program in Puerto Rico and the U.S. Virgin Islands were used in analyses. This work will serve as a working example to demonstrate the application of the methodology to continuously monitor forest resources for the vulnerable ecosystems in Caribbean as well as for other mixed-species forests.

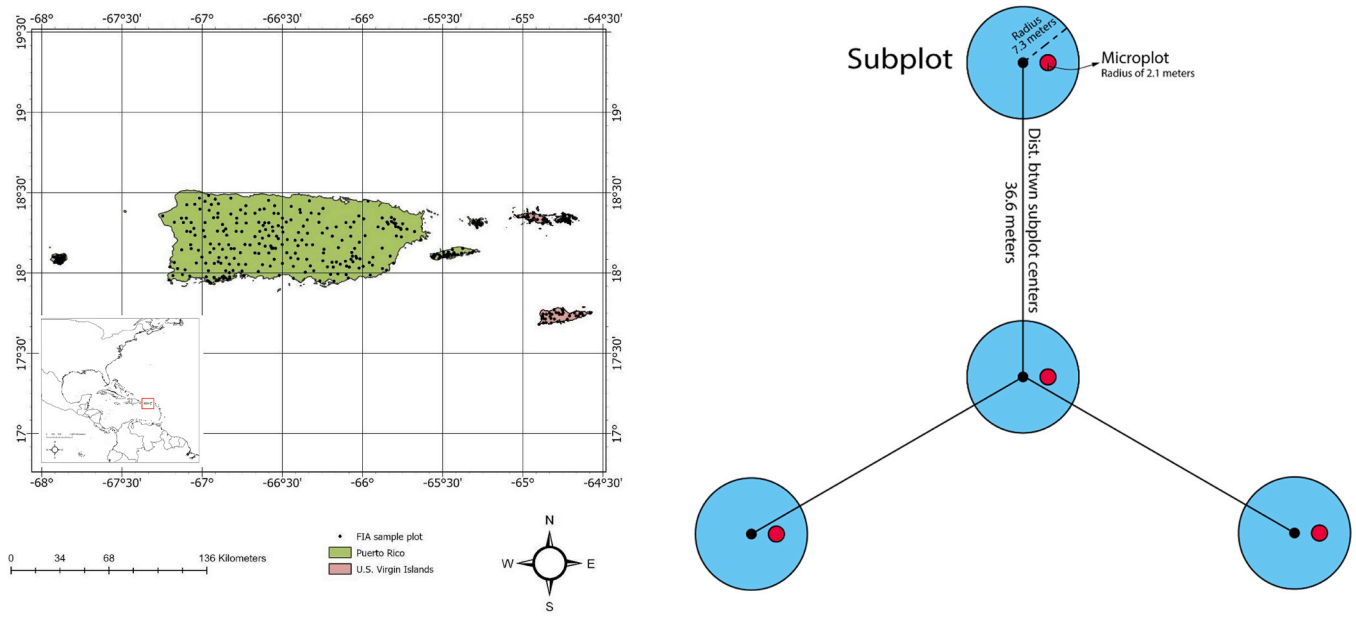
2. Materials and Methodology

2.1. Study area and tree data

2.1.1. FIA long-term sample plots

Data used in this study were collected from the long-term permanent forest inventory and monitoring plots established and maintained by the FIA program. Plots were first established on mainland Puerto Rico in 1980 (Birdsey and Weaver, 1982) although those data and their remeasurement data from 1990 (Franco et al., 1997) could not be used in this study due to compatibility problems introduced when the sampling design was modified and expanded in 2001 to better capture the full range of the diverse forest types (Brandeis, 2003). The plot network now covers the whole mainland of Puerto Rico and includes the outlying Puerto Rican islands of Culebra, Mona and Vieques. The plot network was also expanded in 2004 to include the three U.S. Virgin Islands of St. Croix, St. John and St. Thomas (Fig. 1a). The spatial distribution of the ecological life zones of Puerto Rico and the U.S. Virgin Islands can be found in Ewel and Whitmore (1973).

All sampling points in the systematic hexagonal grid were evaluated every 5 years (See Bechtold and Patterson (2005) and Brandeis 2003 for details on the sampling scheme). For those sampling points with at least 10 percent canopy cover of woody vegetation capable of reaching 5 m in height in-situ over a minimum area of 0.4 ha at the time of assessment, a forest inventory and monitoring plot was installed (details on the forest land definition can be found in USDA, 2023). Each permanent plot contained four 7.3-m (24-ft) fixed-radius subplots and four 2.1-m (6.8-ft) nested, fixed-radius microplots (see Fig. 1b, more detailed on field plot layout and data collection can be found in USDA (2023)). On each forested subplot, all trees with diameter at breast height (dbh) greater than or equal to 12.7 cm (5 in.) were measured, whereas on each nested microplot, trees from 2.5 to 12.6 cm (1.0–4.9 in.) dbh were measured. At each installation, the sampled area (the area of the permanent plot) was classified based on site conditions such as land use, forest type, stand size, regeneration status, tree density, stand origin, ownership group, and disturbance history. If there was more than one



(a) Plot network (b) Field plot layout

Fig. 1. Information of the sampling design. Fuzzed plot coordinates were used to illustrate the distribution of sample plots given the confidential information of the location.

distinct condition, the plot would be mapped by condition class even though the boundaries cut through subplots or microplots. In Puerto Rico and the U.S. Virgin Islands, each permanent plot was revisited every 5 years and remeasured if still forested. Sampling points which were previously not forested but had reverted to forest had plots installed. Each tree is individually identified, relocated and remeasured. Tree mortality or removal was accounted for, as was the recruitment of new trees on the plot. The long-term repeated measurements on individual trees (four measurements on most trees in Puerto Rico and three on trees in the U.S. Virgin Islands after the planned 2019 remeasurement there had to be skipped due to the COVID-19 pandemic) provided a statistically robust database for the development of quantitative models built on individual tree data (see Yang et al., 2022; Brandeis et al., 2007; Brandeis et al., 2009b, Brandeis and Randolph, 2010; for examples of similar work done using this or earlier versions of this dataset).

2.1.2. Tree data and characteristics

Tree data collected in Puerto Rico and the U.S. Virgin Islands were queried from the FIA database. To capture general trends of diameter growth for merchantable woody plants, individual trees were excluded if they met any of the following criteria: (1) they were palms or herbaceous plants, including *Acrocomia*, *Coccothrinax*, *Cocos*, *Prestoea*, *Roystonea*, *Columnea* and *Dracaena* spp.; (2) they were classified as rotten cull (indicating severe damage and decay which would affect the tree's form and growth; tree class code = 4 in the FIA database); (3) they were located on non-stocked lands (areas which were considered forest land but were still in an early or arrested stage of reforestation; forest type code = 999 in the FIA database); (4) their height or diameter measurements were missing. This resulted in a total of 10,850 live trees with at least two or more remeasurements from 382 permanent sample plots across islands, including 334 species in Puerto Rico and 101 species in U. S. Virgin Islands.

Total tree height (h) and dbh of each sample tree were used to compute the height-diameter ratio (h-d ratio = h/dbh) for quantifying individual tree form. To account for the competition effect among individual trees, diameter ratio (the ratio of tree dbh to the average dbh on a plot) was used, denoted as competition index (ci).

2.2. Response variable – species-specific growth rate at the plot level

Trees with at least two repeated measurements were selected to compute growth rate. Growth rate of an individual tree (g, cm/yr) was calculated as:

$$g = \frac{dbh_2 - dbh_1}{t_2 - t_1} \quad (1)$$

where dbh_1 and dbh_2 are diameter at breast height in cm at times 2 and 1 (t_2 and t_1), respectively. For example, if tree A's dbh was measured as 25.0 cm in 2003 and 28.2 cm in 2008, the growth rate of tree A during 2003–2008 is 0.64 cm/year (i.e., (28.2–25.0) / (2008–2003) = 0.64).

For a given species and year on an island, the average growth rate (G, cm/year) was calculated by averaging the growth rate of all sample trees on a plot. That is,

$$G = \frac{\sum g}{n} \quad (2)$$

where n is the total number of sample trees for a given species and year on an island. For example, two trees of species A are recorded on plot I in 2003 and 2008. Their growth rates are 0.64 and 0.50 cm/yr, respectively. The average growth rate for species A on plot I is 0.57 (0.57 = (0.64 +0.50)/2 cm/yr). Given the low number of data points for high growth rate plots, observations with G > 2 (cm/yr) in Puerto Rico and G > 1 (cm/yr) in the U.S. Virgin Islands were excluded. Distributions of growth rate by islands are given in Fig. 2.

2.3. Predictor variables

Past studies indicated that including stand or environmental variables in the models can improve predictability (e.g., Adame et al., 2014; Ankori-Karlinsky et al., 2024; Helmer et al., 2023; Jha et al., 2023; Gutierrez et al., 2023; Vanclay, 1995; Yang et al., 2022; Zhang et al., 2022). Based on literature and our knowledge, a total of 30 variables were selected as predictor variables listed in Table 1.

2.3.1. Stand variables

Seven stand characteristics were computed on a plot, including mean

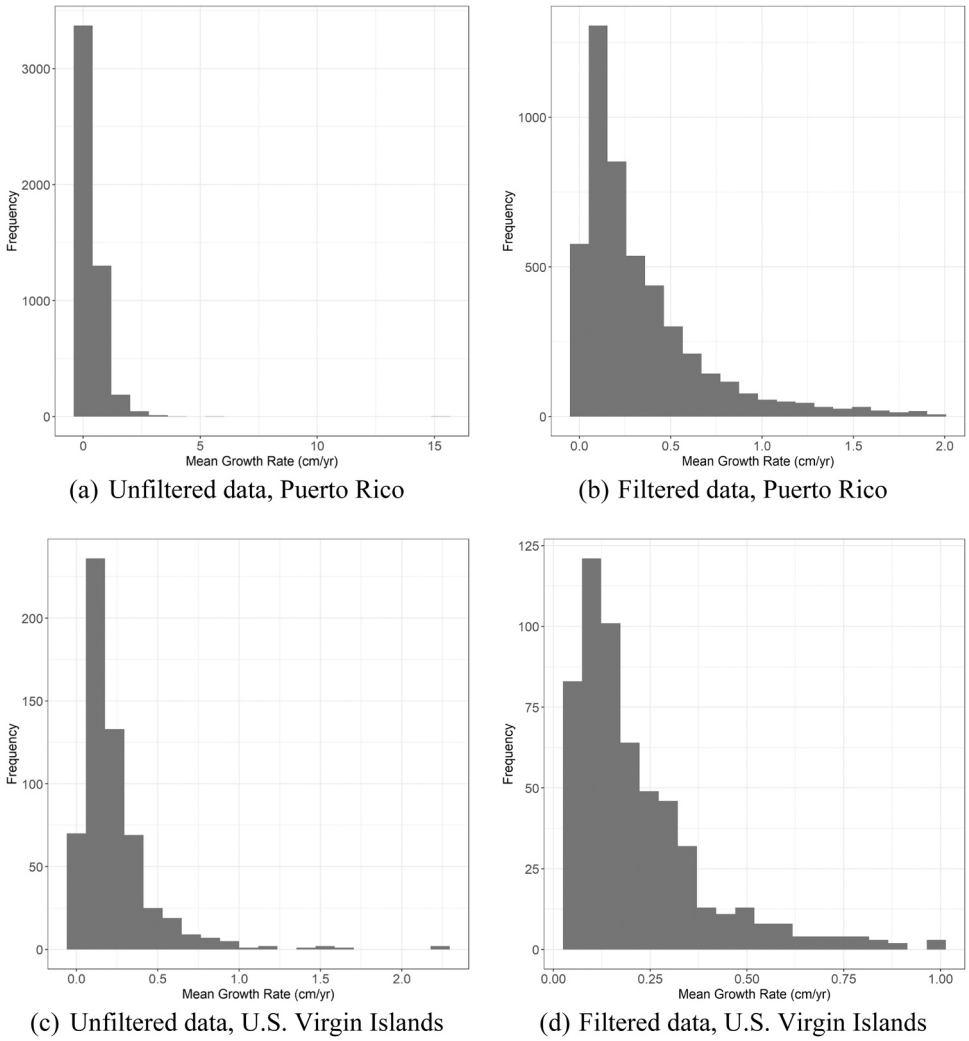


Fig. 2. Distribution of growth rate (cm/yr) by islands.

dbh (D), mean total tree height (H), mean h-d ratio (H-D ratio), mean ci (CI), diversity index, number of trees per ha (TPH), and basal area per ha (BA). For a given species and year, mean dbh (D) was computed as:

$$D = \frac{\sum dbh}{n} \tag{3}$$

where all variables have been defined as above. Similarly, H, H-D ratio and CI were calculated by averaging h, h-d ratio and ci of all live trees on a plot, respectively. For diversity index, Shannon-Wiener species diversity index (diversity index) was used, which was calculated as:

$$Diversity\ index = \exp(-\sum p \ln p) \tag{4}$$

where p is the ratio of the count of a species to the total count of all species on a given plot, exp is Euler's number, and ln is the natural logarithm.

2.3.2. Environmental variables

A total of 23 potential environmental covariates were selected. Environmental variables include topographic and climatic variables. Topographic variables include elevation in meters, sine and cosine of aspect, and slope in percent which came from Shuttle Radar Topography Mission 90-m digital elevation model data (Jarvis et al., 2008), a spatial resolution compatible with the plot sizes. Climate metrics include 19 bioclimatic variables extracted from PRISM climate monthly normals for Puerto Rico and the Virgin Islands for the periods 1963–1965 and 1971–2000, respectively (Daly et al., 2003; PRISM Climate Group, 2005) with the dismo package (v 1.3–14, (2023), Hijmans et al., 2017). The PRISM data for Puerto Rico were resampled with bilinear interpolation from ~450 m cell size to the native resolution of the Virgin Islands climate data cell size of about ~250 m. The bioclimatic variables are calculated from monthly normals and include average annual temperature and total annual precipitation as well as metrics like precipitation over the driest quarter of the year, or temperature of the coldest quarter.

2.4. Model building and evaluation

In this study, a mixed-effects XGBoost model was used to make species-specific predictions of growth rates. This hybrid model includes two components: a mixed model and an XGBoost model. Specifically, a linear mixed-effects model was built with all predictor variables listed in Table 1 as fixed-effects variables and species as a random-effect variable, which can be written as:

$$G = \sum x\beta + b + \epsilon \tag{5}$$

where $\sum x\beta$ is a linear combination of predictors (x) and their corresponding coefficients (β), b is a species-specific random effect, and ϵ is the error term. In this study, we assumed that b follows a normal distribution with mean zero and variance σ^2_{sp} (i.e., $b \sim N(0,\ \sigma^2_{sp})$), and ϵ follows a normal distribution with mean zero and variance σ^2 (i.e., $\epsilon \sim$

Table 1
Summary of all predictor candidates used in model building for Puerto Rico (PR) and U.S. Virgin Islands (VI).

Category	Label	Description	PR	VI
Stand	D	Mean dbh (cm)	15.1	10.2
	H	Mean total tree height (m)	9.4	6.3
	H-D ratio	Mean height-dbh ratio	86.8	88.7
	CI	Mean competition index (ratio of tree diameter to quadratic mean diameter)	1.1	1.1
	Diversity index	Shannon-Wiener species diversity index	7.6	7.1
	TPH	Trees per ha (stem/ha)	3094.8	5383.1
	BA	Basal area per ha (m^2/ha)	19.6	18.2
Temperature	Ann_t	Annual Mean Temperature (°C)	23.4	26.1
	Diurnal_range	Mean Diurnal Range (Mean of monthly (max temp - min temp)) (°C)	10.1	7.4
	Iso_therm	Isothermality (MDR/TAR) (×100) (unitless ratio)	73.1	67.9
	Season_t	Temperature Seasonality (standard deviation ×100) (°C x 100)	114.5	118.1
	Warm_mon_tmax	Max Temperature of Warmest Month (°C)	29.8	31.4
	Cold_mon_tmin	Min Temperature of Coldest Month (°C)	16.4	20.7
	Ann_t_range	Temperature Annual Range (MaxTWM - MinTCM) (°C)	13.4	10.7
	Wet_qtr_t	Mean Temperature of Wettest Quarter (°C)	24.3	26.7
	Dry_qtr_t	Mean Temperature of Driest Quarter (°C)	22	24.6
	Warm_qtr_t	Mean Temperature of Warmest Quarter (°C)	24.7	27.5
	Cold_qtr_t	Mean Temperature of Coldest Quarter (°C)	21.9	24.6
Precipitation	Ann_ppt	Annual Precipitation (mm)	1731.1	1147.3
	Wet_mon_ppt	Precipitation of Wettest Month (mm)	241.0	174.8
	Dry_mon_ppt	Precipitation of Driest Month (mm)	63.8	45.7
	Season_ppt	Precipitation Seasonality (Coefficient of Variation) (unitless ratio)	40.6	41.4
	Wet_qtr_ppt	Precipitation of Wettest Quarter (mm)	641.6	465.7
	Dry_qtr_ppt	Precipitation of Driest Quarter (mm)	221.3	164.5
	Warm_qtr_ppt	Precipitation of Warmest Quarter (mm)	508.8	321.1
	Cold_qtr_ppt	Precipitation of Coldest Quarter (mm)	234.1	181.3
Topography	Elevation	Elevation (m)	305.0	126.4
	Aspect_sin	Aspect Sine (Sine of the aspect angle in radians)	−0.1	0.0
	Aspect_cos	Aspect Cosine (Cosine of the aspect angle in radians)	0.1	−0.1
	Slope	Slope (%)	18.2	25.0

$N(0, \sigma^2)$).

Then, an XGBoost model was trained with the same predictors and the residuals (e_G, observed G – predicted G) calculated from the mixed model as the response variable. The final prediction of growth rate ($\widehat{G}_{final}$) is made by combining the predicted growth rate ($\widehat{G}$) from the linear mixed model and the predicted residual ($\widehat{e}_G$) from the XGBoost model. That is,

$$\widehat{G}_{final} = \widehat{G} + \widehat{e}_G \quad (6)$$

To build a mixed-effects XGBoost model, R packages "lme4" and "XGBoost" (Bates et al., 2024; Chen et al., 2024) were used. Specifically, the "lmer" function was used to construct the linear mixed model, whereas the "xgboost" function was used to train the XGBoost algorithm. A 10-fold cross validation was used to determine the optimal number of boosting rounds (decision trees or iterations) using "xgb.cv". Other hyperparameters were set as default or based on common values suggested in the literature (e.g., Bentéjac et al., 2019; Boehmke and Greenwell, 2019). Specifically, the learning rate (eta) was set to 0.1, each tree was allowed to split up to 6 times (max_depth), an 80 % subsampling ratio was applied to the training data for each boosting round (subsample), and 80 % of the features were randomly sampled for each tree (colsample_bytree).

2.4.1. Model building procedure – full and sub-models

The full and reduced models were built and evaluated based on prediction accuracy. The full model includes all predictors listed in Table 1, whereas the reduced models were built with a reduced number of predictors. Detailed steps of model building are given as:

1. Step 1: The full mixed model was built with all predictors included. Residuals were then calculated from the fitted model.
2. Step 2: An XGBoost model was trained with the residuals from step 1 and all predictors.
3. Step 3: The importance score of each predictor was recorded, which measures the average improvement in prediction accuracy brought by a feature when it is used for splitting the data in decision trees. A higher score implies the feature is more important for contributing to the predictive power of the model. More detailed description of the importance score can be found in Chen et al. (2024).
4. Step 4: Steps 1–3 were repeated 1000 times using a cluster bootstrap where resampling was done at the plot level. Cluster bootstrap was used to account for the potential temporal correlation among repeated measurements on a given plot. If a plot is sampled, all repeated measurements on a plot are selected in model building.
5. Step 5: All predictors were then arranged in descending order of the median of the importance score. The medians were calculated from 1000 bootstrap samples.
6. Step 6: Sub-models were then sequentially fitted with variables ranked by the importance score. For each sub-model, both mixed and XGBoost sub-models were built with the same set of predictors. For example, a mixed sub-model is constructed with D and H as predictors. Then, D and H are used to train an XGBoost sub-model with the residuals computed from the mixed sub-model. The procedure began with the model containing the two most important variables, and then the variable with the highest importance score was added each time in the sequence until all variables were added in the model.
7. Step 7: The final prediction of growth rate was made by combining the predicted values from both mixed and XGBoost models (see Eq. 6). Statistics used in model evaluation were calculated using the observed and final predicted growth rates (see Section "Evaluation of prediction accuracy").

Following the same procedure, the prediction accuracy of the models with and without environmental variables as predictors was examined. Only stand variables (listed in Table 1) were used as predictors when fitting models without environmental variables.

2.4.2. Evaluation of prediction accuracy

The prediction accuracy from the mixed-effects XGBoost models was compared using mean bias (MB) and root-mean-square error (RMSE), which were computed as:

$$MB = \frac{\sum e}{n} \tag{7}$$

$$RMSE = \left[\frac{\sum e^2}{n}\right]^{\frac{1}{2}} \tag{8}$$

where $\overline{G}$ is the mean of G, n is total number of observations, and e is prediction residual, which was defined as:

$$e = G - \widehat{G}_{final} \tag{9}$$

where all variables have been defined as above.

3. Results

The prediction accuracy of the mixed-effects XGBoost model varies by island and number of predictors (Fig. 2). According to mean bias, the mixed-effects XGBoost model can generally produce unbiased predictions of growth rate for Caribbean species in Puerto Rico and the U.S. Virgin Islands regardless of the inclusion of environmental variables as predictors (see mean bias in Fig. 2). The lowest RMSE of 0.23 was recorded for Puerto Rico, which is higher than that of 0.06 for the U.S. Virgin Islands. Models with both stand and environmental predictors can make more reliable predictions than those with only stand predictors (see RMSE in Fig. 2b). As shown in Table 2, the hybrid model generated more precise predictions (lower RMSE) compared to using linear models alone, especially for species in U.S. Virgin Islands. Among all predictor variable candidates, stand variables have higher importance scores than environmental variables, as D, H, H-D ratio and CI are the top four predictors for both islands (Fig. 3). Based on prediction accuracy and number of predictors (Fig. 2), the models with 24 and 17 predictors performed best, denoted as the best sub-model, for Puerto Rico and the U.S. Virgin Islands, respectively. Specifically, the Puerto Rico submodel includes all stand and topographic variables, four temperature variables and eight precipitation variables, whereas the U.S. Virgin Islands submodel includes all stand and topographic variables, two temperature variables and four precipitation variables (Table 3). Similar patterns of observed and predicted growth rates were found for the full model and the best sub-model regardless of island (Fig. 4). Higher observed growth rates are underpredicted, especially when using the models with only two predictors (Fig. 4). Compared to Puerto Rican species, the growth rates of the top five most abundant species in the U. S. Virgin Islands were more accurately predicted by the mixed-effects XGBoost model (Fig. 5).

4. Discussion

4.1. Evaluation of species-specific growth rate predicted by mixed-effects XGBoost

These results suggest that mixed-effects XGBoost models can produce reliable species-specific predictions of diameter growth rate, but the prediction accuracy varies by island and species (Figs. 4 and 5). By adding the machine learning algorithm, the hybrid model can reduce the uncertainty of predictions (lower RMSE as shown in Table 2). This modeling approach leverages the strengths of both the mixed-effects model and the XGBoost algorithm. The mixed-effects model generates interpretable results, while XGBoost captures the remaining variability in the data that the mixed-effects model does not fully explain.

This hybrid model is advantageous because it avoids the need for data grouping and can handle species with only one or a few observations. However, this same characteristic can also be a disadvantage, as it may compromise prediction accuracy by accounting for all species together. For example, growth rate was more accurately predicted for *Tabebuia heterophylla* than for *Guarea guidonia* and *Spathodea campanulata* in Puerto Rico (Fig. 5). By pooling all data together, model predictions may not be flexible enough for each individual species. Suppose that the variability of species A is best explained by variables I, II, and IV, but due to constraints imposed by all data points, variable IV may not be selected in the final model. Consequently, species A may not be well predicted due to the lack of a key predictor. A possible solution is to select a number of species that are of interest for a particular management goal and only use the selected species in model building.

Despite several advantages of using the proposed method, underprediction was found for higher growth rates, especially in Puerto Rico. This result may be due to the majority of the "near-zero" observations (Figs. 2 and 3) driving the overall predictions down. Notably, instead of using the growth rate of individual trees, we calculated growth rate by averaging all trees on a plot for a given species, which might reduce the level of variation. The high variability of growth rate may reflect not only the large number of species but also the complexity and diversity of the ecosystems. The U.S. Virgin Islands landscapes are less environmentally diverse than Puerto Rico, and the growth model was stronger. Notably, Puerto Rico has greater both σ_{sp} and σ than the U.S. Virgin Islands, which indicate that (1) Puerto Rico has greater variation among species than U.S. Virgin Islands, and (2) the unexplained variability in growth rate is higher for Puerto Rican forests than those in U.S. Virgin Islands.

4.2. Key variables identified in predicting growth rate

Stand variables contribute more to explaining the variability of growth rate than environmental variables. It may be because stand variables directly calculated from individual trees better reflect the site growing conditions, while environmental metrics were extracted from external sources and the precision level of GPS may add additional uncertainty in environmental variables. Among all variables, D, H, H-D ratio and CI play the most important roles for both islands. Their importance scores are significantly greater than the remaining variables (see 95 % confidence intervals in Fig. 3). Table 3 shows that D and H-D ratio have negative relationships with growth rate while H and CI have positive relationships. This implies that trees with large dbh have a smaller annual increment than small-sized trees. When the tree taper is greater, fewer resources are allocated to horizontal growth. Taller trees and those with higher competition indices are more dominant and acquire more resources for growth than other trees in a stand. Yang et al. (2022) and Jha et al. (2023) found that a measure of competition improves the prediction accuracy for height-diameter relationships for pantropical trees. Our results confirm that the inclusion of competition index is important in predicting the diameter growth rate for Caribbean species.

Among all environmental variables examined, topographic variables tended to be more important than climatic variables. Both aspect sine and aspect cosine, and elevation and slope, were in the top six environmental variables in both models. Topography affects solar radiation and exposure to prevailing winds and severe wind events and various topographic variables have been shown to be important predictors of tree mortality and canopy height during inventory intervals that include hurricanes and those that do not (Zhang et al., 2022; Helmer et al., 2023;

Table 2
Evaluation statistics of mean bias (MB) and root mean square error (RMSE) for the full models, best and poorest submodels for Puerto Rico (PR), and U.S. Virgin Islands (VI).

		PR		VI	
Type	Method	MB	RMSE	MB	RMSE
Full model	Linear mixed model	0.00	0.28	0.00	0.13
	Hybrid model	0.01	0.23	0.01	0.08
Best submodel	Linear mixed model	0.00	0.28	0.00	0.13
	Hybrid model	0.00	0.23	0.00	0.06
Poorest submodel	Linear mixed model	0.00	0.29	0.00	0.13
	Hybrid model	0.02	0.28	0.01	0.10

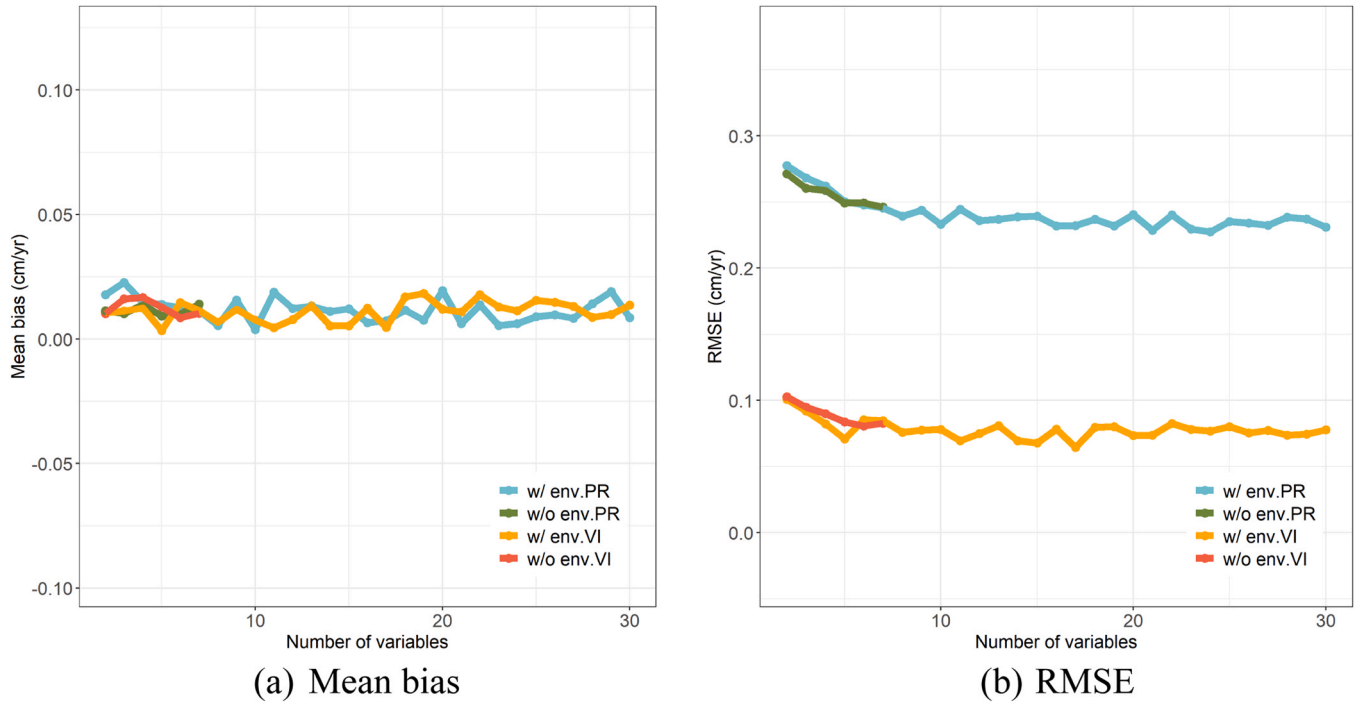


Fig. 3. Performance of mixed-effects XGBoost model by number of variables in Puerto Rico (PR), and U.S. Virgin Islands (VI). Models are built with and without environmental (w/ env. vs w/o env.) variables as predictors.

Table 3
Estimated coefficients for fixed- and random-effects of the final models (best submodels) for Puerto Rico (PR) and U.S. Virgin Islands (VI).

Puerto Rico		US Virgin Islands	
Variable	Estimated coefficient	Variable	Estimated coefficient
Intercept	0.295329	Intercept	−0.3376292
D	−0.000985	D	−0.0033869
H	0.013189	CI	0.0406674
CI	0.011861	H-D ratio	−0.0007923
H-D ratio	−0.001126	H	0.0146508
Diversity index	−0.010661	TPH	−0.0000003
TPH	−0.000008	Diversity index	0.0001477
BA	−0.002389	BA	−0.0026039
Aspect_cos	−0.007375	Aspect_sin	−0.0014515
Aspect_sin	−0.001060	Diurnal_range	−0.1603110
Slope	−0.001275	Aspect_cos	−0.0188059
Season_t	0.003581	Elevation	0.0002077
Elevation	0.000013	Slope	−0.0008282
Diurnal_range	0.060337	Dry_mon_ppt	−0.0073862
Warm_qtr_ppt	−0.000061	Ann_t_range	0.1304819
Ann_t_range	−0.066040	Wet_mon_ppt	−0.0053278
Cold_mon_tmin	0.001689	Ann_ppt	0.0014429
Ann_ppt	0.000137	Warm_qtr_ppt	−0.0001067
Dry_mon_ppt	0.000939	σ_{sp}	0.0740127
Wet_mon_ppt	0.000156	σ	0.1350037
Cold_qtr_ppt	−0.000046		
Wet_qtr_ppt	−0.000080		
Season_ppt	−0.001261		
Dry_qtr_ppt	−0.000777		
σ_{sp}	0.106944		
σ	0.287894		

Ankori-Karlinsky et al., 2024; Gutierrez et al., 2023). Elevation and slope, and to some extent aspect, may also better represent fine-scale differences in climate normal among plots. Elevation and slope are strongly positively correlated with precipitation and negatively correlated with temperature while having a finer spatial resolution, at ~90 m, than the climate data cell size of ~250 m. Both models also included diurnal temperature range among the top six environmental variables, which is lower both around urban areas, which stay warmer, and in higher-elevation areas, which stay cooler, sharply contrasting with moist forest areas that have larger diurnal temperature ranges than other areas. Temperature seasonality was slightly important in the Puerto Rico model, where it is lowest at high elevations in Eastern Puerto Rico, which are the areas receiving the most rainfall.

4.3. *Further discussion on the application of mixed-effects XGBoost*

In mixed-effects XGBoost, the predictors used in building the mixed-effects model can be the same as or different from those used in training the XGBoost model. Using different sets of predictors allows mixed and XGBoost models to select their own important variables in prediction. However, this approach may not be practically useful since it may require additional data for certain variables. For example, if D, H and CI are selected by a mixed model, but D, H and slope are selected by XGBoost, additional data on slope are needed. Notably, in our preliminary study, the prediction accuracy does not improve when different sets of predictors are used in fitting mixed and XGBoost models. Thus, the same set of predictors was used in the formal analysis.

The main goal of this work was to provide a working example for modeling diameter growth rate for a large number of species and variables. The same modeling framework can be applied to other machine learning algorithms, but the efficacy of the model needs to be further investigated. Using a 10-fold cross validation approach in model training provides a satisfactory level of prediction accuracy. Predication accuracy may be further improved if the hyperparameters are tuned deliberately, such as using a grid search. The models presented in this study may not be the most optimal models for this dataset. A comprehensive analysis of the optimal hyperparameters is warranted for future studies. In addition, the mixed-effects machine learning approach may require significant time and computational power due to the iterative process involved (Yang et al., 2022). We did not observe a long processing time (all analyses were done within 15–30 minutes), possibly due to the data size and the algorithm selected. XGBoost is well-known for its computational efficiency (Bentéjac et al., 2019; Chen and Guestrin, 2016; Boehmke and Greenwell, 2019). Lastly, models with a

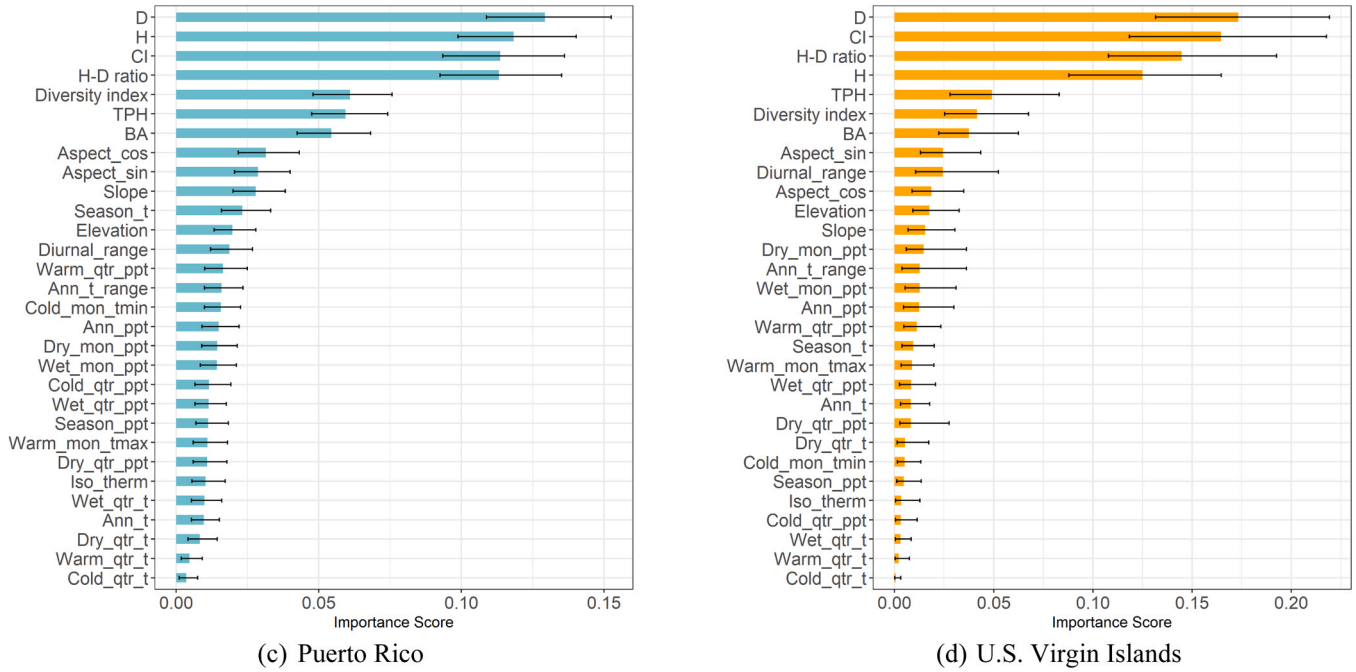


Fig. 4. Summary of importance score and its 95 % confidence interval by predictor variables for Puerto Rico and U.S. Virgin Islands.

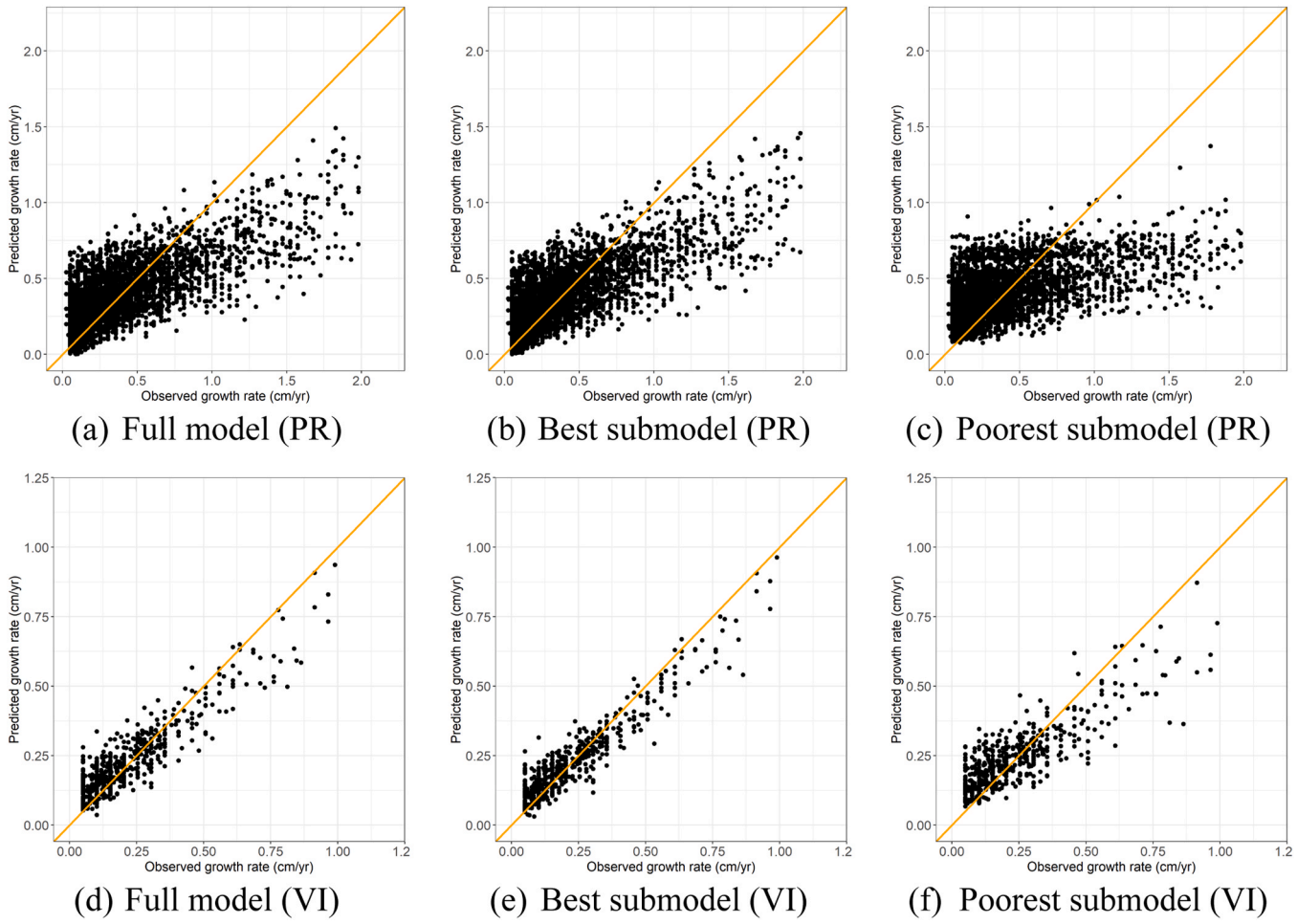
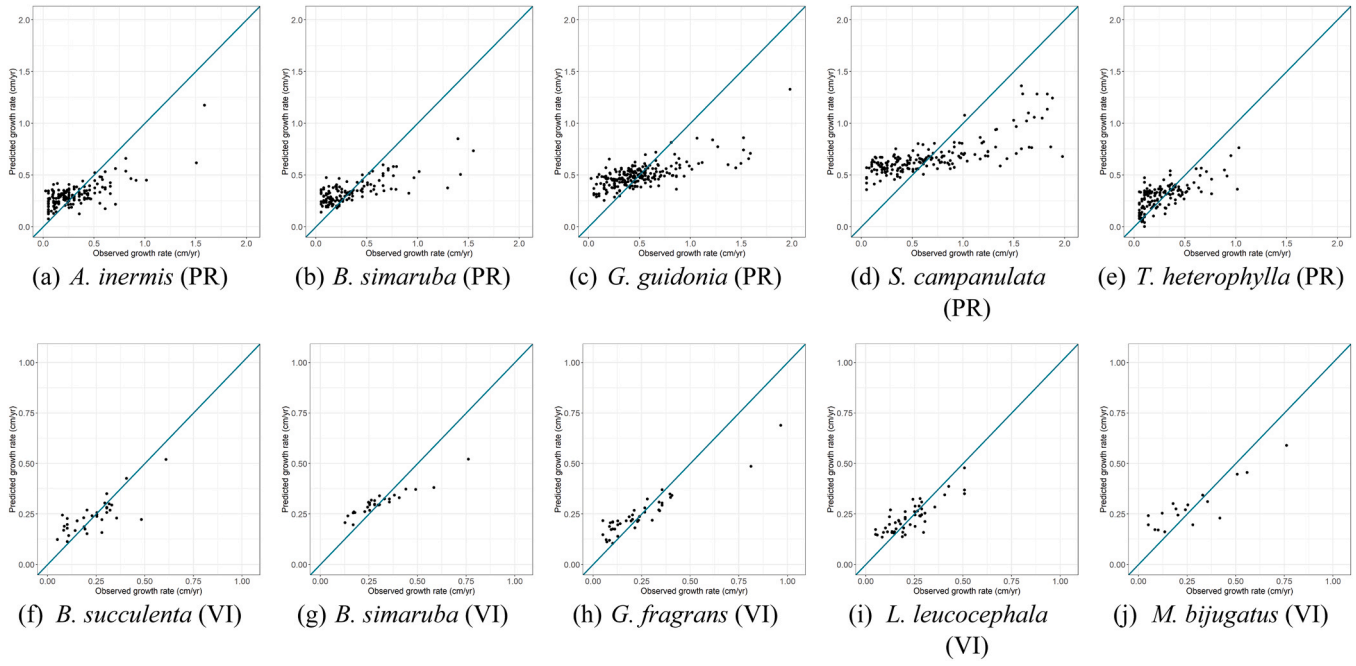


Fig. 5. Comparison of observed and predicted growth rate in Puerto Rico (PR) and U.S. Virgin Islands (VI) among the full model, the best and the poorest submodels. All predictors are considered (i.e., stand and environmental variables). The best model (highest prediction accuracy) was identified with 24 and 17 predictors for PR and VI, respectively. Models with only two predictors were found to provide the lowest prediction accuracy for both islands.



(a) *A. inermis* (PR) (b) *B. simaruba* (PR) (c) *G. guidonia* (PR) (d) *S. campanulata* (PR) (e) *T. heterophylla* (PR)

(f) *B. succulenta* (VI) (g) *B. simaruba* (VI) (h) *G. fragrans* (VI) (i) *L. leucocephala* (VI) (j) *M. bijugatus* (VI)

Fig. 6. Comparison of observed and predicted growth rate for the top five abundant (i.e., total number of observed plots) species in Puerto Rico (PR) and U.S. Virgin Islands (VI). The growth rate was predicted from the best submodel for each island. *A. inermis, B. succulenta, B. simaruba, G. fragrans, G. guidonia, L. leucocephala, M. bijugatus S. campanulata, T. heterophylla* represent *Andira inermis, Bourreria succulenta, Bursera simaruba, Guapira fragrans, Guarea guidonia, Leucaena leucocephala, Melicoccus bijugatus, Spathodea campanulata, Tabebuia heterophylla.*

machine learning component may not be immediately applicable to existing growth and yield systems in forestry practice, as these systems are primarily built on parametric models. This limitation may slow the implementation of the models introduced in this study. However, with the growing popularity of programming languages and publicly-available databases in forestry, implementing these models is expected to become increasingly feasible.

5. Conclusion

In summary, mixed-effects XGBoost as a hybrid modeling approach is advantageous for providing species-specific predictions from both fixed and random effects, as well as capturing the remaining variability from XGBoost. Models produced more accurate predictions of growth rate for forests in the U.S. Virgin Islands than Puerto Rican forests, maybe due to higher environmental and species diversity in Puerto Rico. Among all variables examined, average diameter at breast height (D), average total tree height (H), average height-diameter ratio (H-D ratio) and average competition index (CI) play the most important roles for both islands. Adding topography- and climate-related variables can improve the prediction accuracy of annual diameter increment.

The models presented in this study can be used to answer various questions in Caribbean forest management, which provide a better assessment of tree health and stand conditions after disturbances such as hurricanes. More importantly, the diameter growth rate predicted can be used to assess the growth of forest carbon stocking in the Caribbean forests. The proposed methodology will not only provide species-specific diameter growth models for Caribbean trees, but also advance knowledge in modeling growth rate with broad application for mixed-species forests. Further investigations on applying the proposed method to different forest types or environmental conditions in other Caribbean forests are needed.

CRediT authorship contribution statement

Yang Sheng-I: Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Brandeis Thomas J:** Writing – original draft, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Helmer Eileen H:** Writing – original draft, Methodology, Funding acquisition, Data curation, Conceptualization. **Marcano-Vega Humfredo:** Writing – review & editing, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Support from the USDA Forest Service (project # 24-JV-11330145–057) is greatly appreciated. We are grateful for the valuable comments from two anonymous reviewers.

Data availability

The FIA data can be found at https://www.fia.fs.fed.us/.

References

Adame, P., Brandeis, T.J., Uriarte, M., 2014. Diameter growth performance of tree functional groups in Puerto Rican secondary tropical forests. For. Syst. 23 (1), 52–63.

Ankori-Karlinsky, R., Hall, J., Murphy, L., Muscarella, R., Martinuzzi, S., Fahey, R., Zimmerman, J.K., Uriarte, M., 2024. Chronic winds reduce tropical forest structural complexity regardless of climate, topography, or forest age. Ecosystems 1–13.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R.H.B., Singmann, H., Dai, B. Scheipl, F., Grothendieck. G., Green, P., Fox, J., Bauer, A., Krivitsky, P.N., Tanaka, E. and Jagan, M. (2024) Package "lme4" - Linear Mixed-Effects Models using 'Eigen' and S4. 129 p.

Bechtold, W.A. and Patterson, P.L. (2005) The enhanced forest inventory and analysis program - national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p. (⟨https://research.fs.usda.gov/treesearch/20371⟩).

Bentéjac, C., Csörgő, A., and Martínez-Muñoz, G. (2019) A Comparative Analysis of XGBoost. arXiv:1911.01914.

Birdsey, R.A., Weaver, P.L., 1982. The forest resources of Puerto Rico. Resour. Bull. SO 85, 56.
Boehmke, B., Greenwell, B.M., 2019. Hands-On Machine Learning with R. Chapman and Hall/CRC, p. 484.
Brandeis, T.J., 2003. Puerto Rico's forest inventory: adapting the forest inventory and analysis program to a Caribbean island. J. For. (1), 8–13. ⟨https://research.fs.usda.gov/treesearch/30105⟩.
Brandeis, T.J.; Helmer, E.H.; Oswalt, S.N. (2007) The status of Puerto Rico''s forests, 2003. Resour. Bull. SRS-119. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 72 p. https://doi.org/10.2737/SRS-RB-119.
Brandeis, T.J., Helmer, E.H., Marcano-Vega, H., Lugo, A.E., 2009a. Climate shapes the novel plant communities that form after deforestation in Puerto Rico and the U.S. Virgin Islands. For. Ecol. Manag. 258 (7), 1704–1718. https://doi.org/10.1016/j.foreco.2009.07.030.
Brandeis, T.J., Randolph, K.C., Strub, M.R., 2009b. Modeling Caribbean tree stem diameters from tree height and crown width measurements. Int. J. Math. Comput. For. Nat. -Resour. Sci. 1 (2), 78–85. ⟨https://research.fs.usda.gov/treesearch/33817⟩.
Brandeis, T.J., Randolph, K.C., 2010. Modeling Caribbean tree heights and crown widths. Caribb. J. Sci. 46 (2–3), 176–185. https://doi.org/10.18475/cjos.v46i2.a6.
Brandeis, T.J.; Turner, J.A. (2013a) Puerto Rico's forests, 2009. Resour. Bull. SRS-RB-191. Asheville, NC. U.S. Department of Agriculture Forest Service, Southern Research Station. 85 p. https://doi.org/10.2737/SRS-RB-191.
Brandeis, T.J.; Turner, J.A. (2013b) U.S. Virgin Islands, 2009—forest inventory and analysis factsheet. e-Science Update SRS-SU-077. Asheville, NC: USDA-Forest Service, Southern Research Station. 5 p. ⟨https://research.fs.usda.gov/treesearch/45105⟩.
Burkhart, H.E., Tomé, M., 2012. Modeling Forest Trees And Stands. Springer Netherlands, p. 458.
Chen, T., and Guestrin, C. (2016) XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA. ACM. pp. 785–794.
Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Yuan, J., 2024. R Package 'xgboost' - Extreme Gradient Boosting. 68 p.
Chinea, D., Helmer, E., 2003. Diversity and composition of moist and wet secondary forests of Puerto Rico: analysis of the 1990 forest inventory. Forest Ecology and Management 180, 227–240.
Cornelio, J., Mohd Razak, S., Cho, Y., Liu, H., Vaidya, R., Behnam, J., 2022. Residual learning to integrate neural network and physics-based models for improved production prediction in unconventional teservoirs. SPE Joural 22, 3328–3350. https://doi.org/10.2118/210559-PA.
Daly, C., Helmer, E.H., Quiñones, M., 2003. Mapping the climate of Puerto Rico. Vieques Culebra. Int. J. Clim. 23, 1359–1381.
de Mattos Neto, P.S.G., Cavalcanti, G.D.C., de O. Santos Júnior, D.S., 2022. Hybrid systems using residual modeling for sea surface temperature forecasting. Sci. Rep. 12, 487. https://doi.org/10.1038/s41598-021-04238-z.
Ewel, J.J. and Whitmore, J.L. (1973) The Ecological Life Zones of Puerto Rico and the U. S. Virgin Islands. USDA Forest Service, Institute of Tropical Forestry, Research Paper ITF-018.
Franco, P.A.; Weaver, P.L. and Eggen-McIntosh, S. (1997). Forest resources of Puerto Rico, 1990. Resour. Bull. SRS-22. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 4.5 p.(⟨https://research.fs.usda.gov/treesearch/30327⟩).
Gutierrez, G.V., Marcano, H., Ruzycki, T., Wood, T., Anderegg, W., Powers, J., Helmer, E., 2023. Aridity and forest age mediate landscape scale patterns of tropical forest resistance to cyclonic storms. Author Prepr.
Hall, J., Muscarella, R., Quebbeman, A., Arellano, G., Thompson, J., Zimmerman, J.K., Uriarte, M., 2020. Hurricane-Induced rainfall is a stronger predictor of tropical forest damage in Puerto Rico than maximum wind speeds. Sci. Rep. 10 (2020), 4318. https://doi.org/10.1038/s41598-020-61164-2.
Heartsill-Scalley, T., Gonzalez, G., 2016. Introduction: Caribbean forest dynamics and regional forestry initiatives. The Caribbean Foresters: A collaborative network for forest dynamics and regional forestry initiatives. Caribbean Naturalist (1), 1–12.
Helmer, E.H., Kay, S., Marcano-Vega, H., Powers, J.S., Wood, T.E., Zhu, X., Gwenzi, D., Ruzycki, T.S., 2023. Multiscale predictors of small tree survival across a heterogeneous tropical landscape. Plos One 18 (3), e0280322.
Helmer, E., Ramos, O., Lopez, T.D.M., Quinones, M., Diaz, W., 2002. Mapping the forest type and land cover of Puerto Rico, a component of the Caribbean biodiversity hotspot. Caribb. J. Sci. 38 (3-4), 165–183.
Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2017. Package 'dismo.' Species distribution modeling. Circles 9 (1), 1–68. Version 1.3-14 (2023).
Hu, T., Smith, R.B., 2018. The impact of Hurricane Maria on the vegetation of Dominica and Puerto Rico using multispectral remote sensing. Remote Sens 10, 827. https://doi.org/10.3390/rs10060827.
Jarvis, A., Reuter, H.I., Nelson, A. and Guevara E. (2008) Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (⟨http://srtm.csi.cgiar.org⟩).
Jha, S., Yang, S., Brandeis, T.J., Kuegler, O., Marcano-Vega, H., 2023. Evaluation of regression methods and competition indices in characterizing height-diameter relationships for temperate and pantropical tree species. Front. For. Glob. Change 6, 1282297. https://doi.org/10.3389/ffgc.2023.1282297.
Kuehne, C., Russell, M.B., Weiskittel, A.R., Kershaw, J.A., 2020. Comparing strategies for representing individual-tree secondary growth in mixed-species stands in the Acadian Forest region. For. Ecol. Manag. 459, 117823.
Li, J., Prentice, I.C., 2024. Global patterns of plant functional traits and their relationships to climate. Commun. Biol. 7, 1136. https://doi.org/10.1038/s42003-024-06777-3.
López-Marrero, T., Heartsill-Scalley, T., Rivera-López, C.F., Escalera-García, I.A., Echevarría-Ramos, M., 2019. Broadening our understanding of hurricanes and forests on the Caribbean island of Puerto Rico: where and what should we study now? Forests 10, 710. https://doi.org/10.3390/f10090710.
López-Marrero, T., Heartsill-Scalley, T., 2012. Get up, stand up:environmental situation, threats, and opportunities in the insular Caribbean. Caribbean Studies 40, 3–14.
Lugo, A.E., Carlo, T.A., Wunderle Jr, J.M., 2012. Natural mixing of species: novel plant-animal communities on Caribbean Islands. Anim. Conserv.
MacMillan, R., Sun, L., Taylor, S.W., 2022. Modeling individual extended attack wildfire suppression expenditures in British Columbia. For. Sci. 68, 376–388, 2022.
Marcano Vega, H. (2019). Los bosques de Puerto Rico, 2014. Boletín de Recursos SRS–224. Asheville, NC: Departamento de Agricultura de los Estados Unidos Servicio Forestal, Estación de Investigación del Sur. 90 p. (⟨https://research.fs.usda.gov/treesearch/58591⟩).
Marcano-Vega, H. (2023). Forests of Puerto Rico, 2019. Resource Update FS-461. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 4 p. https://doi.org/10.2737/FS-RU-461.
Marcano-Vega, H., Brandeis, T.J., and Turner, J.A. (2015). Los bosques de Puerto Rico, 2009. Resour. Bull. SRS–202. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 115 p. (⟨https://research.fs.usda.gov/treesearch/47948⟩).
Marcano-Vega, H., and Williamson, J.R. (2017) Forests of US Virgin Islands, 2014. Resource Update FS–144. Asheville, NC: US Department of Agriculture Forest Service, Southern Research Station, 144, 1-4.
Moisen, G.G., Freeman, E.A., Blackard, J.A., Frescino, T.S., Zimmermann, N.E., Edwards, T.C., 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. Ecol. Model. 199 (2), 176–187.
Petter, G., Kreft, H., Ong, Y., Zotz, G., Cabral, J.S., 2021. Modelling the long-term dynamics of tropical forests: From leaf traits to whole-tree growth patterns. Ecol. Model. 460, 109735.
Phillips, P.D., Yasman, I., Brash, T.E., van Gardingen, P.R., 2002. Grouping tree species for analysis of forest data in Kalimantan (Indonesian Borneo). For. Ecol. Manag. 157 (1–3), 205–216.
PRISM Climate Group, Oregon State University, ⟨https://prism.oregonstate.edu⟩, Data Created 2005.
Ramos-Scharrón, C.E., Arima, E.Y., Hughes, K.S., 2022. An assessment of the spatial distribution of shallow landslides induced by Hurricane María in Puerto Rico. Phys. Geogr. 43 (2), 63–191. https://doi.org/10.1080/02723646.2020.1801121.
Russell, M., Weiskittel, A.R., Kershaw, J.A., 2014. Comparing strategies for modeling individual-tree height and height-to-crown base increment in mixed-species Acadian forests of northeastern North America. Eur. J. For. Res. 133, 1121–1135.
United States Department of Agriculture Forest Service (2023) National Core Field Guide, Version 9.3. 525 pp. Available on-line at ⟨https://research.fs.usda.gov/understory/nationwide-forest-inventory-field-guide⟩. Accessed July 19, 2023.
Uriarte, M., Thompson, J., Zimmerman, J.K., 2019. Hurricane María tripled stem breaks and doubled tree mortality relative to other major storms. Nat. Commun. 10, 1362. https://doi.org/10.1038/s41467-019-09319-2.
Vanclay, J.K., 1991. Aggregating tree species to develop diameter increment equations for tropical rainforests. For. Ecol. Manag. 42, 143–168.
Vanclay, J.K., 1995. Synthesis: growth models for tropical forests: a synthesis of models and methods get access arrow. For. Sci. 41, 7–42.
Yang, S., Brandeis, T.J., Helmer, E.H., Oatham, M.P., Heartsill-Scalley, T., Marcano-Vega, H., 2022. Characterizing height-diameter relationships for Caribbean trees using mixed-effects random forest algorithm. For. Ecol. Manag. 524, 120507.
Zhang, J., Heartsill-Scalley, T., Bras, R.L., 2022. Forest structure and composition are critical to hurricane mortality. Forests 13 (2), 202.
Zimmerman, J.K., Wood, T.E., González, G., Ramirez, A., Silver, W.L., Uriarte, M., Willig, M.R., Waide, R.B., Lugo, A.E., 2021. Disturbance and resilience in the Luquillo Experimental Forest. Biol. Conserv. 253. https://doi.org/10.1016/j.biocon.2020.108891.