sustainability

MDPI

Article

Disentangling the Factors That Contribute to the Growth of *Betula* spp. and *Cunninghami lanceolata* in China Based on Machine Learning Algorithms

Hanyue Zhang [1], Zhongke Feng [1,2,3,*], Shan Wang [1] and Wenxu Ji [1]

[1] Precision Forestry Key Laboratory of Beijing, Beijing Forestry University, Beijing 100083, China; zhanghanyue@bjfu.edu.cn (H.Z.); wangshan@bjfu.edu.cn (S.W.); wenxuji@126.com (W.J.)
[2] Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants, Ministry of Education, College of Forestry, Hainan University, Haikou 570228, China
[3] Intelligent Forestry Key Laboratory of Haikou City, College of Forestry, Hainan University, Haikou 570228, China
* Correspondence: zhongkefeng@bjfu.edu.cn; Tel.: +86-138-1030-5579

Abstract: Forests are indispensable materials and spiritual foundations for promoting ecosystem circulation and human survival. Exploring the environmental impact mechanism on individual-tree growth is of great significance. In this study, the effects of biogeoclimate, competition, and topography on the growth of *Betula* spp. and *Cunninghamia lanceolata* (Lamb.) Hook., two tree species with high importance value in China, were explored by gradient boosting regression tree (GBRT), k-nearest neighbor (KNN), and random forest (RF) machine learning (ML) algorithms. The results showed that the accuracy of RF was better than KNN, which was better than GBRT. All ML algorithms performed well for future diameter at breast height (DBH) predictions; the Willmott's indexes of agreement (WIA) of each ML algorithm in predicting the future DBH were all higher than 0.97, and the R^2 was higher than 0.98 and 0.90, respectively. The individual tree annual growth rate is mainly affected by the single-tree size, and the external environment can promote or inhibit tree growth. Climate and stand structure variables were relatively more important for tree growth than the topographic factors. Lower temperature and precipitation, higher stand density, and canopy closure were more unfavorable for their growth. In afforestation, the following factors should be considered in order: geographic location, meteorological climate, stand structure, and topography.

Keywords: gradient boosting regression tree; k-nearest neighbor; random forest; environmental impact mechanism; individual-tree annual growth rate; future DBH prediction

check for updates

Citation: Zhang, H.; Feng, Z.; Wang, S.; Ji, W. Disentangling the Factors That Contribute to the Growth of *Betula* spp. and *Cunninghami lanceolata* in China Based on Machine Learning Algorithms. *Sustainability* **2022**, *14*, 8346. https://doi.org/10.3390/su14148346

Academic Editor: Carmelo Maria Musarella

Received: 25 May 2022
Accepted: 5 July 2022
Published: 7 July 2022

1. Introduction

Known as the "lungs of the earth", forests are indispensable materials and spiritual foundations for promoting ecosystem circulation and human survival, and they play an important role in climate regulation, the deposition of organic matter, and carbon dioxide fixation [1–3]. The vitality of trees, reflecting changes of forest system dynamics, is one of the most important indicators of forest conditions [4]. Tree growth is known to be related to biophysical site attributes such as available sunlight, climate, and soil nutrient and competition, and the relationships between them is complex [5]. Quantifying the impact of different factors on tree growth is of great significance to achieve the efficient management of forest resources, to improve forest carbon sink capacity, and to promote the construction of ecological civilization [6].

Forest growth is an important indicator of sustainable forest management, and the tree growth rate, a transformed form of tree growth, will not change the connotative information of tree growth [7]. The individual tree model, individual-tree, is a modeling unit that considers the stand environment and competition and is an effective tool that may guide long-term planning and management decisions, offering technical subsides for the sustainable development of the forest [8–11]. In field surveys, diameter at breast height (DBH) is the simplest and most easily measurable factor, and its accuracy can be effectively guaranteed during the measurement process [12,13]. Tree growth provides an effective reference for forest growth and harvesting research, and the DBH growth models are more effective in future forest forecasting [14,15]. There are more obvious advantages in terms of growth rate trends than growth increment trends; thus, it is advantageous to establish a DBH growth rate model to explore the influence of the factors on tree growth.

With the rapid development of computer science and technology, machine learning (ML) algorithms have been widely used in forestry model simulation research to efficiently deal with nonlinear and interactive problems. ML algorithms have no prior assumptions about the independence of predictors, are able to fit complex nonlinear relationships, and are highly resistant to the inclusion of a large number of uncorrelated predictors. ML algorithms such as gradient boosting regression tree (GBRT), k-nearest neighbor (KNN), and random forest (RF) have significant advantages in prediction problems. GBRT models address complex nonlinear relationships and interactions between variables, and they are commonly used in ecological studies [16–19]. A. Beaudoin mapped attributes of Canada's forests at a moderate resolution through KNN with 26 geospatial data layers, including MODIS spectral data and climatic and topographic variables [20]. Kilham et al. demonstrated that the replication of NFI (National Forest resources Inventory) harvest patterns can be improved with a stratified procedure based on RF algorithms, which can become important components in generating global wood supply scenarios [21]. GBRT uses a linear combination of multiple learners to predict, avoiding the need for the limited ability of a single learner; the prediction effect is not good, and it has good generalization ability. KNN is an instance-based lazy learning algorithm that learns complex target functions quickly without losing information. RF can handle a large number of automatic variables and select important variables automatically, is not affected by multicollinearity among variables, and is more flexible to evaluate complex interactions between variables.

Arbor is the main body of forest vegetation, and its function and status are usually reflected by the important value of arbor species. The certain tree species with the greater the number, the wider distribution, and the larger amount in the forest would obtain the higher the importance value. *Betula* spp. and *C. lanceolata* are the two species groups with high importance value in China, accounting for 8.79% and 6.36% of the national forest stock, respectively. Relevant studies have shown that many factors such as the tree itself, climate, site, soil, spatial location, and competition would affect their growth [14,22]. In this study, three different machine learning data processing techniques (GBRT, KNN, and RF) were used to estimate the individual annual growth rate of *Betula* spp. and *C. lanceolata* in China using National Forest Inventory (NFI) data. The specific objectives were to (1) explore the influence mechanism of environmental factors such as meteorological climate, topography, and stand structure on their growth, (2) and provide a scientific basis for the growth predictions and precise management of the two species.

2. Materials and Methods

2.1. Study Area

China is a vast country with a complex and diverse topography, situated within six climatic zones, with a high latitudinal difference between the north and south of the country and a high western and low eastern terrain. The rich and diverse climate and physical geography could support a wide range of climatic zone plants, from boreal forests to tropical rainforest vegetation, providing abundant material resources for human beings [23,24]. The NFIs have been conducted in China every five years since the 1970s in China, on 415,000 plots plotted on kilometer grids based on topographic maps at 1:50,000 or 1:100,000 scales, with the general plot of 0.0667 hectares, but some plots can be up to 0.08 or 0.06 hectares [24,25]. In those permanent plots, all the trees with DBH greater than 5 cm were measured; the retest accuracy rate of the plots is more than 98% and that of the trees is more than 95% [26]. In this study, the distribution of the surveying data was shown in Figure 1. *Betula* spp. is distributed throughout the country, mainly in northern regions such as Heilongjiang, Jilin, and Inner Mongolia, and *C. lanceolata* is mainly distributed in southern regions of China, such as Fujian, Hunan, Jiangxi, and Guangxi provinces.

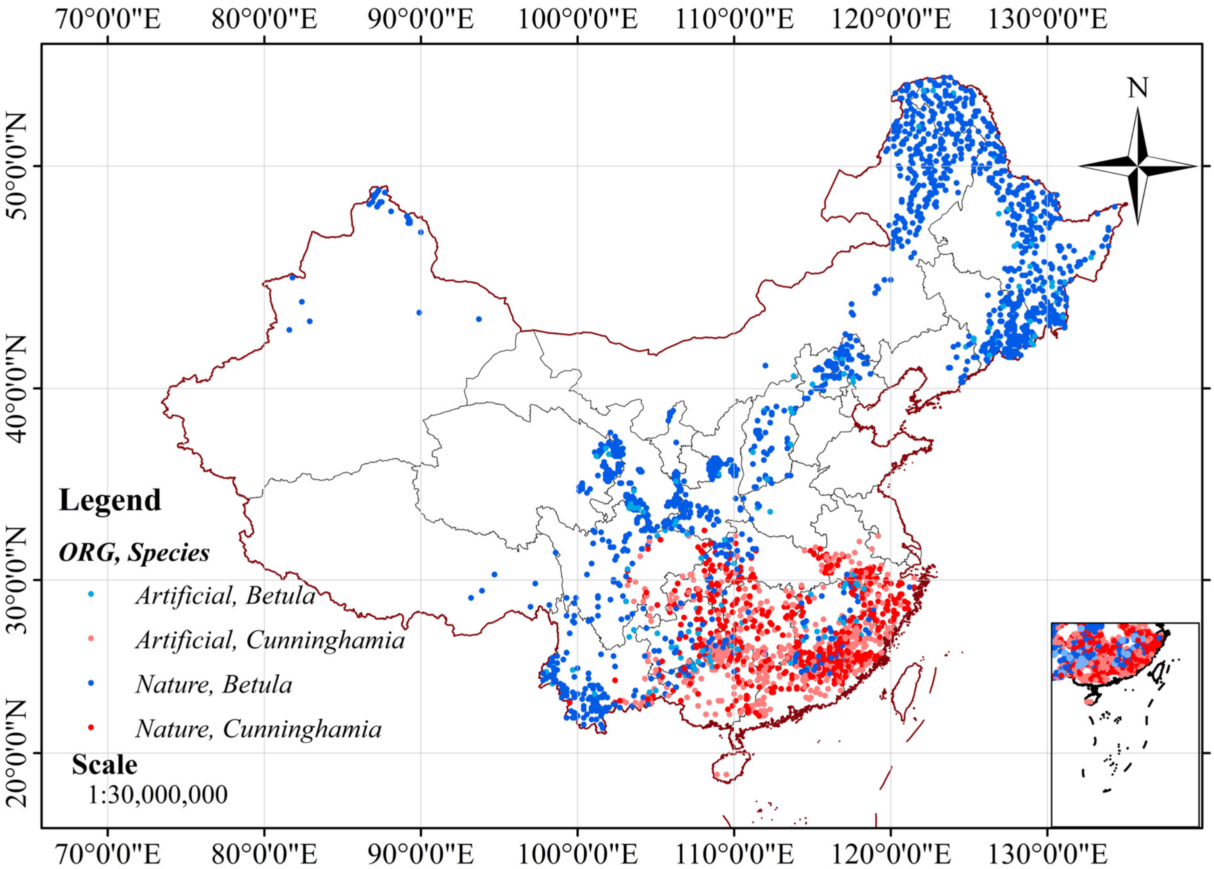

Figure 1. Distribution of the surveying data of *Betula* spp. and *Cunninghamia lanceolata* (Lamb.) Hook. in this study.

2.2. Data

In this study, the latest partial NFI data from 1999–2003, 2004–2008, 2009–2013, and 2014–2018 for *C. lanceolata* and *Betula* spp. and their meteorological data were used. NFI data mainly consisted of the tree species and size, spatial location, topographical conditions, soil information, and stand structure. In addition, the meteorological data of the sample plots during the inventory period were extracted from the annual average temperature and precipitation maps in China, which were obtained by interpolating the meteorological station data provided on the website (https://gis.ncdc.noaa.gov (accessed on 10 December 2019) and http://data.cma.cn (accessed on 22 December 2019)). In this research, the DBH growth rates of 46,351 *C. lanceolata* samples and 39,301 *Betula* spp. samples in China were explored based on the 2796 permanent plots. Among them, there were 2177 natural forest permanent plots, that included 12,512 *C. lanceolata* samples and 37,939 *Betula* spp. samples, and 619 artificial forest permanent plots, with a total of 33,839 *C. lanceolata* samples and 1342 *Betula* spp. samples. The statistical results of the NFI data used in this study are shown in Table 1. The latitudes of the permanent plots of the *Betula* spp. and *C. lanceolata* were from 18.7° N to 32.7° N and 21.5° N to 35.5° N, and the average annual temperature was from −3.9 °C to 24.6 °C and 12.6 °C to 25.6 °C, respectively. Besides, the DBH of *Betula* spp. ranged from 5.0 cm to 83.0 cm and that of *C. lanceolata* ranged from 5.0 cm to 49.0 cm.

Table 1. Statistics of trees and stand variables of *Betula* spp. and *Cunninghamia lanceolata* (Lamb.) Hook. for model calibration and validation.

Statistics	Unit	*Betula* spp.				*Cunninghamia lanceolata* (Lamb.) Hook.			
		Min	Median	Mean	Max	Min	Median	Mean	Max
Annual growth rate	%	0.03	1.33	2.03	45.08	0.06	3.44	4.69	49.31
DBH	cm	5.0	10.1	11.8	83.0	5.0	9.7	10.7	49.0
Latitude	°	21.5	41.1	41.9	53.5	18.7	26.3	26.9	32.7
Longitude	°	81.6	116.0	113.6	134.3	98.3	114.6	113.5	121.5
Elevation	m	70	1340	1679	4100	20	580	628	2160
Annual average temperature	°C	−3.9	4.6	4.6	24.6	12.6	19.0	19.1	25.6
Annual precipitation	mm	56.7	439.3	491.8	2053.1	666.0	1488.0	1476.0	2274.0
Soil thickness	cm	2	45	45	200	15	80	77	150
Humus thickness	cm	0	4	6	70	0	5	8	60
SinSL	-	0.0	0.3	0.3	0.8	0.0	0.5	0.5	0.8
(CosAZ + 1)/2	-	0.0	0.5	0.4	1.0	0.0	0.5	0.5	1.0
Stand canopy density	-	0.2	0.7	0.7	1.0	0.2	0.7	0.7	1.0
Simpson's diversity index	-	0.0	0.4	0.4	0.9	0.0	0.3	0.3	0.9
Stand density index	-	23.0	414.5	417.9	1365.7	29.0	610.1	631.8	2060.9

Note: SL is the slope angle, AZ is the azimuth of the aspect, Sin and Cos are the sine transform and cosine transform functions.

2.3. Model Development

There are many research equations on DBH growth, such as the study of DBH increment, DBH growth rate, squared DBH increase, and the natural logarithm of each increase [27]. In this study, the influence mechanism of the individual tree growth, reflected by the annual DBH growth rate, was explored, and the formula for calculating the annual growth rate of the individual tree is as follows:

$$p = \frac{d_2 - d_1}{n \cdot d_1} \bullet 100\% \tag{1}$$

where p is the annual growth rate of individual tree, d_1 is the pre-observed DBH, d_2 is the DBH after n years, and n is the interval between re-measurements, which was 5 years in this study.

The ML algorithm is characterized by its extraordinary performance, which is better than traditional regression methods in predicting outcomes within large data bases [28]. Although tree growth changes in forest ecosystems are related to the ecological, biological, and physiological properties of plants and various disturbances, the optimal model can perform associated topology with other resemblance stands on a local, regional, and global scale. We developed three types of ML algorithms to model our data: GBRT, KNN and RF. The factors that contributed to the growth of *C. lanceolata* and *Betula* spp. in China were disentangled by these ML algorithms. This research split our dataset into two groups randomly, namely the training sets (80%) for ML model development and the validation sets (20%) for performance evaluation. In the training process, tuning was considered for ML-based models to avoid overfitting, and the best hyper-parameter for ML models was a 10-fold cross-validation. The optimal parameters of the model were with the minimum *RMSE*, and all data were preprocessed by scaling and centering. Pre-processing transformation can be estimated from the training data and applied to any data set with the same variables. Tree growth is influenced by a combination of tree characteristics and external environmental factors, such as tree size, bioclimate, stand structure, topography, etc. The general expression of the model is as follows:

$$p = f(Size, Competition, Bioclimate, Topographical, Soil) \tag{2}$$

where $f()$ represents the GBRT, KNN, and RF model, and the explanatory factors affecting the tree growth mainly include the following aspects:

1. Individual tree size: DBH of the individual tree.
2. Competition factors: stand density index (SDI), stand origin (ORG), stand canopy density (SCD), and Simpson's diversity index (SIDI). ORG was divided into 2 categories: plantation and natural [2].
3. Bioclimatic factors: Geographical factors, including latitude, LAT (°), longitude, LONG (°), and elevation, ELEV (m). Meteorological factors, including average annual temperature, TEMP (°C) and annual precipitation, PREC (mm).
4. Topographical factors: Slope, (SL), aspect, (AZ), and slope position (TPI). There are nine slope directions in the aspect: flat land without slope direction, eight slope directions from north to south and east to west. TPI was categorized into six types based on the topography: ridge, uphill, mid-slope, downhill, valley, and flat ground.
5. Soil factors: Soil thickness, ST (cm), and humus thickness, HT (cm).

Partial variables of the competition and topographical variables were quantified by some preprocessing, such as SL, AZ, SDI, and SIDI, and the calculation criteria are showed in Table 2. The data type of ORG and TPI, the categorical variables, were transformed as the factor type.

Table 2. Calculation basis of environmental factors for tree growth.

Variable	Description	Calculation Criteria
SL	Sine of slope	SinSL
AZ	Cosine variation of aspect	(CosAZ + 1)/2
SDI	Stand density index	$N \cdot \left(\frac{25}{\mathrm{d}_g}\right)^{-1.605}$
SIDI	Simpson diversity index	$1 - \sum \mathrm{p}_i^2$

Note: where N is the number of trees (N/ha), dg is the average DBH of the plot (cm), P_i is the proportion of species i.

Apart from regression methods, machine learning models do not have a mathematical scale parameter, but their structure can be determined based on changes of the input node range. GBRT is an advanced statistical learning method for handling predictor variables of different types and distributions based on classification and regression trees [29]. For the GBRT model, the main parameters were the number of iterations (n.tree), the complexity of the tree (interaction.depth), the learning rate (shrinkage), and the minimum number of training samples (n.minobsinnode). KNN, a nonparametric regression, does not make any assumption on the distribution of data, thereby stimulating a training phase. For KNN, K observations in the proximity are taken into account and so are the average of the response of those K independent variables [30]. RF is based on the classification and regression tree (CART), which is essentially an improvement on the decision tree. For RF, ntree and mtry were the two tuning parameters. The ntree defined the total number of independent trees and decided how many trees will grow. The mtry was the number of predictive variables and determined the correlations between trees, where a decrease in mtry would result in a decrease in correlation between trees [31].

To interpret the ML algorithms, variable importance (VI) was used to offer some information about which attributes played major roles in the generation of the model [28]. The generic function varImp was used to characterize the general effect of predictors on the model in this research, which works with objects produced by train, but this is a simple wrapper for the specific models previously listed. VI values were scaled to percentages to provide a better comparison [32]. After determining which variables were most important, their effects on the output were examined to improve understanding of processes in the tree growth. This was performed using Partial Dependence Plots (PDPs), which can help to visualize the dependence of the dependent variable on the independent variable [33]. It should be noted that that the partial dependence of a dependent variable on an independent variable is calculated by considering the average effect of other variables on the dependent variable, rather than by ignoring the influence of other variables on the dependent variable. In this study, all statistical analyses were performed by using R software, version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). The R packages "caret," "e1071," "random-forest," "gbm," "DALEX," "lattice," "ggplot2", etc., were used for ML algorithms and visualization.

2.4. Model Validation

One of our main concerns was the actual prediction effect of the function model on future values. To validate the three ML algorithms, we examined the prediction accuracy of the GBRT, KNN, and RF models in estimating the annual growth rate of trees and their future DBH. Typically, accuracy validation using independent sample datasets is considered to be the best evaluation of a model [34]. To evaluate model performance, 80% of the dataset was randomly selected for constructing the machine learning models and was evaluated by coefficient of correlation (R^2), root mean square error (*RMSE*), and mean absolute error (*MAE*); meanwhile, the remaining 20% was also used to validate the models. Through the application of this method and ratio, many forestry models were developed. For example, Zeng et al. developed equations for individual tree crown biomass, and Zhang et al. developed a tree growth difference equation for individual tree DBH and their age [35,36].

In the evaluation of the data, the statistical indices are as follows: bias (*BIAS*), *RMSE*, *MAE*, R^2, Willmott's index of agreement (*WIA*), and total relative error (*TRE*) [37,38]. *BIAS* reflects the error between the output of the model on the sample and the true value, *MAE* explains the model stability, R^2 describes the quality of the model, and *TRE* and *RMSE* reflect the model in a straightforward manner [39,40]. Besides, *WIA* is intended to be a descriptive measure, which provides a measure of whether the external predictive of the model is statistically accurate, and it is both a relative and bounded measure that can be widely applied in order to make cross-comparisons between models [41,42]. Their equations are shown in the Table 3.

Table 3. Calculation equations for model validation.

Statistical Indices	Equation	Ideal
BIAS:	$BIAS = \sum_{i=1}^{n} \frac{y_i - \hat{y}_i}{n}$	0
Root mean square error (*RMSE*):	$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}$	0
Mean absolute error (*MAE*):	$MAE = \sum_{i=1}^{n} \left\lvert \frac{y_i - \hat{y}_i}{n} \right\rvert$	0
Coefficient of correlation (R^2):	$R^2 = 1 - \left[\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y}_i)^2} \right]$	1
Willmott's Index of Agreement (*WIA*):	$WIA = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (\lvert y_i - \overline{y}_i \rvert + \lvert \hat{y}_i - \overline{y}_i \rvert)^2}$	1
Total relative error (*TRE*):	$TRE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)}{\sum_{i=1}^{n} \hat{y}_i} \times 100\%$	0

Note: where y_i and $\hat{y}_i$ are the ith observation and prediction value, respectively, $\overline{y_i}$ is the mean of the observation values, and n is the number of samples.

3. Results

3.1. ML Algorithms

3.1.1. Model Construction

In this study, a 10-fold cross-validation and grid search were used to find the parameters, the *RMSE*, R^2, and the *MAE* were used to evaluate the model, and *RMSE* was used to select the optimal model using the smallest value. For the GBRT, the tuning parameter 'shrinkage' and 'n.minobsinnode' were held constant at a value of 0.1 and 10, respectively, and the final model was that the n.trees was 150 and the interaction.depth was 3. For the KNN, the final models were that the k were 7 and 9 for *Betula* spp. and *C. lanceolata*, respectively. In addition, for the RF, the tuning parameter 'ntree' was held constant at a value of 500 to search the mtry, and the final model was that the mtry was 2. The optimal model fitting results are shown in Table 4, the R^2 of the RF model was 0.560 and 0.526 for *Betula* spp. and *C. lanceolata*, respectively, and for both of the species groups, the accuracy of RF was better than KNN, which was better than GBRT.

Table 4. The optimal model fitting results of GBRT, KNN, and RF.

Model	***Betula* spp.**			***Cunninghamia lanceolata* (Lamb.) Hook.**		
	RMSE	R^2	***MAE***	***RMSE***	R^2	***MAE***
GBRT	1.865	0.345	1.196	3.626	0.327	2.594
KNN	1.599	0.519	0.946	3.144	0.491	2.114
RF	1.537	0.560	0.937	3.059	0.526	2.108

3.1.2. Model Evaluation

One significant target of the research on tree growth is to predict the future condition of the forest; thus, the prediction of the future condition of the trees is also a significant evaluation index of the models. In this study, the models were evaluated by the predicting results of annual growth rate and future DBH after 5 years of the 20% samples, with the indicators of *BIAS*, *RMSE*, *MAE*, R^2, *WIA*, and *TRE*. The results are shown in Table 5. The results showed that the validation accuracy were basically consistent with the model results, and the prediction accuracy of the future DBH was consistent with the prediction accuracy of annual growth rate. R^2 of the RF in predicting the annual growth rate for *Betula* spp. and *C. lanceolata* was higher than 0.51; the *WIA* of each ML algorithm in predicting the future DBH were all higher than 0.97, and the R^2 was higher than 0.98 and 0.90 for the two species, respectively, indicating that the ML algorithms can better achieve the future DBH prediction.

Table 5. Validation precision of GBRT, KNN, and RF for *Betula* spp. and *Cunninghamia lanceolata* (Lamb.) Hook.

Species	Verify	Annual Growth-Rata (%)			Future DBH (cm)		
		GBRT	**KNN**	**RF**	**GBRT**	**KNN**	**RF**
Betula spp.	*BIAS*	0.008	−0.004	−0.005	−0.01	−0.02	−0.04
	RMSE	1.888	1.582	1.528	0.90	0.74	0.75
	MAE	1.200	0.933	0.932	0.60	0.47	0.48
	R^2	0.345	0.540	0.571	0.981	0.987	0.987
	WIA	0.675	0.840	0.832	0.995	0.997	0.997
	TRE	0.398	−0.186	−0.253	−0.070	−0.137	−0.324
Cunninghamia lanceolata (Lamb.) Hook.	*BIAS*	−0.049	0.001	−0.028	−0.04	−0.03	−0.08
	RMSE	3.612	3.127	3.048	1.53	1.31	1.30
	MAE	2.597	2.097	2.091	1.18	0.95	0.96
	R^2	0.319	0.489	0.515	0.900	0.926	0.928
	WIA	0.655	0.815	0.798	0.974	0.981	0.982
	TRE	−1.049	0.013	−0.607	−0.312	−0.222	−0.588

Note: The prediction precision of the future DBH was the DBH prediction accuracy after 5 years.

3.2. Disentangling the Factors That Contribute to Tree Growth

3.2.1. Variation Importance Analysis

The orders of the relative importance of variables in each annual growth rate predicting ML algorithm for *Betula* spp. and *C. lanceolata* were different, and the results are shown in Figure 2. Although the general trends were slightly different among those ML algorithms, DBH, TEMP, and PREC ranked as the top three for *Betula* spp., and DBH, SDI, and SIDI ranked as the top three for *C. lanceolata*, except for KNN model. On the other hand, variables such as slope position, aspect, origin, and humus thickness made little contribution to the annual growth rate prediction. Furthermore, RF, as the highest accuracy of the three ML algorithms, showed that the single-tree size, stand structure, and climate would be more important for the tree growth and that the topographic factors would have less contribution.

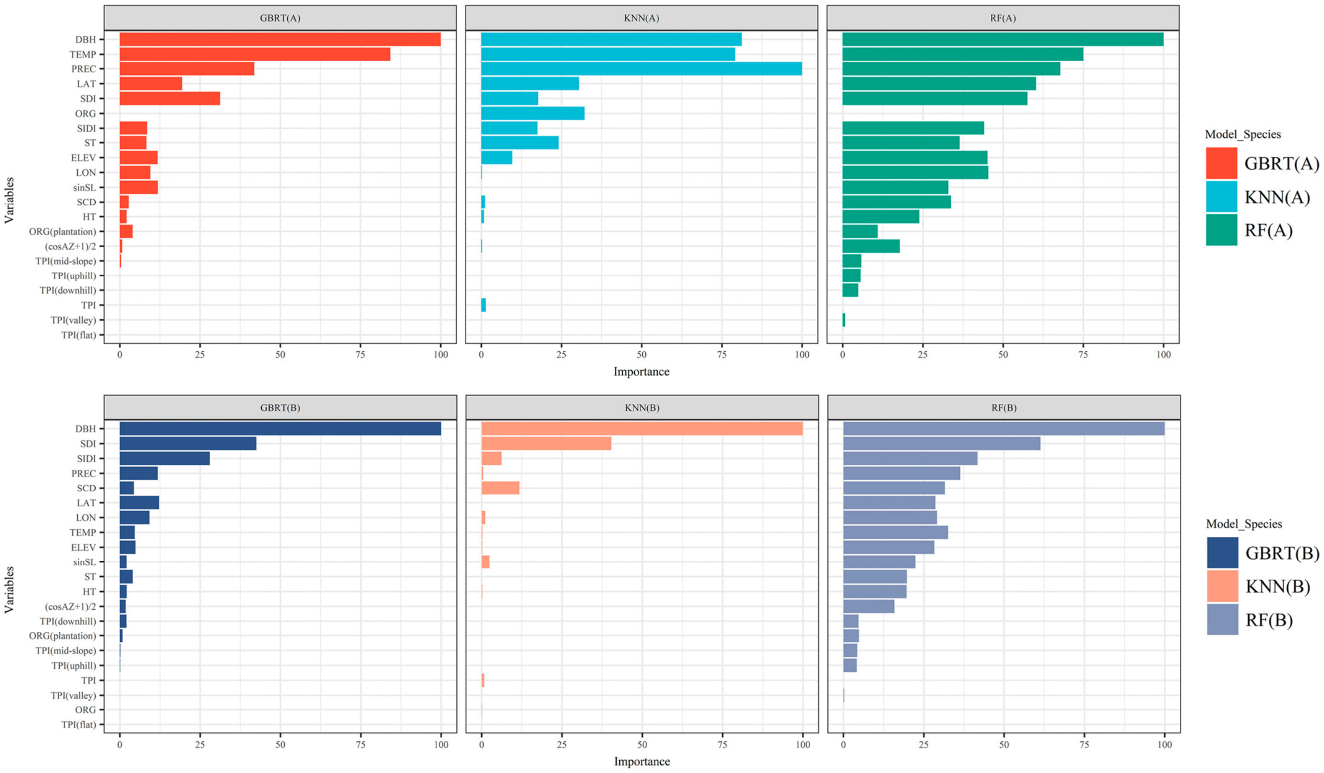

Figure 2. Variation Importance Analysis Results for *Betula* spp. and *Cunninghamia lanceolata* (Lamb.) Hook. (A) and (B) means *Betula* spp. and *Cunninghamia lanceolata* (Lamb.) Hook., respectively.

3.2.2. Partial Dependency Analysis

The importance of influencing factors reflects the relative importance of the influence of each variable on the growth rate and the main factors affecting the growth rate, but it could not quantitatively analyze the influence of the changes of the influencing factors on the growth rate change trend. The partial dependence can measure the marginal impact of one or multiple factors on the output of the ML algorithms, so as to quantitatively analyze the impact of each factor on the growth rate. In this study, PDP for *Betula* spp. and *C. lanceolata*, the partial dependencies of the relationship between each explanatory variable and the response variable of the ML algorithms are shown in Figures 3 and 4.

The PDP for *Betula* spp. is shown in Figure 3. The annual growth rate decreased with the increase of the DBH and showed an obvious inverse "J" curve. The temperature and precipitation were both positively correlated with the growth, and the areas with higher temperatures and more precipitation were more suitable for their growth. Areas in the lower latitude, the higher longitude, and lower altitudes would be more effective for their growth. Considering the topography, we found that the north aspect was more suitable for the *Betula* spp. growth than the south aspect, and the higher slope would achieve a greater restriction on growth, resulting in a lower growth rate. The stand, with lower SDI, lower SCD, and higher SIDI, would achieve a relatively higher growth rate. Furthermore, the annual growth rate of the *Betula* spp. in the plantation stands was higher than that in the natural stand.

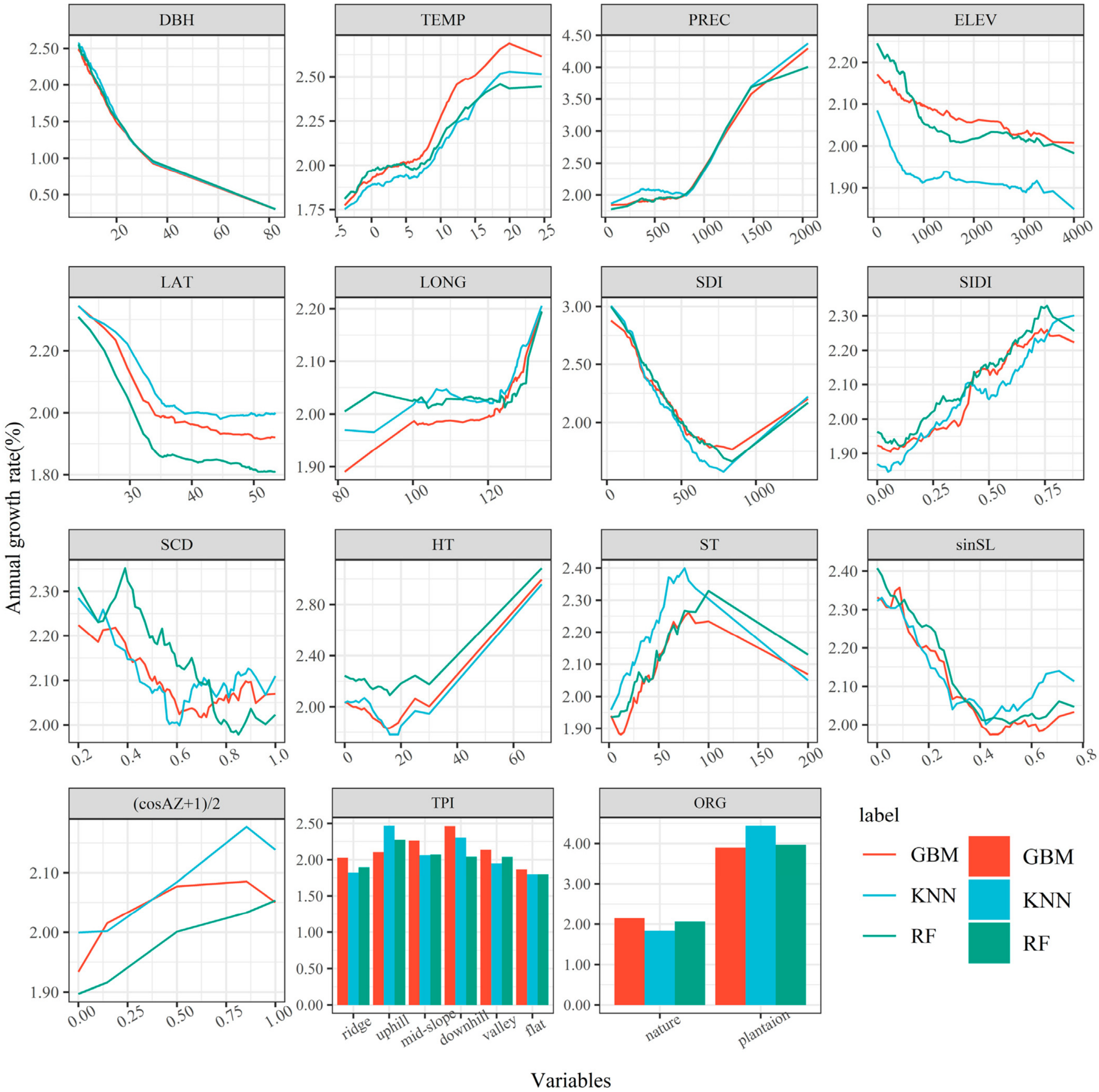

Figure 3. Partial dependence profile of the GBRT, KNN, and RF of *Betula* spp.

The PDP for *C. lanceolata* is shown in Figure 4. The annual growth rate also decreased with the increase of the DBH and showed an obvious inverse "J" curve. In the geographical space location, the region of the higher latitude and the greater longitude was more unfavorable for the *C. lanceolata* growth. Besides, *C. lanceolata* mainly grows in areas where the elevation is less than 2250 m and the annual average temperature is higher than 10 °C; the higher altitude and higher temperature are more conducive to their growth. However, it showed a trend of first increasing and then decreasing with the precipitation on the growth rate, indicating that precipitation is beneficial to the growth of *C. lanceolata* within a certain range, but when it exceeds a certain threshold, it inhibits the growth of them. The same happened with *Betula* spp.; it also showed that the smaller slope would be more favorable for the growth of *C. lanceolata*. Among all the slope positions, the flat land was the most favorable for their growth. When it comes to the stand structure, we found SDI, SCD, and SIDI had negative impacts on the tree growth rate, and the annual growth rate would decrease with the increase of them. In addition, soil thickness and humus thickness had positive effects on the annual growth rate of *C. lanceolata*, the trees that with thicker soil and humus will achieve a higher growth rate.

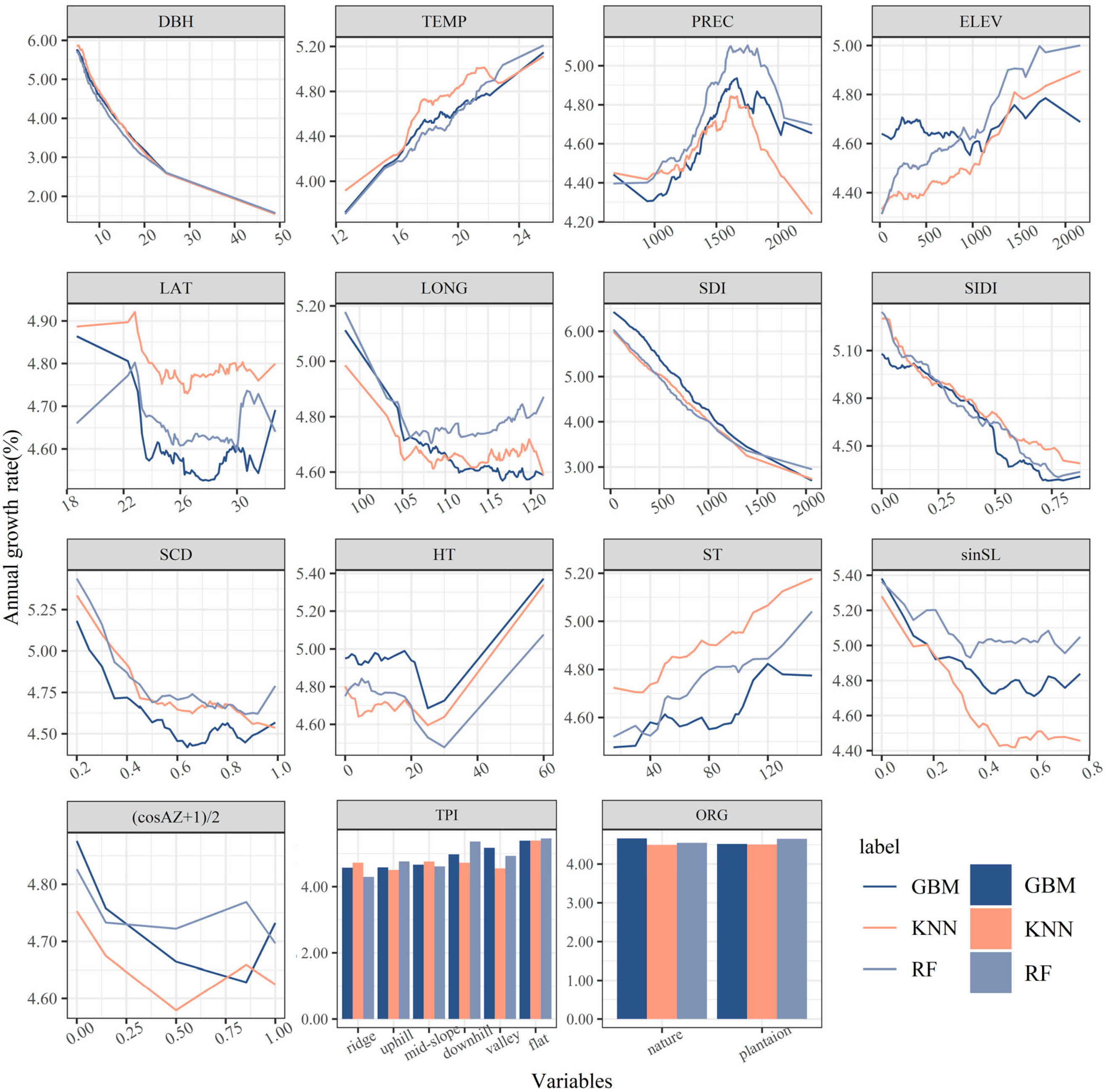

Figure 4. Partial dependence profile of GBRT, KNN, and RF of *Cunninghamia lanceolata* (Lamb.) Hook.

4. Discussion

Tree growth is one of the references for research on forest quantity, quality, and sustainable development [43]. In this study, annual growth rate models, the ML algorithms for the relationship between the DBH growth and explanatory variables, were developed for *Betula* spp. and *C. lanceolata* in China. Generally, there was some commonality results between the three modeling procedures. We found that DBH was the most important variable for the annual growth rate, which decreased with the increase of DBH and conformed to the inverse 'J' curve distribution [44]. Individuals with the same diameter but with very different growth rates were frequently observed due to the competition and the forest succession, which would affect the establishment and development of trees [7,45].

Model validation was considered an important part of model evaluation and could demonstrate the dependability of the model [46]. As a basis for forest management, many scholars have developed many models to predict individual tree growth. Lhotka et al. developed an individual-tree model based on a mixed-effects regression and presented R^2 values from 0.26 to 0.57 [47]. Moreno et al. found that the R^2 of annual DBH growth predictions of the AIDBH model was 0.56 and DBH prediction after six and twelve years were more than 0.97 [37]. Giovanni Correia Vieira et al. found that the ANNs and ANFIS, the ML algorithms, had higher accuracies than regression models for the prognosis of growth of DBH for individual trees [48]. The similar species of the genera *Betula* spp. or *C. lanceolata* were aggregated, respectively, which would obtain more modeling data and improve the model applicability [49,50]. The R^2 of RF, the best performing model, in predicting the annual growth rate was higher than 0.51. The *WIA* of each ML algorithm in predicting the future DBH were all higher than 0.97, the R^2 were higher than 0.90, and the values of the other statistical indices were relatively small; these indicated that the ML algorithms can better achieve the future DBH prediction.

Excavating the influence mechanism of external environment on tree growth is critical in improving forest cover and forest management. Our research showed that the climate variables, average annual temperature and precipitation, are two of the top five environmental factors that have the most impact on tree growth. The growth rate of *Betula* spp. increased with increasing temperature and precipitation and decreased with increasing elevation. Temperature and elevation were positively correlated with tree growth, whereas the effect of precipitation on *C. lanceolata* was first positively and then negatively affected. These were also in accordance with the characteristics of tree growth and other research findings [51]. For instance, Way et al. found that temperature is generally positively correlated with tree growth, except for tropical species, and the growth of broad-leaved species increased with increasing temperature [52]. The research of GÓMEZ-APARICIO et al. showed that the higher temperature would be more favorable for the growth of broad-leaved species [53]. Hart et al. found that temperature was more important to the growth of the two tree species than precipitation, and current and previous growing season temperatures were the key to radial growth [54]. Besides, Toledo et al. also concluded that higher temperatures and more precipitation would contribute to a higher growth rate [55].

The tree growth, affected by topography, climate, and soil condition, and the three factors interacted with each other [56]. According to topographic heterogeneity, there are great differences in ecological and resource gradients across different geomorphic positions, manifested by flat areas owing to higher soil nutrient content than the steeper sites [57]. Our analysis revealed that the tree growth tended to increase under elevated humus and soil depth, which has been frequently attributed to thicker humus and soil providing sufficient nutrients [58]. Additionally, consistent with previous studies, our work also found that slope showed a significant negative correlation with tree growth [44]. The lower resource availability in areas with higher slopes may be due to steep slopes being prone to soil erosion [59]. Besides, light, temperature, wind speed, etc., are closely related to the slope aspect, which significantly affected forest growth. North-facing slopes were more favorable for the growth of *Betula* spp. than south-facing slopes for less light time on the north-facing slopes, which would have small soil evaporation and good water and fertilizer conditions to obtain more conducive site conditions for the *Betula* spp. growth. However, we found that the effect of the slope aspect on the growth of *C. lanceolate* is different from that of *Betula* spp.; their growth on the southeast slope is better than that on the northwest slope due to the relatively longer sunlight on the southeast slope than the northwest slope, which was helpful for photosynthesis and tree growth.

Tree growth is influenced by competitive interactions, especially stand structure such as stand density, stand canopy density, species composition, etc. [60]. Especially for *C.*

lanceolata, which are mainly distributed in southern China where the average annual temperature and precipitation are relatively high (Table 1, Figure 2), the importance of SDI and SIDI is even higher than climate on their growth. Generally, high SDI and SCD would inhibit tree growth, and mixed forests are more conducive to tree growth than pure forests. However, the growth rate of *C. lanceolata* decrease with the increase of the SIDI, which may be caused by many factors, i.e., *C. lanceolata* in the stands with high species diversity was not the dominant tree species group, and the stands with low SIDI are mostly plantations, where human intervention has boosted their growth. This was also consistent with the previous studies, which showed that the total ecosystem production in monocultures of the faster-growing species groups grow more than mixed-species forests [61].

For the sustainable development of forest quantity and quality, forest management should take into account a variety of factors such as climate, topography, and stand structure. High stand density inhibits the growth of trees, but low stand density is detrimental to the strategic needs of the national ecological sustainable development. In addition, the NFI mainly records the coding of tree species or tree species groups, which may cause some errors in the models. In future research, the study of rational stand structure and the influence of finer interspecific differences on tree growth will be some of the key issues.

5. Conclusions

Based on the partial NFI data of *C. lanceolata* and *Betula* spp., this study developed three ML algorithms and verified that the growth rate of trees was mainly affected by their single-tree size, and the external environment promoted or inhibited their growth. The accuracy of RF was better than KNN, which was better than GBRT, and all the models had high *WIA* and R^2 in predicting future DBH. Among all environment variables, climatic and competing variables related to spatial location and stand structure were more important to tree growth than topography. Temperature had a positive effect on the growth of *Betula* spp. and *C. lanceolata*, and with the increase of precipitation, the growth rate of *Betula* spp. showed an increasing trend; that of *C. lanceolata* first increased and then decreased, showing a "parabolic" trend. Furthermore, higher stand density and canopy closure would be more unfavorable for their growth. In forest resource management and future afforestation planning, the factors that should be considered are as follows: firstly, the geographical location, meteorology, and climate; secondly, the stand structure, especially the rationality of stand density; finally, the topographic structure, including slope, aspect, and slope position. In addition, the machine learning algorithms for disentangling the factors that contribute to the growth of trees provided a quantitative basis for forest resource management and provided theoretical support for future research on sustainable forest management and the prediction of forest carbon sink and forest carbon cycles.

Author Contributions: Conceptualization, H.Z. and Z.F.; methodology, H.Z.; validation, S.W. and W.J.; formal analysis, H.Z. and S.W.; data curation, H.Z. and W.J.; writing—original draft preparation, H.Z.; writing—review and editing, H.Z., S.W. and W.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key R&D Projects in Hainan Province (ZDYF2021SHFZ256); Natural Science Foundation of Hainan University, grant numbers KYQD(ZR)21115; and Science and Technology Project of Haikou City, China (2020-057).

Acknowledgments: The authors would like to acknowledge the forest resource data support drawn from the Academy of Forest Inventory and Planning, National Forestry and Grassland Administration. The authors would also like to acknowledge the Geospatial Data Cloud site and National Meteorological Administration site for basic data support.

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

References

1. Zhu, Y.; Feng, Z.; Lu, J.; Liu, J. Estimation of Forest Biomass in Beijing (China) Using Multisource Remote Sensing and Forest Inventory Data. *Forests* **2020**, *11*, 163. [CrossRef]
2. Laubhann, D.; Sterba, H.; Reinds, G.J.; De Vries, W. The impact of atmospheric deposition and climate on forest growth in European monitoring plots: An individual tree growth model. *For. Ecol. Manag.* **2009**, *258*, 1751–1761. [CrossRef]
3. Zeng, W.; Zhang, L.; Chen, X.; Cheng, Z.; Ma, K.; Li, Z. Construction of compatible and additive individual-tree biomass models for Pinustabulaeformis in China. *Can. J. For. Res.* **2017**, *47*, 467–475. [CrossRef]
4. Dobbertin, M. Tree growth as indicator of tree vitality and of tree reaction to environmental stress: A review. *Eur. J. For. Res.* **2005**, *124*, 319–333. [CrossRef]
5. Woollons, R.C.; Snowdon, P.; Mitchell, N.D. Augmenting empirical stand projection equations with edaphic and climatic variables. *For. Ecol. Manag.* **1997**, *98*, 267–275. [CrossRef]
6. Ashraf, M.I.; Zhao, Z.; Bourque, C.P.A.; MacLean, D.A.; Meng, F. Integrating biophysical controls in forest growth and yield predictions with artificial intelligence technology. *Can. J. For. Res.* **2013**, *43*, 1162–1171. [CrossRef]
7. Reis, L.P.; de Souza, A.L.; Mazzei, L.; Dos Reis, P.C.M.; Leite, H.G.; Soares, C.P.B.; Torres, C.M.M.E.; Da Silva, L.F.; Ruschel, A.R. Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks. *For. Ecol. Manag* **2016**, *382*, 161–167. [CrossRef]
8. Condés, S.; Sterba, H. Comparing an individual tree growth model for *Pinus halepensis* Mill. in the Spanish region of Murcia with yield tables gained from the same area. *Eur. J. For. Res.* **2008**, *127*, 253–261. [CrossRef]
9. Choi, J.; An, H. A Forest Growth Model for the Natural Broadleaved Forests in Northeastern Korea. *Forests* **2016**, *7*, 288. [CrossRef]
10. Pdp, A.; Cpda, B.; Bd, C.; Ist, D.; Jnms, E.; Prvg, F. An individual-based spatially explicit simulation model for strategic forest management planning in the eastern Amazon. *Ecol. Model.* **2004**, *173*, 335–354.
11. Valle, D.; Phillips, P.; Vidal, E.; Schulze, M.; Grogan, J.; Sales, M.; van Gardingen, P. Adaptation of a spatially explicit individual tree-based growth and yield model and long-term comparison between reduced-impact and conventional logging in eastern Amazonia, Brazil. *For. Ecol. Manag.* **2007**, *243*, 187–198. [CrossRef]
12. Zhao, X.; Feng, Z.; Zhou, Y.; Lin, Y. Key Technologies of Forest Resource Examination System Development in China. *Engineering* **2020**, *6*, 491–494. [CrossRef]
13. Lu, J.; Feng, Z.; Zhu, Y. Estimation of Forest Biomass and Carbon Storage in China Based on Forest Resources Inventory Data. *Forests* **2019**, *10*, 650. [CrossRef]
14. Thurnher, C.; Klopf, M.; Hasenauer, H. MOSES—A tree growth simulator for modelling stand response in Central Europe. *Ecol. Model.* **2017**, *352*, 58–76. [CrossRef]
15. Qiu, Z.X.; Feng, Z.K.; Song, Y.N.; Li, M.L.; Zhang, P.P. Carbon sequestration potential of forest vegetation in China from 2003 to 2050: Predicting forest vegetation growth based on climate and the environment. *J. Clean Prod.* **2020**, *252*, 119715. [CrossRef]
16. Moisen, G.G.; Freeman, E.A.; Blackard, J.A.; Frescino, T.S.; Zimmermann, N.E.; Edwards, T.C. Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecol. Model.* **2006**, *199*, 176–187. [CrossRef]
17. Martin, M.P.; Wattenbach, M.; Smith, P.; Meersmans, J.; Jolivet, C.; Boulonne, L.; Arrouays, D. Spatial distribution of soil organic carbon stocks in France. *Biogeosciences* **2011**, *8*, 1053–1065. [CrossRef]
18. Zhang, Y.; Chen, H.Y.H.; Reich, P.B. Forest productivity increases with evenness, species richness and trait variation: A global meta-analysis. *J. Ecol.* **2012**, *100*, 742–749. [CrossRef]
19. Ren, Y.; Chen, S.; Wei, X.; Xi, W.; Luo, Y.; Song, X.; Zuo, S.; Yang, Y. Disentangling the factors that contribute to variation in forest biomass increments in the mid-subtropical forests of China. *J. For. Res.* **2016**, *27*, 919–930. [CrossRef]
20. Beaudoin, A.; Bernier, P.Y.; Guindon, L.; Villemaire, P.; Guo, X.J.; Stinson, G.; Bergeron, T.; Magnussen, S.; Hall, R.J. Mapping attributes of Canada's forests at moderate resolution through KNN and MODIS imagery. *Can. J. For. Res.* **2014**, *44*, 521–532. [CrossRef]
21. Kilham, P.; Hartebrodt, C.; Kändler, G. Generating Tree-Level Harvest Predictions from Forest Inventories with Random Forests. *Forests* **2019**, *10*, 20. [CrossRef]
22. Sharkovsky, A.N.; Maistrenko, Y.L.; Romanenko, E.Y. *Difference Equations and Their Applications*; Ser. Mathimatics and Its Applications; Kluwer Academic: Dordrecht, The Netherlands, 1993.
23. Fang, J.; Shen, Z.; Tang, Z.; Wang, X.; Wang, Z.; Feng, J.; Liu, Y.; Qiao, X.; Wu, X.; ZhengJ, C. Forest community survey and the structural characteristics of forests in China. *Ecography* **2012**, *35*, 1059–1071. [CrossRef]
24. State Forestry and Grassland Administration (SFGA). *Forest Resources Report of China: The 9th National Forest Resource Inventory*; China Forestry Publishing House: Beijing, China, 2019. (In Chinese)
25. Zeng, W.; Tomppo, E.; Healey, S.P.; Gadow, K.V. The national forest inventory in China: History—Results—International context. *Ecosyst* **2015**, *2*, 23. [CrossRef]
26. State Forestry and Grassland Administration (SFGA). *Technical Specifications on National Continuous Forest Inventory*; State Forestry and Grassland Administration (SFGA): Beijing, China, 2014. (In Chinese)
27. Adame, P.; Hynynen, J.; Ca Ellas, I.; Río, M.D. Individual-tree diameter growth model for rebollo oak (*Quercus pyrenaica* Willd.) coppices. *For. Ecol. Manag.* **2008**, *255*, 1011–1022. [CrossRef]

28. Zhu, J.; Zheng, J.; Li, L.; Huang, R.; Ren, H.; Wang, D.; Dai, Z.; Su, X. Application of Machine Learning Algorithms to Predict Central Lymph Node Metastasis in T1-T2, Non-invasive, and Clinically Node Negative Papillary Thyroid Carcinoma. *Front. Med.* **2021**, *8*, 635771. [CrossRef]
29. Elith, J.; Hastie, J. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [CrossRef]
30. Goyal, R.; Chandra, P.; Singh, Y. Suitability of KNN Regression in the Development of Interaction based Software Fault Prediction Models. *IERI Procedia* **2014**, *6*, 15–21. [CrossRef]
31. Ou, Q.; Lei, X.; Shen, C. Individual Tree Diameter Growth Models of Larch–Spruce–Fir Mixed Forests Based on Machine Learning Algorithms. *Forests* **2019**, *10*, 187. [CrossRef]
32. Wang, D.; Thunéll, S.; Lindberg, U.; Jiang, L.; Trygg, J.; Tysklind, M.; Souihi, N. A machine learning framework to improve effluent quality control in wastewater treatment plants. *Sci. Total Environ.* **2021**, *784*, 147138. [CrossRef]
33. Friedman, J. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
34. Huang, S.; Yang, Y.; Wang, Y. A Critical Look at Procedures for Validating Growth and Yield Models. In *Modelling Forest Systems*; Amaro, A., Reed, D., Soares, P., Eds.; CAB International: Wallingford, UK, 2003.
35. Zeng, W.; Duo, H.; Lei, X.; Chen, X.; Wang, X.; Pu, Y.; Zou, W. Individual tree biomass equations and growth models sensitive to climate variables for Larix spp. in China. *Eur. J. For. Res.* **2017**, *136*, 233–249. [CrossRef]
36. Zhang, H.; Feng, Z.; Chen, P.; Chen, X. Development of a Tree Growth Difference Equation and Its Application in Forecasting the Biomass Carbon Stocks of Chinese Forests in 2050. *Forests* **2019**, *10*, 582. [CrossRef]
37. Moreno, P.; Palmas, S.; Escobedo, F.; Cropper, W.; Gezan, S. Individual-Tree Diameter Growth Models for Mixed Nothofagus Second Growth Forests in Southern Chile. *Forests* **2017**, *8*, 506. [CrossRef]
38. Han, Y.; Wu, B.; Wang, K.; Guo, E.; Dong, C.; Wang, Z. Individual-tree form growth models of visualization simulation for managed *Larix principis-rupprechtii* plantation. *Comput. Electron. Agric.* **2016**, *123*, 341–350. [CrossRef]
39. Saud, P.; Lynch, T.B.; Anup, K.C.; Guldin, J.M. Using quadratic mean diameter and relative spacing index to enhance height–diameter and crown ratio models fitted to longitudinal data. *Forestry* **2016**, *89*, 215–229. [CrossRef]
40. Scolforo, H.F.; Scolforo, J.R.S.; Thiersch, C.R.; Thiersch, M.F.; McTague, J.P.; Burkhart, H.; Ferraz Filho, A.C.; de Mello, J.M.; Roise, J. A new model of tropical tree diameter growth rate and its application to identify fast-growing native tree species. *For. Ecol. Manag.* **2017**, *400*, 578–586. [CrossRef]
41. Legates, D.R.; Mccabe, G.J. Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **1999**, *35*, 233–241. [CrossRef]
42. Willmott, C.J. Some Comments on the Evaluation of Model Performance. *Bull. Am. Meteorol. Soc.* **1982**, *63*, 1309–1369. [CrossRef]
43. Webster, C.R.; Lorimer, C.G. Minimum Opening Sizes for Canopy Recruitment of Midtolerant Tree Species: A Retrospective Approach. *Ecol. Appl.* **2005**, *15*, 1245–1262. [CrossRef]
44. Zhang, H.; Feng, Z.; Shen, C.; Li, Y.; Feng, Z.; Zeng, W.; Huang, G. Relationship between the geographical environment and the forest carbon sink capacity in China based on an individual-tree growth-rate model. *Ecol. Indic.* **2022**, *138*, 108814. [CrossRef]
45. Farrior, C.E.; Bohlman, S.A.; Hubbell, S.; Pacala, S.W. Dominance of the suppressed: Power-law size structure in tropical forests. *Science* **2016**, *351*, 155. [CrossRef]
46. Kozak, A.; Kozak, R. Does cross validation provide additional information in the evaluation of regression models? *Can. J. For. Res.* **2003**, *33*, 976–987. [CrossRef]
47. Lhotka, J.M.; Loewenstein, E.F. An individual-tree diameter growth model for managed uneven-aged oak-shortleaf pine stands in the Ozark Highlands of Missouri, USA. *For. Ecol. Manag* **2011**, *261*, 770–778. [CrossRef]
48. Vieira, G.C.; de Mendonça, A.R.; Da Silva, G.F.; Zanetti, S.S.; Da Silva, M.M.; Dos Santos, A.R. Prognoses of diameter and height of trees of eucalyptus using artificial intelligence. *Sci. Total Environ.* **2018**, *619–620*, 1473–1481. [CrossRef] [PubMed]
49. Mehtatalo, L.; De-Miguel, S.; Gregoire, T.G. Modeling height-diameter curves for prediction. *Can. J. For. Res.* **2015**, *45*, 826–837. [CrossRef]
50. Gollob, C.; Ritter, T.; Vospernik, S.; Wassermann, C.; Nothdurft, A. A Flexible Height–Diameter Model for Tree Height Imputation on Forest Inventory Sample Plots Using Repeated Measures from the Past. *Forests* **2018**, *9*, 368. [CrossRef]
51. Troll, C. The upper timberlines in different climatic zones. *Arct. Alp. Res.* **1973**, *5*, A3–A18.
52. Way, D.A.; Oren, R. Differential responses to changes in growth temperature between trees from different functional groups and biomes: A review and synthesis of data. *Tree Physiol.* **2010**, *30*, 669–688. [CrossRef]
53. Gómez-Aparicio, L.; García-Valdés, R.; Ruíz-Benito, P.; Zavala, M.A. Disentangling the relative importance of climate, size and competition on tree growth in Iberian forests: Implications for forest management under global change. *Glob. Chang. Biol.* **2011**, *17*, 2400–2414. [CrossRef]
54. Hart, S.J.; Laroque, C.P. Searching for thresholds in climate–radial growth relationships of Engelmann spruce and subalpine fir, Jasper National Park, Alberta, Canada. *Dendrochronologia* **2013**, *31*, 9–15. [CrossRef]
55. Toledo, M.; Poorter, L.; Peña-Claros, M.; Alarcón, A.; Balcázar, J.; Leaño, C.; Licona, J.C.; Llanque, O.; Vroomans, V.; Zuidema, P.; et al. Climate is a stronger driver of tree and forest growth rates than soil and disturbance. *J. Ecol.* **2011**, *99*, 254–264. [CrossRef]
56. Adams, H.R.; Barnard, H.R.; Loomis, A.K.; University of Colorado. Topography alters tree growth-climate relationships in a semi-arid forested catchment. *Ecosphere* **2014**, *5*, t116–t148. [CrossRef]

57. Scholten, T.; Goebes, P.; Kühn, P.; Seitz, S.; Assmann, T.; Bauhus, J.; Bruelheide, H.; Buscot, F.; Erfmeier, A.; Fischer, M.; et al. On the combined effect of soil fertility and topography on tree growth in subtropical forest ecosystems—A study from SE China. *J. Plant Ecol.* **2017**, *10*, 111–127. [CrossRef]
58. Myburgh, P.A.; Van Zyl, J.L.; Conradie, W.J. Effect of Soil Depth on Growth and Water Consumption of Young *Vitis vinifera* L. cv *Pinot noir*. *S. Afr. J. Enol. Vitic.* **2017**, *17*, 53–62. [CrossRef]
59. Liu, J.; Yunhong, T.; Slik, J.W.F. Topography related habitat associations of tree species traits, composition and diversity in a Chinese tropical forest. *For. Ecol. Manag* **2014**, *330*, 75–81. [CrossRef]
60. Coomes, D.A.; Allen, R.B. Effects of size, competition and altitude on tree growth. *J. Ecol.* **2007**, *95*, 1084–1097. [CrossRef]
61. Binkley, D.; Stape, J.L.; Ryan, M.G.; Barnard, H.R.; Fownes, J. Age-related Decline in Forest Ecosystem Growth: An Individual-Tree, Stand-Structure Hypothesis. *Ecosystems* **2002**, *5*, 58–67. [CrossRef]