Variance modelling of longitudinal height data from a *Pinus radiata* progeny test

Luis A. Apiolaza, Arthur R. Gilmour, and Dorian J. Garrick

Abstract: Variance components were estimated using alternative structures for the additive genetic covariance matrix ($\mathbf{G}_0$), for height (m) of trees measured at 10 unequally spaced ages in an open-pollinated progeny test. These structures reflected unstructured, autoregressive, banded correlation and random regressions models. The residual matrix ($\mathbf{R}_0$) was unstructured, and the block and plot strata matrices were autoregressive. The best model for $\mathbf{G}_0$ considering the likelihood value and number of parameters was the autoregressive correlation form with age-specific variances and time on a natural logarithm basis. The genetic correlation between successive measures ranged from 0.93 at age 1 to 0.99 at age 14 years. Heritability increased with age from 0.09 (age 1) to 0.24 (age 7) and then declined to 0.13 at age 15. Heritabilities from the unstructured model were similar, while heritabilities assuming banded correlations were lower after age 7. The covariance structure implicit in the random regressions model was considered unsatisfactory. Using structures in $\mathbf{G}_0$ facilitated model fitting and convergence of the likelihood maximisation algorithm. Fitting a structured matrix that reflects the relationships present in repeated measures may overcome problems of nonpositive definiteness of unstructured matrices from longitudinal data, especially when genetic variation is small.

Résumé : Les auteurs ont estimé les composantes de la variance à l’aide de structures d’ajustement à la matrice de covariances génétiques additives ($\mathbf{G}_0$) pour la hauteur (m) d’arbres mesurés à 10 intervalles non réguliers au sein d’un test de descendances issues de pollinisation libre. Ces structures reflétaient des modèles non structurés, auto-régressifs, par corrélations stratifiées et par régressions aléatoires. La matrice résiduelle ($\mathbf{R}_0$) n’était pas structurée, et les matrices à l’échelle des blocs et des parcelles étaient auto-régressives. En considérant la valeur de maximum de vraisemblance et le nombre de paramètres, le meilleur modèle d’ajustement de $\mathbf{G}_0$ était celui des corrélations auto-régressives avec les variances spécifiques à chaque âge et le temps suivant l’échelle logarithmique naturelle. Les corrélations génétiques entre les mesures successives variaient de 0,93 à 1 an jusqu’à 0,99 à 14 ans. L’héritabilité augmentait avec l’âge, allant de 0,09 (à 1 an) à 0,24 (à 7 ans) pour décroître à 0,13 à 15 ans. Les héritabilités découlant du modèle non structuré étaient similaires, alors que les héritabilités découlant du modèle par corrélations stratifiées étaient plus faibles après 7 ans. La structure de covariances implicite au modèle de régressions aléatoires n’a pas été jugée satisfaisante. La structuration de $\mathbf{G}_0$ a facilité l’ajustement des modèles ainsi que la convergence de l’algorithme de maximisation de la vraisemblance. Les auteurs en concluent que l’ajustement d’une matrice structurée réflétant les liens de dépendance parmi les mesures répétées peut permettre de résoudre les problèmes de manque de d’exactitude présents dans les matrices non structurées découlant de données longitudinales, spécialement lorsque la variabilité génétique est faible.

[Traduit par la Rédaction]

Introduction

The success of tree breeding programmes relies on their ability to identify and deploy superior trees, progressively increasing profit. Decisions on how, when, and what to select are made (or should be made) taking into account genetic and economic information. Selection in tree breeding programmes is based upon genetic information generated from progeny tests.

The net present value of genetic gain depends on the total genetic gain (ΔG) and the time when improved material is deployed and costs (L) are incurred (Newman and Williams 1991). Both ΔG and L contain constraints and opportunities for breeding programmes to make rapid gains. Faster gains can be achieved by increasing the selection differentials and (or) the accuracy of prediction. Furthermore, using overlapping generations and early selection can reduce long generation intervals. Efficient selection at an early age requires high correlation with rotation-age production traits and reasonably high heritabilities of both. Knowledge of the expected covariance structure across ages enables prediction of the response to early selection.

“Longitudinal” data arise when individuals are assessed for the same outcome at several ages (Diggle et al. 1994; see Cnaan et al. 1997 for a review). Breeders use longitudinal data from progeny tests to compare development patterns of genotypes and look at changes in genetic parameters over

Received May 18, 1999. Accepted December 1, 1999.

L.A. Apiolaza.[1] Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, and New Zealand Forest Research Institute, Private Bag 3020, Rotorua, New Zealand. e-mail: luis.apiolaza@utas.edu.au
A.R. Gilmour. New South Wales Agriculture, Orange 2800, Australia. e-mail: arthur.gilmour@agric.nsw.gov.au
D.J. Garrick. Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand. e-mail: d.garrick@massey.ac.nz

[1] Corresponding author. Current address: CRC for Sustainable Production Forestry, School of Plant Science, University of Tasmania, GPO Box 252-55, Hobart, Tasmania 7001, Australia.

time. Understanding the genetics of development allows determination of the optimum evaluation time(s) for fine-tuning breeding programmes and the use of multiple assessments for genetic evaluations. One of the features of longitudinal data is the covariance that exists between observations of the same individual (Diggle et al. 1994; Hand and Crowder 1996). Covariance matrices typically contain pattern or structure, which can be modelled with a reduced number of parameters. Diggle et al. (1994) identify three sources of variation in longitudinal data: serial correlations, random effects, and measurement error. These act simultaneously.

Previous researchers in tree breeding have used several variance models for longitudinal data. Early studies used univariate analysis at each age, sometimes fitting a curve through the estimates to smooth and interpolate the results (e.g., Foster 1986). This procedure may produce unbiased estimates of heritability in absence of selection but ignores the dependence (covariance) between times. Other studies used bivariate analyses of each pair of times (e.g., Balocchi et al. 1993), increasing the understanding of the association between measures. Another procedure uses the correlations between family means of the same genetic material grown at different sites, i.e., based on type B correlation (Burdon 1977; see Hodge and White 1992 for an extensive application), with the intention of avoiding error correlation between the measures.

Longitudinal analyses are more efficient using all available information, especially when missing observations are a problem. Progeny tests, as other long-term forestry experiments, do not maintain the original design. Even when a test starts as a balanced experiment, mortality generates temporal imbalance; hence, early measures contain more records than later ones (Gregoire et al. 1995). Some authors choose to eliminate temporal unbalance, keeping only individuals with a full history of measures (e.g., Balocchi et al. 1993), but this approach omits useful data and does not consider that biases may arise if mortality is not random.

There are recent attempts with forest trees to use two contrasting models for genetic correlation: a repeatability (univariate) model (e.g., Wei and Borralho 1996) and a full multivariate model (e.g., Wei and Borralho 1998). The repeatability model assumes that all measures represent the same trait. This implies a genetic correlation of one between all pairs of records, equal variance for all records and equal environmental correlation between all pairs of records. The model can be represented by $\mathbf{G} = \sigma_a^2\mathbf{J}$ and $\mathbf{R} = \sigma_e^2(I + r\mathbf{J})$ where $\mathbf{I}$ is the identity matrix, $\mathbf{J}$ is a square matrix with all elements equal to 1, σ_a^2 is the additive genetic variance component, σ_e^2 is the error variance component and $r/(1 + r)$ is the correlation between residuals. With height increasing with age over many years, the equal variance and genetic correlation assumptions are often unrealistic, although variance heterogeneity may be crudely removed through standardization. In contrast, the full multivariate model considers each age as the realization of a different trait. It was originally applied to balanced and complete data, but modern computing techniques allow its application to incomplete data sets.

Unstructured covariance matrices in a full multivariate analysis are feasible and reasonable with a small number, t, of successive measures. However, since these matrices have $t(t - 1)/2$ covariance components, more measures implies a large number of poorly estimated parameters and may be considered an overparameterization (Hand and Crowder 1996). That is, there may not be enough information in the data to estimate each variance and covariance with sufficient accuracy for the resulting matrix to be coherent (positive definite), and the likelihood maximization algorithm may even fail to converge. This is less a problem with some structured matrices, making appealing the assumption of more parsimonious covariance structures as t increases.

Some researchers have recognized the need for structured covariance matrices, yet only in isolated cases. For example Quaas et al. (1984, p. 34) proposed the use of an autoregressive error structure in a repeatability model, relaxing the equal correlation assumption. Kremer (1992) explicitly recognized the role of error serial correlation for the analysis of height increments. Coelli et al. (1998) compared several different structures for the multivariate analysis of additive genetic effects of repeated measures in a sheep breeding context.

An alternative approach for modelling covariance structures, regression models with random coefficients, was introduced by Rao (1965) in the context of growth models. Laird and Ware (1982) generalized the theory to include mixed models, with fixed parameters at the population level and random parameters at the individual level. Under this framework, unbalanced and incomplete data sets are readily handled, and the correlation among successive measures is implicitly modelled by the random regressions (Louis 1988; Vonesh and Chinchilli 1997). Schaeffer and Dekkers (1994), Jamrozik and Schaeffer (1997), and Jamrozik et al. (1997) applied random regression models to the analysis of lactation records, while Gregoire et al. (1995) advocated the use of random regressions to model growth in permanent plots in forest mensuration. The use of random regressions in these studies allowed for individuals with heterogeneous ages to be included in the analyses, and a reduction of computational requirements compared with unstructured multivariate analyses. Random regression models directly define covariance functions that are the continuous (infinitesimal) equivalent of a covariance matrix for a given trait and fixed ages (Kirkpatrick et al. 1990, 1994). These functions permit us to calculate the covariance between any two ages, as can also be done with the distance-based autocorrelation model. So, the association among measurements may be modelled either through random regressions or through specification of covariance structures. In some cases both methods are used together (e.g., Chi and Reinsel 1989; Jones 1990).

In this paper we model longitudinal data from a progeny test. We compare estimated genetic parameters obtained from traditional approaches with those from various variance models under the general mixed model. First, we introduce the general model in a tree breeding context. Then we present alternative structures for the additive genetic covariance matrix. Finally, we discuss opportunities and limitations for the use of these models.

Materials and methods

Data set

The forestry company Bosques Arauco S.A. established trial FA8102 in the VIII Región of Chile in 1981 to progeny test 45 radiata pine (*Pinus radiata* Donn ex D. Don)

Table 1. Summary statistics by age.

Age	No. of individuals	Percentage of trees	Height (m)
1	1522	88.4	0.48 (0.12)
2	1525	88.6	1.00 (0.22)
4	1525	88.6	2.98 (0.60)
5	1524	88.6	4.60 (0.85)
6	1511	87.8	6.45 (1.02)
7	1515	88.0	8.43 (1.11)
8	1510	87.7	10.43 (1.22)
9	1505	87.4	12.11 (1.36)
12	1351	78.5	17.97 (1.79)
15	1284	74.6	22.34 (2.46)

Note: Values for height are means, with SD given in parentheses. The percentage of trees refers to the individuals included in the analyses relative to the number initially established (1721, without considering controls and fillers) discounting natural mortality, mechanical damage and inconsistent measures.

open-pollinated first-generation selections with 9 controls. These were planted in five-tree plots within eight randomized complete blocks, a total of 2160 trees. Control plots were planted with mixed seed of unknown pedigree and have been omitted from the analysis. Trees that were suppressed by early competition never reached 5 cm diameter at breast height (DBH) and were also omitted. Consequently, a total of 1526 trees in 353 plots were included in the analysis, each tree being from seed collected from one of 45 mother trees.

The trees were assessed for height at 1, 2, 4, 5, 6, 7, 8, 9, 12, and 15 years of age. In the case of trees with mechanical damage (as those broken by wind at age 12), observations after the damage occurred were eliminated. Summary statistics by age are presented in Table 1. Small fluctuations of the number of trees between ages 1–2 and 6–7 are caused by the elimination of inconsistent assessments. It is only after 9 years that there is appreciable mortality among the trees. The regression of log(standard deviation) on log(mean height) has a slope of 0.74 ± 0.03 suggesting that a power transformation of the data to height$^{0.26}$ would stabilize the variance (Box and Cox 1964). However, we will analyse the data on the measurement scale using models with heterogeneous variances to account for the increase of variance with age.

General linear mixed model

An individual tree ("animal") linear mixed-effects model equation for longitudinal data of tree i can be expressed as

$$[1] \qquad \mathbf{y}_i = \mathbf{X}_i\mathbf{m} + \mathbf{Z}_{\mathrm{B}_i}\mathbf{b} + \mathbf{Z}_{\mathrm{P}_i}\mathbf{p} + \mathbf{Z}_{\mathrm{T}_i}\mathbf{a}_i + \mathbf{e}_i$$

where $\mathbf{y}_i$ is the vector of s_i observations for the individual indexed by age, $\mathbf{m}$ is the vector of fixed effects (which may include regression coefficients at population level), $\mathbf{b}$ is the vector of random block by age effects, $\mathbf{p}$ is the vector of random plot by age effects, $\mathbf{a}_i$ is the vector of individual random additive genetic effects, and $\mathbf{e}_i$ is the vector of random residuals. $\mathbf{X}_i$, $\mathbf{Z}_{\mathrm{B}_i}$, $\mathbf{Z}_{\mathrm{P}_i}$, and $\mathbf{Z}_{\mathrm{T}_i}$ are incidence matrices relating $\mathbf{m}$, $\mathbf{b}$, $\mathbf{p}$, and $\mathbf{a}_i$ to $\mathbf{y}_i$. Thus the expected value and dispersion matrices assuming a multivariate normal distribution (MND) are

$$[2] \qquad \begin{bmatrix}\mathbf{y}_i\\ \mathbf{b}\\ \mathbf{p}\\ \mathbf{a}_i\\ \mathbf{e}_i\end{bmatrix} \sim \mathrm{MND}\left(\begin{bmatrix}\mathbf{X}_i\mathbf{m}\\ \mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{V}_i & \mathbf{Z}_{\mathrm{B}_i}\mathbf{B}_0 & \mathbf{Z}_{\mathrm{P}_i}\mathbf{P}_0 & \mathbf{Z}_{\mathrm{T}_i}\mathbf{G}_0 & \mathbf{R}_0\\ \mathbf{B}_0\mathbf{Z}'_{\mathrm{B}_i} & \mathbf{B}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{P}_0\mathbf{Z}'_{\mathrm{P}_i} & \mathbf{0} & \mathbf{P}_0 & \mathbf{0} & \mathbf{0}\\ \mathbf{G}_0\mathbf{Z}'_{\mathrm{T}_i} & \mathbf{0} & \mathbf{0} & \mathbf{G}_0 & \mathbf{0}\\ \mathbf{R}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_0\end{bmatrix}\right)$$

where $\mathbf{B}_0$, $\mathbf{P}_0$, $\mathbf{G}_0$, and $\mathbf{R}_0$ are the block, plot, additive genetic, and residual covariance matrices, respectively, and $\mathbf{0}$ is a null matrix (with all elements equal to 0). The corresponding characteristic elements (for measures j and k) are $\sigma_{b_{jk}}$, $\sigma_{p_{jk}}$, $\sigma_{a_{jk}}$ and $\sigma_{e_{jk}}$. The number of observations per individual may vary in which case the corresponding rows and columns of $\mathbf{R_0}$ are deleted. Finally, the phenotypic covariance matrix is

$$[3] \qquad \mathbf{V}_i = \mathbf{Z}_{\mathrm{B}_i}\mathbf{B}_0\mathbf{Z}'_{\mathrm{B}_i} + \mathbf{Z}_{\mathrm{P}_i}\mathbf{P}_0\mathbf{Z}'_{\mathrm{P}_i} + \mathbf{Z}_{\mathrm{T}_i}\mathbf{G}_0\mathbf{Z}'_{\mathrm{T}_i} + \mathbf{R}_0$$

Extending the model equation to the n subjects of a progeny test we obtain

$$[4] \qquad \mathbf{y} = \mathbf{Xm} + \mathbf{Z}_{\mathrm{B}}\mathbf{b} + \mathbf{Z}_{\mathrm{P}}\mathbf{p} + \mathbf{Z}_{\mathrm{T}}\mathbf{a} + \mathbf{e}$$

for $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2, \ldots, \mathbf{y}'_n)'$, $\mathbf{a} = (\mathbf{a}'_1, \mathbf{a}'_2, \ldots, \mathbf{a}'_n)'$, $\mathbf{e} = (\mathbf{e}'_1, \mathbf{e}'_2, \ldots, \mathbf{e}'_{\boldsymbol{n}})'$, $\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2, \ldots, \mathbf{X}'_n)'$, $\mathbf{Z}_{\mathrm{B}} = (\mathbf{Z}'_{\mathrm{B}_1}, \mathbf{Z}'_{\mathrm{B}_2}, \ldots, \mathbf{Z}'_{\mathrm{B}_n})'$, $\mathbf{Z}_{\mathrm{P}} = (\mathbf{Z}'_{\mathrm{P}_1}, \mathbf{Z}'_{\mathrm{P}_2}, \ldots, \mathbf{Z}'_{\mathrm{P}_n})'$, and $\mathbf{Z}_{\mathrm{T}} = \Sigma_{\oplus}\mathbf{Z}_{\mathrm{T}_i}$. Hence, in the dispersion matrix $\mathbf{B} = \Sigma_{\oplus}\mathbf{B}_0$, $\mathbf{P} = \Sigma_{\oplus}\mathbf{P}_0$, $\mathbf{G} = \mathbf{A}_{\mathrm{n}}\otimes\mathbf{G}_0$, and $\mathbf{R} = \Sigma_{\oplus}\mathbf{R}_0$, where $\mathbf{A}_{\mathrm{n}}$ is the numerator relationship matrix, $\Sigma_{\oplus}$ denotes direct sum, and ⊗ represents direct product operation (Searle 1982). Since the genetic relationships in our study are limited to half-sib information, it is possible to fit the equivalent half-sib ("sire") model with family rather than tree as the random factor. We keep the tree model notation for the sake of generality.

Parameterizations of the model

The expected value of $\mathbf{y}_i$ is $\mathbf{X}_i\mathbf{m}$ (eq. 2). This is used to model the average performance of trees as fixed effects and, in this case, is the mean at each age. From eq. 3, the dependence of the variance of $\mathbf{y}_i$ on the specification of $\mathbf{B}_0$, $\mathbf{P}_0$, $\mathbf{G}_0$, and $\mathbf{R}_0$ is clear. Since our main interest in the analysis is $\mathbf{G}_0$, thus following Coelli et al. (1998), we will fit an unstructured error covariance matrix ($\mathbf{R}_0$) while examining various forms for the additive genetic covariance matrix. For the block ($\mathbf{B}_0$) covariance matrix we use an autoregressive correlation structure with separate variances at each time, because it matches our general expectation that the correlation would reduce as the time interval increases. Since there are only eight blocks, an unstructured form for $\mathbf{B}_0$ would be singular and not estimable. For the plot ($\mathbf{P}_0$) matrix we will primarily use a similar autoregressive correlation structure with heterogeneous variance but will also report some results from fitting an unstructured form.

Fig. 1. Covariance structures fitted in this study (example using only four ages or measures). Different letters (*a*, *b*, etc.) represent different values of correlation. US, unstructured; BC, banded correlations; AR, autoregressive (t_j is age at measurement *j*), RR, random regressions expressed as the product $\mathbf{Q}_i\Lambda_0\mathbf{Q}_i'$ ($\mathbf{z}_{ij}$ is the *i*th orthogonal polynomial vector evaluated at age *j*); UC, uncorrelated. See text for more detail in the explanation.

$$\begin{bmatrix} 1 & a & b & c \\ a & 1 & d & e \\ b & d & 1 & f \\ c & e & f & 1 \end{bmatrix}$$

US

$$\begin{bmatrix} 1 & a & b & c \\ a & 1 & a & b \\ b & a & 1 & a \\ c & b & a & 1 \end{bmatrix}$$

BC

$$\begin{bmatrix} 1 & a^{|t_2-t_1|} & a^{|t_3-t_1|} & a^{|t_4-t_1|} \\ a^{|t_2-t_1|} & 1 & a^{|t_3-t_2|} & a^{|t_4-t_2|} \\ a^{|t_3-t_1|} & a^{|t_3-t_2|} & 1 & a^{|t_4-t_3|} \\ a^{|t_4-t_1|} & a^{|t_4-t_2|} & a^{|t_4-t_3|} & 1 \end{bmatrix}$$

AR

$$\begin{bmatrix} z_{01} & z_{11} & z_{21} & z_{31} \\ z_{02} & z_{12} & z_{22} & z_{32} \\ z_{03} & z_{13} & z_{23} & z_{33} \\ z_{04} & z_{14} & z_{24} & z_{34} \end{bmatrix} \begin{bmatrix} 1 & a & b & c \\ a & 1 & d & e \\ b & d & 1 & f \\ c & e & f & 1 \end{bmatrix} \begin{bmatrix} z_{01} & z_{02} & z_{03} & z_{04} \\ z_{11} & z_{12} & z_{13} & z_{14} \\ z_{21} & z_{22} & z_{23} & z_{24} \\ z_{31} & z_{32} & z_{33} & z_{34} \end{bmatrix}$$

RR

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

UC

We examine the following forms for the additive genetic covariance $\mathbf{G}_0$ (examples of the structures are displayed in Fig. 1):

(1) Full multivariate model (US). Here $\mathbf{G}_0$ is an unstructured matrix.

(2) Banded correlation model (BC) with heterogeneous variances. Here $\mathbf{G}_0 = \mathbf{SC}_{\mathrm{BC}}\mathbf{S}$, where $\mathbf{S}$ is diagonal matrix of the square roots of the genetic variance components at each age and $\mathbf{C}_{\mathrm{BC}}$ is a banded correlation matrix with a specific correlation for each particular age interval.

(3) Autoregressive model (AR) with heterogeneous variances. Here $\mathbf{G}_0 = \mathbf{SC}_{\mathrm{AR}}\mathbf{S}$, where $\mathbf{S}$ is as above and $\mathbf{C}_{\mathrm{AR}}$ has an autoregressive correlation structure. We use a power formulation (Diggle et al. 1994) which allows for the unequal age intervals, and compare submodels where the age is expressed on the natural ($\rho^{|k-j|}$, ARnat), square root ($\rho^{\left|\sqrt{k}-\sqrt{j}\right|}$, ARsqr), and logarithmic ($\rho^{|\log(k)-\log(j)|} = \rho^{|\log(k/j)|}$, ARlog) scales, where ρ is a correlation coefficient and *j* and *k* are the two ages. Note that when lag (age interval) is expressed in a logarithmic scale the correlation depends upon an age ratio.

(4) Random regression model (RR) using orthogonal polynomials. Here $\mathbf{G}_0 = \mathbf{Q}_i\Lambda_0\mathbf{Q}_i'$, where Λ_0 is the random regressors covariance matrix and $\mathbf{Q}_i$ has $q + 1$ columns containing $z_0, z_1, z_2, \ldots, z_q$, respectively, where *q* is the order of the polynomial and z_i is the *i*th orthogonal polynomial vector. Additionally, $\mathbf{a}_i = \mathbf{Q}_i\lambda_i$, where λ_i is the vector of random regression coefficients.

(5) Lastly, an uncorrelated model (UC) (for estimating heritabilities only) is fitted, which is equivalent to a traditional univariate analysis by age. Here all covariances in $\mathbf{G}_0$ (as well as in $\mathbf{R}_0$) are zero.

All models are fitted by restricted maximum likelihood (REML; Patterson and Thompson 1971) using the average information algorithm (Gilmour et al. 1995) implemented in ASReml (Gilmour et al. 1998).

Model selection

Adding variance parameters to a model may result in a better fit and hence increase the likelihood value. Several criteria may be used to judge whether additional parameters are making an important contribution to the fit. The likelihood ratio test formally tests whether the increase is statistically significant. Akaike's information criterion and Bayesian information criterion (Akaike 1974; Jones 1993; Carlin and Louis 1996) penalize the likelihood by the number of independently fitted parameters used in the model. Based on previous experiences in genetic analyses (Wada and Kashiwagi 1990) we will use Akaike's criterion (AIC), which penalizes likelihood values in such a way that any extra parameter must increase the likelihood by at least one unit to be included in the model:

$$\mathrm{AIC} = -2\ \mathrm{Log}\ L + 2p$$

where –2 Log *L* is twice the negative log-likelihood value for the model and *p* is the number of estimated parameters. Smaller values of AIC reflect an overall better fit.

Genetic parameters

Estimates of heritability (h_j^2) at age *j* and genetic correlation (r_{jk}) between ages *j* and *k* are calculated as

$$\hat{h}_j^2 = \frac{\hat{\sigma}_{a_j}^2}{\hat{\sigma}_{a_j}^2 + \hat{\sigma}_{b_j}^2 + \hat{\sigma}_{p_j}^2 + \hat{\sigma}_{e_j}^2}$$

$$\hat{r}_{jk} = \frac{\hat{\sigma}_{a_{jk}}}{\sqrt{\hat{\sigma}_{a_j}^2 \times \hat{\sigma}_{a_k}^2}}$$

Fig. 2. Box plot of height versus age. The midpoint, box, and whiskers represent the mean, mean ± SD, and minimum–maximum, respectively. A cubic polynomial (······) is fitted to the data.

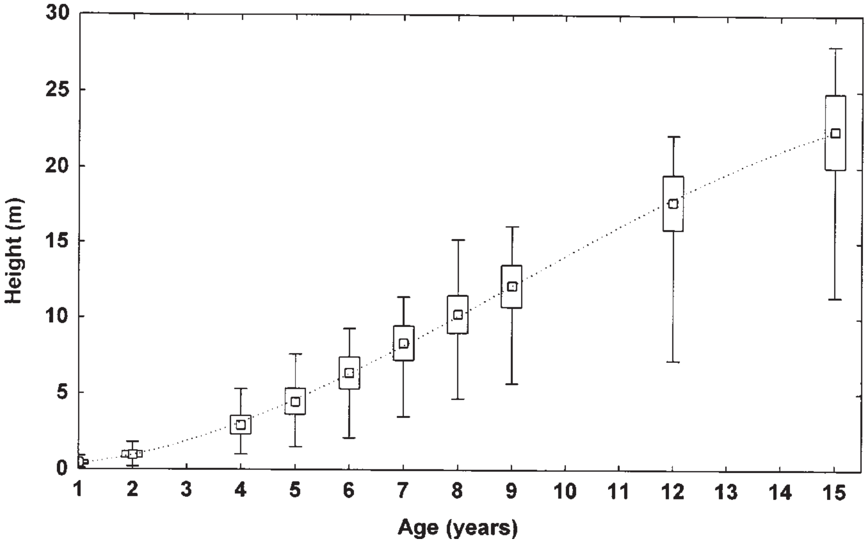


Table 2. Comparison of multivariate models.

Model	No. of parameters ($\mathbf{B}_0$ + $\mathbf{P}_0$ + $\mathbf{G}_0$ + $\mathbf{R}_0$)	Log likelihood	AIC[a]
Autoregressive $\mathbf{P}_0$: base model	11 + 11 + 0 + 55 = 77	7062.51	–13 971.02
Full multivariate (US)[b]	11 + 11 + 55 + 55 = 132	7098.87	–13 933.74
Banded correlations (BC)	11 + 11 + 21 + 55 = 98	7074.94	–13 953.88
Autoregressive (ARnat – age)	11 + 11 + 11 + 55 = 88	7067.04	–13 958.08
Autoregressive (ARsqr – $\sqrt{\text{age}}$)	11 + 11 + 11 + 55 = 88	7067.85	–13 959.70
Autoregressive (ARlog – log(age))[c]	11 + 11 + 11 + 55 = 88	7068.60	–13 961.20
Random regression (RR)[d]	11 + 11 + 10 + 55 = 87	7078.29	–13 982.58
Unstructured $\mathbf{P}_0$: base model	11 + 55 + 0 + 55 = 121	7124.78	–14 007.56
Full multivariate (US)[e]	11 + 55 + 55 + 55 = 176	7153.17	–13 954.34
Autoregressive (ARlog – log(age)) [f]	11 + 55 + 11 + 55 = 132	7127.84	–13 991.68
Uncorrelated (UC)	10 + 10 + 10 + 10 = 40	–3418.23	6 916.46

[a] AIC = –2 × log likelihood + 2 × number of parameters.
[b] $\mathbf{G}_0$ is nonpositive definite; log likelihood reduces to 6854.05 after bending.
[c] Best model including $\mathbf{G}_0$ and considering autoregressive $\mathbf{P}_0$.
[d] Reduced rank version: the model converged only by fixing the variance for the quadratic component.
[e] $\mathbf{G}_0$ is nonpositive definite.
[f] Best model including $\mathbf{G}_0$ and considering unstructured $\mathbf{P}_0$.

using the appropriate variance and covariance estimates from $\mathbf{G}_0$, $\mathbf{B}_0$, $\mathbf{P}_0$, and $\mathbf{R}_0$. Standard errors of the estimates are calculated by ASReml from the average information matrix, using a standard Taylor series approximation (Gilmour et al. 1998).

Results and discussion

Exploratory analysis

Figure 2 shows a box plot of height versus age. The midpoint, box, and whiskers represent the mean, mean ± SD, and minimum–maximum, respectively. This plot depicts an increase of variance with time, which is typical of longitudinal data for growth. It also reveals that the distribution of heights at particular ages is not symmetric after year 5. This asymmetry would be aggravated if a variance stabilizing transformation was applied to the data. The figure suggests that a cubic polynomial (broken line) is a reasonable model to form the basis for modelling genetic effects as random regressions.

Variance models

The log-likelihood values for a range of models are in Table 2. They range from –3418.2 for the UC model to 7098.9 (7153.2) for the US model with autoregressive (unstructured) plot error structure. Using AIC, the best model

Table 3. Variance parameters estimated from the autoregressive (ARlog) model with autoregressive plot error: block (b^2), plot (p^2), additive genetic (h^2), and residuals (e^2) variances expressed as proportion of phenotypic variance ($\hat{\sigma}_p^2$) and correlations among residuals (below diagonal) and genetic correlations (above diagonal).

	Variance parameters					Correlation at age									
Age	b^2	p^2	h^2	e^2	$\hat{\sigma}_p^2$	1	2	4	5	6	7	8	9	12	15
1	0.049	0.180	0.091	0.680	0.016		0.932	0.868	0.849	0.833	0.820	0.809	0.799	0.776	0.759
2	0.068	0.193	0.082	0.658	0.051	0.697		0.932	0.911	0.894	0.880	0.868	0.858	0.833	0.814
4	0.057	0.178	0.170	0.596	0.380	0.607	0.758		0.978	0.960	0.945	0.932	0.921	0.894	0.874
5	0.050	0.162	0.168	0.620	0.754	0.560	0.683	0.836		0.982	0.966	0.953	0.942	0.915	0.894
6	0.057	0.166	0.209	0.568	1.108	0.501	0.627	0.786	0.887		0.984	0.971	0.960	0.932	0.911
7	0.060	0.141	0.238	0.561	1.309	0.469	0.581	0.731	0.831	0.916		0.987	0.975	0.947	0.925
8	0.040	0.125	0.228	0.606	1.573	0.419	0.528	0.675	0.763	0.855	0.901		0.988	0.960	0.938
9	0.055	0.095	0.197	0.653	2.016	0.399	0.492	0.633	0.714	0.820	0.882	0.890		0.971	0.949
12	0.026	0.067	0.171	0.737	3.399	0.361	0.412	0.533	0.620	0.719	0.779	0.812	0.832		0.978
15	0.006	0.015	0.129	0.849	6.959	0.362	0.414	0.529	0.611	0.708	0.770	0.811	0.820	0.898	

(i.e., the one with the lowest value) for both plot error structures is the base model, i.e., the one with no tree effects fitted. Of those with tree effects fitted, the best is the ARlog structure having age on a natural logarithm basis with AIC = –13 961.2. It is followed by the ARsqr, the ARnat, the BC, and the US models (Table 2). The RR model is considered later. The ARnat and ARsqr models have slightly lower heritabilities and slightly higher genetic correlations than the ARlog model, and so their parameters are not included in Figs. 3 and 4. In the following discussion we refer to models fitted with an AR plot variance, since not all models would converge to positive definite solution when fitted with a US plot variance model. The genetic parameters with AR plot variance were not much different from those with US plot variance.

The proportions by age for all variance components in the ARlog model are in Table 3. Plot variances are the largest component (18%) in the early years slowly declining in relative magnitude about 1% per year. The early plot variation might reflect carry-over effects from the nursery. The block variance component is around 5% before declining after age 9. The residual variance is stable at around 60% of phenotypic variance until age 9 increasing to 85% at age 15.

Heritability estimates under the ARlog model (Table 3) increase with age from 0.091 at age 1 to 0.238 at age 7 and then decline to 0.129 at age 15. The BC model gave higher heritability before age 8 and lower heritability for later measurements (Fig. 3). However, compared with the ARlog model, an increase in the likelihood of 6.3 with 10 extra parameters indicates that the model does not fit the data significantly ($P > 0.05$) better. Heritabilities from the US and UC models are very similar to the values from the ARlog model, reaching a maximum of 0.241 and 0.224, respectively (Fig. 3). The figure also includes the asymptotic estimate of the standard error of the heritability from the ARlog model. There is little difference in the standard error of the heritability estimates from the AR (with any time scale), BC, US, and UC models.

The heritability differences among these models are quite small until age 6, when the BC model starts underestimating later values. The benefits of a multivariate versus a univariate approach are clearer when covering traits with large differences between genetic and residual correlations, contrasting heritabilities (Thompson and Meyer 1986) or there are many missing values for some traits, especially if not missing at random. Here however, the correlations are mostly high, all heritabilities are moderate and only 16% of trees have missing values, particularly at ages 12 and 15.

Figure 4 shows the values of genetic correlation plotted against lag for the ARlog, BC, and US models. The simple regressions of the correlations on lag shown in Fig. 4 had R^2 coefficients of 0.58, 0.97, and 0.53 for ARlog, BC, and US, respectively. The correlation between successive ages under the ARlog model increased from 0.932 at age 1 to 0.993 at 14 years of age. The BC model has a single value for each lag whereas the ARlog and US models have several values for each lag, although the ARnat model has a single value per lag.

The US estimates of $\mathbf{G}_0$ were both nonpositive definite and with some correlations above 1 (Fig. 4). Such a solution is not easily used in practice but commonly occurs because

Fig. 3. Heritability estimates based on fitting the unstructured (US; —○—), banded correlation (BC; ···●···), autoregressive (ARlog; ---Δ---), random regressions (RR; —●—), and uncorrelated (UC; ···+···) models considering autoregressive plot errors, and standard error of the heritability for the autoregressive (······) model.

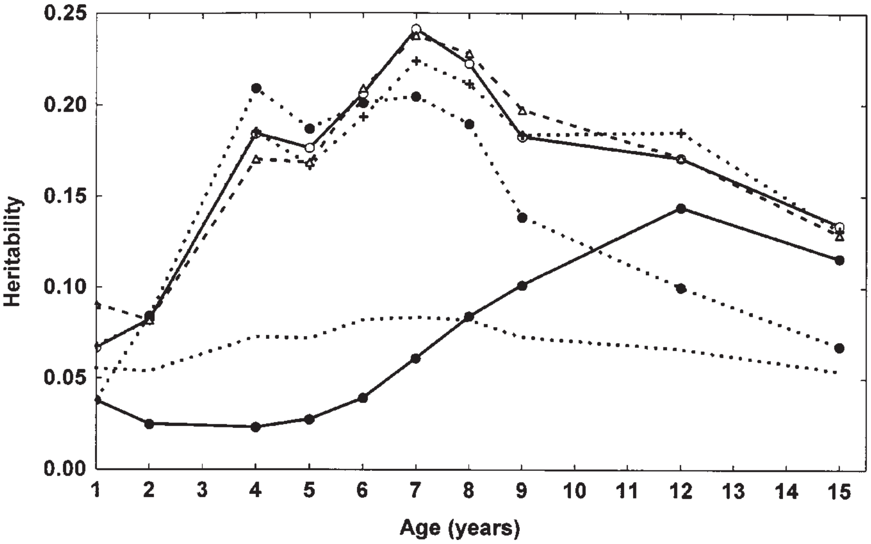
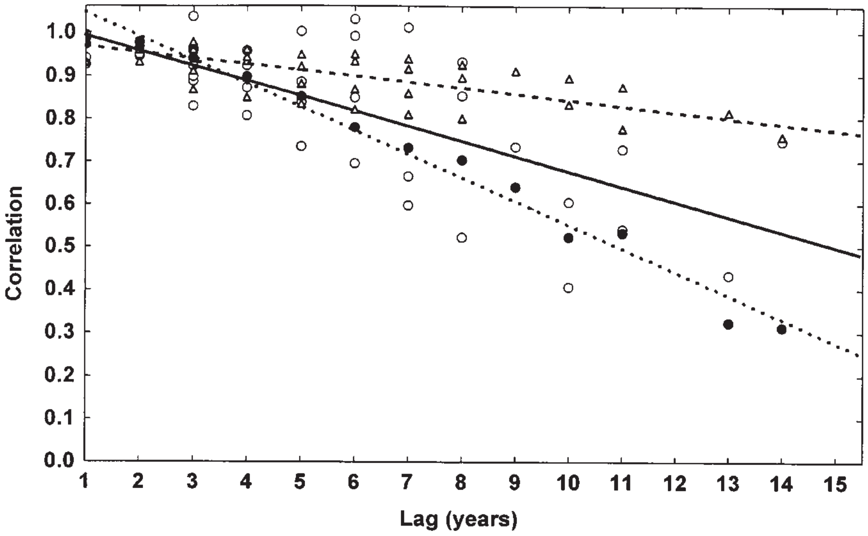


Fig. 4. Genetic correlations for different lags between measures for unstructured (US; —○—), banded correlations (BC; ···●···), and autoregressive (ARlog; —Δ—) models considering autoregressive plot errors.



of sampling variation when there are small numbers of families. However, its correlation values help us interpret the other models, since they are the least constrained values. The US model gave very high correlations of age 1 with ages 2–8 as well as high correlations between successive years, higher than expected under an AR model if the AR parameter was estimated only from more distant ages. Except for the correlations of age 1 noted above, the correlations tended to be lower between younger ages than between higher ages. This is not unexpected when looking at correlations, because new growth each year is a decreasing proportion of the height at the beginning of the year.

Genetic correlations for the ARlog model are shown in Table 3. The ARlog model essentially estimates one autocorrelation coefficient and projects the value to higher lags. The projected correlations are higher than the median of the US model values. The high lag correlations from the BC model agree better with the US model values than do the projected ARlog correlations. The US model regression is strongly influenced by the high correlations at all lags associated with age 1. Doing all combinations of bivariate analysis would produce very similar (nonpositive definite) results. While the US structure puts no constraints on the covariances, the AR and BC structures impose strong constraints, generating more parsimonious models. The variation observed in the US correlations at a given lag represents sampling effects and (or) is the result of growth patterns.

The preferred model (ARlog) has an intuitively appealing structure, where the breeding value of tree *i* observed at time *j* (a_{ij}) is a function of genes acting at time $j - 1$ (a_{ij-1}) plus genes acting on the new measurement (α_j), thus $a_{ij} = \rho^{|\log(j)-\log(j-1)|} a_{ij-1} + \alpha_j$. However, it is necessary to confirm if it is an adequate description of the pattern present in the US correlation structure. The main question posed by the

ARlog model is whether the genetic correlation between ages at higher lag is really so high. As an example, estimates of genetic correlation with the AR model are similar to those obtained for volume by Apiolaza et al. (1994) until age 9 but overestimate the values for older ages. However, the values are substantially different from the estimates based on cloned material by Burdon et al. (1992). This issue can only be resolved from a much larger experiment. In spite of this, the ARlog model does provide the highest correlations between the ranking of breeding values at different ages and the phenotypic ranking at age 15.

The implications of the high correlation at low lag is that measurement of height in consecutive years contributes little. However, after say 5 years according to the US and BC models, the correlation has dropped sufficiently to justify remeasurement. The high correlation of age 1 with later ages in the US model (see Fig. 4), if real, suggests that early growth is a good predictor of the genetic component of later height. However, the low heritability at age 1 would mitigate against depending on this trait, especially since these could be carry-over nursery effects and not genetic effects at all. Further, the high correlations of age 1 with later ages under the US model suggest that, whatever these early differences are, they do persist and are scaled up as the tree grows.

The heritability pattern under the RR model (see Fig. 3) did not follow the pattern of the previous models. The correlation pattern is also quite different to the other models with lag 1 correlations ranging from 0.71 at young ages to 0.99 at older ages. One potential explanation was that changes of scale dominated the model, and while the other four models had separate variance parameters for each age, the RR model had only a function over time. However, a Box–Cox data transformation did not improve the fitting greatly. The changes in heritability are related to the function used to model $\mathbf{G}_0$. Random regressions model the trajectory of breeding values, which deviate from other fixed and random effects included in the model. Hence, a simple polynomial (in this case a third-order one) may not be enough to model those deviations. The use of more flexible functions, e.g., higher order polynomials or cubic splines, suggested when there is no previous knowledge about the underlying biological model (e.g., Verbyla et al. 1997; White et al. 1999), might improve the estimation. Other examples of RR poorly reconstructing the $\mathbf{G}_0$ matrix obtained from a US multivariate analysis are in Van der Werf and Schaeffer (1997) and Van der Werf et al. (1998). One problem is that the polynomials are by nature highly correlated.

Number and distribution of measures

Often the frequency of measurement of a progeny test depends more on budget restrictions than on genetic considerations. Covariance structure modelling may not be possible when only a few measures are available. On the other hand, using 6 measures (1, 4, 6, 9, 12, and 15 years) produced similar results to those obtained using 10 measures (details not shown).

For most longitudinal variance models, equidistant measures are easier to analyse. The presence of unequal intervals involves either the manual specification of the bands (BC) or the use of a distance-based power model (AR). The US model is less likely to converge to a positive definite solution as the number of measures increases or when they are highly correlated. Close measures (as in our data set) result in highly correlated traits increasing the risk that the result might be nonpositive definite and convergence more difficult (Gilmour 1999).

The process of running the UC analysis (equivalent to univariate analyses by age) was very straightforward from both a modelling and computational perspective. Running the US analysis was less straightforward because of the high correlations between measures (traits). The estimated $\mathbf{G}_0$ was close to singular and the algorithm failed to converge from the naïve starting values initially supplied. Using the results from the BC model as starting values, the algorithm converged to the negative definite solution reported for the US model.

The RR model was quite difficult to fit, also requiring a multistage process to achieve convergence. We first fitted a RR model with intercept and slope, then added the quadratic term and finally the cubic term. The Log L values were 7063.71, 7069.12, and 7078.29, respectively. However, the variance matrix for the final model was negative definite and only converged after fixing the variance for the quadratic component. The correlations among the components were very high despite the definition of $\mathbf{Q}_i$ using orthogonal polynomials. An earlier attempt using starting values derived from the matrix $\mathbf{Q}_i'\mathbf{G}_0\mathbf{Q}_i$, where $\mathbf{G}_0$ is the additive genetic variance matrix from fitting the US model and $\mathbf{Q}_i$ is the 10 × 4 matrix of orthogonal polynomial coefficients also failed to converge. It was based on the assumption that $\mathbf{G}_0$ is an estimate of the matrix we want to approximate with a structure $\mathbf{Q}_i\Lambda_0\mathbf{Q}_i'$, where Λ_0 is the variance matrix for the random regression coefficients and noting that $\mathbf{Q}_i'\mathbf{Q}_i$ is $\mathbf{I}$. That is, if $\mathbf{G}_0 = \mathbf{Q}_i\Lambda_0\mathbf{Q}_i'$ then $\mathbf{Q}_i'\mathbf{G}_0\mathbf{Q}_i = \mathbf{Q}_i'\mathbf{Q}_i\Lambda_0\mathbf{Q}_i'\mathbf{Q}_i = \Lambda_0$. We did not try using other polynomials that might change the convergence behaviour.

Further considerations

The use of a multivariate approach takes into account nonrandom missing observations caused by mortality, thinning, or sampling of trees, which can bias parameter estimates (e.g., Apiolaza et al. 1998). Nonetheless nonpositive definite matrices, frequently found in forest genetics literature, may still be an issue. "Bending" (Hayes and Hill 1981) or other techniques for restricting genetic parameter matrices to the parameter space may still be necessary, especially if the US structure is used and the result is negative definite as here. Where there are large scale effects as in this data set, it may be preferable to bend the correlation matrix rather than the covariance matrix unless the data is transformed to stabilize the variance. The log likelihood for the bent US model (bent and fixed $\mathbf{G}_0$, other parameters iterated to convergence) reduces to 6854.05, which is lower than the results for the ARlog and BC models, confirming the adequacy of using simplified covariance models. The best way to reduce the chance of getting a nonpositive definite result is to increase the number of families sampled, reducing the sampling error of the variance components.

An alternative approach to the analysis of highly correlated observations, although beyond the scope of this paper, is the use of canonical transformation. This technique creates independent traits that are analyzed separately, and the results are transformed back to obtain the full parameter matrix with the original traits (measures) (Jensen and Mao 1988). It has restrictions on the number of random effects used in the model and the pattern of missing observations (Lin and Smith 1990; Ducrocq and Besbes 1993).

Conclusions

The use of structured covariance matrices for longitudinal data constrains the correlations to a pattern dependent on the form of the model, potentially smoothing the estimates of heritability and genetic correlation. It also facilitates model fitting and convergence of the likelihood-maximization algorithm. Models that take into account the ordering implicit in successive measures are preferred to the unstructured covariance model when assessing the changes of likelihood relative to the number of parameters. The results presented in this paper suggest that the ARlog model reproduced the results from the US model well enough, while simplifying the analysis; therefore, the US covariance structure is probably not the "best" model for longitudinal data. Equally important is that, assuming AIC is an appropriate model selection criterion, small data sets might not provide enough information to discriminate between some of the models (e.g., RR). On the other hand, if the data set is appropriate AIC appears to be insensitive to substantial differences of genetic parameters.

The results from this study are from a small sample (45 families with up to 40 trees and 10 longitudinal measures of 1 trait, for a total of 15 260 records), but they provide a good starting point for analyses involving larger data sets. Once $\mathbf{G}_0$ and $\mathbf{R}_0$ are estimated, covariance functions can be easily developed (e.g following Kirkpatrick's methodology) allowing for a more detailed study of early selection procedures. Further research might contemplate the use of several measurements to improve early prediction of breeding values (currently under preparation by the authors of this study) and modelling the simultaneous change of other growth traits and wood properties, using multitrait models (Van der Werf et al. 1998).

Although recent literature has emphasized the use of random regression for the analysis of longitudinal data (e.g., Jamrozik and Schaeffer 1997; Jamrozik et al. 1997; Meyer 1998; Van der Werf et al. 1998; White et al. 1999), random regressions are not necessarily suitable for all data sets. Given the potential reduction of number of parameters, other functional relationships (e.g., growth models) that could be used with random regressions within a linear model framework to model $\mathbf{G}_0$ should be studied.

Finally, the "best" parameterizations can be trait specific, i.e., different traits may require different structures (e.g., Coelli et al. 1998). Thus, it is necessary to identify the most appropriate model for each trait considered in a breeding programme. Nevertheless, simple models like AR seem to be flexible enough to hold in many common tree breeding situations.

Acknowledgements

This paper arose from discussions with Nuno Borralho and Sue Jarvis during the Sixth World Congress of Genetics Applied to Livestock Production, Jan. 1998, Armidale, Australia. Their encouragement is much appreciated. Many thanks to Claudio Balocchi and the breeding team of Bosques Arauco S.A., for measuring the trees over such a long period and kindly providing the data set. Valuable comments by Rowland Burdon, Mark Dieters, Greg Dutkowski, Igor Elsabio, Tore Ericsson, and Christian de Veer and by the reviewers Nuno Borralho, Gunnar Jansson, Steen Magnussen, and Per Ståhl greatly improved this paper. L.A.A. was supported with NZODA (New Zealand Overseas Development Assistance) and New Zealand Forest Research Institute scholarships during the completion of this study.

References

Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Automat. Control, **19**: 716–723.

Apiolaza, L.A., White, T.L., and Hodge, G.R. 1994. Análisis genético de *Pinus radiata* en la CMG: estimación de heredabilidad, interacción genotipo-ambiente, correlación edad-edad y correlación entre características en los ensayos de progenie de polinización abierta. School of Forest Resources and Conservation, University of Florida, Gainesville.

Apiolaza, L.A., Burdon, R.D., and Garrick, D.J. 1998. Effects of sampling on open-pollinated bivariate progeny tests. *In* Proceedings of the 6th World Congress of Genetics Applied to Livestock Production, 11–16 Jan. 1998, Armidale, New South Wales, Australia. Vol. 27. pp. 491–494.

Balocchi, C.E., Bridgewater, F.E., Zobel, B.J., and Jahromi, S. 1993. Age trends in genetic parameters for tree height in a nonselected population of loblolly pine. For. Sci. **39**: 231–251.

Box, G.E.P., and Cox, D.R. 1964. An analysis of transformations. J. R. Stat. Soc. B, **26**: 211–243.

Burdon, R.D. 1977. Genetic correlation as a concept for studying genotype–environment interaction in forest tree breeding. Silvae Genet. **26**: 168–175.

Burdon, R.D., Bannister, M.H., and Low, C.B. 1992. Genetic survey of *Pinus radiata*. 5: between-trait and age-age correlations for growth rate, morphology, and disease resistance. N.Z. J. For. Sci. **22**: 211–227.

Carlin, B.P., and Louis, T.A. 1996. Bayes and empirical Bayes methods for data analysis. Chapman & Hall, London.

Chi, E.M., and Reinsel, G.C. 1989. Models for longitudinal data with random effects and AR(1) errors. J. Am. Stat. Assoc. **84**: 452–459.

Cnaan, A., Laird, N.M., and Slasor, P. 1997. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. Stat. Med. **16**: 2349–2380.

Coelli, K.A., Gilmour, A.R., and Atkins, K.D. 1998. Comparison of genetic covariance models for annual measurements of fleece weight and fibre diameter. *In* Proceedings of the 6th World Congress of Genetics Applied to Livestock Production, 11–16 Jan. 1998, Armidale, New South Wales, Australia. Vol. 24. pp. 31–34.

Diggle, P.J., Liang, K.-Y., and Zeger, S.L. 1994. Analysis of longitudinal data. Clarendon Press, Oxford.

Ducrocq, V., and Besbes, B. 1993. Solution of multiple trait animal models with missing data on some traits. J. Anim. Breed. Genet. **110**: 81–92.

Foster, G. 1986. Trends in genetic parameters with stand development and their influence on early selection for volume growth in loblolly pine. For. Sci. **32**: 944–959.

Gilmour, A.R. 1999. Variance structures available in ASREML. *In* Proceedings of the 13th Conference of the Association for the Advancement of Animal Breeding and Genetics, 4–7 July 1999, Mandurah, Western Australia, Australia. **13**: 416–419.

Gilmour, A.R., Thompson, R., and Cullis, B.R. 1995. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. Biometrics, **51**: 1440–1450.

Gilmour, A.R., Cullis, B.R., Welham, S.J., and Thompson, R. 1998. ASReml users' manual. New South Wales Agriculture, Orange, Australia.

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

Hand, D., and Crowder, M. 1996. Practical longitudinal data analysis. Chapman & Hall, London.

Hayes, J.F., and Hill, W.G. 1981. Modification of estimates of parameters in the construction of selection indices ('bending'). Biometrics, **37**: 483–493.

Hodge, G.R., and White, T.L. 1992. Genetic parameter estimates for growth traits at different ages in slash pine and some implications for breeding. Silvae Genet. **41**: 252–262.

Jamrozik, J., and Schaeffer, L.R. 1997. Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. J. Dairy Sci. **80**: 762–770.

Jamrozik, J., Kistemaker, G.J., Dekkers, J.C.M., and Schaeffer, L.R. 1997. Comparison of possible covariates for use in random regression model for analyses of test day yields. J. Dairy Sci. **80**: 2550–2556.

Jensen, J., and Mao, I.L. 1988. Transformation algorithms in analysis of single trait and of multitrait models with equal design matrices and one random factor per trait: a review. J. Anim. Sci. **66**: 2750–2761.

Jones, R.H. 1990. Serial correlation or random subject effects? Commun. Stat. Simul. Comput. B, **19**: 1105–1123.

Jones, R.H. 1993. Longitudinal data with serial correlation: a state-space approach. Chapman & Hall, London.

Kirkpatrick, M., Lofsvold, D., and Bulmer, M. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics, **124**: 979–993.

Kirkpatrick, M., Hill, W.G., and Thompson, R. 1994. Estimating the covariance structure of traits during growth and ageing, illustrated with lactation in dairy cattle. Genet. Res. **64**: 57–69.

Kremer, A. 1992. Predictions of age-age correlations of total height based on serial correlations between height increments in maritime pine (*Pinus pinaster* Ait.). Theor. Appl. Genet. **85**: 152–158.

Laird, N.M., and Ware, J.H. 1982. Random-effects models for longitudinal data. Biometrics, **38**: 963–974.

Lin, C.Y., and Smith, S.P. 1990. Transformation of multitrait to unitrait mixed model analysis of data with multiple random effects. J. Dairy Sci. **73**: 2494–2502.

Louis, T.A. 1988. General methods for analysing repeated measures. Stat. Med. **7**: 29–45.

Meyer, K. 1998. Modeling 'repeated' records: covariance functions and random regression models to analyse animal breeding data. *In* Proceedings of the 6th World Congress of Genetics Applied to Livestock Production, 11–16 Jan. 1998, Armidale, New South Wales, Australia. Vol. 25. pp. 517–520.

Newman, D.H., and Williams, C.G. 1991. The incorporation of risk in optimal selection age determination. For. Sci. **37**: 1350–1364.

Patterson, H.D., and Thompson, R. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika, **58**: 545–554.

Quaas, R.L., Anderson, R.D., and Gilmour, A.R. 1984. BLUP school handbook. 5–7 Feb. 1984, Animal Genetics and Breeding Unit, University of New England, Armidale, New South Wales, Australia.

Rao, C.R. 1965. The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. Biometrika, **52**: 447–458.

Schaeffer, L.R., and Dekkers, J.C.M. 1994. Random regressions in animal models for test-day production in dairy cattle. *In* Proceedings of the 5th World Congress of Genetics Applied to Livestock Production, Aug. 1994, Guelph, Ont. Vol. XVIII. pp. 443–446.

Searle, S.R. 1982. Matrix algebra useful for statistics. John Wiley & Sons, Inc. New York.

Thompson, R., and Meyer, K. 1986. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. Lives. Prod. Sci. **15**: 299–313.

Van der Werf, J.H.J., and Schaeffer, L.R. 1997. Random regression in animal breeding. Course notes. 25–28 June 1997, Centre for the Genetic Improvement of Livestock, Guelph, Ont.

Van der Werf, J.H.J., Goddard, M.E., and Meyer, K. 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. J. Dairy Sci. **81**: 3300–3308.

Verbyla, A.P., Cullis, B.R., Kenward, M.G., and Welham, S.J. 1997. The analysis of designed experiments and longitudinal data using smoothing splines. Department of Statistics, The University of Adelaide, Adelaide, Australia. Res. Rep. No. 97/4.

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

Wada, Y., and Kashiwagi, N. 1990. Selecting statistical models with information statistics. J. Dairy Sci. **73**: 3575–3582.

Wei, X., and Borralho, N.M.G. 1996. A simple model to describe age trends in heritability in short rotation tree species. *In* Proceedings, Tree Improvement for Sustainable Tropical Forestry, 27 Oct. – 1 Nov. 1996, Caloundra, Queensland, Australia.

Wei, X., and Borralho, N.M.G. 1998. Use of individual tree mixed models to account for mortality and selective thinning when estimating base population genetic parameters. For. Sci. **44**: 246–253.

White, I.M.S., Thompson, R., and Brotherstone, S. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. J. Dairy Sci. **82**: 632–638.