

Article

Individual Tree Diameter Growth Models of Larch–Spruce–Fir Mixed Forests Based on Machine Learning Algorithms

Qiangxin Ou ^{1,2}, Xiangdong Lei ^{1,2,*}  and Chenchen Shen ³

¹ Institute of Forest Resources Information Techniques, Chinese Academy of Forestry, Beijing 100091, China; jonsin_ou@outlook.com

² Key Laboratory of Forest Management and Growth Modelling, State Forestry and Grassland Administration, Beijing 100091, China

³ Department of Forest, Rangeland, and Fire Sciences, University of Idaho, Moscow, ID 83844, USA; cshen@uidaho.edu

* Correspondence: xdlei@ifrit.ac.cn; Tel.: +86-10-6288-9178

Received: 20 December 2018; Accepted: 18 February 2019; Published: 20 February 2019



Abstract: Individual tree growth models are flexible and commonly used to represent growth dynamics for heterogeneous and structurally complex uneven-aged stands. Besides traditional statistical models, the rapid development of nonparametric and nonlinear machine learning methods, such as random forest (RF), boosted regression tree (BRT), cubist (Cubist) and multivariate adaptive regression splines (MARS), provides a new way for predicting individual tree growth. However, the application of these approaches to individual tree growth modelling is still limited and short of a comparison of their performance. The objectives of this study were to compare and evaluate the performance of the RF, BRT, Cubist and MARS models for modelling the individual tree diameter growth based on tree size, competition, site condition and climate factors for larch–spruce–fir mixed forests in northeast China. Totally, 16,619 observations from long-term sample plots were used. Based on tenfold cross-validation, we found that the RF, BRT and Cubist models had a distinct advantage over the MARS model in predicting individual tree diameter growth. The Cubist model ranked the highest in terms of model performance (RMSE_{cv} [0.1351 cm], MAE_{cv} [0.0972 cm] and R²_{cv} [0.5734]), followed by BRT and RF models, whereas the MARS ranked the lowest (RMSE_{cv} [0.1462 cm], MAE_{cv} [0.1086 cm] and R²_{cv} [0.4993]). Relative importance of predictors determined from the RF and BRT models demonstrated that the competition and tree size were the main drivers to diameter growth, and climate had limited capacity in explaining the variation in tree diameter growth at local scale. In general, the RF, BRT and Cubist models are effective and powerful modelling methods for predicting the individual tree diameter growth.

Keywords: random forest; boosted regression tree; cubist; multivariate adaptive regression splines; tree diameter growth

1. Introduction

Forest growth models are important tools in providing quantitative and reliable information for forest management decisions making [1,2]. These models can be developed with different resolutions including stand, diameter class and individual tree levels [3]. Both stand and diameter class growth models are unable to represent the growth dynamics of individual trees and are only applicable to homogeneous and even-aged stands [2]. On the other hand, individual tree growth models are flexible and able to represent the growth dynamics of individual trees, which are commonly used to represent growth dynamics for the heterogeneous and structurally complex mixed uneven-aged stands [2,4,5].

Tree growth is often expressed as a function of tree size, competition and site condition [5–9]. Besides these aforementioned factors, climate factors such as temperature, light and precipitation are not negligible in determining tree growth [10]. Climate-sensitive tree growth models have also been developed recently [11–13]. The climate, site condition and competition interactively influence individual tree growth. Meanwhile, these influences may vary with tree species and size. However, so far information about contributions of different factors to individual tree growth in mixed forests is still insufficient, and it is necessary to identify the dominant factors affecting tree growth with appropriate methods [14].

Traditionally, individual tree diameter growth models related tree growth to stand- and tree-level variables using statistical models such as linear and nonlinear regression models with or without random effects [15–18]. These statistical models are often applied under certain statistical assumptions, such as the data are normally distributed, homoscedastic and independent. However, due to continuous observations, hierarchy and intrinsic variability of forest growth data, the above assumptions are usually difficult to satisfy [4]. Moreover, the selection of specific mathematical equation (model form) and convergence is rather difficult when modelling individual tree growth in nonlinear form with multiple variables. Therefore, data-driven methods were also applied and showed their potential such as generalized additive model (GAM) and artificial neural networks (ANN) [4,19,20].

In contrast to traditional statistical models, machine learning (ML) has the ability to model complex and nonlinear interactive relationship between the predictor variables and the response variable without statistical assumptions and predetermined mathematical equations [21]. Many state-of-the-art ML algorithms have been developed, including classification and regression trees (CART), random forest (RF), boosted regression tree (BRT), ANN, support vector machine (SVM), cubist (Cubist) and multivariate adaptive regression splines (MARS). These data-driven methods have already been successfully applied to ecology and remote sensing to perform tasks such as soil organic carbon mapping [22], carbon and energy fluxes [23,24], and species distribution [25,26]. However, the applications in forest growth and yield prediction are still limited. Few applications of these methods have been found in individual tree growth prediction [4,19]. With the increasing number of the ML algorithms, the comparison among these methods becomes more important. A direct comparison of ML algorithms has not yet been attempted in individual tree growth predicting. Unlike ANN and SVM which are often adversely affected by noninformative predictors and high correlation among the predictors, RF, BRT, Cubist and MARS models are naturally resistant to noninformative predictors due to the built-in feature selection mechanisms [27,28]. Highly correlated predictors do not drastically affect model performance in RF, BRT, Cubist and MARS [27]. In addition, unlike the ANN and SVM models, the RF, BRT, Cubist and MARS models do not need to pre-process predictors (such as centralization and standardization), and can well handle the case of many qualitative predictive variables or qualitative predictive variables with many levels. It is noted that both RF and BRT have the ability to provide quantitative and reliable measurements of relative importance of variables. Thus, the information about contributions of different factors to individual tree growth in mixed forests can be determined using both BRT and RF.

Therefore, the goal of this study is to use machine learning algorithms to predict the individual tree diameter growth with long-term plot observations of semi-natural mixed larch–spruce–fir–broadleaved forests in northeast China. The specific objectives were: (1) to compare and evaluate the performance of the RF, BRT, Cubist and MARS models for tree diameter growth modelling based on tree size, competition, site condition and climate factors; and (2) to identify the contributions of different factors to the individual tree diameter growth.

2. Data and Methods

2.1. Data

2.1.1. Study Area

This study was conducted in long-term permanent plots of larch–spruce fir mixed forests located in the Wangqing Forestry Bureau, Jilin Province, northeast China (43°05′–43°49′ N, 126°05′–126°05′ E). The study area belongs to middle or low hilly area of the Changbai Mountains with elevation ranging between 360–1477 m above sea level. The mean annual precipitation and temperature were 547 mm and 3.9 °C, respectively.

2.1.2. Permanent Plot Data

Plot data used to develop individual tree diameter growth models were obtained from 20 remeasured permanent plots of semi-natural conifer-broadleaf mixed forests. The area of plots ranges in size from 0.075 to 0.25 hm⁻². All plots, which originated from pure larch (*Larix olgensis* Henry) plantation planted in 1962–1964, had become seminatural conifer-broadleaf mixed forests of larch–spruce–fir. Larch, spruce (*Picea jezoensis* Carr.) and fir (*Abies nephrolepis* (Trautv.) Maxim.) were the dominant tree species. The main associated tree species were classified as three tree species (groups): Korean pine (*Pinus koraiensis* Sieb. et Zucc.), slow growing tree species group (acer mono (*Acer mono* Maxim.), Manchurian ash (*Fraxinus mandshurica* Rupr.), tilia amurensis (*Tilia amurensis* Rupr.), betula costata (*Betula costata* Trautv.), amur corktree (*Phellodendron amurense* Rupr.) and medium growing tree species group (birch (*Betula platyphylla* Suk), elm (*Ulmus propinqua* Koidz.)). Beginning with the first inventory in the year of 1986, tree- and stand-level characteristics of these permanent plots have been remeasured periodically with an interval of two or three years from 1986 to 2010. Characteristics recorded for each tree include DBH above 5 cm, tree species, and status (live, dead, cut and damage). Site characteristics recorded for each plot include slope, aspect and elevation. Our analyses focused on the plot data within an interval of any five consecutive years in 25 years (1986–2010). In total, we received 16,619 observations. Summary statistics of tree- and stand-level variables were listed in Table 1.

Table 1. Summary statistics of permanent plot data within an interval of any five consecutive years ($n = 16,619$).

Variable	Mean	Sd	Minimum	Maximum	Description
ΔD (cm)	0.28	0.21	0	2.04	Individual tree mean annual DBH increment within an interval of any five consecutive years
N (trees·hm ⁻²)	1017	245	395	1585	The number of trees per hectare
BA (m ² ·hm ⁻²)	26.26	5.42	14.22	37.37	Basal area per hectare
D _g (cm)	18.31	2.13	13.01	22.95	Average DBH
Slope (°)	9.25	3.54	6.00	18.00	Slope
Elevation (m)	691.10	62.79	600.00	800.00	Elevation

2.1.3. Climate Data

Basic climate data of 43 weather stations in and around Jilin Province from 1961 to 2010 are provided by the China Meteorological Data Sharing Service System (<http://cdc.cma.gov.cn>). The climate data were spatially interpolated using ANUSPLIN software with a spatial resolution of 300 m × 300 m [29]. For each plot, 44 candidate climatic variables corresponding to individual tree mean annual DBH increment (ΔD) within any five years' interval were calculated including temperature, precipitation, illumination and composite climatic factors generated by their interactions (Table 2). The values of candidate climate variables were the mean of corresponding climatic factors during the five-year interval.

Table 2. Description of 44 candidate climatic variables.

Type	Variables	Description	Variables	Description
Temperature	Amxt (°C)	Annual maximum temperature	Amint (°C)	Annual minimum temperature
	Amt (°C)	Annual mean temperature	Gsdd5 (°C)	The accumulated temperature is greater than 5 °C in growing season (April–September)
	Gsmxt (°C)	Maximum temperature in growing season (April–September)	Gsmint (°C)	Minimum temperature in growing season (4–9 months)
	Maxtwm (°C)	The highest temperature in the hottest month (July)	Mmincm (°C)	The lowest temperature of the coldest month (January)
	Mtcm (°C)	The mean temperature of the coldest month (January)	Mtwm (°C)	The mean temperature of the hottest month (July)
Precipitation	Gsp (mm)	Average precipitation in growing season (April–September)	Map (mm)	Annual mean precipitation
	Msp (mm)	Monthly total precipitation	Sp (mm)	Summer precipitation (June–August)
	Gsrh (%)	Relative humidity in growing season (April–September)	Marh (%)	Annual mean relative humidity
Illumination	Asr (h)	Annual total solar radiation duration	Gssr (h)	Solar radiation duration in growing season (April–September)
Composite climatic factors generated by the interaction of basic climatic factors	ahm (°C·mm ⁻¹)	1000 × ((Amt + 10)/Map)	shm (°C)	Gsmxt + Maxtwm
	amaxtmap (°C·mm ⁻¹)	Amxt/Map	amtmap (°C·mm ⁻¹)	Amt/Map
	gsmxtgsrh (100 °C)	(Gsmxt × Gsrh)/1000	gsmintgsp (°C·mm ⁻¹)	Gsmint/Gsp
	gsmintgsrh (100 °C)	Gsmint/Gsrh	gsmintmap (°C·mm ⁻¹)	Gsmint/Map
	gspdd5 (mm·°C)	(Gsp × Gsdd5)/1000	gspgsrh (100 mm)	Gsp/Gsrh
	gspmtcm (mm·°C)	(Gsp × Mtcm)/1000	gspgsmint (mm·°C)	(Gsp × Gsmint)/1000
	gsrhgsmint (100 °C)	(Gsrh × Gsmint)/1000	mapdd5 (mm·°C)	(Map × Gsdd5)/1000
	mapgsmint (mm·°C)	(Map × Gsmint)/1000	mapmtcm (mm·°C)	(Map × Mtcm)/1000
	maxtwmsp (°C·mm ⁻¹)	Maxtwm/Sp	mtcmgsp (°C·mm ⁻¹)	Mtcm/Gsp
	mtcmmap (°C·mm ⁻¹)	Mtcm/Map	pratio	Gsp/Map
	prdd5 (°C)	pratio × Gsdd5	tdiff (°C)	Mtwm – Mtcm
	tdgsp (°C·mm ⁻¹)	tdiff/Gsp	gsptd (°C·mm)	(Gsp × tdiff)/1000
	maptd (°C·mm)	(Map × tdiff)/1000	tdmap (°C·mm ⁻¹)	tdiff/Map

2.2. Modelling Techniques

2.2.1. Modelling Methods and Development

Individual tree diameter growth is often modeled as the function of tree size, competition, site condition and climate factors. In our study, the individual tree mean-annual DBH increment (ΔD) as an output were modeled using the RF, BRT, Cubist and MARS models respectively with the mentioned variables below as inputs. As climate, site condition and competition interactively influence individual tree growth, these influences may vary with tree species and size. In order to analyze the influences of tree species on ΔD , tree species were treated as a categorical variable in our study. Thus, the model form was expressed as follows:

$$\Delta D = f(\text{Size, Competition, Site, Climate, Species}). \quad (1)$$

where $f()$ represents the RF, BRT, Cubist or MARS model; Size represents tree size variables including initial tree DBH at the beginning of five-year interval (D_0) and stand average DBH (D_g); Competition represents competition variables (Equations (2)–(5)) including the sum of basal area larger than the subject tree (BAL), the sum of square DBH in trees which are larger than subject tree (DL), the ratio of DBH of subject tree to average DBH (RD), the ratio of DBH of subject tree to the maximal DBH (DDM), the number of trees per hectare (N) and basal area per hectare (BA); Species represents a tree species category including six tree species (groups): larch, spruce, fir, Korean pine, slow growing tree species group and medium growing tree species group; Site represents site condition variables including aspect, slope and altitude; the “Climate” represents climate variables including 44 climatic factors (see Table 2). Totally there were 56 input variables.

$$\text{BAL} = \frac{\pi}{40000} \frac{\sum_{i=1}^{n_{\text{larger}}} D_i^2}{\text{Area}}. \quad (2)$$

$$\text{DL} = \frac{\sum_{i=1}^{n_{\text{larger}}} D_i^2}{\text{Area}}. \quad (3)$$

$$\text{RD} = \frac{D_{\text{subject}}}{D_{\text{average}}}. \quad (4)$$

$$\text{DDM} = \frac{D_{\text{subject}}}{D_{\text{maximum}}}. \quad (5)$$

where n_{larger} is the number of the trees with the basal area larger than that of the subject tree, D_i is the DBH of the i -th tree, D_{subject} is the DBH of the subject tree, D_{average} is the stand average DBH, D_{maximum} is the maximal DBH of the stand, Area is the plot area.

(1) Random Forest Model

The RF algorithm was first proposed by Breiman [30]. It is a decision tree model of ensemble learning. For regression prediction, the base learner of RF is regression tree. Each regression tree of RF is constructed by drawing bootstrap sample of the original dataset with replacement. When a bootstrap sample is drawn, approximately 36.8% of the original dataset is not included in the sample. The portion of the data not drawn into the sample composes the “out-of-bag” data, used to assess the generalization errors of the selected regression tree, whereas the bootstrap sample is used to train a regression tree. For each split node of the regression tree, a randomized subset of predictors is selected: randomly select k ($<P$) of the original predictors (P is number of original predictors). The best predictor to split a “parent” split node into two “child” nodes is the predictor among the k predictors leading to the largest decrease in the overall sums of squares error. A regression tree is completed until the terminal nodes contain only individuals belonging to the same classes or until they contain no more than five individuals. Thus, multiple trees are generated by repeating the partition process for a pre-specified number of trees. The RF predictions are the averaged predictions over all trees. The importance score of predictors can be estimated based on the total decrease in residual sum of squares over all nodes for which it was selected as the splitting predictor and averaged over all trees.

The RF modeling was implemented by the package “randomForest” [31] in R software [32]. Two tuning parameters require users to make choices: “ntree” and “mtry” [33]. The ntree, is the number of trees to grow and governs the total number of independent trees. The mtry is the number of predictive variables randomly sampled as candidates at each split and controls the correlations between trees, and decreasing mtry will decrease the correlations between trees. The default value of ntree is 500. In general, the overall error rate tends towards stability while the ntree is greater than 500 [34]. In order to ensure the reliability of the prediction results without affecting the computational efficiency, the value of ntree in our case was set to be 1000. For mtry, the default value is recommended set to be the number of total predictor variables divided by three in the regression context [30]. However, the default value of mtry is not always able to obtain the optimal model. It is necessary to tune the

parameter. Since there were 56 predictive variables in our case, we tested 56 possible values for $mtry$: $mtry = 1, 2, 3, \dots, 56$. A resampling technique—tenfold cross-validation—was implemented to train and evaluate these 56 RF models ($n_{trees} = 1000$, $mtry = 1, 2, 3, \dots, 56$). The R package “caret” [35] and a random seed were used to tune these RF models.

(2) Boosted Regression Tree Model

The BRT is also an ensemble learning method combining regression trees and “gradient boosting” algorithm. Unlike RF, BRT develops multiple regression trees sequentially from the gradient of loss function of the previous regression tree, whereas RF is a simple ensemble of independent trees [33]. More specifically, at the first step, a regression tree is fitted to a subset of data for minimizing the loss function (e.g., the squared error for regression). Then the second tree is fitted to the gradient (e.g., residual) of the first tree [36]. The two regression trees are combined and the gradient is computed. A new tree is fitted, and so on. The above procedure continues until a pre-specified number of trees is reached. In order to reduce overfitting and improve model prediction performance, each regression tree of BRT is constructed by drawing bootstrap sample of the original dataset without replacement (usually 50% of the original sample) [37]. The BRT predictions are the weighted average of predictions over all trees, the weights are proportional to regression tree performances [36]. The importance score of a predictor can be constructed for BRT in exactly the same way as it was for RF.

The BRT modeling was implemented by the R package “gbm” [38]. Five tuning parameters require users to make choices: “distribution”, “bag.fraction”, “shrinkage”, “n.trees” and “interaction.depth”. As the tuning parameters are multitudinous and changing any of the parameters can influence the optimal values of the other parameters, tuning BRT model is complex [33]. The distribution is the form of loss function. For regression prediction, the distribution is set to be “gaussian” (squared error). The bag.fraction ($0 < \text{bag.fraction} \leq 1$) is the sampling fraction of the training data randomly selected to create each tree. Lower bag.fraction can introduce more randomness into the model and counteract overfitting while also reducing the required computing costs [37]. However, increased randomness can lead to higher between-model variability [33,39]. Friedman [37] recommended using the bag.fraction of around 0.5. The shrinkage ($0 < \text{shrinkage} \leq 1$) is the step-size reduction or learning rate. The n.trees, is the total number of trees to grow. The parameter shrinkage affects the optimal n.trees. Decreasing shrinkage increases the value of n.trees required, and a lower shrinkage improves model performance but also increases computing costs (both storage and time) [38]. Balancing computing costs and model predictive performance, Ridgeway [38] recommended setting shrinkage at a range from 0.01 to 0.001 with n.trees between 3000 and 10,000. Thus, we consider three possible values for shrinkage (0.01, 0.005, 0.001) and 15 possible values for n.trees (3000, 3500, \dots , 10,000). The interaction.depth, is the maximum depth of variable interactions for each tree, with the default value of 1. However, the parameter interaction.depth affects the optimal setting for both n.trees and shrinkage. As interaction.depth is increased, shrinkage needs to be decreased if sufficient trees are grown [39]. Considering the trade-off between computing costs and model predictive performance, we tried 5 possible values for interaction.depth: 1, 3, 5, 7, 9. The tenfold cross-validation was also implemented to train and evaluate these 225 BRT models (distribution = “gaussian”, bag.fraction = 0.5, shrinkage = 0.01, 0.005, 0.001, n.trees = 3000, 3500, \dots , 10,000, and interaction.depth = 1, 3, 5, 7, 9). The R package “caret” [35] and a random seed, which is identical to the seed used for RF models, were used to tune BRT models.

(3) Cubist Model

The Cubist can be an ensemble learning method combining model tree and “committees” (boosting-like) algorithm. The model tree used in Cubist works in a similar way as the classical regression tree does. However, unlike the regression tree, the terminal nodes of the model tree use linear regression models rather than discrete values, and the splitting criterion of the model tree uses the reduction in standard deviation instead of the reduction in squared error [27]. Each model tree

of Cubist is affected by the result of the previous model tree. More specifically, subsequent model trees of Cubist are built as follows: if a data point is over-predicted by the prior model tree, then the same value is adjusted downward for the next model tree. Similarly, underpredicted data points are adjusted so that the next model tree will produce a larger prediction in the next iteration. The above procedure continues until a pre-specified number of committee model trees is reached. Unlike BRT, the predictions of Cubist are simple average of the predictions from each model tree instead of the weighted average of predictions over all trees.

The Cubist modeling was implemented by the R package “Cubist” [40]. Two tuning parameters require users to make choices: “committees” and “neighbors”. The committees is the number of iterative model trees to grow in sequence whose maximal value is limited to 100 [41]. The neighbors is the number of neighboring training points which can be used to adjust the final predictions and is an integer with range 0–9 [42]. Considering the trade-off between computing costs and model predictive performance, we tried 20 possible values for committees (1, 6, 11, . . . , 96) and 6 possible values for neighbors (0, 1, 3, 5, 7, 9). The tenfold cross-validation was also implemented to train and evaluate these 120 Cubist models (committees = 1, 6, 11, . . . , 96, and neighbors = 0, 1, 3, 5, 7, 9). The R package “caret” [35] and a random seed, which is identical to the seed used for RF models, were used to tune Cubist models.

(4) Multivariate Adaptive Regression Splines Model

The MARS, which was firstly proposed by Friedman [43], can be deemed to be an extension of linear models that automatically captures nonlinearity and interactions between variables. The modeling procedures of MARS are inspired by the recursive partitioning approach such as CART [43,44]. However, unlike CART, MARS creates continuous models with continuous derivatives instead of discontinuous branching at tree nodes [28,43]. More specifically, MARS creates flexible regression models by using basis function to fit piecewise linear regression models to distinct intervals of the independent variable space [28,45]. Both the knots (the break values of the intervals) and the variables to use are selected by using basis functions with exhaustive search. After the initial model is created with the first two basis functions, the model keeps on conducting next exhaustive search to get another set of basis functions that yield the best model fit. The above procedure continues until a stopping point is reached. The model form of MARS can be presented as follows:

$$\hat{f}(x) = \beta + \sum_{n=1}^N \beta_n B_n(x) \quad (6)$$

where, $\hat{f}(x)$ represents the dependent variable; x represents the independent variables; β is the intercept; $B_n(x)$ is the basis function, which may be a single hinge function or a product of two or more hinge functions; β_n represents the coefficient of the n -th basis function; N is the number of basis functions of the model. Hinge functions are a key part of MARS models. A hinge function takes the form as follows:

$$\max(0, x - c) \text{ or } \max(0, c - x) \quad (7)$$

where c is a constant, also called the knot.

The MARS modeling was implemented by the R package “earth” [46]. Two tuning parameters require users to make choices: “degree” and “nprune” [27]. The parameter degree is the maximum degree of interaction. The default of degree is 1, which means an additive MARS model without interaction terms is built. Although MARS can build models with two or more degree interactions, Hastie et al. [47] recommended setting an upper bound on the degree of interaction. The lower degree interaction can aid in the interpretation of the final model, whereas the higher degree interaction can lead to occasional instabilities in the model predictions (perhaps an order of magnitude of the true value) [27,47]. Thus, we tried 3 possible values for degree: 1, 2, and 3. Parameter nprune is the maximum number of terms (including intercept). The value of nprune is equal or greater than 2

and less than “nk”. The nk is the maximum number of model terms before pruning ($nk = \min(200, \max(20, 2 \times ncol(x))) + 1$, where $ncol(x)$ is the number of predictive variables). Since there were 56 predictive variables in our case, we considered 112 possible values for nprune: $nprune = 2, 3, \dots, 113$. The tenfold cross-validation was also implemented to train and evaluate these 360 MARS models (degree = 1, 2, 3, and nprune = 2, 3, \dots , 113). The R package “caret” [35] and a random seed, which is identical to the above ML methods, were used.

2.2.2. Model Evaluation

The performance of the RF, BRT, Cubist and MARS models was evaluated using tenfold cross-validation resampling technique. Three measures for model performance were computed for each fold: coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE). Then the 10 resampled validation measurements of model performance were averaged as follows.

Coefficient of determination of cross-validation (R^2_{cv}):

$$R^2_{CV} = \frac{1}{k} \sum_{j=1}^k (R^2_j) = \frac{1}{k} \sum_{j=1}^k \left(1 - \frac{\sum_{i=1}^{n_j} (O_{ij} - P_{ij})^2}{\sum_{i=1}^{n_j} (O_{ij} - \bar{O}_j)^2} \right) \quad (8)$$

Root mean square error of cross-validation ($RMSE_{cv}$):

$$RMSE_{CV} = \frac{1}{k} \sum_{j=1}^k (RMSE_j) = \frac{1}{k} \sum_{j=1}^k \left(\sqrt{\frac{1}{n_j} \sum_{i=1}^{n_j} (O_{ij} - P_{ij})^2} \right) \quad (9)$$

Mean absolute error of cross-validation (MAE_{cv}):

$$MAE_{CV} = \frac{1}{k} \sum_{j=1}^k (MAE_j) = \frac{1}{k} \sum_{j=1}^k \left(\frac{1}{n_j} \sum_{i=1}^{n_j} |P_{ij} - O_{ij}| \right) \quad (10)$$

where, k is the number of fold, and $k = 10$ in our case; O_{ij} is the i -th observed value of the j -th fold; P_{ij} is the i -th predicted value of the j -th fold; \bar{O}_j is the average observed value of the j -th fold; n_j is the number of samples in j -th fold; R^2_j , $RMSE_j$ and MAE_j are the R^2 , RMSE and MAE of the j -th fold, respectively.

2.3. Model Comparison

Once model performance has been quantified through the tuning process described above, the optimal RF, BRT, Cubist and MARS models were determined by selecting parameter settings associated with the numerically best performance estimates (the smallest $RMSE_{cv}$ in our case). Then the comparisons between the four optimal models were made by assessing the performance measures on each fold. As the performances were measured using identically resampled datasets, statistical methods for paired comparisons can be used to detect pairwise differences in the performance among models [27,48]. Therefore, paired t -test was applied to verify if the differences in performance among the four optimal models with the significant level of 0.05.

Moreover, the four models in our study could provide estimations of importance score. For limited space of the manuscript and wide use, only the results of RF and BRT models were reported after the pre-examination, which showed similar results. The importance score of a predictor was estimated based on the total decrease in residual sum of squares over all nodes for which it was selected as the

splitting predictor and averaged over all trees. It was normalized to 100% by its relative value (rIM_i), which was expressed as following for the RF and BRT models:

$$rIM_i = \frac{IM_i}{\sum_{q=1}^P IM_q} \times 100\%. \quad (11)$$

where, the IM_i and IM_q are the importance score of the i -th and q -th predictor, P is the total number of the predictors.

3. Results

3.1. Individual Tree Diameter Prediction Based on ML Algorithms

3.1.1. RF

The effect of $mtry$ on RF model performance is presented in Figure 1. Parameter values of $mtry$ at 1 and 9 had the lowest and highest R^2_{cv} (0.4460 and 0.5565), respectively. For $RMSE_{cv}$, $mtry$ at 9 and 1 resulted in the lowest and highest values (0.1377 and 0.1751 cm). For MAE_{cv} , the lowest and highest MAE_{cv} (0.0996 and 0.1341 cm) were at 11 and 1, respectively. As $mtry$ increased, R^2_{cv} initially increased and then stabilized at the values larger than 6, and both $RMSE_{cv}$ and MAE_{cv} initially decreased and then stabilized at values between 6 and 56. Thus, the optimal model of RF ($n_{tree} = 1000$ and $mtry = 9$) was determined using the smallest $RMSE_{cv}$.

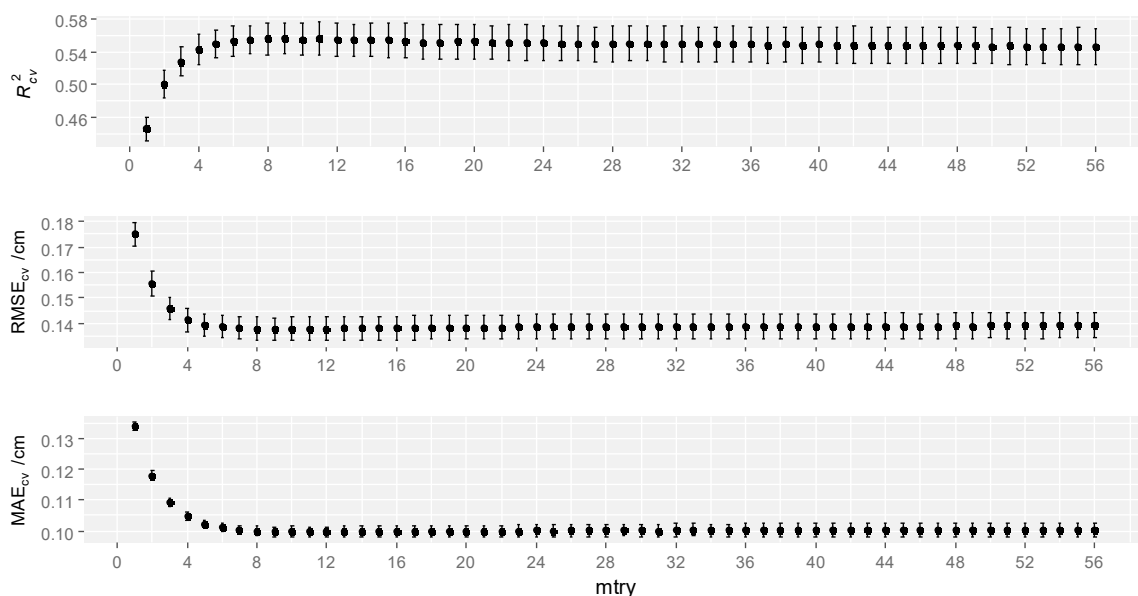


Figure 1. Tuning random forest (RF)—the effect of $mtry$ on RF model performance. The tenfold cross-validation was used for model validating. The error bar is standard deviation. The parameter, $mtry$, is the number of predictive variables randomly sampled as candidates at each split and controls the correlations between trees.

3.1.2. BRT

Figure 2 shows the effect of n_{trees} , $shrinkage$ and $interaction.depth$ on BRT model performance. Parameter values of n_{trees} , $shrinkage$ and $interaction.depth$ at (3000, 0.001, 1) and (8500, 0.01, 9) resulted in the lowest and highest R^2_{cv} (0.4171 and 0.5714). For $RMSE_{cv}$, n_{trees} , $shrinkage$ and $interaction.depth$ at (7000, 0.01, 9) and (3000, 0.001, 1) had the lowest and highest $RMSE_{cv}$ (0.1353 and 0.1611 cm). For MAE_{cv} , the lowest and highest MAE_{cv} (0.0982 and 0.1225 cm) were at (9500, 0.01, 9) and (3000, 0.001, 1), respectively. In general, as n_{trees} increased, R^2_{cv} initially increased and then stabilized at higher number of n_{trees} , both $RMSE_{cv}$ and MAE_{cv} initially decreased and then stabilized

at higher number of n.trees. Overall, shrinkage of 0.01 had better model performance than shrinkage of 0.001 and 0.005. In addition, interaction.depth of 9 had the best model performance. Thus, the optimal model of BRT (distribution = “gaussian”, bag.fraction = 0.5, shrinkage = 0.01, n.trees = 7000 and interaction.depth = 9) was determined using the smallest RMSE_{cv}.

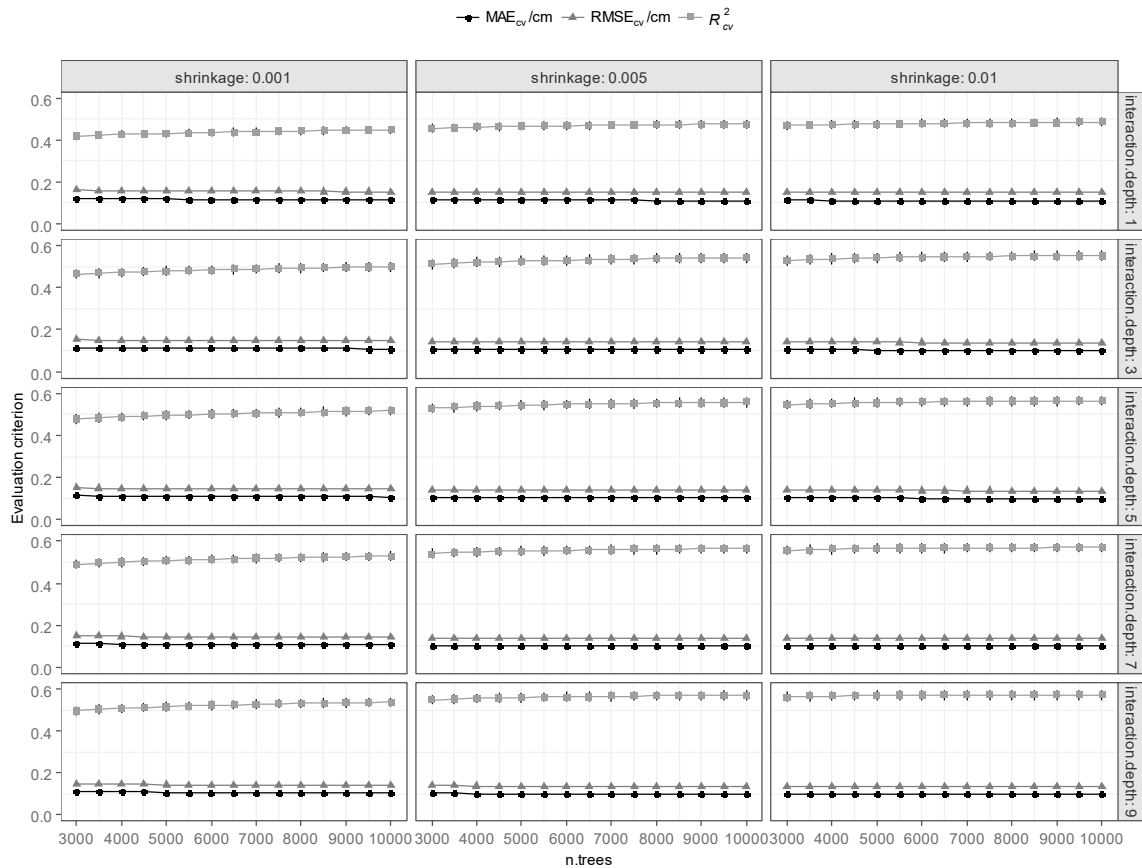


Figure 2. Tuning boosted regression tree (BRT)—the effect of n.trees, shrinkage and interaction.depth on BRT model performance. The tenfold cross-validation was used for validating. The error bar is standard deviation. The parameters, n.trees, shrinkage and interaction.depth are the total number of trees to grow, the step-size reduction or learning rate and the maximum depth of variable interactions for each tree, respectively.

3.1.3. Cubist

The lowest and highest R^2_{cv} (0.4564 and 0.5734) were at committees values of 1 and 86 and neighbors values of 1 and 9, respectively (Figure 3). For RMSE_{cv}, the lowest and largest values (0.1351 and 0.1650 cm) were at committee values of 31 and 1 and neighbors values of 9 and 1. For MAE_{cv}, the lowest and highest (0.0970 and 0.1077 cm) were at committee values of 31 and 1 and neighbors values of 5 and 1. In generally, as committees increased, R^2_{cv} initially increased and then stabilized at higher number of committees, and both RMSE_{cv} and MAE_{cv} initially decreased and then stabilized at higher number of committees. Increasing neighbors would increase R^2_{cv} and decrease both RMSE_{cv} and MAE_{cv} with the exception of the neighbors’ value of 0. Thus, the optimal model of Cubist (committees = 31 and neighbors = 9) was determined using the smallest RMSE_{cv}.

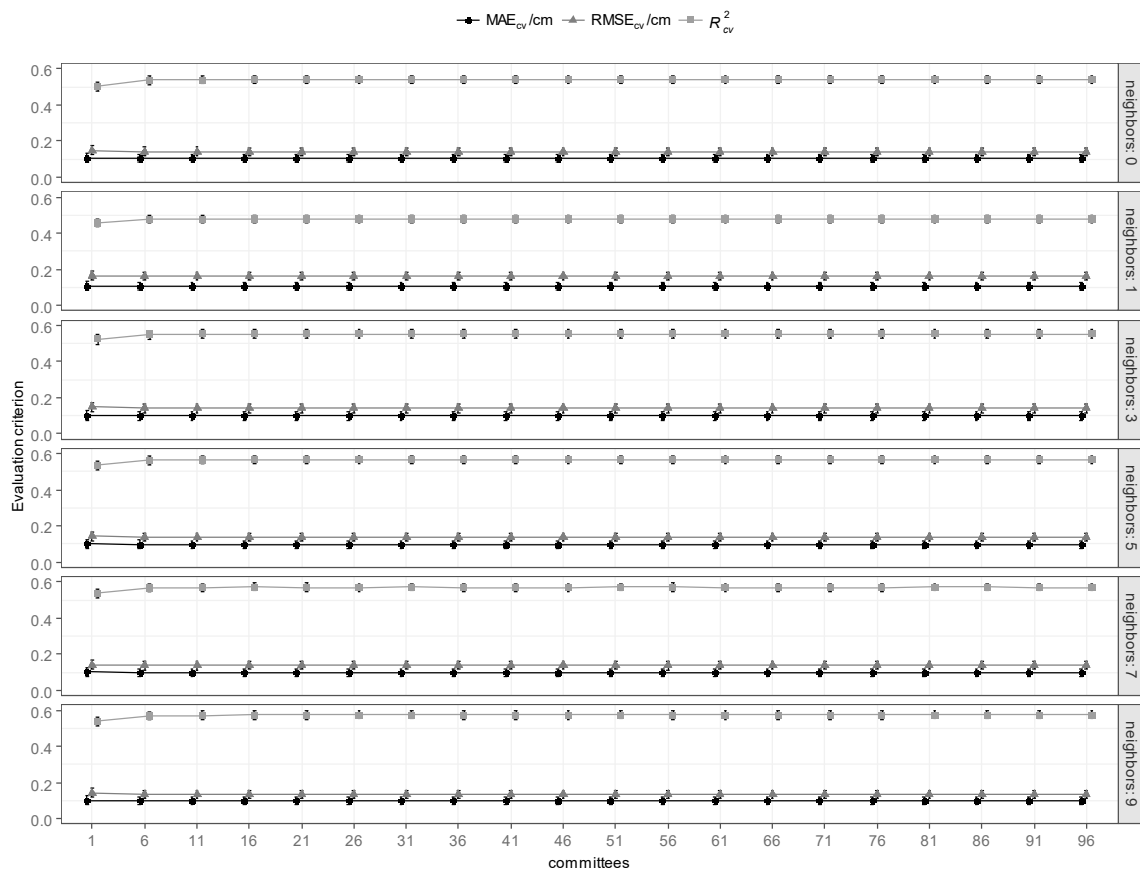


Figure 3. Tuning Cubist—the effect of neighbors and committees on Cubist model performance. The tenfold cross-validation was used for validating. The error bar is standard deviation. The parameters, committees and neighbors, are the number of iterative model trees to grow in sequence and the number of neighboring training points which can be used to adjust the optimal predictions, respectively.

3.1.4. MARS

Parameter values of $nprune$ and $degree$ at (2, 3) and (51, 2) resulted in the lowest and highest R^2_{cv} (0.3062 and 0.4993), respectively (Figure 4). The lowest and highest $RMSE_{cv}$ (0.1462 and 0.1721 cm) were at (39, 2) and (2, 3). Parameters $nprune$ and $degree$ at (45, 3) and (2, 3) had the lowest and highest MAE_{cv} (0.1085 and 0.1278 cm). Generally, as $nprune$ increased, R^2_{cv} initially increased and then stabilized at values between 25 and 113, both $RMSE_{cv}$ and MAE_{cv} initially decreased and then stabilized at the same value. The $degree$ of 2 had better model performance than $degree$ of 1 and 3. Thus, the optimal model of MARS ($nprune = 39$ and $degree = 2$) was determined using the smallest $RMSE_{cv}$.

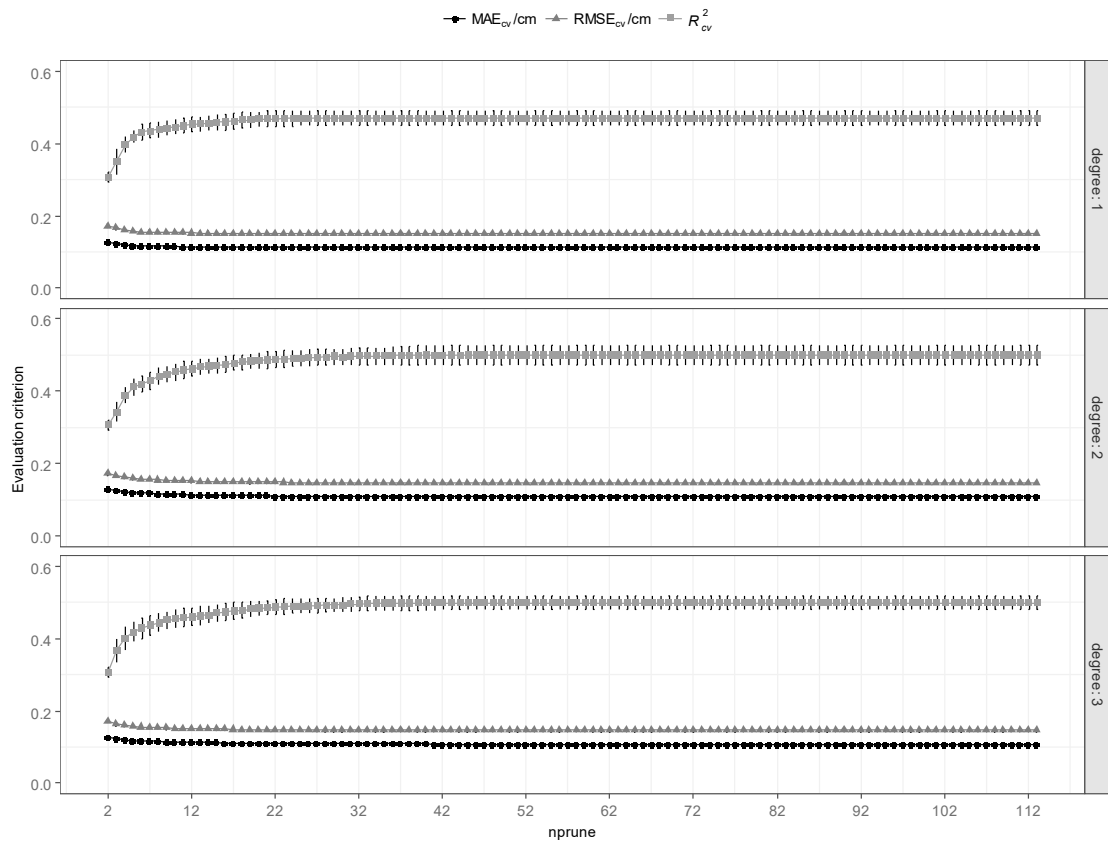


Figure 4. Tuning multivariate adaptive regression splines (MARS)—the effect of degree and nprune on MARS model performance. The tenfold cross-validation was used for validating. The error bar is standard deviation. The parameters, degree and nprune, are the maximum degree of interaction and the maximum number of terms (including intercept), respectively.

3.1.5. Model Evaluation and Comparisons

The predictive performance of four selected optimal models is presented in Table 3. The results of tenfold cross-validation indicated that the RF, BRT and Cubist models significantly outperformed the MARS model in terms of three performance measures (the *t*-test *p* value for corresponding pairs <0.05). The highest R^2 , the lowest RMSE and the lowest MAE were observed for Cubist model, followed by the BRT and RF model, while the lowest R^2 , the highest RMSE and the highest MAE were found for the MARS model.

Table 3. The predictive performance measures of four optimal models based on tenfold cross-validation.

Evaluation Criteria	Models			
	RF	BRT	Cubist	MARS
R^2_{cv}	0.5565(0.0186) a	0.5711(0.0232) a	0.5734(0.0215) a	0.4993(0.0262) b
RMSE _{cv} (cm)	0.1377(0.0044) b	0.1353(0.0050) b	0.1351(0.0048) b	0.1462(0.0047) a
MAE _{cv} (cm)	0.0996(0.0015) b	0.0985(0.0018) bc	0.0972(0.0015) c	0.1086(0.0022) a

Values in the bracket were standard deviation. Different lowercase letters (a, b, c) indicated significant difference between models at 0.05 level.

3.2. Variable Importance Based on the Optimal RF and BRT Models

Relative importance of each variable as determined from the optimal RF and BRT models is presented in Figure 5.

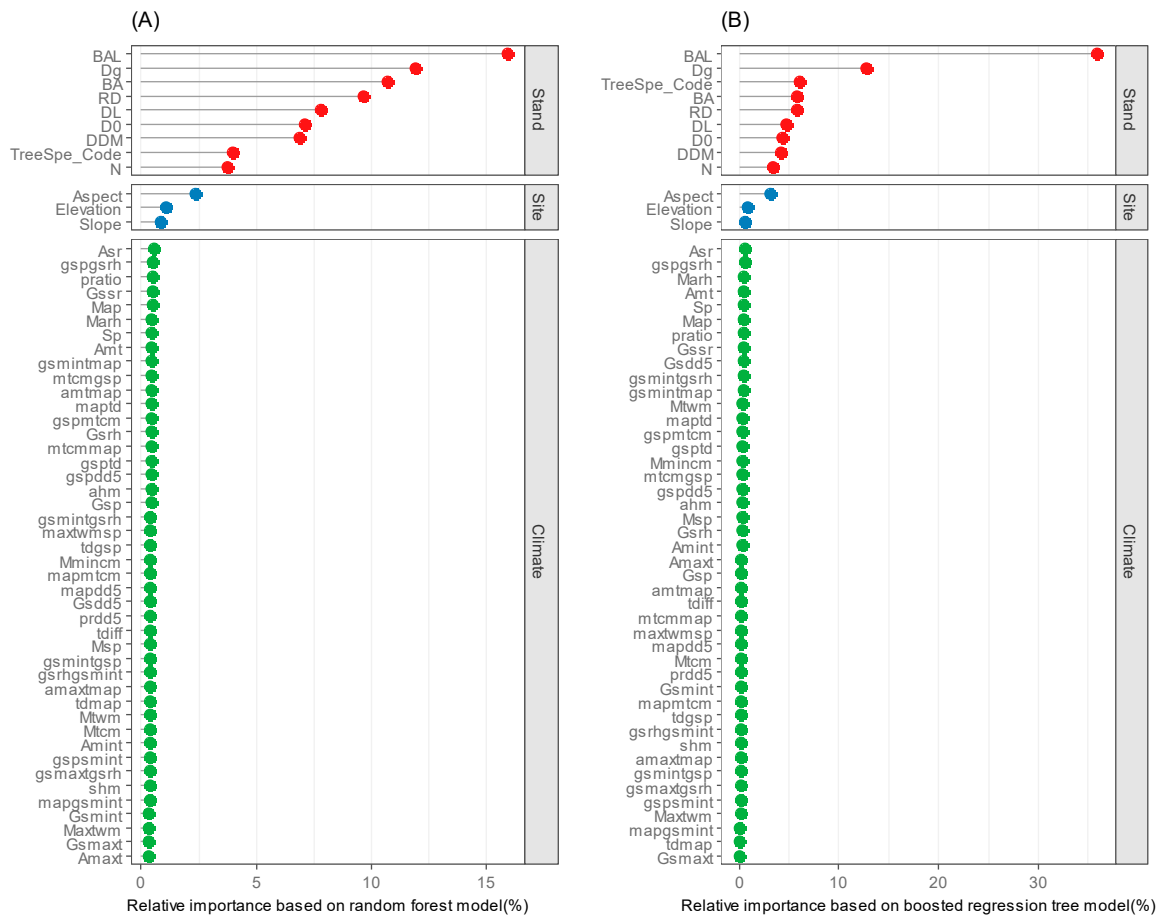


Figure 5. Relative importance of each variable as determined from (A) random forest (RF), and (B) boosted regression tree (BRT), which were normalized to 100%. The explanations of stand factors: BAL is the sum of basal area larger than the subject tree, Dg is average DBH, BA is the basal area per hectare, RD is the ratio of DBH of subject tree to average DBH, DL is the sum of square DBH in trees which are larger than subject tree, D0 is the individual tree DBH at the beginning of 5-year interval, DDM is the ratio of DBH of subject tree to the maximal DBH, TreeSpe_Code is the tree species, and N is the number of trees per hectare. The explanations of site factors: Aspect is aspect, Slope is slope and Altitude is altitude. The explanations of climate factors were presented in Table 2.

The RF model indicated that the variability in individual tree diameter growth was mainly explained by stand factors (accumulating relative importance of 77.69%), followed by climate factors (18.03%) and site factors (4.28%). Among nine stand factors, competition factors had 54.67% relative importance—BAL (15.91%), BA (10.69%), RD (9.66%), DL (7.80%), DDM (6.86%) and N (3.75%)—size factors had 19.02% relative importance—D_g (11.91%) and D₀ (7.11%)—and tree species had 4.00% relative importance (TreeSpe_Code). Among 44 climate factors, Asr (0.54%), gspgsrh (0.50%), pratio (0.49%), Gssr (0.48%) and Map (0.48%) were the top five important variables, and the relative importance of other 39 variables ranged from 0.45% to 0.33%. In addition, among 3 site factors, aspect had the highest influence on individual tree diameter growth (relative importance of 2.38%), followed by elevation (1.05%) and slope (0.85%).

On the other hand, stand factors, climate factors, and site factors had 82.89%, 12.57% and 4.54% of individual tree diameter growth relative importance in the BRT model, respectively. Among nine stand factors, competition factors had 59.77% relative importance—BAL (35.92%), BA (5.80%), RD (5.74%), DL (4.78%), DDM (4.19%) and N (3.34%). Size factors had 17.09 relative importance—D_g (12.76%) and D₀ (4.33%)—and tree species had 6.03% relative importance (TreeSpe_Code). Among 44 climate factors, Asr (0.58%), gspgsrh (0.52%), Marh (0.48%), Amt (0.47), Sp (0.45%) were the top five important

variables, and the relative importance of other 39 variables ranged from 0.45% to 0.00%. In addition, among three site factors, aspect had the highest influence on individual tree diameter growth (relative importance of 3.15%), secondly elevation (0.82%) and slope (0.57%).

Generally, it was shown that the relative importance of predictors as determined from the RF model was basically identical with that from the BRT model: competition and size factors were the most important variables (accumulating relative importance was over 73.00%), followed by climate factors (not more than 18.03%) and site factors (not more than 4.54%).

4. Discussion

4.1. ML Model Development

In the process of ML model evaluation and selection, it is not only required to select suitable learning algorithms, but also to set the corresponding parameters, which is commonly referred to as “parameter tuning” [49]. The performance of models can vary drastically depending on the value selected for the parameters. Therefore, parameter tuning was implemented and tested for the four ML algorithms in the study.

In RF modeling, two parameters were required to be tuned: “ntree” and “mtry” [33]. The default value of ntree is 500 and the overall error rate tends towards stability when the ntree is greater than 500 [34]. Breiman [30] proved that RF is less vulnerable to over-fitting and RF model will not be negatively impacted even though a large ntree is set. Thus, the only limitation of ntree is computing costs (both storage and time) [30]. In our study, the value of ntree was set to be 1000 in order to balance model predictive performance and computing costs. The parameter mtry usually performed well at the recommended value (one-third the number of total predictors for regression) [33]. In addition, both Kuhn and Johnson [27] and Liaw and Wiener [31] found that RF was relatively insensitive to the selection of mtry. Our study found that the performance of RF would be adversely affected by the selection of mtry when the mtry was small (model performance increased with the increasing of mtry when $mtry < 6$). Whereas, model performance of RF was relatively insensitive to the selection of mtry when mtry was large (model performance stabilized at between 6 and 56 mtry). Recommended default mtry of 18 was almost equivalent to that with mtry of 9 in the study in terms of model performance. Our study supported the conclusion that recommended default mtry for regression tended to perform well [27,33].

In BRT modeling, we tested parameters “shrinkage”, “n.trees” and “interaction.depth”. The shrinkage controls the contribution of each tree on the optimal predictions [33,47], and contrary results were found including lower and large shrinkage could improve model performance [27,38]. In our study, we found that shrinkage of 0.01 had better model performance than those of 0.001 and 0.005. The n.trees controls the total number of iterations. Our results were consistent with Kuhn and Johnson [27] that $RMSE_{cv}$ decreased as n.trees increased and then stabilized at higher values. The interaction.depth parameter controls the number of nodes in a tree [39]. Freeman et al. [33] found that model performed better with a fairly large interaction.depth. Our results were in agreement with them.

The maximal value of committees is limited to 100 in Cubist modeling [41]. The other parameter neighbors, is an integer with 0–9 range [42]. We found that the $RMSE_{cv}$ initially significantly reduced as the committees increased and then stabilized at a higher committees, and a model with neighbors of 0 performed better than the model with neighbors of 1. These findings were consistent with the previous study [27].

In MARS modeling, our study found that degree of 1 in predicting individual tree diameter growth performed better than degree of 1 and 3 in terms of $RMSE_{cv}$, MAE_{cv} and R^2_{cv} . Kuhn and Johnson [27] noted that models with degree of 1 in predicting solubility were identical to models with degree of 2 in terms of $RMSE_{cv}$. The result that $RMSE_{cv}$ initially significantly reduced as the nprune increased and then stabilized at a higher nprune in the literature [27] was also supported by our study.

In general, as the RF, Cubist and MARS models had fewer parameters to be set by users and were less sensitive to tuning of these parameters; they were more user-friendly than BRT which had many parameters to be tuned. In contrast to RF, parameters of BRT had stronger influence on model quality [33].

4.2. Performance Evaluations of Four Optimal Models

In this study, the four ML methods, including RF, BRT, Cubist and MARS, were developed and compared for modelling individual tree periodical annual DBH increment with the inputs of stand, climate and site factors of larch–spruce–fir mixed forests in China. Unlike traditional statistical models, collinearity test and the filtering of predictors were not performed in our study. All four models (the RF, BRT, Cubist and MARS models) have the ability to model complex and nonlinear interactive relationship between the predictor variables and the response variable without statistical assumptions and predetermined mathematical equations. In addition, these models are naturally resistant to noninformative predictors due to the built-in feature selection mechanisms while the number of predictors is high [27,28]. In addition, these models are insensitive to the collinearity between independent variables, and model results are robust to unbalanced data and missing data [27,28,50]. Based on the aforementioned advantages, Prasad et al. [28] evaluated the MARS and RF models in reproducing current tree importance values distribution for four tree species without pre-processing of the data. They noted that RF was superior in reproducing current tree importance values distribution. Besides, Yang et al. [22] compared BRT and RF to predict soil organic carbon content based on 12 environment variables, and found that BRT and RF had similar performance. Based on our data, of all models, the RF, BRT and Cubist models had a distinct advantage over MARS in predicting individual tree diameter growth (Table 3). The Cubist model ranked the highest in terms of model performance (RMSE_{cv} [0.1351 cm], MAE_{cv} [0.0972 cm] and R^2_{cv} [0.5734]), followed by BRT and RF model, whereas the MARS ranked the lowest (RMSE_{cv} [0.1462 cm], MAE_{cv} [0.1086 cm] and R^2_{cv} [0.4993]). In generally, RF, BRT and Cubist models had better generalization ability and statistical reliability than MARS in our study. The reason is that the RF, BRT and Cubist models are ensemble learning methods, which construct a set of base learners and then produce new predictions by taking a weighted average of their predictions. Ensemble learning methods often obtain better predictive performance than those obtained from any of the constituent base learners alone. In previous study with the same data, Yu et al. [19] used generalized additive model (GAM) to investigate the effects of stand factors and climate factors on individual tree periodical annual DBH increment of larch–spruce–fir mixed forests in China. They found that the RMSE, MAE and R^2 of GAM were 0.1500 cm, 0.1100 cm and 0.5080, respectively. Comparing the models performance between our study and Yu et al. [19], we found that GAM was slightly better than MARS, but was far below RF, BRT and Cubist. However, when these models were used in different research areas, it was noted that there was no single model that would always do better than any other models. Moisen and Frescino [51] found that MARS and GAM had similar performance for predicting forest characteristics. Aertsens et al. [52] found that the GAM model performed better than BRT for the prediction of site index. On the other hand, Leathwick et al. [53] found that BRT performed better than GAM for predicting variation in demersal fish species richness.

Both the ML models and traditional statistical methods have strengths and weaknesses. The traditional statistical methods can give clear model formulas and corresponding parameters through predetermined mathematical equations (model form), and thus they are intuitive and easy to be understood in terms of model form and the relationship between dependent and independent variables. However, these models also have some limitations. They are often applied under certain statistical assumptions, such as the data are normally distributed, homoscedastic and independent. However, due to continuous observations, hierarchy and intrinsic variability of forest growth data, the above assumptions are usually difficult to satisfy. What's more, the selection of specific mathematical equation (model form) and convergence is rather difficult when modelling individual

tree growth in nonlinear form and with multiple variables. In contrast to the traditional statistical models, the ML methods have the ability to model complex and nonlinear interactive relationship between the predictor variables and the response variable without the test of statistical assumptions and the selection of predetermined mathematical equations. As a data-driven method, it is considered as “black box” model, however, the relative importance and partial dependence plot produced from RF and BRT also provide a clear explanation as traditional methods. When there were a large number of predictors, ML could help select the most important ones. The combinations between ML and traditional statistical methods may be a better option.

4.3. Variable Importance in Predicting Individual Tree DBH Growth

Climate, site condition and competition factors interactively influence individual tree growth. Meanwhile, these influences may vary with tree species and size [14]. In this study, nine stand factors, three site condition factors and 44 climate variables were selected as predictors for predicting individual tree diameter growth because these variables were frequently used as key predictors in previous diameter growth research [5,9,19,54,55]. Our results showed that stand factors were the most influential factors in BRT and RF models (Figure 5). More specially, both BRT and RF showed that competition including stand- (N and BA) and tree-level competition indices (BAL, DL, RD, DDM) was the primary factor affecting individual tree diameter growth, followed by tree size (including D_0 and D_g) and stand characteristic (including tree species). These findings were consistent with Yu et al. [19] which found that stand factors were the main drivers to diameter growth. Similar results have been found in other research, for example, Jiang et al. [14] found that competition indices were the primary factors affected individual tree diameter growth for trembling aspen (*Populus tremuloides* michx.), followed by tree size, and that diameter growth of white spruce (*Picea glauca* (Moench.) Voss) was primarily influenced by tree size, followed by competition. Alam et al. [56] found for white spruce that diameter growth of trees was primary constrained by competition. Our study showed that both tree size and competition were the main drivers to diameter growth. Stand-level competition indices can indicate the average crowding degree of the whole stand or the full utilization degree of the stand to the site. The bigger the competition index of BA and N stands, the smaller average occupancy of each tree to growth space is, so the smaller the average growth of DBH of individual trees [19]. Generally speaking, the growth space occupied by individual trees varied with trees sizes, and trees with different sizes bear different competitive pressures. The bigger the competition index of BAL and DL, the larger competition pressure is, so the smaller average DBH increment is. In contrast, the bigger the competition index of RD and DDM, the smaller the competition pressure is, so the bigger average DBH increment is. The D_g is a basic index reflecting the average size of a stand. The bigger D_g , the smaller average occupancy of each tree to the growth space is, so the smaller the average DBH increment is. The larger D_0 , the greater growth advantage of individual tree is. When D_0 reaches a certain maximum, tree growth gradually slows down. Tree growth is co-influenced by these factors. Although age is one of the most important drivers of tree or stand growth, it was not included as an input in our study since the measurement of tree age for uneven-aged mixed forest was difficult and tree size could be as a proxy in terms of model application. All inputs in the ML methods in the study only explain about 60% variation of tree DBH growth of larch–spruce–fir mixed forests, which showed the complexity of uneven-aged mixed forests growth modelling, and thus more predictors or further algorithms are needed in the future.

In our study, the relative importance of climate factors was less than that of stand factors. This finding was consistent with Yu et al. [19], which showed climate factors had limited capacity in explaining the variation in diameter growth, while stand factors were the main drivers to diameter growth. Similar results have been found in other research [14,56,57], for example, Jiang et al. [14] found for both trembling aspen and white spruce that the climate factors showed less effects on tree growth than that of stand factors. Alam et al. [56] found for white spruce that diameter growth of trees was primary constrained by competition with minor effects from climate factors. Besides, our results

noted that the relative importance of site factors was less than that of stand factors and climate factors. We found that aspect was a critical site factor in BRT and RF models, suggesting that aspect exerted a larger influence on individual tree diameter growth than elevation and slope did.

At first sight the relative importance of each variable as determined from BRT seemed to have a much steeper importance slope than that from RF (Figure 5). It is due to the fact that the base learners from BRT are dependent on each other and hence will have correlated structures among base learners. Thus, the same predictor will be selected across the base learners, increasing its contribution to the variable importance [27].

5. Conclusions

Four ML methods were applied and compared for predicting the individual tree diameter growth of larch–spruce–fir mixed forests in northeast China. Based on tenfold cross-validation resampling, we found that the RF, BRT and Cubist models had a distinct advantage over MARS in predicting individual tree diameter growth. The Cubist model ranked the highest in terms of model performance (RMSE_{cv} [0.1351 cm], MAE_{cv} [0.0972 cm] and R²_{cv} [0.5734]), followed by the BRT and RF models, whereas the MARS ranked the lowest (RMSE_{cv} [0.1462 cm], MAE_{cv} [0.1086 cm] and R²_{cv} [0.4993]). In general, the RF, BRT and Cubist models were effective and powerful modelling methods for predicting the individual tree diameter growth. In addition, as the RF and Cubist models had fewer parameters to be set by users and were less sensitive to tuning of these parameters, they were more user-friendly than BRT.

Relative importance of each predictor variable was determined from the optimal RF model and the optimal BRT model, respectively. Our results showed that both the RF and BRT models can be used to reasonably analyze the complex relationship between predictor variables and individual tree diameter growth, and to identify the dominant factors affecting tree growth. The stand factors were the main drivers to diameter growth, while both site factors and climate factors had limited capacity in explaining the variation in diameter growth at local scale.

Although the ML methods were examined in the case of individual tree DBH growth prediction of larch–spruce–fir mixed forests in northeast China, they could be generally applicable to growth and yield predictions of other forest types.

Author Contributions: Writing—original draft preparation and revising: Q.O.; Supervision, review and modification: X.L.; data preparation: C.S.

Acknowledgments: This research was supported by the National Natural Science Foundation of China (Grant No. 31870623, 61331018). We thank two anonymous reviewers for their valuable suggestions, which improved the manuscript.

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

References

1. Peng, C. Growth and yield models for uneven-aged stands: Past, present and future. *For. Ecol. Manag.* **2000**, *132*, 259–279. [[CrossRef](#)]
2. Gyawali, A.; Sharma, R.P.; Bhandari, S.K. Individual tree basal area growth models for Chir pine (*Pinus roxburghii* Sarg.) in western Nepal. *J. For. Sci.* **2015**, *61*, 535–543. [[CrossRef](#)]
3. Porté, A.; Bartelink, H.H. Modelling mixed forest growth: A review of models for forest management. *Ecol. Model.* **2002**, *150*, 141–188. [[CrossRef](#)]
4. Ashraf, M.I.; Zhao, Z.; Bourque, C.P.A.; Maclean, D.A.; Meng, F.R. Integrating biophysical controls in forest growth and yield predictions with artificial intelligence technology. *Can. J. For. Res.* **2013**, *43*, 1162–1171. [[CrossRef](#)]
5. Ritchie, M.W.; Hann, D.W. Implications of disaggregation in forest growth and yield modeling. *For. Sci.* **1997**, *43*, 223–233.
6. Mabvurira, D.; Miina, J. Individual-tree growth and mortality models for *Eucalyptus grandis* (Hill) Maiden plantations in Zimbabwe. *For. Ecol. Manag.* **2002**, *161*, 231–245. [[CrossRef](#)]

7. Zhao, D.; Borders, B.; Wilson, M. Individual-tree diameter growth and mortality models for bottomland mixed-species hardwood stands in the lower Mississippi alluvial valley. *For. Ecol. Manag.* **2004**, *199*, 307–322. [[CrossRef](#)]
8. Biging, G.S.; Dobbertin, M. Evaluation of competition indices in individual tree growth models. *For. Sci.* **1995**, *41*, 360–377.
9. Ma, W.; Lei, X. Nonlinear simultaneous equations for individual-tree diameter growth and mortality model of natural Mongolian oak forests in Northeast China. *Forests* **2015**, *6*, 2261–2280. [[CrossRef](#)]
10. Feeley, K.J.; Wright, S.J.; Supardi, M.N.N.; Kassim, A.R.; Davies, S.J. Decelerating growth in tropical forest trees. *Ecol. Lett.* **2007**, *10*, 461–469. [[CrossRef](#)]
11. Sánchez-Salguero, R.; Linares, J.C.; Camarero, J.J.; Madrigal-González, J.; Hevia, A.; Sánchez-Miranda, Á.; Ballesteros-Cánovas, J.A.; Alfaro-Sánchez, R.; García-Cervigón, A.I.; Bigler, C. Disentangling the effects of competition and climate on individual tree growth: A retrospective and dynamic approach in Scots pine. *For. Ecol. Manag.* **2015**, *358*, 12–25. [[CrossRef](#)]
12. Trasobares, A.; Zingg, A.; Walthert, L.; Bigler, C. A climate-sensitive empirical growth and yield model for forest management planning of even-aged beech stands. *Eur. J. For. Res.* **2016**, *135*, 263–282. [[CrossRef](#)]
13. Zang, H.; Lei, X.; Ma, W.; Zeng, W. Spatial heterogeneity of climate change effects on dominant height of larch plantations in northern and northeastern China. *Forests* **2016**, *7*, 151. [[CrossRef](#)]
14. Jiang, X.; Huang, J.G.; Cheng, J.; Dawson, A.; Stadt, K.J.; Comeau, P.G.; Chen, H.Y.H. Interspecific variation in growth responses to tree size, competition and climate of western Canadian boreal mixed forests. *Sci. Total Environ.* **2018**, *631*, 1070–1078. [[CrossRef](#)] [[PubMed](#)]
15. Subedi, N.; Sharma, M. Climate-diameter growth relationships of black spruce and jack pine trees in boreal Ontario, Canada. *Glob. Chang. Biol.* **2013**, *19*, 505–516. [[CrossRef](#)] [[PubMed](#)]
16. Chen, L.; Huang, J.G.; Stadt, K.J.; Comeau, P.G.; Zhai, L.; Dawson, A.; Alam, S.A. Drought explains variation in the radial growth of white spruce in western Canada. *Agric. Forest Meteorol.* **2017**, *233*, 133–142. [[CrossRef](#)]
17. Cortini, F.; Filipescu, C.N.; Groot, A.; MacIsaac, D.A.; Nunifu, T. Regional models of diameter as a function of individual tree attributes, climate and site characteristics for six major tree species in Alberta, Canada. *Forests* **2011**, *2*, 814–831. [[CrossRef](#)]
18. Toledo, M.; Poorter, L.; Peña-Claros, M.; Alarcón, A.; Balcázar, J.; Leñaño, C.; Licona, J.C.; Llanque, O.; Vroomans, V.; Zuidema, P. Climate is a stronger driver of tree and forest growth rates than soil and disturbance. *J. Ecol.* **2011**, *99*, 254–264. [[CrossRef](#)]
19. Yu, L.; Lei, X.; Wang, Y.; Yang, Y.; Wang, Q. Impact of climate on individual tree radial growth based on generalized additive model. *J. Beijing For. Univ.* **2014**, *36*, 22–32, (In Chinese with English abstract).
20. 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](#)]
21. De’Ath, G. Boosted trees for ecological modeling and prediction. *Ecology* **2007**, *88*, 243–251. [[CrossRef](#)]
22. Yang, R.; Zhang, G.; Liu, F.; Lu, Y.; Yang, F.; Yang, F.; Yang, M.; Zhao, Y.; Li, D. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecol. Indic.* **2016**, *60*, 870–878. [[CrossRef](#)]
23. Tramontana, G.; Ichii, K.; Camps-Valls, G.; Tomelleri, E.; Papale, D. Uncertainty analysis of gross primary production upscaling using Random Forests, remote sensing and eddy covariance data. *Remote. Sens. Environ.* **2015**, *168*, 360–373. [[CrossRef](#)]
24. Papale, D.; Black, T.A.; Carvalhais, N.; Cescatti, A.; Chen, J.; Jung, M.; Kiely, G.; Lasslop, G.; Mahecha, M.D.; Margolis, H.; et al. Effect of spatial sampling from European flux towers for estimating carbon and water fluxes with artificial neural networks. *J. Geophys. Res. Biogeosci.* **2015**, *120*, 1941–1957. [[CrossRef](#)]
25. Périé, C.; de Blois, S. Dominant forest tree species are potentially vulnerable to climate change over large portions of their range even at high latitudes. *Peer J.* **2016**, *4*, e2218. [[CrossRef](#)] [[PubMed](#)]
26. Pouteau, R.; Meyer, J.Y.; Taputuarai, R.; Stoll, B. Support vector machines to map rare and endangered native plants in Pacific islands forests. *Ecol. Inform.* **2012**, *9*, 37–46. [[CrossRef](#)]
27. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013.
28. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **2006**, *9*, 181–199. [[CrossRef](#)]

29. Shen, C. Climate-Sensitive Site Index Model of *Larix olgensis* Henry. Master's Thesis, Chinese Academy of Forestry, Beijing, China, 2012. (In Chinese)
30. Breiman, L. Random Forest. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
31. Liaw, A.; Wiener, M. Classification and Regression by random Forest. *R. News* **2002**, *2*, 18–22.
32. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015; Available online: <http://www.r-project.org> (accessed on 21 April 2017).
33. Freeman, E.A.; Moisen, G.G.; Coulston, J.W.; Wilson, B.T. Random forests and stochastic gradient boosting for predicting tree canopy cover: Comparing tuning processes and model performance1. *Can. J. For. Res.* **2016**, *46*, 323–339. [[CrossRef](#)]
34. Zhang, R.; Pang, Y.; Li, Z.; Bao, Y. Canopy closure estimation in a temperate forest using airborne LiDAR and LANDSAT ETM+ data. *Chin. J. Plant Ecol.* **2016**, *40*, 102–115, (In Chinese with English abstract).
35. Kuhn, M. Classification and Regression Training. Available online: <https://cran.r-project.org/web/packages/caret/caret.pdf> (accessed on 20 November 2018).
36. Mansiaux, Y.; Carrat, F. Detection of independent associations in a large epidemiologic dataset: A comparison of random forests, boosted regression trees, conventional and penalized logistic regression for identifying independent factors associated with H1N1pdm influenza infection. *BMC Med. Res. Methodol.* **2014**, *14*, 9. [[CrossRef](#)] [[PubMed](#)]
37. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [[CrossRef](#)]
38. Ridgeway, G. Generalized Boosted Models: A Guide to the gbm Package. Available online: <https://cran.r-project.org/web/packages/gbm/vignettes/gbm.pdf> (accessed on 14 September 2018).
39. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [[CrossRef](#)] [[PubMed](#)]
40. Kuhn, M.; Weston, S.; Keefer, C.; Coulter, N.; Quinlan, R.; Rulequest Research Pty Ltd. Rule- and Instance-Based Regression Modeling. Available online: <https://cran.r-project.org/web/packages/Cubist/Cubist.pdf> (accessed on 21 May 2018).
41. Sun, B.; Lam, D.; Yang, D.; Grantham, K.; Zhang, T.; Mutic, S.; Zhao, T. A machine learning approach to the accurate prediction of monitor units for a compact proton machine. *Med. Phys.* **2018**, *45*, 2243–2251. [[CrossRef](#)] [[PubMed](#)]
42. Ferlito, S.; Adinolfi, G.; Graditi, G. Comparative analysis of data-driven methods online and offline trained to the forecasting of grid-connected photovoltaic plant production. *Appl. Energy* **2017**, *205*, 116–129. [[CrossRef](#)]
43. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Stat.* **1991**, *19*, 1–67. [[CrossRef](#)]
44. Lee, T.S.; Chiu, C.C.; Chou, Y.C.; Lu, C.J. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Comput. Stat. Data Anal.* **2006**, *50*, 1113–1130. [[CrossRef](#)]
45. Heddam, S.; Kisi, O. Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *J. Hydrol.* **2018**, *559*, 499–509. [[CrossRef](#)]
46. Milborrow, S. Multivariate Adaptive Regression Splines. Available online: <https://cran.r-project.org/web/packages/earth/earth.pdf> (accessed on 3 January 2019).
47. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009.
48. Smoliński, S.; Radtke, K. Spatial prediction of demersal fish diversity in the Baltic Sea: Comparison of machine learning and regression-based techniques. *ICES J. Mar. Sci.* **2017**, *74*, 102–111. [[CrossRef](#)]
49. Zhou, Z. *Machine Learning*; Tsinghua University Press: Beijing, China, 2016. (In Chinese)
50. Breiman, L. Statistical modeling: The two cultures. *Stat. Sci.* **2001**, *16*, 199–231. [[CrossRef](#)]
51. Moisen, G.G.; Frescino, T.S. Comparing five modelling techniques for predicting forest characteristics. *Ecol. Model.* **2002**, *157*, 209–225. [[CrossRef](#)]
52. Aertsen, W.; Kint, V.; Orshoven, J.V.; Özkan, K.; Muys, B. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecol. Model.* **2010**, *221*, 1119–1130. [[CrossRef](#)]
53. Leathwick, J.R.; Elith, J.; Francis, M.P.; Hastie, T.; Taylor, P. Variation in demersal fish species richness in the oceans surrounding New Zealand: An analysis using boosted regression trees. *Mar. Ecol. Prog. Ser.* **2006**, *321*, 267–281. [[CrossRef](#)]

54. Adame, P.; Hynynen, J.; Cañellas, I.; del Río, M. Individual-tree diameter growth model for rebollo oak (*Quercus pyrenaica* Willd.) coppices. *For. Ecol. Manag.* **2008**, *255*, 1011–1022. [[CrossRef](#)]
55. 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](#)]
56. Alam, S.A.; Huang, J.G.; Stadt, K.J.; Comeau, P.G.; Dawson, A.; Geaizquierdo, G.; Aakala, T.; Hölttä, T.; Vesala, T.; Mäkelä, A.; et al. Effects of competition, drought stress and photosynthetic productivity on the radial growth of white Spruce in Western Canada. *Front. Plant Sci.* **2017**, *8*, 1915. [[CrossRef](#)]
57. Lo, Y.H.; Blanco, J.A.; Seely, B.; Welham, C.; Kimmins, J.P. Relationships between climate and tree radial growth in interior British Columbia, Canada. *For. Ecol. Manag.* **2010**, *259*, 932–942. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).