



Estimating a Nonlinear Mixed Volume-age Model with and without Taking into Account Serially-correlated Errors: Differences and Implications

Shongming Huang, Shawn X. Meng & Yuqing Yang

Biometrics Unit, Forest Management Branch

Alberta Department of Sustainable Resource Development

8th Floor, 9920-108 Street, Edmonton, Alberta T5K 2M4, Canada

Tel: 1-780-422-5281 E-mail: shongming.huang@gov.ab.ca

Abstract

In this study we estimated a lodgepole pine (*Pinus contorta* var. *latifolia* Engelm.) volume-age model with and without taking into account serially-correlated errors arisen from permanent sample plots. The estimations were based on the first-order (FO) and first-order conditional expectation (FOCE) methods of the nonlinear mixed model technique. Among the correlated error structures considered, the spatial power structure was found to be the most appropriate. Model predictions were obtained and evaluated on model fitting data, as well as on independent validation data collected from a different ecoregion. Results showed that the model estimated with the independent and identically distributed (iid) error structure performed much better than the model estimated with the correlated error structure. This is true on both model fitting and validation data sets, and for both FO and FOCE methods. It implies that, if the main purpose of a study is to develop models for predictions, there is no real benefit to consider more elaborate and complex error structures to account for the correlated errors. The iid error structure is a sound choice for dealing with correlated errors under the nonlinear mixed model framework.

Keywords: Longitudinal data, Error structure, Correlation, Model prediction, First-order (FO) method, First-order conditional expectation (FOCE) method, Volume-age model

1. Introduction

For repeatedly measured, unequally spaced and unbalanced longitudinal data collected from permanent sample plots (or stem analysis trees), various within- and between-plot correlations and heterogeneous variances may occur. In addition, the correlations can be both temporal and spatial.

Violation of the independent and identically distributed (iid) error assumption could have important statistical consequences for analyses based on the least squares principle. It may invalidate hypothesis testing and interval estimation. The problem may be solved using the generalized least squares (GLS) method (Judge et al. 1985) or mixed model technique (Davidian and Giltinan 1995). Examples of GLS in forestry have been shown by many (e.g., Monserud 1984, Gregoire 1987). One issue related to GLS is that it generally applies to equally spaced and balanced data. It may create other problems when used for unequally spaced and unbalanced data. Lappi and Bailey (1988) stated that ordinary least squares (OLS) are “deemed to be appropriate” for large samples, even though GLS could be used. Borders et al. (1988), West (1995), and Huang (1997) cited other studies which concluded that OLS is adequate, and it is unnecessary to apply more elaborate procedures when faced with correlated errors. Knowledge of the correlation parameter in GLS is typically ignored in predictions (Monserud 1984).

In more recent years, more researchers have chosen the mixed model approach to deal with correlated errors and heterogeneous variances (e.g., Lappi 1986, Ojansuu 1987, Gregoire et al. 1995, Huang 1997, Fang and Bailey 2001, Schabenberger and Pierce 2002, Wang et al. 2007). This approach allows for a more flexible error covariance structure. More importantly, knowledge of the correlation parameter and the entire error covariance structure from the mixed model estimation can be used in predictions.

In spite of the large amount of effort invested in addressing correlated errors, many questions remain. The primary objectives of this study are to: (1) compare a volume-age model for lodgepole pine (*Pinus contorta* var. *latifolia*

Engelm.) estimated using nonlinear mixed model (NLMM) technique with and without taking into account correlated errors; (2) evaluate the impacts of different error structures on model predictions based on model fitting data as well as on independent model validation data; (3) clarify some methodological confusions and miscues about the NLMM technique in forest modeling; and (4) recommend a sound and practical approach to use when faced with correlated errors.

2. Data and base model

The lodgepole pine data used in this study were collected by the Alberta Forest Service over the last 45 years as a part of the provincial Permanent Sample Plots (PSP) system. They were obtained from pine dominated forests in the upper foothills and lower foothills ecoregions of Alberta. The most common PSP size is 1000 m², but it can range from 200 to 2000 m² depending on the density of a stand. Within each PSP, diameters (always at 1.3 m above ground) were measured for all trees taller than 1.3 m, and heights were measured for all trees with diameters or portions of representative trees (Alberta Forest Service 2000). The missing tree heights were predicted using the “bias-free” height-diameter models (Huang et al. 2008a). Individual tree volumes were calculated following the standard taper functions (Huang 1994), and were summarized to give plot volume. Plot age (at breast height) was obtained from cored trees within the plot or sectioned trees in the buffer areas around the plot.

The volume-age data from the upper foothills ecoregion were used as model fitting data, and the data from the lower foothills ecoregion were used as model validation data. The two data sets (Figure 1) can be considered independent of each other because of the geographic separation. The model fitting data consist of 1,249 measurements from 288 plots (subjects), with a mean (minimum-maximum) volume (m³/ha) of 302.694 (13.909-642.471) and a standard deviation (SD) of 108.902 (m³/ha). The validation data consist of 1,011 measurements from 265 plots, with a mean volume of 287.611 (1.748-564.329) and an SD of 108.581. Both data sets are unequally spaced and unbalanced longitudinal data.

Based on a detailed examination of the overall and subject-specific (SS) volume-age trends exhibited by the data in Figure 1, two NLMMs were selected for evaluation:

$$(1) \quad \text{Vol}_i = (b_1 + u_{1i}) \text{Age}_i^{(b_2 + u_{2i})} \exp[-(b_1 + u_{1i}) \text{Age}_i] + \varepsilon_i$$

$$(2) \quad \text{Vol}_i = (b_1 + u_{1i}) \text{Age}_i^{(b_2 + u_{2i})} \exp[-(b_3 + u_{3i}) \text{Age}_i] + \varepsilon_i$$

where Vol_i and Age_i are observed volumes and ages for the i th plot, b_1 , b_2 and b_3 are fixed parameters common to all plots, u_{1i} , u_{2i} and u_{3i} are random parameters unique to the i th plot, and ε_i is a normally distributed within-plot error term. Both (1) and (2) can be sigmoidal. They can also describe declining volumes at old ages, which were observed in some plots likely due to mortality exceeding growth at old ages (Figure 1). Preliminary analyses indicated that (2) failed to achieve convergence in some cases because of over-parameterization. Therefore, (1) was chosen in this study for further analyses.

3. Error structure and parameter estimation

Model (1) can be written as a generalized NLMM of the form:

$$(3) \quad \mathbf{y}_i = f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i) + \varepsilon_i$$

where $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{in_i}]'$ is a vector of plot volumes for plot i , \mathbf{x}_i is a design matrix of the covariate age, \mathbf{b} is a vector of fixed parameters, $\mathbf{u}_i = [u_{1i}, u_{2i}]'$ is a vector of subject-specific random parameters, and ε_i is a vector of within-plot errors. The \mathbf{u}_i and ε_i are assumed to be uncorrelated and normally distributed with mean zero and variance-covariance matrices \mathbf{D} and \mathbf{R}_i , respectively. The \mathbf{D} is typically assumed to be unstructured and identical for all subjects within the population:

$$(4) \quad \mathbf{D} = \begin{bmatrix} \sigma_{u_1}^2 & \sigma_{u_1 u_2} \\ \sigma_{u_1 u_2} & \sigma_{u_2}^2 \end{bmatrix}$$

where $\sigma_{u_1}^2$, $\sigma_{u_2}^2$, and $\sigma_{u_1 u_2}$ are variances and covariance for random parameters u_1 and u_2 .

As noted earlier, for the unequally spaced and unbalanced longitudinal data collected from the PSPs, the errors may not be iid. Instead, they could be correlated and unequally varied. To account for these, a total of more than 20 error covariance structures described in SAS Institute Inc. (2004) and Littell et al. (2006) were examined. The spatial power error structure produced the most reasonable fit to the data. It was chosen for \mathbf{R}_i . As an example, for a plot repeatedly measured five times, the \mathbf{R}_i is defined as:

$$(5) \quad \mathbf{R}_i = \sigma^2 \Psi = \sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} & \rho^{d_{15}} \\ \rho^{d_{21}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} & \rho^{d_{25}} \\ \rho^{d_{31}} & \rho^{d_{32}} & 1 & \rho^{d_{34}} & \rho^{d_{35}} \\ \rho^{d_{41}} & \rho^{d_{42}} & \rho^{d_{43}} & 1 & \rho^{d_{45}} \\ \rho^{d_{51}} & \rho^{d_{52}} & \rho^{d_{53}} & \rho^{d_{54}} & 1 \end{bmatrix} = \text{SP(POW)} = \text{power}$$

where σ^2 is the overall error variance, Ψ is a generalized error covariance structure, $d_{\eta\eta}$ is the distance between two measurements at times η and η ($d_{\eta\eta} = |t_\eta - t_\eta|$), and ρ is the correlation parameter. The spatial power error structure captured the unequally spaced and unbalanced nature of the volume-age data. Other error structures examined but not presented here, e.g., compound symmetry, heterogeneous Toeplitz, Toeplitz with different bands, unstructured correlation, first-order ante-dependent, and several spatial correlation structures evaluated in Yang and Huang (2008), were not converged or performed poorer than the SP(POW). The time-series structures such as AR(1), MA(2) and ARMA(1,1), which assume that the measurements are taken at equally spaced time intervals, may not be adequate without further adjustments, even though some of them produced seemingly reasonable results.

For models assumed to follow an iid error structure, the R_i is simplified to: $R_i = \sigma^2 \Psi = \sigma^2 I_{n_i}$, where I_{n_i} is a $n_i \times n_i$ identity matrix of 1s.

Different methods have been developed to estimate the generalized NLMM (3). They are described in detail in Davidian and Giltinan (1995), and Vonesh and Chinchilli (1997). Since the estimation of the fixed parameters is readily available in most statistical software packages, we focused on the estimation of subject-specific random parameters based on the first-order (FO) method of Beal and Sheiner (1982) and the first-order conditional expectation (FOCE) method of Lindstrom and Bates (1990).

For the FO method, a first-order Taylor series expansion of (3) around zero, the expected value of the random parameters, is used to approximate (3). The random parameters are predicted by:

$$(6) \quad \hat{\mathbf{u}}_i = \hat{\mathbf{D}}\mathbf{Z}_i'(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i' + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0})]$$

where $\hat{\mathbf{D}}$ and $\hat{\mathbf{R}}_i$ are estimates of \mathbf{D} and \mathbf{R}_i , respectively, $\hat{\mathbf{b}}$ is an estimate of the fixed parameters \mathbf{b} , and \mathbf{Z}_i is the derivative matrix with respect to the random parameters, defined by

$$(7) \quad \mathbf{Z}_i = \left. \frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{0})}{\partial \mathbf{u}_i'} \right|_{\hat{\mathbf{b}}, \mathbf{0}}$$

For the FOCE method, the first-order Taylor series expansion of model (3) is iterated around a current predictor of the random parameters (and a current \mathbf{Z}_i) until the convergence criterion is met. The final random parameters are obtained numerically through the following equation:

$$(8) \quad \hat{\mathbf{u}}_i = \hat{\mathbf{D}}\mathbf{Z}_i'(\mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i' + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i) + \mathbf{Z}_i\hat{\mathbf{u}}_i]$$

where $\hat{\mathbf{D}}$, $\hat{\mathbf{R}}_i$ and $\hat{\mathbf{b}}$ are as defined above, but \mathbf{Z}_i is given by

$$(9) \quad \mathbf{Z}_i = \left. \frac{\partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i)}{\partial \mathbf{u}_i'} \right|_{\hat{\mathbf{b}}, \hat{\mathbf{u}}_i}$$

Since $\hat{\mathbf{u}}_i$ appears on both sides of (8), for the FOCE method, $\hat{\mathbf{u}}_i$ must be solved iteratively. This is the fundamental concept innate to the FOCE method (Schabenberger 1994, Vonesh and Chinchilli 1997, Huang 2008, Temesgen et al. 2008, Meng et al. 2008, Meng and Huang 2009). Unfortunately, many forest modelers used (6) from the FO method to predict $\hat{\mathbf{u}}_i$ while implementing the FOCE method, without realizing that this could seriously bias the outcome. The methodological mixing should be avoided in future studies as it may lead to erroneous conclusions.

All model fittings were carried out using the SAS macro %NLINMIX, with expand = ZERO for the FO method and expand = EBLUP for the FOCE method (Littell et al. 2006). The default fittings assume iid error structure (i.e., $\mathbf{R}_i = \sigma^2 \mathbf{I}_{n_i}$). For the spatial power error structure illustrated in (5), the fittings must be implemented using the keyword TYPE = SP(POW)(age) within the REPEATED statement.

4. Predictions

Once the $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}_i$ are estimated, subject-specific volume predictions for plot i can be made:

$$(10) \quad \hat{\mathbf{y}}_i = f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0}) + \mathbf{Z}_i\hat{\mathbf{u}}_i \quad (\text{FO method})$$

$$(11) \quad \hat{\mathbf{y}}_i = f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i) \quad (\text{FOCE method})$$

where \mathbf{Z}_i for FO is defined in (7), and \mathbf{x}_i is the \mathbf{x} matrix for new or old (modeling) observations. Note the $\hat{\mathbf{y}}_i$ s for FO and FOCE are different (Vonesh and Chinchilli 1997). Mixing them will likely result in biased predictions (Huang 2008, Meng and Huang 2009). It is possible that, following Judge et al. (1985)'s descriptions on prediction, we can also derive the following alternatives to obtain the adjusted predictions from the use of SP(POW) or any other generalized error covariance structure:

$$(10a) \quad \hat{\mathbf{y}}_{0i} = f(\mathbf{x}_{0i}, \hat{\mathbf{b}}, \mathbf{0}) + \mathbf{Z}_i\hat{\mathbf{u}}_i + \hat{\mathbf{V}}'\hat{\Psi}^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0}) - \mathbf{Z}_i\hat{\mathbf{u}}_i] \quad (\text{FO method})$$

$$(11a) \quad \hat{\mathbf{y}}_{0i} = f(\mathbf{x}_{0i}, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i) + \hat{\mathbf{V}}'\hat{\Psi}^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i)] \quad (\text{FOCE method})$$

where \mathbf{x}_{0i} is a known \mathbf{x} matrix of new observations for plot i , $\hat{\mathbf{y}}_{0i}$ corresponds to \mathbf{x}_{0i} , \mathbf{x}_i is an \mathbf{x} matrix from the old (existing) data, $\hat{\mathbf{V}}$ is the estimated covariances between elements of old and new errors, and $\hat{\Psi}$ is the estimated

generalized within-plot error covariance structure. For the SP(POW) structure, for instance, more explicit prediction equations can be derived following Judge et al. (1985) (detailed derivations are omitted here. Interested readers may contact the senior author to obtain a copy):

$$(10b) \quad \hat{\mathbf{y}}_{i,\kappa} = f(\mathbf{x}_{i,\kappa}, \hat{\mathbf{b}}, \mathbf{0}) + \mathbf{Z}_{i,\kappa} \hat{\mathbf{u}}_i + \rho^{d_{l(\kappa+1)}} [\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0}) - \mathbf{Z}_i \hat{\mathbf{u}}_i] \quad (\text{FO method})$$

$$(11b) \quad \hat{\mathbf{y}}_{i,\kappa} = f(\mathbf{x}_{i,\kappa}, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i) + \rho^{d_{l(\kappa+1)}} [\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i)] \quad (\text{FOCE method})$$

where κ is the number of projection intervals (or periods), $(\kappa+1)$ is the number of measurements, $\mathbf{x}_{i,\kappa}$ and $\mathbf{Z}_{i,\kappa}$ are the \mathbf{x}_i and \mathbf{Z}_i values κ intervals ahead, and $d_{l(\kappa+1)}$ is the distance (or time) between the first and $(\kappa+1)$ th measurements (e.g., $d_{l(\kappa+1)} = d_{15}$ for the 4th projection period of the SP(POW) structure given in (5)). For a simpler structure such as AR(1), the above equations reduce to:

$$(10c) \quad \hat{\mathbf{y}}_{i,\kappa} = f(\mathbf{x}_{i,\kappa}, \hat{\mathbf{b}}, \mathbf{0}) + \mathbf{Z}_{i,\kappa} \hat{\mathbf{u}}_i + \rho^\kappa [\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0}) - \mathbf{Z}_i \hat{\mathbf{u}}_i] \quad (\text{FO method})$$

$$(11c) \quad \hat{\mathbf{y}}_{i,\kappa} = f(\mathbf{x}_{i,\kappa}, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i) + \rho^\kappa [\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_i)] \quad (\text{FOCE method})$$

For the reasons discussed at the end of Section 5, however, we opted not to implement the alternative equations in the present study before further assessment and simulations are conducted.

There are different ways to compare the differences between the observed and predicted volumes, and various statistics exist for reporting the goodness-of-fit of NLMMs (Vonesh and Chinchilli 1997). Here, we are mostly interested in the criteria to evaluate the prediction errors (residuals), their average and variations. Thus, the following statistics were computed:

$$(12) \quad \bar{e} = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{n_i} e_{ij} = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_{ij})$$

$$(13) \quad SD = \sqrt{\frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{n_i} (e_{ij} - \bar{e})^2}$$

$$(14) \quad \delta = \bar{e}^2 + SD^2$$

$$(15) \quad \bar{e}\% = \frac{\bar{e}}{\bar{y}} \times 100\%$$

where \bar{e} is the overall mean bias, y_{ij} and \hat{y}_{ij} are the j th observed and predicted values for subject i , m is the number of subjects, N is the total number of observations from m subjects ($N = \sum_{i=1}^m n_i$), SD is standard deviation, δ is an overall accuracy measure combining the bias (\bar{e}) and precision (SD) of the errors (Cochran 1977), $\bar{e}\%$ is percent bias, and \bar{y} is the arithmetic mean of the observed values.

Because the predominant interest of a NLMM is in subject-specific predictions, it is more important that we also compute the above statistics by subject:

$$(16) \quad \bar{e}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} e_{ij} = \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_{ij})$$

$$(17) \quad SD_i = \sqrt{\frac{1}{n_i} \sum_{j=1}^{n_i} (e_{ij} - \bar{e}_i)^2}$$

$$(18) \quad \delta_i = \bar{e}_i^2 + SD_i^2$$

$$(19) \quad \bar{e}_i\% = \frac{\bar{e}_i}{\bar{y}_i} \times 100\%$$

where all variables are as defined before, except that they apply to subject i , not the entire population.

For a population of m subjects, the frequency distribution for any one of the SS measures expressed in (16)-(19) can be constructed, and the percentage of subjects that fall below or above certain thresholds such as the 5th or the 95th percentile, can be determined. As will be shown later, examination of such frequency distributions is exceedingly useful in helping modelers obtaining a realistic understanding of the goodness-of-fit a NLMM when it is used for subject-specific predictions. More importantly, it also helps revealing the differences between the models estimated with different error structures.

It is worthwhile to note here that many forest modelers have used the fixed parameters estimated as a part of a NLMM to predict the population-averaged (PA) responses, i.e., by setting $\hat{\mathbf{u}}_i = \mathbf{0}$ and using $\hat{\mathbf{y}}_i = f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0})$. This is likely a result of some misunderstanding. The fixed parameters estimated as a part of a NLMM do not fully characterize the mean responses of the population, especially when the inter-subject variability is large (Davidian and Giltinan 2003, Fitzmaurice et al. 2004). Consequently, they generally provide a biased partial representation of the true population mean responses. It is very important to recognize that the predictions obtained from the fixed parameters estimated as a

part of a NLMM generally do not represent the “unbiased” PA responses (Huang 2008).

5. Results and discussion

Results of the estimated parameters for model (1) with iid and spatial power error structures are shown in Table 1. The Akaike's information criterion (AIC) is computed by (Littell et al. 2006):

$$(20) \text{ AIC} = -2\text{LL} + 2p$$

where -2LL is -2 times the log-likelihood function, and p is the total number of estimated parameters.

Studentized residual plots were obtained from the FO and FOCE methods using the iid and correlated (spatial power) error structures. Only those from the FOCE method are shown here (Figure 2), as those from the FO method display similar patterns. The differences in the studentized residual plots between the iid errors and the correlated errors are easy to detect. The studentized residuals from the iid errors are scattered evenly around the zero line, whereas the studentized residuals from the correlated errors are not. This suggests that judged by visual means, the iid error structure is more satisfactory.

To facilitate the discussion of the upcoming results, and to help readers grasp the essences of the issues, we first demonstrate the computations involved in predicting the random parameters and in deriving the predicted volumes for a “new” plot measured five times ($j = 5$) from the validation data (Table 2a), based on the FOCE method with the SP(POW) error structure. Actual computation algorithms for this demonstration are provided in Appendix 1, and were explained in detail in Huang (2008), Meng et al. (2008), and Meng and Huang (2009). The computations for the iid errors and for the FO method are relatively simpler and can be derived directly from the algorithms given in Appendix 1 (Huang 2008).

For the FOCE method, the derivatives of model (1) with respect to the two random parameters are:

$$(21) \text{ der_}u_1 = \partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i) / \partial u_1 = \mathbf{Age}_i^{(\hat{b}_2 + \hat{u}_2)} (1 - (\hat{b}_1 + \hat{u}_1) \mathbf{Age}_i) \exp(-(\hat{b}_1 + \hat{u}_1) \mathbf{Age}_i)$$

$$(22) \text{ der_}u_2 = \partial f(\mathbf{x}_i, \mathbf{b}, \mathbf{u}_i) / \partial u_2 = (\hat{b}_1 + \hat{u}_1) \exp(-(\hat{b}_1 + \hat{u}_1) \mathbf{Age}_i) \mathbf{Age}_i^{(\hat{b}_2 + \hat{u}_2)} \ln(\mathbf{Age}_i)$$

Therefore, for the i th subject with j measurements, the estimated \mathbf{Z}_i matrix is constructed as follows:

$$(23) \mathbf{Z}_i = \begin{bmatrix} \frac{\partial f(x_{i1}, \mathbf{b}, \mathbf{u}_i)}{\partial u_1} & \frac{\partial f(x_{ij}, \mathbf{b}, \mathbf{u}_i)}{\partial u_2} \\ \vdots & \vdots \\ \frac{\partial f(x_{ij}, \mathbf{b}, \mathbf{u}_i)}{\partial u_1} & \frac{\partial f(x_{ij}, \mathbf{b}, \mathbf{u}_i)}{\partial u_2} \end{bmatrix} = \begin{bmatrix} \text{der_}u_{11} & \text{der_}u_{21} \\ \vdots & \vdots \\ \text{der_}u_{1j} & \text{der_}u_{2j} \end{bmatrix}$$

where $\text{der_}u_{11}, \dots, \text{der_}u_{1j}$ are given by (21) and $\text{der_}u_{21}, \dots, \text{der_}u_{2j}$ are given by (22), and $j = 5$. The estimated \mathbf{D} matrix for any subject i is (from eq. (4) and Table 1):

$$(24) \hat{\mathbf{D}} = \begin{bmatrix} \sigma_{u_1}^2 & \sigma_{u_1 u_2}^2 \\ \sigma_{u_1 u_2}^2 & \sigma_{u_2}^2 \end{bmatrix} = \begin{bmatrix} 0.00030 & 0.000385 \\ 0.000385 & 0.007794 \end{bmatrix}$$

The estimated \mathbf{R}_i matrix for the example subject is (from eq. (5) and Tables 1 and 2a):

$$(25) \hat{\mathbf{R}}_i = 6361.73 \begin{bmatrix} 1 & 0.9977^7 & 0.9977^{21} & 0.9977^{31} & 0.9977^{41} \\ 0.9977^7 & 1 & 0.9977^{14} & 0.9977^{24} & 0.9977^{34} \\ 0.9977^{21} & 0.9977^{14} & 1 & 0.9977^{10} & 0.9977^{20} \\ 0.9977^{31} & 0.9977^{24} & 0.9977^{10} & 1 & 0.9977^{10} \\ 0.9977^{41} & 0.9977^{34} & 0.9977^{20} & 0.9977^{10} & 1 \end{bmatrix}$$

For the FOCE method, there is no simple algebraic solution for $\hat{\mathbf{u}}_i$ from (8). An iterative procedure involving the following computation steps is used to numerically solve for $\hat{\mathbf{u}}_i$:

Step 1: Obtain an initial estimate, termed $\hat{\mathbf{u}}_{i,0}$, of the random parameters. A reasonable first guess of this estimate is $\hat{\mathbf{u}}_{i,0} = \mathbf{E}(\mathbf{u}_i) = \mathbf{0}$, the expected value of the random parameters for the FO method. Thus, $\hat{\mathbf{u}}_{i,0}$ can be computed directly using eq. (6) for the FO method, as follows:

$$(26) \hat{\mathbf{u}}_{i,0} = \hat{\mathbf{D}} \mathbf{Z}_{i,0}' (\mathbf{Z}_{i,0}' \hat{\mathbf{D}} \mathbf{Z}_{i,0} + \hat{\mathbf{R}}_i)^{-1} [\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \mathbf{0})]$$

where $\mathbf{Z}_{i,0}$ is a first estimate of the \mathbf{Z}_i matrix (equivalent to the \mathbf{Z}_i in (7) for the FO method):

$$\mathbf{Z}_{i,0} = \begin{bmatrix} \frac{\partial f(\mathbf{x}_{i1}, \mathbf{b}, \mathbf{0})}{\partial u_1} & \frac{\partial f(\mathbf{x}_{ij}, \mathbf{b}, \mathbf{0})}{\partial u_2} \\ \vdots & \vdots \\ \frac{\partial f(\mathbf{x}_{ij}, \mathbf{b}, \mathbf{0})}{\partial u_1} & \frac{\partial f(\mathbf{x}_{ij}, \mathbf{b}, \mathbf{0})}{\partial u_2} \end{bmatrix} = \begin{bmatrix} Age_{i1}^{\hat{b}_1} (1 - \hat{b}_1 Age_{i1}) \exp(-\hat{b}_1 Age_{i1}) & \hat{b}_1 \exp(-\hat{b}_1 Age_{i1}) Age_{i1}^{\hat{b}_1} \ln(Age_{i1}) \\ \vdots & \vdots \\ Age_{ij}^{\hat{b}_j} (1 - \hat{b}_j Age_{ij}) \exp(-\hat{b}_j Age_{ij}) & \hat{b}_j \exp(-\hat{b}_j Age_{ij}) Age_{ij}^{\hat{b}_j} \ln(Age_{ij}) \end{bmatrix}$$

Step 2: Once the $\hat{\mathbf{u}}_{i,0}$ is calculated, the next estimation of $\hat{\mathbf{u}}_i$, termed $\hat{\mathbf{u}}_{i,1}$, is obtained based on (8), with $\hat{\mathbf{u}}_i$ on the right-hand side replaced by $\hat{\mathbf{u}}_{i,0}$ from Step 1:

$$(27) \quad \hat{\mathbf{u}}_{i,1} = \hat{\mathbf{D}}\mathbf{Z}'_{i,1}(\mathbf{Z}_{i,1}\hat{\mathbf{D}}\mathbf{Z}'_{i,1} + \hat{\mathbf{R}}_i)^{-1}[\mathbf{y}_i - f(\mathbf{x}_i, \hat{\mathbf{b}}, \hat{\mathbf{u}}_{i,0}) + \mathbf{Z}_{i,1}\hat{\mathbf{u}}_{i,0}]$$

where $\mathbf{Z}_{i,1}$ is the \mathbf{Z}_i expressed in (23) evaluated at $\hat{\mathbf{u}}_{i,0}$, with its elements given by:

$$(28) \quad der_u_1 = \partial f(\mathbf{x}_i, \mathbf{b}, \hat{\mathbf{u}}_{i,0}) / \partial u_1 = \mathbf{Age}_i^{(\hat{b}_2 + \hat{u}_{2,0})} (1 - (\hat{b}_1 + \hat{u}_{1,0}) \mathbf{Age}_i) \exp(-(\hat{b}_1 + \hat{u}_{1,0}) \mathbf{Age}_i)$$

$$(29) \quad der_u_2 = \partial f(\mathbf{x}_i, \mathbf{b}, \hat{\mathbf{u}}_{i,0}) / \partial u_2 = (\hat{b}_1 + \hat{u}_{1,0}) \exp(-(\hat{b}_1 + \hat{u}_{1,0}) \mathbf{Age}_i) \mathbf{Age}_i^{(\hat{b}_2 + \hat{u}_{2,0})} \ln(\mathbf{Age}_i)$$

Step 3: Having the calculated $\hat{\mathbf{u}}_{i,1}$ from (27), the new estimation for $\hat{\mathbf{u}}_i$, termed $\hat{\mathbf{u}}_{i,2}$, is computed using (8) again, with the $\hat{\mathbf{u}}_i$ on the right-hand side replaced by the updated $\hat{\mathbf{u}}_{i,1}$ from Step 2:

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

where $\mathbf{Z}_{i,2}$ is evaluated at $\hat{\mathbf{u}}_{i,1}$, with its elements defined in eqs. (28)-(29) updated using $\hat{\mathbf{u}}_{i,1}$.

This process is iterated k times until a desired precision, e.g., $|\hat{\mathbf{u}}_{i,k} - \hat{\mathbf{u}}_{i,(k-1)}| < 0.0000001$, is achieved. The predicted final random parameters for subject i is: $\hat{\mathbf{u}}_i = \hat{\mathbf{u}}_{i,k}$.

For the example plot shown in Table 2a, the predicted final random parameters for the FOCE-pow are: $u_1 = -0.00338906$ and $u_2 = -0.059555$. The final elements (der_u_1 and der_u_2) of the \mathbf{Z}_i matrix are listed in Table 2a, along with the \mathbf{Z}_i matrices for the other error structure and the FO method.

Once the $\hat{\mathbf{u}}_i$ are known, the predicted volumes for the FOCE method are obtained directly using (11). Results are shown in Table 2a. For the FO method, the predicted volumes are obtained differently using (10). Results are also shown in Table 2a. Once the prediction errors and their SD are known, various intervals can be constructed (Hahn and Meeker 1991). Some common ones are:

$$(31) \quad \bar{e}_i \pm t_{(1-\alpha/2), (n_i-1)} SD \sqrt{1 + \frac{1}{n_i}} \quad (\text{prediction interval for a future observation})$$

$$(32) \quad \bar{e}_i \pm z_{\frac{1+\gamma}{2}} \left(1 + \frac{1}{2n_i} \right) \sqrt{\frac{n_i-1}{\chi^2_{\alpha, (n_i-1)}}} \quad (\text{tolerance interval to contain at least proportion } \gamma \text{ of the subject})$$

$$(33) \quad \bar{e}_i \pm t_{(1-\alpha/2), (n_i-1)} \frac{SD}{\sqrt{n_i}} \quad (\text{confidence interval for the subject mean})$$

$$(34) \quad SD \sqrt{\frac{n_i-1}{\chi^2_{(1-\alpha/2), (n_i-1)}}} \rightarrow SD \sqrt{\frac{n_i-1}{\chi^2_{(\alpha/2), (n_i-1)}}} \quad (\text{confidence interval for the subject standard deviation})$$

where z is the standard normal distribution (and $\alpha = 0.05$ throughout this study).

Table 2b lists the normality test results and the calculated intervals for the example plot. All four tests conducted following Yang et al. (2004) showed that normality was met. However, it must be noted that since the sample size for this example plot is small, the interval calculations serve only as illustrations. The “preferred” calculations in fact require that the sample size should not be smaller than 8 (Hahn and Meeker 1991). In the cases where the normality assumption is not met, distribution-free intervals could be constructed (Hahn and Meeker 1991). One could also use a 10% trimmed mean as an estimate of \bar{e}_i and a jackknifed standard deviation expressed in (35) as an estimate of SD to construct approximated intervals (Efron and Tibshirani 1986, Hahn and Meeker 1991):

$$(35) \quad SD_{jack} = \sqrt{\frac{n_i-1}{n_i} \sum_{j=1}^{n_i} (\bar{e}_{i(j)} - \bar{e}_i)^2}$$

where $\bar{e}_{i(j)}$ is the sample average of the data set without the j th point. It is straightforward to infer that (31)-(35) can also be applied to the entire population of N observations ($N = \sum_{i=1}^m n_i$).

A fundamental difference between the FOCE and FO methods is that, the random parameters for the FOCE method must be solved numerically, whereas they can be obtained directly from (6) for the FO method. This fundamental difference appeared to have largely been neglected in previous applications of NLMs in forestry. It is essential to recognize that when implementing the FOCE method, the $\hat{\mathbf{u}}_i$ calculated by (6) is just a first initial “guesstimate” (and sometimes an inappropriate guesstimate) in the search of a true empirical best linear unbiased predictor of \mathbf{u}_i .

Table 3 lists the overall prediction errors (e_{ij}) on both model fitting and validation data sets. For both FO and FOCE methods, the prediction errors (absolute values) and their variations from the correlated error structure are much larger than their counterparts from the iid error structure. For example, for the FO method on the model fitting data, the δ and $\bar{e}\%$ values for the iid error structure are 107.965 and 0.226%, respectively, whereas for the spatial power error structure, they are 4696.876 and 10.906%, respectively. The differences between the values from the iid and correlated error structures are striking. They are also consistent across the methods and data sets. All these suggest that the iid error structure produced much more accurate and precise predictions than the correlated error structure.

For the more important SS predictions, the statistics calculated according to (16)-(19) by different error structures and data sets are summarized in Table 4. The number of plots where the percent bias ($\bar{e}_i\%$) exceeded $\pm 2.5\%$ of the observed mean were identified and listed separately in Table 4. The frequency distributions of the mean biases, percent biases and standard deviations from all 288 plots of the model fitting data and 265 plots of the model validation data were obtained. For the sake of brevity, only the frequency distributions for $\bar{e}_i\%$ from different error structures and data sets are shown (Figure 3).

Judging from the results shown in Table 4, the contrasts between the iid and power error structures are remarkable. The power error structure produced much larger biases, SDs and δ s. More specifically, for instance, for the FO method, among the 288 plots of the modeling data, only two plots produced biases exceeding $\pm 2.5\%$ of the observed means when the iid error structure was used. But when the power error structure was used, 272 plots produced biases exceeding $\pm 2.5\%$ of the observed means. Similar results are also apparent on the validation data, and for the FOCE method. The frequencies of poorer predictions from the power error structure are much greater than those from the iid error structure.

The poorer performance of the power error structure is more readily seen in Figure 3, where the percent biases from different plots are centered around zero but those from the power error structure are spread out much more than those from the iid error structure, indicating lower accuracies and larger variations from the power error structure. This is true on both model fitting and validation data sets, and for both FO and FOCE methods.

Figure 4 shows the observed volumes against the volumes predicted from different methods with iid and power error structures. We only show the plots from the validation data because the plots from the modeling data are similar. More importantly, it is more telling on validation data than on modeling data because the validation data are “new” independent data that a fitted model is most likely to be applied to make predictions in real-world applications. The poorer performance of the power error structure is rather apparent in Figure 4. It is consistent for both FO and FOCE methods.

An examination of the “spaghetti” or “chow-mein” plots (Figure 5) of the prediction trajectories of the validation data indicates that the prediction trajectories from the iid error structure closely mimic the observed trajectories shown in Figure 1(b), whereas the prediction trajectories from the power error structure “compress” the observed trajectories. This is again consistent for both FO and FOCE, even though the degree of the compression varies slightly.

Since much of the data collected for forest modeling is unequally spaced and unbalanced longitudinal data, where repeated measurements are taken on the same experimental units, the lack of independence of such data is inborn. The correlated nature of the data could violate an important assumption required by certain statistical techniques, mostly notably the least squares principle, for “optimal” performance and inference. Forest modelers have studied for many years, and have used the old and new techniques brought to light by statisticians, to address the dependence problem, believing in many cases, that the successful removal of the correlation would result in a “better” model.

However, contrary to the common belief, the results obtained in this study from the NLMM methods showed that although accounting for the serial correlation appeared to have produced better fits when judged by the AIC values in Table 1 (a smaller AIC value is said to mean a better fit), it produced much worse predictions on both the model fitting and validation data sets. The predictions obtained with the iid error structure had consistently much smaller and better distributed biases than those obtained with the correlated (spatial power) error structure. This created a dilemma, as the successful accounting of the correlated errors produced much poorer predictions, while ignoring the correlated errors produced much better predictions but could potentially invalid hypothesis testing and interval estimation. It also asked for the question of whether the common practice of using AIC (and its modified forms) to select a preferred model and/or a preferred error covariance structure is correct.

While there is no “best” answer to the apparent contradiction, and opinions may understandably vary depending on the focus of a study or a particular researcher, we believe in general that, because the vast majority of forestry models are developed to be used as predictive tools once the parameters have been estimated, and because much of the practical, real-world emphasis of forest modeling has been on the predictive capabilities and biases from the applications of the models, the prediction results, preferably obtained on independent validation data, should be given the predominant consideration. They should be used as the deciding factor in determining the appropriateness of a fitted model, and in

comparing alternative models and/or error structures.

We also believe that, there may be some debatable issues and misconceptions about hypothesis testing and interval estimation in forest modeling. We understand that, in the presence of correlated (and/or heteroscedastic) errors, the ordinary estimators of coefficients ignoring the correlation are still unbiased. However, the standard errors of the coefficients are biased and inconsistent, and the estimators are not efficient (Judge et al. 1985). When a correct error covariance structure has been identified and a consistent estimator of this error covariance structure is obtained, it can be used to obtain a consistent estimator of the standard errors of the estimated coefficients that are also efficient. The question though is that, for a forest model developed as predictive tools, model users are (arguably) not really concerned much about the hypothesis testing and interval estimation on the coefficients that have been estimated in an unbiased manner. Instead, they are generally more concerned about the hypothesis testing and interval estimation on the predictions illustrated in eqs. (31)-(35) and Table 2b, which have little to do with the standard errors of the coefficients obtained on the modeling data. To ward off some potential statistical traps, forest modelers should be aware of and be mindful of some common miscues when conducting hypothesis testing and interval estimation under normal or non-normal error assumption, on different data sets (i.e., modeling vs validation/application), at different levels (i.e., population average vs SS), and for different purposes (e.g., model coefficients vs model predictions).

For the vast majority of forest models built on repeated measurement data and to be used as predictive tools on new data or data ranges beyond those used in modeling, one needs to understand the pertinent prediction scenario first before making any prediction and judging the goodness of the prediction. As illustrated in Figure 6, various prediction scenarios exist, and different evaluation measures can be used to judge the goodness of predictions. We examined the scenarios and conducted simulations based on the fitted volume-age model, and found that accurate local predictions could be achieved from 2, 3 or more prior observations. We also found that, under the NLMM framework, with or without adjusting the predictions from the use of SP(POW) produced varied outcome, dependent on the 'best' covariance structure chosen, the direction of the predictions, the length and number of the prediction intervals, the level on which the predictions are made, the estimation technique (FO vs FOCE) used, the evaluation measures selected, and more importantly, the number of prior observations available for predicting the SS random parameters (we used all available prior observations in our analyses). In general, however, when SS random parameters are obtained from 2 or more prior observations, adjusting the predictions degenerated the predictions for our data. While the results from our study are obvious, we recognize that this is just one such study to demonstrate that the iid error structure is a sound choice for dealing with correlated data under the NLMM framework, if prediction is the primary focus of the study. We realize the potential options and variations in making predictions (as illustrated in Figure 6), and plan to conduct additional studies to further test this phenomenon on larger data sets, and for other tree species and other types of models.

6. Conclusions

Based on the FO and FOCE methods, we evaluated the volume-age model for lodgepole pine estimated with and without taking into account the serially-correlated errors. We found that the model estimated with the iid error structure outperformed the model estimated with the correlated (spatial power) error structure by a large margin. This observation was consistent for both FO and FOCE methods, and on both model fitting and validation data sets. It means that a NLMM estimated with the iid error structure is better in predictions than the model estimated with the correlated error structure. The result of this study can have some important practical implications, as a better model can be estimated using a simpler analysis without the need to account for the correlated error structure. In fact, accounting for the correlated error structure within the NLMM context not only did not improve model predictions, it actually degenerated the predictions substantially. Unless the main objective of modeling is not to develop models for prediction purposes, we believe that there is little or no concrete benefit to consider more elaborate error structures that account for the correlated errors. The iid error structure is simply a better choice for the lodgepole pine model, and for a similar model for black spruce (*Picea mariana*) (Huang et al. 2008b). Future studies should be conducted to see if the conclusions reached in this study still hold for other types of models and species, particularly on independent validation data sets.

Acknowledgements

This research was supported by the Alberta Government and Forest Resource Improvement Association of Alberta (the FRIAA-GYPSY project). Special thanks to the 10 project partners led by West Fraser Mills, Weyerhaeuser Canada and Canadian Forest Products, and to Mr. R. Briand, Dr. W.R. Dempster and Mr. D.J. Morgan for their help and contributions to this project.

References

Alberta Forest Service. (2000). Permanent sample plots field procedures manual. Forest Management Branch, Alberta Sustainable Resource Development, Edmonton, Alberta, Pub. FMOPC 83-03, 106 p.

- Beal, S. L., & Sheiner, L.B. (1982). Estimating population kinetics. *Critical Reviews in Biomedical Engineering* 8: 195-222.
- Borders, B.E., Bailey, R.L., & Clutter, M.L. (1988). Forest growth models: parameter estimation using real growth series. USDA For. Serv. Gen. Tech. Rep. NC-120. pp. 660-667.
- Cochran, W.G. (1977). *Sampling techniques* (3rd ed.). Wiley, New York, 428 p.
- Davidian, M., & Giltinan, D.M. (1995). *Nonlinear models for repeated measurement data*. Chapman & Hall, New York, 360 p.
- Davidian, M., & Giltinan, D.M. (2003). Nonlinear models for repeated measurement data: an overview and update. *J. Agricultural, Biological, and Environmental Statistics* 8: 387-419.
- Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* 1: 54-77.
- Fang, Z., & Bailey, R.L. (2001). Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. *For. Sci.* 47: 287-300.
- Fitzmaurice, G.M., Laird, N.M., & Ware, J.H. (2004). *Applied longitudinal analysis*. Wiley, New York, 506 p.
- Gregoire, T.G. (1987). Generalized error structure for forestry yield models. *For. Sci.* 33: 423-444.
- Gregoire, T.G., Schabenberger, O., & Barrett, J. (1995). Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Can. J. For. Res.* 25: 137-156.
- Hahn, G.J., & Meeker, W.Q. (1991). *Statistical intervals: a guide for practitioners*, John Wiley & Sons, New York. 392 p.
- Huang, S. (1994). Individual tree volume estimation procedures for Alberta: methods of formulation and statistical foundations. Alberta Environmental Protection Tech. Rep. Pub. No. T/288. 80 p.
- Huang, S. (1997). Development of compatible height and site index models for young and mature stands within an ecosystem-based management framework. In *Empirical and process-based models for forest tree and stand growth simulation*. Lisbon, Portugal, pp. 61-98.
- Huang, S. (2008). A generalized procedure for bias-free predictions at population and local levels. A discussion paper, Pub-Feb-22-2008, Forest Management Branch, ASRD, Edmonton, Alberta, 61 p.
- Huang, S., Meng, S.X., & Yang, Y. (2008a). Selected GYPSY components released for evaluation. Pub-Oct-23-2008, Forest Management Branch, ASRD, Edmonton, Alberta, 43 p.
- Huang, S., Meng, S.X., & Yang, Y. (2008b). Selected GYPSY components released for evaluation: total volume models. Pub-Dec-19-2008, Forest Management Branch, ASRD, Edmonton, Alberta, 18 p.
- Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H., & Lee, T.C. (1985). *The theory and practice of econometrics*. 2nd ed., John Wiley & Sons, New York. 1056 p.
- Lappi, J. (1986). Mixed linear models for analyzing and predicting stem form variation of Scots pine. *Communications Inst. Forestalis Fennicae* 134. 69 p.
- Lappi, J., & Bailey, R.L. (1988). A height prediction model with random stand and tree parameters: an alternative to traditional site index methods. *For. Sci.* 34: 907-927.
- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., & Schabenberger, O. (2006). *SAS system for mixed models* (2nd ed.), SAS Institute Inc., Cary, NC, 814 p.
- Lindstrom, M.J., & Bates, D.M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* 46: 673-687.
- Meng, X.S., Huang, S., Lieffers, V.J., Nunifu, T., & Yang, Y. (2008). Wind speed and crown class influence the height-diameter relationship of lodgepole pine: nonlinear mixed effects modeling. *For. Ecol. Manage.* 256: 570-577.
- Meng, X.S., & Huang, S. (2009). Improved calibration of nonlinear mixed-effects models demonstrated on a height growth function. *For. Sci.* (in revision).
- Monserud, R.A. (1984). Height growth and site index curves for inland Douglas-fir based on stem analysis and forest habitat type. *For. Sci.* 30: 943-965.
- Ojansuu, R. (1987). Mixed linear models for stem size and form development. USDA For. Serv. Gen. Tech. Rep. NC-120. pp. 724-730.
- SAS Institute Inc. (2004). *SAS/STAT 9.1 User's Guide*. SAS Institute Inc., Cary, NC. 5136 p.

- Schabenberger, O. (1994). Nonlinear mixed effects growth models for repeated measures in ecology. In *Proc. of ASA, Section on Stats. & Environment*, Toronto, Canada, pp.156-161.
- Schabenberger, O., & Pierce, F.J. (2002). *Contemporary statistical models for the plant and soil sciences*. CRC Press, Boca Raton, FL, 760 p.
- Temesgen, H., Monleon, V.J., & Hann, D.W. (2008). Analysis and comparison of nonlinear tree height prediction strategies for Douglas-fir forests. *Can. J. For. Res.* 38: 553–565.
- Vonesh, E.F., & Chinchilli, V.M. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. Marcel Dekker, New York, 560 p.
- Wang, Y., Lemay, V.M., & Baker, T.G. (2007). Modeling and prediction of dominant height and site index of eucalyptus *globulus* plantations using a nonlinear mixed-effects models approach. *Can. J. For. Res.* 37: 1390-1403.
- West, P.W. (1995). Application of regression analysis to inventory data with measurements on successive occasions. *For. Ecol. & Manage.* 71 (1995): 227-234.
- Yang, Y., & Huang, S. (2008). Modeling percent stocking changes for lodgepole pine stands in Alberta. *Can. J. For. Res.* 38: 1042-1052.
- Yang, Y., Monserud, R.A., & Huang, S. (2004). An evaluation of diagnostic tests and their roles in validating forest biometric models. *Can. J. For. Res.* 34: 619-629.

Appendix 1. A generalized SAS program for predicting random parameters based on the FOCE method with the serially-correlated (spatial power) error structure

-This program applies to unequally spaced and unbalanced longitudinal data for any number of plots.

-The spatial power error structure is illustrated in (5) for a plot (subject) measured five times.

*Input the data, where variable 'measure' refers to the measurement number;

```
data tt1;
```

```
input plotid measure age vol;
```

```
cards;
```

```
38001    1    53.234    155.057
```

```
38001    2    60.234    163.343
```

```
38001    3    74.234    226.229
```

```
38001    4    84.234    268.278
```

```
38001    5    94.234    290.151
```

```
more datalines...;
```

```
run;
```

```
proc sort data=tt1; by plotid;
```

```
run;
```

```
data wed3;
```

```
set tt1;
```

```
by plotid;
```

```
j+1;
```

```
if first.plotid then do; i+1; j=1; end;
```

```
run;
```

```
proc means data=wed3; var i j;
```

```
run;
```

*Input the estimated parameters (from Table 1), and execute the proc iml;

```
proc iml;
```

```

use wed3;
read all var {j} into tobs; read all var {i age} into age; read all var {vol} into vol;
covar={0.000030, 0.000385, 0.007794, 0.9977, 6361.73};
fixp={0.02396 2.5972};
d=j(2,2,.);
d[1,1]=covar[1]; d[1,2]=covar[2]; d[2,1]=covar[2]; d[2,2]=covar[3];
s=covar[5]; cor=covar[4]; bb=fixp[1,]; b=bb`;
tn=max(age[,1]);/*number of plots*/
q=2; /*number of parameters*/
bx={1 1}; /*number of random effect parameters*/
nn=nrow(vol); u=j(tn,q,0); mc=max(tobs);
start sm1 (tn,age,s,nn,cor,cov,mc);
cov=j(nn,mc,0); start=j(nn,1,.);
do i=1 to tn;
c=0; start[,1]=.;
do j=1 to nn; if age[j,1]=i then;
do;
start[j]=j; c=c+1;
end; end;
st=min(start);
do h=1 to c;
do k=1 to c;
a=st+h-1; b=st+k-1;
cv=s*cor*(abs(age[a,2]-age[b,2]));
cov[a,k]=cv;
end; end;
end;
finish sm1;
run sm1 (tn,age,s,nn,cor,cov,mc);
start sm (tn,bx,q,u,z,b,s,d,age,vol,nn,res,uv,cov);
z=j(nn,q,0); res=j(nn,1,0); uv=j(tn,q,0); xx=1;
do k=1 to tn;
z[1:nn]=.; 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)+zb1*u1+zb2*u2;
res[j]=re;
end; end;
end; end;

```

```

r1=z;
r2=r1[loc(r1[,1]^=.),]#bx;
w1=res; w2=w1[loc(w1[,1]^=.),];
mm=nrow(w2); yy=xx+mm-1;
rr=cov[xx:yy,1:mm]; xx=mm+xx;
uu=d*r2`*INV(r2*d*r2`+rr)*w2; uk=uu`;
uv[k,]=uk;
end;
finish sm;

diff=j(tn,q,0); diff1=1; diff2=1;
eps=1E-7;
do iter=1 to 200 until (diff1<eps & diff2<eps);
run sm (tn,bx,q,u,z,b,s,d,age,vol,nn,res,uv,cov);
diff=abs(uv-u);
diff1=max(diff[,1]);
diff2=max(diff[,2]);
u=uv;
end;

*u is the predicted random parameters;
ubu=u;
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={ub1i,ub2i,b1,b2};
create rpm from ub[colname=cnm];
append from ub;
quit;

*predicting the volumes for all plots;
data rpm1;
set rpm;
i=_n_;
run;
proc sort data=wed3;by i;run;
proc sort data=rpm1;by i;run;

*y_pred and y_res are predicted volume and residual from the FOCE method with spatial power errors;
data allx ;
merge wed3 rpm1;
by i ;

```

```

y_pred = (b1+ub1i)*age**(b2+ub2i)*exp(-(b1+ub1i)*age);
y_res= vol - y_pred;
run;

```

Table 1. Parameter estimates and fit statistics for model (1) based on the first-order (FO) and first-order conditional expectation (FOCE) methods assuming iid or correlated (spatial power) error structures.

Method	b_1	b_2	$\sigma_{u_1}^2$	$\sigma_{u_1 u_2}$	$\sigma_{u_2}^2$	ρ	σ^2	AIC
FO-iid	0.02758	2.6258	0.000149	0.001766	0.02693		188.62	12448.0
FO-pow	0.02329	2.5799	0.000028	0.000364	0.008311	0.9980	6952.33	11890.9
FOCE-iid	0.02691	2.6261	0.000098	0.001176	0.01947		171.97	12367.2
FOCE-pow	0.02396	2.5972	0.000030	0.000385	0.007794	0.9977	6361.73	11913.6

Note: iid refers to independent and identically distributed error structure, power refers to spatial power error structure, b_1 and b_2 are fixed parameters, $\sigma_{u_1}^2$ and $\sigma_{u_2}^2$ are variances for random parameters u_1 and u_2 , respectively, and $\sigma_{u_1 u_2}$ is the covariance between u_1 and u_2 , ρ is correlation parameter, σ^2 is residual variance, and AIC is Akaike's information criterion (defined in eq. (20)).

Table 2a. Example computations of prediction errors for a plot (subject) measured five times.

Method	Age	Vol_o	der_u ₁	der_u ₂	Vol_p	Res	\bar{e}_i	SD
FO-iid	53.234	155.057	-3676.41	860.79	144.299	10.759	0.021	8.434
	60.234	163.343	-5921.11	1012.11	174.680	-11.337		
	74.234	226.229	-11034.88	1251.58	230.671	-4.442		
	84.234	268.278	-14744.17	1362.55	263.790	4.488		
	94.234	290.151	-18154.98	1423.50	289.511	0.640		
FO-pow	53.234	155.057	-1973.51	761.18	155.821	-0.764	-7.351	9.475
	60.234	163.343	-3872.25	917.03	185.356	-22.013		
	74.234	226.229	-8668.21	1192.64	238.085	-11.856		
	84.234	268.278	-12553.09	1347.42	268.531	-0.253		
	94.234	290.151	-16498.20	1461.78	292.020	-1.869		
FOCE-iid	53.234	155.057	302.23	579.59	145.821	9.236	-0.144	8.116
	60.234	163.343	-862.75	717.57	175.093	-11.750		
	74.234	226.229	-4343.18	988.36	229.466	-3.237		
	84.234	268.278	-7603.33	1165.34	262.842	5.436		
	94.234	290.151	-11310.55	1320.80	290.555	-0.404		
FOCE-pow	53.234	155.057	-763.71	656.79	165.244	-10.187	-18.020	9.564
	60.234	163.343	-2275.16	802.30	195.767	-32.424		
	74.234	226.229	-6391.37	1074.44	249.451	-23.222		
	84.234	268.278	-9968.73	1240.74	279.850	-11.572		
	94.234	290.151	-13816.33	1376.67	302.845	-12.694		

Note: the methods are defined in Table 1, Vol_o and Vol_p are observed and predicted volumes (m³/ha), der_u₁ and der_u₂ are derivatives with respect to random parameters u_1 and u_2 , respectively, Res is residual, \bar{e}_i is the arithmetic mean of the residuals, and SD is the standard deviation of the residuals. The algorithms given in Appendix 1 show the computations for the FOCE-pow method.

Table 2b. Lower and upper limits of confidence intervals for the example plot shown in Table 2a.

Md.	Normality test				Prediction		Tolerance		CL for \bar{e}_i		CL for SD	
	W	KS	W ²	A ²	lower	upper	lower	upper	lower	upper	lower	upper
1	0.997	0.129	0.018	0.135	-25.63	25.67	-43.12	43.16	-10.45	10.49	5.05	24.24
2	0.818	0.319	0.086	0.492	-36.17	21.47	-55.81	41.11	-19.12	4.41	5.68	27.23
3	0.977	0.154	0.024	0.171	-24.83	24.54	-41.65	41.37	-10.22	9.93	4.86	23.32
4	0.846	0.311	0.076	0.437	-47.11	11.07	-66.94	30.90	-29.90	-6.15	5.73	27.48

Note: Md. (method) is defined in Table 1, W, KS, W² and A² are Shapiro-Wilk, Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling statistics, respectively, for testing normality. Prediction intervals, tolerance intervals, and confidence limits (CL) containing the mean (\bar{e}_i) and the standard deviation (SD) are computed according to eqs. (31)-(34), respectively.

Table 3. A summary of overall prediction errors from different methods and error structures.

Data	Method	N	Mean	Min.	Max.	SD	δ	$\bar{e} \%$
Modelling	FO-iid	1249	0.685	-45.859	50.767	10.368	107.965	0.226
	FO-pow	1249	33.012	-119.206	195.721	60.059	4696.876	10.906
	FOCE-iid	1249	0.288	-50.877	42.649	9.922	98.529	0.095
	FOCE-pow	1249	16.243	-168.589	166.271	61.505	4046.700	5.366
Validation	FO-iid	1011	0.333	-38.308	38.004	8.157	66.648	0.116
	FO-pow	1011	12.438	-90.274	134.873	42.618	1970.998	4.325
	FOCE-iid	1011	-0.003	-38.336	35.250	7.769	60.357	-0.001
	FOCE-pow	1011	-2.211	-124.540	120.759	43.656	1910.735	-0.769

Note: the methods are defined in Table 1, N is the total number of observations, min., max. and SD are minimum, maximum and standard deviation, δ is an overall accuracy measure calculated by eq. (14), and $\bar{e} \%$ is the percent bias calculated by eq. (15).

Table 4. A summary of subject-specific prediction errors from different methods and error structures.

Data	Method	m	Mean	Min.	Max.	SD	δ	Freq. of $ \bar{e}_i \% > 2.5$
Modelling	FO-iid	288	0.250	-1.410	25.603	1.574	2.54	2
	FO-pow	288	6.181	-106.839	53.433	23.357	583.754	272
	FOCE-iid	288	0.085	-2.659	25.864	1.615	2.615	3
	FOCE-pow	288	-0.540	-145.735	50.470	28.404	807.079	260
Validation	FO-iid	265	-0.090	-16.637	5.187	1.549	2.408	12
	FO-pow	265	-0.364	-86.454	66.863	20.288	411.735	228
	FOCE-iid	265	-0.289	-20.144	3.256	1.820	3.396	13
	FOCE-pow	265	-6.627	-119.641	60.429	24.014	620.589	226

Note: the methods are defined in Table 1, m is the number of subjects (plots), min., max. and SD are minimum, maximum and standard deviation, δ_i is the subject-specific accuracy measure calculated by eq. (18), and $\bar{e}_i \%$ is the subject-specific percent bias calculated by eq. (19). The frequency (freq.) of $|\bar{e}_i \%| > 2.5$ refers to the number of plots whose $\bar{e}_i \%$ exceeded ± 2.5 (the 2.5 threshold was chosen by dividing the one-sided 5% significance level commonly used in statistical inference into two-sides).

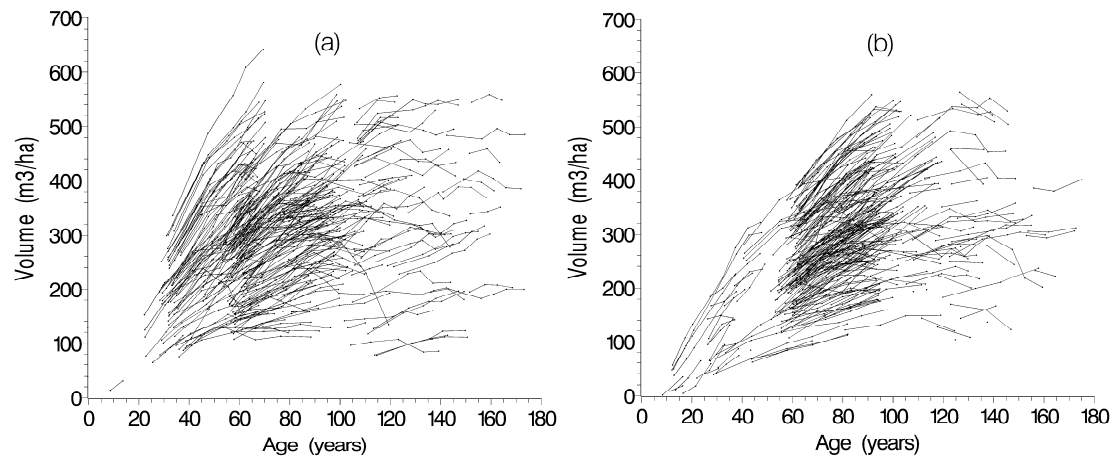


Figure 1. Model fitting (a) and validation (b) data sets. The model fitting data are from the upper foothills ecoregion. The model validation data are from the lower foothills ecoregion.

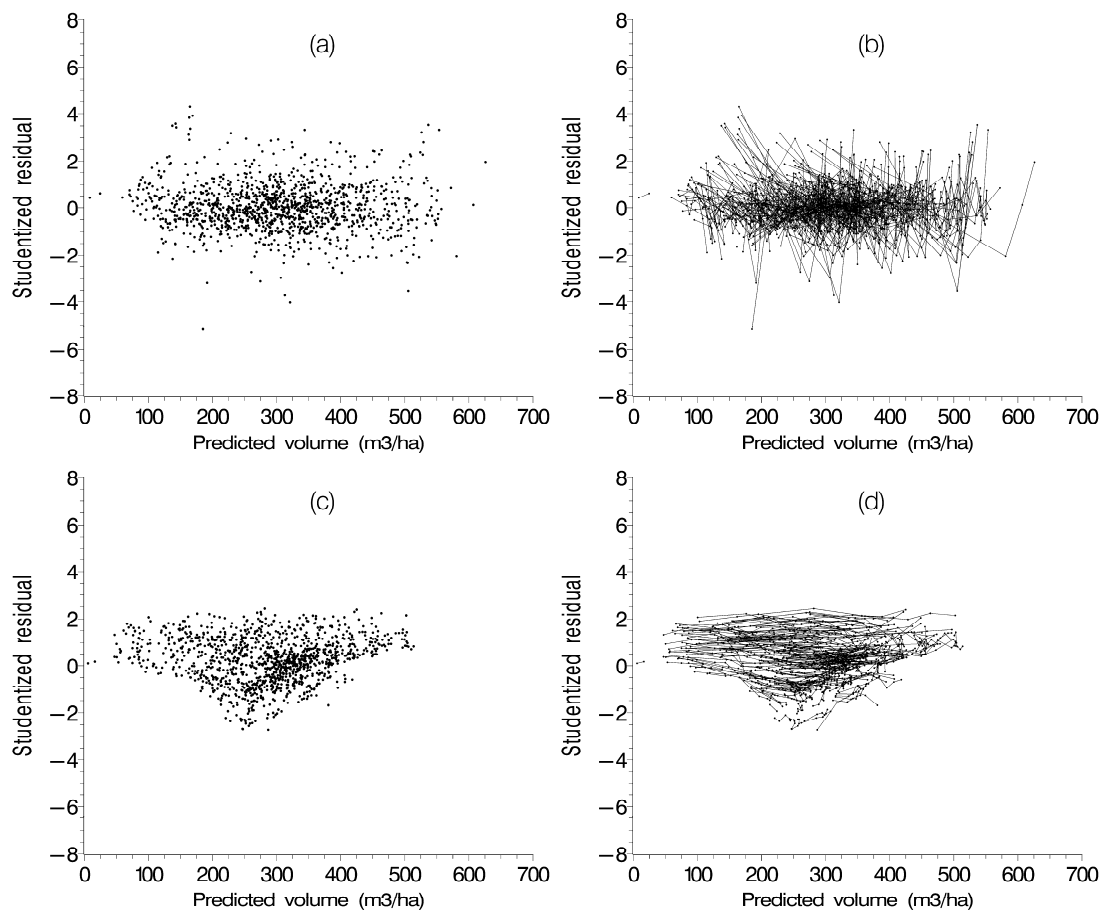


Figure 2. Studentized residual plots from the first-order conditional expectation (FOCE) method with the iid error structure (a, b), or correlated (spatial power) error structure (c, d). The right-hand side graphs connect the scatter points of the left-hand side graphs by subject.

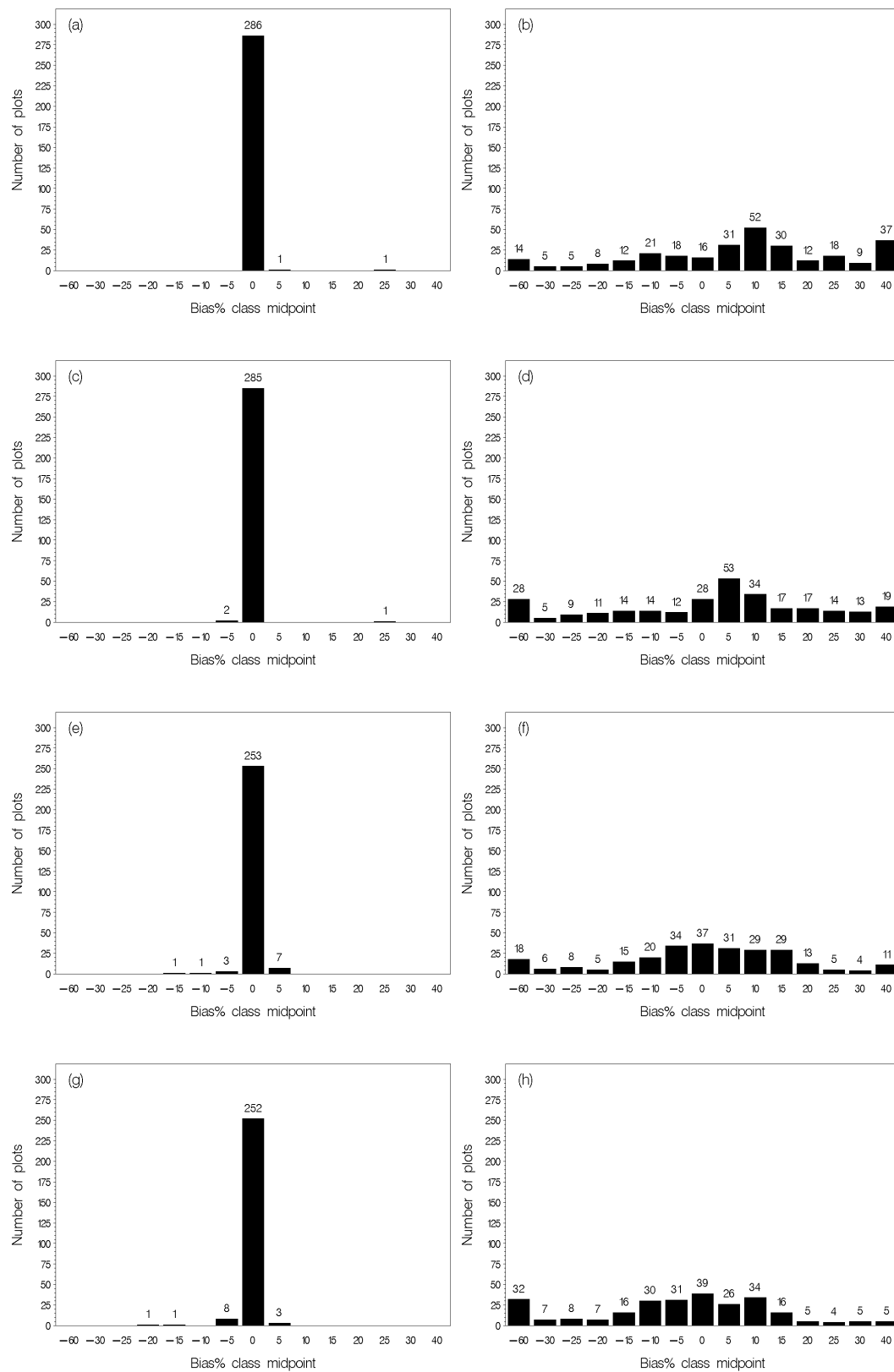


Figure 3. Frequency distributions of subject-specific percent biases ($\bar{e}_i\%$) for the model fitting data (a, b, c, d) and model validation data (e, f, g, h), based on the first-order (FO) and first-order conditional expectation (FOCE) methods: (a) and (e) – FO with iid errors; (b) and (f) - FO with correlated (spatial power) errors; (c) and (g) – FOCE with iid errors; and (d) and (h) - FOCE with spatial power errors.

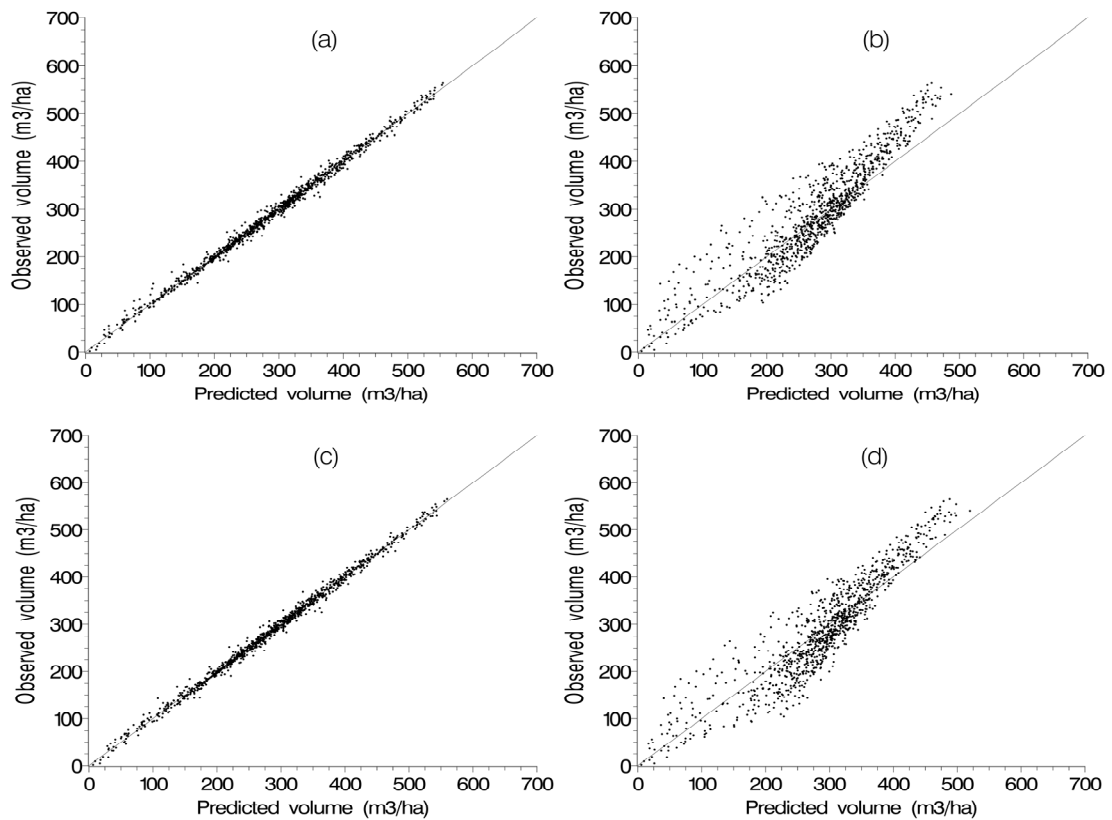


Figure 4. Observed versus predicted volumes for the validation data: (a) – first-order (FO) method with iid errors; (b) – FO method with spatial power errors; (c) – first-order conditional expectation (FOCE) method with iid errors; and (d) – FOCE method with spatial power errors. The line represents equality.

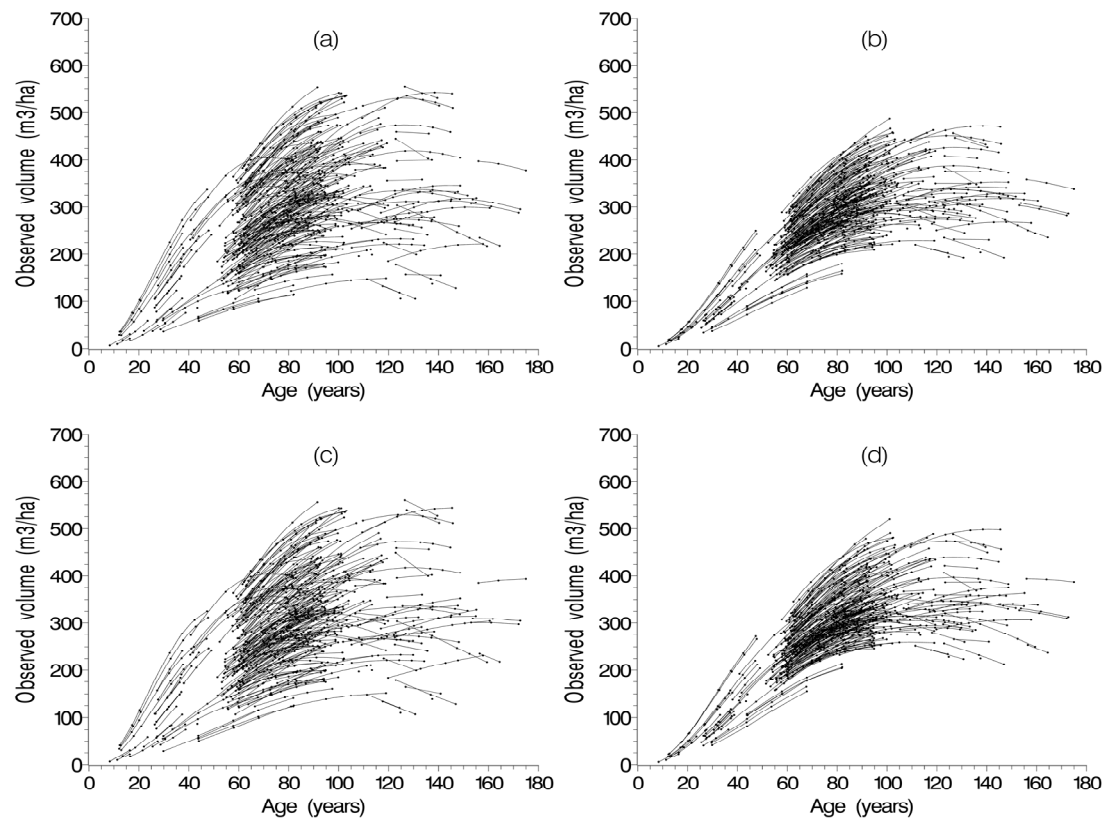


Figure 5. Predicted volume-age trajectories for the validation data: (a) – FO with iid errors; (b) – FO with spatial power errors; (c) – FOCE with iid errors; and (d) – FOCE with spatial power errors.

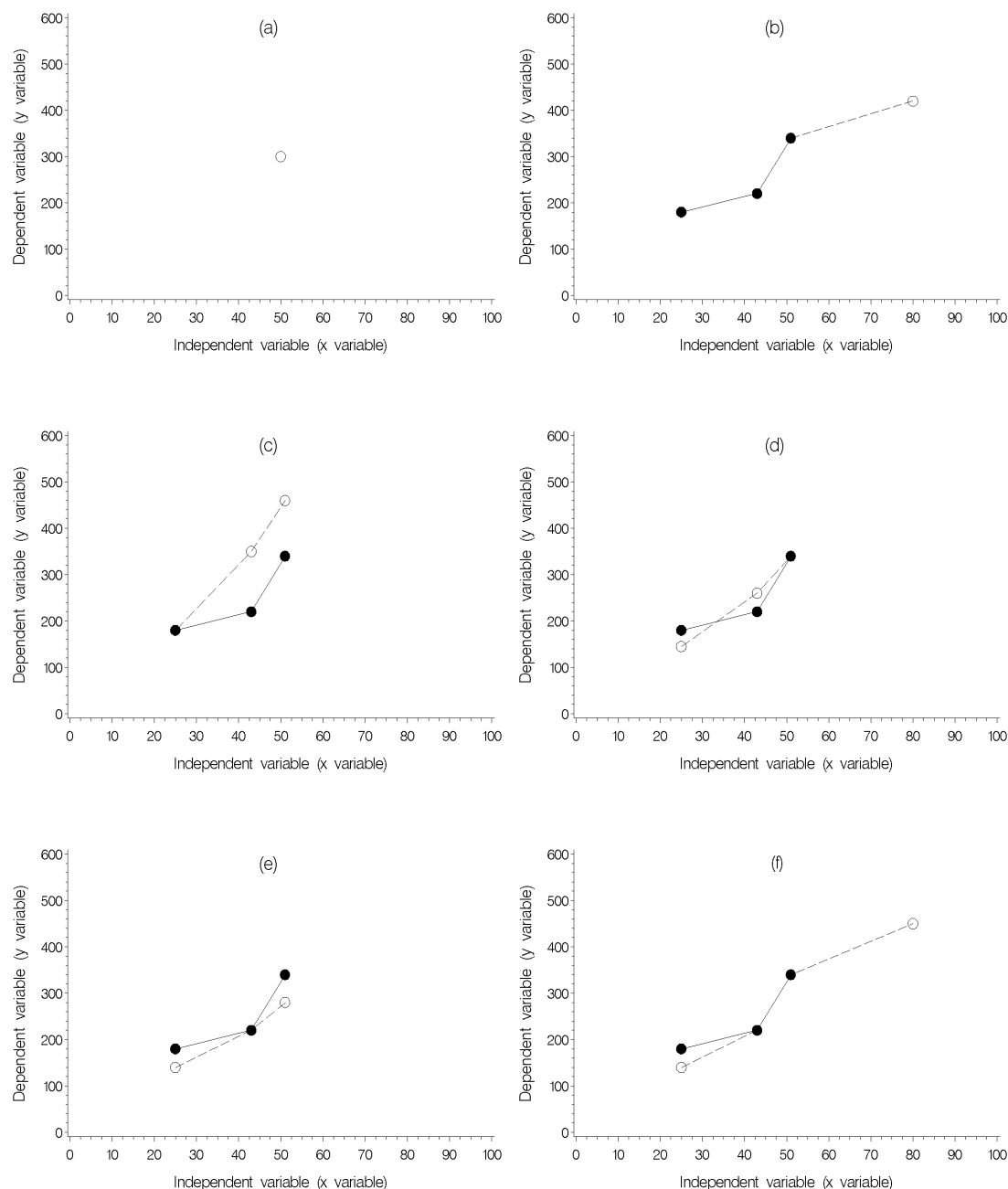


Figure 6. An illustration of selected prediction scenarios, where ‘•’ represents an observed value and ‘○’ represents a predicted value: (a) – predicting a y based on an observed x (equivalent to an overall population mean prediction); (b) – predicting a new (future) y based on the last y - x pair of an observed trajectory; (c) – predicting y s based on the first observed y - x pair; (d) – predicting previous y s based on the last observed y - x pair; (e) – predicting y s based on an observed y - x pair in the middle of a trajectory; and (f) – predicting a y or y s based on two observed y - x pairs. Note that (b), (c), (d) and (e) correspond to SS predictions based on one prior observation, and (f) corresponds to the predictions based on two prior observations. Depending on the number of observations available from an observed trajectory, forward and backward predictions can also be made using 3, 4, ..., k observations randomly selected from the trajectory. In our analyses, using (c) to illustrate, we used all 3 observations to get the random parameters. We then made predictions using (10) or (11) (we also examined the adjusted predictions, starting with the 1st observation and taking into account the spatial power structure when predicting the 2nd and 3rd observations through (10b) or (11b)).