

**Evaluation of regression methods and competition indices in characterizing
height-diameter relationships for temperate and pantropical tree species**

A Thesis Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Sakar Jha

May 2024

Copyright © 2023 Jha, Yang, Brandeis, Kuegler and Marcano-Vega. This research was published as an open access research article by Frontiers in Forests and Global Change as: Jha, S., Yang, S.-I., Brandeis, T.J., Kuegler, O. and Marcano-Vega, H. 2023 Evaluation of regression methods and competition indices in characterizing height-diameter relationships for temperate and pantropical tree species. *Frontiers in Forests and Global Change*, **6**. The permission to include this article in this thesis from the Frontiers in Forests and Global Change is gratefully acknowledged.

ACKNOWLEDGEMENTS

This thesis would not be possible without the encouragement, motivation and support of a number of people in my life. I would like to extend my deepest gratitude to my advisor Dr. Sheng I Yang. His guidance, mentorship and motivation has helped me develop my technical expertise in forest biometrics and made me fall in love with tackling research problems through data and statistics driven solutions. I would like to express my appreciation to Dr. Thomas J. Brandeis who has played an instrumental role in the progress of my research and helped me refine it. Similarly, I would like to thank Dr. David Buckley for serving on my graduate committee and for teaching dendrology and advanced forest ecology.

I would like to thank the USDA Forest Service Forest Inventory and Analysis program for providing the long term remeasurement dataset for the eastern US and the Caribbean and Pacific islands. These datasets have been tremendously valuable for addressing research questions. I would like to thank my lab mates Tara Skiba, Dinuka Senevirathne and Samjhana Wagle for always engaging in academic discussions about contemporary topics in forestry.

I would like to thank my mother, Madhu Jha, my grandmother, Shanti Jha and my brother, Nirakar Jha who have always supported me in my life and my decisions. My deepest gratitude to my sister Dr. Akriti Khadka and brother-in-law Dr. Saroj Dhital who have always guided me in tough situations and made me feel like home in the US.

ABSTRACT

Height-diameter relationship models, denoted as H-D models, have important applications in sustainable forest management which include studying the vertical structure of a forest stand, understanding habitat heterogeneity for wildlife niches and analyzing growth rate patterns for making decisions regarding silvicultural treatments. Compared to monocultures, characterizing allometric relationships for uneven-aged, mixed-species forests, especially tropical forests, is more challenging and has historically received less attention. Modeling how the competitive interactions between trees of varying sizes and multiple species affects these relationships adds a high degree of complexity. In this study, five regression methods and five distance-independent competition indices were evaluated for temperate and pantropical tree species in different physiographic regions. A total of 163,922 individual tree measurements from the U.S. Department of Agriculture, Forest Inventory and Analysis (FIA) database were used in the analyses, which covered the Appalachian plateau (AP) and Ridge and Valley (VR) in the southeastern US, as well as Caribbean (CAR) and Pacific (PAC) islands. Results indicated that the generalized additive model (GAM) and the Pearl and Reed model provided more accurate predictions than other regression methods examined. Models with competition indices had a varying level of predictability, while diameter ratio, cumulative distribution function and partitioned stand density index (PSDI) were found to improve the prediction accuracy for AP, VR and CAR. The results of this work provide additional insights on modeling H-D relationships for a variety of species in temperate and pantropical forests.

Keywords: Temperate forest, Caribbean islands, Pacific islands, Tree allometry, Generalized additive model, Machine learning

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
1.1 Background	1
1.2 Objectives	4
CHAPTER 2: LITERATURE REVIEW	5
2.1 Tree allometry	5
2.2 Height Diameter relationship	5
2.3 Height-Diameter modelling in mixed forests	6
2.4 Resource competition.....	7
2.5 Advanced modelling methods.....	8
CHAPTER 3: METHODS	9
3.1 Study Area.....	9
3.1.1 Temperate forests	9
3.1.2 Pantropical forests.....	9
3.2 Tree Data.....	11
3.3 Regression methods	12
3.3.1 Parametric model 1 – Pearl & Reed (1920)	12
3.3.2 Parametric model 2 – Wykoff et al. (1982).....	15
3.3.3 Semi-parametric model 1 – General additive model (GAM)	15
3.3.4 Nonparametric model 1 – Extreme Gradient Boosting (XGBoost).....	16

3.4	Competition index (CI)	18
3.4.1	Diameter ratio	19
3.4.2	Basal area ratio.....	19
3.4.3	Partitioned stand density index	19
3.4.4	Cumulative distribution function	20
3.4.5	Basal area of larger trees	20
CHAPTER 4: RESULTS		23
4.1	Predictability of the base models on temperate trees.....	23
4.2	Predictability of base models on pantropical trees	23
4.3	Predictability of models with and without competition index	32
4.4	Best combination of regression method and competition index for different regions..	32
CHAPTER 5: DISCUSSION.....		43
5.1	Comparison among regression methods	43
5.2	Inclusion of a competition index in H-D models.....	45
5.3	Further discussion	49
CHAPTER 6: CONCLUSION		55
REFERENCES		56
APPENDIX.....		63
VITA		104

LIST OF TABLES

Table 1. Summary of tree data among temperate (Appalachian Plateau, AP, and Valley and Ridge, VR) and pantropical forests (Caribbean islands, CAR, and Pacific islands, PAC). Number of observations ($N_{\text{obs.}}$) is given for each species. Average (Avg), standard deviation (Std), minimum (Min) and maximum (Max) of DBH (cm) and total tree height (m) were calculated.	13
Table 2. Median of the evaluation statistics (coefficients of determination, R^2 ; mean bias, MB; root mean squared error, RMSE) for the base models (i.e., no competition index included in the model) and five competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) using five regression models (Pearl & Reed (1920), PR; Wykoff et al. (1982), WK; Generalized additive model, GAM; Extreme Gradient Boosting, XGboost; Artificial Neural Network, ANN) for the six temperate trees species (eastern white pine, EWP; hickory species, HICK; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar, YP) in Appalachian plateau (AP).	26
Table 3. Median of the evaluation statistics (coefficients of determination, R^2 ; mean bias, MB; root mean squared error, RMSE) for the base models (i.e., no competition index included in the model) and five competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) using five regression models (Pearl & Reed (1920), PR; Wykoff et al. (1982), WK, WK; Generalized additive model, GAM; Extreme Gradient Boosting, XGboost; Artificial Neural Network, ANN) for the six temperate trees species (eastern white pine, EWP; hickory species, HICK; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar, YP) in Ridge and Valley (VR).	29

Table 4. Median of the evaluation statistics (coefficients of determination, R^2 ; mean bias, MB; root mean squared error, RMSE) for the base models (i.e., no competition index included in the model) and five competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) using five regression models (Pearl & Reed (1920), PR; Wykoff et al. (1982), WK; Generalized additive model, GAM; Extreme Gradient Boosting, XGboost; Artificial Neural Network, ANN) for four pantropical trees species in Caribben islands (CAR) (pink trumpet tree, PTT; muskwood, MW; African tulip tree, ATT; acacia, ACA) and three pantropical species in Pacific islands (PAC) (African tulip tree, ATT; ‘Ōhi‘a lehua, OL, acacia, ACA).	33
Table 5. Percentage change of evaluation statistics (coefficients of determination, R^2 ; mean bias, MB; root mean squared error, RMSE) between the best performing and the corresponding base model for temperate species (eastern white pine, EWP; hickory species, HICK; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar, YP) in Appalachian plateau (AP) and Ridge and Valley (VR) and pantropical species in Caribbean islands (CAR) (pink trumpet tree, PTT; muskwood, MW; African tulip tree, ATT; acacia, ACA) and Pacific islands (PAC) (African tulip tree, ATT; ‘Ōhi‘a lehua, OL, acacia, ACA).....	38
Table 6. Range of evaluation statistics (coefficients of determination, R^2 ; mean bias, MB; root mean squared error, RMSE) between the best performing and the corresponding base model for temperate species (eastern white pine, EWP; hickory species, HICK; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar) in Appalachian plateau (AP) and Ridge and Valley (VR) and pantropical species in Caribbean islands (CAR) (pink trumpet tree, PTT; muskwood, MW; African tulip tree, ATT; acacia, ACA) and Pacific islands (PAC) (African tulip tree, ATT; ‘Ōhi‘a lehua, OL, acacia, ACA)	46

LIST OF FIGURES

Figure 1. Map of the temperate physiographic regions (Appalachian Plateau and Ridge and Valley) (A) and pantropical islands (Puerto Rico and U.S. Virgin Islands in the Caribbean (B) and Hawai‘i, Federated States of Micronesia, Guam and Northern Mariana Islands in the Pacific (C, D, E)).....	10
Figure 2. Overall ranking (Lowest rank is the best) of five base regression methods (Pearl & Reed (1920), PR; Wykoff et al., 1982, WK; Generalized additive model, GAM; Extreme gradient boost, XGB; Artificial neural network, ANN) across all the species and physiographic region in (A) and across all species in four physiographic regions (Apalachian plateau, AP; Ridge and Valley, VR; Caribbean islands, CAR; Pacific islands, PAC) in (B).....	24
Figure 3. Average rank (Lowest rank is the best) of five base regression methods for species in Appalachian plateau (AP) in (A) (eastern white pine, EWP; hickory species, HICK; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar, YP), Ridge and Valley (VR) in (B) (eastern white pine, EWP; hickory species, HICK; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar, YP), Caribbean islands (CAR) in (C) (pink trumpet tree, PTT; muskwood, MW; African tulip tree, ATT; acacia species, ACA) and Pacific islands (PAC) in (D) (African tulip tree, ATT; ‘Ōhi‘a lehua, OL; acacia species, ACA).....	25
Figure 4. Average ranking (Lowest rank is the best) of competition indices (diameter ratio, <i>Rdia</i> ; basal area ratio, <i>RBA</i> ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) for five regression models (Pearl & Reed (1920), PR (A); Wykoff et al. (1982), WK (B); Generalized additive model, GAM (C); Extreme Gradient Boosting, XGboost (D); Artifical Neural Network, ANN (E)) across all species and physiographic regions	33

Figure 5. Best combination of regression method (Pearl & Reed (1920), PR; Wykoff et al. (1982), WK; Generalized additive model, GAM; Extreme Gradient Boosting, XGboost; Artificial Neural Network, ANN) and competition index (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) for all the species across four physiographic regions (Appalachian plateau, AP (A); Ridge and Valley, VR (B); Caribbean islands, CAR (C); Pacific islands, PAC (D)).....	41
Figure 6. Height diameter relationship of the testing dataset (20% of the dataset) for each of the four physiographic region (Appalachian plateau, AP; Ridge and Valley, VR (B); Caribbean islands, CAR (C); Pacific islands, PAC (D)). The curves show the predicted relationship for the highest and the lowest ranked models for each of the physiographic regions.....	42
Figure 7. Comparison of the predicted tree height and the residual height for the highest ranking and lowest ranking combination of regression method and competition index for species in Appalachian plateau (AP; eastern white pine, EWP, A,B; hickory species, HICK, C,D; red maple, RM, E,F; sweetgum, SWG, G,H, white oak, WO, I,J; yellow poplar, YP, K,L).	51
Figure 8. Comparison of the predicted tree height and the residual height for the highest ranking and lowest ranking combination of regression method and competition index for species in Ridge and Valley (VR; eastern white pine, EWP, A,B; hickory species, HICK, C,D; red maple, RM, E,F; sweetgum, SWG, G,H, white oak, WO, I,J; yellow poplar, YP, K,L).	52
Figure 9. Comparison of the predicted tree height and the residual height for the highest ranking and lowest ranking combination of regression method and competition index for species in Caribbean (CAR) and Pacific (PAC) islands (pink trumpet tree, PTT, A,B; muskwood, MW, C,D; African tulip tree, ATT, E–H; acacia species, ACA, I–L; ‘Ōhi‘a lehua, OL, M,N).	53

CHAPTER 1: INTRODUCTION

1.1 Background

As vital components of the global carbon cycle, forest ecosystems play an important role in maintaining ecological balance and providing several ecosystem services. Temperate forests in the eastern US and tropical forests in the Caribbean and Pacific regions are provide space to large amount of forestland that conserve and maintain critical habitats and ecosystems.

Accurately assessing current forest resources is required for the development of effective silvicultural and forest management practices for wood products, wildlife habitat and ecosystem services (Burkhart & Tomé, 2012). Total tree height (H) and diameter at breast height (DBH) are important components in estimating tree carbon, predicting tree growth and understanding stand vertical structure. However, measuring tree height in a forest is often more time consuming and labor intensive than DBH. When faced with resource constraints, a subsampling strategy is commonly combined with height-diameter (H-D) relationship models to provide an efficient approach to obtaining these data in forest inventories (Yang & Burkhart, 2020). In addition, H-D relationship models can also be applied to predict the intact length of top-broken trees, thereby improving the estimation of tree volume loss from wind disturbance such as hurricanes (Yang et al., 2022).

Due to important applications of H-D relationship models in natural resources, several models have been extensively explored and proposed according to forest types and regions (e.g., Curtis, 1967, Huang et al., 2000, Soares & Tomé, 2002). Temesgen et al. (2014) used nonlinear mixed-effects models to quantify the H-D relationships for 23 tree species in northeastern China. Mehtätalo et al. (2015) examined the performance of 16 nonlinear functions using 126,000 trees collected from different parts of the world. It was found that the model performance varied

among tree species and regions (Mehtätalo et al., 2015). With increased computing power in recent decades, alternative modeling methods (e.g., semi-parametric and nonparametric approaches) have been applied to model H-D relationships. Semi-parametric models like generalized additive model (GAM) have shown to have comparable and sometimes better performance than parametric models in forestry research (Adamec & Drápela, 2015, Robinson et al., 2011, Yang et al., 2023, Zang et al., 2016). Adamec & Drápela (2015) found that GAM models are suitable for modelling the H-D relationship for Norway spruce. Similarly, Zang et al. (2016) found that GAM models had comparable performance to parametric mixed models. Furthermore, Özçelik et al. (2013) applied back-propagation artificial neural networks (ANN) models to predict total tree height of Crimean pines in southwestern Turkey. The prediction accuracy of the back-propagation artificial neural network models was comparable to the parametric methods (Özçelik et al., 2013). Chen et al. (2020) built H-D relationship models with 31 stand, environmental and climatic variables using ANN models for six boreal species in Ontario, Canada. Their models can explain more than 90% of the height variation (Chen et al., 2020). Ogana & Ercanli (2022) compared deep learning algorithms to model H-D relationships for tropical rain forest species and found that they performed better than nonlinear models. Decision tree-based algorithms such as random forests have also been found to perform well for diverse-species forests in the insular Caribbean (Yang et al., 2022). Extreme gradient boosting (XGBoost) is another tree-based algorithm which gained popularity after winning several data-mining competitions hosted by Kaggle in 2015 (Chen & Guestrin, 2016) and has been a go-to machine learning algorithm among data scientists since then. XGBoost is based on the ensemble learning method where multiple weak decision trees are combined into a strong decision tree. At each iteration, a new decision tree is trained on the residual errors of the previous tree. These

decision trees are generated based on the objective function. Different from traditional gradient boosting algorithm, the objective function in XGBoost is penalized with regularization techniques to prevent overfitting (Chen & Guestrin, 2016). XGBoost has been applied in several areas of forestry, especially in using remote sensing for forestry research (e.g., Wang et al., 2022, Xu et al., 2022, Zhang et al., 2022). However, to our understanding, applying XGBoost in modeling H-D relationships for a variety of species has not been explicitly examined in the literature.

In addition to DBH and H, distance-independent competition index is another important variable in individual tree growth and allometric models. The distance-independent indices are used to quantify the influence of competition by the local neighbors of the subject tree, which do not require individual tree coordinates (Burkhart & Tomé, 2012). Sun et al. (2019) evaluated six different distance-independent competition indices for loblolly pine (*Pinus taeda* L.) in predicting tree growth and mortality. Diameter ratio ranked highest to predict tree survival while relative position indices performed the best for predicting diameter growth (Sun et al., 2019). For quantifying H-D relationships, Temesgen et al. (2014) indicated that the inclusion of the relative size position of trees in the stand (sum of basal area of larger trees greater than the subject tree) in the models improved the prediction precision compared to the original models. Yang et al. (2022) pointed out that the ratio of the subject tree DBH to the sampling-unit-level quadratic mean diameter (QMD) is one of the key predictors in H-D models for multi-species subtropical forests.

In this region, significant efforts were dedicated to modeling H-D relationships for single-species, even-aged forests. Compared to monocultures, characterizing allometric relationships for uneven-aged, mixed-species forests, especially tropical forests, is more challenging and has

historically received less attention. Modelling how the competitive interactions between trees of varying sizes and multiple species affects these relationships adds a high degree of complexity. Further, generic models and default values developed for temperate species are currently applied in the Caribbean and Pacific regions, which may not provide the accurate quantitative information for assessing the impact of natural disturbances and forest health.

1.2 Objectives

The main objective of this study was to evaluate the performance of different regression methods and distance independent competition indices in characterizing H-D relationships. The specific objectives of this study were as follows:

- (A) Compare the predictive capabilities of two commonly used parametric, one semi-parametric and two machine learning regression methods.
- (B) Evaluate the prediction capabilities of regression methods after incorporating distance independent competition indices as predictors in the regression methods.

CHAPTER 2: LITERATURE REVIEW

2.1 Tree allometry

Tree allometry delves into establishing relationships between the size of trees and several structural and functional attributes, such as height, diameter, biomass and leaf area. Establishing these relationships between the different attributes of trees aids in estimating parameters such as volume, biomass, carbon storage that are not easily measured in the field. This knowledge is instrumental in efficiently carrying out forest inventory and designing management plans to meet management objectives. Allometric equations are used to express mathematical relationships between tree characteristics such as height-diameter equations (Mehtätalo et al., 2015), tree biomass equations (Jenkins et al., 2003) and crown structure equations (Gering & May, 1995). With the advancement of improved forest measurement and statistical modelling, the study of tree allometry has significantly expanded. Jenkins et al. (2003) developed a comprehensive database of 2640 diameter-based biomass equations for tree species in north America. Availability of large-scale database of tree measurements such as the FIA database has allowed the development of allometric equations that are suited for a particular species in a geographic location.

2.2 Height Diameter relationship

The strong relationship between height and diameter of individual trees has been known to foresters for a long time. Both diameter and height measurements are necessary for estimating the volume of a tree and height is also used to characterize the quality of the site's growing capacity. DBH measurements of trees are relatively easily obtained during forest measurements compared to tree height, which is more labor intensive. Hence, height-diameter models are useful tools for building a decision support system for forest management. Different forms of height-

diameter equations have been tested since the late 1960s (e.g. Curtis, 1967) and followed up by several researchers throughout the decades (e.g. Adame et al., 2008, Lam et al., 2017, Özçelik et al., 2013, Sharma et al., 2019, Soares & Tomé, 2002, Temesgen et al., 2014, Yang et al., 2022). Generally, the models that describe the height-diameter relationship for tree are of two types; models using only DBH as a predictor and models using DBH and additional predictors that describe stand characteristics and competition. The earlier height-diameter models had linear form, but it does not hold true for most of the species as lateral growth of trees continue even when height growth has plateaued. Hence, non-linear models provide a better fit. Mehtätalo et al. (2015) evaluated sixteen different nonlinear model forms using an extensive dataset for several species. In addition, the height diameter relationship tends to vary in mixed species forest and is dependent on habitat specific characteristics such as light, elevation, slope, aspect (Schmidt et al., 2011). Further, the density of the stand also determines the growth behavior of trees. For example, trees growing in a dense stand have smaller DBH in proportion to a given height compared to stands with lower density.

2.3 Height-Diameter modelling in mixed forests

Developing models for H-D relationships in mixed forest ecosystems is often more difficult because of multiple species in different size classes and complex competition dynamics between the species (Schmidt et al., 2011, Temesgen et al., 2014). Feldpausch et al. (2011) studied the variation in H-D relationship for tropical species in different forest types and geographic regions. Temesgen et al. (2014) evaluated twelve H-D models using non-linear mixed effects approach among different habitat types. Yang et al. (2022) mixed effect random forest models where the forest types in the pantropical forests were treated as random and found improvement in model

performance. Similarly, Ogana & Ercanli (2022) found deep-learning algorithms were able to better predict the H-D relationship for tropical rain forests in Nigeria.

2.4 Resource competition

The struggle for resources such as light, nutrients, space, water between individual trees in a stand determines the growth and dynamics of that stand. The competition between individual trees is highly dynamic and depends on various factors such as stage of stand development, geographic location, silvical characteristics of the tree species and site quality, among others. The competition also varies between individual trees of the same species (intraspecific) and between different species (interspecific) in the stand. Interspecific competition is significant in mixed species stand whereas intraspecific competition is generally considered in plantations. However, Zhao et al. (2006) found that intraspecific competition in a mixed stand was more significant than interspecific competition. Similarly, competition can also be characterized as one-sided competition where the larger and better-established trees in the stand are not affected by smaller trees in the stand and two-sided competition where both larger and smaller trees affect each other. It is commonly believed that one-sided competition generally occurs for aboveground resources (e.g. light) and two-sided competition occurs for belowground resources (e.g. water, nutrients) (Casper & Jackson, 1997). Competition indices are used to mathematically represent the biological implications of competition in tree growth. The competition indices generally used for research take into account the size of the individual trees, the number of the trees and distribution of the trees based on size of the trees (Weiskittel et al., 2011). Competition indices are also generally classified into distance dependent indices where the individual geographic coordinates of the trees are used for calculation of the index and distance independent competition indices where the distance between the trees are not taken into consideration.

2.5 Advanced modelling methods

The advancement of computing power has allowed for the development of several algorithms that have shown promising results in solving classification and prediction problems. Semi-parametric and non-parametric algorithms such as generalized additive model, support vector machine, random forest, artificial neural network have become mainstream for modelling complex systems. These algorithms do not make any assumptions about the underlying distribution of the data and use complex design architectures and calculations to make predictions. This property makes them more flexible for fitting complex datasets and improves their performance as more data becomes available (Robert, 2014). However, these types of models are less interpretable and generalizable as they may not always perform well for new dataset if it falls outside the range of the training dataset (Robert, 2014). The use of such advanced non-parametric algorithm in modelling H-D relationship was first done by Özçelik et al. (2013), where an ANN model was used to model H-D relationship of Crimean juniper. Since then several researchers have explored the use of ANNs and other such complex algorithms for modelling H-D relationships. For example, Ogana & Ercanli (2022) used deep learning algorithms to model the H-D relationship of tropical rain forests in Nigeria. Yang et al. (2022) used mixed random forest model for modelling H-D relationship for pantropical trees in Caribbean islands.

CHAPTER 3: METHODS

3.1 Study Area

A total of 163,922 observations used in analyses were collected from temperate forests in the eastern U.S. and pantropical forests in Caribbean and Pacific islands. Spatial distribution of the study area is shown in Figure 1. Description of all forests is given below.

3.1.1 *Temperate forests*

The oak-hickory forests located in the Appalachian Plateau (AP), and Ridge and Valley (VR) physiographic regions were selected, which provide critical forest ecosystems and important wildlife habitats in the eastern US. The boundaries of both physiographic zones (AP and VR) were delineated with the classification made by Fenneman & Johnson (1946) and the map of the FIA survey units listed in Burrill et al. (2021). The study area starts from northern Alabama and extends northeast towards New York covering parts of several eastern states, including Georgia, South Carolina, North Carolina, Tennessee, Kentucky, Virginia, West Virginia, Ohio, Maryland, and Pennsylvania. In this study, six ecologically and economically important species were selected in model building, which include eastern white pine (*Pinus strobus* L.), red maple (*Acer rubrum* L.), sweetgum (*Liquidambar styraciflua* L.), white oak (*Quercus alba* L.), yellow poplar (*Liriodendron tulipifera* L.) and hickory spp. (*Carya* spp.). The models were constructed at the genus level for hickories due to the limited number of trees for each individual species in the database.

3.1.2 *Pantropical forests*

The Caribbean islands of Puerto Rico and U.S. Virgin Islands and the Pacific islands of Guam, Hawai'i, Federated States of Micronesia, and Northern Mariana Islands were selected. The

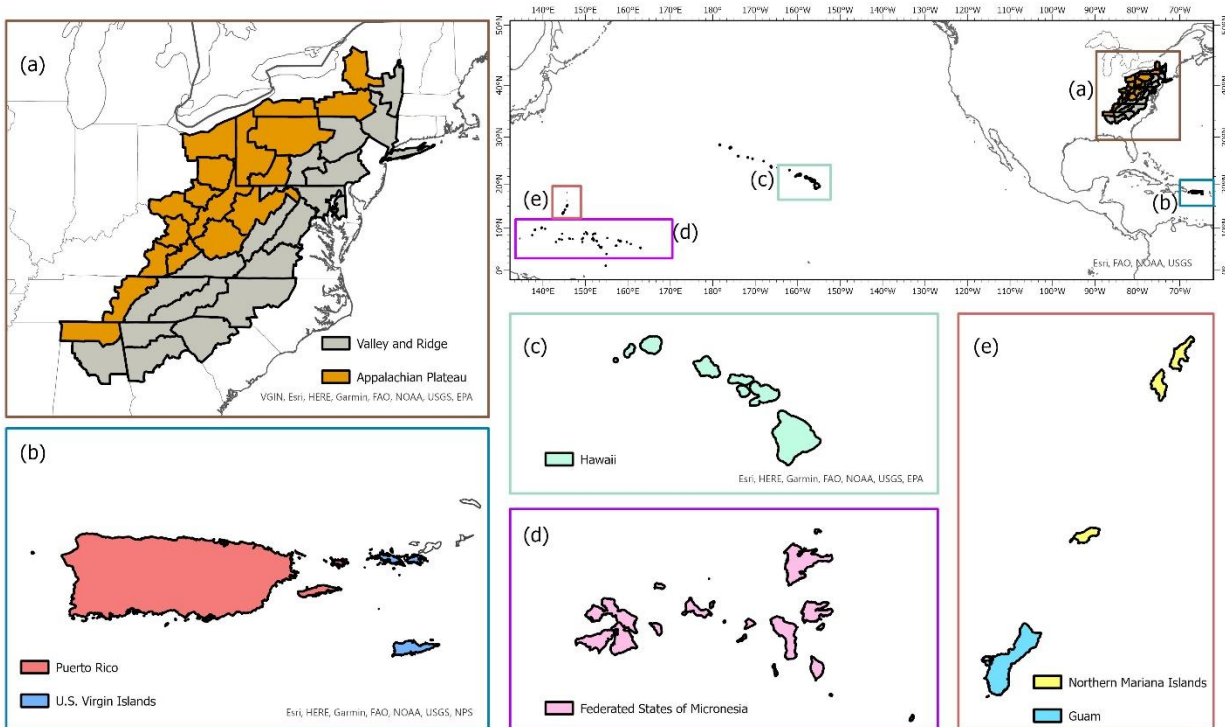


Figure 1. Map of the temperate physiographic regions (Appalachian Plateau and Ridge and Valley) (A) and pantropical islands (Puerto Rico and U.S. Virgin Islands in the Caribbean (B) and Hawai'i, Federated States of Micronesia, Guam and Northern Mariana Islands in the Pacific (C, D, E)).

Caribbean and Pacific islands have unique climatic conditions which allow for diverse forests to grow (Whiffin, 1992). Four common species groups in Puerto Rico and U.S. Virgin Islands were selected, including pink trumpet tree (*Tabebuia heterophylla* DC.), muskwood (*Guarea guidonia* L.), African tulip tree (*Spathodea campanulata* P.Beauv) and Acacia (*Acacia* spp.). For the Pacific Island forests, common species groups: African tulip tree (*Spathodea campanulata* P.Beauv), 'Ōhi'a lehua (*Metrosideros polymorpha* Gaudich.) and Acacia (*Acacia* spp.), were selected. Similar to hickories in temperate forests, models were constructed at the genus level for Acacia.

3.2 Tree Data

Tree data were obtained from the US national forest inventory established and maintained by the Forest Inventory and Analysis (FIA) program under the US Department of Agriculture (USDA) Forest Service. The permanent sample plots were installed throughout the country with 10 to 20 % of the sample plots being measured annually depending on the region (Burrill et al., 2018). The FIA program has regularly monitored forest resources in the US commonwealth of Puerto Rico since 1980, with US Virgin Islands, associated Pacific Islands and the state of Hawai'i being incorporated into the inventory network more recently. Each FIA permanent plot consisted of four fixed-radius subplots, each measuring 7.3-m (24 ft) in radius, and four fixed-radius microplots nested within each subplot, each measuring 2.1-m (6.8 ft) in radius. On each subplot, DBH and H were measured for all trees with $DBH \geq 12.7$ cm (5 in.). On each nested microplot, tree with 2.5 cm (1 in.) $\leq DBH < 12.7$ cm (5 in.) were measured. The permanent plots were classified based on site condition including land use, forest type, stand origin, regeneration status, ownership group and disturbance history. An FIA permanent plot can have multiple condition classes based on distinct differences between forested growing conditions. Throughout this study, a distinct forested condition class within a plot was considered as the fundamental sampling unit (i.e., an individual

sample of trees). Stand characteristics (e.g., QMD, stand density index) used in computing competition indices were computed from all live trees on a “condition class” rather than a whole “plot”. Irregular observations (e.g., broken-top individuals, seriously-damaged stems) were removed from analysis. More details of the FIA sampling design and measurement procedures are available in (Bechtold & Patterson, 2005). A summary of tree characteristics for all species in temperate and pantropical forests is given in Table 1.

3.3 Regression methods

Five regression methods: two parametric functions (Pearl & Reed, 1920, Wykoff et al., 1982), one semi-parametric algorithm (GAM) and two nonparametric machine learning algorithms (XGBoost and ANN), were used to quantify H-D relationships. The regression methods were applied to each combination of species and region. For a given regression method, two types of H-D models were built. The first type included only DBH, denoted as base model, while the other type was composed of DBH and a competition index. In this study, five different competition indices were evaluated, resulting in a total of five H-D models in the second type. Detailed descriptions of the methods and competition indices are given below.

3.3.1 Parametric model 1 – Pearl & Reed (1920)

The Pearl & Reed (1920) model, denoted as PR model, was first introduced for predicting population growth in the US. It has been adopted in forestry to predict tree height for different species, such as oaks (Misik et al. 2016) and subtropical trees (Yang et al., 2022). Similar to Yang et al. (2022), the competition index was added in the exponent with DBH in the model. That is,

Table 1. Summary of tree data among temperate (Appalachian Plateau, AP, and Ridge and Valley, VR) and pantropical forests (Caribbean islands, CAR, and Pacific islands, PAC). Number of observations ($N_{\text{obs.}}$) is given for each species. Average (Avg), standard deviation (Std), minimum (Min) and maximum (Max) of DBH (cm) and total tree height (m) were calculated.

Forest	Reg.	Species	Species	$N_{\text{obs.}}$	DBH (cm)				Total height (m)			
		Common name	Scientific name		Avg	Std.	Min	Max	Avg	Std	Min	Max
Temperate	AP	Eastern white pine	<i>Pinus strobus</i>	1719	22.90	12.04	2.54	79.75	14.44	6.13	2.74	36.87
		Hickory spp.	<i>Carya spp.</i>	9037	23.85	10.31	2.54	90.67	19.37	6.36	1.82	41.14
		Red maple	<i>Acer rubrum</i>	27184	20.78	9.42	2.54	96.77	17.06	5.19	1.52	36.87
		Sweetgum	<i>Liquidambar styraciflua</i>	1030	19.36	10.52	2.54	68.32	17.29	7.13	2.43	38.09
		White oak	<i>Quercus alba</i>	10884	28.26	12.30	2.54	113.03	19.99	5.81	1.52	41.45
	VR	Yellow poplar	<i>Liriodendron tulipifera</i>	10190	27.07	13.71	2.54	111.76	21.66	7.52	2.43	45.71
		Eastern white pine	<i>Pinus strobus</i>	3948	20.25	12.83	2.54	108.45	13.55	6.66	1.82	44.80
		Hickory spp.	<i>Carya spp.</i>	12264	21.32	11.03	2.54	88.90	17.85	6.82	1.52	41.75
		Red maple	<i>Acer rubrum</i>	27205	18.91	10.23	2.54	88.90	16.00	5.77	1.52	38.40
		Sweetgum	<i>Liquidambar styraciflua</i>	8023	17.88	10.45	2.54	77.21	16.03	7.29	1.52	42.66
Pantropical	CAR	White oak	<i>Quercus alba</i>	13282	25.99	13.00	2.54	102.61	19.13	6.43	2.74	53.03
		Yellow poplar	<i>Liriodendron tulipifera</i>	19869	25.84	14.92	2.54	129.54	21.62	8.22	1.52	50.59
		Acacia	<i>Acacia spp.</i>	716	9.82	6.85	2.54	43.68	7.00	3.54	0.91	18.28
		Muskwood	<i>Guarea guidonia</i>	3151	18.49	15.13	2.54	140.20	10.49	4.91	0.30	31.08

Table 1 Continued.

	Pink trumpet tree	<i>Tabebuia heterophylla</i>	1998	13.18	8.68	2.54	58.42	9.07	4.58	0.60	23.46
	African tulip tree	<i>Spathodea campanulata</i>	2987	21.15	13.74	2.54	130.30	12.30	5.41	0.30	34.13
	Acacia	<i>Acacia spp.</i>	1899	24.87	18.11	2.54	161.29	13.25	5.35	2.43	35.96
PAC	African tulip tree	<i>Spathodea campanulata</i>	285	20.13	12.21	2.54	70.86	12.22	5.32	2.74	27.73
	‘Ōhi‘a lehua	<i>Metrosideros polymorpha</i>	8251	19.13	13.79	2.54	157.73	11.66	5.48	1.52	39.92

$$h = \frac{\beta_0}{1 + \beta_1 e^{(\beta_2 DBH + \beta_3 CI)}} \quad (1)$$

where h is total tree height in m, DBH is diameter at breast height in cm, CI is competition index, β_0 , β_1 , β_2 and β_3 are model coefficients. To construct the base model, $\beta_3 CI$ in Equation (1) was removed and other variables remained the same.

3.3.2 Parametric model 2 – Wykoff et al. (1982)

Wykoff et al. (1982) developed a statistical model to estimate the missing heights of trees, which is another commonly used H-D model in forestry. In our preliminary analysis, the original form of the Wykoff et al. (1982) model, denoted as WK model, did not fit well for our data. Thus, a modified version of the WK model was used in the formal analysis, which can be written as:

$$h = e^{\beta_0 - \frac{1}{|\beta_1 DBH + \beta_2 CI|}} \quad (2)$$

where all variables and symbols have been defined as above. Similar to the PR model, the base model was built using Equation 2 without $\beta_2 CI$ in the model.

The ‘nlsLM’ function from the R package ‘minpack.lm’ (Elzhov et al., 2022) was used to estimate the model coefficients. Initial values of the model coefficients were selected based on previous studies (e.g., Misik et al., 2016, Yang et al., 2022).

3.3.3 Semi-parametric model 1 – General additive model (GAM)

GAM is the blended product of generalized linear models and additive models that uses smoothing functions of the predictor variables to capture any non-linear relationship between the response and the predictor variables. The relationship between the response variable (tree height) and the predictor variables (DBH and competition index) is modeled as a sum of smooth functions of the predictor variables which can be represented as:

$$h = \beta_0 + f_1x_1 + f_2x_2 + \dots + f_kx_k \quad (3)$$

where h is the total tree height (m), β_0 is the model intercept and f_1 to f_k are the smooth functions for predictors x_1 to x_k . The smooth functions f_1 to f_k are used to find the signal in predictor variables to make predictions for the tree height without being constrained by predefined assumptions about the distribution of the data. The ‘gam’ function from the R package ‘mgcv’ (Wood, 2017) was used to fit the model.

3.3.4 Nonparametric model 1 – Extreme Gradient Boosting (XGBoost)

The working mechanism of XGBoost involves minimizing the loss function using gradient descent optimization, while incorporating regularization techniques to prevent overfitting and improve generalization performance (Chen & Guestrin, 2016). The input data is a matrix X of dimension $n \times p$ where n is the number of observations and p is the number of predictor variables (DBH and CI in this case). X can be represented as $X = [x_1, x_2, \dots, x_p]$ where x_i is a column vector of size n containing the values of the i th predictor variable for all n observations. The predicted height ($\hat{h}$) can be represented as:

$$\begin{aligned} \hat{h} = w_0 + w_1d_1(X) + w_2d_2(X) \\ + \dots w_kd_k(X) \end{aligned} \quad (4)$$

where $\hat{h}$ is the predicted tree height $w_0, w_1, \dots, w_k$ are the weights assigned to each decision tree, $d_i(X)$ is the output of the i th decision tree for input X , and k is the number of decision trees in the model. The optimal weights and decision trees are computed using the objective function. The objective function of XGBoost is composed of a loss function and regularization parameters, given by,

$$Obj = L(h, \hat{h}) + \Omega(d_k(X)) \quad (5)$$

where $L(h, \hat{h}) = \sum l(h_i, \hat{h}_i)$ is the sum of differentiable loss function for the i th observation that measures the difference between the predicted tree height ($\hat{h}_i$) and observed tree height (h_i), Ω represents the regularization parameters that penalizes the complex model and $d_k(X)$ is the k th decision tree in the ensemble model. The loss function is minimized using the gradient descent optimization technique where the gradient of the loss function with respect to the predicted values for each observation is represented as:

$$g_i = \frac{d(l(h_i, \hat{h}_i))}{d\hat{h}_i} \quad (6)$$

The gradient descent algorithm then updates the predicted height ($\hat{h}_{k+1}$) by adding a shrinkage parameter (λ) multiplied by the output of the i th decision tree ($d_i(X)$) with the predicted height from previous decision ($\hat{h}_k$) tree given as:

$$\hat{h}_{k+1} = \hat{h}_k + \lambda d_i(X) \quad (7)$$

The XGBoost model has seven hyperparameters that need to be tuned for optimal model performance. They are: nrounds, eta, max_depth, colsample_bytree, subsample, gamma and min_child_weight for every given species in a region. These hyperparameters were tuned using 5-fold cross-validation method using the ‘caret’ package in R (Kuhn, 2008). The results of parameter tuning for the XGBoost model for a given species in a region are shown in Supplementary Table 1.

3.3.5 Nonparametric model 2 – Artificial Neural Networks (ANN)

Artificial neural network (ANN) is a powerful machine learning algorithm that has been used to solve highly complex prediction problems (Abiodun et al., 2018). ANNs are based on a simplified

model of biological neural systems where the output from a layer of nodes serves as input for the next layer of nodes (Walczak & Cerpa, 2003). ANNs can have multiple layers of nodes which includes the input and output layers that have same number of nodes as the number of input and output variables respectively (Walczak & Cerpa, 2003). The layers between the input and output layers are called hidden layers. The hidden layers are generally composed of multiple dense layers and dropout layers specified by the researcher. Dense layers have specified number of nodes for training the model and dropout layers are used to drop some of the nodes from the previous layer based on a specified dropout rate to prevent overfitting. The output of a particular node is given by an activation function with the sum of the weighted inputs and biases from the previous dense layer. ANN uses back propagation training algorithm that allows the ANN to iteratively adjust the weight and biases of the nodes by propagating the error from the output layer back through the network.

We employed an ANN model with four hidden layers between the input and output layers. The input layer was passed through a dense layer followed by a dropout layer, followed by another dense layer and a dropout layer and finally an output layer (see Supplementary Figure 1). The rectified linear unit (ReLU) activation function was used for the dense layers. For each of the ANN models, the number of nodes in dense layers, the dropout rate of the dropout layers and the batch size was tuned for optimal model performance using the ‘tfruns’ package in R (Kalinowski et al., 2022). The result of parameter tuning for ANN model for a given species in a region are shown in Supplementary Table 2.

3.4 Competition index (CI)

Competition between neighboring trees is a primary determinant of resource availability for tree establishment and growth (Huang & Titus, 1995, Stage, 1973). Several distance-dependent and

distance-independent competition indices have been developed to mathematically represent the level of competition for a target tree (Burkhart & Tomé, 1989). In this study, we selected five commonly used distance-independent competition indices (CI's).

3.4.1 Diameter ratio

Diameter ratio (R_{dia}) is the ratio of DBH of subject tree to the average DBH on a sampling unit. The quadratic mean diameter (QMD) was used as the average DBH instead of arithmetic mean diameter because QMD tends to better represent the average diameter in a stand with a wide range of diameters (Curtis & Marshall, 2000).

$$R_{dia} = \frac{DBH}{QMD} \quad (8)$$

where all variables have been defined as above.

3.4.2 Basal area ratio

Similar to R_{dia} , basal area ratio (R_{BA}) is the ratio of individual tree basal area (BA) to the mean tree BA ($\overline{BA}$) calculated from all sample trees on a sampling unit, which can be expressed as:

$$R_{BA} = \frac{BA}{\overline{BA}} \quad (9)$$

where BA is the basal area of tree and $\overline{BA}$ is the mean basal area in the condition class.

3.4.3 Partitioned stand density index

Stand density index (SDI) proposed by Reineke (1933) was a standardized measure of stand density. Stage (1968) showed that Reineke's SDI can be partitioned into additive components at the individual tree and group levels. Partitioned SDI (PSDI) divides the stand density index at the individual tree level and was used as an individual tree competition index by Sun et al. (2019), which is expressed as:

$$PSDI = EF(a + bd^2) \quad (10)$$

where EF is the expansion factor which is constant for FIA subplots = 6.018046 (Burrill et al., 2021), d is the diameter at breast height of subject tree, a is the weight of the presence of subject tree, b is the weight of size of subject tree (Lu et al., 2017). They are given as:

$$a = 25^{-c} \left(1 - \frac{c}{2}\right) QMD \quad (11)$$

$$b = 25^{-c} \left(\frac{c}{2}\right) QMD^{c-2} \quad (12)$$

where c is the self-thinning constant = 1.605 (Reineke, 1933).

3.4.4 Cumulative distribution function

The cumulative distribution function provides the observed cumulative probability of the individual tree in a sampling unit (Poudel & Cao, 2013). It was used as competition index by Sun et al. (2019) to illustrate the relative position of a loblolly pine tree in a population, which can be written as:

$$CDF = \frac{r - 0.5}{n} \quad (13)$$

where r is the rank of tree in terms of DBH (from largest to smallest) and n is the number of trees in the condition class.

3.4.5 Basal area of larger trees

Proposed by Wykoff et al. (1982), basal area in larger trees (BAL) was calculated by summing up the BA of all trees larger than the subject tree based on DBH. BAL can be used to quantify the relative dominance of the subject tree in the population. A greater BAL implies that the subject tree has a smaller diameter than other trees.

3.5 Model building and evaluation

For a given species in a region, all sampling units were randomly divided into fit and test datasets with an 80/20 split. Specifically, fit dataset includes 80% of the sampling units to construct the models, and the test dataset was composed of the remaining 20% of the data for model evaluation. To account for the correlation among sample trees within the same sampling unit, the cluster bootstrap technique was used to estimate the confidence intervals of the model parameters and evaluation statistics. The cluster bootstrap was applied at the sampling unit (condition class) level, rather than at the individual tree level. Sampling units in the fit and test data were sampled with replacement, respectively, and all trees within the selected sampling units were included in model building and evaluation. A total of 500 bootstrapped samples were drawn for a given combination of species and region. The 50% quantile of the 500 bootstrapped samples served as the point estimate of the model parameters and evaluation statistics. The 97.5% and 2.5% quantiles of the 500 bootstrapped samples were used as the upper and lower limits of the 95% confidence intervals for the model parameters and evaluation statistics.

Evaluation statistics were computed using the test data, including coefficient of determination (R^2) which represents the amount of variance explained, mean bias (MB) represents precision and root-mean-square error (RMSE) represents accuracy of the model. They were computed as:

$$R^2 = 1 - \frac{\sum Res^2}{\sum (h - \bar{h})^2} \quad (14)$$

$$MB = \frac{\sum Res}{n} \quad (15)$$

$$RMSE = \left[\frac{\sum Res^2}{n} \right]^{\frac{1}{2}} \quad (16)$$

where n is total number of observations for a given species in a region, and Res is prediction residual, which was defined as:

$$Res = h - \hat{h} \quad (17)$$

where h is the observed total tree height and $\hat{h}$ is the predicted total tree height of the individual tree.

The relative rank proposed by Poudel & Cao (2013) was used to rank the models. The evaluation statistics computed for every combination of regression method and competition index was used to rank the models for a given species in a region. The relative rank for absolute MB and RMSE is given as:

$$Rank_i = 1 + \frac{(m - 1)(ES_i - ES_{min})}{ES_{max} - ES_{min}} \quad (18)$$

$$Rank_i = m - \frac{(m - 1)(ES_i - ES_{min})}{ES_{max} - ES_{min}} \quad (19)$$

where $Rank_i$ is the rank of i th combination of regression method and competition index for a given species in a region, m is the number of models evaluated. ES_i is the evaluation statistics of the i th combination of regression method and competition index, ES_{min} and ES_{max} are the minimum and maximum values of evaluation statistics for a given species in a region. Similarly, the relative rank for R^2 is given Equation (19) where the symbols are same as defined above. The ranks among the three evaluation statistics were averaged to compute the final rank of the model. The base models were ranked separately to identify the best regression methods, and all models were ranked together to identify the best combination of regression method and competition index.

CHAPTER 4: RESULTS

4.1 Predictability of the base models on temperate trees

In general, the GAM model was found to have the highest rank for all the species across the regions (Figure 2A). The PR model was found to have the highest rank for the temperate species in AP and XGBoost and GAM were found to have the highest rank for species in VR (Figure 2B). The GAM model provided the least biased predictions for two species in AP and three species in VR and the lowest RMSE for two species in AP and three species in VR (see Table 2, 3). The evaluation statistics for GAM, XGBoost and PR models are similar in VR (see Table 3). The PR model produced the least biased predictions for two species in AP and two species in VR. The WK model provided the least biased predictions for only one species and did not provide the highest R^2 or least RMSE for any of the species. However, it did provide a higher R^2 for three species and a lower RMSE for two species in AP compared to either XGBoost or ANN models. The predictive performance of XGBoost and ANN models had more variation compared to other regression methods in the temperate region. Although XGBoost and ANN produced the highest average rank for two species in AP and two species in VR respectively, they also produced the lowest overall ranks for other species in the temperate region (Figure 3A, 3B).

4.2 Predictability of base models on pantropical trees

The XGBoost model performed the best for species in CAR and GAM performed the best for species in PAC (Figure 2B). XGBoost produced the least biased prediction for one species and most accurate (lowest RMSE) predictions for two species in CAR. It also had the lowest overall rank for all the species in PAC (Figure 3D). GAM produced the most precise predictions for one species in PAC and most accurate predictions for two species in CAR and two species in PAC.

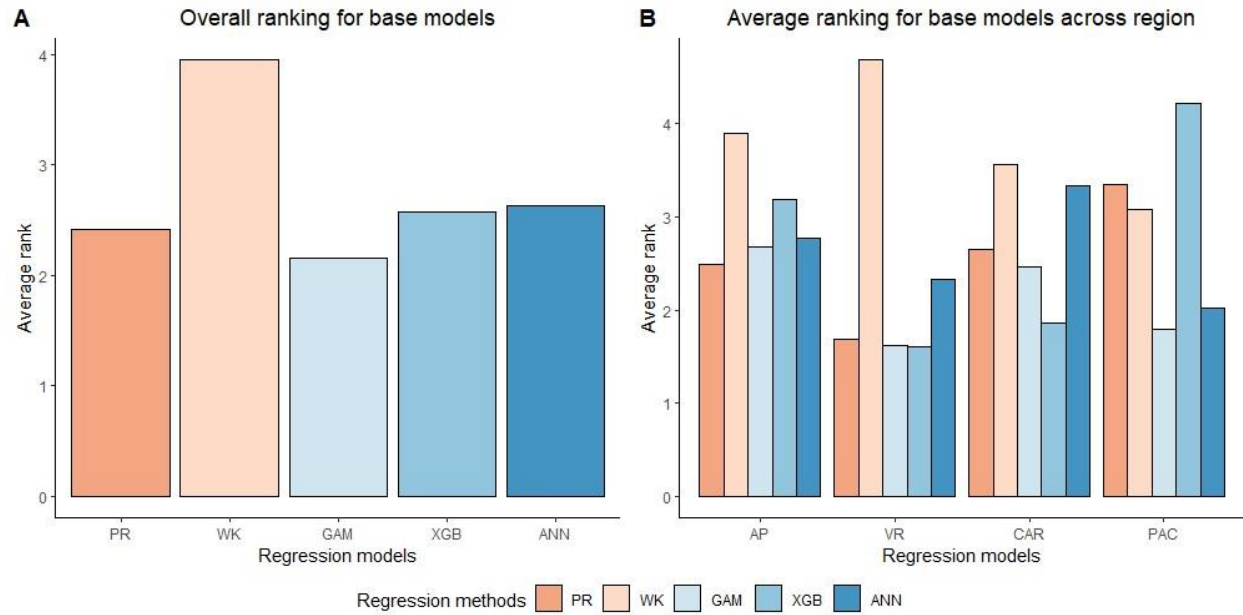


Figure 2. Overall ranking (Lowest rank is the best) of five base regression methods (Pearl & Reed (1920), PR; Wykoff et al., 1982, WK; Generalized additive model, GAM; Extreme gradient boost, XGB; Artificial neural network, ANN) across all the species and physiographic region in (A) and across all species in four physiographic regions (Apalachian plateau, AP; Ridge and Valley, VR; Caribbean islands, CAR; Pacific islands, PAC) in (B).

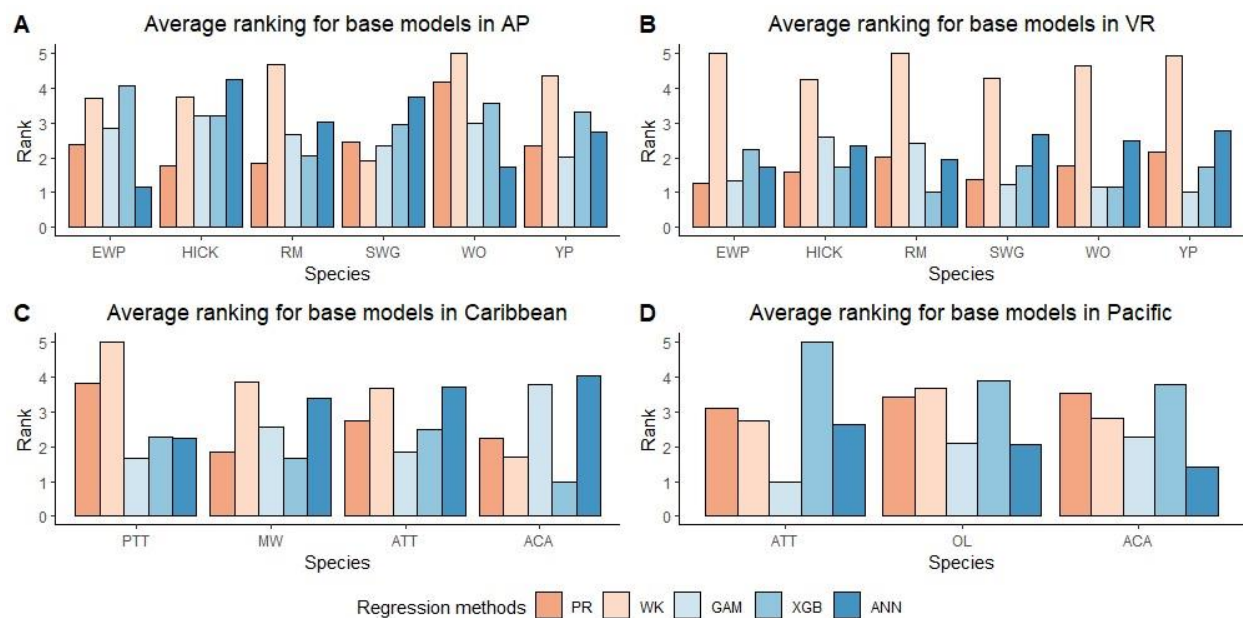


Figure 3. Average rank (Lowest rank is the best) of five base regression methods for species in Appalachian plateau (AP) in (A) (eastern white pine, EWP; hickory species, HICK; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar, YP), Ridge and Valley (VR) in (B) (eastern white pine, EWP; hickory species, HICK; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar, YP), Caribbean islands (CAR) in (C) (pink trumpet tree, PTT; muskwood, MW; African tulip tree, ATT; acacia species, ACA) and Pacific islands (PAC) in (D) (African tulip tree, ATT; ‘Ōhi‘a lehua, OL; acacia species, ACA).

Table 2. Median of the evaluation statistics (coefficients of determination, R^2 ; mean bias, MB; root mean squared error, RMSE) for the base models (i.e., no competition index included in the model) and five competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) using five regression models (Pearl & Reed (1920), PR; Wykoff et al. (1982), WK; Generalized additive model, GAM; Extreme Gradient Boosting, XGboost; Artificial Neural Network, ANN) for the six temperate trees species (eastern white pine, EWP; hickory species, HICK; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar, YP) in Appalachian plateau (AP).

Sp.	Model	R^2						MB						RMSE					
		Base	R_{dia}	R_{BA}	CDF	BAL	PSDI	Base	R_{dia}	R_{BA}	CDF	BAL	PSDI	Base	R_{dia}	R_{BA}	CDF	BAL	PSDI
EWP	PR	0.69	0.71	0.70	0.65	0.65	0.65	0.36	0.35	0.32	0.33	0.34	0.34	3.44	3.28	3.40	3.43	3.44	3.39
	WK	0.65	0.66	0.64	0.50	0.60	0.65	0.50	0.53	0.44	0.46	0.47	0.38	3.55	3.47	3.56	3.55	3.55	3.50
	GAM	0.68	0.69	0.63	0.73	0.72	0.72	0.46	0.57	0.68	- 0.14	- 0.13	-0.12	3.44	3.38	3.91	3.22	3.24	3.24
	XGB	0.60	0.67	0.71	0.71	0.73	0.71	0.21	-0.05	0.18	0.18	- 0.06	0.17	3.76	3.15	3.63	3.64	3.52	3.62
	ANN	0.68	0.69	0.68	0.60	0.67	0.69	0.09	-0.02	0.07	- 0.01	0.07	0.18	3.24	3.19	3.26	0.18	3.29	3.21
HICK	PR	0.72	0.73	0.72	0.71	0.71	0.71	-0.02	-0.01	- 0.02	- 0.02	- 0.03	-0.02	3.66	3.63	3.65	3.66	3.66	3.65

Table 2 Continued.

	WK	0.70	0.70	0.70	0.70	0.68	0.69	0.13	0.15	0.11	0.12	0.13	0.05	3.74	3.72	3.73	3.73	3.74	3.69
	GAM	0.70	0.71	0.64	0.70	0.70	0.70	0.28	0.23	- 1.12	0.28	0.27	0.27	3.57	3.49	3.91	3.57	3.56	3.57
	XGB	0.70	0.70	0.70	0.70	0.71	0.70	0.31	0.32	0.31	0.31	0.28	0.33	3.55	3.56	3.55	3.55	3.51	3.58
	ANN	0.69	0.70	0.69	0.55	0.69	0.69	0.28	0.29	0.27	- 0.01	0.30	0.38	3.65	3.59	3.66	0.19	3.64	3.63
RM	PR	0.61	0.62	0.61	0.61	0.61	0.61	0.01	0.03	0.00	0.01	0.01	0.00	3.26	3.23	3.26	3.26	3.26	3.26
	WK	0.58	0.59	0.60	0.59	0.59	0.61	0.13	0.15	0.08	0.10	0.12	0.04	3.38	3.36	3.32	3.36	3.38	3.29
	GAM	0.61	0.63	0.62	0.62	0.63	0.62	-0.17	-0.16	- 0.19	- 0.02	- 0.02	-0.01	3.19	3.12	3.28	3.25	3.24	3.25
	XGB	0.62	0.67	0.63	0.63	0.63	0.62	0.12	0.08	0.08	0.10	0.10	0.08	3.21	3.09	3.18	3.19	3.15	3.18
	ANN	0.59	0.61	0.60	0.63	0.60	0.60	-0.03	0.01	0.05	0.00	0.04	0.13	3.31	3.25	3.31	0.17	3.30	3.30
SWG	PR	0.81	0.83	0.82	0.71	0.73	0.81	-0.30	-0.27	- 0.32	- 0.32	- 0.25	-0.33	3.32	3.27	3.24	3.32	3.30	3.29
	WK	0.78	0.78	0.79	0.72	0.73	0.78	0.13	0.18	0.03	0.00	0.16	-0.03	3.54	3.55	3.43	3.43	3.57	3.32
	GAM	0.83	0.85	0.72	0.74	0.76	0.75	-0.33	-0.23	- 0.26	- 0.95	- 0.86	-0.93	3.23	3.02	3.69	3.51	3.35	3.46
	XGB	0.74	0.75	0.74	0.74	0.69	0.74	-0.24	0.09	- 0.33	- 0.31	- 0.38	-0.34	3.46	3.78	3.44	3.44	3.74	3.45

Table 2 Continued.

	ANN	0.70	0.71	0.70	0.41	0.71	0.68	0.14	-0.09	0.12	0.06	0.22	0.17	4.23	4.19	4.26	0.22	4.13	4.34
WO	PR	0.65	0.66	0.65	0.65	0.65	0.65	0.22	0.19	0.21	0.22	0.21	0.22	3.48	3.38	3.48	3.48	3.48	3.48
	WK	0.64	0.65	0.65	0.65	0.65	0.64	0.27	0.25	0.26	0.26	0.25	0.25	3.51	3.44	3.50	3.51	3.51	3.49
	GAM	0.64	0.67	0.56	0.66	0.66	0.66	0.02	0.05	1.74	0.22	0.22	0.22	3.42	3.30	3.84	3.40	3.41	3.40
	XGB	0.64	0.70	0.66	0.65	0.66	0.66	-0.04	0.17	-	-	-	-0.03	3.48	3.22	3.40	3.43	3.39	3.40
										0.02	0.04	0.03							
	ANN	0.68	0.70	0.68	0.56	0.68	0.68	0.16	0.16	0.18	0.01	0.14	0.31	3.33	3.24	3.33	0.19	3.34	3.31
YP	PR	0.72	0.73	0.72	0.73	0.73	0.71	-0.13	-0.13	-	-	-	-0.14	3.83	3.76	3.81	3.84	3.83	3.82
										0.14	0.14	0.13							
	WK	0.70	0.70	0.70	0.71	0.71	0.69	0.13	0.17	0.09	0.09	0.08	0.00	4.02	4.01	4.01	3.99	4.01	3.93
	GAM	0.73	0.75	0.69	0.74	0.74	0.74	0.01	0.00	-	-	-	-0.01	3.93	3.83	4.07	3.70	3.56	3.70
										1.12	0.02	0.02							
	XGB	0.71	0.77	0.72	0.72	0.72	0.72	-0.02	-0.10	-	-	-	-0.08	4.01	3.60	3.93	3.93	3.93	3.93
										0.08	0.03	0.07							
	ANN	0.74	0.75	0.74	0.55	0.74	0.74	0.24	0.23	0.24	0.02	0.24	0.35	3.89	3.81	3.89	0.19	3.89	3.88

Note: Bold evaluation statistics indicate the combination (model + competition index) with the greatest R^2 value, the closest to zero

MB or the lowest RMSE for a given species.

Table 3. Median of the evaluation statistics (coefficients of determination, R^2 ; mean bias, MB; root mean squared error, RMSE) for the base models (i.e., no competition index included in the model) and five competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) using five regression models (Pearl & Reed (1920), PR; Wykoff et al. (1982), WK, WK; Generalized additive model, GAM; Extreme Gradient Boosting, XGboost; Artificial Neural Network, ANN) for the six temperate trees species (eastern white pine, EWP; hickory species, HICK; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar, YP) in Ridge and Valley (VR).

Sp.	Model	R^2						MB						RMSE					
		Base	R_{dia}	R_{BA}	CDF	BAL	PSDI	Base	R_{dia}	R_{BA}	CDF	BAL	PSDI	Base	R_{dia}	R_{BA}	CDF	BAL	PSDI
EWP	PR	0.83	0.83	0.78	0.80	0.80	0.80	-0.08	-0.07	-	-	-	-0.43	2.86	2.81	3.02	2.98	2.92	2.89
										0.13	0.46	0.47							
	WK	0.77	0.77	0.73	0.75	0.73	0.78	0.54	0.57	0.07	-	0.03	-0.27	3.29	3.28	3.40	3.29	3.36	3.08
											0.06								
	GAM	0.82	0.83	0.79	0.82	0.82	0.82	0.08	0.06	0.68	0.08	0.10	0.08	2.81	2.79	3.04	2.82	2.81	2.82
	XGB	0.80	0.77	0.80	0.79	0.80	0.80	-0.18	-0.24	-	-	-	-0.13	2.87	3.05	2.88	2.89	2.84	2.88
										0.19	0.16	0.14							
	ANN	0.80	0.81	0.80	0.80	0.80	0.80	0.03	0.01	0.00	-	-	-0.10	2.84	2.78	2.85	2.85	2.85	2.84
											0.07	0.07							
HICK	PR	0.77	0.78	0.77	0.77	0.77	0.77	0.05	0.02	0.05	0.04	0.04	0.05	3.27	3.22	3.25	3.27	3.27	3.26

Table 3 Continued.

	WK	0.74	0.74	0.75	0.74	0.74	0.76	0.38	0.38	0.27	0.30	0.35	0.20	3.50	3.47	3.41	3.46	3.49	3.33
	GAM	0.76	0.77	0.73	0.76	0.76	0.76	-0.05	-0.07	-	-	-	-0.06	3.40	3.28	3.62	3.40	3.40	3.40
										0.85	0.06	0.05							
	XGB	0.77	0.76	0.77	0.77	0.77	0.79	-0.07	-0.09	-	-	-	0.19	3.29	3.34	3.29	3.30	3.28	3.24
										0.07	0.06	0.07							
	ANN	0.79	0.80	0.79	0.79	0.79	0.79	0.31	0.33	0.24	0.20	0.21	0.21	3.26	3.14	3.26	3.25	3.25	3.23
RM	PR	0.70	0.71	0.70	0.69	0.69	0.69	-0.05	-0.04	-	-	-	-0.09	3.16	3.13	3.16	3.14	3.14	3.14
										0.04	0.10	0.10							
	WK	0.64	0.64	0.66	0.66	0.65	0.67	0.08	0.08	0.00	0.03	0.04	-0.02	3.39	3.40	3.29	3.32	3.37	3.24
	GAM	0.69	0.70	0.69	0.69	0.69	0.69	0.05	0.07	0.11	0.05	0.04	0.05	3.20	3.12	3.22	3.20	3.19	3.19
	XGB	0.71	0.70	0.71	0.71	0.71	0.71	-0.02	-0.03	-	-	-	0.07	3.13	3.17	3.13	3.13	3.12	3.13
										0.02	0.02	0.02							
	ANN	0.71	0.72	0.71	0.71	0.71	0.71	0.06	0.10	0.02	0.00	-	0.04	3.14	3.07	3.13	3.14	3.13	3.13
												0.01							
SWG	PR	0.81	0.83	0.70	0.82	0.83	0.82	-0.13	-0.08	-	-	-	-0.05	3.04	2.94	3.18	3.11	3.09	3.12
										0.13	0.06	0.04							
	WK	0.76	0.76	0.78	0.78	0.77	0.79	0.19	0.19	0.18	0.42	0.42	0.28	3.46	3.46	3.51	3.50	3.57	3.41
	GAM	0.81	0.82	0.77	0.81	0.81	0.81	-0.04	-0.04	-	-	-	-0.03	3.11	2.99	3.44	3.11	3.09	3.11
										1.11	0.03	0.04							

Table 3 Continued.

	XGB	0.81	0.80	0.81	0.81	0.81	0.81	0.16	0.17	0.17	0.17	0.13	0.25	3.13	3.23	3.13	3.14	3.14	3.13
	ANN	0.81	0.82	0.81	0.81	0.81	0.81	0.37	0.35	0.34	0.28	0.24	0.23	3.15	3.06	3.15	3.14	3.14	3.13
WO	PR	0.73	0.75	0.74	0.72	0.72	0.72	0.19	0.15	0.17	0.13	0.13	0.13	3.25	3.18	3.33	3.40	3.40	3.39
	WK	0.71	0.72	0.72	0.71	0.71	0.71	0.39	0.37	0.35	0.29	0.29	0.24	3.37	3.33	3.46	3.50	3.49	3.45
	GAM	0.74	0.76	0.73	0.74	0.74	0.74	0.22	0.19	0.35	0.22	0.22	0.21	3.21	3.10	3.26	3.21	3.20	3.20
	XGB	0.74	0.73	0.74	0.74	0.75	0.74	0.22	0.25	0.22	0.22	0.22	0.31	3.21	3.28	3.22	3.23	3.17	3.20
	ANN	0.74	0.75	0.74	0.74	0.74	0.74	0.46	0.40	0.38	0.34	0.33	0.36	3.23	3.13	3.22	3.22	3.22	3.19
YP	PR	0.79	0.80	0.81	0.81	0.81	0.82	0.05	0.01	-	-	-	-0.17	3.68	3.59	3.53	3.60	3.60	3.58
										0.11	0.17	0.18							
	WK	0.76	0.76	0.77	0.78	0.77	0.79	0.32	0.34	0.14	0.15	0.24	0.08	3.95	3.95	3.87	3.92	3.99	3.79
	GAM	0.82	0.83	0.81	0.82	0.82	0.82	-0.02	0.02	-	-	-	-0.01	3.58	3.44	3.62	3.58	3.58	3.58
										0.17	0.02	0.03							
	XGB	0.81	0.80	0.81	0.81	0.82	0.82	0.11	0.12	0.13	0.12	0.12	0.26	3.62	3.74	3.62	3.63	3.60	3.57
	ANN	0.81	0.83	0.81	0.81	0.81	0.82	0.34	0.35	0.28	0.21	0.22	0.24	3.64	3.46	3.64	3.63	3.62	3.55

Note: Bold evaluation statistics indicate the combination (model + competition index) with the greatest R^2 value, the closest to zero MB or the lowest RMSE for a given species.

The ANN model produced the most precise predictions for one species in CAR and most accurate predictions for one species in PAC, respectively. The WK model had comparatively closer to zero MB values for species in PAC than other regions but also had the highest RMSE values for two of the species in PAC (see Table 4).

4.3 Predictability of models with and without competition index

The inclusion of competition index improved the predictive abilities of most of the models for the temperate and pantropical species. On average, R_{dia} was found to have the highest rank for PR and GAM models (Figure 4A, 4C). PSDI was found to have the highest rank for WK models (Figure 4B). Similarly, CDF was found to have the highest rank for XGBoost and ANN models (Figure 4D, 4E). The R^2 values for every species increased by an average of 4.8% and absolute MB and RMSE decreased by an average of 79.3% and 48.8% respectively in the temperate region. Similarly, the R^2 value increased by an average of 4.0% and absolute MB and RMSE decreased by an average of 43.3% and 10.3% respectively in the pantropical region (see Table 5). However, the magnitude of improvement varied by species. The inclusion of competition index increased the R^2 value by a minimum of 1.3% for hickory spp. in AP and 0.0% for eastern white pine in VR to a maximum of 21.6% for eastern white pine in AP and 2.7% for white oak in VR (see Table 5). Among the pantropical species, R^2 values did not increase for muskwood in CAR and acacia in PAC (see Table 5).

4.4 Best combinations of regression method and competition index for different regions

The combination of regression method and competition index that had the highest rank was different for the four physiographic regions. The combination of ANN and CDF ranked the

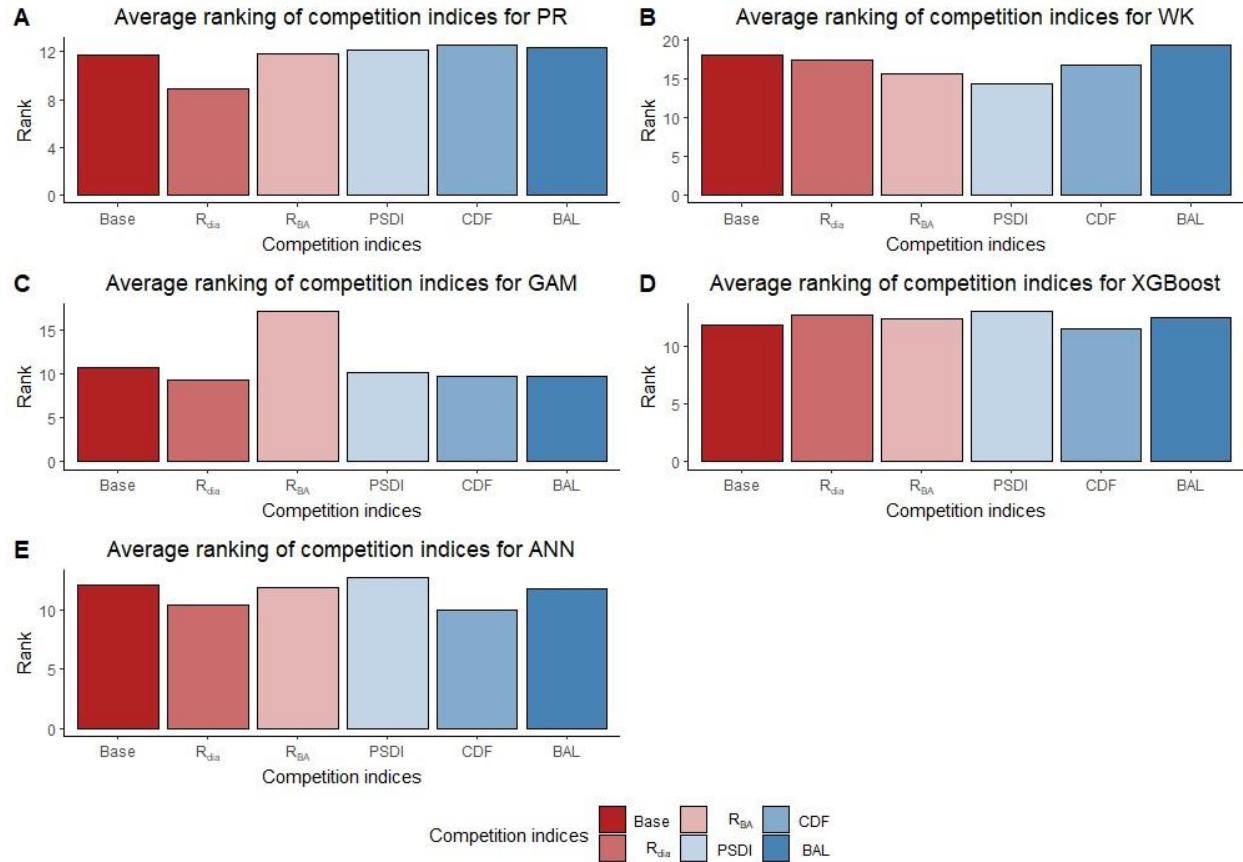


Figure 4. Average ranking (Lowest rank is the best) of competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) for five regression models (Pearl & Reed (1920), PR (A); Wykoff et al. (1982), WK (B); Generalized additive model, GAM (C); Extreme Gradient Boosting, XGboost (D); Artificial Neural Network, ANN (E)) across all species and physiographic regions.

Table 4. Median of the evaluation statistics (coefficients of determination, R^2 ; mean bias, MB; root mean squared error, RMSE) for the base models (i.e., no competition index included in the model) and five competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) using five regression models (Pearl & Reed (1920), PR; Wykoff et al. (1982), WK; Generalized additive model, GAM; Extreme Gradient Boosting, XGboost; Artificial Neural Network, ANN) for four pantropical trees species in Caribben islands (CAR) (pink trumpet tree, PTT; muskwood, MW; African tulip tree, ATT; acacia, ACA) and three pantropical species in Pacific islands (PAC) (African tulip tree, ATT; ‘Ōhi‘a lehua, OL, acacia, ACA).

		R ²							MB							RMSE						
Reg.	Sp.	Model	Base	R _{dia}	R _{BA}	CDF	BAL	PSDI	Base	R _{dia}	R _{BA}	CDF	BAL	PSDI	Base	R _{dia}	R _{BA}	CDF	BAL	PSDI		
CAR	PTT	PR	0.67	0.72	0.70	0.70	0.72	0.70	0.91	0.10	0.14	0.13	0.13	0.14	2.80	2.36	2.46	2.45	2.39	2.44		
		WK	0.61	0.66	0.66	0.66	0.65	0.67	0.99	0.29	0.30	0.24	0.27	0.29	3.05	2.62	2.63	2.63	2.66	2.59		
		GAM	0.73	0.68	0.71	0.73	0.75	0.72	-0.55	-0.50	-	-	-	-0.55	2.19	2.38	2.28	2.19	2.12	2.23		
											0.50	0.52	0.45									
		XGB	0.68	0.67	0.61	0.68	0.61	0.63	-0.52	-0.47	-	-	-	-0.42	2.26	2.31	2.51	2.29	2.50	2.43		
											0.53	0.53	0.58									
	ANN	0.66	0.68	0.65	0.66	0.68	0.66	0.12	-0.04	0.23	0.29	0.04	0.11	2.48	2.40	2.51	2.48	2.42	2.47			
	MW	PR	0.70	0.64	0.64	0.64	0.64	0.64	-0.11	-0.02	0.00	0.00	0.00	0.01	2.87	2.30	3.01	2.99	2.99	3.00		
		WK	0.64	0.61	0.62	0.61	0.61	0.42	0.17	0.29	0.29	0.28	0.28	0.69	3.14	3.15	3.11	3.13	3.15	3.82		

Table 4 Continued.

	GAM	0.74	0.73	0.72	0.73	0.73	0.74	-0.25	-0.22	-	-	-	-0.24	2.29	2.32	2.53	2.30	2.29	2.29
										0.20	0.26	0.23							
	XGB	0.61	0.60	0.55	0.61	0.55	0.56	-0.36	-0.38	-	-	-	-0.41	3.05	3.09	3.28	3.05	3.26	3.23
										0.38	0.39	0.38							
	ANN	0.66	0.66	0.73	0.67	0.66	0.66	-0.15	-0.18	6.45	0.04	-	-0.18	2.54	2.56	8.23	2.51	2.54	2.55
												0.17							
ATT	PR	0.38	0.40	0.39	0.39	0.40	0.39	-1.71	-1.71	-	-	-	-1.66	4.26	4.23	4.25	4.26	4.23	4.25
										1.71	1.71	1.75							
	WK	0.35	0.36	0.37	0.36	0.35	0.24	-1.53	-1.49	-	-	-	-1.65	4.39	4.36	4.35	4.35	4.38	4.75
										1.59	1.54	1.53							
	GAM	0.64	0.66	0.62	0.64	0.63	0.65	0.29	0.34	0.32	0.34	0.44	0.26	3.12	3.03	3.20	3.12	3.13	3.09
	XGB	0.70	0.69	0.67	0.69	0.64	0.67	0.60	0.72	0.49	0.65	0.59	0.55	2.90	2.91	3.03	2.92	3.16	3.04
	ANN	0.69	0.70	0.69	0.69	0.69	0.68	-0.29	-0.33	-	0.03	-	-0.28	2.76	2.72	2.79	2.78	2.79	2.81
										0.23		0.14							
ACA	PR	0.40	0.58	0.50	0.50	0.47	0.50	-0.33	0.01	0.61	0.63	0.70	0.64	1.96	1.65	2.94	2.96	3.02	2.95
	WK	0.46	0.58	0.53	0.49	0.48	0.49	-0.30	0.58	0.55	0.65	0.70	0.65	1.86	2.95	2.86	2.99	3.00	2.99
	GAM	0.53	0.62	0.53	0.53	0.54	0.53	1.12	1.24	1.09	1.14	1.04	1.15	2.82	2.54	2.82	1.14	2.80	2.84
	XGB	0.55	0.42	0.38	0.54	0.39	0.30	0.22	-0.15	0.20	0.20	0.20	-0.09	1.78	2.02	2.09	1.80	2.07	2.21

Table 4. Continued.

		ANN	0.31	0.20	0.31	0.36	0.28	0.20	-0.93	-1.13	-	-	-	-0.99	2.30	2.48	2.30	2.21	2.41	2.68
											0.94	0.72	0.93							
PAC	ATT	PR	0.57	0.61	0.60	0.56	0.55	0.63	-1.62	1.20	1.33	1.35	1.38	1.20	2.83	3.23	3.24	3.40	3.45	3.12
		WK	0.48	0.53	0.53	0.55	0.53	0.46	-1.00	1.35	1.56	1.40	1.40	1.08	3.10	3.54	3.52	3.46	3.55	3.79
		GAM	0.68	0.70	0.64	0.68	0.56	0.70	-0.93	-0.75	-	-	-	-0.76	2.07	2.01	2.24	2.09	2.47	2.00
											0.91	0.88	0.86							
		XGB	0.35	0.50	0.35	0.49	0.40	0.24	-1.81	-0.24	-	-	0.13	-1.74	3.70	3.91	3.72	3.25	4.25	5.94
											1.83	1.25								
		ANN	0.51	0.54	0.50	0.35	0.30	0.51	-1.02	-0.61	1.64	2.97	1.97	3.05	3.07	3.12	3.55	4.68	4.44	4.73
OL	PR		0.60	0.55	0.55	0.54	0.54	0.54	-0.30	-0.19	-	-	-	-0.22	3.35	3.35	3.37	3.39	3.39	3.40
											0.22	0.24	0.22							
		WK	0.54	0.51	0.52	0.51	0.53	0.52	-0.04	-0.07	-	-	-	-0.13	3.57	3.50	3.46	3.49	3.42	3.45
											0.15	0.13	0.18							
		GAM	0.66	0.69	0.66	0.66	0.67	0.66	0.22	0.19	0.20	0.24	0.22	0.24	3.29	3.13	3.30	3.29	3.25	3.29
		XGB	0.57	0.58	0.57	0.57	0.55	0.55	0.27	0.78	0.28	0.28	0.90	0.24	3.42	3.63	3.42	3.42	3.74	3.48
		ANN	0.63	0.64	0.63	0.63	0.64	0.64	0.18	0.26	-	-	-	-0.13	3.25	3.24	3.15	3.13	3.13	3.11
											0.18	0.13	0.22							
ACA	PR		0.62	0.62	0.62	0.62	0.63	0.62	-0.65	-0.62	-	-	-	-0.64	3.90	3.90	3.89	3.91	3.85	3.90
											0.59	0.66	0.59							

Table 4 Continued.

WK	0.57	0.57	0.57	0.58	0.57	0.59	-0.16	-0.19	-	-	-	-0.16	4.17	4.18	4.16	4.11	4.19	4.07
									0.16	0.30	0.32							
GAM	0.62	0.60	0.62	0.63	0.63	0.62	0.57	0.68	0.07	0.01	-	0.13	3.02	3.11	3.38	3.35	3.37	3.58
											0.03							
XGB	0.43	0.36	0.53	0.54	0.43	0.47	0.50	-0.62	-	-	-	-0.65	3.46	4.16	3.55	3.54	3.93	3.79
									0.81	0.84	0.62							
ANN	0.65	0.64	0.64	0.65	0.63	0.65	-0.30	-0.11	-	-	-	-0.16	3.04	3.09	3.11	3.06	3.16	3.07
									0.10	0.23	0.32							

Note: Bold evaluation statistics indicate the combination (model + competition index) with the greatest R^2 value, the closest to zero MB or the lowest RMSE for a given species.

Table 5. Percentage change of evaluation statistics (coefficients of determination, R^2 ; mean bias, MB; root mean squared error, RMSE) between the best performing and the corresponding base model for temperate species (eastern white pine, EWP; hickory species, HICK; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar, YP) in Appalachian plateau (AP) and Ridge and Valley (VR) and pantropical species in Caribbean islands (CAR) (pink trumpet tree, PTT; muskwood, MW; African tulip tree, ATT; acacia, ACA) and Pacific islands (PAC) (African tulip tree, ATT; ‘Ōhi‘a lehua, OL, acacia, ACA).

Region	Species	R^2 change	MB change	RMSE change
AP	EWP	21.6% and 7.3%	-111.1%	-94.4%
	HICK	1.3%	-50.0%	-94.7%
	RM	8.0%	-100.0%	-94.8%
	SWG	2.4%	-100.0%	-94.7%
	WO	9.3% and 2.9%	-93.7%	-94.2%
	YP	8.4%	-100.0%	-95.1%
VR	EWP	0.0%	-100.0%	-2.1%
	HICK	1.2%	-60.0%	-3.6%
	RM	1.4%	-100%	-2.2%
	SWG	2.4%	-25%	-3.2%
	WO	2.7%	-31.5%	-3.4%
	YP	1.2% and 2.4%	-80.5%	-3.9%

Table 5 Continued.

	Average change	4.8%	-79.3%	-48.8%
CAR	PTT	2.7%	-133.3% and 66.6%	-3.1%
	MW	0%	-14.3%	0%
	ATT	1.4%	-66.6 and 66.6%	-1.8%
	ACA	16.9%	-103.0%	-59.5%
PAC	ATT	2.9%	-107.1%	-3.3%
	OL	4.5%	0%	-4.3%
	ACA	0%	-98.2%	0%
	Average change	4.0%	-43.2%	-10.3%

highest for AP (see figure 5A). Specifically, the ANN model with CDF as competition index provided the least biased predictions for four species and most accurate predictions for all the species in AP (see Table 2). For the temperate species in VR, GAM with R_{dia} as competition index performed the best (see Figure 5B). This combination of regression method and competition index provided the highest coefficient of determination for three species and lowest RMSE for two species (see Table 3). For the pantropical species in CAR, PR with R_{dia} was found to perform the best (see Figure 5C). This combination provided the least biased prediction for two of the CAR species. Similarly, GAM with CDF performed the best for species in PAC region (see Figure 5D). The height-diameter curves of the highest and lowest ranked model for the four physiographic regions are given in Figure 6. For some of the species, more than one combination of regression method and competition index provided the highest rank for one of the three evaluation statistics. For example, PR with R_{BA} and PSDI provided the least biased predictions for red maple in AP (see Table 2). Further, some base models had the same evaluation statistics as models with competition index which was best for the species. For example, base XGB, XGB with R_{dia} and XGB with CDF had the highest amount of variation explained for muskwood in CAR.

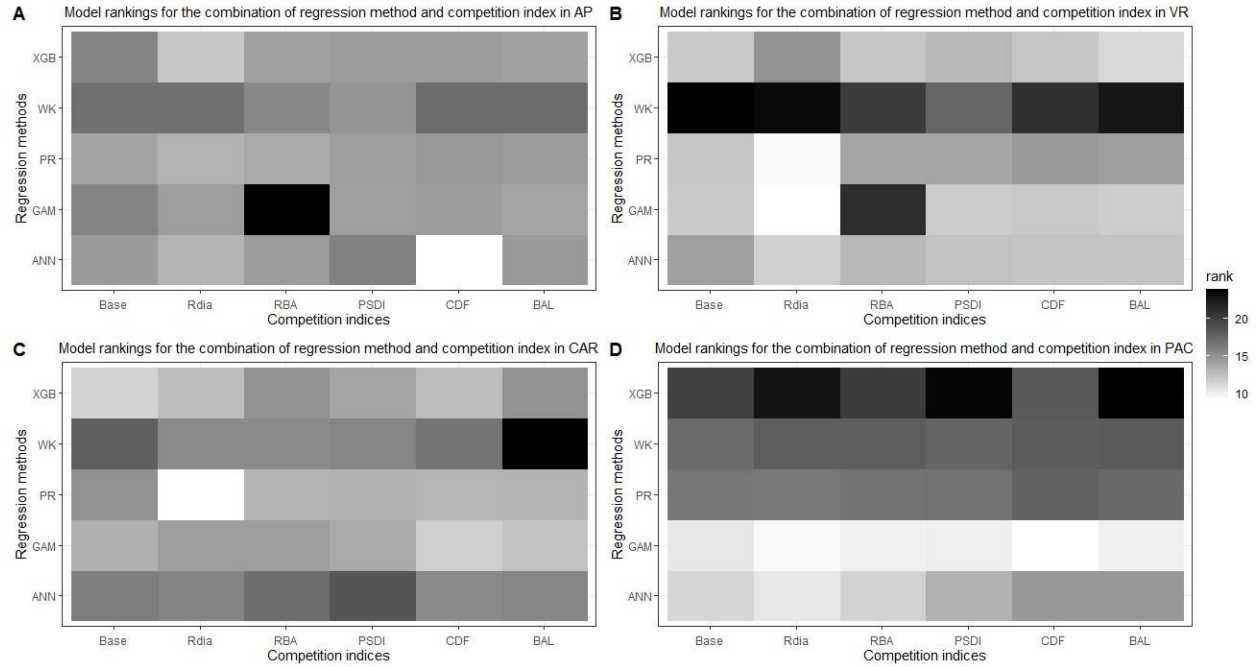


Figure 5. Best combination of regression method (Pearl & Reed (1920), PR; Wykoff et al. (1982), WK; Generalized additive model, GAM; Extreme Gradient Boosting, XGboost; Artificial Neural Network, ANN) and competition index (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) for all the species across four physiographic regions (Appalachian plateau, AP (A); Ridge and Valley, VR (B); Caribbean islands, CAR (C); Pacific islands, PAC (D)).

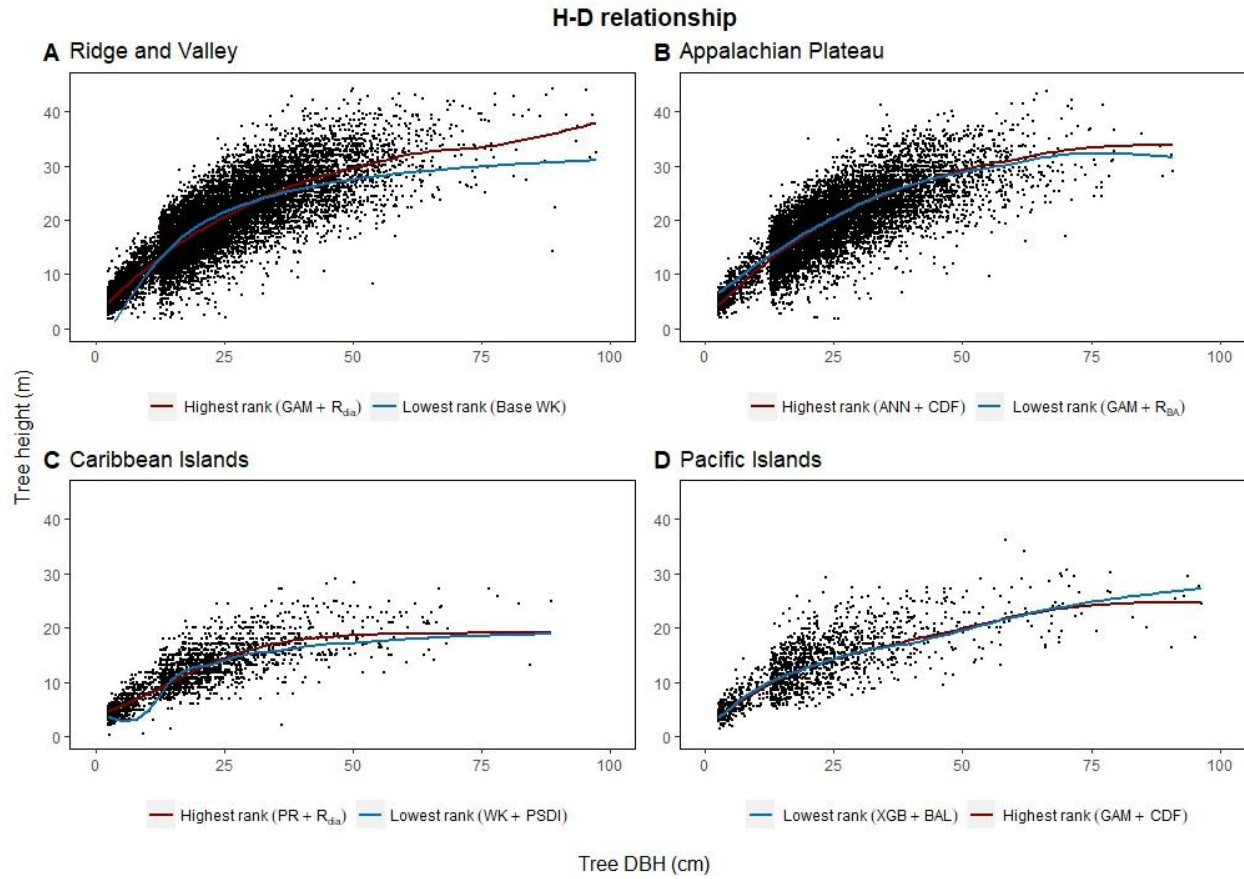


Figure 6. Height diameter relationship of the testing dataset (20% of the dataset) for each of the four physiographic regions (Appalachian plateau, AP; Ridge and Valley, VR (B); Caribbean islands, CAR (C); Pacific islands, PAC (D)). The curves show the predicted relationship for the highest and the lowest ranked models for each of the physiographic regions.

CHAPTER 5: DISCUSSION

5.1 Comparison of regression methods

Among the five regression methods examined for this study, all of them were found to have adequate performance for predicting individual tree height from diameter at breast height. However, a universally superior H-D model was not found for all nineteen species in the four physiographic regions, which is consistent with the findings from Mehtätalo et al. (2015). Although none of the models stood out with the best performance, GAM was found to perform relatively better than the rest of the models. On average, GAM ranks the highest among the parametric and nonparametric models examined based on R^2 , MB and RMSE (see Figure 2A). The base GAM provided the most precise predictions for seven of the nineteen species (see table 2-4). We confirmed the findings of Adamec & Drápela (2015) that GAM can be used to provide precise predictions of H using DBH as a predictor. The flexibility of the model was able to characterize variable shapes of H-D relationships for different species and region (see Figure 6). However, it should be noted that GAM may not be optimal model for a given species or region. As shown in Figure 4, the GAM model ranks lower than the PR for eastern white pine, hickory spp. and red maple in AP and VR. Specifically, GAM yielded the least precise predictions for red maple and sweetgum in AP and Acacia spp. in CAR (see Table 2, 4).

The ANN models rank the highest for four of the nineteen species while ranking the lowest for three of the nineteen species (see Figure 3). Özçelik et al. (2013) and Castaño-Santamaría et al. (2013) found that ANN models without considering diameter variation for each sampling unit performed worse than those accounting for it. This is only partially supported by our results because five of the base ANN models in this study performed the best compared to other regression methods, albeit the diameter variation of each condition class was not taken into

consideration. This may be because the network architecture used in this study was more complex compared to Özçelik et al. (2013) and Castaño-Santamaría et al. (2013). More hidden layers (two dense layers and two dropout layers) and more nodes per dense layer were incorporated in this study.

Misik et al. (2016) and Cui et al. (2022) reported that the PR and WK model generally had lower RMSE for temperate species in Turkey and China, respectively. However, our results indicate that the PR model performs well for the nineteen species, but the WK model generally had poor prediction accuracy. Specifically, the PR model had the highest rank of RMSE for two species whereas the WK model did not have the highest rank of RMSE for any species examined. Notably, the WK model tends to underpredict the height of the tallest and shortest trees in the stand (see supplementary figure 9B). Varying degrees of performance have been reported when employing the WK model in predicting tree height (e.g., Temesgen et al., 2014, Yang et al., 2022). Temesgen et al. (2014) found that this model provided the lowest MB and RMSE for most of the tree species in Northern China. However, Yang et al. (2022) reported that the WK model had the highest overall RMSE for pantropical tree species in Trinidad and Tobago. They also found that the WK model tends to be more inaccurate for trees in large DBH classes (see Table 3 of Yang et al. (2022)).

This study aimed to evaluate commonly used regression methods and competition indices to characterize the H-D relationships of temperate and pantropical tree species. The results of this study showed that the semiparametric model (GAM) produced more accurate predictions than parametric (PR and WK) and nonparametric models (XGBoost and ANN). The smoothing functions used by GAM for model fitting increases the flexibility of the curves while still retaining the structure of the curves like parametric models (see Figure 6A, 6D). However, GAM

is a data-driven model similar to XGBoost and ANN which may not produce satisfactory results when there are not many data points available for training the model. GAM is also often criticized for its tendency to overfit (Adamec & Drápela, 2015, Robinson et al., 2011). Although such an issue was not found in this study, it needs to be taken into consideration when building H-D models.

Lastly, the predictive performance of the H-D models was found to be better for the temperate species than the pantropical species. The range of the evaluation statistics was wider with larger extremes values for the pantropical species compared to the temperate species (see Table 6). This may be because the pantropical forests have higher tree species diversity which causes the trees to interact in a different manner compared to temperate forests. It is worth mentioning that the models presented in this study may not be optimal for a given species and region, but the models selected in comparison are commonly used in forest practices. Besides, in the preliminary analysis, random forest algorithm was tested, but the predictive performance of the models was poorer than ANN and XGBoost which was also reported by Li et al. (2020) and Raczko & Zagajewski (2017).

5.2 Inclusion of a competition index in H-D models

Inclusion of a competition index improved the predictive abilities of the models for most of the species in the AP, VR and CAR regions compared to the corresponding base model. Competition indices are generally included in individual tree models to represent the effect of neighboring trees have on a subject tree (Burkhardt & Tomé, 2012). The results of this study confirm that inclusion of distance-independent competition indices can improve the performance of height-diameter models when spatially explicit information at the individual tree level is not available. However, for the PAC region, the base model on average was found to perform better than the

Table 6. Range of evaluation statistics (coefficients of determination, R^2 ; mean bias, MB; root mean squared error, RMSE) between the best performing and the corresponding base model for temperate species (eastern white pine, EWP; hickory species, HICK; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar) in Appalachian plateau (AP) and Ridge and Valley (VR) and pantropical species in Caribbean islands (CAR) (pink trumpet tree, PTT; muskwood, MW; African tulip tree, ATT; acacia, ACA) and Pacific islands (PAC) (African tulip tree, ATT; ‘Ōhi‘a lehua, OL, acacia, ACA)

Region	Species	R^2 range	MB range	RMSE range
AP	EWP	0.50-0.73	-0.14-0.53	0.18-3.91
	HICK	0.55-0.73	-1.12-0.38	0.19-3.91
	RM	0.58-0.67	-0.19-0.15	0.17-3.38
	SWG	0.41-0.85	-0.95-0.22	0.22-4.34
	WO	0.56-0.70	-0.04-1.74	0.19-3.84
	YP	0.55-0.77	-1.12-0.35	0.19-4.07
VR	EWP	0.73-0.83	-0.47-0.68	2.78-3.40
	HICK	0.74-0.80	-0.85-0.38	3.14-3.50
	RM	0.64-0.72	-0.10-0.11	3.07-3.40
	SWG	0.76-0.83	-1.11-0.42	2.94-3.57
	WO	0.71-0.76	0.13-0.46	3.10-3.50

Table 6 Continued.

	YP	0.76-0.83	-0.18-0.35	3.44-3.99
	PTT	0.61-0.75	-0.58-0.99	2.12-3.05
CAR	MW	0.45-0.78	-0.15-0.48	2.30-3.72
	ATT	0.53-0.73	-0.25-0.85	2.69-3.88
	ACA	0.20-0.62	-1.13-1.24	1.14-3.02
	ATT	0.24-0.70	-1.83-3.05	2.00-5.94
PAC	OL	0.51-0.69	-0.30-0.90	3.11-3.74
	ACA	0.36-0.65	-0.84-0.68	3.02-4.19

models with competition indices across all the species and regression methods based on the evaluation statistics (See Supplementary Figure 6). The forested areas in PAC are severely depleted due to large-scale harvesting in the 1970s and severe cyclones in 1990s (FAO, 2010). In addition to that, the PAC islands are infested with various insects, pathogens and invasive species which has created a significant threat to the forest health (USDAFS, 2022). Frequent anthropogenic and natural disturbance may cause the large variation of tree allometry in the region. Thus, a single competition index may not be adequate to predict total tree height for Pacific islands' forests. Adding additional predictor variables (e.g., site or environmental variables) may help improve the prediction accuracy of the models, which is suggested as the next step of this study.

In general, R_{dia} , PSDI and CDF performed best for the five-regression method across the physiographic regions and species. PR and GAM models with R_{dia} as competition index tended to have the highest rank. Similarly, PSDI had the highest rank for the WK models. CDF worked well with non-parametric regression methods (XGBoost and ANN). These competition indices are used to mathematically represent the ability of a particular tree to grow without explicit spatial information about the neighboring trees (Burkhardt & Tomé, 1989). Diameter ratio and basal area ratio (size ratios) represented by the ratio of diameter and basal area of subject tree to the average diameter and basal area in the stand, respectively, provide information about the relative position of the subject tree in the stand. Basal area ratio emphasizes more weight on the size of the tree compared to diameter ratio. The size ratios assume that the competition is two-sided and smaller trees can also affect larger trees. Cumulative distribution function and basal area of larger trees also express the relative position of subject tree in the stand but the assumption for these indices is that competition is only one-sided and larger trees put the subject trees at a disadvantageous position while smaller trees do not affect them. The partitioned stand density index represents the general environment of competition by dividing the stand density index at the individual tree level. R_{dia} performed the best for AP, VR and CAR. Yang

et al. (2022) reported that R_{dia} was consistently the second most important predictor (after DBH) with RF and MERF models. It has also been reported to work well for other individual tree level models such as tree survival model (Sun et al., 2019), diameter growth models (Holmes and Reed, 1991). Similarly, Temesgen et al. (2007) found that inclusion of relative position and stand density competition indices to the base model performs better for predicting tree height. CDF was found to perform well for the species in AP and CAR and PSDI was found to provide the least biased estimates for species in VR. Sun et al. (2019) pointed out that CDF ignores the variation in relative competition among mixed forests and might not have a good performance. However, it was not supported by our study because all four of the physiographic regions are mixed forest and the CDF was calculated at the condition class level. In addition to this, the competition indices that had similar characteristics (e.g., size ratios) yielded different results. For most of the cases, R_{dia} performed better than R_{BA} and CDF performed better than BAL. Sun et al. (2019) also reported similar results for the diameter growth and survival equations for even-aged loblolly pine stands.

5.3 Further discussion

Height-diameter relationship models have important applications in sustainable forest management which include studying the vertical structure of a forest stand. The models in this study were evaluated using data from natural temperate and pantropical mixed forests which incorporate multiple biomes and forest types. The results of this study indicate that H-D relationships are variable for the same species in different regions, and a universally applicable method is difficult to ascertain. The graphs of the predicted height values and residuals are given in Figure 7-9. The forests in the Caribbean and Pacific regions are frequently affected by severe wind disturbances. These intense disturbance events severely affect tree growth and yield, causing widespread tree damage and mortality. The models developed in this study can be used to estimate the original length of trees in order to quantify the impact of wind damage in the regions. Although none of the models were found

to perform well for all 19 species, we recommend the highest ranked model for each of the physiographic regions (Figure 6) would provide adequate results.

The competition between species in a mixed forest has complex relationships that can be within the same species and between different species. This complex relationship was demonstrated by our study as the highest-ranking competition indices varied according to the species, physiographic region and regression method. The highest ranked competition indices (R_{dia} , PSDI and CDF for AP, VR and CAR) for the physiographic regions show that the influence of tree growth varies according to the region. Although this study did not use more than one competition index for evaluating the H-D models, a future study that uses more than one competition index as predictors is warranted to further understand the effect of competition in natural mixed forests.

Lastly, the GAM, ANN and XGBoost models can be transferred while the transferring approach is different from the classic parametric models. For parametric models, the transfer is done by sharing the equations with the estimated coefficients. In GAM, ANN and XGBoost, the transfer can be done by sharing the preprocessing script (or even an R package) and data used to build the models. Then the user can re-train (i.e., calibrate) the models with their own dataset. Another approach is to build an app that runs the algorithm in the background. Users can input the values of DBH and competition index, and then the predicted tree heights can be generated. Users can

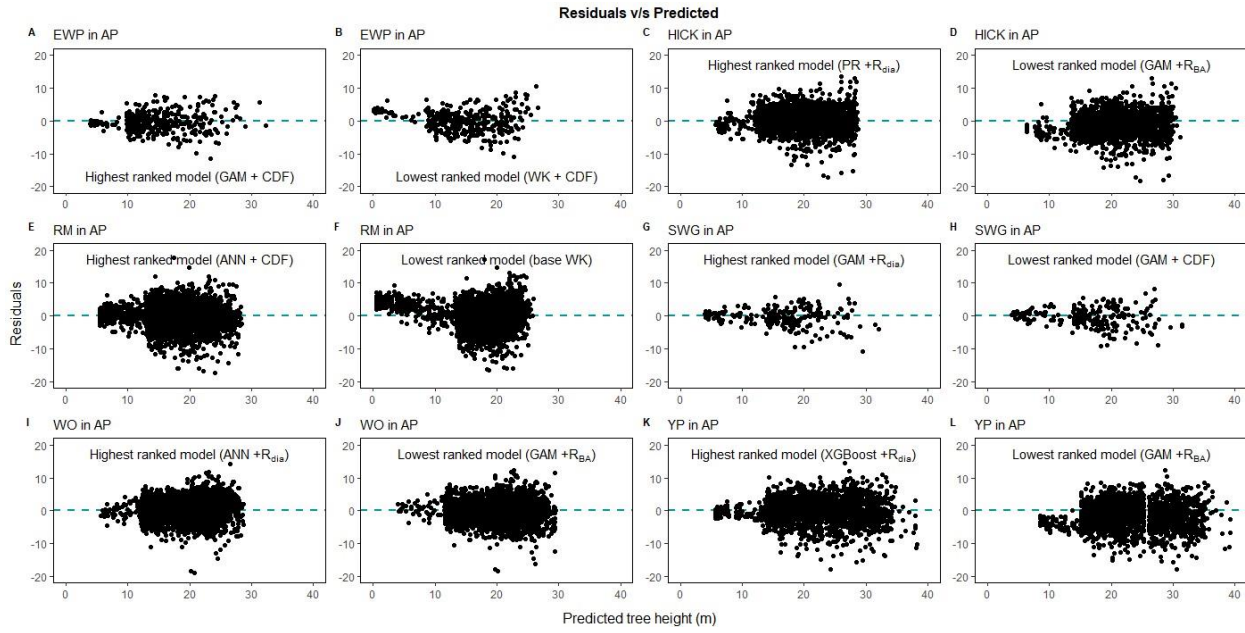


Figure 7. Comparison of the predicted tree height and the residual height for the highest ranking and lowest ranking combination of regression method and competition index for species in Appalachian plateau (AP; eastern white pine, EWP, A,B; hickory species, HICK, C,D; red maple, RM, E,F; sweetgum, SWG, G,H; white oak, WO, I,J; yellow poplar, YP, K,L).

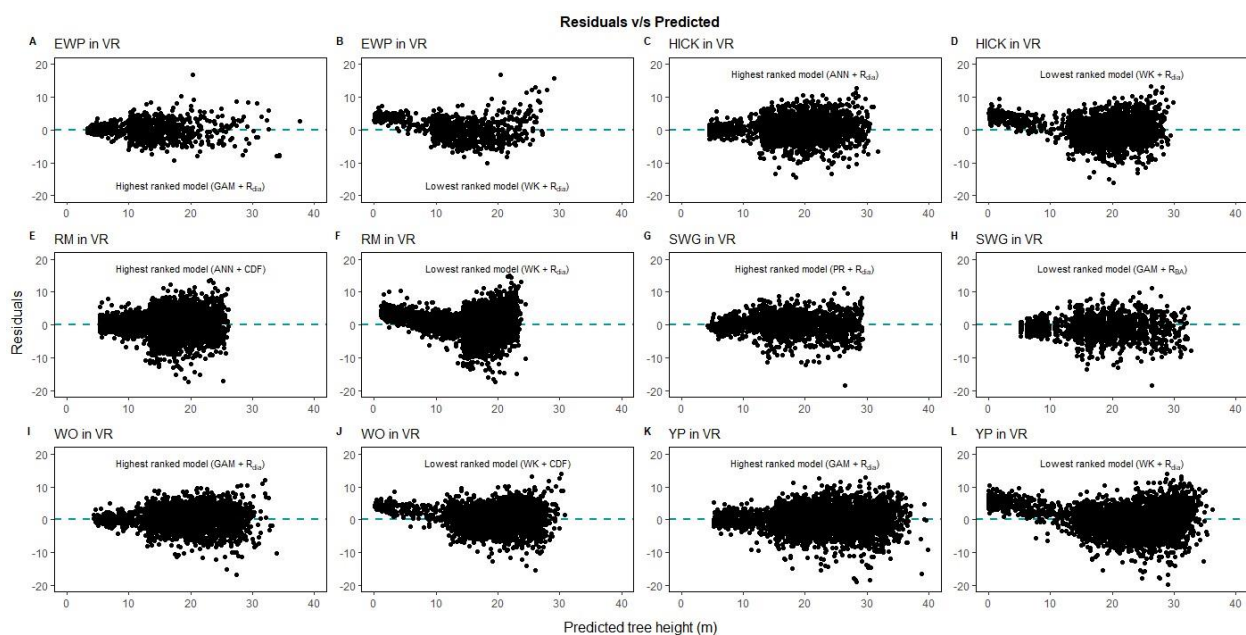


Figure 8. Comparison of the predicted tree height and the residual height for the highest ranking and lowest ranking combination of regression method and competition index for species in Ridge and Valley (VR; eastern white pine, EWP, A,B; hickory species, HICK, C,D; red maple, RM, E,F; sweetgum, SWG, G,H, white oak, WO, I,J; yellow poplar, YP, K,L).

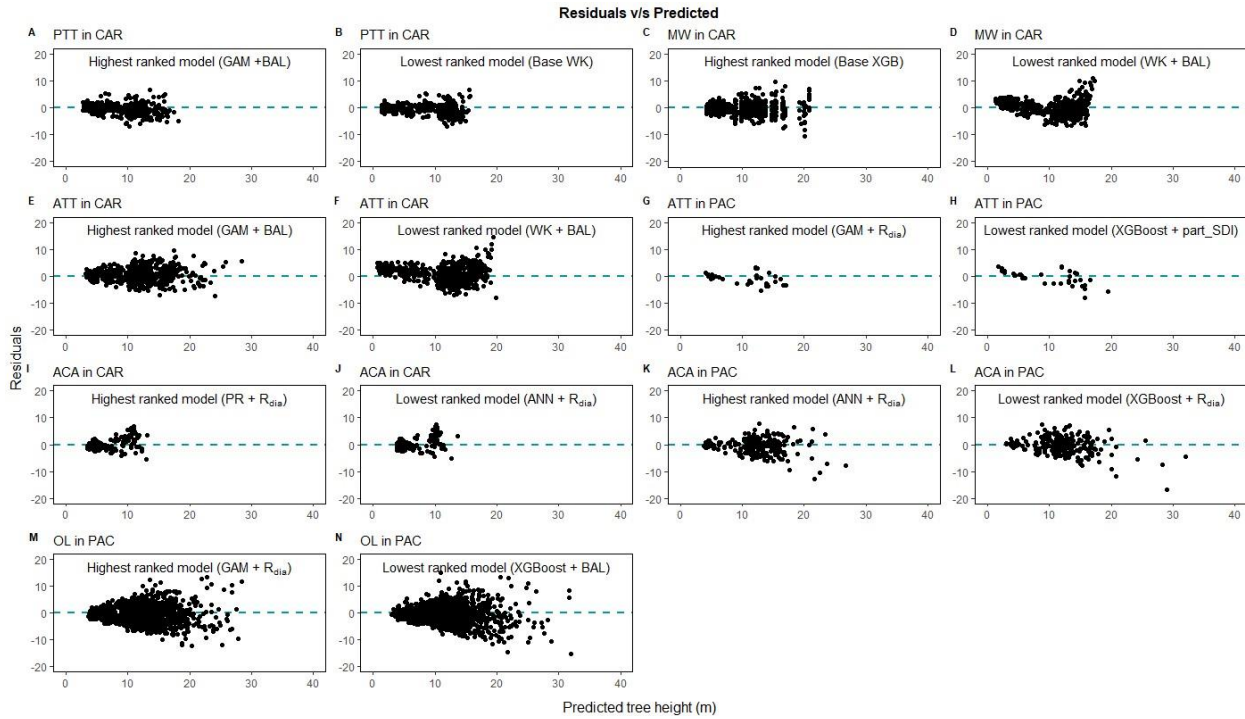


Figure 9. Comparison of the predicted tree height and the residual height for the highest ranking and lowest ranking combination of regression method and competition index for species in Caribbean (CAR) and Pacific (PAC) islands (pink trumpet tree, PTT, A,B; muskwood, MW, C,D; African tulip tree, ATT, E–H; acacia species, ACA, I–L; ‘Ōhi‘a lehua, OL, M,N).

upload their own dataset, allowing the app to re-train the models for localized predictions of tree height.

CHAPTER 6: CONCLUSION

This study shows that all the regression methods examined in this study were capable of predicting tree height from tree DBH. Among them, GAM can be used as the initial model when fitting H-D relationships, although it may not produce optimal predictions for a given species and physiographic region. However, it can be used to provide preliminary information about total tree length, especially to provide a quick assessment of the wind damage on total tree height. The results also showed that different competition indices are suitable for different regression methods. Diameter ratio, partitioned stand density index and cumulative distribution function were found to improve the predictive abilities of the regression methods for species in AP, VR and CAR. These competition indices are easy to calculate and hence we recommend including competition indices when building H-D models.

REFERENCES

- Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A. and Arshad, H. 2018 State-of-the-art in artificial neural network applications: A survey. *Heliyon*, **4** (11), e00938.
- Adame, P., del Río, M. and Cañellas, I. 2008 A mixed nonlinear height-diameter model for pyrenean oak (*Quercus pyrenaica* Willd.). *Forest Ecology and Management*, **256** (1-2), 88-98.
- Adamec, Z. and Drápela, K. 2015 Generalized additive models as an alternative approach to the modelling of the tree height-diameter relationship. *Journal of Forest Science*, **61** (6), 235-243.
- Bechtold, W.A. and Patterson, P.L. 2005 The Enhanced Forest Inventory and Analysis Program National Sampling Design and Estimation Procedures. U.S. Department of Agriculture, Forest Service, Southern Research Station. Asheville, NC.
- Burkhart, H.E. and Tomé, M. 1989 Distance-dependent competition measures for predicting growth of individual trees. *Forest Science*, **35** (3), 816-831.
- Burkhart, H.E. and Tomé, M. 2012 *Modeling forest trees and stands*. Springer.
- Burrill, E.A., DiTommaso, A.M., Turner, J.A., Pugh, S.A., Christensen, G., Perry, C.J. and Conkling, B.L. 2021 The Forest Inventory and Analysis Database: Database Description and User Guide version 9.0.1 for Phase 2. U.S. Department of Agriculture, Forest Service, p. 1026.
- Burrill, E.A., Wilson, A.M., Turner, J.A., Pugh, S.A., Menlove, J., Christensen, G., Conkling, B.L. and David, W. 2018 The Forest Inventory and Analysis Database: database description and user guide version 8.0 for Phase 2. U.S. Department of Agriculture, Forest Service, p. 946.
- Casper, B.B. and Jackson, R.B. 1997 Plant Competition Underground. *Annual Review of Ecology and Systematics*, **28**, 545-570.
- Castaño-Santamaría, J., Crecente-Campo, F., Fernández-Martínez, J.L., Barrio-Anta, M. and Obeso, J.R. 2013 Tree height prediction approaches for uneven-aged beech forests in northwestern Spain. *Forest Ecology and Management*, **307**, 63-73.

- Chen, J., Yang, H., Man, R., Wang, W., Sharma, M., Peng, C., Parton, J., Zhu, H. and Deng, Z. 2020 Using machine learning to synthesize spatiotemporal data for modelling DBH-height and DBH-height-age relationships in boreal forests. *Forest Ecology and Management*, **466**, 118104.
- Chen, T. and Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge, Discovery and Data Mining*. ACM.
- Cui, K., Wu, X., Zhang, C., Zhao, X. and von Gadow, K. 2022 Estimating height-diameter relations for structure groups in the natural forests of Northeastern China. *Forest Ecology and Management*, **519**, 120298.
- Curtis, R. and Marshall, D.D. 2000 Why Quadratic Mean Diameter? *Western Journal of Applied Forestry*, **15** (3), 137-139.
- Curtis, R.O. 1967 Height-Diameter and Height-Diameter-Age Equations For Second-Growth Douglas-Fir. *Forest Science*, **13** (4), 365-375.
- Elzhov, V., Mullen, K.M., Spiess, A.N. and Bolker, B. 2022 minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds. 1.2-4 Ed., CRAN.
- FAO. 2010 Pacific forests and forestry to 2020: Subregional report of the second asia-pacific forestry sector outlook study. Food and agriculture organization of the United Nations. Bangkok, Thailand.
- Feldpausch, T.R., Banin, L., Phillips, O.L., Baker, T.R., Lewis, S.L., Quesada, C.A., Affum-Baffoe, K., Arets, E.J.M.M., Berry, N.J., Bird, M., Brondizio, E.S., de Camargo, P., Chave, J., Djangbletey, G., Domingues, T.F., Drescher, M., Fearnside, P.M., França, M.B., Fyllas, N.M., Lopez-Gonzalez, G., Hladik, A., Higuchi, N., Hunter, M.O., Iida, Y., Salim, K.A., Kassim, A.R., Keller, M., Kemp, J., King, D.A., Lovett, J.C., Marimon, B.S., Marimon-Junior, B.H., Lenza, E., Marshall, A.R., Metcalfe, D.J., Mitchard, E.T.A., Moran, E.F., Nelson, B.W., Nilus, R.,

- Nogueira, E.M., Palace, M., Patiño, S., Peh, K.S.H., Raventos, M.T., Reitsma, J.M., Saiz, G., Schrod, F., Sonké, B., Taedoumg, H.E., Tan, S., White, L., Wöll, H. and Lloyd, J. 2011 Height-diameter allometry of tropical forest trees. *Biogeosciences*, **8** (5), 1081-1106.
- Fenneman, N.M. and Johnson, D.W. 1946 Physiographic divisions of the conterminous U.S., U.S. Geological Survey, Reston, VA.
- Gering, L.R. and May, D.M. 1995 The Relationship of Diameter at Breast Height and Crown Diameter for Four Species Groups in Hardin County, Tennessee. *Southern Journal of Applied Forestry*, **19** (4), 177-181.
- Huang, S., Price, D. and Titus, S.J. 2000 Development of ecoregion-based height-diameter models for white spruce in boreal forests. *Forest Ecology and Management*, **129** (2000), 125-141.
- Huang, S. and Titus, S.J. 1995 An individual tree diameter increment model for white spruce in Alberta. *Canadian Journal of Forest Research*, **25** (9), 1455-1465.
- Jenkins, J.C., Chojnacky, D.C., Heath, L.S. and Birdsey, R.A. 2003 National scale biomass estimators for United States tree species. *Forest Science* (49), 12-35.
- Kalinowski, T., Allaire, J.J., Falbel, D., Bostock, M., Masayuki, T., Bowe, S., You, Y., Decker, K., Sagalaev, I. and Pakala, Y. 2022 tfruns: Training Run Tools for 'TensorFlow'. 1.5.1 Ed., CRAN.
- Kuhn, M. 2008 Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, **28** (5), 1 - 26.
- Lam, T.Y., Kershaw, J.A., Hajar, Z.S.N., Rahman, K.A., Weiskittel, A.R. and Potts, M.D. 2017 Evaluating and modelling genus and species variation in height-to-diameter relationships for Tropical Hill Forests in Peninsular Malaysia. *Forestry*, **90** (2), 268-278.
- Li, Y., Li, M., Li, C. and Liu, Z. 2020 Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. *Scientific Reports*, **10** (1), 9952.

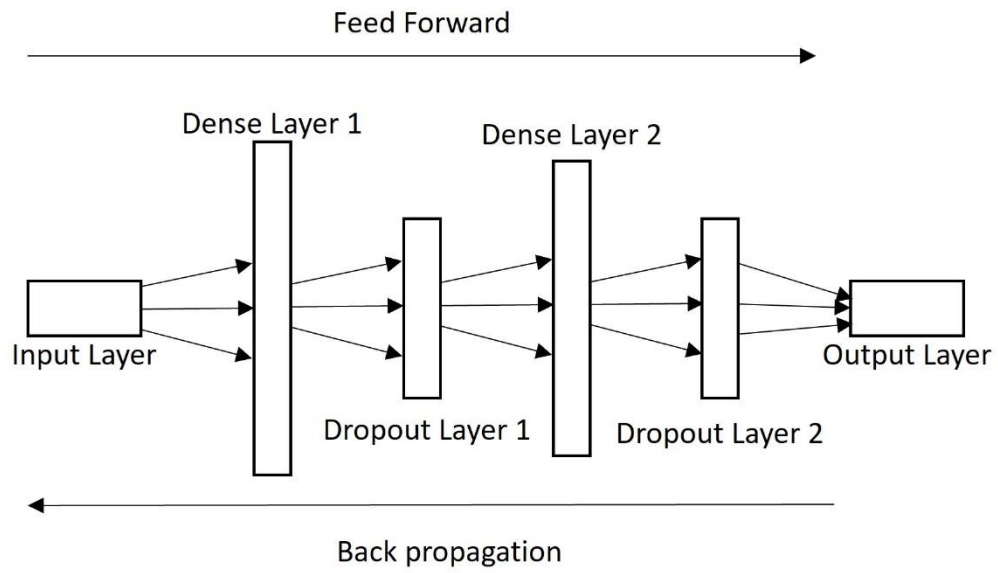
- Lu, H.-c.B., Martin, F. and Johnson, R. Theoretical Foundation of Stage's Formulation of Stand Density Index. In *Proceedings of the 2017 Forest Vegetation Simulator (FVS) e-Conference*. U.S. Department of Agriculture, Forest Service, Southern Research Station, pp. 64-69.
- Mehtätalo, L., de-Miguel, S. and Gregoire, T.G. 2015 Modeling height-diameter curves for prediction. *Canadian Journal of Forest Research*, **45** (7), 826-837.
- Misik, T., Antal, K., Kárász, I. and Tóthmérész, B. 2016 Nonlinear height-diameter models for three woody, understory species in a temperate oak forest in Hungary. *Canadian Journal of Forest Research*, **46** (11), 1337-1342.
- Ogana, F.N. and Ercanli, I. 2022 Modelling height-diameter relationships in complex tropical rain forest ecosystems using deep learning algorithm. *Journal of Forestry Research*, **33** (3), 883-898.
- Özçelik, R., Diamantopoulou, M.J., Crecente-Campo, F. and Eler, U. 2013 Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. *Forest Ecology and Management*, **306**, 52-60.
- Pearl, R. and Reed, L.J. 1920 On the rate of growth of population of the United States since 1790 and its mathematical representation. *PNAS*, **6**, 275-288.
- Poudel, K.P. and Cao, Q.V. 2013 Evaluation of methods to predict weibull parameters for characterizing diameter distributions. *Forest Science*, **59** (2), 243-252.
- Raczko, E. and Zagajewski, B. 2017 Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *European Journal of Remote Sensing*, **50** (1), 144-154.
- Reineke, L.H. 1933 Perfecting a stand-density index for even-aged forests. *Journal of Agricultural Research*, **46** (7), 627-638.
- Robert, C. 2014 Machine Learning, a Probabilistic Perspective. *CHANCE*, **27** (2), 62-63.

- Robinson, A.P., Lane, S.E. and Thérien, G. 2011 Fitting forestry models using generalized additive models: A taper model example. *Canadian Journal of Forest Research*, **41** (10), 1909-1916.
- Schmidt, M., Kiviste, A. and von Gadow, K. 2011 A spatially explicit height–diameter model for Scots pine in Estonia. *European Journal of Forest Research*, **130** (2), 303-315.
- Sharma, R.P., Vacek, Z., Vacek, S. and Kučera, M. 2019 Modelling individual tree height–diameter relationships for multi-layered and multi-species forests in central Europe. *Trees - Structure and Function*, **33** (1), 103-119.
- Soares, P. and Tomé, M. 2002 Height-diameter equation for first rotation eucalypt plantations in Portugal. *Forest Ecology and Management*, **166** (2002), 99-109.
- Stage, A.R. 1968 A tree-by-tree measure of site utilization for Grand fir related to Stand Density Index. Ogden, Utah.
- Stage, A.R. 1973 PROGNOSIS MODEL FOR STAND DEVELOPMENT. Ogden, Utah.
- Sun, S., Cao, Q.V. and Cao, T. 2019 Evaluation of distance-independent competition indices in predicting tree survival and diameter growth. *Canadian Journal of Forest Research*, **49** (5), 440-446.
- Temesgen, H., Hann, D.W. and Monleon, V.J. 2007 Regional Height-Diameter Equations for Major Tree Species of Southwest Oregon. *Western Journal of Applied Forestry*, **22** (3), 213-219.
- Temesgen, H., Zhang, C.H. and Zhao, X.H. 2014 Modelling tree height-diameter relationships in multi-species and multi-layered forests: A large observational study from Northeast China. *Forest Ecology and Management*, **316**, 78-89.
- USDAFS. 2022 Pacific Islands Forest Health Highlights 2021. U.S. Department of Agriculture Forest Service.
- Walczak, S. and Cerpa, N. 2003 Artificial Neural Networks. In *Encyclopedia of Physical Science and Technology*. R.A. Meyers (ed.). Third Edition Ed., Academic Press, pp. 631-645.

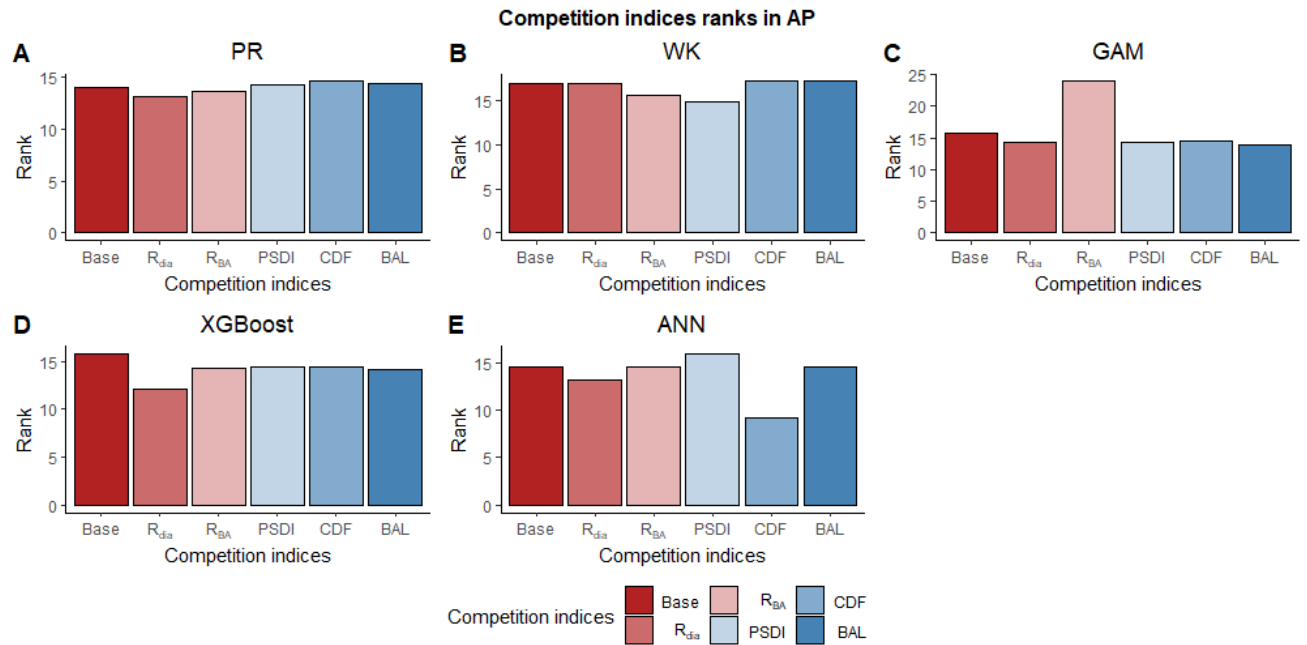
- Wang, Y., Liu, H., Sang, L. and Wang, J. 2022 Characterizing Forest Cover and Landscape Pattern Using Multi-Source Remote Sensing Data with Ensemble Learning. *Remote Sensing*, **14** (21), 54.
- Weiskittel, A.R., Hann, D.W., Kershaw, J.A. and J.K., V. 2011 *Forest Growth and Yield Modeling*. John Wiley & Sons, Ltd.
- Whiffin, T. 1992 The status of forest biodiversity in Oceania. *Journal of Tropical Forest Science*, **5** (2), 155-172.
- Wood, S.N. 2017 *Generalized additive models: An introduction with R, second edition*. CRC Press, 1-476 p.
- Wykoff, W.R., Crookston, N.L. and Stage, A.R. 1982 User's Guide to the Stand Prognosis Model. Ogden, Utah.
- Xu, Z., Zhang, Q., Xiang, S., Li, Y., Huang, X., Zhang, Y., Zhou, X., Li, Z., Yao, X., Li, Q. and Guo, X. 2022 Monitoring the Severity of *Pantana phyllostachysae* Chao Infestation in Moso Bamboo Forests Based on UAV Multi-Spectral Remote Sensing Feature Selection. *Forests*, **13** (3).
- Yang, S.-I., Burkhardt, H.E. and Seki, M. 2023 Evaluating semi- and nonparametric regression algorithms in quantifying stem taper and volume with alternative test data selection strategies. *Forestry: An International Journal of Forest Research*.
- Yang, S.I., Brandeis, T.J., Helmer, E.H., Oatham, M.P., Heartsill-Scalley, T. and Marciano-Vega, H. 2022 Characterizing height-diameter relationships for Caribbean trees using mixed-effects random forest algorithm. *Forest Ecology and Management*, **524**.
- Yang, S.I. and Burkhardt, H.E. 2020 Evaluation of total tree height subsampling strategies for estimating volume in loblolly pine plantations. *Forest Ecology and Management*, **461**.

- Zang, H., Lei, X. and Zeng, W. 2016 Height-diameter equations for larch plantations in northern and northeastern China: A comparison of the mixed-effects, quantile regression and generalized additive models. *Forestry*, **89** (4), 434-445.
- Zhang, N., Chen, M., Yang, F., Yang, C., Yang, P., Gao, Y., Shang, Y. and Peng, D. 2022 Forest Height Mapping Using Feature Selection and Machine Learning by Integrating Multi-Source Satellite Data in Baoding City, North China. *Remote Sensing*, **14** (18).
- Zhao, D., Borders, B., Wilson, M. and Rathbun, S.L. 2006 Modeling neighborhood effects on the growth and survival of individual trees in a natural temperate species-rich forest. *Ecological Modelling*, **196** (1), 90-102.

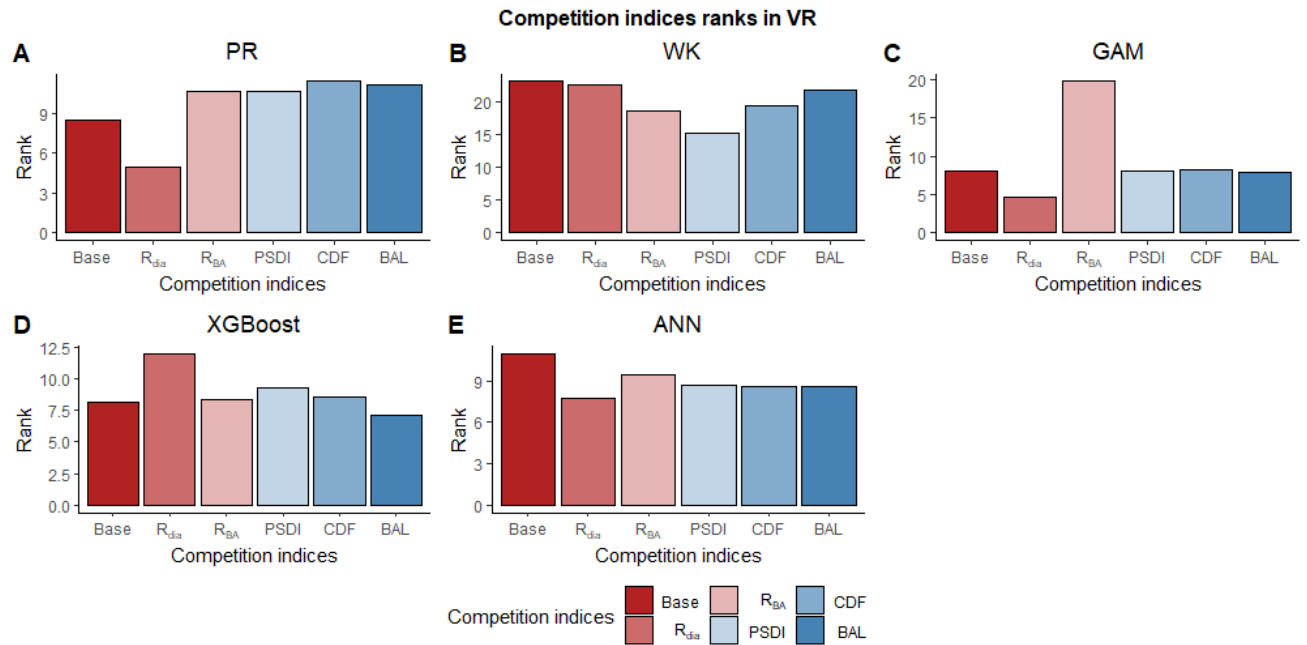
APPENDIX



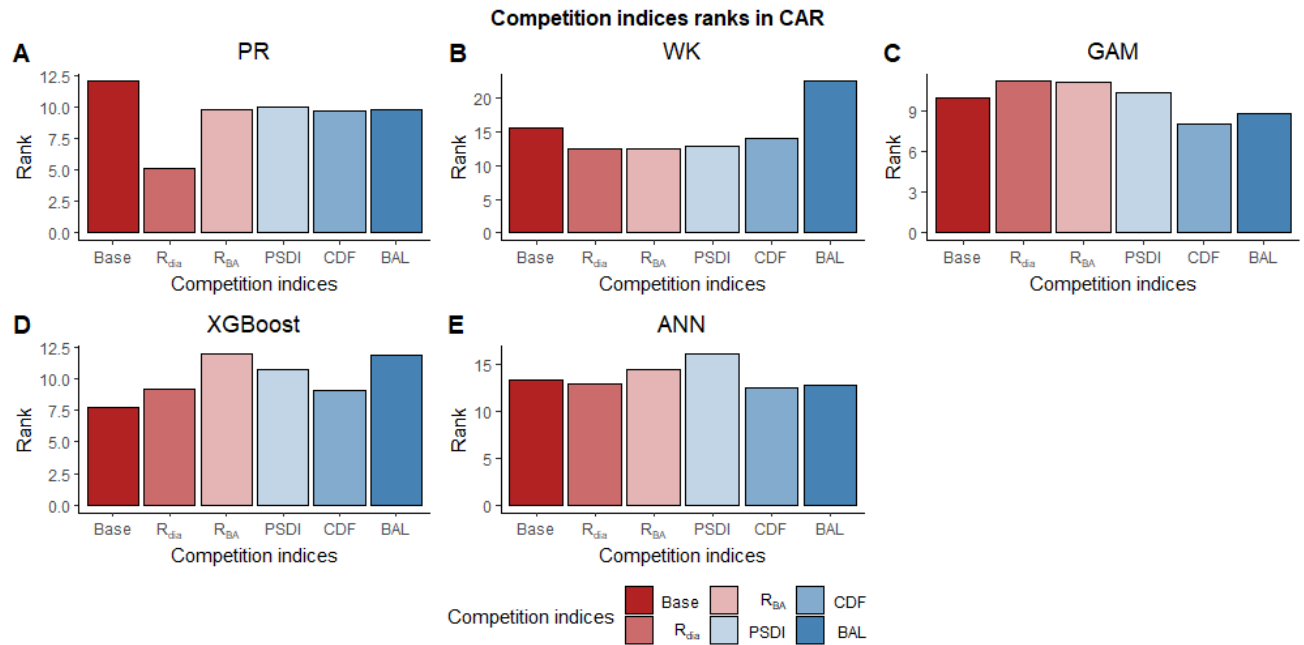
Supplementary Figure 1. Schematic representation of artificial neural network model architecture used for this study



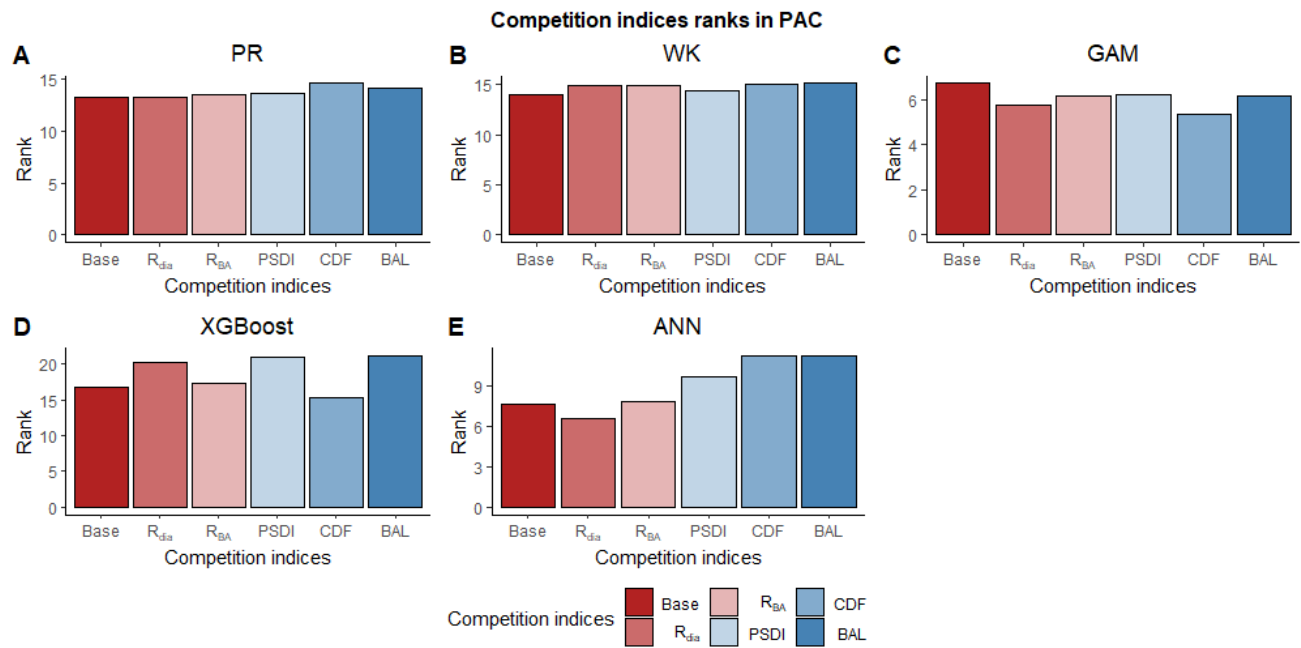
Supplementary Figure 2. Average ranking of competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) for the five regression methods (Pearl & Reed (1920), PR (A); Wykoff et al. (1982), WK (B); Generalized additive model, GAM (C); Extreme Gradient Boosting, XGBoost (D); Artificial Neural Network, ANN (E)) in Appalachian plateau (AP)



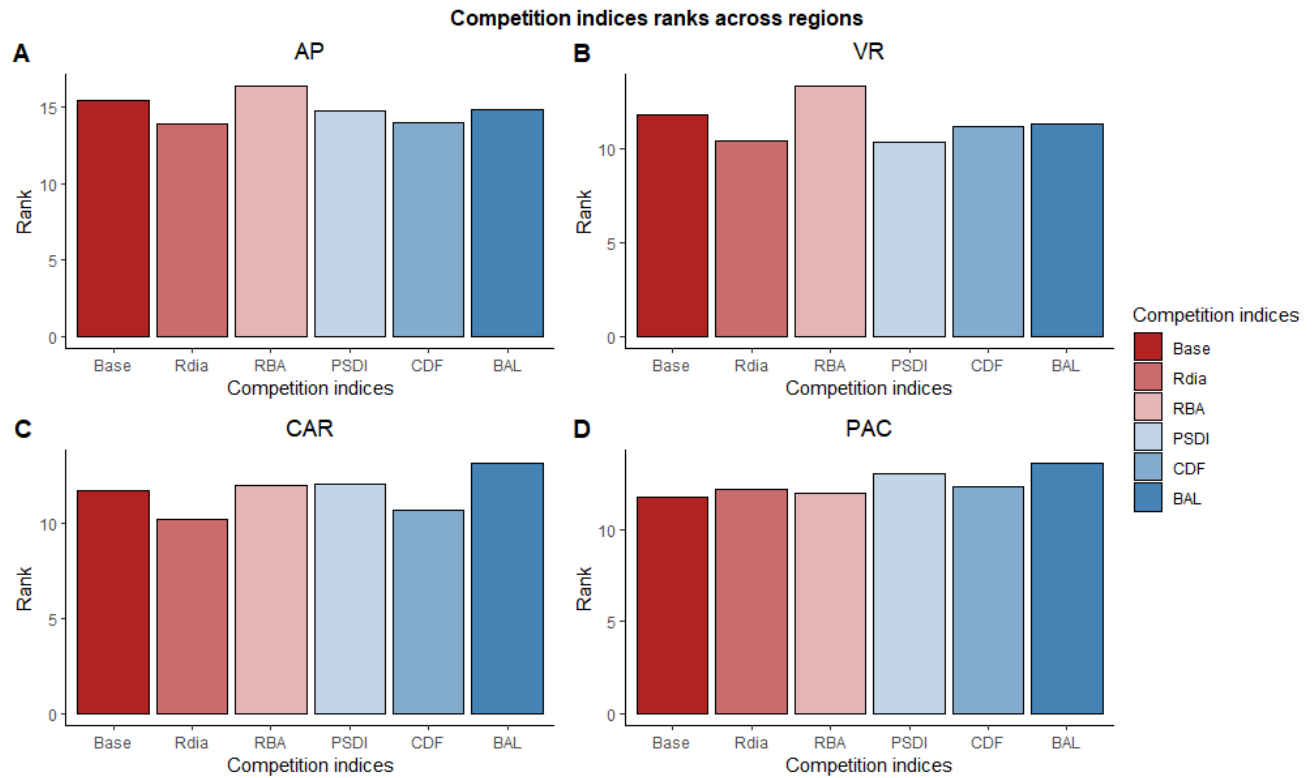
Supplementary Figure 3. Average ranking of competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) for the five regression methods (Pearl & Reed (1920), PR (A); Wykoff et al. (1982), WK (B); Generalized additive model, GAM (C); Extreme Gradient Boosting, XGBoost (D); Artificial Neural Network, ANN (E)) in Ridge and Valley (VR)



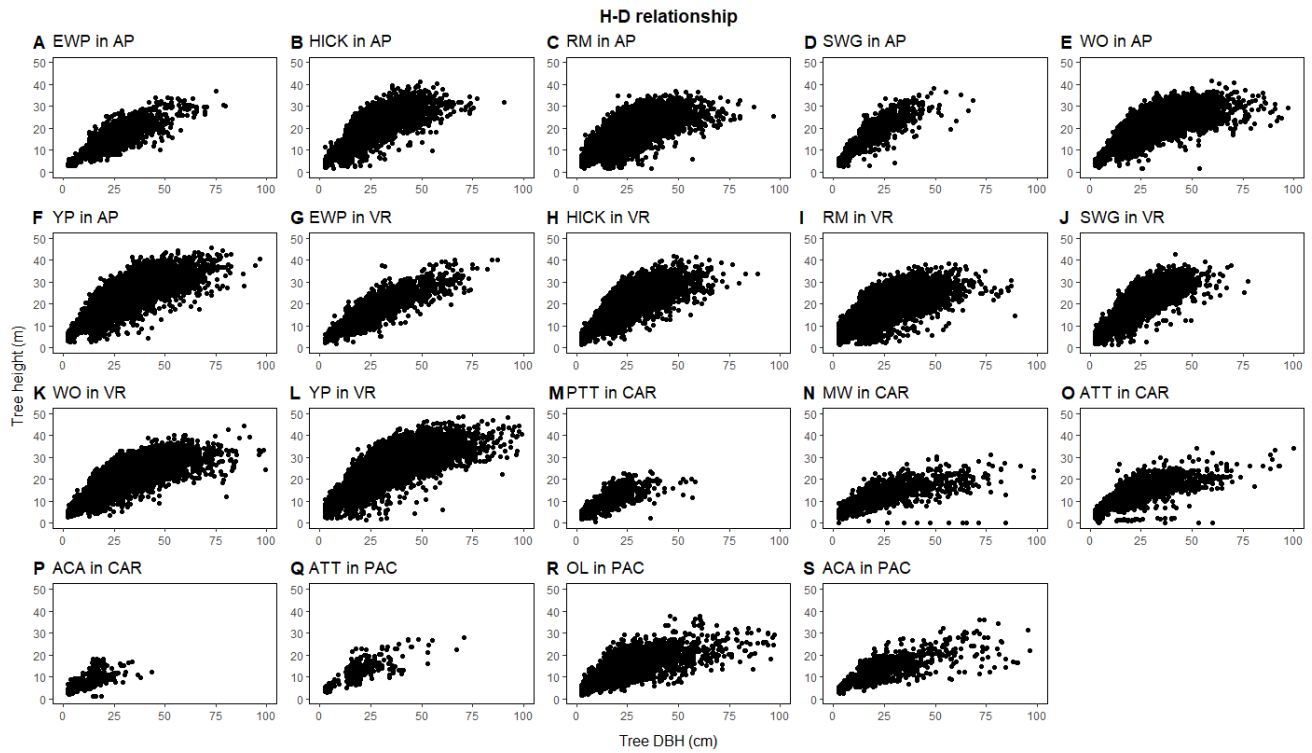
Supplementary Figure 4. Average ranking of competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) for the five regression methods (Pearl & Reed (1920), PR (A); Wykoff et al. (1982), WK (B); Generalized additive model, GAM (C); Extreme Gradient Boosting, XGBoost (D); Artificial Neural Network, ANN (E)) in Caribbean islands (CAR)



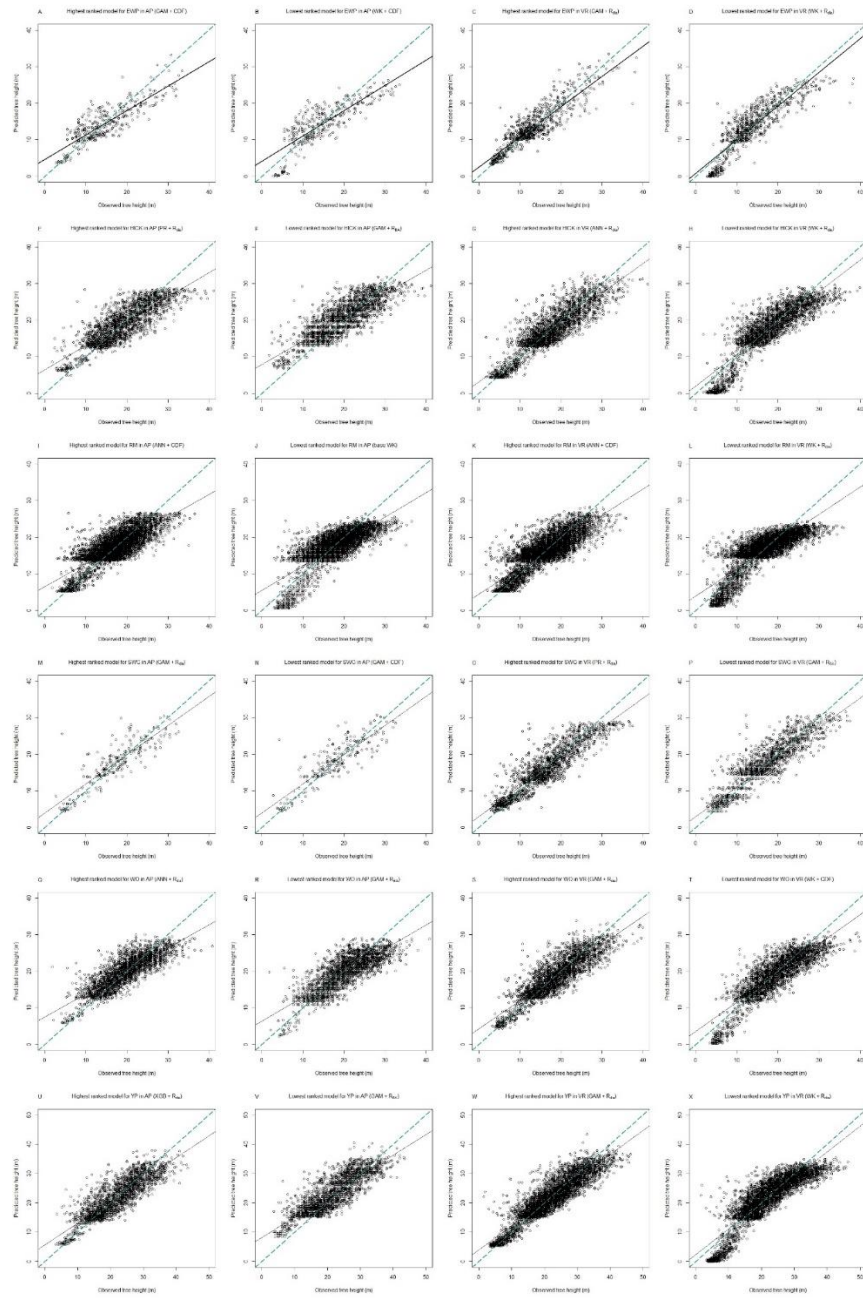
Supplementary Figure 5. Average ranking of competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) for the five regression methods (Pearl & Reed (1920), PR (A); Wykoff et al. (1982), WK (B); Generalized additive model, GAM (C); Extreme Gradient Boosting, XGBoost (D); Artificial Neural Network, ANN (E)) in Pacific islands (PAC)



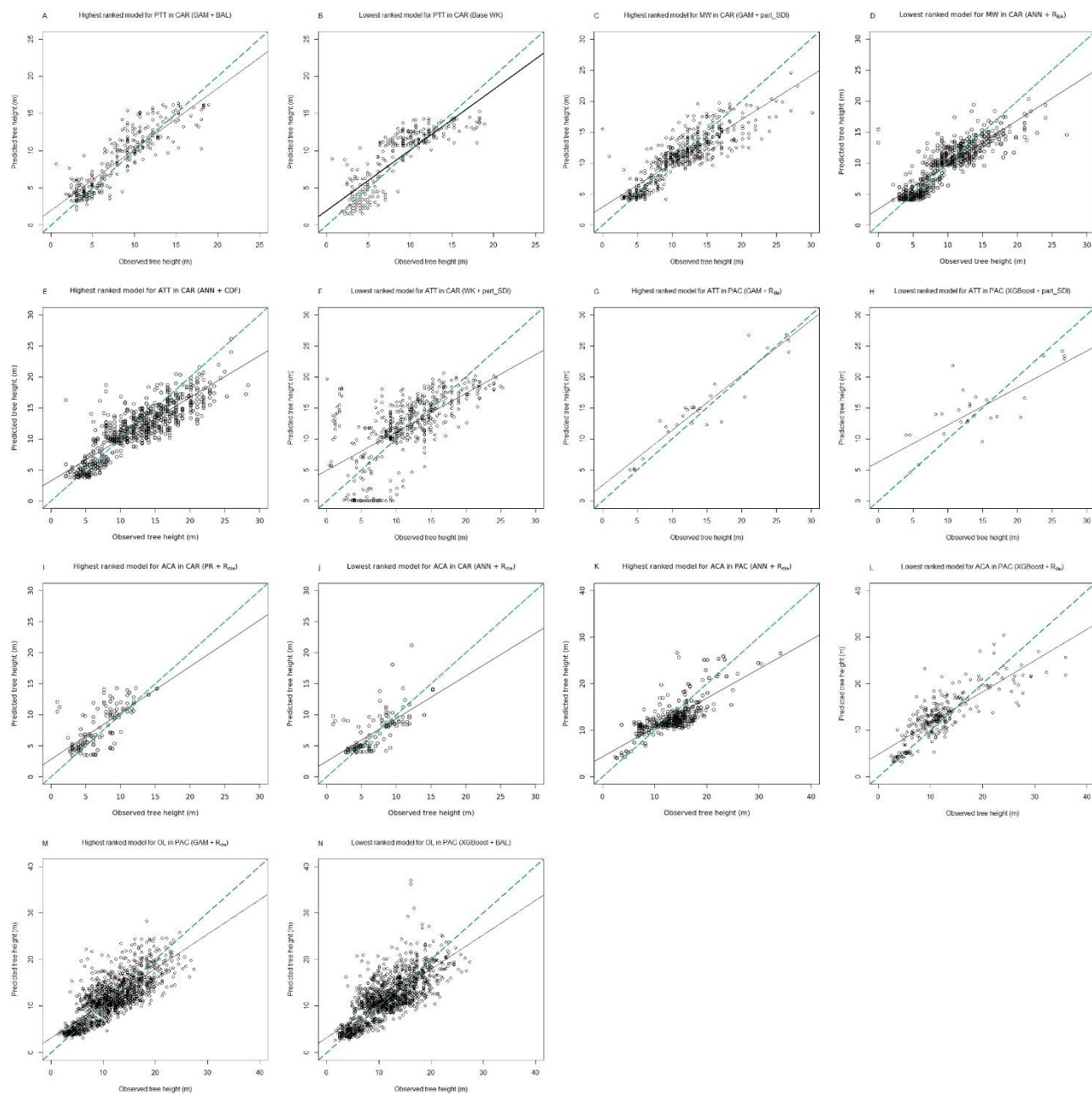
Supplementary Figure 6. Average ranking of competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL) across the four physiographic region (Appalachian Plateau (AP), Ridge and Valley (VR), Caribbean Islands (CAR) and Pacific islands (PAC))



Supplementary Figure 7. Height-Diameter relationship of temperate species in Appalachian Plateau (AP) and Ridge and Valley (VR) (eastern white pine, EWP (A,G); hickory species, Hick (B, H); Red maple, RM (C, I); sweetgum, SWG (D, J); white oak, WO (E, K); yellow poplar, YP (F, L)) and pantropical species in Caribbean islands (CAR) (pink trumpet tree, PTT (M); muskwood, MW (N); African tulip tree, ATT (O); acacia species, ACA (P)) and Pacific islands (PAC) (African tulip tree, ATT (Q); ‘Ōhi‘a lehua, OL (R); acacia species, ACA(S))



Supplementary Figure 8. Comparison of the observed tree height and predicted tree height for the highest ranking and lowest ranking combination of regression method and competition index for species in Appalachian plateau (AP) and Ridge and Valley (VR) (eastern white pine, EWP, A-D; hickory species, HICK, E-H; red maple, RM, I-L; sweetgum, SWG, M-P; white oak, WO, Q-T; yellow poplar, YP, U-X)



Supplementary Figure 9. Comparison of the observed tree height and predicted tree height for the highest ranking and lowest ranking combination of regression method and competition index for species in Caribbean (CAR) islands (pink trumpet tree, PTT, A-B; muskwood, MW, C-D; African tulip tree, ATT, E-F; acacia species, ACA, I-J) and Pacific (PAC) islands (African tulip tree, ATT, G-H; ‘Ōhi‘a lehua, OL, M-N; acacia species, ACA, K-L)

Supplementary Table 1. Hyperparameters for extreme gradient boosting (XGBoost) models for the four physiographic regions (Appalachian plateau, AP; Ridge and Valley, VR; Caribbean islands, CAR; Pacific islands, PAC) for the base model and models with competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL). The hyperparameter values were same for all the species for a given XGBoost model in a region

	Region	Nrounds	eta	Max_depth	Colsample_ bytree	subsample	gamma	Min_child_w eight
Base	AP	150	0.05	2	0.50	0.50	100	20
	VR	150	0.05	2	1	0.50	0.00	0.00
	CAR	100	0.05	2	0.50	1	0.00	20
	PAC	150	0.05	2	0.50	0.50	0.00	0.00
R_{dia}	AP	150	0.05	4	1	0.50	100	0.00
	VR	150	0.05	3	0.50	0.50	100	20
	CAR	50	0.1	4	1	1	0.00	20
	PAC	250	0.1	5	1	1	0.00	0.00
R_{BA}	AP	150	0.05	2	1	0.50	100	20
	VR	150	0.05	2	1	0.50	0.00	20
	CAR	250	0.05	5	1	1	0.00	0.00

Supplementary Table 1 Continued.

PSDI	PAC	150	0.05	2	1	0.50	0.00	0.00
	AP	150	0.05	5	1	0.50	0.00	20
	VR	100	0.05	5	1	0.50	0.00	20
	CAR	250	0.05	5	1	0.50	0.00	0.00
CDF	PAC	300	0.05	5	1	0.50	0.00	0.00
	AP	150	0.05	3	1	0.50	0.00	20
	VR	150	0.05	3	1	0.50	0.00	0.00
	CAR	100	0.05	2	1	0.50	0.00	20
BAL	PAC	250	0.05	2	1	0.50	100	0.00
	AP	300	0.10	5	1	1	0.00	0.00
	VR	300	0.05	5	1	1	0.00	0.00
	CAR	250	0.05	5	1	0.50	0.00	0.00
	PAC	300	0.10	5	1	1	0.00	0.00

Supplementary Table 2. Hyperparameters for artificial neural network (ANN) models for the four physiographic regions (Appalachian plateau, AP; Ridge and Valley, VR; Caribbean islands, CAR; Pacific islands, PAC) for the base model and models with competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL). The hyperparameter values were same for all the species for a given ANN model in a region

Competition index	Region	Dense units 1	Dense units 2	Dropout 1	Dropout 2	Batch size
Base	AP	64	32	0.20	0.10	32
	VR	64	32	0.20	0.10	64
	CAR	64	32	0.20	0.10	32
	PAC	64	32	0.10	0.10	32
R_{dia}	AP	64	32	0.10	0.10	32
	VR	64	16	0.10	0.10	64
	CAR	64	32	0.10	0.10	64
	PAC	64	32	0.10	0.10	32
R_{BA}	AP	64	32	0.10	0.10	32
	VR	32	16	0.10	0.10	32
	CAR	32	16	0.10	0.10	32
	PAC	32	16	0.20	0.10	32

Supplementary Table 2 Continued.

PSDI	AP	64	16	0.10	0.10	64
	VR	64	32	0.20	0.20	32
	CAR	64	32	0.10	0.10	64
	PAC	64	16	0.10	0.10	64
CDF	AP	64	16	0.10	0.10	32
	VR	64	32	0.20	0.10	64
	CAR	64	16	0.20	0.20	64
	PAC	64	16	0.20	0.10	64
BAL	AP	64	32	0.10	0.10	32
	VR	64	16	0.20	0.20	32
	CAR	64	32	0.20	0.10	64
	PAC	32	16	0.10	0.10	32

Supplementary Table 3. Parameter estimates (50% quantile) with lower (2.5% quantile) and upper (97.5% quantile) limits of the confidence interval from the 500 bootstrap samples for the parametric models (Pearl & Reed (1920), PR; Wykoff et al. (1982), WK) for species (eastern white pine, EWP; hickory species, HIC; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar) in Appalachian plateau (AP) for the base model and models with competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL)

Species	Model	Competition index	β_0	β_1	β_2	β_3
EWP	PR	Base	28.2280 (25.8965, 30.4671)	4.7647 (4.3687, 5.2412)	-0.0721 (-0.0824, -0.0633)	-
		R_{dia}	28.2059 (25.9630, 30.5846)	4.4783 (4.0917, 4.8595)	-0.0965 (-0.1115, -0.0815)	0.6202 (0.3723, 0.8827)
		R_{BA}	30.1787 (27.6300, 32.8946)	6.2358 (5.6410, 6.8369)	-0.0618 (-0.0711, -0.0535)	-5.4792 (-7.1631, -3.8308)
		PSDI	31.2487 (27.6866, 36.5106)	4.7450 (4.3029, 5.2358)	-0.0668 (-0.0767, -0.0583)	0.0184 (0.0115, 0.0273)
		CDF	28.2306 (26.0767, 30.7713)	4.7674 (4.3522, 5.2137)	-0.0722 (-0.0840, -0.0634)	0.0182 (-0.0841, 0.0940)
		BAL	28.1042 (25.7744, 30.4199)	4.6589 (4.1548, 5.2144)	-0.0727 (-0.0832, -0.0636)	0.0004 -0.0002, 0.0011

Supplementary Table 3 Continued.

	WK	Base	3.4717 3.4120, 3.5301	-0.0610 -0.0656, -0.0566	-	-
		R _{dia}	3.4595 3.4024, 3.5086	-0.0818 -0.0921, -0.0727	0.4480 0.2542, 0.6350	-
		R _{BA}	3.4206 3.3500, 3.4983	0.0736 0.0596, 0.0865	-2.3295 -4.5059, -0.2333	-
		PSDI	3.4830 3.4263, 3.5334	-0.0563 -0.0605, -0.0529	-0.0133 -0.0162, -0.0110	-
		CDF	3.5017 3.4441, 3.5565	-0.0557 -0.0599, -0.0517	-0.1253 -0.1789, -0.0706	-
		BAL	3.4757 3.4116, 3.5375	-0.0610 -0.0671, -0.0558	-3.458e-05 -5.115e-04, 4.059e-04	-
HICK	PR	Base	28.6712 (28.1673, 29.2091)	5.0306 (4.7529, 5.3192)	-0.1071 (-0.1122, -0.1020)	-
		R _{dia}	28.7404 (28.1702, 29.2923)	4.7319 (4.4940, 4.9792)	-0.1233 (-0.1305, -0.1163)	0.4749 (0.3273, 0.6135)
		R _{BA}	29.1334 (28.5705, 29.6847)	5.6880 (5.4336, 5.9405)	-0.0999 (-0.1063, -0.0946)	-3.2810 (-4.2993, -2.3022)

Supplementary Table 3 Continued.

		PSDI	29.0391 28.3687, 29.6476	4.7645 4.5181, 5.1036	-0.1048 -0.1111, -0.0999	0.0083 0.0048, 0.0120
		CDF	28.6678 (28.2050, 29.1995)	5.0214 4.7608, 5.2865	-0.1071 -0.1121, -0.1022	0.0023 -0.0412, 0.0442
		BAL	28.5857 28.0924, 29.0974	4.7537 4.4580, 5.0897	-0.1076 -0.1124, -0.1025	0.0006 0.0002, 0.0012
	WK	Base	3.6317 3.6131, 3.6460	-0.0700 -0.0723, -0.0682	-	-
		R _{dia}	3.6216 3.6048, 3.6366	-0.0816 -0.0858, -0.0771	0.2506 0.1478, 0.3428	-
		R _{BA}	3.5641 3.5448, 3.5842	0.0911 0.0860, 0.0965	-3.8842 -4.8227, -3.0837	-
		PSDI	3.6354 3.6199, 3.6506	-0.0666 -0.0684, -0.0650	-0.0111 -0.0126, -0.0098	-
		CDF	3.6438 3.6269, 3.6590	-0.0667 -0.0689, -0.0649	-0.0807 -0.1065, -0.0496	-
		BAL	3.6317 3.6118, 3.6492	-0.0699 -0.0733, -0.0671	-1.792e-05 -0.0002, 0.0002	-

Supplementary Table 3 Continued.

RM	PR	Base	24.9907 24.6101, 25.3819	3.4312 3.3181, 3.5503	-0.1043 -0.1083, -0.1005	-
		R _{dia}	25.1939 24.8207, 25.6302	3.1344 3.0269, 3.2433	-0.1221 -0.1272, -0.1170	0.5742 0.4743, 0.6772
		R _{BA}	25.9333 25.4247, 26.4220	3.7234 3.5883, 3.8542	-0.0883 -0.0935, -0.0832	-3.9445 -4.7927, -3.1355
		PSDI	29.0391 28.3687, 29.6476	4.7645 4.5181, 5.1036	-0.1048 -0.1111, -0.0999	0.0083 0.0048, 0.0120
		CDF	24.9929 24.6113, 25.3742	3.3805 3.2743, 3.5006	-0.1043 -0.1086, -0.1006	0.0356 0.0085, 0.0594
		BAL	24.9057 24.5565, 25.2907	3.2884 3.1515, 3.4326	-0.1055 -0.1094, -0.1013	0.0007 0.0005, 0.0009
	WK	Base	3.3702 3.3548, 3.3855	-0.1030 -0.1063, -0.0997	-	-
		R _{dia}	3.3630 3.3477, 3.3767	-0.1192 -0.1256, -0.1117	0.3438 0.1292, 0.4700	-
		R _{BA}	3.3203 3.3064, 3.3327	0.1448 0.1394, 0.1501	-8.4261 -9.1671, -7.5374	-

Supplementary Table 3 Continued.

		PSDI	3.4155 3.4040, 3.4286	-0.0892 -0.0914, -0.0871	-0.0159 -0.0170, -0.0149	-
		CDF	3.3939 3.3792, 3.4067	-0.0926 -0.0953, -0.0904	-0.2022 -0.2260, -0.1791	-
		BAL	3.3788 3.3620, 3.3941	-0.0993 -0.1031, -0.0962	-0.0005 -0.0007, -0.0002	-
SWG	PR	Base	27.6184 25.7726, 29.3909	5.1866 4.6136, 5.7766	-0.1232 -0.1413, -0.1080	-
		R _{dia}	27.6526 25.6456, 29.4505	4.6805 4.2018, 5.1971	-0.1424 -0.1662, -0.1253	0.5263 0.3030, 0.8233
		R _{BA}	28.7913 26.2803, 31.6479	5.2315 4.8033, 5.7403	-0.1019 -0.1276, -0.0813	-4.2833 -6.7276, -1.6055
		PSDI	28.0625 26.2333, 30.2547	4.6790 4.1057, 5.3574	-0.1169 -0.1342, -0.1011	0.0078 0.0016, 0.0142
		CDF	27.5825 25.8285, 29.3928	5.1427 4.4988, 5.7216	-0.1236 -0.1422, -0.1089	0.0232 -0.1061, 0.1412
		BAL	27.4006 25.8383, 29.1045	4.2844 3.7573, 4.8068	-0.1271 -0.1451, -0.1122	0.0032 0.0021, 0.0043

Supplementary Table 3 Continued.

	WK	Base	3.5038 3.4319, 3.5567	-0.0951 -0.1072, -0.0859	-	-
		R _{dia}	3.4873 3.4191, 3.5535	-0.1073 -0.1263, -0.0893	0.1737 -0.0550, 0.5018	-
		R _{BA}	3.4941 3.4428, 3.5455	0.1116 0.1002, 0.1204	-4.6515 -6.0882, -3.0530	-
		PSDI	3.5677 3.5033, 3.6249	-0.0803 -0.0905, -0.0729	-0.0099 -0.0124, -0.0075	-
		CDF	3.5387 3.4749, 3.5963	-0.0832 -0.0932, -0.0745	-0.2079 -0.2575, -0.1528	-
		BAL	3.4950 3.4365, 3.5640	-0.0981 -0.1115, -0.0839	0.0002 -0.0004, 0.0013	-
WO	PR	Base	27.8072 27.2805, 28.2779	3.9622 3.7503, 4.1885	-0.0902 -0.0955, -0.0858	-
		R _{dia}	27.9843 27.5590, 28.5163	3.3935 3.2010, 3.5775	-0.1100 -0.1165, -0.1040	0.6586 0.5435, 0.7695
		R _{BA}	27.9973 27.4516, 28.5290	4.5212 4.2584, 4.8021	-0.0869 -0.0926, -0.0822	-2.6981 -3.9203, -1.5748

Supplementary Table 3 Continued.

		PSDI	28.1145 27.5309, 28.7315	3.8587 3.6366, 4.0831	-0.0898 -0.0944, -0.0857	0.0097 0.0052, 0.0143
		CDF	27.8083 27.2923, 28.3202	3.9513 3.7271, 4.1751	-0.0903 -0.0946, -0.0860	0.0058 -0.0345, 0.0463
		BAL	27.8533 27.3922, 28.3246	4.0527 3.8094, 4.3153	-0.0903 -0.0946, -0.0861	-0.0002 -0.0006, 0.0001
	WK	Base	3.5558 3.5408, 3.5697	-0.0712 -0.0730, -0.0697	-	-
		R _{dia}	3.5354 3.5208, 3.5488	-0.0903 -0.0947, -0.0865	0.3847 0.3142, 0.4619	-
		R _{BA}	3.5109 3.4903, 3.5312	0.0869 0.0814, 0.0919	-3.0149 -3.8761, -2.1393	-
		PSDI	3.5530 3.5399, 3.5671	-0.0682 -0.0700, -0.0666	-0.0122 -0.0138, -0.0109	-
		CDF	3.5647 3.5488, 3.5800	-0.0686 -0.0710, -0.0666	-0.0671 -0.0962, -0.0370	-
		BAL	3.5643 3.5469, 3.5821	-0.0694 -0.0722, -0.0667	-2.557e-04 -5.115e-04, 1.313e-06	-

Supplementary Table 3 Continued.

YP	PR	Base	33.4603 32.7430, 34.1088	3.9643 3.7864, 4.1439	-0.0803 -0.0846, -0.7699	-
		R _{dia}	33.3744 32.6181, 34.1362	3.4285 3.2711, 3.5814	-0.0971 -0.1025, -0.0915	0.5232 0.4489, 0.6043
		R _{BA}	34.2722 33.4269, 35.0185	4.7131 4.4715, 4.9472	-0.0730 -0.0782, -0.0691	-4.1554 -5.0728, -3.1185
		PSDI	34.3319 33.5567, 35.2729	3.7980 3.6366, 3.9657	-0.0787 -0.0823, -0.0753	0.0130 0.0089, 0.0172
		CDF	33.4614 32.7558, 34.1837	3.8722 3.6928, 4.0446	-0.0805 -0.0849, -0.0766	0.0529 0.0170, 0.0908
		BAL	33.4518 32.7326, 34.1687	3.9395 3.7062, 4.1486	-0.0804 -0.0845, -0.0763	8.134e-05 -3.018e-04, 4.707e-04
	WK	Base	3.6959 3.6785, 3.7133	-0.0698 -0.0720, -0.0676	-	-
		R _{dia}	3.6712 3.6522, 3.6881	-0.0855 -0.0906, -0.0799	0.2807 0.1900, 0.3544	-
		R _{BA}	3.6421 3.6191, 3.6674	0.0874 0.0805, 0.0933	-3.3766 -4.3523, -2.3591	-

Supplementary Table 3 Continued.

		PSDI	3.6989 3.6830, 3.7135	-0.0645 -0.0661, -0.0626	-0.0172 -0.0187, -0.0156	-
		CDF	3.7129 3.6930, 3.7316	-0.0649 -0.0671, -0.0627	-0.1257 -0.1575, -0.0939	-
		BAL	3.4950 3.4365, 3.5640	-0.0981 -0.1115, -0.0839	0.0002 -0.0004, 0.0013	-

Supplementary Table 4. Parameter estimates (50% quantile) with lower (2.5% quantile) and upper (97.5% quantile) limits of the confidence interval from the 500 bootstrap samples for the parametric models (Pearl & Reed (1920), PR; Wykoff et al. (1982), WK) for species (eastern white pine, EWP; hickory species, HIC; red maple, RM; sweetgum, SWG; white oak, WO; yellow poplar) in Ridge and Valley (VR) for the base model and models with competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL)

Species	Model	Competition index	β_0	β_1	β_2	β_3
EWP	PR	Base	30.0861 28.6291, 31.7684	5.5853 5.2847, 5.8536	-0.0769 -0.0824, -0.0712	-
		R_{dia}	29.8979 28.6033, 31.4388	5.4850 5.2247, 5.7448	-0.0898 -0.0974, -0.0824	0.3181 0.2021, 0.4271
		R_{BA}	33.3657 31.1031, 36.0034	7.0894 6.6302, 7.5427	-0.0602 -0.0669, -0.0539	-6.0601 -7.0188, -5.0550
		PSDI	32.5921 30.3817, 35.8630	5.1296 4.8618, 5.3708	-0.0698 -0.0762, -0.0634	0.0156 0.0116, 0.0196
		CDF	29.9204 28.5713, 31.3462	5.5221 5.2254, 5.7757	-0.0777 -0.0839, -0.0721	0.0197 -0.0314, 0.0636
		BAL	29.7752 28.3560, 31.3014	5.4270 5.1394, 5.7249	-0.0783 -0.0845, -0.0726	0.0002 -0.0000, 0.0006

Supplementary Table 4 Continued.

	WK	Base	3.5161 3.4752, 3.5546	-0.0620 -0.0650, -0.0588	-	-
		R _{dia}	3.5095 3.4673, 3.5470	-0.0699 -0.0754, -0.0654	0.1689 0.0977, 0.2712	-
		R _{BA}	3.4409 3.4011, 3.4697	0.08522 0.07975, 0.0920	-4.6507 -5.6712, -3.8213	-
		PSDI	3.5635 3.5330, 3.5925	-0.0543 -0.0565, -0.0522	-0.0121 -0.0130, -0.0114	-
		CDF	3.5561 3.5139, 3.5905	-0.0543 -0.0574, -0.0519	-0.1913 -0.2205, -0.1658	-
		BAL	3.5224 3.4805, 3.5611	-0.0616 -0.0654, -0.0579	-8.100e-05 -3.028e-04, 1.621e-04	-
HICK	PR	Base	28.4448 27.9824, 28.9344	4.8839 4.7331, 5.0620	-0.1075 -0.1115, -0.1036	-
		R _{dia}	28.5626 28.1259, 28.9907	4.6058 4.4481, 4.7595	-0.1257 -0.1310, -0.1208	0.5324 0.4447, 0.6171
		R _{BA}	29.6419 29.0021, 30.2742	5.3134 5.1401, 5.4661	-0.0905 -0.0961, -0.0854	-4.5912 -5.4287, -3.7263

Supplementary Table 4 Continued.

		PSDI	29.0262 28.5348, 29.5912	4.4224 4.2303, 4.6327	-0.1015 -0.1059, -0.0976	0.0088 0.0061, 0.0110
		CDF	28.4603 28.0241, 28.8856	4.7728 4.6162, 4.9438	-0.1077 -0.1114, -0.1036	0.0541 0.0109, 0.0870
		BAL	28.4214 27.9411, 28.8918	4.8135 4.6430, 4.9873	-0.1079 -0.1121, -0.1038	0.0001 -9.3395, 0.0003
	WK	Base	3.5536 3.5379, 3.5698	-0.0811 -0.0833, -0.0788	-	-
		R _{dia}	3.5472 3.5321, 3.5621	-0.0938 -0.0984, -0.0906	0.2838 0.2185, 0.3608	-
		R _{BA}	3.4983 3.4842, 3.5141	0.1090 0.1050, 0.1128	-5.7995 -6.4584, -5.2060	-
		PSDI	3.5964 3.5835, 3.6099	-0.0713 -0.0729, -0.0698	-0.0116 -0.0123, -0.0109	-
		CDF	3.5784 3.5626, 3.5943	-0.0732 -0.0751, -0.0712	-0.1760 -0.1947, -0.1548	-
		BAL	3.5754 3.5573, 3.5936	-0.0753 -0.0778, -0.0729	-0.0004 -0.0005, -0.0003	-

Supplementary Table 4 Continued.

RM	PR	Base	24.7433 24.4433, 25.0963	3.7144 3.6344, 3.8076	-0.1114 -0.1144, -0.1081	-
		R _{dia}	24.9376 24.6039, 25.2691	3.4410 3.3606, 3.5261	-0.1314 -0.1355, -0.1275	0.6115 0.5287, 0.6850
		R _{BA}	25.6806 25.2422, 26.0913	3.7674 3.6733, 3.8509	-0.0932 -0.0984, -0.0888	-3.4425 -4.1517, -2.8113
		PSDI	24.8903 24.5172, 25.2642	3.5620 3.4386, 3.6892	-0.1090 -0.1128, -0.1055	0.0033 0.0017, 0.0050
		CDF	24.7254 24.3637, 25.0195	3.6407 3.5530, 3.7353	-0.1122 -0.1156, -0.1093	0.0565 0.0284, 0.0816
		BAL	24.6457 24.3042, 24.9594	3.6163 3.5162, 3.7050	-0.1136 -0.1172, -0.1105	0.0004 0.0002, 0.0006
	WK	Base	3.2859 3.2706, 3.3005	-0.1277 -0.1321, -0.1239	-	-
		R _{dia}	3.2878 3.2722, 3.3031	-0.1232 -0.1277, -0.1185	-0.0909 -0.1665, -0.0205	-
		R _{BA}	3.2922 3.2800, 3.3022	0.1607 0.1565, 0.1648	-10.2311 -10.9057, -9.6060	-

Supplementary Table 4 Continued.

		PSDI	3.5964 3.5835, 3.6099	-0.0713 -0.0729, -0.0698	-0.0116 -0.0123, -0.0109	-
		CDF	3.3398 3.3239, 3.3534	-0.1041 -0.1075, -0.1015	-0.2830 -0.2968, -0.2689	-
		BAL	3.5224 3.4805, 3.5611	-0.0616 -0.0654, -0.0579	-8.100e-05 -3.028e-04, 1.621e-04	-
SWG	PR	Base	29.1202 28.4541, 29.6843	5.6413 5.4365, 5.8347	-0.1148 -0.1189, -0.1111	-
		R _{dia}	29.1628 28.5065, 29.7364	4.9867 4.8176, 5.1695	-0.1307 -0.1356, -0.1255	0.4686 0.3948, 0.5492
		R _{BA}	30.2695 29.4534, 31.0752	5.7347 5.5445, 5.9253	-0.0973 -0.1045, -0.0917	-3.6599 -4.5220, -2.7758
		PSDI	29.3094 28.3318, 30.1800	5.4215 5.0547, 5.7536	-0.1125 -0.1206, -0.1065	0.0035 0.0009, 5.1281
		CDF	29.1150 28.5335, 29.7356	5.4832 5.2737, 5.7149	-0.1147 -0.1194, -0.1105	0.0555 0.0110, 0.1036
		BAL	28.9883 28.3953, 29.5401	5.2212 4.9848, 5.4439	-0.1156 -0.1198, -0.1115	0.0007 0.0005, 0.0009

Supplementary Table 4 Continued.

	WK	Base	3.5264 3.4925, 3.5561	-0.0886 -0.0939, -0.0846	-	-
		R _{dia}	3.5234 3.4897, 3.5518	-0.0900 -0.0958, -0.0848	0.0154 -0.0553, 0.1056	-
		R _{BA}	3.5024 3.4807, 3.5221	0.1140 0.1095, 0.1180	-6.5034 -7.1826, -5.8506	-
		PSDI	3.6111 3.5870, 3.6395	-0.0715 -0.0739, -0.0685	-0.0134 -0.0145, -0.0119	-
		CDF	3.5765 3.5497, 3.6030	-0.0747 -0.0781, -0.0716	-0.2116 -0.2312, -0.1898	-
		BAL	3.5545 3.5228, 3.5787	-0.0806 -0.0853, -0.0769	-0.0005 -0.0006, -0.0004	-
WO	PR	Base	28.9254 28.4348, 29.4034	3.9917 3.8409, 4.1389	-0.0866 -0.0903, -0.0830	-
		R _{dia}	28.9073 28.4718, 29.3904	3.5129 3.3716, 3.6546	-0.1089 -0.1140, -0.1041	0.6419 0.5763, 0.7099
		R _{BA}	29.9035 29.2977, 30.5503	4.6301 4.4789, 4.7826	-0.0743 -0.0783, -0.0697	-4.9346 -5.7415, -4.2623

Supplementary Table 4 Continued.

		PSDI	29.5717 28.9763, 30.2938	3.7058 3.5632, 3.8571	-0.0831 -0.0874, -0.0793	0.0109 0.0081, 0.0136
		CDF	28.8878 28.4159, 29.4418	3.9369 3.7856, 4.0937	-0.0865 -0.0903, -0.0831	0.0150 -0.0138, 0.0488
		BAL	28.9821 28.5588, 29.4749	4.1007 3.9218, 4.2656	-0.0860 -0.0895, -0.0824	-0.0003 -0.0005, -0.0001
	WK	Base	3.5566 3.5410, 3.5689	-0.0739 -0.0759, -0.0722	-	-
		R _{dia}	3.5345 3.5210, 3.5481	-0.0946 -0.0984, -0.0908	0.4007 0.3324, 0.4636	-
		R _{BA}	3.5037 3.4881, 3.5200	0.0936 0.0889, 0.0979	-3.8420 -4.5332, -3.1689	-
		PSDI	3.5662 3.5527, 3.5780	-0.0685 -0.0704, -0.0670	-0.0131 -0.0143, -0.0121	-
		CDF	3.5728 3.5575, 3.5882	-0.0683 -0.0701, -0.0664	-0.1517 -0.1709, -0.1294	-
		BAL	3.5748 3.5586, 3.5894	-0.0690 -0.0714, -0.0668	-0.0005 -0.0006, -0.0003	-

Supplementary Table 4 Continued.

YP	PR	Base	34.1086 33.6381, 34.5364	4.1688 4.0438, 4.3019	-0.0855 -0.0880, -0.0830	-
		R _{dia}	34.0084 33.6053, 34.4720	3.6765 3.5650, 3.7880	-0.1000 -0.1038, -0.0963	0.4607 0.4026, 0.5189
		R _{BA}	36.0678 35.4507, 36.5987	5.0274 4.8615, 5.1620	-0.0677 -0.0707, -0.0651	-6.6262 -7.2231, -6.0018
		PSDI	35.0931 34.4309, 35.6864	3.7874 3.6541, 3.9198	-0.0819 -0.0849, -0.0793	0.0157 0.0132, 0.0182
		CDF	34.0082 33.5423, 34.4695	4.0677 3.9422, 4.2032	-0.0859 -0.0886, -0.0832	0.0415 0.0123, 0.0652
		BAL	33.9283 33.4468, 34.4320	4.0330 3.8890, 4.1810	-0.0863 -0.0889, -0.0836	0.0002 0.0001, 0.0004
	WK	Base	3.6913 3.6740, 3.7050	-0.0766 -0.0790, -0.0748	-	-
		R _{dia}	3.6847 3.6715, 3.6968	-0.0816 -0.0863, -0.0774	0.0866 0.0057, 0.1728	-
		R _{BA}	3.6384 3.6231, 3.6544	0.1015 0.0970, 0.1060	-5.2462 -5.9702, -4.4787	-

Supplementary Table 4 Continued.

		PSDI	3.7109 3.6988, 3.7237	-0.0686 -0.0702, -0.0672	-0.0163 -0.0177, -0.0150	-
		CDF	3.7194 3.7062, 3.7316	-0.0675 -0.0691, -0.0661	-0.2282 -0.2464, -0.2079	-
		BAL	3.7089 3.6939, 3.7232	-0.0715 -0.0737, -0.0694	-0.0006 -0.0007, -0.0004	-

Supplementary Table 5. Parameter estimates (50% quantile) with lower (2.5% quantile) and upper (97.5% quantile) limits of the confidence interval from the 500 bootstrap samples for the parametric models (Pearl & Reed (1920), PR; Wykoff et al. (1982), WK) for four pantropical species in Caribbean islands (CAR) (pink trumpet tree, PTT; muskwood, MW; African tulip tree, ATT; acacia, ACA) and three pantropical species in Pacific islands (PAC) (African tulip tree, ATT; ‘Ōhi‘a lehua, OL, acacia, ACA) for the base model and models with competition indices (diameter ratio, R_{dia} ; basal area ratio, R_{BA} ; partitioned stand density index, PSDI; cumulative distribution function, CDF; basal area of larger trees, BAL)

Species	Model	Competition index	β_0	β_1	β_2	β_3
PTT	PR	Base	16.2185 14.3567, 18.3359	4.7370 4.0537, 5.4184	-0.1399 -0.1595, -0.1239	-
		R_{dia}	16.6687 14.5692, 18.4788	4.4380 3.9900, 4.8670	-0.1784 -0.2058, -0.1594	1.2528 0.7597, 1.7598
		R_{BA}	16.5661 (14.7181, 18.3507)	5.2721 (4.5288, 5.9659)	-0.1604 (-0.1972, -0.1328)	3.6506 (-1.0742, 8.7450)
		PSDI	16.6224 14.7668, 18.3859	5.8834 4.8823, 7.0951	-0.1507 -0.1681, -0.1357	-0.0035 -0.0058, -0.0012
		CDF	16.8296 14.8223, 18.6243	4.8270 4.2917, 5.4740	-0.1423 -0.1607, -0.1301	0.0761 -0.0489, 0.2003

Supplementary Table 5 Continued.

		BAL	16.8855 14.8947, 18.6255	4.5270 4.0333, 5.0761	-0.1426 -0.1609, -0.1290	0.0013 0.0007, 0.0019
	WK	Base	2.7829 2.7047, 2.8847	-0.1714 -0.1933, -0.1510	-	-
		R _{dia}	2.8143 2.7113, 2.9158	-0.2097 -0.2405, -0.1839	1.0200 0.4158, 1.6379	-
		R _{BA}	2.8860 2.7272, 3.0096	-0.1642 -0.1846, -0.1458	6.2446 -4.6078, 9.5717	-
		PSDI	2.9523 2.7793, 3.0953	0.1156 0.0940, 0.1563	0.0041 0.0025, 0.0052	-
		CDF	2.9088 2.7900, 3.0233	-0.1264 -0.1502, -0.1072	-0.2377 -0.2850, -0.1902	-
		BAL	2.8545 2.7518, 2.9642	0.0003 -0.0000, 0.0006	2.8545 2.7518, 2.9628	-
MW	PR	Base	19.2050 17.7013, 20.9978	3.3393 3.1392, 3.5224	-0.0850 -0.0978, -0.0731	-
		R _{dia}	19.1355 17.6342, 20.8713	3.3087 3.1071, 3.4955	-0.0936 -0.1192, -0.0738	0.2968 -0.1497, 0.8441

Supplementary Table 5 Continued.

		R _{BA}	19.9984 18.0977, 21.8564	3.3739 3.1365, 3.5684	-0.0655 -0.0858, -0.0534	-4.4545 -6.0307, -2.5037
		PSDI	19.5678 17.7935, 21.9750	3.0792 2.7522, 3.3834	-0.0795 -0.0982, -0.0639	0.0012 -0.0002, 0.0026
		CDF	19.2042 17.7111, 20.9857	3.2638 3.0677, 3.4382	-0.0858 -0.0985, -0.0729	0.0719 -0.0234, 0.1511
		BAL	19.1658 17.6979, 20.8648	3.2659 3.0594, 3.4627	-0.0868 -0.1007, -0.0734	0.0006 -0.0000, 0.0013
	WK	Base	2.8720 2.8211, 2.9194	-0.1639 -0.1796, -0.1483	-	
		R _{dia}	2.8716 2.8197, 2.9236	-0.1675 -0.2079, -0.1358	0.1562 -0.7703, 0.9144	-
		R _{BA}	2.9000 2.8548, 2.9421	-0.1920 -0.2153, -0.1664	15.9246 5.7194, 21.2409	-
		PSDI	2.9959 2.9190, 3.0685	0.1178 0.1019, 0.1368	-0.0203 -0.0234, -0.0173	-
		CDF	2.9332 2.8664, 2.9840	-0.1242 -0.1448, -0.1098	-0.3778 -0.4257, -0.3218	-

Supplementary Table 5 Continued.

		BAL	2.8883 2.8349, 2.9478	0.1495 0.1237, 0.1672	0.0013 0.0006, 0.0019	-
ATT	PR	Base	19.2050 18.5210, 21.2858	3.5446 3.2446, 3.8896	-0.0942 -0.1094, -0.0771	-
		R _{dia}	19.8947 18.5846, 21.2691	3.4395 3.1596, 3.7615	-0.1187 -0.1417, -0.0925	0.9545 0.4285, 1.4630
		R _{BA}	20.7893 19.2200, 23.5668	3.7400 3.4007, 4.0884	-0.0731 -0.0935, -0.0515	-5.0747 -7.6644, -2.7156
		PSDI	20.6565 19.0081, 25.0426	2.9305 2.5909, 3.3865	-0.0811 -0.1002, -0.0527	0.0032 0.0017, 0.0061
		CDF	19.8161 18.5228, 21.2917	3.5482 3.1949, 3.9441	-0.0942 -0.1091, -0.0769	-0.0054 -0.1035, 0.0957
		BAL	19.8129 18.5158, 21.2486	3.3707 3.0711, 3.7241	-0.0968 -0.1117, -0.0802	0.0016 0.0004, 0.0033
	WK	Base	3.0370 2.9708, 3.0903	-0.1203 -0.1361, -0.1088	-	
		R _{dia}	3.0307 2.9662, 3.0806	-0.1541 -0.1819, -0.1305	1.1309 0.4677, 1.1770	-

Supplementary Table 5 Continued.

		R _{BA}	3.0392 2.9699, 3.0901	-0.1435 -0.1660, -0.1315	7.6601 5.5651, 10.0090	-
		PSDI	3.1503 3.0712, 3.2139	0.0981 0.0860, 0.1132	-0.0157 -0.0192, 0.0038	-
		CDF	3.0997 3.0293, 3.1522	-0.0937 -0.1058, -0.0840	-0.3333 -0.3870, -0.2848	-
		BAL	3.0365 2.9703, 3.0922	0.1220 0.1100, 0.1386	-0.0003 -0.0027, 0.0006	-
ACA	PR	Base	13.2232 10.3070, 15.3875	4.1380 2.7281, 5.5748	-0.1751 -0.2237, -0.1168	-
		R _{dia}	13.8366 10.8052, 15.5474	3.5227 2.6057, 4.3507	-0.2328 -0.3013, -0.1532	1.9133 0.7062, 3.0628
		R _{BA}	13.3886 10.0454, 16.7967	3.4240 2.4783, 4.4913	-0.1327 -0.1875, -0.0937	-2.1309 -8.8509, 3.7132
		PSDI	13.1338 10.1304, 16.3661	3.3364 2.2494, 4.2675	-0.1449 -0.1981, -0.0964	0.0014 -0.0040, 0.0072
		CDF	13.3390 9.9816, 17.4577	3.5208 2.4424, 4.5695	-0.1405 -0.1875, -0.0896	-0.0156 -0.1801, 0.1222

Supplementary Table 5 Continued.

		BAL	14.2332 10.4440, 18.1797	2.8911 1.9494, 3.7632	-0.1325 -0.1743, -0.0936	0.0025 0.0013, 0.0038
	WK	Base	2.6037 2.3702, 2.7894	-0.2131 -0.2848, -0.1762	-	-
		R _{dia}	2.5107 2.3247, 2.6724	-0.2776 -0.3949, -0.2166	0.6268 -0.2159, 1.5297	-
		R _{BA}	2.4409 2.2874, 2.5891	-0.1745 -0.2609, -0.1221	-65.6300 -100.1438, -21.6672	-
		PSDI	2.5268 2.3377, 2.6973	-0.2286 -0.3127, -0.1798	-0.0013 -0.0070, 0.0023	-
		CDF	2.5323 2.2609, 2.6996	-0.2138 -0.6059, -0.1671	-0.2124 -0.3085, 2.7653	-
		BAL	2.4931 2.2924, 2.6683	-0.2523 -0.3491, -0.1996	0.0002 -0.0003, 0.0008	-
ATT	PR	Base	19.6211 15.0505, 27.4414	3.7887 2.9559, 4.6406	-0.1054 -0.1842, -0.0509	-
		R _{dia}	20.7904 14.5396, 31.8210	3.7372 2.7299, 5.2252	-0.1031 -0.2110, -0.0603	0.2317 -0.7440, 0.6648

Supplementary Table 5 Continued.

		R _{BA}	27.0456 14.0685, 2445619	5.2416 3.4142, 550620.1	-0.0421 -0.2238, -0.0137	-8.7997 -13.3689, 9.3110
		PSDI	5.290e+05 14.2795, 1.411e+07	8.775e+04 3.3511, 2.781e+06	-0.0541 -0.1871, -0.0442	0.0198 -0.0153, 0.0282
		CDF	18.1755 14.3376, 26.3378	3.2713 2.0834, 4.5306	-0.1123 -0.1851, -0.0558	0.2942 -0.0910, 0.9396
		BAL	21.1477 14.4421, 30.6280	3.4696 2.4420, 4.6014	-0.0845 -0.1735, -0.0467	0.0266 -0.0071, 0.0754
	WK	Base	3.0067 2.8237, 3.2532	-0.1402 -0.1955, -0.0852	-	-
		R _{dia}	2.9398 2.7833, 3.2150	-0.1461 -0.2427, -0.0758	-0.0037 -0.5372, 0.6061	-
		R _{BA}	2.9613 2.7681, 3.1516	-0.1742 -0.2105, -0.1285	10.2660 -2.9846, 18.1355	-
		PSDI	3.0784 2.7636, 3.3138	-0.1156 -0.2919, -0.0603	-0.0106 -0.0218, 0.1520	-
		CDF	3.0043 2.7834, 3.2410	-0.1186 -0.1972, -0.0720	-0.2462 -0.3675, -0.0345	-

Supplementary Table 5 Continued.

OL		BAL	2.9654 2.7573, 3.2489	0.1459 0.0679, 0.2114	0.0016 -0.1161, 0.0208	-
	PR	Base	21.2951 19.7954, 22.8454	4.0541 3.6863, 4.4333	-0.0915 -0.1053, -0.0800	-
		R _{dia}	21.4956 20.4900, 23.7074	3.7185 3.4267, 3.9856	-0.1025 -0.1200, -0.0876	0.2536 0.1653, 0.3582
		R _{BA}	24.4674 22.1153, 26.6676	4.5313 4.0751, 4.9687	-0.0595 -0.0754, -0.0495	-5.9938 -7.7234, -4.1295
		PSDI	22.9081 20.7115, 25.5317	3.8223 3.4930, 4.1964	-0.0808 -0.0966, -0.0696	0.0046 0.0010, 0.0084
		CDF	22.0535 20.2636, 23.9228	4.0528 3.7062, 4.3845	-0.0858 -0.1005, -0.0736	-0.0128 -0.0698, 0.0495
		BAL	22.2194 20.6708, 23.8639	3.4488 3.0042, 4.0084	-0.0826 -0.0935, -0.0732	0.0135 0.0020, 0.0229
	WK	Base	3.0873 3.0156, 3.1601	-0.1083 -0.1229, -0.0936	-	-
		R _{dia}	3.0861 3.0076, 3.1626	-0.1252 -0.1480, -0.1050	0.1940 0.0863, 0.3124	-

Supplementary Table 5 Continued.

		R _{BA}	3.0970 3.0465, 3.1562	-0.1363 -0.1510, -0.1237	9.3312 7.7628, 10.9146	-
		PSDI	3.1874 3.1241, 3.2460	-0.0786 -0.0980, -0.0710	-0.0111 -0.0128, 0.0382	-
		CDF	3.1565 3.0860, 3.2194	-0.0858 -0.0985, -0.0760	-0.2739 -0.3134, -0.2331	-
		BAL	3.1997 3.1215, 3.2667	0.0780 0.0680, 0.0906	0.0155 0.0127, 0.0183	-
ACA	PR	Base	22.6600 19.2965, 25.7042	2.6950 2.1612, 3.3612	-0.0602 -0.0912, -0.0459	-
		R _{dia}	22.6339 19.1517, 25.9851	2.7105 2.1774, 3.2853	-0.0623 -0.1094, -0.0438	0.0367 -0.1636, 0.2840
		R _{BA}	25.3671 22.2891, 28.4247	4.4140 3.5910, 5.2479	-0.0391 -0.0509, -0.0307	-10.9925 -13.8392, -8.1018
		PSDI	24.6027 20.6974, 31.0239	2.8203 2.2478, 3.7270	-0.0601 -0.0813, -0.0460	0.0094 0.0057, 0.1284
		CDF	22.9876 19.0835, 26.0529	2.6818 2.2262, 3.3306	-0.0588 -0.0967, -0.0455	0.0234 -0.0723, 0.1159

Supplementary Table 5 Continued.

		BAL	22.8799 19.6496, 25.8125	2.4774 1.9119, 3.3056	-0.0590 -0.0864, -0.0450	0.0110 -0.0102, 0.0317
	WK	Base	3.1247 3.0383, 3.2192	-0.0971 -0.1178, -0.0781	-	-
		R _{dia}	3.1286 3.0421, 3.2208	-0.0945 -0.1303, -0.0669	-0.0287 -0.3028, 0.2843	-
		R _{BA}	3.1200 3.0208, 3.2348	-0.1002 -0.1318, -0.0721	0.6603 -2.4319, 4.4004	-
		PSDI	3.1425 3.0565, 3.2391	-0.0897 -0.1270, -0.0724	-0.0054 -0.0092, 0.0414	-
		CDF	3.1576 3.0572, 3.2520	-0.0847 -0.1038, -0.0675	-0.2612 -0.3275, -0.1818	-
		BAL	3.1596 3.0672, 3.2610	0.0846 0.0630, 0.1076	0.0142 0.0025, 0.0224	-

VITA

Sakar Jha is originally from Nepal where he received a Bachelor of Science in Forestry at Tribhuvan University, Institute of Forestry, Pokhara Campus, Nepal. As an undergraduate, Sakar actively worked on several research projects that helped him gain the knowledge and skill necessary to become a researcher. He was also interested in communicating scientific knowledge through effective mediums, which led him to write several articles in newspapers and online magazines.

Sakar joined the Forest Biometrics lab at the University of Tennessee, Knoxville as a graduate student to complete his MS in Forestry under Dr. Sheng I Yang. As a student at UTK, Sakar has worked on several research projects that involved using advanced statistical techniques and modelling techniques to answer contemporary questions in the field of forest biometrics. He has had the opportunity to present his research in front of the scientific community in the form of original research articles and oral presentations at several conferences like the 2nd North American forest mensurationists conference, 22nd Biennial Southern Silvicultural conference and the Joint Southern and Northern Mensurationists conference. Other than schoolwork, Sakar has also participated in timbersports competition with the UTK forestry club and JiuJitsu with the UTK Martial arts club.