Forest Ecology and Management 524 (2022) 120507

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Characterizing height-diameter relationships for Caribbean trees using mixed-effects random forest algorithm

Sheng-I Yang [a,*], Thomas J. Brandeis [b], Eileen H. Helmer [c], Michael P. Oatham [d], Tamara Heartsill-Scalley [c], Humfredo Marcano-Vega [e]

[a] Department of Forestry, Wildlife and Fisheries, University of Tennessee, 427 Plant Biotechnology Building, Knoxville, TN 37996, 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] Department of Life Sciences, University of the West Indies St. Augustine, St Augustine Circular, St Augustine, Trinidad and Tobago
[e] 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
Trinidad and Tobago
Tree allometry
Environmental variables
Machine learning

ABSTRACT

Accurately characterizing the relationships between tree height and diameter at breast height (DBH) is vital to quantify productivity and predict tree growth for forest management. This study makes use of data from forest inventories in the Caribbean islands of Puerto Rico, Trinidad and the U.S. Virgin Islands to compare the efficacy of parametric mixed models, random forest and mixed-effects random forest algorithms (MERF) for modeling tree species-specific H-D relationships. MERF as a semi-parametric approach is advantageous for handling high dimensional and complex datasets using the random forest algorithm, as well as to provide species-specific predictions with random effects in the mixed model framework. A total of 37,124 observations used in analyses were collected from 645 permanent sample plots across three Caribbean islands, including 309 species in Puerto Rico, 121 species in the U.S. Virgin Islands and 34 species in Trinidad.

Results showed that MERF can be used to reliably predict total tree height for diverse species, which produced more accurate predictions than the other two models. In addition to DBH, competition index was found to be the most important variable among all other stand and environmental variables in all islands. Adding climate-related variables can improve the prediction accuracy of total tree height in parametric mixed models, especially when sample size is small. Precipitation variables were selected most often, confirming the overall importance of precipitation to forest height. The results of this work will not only provide species-specific H-D relationship models for Caribbean trees, but also advance knowledge in modeling H-D relationships with broad application for mixed-species forests.

1. Introduction

Accurately characterizing the relationships between tree height and diameter at breast height (DBH) is vital to quantify productivity and predict tree growth, important components in forest management. In forest inventories, a combination of subsampling strategy and height-diameter (H-D) relationship modeling provides an efficient approach to estimate the per-unit-area stand variables of interest (Yang and Burkhart, 2020a). Due to the importance of understanding H-D relationships for monitoring forest dynamics and succession, as well as for estimating biomass and carbon sequestration, various H-D relationship models have been extensively examined and proposed for different forest types in different regions (e.g., Burkhart and Tomé, 2012; Temesgen et al., 2006; Thomas et al., 2015).

However, developing species-specific H-D relationship models is challenging in species-abundant forest ecosystems (e.g., tropical forests) due to small sample size of uncommon species (Adame et al., 2014, Lam et al., 2017). Past modeling efforts were focused on aggregating data based on similar tree characteristics (e.g., taxonomic rank, tree size), and functional or ecological traits (e.g., shade tolerance, successional plant functional types) (e.g., Vanclay, 1991, Phillips et al., 2002, Zhang et al., 2022). Alternatively, mixed-effects models with species as a

* Corresponding author.
E-mail addresses: syang47@utk.edu (S.-I. Yang), thomas.j.brandeis@usda.gov (T.J. Brandeis), eileen.helmer@usda.gov (E.H. Helmer), Mike.Oatham@sta.uwi.edu (M.P. Oatham), tamara.heartsill-scalley@usda.gov (T. Heartsill-Scalley), humfredo.marcano@usda.gov (H. Marcano-Vega).

https://doi.org/10.1016/j.foreco.2022.120507
Received 11 July 2022; Received in revised form 25 August 2022; Accepted 29 August 2022
Available online 18 September 2022

random effect were proposed to provide species-specific equations without data grouping (e.g., Russell et al., 2014, Lam et al., 2017, Kuehne et al., 2020). Mixed-effects models are composed of two components: fixed and random effects. Fixed effects represent the mean response while random effects provide the adjusted response (e.g., species-specific in this case) (Burkhart and Tomé, 2012). Although mixed models have proven useful for producing reliable predictions of total tree height for diverse species, determining appropriate model form and initial parameter values can become challenging as the number of covariates increases, such as when stand and environmental variables are included in the model.

In machine learning, the random forest algorithm as an ensemble learning method is efficient for highly dimensional and nonlinear data because subsets of observations and variables are randomly selected to grow decision trees (Cutler et al., 2007). Although adding species as a categorical variable in the algorithm can be used to build species-specific models, one major limitation is that high cardinality of the categorical variable will greatly increase computation time because the number of possible splits grows nonlinearly with the cardinality (Zhu, 2020). High cardinality can also lead to overfitting the random forest models. To build species-specific models with multiple variables included, mixed-effects random forest (MERF) model (Hajjem et al., 2014) provides a possible solution. MERF is a combination of the random forest algorithm and a mixed model, which include both fixed and random effects. Specifically, the random forest algorithm is used to estimate the fixed effects, but the random component in the model is retained. By implementing random forest algorithm, modelers do not need to specify a model from or identify proper initial values in model fitting. In the meanwhile, species-specific predictions can be made by MERF because the random component is included. To our understanding, the efficacy of MERF on predicting the total heights of diverse species has not been explicitly examined in forestry literature. This study makes use of data from forest inventories in the Caribbean islands of Puerto Rico, Trinidad and the U.S. Virgin Islands to explore the usefulness of parametric mixed models, random forest algorithm and MERF for developing tree species-specific H-D relationship models.

Caribbean islands encompass various forest ecosystems, ranging from coastal mangrove forests, freshwater swamp forests, subtropical dry forests, moist forests to lower montane wet and rain forests in a complex social and environmental setting (Helmer et al., 2002; López-Marrero and Heartsill-Scalley, 2012, Brandeis and Turner, 2013a, 2013b; Heartsill-Scalley and Gonzalez, 2016). Caribbean forests are highly diverse, species rich, and contain many endemic species (Santiago-Valentin and Olmstead, 2004, Lugo et al., 2012a). Beginning in the early 16th century with European colonization throughout the Caribbean, most of the forests on these islands were cleared for export-oriented agricultural use or establishment of urban areas. In the late 20th century, the islands experienced large-scale agricultural abandonment due to the loss of export markets for agricultural products and economic development policy changes (Rudel et al., 2000). The result has been a gradual transition of some of these deforested areas to secondary forests, resulting in a mosaic of land cover types on the islands, often highly fragmented and frequently experiencing natural and anthropogenic disturbances (Brandeis, 2003; Lugo et al., 2012a). Some of these secondary forests have been described as “novel” forest communities due to the naturalization of many non-native tree species introduced for timber or agricultural purposes (Brandeis et al., 2009; Lugo et al., 2012b). In the specific case of Puerto Rico, forest inventories at the national level show that non-native tree species account for 38 percent of the total live aboveground carbon in trees within forest areas (Lugo et al., 2022). In the U.S. Virgin Islands, the introduced trees *Leucaena leucocephala* have been consistently recorded as dominant in secondary forests, accounting for the species with highest biomass storage (Brandeis and Turner, 2013b, Marcano-Vega and Williamson, 2017).

The islands of Trinidad and Tobago are continental in origin so despite current macro-climatic conditions being more similar to the insular Caribbean, the forests are more similar biogeographically to the lowland forests found on the mainland of northern South America (Clubbe and Jhilmit, 1992). The Forestry Division in Trinidad and Tobago was established in the early 1900s with initial forest management goals of timber production, with the main production systems employed being a selective cutting systems and plantations primarily of teak (*Tectona grandis*) and pine species (*Pinus* spp.) (Fairhead and Leach, 2002). Almost 40 % of the land area of Trinidad and Tobago was under forest cover by 1962 when the islands gained independence (Gibbes et al., 2009). In the 1980s, in response to increasing costs of plantation formation, a tightly managed polycyclic system was initiated in the natural forests of southeast Trinidad (the Periodic Block System-PBS) instead of further conversion of natural forest to plantation (Fairhead and Leach, 2002). By 2000 the only forest still producing native timber was managed under the PBS, and all other native forest managed under other systems were closed to timber extraction due to lack of saleable timber (Fairhead and Leach, 2002), which has prompted the Forestry Division to consider plans for PBS management of all natural forest reserves when the native forests have recovered enough for timber extraction (Roberts pers comm).

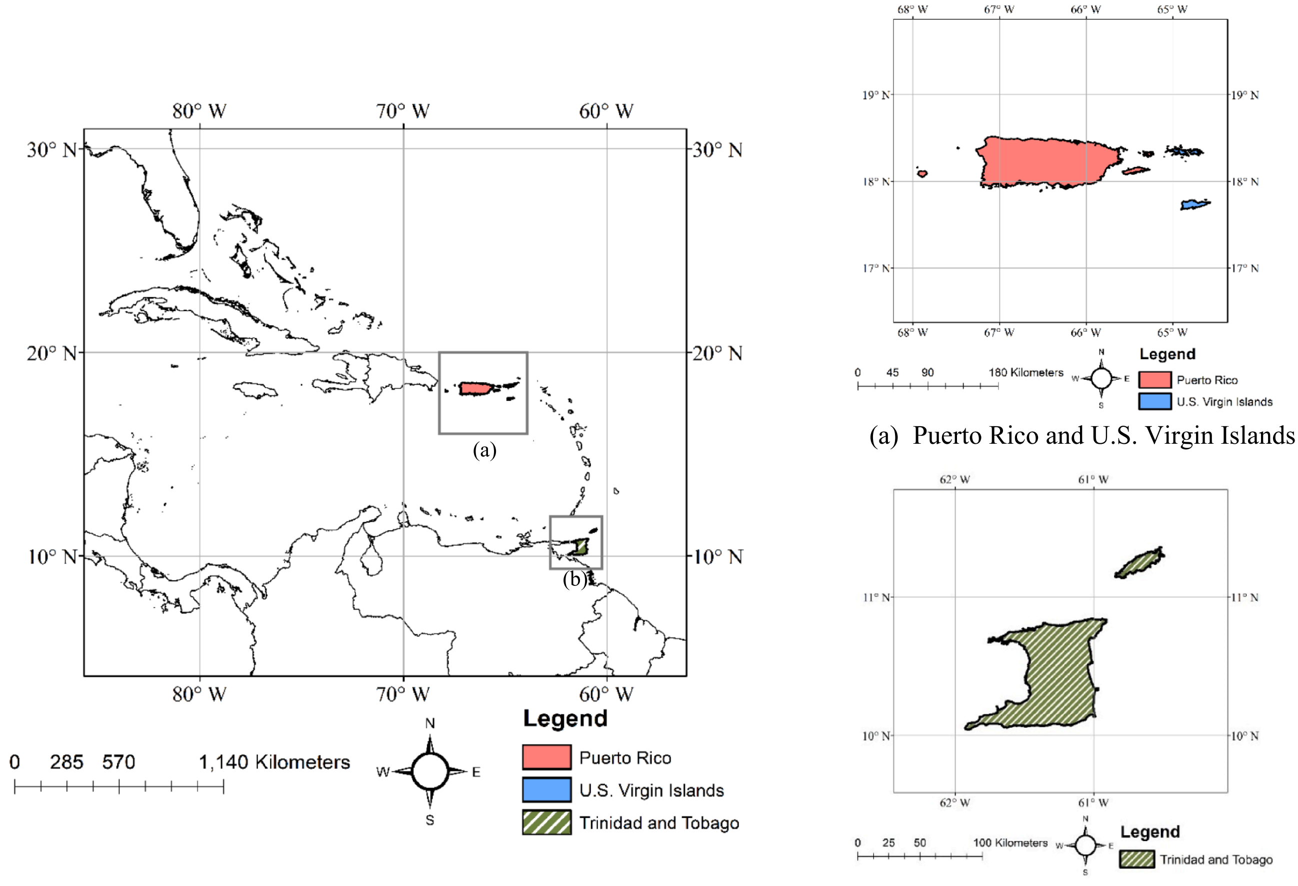
While greatly changed from their pre-European colonization structure and species composition, Caribbean forests continue to provide local populations with timber and valuable ecosystem services like watershed protection, carbon sequestration, soil erosion control, hurricane impact mitigation, biodiversity maintenance, recreation opportunities, and other non-material cultural and social benefits. Quantifying the ecosystem services of Caribbean forests requires the ability to accurately quantify the trees contained within them. Allometric models describing the relationships between different dimensions of individual trees allow estimation of carbon storage and tree canopy heights, and also serve to reconstruct those values when forests are damaged by natural disturbances such as Caribbean hurricanes. There have been efforts to develop the allometric equations needed to estimate aboveground biomass and carbon from total height and DBH measurements (Scatena et al., 1996; Weaver and Gillespie, 1992; Brandeis and Oswalt, 2007), and the relationships between crown width and DBH (Brandeis and Randolph, 2010; Brandeis et al., 2009). Early efforts to model H-D models for Caribbean species from the forest inventory data collected on Puerto Rico were limited by the small sizes available at the early stages of data collection and did not include environmental or climatic factors in the models (Brandeis and Randolph, 2010; Brandeis et al., 2009).

Therefore, the objective of this study was to develop reliable statistical models to characterize H-D relationships for Caribbean trees. This was accomplished by comparing the performance of MERF with the original random forest algorithm and parametric mixed models. In addition to DBH, a total of 27 tree, stand and environmental variables were included in model building. Tree data were collected from the U.S. national forest inventory in Puerto Rico and U.S. Virgin Islands, as well as permanent sample plots on the island of Trinidad in the nation of Trinidad and Tobago (Fig. 1). The results of this work will not only provide species-specific H-D relationship models for Caribbean trees, but also advance knowledge in modeling H-D relationships with broad application for mixed-species forests.

2. Materials and methodology

2.1. Study area and tree data

A total of 37,124 observations used in analyses were collected from 645 permanent sample plots across three islands, including 309 species in Puerto Rico, 121 species in U.S. Virgin Islands and 34 species in Trinidad (Table 1). Data for Trinidad and Tobago were collected from 27 permanent sample plots in the Victoria-Mayaro Forest Reserve (VMFR). These plots were a sub-set of a larger network of plots established and were chosen for lack of harvesting activity or fire disturbance.



(a) Puerto Rico and U.S. Virgin Islands

(b) Trinidad and Tobago

Fig. 1. Map of the Caribbean islands of Puerto Rico, U.S. Virgin Islands, and Trinidad and Tobago.

Table 1
Summary of tree data among the Caribbean islands of Puerto Rico (PR), U.S. Virgin Islands (VI), and Trinidad and Tobago (TT). N_{plot}, $N_{obs.}$ and $N_{sp.}$ represent total number of plots, observations and species for a given island. Mean, standard deviation (Std.), minimum (Min.) and maximum (Max.) of diameter at breast height (DBH, cm), total tree height (Total height, m) and competition index (CI) are given.

				DBH (cm)				Total height (m)				CI			
Island	N_{plot}	$N_{obs.}$	$N_{sp.}$	Mean	Std.	Min.	Max.	Mean	Std.	Min.	Max.	Mean	Std.	Min.	Max.
PR	517	27,283	309	13.9	12.0	2.5	147.1	8.9	4.7	0.3	35.1	1.3	1.0	0.1	13.9
VI	101	6,083	121	9.6	8.4	2.5	118.6	6.5	2.7	0.9	21.3	1.3	1.0	0.1	10.3
TT	27	3,758	34	31.5	14.0	10.6	124.1	20.1	6.6	2.3	50.0	2.6	1.2	0.9	11.0

2.2. Puerto Rico and U.S. Virgin islands

Tree data from Puerto Rico and U.S. Virgin Islands were obtained from the U.S. national forest inventory database established and maintained by the USDA Forest Service's Forest Inventory and Analysis (FIA) program. This program continuously monitors forest resources and health in the United States and its associated commonwealths and territories. Permanent sample plots were first established on mainland Puerto Rico in 1980 (Birdsey and Weaver, 1982). The sampling design was modified and expanded in 2001 to better capture the full range of diverse forest types and now includes the outlying Puerto Rican islands of Culebra, Mona and Vieques, as well as the three U.S. Virgin Islands of St. Croix, St. John and St. Thomas (Brandeis, 2003). The long-term repeated measurements provide a sound database for various research objectives, such as assessing the relationships between environmental factors and species composition (e.g., Brandeis et al., 2009, Franklin et al., 2015).

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. The FIA program defines forest as land with at least 10 percent canopy cover of trees at least 30 cm in height within a minimum area of 0.4 ha (Brandeis et al., 2006). 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 distinct condition, the plot would be mapped by condition class even though the boundaries cut through subplots or microplots. Each permanent plot was revisited every 5 years and remeasured if forested. Each tree is individually identified, relocated and remeasured. Tree mortality or removal is accounted for, as is the recruitment of new trees on the plot. The details of sampling design, plot layout, and estimation procedures can be found in Bechtold and Patterson (2005).

2.3. Trinidad and Tobago

Due to lack of information on timber volumes and species-specific growth rates needed to set cutting cycle lengths, a network of one-hectare permanent sample plots was established throughout Trinidad but most intensively in the forests where the PBS was the main management system. The Trinidad data were taken from information gathered by the Forest Resource Inventory and Management (FRIM) Section of the Forestry Division of the Government of the Republic of Trinidad and Tobago. In all, the FRIM established 144 one-hectare plots (100 m × 100 m) across the island of Trinidad for the purpose of monitoring forest timber resource turnover and growth. The majority of the plots were established in the southeast of the country in the Victoria-Mayaro Forest Reserve (VMFR) where native timber extraction was most active. Most plots were established in 1983–84 (109 plots) and the last were established in 1998. Plots have been inventoried up to 9 times, at intervals of approximately 3 years. Most tree species were measured down to 10 cm DBH with some (non commercial species such as *Cecropia peltata*, or *Pachira insignis*) measured to 20 cm. Palm species were not measured. Data initially gathered for each stem encountered at establishment of the plot included species identification (common names were used and some species, such as members of the *Lauraceae*, were lumped under one common name), DBH, and height to the top of the crown. Each tree was permanently marked using paint markings with an individual stem number which was sufficient to relocate the stem at the next inventory. At subsequest inventories, previously recorded stems only had their DBH recorded and any new stems that had attained the minium DBH size for measurement had species ID, height and DBH recorded. Data were stored in paper data forms up to 1996 when the records were progressively digitized by FRIM. The FRIM digitized records only included DBH and species ID, but a subset of the PSPs in the VMFR were digitized in full (including initial heights) for another project (Oatham and Ramnarine, 2006). A subset of the data was used in this study.

2.4. Tree and stand variables

In addition to DBH and total tree height, competition index (CI) was calculated for each tree, which was defined as the ratio of the subject tree DBH to the sampling-unit-level quadratic mean diameter (QMD). Stand characteristics, including QMD, number of trees per ha (TPH), basal area per ha (BA), were computed from all live trees on a plot, or on a “condition class” rather than a “plot” in the case of the FIA plots. For example, two distinct condition classes on a FIA plot would have two separate values for a given stand characteristics (i.e., two observations). In other words, each plot’s condition was treated as an individual sample of trees. In Puerto Rico and U.S. Virgin Islands, trees damaged by hurricanes were recorded in the database, which were removed when computing stand characterisitcs.

2.5. Environmental variables

We tested 23 potential covariates of H-D relationships representing climate, topography and surficial geology. Climate metrics included 19 bioclimatic variables from Worldclim 2.1 climate data, which are gridded to a spatial resolution of ∼ 1-km (Fick and Hijmans, 2017). The bioclimatic variables were calculated from monthly normals for the period 1970–2000 and included average annual temperature and total precipitation as well as metrics like precipitation over the driest quarter of the year, or temperature of the coldest quarter. Three topographic variables (elevation, aspect and slope) came from Shuttle Radar Topography Mission 90-m digital elevation model data (Jarvis et al., 2008), a spatial resolution compatible with the plot sizes. Geology was obtained from a Caribbean-wide map of surficial geology (French and Schenk, 2004), which is a categorical variable. Detailed descriptions of the stand, bioclimatic and topographic variables are given in Table 2.

2.6. Statistical methods

Three methods were applied in this study. In methods 1 and 2, total tree height was predicted using the random forest algorithm and MERF, respectively, whereas in method 3, predictions were made with parametric mixed models with and without environmental variables included.

2.6.1. Method 1 - random forest algorithm (RF)

Random forest (RF) was learnt from an ensemble of 500 regression trees. DBH and all variables mentioned before were included as predictors and total tree height served as the response variable. The RF

Table 2
Description and average of the stand and enviromental variables (standard deviation is given in parentheses) among the Caribbean islands of Puerto Rico (PR), U.S. Virgin Islands (VI) and Trinidad and Tobago (TT).

Category	Label	Description	PR	VI	TT
Stand	QMD	Quadratic mean diameter (cm)	13.3 (11.1)	8.5 (4.8)	12.9 (2.7)
	TPH	Trees per ha (stem/ha)	2395.1 (1950.3)	4107.5 (2601.7)	213.0 (29.2)
	BA	Basal area per ha (m^2/ha)	19.6 (16.3)	17.4 (9.5)	27.8 (9.4)
Temp.	Ann_t	Annual Mean Temperature (°C)	24.3 (1.6)	26.1 (0.4)	26.2 (0.0)
	Diurnal_range	Mean Diurnal Range (Mean of monthly (max temp - min temp)) (°C)	10.5 (1.9)	7.0 (0.2)	7.4 (0.1)
	Iso_therm	Isothermality (MDR/TAR) (×100) (unitless ratio)	74.3 (3.3)	69.1 (0.8)	79.9 (0.3)
	Season_t	Temperature Seasonality (standard deviation × 100) (°C × 100)	131.9 (5.5)	122.1 (1.9)	57.3 (0.5)
	Warm_mon_tmax	Max Temperature of Warmest Month (°C)	31.0 (1.9)	31.1 (0.4)	30.6 (0.0)
	Cold_mon_tmin	Min Temperature of Coldest Month (°C)	16.9 (2.0)	20.9 (0.4)	21.3 (0.1)
	Ann_t_range	Temperature Annual Range (MaxTWM - MinTCM) (°C)	14.0 (2.0)	10.2 (0.2)	9.3 (0.1)
	Wet_qtr_t	Mean Temperature of Wettest Quarter (°C)	25.4 (1.5)	26.7 (0.4)	26.4 (0.1)
	Dry_qtr_t	Mean Temperature of Driest Quarter (°C)	22.7 (1.7)	24.6 (0.4)	25.9 (0.0)
	Warm_qtr_t	Mean Temperature of Warmest Quarter (°C)	25.7 (1.6)	27.5 (0.5)	26.7 (0.0)
	Cold_qtr_t	Mean Temperature of Coldest Quarter (°C)	22.6 (1.7)	24.6 (0.4)	25.4 (0.0)
Precip.	Ann_ppt	Annual Precipitation (mm)	1686.1 (394.0)	1125.9 (62.9)	2127.3 (54.3)
	Wet_mon_ppt	Precipitation of Wettest Month (mm)	222.3 (50.0)	146.6 (5.9)	278.8 (14.8)
	Dry_mon_ppt	Precipitation of Driest Month (mm)	61.5 (18.7)	43.5 (3.3)	62.3 (7.8)
	Season_ppt	Precipitation Seasonality (Coefficient of Variation) (unitless ratio)	40.7 (9.8)	39.0 (2.8)	44.7 (2.1)
	Wet_qtr_ppt	Precipitation of Wettest Quarter (mm)	621.8 (143.8)	426.2 (18.8)	764.4 (38.4)
	Dry_qtr_ppt	Precipitation of Driest Quarter (mm)	212.7 (67.2)	154.5 (17.2)	215.7 (14.2)
	Warm_qtr_ppt	Precipitation of Warmest Quarter (mm)	517.8 (142.5)	287.9 (45.0)	679.7 (45.9)
	Cold_qtr_ppt	Precipitation of Coldest Quarter (mm)	214.9 (67.4)	176.8 (27.1)	394.6 (37.8)
Topo.	Elev	Elevation (m)	257.1 (246.7)	103.6 (75.8)	57.9 (10.3)
	Aspect	Aspect (°)	180.9 (106.6)	175.7 (98.1)	173.1 (108.8)
	Slope	Slope (%)	15.2 (12.3)	15.2 (13.1)	2.7 (2.4)

model was built using the R function “ranger” with all hyperparameters kept at the default values (Wright and Ziegler, 2017). “Ranger” function was proposed to be a faster and more memory efficient implementation of random forests to analyze data than other commonly used random forest packages in R (Wright and Ziegler, 2017).

2.6.2. Method 2 - mixed-effects random forest (MERF)

Hajjem et al. (2014) proposed MERF where the random forest algorithm was incorporated in mixed-effects model framework. Similar to mixed models, MERF includes both fixed- and random-effects components. For the j^{th} observation given the i^{th} species, the total tree height (h) was predicted as:

$$h_{ij} = f\left(x_{ij}\right) + b_i + \epsilon_{ij} \quad (1)$$

where $f(x)$ is a general nonlinear function of predictors (x) as fixed-effects, b is species-specific random effect, and ϵ is error term. In this study, we assumed that b follows a normal distribution with mean zero and variance σ^2_{sp} (i.e., $b\ N(0, \sigma^2_{sp})$), and ϵ follows a normal distribution with mean zero and variance σ^2 (i.e., $\epsilon\ N(0, \sigma^2)$). In MERF, parameters were estimated through an iterative procedure. At each iteration, the first step was to build random forest with the transformed response (h^*_{ij}) data where the current estimate of the random effect component is removed from the original response (i.e., $h^*_{ij} = h_{ij} - b_i$). Secondly, the residuals calculated from the fitted and observed responses were used to estimate the random effects. Then, the estimates were updated to contruct a new transformed response data in the first step. The two steps were repeated until the parameter convergence criterion met (i.e., the generalized log-likelihood was below a predefined threshold). More details on steps can be found in Hajjem et al. (2014). In this study, the source code of the R package “MixRF” (Wang and Chen, 2016) was modified to accommodate the data structure and variables of this work, and the default random forest function in the package was replaced with the “ranger” function (Wright and Ziegler, 2017) to run the random forest algorithm.

2.6.3. Method 3 – mixed models with and without environmental variables as predictors

Given that parametric allometric equations are still commonly used in practice, we further built the parametric mixed-models with and without enviromental variables as predictors. In the preliminary analysis, it was found that the models with the base form of Pearl and Reed (1920) provided the most reliable predictions for Puerto Rican and U.S. Virgin Islands’ forests, which is given as:

$$h_{ij} = \frac{(\beta_0 + b_i)}{1 + \beta_1 e^{\beta_2 DBH_{ij}} + \sum x_{ij}\beta_k} + \epsilon_{ij}$$

where β_0, β_1, β_2 and β_k are fixed-effects, e is Euler’s number, $\sum$ is summation, other variables and symbols are the same as above. For Trinidad, the base model proposed by Wykoff et al. (1982) performed better than other candidates. The extended version of the model was written as:

$$h_{ij} = e^{(\beta_0 + b_i) - \frac{\sum x_{ij}\beta_k}{(DBH+1)}} + \epsilon_{ij}$$

where all variables and symbols have been defined as above. Both models included a linear combination of predictors (i.e., $\sum x_{ij}\beta_k$), which were obtained from the best MERF model in each island. For example, the best MERF model for Puerto Rican trees includes six predictors: DBH, three stand variables and two environmental variables. To build the mixed model with enviromental variables as predictors, all six variables were included, whereas only DBH and three stand variables were included in fitting the mixed model without environmental variables. To determine the initial values in parameter estimation, the models with only fixed-effects were fitted by the “nlsLM” function in R (Elzhov et al., 2022). All model coefficients were then estimated using the “nlme” function in R (Pinheiro et al., 2021).

2.6.4. Model building

The full model with all predictors (DBH + all tree, stand and environmental variables) was first fitted using methods 1 and 2, respectively. All predictors were then arranged in descending order of the importance score calculated using the permutation approach similar to the procedure used in Sabatia and Burkhart (2014). The importance score was used to assess if a variable has a positive impact on the prediction performance through breaking the relationship between the predictor of interest and the response variable. In the permutation approach, the importance score of a predictor was calcaluted by averaging the differences of prediction accuracy between the reference and permutated-predictor trees. For example, predictor A is the variable of interest. The reference tree is grown with the orginial set of predictors and response, and the prediction accuracy of the tree is calculated using out-of-bag observations. Then, a new dataset is constructed by randomly shuffling the values of the predictor A and keeping pairs of all other predictors and response unchanged. Another tree is built using the new dataset, and the prediction accuracy of the permutated-predictor tree is computed as the reference tree. Thus, the greater score (i.e., larger difference) indicates the predictor is more important. More details of steps can be found in Wright et al. (2016). The importance score was obtained from the “ranger” function (Wright and Ziegler, 2017).

After the full model was built, sub-models were then sequentially fitted with variables ranked by the importance score. The procedure began with the model containing the two most important variables, and then the variable with the highest importance score was added at every step in the sequence until all variables were added in the model. A total of 27 sub-models were built. For a given method and island, the performance of all sub-models (i.e., models with varying number of predictors) were evaluated using psuedo R^2, which was calculated from out-of-bag sample. Higher psuedo R^2 implies more variability in total tree height is explained by the model. The final model was selected based on psuedo R^2 and parsimony (i.e., number of predictors). Then, the variables selected in the final MERF model were used to build parametric mixed models with and without environmental variables as predictors.

2.6.5. Evaluation of prediction accuracy

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

$$MB = \frac{\sum Res}{n}$$

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

where n is total number of observations, and Res is prediction residual, which was defined as:

$$Res = h - \widehat{h}$$

where h and $\widehat{h}$ are the observed and predicted total tree heights in m, respectively.

In addition to the overall MB and RMSE, the prediction accuracy on different tree size classes was evaluated. Specifically, sample trees in each island were divided into three size classes based on DBH: small- (DBH $\leq$ 1/3% quantile of DBH), medium- (1/3% quantile of DBH $<$ DBH $\leq$ 2/3% quantile of DBH) and large-sized (2/3% quantile of DBH $<$ DBH) classes. MB and RMSE were calculated for each size class.

3. Results and discussion

3.1. Comparison of prediction accuracy among models, islands and tree sizes

As shown in Fig. 2, RF and MERF produced similar pseudo R^2 values for a given island and number of predictors. Pseudo R^2 increased with increasing number of predictors, but no obvious increase or decrease was found after about six-eight variables were added in the models (see Fig. 2). The trend is consistent with the finding in Weiskittel et al. (2011), but different from the report of Sabatia and Burkhart (2014). As pointed out by Sabatia and Burkhart (2014), the differences may be because smaller numbers of observations were used in model training in their study. For the final models of RF and MERF, pseudo R^2 values for Puerto Rico and U.S. Virgin Islands were around 0.81, while pseudo R^2 for Trinidad was estimated at about 0.30.

Overall, the total tree heights were predicted without bias by RF, MERF and parametric mixed models in all islands (see MB in Table 3). However, the precision of the predictions was improved when the MERF models were implemented (see RMSE in Table 3). As Table 3 shows, the magnitude of improvement varied among islands. Trinidad yielded the largest difference of RMSE between MERF and other models compared to Puerto Rico and U.S. Virgin Islands. As discussed in the methods section, the sample trees were divided into three size classes by DBH to further examine the predictability of the models. Both RF and MERF provided unbiased predictions of all size classes for Puerto Rican and U. S. Virgin Islands trees, while the predictions of the small and large trees in Trinidad were slightly biased. The poorer predictions in Trinidad may be due to the smaller sample size or damaged trees included in analyses. No damage code is recorded in the Trinidad data. Mixed-models produced slightly different prediction accuracy of total tree heights for different size classes.

For all islands, total tree heights of the very tall trees were underpredicted (Figs. 3-5), while overpredictions were found for the very short trees in Puerto Rico and Trinidad. This result may stem from large variation in total tree height at both ends of the data. The prediction accuracy may be improved if more data points are included.

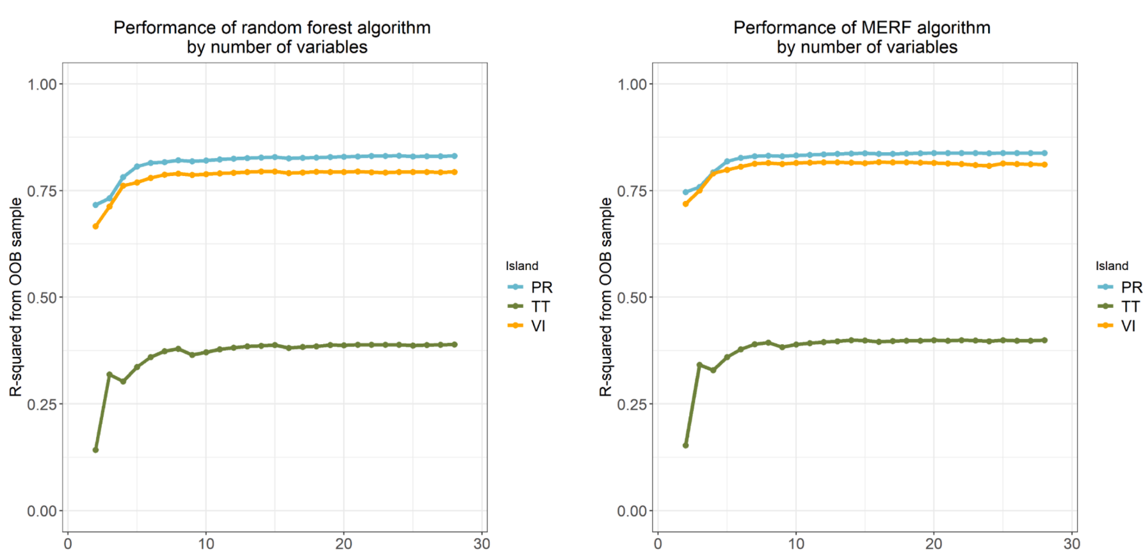
Table 3
Fit statistics of random forest (RF), mixed-effects random forest (MERF) and mixed-models with and without environmental variables as predictors (MM envir. vs MM no envir.) for the Caribbean islands of Puerto Rico (PR), U.S. Virgin Islands (VI), and Trinidad and Tobago (TT). Mean bias (MB, m) and root-mean-squared error (RMSE, m) for overall and three size classes are given.

		Overall			Small		
Statistics	Model	PR	VI	TT	PR	VI	TT
MB	RF	0.0	0.0	0.0	0.0	0.0	−0.2
	MERF	0.0	0.0	0.0	0.0	0.0	−0.2
	MM (envir.)	0.0	0.0	0.0	−0.2	−0.1	0.3
	MM (no envir.)	0.0	0.0	0.0	−0.2	−0.1	−0.9
RMSE	RF	2.0	1.2	5.2	1.0	0.9	4.9
	MERF	1.9	1.2	4.8	1.0	0.8	4.6
	MM (envir.)	2.2	1.4	5.1	1.1	0.9	4.7
	MM (no envir.)	2.2	1.4	5.7	1.1	0.9	4.9

		Medium			Large		
Statistics	Model	PR	VI	TT	PR	VI	TT
MB	RF	0.0	0.0	0.0	0.0	0.0	0.3
	MERF	0.0	0.0	−0.1	0.0	0.0	0.3
	MM (envir.)	0.3	0.1	−0.3	−0.1	0.0	0.0
	MM (no envir.)	0.3	0.1	−0.3	−0.1	0.0	1.1
RMSE	RF	1.9	1.1	4.9	2.8	1.6	5.8
	MERF	1.8	1.0	4.5	2.6	1.6	5.3
	MM (envir.)	2.0	1.2	4.8	3.0	1.9	5.7
	MM (no envir.)	2.0	1.2	5.2	3.1	1.9	6.7

3.2. Variables included in the final models

The variables included in the final models are given in Table 4, and the estimates of fixed and random effects are given in Tables 5 and 6. The estimate of -standard deviation σ for MERF was smaller than mixed-models, which implied that the variation of heights can be explained more by the MERF models (Table 6). Consistently among all islands, DBH and competition index (CI) had the greatest importance scores among all stand and environmental variables in the RF and MERF models (see Table 4). A measure of competition is usually included in distance-independent, individual tree models, which implies the relative competitive status of the tree in the population to compensate the lack of spatially explicit information (Burkhart and Tomé, 2012). Our results



(a) Random forest (b) Mixed-effects random forest (MERF)

Fig. 2. Performance of random forest (RF) and mixed-effects random forest (MERF) algorithms by number of variables among the Caribbean islands of Puerto Rico (PR), U.S. Virgin Islands (VI), and Trinidad and Tobago (TT).

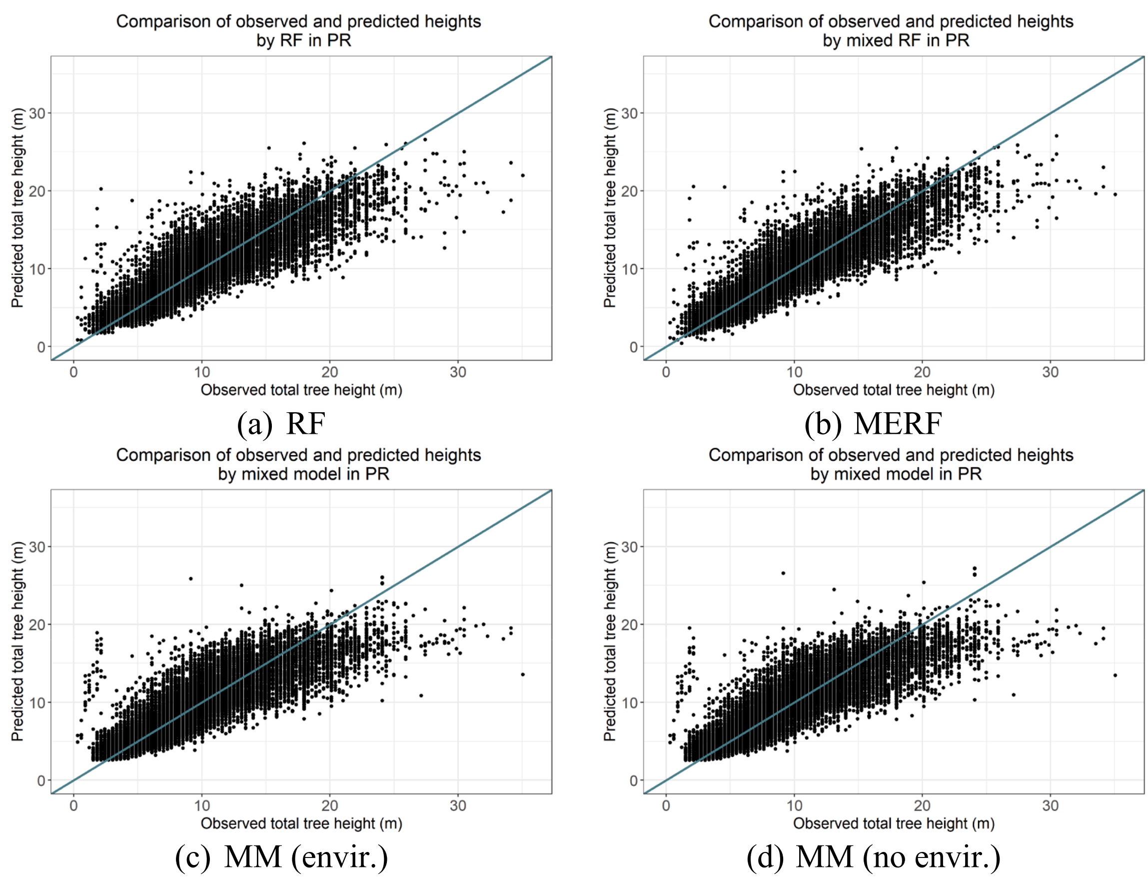


Fig. 3. Comparison of observed and predicted total tree heights in Puerto Rico among four methods, including (a) random forest, RF, (b) mixed-effects random forest, MERF, (c) mixed model with environmental variables as predictors, MM (envir.), and (d) mixed model without environmental variables as predictors, MM (no envir.).

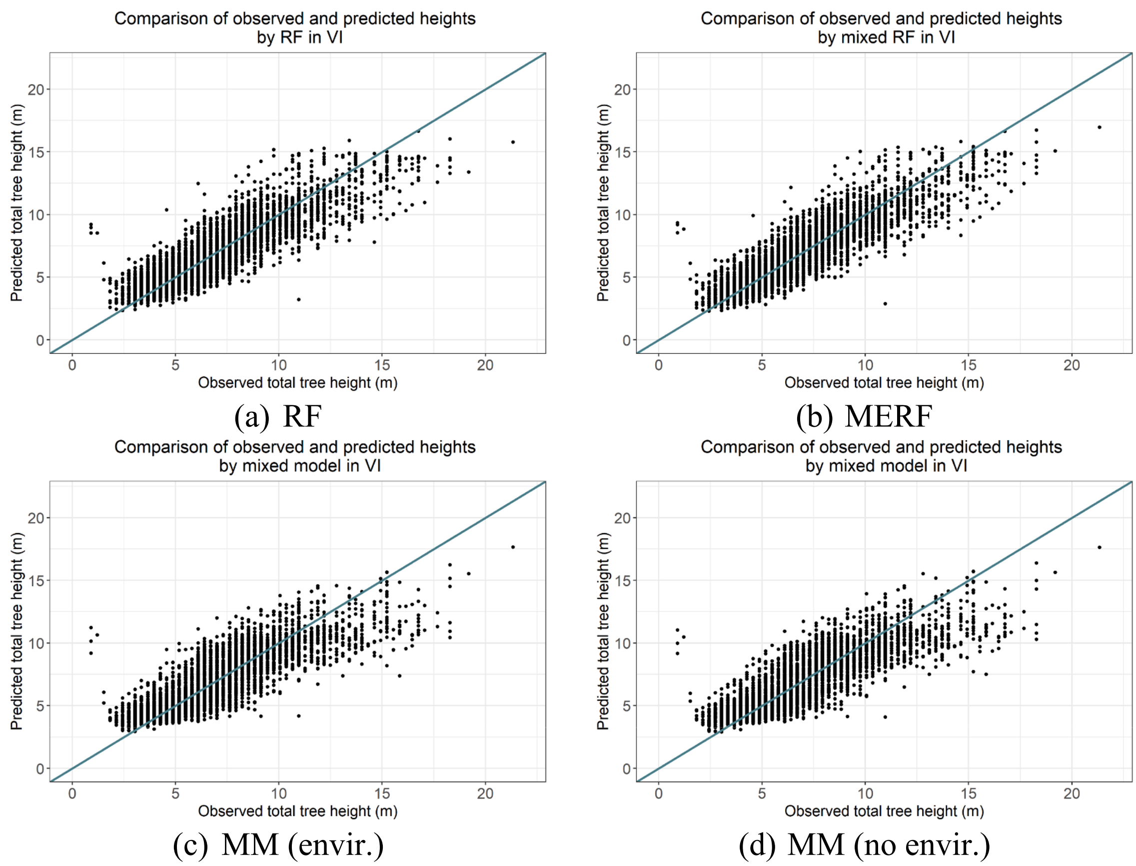


Fig. 4. Comparison of observed and predicted total tree heights in U.S. Virgin Islands among four methods, including (a) random forest, RF, (b) mixed-effects random forest, MERF, (c) mixed model with environmental variables as predictors, MM (envir.), and (d) mixed model without environmental variables as predictors, MM (no envir.).

confirm the validity of using the ratio of the subject tree DBH to the plot-level QMD as a distance-independent competition index suggested by Amateis et al. (1989). Examining the predictive power of other competition quantities, such as the sum of tree basal area greater than the subject tree (e.g., Lam et al., 2017; Temesgen et al., 2006), is suggested for further studies. In addition to DBH and CI, climatic variables, for example annual precipitation (Ann_ppt) or precipitation during the warmest quarter (Warm_qtr_ppt), were also important predictors associated with total tree height (see Table 4). Compared to temperature and topography, precipitation-related variables explained more variability of total tree height. The importance of precipitation in predicting H-D relationships agrees with global patterns in tropical forest height. All else equal, tropical forests receiving more annual precipitation tend to be taller (Yang et al., 2016). Based on fit statistics shown in Table 3, adding environmental variables in parametric mixed models can improve the prediction accuracy of total tree height, especially for small sample sizes. An example of this improved prediction accuracy was found for tree height in Trinidad. The predicted heights were closer to the observed heights when five environmental variables were included in the models (see Fig. 5).

Notably, height-based variables were not considered as predictors in model building, because we assumed that often information on tree height will not be available when implementing the models, which is a common situation in forest and natural resources data collection practices. Plot-level height information could be obtained from satellite imagery or remote sensing data that is becoming more readily available. In such cases, height related variables may be useful to improve the predictability of the models. For example, Lam et al. (2014) added the maximum tree height in mixed models as an indication of stage in stand development. This topic is worthwhile for further investigations.

3.3. Further discussion

As the results show, MERF can produce reliable species-specific predictions of total tree height without a specified form of fixed effects. Without the input of species in the model, RF can still achieve a certain level of prediction accuracy, but the model is not able to provide species-specific predictions, nor can the variation among species groups be assessed. In the mixed model framework, species as a high cardinality categorical variable can be included as a random effect, which compensates for the shortcoming of the random forest algorithm. Here we found that MERF produced more precise predictions than RF, especially when sample size was small (e.g., Trinidad in this study).

In this study, the same species groups were used in model training and testing, so the response variable (i.e., total tree height) was predicted using both estimated fixed and random effects. MERF models can be implemented to predict heights for new species or new populations. Predictions can be made with the fixed effects of the existing models, or with the models calibrated by a subset of observations from the new species or target populations (Burkhart and Tomé, 2012). The calibration procedure is similar to parametric mixed models (Sabatia and Burkhart, 2015). Notably, MERF as an ensemble algorithm may result in poor predictions outside the range of the training dataset (i.e., extrapolation) (Yang and Burkhart, 2020b), so model calibration is recommended when the MERF models are used for new species or populations. The impact of sample size and tree selection procedure for calibration on prediction accuracy is suggested for a follow-up investigation.

In preliminary analyses for the current study, we examined adding other random effects in the models, but the parameters failed to converge in most of the cases. Adding species as a random effect was found to produce the most stable model among all candidates examined.

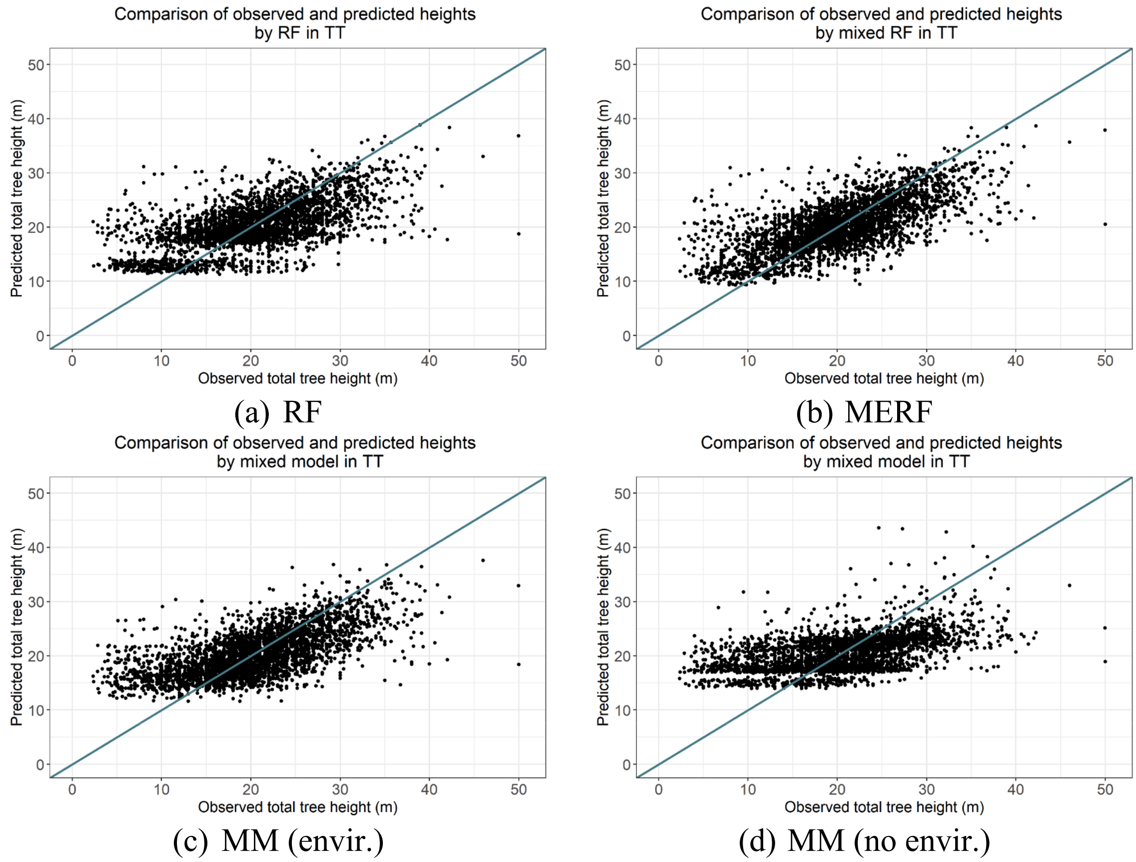


Fig. 5. Comparison of observed and predicted total tree heights in Trinidad and Tobago among four methods, including (a) random forest, RF, (b) mixed-effects random forest, MERF, (c) mixed model with environmental variables as predictors, MM (envir.), and (d) mixed model without environmental variables as predictors, MM (no envir.).

Table 4
Model variables selected for the Caribbean islands of Puerto Rico (PR), U.S. Virgin Islands (VI), and Trinidad and Tobago (TT) using random forest (RF) and mixed-effects random forest (MERF) methods. Variables included in models were sorted by importance score. Psuedo R^2 were calculated from out-of-bag samples of 500 regression trees.

	RF			MERF		
Island	Variable	Importance score	Psuedo R^2	Variable	Importance score	Psuedo R^2
PR	DBH	17.1	0.82	DBH	17.4	0.83
	CI	7.0		CI	6.1	
	QMD	5.7		QMD	4.9	
	Ann_ppt	3.5		Wet_qtr_t	2.6	
	BA	3.5		BA	2.6	
	Warm_qtr_ppt	2.7		Ann_t	2.5	
	TPH	2.3				
VI	DBH	5.1	0.79	DBH	5.5	0.81
	CI	2.6		CI	2.5	
	QMD	1.6		QMD	1.6	
	BA	1.1		BA	1.2	
	Cold_qtr_ppt	0.8		Cold_qtr_ppt	0.8	
	Wet_mon_ppt	0.8		TPH	0.7	
	TPH	0.7				
TT	DBH	15.0	0.38	DBH	14.7	0.39
	CI	11.9		CI	11.1	
	BA	8.2		BA	7.2	
	Warm_qtr_ppt	7.7		Warm_qtr_ppt	6.1	
	Wet_qtr_ppt	5.6		Cold_mon_tmin	5.0	
	Wet_mon_ppt	4.6		Wet_qtr_ppt	4.9	
	TPH	4.4		Cold_qtr_t	4.2	
	Diurnal_range	3.9		Diurnal_range	3.9	

Table 5
The estimates of standard deviations σ_{sp} and σ by mixed-effects random forest (MERF) and mixed-models with and without environmental variables as predictors (MM envir. vs MM no envir.) for the Caribbean islands of Puerto Rico (PR), U.S. Virgin Islands (VI) and Trinidad and Tobago (TT).

	MERF		MM (envir.)		MM (no envir.)	
	σ_{sp}	σ	σ_{sp}	σ	σ_{sp}	σ
PR	1.37	1.92	1.20	2.22	2.39	2.23
VI	0.54	1.21	0.69	1.40	0.92	1.40
TT	2.02	4.79	0.11	5.12	0.13	5.67

Table 6
Estimated fixed effect coefficients by parametric mixed-models with and without environmental variables as predictors (MM envir. vs MM no envir.) for the Caribbean islands of Puerto Rico (PR), U.S. Virgin Islands (VI) and Trinidad and Tobago (TT). Corresponding variables and p-values of the estimates are given.

Island	Model	Coefficients	Cor. Variable	Estimate	P-value
PR	MM (no envir.)	β_0	–	14.4917	<0.0001
		β_1	–	3.1382	<0.0001
		β_2	DBH	0.1169	<0.0001
		β_3	CI	−0.0034	0.15
		β_4	QMD	0.0001	0.76
		β_5	BA	−0.0024	<0.0001
	MM (envir.)	β_0	Intercept	7.5182	<0.0001
		β_1	–	1.5949	<0.0001
		β_2	DBH	0.1158	<0.0001
		β_3	CI	−0.0018	0.15
		β_4	QMD	0.0001	0.41
		β_5	Wet_qtr_t	−0.0013	<0.0001
		β_6	BA	−0.0130	<0.0001
		β_7	Ann_t	−0.0001	<0.0001
VI	MM (no envir.)	β_0	Intercept	8.6573	<0.0001
		β_1	–	1.8659	<0.0001
		β_2	DBH	0.1643	<0.0001
		β_3	CI	−0.0188	<0.0001
		β_4	QMD	0.0026	0.17
		β_5	BA	−0.0098	<0.0001
		β_6	TPH	0.00002	<0.0001
	MM (envir.)	β_0	Intercept	6.4676	<0.0001
		β_1	–	1.3717	<0.0001
		β_2	DBH	0.1696	<0.0001
		β_3	CI	−0.0187	<0.0001
		β_4	QMD	0.0020	0.16
		β_5	BA	−0.0068	<0.0001
		β_6	TPH	−0.0015	<0.0001
		β_7	Cold_qtr_ppt	0.00001	<0.0001
TT	MM (no envir.)	β_0	Intercept	2.6651	<0.0001
		β_1	CI	−0.0920	<0.0001
		β_2	BA	−0.1156	<0.0001
	MM (envir.)	β_0	Intercept	−21.3193	<0.0001
		β_1	CI	−0.0465	<0.0001
		β_2	BA	−0.0043	<0.0001
		β_3	Warm_qtr_ppt	0.0013	<0.0001
		β_4	Cold_mon_tmin	0.0061	<0.0001
		β_5	Wet_qtr_ppt	−0.0002	0.53
		β_6	Cold_qtr_t	−1.0053	<0.0001
		β_7	Diurnal_range	0.9123	<0.0001

It is possible that the models presented in this study are not the most optimal models for the datasets. For simplicity, we used the default values of the hyperparameters in building the RF and MERF models, and a linear combination of predictors was included as auxiliary information to predict total tree height in parametric mixed models. The main goal of this work was to provide an alternative solution for modelling H–D relationships for a large number of species and variables. A comprehensive analysis of the model sensitivity and efficacy with various forms and values of hyperparameters is warranted for future studies.

We noted that a few coefficients were not significantly different from zero (Table 6), which implies that a model with fewer parameters could predict total tree height equally well as some of the models examined here. However, to be comparable with the results from other methods, we still included all variables in constructing parametric equations. A simplified form of H-D models could be further investigated, but as indicated before, specifying a functional form for a high number of parameters is a challenging task in parametric mixed-effects modeling.

Although MERF is advantageous to predict total tree height for diverse species, running the algorithm with a large dataset can be time consuming. In this study, it took about three to four hours to run all submodels for the U.S. Virgin Islands and Trinidad, and around 12 hours for Puerto Rican trees, with the use of parallel computation in R. Applying such an algorithm can be challenging if computation power or time is limited. Recently, a variety of tree-based algorithms, such as Gaussian process boosting (Sigrist, 2021) and stochastic MERF (Capitaine et al., 2021) have been proposed in the literature, but the efficacy of the methods in modeling H-D relationships has not been fully investigated. Comparing alternative tree-based algorithms for characterizing H-D relationships is a proposed next step of this work.

Lastly, the models presented in this study can be used in forest inventory, growth, and yield estimation as the traditional applications of the H-D models. More importantly, the allometric models can also be used to estimate the original height of top-broken trees, which can enhance the quantification of the loss of tree volume from wind damage and provide a better assessment of tree health and stand conditions after disturbances such as hurricanes.

4. Conclusion

In short, MERF as a semi-parametric approach is advantageous for handling high dimensional and complex datasets using the random forest algorithm, as well as to provide species-specific predictions with random effects in the mixed model framework. MERF can be used to reliably predict total tree height for diverse species without a specified form of fixed effects. However, applying MERF can be challenging if the computation power or time is limited.

We found that in addition to DBH, competition index, defined as the ratio of the subject tree DBH to the plot-level QMD, was the most important variable among all other stand and environmental variables in all islands. Adding climate-related variables can improve the prediction accuracy of total tree height in parametric mixed models, especially when sample size is small. Precipitation variables were selected most often, confirming the overall importance of precipitation to forest height. Further investigations on applying tree-based algorithms to different forest types or environmental conditions in Caribbean forests are needed.

CRediT authorship contribution statement

Sheng-I Yang: Conceptualization, Methodology, Writing – original draft. **Thomas J. Brandeis:** Conceptualization, Data curation, Methodology, Writing – original draft. **Eileen H. Helmer:** Conceptualization, Data curation, Methodology, Writing – review & editing. **Michael P. Oatham:** Conceptualization, Data curation, Writing – review & editing. **Tamara Heartsill-Scalley:** Conceptualization, Data curation, Writing – review & editing. **Humfredo Marcano-Vega:** Conceptualization, Writing – review & editing.

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.

Data availability

The FIA data can be found at https://www.fia.fs.fed.us/. The Trinidad data may be requested from the Forestry Division of the Ministry of Agriculture, Land and Fisheries, Trinidad and Tobago.

Acknowledgement

The findings and conclusions in this publication are those of the authors and should not be construed to represent any official U.S. Department of Agriculture (USDA) or U.S. Government determination or policy. This research was supported in part by the USDA, Forest Service. The USDA Forest Service International Institute of Tropical Forestry works in collaboration with the University of Puerto Rico.

References

Adame, P., Brandeis, T.J., Uriarte, M., 2014. Diameter growth performance of tree functional groups in Puerto Rican secondary tropical forests. Forest Systems 23 (1), 52–63.
Amateis, R.L., Burkhart, H.E., Walsh, T.A., 1989. Diameter increment and survival equations for loblolly pine trees growing in thinned and unthinned plantations on cutover, site-prepared lands. South. J. Appl. For. 13, 170–174.
Bechtold, W.A., Patterson, P.L., 2005. The enhanced forest inventory and analysis program—National sampling design and estimation procedures. USDA Forest Service, Southern Research Station, General Technical Report SRS–80. 98 p.
Birdsey, R. A., Weaver, P.L., 1982. The forest resources of Puerto Rico. (Resource Bulletin SO-85). 56 p.
Brandeis, T.J., Turner, J.A., 2013a. Puerto Rico's forests, 2009. USDA Forest Service Southern Research Station Resource Bulletin SRS–191. 85 p.
Brandeis, T.J., Turner, J.A., 2013b. U.S. Virgin Islands' forests, 2009. USDA Forest Service Southern Research Station Resource Bulletin SRS–191. 56 p.
Brandeis, T.J., Delaney, M., Parresol, B.R., Royer, L., 2006. Development of equations for predicting Puerto Rican subtropical dry forest biomass and volume. For. Ecol. Manage. 233 (1), 133–142.
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. Manage. 258, 1704–1718.
Brandeis, T.J., Randolph, K., 2010. Modeling Caribbean tree heights and crown widths. Carib. J. Sci. 46 (2–3), 176–185.
Brandeis, T.J., Randolph, K.C., Strub, M.R., 2009b. Modeling Caribbean tree stem diameters from tree height and crown width measurements. Math. Comput. For. Nat.-Resource Sci. 1 (2), 78–85.
Brandeis, T.J., Oswalt, S.N., 2007. The status of U.S. Virgin Islands' forests, 2004. Resource Bulletin SRS–122. U.S. Department of Agriculture Forest Service, Southern Research Station, Asheville, NC. 61 p.
Brandeis, T.J., 2003. Puerto Rico's forest inventory: adapting the forest inventory and analysis program to a Caribbean island. J. For. Jan/Feb (1), 8–13.
Burkhart, H.E., Tomé, M., 2012. Modeling Forest Trees and Stands. Springer, Netherlands, p. 458.
Capitaine, L., Genuer, R., Thiébaut, R., 2021. Random forests for high-dimensional longitudinal data. Stat. Methods Med. Res. 30 (1), 166–184.
Clubbe, C., Jhilmit, S., 1992. Integrating forestry and biodiversity conservation in tropical forests in trinidad. In: Maunder, M., Clubbe, C., Hankemer, C., Groves, M. (Eds.), Plant Conservation in the Tropics: Perspectives and Practice. The Royal Botanic Gardens, Kew. UK, pp. 185–206.
Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792.
Elzhov, V., Mullen, K.M., Spiess, A.N., Bolker, B., 2022. R package 'minpack.lm' - Interface to the Levenberg-Marquardt nonlinear least-squares algorithm Found in MINPACK, Plus Support for Bounds. 14 p.
Fairhead, J., Leach, M., 2002. 'Sustainable' timber production and science/policy processes in Trinidad. Int. Dev. Services (IDS) Bull. 33, 75–83.
Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37 (12), 4302–4315. Available from: <https://www.worldclim.org/data/index.html>.
Franklin, J., Ripplinger, J., Freid, E.H., Marcano-Vega, H., Steadman, D.W., 2015. Regional variation in Caribbean dry forest tree species composition. Plant Ecol. 216 (6), 873–886.
French, C.D., Schenk, C.J., 2004. Map showing geology, oil and gas fields, and geologic provinces of the Caribbean region (No. 97-470-K). US Geological Survey. doi: 10.3133/ofr97470K.
Gibbes, C., Southworth, J., Keys, E., 2009. Wetland conservation: change and fragmentation in Trinidad's protected areas. Geoforum 40, 91–104.
Hajjem, A., Bellavance, F., Larocque, D., 2014. Mixed effects random forest for clustered data. J. Stat. Comput. Simul. 84 (6), 1313–1328.
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. Special Issue No. 1, pp. 1-12.
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. Carib. J. Sci. 38 (3–4), 165–183.
Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org).
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. Manage. 459, 117823.
Lam, T.Y., Kershaw Jr., J.A., Hajar, Z.S.N., Rahman, K.A., Weiskittel, A.R., Potts, M.D., 2017. Evaluating and modelling genus and species variation in height-to-diameter relationships for Tropical Hill Forests in Peninsular Malaysia. For.: Int. J. For. Res. 90 (2), 268–278.
López-Marrero, T., Heartsill-Scalley, T., 2012. Get up, stand up: environmental situation, threats, and opportunities in the insular Caribbean. Caribbean Studies 40 (1), 3–14.
Lugo, A.E., Carlo, T.A., Wunderle, Jr., J.M., 2012b. Natural mixing of species: novel plant-animal communities on Caribbean Islands. Animal Conservation.
Lugo, A.E., Smith, J.E., Potter, K.M., Marcano-Vega, H., Kurtz, C.M., 2022. The contribution of nonnative tree species to the structure and composition of forests in the conterminous United States in comparison with tropical islands in the Pacific and Caribbean. Gen. Tech. Rep. IITF-54. Río Piedras, PR: U.S. Department of Agriculture Forest Service, International Institute of Tropical Forestry. 81 p. doi: 10.2737/IITF-GTR-54.
Lugo, A.E., Helmer, E.H., Santiago Valentín, E., 2012a. Caribbean landscapes and their biodiversity. Interciencia 37, 705–710.
Marcano-Vega, H., 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, pp. 1–4.
Oatham, M.P., Ramnarine, S., 2006. Dynamics of pioneer and primary successional stage trees in a logged Trinidadian tropical rainforest and the influence of drought. Trop. Ecol. 47 (1), 13–26.
Pearl, R., Reed, L.J., 1920. On the rate of growth of the population of the United States since 1790 and its mathematical representation. PNAS 6 (6), 275–288.
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. Manage. 157 (1–3), 205–216.
Pinheiro, J., Bates, D., DebRoy, Saikat., Sarkar, D., EISPACK authors, Heisterkamp, S., Van Willigen, B., Ranke, J., R Core Team, 2021. R package "nlme" - Linear and Nonlinear Mixed Effects Models. 338 p.
Rudel, T.K., Pérez-Lugo, M., Zichal, H., 2000. When fields revert to forest: development and spontaneous reforestation in post-war Puerto Rico. Professional Geographer 52 (3), 386–397.
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.
Sabatia, C.O., Burkhart, H.E., 2014. Predicting site index of plantation loblolly pine from biophysical variables. For. Ecol. Manage. 326, 142–156.
Sabatia, C.O., Burkhart, H.E., 2015. On the use of upper stem diameters to localize a segmented taper equation to new trees. For. Sci. 61 (3), 411–423. https://doi.org/10.5849/forsci.14-039.
Santiago-Valentin, E., Olmstead, R.G., 2004. Historical biogeography of Caribbean plants: introduction to current knowledge and possibilities from a phylogenetic perspective. Taxon 53 (2), 299–319. https://doi.org/10.2307/4135610.
Scatena, F.N., Moya, S., Estrada, C., Chinea, J.D., 1996. The first five years in the reorganization of aboveground biomass and nutrient use following hurricane Hugo in the Bisley Experimental Watersheds, Luquillo Experimental Forest, Puerto Rico. Biotropica 28 (4a), 424–440.
Sigrist, F., 2021. Gaussian process boosting. arXiv:2004.02653v3 [cs.LG]. 41 p. doi: 10.48550/arXiv.2004.02653.
Temesgen, H., Hann, D.W., Monleon, V.J., 2006. Regional height-diameter equations for major tree species of Southwest Oregon. West. J. Appl. For. 22 (3), 213–219.
Thomas, S.C., Martin, A.R., Mycroft, E.E., 2015. Tropical trees in a wind-exposed island ecosystem: height-diameter allometry and size at onset of maturity. J. Ecol. 103 (3), 594–605.
Vanclay, J.K., 1991. Aggregating tree species to develop diameter increment equations for tropical rainforests. For. Ecol. Manage. 42, 143–168.
Wang, J., Chen, L.S., 2016. R Package 'MixRF' - A random-forest-based approach for imputing clustered incomplete data. 6 p.
Weaver, P.L., Gillespie, A.J., 1992. Tree biomass equations for the forests of the Luquillo Mountains, Puerto Rico. Commonwealth For. Rev. 71 (1), 35–39.
Weiskittel, A.R., Crookston, N.L., Radtke, P.J., 2011. Linking climate, gross primary productivity, and site index across forests of the western United States. Can. J. For. Res. 41 (8), 1710–1721.
Wright, M.N., Ziegler, A., König, I.R., 2016. Do little interactions get lost in dark random forests? BMC Bioinf. 17, 145.
Wright, M.N., Ziegler, A., 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. 77 (1), 1–17.
Wykoff, W.R., Crookston, N.L., Stage, A.R., 1982. User's guide to the stand prognosis model. In: Gen. Tech. Rep. INT-122. U.S. Department of Agriculture, Forest Service. Intermountain Forest and Range Experiment Station, Ogden, UT, USA, 113 pp.
Yang, S., Burkhart, H.E., 2020a. Evaluation of total tree height subsampling strategies for estimating volume in loblolly pine plantations. For. Ecol. Manage. 461, 117878.
Yang, S., Burkhart, H.E., 2020b. Robustness of parametric and nonparametric fitting procedures of tree-stem taper with alternative definitions for validation data. J. Forest. 118 (6), 576–583.

Yang, Y., Saatchi, S.S., Xu, L., Yu, Y., Lefsky, M.A., White, L., Knyazikhin, Y., Myneni, R. B., 2016. Abiotic controls on macroscale variations of humid tropical forest height. Remote Sensing 8 (6), 494. https://doi.org/10.3390/rs8060494.

Zhang, J., Bras, R.L., Longo, M., Heartsill Scalley, T., 2022. The impact of hurricane disturbances on a tropical forest: implementing a palm plant functional type and hurricane disturbance module in ED2-HuDi V1. 0. Geosci. Model Dev. 15 (13), 5107–5126. https://doi.org/10.5194/gmd-15-5107-2022.

Zhu, T., 2020. Analysis on the applicability of the random forest. J. Phys. Conf. Ser. 1607, 012123.

Further reading

Dietz, J.L., 1986. Economic History of Puerto Rico. Institutional Change and Capitalist Development. Princeton, NJ, Princeton University Press, 337 p.