Forestry An International Journal of Forest Research

Institute of Chartered Foresters

Forestry, Vol. 85, No. 1, 2012. doi:10.1093/forestry/cpr050
Advance Access publication date: 19 October 2011

A Bayesian approach for modelling non-linear longitudinal/hierarchical data with random effects in forestry

RONGXIA LI[1]*, BROCK STEWART[2] AND AARON WEISKITTEL[1]

[1] School of Forest Resources, University of Maine, 5755 Nutting Hall, Orono, ME 04469-57552, USA
[2] Centers for Disease Control and Prevention, 1600 Clifton Rd NE, Atlanta, GA 30333, USA
*Corresponding author. E-mail: rongxia.li@umit.maine.edu

Summary

Longitudinal or hierarchical data are often observed in forestry, which can pose both challenges and opportunities when performing statistical analyses. The current standard approach for analysing these types of data is mixed-effects models under the frequentist paradigm. Bayesian techniques have several advantages when compared with traditional approaches, but their use in forestry has been relatively limited. In this paper, we propose a Bayesian solution to non-linear mixed-effects models for longitudinal data in forestry. We demonstrate the Bayesian modelling process using individual tree height–age data for balsam fir (*Abies balsamea* (L.)) collected from eastern Maine. Due to its frequent utilization in modelling dominant tree height growth over time, we choose to examine models based on the Chapman–Richards function. We established four different model formulations, each having varying subject-specific parameters, which we estimated using both frequentist and Bayesian approaches. We found the estimation results to be quite close between the two methods. In addition, an important feature of the Bayesian method is the unified manner in which estimation and prediction are handled. Specifically, local parameters can be predicted for a new dataset after setting the posterior distributions from the estimation stage as new priors in the prediction phase.

Introduction

Longitudinal or hierarchical data are often observed in forestry since multiple measurements are frequently collected on the same tree, plot, or stand over time. For example, in growth and yield modelling, stand dominant tree height is generally measured at multiple times in order to develop a growth curve for future height predictions. In stem taper profile modelling, multiple measurements of diameters (inside or outside bark) are gathered from the base to the tip of an individual tree along the bole. These types of data are sometimes referred to as repeated measurements data. A distinctive feature of these types of data is that observations of the same individual subject are often correlated, which should be accounted for in the modelling process (Fitzmaurice *et al.*, 2004).

A standard approach to model clustered data is through mixed-effects modelling techniques. The theory of mixed-effects modelling has been well founded in statistics for both linear and non-linear regression (e.g. Lindsey, 1993; Davidian and Giltinan, 1995; Vonesh and Chinchilli, 1997). Applications using mixed-effects modelling approaches under the frequentist paradigm have been presented in the forestry literature for many data types, such as stem taper (e.g. Garber and Maguire, 2003), dominant height–age curves (e.g. Hall and Clutter, 2004) and stand basal area growth functions (e.g. Gregoire *et al.*, 1995). In a mixed-effects modelling framework, both within- and between-subject variation can be handled by specifying local (subject-specific) and global (population-level) parameters, respectively, while any within-subject correlation not accounted for by local parameters can be modelled in a straightforward manner using, for example, an autocorrelation structure on the error term variance–covariance matrix. Future predictions for a single subject can be obtained through the Estimated Best Linear Unbiased Predictor (EBLUP) formula (Hall and Bailey, 2001). The underlying estimation process to classical mixed-effects models is the maximum likelihood (ML) method.

However, dealing with non-linear mixed-effects models under the frequentist paradigm can be difficult. This is due in part to statistical inference for non-linear models depending on linear approximation methods (such as Taylor expansion) and large-sample asymptotic results based on normal theory. For a given dataset, it is often difficult to ensure such approximations are valid (De la Cruz-Mesia and Marshall, 2006). In addition, the uncertainty of parameter estimates and future predictions is relatively difficult to assess under this type of inferential set-up, especially in cases when general model assumptions such as independent errors and normality are relaxed. An alternative to classical ML mixed-effects models is to use Bayesian techniques.

Although Bayesian and frequentist approaches have been debated on the philosophical level in many scientific fields (Dennis, 1996; Edwards, 1996; Ellison, 2004), it has been shown that the Bayesian modelling method has unique advantages in many situations, which are facilitated by the dramatic increase recently seen in computing power. The primary benefit of using Bayesian approaches lies in their ability to generate the full posterior distribution of estimated parameters, thus various statistics (mean, mode, median, etc) can be easily calculated from samples generated during model fitting; since it is based on sample simulation algorithms, tedious inference is avoided. Moreover, the Bayesian approach does not impose additional constraints or assumptions on posterior distributions as in the frequentist ML method and variables are allowed to depart from the standard normal paradigm (Davidian and Giltinan, 1995).

In forestry, Bayesian methods have been applied to several applications, including diameter distribution estimates (Bullock and Boone, 2007), stand-level height and volume growth models (Green and Strawderman, 1996), linear models with heterogeneous variance for estimating tree foliar dry matter (Green and Valentine, 1998), individual tree mortality (Metcalf *et al.*, 2009) and a comparison between empirical and hierarchical Bayes inventory estimation methods (Green and Strawderman, 1992). Despite aforementioned studies, there is still a relative shortage of publications on the utilization of Bayesian methods in the forestry applications, compared with the rapid increase in the use of Bayesian techniques in other fields. Furthermore, we are unaware of the use of Bayesian techniques to model clustered data via non-linear models with random effects in the forestry literature, whereas in other fields it is not uncommon to apply Bayesian techniques to non-linear mixed-effects models (e.g. Wakefield, 1996; Broemeling and Cook, 1997).

The primary goal of this paper is to present a Bayesian approach for non-linear mixed-effects models to forestry longitudinal or hierarchical data. We first briefly introduce Bayes' Theorem, the foundation of the Bayesian modelling approach, and then we formulate a Bayesian modelling framework for non-linear mixed-effects models. In the end, we demonstrate the use of the Bayesian approach by modelling repeated measures of individual tree height–age data using both the classical ML method and the Bayesian approach to estimate the parameters, including random effects.

Methods

Bayes' Theorem

Readers who are interested in more details can consult (among others) Box and Tiao (1992) and Gelman *et al.* (1995). Here, we only present results that are essential for the current presentation.

Let $\mathbf{y} = (y_1, y_2, \ldots y_j, \ldots)$ represent a vector of data and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots \theta_j, \ldots)$ be a vector of parameters. The Bayesian Theorem can be summarized as follows:

$$p(\mathbf{y}, \boldsymbol{\theta}) = p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \mid \mathbf{y}) p(\mathbf{y}), \quad (1)$$

where p refers to the probability distribution or density function. In words, both the data and parameters are treated as random variables, and the joint probability function can be expressed as the product of the conditional distribution of one and the marginal distribution of the other. This equation can be rearranged as follows:

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{y})}, \quad (2)$$

where $p(\mathbf{y}) = E_{\boldsymbol{\theta}}(p(\mathbf{y} \mid \boldsymbol{\theta})) = \int p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$ for continuous θ. Since it is the integration of all admissible values of θ, p (y) does not depend on θ and can be viewed as a constant for fixed y, which yields the following:

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}). \quad (3)$$

Note that the conditional distribution of θ given data y (i.e. $p(\boldsymbol{\theta} \mid \mathbf{y})$) is what we are interested in estimating and is referred to as the posterior distribution in a Bayesian framework. $p(\mathbf{y} \mid \boldsymbol{\theta})$ tells us the distribution of y assuming θ is known, which is a likelihood function when viewed as a function of the parameters. p (θ) is the distribution of θ without any knowledge of data and is referred to as prior distribution of θ. Therefore, equation (3) says that the posterior distribution of θ is proportional to the product of the likelihood of y given θ and the prior distribution of θ.

Non-linear mixed-effects model specification in a Bayesian framework

The mathematical formulation of a non-linear mixed-effects model in a Bayesian framework is very similar as in a classical sampling theory context, except that all parameters in a Bayesian model are treated as random variables. In fact, mixed models can be considered a compromise between the frequentist and Bayesian approaches (Demidenko, 2004). A general non-linear mixed-effects model takes the following form:

$$y_{ij} = f(\boldsymbol{\theta}_i, \mathbf{x}_{ij}) + \varepsilon_{ij}, \quad (4)$$

where $\boldsymbol{\varepsilon}_{ij} \sim (0, \mathbf{R}_i)$, $\boldsymbol{\theta}_i = \boldsymbol{d}(\boldsymbol{\lambda}, \mathbf{b}_i)$, $\mathbf{b}_i \sim (0, \mathbf{D})$ and Corr(ε_{ij},b_i)=0.

y_{ij} denotes the j^{th} response value for i^{th} subject and $\boldsymbol{\theta}_i = (\theta_{1i}, \theta_{2i}, \ldots, \theta_{ki})$, assuming we have k parameters for any i^{th} subject. $\mathbf{x}_{ij}$ is a design matrix of explanatory variables, and ε_{ij} is an error term. Note that ε_{ij} does not have to follow a normal distribution. Further, $\boldsymbol{\theta}_i$ is formulated as a

function, d, of a global parameter vector $\boldsymbol{\lambda}$ whose value remains constant over the entire population and a random effect vector $\mathbf{b}_i$ that takes values specific to each subject. We take b_i to have mean $\mathbf{0}$ and covariance matrix $\mathbf{D}$; in general, the mean zero condition can be relaxed. Commonly, the random effects are assumed to follow a Gaussian or normal distribution: $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$. In a classical model, $\boldsymbol{\lambda}$ is generally referred to as a vector of fixed unknown parameters, whereas under a Bayesian set-up $\boldsymbol{\lambda}$ is viewed as a vector of random variables. In the Bayesian approach, the only difference between the global and local parameter vectors is that $\boldsymbol{\lambda}$ is used to depict the population trend, while $\mathbf{b}_i$ is used to describe each subject's departure from the trend given by $\boldsymbol{\lambda}$.

Prior distribution specification

In the above model specification, we need to choose appropriate prior distributions for all parameters, including λ, D and R_i. Non-informative priors are generally chosen if little knowledge about values of those parameters exists. For $\boldsymbol{\lambda}$, a common prior is $\boldsymbol{\lambda} \sim N(\boldsymbol{\lambda}^*, \mathbf{H})$, where $\boldsymbol{\lambda}^*$ is an *a priori* mean, and the variance–covariance matrix H should be large enough to ensure that $\mathbf{H}^{-1} \approx \mathbf{0}$. A Wishart distribution (a generalization to multiple dimensions of the chi-square distribution) is generally used as a prior for $\mathbf{D}^{-1}$ (Davidian and Giltinan, 1995). For $\mathbf{R}_i$, an identity matrix $\mathbf{I}_i$ multiplied by σ^2 is often used to reduce the number of parameters that need to be estimated. The variance–covariance matrices $\mathbf{H}$ and $\mathbf{D}$ can both be diagonal in order to further reduce the degree of model complexity.

Height–age modelling using a Bayesian approach

As an example, in this paper, we present the analysis of individual tree height–age models in a non-linear mixed-effects framework. We used both a classical ML and a Bayesian approach to estimate parameters.

Data

Most height–age data come from monumented plot, tree remeasurement or stem analysis data, which has important consequences on the model applicability (e.g. Raulier *et al.*, 2003). Our study used stem analysis data, which were originally collected between 1970 and 1977 (Vicary *et al.*, 1984). Stands were located throughout eastern, northern and western sections of Maine. Trees that were younger than 50 years old at breast height were sampled from even-aged spruce-fir stands. The trees were felled and stem analysis data were taken at stump height, breast height, and every 1.22 m (4 ft) above breast height with diameter inside bark and diameter outside bark measured at each section. The trees were systematically sampled to cover the range of diameters within a stand with emphasis on dominant and co-dominant individuals. Age at each stem section height was also recorded. Originally, sample trees included both balsam fir (*Abies balsamea* (L.)) and red spruce (*Picea rubens* (Sarg.)) in three regions in Maine. However, our analysis here is based only on the balsam fir data from the eastern region of Maine. This analysis can be easily expanded to other species and other regions. In total, 1121 observations of 121 trees from 23 plots were used for this study. The response variable height ranges from 0.11 to 56.60 m with a mean of 20.64 m and standard deviation (SD) of 12.98 m; the explanatory variable age ranges from 1 to 88 years with a mean of 26 years and an SD of 16 years (Figure 1).

Base model form

We opted to use the Chapman–Richards function (Richards, 1959) as the base function because it has been tested in numerous height–age applications for multiple species and shown to perform well in most situations (Garcia, 1983; Rennolls, 1995; Wang *et al.*, 2008; Weiskittel *et al.*, 2009; etc.). This function can be represented as follows:

$$H = \alpha(1 - \mathrm{e}^{-\beta A})^{\gamma}, \tag{5}$$

where H is dominant tree height, A is the corresponding breast-height age at that height, and α, β and γ are parameters that need to be estimated. This function produces sigmoidally shaped curves having an asymptote of α, while the parameters β and γ denote scale and shape, respectively. In order to model subject-level growth curves, one or more of the function's three parameters are allowed to take values that vary across individuals.

Model specification within Bayesian framework

We chose four models with different subject-specific parameters for evaluation, which are listed in Table 1 as M1–M4. Here, subject-specific mean parameters should differ among the 23 field plots to reflect plot-to-plot variation in the tree growth. These four model specifications were previously tested by Wang *et al.* (2008). Model M1 has a single

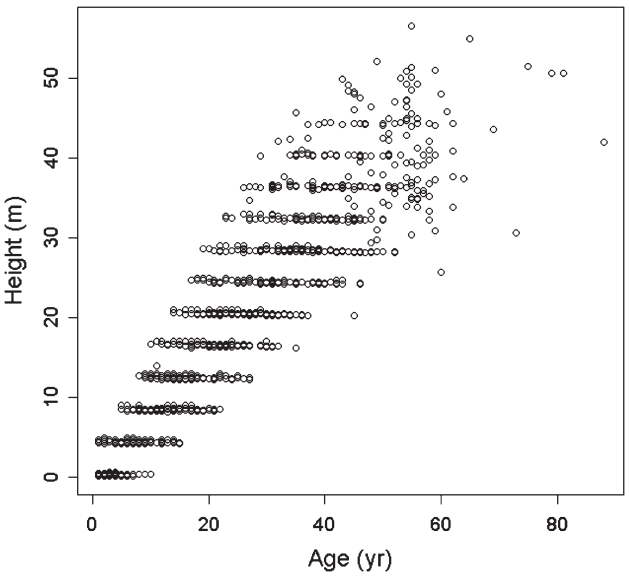

Figure 1. Raw data plot of tree height and age for 121 balsam fir trees in eastern Maine.

Table 1: Four height–age models with different subject-specific parameters

Model	Model form	Subject-specific parameter
M1	$H_{ij} = \alpha_i(1-e^{-\beta A_{ij}})^{\gamma} + \varepsilon_{ij}$	$\mathbf{b}_i = (a_i,0,0)$, local parameter: α_i
M2	$H_{ij} = \alpha(1-e^{-\beta A_{ij}})^{\gamma_i} + \varepsilon_{ij}$	$\mathbf{b}_i = (0,0,c_i)$, local parameter: γ_i
M3	$H_{ij} = \alpha_i(1-e^{-\beta A_{ij}})^{\gamma_i} + \varepsilon_{ij}$	$\mathbf{b}_i = (a_i,0,c_i)$, local parameter: α_i,γ_i
M4	$H_{ij} = e^{\chi_i}(1-e^{-\beta A_{ij}})^{\rho_1+\frac{\rho_2}{\chi_i}} + \varepsilon_{ij}$	Local parameter: χ_i

local quantity associated with parameter α, which provides growth curves having multiple asymptotes. Model M2 also has a single local quantity associated with parameter γ, which provides polymorphic curves having the same asymptote. Model M3 specifies that both α and γ have values specific to each subject in a functionally independent manner, which allows for polymorphic growth curves and multiple asymptotes. Model M4 is of the type suggested by Cieszewski and Bailey (2000) and tested by Dieguez-Aranda *et al*. (2006) and Wang *et al*. (2008), which imposes an *a priori* functional relationship between the values that α and γ are allowed to take across subjects. For the j^{th} observation of the i^{th} tree, the model forms for M1–M3 are as follows:

$$H_{ij} = f(\boldsymbol{\theta}_i, A_{ij}) = \alpha_i(1-e^{-\beta_i A_{ij}})^{\gamma_i} + \varepsilon_{ij}, \quad (6)$$

$$\text{where } \boldsymbol{\theta}_i = (\alpha_i,\beta_i,\gamma_i) = d(\boldsymbol{\lambda},\mathbf{b}_i) = \begin{bmatrix} \alpha + a_i \\ \beta + b_i \\ \gamma + c_i \end{bmatrix},$$

$$\boldsymbol{\lambda} = (\alpha,\beta,\gamma) \text{ and } \mathbf{b}_i = (a_i,b_i,c_i).$$

In M1, both b_i and c_i equal to zero; in M2, both a_i and b_i equal to zero; in M3, b_i is zero.

The form for M4 is as follows:

$$H_{ij} = f(\boldsymbol{\theta}_i, A_{ij}) = e^{X_i}(1-e^{-\beta A_{ij}})^{\rho_1+\frac{\rho_2}{\chi_i}} + \varepsilon_{ij}, \quad (7)$$

$$\text{where } \boldsymbol{\theta}_i = (\alpha_i,\beta_i,\gamma_i) = d(\boldsymbol{\lambda},\mathbf{b}_i) = \begin{bmatrix} e^{X_i} \\ \beta \\ \dfrac{\rho_1+\rho_2}{\chi_i} \end{bmatrix},$$

$$\boldsymbol{\lambda} = (\beta,\rho_1,\rho_2) \text{ and } \mathbf{b}_i = \chi_i.$$

The distributions for ε_{ij}, b_i, and λ are as follows:

$$\varepsilon_{ij} \sim N(0,\mathbf{R}_i),$$

$$\mathbf{b}_i \sim N(0,\mathbf{D}) \text{ and}$$

$$\boldsymbol{\lambda} \sim N(\boldsymbol{\lambda}^*,\mathbf{H}).$$

We assume $\mathbf{R}_i = \sigma^2\mathbf{I}_i$, which means we can use one single parameter σ to represent the intra-variation for all the subjects.

Table 2: Prior distributions of each parameter in M1–M4

Model	Parameter	Prior distribution
M1	α	$\alpha \sim N(57,1.0E-6)$
	β	$\beta \sim N(0.1,1.0E-6)$
	γ	$\gamma \sim N(1,1.0E-6)$
	σ_α	$\left(\frac{1}{\sigma_a}\right)^2 \sim$ Gamma(0.001,0.001)
	σ	$\left(\frac{1}{\sigma}\right)^2 \sim$ Gamma(0.001,0.001)
M2	α	$\alpha \sim N(57,1.0E-6)$
	β	$\beta \sim N(0.1,1.0E-6)$
	γ	$\gamma \sim N(1,1.0E-6)$
	σ_γ	$\left(\frac{1}{\sigma_\gamma}\right)^2 \sim$ Gamma(0.001,0.001)
	σ	$\left(\frac{1}{\sigma}\right)^2 \sim$ Gamma(0.001,0.001)
M3	α	$\alpha \sim N(57,1.0E-6)$
	β	$\beta \sim N(0.1,1.0E-6)$
	γ	$\gamma \sim N(1,1.0E-6)$
	σ_α	$\left(\frac{1}{\sigma_a}\right)^2 \sim$ Gamma(0.001,0.001)
	σ_γ	$\left(\frac{1}{\sigma_\gamma}\right)^2 \sim$ Gamma(0.001,0.001)
	σ	$\left(\frac{1}{\sigma}\right)^2 \sim$ Gamma(0.001,0.001)
M4	β	$\beta \sim N(0.1,1.0E-6)$
	ρ_1	$\rho_1 \sim N(1,1.0E-6)$
	ρ_2	$\rho_2 \sim N(1,1.0E-6)$
	χ	$\chi \sim N(1,1.0E-6)$
	σ_γ	$\left(\frac{1}{\sigma_x}\right)^2 \sim$ Gamma(0.001,0.001)
	σ	$\left(\frac{1}{\sigma}\right)^2 \sim$ Gamma(0.001,0.001)

For M1–M3,

$$\mathbf{D} = \begin{bmatrix} \sigma_\alpha & 0 & 0 \\ 0 & \sigma_\beta & 0 \\ 0 & 0 & \sigma_\gamma \end{bmatrix},$$

where σ_α, σ_β and σ_γ can be set as zero depending on the specific model form.

For M4, σ_x is the only parameter in the matrix **D**. The specific prior distributions for each parameter in M1–M4 are listed in Table 2. Based on the preliminary study, the priors for parameters α, β and γ were set as normal distributions with a specified mean. All other parameters have non-informative priors.

In addition to the parameter estimates, the diagnostic statistics root mean square error (RMSE), Akaike's Information Criterion (AIC) and Deviance Information Criterion (DIC) were also calculated for model performance evaluation. DIC is a generalization of AIC (Spiegelhalter *et al.*, 2002) and it is particularly useful in the Bayesian model selection. As with AIC, a smaller DIC number is preferred than a larger DIC to select a good-fitted model. DIC is defined as follows:

$$\mathrm{DIC} = \mathrm{Dbar} + \mathrm{pD}, \tag{8}$$

where Dbar is the posterior mean of the deviance, $\mathrm{Dbar} = E_\theta\left(-2\log\left(p(\mathrm{y} \mid \boldsymbol{\theta})\right)\right)$ and pD is the effective number of parameters, pD = Dbar – Dhat. Dhat is a point estimate of deviance given by $\mathrm{Dhat} = -2\log\left(p(\mathrm{y} \mid \bar{\boldsymbol{\theta}})\right)$, where y is the observed data and θ is the unknown parameter vector of the model.

Prediction

In this study, we randomly selected 3 plots from total 23 plots for prediction and used the remaining 20 plots for estimation. There are five to seven sampled trees located in each of the three randomly selected plots. We randomly chose one tree from each of the three plots to estimate plot-level random effects. Since our purpose for prediction in this study is not to validate the fitting results, we believe three randomly selected plots are sufficient to demonstrate the effectiveness of prediction using a Bayesian approach. In the frequentist paradigm, EBLUP was used to estimate the plot-level random effects. The EBLUP calculation can be found throughout the literature (Davidian and Giltinan, 1995; Vonesh and Chinchilli, 1997; Hall and Bailey, 2001). In the Bayesian analysis, the random effects were estimated by the same procedure we used to estimate the fixed effects. We can set the fixed parameters as fixed values estimated in the estimation procedure. Alternatively, we can set the previously estimated posterior distribution as the new prior distribution in the estimation of random effects. In this study, we chose to set the parameters α, β and γ as fixed values calculated from the posterior distribution in the estimation procedure.

Among the previously described four models M1–M4, we chose M1 and M3 to examine the models' prediction ability since M1 represents anamorphic height–age curves and M3 is a typical polymorphic model. The prior distribution of the random effects of a_i and c_i were set as normal. The mean absolute bias (MAB) and RMSE were used to examine the models' performance regarding the prediction. MAB is defined as follows:

$$\mathrm{MAB} = \frac{1}{n}\sum_{i=1}^{n} \mid \hat{Y}_i - Y_i \mid,$$

where Y_i refers to the observed or measured i^{th} response value (height), $\hat{Y}_i$ is the corresponding predicted response value and n is the number of observations.

We estimated Bayesian parameters using the WinBUGS software (Spiegelhalter *et al.*, 2003), which uses Markov chain Monte Carlo algorithms to solve for the optimal parameter values. We used the R packages BRugs (Thomas, 2004) and R2WinBUGS (Sturtz *et al.*, 2005) to link R and WinBUGS for data input and output and graph generation. We estimated traditional mixed-effects model parameters through the ML approach as implemented in the nlme function of the nlme library (Pinherio and Bates, 2000) in R.

Results

For models M1–M3, 500 000 iterations were set to run to ensure the obtainment of maximum convergence and satisfied posterior distributions of estimated parameters. Among those 500 000 iterations, the initial 50 000 iterations were discarded from analysis as burn-in iterations. For model M4, it only began to reach convergence after 300 000 iterations, therefore we set the total running iteration as 700 000 with the first 300 000 as burn-in iterations. The thinning parameter in M1–M4 was set to 3 to reduce the impact due to the correlation between neighbouring iterations. After the samples of posterior distributions were generated for each estimated parameter, the modes, SD and 95 per cent credible intervals were then calculated from the samples. The estimated values of those parameters in M1–M4 are presented in Table 3. The corresponding ML solutions under the frequentist paradigm are also presented in Table 3 as a comparison.

We found that M3, which specifies varying asymptote and shape parameters, consistently outperformed the other three models regardless of the modelling approach. Although model M4 also has varying asymptotes and shapes, the restriction imposed by the *ad hoc* functional parameter relationship restricts this model to a one-dimensional parameter space (Stewart *et al.*, 2011). Comparing the parameter estimates from frequentist solutions with the ones from Bayesian solutions, we found they are quite close for almost all the parameters. For example, the estimates of α, β and γ for M3 using the Bayesian approach are within 1.5, 5.4 and 3.8 per cent of frequentist ML solutions, respectively. Figure 2 shows the posterior distributions of each estimated parameter in M3 using the Bayesian approach and the classical ML parameter estimates. It is easy to see from the figure that the modes of all five parameters including both global parameters α, β and γ and local parameters σ_α and σ_γ are close to the ML estimates. For parameter α (asymptote), the modes in the Bayesian models centred at height (m) 51, 63, 48 and 51 for M1–M4, respectively, whereas the best ML solution for α is 49 m in M3. For parameter β, the difference between models exists only to three decimal places. Furthermore, the 95 per cent credible intervals from Bayesian solutions overlapped with the 95 per cent confidence interval to a large extent (Figure 3). RMSE of M3 from ML solutions is only 1.7 per cent lower than the one using the Bayesian approach.

Among our four models, M4 took the longest computing time (2.75 h for 700 000 iterations; all computations were performed on an Intel (R) e8600 3.33 GHz CPU). Models M1 and M2, which are subsets of model M3 obtained by setting one of the random effects in M3 to zero, required

Table 3: Parameter estimates of four models using both ML and Bayesian approaches

		Classical model (ML)						Bayesian model					
		Parameter estimate		95% interval				Parameter estimate		95% interval			
Model		Mean	SD	Lower	Higher	RMSE	AIC	Mode	SD	Lower	Higher	RMSE	DIC
M1	α	50.41	2.49	45.53	55.30	3.38	6039	51.02	2.76	46.03	56.87	3.42	5970
	β	0.034	0.002	0.029	0.038			0.032	0.002	0.028	0.037		
	γ	1.62	0.06	1.49	1.74			1.59	0.06	1.46	1.71		
	σ_α	9.36		6.89	12.72			9.67	1.70	7.12	13.55		
M2	α	71.92	3.62	64.83	79.02	3.51	6120	62.37	2.94	57.70	69.23	3.54	6047
	β	0.02	0.002	0.02	0.02			0.03	0.002	0.02	0.03		
	γ	1.46	0.08	1.30	1.62			1.67	0.11	1.44	1.87		
	σ_α	0.27		0.18	0.40			0.35	0.08	0.24	0.53		
M3	α	49.46	2.15	45.24	53.67	2.91	5777	48.89	2.37	44.80	53.59	2.96	5665
	β	0.037	0.002	0.033	0.041			0.039	0.002	0.035	0.042		
	γ	1.84	0.10	1.64	2.05			1.91	0.13	1.68	2.17		
	σ_α	8.32		6.02	11.49			7.97	1.52	5.78	11.34		
	σ_c	0.36		0.26	0.52			0.39	0.08	0.27	0.57		
M4	β	0.034	0.002	0.029	0.038	3.34	6011	0.034	0.002	0.030	0.039	3.37	5937
	ρ_1	−6.00	1.54	−9.03	−2.98			−2.94	1.25	−5.96	−1.19		
	ρ_2	30.75	6.23	18.55	42.96			18.23	5.04	11.22	30.55		
	χ	3.96	0.04	3.88	4.04			3.93	0.04	3.86	4.01		
	σ_x	0.11		0.07	0.16			0.13	0.02	0.09	0.18		

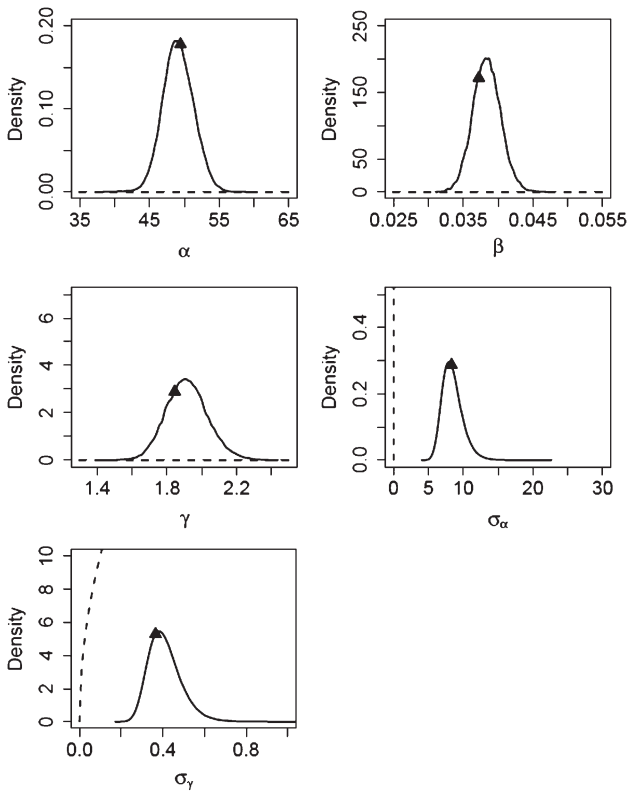

Figure 2. ML and Bayesian estimates for the parameters in M3. Solid lines are posterior density curves, whereas dashed lines are prior density curves. ML estimates are shown as solid triangle on the corresponding Bayesian posterior curves.

42 and 29 per cent less running time than model M3, respectively: models M1, M2 and M3 required 1.80, 1.98 and 2.56 h to complete 500 000 iterations, respectively. As a comparison, we also fitted the base function of models M1–M3 by specifying all three parameters as fixed across all subjects (no random effects); this model required only 0.71 h to complete 500 000 iterations. It should also be noted that factors other than the number of random effects in the model can affect running times. However, running time was not the primary interest in our study; perhaps future research could more fully examine computational times of frequentist and Bayesian modelling approaches as they relate to forest growth and yield modelling.

For the predictions, M1 showed a bias of 3.83 m and an RMSE of 4.97 m using Bayesian techniques, comparing with an MAB of 3.14 m and an RMSE of 4.48 m using traditional EBLUP method. For the model M3 with both α and γ as random effects, the MAB was 3.44 and 3.21 m and the RMSE was 4.32 and 4.21 m using Bayesian and EBLUP methods, respectively.

Discussion

We have presented a Bayesian solution to non-linear mixed-effects models in a context of individual tree height–age modelling. Although the Bayesian approach was introduced to forest modelling four decades ago (Swindel, 1972), applications using Bayesian techniques in forestry are still relatively uncommon in the literature. This might be due to the extensive computing power required by the Bayesian simulations or perhaps a lack of familiarity of Bayesian

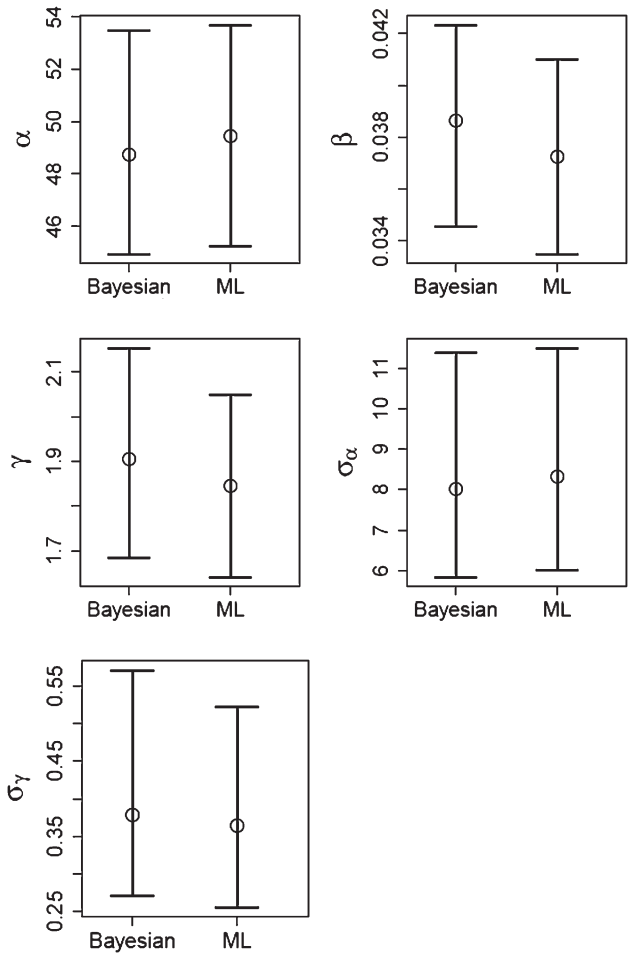

Figure 3. Bayesian 95% credible intervals and ML 95% confidence intervals for parameters in M3.

techniques. In this paper, we do not intend to argue whether Bayesian modelling techniques are superior to the classical frequentist approach nor do we attempt to bring any kind of debate regarding the appropriate use of Bayesian models in forest growth and yield modelling. We simply propose an alternative way to solve the non-linear mixed-effects models for longitudinal or hierarchical data commonly found in forestry. We think both frequentist and Bayesian approaches have their own unique features. Although the frequentist ML method guarantees an optimal solution in most situations, Bayesian approaches have the ability to generate full posterior samples and thus allow for a straightforward assessment of uncertainties of each parameter estimate.

Also, Bayesian parameter estimates can be used as input data for further modelling process, e.g. predictions. Another potential feature of Bayesian methods is that inference does not rely on large-sample, asymptotic approximate normal theory results. It is in our expectation that the Bayesian results did not outperform the frequentist in estimation and prediction in this particular study since a Gaussian framework is more likely to favour ML estimates. However, it shows that the Bayesian approach can be compatible with traditional regression methods in certain situations and should be adopted into forestry modelling for longitudinal or hierarchical data as an alternative to the frequentist ML method. In addition, in this study, we only provided one tree to estimate the plot-level random effects for prediction; further studies are needed to examine whether multiple trees or more observations can, and to what extent, improve Bayesian prediction accuracy.

In our study, we chose four model types with varying subject-specific parameters based on the Chapman–Richards function to model forest stand height–age data. All four model forms have been examined for different species in the literature. For example, Krumland and Eng (2005) used M1 and M4 in their report to predict site index for multiple species, including Douglas-fir (*Pseudotsuga menziesii* Franco.), lodgepole pine (*Pinus contorta* Dougl.), red alder (*Alnus rubra* Bong.) and others in California. Wang *et al.* (2008) tested and compared the performance of M1–M4 for loblolly pine trees using different estimation methods. Although Bayesian estimation methods have not been employed to solve mixed-effects models in the literature given above, we found Bayesian techniques are relatively common in other scientific fields for longitudinal data with random effects. For example, a Bayesian zero-inflated Poisson model was established for count data to study the effects of pesticide on immature insects (Dagne, 2010); Wagner and Tüchler (2010) described a Bayesian approach to model multivariate responses of mixed data; Bayesian methodology for repeated measurements and longitudinal data was also widely adopted in epidemiology (e.g. Horrocks and van Den Heuvel, 2009) and pharmaceuticology (e.g. Cauchemez *et al.*, 2004; Hu *et al.*, 2009) fields. We believe that the field of forest growth and yield modelling could benefit from further explorations of the use of Bayesian techniques for non-linear modelling of longitudinal data.

One of the most significant features of Bayesian methods for mixed-models analysis is perhaps that they provide a unified framework for estimation and predictions (De la Cruz-Mesia and Marshall, 2006). This is very important in subject-specific mixed-effects models because the subject-specific local parameters can be easily predicted on a new dataset with the same procedures used in estimation. In the prediction stage, the estimated posterior distributions for global parameters can be set as either fixed values or new priors in predicting local parameters. That is, informative new priors obtained from the estimation stage are used in the model calibration process to incorporate estimation results from previously fitted models. It is the nature of Bayesian statistics to update an old model with new data. This aspect of Bayesian approaches is also of great importance when fitted models need to be frequently updated with newly collected data, which is common for forestry applications. For the future research, we need to develop software packages to smoothly integrate Bayesian modelling techniques into forestry modelling practice.

Funding

Funding was provided by the University of Maine, School of Forest Resources, Cooperative Forestry Research Unit, and the Forest Bioproducts Research Institute.

Conflict of interest statement

None declared.

References

Box, G. and Tiao, G. 1992 *Bayesian Inference in Statistical Analysis*. Wiley, New York, NY.

Broemeling, L.D. and Cook, P. 1997 A Bayesian analysis of regression models with continuous errors with application to longitudinal studies. *Stat. Med.* **16**, 321–332.

Bullock, B.P. and Boone, E.L. 2007 Deriving tree diameter distributions using Bayesian model averaging. *For. Ecol. Manag.* **242**, 127–132.

Cauchemez, S., Carrat, F., Viboud, C., Valleron, A.J. and Boëlle, P.Y. 2004 A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat. Med.* **23**, 3469–3487.

Cieszewski, C.J. and Bailey, R.L. 2000 Generalized algebraic difference approach: theory based derivation of dynamic site equations with polymorphism and variable asymptotes. *For. Sci.* **46** (1), 116–126.

Dagne, G.A. 2010 Bayesian semiparametric zero-inflated Poisson model for longitudinal count data. *Math. Biosci.* **224**, 126–130.

Davidian, M. and Giltinan, D.M. 1995 *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall, London, UK.

De la Cruz-Mesía, R. and Marshall, G. 2006 Non-linear random effects models with continuous time autoregressive errors: a Bayesian approach. *Stat. Med.* **25**, 1471–1484.

Demidenko, E. 2004 *Mixed Models: Theory and Applications*. Wiley, Hoboken, NJ. pp. P8–P9.

Dennis, B. 1996 Discussion: should ecologists become Bayesians? *Ecol. Appl.* **6**, 1095–1103.

Dieguez-Aranda, U., Burkhart, H.E. and Amateis, R.L. 2006 Dynamic site models for loblolly pine (*Pinus taeda* L.) plantations in the United States. *For. Sci.* **52**, 262–272.

Edwards, D. 1996 Comment: the first data analysis should be journalistic. *Ecol. Appl.* **6**, 1090–1094.

Ellison, A.M. 2004 Bayesian inference in ecology. *Ecol. Lett.* **7**, 509–520.

Fitzmaurice, G.M., Laird, N.M. and Ware, J.H. 2004 *Applied Longitudinal Analysis*. Wiley, Hoboken, NJ. pp. P2–P4.

Garber, S.M. and Maguire, D.A. 2003 Modeling stem taper of three central Oregon species using nonlinear mixed effects models and autoregressive error structures. *For. Ecol. Manag.* **179**, 507–522.

Garcia, O. 1983 A stochastic differential equation model for the height growth of forest stands. *Biometrics*. **39**, 1059–1072.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. 1995 *Bayesian Data Analysis*. Chapman & Hall, London.

Green, E.J. and Strawderman, W.E. 1992 A comparison of hierarchical Bayes and empirical Bayes methods with a forestry application. *For. Sci.* **38**, 350–366.

Green, E.J. and Strawderman, W.E. 1996 Predictive posterior distributions from a Bayesian version of a slash pine yield model. *For. Sci.* **42**, 456–464.

Green, E.J. and Valentine, H.T. 1998 Bayesian analysis of the linear model with heterogeneous variance. *For. Sci.* **44**, 134–138.

Gregoire, T.G., Schabenberger, O. and Barrett, J.P. 1995 Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Can. J. For. Res.* **25**, 137–156.

Hall, D.B. and Bailey, R.L. 2001 Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. *For. Sci.* **47**, 311–321.

Hall, D.B. and Clutter, M. 2004 Multilevel multivariate nonlinear mixed effects models for timber yield predictions. *Biometrics*. **60**, 16–24.

Horrocks, J. and van Den Heuvel, M.J. 2009 Prediction of pregnancy: a joint model for longitudinal and binary data. *Bayesian Anal.* **4**, 523–538.

Hu, W., Li, G. and Li, N. 2009 A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. *Stat. Med.* **28**, 1601–1619.

Krumland, B. and Eng, H. 2005 *Site Index Systems for Major Young-Growth Forest and Woodland Species in Northern California. California Forestry Report 4*. Department of Forestry and Fire Protection, State of California Resources Agency, Sacramento, CA. 219 pp.

Lindsey, J.K. 1993 *Models for Repeated Measurements*. Oxford University Press Inc., New York, NY.

Metcalf, C.J.E., McMahon, S.M. and Clark, J.S. 2009 Overcoming data sparseness and parametric constraints in modeling of tree mortality: a new nonparametric Bayesian model. *Can. J. For. Res.* **39**, 1677–1687.

Pinherio, J.C. and Bates, D.M. 2000 *Mixed-Effects Models in S and S-Plus*. Springer-Verlag, New York, NY.

Raulier, F., Lambert, M.C., Pothier, D. and Ung, C.H. 2003 Impact of dominant tree dynamics on site index curves. *For. Ecol. Manag.* **184**, 65–78.

Rennolls, K. 1995 Forest height growth modeling. *For. Ecol. Manag.* **71**, 217–225.

Richards, F.J. 1959 A flexible growth function for empirical use. *J. Exp. Bot.* **10**, 290–300.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. 2003 *WinBUGS User Manual*. http://www.mrc-bsu.cam.ac.uk/bugs (accessed on 2 October, 2011).

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van der Linde, A. 2002 Bayesian measures of model complexity and fit (with discussion). *J. Roy. Stat. Soc. B*. **64**, 583–616.

Stewart, B., Jordan, L. and Li, R. 2011 Comparing ADA, GADA, and site index models versus effects models: restrictions in model formulation. *For. Sci.* (in press).

Sturtz, S., Ligges, U. and Gelman, A. 2005 R2WinBUGS: A Package for Running WinBUGS from R. *J. Stat. Softw.* **12**, 1–16.

Swindel, B.F. 1972 *The Bayesian Controversy*. USDA Forest Serv. Res. Pap. SE-95, 12 pp.

Thomas, A. 2004 *BRugs User Manual: The R Interface to BUGS*. http://www.openbugs.info/w/ (accessed on 2 October, 2011)

Vicary, B.P., Brann, T.B. and Griffin, R.H. 1984 *Polymorphic Site Index Curves for Even-Aged Spruce-Fir Stands in Maine*. Maine Agricultural Experiment Station Bulletin 802. University of Maine, Orono, ME, 33 pp.

Vonesh, E.F. and Chinchilli, V.M. 1997 *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker, New York, NY.

Wagner, H. and Tüchler, R. 2010 Bayesian estimation of random effects models for multivariate responses of mixed data. *Comput. Stat. Data An.* **54**, 1206–1218.

Wakefield, J. 1996 The Bayesian analysis of population pharmacokinetic models. *J. Am. Stat. Assoc.* **92**, 62–75.

Wang, M., Borders, B.E. and Zhao, D. 2008 An empirical comparison of two subject-specific approaches to dominant heights modeling: the dummy variable method and the mixed model method. *For. Ecol. Manag.* **255**, 2659–2669.

Weiskittel, A.R., Hann, D.W., Hibbs, D.E., Lam, T.Y. and Bluhm, A.A. 2009 Modeling top height growth of red alder plantations. *For. Ecol. Manag.* **258**, 323–331.

Received 22 February 2011