Studying the Growth Prediction Model of Eucalyptus Research based on Multivariate Statistical Analysis

Zhe Yang*, Weikun Wu, Jingshang Chen, Yongsheng Lu and Jianan Yao
Zhaoqing Power Supply Bureau of Guangdong Power Grid Company Limited, Zhaoqing, Guangdong, China

*Corresponding author: zhe_yang@gyzq.csg.cn

Abstract. Forests have ecological functions such as purifying the atmosphere, fixing carbon and releasing oxygen, conserving water sources, preventing wind and sand invading, and harbouring biodiversity. At the same time, they provide human beings with materials and energy, and thus are important for economic and societal development. However, in our national power grid system, trees might cause interference on power transformation lines. Furthermore, if these circumstances cannot be responded timely, huge economic losses would happen. Therefore, research of forest growth models can more accurately predict the growth and the dynamics of forests, as well as provide a reference for the power warning system. This study conducts a forecast simulation study on the growth of eucalyptus in Guangzhou province, China, proposes a multivariate statistical analysis to predict tree species growth, and then compares its performances with traditional Logistic model results. We found that our multivariate statistical model can effectively consider various environmental influencing factors, thus predict tree height excellently ($R^2=0.9776$), so as provide new foundations for tree height prediction.

1. Introduction
The forest ecosystem has the most complex structure, harbors the richest species, and produces the most effective functions in our earth planet. It plays a decisive role in maintaining the global ecological and environmental balance. As main body of forest ecosystem, forests provide material foundation and resources for not only mankind but for economic and social development [1]. Yet, human’s land cover transformation from agricultural civilization to industrial civilization, as well as other factors such as wars and population explosions, had caused forest resources been exploited and utilized in a plundered manner in the past few centuries. Those severe damage towards land cover has accompanied by a series of ecological and environmental problems nowadays. Facing with current severely situation of tightening resource constraints, serious environmental pollution, and degradation of the ecosystem, it is necessary to establish an ecological civilization awareness of respecting, and protecting nature, and to follow a path of sustainable development. Thus, an important aspect of solving ecological environmental problems and building ecological civilization is to understand the growth laws of forests, to scientifically manage forest resources, and to realize the sustainable development of forest resources [2].
Eucalyptus belongs to the Myrtaceae tree family, which is one of the three most fast-growing tree species in the world. It has advantages of strong adaptability, short cultivation cycle, high yield, and wide utilizing ranges. It provides a large amount of wood raw materials for world timber market and remains an important tree species for commercial forest production in China. Studying the growth model of eucalyptus means that to describe the changing process of the forest, to calculate mature ages, and to provide a better management plan for forestry development. Specially, for the Southern Power Grid system in China, eucalyptus will cause damages to the high-voltage power line systems in the wild environment, partly due to the rapid growth of eucalyptus trees, thus causing serious effects to the power transmission system, and further resulting in serious economic losses. Therefore, based on lidar point cloud data, fully studying the growth predicting model of eucalyptus can effectively reflect the growth of eucalyptus.

When analyzing whole growth process of eucalyptus, previous studies reveal that their growth follows a stable S curve, although there exist some small fluctuations influenced by the surrounding environment. Because in a specific forest stand, the growth of a single tree in the stand is theoretically an "S"-curve, the overall growth of the stand composed of a single plant also shows a "slow-fast-slow" trend. In addition, the cumulative distribution of the percentage of the number of trees in the diameter class caused by the differentiation also showed an "S"-shaped state. Therefore, the theoretical growth equation of "S" can be applied to the growth or distribution of individual trees, stand, and diameter levels, and can be used to construct first, second and third types of models. Different theoretical growth equations are suitable for describing different objects. As far as the same object is concerned, the number of parameters, structure composition and inflection point of the "S" type equation have a crucial effect on the simulation accuracy of the equation. At present, the application field of the theoretical growth equation is very broad, and the main research applications are: Gompertz equation [3], Logistic equation [4], Bertalanffy equation [5], Richards equation [6], Schumacher equation [7] and Korf equation [8]. Some researchers used Weibull distribution, reverse-Weibull distribution, Logistic equation and other models to fit the diameter distribution of Eucalyptus grandis plantations, and found that the diameter distribution model established by Logistic equation can be effectively utilized to the growth prediction of eucalyptus. They used the Richards growth function as a model to fit a eucalyptus clone stand growth and harvest prediction model. Others also established a relative tree height curve model for eucalyptus plantations, or fitted eucalyptus Site index model.

On the other hand, the theoretical growth model has limited consideration factors, and the eucalyptus growth model based on multivariate statistical analysis also needs to be further developed. Therefore, this study uses multivariate statistical analysis to analyze the impact of different environmental factors on the growth of eucalyptus, thereby constructs a eucalyptus growth model, and compares with the conventional Logistic model results, so as provides new ideas for the growth prediction of eucalyptus.

2. Material and Methods

2.1. Data
This study used simulated datasets, and all of simulated parameters based on characteristics in Guangdong Province. This region harbors a tropical climate, where annual accumulated temperature is about 8382.3°C, the annual average temperature is 23°C, the annual average rainfall fluctuates between 1400 and 1700 mm, and the annual average relative humidity is 84%. There also exist distinct dry and wet seasons. The rainy seasons long from June to September, and the dry season is from November to March every year. Through overall analysis of the collected characteristics, the initial parameters of the simulation are given, and thus simulation data follow Gaussian normal distributions. In our study, 420 sets of samples are simulated, and the relevant meteorological data is obtained directly from the website: http://data.sheshiyuanyi.com/WeatherData/.
2.2. Method
When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. Should authors use tables or figures from other Publications, they must ask the corresponding publishers to grant them the right to publish this material in their paper.

2.2.1. Logistic Model. Logistic method, initially describing the law of population growth, was created by a Belgian mathematician called Verhulst [4]. Later, it was used in ecology and thus gradually became one of the most famous and excellent equation in this field. Nowadays, scholars have paid extensive attention towards it, some researchers had applied it into the study of forest structure. This equation reflects that there exists a linear relationship between the growth rate of biological populations and the surrounding nutrient contents, through which an extremely complex relationship in biology could be expressed approximately [9, 10]:

\[
\frac{dy}{dx} = y(1 - y/Y_{\text{max}}) \quad (1)
\]

\[
y = Y_{\text{max}} \left[1 + \left(\frac{Y_{\text{max}} - y_0}{y_0}\right)e^{-d(x - x_0)}\right] \quad (2)
\]

\[
y = C / \left(1 + e^{p-q\theta}\right) \quad (3)
\]

In (3), where C and (2) are both upper equally asymptotic values; \( p = \ln\left(\left(\frac{Y_{\text{max}} - y_0}{y_0}\right)+dx_0, q=d, \right. \)

\( q=Y_{\text{max}} \), which is the intrinsic growth rate; \( x_0, \ y_0 \), are two initial values of \( x \) and \( y \) respectively, and \( C \), \( q>0 \). Logistician is widely utilized successfully in ecology and economics fields, mainly due to its theoretical base of nutrition dynamics, and its standard S-curve law, to which lots of natural and societal changes hew. However, although the Logistic equation is conformed with the reality of nature, it cannot be established when nutrients are abundant (such as in field cultivation and microorganism fermentation). To solve this problem, researchers then proposed a new theoretical growth model, which is the fusion and expansion of the Malthusian exponential equation and the logistic equation [11]. As following equation (4):

\[
y = C / \left(1 + e^{p-q\theta}\right)^\theta \quad (4)
\]

Where the parameter \( \theta \) in the equation is a constant. \( C, q>0 \), the inflection point of (4) can be obtained at about 1/2, which greatly improves the flexibility of the logistic equation in theory.

2.2.2. Back Propagation Neural Network. The BPNN models include at least three layers: input layer, hidden layer, and output layer [12]. In the forward propagation process, the relationship between the input layer and the output layer can be expressed in the following form:

\[
Y_i^k = f(x_i^k) \quad (5)
\]

\[
x_i^k = \sum_j W_{ij}Y_j^{k-1} \quad (6)
\]

Where \( x_i^k \) is the input of the \( i \)-th node of the \( k \)-th layer; \( Y_i^k \) is the output of the \( i \)-th node of the \( k \)-th layer; \( W_{ij} \) is the weight from the node of the \( j \)-th unit of the \( k-1 \) layer to the \( i \)-th node of the \( k \) layer. If the desired output value is not obtained in the output layer, the error will be backpropagated, and the
errors of each node in the output layer will be returned layer by layer along the connected place, backpropagated to the input layer and then equally distributed to each node. In order to obtain the reference error of each layer node, to adjust the weight of each node, it is hoped that the error signal tends to be the smallest [13]. The average error function is:

\[ E = \frac{1}{2} \sum_{j} (Y_j^m - y_j)^2 \]  

(7)

2.3. Performance analysis
The validation and evaluation process are carried out through the measured dataset, and the retrieval results of different models are compared. The evaluation indicators are \( R^2 \) which is the coefficient of determination. Sets \( y_j \) and \( y'_j \) denote the measured and predicted trait content, respectively. \( \bar{y} \) refers to the average of \( y_j \), and \( n \) is the number of measurements:

\[ R^2 = 1 - \frac{\sum_{j=1}^{n} (y'_j - y_j)^2}{\sum_{j=1}^{n} (y'_j - \bar{y})^2} \]  

(8)

3. Results and discussion

3.1. Correlation between tree ages and heights
The heights of trees at different growth stages were analyzed, and then illustrating the relationship between tree heights and ages.

![Figure 1](image)

**Figure 1.** The relationship between tree heights and tree ages for eucalyptus.

As shown in Figure 1, tree heights show a "S" change with tree ages, conformed with previous researches. From the beginning, trees grow slowly, but following a rapid growth process. With the time goes by, when reaching a certain age threshold, trees grow slowly again. Therefore, the Logistic model can be used for simulation analysis. What’s more, tree height corresponding to different tree ages also has a certain correlation, which provides a basis for the use of multivariate statistical analysis to analyze the change of tree height.
3.2. Retrieval Results Based on Logistic Eucalyptus Growth Model

Taking into account the fast-growing characteristics of eucalyptus, the annual growth and harvest prediction model has a large gap among adjacent ages. The eucalyptus growth prediction model constructed in this research takes 6 months, tantamount to 0.5 years, as the node, which can reflect multiple growth processes at the time node. Based on the simulated data, the Logistic eucalyptus growth model can be obtained as:

\[
L = \frac{13.98 \cdot H_T}{17.8158 / (1 + 7.9085 \exp(-1.2351t))}
\] (9)

According to statistical analysis, the Logistic curve has an achievable effect as a position index guide curve.

\[
H_T = \frac{17.7184}{1 + 7.9085 \exp(-1.2351t)}
\] (10)

Based on the constructed Logistic model, tree heights are further analyzed to be fitted and compared with the measured tree height, as shown in Figure 2. Through analysis, it is found that the predicted tree heights have an excellent correlation with the measured tree heights. \(R^2\) reaches 0.9386, which is consistent with than revealed in previous literature, proved that the Logistic model can simulate a tree growth process effectively.

![Figure 2. The relationship between the predicted tree height based on the Logistic model and the measured tree height.](image)

\[\begin{align*}
y &= 0.9586x + 0.18 \\
R^2 &= 0.9386
\end{align*}\]

3.3. Retrieval Results based on BPNN model

In BPNN model, 70% of the data is partitioned as the training dataset and the residual 30% as the verification dataset. The input parameters are tree ages and climatic conditions of the year such as annual temperature, annual precipitation, and annual average relative humidity, whereas the output parameters are tree heights. In this way, the correlation between the predicted and measured tree heights can be obtained as shown in the following Figure 3:
Figure 3. The relationship between the predicted tree height based on the BPNN model and the measured tree height.

It can be seen from the figure that the predicted tree heights have a good correlation with the measured tree heights. \( R^2 \) reaches 0.9776, which shows an effectiveness of the constructed multivariate statistical model in predicting tree heights. Compared with traditional Logistic model, the newly proposed multivariate statistical model can also be effectively applied to the prediction of tree height, thus providing a new idea for the development of tree height prediction.

The expression of the tree growth prediction model is generally a set of mathematical functions, which represents the growth process of the trees, also including the relationship between stand ages, site factors, stand density index, stand areas, and stand volumes. After maturing, trees’ growth process would slow down. Therefore, in the power lines’ early warning research, it is necessary to fully consider the impact of the tree growth process towards the power lines. Therefore, fully obtaining the tree growth process is vitally crucial for power lines early warning. This study proposes a multivariate statistical analysis method, which can effectively and comprehensively consider the tree species themselves and various meteorological factors to predict and analyze the tree heights. However, further research such as the optimal intermediate layers and the selection of the optimal functions will further affect the accuracy of the model, which is also a problem that needs to be considered in the future.

4. Conclusion
This study compared and analyzed different growth models, respectively, the multivariate statistical analysis model and the Logistic model to predict eucalyptus growth. Results demonstrated that Logistic model can predict the growth of eucalyptus to a certain extent and the performances are still worse than the proposed multivariate statistical analysis model. Thus, multivariate statistical analysis can effectively improve the accuracy of eucalyptus growth prediction. In addition, this study considered other influencing factors in the model construction, thus providing new reference for eucalyptus growth prediction. What’s more, this study only regarded the feasibility of the multivariate statistical model without considering the parameter settings and whether it is applicable to other tree species. Therefore, relating research is still needed for further investigation in the future.

Acknowledgments
This work was financially supported by Science and Technology Project of China Southern Power Grid Corporation (No. (031200KK52190099) GDKJXM20198220).
References
[1] B. Zeide, Accuracy of equations describing diameter growth. Can. J. For. Res. 19 (1989) 1283-1286.
[2] H. Wakt. Sustainable development and sustainable forestry: analyzes, differences and the role of flexibility. Eur. J. Forest Res. 129 (2010) 781-801.
[3] B. Gompertz, On the nature of the function expressive of human mortality, and on a new mode of determining the value of life contingencies. Phil. Transac. Roy. Soci. London 115 (1825) 513-585.
[4] M. Vogels, R. Zoeckler, D.M. Stasiw, L.C. Cerny, Verhulst's "notice sur la loi que la populations suit dans son accroissement" from correspondence mathematique et physique. Ghent, vol. X, 1838. J. Biol. Phys. 3 (1975) 183-192.
[5] Bertalanffy L. Von. Quantitative laws in metabolism and growth. Quart. Rev. Biol. 32 (1957) 217-231.
[6] F.J. Richards, A flexible growth function for empirical use. J. Exp. Bot. 29 (1959) 290-300.
[7] F.X. Schumacher, A new growth curve and its application to timber-yield studies. J. For. 37 (1939) 819-820.
[8] A.K. Kiviste, Mathematical functions of forest growth. Estonian Agricultural Academy, Tartu, USSR. 1988.
[9] A.G. Duan, J.G. Zhang, S.Z. Tong, Application of six growth equations on stands diameter structure of Chinese fir plantations. Forest Res. 16 (2003) 423-429.
[10] Q.Li, X.B. Tang, Y.H. Ma, Logistic Model Based on the Gray Tree Growth Studies. J. Hubei Univer. Nation., 29 (2011) 473-475.
[11] B. Mondal, A. Garai, A. Mukhopadhyay, S.K. Majumder, Inventory policies for seasonal items with logistic-growth demand rate under fully permissible delay in payment: a neutrosophic optimization approach. Soft Comput. 25 (2021) 3725-3750.
[12] L.E. Keiner, X.H. Yan, A neural network model for estimating sea surface chlorophyll and sediments from thematic mapper imagery. Remote Sens. Environ. 66 (1998) 153-165.
[13] Q.X. Yi, J.F. Huang, F.M. Wang, X.Z. Wang, Z.Y. Liu, Monitoring rice nitrogen status using hyperspectral reflectance and artificial neural network. Environ. Sci. technol. 41 (2007) 6770-6775.