

Research Article

Algorithm for Tree Growth Modeling Based on Random Parameters and ARMA

¹Lichun Jiang, ¹Fengri Li and ²Yaoxiang Li

¹College of Forestry,

²College of Engineering and Technology, Northeast Forestry University, Harbin 150040, P.R. China

Abstract: Chapman-Richards function is used to model growth data of dahurian larch (*Larix gmelinii* Rupr.) from longitudinal measurements using nonlinear mixed-effects modeling approach. The parameter variation in the model was divided into random effects, fixed effects and variance-covariance structure. The values for fixed effects parameters and the variance-covariance matrix of random effects were estimated using NLME function in S-plus software. Autocorrelation structure was considered for explaining the dependency among multiple measurements within the individuals. Information criterion statistics (AIC, BIC and Likelihood ratio test) are used for comparing different structures of the random effects components. These methods are illustrated using the nonlinear mixed-effects methods in S-Plus software. Results showed that the Chapman-Richards model with three random parameters could typically depict the dahurian larch tree growth in northeastern China. The mixed-effects model provided better performance and more precise estimations than the fixed-effects model.

Keywords: Fixed effects, modeling algorithm, nonlinear mixed effects, random effects, tree growth

INTRODUCTION

Analysis of repeated measurement data is a recurrent challenge to statisticians engaged in biological and biomedical applications. A common type of repeated measurement data is longitudinal data. Longitudinal data can be defined as repeated measurement data where the observations within individuals could not have been randomly assigned to the levels of a treatment of interest. Repeated measurements data are data in which multiple individuals have multiple measurements over time or space. Thus the method of analysis for this type of data needs to recognize and estimate two distinct types of variability: between-individual variability and within-individual variability. Nonlinear mixed effects models provide a tool for analyzing repeated measurements data and give an unbiased and efficient estimation of the fixed parameters of the model. Furthermore, nonlinear mixed models improve predictive ability if we are able to predict the value of the random parameters for an unsampled location. This is possible if complementary observations of the dependent variable are available. There has been a great deal of recent interest in mixed effects models for repeated measures data in forestry (Calama and Montero, 2004; Gregoire *et al.*, 1995; Jiang and Li, 2008a, b, c).

Forest biometricians have been developing and adapting statistical techniques to improve tree growth prediction to meet particular objectives for forest

growth and yield system. The aim of this study was to investigate the algorithm for dahurian larch (*Larix gmelinii* Rupr.) tree grow modeling using nonlinear mixed-effects modeling approach with random parameters and autocorrelation.

MODELING ALGORITHM

Algorithm for nonlinear mixed modeling: The general expression for a nonlinear mixed effects model can be written as Pinheiro and Bates (2000):

$$y_{ij} = f(\beta_i, t_{ij}) + \varepsilon_{ij} \quad i = 1, \dots, m, \quad j = 1, \dots, n_i \quad (1)$$

where,

- y_{ij} = A diameter measurement at time j for the i th tree
- t_{ij} = Age (years) for the i th tree on time j
- m = The number of trees, n_i is the total number of measurements on tree i
- f = A nonlinear function, β_i is a parameter vector, ε_{ij} is a normally distributed within-plot error term

The parameter β_i varies from tree to tree to account for between-individual variation. The parameter β_i can be expressed as:

$$\beta_i = A_i\beta + B_i b_i \quad b_i \sim N(0, D) \quad (2)$$

where β is a $p \times 1$ vector of fixed effects parameters, b_i is a $q \times 1$ vector of random effects parameters, p is the

Corresponding Author: Yaoxiang Li, College of Engineering and Technology, Northeast Forestry University, Harbin 150040, P.R. China

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

number of fixed parameters in the model, q is the number of random parameters in the model, D is the variance-covariance matrix for the random effects. In the basic assumptions, the within-tree errors are independent and the ε_i follow a $N(0, R_i)$ distribution and are independent of b_i , A_i and B_i are design matrices for the fixed and random effects respectively. These design matrices contain zeros and ones associated with the fixed and random effects.

After initial screening of some growth models, such as Chapman-Richards, Logistic, Weibull and Gompertz, the Chapman-Richards model was selected for this study due to its flexibility and biologically interpretable coefficients. The form of the Chapman-Richards model is:

$$y = \beta_1(1 - \exp(-\beta_2 t))^{\beta_3} \quad (3)$$

where, y is tree diameter (cm), t is age (years), β_1 , β_2 , β_3 are regression coefficients.

Model development: Commonly, three steps were involved for mixed-effects model development (Pinheiro and Bates, 2000). First, the parameters of the model were specified either as mixed (both random and fixed) or purely fixed. Then the structure of the among-tree variance-covariance matrix (D) was determined. Finally, the within-tree variance-covariance structure (R_i) was determined to account for heteroscedasticity and residual correlation.

Data from stem analysis of 40 trees were collected from Daxinganling region in Heilongjiang Province, northeastern China. This species is common and commercially important in northeast China. Sample trees were selected from the dominant and codominant crown class, undamaged and unsuppressed. Each sample tree was felled and disks were extracted at breast height (1.3 m). The number of annual rings and ring width were measured for each disk. The 40 sample trees yielded 528 diameter-age pairs.

Parameter effects were analyzed first. For the construction of a mixed model, the first question that should be addressed is determining which effects should be considered mixed and which should be considered purely fixed. An intuitive approach is to fit diameter-age models to each individual tree and assess the variability of estimated parameters by considering the individual confidence intervals (Lindstrom and Bates, 1990). The parameters with high variability and less overlap in confidence intervals across trees should be considered as mixed effects. This approach requires sufficient observations on each individual to give meaningful parameter estimates by the individual fits. In our case there are over 15 observations for each tree and thus it is possible to get the confidence intervals of the parameters for each individual fit.

Variance-covariance structure was further investigated. The among-tree variance-covariance matrix for the random-effects (D), common to all trees, defines the existing variability among tree. First, we assume that the random effects are independent of each other which make a diagonal random effects variance-covariance matrix to reduce the number of parameters in the model, then different random effects variance-covariance structures (e.g. diagonal and general positive-definite matrices) will be explored to determine whether a correlated variance-covariance structure is needed for random effects.

To specify the within-tree variance-covariance structure (R_i), two components of the heteroscedasticity and the autocorrelation structure should be addressed (Pinheiro and Bates, 1998). The expression for the within-tree variance-covariance matrix is then having the following form:

$$R_i = \sigma^2 G_i^{0.5} \Gamma_i G_i^{0.5} \quad (4)$$

where, σ^2 is a scaling factor for the error dispersion given by the value of the residual variance of the model and G_i is a $n_i \times n_i$ diagonal matrix that specifies within-tree variance, Γ_i is correlation structure, i is the tree number and n_i indicating total number of diameter-age measurements in the tree.

Model evaluation and prediction: To evaluate model performance, Bias, root mean square error (RMSE) and coefficient of determination (R^2) were employed. These evaluation statistics are defined as:

$$Bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}}$$

$$R^2 = 1 - \left[\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right]$$

where, y_i is observed value for the i^{th} observation, \hat{y}_i is predicted value for the i^{th} observation, \bar{y} is the mean of the y_i , n is the total number of observations used to fit the model. The model with the smallest values of Bias, RMSE and higher R^2 was considered the best. Nonlinear mixed-effects models were fitted with the NLME procedure in S-Plus 2000 (Mathsoft, Seattle, WA).

The main purpose of final model involves diameter prediction for unmeasured trees. An important characteristic, compared to traditional regression, is that mixed-effects models allow for both mean response and calibrated prediction.

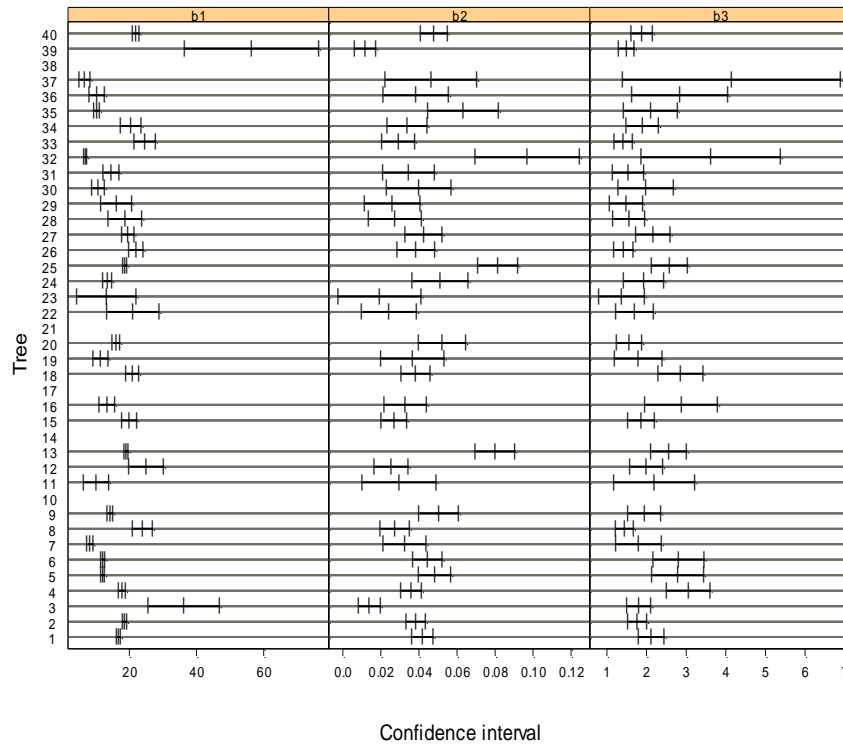


Fig. 1: Ninety-five percent confidence intervals on the parameters in the Chapman-Richards model based on individual fits

If no sub-sample tree diameters have been measured for a particular year, the parameters of random effects is not predicted. Thus, the expected value 0 is used for all random parameters. The mean diameter prediction is obtained using only the fixed parameters estimates, where the mean behavior of the diameter variation for a certain tree is represented.

A calibrated response requires prior measured sub-sample diameter information for a tree. It could significantly increase the accuracy of tree diameter prediction. If both diameter and age have been measured for a sub-sample, the vector of random effects b_k at the tree level can be estimated using this additional information. Calculation is carried out using an approximate Bayes estimator of b_k (Vonesh and Chinchilli, 1997; Fang and Bailey, 2001):

$$\hat{b}_k \approx \hat{D}\hat{Z}_k^T(\hat{Z}_k\hat{D}\hat{Z}_k^T + \hat{R}_k)^{-1}\hat{e}_k \quad (5)$$

where \hat{D} is the $q \times q$ variance-covariance matrix (q is number of random-effects parameters included in the model) for the among tree variability, \hat{R}_k is the $k \times k$ variance-covariance matrix for within-individual variability, \hat{Z}_k is the design matrix for the random components specific to the additional observations, \hat{e}_k is the difference between the observed diameter and the predicted diameter using the fixed effects parameters.

RESULTS AND DISCUSSION

Determining parameter effects: Individual fits approach was first used to determine parameter effect either as mixed or purely fixed. Confidence intervals were obtained on the parameters in Chapman-Richards model showed in Eq. (3) based on individual fits using nlsList function in S-Plus. Figure 1 gives the approximate 95% confidence intervals for three parameters of b_1 , b_2 and b_3 in Eq. (3) for each tree except that several trees failed to converge. It was noticed that, for parameters b_1 , b_2 and b_3 , the confidence intervals of the thirty five trees showed more among-tree variability. Therefore, parameters b_1 , b_2 and b_3 were considered mixed effects.

In order to get reliable random parameters, The Chapman-Richards model with different combinations of fixed and random parameters was fitted using nonlinear mixed-effects modeling procedure with NLME function in S-Plus. Summary of fit statistics obtained from fitting the Chapman-Richards model are presented in Table 1.

It indicates that there are significant differences between models containing random parameters and model with only fixed parameters. These statistics show that addition of random-effects improves the model fit. Chapman-Richards model with three random parameters is the best with the smallest AIC and BIC values of 900.02 and 942.77, respectively.

Table 1: AIC, BIC, log-likelihood, and Likelihood Ratio Tests (LRT) with different fixed and random effects components

Model	Random parameters	No. of parameters	AIC	BIC	Log-likelihood	LRT	p-value
1	b_1, b_2, b_3	10	900.02	942.76	-440.01	77.51	<0.0001
2	b_1, b_2	7	971.53	1001.46	-478.77	601.40	<0.0001
3	b_2, b_3	7	1108.67	1138.59	-547.33		
4	b_1	5	1631.42	1652.79	-810.71		
5	b_2	5	1568.89	1590.26	-779.44	1191.48	<0.0001
6	None	4	2758.42	2775.51	-1375.21		

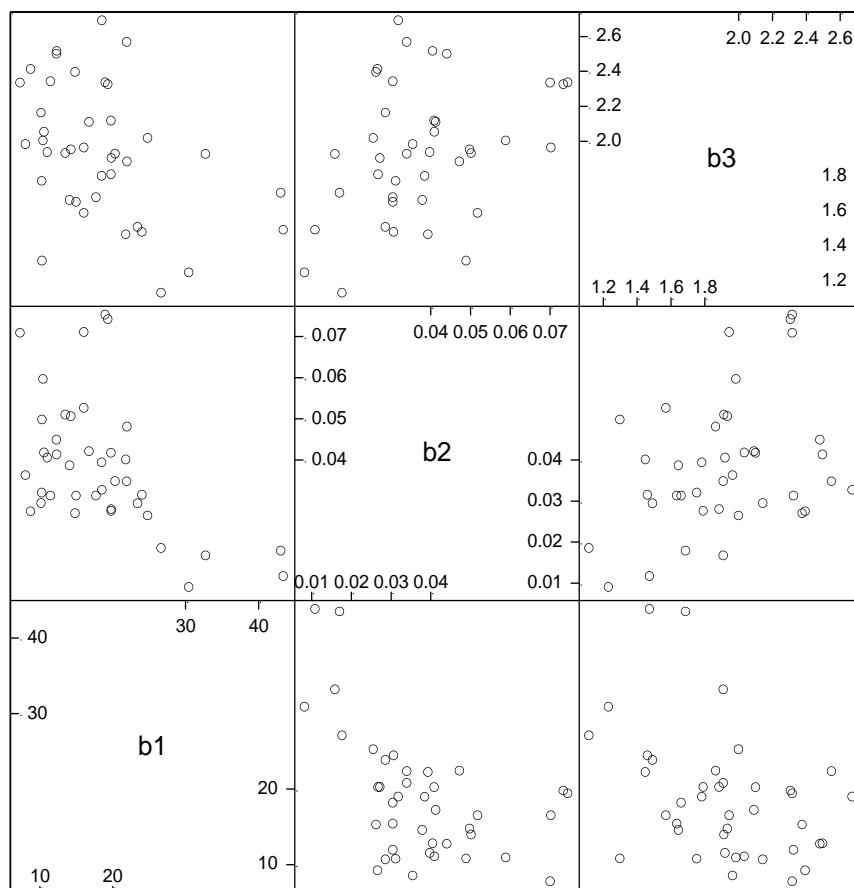


Fig. 2: Scatterplot of the estimated random effects from the fixed model

Table 2: Comparison of mixed-effects model performance with different variance-covariance structures (Diag = Diagonal, Sym = General positive-definite)

Var-Cov Structure	No of parameters	AIC	BIC	Log-likelihood	LRT	p-value
Sym	10	900.02	942.76	-440.01	84.67	<0.0001
Diag	7	976.70	1002.35	-482.35		

To avoid over-parameterization and ill-conditioning problems of variance-covariance matrix, the scatter-plot matrix of the estimated random parameters was produced (Fig. 2). Random parameters b_1 , b_2 and b_3 did not show high correlation among random effects.

Variance-covariance structure: To get the among-tree variance-covariance structure, the Chapman-Richards model with three random parameters was fitted to the data using NLME function in S-Plus. Both diagonal (Diag) and general positive-definite symmetric (Sym)

variance-covariance structures were examined for the random effects (D) using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Likelihood ratio test (LRT) (Table 2).

It indicated that two variance-covariance structures are significant different (LRT = 84.6, $p < 0.0001$). Chapman-Richards model with symmetric variance-covariance structure resulted in lower values of AIC (900.02) and BIC (942.77), indicating that a correlated variance-covariance structure was preferred in the model fitting.

Table 3: Comparisons of mixed-effects model performance with different correlation structures.

Model	Correlation structure	No. of parameters	AIC	BIC	Log-likelihood	LRT	p-value
1	None	10	900.02	942.76	-440.01		
1.1	AR(2)	12	631.29	682.59	-303.64	272.72	<0.0001
1.2	ARMA(1,1)	12	648.58	699.88	-312.29	255.43	<0.0001
1.3	ARMA(1,2)	13	640.13	695.70	-307.06	265.88	<0.0001
1.4	ARMA(2,1)	13	613.64	669.21	-293.82	292.37	<0.0001
1.5	ARMA(2,2)	14	615.35	675.19	-293.67	292.67	<0.0001

Table 4: Parameter estimates and variance components

	Parameter	Fixed model				Mixed model			
		Estimate	S.E.	t-value	p-value	Estimate	S.E.	t-value	p-value
Fixed parameters	β_1	15.1778	0.8804	17.2390	<0.0001	18.0089	1.0604	16.9832	<0.0001
	β_2	0.0411	0.0072	5.6782	<0.0001	0.0371	0.0027	13.6444	<0.0001
	β_3	1.9025	0.3183	5.9765	<0.0001	1.9200	0.0740	25.9477	<0.0001
Variance components	σ^2_{b1}					35.5281			
	σ^2_{b2}					0.0002			
	σ^2_{b3}					0.0040			
	σ_{b2b3}					-0.0271			
	σ_{b1b3}					0.1766			
	σ_{b2b3}					-0.0007			
	σ^2	10.4592				0.2038			
Correlation parameters	ϕ_1					1.5466			
	θ_1					-0.8282			
	θ_2					-0.9967			
Goodness-of-fit	Bias	-0.0135				-0.0043			
	RMSE	3.2249				0.3012			
	R ²	0.6312				0.9968			

Correlation structures: Correlation structures are used to model dependence among observations. In the context of mixed-effects models, they are used to model dependence among the within-group errors. Historically, correlation structures have been developed for two main classes of data: time-series data and spatial data. The former is generally associated with observations indexed by an integer-valued time variable, while the latter refers primarily to observations indexed by a two-dimensional spatial location vector. In this study, the sample data is tree growth data in which serially correlated errors usually arise in the fitting of growth curves to the sample data. Correlation structures can be incorporated into mixed-effects models (Chi and Reinsel, 1989; Lindstrom and Bates, 1990). In this study, the AR (p), MA (q) and ARMA (p, q) correlation structures were tested for inclusion in the mixed-effects model (Table 3). The mixed model with different correlation structures produced significant improvements in model fitting except that several models failed to converge (p<0.0001). ARMA (2, 1) correlation structure was selected as the best one because the mixed model with ARMA (2, 1) correlation structure resulted in smaller AIC and BIC values, the ARMA (2, 1) could explain more dependency among repeated measurements within the individual. The final model is mixed-effects model with autocorrelation structure ARMA (2, 1). Parameters estimates for the final model are presented in Table 4.

Scatter plots of Studentized residuals were constructed for mixed-effects and fixed-effects models (Fig. 3). The Scatter plots showed that the mixed-

effects model showed more homogeneous residual variance over the full range of the predicted values and no systematic pattern in the variation of the residuals. It indicated that mixed-effects model significantly improve the model performances compared to fixed-effects model. In this case, the within-tree variance-covariance matrix is the diagonal matrix:

$$R_i = \sigma^2 \Gamma_i I_{n_i} \tag{6}$$

where I_{n_i} is the identity matrix ($n_i \times n_i$) and other variables as previously defined.

Based on the above considerations, the resulting mixed diameter-age model was:

$$\begin{cases} y_{ij} = (\beta_1 + b_{1i})[1 - \exp[-(\beta_2 + b_{2i})t_{ij}]]^{(\beta_3 + b_{3i})} + \varepsilon_{ij}, \\ (b_{1i} \ b_{2i} \ b_{3i})^T \sim N(0, D), \\ D = \begin{bmatrix} \sigma_{b_1}^2 & \sigma_{b_1 b_2} & \sigma_{b_1 b_3} \\ \sigma_{b_1 b_2} & \sigma_{b_2}^2 & \sigma_{b_2 b_3} \\ \sigma_{b_1 b_3} & \sigma_{b_2 b_3} & \sigma_{b_3}^2 \end{bmatrix}, \\ \varepsilon_{ij} \sim N(0, R_i), \\ R_i = \sigma^2 \Gamma_i(\phi_1, \theta_1, \theta_2), \\ \Gamma_i(\phi_1, \theta_1, \theta_2) = ARMA(2, 1). \end{cases} \tag{7}$$

Model diagnosis: Before making inferences about a fitted mixed-effects model, we should check whether the underlying distributional assumptions are valid for the data. There are two basic distributional assumptions

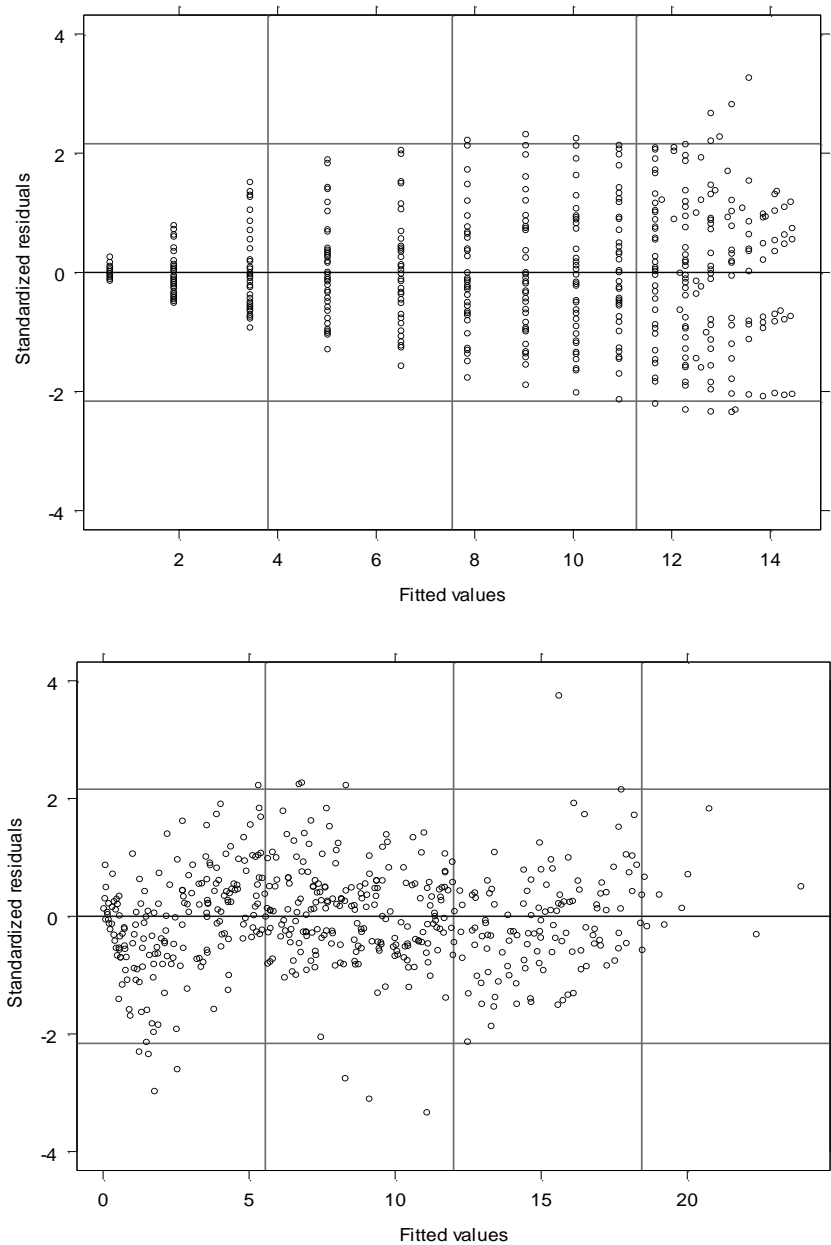


Fig. 3: Studentized residuals for the fixed-effects model (left) and mixed-effects model (right)

for the mixed-effects Richards model considered in this study: The within-group errors are independent and identically distributed, with mean zero and variance σ^2 and they are independent of the random effects; The random effects are normally distributed, with mean zero and variance σ^2_i (not depending on the group) and are independent for different groups.

Scatter plots of Studentized residuals is showed in Fig. 3. It indicates that the residuals are approximately centered at zero, though we observe several outlying observations and large residuals.

Normal plot of estimated random effects was used to investigate departures for the assumption of normality of random effects (Fig. 4).

Parameter estimates and fit statistics for basic and mixed models are presented in Table 4. Compared to the basic model, the mixed-effects model showed the better performance with the lower bias, RMSE and higher R^2 . The performance of the mixed model was also visualized by displaying the predicted and observed values in the same tree (Fig. 5). Both the fixed-effects model with random effects set to zero and mixed-effects model are compared. Diameter growth predictions for all trees obtained from the fixed-effects model differ substantially from the actual diameter. The mixed-effects model more closely followed the actual values for most trees and indicated that mixed effects model described the diameter growth data well.

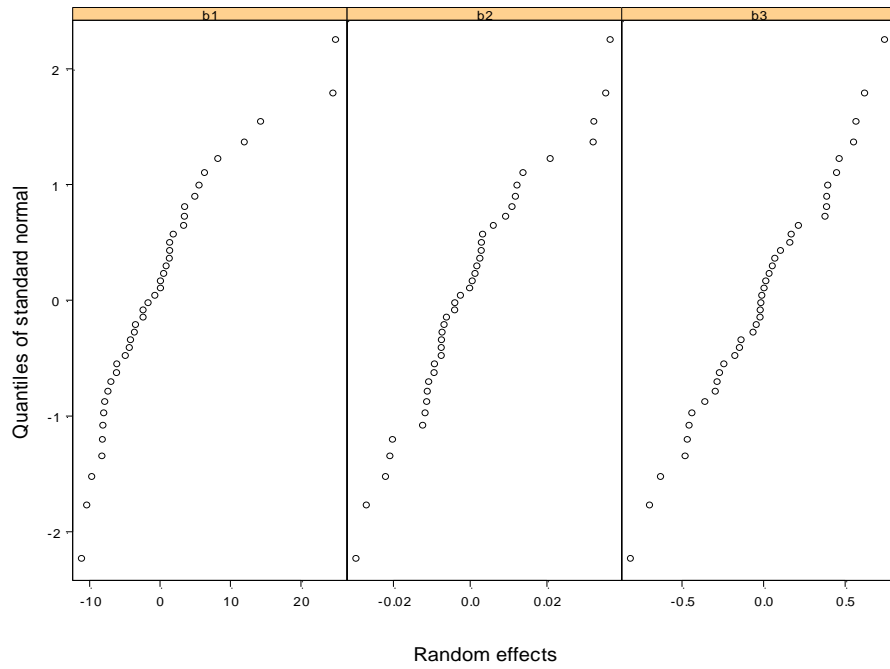


Fig. 4: Normal plot for the estimated random effects

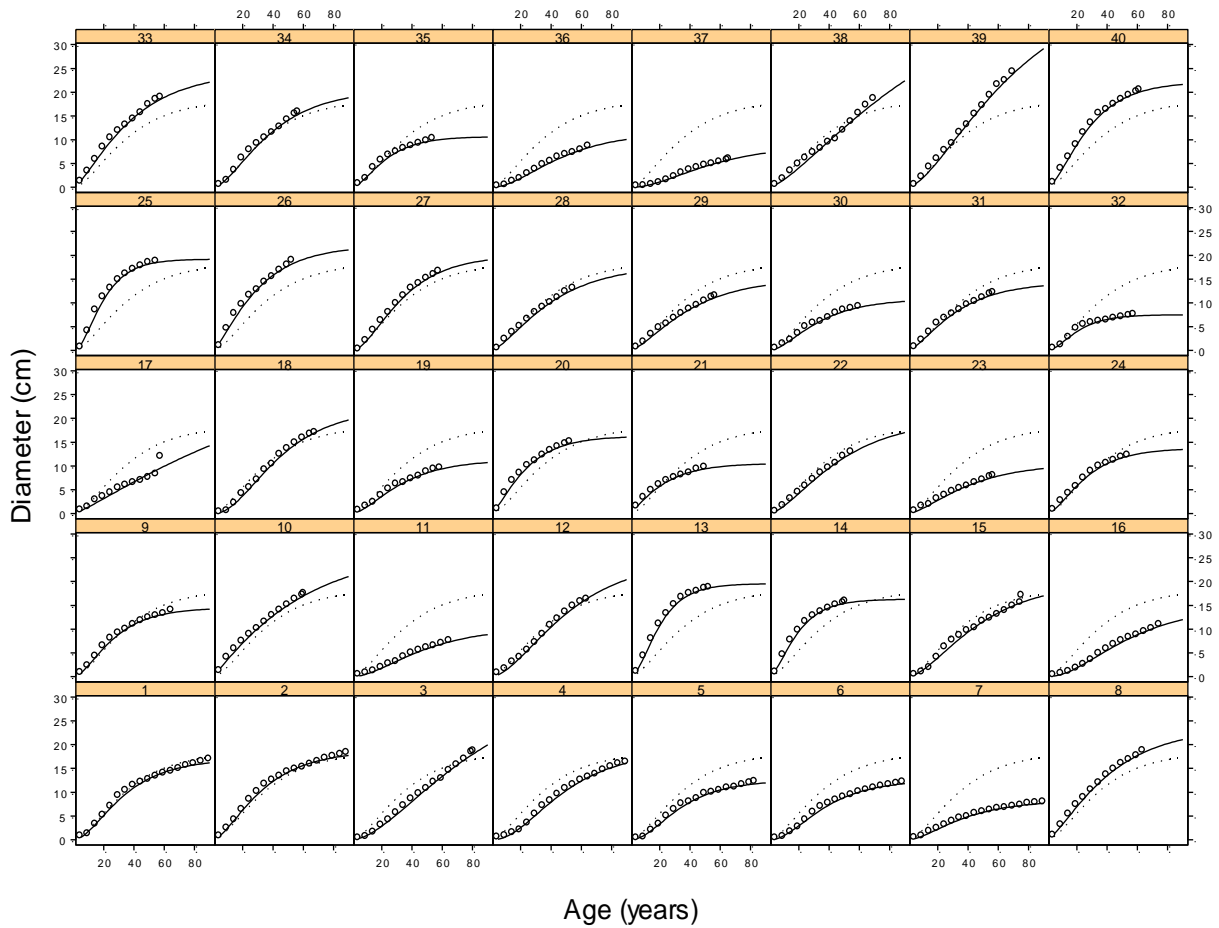


Fig. 5: Predicted versus actual diameters (circle) with (solid lines) and without (dotted lines) random parameters

CONCLUSION

In this study, a nonlinear mixed-effects diameter growth model was developed for dahurian larch in northeastern China. Nonlinear mixed-effects modeling techniques were used to estimate fixed and random-effect parameters for Chapman-Richards model. The results showed that the Chapman-Richards model with three random parameters was found to be the best in terms of goodness-of-fit criteria. The mixed-effects model provided better model fitting and more precise estimations than the fixed-effects model. The fixed-effects parameters alone can be used to obtain the mean diameter curve for dahurian larch plantation based on all trees sampled from the population of plantation in Dailing forest bureau. The random parameters for a tree of interest could be predicted with an approximate Bayes estimator of b_k using the diameter and age measurements from that tree together with estimates of the fixed-effects parameters, the residual variance and the estimated variance-covariance matrix for the random-effects parameters.

The developed models can be implemented in forest inventories by measurement of two tree heights per plot. The process of calibration apparently accounted indirectly for effects of stand density on the height-diameter relationship. The estimate of random-effects parameters makes the inclusion of additional predictor variables in the model unnecessary for many forest inventory applications that consider measurement of sub-sample trees. In addition, the use of the mixed-effects model through a sub-sample of trees for height measurement allows maintenance of a simple model structure without including additional predictor variables.

It should be recognized that the data used in this study were collected from a narrow geographic range, severely restricting the use of the model. Thus, caution should be used when extrapolating beyond the natural range of the data on which the model is based.

ACKNOWLEDGMENT

This study was supported by the Twelfth Five-Year Plan of National Science and Technology Project (2012BAD22B0202), Fundamental Research Funds for the Central Universities (DL12DA01, DL12EB07-2) and NSFC (31170591).

REFERENCES

- Calama, R. and G. Montero, 2004. Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain. *Candian J. Forest Res.*, 34: 150-163.
- Chi, E.M. and G.C. Reinsel, 1989. Models for longitudinal data with random effects and AR (1) errors. *J. Am. Stat. Assoc.*, 84: 452-459.
- Fang, Z. and R.L. Bailey, 2001. Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. *Forest Sci.*, 47: 287-300.
- Gregoire, T.G., O. Schabenberger and J.P. Barrett, 1995. Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent plot measurements. *Canadian J. Forest Res.*, 25: 137-156.
- Jiang, L. and Y. Li, 2008a. Fitting nonlinear mixed models with the NLME procedure in S-Plus. *Proceeding of 2nd International Symposium on Computer Science and Computational Technology. Shanghai, China*, 2: 251-254.
- Jiang, L. and Y. Li, 2008b. Nonlinear mixed modeling approach-an application to tree growth data. *Proceeding of IEEE International Conference on Intelligent Computation Technology and Automation. Changsha, China*, pp: 646-649.
- Jiang, L. and F. Li, 2008c. Modeling tree diameter growth using nonlinear mixed-effects models. *Proceeding of International Seminar on Future Information Technology and Management Engineering (FITME '08), Leicestershire, United Kingdom*, pp: 636-639.
- Lindstrom, M.J. and D.M. Bates, 1990. Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46: 673-687.
- Pinheiro, J.C. and D.M. Bates, 1998. *Model Building for Nonlinear Mixed Effects Models*. Tech. rep., Dep. of Stat., Univ., of Wisconsin.
- Pinheiro, J.C. and D.M. Bates, 2000. *Mixed-Effects Models in S and S-PLUS*. Springer, New York, pp: 528.
- Vonesh, E.F. and V.M. Chinchilli, 1997. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker, New York.