

PREDICTING THE HEIGHT BASED ON THE MAIN INFLUENCE FACTORS FOR DAHURIAN LARCH (*LARIX GMELINII*) IN THE GREAT KHING'AN MOUNTAINS, CHINA

ZHANG, C.¹ – LIU, Y.¹ – WU, Y. H.¹ – HUA, J.² – TIE, N.^{3*}

¹College of Forestry, Inner Mongolia Agriculture University,
No. 275, Xueyuan East Street, Hohhot, Inner Mongolia, China
(phone: +86-199-8314-3900)

²Inner Mongolia Daqingshan National Nature Reserve
No. 6, Guishu Street, Hohhot, Inner Mongolia, China
(e-mail: dqsxch@126.com)

³Forestry and Grassland Bureau of Inner Mongolia Autonomous Region
No. 23, Xueyuan East Street, Hohhot, Inner Mongolia, China
(e-mail: wangtieniu@126.com)

*Corresponding author
e-mail: wangtieniu@126.com; phone: +86-153-3556-8882

(Received 28th Oct 2022; accepted 10th Jan 2023)

Abstract. It is difficult to meet the needs of forest investigation and management in different stands using an h–d model with the diameter at breast height (DBH) as the only variable to predict the height (h). To predict h accurately, we tested 12 frequently used h–d models based on 5083 Dahurian larch (*Larix gmelinii*) in the Great Khing'an Mountains, and constructed a structural equation model (SEM) with 15 factors to verify their effects on h, the main factors were screened as variables to improve the h–d model. The results demonstrated that the Chapman–Richards model was the best predictor of h. The dominant diameter (DD) and total basal area of all trees with a diameter larger than the target tree (LDTBA) had a significant influence on the observed h. By combining them as independent variables with the Chapman–Richards model terms, we found that modified model *Equation 20* reduced the differential value from -0.044 to -0.010 compared with the Chapman–Richards model. The study proved that using SEM to select critical variables is an effective method to improve the accuracy of the h–d model.

Keywords: model fitting, structural equation model, tree size effects, stand and site quality effects, competition effects

Introduction

Diameter at breast height (DBH) and tree height (h) are the two most critical parameters used to estimate stand volume, biomass, carbon, and other derived parameters in the forest inventory, which are frequently used in forest operation and management (Colbert et al., 2002; Mehtätalo et al., 2015; Schmidt et al., 2018). There is an allometric correlation between h and DBH (Mehtätalo et al., 2015; Mayer, 1936), and forest managers use this relationship to predict h. Typically, using a simple operation and portable equipment can achieve a sufficiently accurate measurement of DBH, whereas the measurement of h is difficult to obtain because the observers' sight is obstructed by the forest canopy. Additionally, the equipment is expensive (Colbert et al., 2002; Adame et al., 2008; Ng'andwe et al., 2019).

At present, many functions have been developed to predict h, including linear functions and nonlinear functions (Bronisz and Mehtätalo, 2020). Nonlinear functions have been used more widely than linear functions because the growth of the diameter and h are always in an

allometric relationship, which is consistent with the basic biological principles of biological growth (Huang et al., 2000). Particularly in natural forests, trees have different ages and crown types, and their h–d relationship varies more than those in plantations (Temesgen et al., 2014). The models typically use an exponential function, power function, and logistic function (Huang et al., 1992). Although the forms of the h–d model differ substantially, they can be divided into two basic types according to the independent variables (Lei et al., 2009): in one, it is assumed that h is completely dependent on the DBH, and in the other, it is considered that the independent variables include DBH and other stand level or individual plant level variables, such as age, site index, stand basal area, stand density, and dominant tree height. These models are also called composite models (Bruchwald and Wróblewski, 1994; Soares and Tomé, 2002; Newton and Amponsah, 2007). The above two types of models whose parameter estimation are sufficiently simple to fit the size distribution for different shapes and skewness have been widely used by scholars and practitioners (Liu et al., 2014).

The influence of trees themselves and their environment on the growth process is complex in nature. Such influencing factors are broadly categorized as structure, site, density, and competition factors according to their characteristics, etc. (Wang et al., 2015). These influence factors restrict or promote each other directly or indirectly. For the prediction of h, to select the main influence factors from many of them, it is necessary to make the simultaneous action clear of multiple factors in the influence system to represent the complexity of the system (Doncaster, 2007). At present, to find the h and its related variables, most of the methods used correlation analysis, or use a dummy variable and quantile model to study the classification of h; few of them determine effective variables based on the interaction of candidate variables, and these studies have failed to empirically examine the integral influence of factors. How do the influencing factors affect each other, and what are the relationships among the different categories to which they belong? Structural equation model (SEM) is a new method to find effective variables based on factors relationship. Therefore, the objectives of this study are as follows: (1) to evaluate whether the commonly used model for larch can accurately fit the DBH and h of Dahurian larch in the Great Khing'an Mountains; (2) by considering the interrelated influence factors, which of them have an important influence on h; and (3) to develop and test a new h–d model for Dahurian larch based on factors classification and relationship of SEM.

Materials and methods

Study site

The data for this study were collected from the north of the Great Khing'an Mountains (121°30'–121°32'E, 50°53'–50°57'N), the altitude was 810–1160 m, the annual average precipitation was 450–550 mm, and the annual average temperature was below –5.4 °C, which belongs to the cold temperate continental monsoon climate. Dahurian larch is not only a dominant species in the Great Khing'an Mountains of Inner Mongolia, but also the dominant tree species in the cold temperate zone of North China. Its stock volume accounts for more than 78% of the forest land (Wang et al., 2012).

Modeling data was from 18 permanent sample plots with an area of 0.16–0.25 ha set up in a typical area in the Chaocha forest farm of the Great Khing'an Mountains, which included 8 primary forest plots and 9 natural secondary forest plots. Additionally, there was 1 primary forest sample plot with an area of 1 ha. To make the investigation easier, each sample was divided into a square sample plot with a side length of 10 m (Fig. 1).

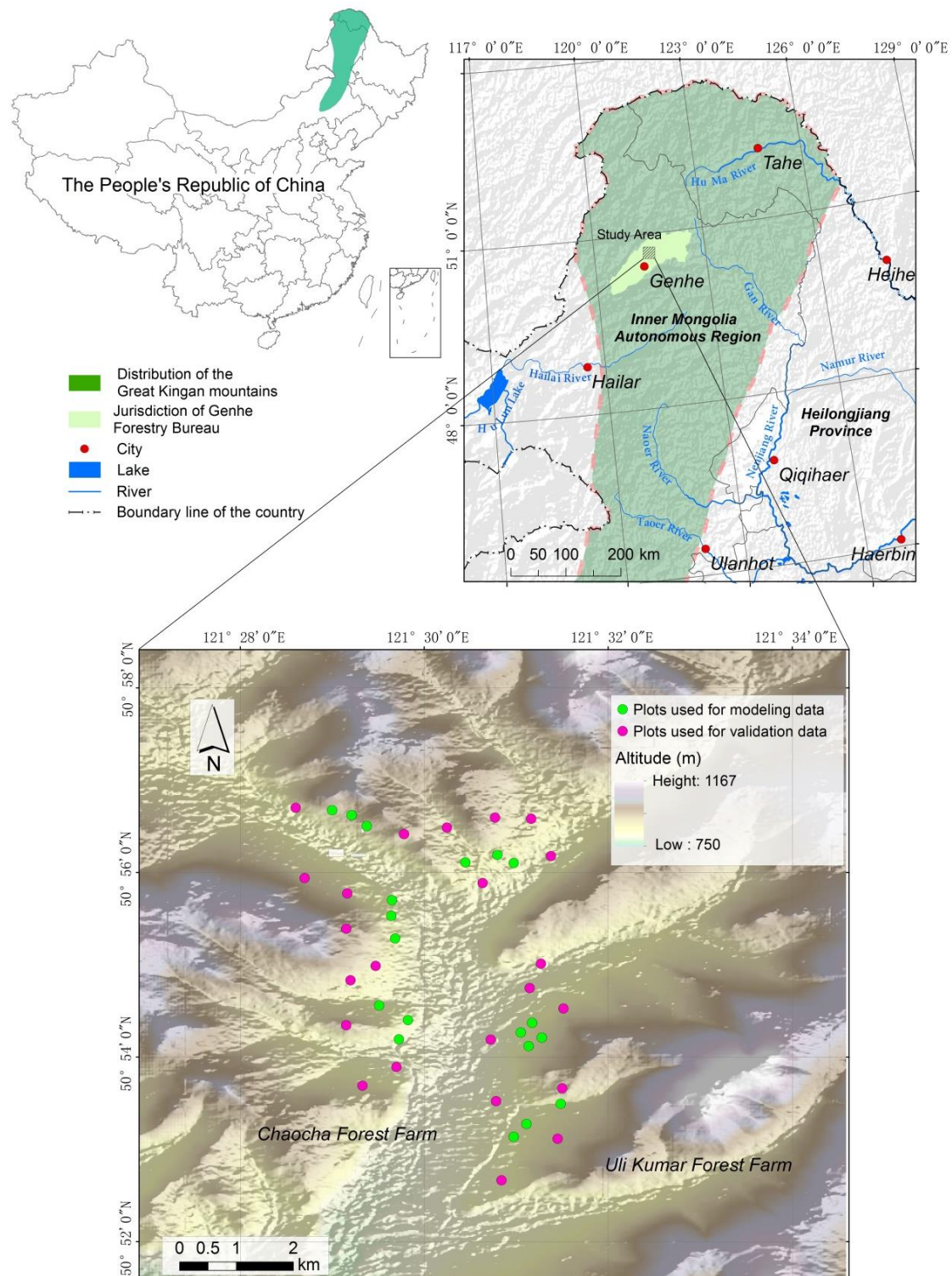


Figure 1. Location of the research area and the distribution of the plots

The DBH and h data were obtained from 5083 Dahurian larch, the information of trees and sites are in *Table 1*. The diameter at 1.3 m (DBH) was measured using a tree measuring tape and electronic caliper, and the h was measured using a Hagl f altimeter. The investigation team was composed of 10 statisticians with forestry investigation

experience. Dahurian larch with a DBH greater than or equal to 5 cm was measured. It is generally accepted in forestry that the dataset used for model verification should preferably be independent of the development data (Bohora and Cao, 2014). To accurately verify the established model, the independent datasets for verification came from 24 complete sample plots with an area of 0.01–0.25 ha, which contained a total of 906 Dahurian larch (Table 1).

Table 1. Characteristics of trees and sites

		H (m)	DBH (cm)	BA (cm ²)	DH (m)	DD (cm)	EN (m)	PD (n/ha)
Fitting data N = 5083	Mean	10.1	9.2	122.1	16.16	18.9	829	2100
	Max	31.9	56.8	2532.7	39.0	41.0	883	3500
	Min	3.1	5	19.6	5.47	7.9	813	400
	Std.	5.5	7.3	246.6	7.7	9.6	24.4	884.8
Verification data N = 906	Mean	10.0	9.3	114.9	15.7	18.2	830	2300
	Max	27.9	49.5	1923.4	28.4	36.4	876	3500
	Min	3.8	5	19.62	7.1	8.3	817	700
	Std.	4.0	6.8	218.4	6.3	8.5	22.5	653.1

BA is the basal area; Std. is the standard deviation; DH is the dominant height of the plot; DD is the dominant diameter of the plot; EN is the elevation; PD is the number of trees per hectare

Candidate *h*–*d* model

This step is to preliminary screening the most appropriate model by evaluating the accuracy, precision, and reliability for Dahurian larch whether they come from primary forest or secondary forest. Candidate models for predicting *h* include the Curtis model, Wykoff model, and Näslund model etc. The root mean square error (RMSE) (Eq. 1) is used to express the degree of dispersion between the model predicted *h* and the measured *h*. The mean absolute percentage error (MAPE) (Eq. 3) is used as the accuracy index, which is based on the absolute percentage error (APE) (Eq. 3) (Sileshi, 2014; Yao et al., 2013). The overall model prediction accuracy (MPA) (Eq. 5) is the precision that combines the reliability of the mean deviation (Eq. 4) and deviation change. The measured *h* is compared with the predicted *h* generated by each model (Huang et al., 2003). The model is considered to be highly reliable when the MAPE is less than 10%; acceptable when the MAPE is 10%–20% (Huang et al., 2003); and unreliable when it is greater than 20%. The parameters are calculated as follows:

$$RMSE = \sqrt{1/(n-k) \sum_{i=1}^n (h_i - h'_i)^2} \quad (Eq.1)$$

$$APE_i = \frac{|h_i - h'_i|}{h_i} \times 100 \quad (Eq.2)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|(h_i - h'_i)|}{h_i} \quad (Eq.3)$$

$$\bar{\varepsilon} = \sum_{i=1}^n \frac{h_i - h_i'}{h_i} \quad (\text{Eq.4})$$

$$\text{MPA} = \bar{\varepsilon}^{-2} + \text{SD}^2 \quad (\text{Eq.5})$$

where k is the number of fixed model parameters; MAPE is the arithmetic average of the sum of the APEs divided by the number n ; h_i is the measured h of the i -th tree; h_i' is the predicted h of the i -th tree; $\bar{\varepsilon}$ is the average prediction deviation, which reflects the deviation between the predicted h of the model and the measured h ; SD is the standard deviation of the prediction deviation.

To compare the predicted h with the measured h , Welch's t-test is selected, which can correct samples with unequal variance (Welch, 1938). The Honest Significant Difference (HSD) method of analysis of variance (ANOVA) is used to compare the predictions of multiple h - d models.

In a comparison of the prediction ability of the model, the model with low MAPE and MPA is preferred. In previous studies, high R^2 , low MAPE, and low MPA were used as the basic criteria for screening models (Huang et al., 2003; Anitha et al., 2015). Although R^2 has been widely used to evaluate the goodness of fit of linear models (Dorado et al., 2006; Chai et al., 2018; Lumbres et al., 2013; Huang et al., 2000), its value sometimes exceeds the range (0–1) when evaluating nonlinear models, and the total variance is also affected, which increases the type II error (Quinn and Keough, 2002; Sokal et al., 1985; Spiess and Neumeyer, 2010). In this study, the models used to fit h - d are all nonlinear models, so R^2 is not selected as the evaluation index (Table 2).

Table 2. Frequently used theoretical functions in the development of the h - d models for Dahurian larch

No.	Function name	K	Function	References
1	Curtis	2	$h = 1.3 + \beta_1 / (1 + d^{-1})^{\beta_2}$	Curtis, 1976
2	Wykoff	2	$h = 1.3 + e^{\beta_1 + \beta_2 / (d+1)}$	Dorado, 2006
3	Näslund	2	$h = 1.3 + (d / (\beta_1 + \beta_2 d))^2$	Laiho et al., 2014
4	Meyer	2	$h = 1.3 + \beta_1 (1 - e^{-\beta_2 d})$	Meyer, 1940
5	Schumacher	2	$h = 1.3 + \beta_1 e^{\beta_2 / d}$	Schumacher, 1939; Bronisz and Mehtätalo, 2020
6	Modified Logistic	3	$h = 1.3 + \beta_1 / (1 + \beta_2^{-1} d^{-\beta_3})$	Lumbres et al., 2013
7	Hossfeld	3	$h = 1.3 + \beta_1 d / (\beta_2 + \beta_3 d^{\beta_1})$	Sharma, 2010
8	Strand	3	$h = 1.3 + d^2 / (\beta_1 + \beta_2 d + \beta_3 d^2)$	Bronisz and Mehtätalo, 2020
9	Weibull	3	$h = 1.3 + \beta_1 (1 - e^{-\beta_2 d^{\beta_3}})$	Huang et al., 1992
10	Zeide	3	$h = 1.3 + \beta_1 e^{-\beta_2 d^{-\beta_3}}$	Zeide, 1989
11	Ratkowsky	3	$h = 1.3 + \beta_1 e^{\beta_2 / (d + \beta_3)}$	Weiss and David, 1990
12	Chapman Richards	3	$h = 1.3 + \beta_1 (1 - e^{-\beta_2 d})^{\beta_3}$	Lumbres et al., 2013

β_1 , β_2 , and β_3 are the parameters of the functions; h is the tree height (m), d is the DBH (cm); and K is the number of parameters

Building a SEM using influence factors

The growth process of trees on the one hand follows their own biological characteristics and stand conditions, such as diameter and plot arithmetic mean diameter; on the other hand it is disturbed by external biological and non-biological factors, such as plant density, elevation, and number of trees with a diameter larger than the target tree. According to previous experience, the factors that represent the above effects are classified as tree size effects, standard and site quality effects, and competition effects (Fu et al., 2016; Wykoff et al., 1982; Monserud and Sterba, 1996), with a total of 15 influence factors (*Table 3*).

Correlation analysis is the most commonly used method to screen the relationship between influence factors and *h*, however, it can only explain the direct influence of relevant variables, because this method does not consider the mutual constraint or promotion relationship among factors. Therefore, there are limitations in explaining the complex effects of many factors on *h*. To control different influence factors simultaneously, a SEM based on the maximum likelihood (ML) can be used as an effective method to process the complex relationship between the influence factors and *h*. It can focus not only on the direct influence of the factors on *h* but also the indirect impact of the influence factors on *h* through other factors, and find the maximum value of the joint probability of continuous sample observations (Lee and Kim, 2018). Amos 21.0 (developed by IBM SPSS Statistics) is used to construct the structural equation path of the relationship between influence factors and *h*.

Table 3. *Candidate variables that affect the relationship between *h* and DBH*

Group	Variables
Tree size effects	Plot arithmetic mean diameter (AMD) Plot quadratic mean diameter (QMD) Plot dominant tree diameter (DD) Plot dominant tree height (DH)
Stand and site quality effects	Canopy density (CD) Plant density (PD) Slope (SE) Elevation (EN)
Competition effects	Total basal area (TBA) Mean basal area (MBA) Number of trees with diameter larger than a target tree (LDN) Mean diameter of all trees with diameter larger than the target tree (LDMD) Total diameter of all trees with diameter larger than the target tree (LDTD) Mean basal area of all trees with diameter larger than the target tree (LDMBA) Total basal area of all trees with diameter larger than the target tree (LDTBA)

According to whether the influence factors can be measured, they are defined as observed variables and latent variables (Grace, 2006). The SEM is a confirmatory model, which cannot be used to define causality. The establishment path must be based on the theoretical background, logical hypothesis, or research design (Sobel, 2008). Tree size is generally considered to be closely coordinated with changes during tree growth (Mackenzie et al., 1997; Seifert et al., 2014), but it is moderately limited by competition, and forest stand and site quality. According to this relationship, stand and site quality factors are considered as exogenous variables; *h* is considered as an

endogenous variable. Tree size factors and competition factors are considered as mediator variables because they are affected by stand and site quality factors, and can directly affect the h. Moreover, tree size factors are also constrained using competition factors. The model is shown in *Figure 2*.

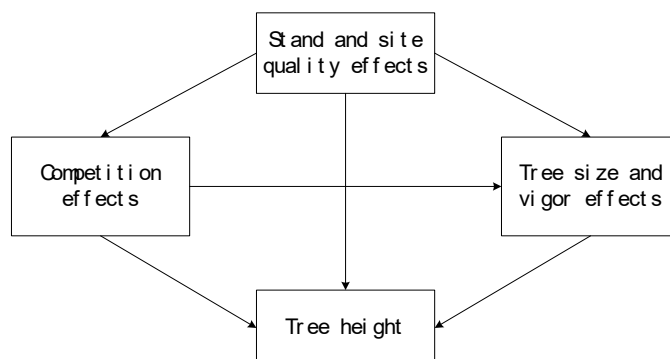


Figure 2. SEM based on the influence relationship

The fitness and accuracy test results of the SEM are evaluated using the following indicators, they are provided automatically after the program runs:

$$\text{RMSEA} = \sqrt{\frac{F_0}{df}} = \sqrt{\max\left(\frac{F_{ML}}{df} - \frac{1}{N-1}, 0\right)} \quad (\text{Eq.6})$$

$$\text{GFI} = 1 - \frac{t_r \left[\sum^{-1} (S - \Sigma) \right]^2}{t_r \left(\sum^{-1} - S \right)^2} \quad (\text{Eq.7})$$

$$\text{AGFI} = 1 - (1 - \text{GFI}) \left[\frac{k(k+1)}{2df} \right] \quad (\text{Eq.8})$$

$$\text{NFI} = \frac{\chi_n^2 - \chi_t^2}{\chi_n^2} \quad (\text{Eq.9})$$

$$\text{TLI} = \left(\frac{\chi_n^2}{df_n} - \frac{\chi_t^2}{df_t} \right) / \frac{\chi_n^2}{df_n - 1} \quad (\text{Eq.10})$$

$$\text{IFI} = \frac{\chi_n^2 - \chi_t^2}{\chi_n^2 - df_n} \quad (\text{Eq.11})$$

$$\text{CFI} = \frac{(\chi_n^2 - df_n) - (\chi_t^2 - df_t)}{\chi_n^2 - df_n} \quad (\text{Eq.12})$$

$$\text{PNFI} = \text{NFI} \frac{df_{pro}}{df_n} \quad (\text{Eq.13})$$

$$\text{PGFI} = \text{GFI} \frac{df}{0.5m(m+1)} \quad (\text{Eq.14})$$

$$\text{PCFI} = \text{CFI} \frac{d_j}{d_d} \quad (\text{Eq.15})$$

where RMSEA is the RMSE of the approximation, GFI is the goodness-of-fit index, AGFI is the adjusted GFI, NFI is the normed fit index, IFI is the incremental fit index, TLI is the Tucker–Lewis index, CFI is the comparative fit index, PNFI is the parsimony-adjusted NFI, PGFI is the parsimony-adjusted GFI, PCFI is the parsimony-adjusted CFI, F_0 is the total difference function value, F_{ML} is the adaptation function, n is the number of samples, t_r is the sum of the diagonal elements in the matrix, Σ^{-1} is the sample covariance matrix, Σ is the hidden covariance matrix of the model, k is the number of model variables, χ^2_n and χ^2_t are the virtual model and hypothesis model, respectively, S is the observation matrix, df is the degrees of freedom, df_{pro} is the adaptation function, m is the number of observation variables, d_j and d_d are the test model and independent model, respectively. Using the observed variable covariance matrix to estimate the unstandardized regression coefficients of the model parameters, and the unregression coefficients of each latent variable and one of the measured variables set to 1, these paths need not be tested for the significance of the path coefficients. The critical ratio is the ratio between the estimated value of the parameter and the standard error. The path that does not meet the inspection requirements is deleted to complete the adjustment of the path analysis. Finally, the influence of an independent variable on a dependent variable is expressed using the regression coefficient.

Adjustment of the h–d model

Two factors that have a significant impact on the change of h are selected as independent variables, and the necessary combinations and transformations are detected in the h–d model. The parameters of the new h–d model are obtained by fitting a subset of the data. The performance of the new model is evaluated using the MAPE (Eq. 3) and MRE (Eq. 5). Using Welch's t-test, the means of the predicted h and measured h are also compared following Bartlett's homogeneity test for variance.

Result

Performance of frequently used h–d models in the Great Khing'an Mountains

The parameters and fitting evaluation of 12 frequently used nonlinear functions are shown (Table 4). The fitting results of β_1 , β_2 , and β_3 of all functions were significant ($P < 0.05$). The fitting curves of all models were drawn in the trend of the h scatter distribution (Fig. 3). The table shows that the fitting results of the three-parameter function were generally better than those of the two-parameter function (Table 4).

Table 4. Estimated parameters and their associated statistic fits for each *h–d* model for Dahurian larch

Function name	Parameter estimates			Modeling data						Validation data					
	β_1	β_2	β_3	RMSE	Rk	MAPE	Rk	MPA	Rk	RMSE	Rk	MAPE	Rk	MPA	Rk
Curtis	22.727(0.144)	9.627(0.072)		1.807	11	12.75	11	12.24	10	0.825	11	2.93	1	9.72	11
Wykoff	3.154(0.006)	-10.533(0.077)		1.748	10	12.38	10	12.02	9	0.811	10	3.04	6	9.45	10
Näslund	1.433(0.011)	0.195(0.001)		1.729	9	11.80	9	11.50	8	0.780	8	3.04	5	8.79	9
Meyer	23.439(0.254)	0.045(0.001)		1.700	8	11.66	5	13.32	12	0.763	1	3.02	3	8.77	8
Schumacher	22.060 (0.139)	-8.783(0.068)		1.834	12	13.01	12	12.47	11	0.840	12	3.14	11	10.00	12
Modified Logistic	40.213(2.490)	0.030(0.001)	-0.946(0.025)	1.688	3	11.60	3	10.27	3	0.775	4	3.04	7	7.65	4
Hossfeld	0.846(0.048)	0.967(0.066)	0.044(0.006)	1.686	2	11.71	8	10.31	5	0.788	9	3.17	12	7.56	2
Strand	-0.123(0.151)	0.924(0.025)	0.028(0.001)	1.689	4	11.59	1	10.30	4	0.777	5	3.03	4	7.70	5
Weibull	29.645(1.496)	0.044(0.001)	0.875(0.018)	1.690	5	11.67	6	10.22	2	0.777	6	3.06	10	7.58	3
Zeide	115.399(22.087)	-5.044(0.105)	0.286(0.024)	1.684	1	11.65	4	10.62	7	0.771	2	3.05	8	7.76	6
Ratkowsky	32.337(0.673)	-23.635(0.994)	7.519(0.440)	1.693	7	11.71	7	10.47	6	0.772	3	3.06	9	7.76	7
Chapman Richards	28.162(1.100)	0.027(0.002)	0.849(0.020)	1.693	6	11.60	2	10.16	1	0.780	7	3.02	2	7.55	1

The standard deviation is in brackets

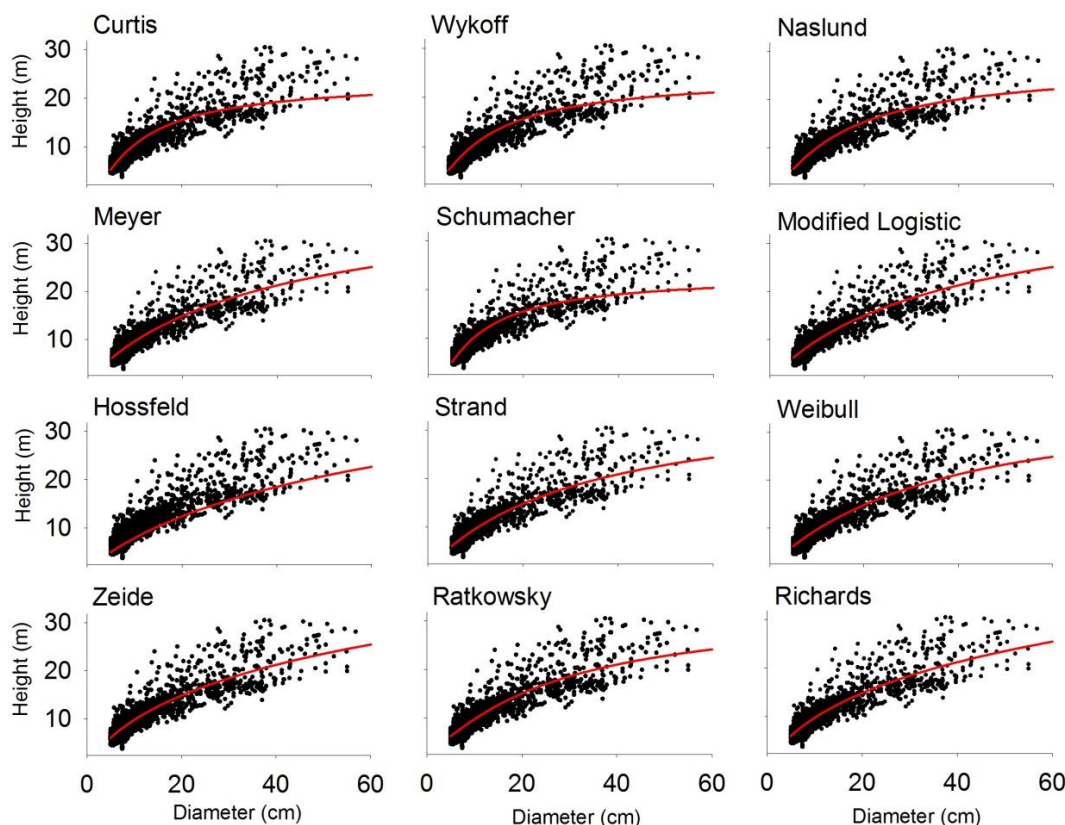


Figure 3. Fitting of the frequently used models on the actual data for Dahurian larch in the Great Khing'an Mountains

For example, the minimum value of the RMSE for the two-parameter function in the modeling data stage was 1.700 (Meyer) and the maximum was 1.834 (Schumacher), whereas the minimum value of RMSE for the three-parameter function was 1.684 (Zeide) and the maximum value was 1.693 (Ratkowsky). Although the Curtis function had the best accuracy in the validation data stage (MAPE 2.93), it had a higher RMSE and MPA for the modeling dataset and validation dataset, and the fitting result was not satisfied compared with other functions. The Chapman–Richards model ranked second for accuracy (MAPE 11.60) and first for precision (MPA 10.16) in the modeling data stage. It also had the same reliable performance in the independent validation data stage, that is, accuracy ranked second (MAPE 3.02) and precision ranked first (MPA 7.55). This model can be regarded as the most appropriate function among the commonly used models in research. Therefore, the Chapman–Richards model can be used to explain the relationship between h and DBH in both the primary forest and natural secondary forest of Dahurian larch in the Great Khing'an Mountains.

The expression with the parameters is

$$h = 1.3 + 28.162(1 - e^{-0.027 \cdot d})^{0.849} \quad (\text{Eq.16})$$

Parameter test and estimation of the SEM

The fitness test based on information theory was used to evaluate the degree of consistency between the path analysis model and the observation data. The RMSEA

(0.073) (Table 5) was between 0.05 and 0.08, which indicates that the model fits well (Lai and Green, 2016; Browne and Cudeck, 1992). The value of the GFI (0.946) and AGFI (0.912) were greater than standard value 0.90, indicating the model passed the goodness of fit test. The value of NFI (0.933), IFI (0.935), TLI (0.926) and CFI (0.934) were all greater than standard value 0.90, which indicates that the model support the good fitting accuracy. The values of the PNFI (0.776), PGFI (0.778), and PCFI (0.791) were all greater than standard value 0.5, which indicates that the model performed well. All Indicators met the requirement.

Table 5. Model accuracy fitness and test table

Indicators	Standard	Value	Indicators	Standard	Value
RMSEA	< 0.08	0.073	TLI	> 0.90	0.926
GFI	> 0.90	0.946	CFI	> 0.90	0.934
AGFI	> 0.90	0.912	PNFI	> 0.50	0.776
NFI	> 0.90	0.933	PGFI	> 0.50	0.778
IFI	> 0.90	0.935	PCFI	> 0.50	0.791

The critical ratio is the ratio between the estimated value of the parameter and the standard error. The absolute value was greater than 2.58, and the P-values were all less than 0.01, which indicates that the observed variables all passed the significance test.

The SEM used the ML estimation method to estimate the regression coefficients of the model parameters. The standardized path coefficient represents the correlation between endogenous variables and exogenous variables (Table 6).

Table 6. Standardized regression coefficients of the latent variable to h and the observation variable to the latent variable

Endogenous variables	Latent variables	Total effects	Observed variables	Total effects	Critical ratio	P
Tree height	Tree size effects	0.330	AMD	0.742		
			QMD	0.626	30.659	**
			DD	0.924	53.051	**
			DH	0.915	47.833	**
	Stand and site quality effects	0.300	CD	0.870		
			PD	-0.317	-14.848	**
			SE	0.318	56.872	**
			EN	0.573	28.916	**
	Competition effects	0.213	TBA	0.971		
			MBA	0.297	14.829	**
			LDN	-0.061	-2.905	**
			LDTD	0.618	36.705	**
			LDMD	0.823	65.529	**
			LDTBA	0.922	231.057	**
			LD MBA	0.619	36.959	**

**is significant codes: $P < 0.01$

Among all the latent variables, tree size effects had the greatest impact on *h* (path coefficient 0.330), followed by stand and site quality effects (0.300). The impact of competition (0.213) on *h* was significantly smaller than other effects, which indicates that *h* was mainly affected by tree size, and stand and site quality. Among the influence factors that represent the size of trees, the dominant diameter (DD) had the largest path coefficient (0.924), which indicates that the DD fully reflected the size of the trees in the stand. Among the influence factors that represent the forest stand and site quality, the path coefficient of CD was the largest (0.870), and the path coefficients of other variables were significantly smaller than that of CD (−0.317–0.573), which indicates that CD was the dominant influence factor in the relationship between stand and site quality affecting *h*. Among the influence factors that represent competition, the path coefficient of the diameter larger than the target tree (LDTBA) was the largest (0.922), which indicates that LDTBA was the most representative of all factors that characterized competition. When selecting the new variables added by the *h*–*d* model, the simplicity of the model should be ensured so that it can be used in practical work. Meanwhile, the variables that can be measured easily and have high accuracy should be considered first; hence, the DD variable and LDTBA variable must be chosen (Fig. 4).

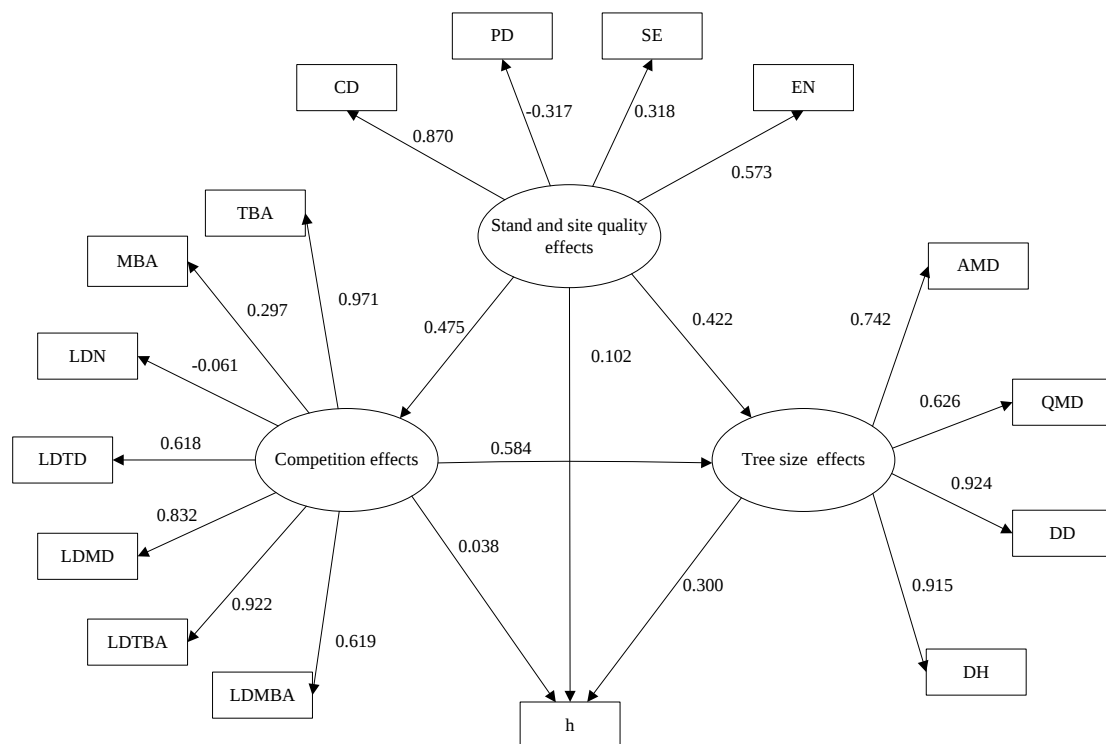


Figure 4. SEM for the influence factors

Model development

To improve the accuracy and precision of model fitting, a combination of DD variable and LDTBA variable was tested at different parameter positions. The formulas are as follows:

$$h = 1.3 + (\beta_1 + \beta_4 * DD + \beta_5 * LDTBA)(1 - e^{-\beta_2 * d})^{\beta_3} \quad (\text{Eq.17})$$

$$h = 1.3 + (\beta_1 + \beta_4 * DD)(1 - e^{-(\beta_2 + \beta_5 * LDTBA) * d})^{\beta_3} \quad (\text{Eq.18})$$

$$h = 1.3 + (\beta_1 + \beta_4 * DD)(1 - e^{-\beta_2 * d})^{(\beta_3 + \beta_5 * LDTBA)} \quad (\text{Eq.19})$$

$$h = 1.3 + (\beta_1 + \beta_5 * LDTBA)(1 - e^{-1 * (\beta_2 + \beta_4 * DD) * d})^{\beta_3} \quad (\text{Eq.20})$$

$$h = 1.3 + (\beta_1 + \beta_5 * LDTBA)(1 - e^{-\beta_2 * d})^{(\beta_3 + \beta_4 * DD)} \quad (\text{Eq.21})$$

$$h = 1.3 + \beta_1(1 - e^{-1 * |\beta_2 + \beta_4 * DD + \beta_5 * LDTBA| * d})^{\beta_3} \quad (\text{Eq.22})$$

$$h = 1.3 + \beta_1(1 - e^{-\beta_2 * d})^{(\beta_3 + \beta_4 * DD + \beta_5 * LDTBA)} \quad (\text{Eq.23})$$

The parameter estimates of the modified model (Eqs. 17–23) were all significant ($P < 0.05$), RMSE (1.451–1.707), MAPE (11.02–12.15), and MPA (9.26–12.38). These values are within the acceptable thresholds for reliable models that accurately predict h . Only the fitting results of Equation 22 (RMSE1.707, MAPE12.15, MPA12.38) were all less than Chapman Richards model (RMSE1.693, MAPE11.60, MPA10.16) (Table 7), because the combination of variables in the defined domain made the value of Equation 22 model less than zero, thereby resulting in invalid values and narrowing the fitting sample. Additionally, the fitting performance of the Chapman–Richards model was not the worst; its accuracy (MAPE11.60) and precision (MPA10.16) were better than those of Equation 22 (MAPE12.15, MPA12.38) and Equation 23 (MAPE11.73, MPA10.78), which indicates that the predictive ability of the model is not the more variables, the better, but determined by both the number and the location of variables. Among all the models, the predicted h in Equation 19 and observed h had the lowest degree of dispersion (RMSE1.693), and the accuracy and precision were the best (MAPE11.02, MPA9.26), which indicates that Equation 19 was the best choice for predicting h . The difference between the predicted mean and the observed mean of Equation 19 (−0.010) was closer to zero than that of the other models (−0.059–0.775), which indicates that the predicted value of Equation 19 was closer to the observed value. Comparing Equation 19 with the Chapman–Richards model (0.042), the prediction mean of the Chapman–Richards model was overestimated, which indicates that Equation 19 had a significantly better prediction effect. This model had higher prediction efficiency, whereas other models overestimated or underestimated the results to varying degrees. Finally, Equation 19 was determined to be the best model for predicting h .

Discussion

We tested two-parameter models, including Curtis and Wykoff etc, and three-parameter models, including modified logistic and Hossfeld etc. Based on the analysis of the functional performance standards, the prediction results should consider the precision, accuracy, and biological interpretation (Huang et al., 1992; Misir, 2010; Stankova and Diéguez-Aranda, 2013). Among frequently used models, the Chapman–

Richards model was the best fitting, and its RMSE, MAPE, and MPA demonstrated good performance compared with the others. Although the fitting evaluation was sufficiently good in some other models, for example, the RMSE of the Zeide model was the smallest, however, on considering MAPE and MPA, this model was not the most appropriate choice. After comprehensive consideration, the Chapman–Richards model was selected for Dahurian larch in the Great Khing'an Mountains (Enzinga and Jiang, 2019). Also, the Chapman–Richards model curve reflected the biological characteristics of h and DBH (Enzinga and Jiang, 2019; Sharma et al., 2019; Ahmadi et al., 2013), and had a wide application for predicting the growth of larch. Dai et al. developed the h of the larch model in Northeast China and found that the Chapman–Richards model was one of the most appropriate models among the 12 classical models (Dai and Jiang, 2015). Generally, the fitting effect of the three-parameter h – d model is better than that of the two-parameter model. However, studies have shown that the number of parameters is not the critical determinant. For example, on the study of predicting planted and natural pine forests in China, two-parameter models (Näslund and Curtis) and three-parameter models (Chapman–Richards and Hossfeld) have the same fitting effect (Chai et al., 2018). In the establishment of a variety of pine growth models in Angola, a two-parameter model (Näslund) performed well (Delgado and Pukkala, 2013). This may have been caused by the difference in stands. Therefore, to improve the prediction ability of the model, it is more important to find the key variables by using scientific analytical methods. When models have similar prediction capabilities, model with less parameters and variables is easier to use because it has more concise model expression.

Table 7. Estimated parameters and their associated statistic fits for the modified h – d models for Dahurian larch in the Great Khing'an Mountains

Equation	β_1	β_2	β_3	β_4	β_5	RMSE	MAPE	MPA	Predicted h	Observed h	Differential value
Chapman Richards	28.162	0.027	0.849			1.693	11.60	10.16	9.170		-0.044
(18)	19.047	0.058	1.298	0.358	-0.876	1.542	11.13	10.01	9.256		0.042
(19)	10.956	0.124	1.856	0.418	-0.009	1.504	11.04	9.58	9.259		0.045
(20)	12.738	0.084	1.240	0.386	0.191	1.451	11.02	9.26	9.204	9.214	-0.010
(21)	34.383	0.025	1.092	0.0001	-0.401	1.483	11.18	9.67	9.155		-0.059
(22)	32.683	0.037	1.272	-0.006	0.119	1.513	11.29	9.96	9.333		0.119
(23)	30.662	0.030	1.172	0.001	-0.003	1.707	12.15	12.38	9.969		0.755
(24)	31.796	0.041	1.395	-0.015	0.070	1.562	11.73	10.78	9.310		0.096

Many factors affect the h and their influences are complex. It is difficult to define which of these factors is absolutely active or passive. According to previous classifications of influence factors (Wykoff et al., 1982; Walters and Hann, 1986; Monserud and Sterba, 1996; Rijal et al., 2012), they are classified into tree size effects, stand and site quality effects, and competition effects. the authors believe that tree size effects are affected by stand and site quality effects, and competition factors (Mackenzie et al., 1997; Seifert et al., 2014; Sharma and Brunner, 2017). The forest stand and site factors are not affected by other types of factors. The SEM is often used in the fields of engineering and medicine (Chang et al., 2019; Tekce, 2020), however, it is rarely used in the forestry to analyze the relationship between h and its influence factors. This study provides a new perspective to show the influence of variables in variously classified. Among all classifications, tree size proved to be the most influential category by using

SEM. The DD and LDTBA are the main factors that affect the h prediction of Dahurian larch in all the investigated stands. The DD has the greatest impact on h prediction. It is a common index used to express the age and growth quality of the stand and has been used in prediction models to improve their prediction ability. For example, DD was selected as a highly significant virtual variable to establish the h to diameter ratio models of Norway spruce (*Picea abies* (L.) Karst.) and European beech (*Fagus sylvatica* L.) located in the Czech Republic, and provided more accurate predictions for models. It was found that h to diameter ratio decreased with increasing DD (Sharma et al., 2016). Some studies have shown that other variables belong to these classifications, such as stand density and diameter distribution percentiles help to improve the estimation of h (Lei et al., 2009; Eerikäinen, 2003; Temesgen et al., 2008). By using SEM, it is proved that the method of screening important influence factors is effective.

Compared with the modified model, the Chapman–Richards model overestimated the prediction of h in Dahurian larch to a certain extent. The prediction ability of the modified model with influence factor variables was better than that of the original model. In a follow-up study, the DD and LDTBA should be used as variables to predict h together with the DBH because the variables can fully reflect the variability of different sampling sites (Hökkä, 1997). However, this method must be supported by a sufficient sample size (Li et al., 2015). In addition to the influence factors within the forest stand, the reason for the difference in h of Dahurian larch may also be caused by the large-scale clear-cutting of the Great Khing'an Mountains' forests in the 1970s. After the forest was thinned by humans, its spatial structure and age composition changed significantly. A large number of young trees grew and competed simultaneously, which was different from the diameter structure of the original forest. In a future study, if the researchers take into account whether the forest has been disturbed by humans, the accuracy of h prediction may be improved further.

Conclusions

In this study, the Chapman–Richards model was shown to overestimate the prediction of the h of Dahurian larch. This model was reasonable when considering it needs only DBH as input variables. However, if used in a specific forest type or smaller area, the improved model performed better. If considering new variables in the model, it is better to introduce them that related to tree size. Because of different site conditions, tree size, and competition in a Dahurian larch forest, the prediction efficiency of the model with a single variable was limited. The main factors that affect h were found in Dahurian larch forest in the Great Khing'an Mountains, which proved that it was reliable to add these variables into the model. In the process of forest resource investigation and forest management planning, local forest characteristics should be taken into account, and influence factors, such as site quality and competition intensity, should be fully considered. This method can be extended to other forests according to the stand characteristics, site characteristics and competition, to achieve accurate and reliable prediction results.

Acknowledgments. This work was supported by the Natural Science Foundation of Inner Mongolia #1 under Grant No. 2021MS03054 and Science and Technology Planning Project of Inner Mongolia Autonomous Region #2 under Grant No. 2020GG0067. We thank Shengli Han, the host of the fund Natural Science Foundation of Inner Mongolia, for giving the study support in finance. We thank Maxine Garcia, PhD, from Liwen Bianji (Edanz) for editing the English text of a draft of this manuscript.

REFERENCES

- [1] Adame, P., del Río, M., Cañellas, I. (2008): A mixed nonlinear height–diameter model for Pyrenean oak (*Quercus pyrenaica* Willd.). – *Forest Ecology and Management* 256(1-2): 88-98.
- [2] Ahmadi, K., Jalil, A. S., Kouchaksaraei, M. T., Aertsen, W. (2013): Non-linear height–diameter models for oriental beech (*Fagus orientalis* Lipsky) in the Hyrcanian forests, Iran. – *Biotechnology, Agronomy, Society and Environment* 17(3): 431-440.
- [3] Anitha, K., Verchot, L., Joseph, S., Herold, M., Manuri, S., Avitabile, V. (2015): A review of forest and tree plantation biomass equations in Indonesia. – *Annals of Forest Science* 72(8): 981-997.
- [4] Bohora, S. B., Cao, Q. V. (2014): Prediction of tree diameter growth using quantile regression and mixed-effects models. – *Forest Ecology and Management* 319: 62-66.
- [5] Bronisz, K., Mehtätalo, L. (2020): Mixed-effects generalized height–diameter model for young silver birch stands on post-agricultural lands. – *Forest Ecology and Management* 460(12): 117901.
- [6] Browne, M. W. Cudeck, R. (1992): Alternative ways of assessing model fit. – *Sociological Methods & Research* 21(2): 230-258.
- [7] Bruchwald, A., Wróblewski, L. (1994): Uniform height curves for Norway spruce stands. – *Folia For. Pol. Ser. A For* 36: 43-47.
- [8] Chai, Z., Tan, W., Li, Y., Yan, L., Yuan, H., Li, Z. (2018): Generalized nonlinear height–diameter models for a *Cryptomeria fortunei* plantation in the Pingba region of Guizhou Province, China. – *Web Ecology* 18(1): 29-35.
- [9] Chang, H. Y., Lo, C. L., Hung, Y. Y. (2019): Development and validation of traditional & complementary medicine (TCM) scales for nurses: using structural equation modelling (SEM). – *BMC Complementary and Alternative Medicine* 19(1): 1-10.
- [10] Colbert, K. C., Larsen, D. R., Lootens, J. R. (2002): Height-diameter equations for thirteen midwestern bottomland hardwood species. – *Northern Journal of Applied Forestry* 19(4): 171-176.
- [11] Curtis, R. O. (1967): Height–diameter and height–diameter–age equations for second-growth Douglas-fir. – *Forest Science* 13(4): 365-375.
- [12] Dai, Z., Jiang, L. (2015): Ecoregion based Height–diameter models for *Larix gmelinii* Rupr. in DaKhing'an Mountains. – *Bulletin of Botanical Research* 35(4): 583-589.
- [13] Delgado, M. C., Pukkala, T. (2013): Growth models based on radial increment observations for eight pine species in Angola. – *Southern Forests* 75(1): 19-27.
- [14] Doncaster, C. P. (2007): Structural equation modeling and natural systems. – *Fish and Fisheries* 8(4): 368-369.
- [15] Dorado, F. C., Diéguez-Aranda, U., Anta, M. B., Rodríguez, M. S., Gadow, K. (2006): A generalized height–diameter model including random components for radiata pine plantations in northwestern Spain. – *Forest Ecology and Management* 229(1-3): 202-213.
- [16] Eerikäinen, K. (2003): Predicting the height $\pm$ diameter pattern of planted *Pinus kesiya* stands in Zambia and Zimbabwe. – *Forest Ecology and Management* 175(1-3): 355-366.
- [17] Enzinga, G. Y., Jiang, L. C. (2019): Evaluation of region and subregion-based height diameter models for Dahurian larch (*Larix gmelinii*) in Dakhing'an mountains in China. – *Applied Ecology and Environmental Research* 17(6): 13567-13591.
- [18] Fu, L., Zhang, H., Sharma, R. P., Pang, L., Wang, G. (2016): A generalized nonlinear mixed-effects height to crown base model for Mongolian oak in northeast China. – *Forest Ecology and Management* 384(1): 34-43.
- [19] Grace, J. B. (2006): Structural equation modeling and natural systems. – Cambridge University Press.
- [20] Hökkä, H. (1997): Height–diameter curves with random intercepts and slopes for trees growing on drained peatlands. – *Forest Ecology and Management* 97(1): 63-72.

- [21] Huang, S., Titus, S. J., Wiens, D. P. (1992): Wiens. comparison of nonlinear height diameter functions for major Alberta tree species. – Canadian Journal of Forest Research 22(9): 1297-1304.
- [22] Huang, S., Price, D., Titus, S. J. (2000): Development of ecoregion-based height–diameter models for white spruce in boreal forests. – Forest Ecology and Management 129(1): 125-141.
- [23] Huang, S., Yang, Y., Wang, Y. (2003): A critical look at procedures for validating growth and yield models. – Modelling Forest Systems Workshop on the Interface Between Reality 271-293.
- [24] Lai, K., Green, S. B. (2016): The problem with having two watches: assessment of fit when RMSEA and CFI disagree. – Multivariate Behavioral Research 51(23): 220-239.
- [25] Laiho, O., Pukkala, T., Lähde, E. (2014): Height increment of understorey Norway spruces under different tree canopies. – Forest Ecosystems 1(1): 4.
- [26] Lee, S. W., Kim, E. J. (2018): Structural equation model for burn severity with topographic variables and susceptible forest cover. – Sustainability 10(7): 2473.
- [27] Lei, X., Peng, C., Wang, H., Zhou, X. (2009): Individual height–diameter models for young black spruce (*Picea mariana*) and jack pine (*Pinus banksiana*) plantations in New Brunswick, Canada. – The Forestry Chronicle 85(1): 43-56.
- [28] Li, Y., Deng, X., Huang, Z., Xiang, W., Yan, W., Lei, P., Zhou, X., Peng, C. (2015): Development and evaluation of models for the relationship between tree height and diameter at breast height for Chinese-fir plantations in subtropical China. – PLoS ONE 10(4): e0125118.
- [29] Liu, F., Li, F., Zhang, L., Jin, X. (2014): Modeling diameter distributions of mixed-species forest stands. – Scandinavian Journal of Forest Research 29(7): 653-663.
- [30] Lumbres, R. I. C., Lee, J. Y., Calora, F. G., Parao, M. R. (2013): Model fitting and validation of six height-DBH equations for *Pinus kesiya* Royle ex Gordon in Benguet Province, Philippines. – Forest Science and Technology 9(1): 45-50.
- [31] Mackenzie, A., Begon, M., Harper, J. L., Townsend, C. R. (1997): Ecology: individuals, populations and communities. – Journal of Applied Ecology 34(1): 261-262.
- [32] Mayer, W. H. (1936): Height Curves for Even-Aged Stands of Douglasfir. – U.S. Dept. of Agriculture, Forest Service, Pacific Northwest Forest Experiment Station.
- [33] Mehtätalo, L., de-Miguel, S., Gregoire, T. G. (2015): Modeling height–diameter curves for prediction. – Canadian Journal of Forest Research 45: 826-837.
- [34] Meyer, H. A. (1940): A mathematical expression for height curves. – Journal of Forestry 38(5): 415-420.
- [35] Misir, N. (2010): Generalized height–diameter models for *Populus tremula* L. stands. – African Journal of Biotechnology 9(28): 4348-4355.
- [36] Monserud, R. A., Sterba, H. (1996): A basal area increment model for individual trees growing in even- and uneven-aged forest stands in Austria. – Forest Ecology and Management 80(1): 57-80.
- [37] Newton, P. F., Amponsah, I. G. (2007): Comparative evaluation of five height–diameter models developed for black spruce and jack pine stand-types in terms of goodness-of-fit, lack-of-fit and predictive ability. – Forest Ecology and Management 247(1): 149-166.
- [38] Ng'andwe, P., Chungu, D., Yambayamba, A. M., Chilambwe, A. (2019): Modeling the height–diameter relationship of planted *Pinus kesiya* in Zambia. – Forest Ecology and Management 447: 11.
- [39] Quinn, G. P., Keough, M. J. (2002): Experimental Design and Data Analysis for Biologists. – Cambridge University Press, Cambridge.
- [40] Rijal, B., Weiskittel, A. R., Kershaw, Jr. J. A. (2012): Development of height to crown base models for thirteen tree species of the North American Acadian Region. – The Forestry Chronicle 88(1): 60-73.

- [41] Schmidt, M., Breidenbach, J., Astrup, R. (2018): Longitudinal height–diameter curves for Norway spruce, Scots pine and silver birch in Norway based on shape constraint additive regression models. – *Forest Ecosystems* 5(1): 9.
- [42] Schumacher, F. X. (1939): New growth curve and its application to timber-yield studies. – *Journal of Forestry* 37(10): 819-20.
- [43] Seifert, T., Seifert, S., Seydack, A., Durrheim, G., Gadow, K. (2014): Competition effects in an afrotemperate forest. – *Forest Ecosystems* 1(1): 1-15.
- [44] Sharma, R. P. (2010): Modelling height–diameter relationship for Chir pine trees. – *Banko Janakari* 19(2): 3-9.
- [45] Sharma, R. P., Brunner, A. (2017): Modeling individual tree height growth of Norway spruce and Scots pine from national forest inventory data in Norway. – *Scandinavian Journal of Forest Research* 32(6): 501-514.
- [46] Sharma, R. P., Vacek, Z., Vacek, S. (2016): Modeling individual tree height to diameter ratio for Norway spruce and European beech in Czech Republic. – *Trees* 30(6): 1969-1982.
- [47] Sharma, R. P., Vacek, Z., Vacek, S., Kučera, M. (2019): Modelling individual tree height–diameter relationships for multi-layered and multi-species forests in central Europe. – *Trees* 33(1): 103-119.
- [48] Sileshi, G. W. (2014): A critical review of forest biomass estimation models, common mistakes and corrective measures. – *Forest Ecology and Management* 329: 237-254.
- [49] Soares, P., Tomé, M. (2002): Height–diameter equation for first rotation eucalypt plantations in Portugal. – *Forest Ecology and Management* 166(1-3): 99-109.
- [50] Sobel, M. (2008): Identification of causal parameters in randomized studies with mediating variables. – *Journal of Educational and Behavioral Statistics* 33(2): 230-251.
- [51] Sokal, R. R., Rohlf, F. J., Cairns, V., Keil, U., Döring, A., Koenig, W., Stieber, J. (1985): *Biometry: the principles and practice of statistics in bio.* – *International Journal of Epidemiology* 14: 389-395.
- [52] Spiess, A. N., Neumeyer, N. (2010): An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. – *BMC Pharmacology* 10(1): 6.
- [53] Stankova, T. V., Diéguez-Aranda, U. (2013): Height–diameter relationships for Scots pine plantations in Bulgaria: optimal combination of model type and application. – *Annals of Forest Research* 56(1): 149-163.
- [54] Tekce, I., Ergen, E., Artan, D. (2020): Structural equation model of occupant satisfaction for evaluating the performance of office buildings. – *Arabian Journal for Science and Engineering* 45(10): 8759-8784.
- [55] Temesgen, H., Monleon, V. J., Hann, D. W. (2008): Analysis and comparison of nonlinear tree height prediction strategies for Douglas-fir forests. – *Canadian Journal of Forest Research* 38(3): 553-565.
- [56] Temesgen, H., Zhang, C., Zhao, X. H. (2014): Modelling tree height–diameter relationships in multi-species and multi-layered forests: a large observational study from Northeast China. – *Forest Ecology and Management* 316: 78-89.
- [57] Walters, D. K., Hann, D. W. (1986): Taper equations for six conifer species in Southwest Oregon. – *Research Bulletin. Oregon State University, Forest Research Laboratory (USA)*.
- [58] Wang, H., Wang, W., Qiu, L., Su, D., An, J., Zheng, G., Zu, Y. (2012): Differences in biomass, litter layer mass and SOC storage changing with tree growth in *Larix gmelinii* plantations in Northeast China. – *Acta Ecologica Sinica* 32(3): 833-843.
- [59] Wang, D., Zhang, Z., Mu, H., Li, Y., Huang, X. (2015): Applications of structural equation model in the management of *Larix principis-rupprechtii* plantations. – *Journal of Beijing Forestry University* 37(3): 69-75.
- [60] Weiss, R. E., David, R. (1990): Handbook of nonlinear regression. – *Journal of the American Statistical Association* 85(412): 1172.

- [61] Welch, B. L. (1938): The significance of the difference between two means when the population variances are unequal. – *Biometrika* 29(3): 350-362.
- [62] Wykoff, W. (1982): User's Guide to the Stand Prognosis. – US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station.
- [63] Yao, x., Fu, B., Lu, Y., Sun, F., Wang, S., Liu, M. (2013): Comparison of four spatial interpolation methods for estimating soil moisture in a complex terrain catchment. – *PLoS ONE* 8(1): e54660.
- [64] Zeide, B. (1989): Accuracy of equations describing diameter growth. – *Canadian Journal of Forest Research* 19(10): 1283-1286.