Development and evaluation of height diameter at breast models for native Chinese Metasequoia

Mu Liu1,2, Zhongke Feng1,* , Zhixiang Zhang3, Chenghui Ma4, Mingming Wang1, Bo-ling Lian4, Renjie Sun1, Li Zhang1

1 Precision Forestry Key Laboratory of Beijing, Beijing Forestry University, Beijing, China, 2 The Key Laboratory for Silviculture and Conservation of Ministry of Education, Beijing Forestry University, Beijing, China, 3 School of Nature Conservation, Beijing Forestry University, Beijing China, 4 Northeast Forestry University, Harbin, China

* fengzhongke@126.com

Abstract

Accurate tree height and diameter at breast height (dbh) are important input variables for growth and yield models. A total of 5503 Chinese Metasequoia trees were used in this study. We studied 53 fitted models, of which 7 were linear models and 46 were non-linear models. These models were divided into two groups of single models and multivariate models according to the number of independent variables. The results show that the allometry equation of tree height which has diameter at breast height as independent variable can better reflect the change of tree height; in addition the prediction accuracy of the multivariate composite models is higher than that of the single variable models. Although tree age is not the most important variable in the study of the relationship between tree height and dbh, the consideration of tree age when choosing models and parameters in model selection can make the prediction of tree height more accurate. The amount of data is also an important parameter what can improve the reliability of models. Other variables such as tree height, main dbh and altitude, etc can also affect models. In this study, the method of developing the recommended models for predicting the tree height of native Metasequoias aged 50–485 years is statistically reliable and can be used for reference in predicting the growth and production of mature native Metasequoia.

Introduction

Tree height and dbh are the two most important factors in surveys, production and management of forest resources and research on forest ecosystems [1,2]. They are usually used to calculate the volume, site index, forest growth and yield [3]; and to estimate forest volume, biomass and carbon stock [4]. Accurate tree height and dbh are necessary conditions for evaluating biomass and are of great importance for the research of forest growth models based on physiological ecology.

Compared with dbh, the observation of tree height is often affected by the complexity of the distribution of forest vegetation, forest density and landform [5,6]. At the same time, it takes time and effort to measure tree height since there are some limitations caused by observational...
error and visual disturbance [7], which increases the cost of the forest survey. Therefore, it is very necessary to construct a simple and accurate tree height-dbh model to estimate the height of trees [8].

The allometry equation of tree height and dbh is usually used in estimating tree height. This method neglects the possible large deviation in estimating biomass by allometry [9]. There are currently many studies on tree height-dbh models, and some tree height-dbh models of common tree species have achieved good effects in application [10–12]. These studies mainly focus on artificial forests, natural forests and pure forests [13–17]. An individual tree growth model is the basis of a forest growth and production forecast [18]. Huang et al. [11] evaluated the main species of Alberta by using 20 non-linear tree height-dbh models. The non-linear function fitting is relatively easier, so the non-linear tree height-dbh functions are widely used in the prediction of tree height [19–22]. However, changes in tree height and dbh may lead to a deviation in the predicted tree height. In tree height and dbh relationships, there is a correlation between tree height and growing conditions, forest density, tree age, basal area and dominant tree height and dbh. The correct choice of allometry models is the key in accurate prediction [23]. In this process, the choice of parameters is a key factor in making errors in models. The proper parameters can improve the data fitting of models [24]. Therefore, we cannot apply the same function to all the tree height-dbh models.

Metasequoia is a relict plant and is known as a “living fossil”. The earliest Metasequoia fossil plants appeared in the sedimentary formation in the Mesozoic Cretaceous period, which was 63~110 million years ago [25]. The ancient Metasequoia plant originated in the Arctic Circle and then gradually expanded southward. In the quaternary Pleistocene period, when temperature dropped dramatically, Metasequoia plants gradually disappeared [26]. The native population of existing Metasequoia is distributed only at the junction of the Hubei, Hunan, Sichuan and Chongqing Provinces of China. Metasequoia in other regions were directly or indirectly introduced from the studied area. This area has become the only existing shelter for the native population of Metasequoia [27].

The tree population of native Metasequoia is one of the 69 endangered species in China [28]. It is the only existing habitat of the most primitive population of Metasequoia, therefore is the only place to study the living situation of the natural population of Metasequoia which preserves the most complete gene pool of the existing Metasequoia and is of great scientific value. This plays an irreplaceable role in the study of the genetic diversity and other biological characteristics of Metasequoia. The study of the native population and its living environment is of great significance to paleobotany, paleoclimatology and paleogeology [29].

Up to now, there has not been any study worldwide on the allometry models of tree height and dbh of native Chinese Metasequoia. This thesis aims to study the allometry equation of native Metasequoia by using tree height and dbh data. It is significant in clarifying the development and structure of this species. Previous studies were mostly conducted on plantations, and the growth time was short. However, predicting the height of native Metasequoia by using area-based tree height-dbh models has high uncertainty. Therefore, this study aims to 1) develop a model that can be used to predict the height-dbh relationship of native Metasequoia in China and 2) determine which variables have a significant effect on the relationship between tree height and dbh.

**Materials and methods**

**Ethics statement**

Our study sites (Xingdoushan National Nature Reserve, Hubei) are owned by the National Forestry Bureau of Enshi Tujia and Miao Autonomous Prefecture. Our research work was
conducted in collaboration with the National Forestry Bureau of Enshi Tujia and Miao Autonomous Prefecture. We are permitted by the Forestry Bureau to collect the related plant sample data. Our research object is Metasequoia, the local protected species.

The profile of the study sites

The study sites are located in Enshi City, Lichuan City and Xianfeng County of southwestern Hubei Province (Fig 1). The protection zone is divided into two districts: the eastern part is the Xingdou mountain area, located at the junction of Enshi City, Lichuan City and Xianfeng County, which is located at north latitude 29˚ 57’ ~ 30˚ 10’ and east longitude 108˚ 57’ ~ 109˚ 27’; and the western part is the Xiaohe area, located in Lichuan City, which is located at north latitude 30˚ 04’ ~ 30˚ 14’ and east longitude 108˚ 31’ ~ 108˚ 48’, with a total area of 68339 hm². The Metasequoia under study is in the Xiaohe area, the western part of the protection zone. The study sites are located in a sub-tropical continental monsoon area. The annual precipitation is 1481 mm. The annual average temperature is 14.9˚C. The annual average frost-free period is 217 days. The trees of Metasequoia under study are distributed at an altitude of 900–1350 m. The soil in the study sites is loam, mainly a mixture of yellow soil and brown soil, with a small amount of purple soil. There are 5763 trees of native Metasequoia in total, and we collected data from all these trees. From the material collection and investigation, we know that the age of the Metasequoia trees is 50–510 years and that these trees are distributed in 4 towns, 16 administrative zones and 45 villages, with a total area of more than 800 km². The trees are scattered, with a maximum of 20 in one district. The oldest trees are 510 years old, with a dbh of 248 cm and a height of 51 m. The average age of the trees is 97 years. The average dbh is 65.24 cm, and the average height is 27.84 m.

Data collection

The data that were collected from the trees are individual measurements, including dbh, tree height, elevation and tree age, collected using a dbh ruler, height finder, Runtastic Altimeter
PRO and vegetative cone, respectively. We chose the 10 toppest trees as the main samples. We also measured other stand-level variables, such as the dominant dbh (D₀ cm), the dominant height (H₀) and the basal area (BA). There are data from a total of 5746 Metasequoia trees available for this study. Some samples are either top dried, top broken, or struck by lightning, burned, or damaged by people. These samples could produce errors in the models; therefore, they were removed from the dataset. There were 5503 samples remaining that were studied.

The data from the 5503 sample trees were divided into two parts at random. The majority of the data (80%) was used for model calibration, and the remainder (20%) was used for verifying the consistency of the calibration data and the models. The calibration data show that the average dbh and height are 57.03 cm and 27.61 m, respectively. However, according to the validation data, the average dbh and height are 57.3 cm and 27.73 m, respectively. Table 1 shows the description of the stand-level variables.

Model selection and evaluation

We chose the linear and non-linear models and compared their performance. In this study, we selected two groups of general tree heightdbh models. There are 53 equations in total, as shown in Table 2. The criteria set for the two groups of models is in accordance with the number of variables, as follows:

Group 1: models that use one independent variable. We only need to measure the dbh.

Group 2: models that use two or more independent variables. We need to measure the dbh and other variables.

Model selection and evaluation are based on the graphical and numerical data analysis, as follows: (1) root of mean square error (RMSE) (Eq 1), used for the accuracy estimation of the analysis (the smaller, the better); (2) adjusted coefficient of determination (R²_adj) (Eqs.2 and 3), which reflects the part of the total variance explained by the model and takes into account the parameters that are necessary in making the estimation (the greater the value, the more inter-related between the variables); (3) deviation (Eq 4) and relative deviation (Eq 5), which is the deviation in the observed results from the evaluation model (the smaller, the better the effect); and (4) Akaike’s information criterion (AIC) (Eqs 6 and 7), which is an information standard commonly used to choose the best model (as a rule, the model whose AIC value is lower is

### Table 1. Regional Metasequoia sample statistics.

|                     | h(m) | d(cm) | BA(cm²) | ASL(m) | T(y) | H₀(m) | D₀(cm) |
|---------------------|------|-------|---------|--------|------|-------|--------|
| Fitting data Mean   | 27.61| 57.03 | 2675.36 | 1187.16| 95   | 43.04 | 122.59 |
| N = 4401 Max        | 46.41| 134.68| 14246.35| 1590   | 485  | 46.41 | 134.68 |
| Min                 | 16.69| 26.35 | 545.21  | 750    | 50   | 40.5  | 110.14 |
| Standard deviation  | 3.78 | 12.43 | 1221.36 | 111.38 | 32.58| 3.26  | 8.79   |
| Validation data Mean| 27.73| 57.3  | 2686.72 | 1185.33| 95.92| 38.11 | 95.68  |
| N = 1102 Max        | 40.09| 109.8 | 9468.78 | 1605   | 325  | 40.09 | 109.8  |
| Min                 | 18.8 | 28.47 | 636.69  | 856    | 50   | 37.12 | 95.68  |
| Standard deviation  | 3.55 | 11.75 | 1124.55 | 11.93  | 32.91| 0.89  | 6.39   |

h: height; d: diameter at breast height (dbh); BA: basal area; ASL: Above Sea Level; T: age of the stand; H₀: dominant height of the stand, m; D₀: dominant dbh of the stand, cm; N: number of trees. doi: 10.6084/m9.figshare.4956284.t001

https://doi.org/10.1371/journal.pone.0182170.t001
Table 2. Selected tree height-db h models.

| No. | Models | References | Group |
|-----|--------|------------|-------|
| 1   | $h = a_0 + a_1 d$ | [1] | 1 |
| 2   | $h = a_0 + a_1 d \log d$ | [10] | 1 |
| 3   | $h = a_0 + a_1 d + a_2 d^2$ | [10] | 1 |
| 4   | $h = a_0 + a_1 d^2 + a_2 d^3$ | [10] | 1 |
| 5   | $h = a_0 + a_1 d^{-1} + a_2 d^2$ | [10] | 1 |
| 6   | $h = a_0 + a_1 (d/Dq) + a_2 Hm$ | [1] | 2 |
| 7   | $h = 1.3 + a_0 + a_1 \log (d/Dq) + a_2 /\log (Hm)$ | [1] | 2 |

Non-linear models

8
$h = 1.3 + a_1 (1 - e^{-\alpha_2 d})^{\alpha_3}$
[30] 1
9
$h = 1.3 + a_1 d^{\alpha_2}$
[31] 1
10
$h = 1.3 + e^{a_1 \alpha_2 / (d-1)}$
[29] 1
11
$h = 1.3 + a_0 d^2 (a_1 + d)$
[21] 1
12
$h = 1.3 + a_1 (1 - e^{-\alpha_2 d})^{\alpha_3}$
[32] 1
13
$h = 1.3 + 10^\alpha_2 d^{\alpha_3}$
[33] 1
14
$h = 1.3 + a_0 d^2 (d + 1) + a_1 d$
[34] 1
15
$h = 1.3 + a_1 (d/(1 + d))^{\alpha_2}$
[10] 1
16
$h = 1.3 + a_1 / (1 + a_2 e^{-\alpha_2 d})$
[35] 1
17
$h = 1.3 + a_1 (1 - e^{-\alpha_2 d})^{\alpha_3}$
[36] 1
18
$h = 1.3 + a_0 d^2 (d^{-\alpha_2})$
[37] 1
19
$h = 1.3 + a_1 e^{-\alpha_2 d^{\alpha_3}}$
[38] 1
20
$h = 1.3 + d^2 (a_0 + a_1 d + a_2 d^2)$
[10] 1
21
$h = 1.3 + a_2 d^2 (d)_{\alpha_2}$
[39] 1
22
$h = 1.3 + a_0 e^{a_1 / d^{\alpha_2}}$
[40] 1
23
$h = 1.3 + a_0 / (1 + a_1 d^{-\alpha_2})$
[41] 1
24
$h = 1.3 + a_0 + a_1 (d + a_2 d)$
[42] 1
25
$h = 1.3 + a_0 e^{-\alpha_2 d^{\alpha_3}}$
[43] 1
26
$h = 1.3 + a_0 (e^{a_1 - a_2 (d-a_3) / d})$
[44] 1
27
$h = 1.3 + a_0 e^{a_1 / d^{\alpha_2}}$
[20] 1
28
$h = a_0 + e^{a_1 / d^{\alpha_2}}$
[45] 1
29
$h = 1.3 + a_1 / (1 + a_2 d^{\alpha_2})$
[46] 1
30
$h = 1.3 + a_0 d e^{-\alpha_2 d}$
[46] 1
31
$h = 1.3 + a_0 d^2 (d-a_2) e^{-\alpha_2 d}$
[46] 1
32
$h = 1.3 + a_0 e^{a_1 / d^{\alpha_2}}$
[46] 1
33
$h = 1.3 + (a_0 + a_1 / d)^{\alpha_2}$
[46] 1
34
$h = 1.3 + d^2 (a_0 + a_1 d)^{\alpha_2}$
[22] 1
35
$h = 1.3 + a_1 d^{\alpha_2}$
[47] 1
36
$h = 1.3 + a_1 e^{a_2 / d^{\alpha_2}}$
[19] 1
37
$h = a_0 e^{a_1 / d^{\alpha_2}}$
[48] 1
38
$h = e^{a_0 + a_1 d^{-\alpha_2} + a_2 d^{-\alpha_3} + a_3 d^{\alpha_3}}$
[48] 1
39
$h = a_0 + e^{a_1 / d^{\alpha_2}}$
[48] 1
40
$h = 1.3 + a_0 (BA)^{\alpha_2} (1 - e^{-a_2 d^{\alpha_3}})$
[49] 2
41
$h = 1.3 + (H_m - 1.3) e^{a_1 / (1 - d/Dq) + a_2 (d/Dq - 1) / d}$
[15] 2
42
$h = H_m (1 + a_1 e^{a_2 / d}) (1 - e^{-a_3 d^{\alpha_3}})$
[30] 2
43
$h = 1.3 + a_1 H_m d^{\alpha_2}$
[50] 2
44
$h = 1.3 + (a_0 + a_1 H_m - a_2 Dq) e^{-a_3 / d}$
[51] 2
45
$h = 1.3 + (a_0 + a_1 H_m - a_2 Dq) e^{-a_3 / d}$
[15] 2

(Continued)
Table 2. (Continued)

| No. | Models                                                                 | References | Group |
|-----|------------------------------------------------------------------------|------------|-------|
| 46  | \( h = 1.3 + (a_0 + a_1 h - a_0 Dq + a_0 BA) e^{-a_1/\sqrt{d}} \)     | [15]       | 2     |
| 47  | \( h = a_0 + a_1 Hm + a_2 Dq^{0.95} + a_3 e^{-0.08d} + a_4 Hm^2 e^{-0.08d} + a_5 Dq^2 e^{-0.08d} \) | [15]       | 2     |
| 48  | \( h = 10^{(a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8 + a_9 + a_{10})} \) | [15], [52] | 2     |
| 49  | \( h = e^{a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8 + a_9 + a_{10}} \) | [53]       | 2     |
| 50  | \( h = a_0 + a_1 Hm + a_2 Dq + a_3 e^{0.08d} + a_4 Hm^2 e^{0.08d} + a_5 Dq^2 e^{0.08d} \) | [52]       | 2     |
| 51  | \( h = e^{a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8 + a_9 + a_{10}} \) | [15], [52] | 2     |
| 52  | \( h = a_0 H_n^m (BA)^n N^p e^{a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8 + a_9 + a_{10}} \) | [54]       | 2     |
| 53  | \( h = 1.3 + (a_0 + a_1 Hm + a_2 H_n) (1 - e^{-a_1/\sqrt{d}}) h \)     | [55]       | 2     |

\( h \): height, \( m \); \( d \): diameter at breast height (dbh), cm; \( BA \): basal area, \( \text{cm}^2 \); \( t \): age of the stand; \( H_d \): dominant height of the stand, cm; \( Dq \): dominant dbh of the stand, cm; \( H_m \): mean height; \( Dq \): quadratic mean dbh; \( a_0-a_9 \) are parameters. The base is 10 for logarithm; \( N \) is number of trees. doi: 10.6084/m9.ﬁgshare.4956284.1002.

The expressions of the statistics are shown as follows:

\[
\text{Root of mean square error } \text{RMSE} = \sqrt{\frac{n}{n-p-1} \sum_{i=1}^{n} (h_i - \hat{h}_i)^2}
\]

(1)

\[
\text{Coefficient of determination } : \quad R^2 = 1 - \frac{\sum_{i=1}^{n} (h_i - \hat{h}_i)^2}{\sum_{i=1}^{n} (h_i - \bar{h})^2}
\]

(2)

Adjusted coefficient of determination : \( R^2_{adj} = 1 - (1 - R^2) \ast \frac{n-1}{n-p-1} \)

(3)

\[
\text{Bias} = \frac{\sum_{i=1}^{n} (\hat{h}_i - h_i)}{n}
\]

(4)

\[
\text{Relative bias(\%)} = \frac{\sum_{i=1}^{n} (\hat{h}_i - h_i)/h_i}{n} \ast 100
\]

(5)

\[
\text{Residual sum of squares } : \quad \text{RSS} = \sum_{i=1}^{n} (h_i - \hat{h}_i)^2
\]

(6)

\[
\text{Akaike's information criterion } AIC = n \ln(\text{RSS}) + 2(p+1) - n \ln(n)
\]

(7)

Where \( h_i \) is the observation value; \( \hat{h}_i \) is the forecast value and; \( \bar{h}_i \) is the average value; \( n \) is the total number of data used to the fitted model; and \( p \) is the number of independent variables.

Data analysis

Because many of the selected models were non-linear, they were fitted with the statistical package SPSS 20.0 (SPSS for Windows version 20.0) and the Levenberg-Marquardt (LM) methods. The initial values of the parameters for starting the iterative procedure were used to fit the data. The regression analysis was made using the SPSS program.
Results

The relationship between tree height and dbh

As shown in Fig 2, which is a scatter diagram of the tree height and dbh of all the 5503 native Metasequoia, the minimum dbh is greater than 26 cm, and with the increase in the dbh, the tree height also slowly increases. The solid line is the trend line, whose formula is \( h = 12.546 + 0.264 \times d (R^2 = 0.7582) \).

Model fitting and selection

Tables 3 and 4 are the prediction and fitting results of the fitting data and validation data from Group 1 models and Group 2 models, respectively. The two groups of models show good results, except for models 44, 47 and 48. Therefore, we can see that these three models are not fit for the prediction of the data, and we will remove their results in later statistics. As shown in Table 3, in the Group 1 models, the changes of the derived value between the adjusted \( R^2 \) and RMES are roughly equivalent. In the Group 1 models, model 4 is the most suitable linear model (the adjusted \( R^2 \) value is the highest, and the RMES value, deviation sum and AIC are the lowest). It derives the most accurate result in the fitting of calibration data and is suitable for estimating the height of most trees in Group 1. Model 22 is the most suitable non-linear model. Both models have only one independent variable (dbh) that is used to predict tree height, and their derived data are in good agreement. When considering the other variables (Group 2 models) the models show different performances. In the calibration data from the Group 2 models, the derived \( R^2 \) value is 0.7456~0.7737, and the RMSE value is 1.7965~1.9050 (Table 4). In addition, most of the Group 2 models have greater adjusted \( R^2 \) values and smaller RMSE values than the Group 1 models. In the models of Group 1, the calibration dataset of the AIC value is relatively higher. The calibration data of the AIC value in Group 2 models is smaller than that in the Group 1 models. The choice of models should be based on the arrangement of the expression of models using calibration data and validation data (the adjusted \( R^2 \),
RMSE, the absolute deviation, the relative deviation and the AIC). In this analysis, the top two or three models are summarized below in Table 5 (the numbers in brackets indicate the ranking of the parameters). The model with the adjusted $R^2$ value closest to the highest value and a deviation (absolute and relative deviation) closest to zero is thought to be the best. The lower the value of RMSE and AIC, the higher the ranking of the model. For each model, its final ranking is the sum of the statistics of the five evaluation values. The model with the minimum
sum (i.e., the highest ranking) is considered to be the functional model that is most suitable for estimating the growth of Metasequoia. According to this analysis, in the models of Group 1, the calibration data show that model 22 is the most suitable, and the validation data show that model 22 and model 4 are both the most suitable. In the Group 2 models, model 52 ranks the same according to both calibration and validation data. Overall, model 4 is the most suitable of the linear models and model 52 is the most suitable of the non-linear models. In the Group 1 models, model 22 is the most suitable. In all models, model 52 and model 51 are the most suitable non-linear models; model 4 is the most suitable linear model. Overall, the non-linear models performed slightly better than the linear models.

Fig 3 shows the predicted values and the observed values obtained by using the model. The straight line between the observed value and the predicted value is the determining factor of the model evaluation standard. Each model has a higher $R^2$ value, so the solid line is closely

### Table 4. Calibration data and validation data adaptive statistics for models in Group 2. doi: 10.6084/m9.figshare.4956284.t004.

| No. | Variables | Fitting data | Validation data |
|-----|-----------|--------------|-----------------|
|     |           | Bias(m) | RMSE(m) | AIC | $R^2$ | Bias(m) | RMSE(m) | AIC | $R^2$ |
|     | Liner models |     |         |     |       |     |         |     |       |
| 6   | h,d,Dq,Hm | 0.000  | 1.8411  | 5379.1 | 0.7623 | 0.0000 | 1.8087 | 1315.3 | 0.7397 |
| 7   | h,d,Dq,Hm | 0.000  | 1.8902  | 5610.7 | 0.7495 | 0.0000 | 1.8578 | 1374.4 | 0.7253 |
|     | Non-linear models |     |         |     |       |     |         |     |       |
| 40  | h,d,BA   | -0.0015 | 1.8419  | 5380.8 | 0.7621 | -0.0017 | 1.8176 | 1324.1 | 0.7371 |
| 41  | h,d,Dq,Hm | 0.0079  | 1.9050  | 5677.3 | 0.7456 | 0.0028  | 1.8286 | 1337.4 | 0.7339 |
| 42  | h,d,Ho   | -0.0159 | 1.9013  | 5660.4 | 0.7465 | -0.0121 | 1.8680 | 1384.5 | 0.7223 |
| 43  | h,d,Ho   | -0.0015 | 1.8419  | 5380.9 | 0.7621 | -0.0017 | 1.8176 | 1324.1 | 0.7371 |
| 44  | h,d,Ho,Dq | -0.0078 | 1.9520  | 5891.8 | 0.7329 | -0.0059 | 1.9068 | 1429.9 | 0.7106 |
| 45  | h,d,Ho,Dq | -0.0048 | 1.8807  | 5564.6 | 0.7520 | -0.0037 | 1.8510 | 1364.3 | 0.7274 |
| 46  | h,d,Ho,Dq,BA | -0.0003 | 1.8426  | 5388.4 | 0.7619 | -0.0001 | 1.8116 | 1320.8 | 0.7388 |
| 47  | h,d,Dq,Hm | 0.0000  | 2.5633  | 8291.3 | 0.5393 | 0.0000  | 2.4307 | 1967.4 | 0.5298 |
| 48  | h,d,t     | -0.4314 | 4.2983  | 12790.1 | 0.2850 | -0.5747 | 4.7089 | 3426.1 | 0.2646 |
| 49  | h,d,Dq,N  | 0.0019  | 1.8341  | 5321.5 | 0.7633 | 0.0000  | 1.8075 | 1309.8 | 0.7397 |
| 50  | h,d,Dq,Hm | 0.0000  | 1.8372  | 5360.7 | 0.7633 | 0.0000  | 1.8086 | 1315.2 | 0.7397 |
| 51  | h,d,Ho,N,t | 0.0000  | 1.8337  | 5321.2 | 0.7636 | 0.0000  | 1.7609 | 1308.2 | 0.7532 |
| 52  | h,d,BA,N,t | 0.0000  | 1.7965  | 5165.4 | 0.7737 | 0.0000  | 1.7198 | 1206.1 | 0.7646 |
| 53  | h,d,Ho,Hm | -0.0015 | 1.8419  | 5382.9 | 0.7621 | -0.0017 | 1.8176 | 1326.1 | 0.7371 |

https://doi.org/10.1371/journal.pone.0182170.t004

### Table 5. Model ranking based on performance. doi: 10.6084/m9.figshare.4956284.t005.

| No. | Group 1 models | Fitting data | Validation data | Group 2 models | Fitting data | Validation data | All models | Fitting data | Validation data |
|-----|----------------|--------------|-----------------|----------------|--------------|----------------|------------|--------------|-----------------|
| RMSE | 22(1);4(2);19, 26,39(3) | 22,4(1);16,39 (2);19,26,38(3) | 52(1);51(2);49(3) | 52(1);51(2);49(3) | 52(1);22(2);51(3) | 52(1);51(2);4,22(3) |
| AIC  | 22(1);4(2);19, 26,39(3) | 22(1);4(2);39(3) | 52(1);51(2);49(3) | 52(1);51(2);49(3) | 52(1);22(2);51(3) | 52(1);51(2);4,22(3) |
| $R^2$ | 22(1);4(2);19, 26,39(3) | 4,22,39,16 (1);19,26,38,31,32(2) | 52(1);51(2);6,49,50(3) | 52(1);51(2);49(3) | 52(1);22(2);51(3) | 52(1);51(2);4,22(3) |
| Absolute bias | 1–5,14,16,18– 28,31–33,37–39 | 1–5,14,21,22,24,25,27–29,31,32,37–39 | 47,50,51,52,6,7,49,41 (1) | 49,50,51,52,6,7,4,47 (1) | 1–7,14,16,18,20–22,24–28,31,32,37–39,41,49,50–52(1) | 1–7,14,21,22,24,25,27–29,31,32,37–39,41,49,50–52(1) |
| Relative bias | 30(1);12,22,11 (3) | 24(1);23(2);18(3) | 42(1);51(2);52(3) | 42(1);51(2);52(3) | 30(1);12,42(2);51(3) | 21(1);42,23(2)18,51(3) |

https://doi.org/10.1371/journal.pone.0182170.t005
surrounded by data points. There is no obvious tendency to overestimate or underestimate the observed height values. Fig 4 shows the residual error of the predicted height obtained from model 4, 22, 51 and 52 by using calibration data. Most data points are distributed around the solid line. These models all show the uniform variance of the predicted value, as well as the independence of the residuals. Fig 5 shows the average deviation value obtained from the diameter by models 4, 22, 51 and 52 using calibration data and validation data. From the above analysis, models 4, 22, 51 and 52 become the best four models in the study. However, the number of parameters in models 51 and 52 is higher than that in models 4 and 22, but with a higher $R^2$ and a lower RMSE and AIC. Therefore, in the case that only the dbh is known, models 4 and 22 can predict tree height; and if tree age, dominant tree height, basal area and the number of samples are known, models 51 and 52 can be used to obtain a more accurate tree height.

**Fig 3.** Graph of predicted values in contrast to observed values in the calibration dataset for the three best models (a: Model 4; b: Model 22; c: Model 51; d: Model 52). The solid line represents the diagonal. doi: 10.6084/m9.figshare.4956284.g003.

https://doi.org/10.1371/journal.pone.0182170.g003
Parameter estimation

Initially, the model estimates all the parameters using the calibration data. Table 6 shows the parameter estimation and fitting statistics of models 4, 22, 51 and 52 using all data. All the parameters were significant (P < 0.05). It is easy to obtain the parameters of each model through calculations.

Discussion

Evaluation and application of models

The choice of the models is based on the degree of optimization, the accuracy and the practical application. The models were compared and ranked after implementation (Table 5). Under this approach, models from Group 1 cannot be clearly differentiated. As shown in Fig 2, the result is that the tree height increases monotonically with the increase in the dbh. A reasonable tree height-dbh model should be consistent with the first order monotone increasing characteristic [56]. In the models of Group 2, the calibration and validation data for models 22, 51,
and 52 are in agreement. Usually, the increase in the new independent variables can reduce the error in the height-dbh model and increase the accuracy of the model. Therefore, through large amounts of sampling and measuring the different variables, we derived the Group 2 models that are more suitable for the prediction of tree height than the Group 1 models.

In this research, model 4 is the only linear model with 2 variables of tree height and dbh, and the others are non-linear models, in which model 22 has 2 variables of tree height and dbh; model 51 has 5 variables of tree height, dbh, main tree height, sample numbers and tree age; and model 52 has 5 variables of tree height, dbh, basal area, sample numbers and tree age. The non-linear models are more accurate. Therefore, we believe that the non-linear models perform better than the linear models in the case of Metasequoia. The non-linear models are more flexible. The forms and parameters of the models are very important for the application of models. We have demonstrated the stability of the structure of the four models. Therefore, the parameter is the determining factor that affects the accuracy of the models in their application.

The development of simple and accurate models makes it possible for forestry workers to predict the height of a tree by relying on the diameter data of a region, which is very important.

Table 6. Parameter estimates and fitting statistics of the final models by using all data. doi: 10.6084/m9.figshare.4956284.t006.

| Parameter | Model 4 | Model 22 | Model 51 | Model 52 |
|-----------|---------|----------|----------|----------|
| a0        | 17.7508 | 128.1449 | 15.3671  | 0.4994   |
| a1        | 0.0042  | -49.9808 | -3.0737  | 0.0507   |
| a2        | 0.0001  | 69.7668  | -9.5212  | 0.3521   |
| a3        |         | -28.3174 | 0.0196   |          |
| a4        |         | 544.2804 | 0.1972   |          |
| a5        |         | 206.7969 | 8.6632   |          |
| Adjusted $R^2$ | 0.7583 | 0.7591  | 0.7314  | 0.7592  |
| RMSE      | 1.8277  | 1.8237  | 1.8186  | 1.7806  |
| Absolute bias (m) | 0.0000 | -0.0032 | 0.0001 | 0.0000 |
| Relative bias (%)  | 0.4542 | 0.3904  | 0.4104  | 0.4630  |

Fig 5. Values of average deviation in relation to dbh in the calibration and validation datasets for the two best models (a: calibration data; b: validation data). doi: 10.6084/m9.figshare.4956284.g005.

https://doi.org/10.1371/journal.pone.0182170.g005
for forest management. In this study, models 51 and 52 are similar. The four models are all consistent with statistical principles. There are a relatively larger number of parameters in models 51 and 52, and the increase in the number of parameters can improve the accuracy of the model in the prediction of tree height. The choice of the four models is not only statistically reliable but is easier for application. The practical application of the models is necessary for the prediction of tree height.

The relationship between tree height and the study area variables

According to the study, bringing the study area variables as independent variables into the height-dbh model can improve the accuracy of tree height predictions [49]. The study area variables mainly include the development of comprehensive factors, such as the dominant tree height, dbh, tree age, basal area, stand density and growth status, etc. In the Group 1 models, only one independent variable of dbh (d) can be used to predict the relationship between height and dbh. Depending on the specific circumstances, if the other variables of the study area can be measured, Group 2 models are the best method. We can accurately assess the direct relationship between tree height and dbh using Group 2 models and can predict the future development.

A significant difference exists in the growth rate of tree height and dbh, and there is an allometric relationship between them. Because the average age of the Metasequoia trees collected in this study is more than 50 years and they are a mature forest, the tree height and dbh growth relationships are relatively stable. Although tree age is not the most important variable in the study of the tree height-dbh relationship, some differences in parameter estimation can be discovered when selecting models. Therefore, considering tree age when selecting models and parameters can make the tree height prediction more accurate.

Conclusion

This study compared and predicted the native population of Metasequoia aged 50–485 in China using 53 height-dbh models. The choice of models is based on the principles of good fit, high accuracy and suitability for practical application. The results show that the composite models which include additional variables can improve the quality of the models. Linear model 4 and non-linear model 22, which have 3 parameters and 2 variables each, and non-linear models 51 and 52, which have 6 parameters and 5 variables each, can better predict the height of natural Metasequoia trees. Although stand age is not the most significant variable in the relationship between height and dbh, knowing the tree age can improve the accuracy of natural Metasequoia tree height prediction. In addition, sample size can also improve the accuracy of tree height prediction to some degree. In this study, forest density will not affect the growth of the trees. However, as a referential maximum value of growth of local Metasequoia, the main tree height can also improve the accuracy of model prediction. In remote or individual regions, because different ecological niches may have an important impact on the estimation of height, it is not suitable for application. In our analysis, the measurement error due to the ground level was not considered. This is because these errors likely cannot affect the allometric models [57].

This research aims to build some models to facilitate the prediction of tree height of local Metasequoia, thus making it easier to evaluate the growth of native Metasequoia and protect it as an important plant resource. Therefore, the sanctuary workers who have mastered the original data of local Metasequoia can predict tree height using models 51 or 52. General scientific researchers, in the case that they have measured the dbh, can use models 4 or 22 to predict tree height. The methods and the models recommended in this study are based on statistics, which
are reliable for forest growth and survival estimates and planting management for Chinese native Metasequoia. After verification, these models can be applied in other regions and to other tree species.

Acknowledgments

We thank lecturer Gang Liu from Huazhong Agricultural University, College of Horticulture and Forestry Sciences, who provided help with survey instruments.

Author Contributions

Conceptualization: Mu Liu, Zhongke Feng, Zhixiang Zhang, Chenghui Ma.

Data curation: Mu Liu.

Formal analysis: Mu Liu, Zhongke Feng, Chenghui Ma, Mingming Wang, Bo-ling Lian.

Investigation: Mu Liu, Chenghui Ma, Renjie Sun, Li Zhang.

Methodology: Mu Liu, Chenghui Ma.

Project administration: Zhongke Feng, Zhixiang Zhang.

Supervision: Mu Liu, Zhongke Feng.

Validation: Mu Liu, Mingming Wang, Bo-ling Lian.

Writing – original draft: Mu Liu, Bo-ling Lian.

Writing – review & editing: Mu Liu, Bo-ling Lian.

References

1. Lei XD, Peng CH, Wang HY, Zhou XL. (2009). Individual height-diameter models for young black spruce (picea mariana) and jack pine (pinus banksiana) plantations in new brunswick, canada. Forestry Chronicle, 85(1), 43–56.

2. Vargaslarreta B., Castedodorado F., ÁlvarezGonzález, Gabriel Juan, Barrioanta M., Cruzcobos F. (2009). A generalized height–diameter model with random coefficients for uneven-aged stands in el salto, durango (mexico). Forestry, 84(4), 445–462.

3. Daniel L, Jeffery G. (2009). A height-diameter curve for longleaf pine plantations in the gulf coastal plain. Southern Journal of Applied Forestry, 33(4), 164–170(7).

4. Li Y, Deng KW, Huang Zh, Xiang W, Yan WD, Lei PF, et al. (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. https://doi.org/10.1371/journal.pone.0125118 PMID: 25905458

5. Xu H, Sun Y, Wang X, Fu Y, Dong Y, Li Y (2014) Nonlinear mixed-effects (NLME) diameter growth models for individual China-fir (Cunninghamia lanceolata) trees in southeast China. PLoS ONE 9:e104012. https://doi.org/10.1371/journal.pone.0104012 PMID: 25084538

6. Temesgen H, Zhang C H, 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 & Management 316(316):78–89.

7. Colbert KC, Larsen DR, Lootens JR (2002) Height-Diameter Equations for Thirteen Midwestern Bottomland Hardwood Species. Northern Journal of Applied Forestry 19(4):171–176.

8. Haiku Li, Lei Fa (2011) Height-Diameter Model for Major Tree Species in China Using the Classified Height Method. Scientia Silvae Sinicae 47(10):83–90.

9. Feldpausche T. R., Banin L., Phillips O. L., Baker T. R., Lewis S. L., & Quesada C. A., et al. (2011). Height-diameter allometry of tropical forest trees. Biogeosciences, 8(5), 7727–7793.

10. Curtis R (1967) Height-Diameter and Height-Diameter-Age Equations for Second-Growth Douglas-Fir. Forest Science 13(4):365–375.

11. Huang S, Titus S J, Wiens D P (1992) Comparison of nonlinear height–diameter functions for major Alberta tree species. Canadian Journal of Forest Research 22(9):1297–1304.
12. Fan YW, Wang XG, Zeng LB. (2011) Allometric relationship between diameter at breast height and height of *Quercus variabilis* plantations in Beijing. Journal of Beijing Forestry University 33(6): 146–151.

13. Zhao JH, Kang XG Liu Y (2009) Optimal height-diameter models for dominant coniferous species in Changbai Mountain, northeastern China. Journal of Beijing Forestry University 31(4): 13–18.

14. Li HK, Fa L (2011) Height-Diameter Model for Major Tree Species in China Using the Classified Height Met. Scientia Silvae Sinicae 47(10): 83–90.

15. Sanchez CA, Varela J, Dorado FC, Alboreca AR, Soalleiro R.R, Gonzalez JG, Rodriguez FS. (2003) A height-diameter model for *Pinus radiata* D. Don in Galicia (Northwest Spain). Annals of Forest Science 60(3):237–245.

16. Calama R, Montero G (2004) Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain. Canadian Journal of Forest Research 34(1):150–163.

17. Dieguezaranda U., Barroioanta M., Gonzalez J. G., Dorado F. C. (2005) Relación altura-diámetro generalizada para masas de *Pinus sylvestris* L. procedentes de repoblación en el noroeste de España. Investigación Agraria Sistemas Y Recursos Forestales 14(2):229–241.

18. Guangyi M., Yujun S., Hao X., Demiguel S. (2015). A mixed-effects model with different strategies for modeling volume in *Cunningh amia lanceolata* plantations. PloS One, 10(10), e0140095. https://doi.org/10.1371/journal.pone.0140095 PMID: 26445505

19. Schreuder HT, Hafley WL, Bennett FA (1979) Yield prediction for unthinned natural slash pine stands. Forest Science 25(1):25–30.

20. Curtis RO, Clendenen GW, Demars DJ (1981) A new stand simulator for coast Douglas-fir: DFSIM user’s guide.

21. Wykoff WR, Crookston NL, Stage AR (1982) User’s Guide to the Stand Prognosis Model.

22. Farr WA, Demars DJ, Dealy JE (2011) Height and crown width related to diameter for open-grown western hemlock and Sitka spruce. Canadian Journal of Forest Research 19(9):1203–1207.

23. Chave J, Condit R, Aguilar S, Hernandez A, Lao S, Perez R. (2004) Error Propagation and Scaling for Tropical Forest Biomass Estimates. Philosophical Transactions of the Royal Society B Biological Sciences, 359(1443):409–420.

24. Ledo A, Cornulier T, Illian J, Iida Y, Kassim AR, Burslem DF. (2016) Re-evaluation of individual diameter: height allometric models to improve biomass estimation of tropical trees. Ecological Applications a Publication of the Ecological Society of America, 26(8):2374–2382. https://doi.org/10.1002/eap.1450 PMID: 27907254

25. Yang Y, Liu B (2011) Who is the "living fossil" *Metasequoia* named person. Fossils (3):44–45.

26. Li YY (2006) Phylogenetic position and recovery assessment of *Metasequoia glyptostroboides*. East China Normal University.

27. Liu SX, Zou JP. (2003) Scientific surver of xingdoushan nature reserve Hubei. Hubei Science & Technology Press.

28. Guo Z, Zang RG. (2013) Evaluation index system of endangered levels of the wild plants with tiny population in China. Scientia Silvae Sinicae 49(6):10–17.

29. Le XW. (2013) Phylogeographic study of a relict plant: *Metasequoia glyptostroboides*. East China Normal University.

30. Vanclay JK. (1995) Growth models for tropical forests: a synthesis of models and methods. Forest Science 41(1):7–42.

31. Stoffels A, van Soest J (1953) The main problems in sample plots. Ned Boschb Tijdschr 25: 190–199.

32. Bates DM, Watts DG. (1980) Relative curvature measures of nonlinearity. Journal of the Royal Statistical Society 42: 1–16.

33. Buford MA (1986) Height-diameter relationships at age 15 in loblolly pine seed sources. Forest Science 32(3):812–818.

34. Larson BC. (2011) Development and growth of even-aged stands of Douglas-fir and grand fir. Canadian Journal of Forest Research 16(2):367–372.

35. Pearl R, Reed LJ (1920) On the rate of growth of the population of the united states since 1790 and its mathematical representation. Proceedings of the National Academy of Sciences 6(6), 275–288.

36. Richards FJ (1959) A Flexible Growth Function for Empirical Use. Journal of Experimental Botany 10 (2):290–301.

37. Bailey RL. (2011) The potential of Weibull-type functions as flexible growth curves: discussion. Canadian Journal of Forest Research 10(1): 117–118.

38. Winsor CP. (1932). The gompertz curve as a growth curve. Proceedings of the National Academy of Sciences of the United States of America, 18(1), 1–8. PMID: 16577417
39. Sibbesen E (1981) Some new equations to describe phosphate sorption by soils. European Journal of Soil Science 32(1):67–74.
40. Ziegel ER (1991) Handbook of Nonlinear Regression Models. Technometrics 33(2):240–241.
41. Ratkowsky DA, Reedy TJ (1986) Choosing near-linear parameters in the four-parameter logistic model for radio ligand and related assays. Biometrics 42(3):575–582. PMID: 3567290
42. Li X, Beijinr C, Yuan G, Ban D (1994) Self-adjusted height-diameter curves and one-entry volume model. Forest Research.
43. Zeide B (2011) Accuracy of equations describing diameter growth. Canadian Journal of Forest Research 19(10):1283–1286.
44. Seber GAF, Wild CJ (1989) Nonlinear regression.
45. Zang H, Lei XD, Zhang HR, Li CM, Lu J. (2016) Nonlinear mixed-effects height-diameter model of Pinus koraiensis. Journal of Beijing Forestry University 38(6):8–16.
46. Wang X, Li F, Jia W, Chen D, University NF. (2013). Height-diameter equation of main broad leaved tree species in natural secondary forest at maershan forestry center. Journal of Northeast Forestry University. (12):116–120.
47. Brack CL (2010) National Forest Inventories. Springer Netherlands.
48. Kozak A (2011) A variable-exponent taper equation. Canadian Journal of Forest Research 18 (11):1363–1368.
49. Sharma Mahadev, Zhang Shu Yin (2004) Height–Diameter Models Using Stand Characteristics for Pinus banksiana and Picea mariana. Scandinavian Journal of Forest Research 19(5):442–451.
50. Harrison WC, Burk TE, Beck DE (1986) Individual tree basal area increment and total height equations for Appalachian mixed hardwoods after thinning. Southern Journal of Applied Forestry 10(2):99–104.
51. Von Gadow K, Hui G (2001) Modelling forest development. Springer Science & Business Media.
52. Cox F (1994) Modelos parametriz a dos de altura Informe de convenio de investigación interempresas. INFORA, Santiago 28.
53. Curtis RO (1970) Stand density measures: an interpretation. Forest Science 16(4):403–414.
54. Burkhart HE, Strub MR (1974) A model for simulation of planted loblolly pine stands. J. Fries. Royal College of Forestry, Stockholm, Sweden 128–135.
55. Páscoa F (1987) Estrutura, crescimento e produção em povoamentos de pinheiro bravo. Um modelo de simulação.
56. Lei Y, Parresol B R (2001) Remarks on Height-Diameter Modeling. 010.
57. Muller-Landau H., Detto M., Chisholm R. A., Hubbell S., Condit R. (2014). Detecting and projecting changes in forest biomass from plot data. Forests and Global Change. 17. 381–416.