Research Article
Fitting Logistic Growth Curve with Nonlinear Mixed-effects Models

[1]Yaoxiang Li and [2]Lichun Jiang
[1]College of Engineering and Technology,
[2]College of Forestry, Northeast Forestry University, Harbin, 150040, P.R. China

Abstract: A nonlinear mixed-effects modeling approach was used to model individual tree diameter increment based on Logistic growth function for dahurian larch (Larix gmelinii Rupr.) plantations in northeastern China. The study involved the estimation of fixed and random parameters, as well as procedures for determining random effects variance-covariance matrices. Results showed that the mixed-effects model provided better model fitting than the fixed-effects model. The logistic model with three random parameters b_1, b_2, b_3 was considered the best mixed model. Time series correlation structures included Autoregressive correlation structure AR (1) and AR (2), Moving Average correlation structure MA (1) and MA (2) and Autoregressive-Moving Average correlation structure [ARMA (1, 1)] and ARMA (2, 2) were incorporated into the best mixed model. The mixed model with MA (2) correlation structure showed lower AIC and BIC values and significantly improved model performance (LRT = 545.6, $p<0.0001$). Techniques for calibrating the diameter growth model for a particular tree of interest were also explored. The results indicated that the mixed-effects model provided better diameter predictions than the models using only fixed-effects parameters.

Keywords: Diameter growth, logistic function, model calibration, nonlinear mixed-effects modeling

INTRODUCTION

Statistical models in which both fixed and random effects enter nonlinearly are becoming increasingly popular (Wolfinger, 1999). These models have a wide variety of applications in many areas such as agriculture, forestry, biology, ecology, biomedicine, sociology, economics, pharmacokinetics and other areas (Pinheiro and Bates, 1998). Mixed effect models are flexible models to analyze grouped data including longitudinal data, repeated measures data and multivariate multilevel data (Lindstrom andBates, 1990). One of the most common applications is nonlinear growth data (Palmer *et al.*, 1991).

A number of growth functions have been used for modeling tree growth variables (Huang and Titus, 1995; Colbert *et al.*, 2002). Ordinary Linear Square (OLS) and Ordinary Non-Linear least Squares (ONLS) regression techniques have been used to fit these functions. However, tree diameter increment data are generally taken from trees growing in plots located in different stands. This hierarchical structure results in a lack of independence between observations, which results in biased estimates for the confidence interval of the parameters if ordinary least squares regression techniques are used. To deal with this problem, mixed model approaches have been used in tree growth modeling (Fang and Bailey, 2001; Jiang and Li, 2008,

2010; Gregoire and Schabenberger, 1996). Mixed effects models estimate both fixed and random parameters simultaneously for the same model, providing consistent estimates of the fixed parameters and their standard errors. Furthermore, the inclusion of random parameters captures more variation among and within individuals. Mixed models may also improve the predictions through calibration of growth models for a specific location and growth period. Prediction of the random parameters using best Unbiased Predictors (BLUP) can be carried out if supplementary observations of the dependent variable are available.
The objectives of this study were to:

- Develop diameter increment models for dahurian larch in northeastern China based on nonlinear mixed-effects modeling approach
- Determine and account for variance-covariance structure within individual
- Evaluate the predictability of the mixed-effects model based on the calibration

MATERIALS AND METHODS

Data used in this study were obtained from 10 stands of dahurian larch plantations located in Wuying forest bureau in Heilongjiang Province, northeastern China (129°06′-129°30′E, 47°54′-48°19′N). The annual

average temperature is 0.2°C. The frost-free period is around 106 days with annual precipitation about 620 mm. First, a plot of 0.04 ha was established for each stand. In each plot, diameter at breast height (DBH, 1.3 m) was measured with a diameter tape and recorded for every living tree to the nearest 0.1 cm. Total tree height was measured with a Hagloff Laser Vertex Hypsometer and recorded to the nearest 0.1 m. Then, three sample trees were selected from each plot. Stem analysis data were collected by felling the trees, sectioning them and counting the ring on each section. A disk was extracted at 1.3 m to measure diameter increment. A total of 30 trees are available for stem analysis in this study with age ranging from 38 to 53 years old. Ranges of these tree characteristics were 6.0-32.2 cm for diameter at breast height and 6.7-26.2 m for total height. Mean values of diameter and total height for these trees were 20.1 cm and 21.2 m with their standard deviations 5.9 cm and 4.2 m, respectively. The 30 sample trees yielded 636 diameter-age pairs.

After initial screening of some growth models, such as Chapman-Richards, Logistic, Weibull and Compertz, the Logistic model was selected for the mixed effects analysis due to its best overall fit. The form of the mixed Logistic model is:

$$y_{ij} = \frac{w_1}{1 + (w_2)\exp\left(-(w_3)t_{ij}\right)} + \varepsilon_{ij} \qquad (1)$$

where

y_{ij}	: The diameter at time j for the ith tree (cm)
t_{ij}	: The age at time j for the ith tree (years)
w_1, w_2 & w_3	: Parameters
ε_{ij}	: Random error

The parameter w_i varies from tree to tree to account for between tree variations. The parameter w_i can be expressed as:

$$w_i = A_i\beta + B_i b_i \qquad b_i \sim N(0, D)$$

where,

β	= The fixed effects parameter
b_i	= The random effects associated with the ith tree
D	= The variance-covariance matrix for the random effects
A_i & B_i	= Design matrices for the fixed and random effects, respectively

Model construction: According to Pinheiro and Bates (2000), model construction techniques for mixed-effects model involve the following steps:

- Specification of the parametersof the model either to be considered mixed (both random and fixed) or purely fixed.

- Determination of the structure of the among-tree variance-covariance matrix (D).
- Determination of the within-tree variance-covariance structure (R_i) to account for heteroscedasticity and residual correlation.

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 (Pinheiro and Bates, 1998). The parameters with high variability and less overlap in confidence intervals across trees should be considered as mixed effects. A more general approach is to fit different prospective models and compare nested models using likelihood ratio tests or information criterion statistics, such as the Akaike Information Criterion (AIC). To avoid over-parameterization problems and ill-conditioning problems of variance-covariance matrix, both methods were used in this study to determine the parameter effects.

The variance-covariance matrix for the random-effects (D), common to all trees, defines the existing variability among trees. 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) are explored to determine whether a correlated variance-covariance structure is needed for random effects.

To specify the within-tree variance-covariance structure (R_i), the autocorrelation structure must be addressed. Since our sample data is repeated-measurements data from the same tree, the expression for the within-tree variance-covariance matrix has the following form:

$$R_i = \sigma^2 \Gamma_i I_{n_i} \qquad (2)$$

where,

I_{ni}	= An identity matrix ($n_i \times n_i$)
Γ_i	= Correlation structures
σ^2	= Residual variance of the model

Model evaluation: To evaluate model performance, Bias, Root Mean Square Error (RMSE) and coefficient of determination (R^2) were employed. These evaluation statistics are defined as:

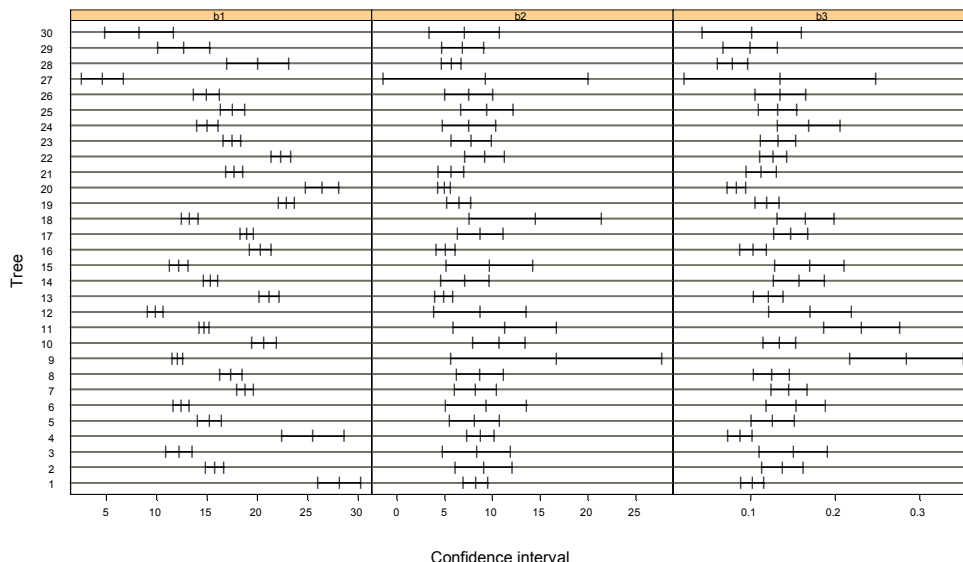$$Bias = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n}$$

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

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

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

where,
y_i = Observed value for the ith observation
$\hat{y}_i$ = Predicted value for the ith observation
$\bar{y}_i$ = The mean of the y_i

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 (Math soft, 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.

If no tree diameter has been measured, the parameters of random effects are not predicted. Thus, the expected value 0 is used for all random parameters. The mean diameter prediction is obtained using the fixed parameters estimates.

A calibrated response requires prior measured diameter information. If both diameter and age have been measured for a subsample of k trees, the vector of random effects b_k can be estimated using this additional information. Calculation is carried out using an approximate Bayes estimator of b_k (Vonesh and Chinchilli, 1997):

$$\hat{b}_k \approx \hat{D}\hat{Z}_k^T\left(\hat{Z}_k\hat{D}\hat{Z}_k^T + \hat{R}_k\right)^{-1}\hat{e}_k \quad (3)$$

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

RESULTS

Individual fits approach was first used to determine parameter effect either as mixed or purely fixed. Confidence intervals were obtained on the parameters in logistic model showed in Eq. (1) 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. (1) for each tree. It was noticed that confidence intervals of three parameters of b_1, b_2 and b_3 showed more among-tree variability. Therefore, parameters b_1, b_2 and b_3 were considered random effects.

To avoid over-parameterization problems and ill-conditioning problems of variance-covariance matrix, the different combinations of fixed and random parameters were also explored (Table 1). The results showed that the best performances were obtained for logistic model with parameters b_1 as mixed effects when considering one random parameter and with parameters b_1 and b_3 as mixed effects when

Table 1: Likelihood ratio tests with different fixed and random effects components

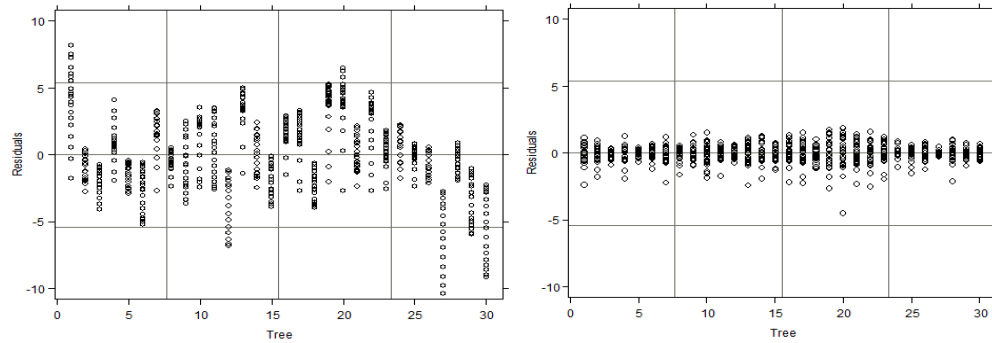| Model	Random parameters	No of parameters	AIC	BIC	Log-likelihood	LRT	p-value
1.0	None	4	3182.5210	3200.3420	-1587.2600		
1.1	b_1	5	1952.8910	1975.1670	-971.4450	1231.6300	<0.0001
1.2	b_2	5	2506.7180	2528.9940	-1248.3590		
1.3	b_3	5	2265.0770	2287.3530	-1127.5380		
1.4	b_1, b_2	7	1853.7650	1884.9510	-919.8820		
1.5	b_1, b_3	7	1806.9610	1838.1480	-896.4810	149.9295	<0.0001
1.6	b_2, b_3	7	2229.1800	2260.3660	-1107.5900		
1.7	b_1, b_2, b_3	8	1756.1300	1800.6820	-868.0649	56.8317	<0.0001

Fig. 2: Scatter plot of residuals for the fixed-effects model (left) and mixed-effects model (right)

considering two random parameters. Likelihood ratio test indicated that the addition of the random parameter b_1 significantly improved model fitting on model 1 (LRT = 1231.6, p<0.0001). The addition of the random parameters b_1, b_3 significantly improved model fitting on model 1.1 (LRT = 149.9, p<0.0001). The addition of the random parameters b_1, b_2, b_3 significantly improved model fitting on model 1.5 (LRT = 56.8, p<0.0001). Therefore, The two methods gave the same results that the three random parameters b_1, b_2, b_3 were needed in mixed-effects logistic model.

To get the among-tree variance-covariance structure, the logistic model with three random parameters was fitted to the sample data using NLME function in S-Plus. Taking b_1, b_2, b_3 as mixed effects, both Diagonal (Diag) and general positive-definite Symmetric (Sym) variance-covariance structures were examined for the random effects (D). It indicated that the two structures for the mixed logistic model were of significant difference (LRT = 16.1, p = 0.0011). The mixed logistic model with symmetric variance-covariance structure resulted in lower values of AIC (1756.13) and BIC (1800.68), indicating that a correlated variance-covariance structurewas preferred in the model fitting.

Sample data are 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). Therefore the AR (1), AR (2), MA (1), MA (2), ARMA (1, 1) and ARMA (2, 2) correlation structures were tested for inclusion in the mixed-effects logistic model. The mixed logistic model with MA (2) correlation structure showed lower values of AIC and

BIC with significant improvement of the model performance (LRT = 545.6, p<0.0001). The MA (2) could explain more dependency among repeated measurements within the individual. The correlation structure for MA (2) is:

$$\Gamma_i(\theta) = \begin{bmatrix} 1 & \dfrac{\theta}{1+\theta^2} & 0 \\ \dfrac{\theta}{1+\theta^2} & 1 & \dfrac{\theta}{1+\theta^2} \\ 0 & \dfrac{\theta}{1+\theta^2} & 1 \end{bmatrix} \tag{4}$$

Scatter plots of residuals were constructed for mixed-effects and fixed-effects models (Fig. 2). The Scatter plots showed that the mixed-effects model showed more homogeneous residual variance and no systematic pattern in the variation of the residuals. It indicated that mixed-effects model reduced the variance and significantly improved the model performances compared to fixed-effects model.

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

$$y_{ij} = \frac{(\beta_1 + b_{1i})}{1 + (\beta_2 + b_{2i})\exp(-(\beta_3 + b_{3i})t_{ij})} + \varepsilon_{ij} \tag{5}$$

$$\begin{bmatrix} b_{1i} \\ b_{2i} \\ b_{3i} \end{bmatrix} \sim N(0, D) \qquad D = \begin{bmatrix} \sigma_{b_1}^2 & \sigma_{b_1 b_2} & \sigma_{b_1 b_3} \\ \sigma_{b_1 b_2} & \sigma_{b_2}^2 & \sigma_{b_2 b_3} \\ \sigma_{b_1 b_3} & \sigma_{b_2 b_3} & \sigma_{b_3}^2 \end{bmatrix}$$

Parameter estimates and fit statistics for basic and mixed models are presented in Table 2. Compared to

Table 2: Estimated parameters and fit statistics for the basic model and mixed-effects model

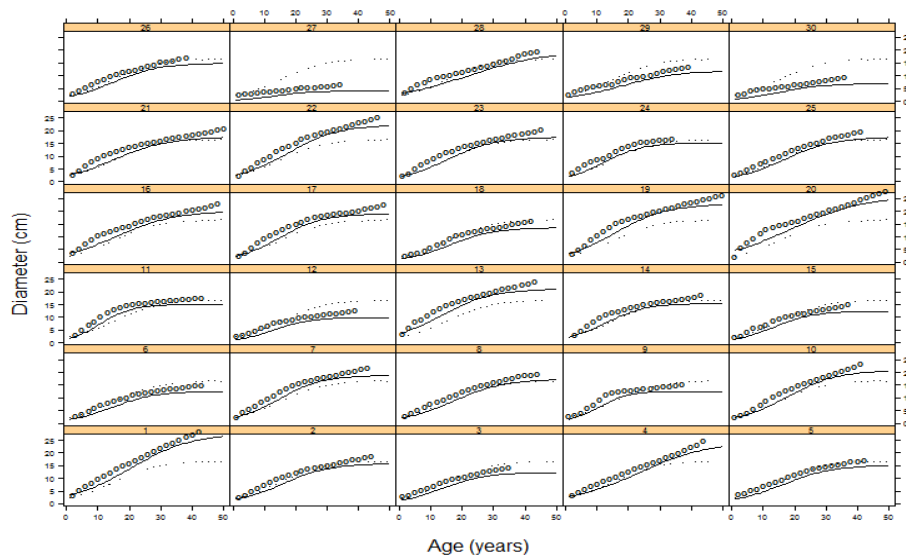|	Parameter	Basic model	Mixed-effects model
Fixed parameters	β_1	20.5736 (0.8050)	16.7122 (0.9794)
	β_2	6.8052 (0.4778)	8.1273 (0.3872)
	β_3	0.0915 (0.0060)	0.2579 (0.0059)
Variance components	σ^2	8.6558	0.4430
	σ^2_{b1}		28.1501
	σ^2_{b2}		2.2529
	σ^2_{b3}		0.0008
	σ_{b1b2}		-2.9624
	σ_{b1b3}		-0.1016
	σ_{b2b3}		0.0216
	θ		0.9976
Goodness-of-fit	Bias	-0.0678	-0.0564
	RMSE	2.9351	0.7130
	R^2	0.7526	0.9854

Fig. 3: Predicted versus actual diameters (circle) with (solid lines) and without (dotted lines) random parameters

the basic model, the mixed-effects model showed better performances with 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. 3). Both the fixed-effects model with random effects set to zero and mixed-effects model are compared. Prediction values of the mixed-effects model more closely followed the actual values for most trees and indicated that mixed model described the diameter growth of larch tree well.

Tree diameter predictions for new observations can be obtained either by using fixed parameters alone or by adding random parameters calculated from prior observations.

Case 1: Prediction with no prior diameter observations: If no prior information is available for a tree, the vectors of random effects are assumed to be equal to their expected value, which are zero by definition. The values of the fixed-effects parameters are from Table 2. Tree diameter is then calculated using Eq. (5). For example, the estimated diameter at age 24

is $\hat{y} = 11.7$ cm, the bias is 4.5 cm and the variance is var $(\hat{y}) = 20.2$. A 95% confidence interval for the prediction is (2.9, 20.9).

Case 2: Prediction with prior diameter observations: Suppose we know the tree diameters at age 2, 4, 6, 8, 10 and 12 to be 1.0, 2.7, 4.4, 6.7, 8.4 and 10.1 cm for a tree. Assume we want to predict the diameter of this tree at age 24.

The vector of random effects parameters $\hat{b}_k$ is calculated by using Eq. (3):

$$\hat{b}_k = \begin{pmatrix} -3.227434 \\ 0.339652 \\ 0.011649 \end{pmatrix}$$

Finally, tree diameter was calculated using Eq. (5). The estimated diameter at age 24 is $\hat{y} = 13.3$ cm, the bias is 2.9 and the variance is var $(\hat{y}) = 8.4$. A 95% confidence interval for the prediction is (7.6, 19.0). Compared with Case 1, Case 2 gives a

narrowerconfidence interval of tree diameterprediction at age 24. Therefore, the predictive ability of the mixed model with calibration is better than that of the fixed-effect model.

CONCLUSION

In this study, a nonlinear mixed-effects diameterincrement model was developed for dahurian larch in northeastern China. Nonlinear mixed-effects modeling techniques were used to estimate fixed and random-effect parameters for logistic growth function. The results showed thatthe logistic model with threerandom 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. Time series correlation structures including Autoregressive Correlation structure AR (1) and AR (2), Moving Average correlation structure MA (1) and MA (2), Autoregressive-Moving Average correlation structure [ARMA (1, 1)] and ARMA (2, 2) were incorporated into the best mixed model. The mixed model with MA (2) correlation structure significantly improved model performance (LRT = 545.6, p<0.0001). The fixed-effects parameters alone can be used to obtain the mean diameter growth curve for dahurian larch plantation based on all trees sampled from the population of plantation in the region. The random parameters for a tree of interest could be predicted with an approximate Bayes estimator of b_k using the prior diameter measurements from sample trees together with estimates of the fixed-effects parameters, the residual variance and the estimated variance-covariance matrix for the random-effects parameters. After calibration of random parameters, mixed model could improve diameter prediction compared to fixed model using fixed parameters.

The models developed in this study can be implemented in forest inventories with measurement of several tree diameters. The process of calibration apparently accounted indirectly for effects of site quality and stand density on the diameter growth. 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.

ACKNOWLEDGMENT

REFERENCES

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.

Colbert, K.C., D.R. Larsen and J.R. Lootens, 2002. Height-diameter equations for thirteen Midwestern bottomland hardwood species. North. J. Appl. For., 19: 171-176.

Fang, Z. and R.L. Bailey, 2001. Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. For. Sci., 47: 287-300.

Gregoire, T. and O. Schabenberger, 1996. A non-linear mixed-effects model to predict cumulative bole volume of standing tree. J. Appl. Statist., 23: 257-271.

Huang, S. and S.J. Titus, 1995. An individual tree diameter increment model for white spruce in Alberta. Can. J. For. Res., 25: 1455-1465.

Jiang, L. and F. Li, 2008. Modeling tree diameter growth using nonlinear mixed-effects models. Proceeding of International Seminar on Future Information Technology and Management Engineering. Leicestershire, United Kingdom, pp: 636-639.

Jiang, L. and F. Li, 2010. Application of nonlinear mixed-effects modeling approach in tree height prediction. J. Comp., 5(10): 1575-1581.

Lindstrom, M.J. and D.M. Bates, 1990. Nonlinear mixed effects models for repeated measures data. Biometrics, 46: 673-687.

Palmer, M.J., B.F. Phillips and G.T. Smith, 1991. Application of nonlinear models with random coefficients to growth data. Biometrics, 47(2): 623-635.

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.

Wolfinger, R., 1999. Fitting nonlinear mixed models with the new NLMIXED procedure. Proceedings of the 24th Annual SAS Users Group International Conference. SAS Institute Inc, Cary, NC, pp: 287.