Nonlinear Mixed Model Methods and Prediction Procedures Demonstrated on a Volume-Age Model

Prepared by:
Shongming Huang, Shawn X. Meng, Yuqing Yang

August 1, 2013

Technical Report Pub. No.: T/601(2013)

ISBN No. 978-1-4601-0755-3 (Printed Edition)
ISBN No. 978-1-4601-0756-0 (On-line Edition)

For additional copies of this publication, please contact:

Forest Management Branch
Alberta Ministry of Environment and Sustainable Resource Development
8th Floor, Great West Life Building
9920–108 Street
Edmonton, Alberta, Canada T5K 2M4

Main reception: Tel: (780) 427-8474 Fax: (780) 427-0085

This publication will be made available on the Alberta Government website. Please check the latest information on the Alberta Government website at:
http://esrd.alberta.ca/lands-forests/forest-management/default.aspx

Edmonton, Alberta, Canada
August 2013

Executive Summary and Acknowledgement

This study describes the nonlinear mixed-effects modeling technique applied to a volume-age model frequently used in Alberta. It demonstrates the procedures for using a fitted mixed model to make subject-specific predictions on data not used in model fitting. Generalized and step-by-step computer programs associated with the predictions are provided to facilitate the computations. Criteria for obtaining the most reasonable predictions under different circumstances are discussed.

Karl Peck, Daryl Price, Darren Aitkin and Dave Morgan reviewed an early draft and provided many constructive comments and suggestions. Glenn Buckmaster at West Fraser's Hinton Wood Products provided some additional data for model testing. Their help is much appreciated.

Table of Contents

Page

Executive Summary and Acknowledgement iii

1. **Introduction** 1

2. **Data and Models** 2
   - 2.1 Data 2
   - 2.2 Models 2

3. **Nonlinear Mixed Model Methods** 6
   - 3.1 Basic Formulation of Nonlinear Mixed Models 6
   - 3.2 The First-Order Method 8
   - 3.3 The First-Order Conditional Expectation Method 9

4. **Model Predictions** 14
   - 4.1 Prediction from the First-Order Method 14
   - 4.2 Prediction from the First-Order Conditional Expectation Method 16
   - 4.3 Goodness-of-Fit Statistics 20
   - 4.4 Choosing the “Best” Prediction 21

5. **Additional Notes** 22
   - 5.1 Adjusting the Predictions 22
   - 5.2 Population Average Predictions 24

6. **References** 27

7. **Appendices** 28
   - Appendix 1. Generalized Program for the First-Order Method 29
   - Appendix 2. Step-by-Step Program for the First-Order Method 32
   - Appendix 3. Generalized Program for the First-Order Conditional Expectation Method 34
   - Appendix 4. Step-by-Step Program for the First-Order Conditional Expectation Method 37
   - Appendix 5. Metric Conversion Chart 42

1. Introduction

Regression models can generally be classified as population-based models and subject-specific models. Traditional regression models estimated from the least squares method are typically population-based models. They predict the population averages, and as such, they are also referred to as “population-average models”, “population models”, or “base models” (when contrasted to subject-specific models or mixed models).

One common problem with the population-based models is that, due to the intrinsic variation and the polymorphic nature of biological growth, the trends exhibited by the data from individual subjects within a population may not always follow the trend exhibited by the population averages. This is illustrated in Figure 1, where the data from individual subjects show differing trends than that of the population averages. Because of the differing trends, it is quite possible that a population-based model may fit or predict the data well on average for the entire population, but it could perform poorly for the individual subjects within the population. Sometime population averages could be meaningless at a subject-specific level.

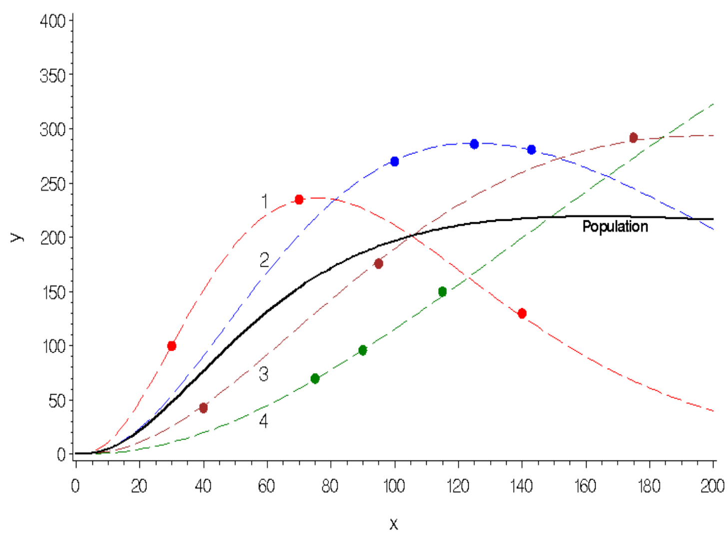

Figure 1. An illustration of population-based (solid line) and subject-specific (dashed-lines) models, where 1, 2, 3 and 4 represent four subjects in the population. The population-based model is obtained from all data combined.

Subject-specific models, on the other hand, describe the mean responses of individual subjects within a population. They can account for the idiosyncrasies of individual subjects within a population (e.g., Figure 1). Subject-specific models are often developed from the mixed-effects modeling technique. Thus, they are often referred to as mixed-effects models, or mixed models. Since the mixed model developed in this study is nonlinear, the term nonlinear mixed model (NMM) is used throughout this study.

The main objective of this study is to demonstrate the NMM technique based on a volume-age model frequently used in Alberta. The emphasis of this study is to show how to use a fitted mixed model to make subject-specific predictions on data not used in model fitting. For the volume-age model, each subject is a sample plot (e.g., a permanent sample plot or a temporary sample plot). The population is an amalgamation of the sample plots.

To facilitate the computation, generalized and step-by-step computer programs associated with the predictions from specific NMM methods are provided in Appendices. A summary of the NMM methods is also provided prior to using them to make predictions.

2. Data and Models

2.1 Data

Black spruce (*picea mariana* (Mill.) B.S.P.) volume-age data from 182 permanent sample plots (PSPs) with various black spruce compositions were used in this study. Among the 182 plots, 103 from the lower foothills natural subregion were used as model fitting data. The rest (79 plots) were used as model application data. Summary statistics for the model fitting and model application data are listed in Table 1. Throughout this study, volume refers to black spruce total volume (m^3/ha). Age refers to black spruce breast height age (years).

Table 1. Summary statistics for model fitting and model application data.

Data	Variable	*N*	*m*	Mean	Min	Max	SD
Model fitting	Volume (m^3/ha)	389	103	116.377	7.513	415.991	76.470
	Age (years)	389	103	90.010	24.000	168.417	33.185
Model application	Volume (m^3/ha)	322	79	85.578	1.439	287.890	68.534
	Age (years)	322	79	101.328	27.000	167.000	31.356

Note: volume and age are total volume and breast height age, *N* is the total number of observations (measurements from PSPs), *m* is the number of plots (subjects), min is minimum, max is maximum, and SD is standard deviation.

Figure 2 shows the model fitting and model application data. It can be seen that there are cross-overs among some of the volume-age trajectories from different plots/subjects in the data.

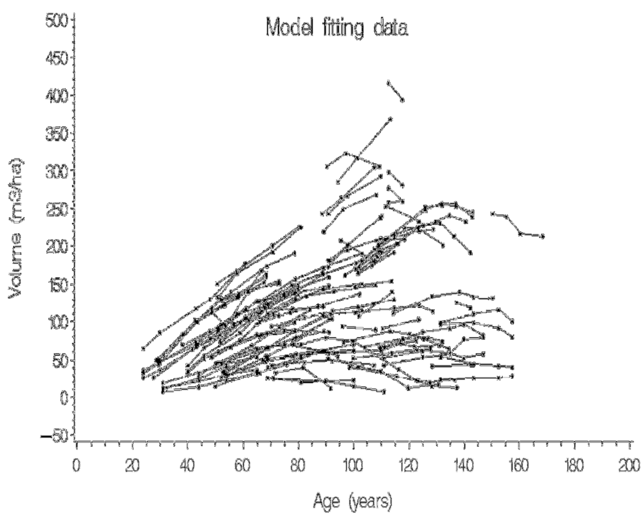

Figure 2. Volume-age trajectories for model fitting and model application data. Summary statistics for the data are listed in Table 1. Each trajectory represents repeated measures from one plot.

2.2 Models

Comparison of alternative model forms suggests that the following base model is appropriate for describing the volume-age relationship for black spruce in Alberta:

[1] $$\mathrm{Vol} = b_1 \mathrm{Age}^{b_2} \exp(-b_1 \mathrm{Age})$$

where Vol is total volume (m^3/ha), age is breast height age (years), b_1 and b_2 are model parameters (also called fixed parameters) applicable to the entire population, and exp denotes the exponential function.

The mixed model derived from the base model takes the following form:

$$\text{[2]} \qquad \mathrm{Vol} = (b_1 + u_1)\mathrm{Age}^{(b_2+u_2)}\exp(-(b_1 + u_1)\mathrm{Age})$$

where b_1 and b_2 are fixed parameters applicable to every plot in the population, and u_1 and u_2 are random parameters used to account for unique characteristics of each plot in the population.

Parameter estimates for the base and mixed models are listed in Table 2. Summary goodness-of-fit statistics associated with the estimates are also listed in Table 2. The parameter estimates for the base model [1] were obtained from the ordinary nonlinear least squares (NLS) method. The parameter estimates for the mixed model [2] were obtained from the first-order (FO) method and first-order conditional expectation (FOCE) method of the NMM technique. Both methods of the NMM technique are detailed in Section 3.

Table 2. Parameter estimates and goodness-of-fit (GOF) statistics obtained on the model fitting data.

Parameter	Model and method			GOF measure	Model and method		
	[1]–NLS	[2]–FO	[2]–FOCE		[1]–NLS	[2]–FO	[2]–FOCE
b_1	0.0206	0.01785	0.01879	$\bar{e}$	-0.035386	0.066026	-0.25955
b_2	2.3622	2.3396	2.3412	$\bar{e}\%$	-0.030406	0.056734	-0.22302
$\sigma^2_{u_1}$		0.0000641	0.0000878	SD	70.2203	4.87835	4.91032
$\sigma_{u_1u_2}$		0.0006674	0.0005871	MAD	56.4252	3.54557	3.57784
$\sigma^2_{u_2}$		0.02049	0.02121	MSE	4943.64	23.8642	24.2413
σ^2		45.0368	44.6634	R^2	0.15678	0.99593	0.99587
AIC		3393.0	3405.7	CC	0.27100	0.99796	0.99792
BIC		3408.8	3421.5	MPE	-65.4912	-0.70345	-2.22422
N	389	389	389	MAPE	94.4057	4.88089	5.70855
m	103	103	103	e_{10}	89.2031	10.7969	12.3393
				δ	4930.90	23.8027	24.1786

Note: *N* is the total number of observations, *m* is the number of plots, $\sigma^2_{u_1}$, $\sigma^2_{u_2}$ and $\sigma_{u_1u_2}$ are variances and covariance for the random parameters, and σ^2 is the residual variance. The goodness-of-fit (GOF) measures are defined in Table 3.

The goodness-of-fit statistics listed in Table 2 were calculated from the *N* observations (*N*=389) in the entire model fitting population, based on the formulas given in Table 3. They were <u>not</u> calculated by averaging the goodness-of-fit statistics from the individual plots in the population.

Historically, different goodness-of-fit measures have been used in different studies to determine the goodness-of-fit of a fitted model. Each measure has its pros and cons, and each usually reflects one aspect of a fitted model. This explains why a variety of goodness-of-fit measures listed in Table 3 were calculated in this study. In general, for most practical purposes, the overall accuracy (δ) value, which combines the mean bias ($\bar{e}^2$) and the variance of the residuals or prediction errors (SD^2), can be considered a good overall indication of model accuracy, for the δ value is a summation of the bias ($\bar{e}^2$) and precision (SD^2).

Figure 3 shows the "spaghetti plots" from the FO and FOCE methods of the NMM technique. The "spaghetti plots" display the plot-specific volume-age predictions for all 103 plots of the model fitting data across the age ranges where the fitted model is likely to be applied. Table 4 illustrates how the predictions are obtained to draw spaghetti plots for three example plots of the model fitting data. All variables and computations listed in Table 4 are described in more details in "Model Predictions" (Section 4) and "Appendices" (Section 7).

Table 3. Goodness-of-fit measures used in this study.

Goodness-of-fit measure	Computation formula
1. Mean bias (or mean error)	$\overline{\mathrm{e}} = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_{ij}) = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\mathrm{e}_{ij}$
2. Percent mean bias	$\overline{\mathrm{e}}\% = \frac{\overline{\mathrm{e}}}{\overline{y}} \times 100$
3. Standard deviation	$\mathrm{SD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\mathrm{e}_{ij} - \overline{\mathrm{e}})^2}$
4. Mean absolute deviation	$\mathrm{MAD} = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\lvert y_{ij} - \hat{y}_{ij}\right\rvert$
5. Mean square error (on model fitting data)	$\mathrm{MSE} = \frac{1}{N\text{-}p}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_{ij})^2$
6. Mean square error (on model application data)	$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_{ij})^2$
7. Coefficient of determination	$\mathrm{R}^2 = 1 - \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_{ij})^2 \Big/ \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \overline{y})^2$
8. Concordance correlation coefficient	$\mathrm{CC} = 1 - \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_{ij})^2}{\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \overline{y})^2 + \sum_{i=1}^{m}\sum_{j=1}^{n_i}(\hat{y}_{ij} - \overline{\hat{y}})^2 + N(\overline{y} - \overline{\hat{y}})^2}$
9. Mean percent error	$\mathrm{MPE} = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\frac{y_{ij} - \hat{y}_{ij}}{y_{ij}}) \times 100$
10. Mean absolute percent error	$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\lvert\frac{y_{ij} - \hat{y}_{ij}}{y_{ij}}\right\rvert \times 100$
11. Number of absolute percent errors >10%	$\mathrm{e}_{10} = \frac{number\ of\ \lvert \mathrm{PE}_{ij}\rvert > 10}{N}$, $\mathrm{PE}_{ij} = (\frac{y_{ij} - \hat{y}_{ij}}{y_{ij}}) \times 100$
12. Overall accuracy	$\delta = \overline{\mathrm{e}}^2 + \mathrm{SD}^2$
13. Akaike information criterion	$\mathrm{AIC} = -2\ln(L) + 2P$
14. Schwarz's Bayesian information	$\mathrm{BIC} = -2\ln(L) + P\ln(m)$
Grand means and total number	$\overline{y} = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}$ $\overline{\hat{y}} = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\hat{y}_{ij}$ $N = \sum_{i=1}^{m} n_i$

Note: y_{ij} and $\hat{y}_{ij}$ are the jth observed and predicted volumes for the ith plot, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n_i$, m is the number of plots in the data, n_i is the number of observations in the ith plot, $\overline{y}$ and $\overline{\hat{y}}$ are the grand means of the observed and predicted volumes, N is the total number of observations, L is the maximized value of the likelihood function for the estimated model, p is the number of fixed parameters, P is the total number of effective parameters in mixed model estimation (includes fixed parameters, variance-covariance components of the random parameters, plus the residual variance component), and AIC and BIC are information criteria used for mixed model on model fitting data only.

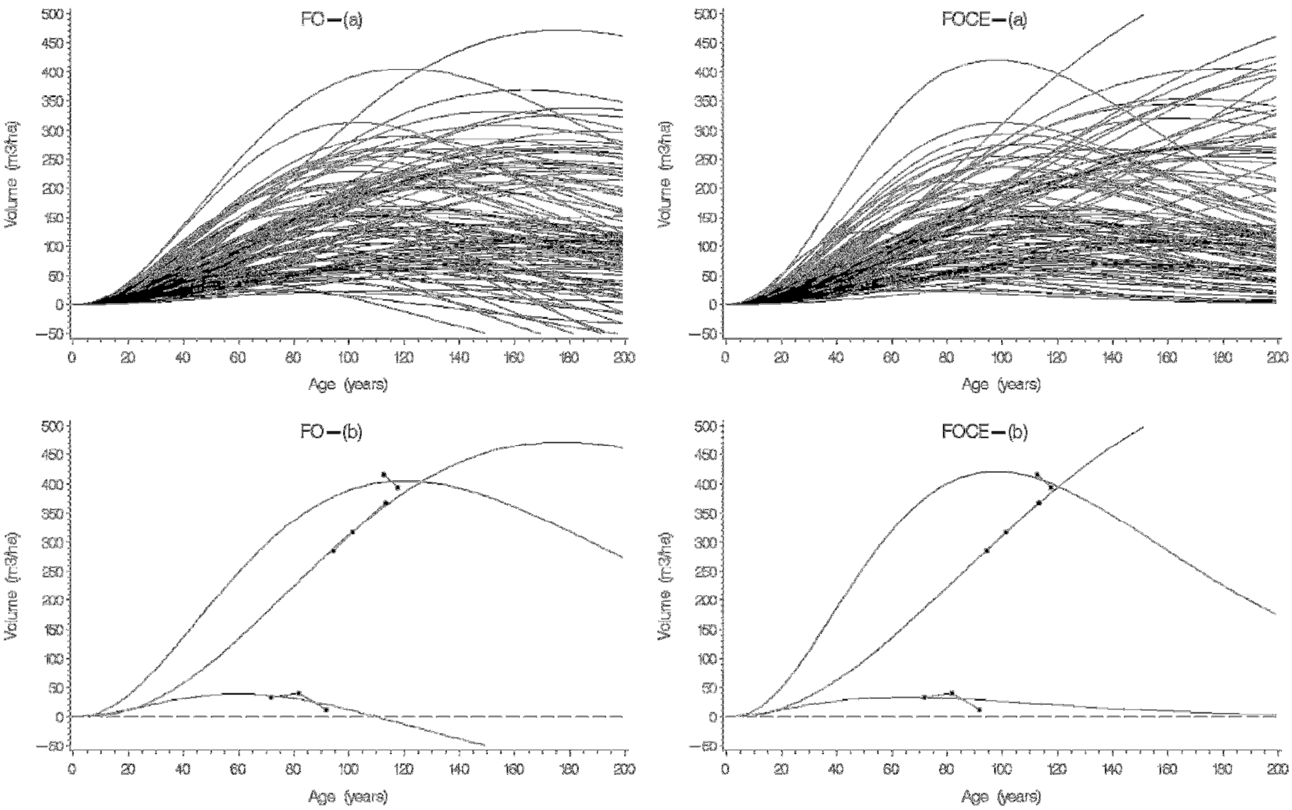

Figure 3. Plot-specific predictions from the FO and FOCE methods across the likely age ranges: (a) for all 103 plots of the model fitting data; and (b) for three example plots listed in Tables 4.

Table 4. Original data and calculations for three example plots of the model fitting data.

Plot	Time	Age	Volume	Vol_{fix}	d_u_1	d_u_2	u_1	u_2	y_pred	y_res
					FO method					
m1	1	94.332	285.380	138.126	-5291.53	628.03	-0.01533	0.11080	288.821	-3.440
m1	2	101.332	317.151	144.126	-6530.28	665.63	-0.01533	0.11080	317.975	-0.824
m1	3	113.332	368.217	151.159	-8662.86	715.03	-0.01533	0.11080	363.170	5.047
m2	1	71.746	33.106	108.962	-1713.28	465.61	0.00711	-0.13121	35.688	-2.582
m2	2	81.746	40.085	123.691	-3181.77	544.69	0.00711	-0.13121	29.600	10.485
m2	3	91.746	12.273	135.543	-4842.09	612.52	0.00711	-0.13121	20.748	-8.475
m3	1	112.667	415.991	150.872	-8546.10	712.79	0.00803	0.44998	402.985	13.006
m3	2	117.667	393.785	152.746	-9415.94	728.27	0.00803	0.44998	404.841	-11.056
					FOCE method					
m1	1	94.332	285.380	134.033	-316.59	1302.54	-0.00806	0.12306	286.473	-1.093
m1	2	101.332	317.151	138.954	-2569.47	1464.11	-0.00806	0.12306	317.017	0.134
m1	3	113.332	368.217	144.126	-7383.55	1737.20	-0.00806	0.12306	367.247	0.970
m2	1	71.746	33.106	107.955	-1368.35	138.83	0.01496	-0.16687	32.488	0.618
m2	2	81.746	40.085	121.427	-1604.53	135.57	0.01496	-0.16687	30.786	9.299
m2	3	91.746	12.273	131.842	-1753.76	127.58	0.01496	-0.16687	28.233	-15.960
m3	1	112.667	415.991	143.942	-31229.07	1932.13	0.00875	0.34883	408.965	7.026
m3	2	117.667	393.785	145.056	-32585.28	1909.53	0.00875	0.34883	400.501	-6.716

Note: m1, m2 and m3 are three example plots. Time is measurement time. All other variables are detailed in Section 4.

3. Nonlinear Mixed Model Methods

Before demonstrating how to use the fitted models to make predictions, some relevant background material on NMM methods is presented here. Readers who are familiar with the NMM methods can skip the background material and go directly to "Model Predictions" (Section 4).

3.1 Basic Formulation of Nonlinear Mixed Models

Using the standard terminology for nonlinear models, a population-based model that describes the population averages can be written as:

[3] $$\mathbf{y} = f(\mathbf{x}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$$

where **y** is the dependent variable (also referred to as response variable), f denotes some nonlinear function, **x** is a known matrix of covariates (also referred to as independent variables, regressors, predictors, or *x*-variables), $\boldsymbol{\beta}$ is a vector of model parameters applicable to the entire population, and $\boldsymbol{\varepsilon}$ is the error term.

For a subject-specific NMM, it can be written as:

[4] $$y_{ij} = f(\mathbf{x}_{ij}, \mathbf{b}, \mathbf{u}_i) + \varepsilon_{ij}$$

where y_{ij} is the *j*th observation in the *i*th subjects, i=1, 2,…, m, j=1, 2,…, n_i, m is the number of subjects in the population, n_i is the number of observations in the *i*th subjects, f is a general expression of a nonlinear function, $\mathbf{x}_{ij}$ is a known vector of covariates for the *j*th observation in the *i*th subjects, **b** is a vector of fixed parameters common to all subjects in the population, $\mathbf{u}_i$ is a vector of random parameters unique for the *i*th subject in the population, and ε_{ij} is a normally distributed within-subject error term.

In this study, each plot is a subject. The measurements at different times within each plot are observations.

The subject-specific NMM [4] describes the mean responses of individual subjects within a population. This is achieved through the inclusion of subject-specific random parameters $\mathbf{u}_i$ in the model. For instance, the subject-specific volume-age model [2] can be written more explicitly as:

[5] $$\mathrm{Vol}_{ij} = (\mathrm{b}_1 + \mathrm{u}_{1i})\mathrm{Age}_{ij}^{(\mathrm{b}_2 + \mathrm{u}_{2i})}\exp(\text{-}(\mathrm{b}_1 + \mathrm{u}_{1i})\mathrm{Age}_{ij})$$

where Vol_{ij} and Age_{ij} are observed volume and age for the *j*th measurement in the *i*th plot, b_1 and b_2 are fixed parameters common to every plot in the population, and u_{1i} and u_{2i} are random parameters unique for the *i*th plot in the population.

In essence, a subject-specific model has a unique set of coefficients for each subject in the population. For the subject-specific volume-age model [5] developed from *m* plots, there are *m* unique sets of coefficients for *m* plots in the population:

$$\begin{array}{llll}
\text{Plot 1:} & \mathrm{Vol}_{1j} = \mathrm{b}_{11}\mathrm{Age}_{1j}^{\mathrm{b}_{21}}\exp(\text{-}\mathrm{b}_{11}\mathrm{Age}_{1j}) & \mathrm{b}_{11} = \mathrm{b}_1 + \mathrm{u}_{11}, & \mathrm{b}_{21} = \mathrm{b}_2 + \mathrm{u}_{21} \\
\text{Plot 2:} & \mathrm{Vol}_{2j} = \mathrm{b}_{12}\mathrm{Age}_{2j}^{\mathrm{b}_{22}}\exp(\text{-}\mathrm{b}_{12}\mathrm{Age}_{2j}) & \mathrm{b}_{12} = \mathrm{b}_1 + \mathrm{u}_{12}, & \mathrm{b}_{22} = \mathrm{b}_2 + \mathrm{u}_{22} \\
\vdots & \vdots \qquad \vdots & \vdots & \vdots \\
\text{Plot } m\text{:} & \mathrm{Vol}_{mj} = \mathrm{b}_{1m}\mathrm{Age}_{mj}^{\mathrm{b}_{2m}}\exp(\text{-}\mathrm{b}_{1m}\mathrm{Age}_{mj}) & \mathrm{b}_{1m} = \mathrm{b}_1 + \mathrm{u}_{1m}, & \mathrm{b}_{2m} = \mathrm{b}_2 + \mathrm{u}_{2m}
\end{array}$$

Because a unique set of coefficients is developed for each subject in the population, rather than assigning the same set of coefficients obtained for the entire population to each subject in the population, a subject-specific model is much more flexible and powerful than a population-based model. It can mimic the data trends exhibited by individual subjects in the population more closely. A subject-specific model typically provides more accurate fits and predictions on a subject-specific level than a population-based model.

In a more compact form, the mixed model [4] can be written for subject *i* as:

[6] $$\mathbf{y}_i = f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i) + \boldsymbol{\varepsilon}_i$$

where i=1, 2,…, m, $\mathbf{y}_i = [y_{i1}, y_{i2}, \ldots, y_{in_i}]'$ is a vector of observations for the *y*-variable from subject *i*, $\mathbf{x}_i$ is a known matrix of the *x*-variables, and $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in_i}]'$ is a vector of within-subject errors. The random parameters vector $\mathbf{u}_i$ and the error vector $\boldsymbol{\varepsilon}_i$ are typically assumed to be uncorrelated and (multivariate) normally distributed with mean zero and variance-covariance matrices **D** and $\mathbf{R}_i$, respectively. That is:

$$Cov[\mathbf{u}_i, \boldsymbol{\varepsilon}_i] = \mathbf{0} \qquad E\begin{bmatrix} \mathbf{u}_i \\ \boldsymbol{\varepsilon}_i \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \qquad Var\begin{bmatrix} \mathbf{u}_i \\ \boldsymbol{\varepsilon}_i \end{bmatrix} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_i \end{bmatrix}$$

which can be simplified to $\mathbf{u}_i \sim N(\mathbf{0}, \mathbf{D})$ and $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$. The variance-covariance matrix **D** of the random parameters is generally assumed to be the same for each and every subject in the population (i.e., $\mathbf{D}_i$=**D** for i=1, 2,…, m). The variance-covariance matrix $\mathbf{R}_i$ of the within-subject errors can take many forms to represent independent and identically distributed (*iid*) errors, correlated errors, heterogeneous errors, and generalized (correlated and heterogeneous) errors. Due to its confusing but often inconsequential nature in predictions (Huang et al. 2009a, 2009b; Meng and Huang 2010; Huang et al. 2011; Yang and Huang 2011a), and to avoid a digression from the main purpose of this study, $\mathbf{R}_i$ was assumed to be *iid*. That is, $\mathbf{R}_i=\sigma^2\mathbf{I}$, where σ^2 is the error variance and **I** is an $n_i \times n_i$ identity matrix (a square matrix with ones on the main diagonal and zeros elsewhere).

The key difference between subject-specific mixed models and population-based models is the inclusion of random parameters in mixed models. Random parameters primarily serve four purposes:

1). Account for the idiosyncrasies of individual subjects within a population;

2). Account for the remnant impacts of the *x*-variables already included in the model – this can be important when the true model specification is unknown;

3). Account for the impacts of other known and unknown *x*-variables left-out by the model without actually requiring these variables to be identified or measured – this can be a good or a bad trait;

4). Alleviate or eliminate entirely the correlation and heteroskedasticity issues commonly occurred in forest modeling from repeatedly measured cross-sectional data.

These and other related topics are discussed elsewhere (e.g., Huang et al. 2009c; Meng and Huang 2010).

Different methods can be used to estimate the parameters of NMMs. They include first-order linearization, Laplace's approximation, adaptive Gaussian quadrature, importance sampling, and Bayesian estimation (Davidian and Giltinan 1995, Vonesh and Chinchilli 1997, Pinheiro and Bates 2004). The two most commonly used methods, which were implemented in this study due mainly to their computational simplicity in making subject-specific predictions on new data not used in model fitting, are the first-order (FO) method of Beal and

Sheiner (1982) and the first-order conditional expectation (FOCE) method of Lindstrom and Bates (1990). Both methods use a first-order Taylor series expansion of the mixed model [6] around a $\mathbf{b}^*$ close to $\mathbf{b}$ and an $\mathbf{u}_i^*$ close to $\mathbf{u}_i$, to linearize [6], with the negligible terms (e.g., quadratics, cubics and cross-products) dropped:

[7] $$\mathbf{y}_i \approx f(\mathbf{x}_i, \mathbf{b}^*, \mathbf{u}_i^*) + \mathbf{X}_i(\mathbf{b} - \mathbf{b}^*) + \mathbf{Z}_i(\mathbf{u}_i - \mathbf{u}_i^*) + \boldsymbol{\varepsilon}_i$$

where $\mathbf{X}_i$ and $\mathbf{Z}_i$, often referred to as design matrices in mixed model idiom, are partial derivatives of $\mathbf{y}_i$ with respect to $\mathbf{b}$ and $\mathbf{u}_i$, respectively. The methods differ on how the $\mathbf{u}_i^*$ is defined in [7].

3.2 The First-Order Method

For the FO method, $\mathbf{u}_i^*$ in [7] is set to its expectation of zero, i.e., $\mathbf{u}_i^*$ = E($\mathbf{u}_i$) = $\mathbf{0}$. Therefore, [7] is reduced to:

[8] $$\mathbf{y}_i \approx f(\mathbf{x}_i, \mathbf{b}^*, \mathbf{0}) + \mathbf{X}_i(\mathbf{b} - \mathbf{b}^*) + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$$

where the design matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ are defined by:

[9] $$\mathbf{X}_i = \left.\frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial \mathbf{b}'}\right|_{\mathbf{b}^*, \mathbf{0}} \qquad \mathbf{Z}_i = \left.\frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial \mathbf{u}_i'}\right|_{\mathbf{b}^*, \mathbf{0}}$$

Rearranging [8] yields:

[10] $$\mathbf{y}_i - f(\mathbf{x}_i, \mathbf{b}^*, \mathbf{0}) + \mathbf{X}_i\mathbf{b}^* = \mathbf{X}_i\mathbf{b} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$$

To create a linearized form of [10], the left-hand side of [10] is defined as the "pseudo-response" function $\mathbf{y}_i^*$:

[11] $$\mathbf{y}_i^* = \mathbf{y}_i - f(\mathbf{x}_i, \mathbf{b}^*, \mathbf{0}) + \mathbf{X}_i\mathbf{b}^*$$

Hence, [10] can be written as a standard linear mixed model:

[12] $$\mathbf{y}_i^* = \mathbf{X}_i\mathbf{b} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$$

Following the standard linear mixed model theory (e.g., see Fitzmaurice et al. 2004), the generalized least squares estimator $\hat{\mathbf{b}}$ of the fixed parameters $\mathbf{b}$ and the random parameters predictor $\hat{\mathbf{u}}_i$ of the $\mathbf{u}_i$ in [12], can be obtained as follows:

[13] $$\hat{\mathbf{b}} = \left(\sum_{i=1}^{m} \mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}\mathbf{X}_i\right)^{-1} \sum_{i=1}^{m} \mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}\mathbf{y}_i^*$$

[14] $$\hat{\mathbf{u}}_i = \hat{\mathbf{D}}\mathbf{Z}_i'\hat{\mathbf{V}}_i^{-1}(\mathbf{y}_i^* - \mathbf{X}_i\hat{\mathbf{b}})$$

where $\hat{\mathbf{D}}$ is an estimate of the variance-covariance matrix $\mathbf{D}$ for the random parameters, and $\hat{\mathbf{V}}_i$ is the estimated (marginal) variance-covariance matrix for the pseudo-response function $\mathbf{y}_i^*$, averaged over the distribution of the random parameters $\mathbf{u}_i$ (Davidian and Giltinan 1995, Vonesh and Chinchilli 1997):

[15] $\hat{\mathbf{V}}_i = Var(\mathbf{y}_i^*) = Var(\mathbf{Z}_i\hat{\mathbf{u}}_i) + Var(\boldsymbol{\varepsilon}_i) = \mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i^{'} + \hat{\mathbf{R}}_i$

where $\hat{\mathbf{R}}_i$ is an estimate of the variance-covariance matrix $\mathbf{R}_i$ for the within-subject error term.

Substituting the $\hat{\mathbf{V}}_i$ in [15] and the pseudo-response function $\mathbf{y}_i^*$ in [11] into [14], and recognizing $\mathbf{u}_i^* = \mathbf{0}$ and $\mathbf{b}^* = \hat{\mathbf{b}}$ once $\hat{\mathbf{b}}$ is estimated, the predictor of random parameters $\hat{\mathbf{u}}_i$ in [14] can be written as:

[16] $\hat{\mathbf{u}}_i = \hat{\mathbf{D}}\mathbf{Z}_i^{'}(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i^{'} + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0})]$

where the $\mathbf{Z}_i$ matrix is defined in [9] (with $\mathbf{b}^* = \hat{\mathbf{b}}$). Equation [16] is the random parameters prediction equation for the FO method.

Once the values of $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}_i$ are available, the pseudo-response function $\mathbf{y}_i^*$ can be predicted from [12]:

[17] $\hat{\mathbf{y}}_i^* = \mathbf{X}_i\hat{\mathbf{b}} + \mathbf{Z}_i\hat{\mathbf{u}}_i$

Substituting the $\mathbf{y}_i^*$ in the pseudo-response function [11] by the $\hat{\mathbf{y}}_i^*$ in [17] and recognizing $\mathbf{b}^* = \hat{\mathbf{b}}$ produce:

$$\mathbf{X}_i\hat{\mathbf{b}} + \mathbf{Z}_i\hat{\mathbf{u}}_i = \mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0}) + \mathbf{X}_i\hat{\mathbf{b}}$$

Rearranging for $\mathbf{y}_i$ gives the predicted $\mathbf{y}_i$ ($\hat{\mathbf{y}}_i$) for subject *i* for the FO method:

[18] $\hat{\mathbf{y}}_i = f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0}) + \mathbf{Z}_i\hat{\mathbf{u}}_i$

where $\mathbf{Z}_i$ is defined in [9] (with $\mathbf{b}^* = \hat{\mathbf{b}}$). Equation [18] is the response variable prediction equation for the FO method. The corresponding residuals or prediction errors ($\mathbf{e}_i$) are calculated by:

[19] $\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i = \mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0}) - \mathbf{Z}_i\hat{\mathbf{u}}_i$

Notice the term $\mathbf{Z}_i\hat{\mathbf{u}}_i$ must be included in predicting the $\hat{\mathbf{y}}_i$ and calculating the $\mathbf{e}_i$ values for the FO method.

3.3 The First-Order Conditional Expectation Method

For the FOCE method, [7] is first written as:

[20] $\mathbf{y}_i - f(\mathbf{x}_i, \mathbf{b}^*, \mathbf{u}_i^*) + \mathbf{X}_i\mathbf{b}^* + \mathbf{Z}_i\mathbf{u}_i^* = \mathbf{X}_i\mathbf{b} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$

where the design matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ are given by:

[21] $\mathbf{X}_i = \left.\dfrac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial \mathbf{b}'}\right|_{\mathbf{b}^*, \mathbf{u}_i^*}$ $\quad$ $\mathbf{Z}_i = \left.\dfrac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial \mathbf{u}_i^{'}}\right|_{\mathbf{b}^*, \mathbf{u}_i^*}$

Define the left-hand side of [20] as the "pseudo-response" function $\mathbf{y}_i^*$ for the FOCE method:

[22] $\mathbf{y}_i^* = \mathbf{y}_i - f(\mathbf{x}_i, \mathbf{b}^*, \mathbf{u}_i^*) + \mathbf{X}_i\mathbf{b}^* + \mathbf{Z}_i\mathbf{u}_i^*$

Then [20] can be written as a standard linear mixed model:

[23] $\mathbf{y}_i^* = \mathbf{X}_i\mathbf{b} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$

Following the standard linear mixed model theory (e.g., see Fitzmaurice et al. 2004), the generalized least squares estimator $\hat{\mathbf{b}}$ of the fixed parameters **b** and the random parameters predictor $\hat{\mathbf{u}}_i$ of the $\mathbf{u}_i$ in [23] can be obtained as described in equations [13] and [14].

For the FOCE method, substituting the $\hat{\mathbf{V}}_i$ in [15] and the pseudo-response function $\mathbf{y}_i^*$ in [22] into [14], and recognizing $\mathbf{u}_i^* = \hat{\mathbf{u}}_i$ and $\mathbf{b}^* = \hat{\mathbf{b}}$, the $\hat{\mathbf{u}}_i$ prediction equation [14] can be written as:

[24] $\hat{\mathbf{u}}_i = \hat{\mathbf{D}}\mathbf{Z}_i'(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i' + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i) + \mathbf{Z}_i\hat{\mathbf{u}}_i]$

where the $\mathbf{Z}_i$ matrix is defined in [21] (with $\mathbf{u}_i^* = \hat{\mathbf{u}}_i$ and $\mathbf{b}^* = \hat{\mathbf{b}}$).

For the FOCE method, the random parameters to be predicted ($\hat{\mathbf{u}}_i$) appear on both sides of [24]. This is very different from that for the FO method, where $\hat{\mathbf{u}}_i$ appears only on the left-hand side of equation [16].

There is no easy algebraic solution for $\hat{\mathbf{u}}_i$ in [24]. Instead, a three-step numerical procedure needs to be implemented to iteratively solve for $\hat{\mathbf{u}}_i$ in [24]:

Step 1. Obtain a first estimate, termed $\hat{\mathbf{u}}_{i,1}$, of the random parameters. This is achieved by assuming the initial $\hat{\mathbf{u}}_i$ appearing on the right-hand side of [24] equal to zero (i.e., $\hat{\mathbf{u}}_i = \mathbf{0}$):

[25] $\hat{\mathbf{u}}_{i,1} = \hat{\mathbf{D}}\mathbf{Z}_{i,0}'(\mathbf{Z}_{i,0}\hat{\mathbf{D}}\mathbf{Z}_{i,0}' + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0})]$

where the initial design matrix $\mathbf{Z}_{i,0}$ is evaluated at the assumed initial $\hat{\mathbf{u}}_i = \mathbf{0}$:

[26] $\mathbf{Z}_{i,0} = \left.\dfrac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial \mathbf{u}_i'}\right|_{\hat{\mathbf{b}}, \mathbf{0}}$

Step 2. Once the $\hat{\mathbf{u}}_{i,1}$ is calculated from step 1, the next $\mathbf{Z}_i$, termed $\mathbf{Z}_{i,1}$, is evaluated at $\hat{\mathbf{u}}_{i,1}$:

[27] $\mathbf{Z}_{i,1} = \left.\dfrac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial \mathbf{u}_i'}\right|_{\hat{\mathbf{b}}, \hat{\mathbf{u}}_{i,1}}$

The next estimation of $\hat{\mathbf{u}}_i$, termed $\hat{\mathbf{u}}_{i,2}$, is obtained based on [24] again, with the $\hat{\mathbf{u}}_i$ and $\mathbf{Z}_i$ on the right-hand side of [24] replaced by $\hat{\mathbf{u}}_{i,1}$ from [25] and $\mathbf{Z}_{i,1}$ from [27], respectively:

[28] $\hat{\mathbf{u}}_{i,2} = \hat{\mathbf{D}}\mathbf{Z}_{i,1}'(\mathbf{Z}_{i,1}\hat{\mathbf{D}}\mathbf{Z}_{i,1}' + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_{i,1}) + \mathbf{Z}_{i,1}\hat{\mathbf{u}}_{i,1}]$

Step 3. Having the calculated $\hat{\mathbf{u}}_{i,2}$ from step 2, the next $\mathbf{Z}_i$, termed $\mathbf{Z}_{i,2}$, is evaluated at $\hat{\mathbf{u}}_{i,2}$:

[29] $$\mathbf{Z}_{i,2} = \left.\frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial \mathbf{u}_i'}\right|_{\hat{\mathbf{b}}, \hat{\mathbf{u}}_{i,2}}$$

The next estimation of $\hat{\mathbf{u}}_i$, termed $\hat{\mathbf{u}}_{i,3}$, is computed based on [24] again, with the $\hat{\mathbf{u}}_i$ and $\mathbf{Z}_i$ on the right-hand side of [24] replaced by the updated $\hat{\mathbf{u}}_{i,2}$ from [28] and $\mathbf{Z}_{i,2}$ from [29], respectively:

[30] $$\hat{\mathbf{u}}_{i,3} = \hat{\mathbf{D}}\mathbf{Z}_{i,2}'(\mathbf{Z}_{i,2}\hat{\mathbf{D}}\mathbf{Z}_{i,2}' + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_{i,2}) + \mathbf{Z}_{i,2}\hat{\mathbf{u}}_{i,2}]$$

This process is iterated *k* times until a user-specified convergence criterion is achieved, such as:

[31] $$\left|\hat{\mathbf{u}}_{i,k} - \hat{\mathbf{u}}_{i,(k-1)}\right| < 0.0000001$$

Once the convergence criterion is achieved, the final predictor of the random parameters is:

[32] $$\hat{\mathbf{u}}_i = \hat{\mathbf{u}}_{i,k}\,.$$

Intrinsically, some readers may already recognize that the three-step procedure is in fact an iteration of equations [27] and [28], with the initial and end conditions given by [25] and [31], respectively. More details were provided in the above descriptions for the sake of other interested readers.

With the known $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}_i$ values, the pseudo-response function $\mathbf{y}_i^*$ in [23] can be predicted:

[33] $$\hat{\mathbf{y}}_i^* = \mathbf{X}_i\hat{\mathbf{b}} + \mathbf{Z}_i\hat{\mathbf{u}}_i$$

Substituting the $\mathbf{y}_i^*$ in the pseudo-response function [22] by the $\hat{\mathbf{y}}_i^*$ in [33], and recognizing $\mathbf{b}^* = \hat{\mathbf{b}}$ and $\mathbf{u}_i^* = \hat{\mathbf{u}}_i$, [22] can be written as:

$$\mathbf{X}_i\hat{\mathbf{b}} + \mathbf{Z}_i\hat{\mathbf{u}}_i = \mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i) + \mathbf{X}_i\hat{\mathbf{b}} + \mathbf{Z}_i\hat{\mathbf{u}}_i$$

Rearranging for $\mathbf{y}_i$ produces the predicted $\mathbf{y}_i$ ($\hat{\mathbf{y}}_i$) for subject *i* for the FOCE method:

[34] $$\hat{\mathbf{y}}_i = f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i)$$

Evidently, for the FOCE method, the response variable for any subject *i* can be predicted directly by simply substituting the known $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}_i$ values into the mixed model without involving the term $\mathbf{Z}_i\hat{\mathbf{u}}_i$. This is fundamentally different from the FO method, which involves $\mathbf{Z}_i\hat{\mathbf{u}}_i$ and uses [18] to predict the response variable. The corresponding residuals or prediction errors ($\mathbf{e}_i$) for the FOCE method are calculated by:

[35] $$\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i = \mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i)$$

which is again different from [19] for the FO method. Table 5 summarizes the formulas explicit to the FO and FOCE methods of the NMM technique.

Table 5. A summary of the first-order (FO) and first-order conditional expectation (FOCE) methods.

Method = FO	Method = FOCE
Taylor series expansion	
$\mathbf{y}_i \approx f(\mathbf{x}_i, \mathbf{b}^*, \mathbf{0}) + \mathbf{X}_i(\mathbf{b}-\mathbf{b}^*) + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$	$\mathbf{y}_i \approx f(\mathbf{x}_i, \mathbf{b}^*, \mathbf{u}_i^*) + \mathbf{X}_i(\mathbf{b}-\mathbf{b}^*) + \mathbf{Z}_i(\mathbf{u}_i - \mathbf{u}_i^*) + \boldsymbol{\varepsilon}_i$
Pseudo-response function	
$\mathbf{y}_i^* = \mathbf{y}_i - f(\mathbf{x}_i, \mathbf{b}^*, \mathbf{0}) + \mathbf{X}_i\mathbf{b}^*$	$\mathbf{y}_i^* = \mathbf{y}_i - f(\mathbf{x}_i, \mathbf{b}^*, \mathbf{u}_i^*) + \mathbf{X}_i\mathbf{b}^* + \mathbf{Z}_i\mathbf{u}_i^*$
Linearized model	
$\mathbf{y}_i^* = \mathbf{X}_i\mathbf{b} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$	$\mathbf{y}_i^* = \mathbf{X}_i\mathbf{b} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$
Design matrices	
$\mathbf{X}_i = \left.\frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial \mathbf{b}'}\right\|_{\hat{\mathbf{b}}, \mathbf{0}}$ $\quad \mathbf{Z}_i = \left.\frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial \mathbf{u}_i'}\right\|_{\hat{\mathbf{b}}, \mathbf{0}}$	$\mathbf{X}_i = \left.\frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial \mathbf{b}'}\right\|_{\hat{\mathbf{b}}, \mathbf{u}_i^*}$ $\quad \mathbf{Z}_i = \left.\frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial \mathbf{u}_i'}\right\|_{\hat{\mathbf{b}}, \mathbf{u}_i^*}$
Predictor of random parameters	
$\hat{\mathbf{u}}_i = \hat{\mathbf{D}}\mathbf{Z}_i'(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i' + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0})]$	$\hat{\mathbf{u}}_i = \hat{\mathbf{D}}\mathbf{Z}_i'(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i' + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i) + \mathbf{Z}_i\hat{\mathbf{u}}_i]$
Marginal prediction from fixed parameters	
$\hat{\mathbf{y}}_{i\_fix} = f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0})$	$\hat{\mathbf{y}}_{i\_fix} = f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0})$
Subject-specific prediction with independent and identically distributed error structure ($\hat{\mathbf{R}}_i = \sigma^2\mathbf{I}$)	
$\hat{\mathbf{y}}_i = f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0}) + \mathbf{Z}_i\hat{\mathbf{u}}_i$	$\hat{\mathbf{y}}_i = f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i)$
Residuals or prediction errors	
$\hat{\mathbf{e}}_i = \mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0}) - \mathbf{Z}_i\hat{\mathbf{u}}_i$	$\hat{\mathbf{e}}_i = \mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i)$
Subject-specific forecast with generalized error structure ($\hat{\mathbf{R}}_i = \sigma^2\boldsymbol{\Psi}$)	
$\hat{\mathbf{y}}_{0i} = f(\mathbf{x}_{0i}, \hat{\mathbf{b}}, \mathbf{0}) + \mathbf{Z}_{0i}\hat{\mathbf{u}}_i + \mathbf{V}'\boldsymbol{\Psi}^{-1}\hat{\mathbf{e}}_i$	$\hat{\mathbf{y}}_{0i} = f(\mathbf{x}_{0i}, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i) + \mathbf{V}'\boldsymbol{\Psi}^{-1}\hat{\mathbf{e}}_i$

Note: $\mathbf{y}_i$ and $\hat{\mathbf{y}}_i$ are observed and predicted values for subject *i*, $\mathbf{x}_i$ is a matrix of covariate(s), $\mathbf{X}_i$ and $\mathbf{Z}_i$ are partial derivatives with respect to fixed parameters $\mathbf{b}$ and random parameters $\mathbf{u}_i$, respectively, $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}_i$ are predictors of $\mathbf{b}$ and $\mathbf{u}_i$, $\hat{\mathbf{D}}$ and $\hat{\mathbf{R}}_i$ are estimated variance-covariance matrices for $\mathbf{u}_i$ and $\boldsymbol{\varepsilon}_i$, respectively, $\hat{\mathbf{y}}_{0i}$, $\mathbf{x}_{0i}$ and $\mathbf{Z}_{0i}$ denote the variables associated with future observations, $\mathbf{V}$ contains the correlations between the elements of past and future errors, and $\boldsymbol{\Psi}$ is the correlation matrix of past errors.

It is worthwhile to reiterate that the formulas summarized in Table 5 for the FO and FOCE methods are quite different, particularly with regard to: a) design matrices; b) predictors of random parameters; c) predictions of the response variable; and d) residual or prediction error calculations. Mixing the formulas from the FO and FOCE methods would be mathematically incorrect.

Indeed, due to the methodological and computational differences between the FO and FOCE methods, when estimating a model and using the estimated model to make predictions, it is very important to ensure that the model estimation procedure is consistent with the model prediction procedure. Otherwise, the results could be inconsistent or simply wrong, and the reported model fitting statistics could lose their intended meaning or could even give a false indication about the model's performance when predictions are made. Extensive evaluation on model fitting and model application data showed that substantial differences occurred when the formulas from the FO and FOCE methods were mixed (Huang 2008, Meng and Huang 2009).

It is questionable that a number of NMM applications in forestry used $\hat{\mathbf{u}}_i = \hat{\mathbf{D}}\mathbf{Z}_i^{'}(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i^{'} + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0})]$ (equation [16]) from the FO method to predict $\hat{\mathbf{u}}_i$, then used $\hat{\mathbf{y}}_i = f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i)$ (equation [34]) from the FOCE method to predict $\hat{\mathbf{y}}_i$ and came up with an unbelievably convincing outcome. Had the correct formulas or a different data set been used in those applications, the outcome could have been different.

In practice, before a fitted mixed model can be used to make predictions on data not used in model fitting, modellers must check two items prior to recommending a prediction procedure:

1). On the model fitting data, the random parameters predicted from the prediction procedure are equivalent to the random parameters obtained from the model estimation procedure; and

2). On the model fitting data, the prediction errors obtained from the prediction procedure are equivalent to the residuals obtained from the model estimation procedure.

Failing to verify any one of the two items could mean that the prediction procedure is incompatible with the model estimation procedure. In cases where the predictions cannot be checked on model fitting data (e.g., model users typically can only access the parameter estimates listed in Table 2, but not the full model fitting data), either the model estimation method should be known, or a mathematically consistent set of formulas corresponding to a specific method should be chosen for predictions after comparing alternative methods.

It is very important to ensure that model estimation procedure is equivalent to model prediction procedure. As a general rule this principle of equivalence between model estimation and model prediction procedures shall be followed whenever possible in any type of model estimation and prediction. Due to the computational difficulties in developing and implementing an equivalent procedure in model estimation and prediction on both model fitting and model application data, other NMM methods (Laplace's approximation, adaptive Gaussian quadrature, importance sampling and Bayesian estimation) are not discussed in this study. Interested readers may wish to read Yang and Huang (2011b) on comparing some of these methods.

Table 5 also lists the formulas for subject-specific forecasts with a generalized error structure ($\hat{\mathbf{R}}_i = \sigma^2\mathbf{\Psi}$). Since the NMM technique often alleviates or eliminates entirely the correlation and heteroskedasticity issues that commonly occur in forest modeling, further details on how to address these issues in NMMs are not presented here. Generalized error structure is used in <u>forecasting</u> future observations directly from the past measurements of the same sequence/trajectory. It has no use in <u>predicting</u> current or future observations that does not directly rely on the past measurements of the same sequence. Interested readers may wish to read Huang et al. (2009a, 2009b, 2011), Meng and Huang (2010), and Meng et al. (2012) on how to use estimated generalized error structures to forecast future observations from the past measurement(s) of the same sequence.

4. Model Predictions

Three plots each from the model fitting data (Table 4) and model application data (Table 6) are used to demonstrate model predictions. Since the prediction procedures on the model fitting and model application data are the same, detailed computations are illustrated here only for the model application data.

Table 6. Example model application data and computations based on the FO method.

Plot (1)	Time (2)	Age (3)	Volume (4)	Vol_{fix} (5)	d_u_1 (6)	d_u_2 (7)	u_1 (8)	u_2 (9)	y_pred (10)	y_res (11)
v1	1	136.000	166.812	154.516	-12357.79	759.08	0.01639	0.29303	174.400	-7.588
v1	2	150.000	154.413	151.355	-14224.01	758.39	0.01639	0.29303	140.448	13.965
v1	3	156.000	135.661	149.051	-14901.80	752.69	0.01639	0.29303	125.365	10.296
v1	4	161.000	114.873	146.768	-15407.33	745.79	0.01639	0.29303	112.773	2.100
v1	5	167.000	79.100	143.647	-15941.56	735.18	0.01639	0.29303	97.789	-18.689
v2	1	49.667	19.999	68.349	434.39	266.93	-0.01516	-0.09195	37.222	-17.223
v2	2	60.667	75.227	89.688	-416.56	368.21	-0.01516	-0.09195	62.146	13.081
v3	1	51.000	32.880	71.008	356.63	279.19	0.00190	-0.16995	24.237	8.643
v3	2	56.000	26.408	80.831	1.81	325.37	0.00190	-0.16995	25.537	0.871
v3	3	62.000	17.720	92.148	-550.82	380.31	0.00190	-0.16995	26.469	-8.749

Note: v1, v2 and v3 are three example plots. Time refers to measurement time (1, 2, 3, etc.). Age and volume refer to black spruce breast height age (years) and total volume (m^3/ha). All other variables are described in the main text.

4.1 Prediction from the First-Order Method

To make plot-specific predictions from the FO method, random parameters unique for each plot must be predicted first based on equation [16]. Using plot v1 of the model application data in Table 6 as an example, volume predictions (Vol_{fix}) based on the fixed parameters only are listed in column 5 of Table 6. They are obtained directly from the mixed model [2], with the random parameters set to zero and the fixed parameters given in Table 2 for the FO method (b_1=0.01785 and b_2=2.3396):

$$\mathbf{Vol}_{i\_fix} = (b_1 + 0)\mathbf{Age}_i^{(b_2+0)} \exp(-(b_1 + 0)\mathbf{Age}_i) = [154.516, 151.355, 149.051, 146.768, 143.647]'$$

Given $\sigma^2_{u_1}$ =0.0000641, $\sigma_{u_1u_2}$ =0.0006674, $\sigma^2_{u_2}$ =0.02049 and σ^2 =45.0368 for the FO method (from Table 2), for plot v1 with five observations and two random parameters, the variance-covariance matrices $\hat{\mathbf{R}}_i$ and $\hat{\mathbf{D}}$ for the errors and random parameters are:

$$\hat{\mathbf{R}}_i = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{bmatrix} = \begin{bmatrix} 45.0368 & 0 & 0 & 0 & 0 \\ 0 & 45.0368 & 0 & 0 & 0 \\ 0 & 0 & 45.0368 & 0 & 0 \\ 0 & 0 & 0 & 45.0368 & 0 \\ 0 & 0 & 0 & 0 & 45.0368 \end{bmatrix}$$

$$\hat{\mathbf{D}} = \begin{bmatrix} \sigma^2_{u_1} & \sigma_{u_1u_2} \\ \sigma_{u_1u_2} & \sigma^2_{u_2} \end{bmatrix} = \begin{bmatrix} 0.0000641 & 0.0006674 \\ 0.0006674 & 0.02049 \end{bmatrix}$$

The partial derivatives of [2] with respect to the two random parameters are:

[36] $$\mathrm{d\_u}_1 = \mathrm{d\_b}_1 = \frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{0})}{\partial \mathrm{b}_1} = \mathbf{Age}_i^{\mathrm{b}_2}(1 - \mathrm{b}_1 \mathbf{Age}_i) \cdot \exp(-\mathrm{b}_1 \mathbf{Age}_i)$$

[37] $$\mathrm{d\_u}_2 = \mathrm{d\_b}_2 = \frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{0})}{\partial \mathrm{b}_2} = \mathrm{b}_1 \cdot \exp(-\mathrm{b}_1 \mathbf{Age}_i) \cdot \mathbf{Age}_i^{\mathrm{b}_2} \ln(\mathbf{Age}_i)$$

The derivatives (columns 6 and 7 of Table 6) constitute the design matrix $\mathbf{Z}_i$ for plot v1:

$$\mathbf{Z}_i = \begin{bmatrix} \mathrm{d\_u}_{11} & \mathrm{d\_u}_{21} \\ \mathrm{d\_u}_{12} & \mathrm{d\_u}_{22} \\ \mathrm{d\_u}_{13} & \mathrm{d\_u}_{23} \\ \mathrm{d\_u}_{14} & \mathrm{d\_u}_{24} \\ \mathrm{d\_u}_{15} & \mathrm{d\_u}_{25} \end{bmatrix} = \begin{bmatrix} -12357.79 & 759.08 \\ -14224.01 & 758.39 \\ -14901.80 & 752.69 \\ -15407.33 & 745.79 \\ -15941.56 & 735.18 \end{bmatrix}$$

Therefore, the two random parameters for plot v1 can be predicted following equation [16]:

$$\hat{\mathbf{u}}_i = \hat{\mathbf{D}}\mathbf{Z}_i'(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i' + \hat{\mathbf{R}}_i)^{-1}(\mathbf{Vol}_i - \mathbf{Vol}_{i\_fix})$$

which produces the following random parameter predictions for plot v1:

$$\hat{\mathbf{u}}_i = [\mathrm{u}_1, \mathrm{u}_2]' = [0.01639, 0.29303]'$$

Random parameter predictions for other plots of the model application data are obtained in a similar manner. They are listed in columns 8 and 9 of Table 6. Once the $\hat{\mathbf{u}}_i$ values are known, plot-specific volume predictions are calculated following equation [18], which becomes $\mathbf{V\hat{o}l}_i = \mathbf{Vol}_{i\_\mathrm{fix}} + \mathbf{Z}_i\hat{\mathbf{u}}_i$. For plot v1, this gives:

$$\mathbf{V\hat{o}l}_i = \begin{bmatrix} 154.516 \\ 151.355 \\ 149.051 \\ 146.768 \\ 143.647 \end{bmatrix} + \begin{bmatrix} -12357.79 & 759.08 \\ -14224.01 & 758.39 \\ -14901.80 & 752.69 \\ -15407.33 & 745.79 \\ -15941.56 & 735.18 \end{bmatrix} \times \begin{bmatrix} 0.01639 \\ 0.29303 \end{bmatrix} = \begin{bmatrix} 174.400 \\ 140.448 \\ 125.365 \\ 112.773 \\ 97.789 \end{bmatrix}$$

They are listed in column 10 (y_pred= $\mathbf{V\hat{o}l}_i$) of Table 6. The corresponding prediction errors (y_res= $\mathbf{e}_i$) are listed in column 11 of Table 6. They are computed following equation [19]:

$$\mathbf{e}_i = \mathbf{Vol}_i - \mathbf{Vol}_{i\_fix} - \mathbf{Z}_i\hat{\mathbf{u}}_i$$

The computations for the model fitting data follow the same procedures. Example results are listed in Table 4.

Appendices 1 and 2 provide two programs associated with the FO method. One is generalized (Appendix 1). The other is step-by-step (Appendix 2). Both programs apply to model fitting as well as model application data. Interested readers could use either one of them to carry out the computations involved in the FO method.

Figure 4 (left-hand side graphs) shows the plot-specific predictions from the FO method for all 79 plots of the model application data. The predictions are made within the observed data range (graph FO-(a)), as well as across the potential age range where the fitted model is likely to be applied in practice (graph FO-(b)). The original data and the predictions for the three example plots listed in Table 6 are displayed in graph FO-(c). Note that the predictions from the FO method could be negative in some cases (e.g., for plot v3 with three observations, when the age is older than about 140 years – see graph FO-(c)).

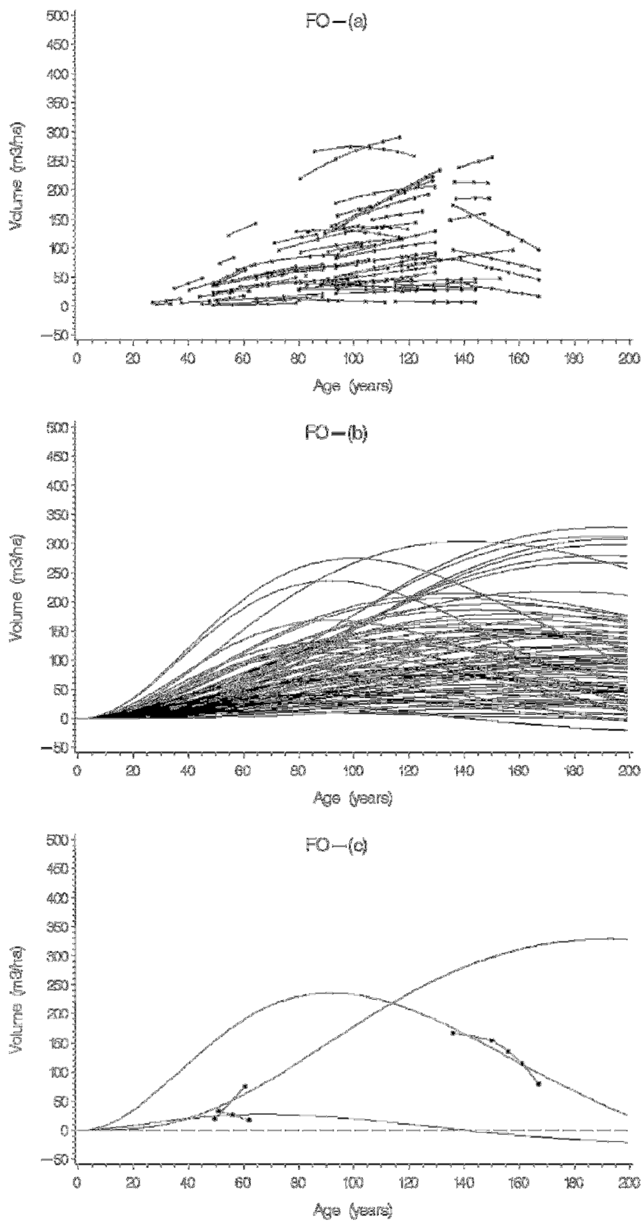

Figure 4. Plot-specific predictions from the FO and FOCE methods for the model application data: (a) within the observed data range, (b) across the potential age range, and (c) for the three example plots listed in Tables 6 (FO method) and 7 (FOCE method). Dashed lines in (c) indicate the line of zero.

4.2 Prediction from the First-Order Conditional Expectation Method

For the FOCE method, the prediction for the random parameters follows the three-step procedure described in [25] to [32]. Plot v1 of the model application data listed in Table 7 is used to demonstrate the computations.

Table 7. Example model application data and computations based on the FOCE method.

Plot (1)	Time (2)	Age (3)	Volume (4)	Vol_{fix} (5)	d_u_1 (6)	d_u_2 (7)	y_pred (8)	y_res (9)	u_1 (10)	u_2 (11)
					Step 1 computation					
v1	1	136.000	166.812	144.263	-11942.11	708.71			0.01665	0.32237
v1	2	150.000	154.413	139.487	-13499.55	698.92			0.01665	0.32237
v1	3	156.000	135.661	136.599	-14039.71	689.81			0.01665	0.32237
v1	4	161.000	114.873	133.883	-14429.90	680.31			0.01665	0.32237
v1	5	167.000	79.100	130.306	-14826.26	666.91			0.01665	0.32237
					Step 2 computation					
v1	1	136.000	166.812	144.263	-14850.90	676.90	137.787	29.025	0.01437	0.32107
v1	2	150.000	154.413	139.487	-13263.79	545.73	108.915	45.498	0.01437	0.32107
v1	3	156.000	135.661	136.599	-12490.57	493.62	97.750	37.911	0.01437	0.32107
v1	4	161.000	114.873	133.883	-11824.83	452.52	89.055	25.818	0.01437	0.32107
v1	5	167.000	79.100	130.306	-11014.71	406.20	79.367	-0.267	0.01437	0.32107
					Step 3 computation					
v1	1	136.000	166.812	144.263	-18491.83	858.32	174.717	-7.905	0.01561	0.34510
v1	2	150.000	154.413	139.487	-17085.61	714.37	142.572	11.841	0.01561	0.34510
v1	3	156.000	135.661	136.599	-16323.01	655.03	129.714	5.947	0.01561	0.34510
v1	4	161.000	114.873	133.883	-15638.78	607.37	119.527	-4.654	0.01561	0.34510
v1	5	167.000	79.100	130.306	-14776.99	552.68	107.988	-28.888	0.01561	0.34510
					Final iteration results					
v1	1	136.000	166.812	144.263	-18376.22	845.05	172.016	-5.204	0.01549	0.34226
v1	2	150.000	154.413	139.487	-16730.03	693.78	138.461	15.952	0.01549	0.34226
v1	3	156.000	135.661	136.599	-15882.74	632.39	125.230	10.431	0.01549	0.34226
v1	4	161.000	114.873	133.883	-15137.21	583.47	114.825	0.048	0.01549	0.34226
v1	5	167.000	79.100	130.306	-14213.15	527.78	103.122	-24.022	0.01549	0.34226
v2	1	49.667	19.999	69.097	4917.70	159.44	40.827	-20.828	-0.01291	-0.00138
v2	2	60.667	75.227	89.765	6689.36	250.91	61.117	14.110	-0.01291	-0.00138
v3	1	51.000	32.880	71.698	1405.50	85.59	21.769	11.111	-0.01014	-0.23744
v3	2	56.000	26.408	81.246	1511.77	102.17	25.381	1.028	-0.01014	-0.23744
v3	3	62.000	17.720	92.115	1598.90	123.20	29.850	-12.130	-0.01014	-0.23744

Note: v1, v2 and v3 are three example plots. Time refers to measurement time (1, 2, 3, etc.). Age and volume refer to black spruce breast height age (years) and total volume (m^3/ha). All other variables are described in the main text.

Step 1. Volume predictions based on the fixed parameters only are listed in column 5 of Table 7. They are obtained directly from the mixed model [2] with the random parameters set to zero and the fixed parameters given in Table 2 for the FOCE method (b_1=0.01879 and b_2=2.3412):

$$\mathbf{Vol}_{i\_fix} = (b_1+0)\mathbf{Age}_i^{(b_2+0)}\exp(-(b_1+0)\mathbf{Age}_i) = [144.263, 139.487, 136.599, 133.883, 130.306]'$$

The partial derivatives of [2] with respect to the random parameters evaluated at $\hat{\mathbf{u}}_i = \mathbf{0}$ are calculated by:

$$d\_u_1 = d\_b_1 = \frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{0})}{\partial b_1} = \mathbf{Age}_i^{(b_2+0)}(1-(b_1+0)\mathbf{Age}_i)\cdot \exp(-(b_1+0)\mathbf{Age}_i)$$

$$d\_u_2 = d\_b_2 = \frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{0})}{\partial b_2} = (b_1+0)\cdot \exp(-(b_1+0)\mathbf{Age}_i)\cdot \mathbf{Age}_i^{(b_2+0)}\ln(\mathbf{Age}_i)$$

They are listed in columns 6 and 7 of Table 7. The derivatives constitute the initial design matrix $\mathbf{Z}_{i,0}$ evaluated at $\hat{\mathbf{u}}_i = \mathbf{0}$:

$$\mathbf{Z}_{i,0} = \left.\frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial \mathbf{u}_i'}\right|_{\hat{\mathbf{b}},\mathbf{0}} = \begin{bmatrix} d\_u_{11} & d\_u_{21} \\ d\_u_{12} & d\_u_{22} \\ d\_u_{13} & d\_u_{23} \\ d\_u_{14} & d\_u_{24} \\ d\_u_{15} & d\_u_{25} \end{bmatrix} = \begin{bmatrix} -11942.11 & 708.71 \\ -13499.55 & 698.92 \\ -14039.71 & 689.81 \\ -14429.90 & 680.31 \\ -14826.26 & 666.91 \end{bmatrix}$$

Given $\sigma^2_{u_1}$ =0.0000878, $\sigma_{u_1u_2}$ =0.0005871, $\sigma^2_{u_2}$ =0.02121 and σ^2 =44.6634 for the FOCE method (from Table 2), for plot v1 with five observations and two random parameters, the variance-covariance matrices $\hat{\mathbf{R}}_i$ and $\hat{\mathbf{D}}$ for the errors and random parameters are:

$$\hat{\mathbf{R}}_i = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{bmatrix} = \begin{bmatrix} 44.6634 & 0 & 0 & 0 & 0 \\ 0 & 44.6634 & 0 & 0 & 0 \\ 0 & 0 & 44.6634 & 0 & 0 \\ 0 & 0 & 0 & 44.6634 & 0 \\ 0 & 0 & 0 & 0 & 44.6634 \end{bmatrix}$$

$$\hat{\mathbf{D}} = \begin{bmatrix} \sigma^2_{u_1} & \sigma_{u_1u_2} \\ \sigma_{u_1u_2} & \sigma^2_{u_2} \end{bmatrix} = \begin{bmatrix} 0.0000878 & 0.0005871 \\ 0.0005871 & 0.02121 \end{bmatrix}$$

Hence, the first estimate $\hat{\mathbf{u}}_{i,1}$ of the random parameters $\hat{\mathbf{u}}_i$ for the FOCE method can be obtained following equation [25]:

$$\hat{\mathbf{u}}_{i,1} = \hat{\mathbf{D}}\mathbf{Z}'_{i,0}(\mathbf{Z}_{i,0}\hat{\mathbf{D}}\mathbf{Z}'_{i,0} + \hat{\mathbf{R}}_i)^{-1}(\mathbf{Vol}_i - \mathbf{Vol}_{i\_fix}) = [\hat{u}_{1,1},\ \hat{u}_{2,1}]' = [0.01665,\ 0.32237]'$$

Step 2. Once the $\hat{\mathbf{u}}_{i,1}$ is calculated from step 1, a new set of the partial derivatives are evaluated at $\hat{\mathbf{u}}_i = \hat{\mathbf{u}}_{i,1}$:

[38] $$d\_u_1 = d\_b_1 = \frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial b_1} = \mathbf{Age}_i^{(b_2+\hat{u}_{2,1})}(1-(b_1+\hat{\mathbf{u}}_{1,1})\mathbf{Age}_i)\cdot \exp(-(b_1+\hat{\mathbf{u}}_{1,1})\mathbf{Age}_i)$$

[39] $$d\_u_2 = d\_b_2 = \frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial b_2} = (b_1+\hat{\mathbf{u}}_{1,1})\cdot \exp(-(b_1+\hat{\mathbf{u}}_{1,1})\mathbf{Age}_i)\cdot \mathbf{Age}_i^{(b_2+\hat{u}_{2,1})}\ln(\mathbf{Age}_i)$$

The partial derivatives (columns 6 and 7 of Table 7) constitute a new design matrix $\mathbf{Z}_{i,1}$ evaluated at $\hat{\mathbf{u}}_{i,1}$:

[40] $$\mathbf{Z}_{i,1} = \left.\frac{\partial f(\mathbf{x}_i,\mathbf{b},\mathbf{u}_i)}{\partial \mathbf{u}_i{'}}\right|_{\hat{\mathbf{b}},\hat{\mathbf{u}}_{i,1}} = \begin{bmatrix} \mathrm{d\_u}_{11} & \mathrm{d\_u}_{21} \\ \mathrm{d\_u}_{12} & \mathrm{d\_u}_{22} \\ \mathrm{d\_u}_{13} & \mathrm{d\_u}_{23} \\ \mathrm{d\_u}_{14} & \mathrm{d\_u}_{24} \\ \mathrm{d\_u}_{15} & \mathrm{d\_u}_{25} \end{bmatrix} = \begin{bmatrix} -14850.90 & 676.90 \\ -13263.79 & 545.73 \\ -12490.57 & 493.62 \\ -11824.83 & 452.52 \\ -11014.71 & 406.20 \end{bmatrix}$$

where the elements $\mathrm{d\_u}_{1j}$ and $\mathrm{d\_u}_{2j}$ in $\mathbf{Z}_{i,1}$ are obtained from equations [38] and [39], respectively.

Volume predictions for the FOCE method from the known $\hat{\mathbf{u}}_{i,1}$ and **b** values, $\mathbf{V\hat{o}l}_{i,1} = f(\mathbf{x}_i,\hat{\mathbf{b}},\hat{\mathbf{u}}_{i,1})$, are obtained following equation [34]. For plot v1, this gives:

[41] $$\mathbf{V\hat{o}l}_{i,1} = (\mathrm{b}_1+\hat{\mathrm{u}}_{1,1})\mathbf{Age}_i^{(\mathrm{b}_2+\hat{\mathrm{u}}_{2,1})}\exp(-(\mathrm{b}_1+\hat{\mathrm{u}}_{1,1})\mathbf{Age}_i) = [137.787, 108.915, 97.750, 89.055, 79.367]'$$

Results are listed in column 8 of Table 7. Column 9 of Table 7 lists the corresponding prediction errors calculated by $\mathbf{e}_{i,1} = \mathbf{y}_i - f(\mathbf{x}_i,\hat{\mathbf{b}},\hat{\mathbf{u}}_{i,1}) = \mathbf{Vol}_i - \mathbf{V\hat{o}l}_{i,1}$ (note that in Table 7, y_pred= $\mathbf{V\hat{o}l}_i$ and y_res=$\mathbf{e}_i$).

Having the $\hat{\mathbf{u}}_{i,1}$ and $\mathbf{Z}_{i,1}$ values, the next estimation of $\hat{\mathbf{u}}_i$, termed $\hat{\mathbf{u}}_{i,2}$, is obtained following equation [28], which gives:

[42] $$\hat{\mathbf{u}}_{i,2} = \hat{\mathbf{D}}\mathbf{Z}_{i,1}^{'}(\mathbf{Z}_{i,1}\hat{\mathbf{D}}\mathbf{Z}_{i,1}^{'}+\hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i,\hat{\mathbf{b}},\hat{\mathbf{u}}_{i,1})+\mathbf{Z}_{i,1}\hat{\mathbf{u}}_{i,1}] = [0.01437, 0.32107]'$$

They are listed in columns 10 and 11 of Table 7.

Step 3. Having the calculated $\hat{\mathbf{u}}_{i,2}$ from step 2, the next set of the partial derivatives are evaluated at $\hat{\mathbf{u}}_i = \hat{\mathbf{u}}_{i,2}$:

[43] $$\mathrm{d\_u}_1 = \mathrm{d\_b}_1 = \frac{\partial f(\mathbf{x}_i,\mathbf{b},\mathbf{u}_i)}{\partial \mathrm{b}_1} = \mathbf{Age}_i^{(\mathrm{b}_2+\hat{\mathbf{u}}_{2,2})}(1-(\mathrm{b}_1+\hat{\mathbf{u}}_{1,2})\mathbf{Age}_i)\cdot\exp(-(\mathrm{b}_1+\hat{\mathbf{u}}_{1,2})\mathbf{Age}_i)$$

[44] $$\mathrm{d\_u}_2 = \mathrm{d\_b}_2 = \frac{\partial f(\mathbf{x}_i,\mathbf{b},\mathbf{u}_i)}{\partial \mathrm{b}_2} = (\mathrm{b}_1+\hat{\mathbf{u}}_{1,2})\cdot\exp(-(\mathrm{b}_1+\hat{\mathbf{u}}_{1,2})\mathbf{Age}_i)\cdot\mathbf{Age}_i^{(\mathrm{b}_2+\hat{\mathbf{u}}_{2,2})}\ln(\mathbf{Age}_i)$$

They (columns 6 and 7 of Table 7) constitute the next design matrix $\mathbf{Z}_{i,2}$ evaluated at $\hat{\mathbf{u}}_{i,2}$:

[45] $$\mathbf{Z}_{i,2} = \left.\frac{\partial f(\mathbf{x}_i,\mathbf{b},\mathbf{u}_i)}{\partial \mathbf{u}_i{'}}\right|_{\hat{\mathbf{b}},\hat{\mathbf{u}}_{i,2}} = \begin{bmatrix} \mathrm{d\_u}_{11} & \mathrm{d\_u}_{21} \\ \mathrm{d\_u}_{12} & \mathrm{d\_u}_{22} \\ \mathrm{d\_u}_{13} & \mathrm{d\_u}_{23} \\ \mathrm{d\_u}_{14} & \mathrm{d\_u}_{24} \\ \mathrm{d\_u}_{15} & \mathrm{d\_u}_{25} \end{bmatrix} = \begin{bmatrix} -18491.83 & 858.32 \\ -17085.61 & 714.37 \\ -16323.01 & 655.03 \\ -15638.78 & 607.37 \\ -14776.99 & 552.68 \end{bmatrix}$$

A new set of volume predictions from the known $\hat{\mathbf{u}}_{i,2}$ and **b** values, are calculated following equation [34] again, which produces (column 8 of Table 7):

[46] $$\mathbf{V\hat{o}l}_{i,2} = (\mathrm{b}_1+\hat{\mathrm{u}}_{1,2})\mathbf{Age}_i^{(\mathrm{b}_2+\hat{\mathrm{u}}_{2,2})}\exp(-(\mathrm{b}_1+\hat{\mathrm{u}}_{1,2})\mathbf{Age}_i) = [174.717, 142.572, 129.714, 119.527, 107.988]'$$

The corresponding prediction errors ($\mathbf{e}_{i,2} = \mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_{i,2}) = \mathbf{Vol}_i - \mathbf{V\hat{o}l}_{i,2}$) are listed in column 9 of Table 7.

With the known $\hat{\mathbf{u}}_{i,2}$ and $\mathbf{Z}_{i,2}$, the next estimation of $\hat{\mathbf{u}}_i$, termed $\hat{\mathbf{u}}_{i,3}$, is obtained following [30], which gives:

[47] $$\hat{\mathbf{u}}_{i,3} = \hat{\mathbf{D}}\mathbf{Z}'_{i,2}(\mathbf{Z}_{i,2}\hat{\mathbf{D}}\mathbf{Z}'_{i,2} + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_{i,2}) + \mathbf{Z}_{i,2}\hat{\mathbf{u}}_{i,2}] = [0.01561, 0.34510]'$$

They are listed in columns 10 and 11 of Table 7.

The process described in equations [43] to [47] is iterated *k* times, with the prior $\hat{\mathbf{u}}_{i,(k-1)}$ in the equations replaced by the newer $\hat{\mathbf{u}}_{i,k}$ calculated from [47]. The process is stopped once the convergence criterion specified in [31] is achieved. For plot v1 of the model application data, the convergence criterion is achieved after six iterations. The final $\hat{\mathbf{u}}_i$ is predicted to be (columns 10 and 11 of Table 7):

$$\hat{\mathbf{u}}_i = [\hat{u}_1, \hat{u}_2]' = [0.01549, 0.34226]'$$

The final volume predictions for plot v1 are (column 8 of Table 7):

$$\mathbf{V\hat{o}l}_i = (b_1 + \hat{u}_1)\mathbf{Age}_i^{(b_2+\hat{u}_2)}\exp(-(b_1 + \hat{u}_1)\mathbf{Age}_i) = [172.016, 138.461, 125.230, 114.825, 103.122]'$$

The associated final prediction errors (column 9 of Table 7) are calculated by $\mathbf{e}_i = \mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i) = \mathbf{Vol}_i - \mathbf{V\hat{o}l}_i$.

The final results from the FOCE method for plots v2 and v3 of the model application data are also listed in Table 7. The final results from the FOCE method for plots m1, m2 and m3 of the model fitting data are listed in Table 4. Interested readers may wish to verify and duplicate the computations.

Appendices 3 and 4 provide two programs associated with the FOCE method. One is generalized (Appendix 3). The other is step-by-step (Appendix 4). Both programs apply to model fitting as well as model application data. Interested readers could use either one of them to carry out the iterations required by the FOCE method.

Figure 4 (right-hand side graphs) shows the plot-specific predictions from the FOCE method for all 79 plots of the model application data. The predictions are made within the observed data range (graph FOCE-(a)), as well as across the potential age range where the fitted model is likely to be applied in practice (graph FOCE-(b)). The original data and the predictions for the three example plots listed in Table 7 are displayed in graph FOCE-(c). Note that the predictions from the FOCE method are always positive across the age range, whereas the predictions from the FO method are not.

Overall, Figure 4 suggests that the predictions from the FO and FOCE methods are similar within the observed data range, but beyond the observed data range, some of the predictions can be quite different. This explains why sometimes two methods or two models with similar statistics can produce very different predictions.

4.3 Goodness-of-Fit Statistics

The goodness-of-fit statistics obtained from the FO and FOCE methods on the model application data are listed in Table 8. For a comparison, the goodness-of-fit statistics obtained from the direct application of model [1] on the model application data are also listed in Table 8. All goodness-of-fit statistics were calculated according to the formulas defined in Table 3, based on the *N*=322 observations in the model application data. They were not averaged from the goodness-of-fit statistics calculated by individual plots in the population.

Table 8. Goodness-of-fit statistics obtained on the model application data.

Model	Goodness-of-fit measure										
	$\bar{e}$	$\bar{e}$%	SD	MAD	MSE	R^2	CC	MPE	MAPE	e_{10}	δ
[1]-NLS	-39.7499	-46.4486	62.5401	64.7558	5479.18	-0.1702	0.1975	-255.44	266.86	90.99	5491.32
[2]-FO	-0.1678	-0.1960	3.9715	2.7604	15.75	0.9966	0.9983	-2.54	6.64	13.66	15.8011
[2]-FOCE	-0.4756	-0.5557	4.4600	3.0639	20.06	0.9957	0.9978	-4.79	9.10	19.57	20.1175

Note: The goodness-of-fit measures are defined in Table 3.

On the model fitting data, the predictions of the random parameters and response variables from the FO and FOCE methods, as well as the computations of the goodness-of-fit statistics for the predictions, follow the same procedures as demonstrated on the model application data. Example predictions on the model fitting data are embedded in the programs provided in the Appendices. They are also listed in Table 4. Interested readers may wish to verify and duplicate the results in the programs. The predictions for all 103 plots of the model fitting data across the potential age range where the fitted model is likely to be applied to make predictions (i.e., the spaghetti plots), are shown in Figure 3. Goodness-of-fit statistics corresponding to the model fitting data are listed in Table 2.

4.4 Choosing the "Best" Prediction

The "best" prediction is chosen based on a combination of statistical, graphical, biological and other considerations, on both model fitting and model application data.

Judging from the goodness-of-fit statistics on the model fitting data (Table 2), and using the overall accuracy measure δ as the example, model [2] estimated from the FO method is the most accurate, with δ=23.8027.

On the model application data (Table 8), model [2] estimated from the FO method is again the most accurate, with δ=15.8011. Therefore, based on the goodness-of-fit statistics alone, model [2] estimated from the FO method provided the "best" prediction.

From the prediction graphs (also referred to as spaghetti plots) shown in Figure 3 for the model fitting data and Figure 4 for the model application data, it can be seen that the predictions from model [2] estimated from the FO method can be negative in some cases when the age is beyond certain ranges. This is caused inherently by the term $\mathbf{Z}_i\hat{\mathbf{u}}_i$ in equation [18] for the FO method. This term could also cause the innate shape of a base model to be altered (Huang et al. 2009a, Yang and Huang 2011b). For the FOCE method, equation [34], which maintains the innate shape of a base model, is used to predict the response variable $\mathbf{y}_i$.

In spite of the negative volume predictions, the prediction graphs from the FO method (in Figures 3 and 4) show no obvious anomaly. In real-world situations, negative volume predictions could be constrained to zero to be biologically meaningful. There are plots where the volume could become zero because of mortality.

For the FOCE method, all volume predictions are non-negative (Figures 3 and 4). But some do not taper-off at older ages, especially on the model application data (Figure 4). This may not coincide well with the expected biological growth. It is also more difficult to establish a meaningful biological up-limit in this case.

Hence, based on the statistical, graphical and biological considerations, and given the available data with their inherent quality, quantity, relevance (e.g., data range and distribution) and limitation, it can be inferred that model [2] estimated from the FO method appears to be the "best" for the volume-age relationship considered in this study. Computation-wise, the FO method is also much simpler than the FOCE method. It does not require iteration. All prediction equations involved in the FO method can be solved algebraically.

5. Additional Notes

5.1 Adjusting the Predictions

For most practical purposes, the nonlinear mixed model methods can generally be considered unbiased. But the unbiasedness property, and indeed, many other properties associated with the NMM methods, hold only in an "asymptotically approximated" sense. When the sample size is not "large enough", or when model specification is problematic, mixed model may produce biased predictions.

The potential bias of a nonlinear mixed model can be removed by different methods. The proportional adjustment method is the simplest and most effective method in many cases. This method is implemented through the calculation of a ratio, called the proportional adjustment ratio PAR_i, from the data in plot *i*:

[48] $$\text{PAR}_i = \frac{\overline{y}_i}{\overline{\hat{y}}_i} = \frac{\frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}}{\frac{1}{n_i}\sum_{j=1}^{n_i} \hat{y}_{ij}} = \frac{\sum_{j=1}^{n_i} y_{ij}}{\sum_{j=1}^{n_i} \hat{y}_{ij}}$$

where PAR_i is the proportional adjustment ratio for plot *i* in the population, y_{ij} is the *j*th observed *y*-value for the *i*th plot in the population and $\hat{y}_{ij}$ is its prediction from the mixed model, $\overline{y}_i$ is the mean of the observed values, $\overline{\hat{y}}_i$ is the mean of the predicted values, and n_i is the number of observations in the *i*th plot. Re-arrange [48] produces:

[49] $$\sum_{j=1}^{n_i} y_{ij} \text{-PAR}_i \cdot \sum_{j=1}^{n_i} \hat{y}_{ij} = 0$$

which implies that the summation of the observed values for plot *i* equals to the PAR_i times the summation of the predicted values. Define the adjusted predictions as:

[50] $$\hat{y}_{ij\_adj} = \text{PAR}_i \cdot \hat{y}_{ij}$$

Then, from [50] and [48]:

[51] $$\overline{\hat{y}}_{ij\_adj} = \frac{1}{n_i}\sum_{j=1}^{n_i} \hat{y}_{ij\_adj} = \frac{\text{PAR}_i}{n_i}\sum_{j=1}^{n_i} \hat{y}_{ij} = \text{PAR}_i \cdot \overline{\hat{y}}_i = \overline{y}_i$$

Hence:

[52] $$\overline{y}_i - \text{PAR}_i \cdot \overline{\hat{y}}_i = 0 \ \ (\text{or } \overline{y}_i - \overline{\hat{y}}_{ij\_adj} = 0)$$

which means that the mean of the observed values for plot *i* equals to the mean of the proportionally adjusted predictions from the mixed model. What is really important about the expressions given in [48]-[52] is that, they all suggest that the mean bias of the adjusted predictions for plot *i* is zero. That is:

[53] $$\overline{e}_{ij\_adj} = \frac{\sum_{j=1}^{n_i} e_{ij\_adj}}{n_i} = \frac{\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_{ij\_adj})}{n_i} = \frac{\sum_{j=1}^{n_i} y_{ij} - \text{PAR}_i \cdot \sum_{j=1}^{n_i} \hat{y}_{ij}}{n_i} = 0$$

where $\bar{e}_{ij\_adj}$ is the mean bias of the adjusted prediction errors for plot *i* in the population. The adjusted predictions obtained for each plot from the plot-specific proportional adjustment method are guaranteed to have a zero mean bias. Since the mean bias for each plot in the population is zero, the mean bias of the adjusted predictions is also guaranteed to be zero for the entire population.

The essence of the proportional adjustment method is to utilize the power of the mixed model method to track the trends of plot-specific data in a population, while simultaneously shifting the predictions up or down proportionally to alleviate the sample size and/or nonlinear approximation and asymptotic issues. This allows for a nonlinear mixed model to fit any plot-specific data as close as possible. With the proportional adjustment method it is always possible to achieve a bias-free fit that closely mimics the data. But the bias-free fit still does not imply the "best" fit, because the "best" fit is not determined by the bias alone.

More specific examples on how to use the proportional adjustment method to adjust plot level and population level predictions are demonstrated elsewhere (Huang 2008, Huang et al. 2013). For most practical purposes, it is recommended that adjusting mixed model predictions shall be done only when the percent mean bias of the unadjusted predictions exceeds ±5% (i.e., $|\bar{e}\%|>5\%$). Otherwise, the gains from adjusting the predictions may not be substantial.

As an example, among the 79 plots of the model application data, four plots have $|\bar{e}\%|$ values that exceed 5% from the FO method. Therefore, adjusted predictions for these plots are obtained following the proportional adjustment method. They are shown in Figure 5. Actual data and computations associated with Figure 5 are listed in Table 9. Relevant plot-specific goodness-of-fit statistics for unadjusted and adjusted predictions are listed in Table 10.

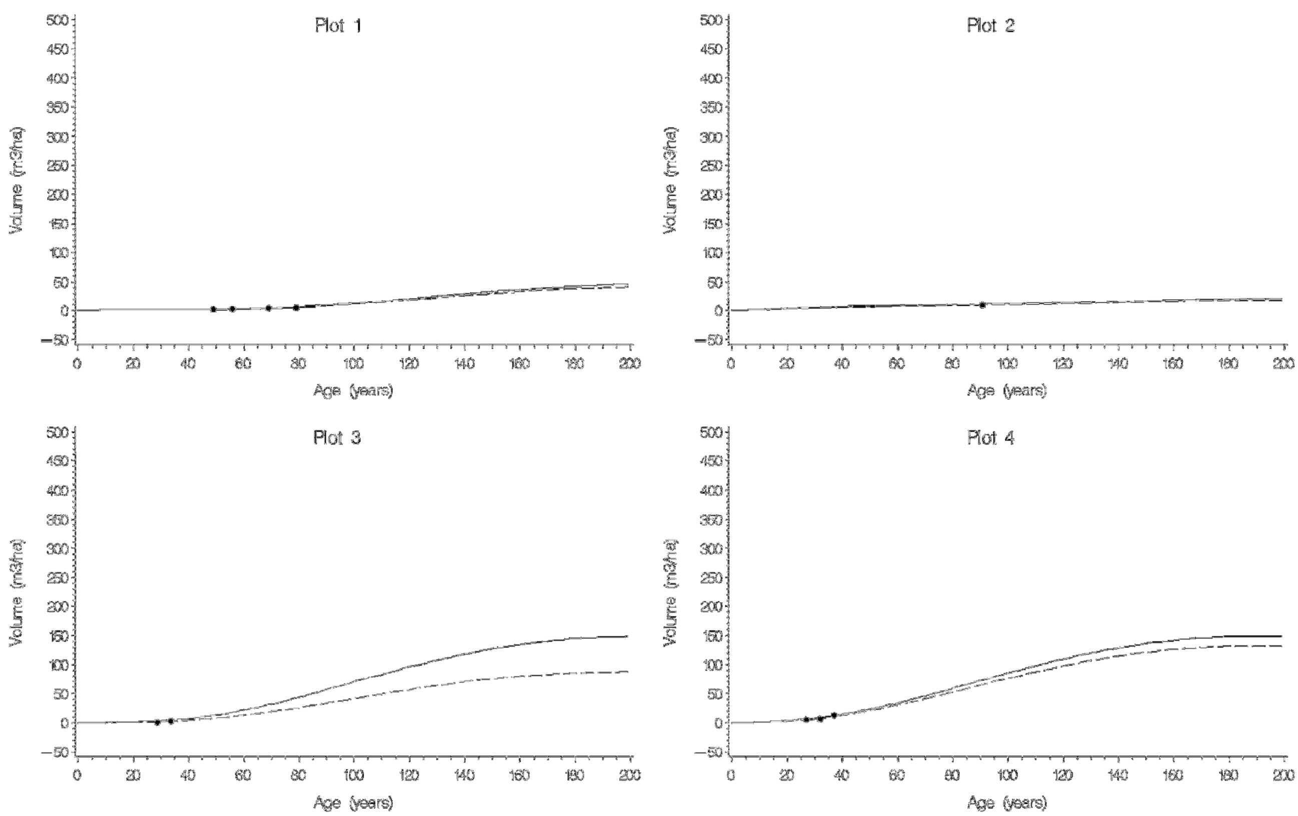

Figure 5. Unadjusted (solid lines) and adjusted (dashed lines) predictions for four plots of the model application data. Actual data and computations are listed in Table 9.

Table 9. Example calculations for proportionally adjusting the predictions from the FO method.

Plot	Time	Age	Volume	$\hat{y}$	e	e%	$\bar{y}$	$\bar{\hat{y}}$	PAR	$\hat{y}_{adj}$	e_{adj}
1	1	49.000	1.664	1.755	-0.091	-10.033	3.013	3.315	0.909	1.595	0.069
1	2	56.000	2.336	2.096	0.240	-10.033	3.013	3.315	0.909	1.905	0.431
1	3	69.000	3.516	3.651	-0.135	-10.033	3.013	3.315	0.909	3.318	0.198
1	4	79.000	4.534	5.757	-1.223	-10.033	3.013	3.315	0.909	5.232	-0.698
2	1	90.822	9.168	10.254	-1.086	-11.843	9.168	10.254	0.894	9.168	0.000
3	1	28.667	1.439	3.116	-1.677	-69.423	2.275	3.854	0.590	1.839	-0.400
3	2	33.667	3.110	4.592	-1.482	-69.423	2.275	3.854	0.590	2.710	0.400
4	1	27.000	5.242	6.537	-1.295	-12.223	8.381	9.405	0.891	5.825	-0.583
4	2	32.000	6.426	9.229	-2.803	-12.223	8.381	9.405	0.891	8.224	-1.798
4	3	37.000	13.474	12.449	1.025	-12.223	8.381	9.405	0.891	11.093	2.381

Note: $\hat{y}$ and e are predicted volume and prediction error, e% is percent mean bias, $\bar{y}$ and $\bar{\hat{y}}$ are means of observed and predicted volumes, PAR is proportional adjustment ratio, and $\hat{y}_{adj}$ and e_{adj} are proportionally adjusted $\hat{y}$ and e.

Table 10. Plot-specific goodness-of-fit statistics from the FO method.

Plot	$\bar{e}$	$\bar{e}$%	SD	MAD	MSE	R^2	CC	MPE	MAPE	e_{10}	δ
					Unadjusted predictions						
1	-0.302	-10.033	0.636	0.422	0.395	0.674	0.896	-6.502	11.634	50.000	0.496
2	-1.086	-11.843	N/A	1.086	1.179	N/A	0.000	-11.843	11.843	100.000	1.179
3	-1.579	-69.423	0.138	1.579	2.503	-2.585	0.330	-82.072	82.072	100.000	2.512
4	-1.024	-12.223	1.928	1.708	3.528	0.733	0.824	-20.239	25.310	66.667	4.767
					Adjusted predictions						
1	0.000	0.000	0.489	0.349	0.179	0.852	0.945	3.209	10.904	50.000	0.239
2	0.000	0.000	N/A	0.000	0.000	N/A	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.566	0.400	0.160	0.771	0.820	-7.466	20.325	100.000	0.320
4	0.000	0.000	2.149	1.587	3.080	0.767	0.827	-7.143	18.923	100.000	4.620

Note: actual data for the four plots are listed in Table 9. The goodness-of-fit measures ($\bar{e}$, $\bar{e}$% ,SD, …, δ) are defined in Table 3 and calculated by plot. N/A denotes "not applicable" (due to a zero denominator).

While visually some of the differences between the unadjusted and adjusted predictions may be hard to see in Figure 5, the plot-specific goodness-of-fit statistics listed in Table 10 clearly indicate that the overall accuracy (δ) of the adjusted predictions is improved for all four plots over their unadjusted counterparts. This suggests that proportionally adjusting the predictions from the FO method is beneficial for all four plots. As expected, the adjusted predictions are guaranteed to be unbiased (i.e., $\bar{e}$ =0), but in general there is no guarantee that they are always more accurate than the unadjusted predictions (Huang et al. 2013).

5.2 Population Average Predictions

The fixed parameters estimated for model [1] from the NLS method (b_1=0.0206 and b_2=2.3622, Table 2) are intended to make population average predictions.

The fixed parameters estimated for model [2] from the FO method (b_1=0.01785 and b_2=2.3396, Table 2) and FOCE method (b_1=0.01879 and b_2=2.3412, Table 2) are intended to make the so-called "marginal prediction" during the estimation of a nonlinear mixed model. They are not intended to make population average predictions. In fact, they generally give biased population average predictions.

The technical aspect of why the fixed parameters estimated as a part of a nonlinear mixed model do not represent the population averages is detailed in Davidian and Giltinan (2003) and Fitzmaurice et al. (2004). Relevant forestry examples are provided in Huang (2008) for the volume-age relationship and Meng et al. (2009) for the height-age relationship.

To briefly illustrate here, the three sets of fixed parameters estimated from the NLS, FO and FOCE methods (Table 2) were used to make population average predictions. The predictions were first made on the model fitting data, then on the model application data. Figure 6 shows the predictions overlaid on the data.

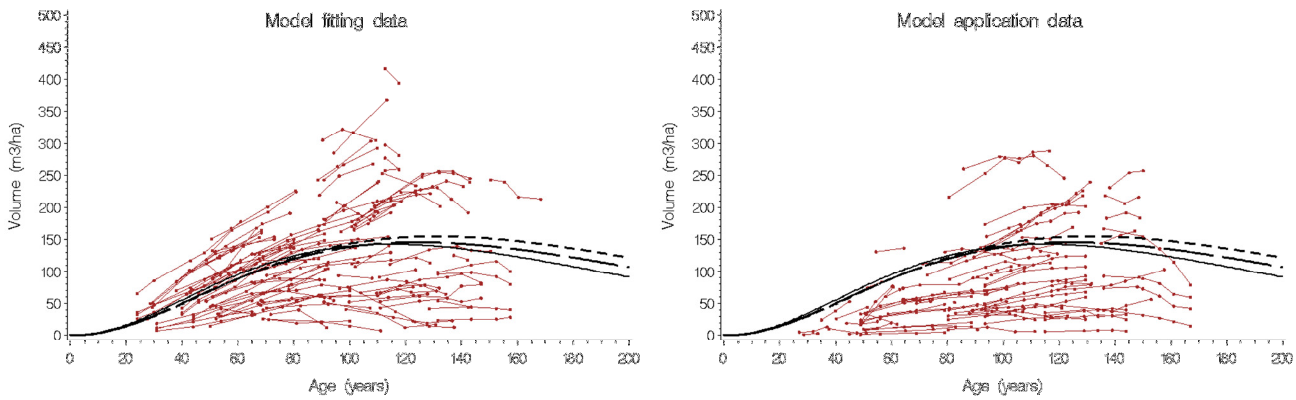

Figure 6. Population average predictions from the fixed parameters of NLS (solid line), FO (short-dash) and FOCE (long-dash) methods. Relevant goodness-of-fit statistics are listed in Table 11.

Alternative Population Average Predictions

On the model application data, alternative population average predictions can be obtained from the FO and FOCE methods, by treating the entire model application data as "one combined plot" with *N* observations (*N*=322). The predictions follow Section 4.1 for the FO method and Section 4.2 for the FOCE method (except that this combined plot has 322 observations). Prediction results across the age range are shown in Figure 7. For interested readers, the predicted random parameters are u_1=-0.00212 and u_2=-0.09594 for the FO method, and u_1=-0.00359 and u_2=-0.11660 for the FOCE method.

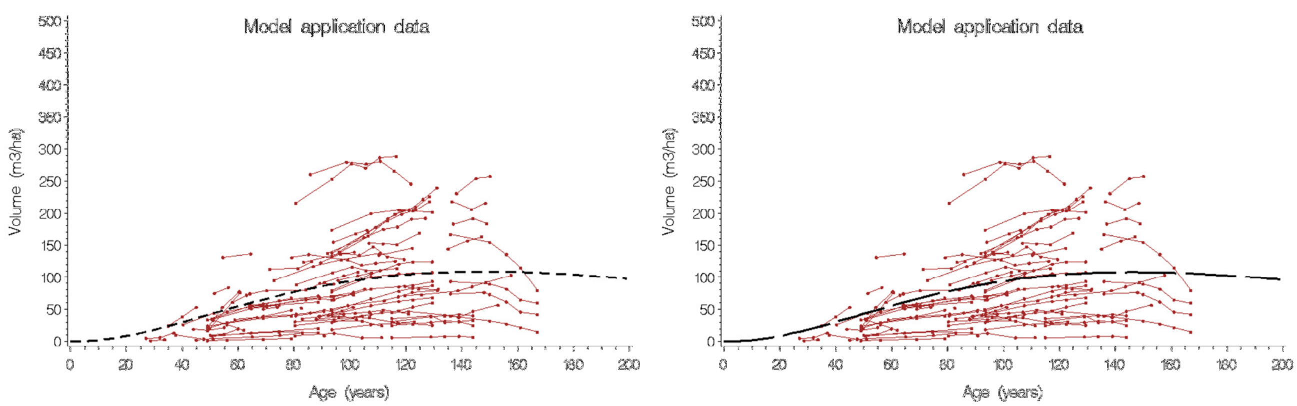

Figure 7. Population average predictions from the FO (left) and FOCE (right) methods. The predictions are obtained by treating the model application data as "one combined plot" with 322 observations.

Table 11 lists the goodness-of-fit statistics associated with different types of population average predictions on model fitting and model application data.

Table 11. Goodness-of-fit statistics from different types of population average predictions.

Model-method	$\bar{e}$	$\bar{e}\%$	SD	MAD	MSE	R^2	CC	MPE	MAPE	e_{10}	δ
	Model fitting data – using fixed parameters only										
[1]-NLS	-0.0354	-0.0304	70.2203	56.4252	4943.64	0.1568	0.2710	-65.49	94.41	89.20	4930.90
[2]-FO	-2.9402	-2.5265	70.8319	57.6746	5038.81	0.1406	0.3032	-68.88	98.53	93.32	5025.80
[2]-FOCE	1.1069	0.9511	70.4146	56.9962	4972.27	0.1519	0.2878	-63.41	94.47	92.29	4959.45
	Model application data – using fixed parameters only										
[1]-NLS	-39.7499	-46.4486	62.5401	64.7558	5479.18	-0.1702	0.1975	-255.44	266.86	90.99	5491.32
[2]-FO	-45.5809	-53.2622	62.1975	66.5746	5934.14	-0.2674	0.2265	-258.94	268.35	90.06	5946.15
[2]-FOCE	-40.1692	-46.9385	62.1745	64.0781	5467.23	-0.1676	0.2212	-248.65	259.55	90.06	5479.24
	Model application data – treating the data as from one combined plot, using fixed and random parameters										
[2]-FO	-1.5724	-1.8374	62.7347	49.6188	3925.90	0.1615	0.2451	-134.08	160.66	93.17	3938.12
[2]-FOCE	-1.6140	-1.8860	62.7362	49.6555	3926.21	0.1615	0.2424	-134.84	161.34	92.86	3938.43

Note: fixed parameters for the three methods are listed in Table 2. The goodness-of-fit measures are defined in Table 3.

Results in Table 11 suggest that:

1. On the model fitting data, the $\bar{e}$ from the NLS method is very close to zero (i.e., $|\bar{e}\%|$ is less than half-a-percent). This is a consequence of the NLS method. A correctly specified and fitted nonlinear model should produce an $\bar{e}$ that is "asymptotically approximately equivalent" to zero on the model fitting data. The $\bar{e}$ values from the FO and FOCE methods using fixed parameters only are much larger. They can be considered biased (i.e., the absolute $\bar{e}\%$ values exceed half-a-percent for both FO and FOCE). Typically, the NLS method is the most accurate method (i.e., with the smallest δ value) in making population average predictions on model fitting data. It is also the preferred method to use on model application data if no measurement is available from the model application data.

2. On the model application data using fixed parameters only, overall the FOCE method (δ=5479.24) is slightly more accurate than the NLS method (δ=5491.32). Both are more accurate than the FO method (δ=5946.15). In general, on model application data using fixed parameters only, the most accurate method varies depending on the specific data involved. There is no generic trend with regard to which method is the best.

3. When treating the model application data as "one combined plot" and using the FO and FOCE methods with fixed and random parameters, the overall accuracies from the FO and FOCE methods are virtually identical (δ=3938.12 for FO and δ=3938.43 for FOCE). Both are substantially better (more accurate) than those on the model application data using fixed parameters only.

Hence, <u>on model application data</u>, population average predictions can be obtained by treating the model application data as "one combined plot", then implementing the FO or FOCE method. The population average predictions obtained in this manner are typically better than those obtained from other methods without adjustment. In practice, such population average predictions can be obtained for any user-defined model application population. They can also be obtained for any sub-population from any specific area.

It may be worthwhile to repeat that, if no *y* measurement is available from the model application data, the full FO and FOCE methods cannot be implemented. In this case the fixed parameters from the NLS method should be used to make population average predictions. If necessary, these predictions can be adjusted to increase the accuracy of the predictions. This can be achieved through the proportional adjustment method (Huang et al. 2013), or some other methods if the proportionality does not hold (Huang 2002).

6. References

Beal, S.L. and Sheiner, L.B. 1982. Estimating population kinetics. *Critical Reviews in Biomedical Engineering* 8: 195-222.

Davidian, M., and Giltinan, D.M. 1995. Nonlinear models for repeated measurement data. Chapman & Hall, New York. 360 p.

Davidian, M., and Giltinan, D.M. 2003. Nonlinear models for repeated measurement data: an overview and update. *Journal of Agric., Biological, and Environmental Statistics* 8: 387–419.

Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. 2004. Applied longitudinal analysis. Wiley, New York. 506 p.

Huang, S. 2002. Validating and localizing growth and yield models: procedures, problems and prospects. An invited paper presented at "Reality, Models and Parameter Estimation", sponsored by IUFRO, Instituto Superior de Gestão and Instituto Superior de Agronomia, June 2-5, 2002, Sesimbra, Portugal, 28 p.

Huang, S. 2008. A generalized procedure for bias-free predictions at population and local levels. A discussion paper, Forest Management Branch, Alberta Ministry of Sustainable Resource Development, Edmonton, Alberta, Pub-Feb-22-2008, 61 p.

Huang, S., Meng, S.X., and Yang, Y. 2009a. Assessing the goodness-of-fit of forest models estimated by nonlinear mixed model methods. *Canadian Journal of Forest Research* 39: 2418-2436.

Huang, S., Meng, S.X. and Yang, Y. 2009b. Prediction implications of nonlinear mixed-effects forest biometric models estimated with a generalized error structure. Proceedings of Joint Statistical Meetings, Section on Statistics and the Environment, August 1–6, 2009, Washington, D.C. American Statistical Association, Alexandria, Va., pp. 1174–1188.

Huang, S., Wiens, D.P., Yang, Y., Meng, S.X., and VanderSchaaf C.L. 2009c. Assessing the impacts of species composition, top height and density on individual tree height prediction of quaking aspen in boreal mixedwoods. *Forest Ecology and Management* 258: 1235-1247.

Huang, S., Yang, Y. and Meng, S.X. 2011. Developing forest models from longitudinal data: a case study assessing the need to account for correlated and/or heterogeneous error structures under a nonlinear mixed model framework. *Journal of Forest Planning* 16: 121–131.

Huang, S., Yang, Y. and Aitkin, D. 2013. Population and plot-specific individual tree height-diameter models for major Alberta tree species. Forest Management Branch, Alberta Environment & Sustainable Resource Development, Tech. Rept. Pub. T/600(2013), Edmonton, Alberta. 81 p.

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

Meng, S.X., and Huang, S. 2009. Improved calibration of nonlinear mixed-effects models demonstrated on a height growth function. *Forest Science* 55: 238-248.

Meng, S.X. and Huang, S. 2010. Incorporating correlated error structure into mixed forest growth models: prediction and inference implications. *Canadian Journal of Forest Research* 40: 977-990.

Meng, S.X., Huang, S., Yang, Y., Trincado, G., and VanderSchaaf C.L. 2009. Evaluation of population-averaged and subject-specific approaches for modeling the dominant or codominant height of lodgepole pine trees. *Canadian Journal of Forest Research* 39: 1148-1158.

Meng, S.X., Huang, S., VanderSchaaf, C.L., Yang, Y. and Trincado, G. 2012. Accounting for serial correlation and its impact on forecasting ability of a fixed- and mixed-effects basal area model: a case study. *European Journal of Forest Research* 131: 541–552.

Pinheiro, J.C. and Bates, D.M. 2004. Mixed effects models in S and S-PLUS. Springer, New York. 528 p.

Vonesh, E.F. and Chinchilli, V.M. 1997. Linear and nonlinear models for the analysis of repeated measurements. Marcel Dekker, New York. 560 p.

Yang, Y. and Huang, S. 2011a. Estimating a multilevel dominant height-age model from nested data with generalized errors. *Forest Science* 57:102–116.

Yang, Y. and Huang, S. 2011b. Comparison of different methods for fitting nonlinear mixed forest models and for making predictions. *Canadian Journal of Forest Research* 41: 1671-1686.

7. Appendices

Four appendices are provided here to demonstrate and facilitate the computations associated with the FO and FOCE methods of the nonlinear mixed-effects modeling technique. Appendices 1 and 2 apply to the FO method. Appendices 3 and 4 apply to the FOCE method. Three plots from the model fitting data (plots m1, m2 and m3) and three plots from the model application data (plots v1, v2 and v3) are used as examples in all four Appendices. They are listed here, where year refers to measurement time/year, and vol and age refer to black spruce total volume (m^3/ha) and breast height age (years), respectively:

Plotid	Year	Vol	Age
m1	1961	285.380	94.332
m1	1968	317.151	101.332
m1	1980	368.217	113.332
m2	1983	33.106	71.746
m2	1993	40.085	81.746
m2	2003	12.273	91.746
m3	1990	415.991	112.667
m3	1995	393.785	117.667
v1	1964	166.812	136.000
v1	1978	154.413	150.000
v1	1984	135.661	156.000
v1	1989	114.873	161.000
v1	1995	79.100	167.000
v2	1990	19.999	49.667
v2	2001	75.227	60.667
v3	1984	32.880	51.000
v3	1989	26.408	56.000
v3	1995	17.720	62.000

Appendix 1. Generalized Program for the First-Order Method

This generalized program for the FO method does not require a user to know or to specify the number of subjects (plots) in the data beforehand. It applies to any number of subjects. The program is designed for experienced SAS/IML users. A more intuitive program is given in Appendix 2.

```
1
2     OPTIONS LS=100 PS=45;
3
4     data comb1;
5      input plotid $ YEAR vol age;
6      cards;
NOTE: DATALINES 7 to 24 (for plots m1, m2, m3, v1, v2 and v3);
25   ;
26   run;
27
28   proc sort data=comb1;
29     by plotid;
30   run;
31
32   data wed3;
33   set comb1;
34   by plotid;
35   j+1;
36   if first.plotid then do; i+1; j=1; end;
37   run;
38
39   proc iml;
40   use wed3;
41   read all var {j} into tobs;
42   read all var {i age} into age;
43   read all var {vol} into vol;
44   fixp={0.01785 2.3396};
45   covar={0.0000641, 0.0006674, 0.02049, 45.0368};
46   d=j(2,2,0);
47   d[1,1]=covar[1];
48   d[1,2]=covar[2];
49   d[2,1]=covar[2];
50   d[2,2]=covar[3];
51   s=covar[4];
52   bb=fixp[1,];
53   b=bb`;
54   tn=max(age[,1]);
55   q=2;
56   bx={1 1};
57   nn=nrow(vol);
58   u=j(tn,q,0);
59   mc=max(tobs);
60   start sm (tn,bx,q,u,z,b,s,d,age,vol,nn,res,uv);
61   z=j(nn,q,0);
62   res=j(nn,1,0);
63   uv=j(tn,q,0);
64   do k=1 to tn;
65   z[1:nn,]=.;
66   res[1:nn,]=.;
67    do j=1 to nn;
68    if age[j,1]=k then;
69    do;
```

```
 agem=age[j,2];
 volm=vol[j,1];
 zb1 = agem**b[2]*(1-b[1]*agem)*exp(-b[1]*agem);
 z[j,1]=zb1;
 zb2 = b[1]*exp(-b[1]*agem)*log(agem)*agem**b[2];
 z[j,2]=zb2;
 re=volm-b[1]*agem**b[2]*exp(-b[1]*agem);
 res[j]=re;
end;
end;
r1=z;
r2=r1[loc(r1[,1]^=.),]#bx;
w1=res;
w2=w1[loc(w1[,1]^=.),];
mm=nrow(w2);
rr=s*I(mm);
uu=d*r2`*INV(r2*d*r2`+rr)*w2;
uk=uu`;
uv[k,]=uk;
end;
finish sm;
run sm (tn,bx,q,u,z,b,s,d,age,vol,nn,res,uv);
ubu=uv;
bf=j(tn,q,0);
do i=1 to tn;
bf[i,1]=b[1];
bf[i,2]=b[2];
end;
ub=ubu||bf;
cnm={u1i,u2i,b1,b2};
create rpm from ub[colname=cnm];
append from ub;
quit;
data rpm1;
set rpm;
i=_n_;
run;
proc sort data=wed3; by i;run;
proc sort data=rpm1; by i;run;
data allP ;
 merge wed3 rpm1 ;
 by i ;
 zb1 = age**b2*(1-b1*age)*exp(-b1*age);
 zb2 = b1*exp(-b1*age)*log(age)*age**b2;
 vol_fix=b1*age**b2*exp(-b1*age);
 res_fix=vol-vol_fix;
 y_pred = vol_fix+u1i*zb1+u2i*zb2;
 Y_res=vol-y_pred;
run;
proc print data=allP(obs=18);
 var plotid i j age vol vol_fix zb1 zb2 u1i u2i y_pred y_res;
run;
```

The print statement produces the following results. They are listed in Table 4 (**FO method**) for the model fitting data (plots m1, m2 and m3) and Table 6 for the model application data (plots v1, v2 and v3).

Obs	plotid	i	j	age	vol	vol_fix	zb1	zb2	U1I	U2I	y_pred	y_res
1	m1	1	1	94.332	285.380	138.126	-5291.53	628.032	-0.015328	0.11080	288.821	-3.4410
2	m1	1	2	101.332	317.151	144.126	-6530.28	665.631	-0.015328	0.11080	317.975	-0.8237
3	m1	1	3	113.332	368.217	151.159	-8662.86	715.031	-0.015328	0.11080	363.170	5.0469
4	m2	2	1	71.746	33.106	108.962	-1713.28	465.610	0.007109	-0.13121	35.688	-2.5817
5	m2	2	2	81.746	40.085	123.691	-3181.77	544.687	0.007109	-0.13121	29.600	10.4846
6	m2	2	3	91.746	12.273	135.543	-4842.09	612.523	0.007109	-0.13121	20.748	-8.4753
7	m3	3	1	112.667	415.991	150.872	-8546.10	712.787	0.008030	0.44998	402.985	13.0056
8	m3	3	2	117.667	393.785	152.746	-9415.94	728.270	0.008030	0.44998	404.841	-11.0559
9	v1	4	1	136.000	166.812	154.516	-12357.79	759.082	0.016390	0.29303	174.400	-7.5881
10	v1	4	2	150.000	154.413	151.355	-14224.01	758.387	0.016390	0.29303	140.448	13.9651
11	v1	4	3	156.000	135.661	149.051	-14901.80	752.688	0.016390	0.29303	125.365	10.2960
12	v1	4	4	161.000	114.873	146.768	-15407.33	745.787	0.016390	0.29303	112.773	2.0998
13	v1	4	5	167.000	79.100	143.647	-15941.56	735.183	0.016390	0.29303	97.789	-18.6886
14	v2	5	1	49.667	19.999	68.349	434.39	266.926	-0.015157	-0.09195	37.222	-17.2229
15	v2	5	2	60.667	75.227	89.688	-416.56	368.205	-0.015157	-0.09195	62.146	13.0805
16	v3	6	1	51.000	32.880	71.008	356.63	279.192	0.001898	-0.16995	24.237	8.6433
17	v3	6	2	56.000	26.408	80.831	1.81	325.373	0.001898	-0.16995	25.537	0.8706
18	v3	6	3	62.000	17.720	92.148	-550.82	380.306	0.001898	-0.16995	26.469	-8.7492

Appendix 2. Step-by-Step Program for the First-Order Method

This step-by-step program for the FO method requires a user to know the exact number of subjects (plots) in the data beforehand. This number is specified in line 52. For the example data, the exact number of subjects is six: three (m1, m2 and m3) from the model fitting data and three (v1, v2 and v3) from the model application data.

For any data set, if the exact number of subjects is known and specified in line 52, this step-by-step program also applies to any number of subjects. Some readers may find this step-by-step program is relatively easier to follow than the generalized program, although both programs produce the same results.

```
1
2     OPTIONS LS=100 PS=45;
3
4     data comb1;
5      input plotid $    YEAR    vol age;
6    cards;
NOTE: DATALINES 7 to 24 (for plots m1, m2, m3, v1, v2 and v3);
25   ;
26   run;
27
28   proc sort data=comb1;
29     by plotid;
30   run;
31
32   data wed3;
33   set comb1;
34   by plotid;
35   j+1;
36   if first.plotid then do; i+1; j=1; end;
37   run;
38
39   data wed4;
40     set wed3; by plotid;
41        b1=0.01785; b2=2.3396;
42    zb1 = age**b2*(1-b1*age)*exp(-b1*age);
43    zb2 = b1*exp(-b1*age)*log(age)*age**b2;
44    vol_fix=b1*age**b2*exp(-b1*age);
45    res_fix=vol-vol_fix;
46   run;
47
48   filename random 'c:\_localdata\random.txt' ;
49   proc iml;
50   file random;
51   use wed4;
52   do k=1 to 6;
53       read all var {zb1 zb2} into Z where (i=k);
54       read all var {res_fix} into RES where (i=k);
55       read all var {j} into MM where (i=k);
56   ss=nrow(mm);
57   R= 45.0368 * I(ss);
58   D= {0.0000641 0.0006674, 0.0006674 0.02049};
59   b=D*Z`*INV(Z * D * Z` + R)*RES;
60       bTrans = b`;
61       u1i= bTrans[1,1] ;
62       u2i= bTrans[1,2] ;
63       put k 5. +2 u1i 15.10 +2 u2i 15.10;
```

```
end;

closefile random ;
quit ;

data prandom ;
infile random;
input i u1i u2i ;
run ;

data allP ;
 merge wed4 Prandom ;
 by i ;
 zb1 = age**b2*(1-b1*age)*exp(-b1*age);
 zb2 = b1*exp(-b1*age)*log(age)*age**b2;
 vol_fix=b1*age**b2*exp(-b1*age);
 res_fix=vol-vol_fix;
 y_pred = vol_fix+u1i*zb1+u2i*zb2;
 Y_res=vol-y_pred;
 run;
proc print data=allP(obs=18);
 var plotid i j age vol vol_fix zb1 zb2 u1i u2i y_pred y_res ;
run;
```

The print statement produces the same results as those from Appendix 1. They are listed in Table 4 (**FO method**) for the model fitting data (plots m1, m2 and m3) and Table 6 for the model application data (plots v1, v2 and v3).

Appendix 3. Generalized Program for the First-Order Conditional Expectation Method

This generalized program for the FOCE method does not require a user to know or to specify the number of subjects (plots) in the data beforehand. It applies to any number of subjects.

A convergence criterion of 10^{-7} is specified in line 97 for the iteration, which is equivalent to a precision of 0.0000001. Readers may wish to choose a different convergence criterion in some cases (such as 10^{-6} or 10^{-5}), but are not recommended to go above 10^{-5}.

```
1
2     OPTIONS LS=100 PS=45;
3
4     data comb1;
5      input plotid $    YEAR    vol age;
6    cards;
NOTE: DATALINES 7 to 24 (for plots m1, m2, m3, v1, v2 and v3);
25   ;
26   run;
27
28   proc sort data=comb1;
29     by plotid;
30   run;
31
32   data wed3;
33   set comb1;
34   by plotid;
35   j+1;
36   if first.plotid then do; i+1; j=1; end;
37   run;
38
39   proc iml;
40   use wed3;
41   read all var {j} into tobs;
42   read all var {i age} into age;
43   read all var {vol} into vol;
44   covar={0.0000878, 0.0005871, 0.02121, 44.6634};
45   fixp={0.01879 2.3412};
46   d=j(2,2,0);
47   d[1,1]=covar[1];
48   d[1,2]=covar[2];
49   d[2,1]=covar[2];
50   d[2,2]=covar[3];
51   s=covar[4];
52   bb=fixp[1,];
53   b=bb`;
54   tn=max(age[,1]);
55   q=2;
56   bx={1 1};
57   nn=nrow(vol);
58   u=j(tn,q,0);
59   start sm (tn,bx,q,u,z,b,s,d,age,vol,nn,res,uv);
60   z=j(nn,q,0);
61   res=j(nn,1,0);
62   uv=j(tn,q,0);
63   do k=1 to tn;
64   z[1:nn,]=.;
65   res[1:nn,]=.;
```

```
 do j=1 to nn;
 if age[j,1]=k then;
 do;
 u1=u[k,1];
 u2=u[k,2];
 agem=age[j,2];
 volm=vol[j,1];
 zb1 = agem**(b[2]+u2)*(1-(b[1]+u1)*agem)*exp(-(b[1]+u1)*agem);
 z[j,1]=zb1;
 zb2 = (b[1]+u1)*exp(-(b[1]+u1)*agem)*log(agem)*agem**(b[2]+u2);
 z[j,2]=zb2;
 re=volm-(b[1]+u1)*agem**(b[2]+u2)*exp(-(b[1]+u1)*agem)+u1*zb1+u2*zb2;
 res[j]=re;
end;
end;
r1=z;
r2=r1[loc(r1[,1]^=.),]#bx;
w1=res;
w2=w1[loc(w1[,1]^=.),];
mm=nrow(w2);
rr=s*I(mm);
uu=d*r2`*INV(r2*d*r2`+rr)*w2;
uk=uu`;
uv[k,]=uk;
end;
finish sm;
run sm (tn,bx,q,u,z,b,s,d,age,vol,nn,res,uv);
bu=uv;
do k=1 to tn;
diff1=1;
diff2=1;
eps1=10E-7;
 do iter=1 to 1000 until ((diff1<eps1) & (diff2<eps1));
run sm (tn,bx,q,u,z,b,s,d,age,vol,nn,res,uv);
diff=abs(uv-u);
diff1=max(diff[,1]);
diff2=max(diff[,2]);
u=uv;
end;
end;
ubu=u;
bf=j(tn,q,0);
do i=1 to tn;
bf[i,1]=b[1];
bf[i,2]=b[2];
end;
ub=bu||ubu||bf;
cnm={b1i,b2i,ub1i,ub2i,b1,b2};
create rpm from ub[colname=cnm];
append from ub;
quit;

data rpm1;
set rpm;
i=_n_;
run;

proc sort data=wed3;by i;run;
proc sort data=rpm1;by i;run;
```

```
data allx ;
 merge wed3 rpm1;
 by i ;
 vol_fix  = (b1     )*age**(b2     )*exp(-(b1     )*age);
 Y_pred = (b1+ub1i)*age**(b2+ub2i)*exp(-(b1+ub1i)*age);
 Y_res= vol - y_pred;
 zb1 = age**(b2+ub2i)*(1-(b1+ub1i)*age)*exp(-(b1+ub1i)*age);
 zb2 = (b1+ub1i)*exp(-(b1+ub1i)*age)*log(age)*age**(b2+ub2i);
proc print data=allx(obs=18);
 var plotid i j age vol vol_fix zb1 zb2 ub1i ub2i y_pred y_res ;
run;
```

The print statement produces the following results. They are listed in Table 4 (**FOCE method**) for the model fitting data (plots m1, m2 and m3) and Table 7 (**Final iteration results**) for the model application data (plots v1, v2 and v3).

Obs	plotid	i	j	age	vol	vol_fix	zb1	zb2	UB1I	UB2I	Y_pred	Y_res
1	m1	1	1	94.332	285.380	134.033	-316.59	1302.54	-0.008063	0.12306	286.473	-1.0932
2	m1	1	2	101.332	317.151	138.954	-2569.47	1464.11	-0.008063	0.12306	317.017	0.1341
3	m1	1	3	113.332	368.217	144.126	-7383.55	1737.20	-0.008063	0.12306	367.247	0.9699
4	m2	2	1	71.746	33.106	107.955	-1368.35	138.83	0.014962	-0.16687	32.488	0.6179
5	m2	2	2	81.746	40.085	121.427	-1604.53	135.57	0.014962	-0.16687	30.786	9.2988
6	m2	2	3	91.746	12.273	131.842	-1753.76	127.58	0.014962	-0.16687	28.233	-15.9595
7	m3	3	1	112.667	415.991	143.942	-31229.07	1932.13	0.008754	0.34883	408.965	7.0260
8	m3	3	2	117.667	393.785	145.056	-32585.28	1909.53	0.008754	0.34883	400.501	-6.7163
9	v1	4	1	136.000	166.812	144.263	-18376.22	845.05	0.015490	0.34226	172.016	-5.2036
10	v1	4	2	150.000	154.413	139.487	-16730.03	693.78	0.015490	0.34226	138.461	15.9524
11	v1	4	3	156.000	135.661	136.599	-15882.74	632.39	0.015490	0.34226	125.230	10.4312
12	v1	4	4	161.000	114.873	133.883	-15137.21	583.47	0.015490	0.34226	114.825	0.0483
13	v1	4	5	167.000	79.100	130.306	-14213.15	527.78	0.015490	0.34226	103.122	-24.0218
14	v2	5	1	49.667	19.999	69.097	4917.70	159.44	-0.012912	-0.00138	40.827	-20.8279
15	v2	5	2	60.667	75.227	89.765	6689.36	250.91	-0.012912	-0.00138	61.117	14.1104
16	v3	6	1	51.000	32.880	71.698	1405.50	85.59	-0.010137	-0.23744	21.769	11.1109
17	v3	6	2	56.000	26.408	81.246	1511.77	102.17	-0.010137	-0.23744	25.381	1.0275
18	v3	6	3	62.000	17.720	92.115	1598.90	123.20	-0.010137	-0.23744	29.850	-12.1301

Appendix 4. Step-by-Step Program for the First-Order Conditional Expectation Method

This step-by-step program for the FOCE method corresponds to the three-step iteration procedure described in Section 3.3 and demonstrated in Section 4.2.

```
1    OPTIONS LS=100 PS=45;
2
3     data comb1;
4      input plotid $    YEAR    vol age;
5    cards;
NOTE: DATALINES 6 to 23 (for plots m1, m2, m3, v1, v2 and v3);
24   ;
25   run;
26
27   proc sort data=comb1;
28     by plotid;
29   run;
30
31   data wed3;
32   set comb1;
33   by plotid;
34   j+1;
35   if first.plotid then do; i+1; j=1; end;
36   run;
37
38   data wed4;
39   set wed3;
40   by i;
41        b1=0.01879; b2=2.3412;
42    z1 = age**b2*(1-b1*age)*exp(-b1*age);
43    z2 = b1*exp(-b1*age)*log(age)*age**b2;
44    vol_fix = b1*age**b2*exp(-b1*age);
45    res_fix = vol - b1*age**b2*exp(-b1*age);
46   run;
47
48   filename random 'c:\_localdata\random.txt' ;
49   proc iml;
50   file random;
51   use wed4;
52   do k=1 to 6;
53       read all var {z1 z2} into Z where (i=k);
54       read all var {res_fix} into RES where (i=k);
55       read all var {j} into MM where (i=k);
56   ss=nrow(mm);
57   R= 44.6634 * I(ss);
58   D= { 0.0000878 0.0005871, 0.0005871 0.02121};
59   b=D*Z`*  INV(Z * D * Z` + R)*RES;
60       bTrans = b`;
61       u1i= bTrans[1,1] ;
62       u2i= bTrans[1,2] ;
63       put k 5. +2 u1i 15.10 +2 u2i 15.10;
64   end;
65   closefile random ;
66   quit ;
67
68   data prandom ;
69   infile random;
70   input     i u1i u2i;
```

```
run ;

data all ;
 merge wed4 Prandom ;
 by i ;
proc print data=all;
  var  plotid i j   age     vol   vol_fix  z1   z2 u1i u2i;
run;
```

This print statement produces the step one results. Table 7 (<u>**Step 1 computation**</u>) lists the results for plot v1 (shaded area) of the model application data.

Obs	plotid	i	j	age	vol	vol_fix	z1	z2	u1i	u2i
1	m1	1	1	94.332	285.380	134.033	-5510.40	609.425	-0.019342	0.07745
2	m1	1	2	101.332	317.151	138.954	-6685.39	641.747	-0.019342	0.07745
3	m1	1	3	113.332	368.217	144.126	-8663.75	681.764	-0.019342	0.07745
4	m2	2	1	71.746	33.106	107.955	-2000.01	461.308	0.008509	-0.11649
5	m2	2	2	81.746	40.085	121.427	-3463.84	534.716	0.008509	-0.11649
6	m2	2	3	91.746	12.273	131.842	-5079.35	595.795	0.008509	-0.11649
7	m3	3	1	112.667	415.991	143.942	-8556.92	680.043	0.003230	0.42059
8	m3	3	2	117.667	393.785	145.056	-9348.43	691.605	0.003230	0.42059
9	v1	4	1	136.000	166.812	144.263	-11942.11	708.714	0.016647	0.32237
10	v1	4	2	150.000	154.413	139.487	-13499.55	698.917	0.016647	0.32237
11	v1	4	3	156.000	135.661	136.599	-14039.71	689.807	0.016647	0.32237
12	v1	4	4	161.000	114.873	133.883	-14429.90	680.312	0.016647	0.32237
13	v1	4	5	167.000	79.100	130.306	-14826.26	666.906	0.016647	0.32237
14	v2	5	1	49.667	19.999	69.097	245.49	269.847	-0.020829	-0.10756
15	v2	5	2	60.667	75.227	89.765	-668.50	368.523	-0.020829	-0.10756
16	v3	6	1	51.000	32.880	71.698	159.16	281.906	0.005873	-0.16514
17	v3	6	2	56.000	26.408	81.246	-225.88	327.044	0.005873	-0.16514
18	v3	6	3	62.000	17.720	92.115	-808.79	380.171	0.005873	-0.16514

```

data wed41;
set all;
by i;
     b1=0.01879; b2=2.3412;
 z11 = age**(b2+u2i)*(1-(b1+u1i)*age)*exp(-(b1+u1i)*age);
 z22 = (b1+u1i)*exp(-(b1+u1i)*age)*log(age)*age**(b2+u2i);
 vol_fix = b1*age**b2*exp(-b1*age);
 y_pred = (b1+u1i)*age**(b2+u2i)*exp(-(b1+u1i)*age) ;
 y_res =vol- y_pred ;
 res_c = vol -(b1+u1i)*age**(b2+u2i)*exp(-(b1+u1i)*age)+z11*u1i+z22*u2i ;
 drop u1i u2i;
run;

filename random1 'c:\_localdata\random.txt' ;
proc iml;
file random1;
use wed41;
do k=1 to 6;
    read all var {z11 z22} into Z where (i=k);
    read all var {res_c} into RES where (i=k);
    read all var {j} into MM where (i=k);
ss=nrow(mm);
R= 44.6634 * I(ss);
D= { 0.0000878 0.0005871, 0.0005871 0.02121};
b=D*Z`*  INV(Z * D * Z` + R)*RES;
```

```
    bTrans = b`;
    u1i= bTrans[1,1] ;
    u2i= bTrans[1,2] ;
     put k 5. +2 u1i 15.10 +2 u2i 15.10;
end;
closefile random1 ;
quit ;

data prandom1;
infile random1;
input     i u1i u2i;
run ;

data all ;
 merge wed41 Prandom1;
 by i;
proc print data=all(obs=18);
 var plotid i j age vol vol_fix z11 z22 y_pred y_res  u1i u2i ;
run;
```

This print statement produces the step two results. Table 7 (**Step 2 computation**) lists the results for plot v1 (shaded area) of the model application data.

Obs	plotid	i	j	age	vol	vol_fix	z11	z22	y_pred	y_res	u1i	u2i
1	m1	1	1	94.332	285.380	134.033	66166.35	-157.77	-34.700	320.080	-0.015225	-0.01019
2	m1	1	2	101.332	317.151	138.954	79267.59	-191.29	-41.419	358.570	-0.015225	-0.01019
3	m1	1	3	113.332	368.217	144.126	105255.82	-258.54	-54.655	422.872	-0.015225	-0.01019
4	m2	2	1	71.746	33.106	107.955	-1818.17	221.25	51.777	-18.671	0.015979	-0.13909
5	m2	2	2	81.746	40.085	121.427	-2376.70	231.99	52.681	-12.596	0.015979	-0.13909
6	m2	2	3	91.746	12.273	131.842	-2856.71	234.23	51.832	-39.559	0.015979	-0.13909
7	m3	3	1	112.667	415.991	143.942	-57504.73	4039.56	855.036	-439.045	0.007605	0.37345
8	m3	3	2	117.667	393.785	145.056	-62389.66	4116.93	863.475	-469.690	0.007605	0.37345
9	v1	4	1	136.000	166.812	144.263	-14850.90	676.90	137.787	29.025	0.014365	0.32107
10	v1	4	2	150.000	154.413	139.487	-13263.79	545.73	108.915	45.498	0.014365	0.32107
11	v1	4	3	156.000	135.661	136.599	-12490.57	493.62	97.750	37.911	0.014365	0.32107
12	v1	4	4	161.000	114.873	133.883	-11824.83	452.52	89.055	25.818	0.014365	0.32107
13	v1	4	5	167.000	79.100	130.306	-11014.71	406.20	79.367	-0.267	0.014365	0.32107
14	v2	5	1	49.667	19.999	69.097	7486.57	-54.14	-13.862	33.861	-0.013579	-0.08255
15	v2	5	2	60.667	75.227	89.765	12213.76	-90.99	-22.163	97.390	-0.013579	-0.08255
16	v3	6	1	51.000	32.880	71.698	-380.91	143.27	36.438	-3.558	0.003886	-0.25433
17	v3	6	2	56.000	26.408	81.246	-610.12	158.92	39.481	-13.073	0.003886	-0.25433
18	v3	6	3	62.000	17.720	92.115	-911.61	175.37	42.492	-24.772	0.003886	-0.25433

```

data wed41;
set all;
by i;
     b1=0.01879; b2=2.3412;
 z11 = age**(b2+u2i)*(1-(b1+u1i)*age)*exp(-(b1+u1i)*age);
 z22 = (b1+u1i)*exp(-(b1+u1i)*age)*log(age)*age**(b2+u2i);
 vol_fix = b1*age**b2*exp(-b1*age);
 y_pred = (b1+u1i)*age**(b2+u2i)*exp(-(b1+u1i)*age) ;
 y_res =vol- y_pred ;
 res_c = vol -(b1+u1i)*age**(b2+u2i)*exp(-(b1+u1i)*age)+z11*u1i+z22*u2i ;
 drop u1i u2i;
run;

filename random1 'c:\_localdata\random.txt' ;
```

```
139  proc iml;
140  file random1;
141  use wed41;
142  do k=1 to 6;
143      read all var {z11 z22} into Z where (i=k);
144      read all var {res_c} into RES where (i=k);
145      read all var {j} into MM where (i=k);
146  ss=nrow(mm);
147  R= 44.6634 * I(ss);
148  D= { 0.0000878 0.0005871, 0.0005871 0.02121};
149  b=D*Z`*  INV(Z * D * Z` + R)*RES;
150      bTrans = b`;
151      u1i= bTrans[1,1] ;
152      u2i= bTrans[1,2] ;
153      put k 5. +2 u1i 15.10 +2 u2i 15.10;
154  end;
155  closefile random1 ;
156  quit ;
157
158  data prandom1;
159  infile random1;
160  input     i u1i u2i;
161  run ;
162
163  data all ;
164   merge wed41 Prandom1;
165   by i;
166  proc print data=all(obs=18);
167   var plotid i j age vol vol_fix z11 z22 y_pred y_res  u1i u2i ;
168  run;
```

This print statement produces the step three results. Table 7 (**<u>Step 3 computation</u>**) lists the results for plot v1 (shaded area) of the model application data.

Obs	plotid	i	j	age	vol	vol_fix	z11	z22	y_pred	y_res	u1i	u2i
1	m1	1	1	94.332	285.380	134.033	19006.24	464.12	102.077	183.303	-0.003373	-0.10487
2	m1	1	2	101.332	317.151	138.954	21080.58	543.31	117.640	199.511	-0.003373	-0.10487
3	m1	1	3	113.332	368.217	144.126	24462.63	692.10	146.311	221.906	-0.003373	-0.10487
4	m2	2	1	71.746	33.106	107.955	-1505.97	149.71	35.035	-1.929	0.015514	-0.16054
5	m2	2	2	81.746	40.085	121.427	-1747.60	145.24	32.983	7.102	0.015514	-0.16054
6	m2	2	3	91.746	12.273	131.842	-1891.89	135.74	30.037	-17.764	0.015514	-0.16054
7	m3	3	1	112.667	415.991	143.942	-37471.02	2367.32	501.079	-85.088	0.009528	0.36545
8	m3	3	2	117.667	393.785	145.056	-39416.78	2355.63	494.065	-100.280	0.009528	0.36545
9	v1	4	1	136.000	166.812	144.263	-18491.83	858.32	174.717	-7.905	0.015611	0.34510
10	v1	4	2	150.000	154.413	139.487	-17085.61	714.37	142.572	11.841	0.015611	0.34510
11	v1	4	3	156.000	135.661	136.599	-16323.01	655.03	129.714	5.947	0.015611	0.34510
12	v1	4	4	161.000	114.873	133.883	-15638.78	607.37	119.527	-4.654	0.015611	0.34510
13	v1	4	5	167.000	79.100	130.306	-14776.99	552.68	107.988	-28.888	0.015611	0.34510
14	v2	5	1	49.667	19.999	69.097	3875.45	106.42	27.250	-7.251	-0.011322	-0.03182
15	v2	5	2	60.667	75.227	89.765	5305.27	165.98	40.430	34.797	-0.011322	-0.03182
16	v3	6	1	51.000	32.880	71.698	-180.15	102.66	26.109	6.771	0.000803	-0.27491
17	v3	6	2	56.000	26.408	81.246	-337.18	114.06	28.335	-1.927	0.000803	-0.27491
18	v3	6	3	62.000	17.720	92.115	-547.42	126.22	30.583	-12.863	0.000803	-0.27491

Carry out the iteration manually by repeatedly running the program lines 125 to 168 for 60 times. This produces the following results:

Obs	plotid	i	j	age	vol	vol_fix	z11	z22	y_pred	y_res	u1i	u2i
1	m1	1	1	94.332	285.380	134.033	-316.59	1302.54	286.473	-1.0932	-0.008063	0.12306
2	m1	1	2	101.332	317.151	138.954	-2569.47	1464.11	317.017	0.1341	-0.008063	0.12306
3	m1	1	3	113.332	368.217	144.126	-7383.55	1737.20	367.247	0.9699	-0.008063	0.12306
4	m2	2	1	71.746	33.106	107.955	-1368.35	138.83	32.488	0.6179	0.014962	-0.16687
5	m2	2	2	81.746	40.085	121.427	-1604.53	135.57	30.786	9.2988	0.014962	-0.16687
6	m2	2	3	91.746	12.273	131.842	-1753.76	127.58	28.233	-15.9595	0.014962	-0.16687
7	m3	3	1	112.667	415.991	143.942	-31229.07	1932.13	408.965	7.0260	0.008754	0.34883
8	m3	3	2	117.667	393.785	145.056	-32585.28	1909.53	400.501	-6.7163	0.008754	0.34883
9	v1	4	1	136.000	166.812	144.263	-18376.22	845.05	172.016	-5.2036	0.015490	0.34226
10	v1	4	2	150.000	154.413	139.487	-16730.03	693.78	138.461	15.9524	0.015490	0.34226
11	v1	4	3	156.000	135.661	136.599	-15882.74	632.39	125.230	10.4312	0.015490	0.34226
12	v1	4	4	161.000	114.873	133.883	-15137.21	583.47	114.825	0.0483	0.015490	0.34226
13	v1	4	5	167.000	79.100	130.306	-14213.15	527.78	103.122	-24.0218	0.015490	0.34226
14	v2	5	1	49.667	19.999	69.097	4917.70	159.44	40.827	-20.8279	-0.012912	-0.00138
15	v2	5	2	60.667	75.227	89.765	6689.36	250.91	61.117	14.1104	-0.012912	-0.00138
16	v3	6	1	51.000	32.880	71.698	1405.50	85.59	21.769	11.1109	-0.010137	-0.23744
17	v3	6	2	56.000	26.408	81.246	1511.76	102.17	25.381	1.0275	-0.010137	-0.23744
18	v3	6	3	62.000	17.720	92.115	1598.89	123.20	29.850	-12.1301	-0.010137	-0.23744

The results obtained after 60 iterations are equivalent to those obtained from Appendix 3 (except occasional decimal places for some intermediate computations).

Notes:

1. Depending on the data and model involved and the number of iterations carried out, some minuscule differences at certain decimal places may occur between the final results from Appendices 3 and 4. This is caused primarily by different ways of computations (e.g., the generalized program in Appendix 3 uses a "do…until" statement with a maximum iteration of up to 1000).

2. In most cases, stable predictions for the random parameters can be achieved in less than 10 iterations, so 60 iterations used in the above example are more than enough for most data. However, there are cases where the number of iterations may need to be increased (e.g., to 100, 200 or even higher) until the final predictions show no practical improvement. Of course, the manual iteration of repeatedly running lines 125 to 168 many times can be programmed using a macro (available to interested readers).

3. The step-by-step program for the FOCE method also applies to any number of subjects, provided that the number of subjects is specified in lines 52, 97 and 142. Some practitioners may find the step-by-step program is relatively easier to follow than the generalized program in Appendix 3. The step-by-step program also allows for easier diagnostics when convergence is not achieved or does not exist for some specific data. The non-convergence problem of the FOCE method, which is often caused by the specific data involved and the model specification used, is an area that deserves further studies.

Appendix 5. Metric Conversion Chart

1 cm	=	0.39370 in.
1 m	=	3.28083 ft.
1 m	=	1.09361 yards
1 ha	=	2.47105 acres
1 m^2	=	10.76385 ft^2
1 m^3	=	35.31435 ft^3
1 m^2/ha	=	4.3560 ft^2/acre
1 m^3/ha	=	14.2913 ft^3/acre
1 ha	=	10000 m^2
1 km	=	1000 m
1 km	=	0.62137 miles
1 km^2	=	100 ha
1 km^2	=	0.3861 $miles^2$
1 in.	=	2.5400 cm
1 ft.	=	0.3048 m
1 acre	=	0.4047 ha
1 ft^2	=	0.09290 m^2
1 ft^3	=	0.02832 m^3
1 ft^2/acre	=	0.2296 m^2/ha
1 ft^3/acre	=	0.06997 m^3/ha
1 mile	=	1.6093 km
1 $mile^2$	=	2.5898 km^2
1 $mile^2$	=	258.9846 ha
1 fbm	=	1 ft. $\times$ 1 ft. $\times$ 1 in.
1 fbm	=	0.0023597 m^3
1 Mfbm	=	1000 foot board measure (fbm)
1 Mfbm	=	2.3597 m^3
1 township	=	6 miles $\times$ 6 miles = 36 $mile^2$
1 township	=	9.6558 km $\times$ 9.6558 km = 93.2345 km^2
1 township	=	9323.45 ha
1 m^3 log	≈	233 board feet lumber (provincial average conversion factor)
1 Mfbm	≈	4.3 m^3 log (provincial average conversion factor)

Alberta

Freedom To Create. Spirit To Achieve.