Bayesian inference for normal multiple-trait individual-tree models with missing records via full conjugate Gibbs

Eduardo P. Cappa and Rodolfo J.C. Cantet

Abstract: In forest genetics, restricted maximum likelihood (REML) estimation of (co)variance components from normal multiple-trait individual-tree models is affected by the absence of observations in any trait and individual. Missing records affect the form of the distribution of REML estimates of genetics parameters, or of functions of them, and the estimating equations are computationally involved when several traits are analysed. An alternative to REML estimation is a fully Bayesian approach through Markov chain Monte Carlo. The present research describes the use of the full conjugate Gibbs algorithm proposed by Cantet et al. (R.J.C. Cantet, A.N. Birchmeier, and J.P. Steibel. 2004. Genet. Sel. Evol. **36**: 49–64) to estimate (co)variance components in multiple-trait individual-tree models. This algorithm converges faster to the marginal posterior densities of the parameters than regular data augmentation from multivariate normal data with missing records. An expression to calculate the deviance information criterion for the selection of linear parameters in normal multiple-trait models is also given. The developments are illustrated by means of data from different crosses of two species of *Pinus*.

Résumé : En génétique forestière, la méthode d'estimation par le maximum de vraisemblance restreinte (REML) des composantes de la variance et de la covariance à partir de modèles normaux à caractères multiples d'arbres individuels est influencée par les observations manquantes pour un caractère ou un individu. Les données manquantes influencent la forme de la distribution des estimations par REML des paramètres génétiques ou des fonctions mathématiques qui les représentent. De plus, les équations d'estimation sont aussi impliquées dans le calcul lorsque plusieurs caractères sont analysés. Une approche bayésienne complète recourrant aux méthodes de Monte Carlo par chaînes de Markov constitue une alternative à la méthode d'estimation par REML. Les auteurs décrivent une utilisation de l'algorithme de Gibbs dans sa version complètement conjuguée, tel que Cantet et al. (R.J.C. Cantet, A.N. Birchmeier et J.P. Steibel. 2004. Genet. Sel. Evol. **36** : 49–64) l'ont proposé, pour estimer les composantes de la variance et de la covariance pour des modèles à caractères multiples d'arbres individuels. Cet algorithme converge plus rapidement vers les densités marginales a posteriori des paramètres que la méthode traditionnelle d'augmentation des données à partir des données normales multivariées comprenant les observations manquantes. Les auteurs fournissent également une équation permettant de calculer le critère d'information de la déviance pour la sélection des paramètres linéaires dans les modèles normaux à caractères multiples. Les développements mathématiques sont illustrés à partir des données provenant de différents croisements chez deux espèces du genre *Pinus*.

[Traduit par la Rédaction]

Introduction

Genetic evaluation in forest trees is usually performed with data from progeny tests on full or half-sib families. The individual-tree mixed model, introduced in forest genetics by Borralho (1995), appropriately takes into account additive relationships, especially for multiple-trait data where (co)variance components are the parameters to estimate. Restricted maximum likelihood (REML; Patterson and Thompson 1971) is frequently employed by forest breeders (e.g., Huber et al. 1994; Dieters et al. 1995) to estimate those parameters. From a Bayesian standpoint, REML is seen as the mode of a joint posterior distribution of all (co)variance components, with noninformative prior densities, once the fixed effects are marginalized by translation invariance functions of the data (Harville 1974). However, there may be differences between estimates obtained from joint or marginal modes if the model contains several parameters and the amount of information present in the data differs among parameters (Sorensen and Gianola 2002). In addition, REML relies on asymptotic theory to obtain a measure of the precision of the estimates, since the estimating equations have to be solved iteratively so that sampling distributions of the estimators are impossible to obtain. An alternative to REML estimation is a full Bayesian approach through Markov chain Monte Carlo (MCMC) methods.

During the last decade, the contribution of Bayesian theory to statistical analyses in most scientific fields has in-

Received 28 January 2005. Resubmitted 31 August 2005. Accepted 13 January 2006. Published on the NRC Research Press Web site at http://cjfr.nrc.ca on 2 May 2006.

E.P. Cappa.[1] Department of Animal Production, University of Buenos Aires, Avenida San Martín 4453, C1417DSQ Buenos Aires, Argentina.
R.J.C. Cantet. Department of Animal Production, University of Buenos Aires – "Consejo Nacional de Investigaciones Científicas y Técnicas" (CONICET), Avenida San Martín 4453, C1417DSQ Buenos Aires, Argentina.

[1]Corresponding author (e-mail: ecappa@mail.agro.uba.ar).

 doi:10.1139/X06-024

creased tremendously because of the feasibility of doing posterior inference by means of MCMC algorithms. These methods allow marginal inferences on each individual parameter and produce measures of precision of the estimators through posterior variances or posterior standard errors (Sorensen and Gianola 2002). In addition, the distributions of complex, meaningful genetic parameters, such as heritabilities and genetic correlations, which are functions of the (co)variance components, are obtained as byproducts of the MCMC sampling scheme. Note that in multiple-trait models with additive relationship matrices and several additive and environmental covariance components, there is no frequentist counterpart to a posterior distribution: there are no small sampling distributions for (co)variance parameters (or functions of them). After the lead of D. Gianola and coworkers, animal breeders have used MCMC techniques such as the Gibbs sampling to estimate (co)variance components since 1994 (Sorensen et al. 1994; Wang et al. 1994). More recently, forest geneticists have become acquainted with Bayesian inference using MCMC algorithms (Soria et al. 1998; Gwaze and Woolliams 2001; Zeng et al. 2004). Soria et al. (1998) applied Gibbs sampling to the genetic analysis of growth from 260 families of *Eucalyptus globulus* and local genetic material using two-trait individual-tree models. Gwaze and Woolliams (2001) employed Gibbs sampling for a decision-making process on the choice of site for locating progeny tests. Finally, Zeng et al. (2004) used a Gibbs block sampler to make inferences about major genes and polygenic effects in a population derived from a half-diallel mating design.

Frequently, multiple-trait estimation of genetic parameters in forest genetics is affected by missing observations due death or damage of trees or practical and technical problems of data recollection. Cost considerations or operational problems with measuring certain characteristics result in a fewer number of records for those traits. Examples of expensive traits are those related to wood quality (Apiolaza et al. 1999) or to "branching" (Shepherd et al. 2002). An example of an operational problem that results in fewer records is the report of Dungey (2000), in which height measures were taken only in the first two trees from each plot. The statistical approach commonly used for analysing multiple-trait models is the "missing data" theory, as long as the subsampling induces a "missing at random" process (Rubin 1976). The estimation of heritabilities and genetic correlations in multiple-trait models with missing or subsampled data is a complex statistical problem, even for Bayesian methods using MCMC techniques. Thus, the data augmentation algorithm first employed by Van Tassell and Van Vleck (1996) tends to be very slow to converge because of the MCMC chain being strongly autocorrelated. The reason for this correlation is that the sampling of the "missing data" (error terms for nonobserved data) and of the (co)variance components depend on each other. To lessen the effect of this correlation, Cantet et al. (2004) proposed an MCMC method for estimating covariance matrices of error effects in multiple-trait normal models: the full conjugate Gibbs (FCG) algorithm. This method reaches faster convergence than do the data augmentation procedures of Van Tassell and Van Vleck (1996) by reducing the correlation between sampled missing errors and their covariance matrix. This is achieved by sampling missing patterns rather than individual missing errors, as will be explained in the next section.

Although in individual-tree models the genetic random effects are well defined, there may be competing classification effects and covariates (either fixed or random) that can provide adequate fit. Therefore, a model selection process is necessary prior to predicting breeding values. For example, when analysing data from purebred and crossbred progeny, the model equation may include terms for mean additive (A), dominance (D), and epistatic (A × A, A × D, D × D) effects (Hill 1982). These parameters are estimated as covariates from the data, and each genotype is a linear combination of them. Also, there may be different ways to block the data or different environmental covariates to include in the model. Spiegelhalter et al. (2002) proposed a Bayesian statistic for model selection that is viewed as the counterpart of the Akaike information criterion: the deviance information criterion (DIC). The DIC is composed of a measure of total fit and a penalization of the complexity of the model.

The goal of this paper is twofold: (1) to apply the FCG algorithm proposed by Cantet et al. (2004) to estimate (co)variance components, or functions of them, from multiple-trait individual-tree models with missing records; (2) to obtain an expression to calculate DIC for model selection in multiple-trait individual-tree models. Developments are illustrated by means of data from different crosses of *Pinus elliottii* var. *elliottii* Engelm. (E) and *Pinus caribaea* var. *hondurensis* (Sénécl) Barrett et Golfari (H).

FCG sampling

Statistical model

Suppose data are collected on q trees scored for r continuous traits and, although desirable, not all individuals have measures in all traits. Individual trees are arranged in groups such that each group has a different mean. By group we mean environmental systematic effects, such as blocks nested within trials, or trials, or any other blocking effect. Let the subscript j index the traits ($j = 1, 2, \ldots, r$), i the trees ($i = 1, 2, \ldots, q$), and l the groups ($l = 1, 2, \ldots, g$). Let y_{ijl} and a_{ijl} be the phenotype and the breeding value, respectively, of individual i for trait j, scored in group l. A mixed model to analyze the data is

$$[1] \qquad y_{ijl} = \boldsymbol{X}'_{ij}\boldsymbol{\beta}_j + a_{ijl} + e_{ijl}$$

where $\boldsymbol{\beta}_j$ is a $p \times 1$ vector of parameters related to trait j and involving the means of the groups, $\boldsymbol{X}_{ij}$ is a column vector that relates y_{ijl} to the elements of $\boldsymbol{\beta}_j$, and e_{ijl} is the error term. The data are ordered such that traits are nested within trees. Thus, by letting $\boldsymbol{y} = [y_{11l}, y_{12l}, \ldots, y_{qrl}]$, we can write the following multiple-trait individual-tree model:

$$[2] \qquad \begin{bmatrix} \boldsymbol{y}_i \\ \cdot \\ \boldsymbol{y}_r \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \cdot & \boldsymbol{0} \\ \cdot & \cdot & \cdot \\ \boldsymbol{0} & \cdot & \mathbf{X}_r \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \cdot \\ \boldsymbol{\beta}_r \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \cdot & \boldsymbol{0} \\ \cdot & \cdot & \cdot \\ \boldsymbol{0} & \cdot & \mathbf{Z}_r \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \cdot \\ \boldsymbol{a}_r \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_1 \\ \cdot \\ \boldsymbol{e}_r \end{bmatrix}$$

or more compactly as

$$[3] \qquad \boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\,\boldsymbol{a} + \boldsymbol{e}$$

Regardless of whether an individual has measures on all r traits, the breeding values of all individuals for all traits are included in $[\boldsymbol{a}_1'|\ldots|\ \boldsymbol{a}_r'] = \boldsymbol{a}$. This vector has zero expectation and a covariance matrix equal to

$$[4]\qquad \text{Var}\begin{bmatrix}\boldsymbol{a}_1\\ \boldsymbol{a}_2\\ \vdots\\ \boldsymbol{a}_r\end{bmatrix} = \begin{bmatrix} g_{1,1}\mathbf{A} & g_{1,2}\mathbf{A} & \cdot & \cdot & \cdot & g_{1,r}\mathbf{A}\\ g_{2,1}\mathbf{A} & g_{2,2}\mathbf{A} & \cdot & \cdot & \cdot & g_{2,r}\mathbf{A}\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & g_{i,j}\mathbf{A} & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ g_{r,1}\mathbf{A} & g_{r,2}\mathbf{A} & \cdot & \cdot & \cdot & g_{r,r}\mathbf{A}\end{bmatrix}$$

$$= \begin{bmatrix} g_{1,1} & g_{1,2} & \cdot & \cdot & \cdot & g_{1,r}\\ g_{2,1} & g_{2,2} & \cdot & \cdot & \cdot & g_{2,r}\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & g_{i,j} & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ g_{r,1} & g_{r,2} & \cdot & \cdot & \cdot & g_{r,r}\end{bmatrix} \otimes \mathbf{A}$$

$$= \mathbf{G}_0 \otimes \mathbf{A}$$

where $g_{jj'}$ is the additive genetic covariance between traits j and j' if $j \neq j'$ and the additive variance of trait j otherwise. The square matrix $\mathbf{A}$ is of order $q \times q$ and contains the additive relationships (Henderson 1984) among all trees: parents without records plus offspring with data in $\boldsymbol{y}$.

For any given tree, missing records on one or more traits induce a pattern of missing data that affects the distribution of the error terms. To define this distribution, order the data by trait within tree. This allows missing data patterns to be accommodated by an indicator matrix $\mathbf{M}_k$ (Dominici et al. 2000; Cantet et al. 2004) having r_k rows and r columns, with $k = 1, 2, \ldots, K$, where K is the number of patterns of missing data in the data set. For example, suppose $r = 3$. Then, the data from trees with all traits recorded is represented as $\mathbf{M}_k = \mathbf{I}_r$, but if the recorded traits are just 1 and 3, we have $\mathbf{M}_k = \begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}$. It will be assumed that the complete pattern is observed in at least r trees, and n_k denotes the number of trees with records in pattern k. To obtain a matrix formulation of the distribution of $\boldsymbol{e}$, the error terms are ordered within the pattern, so that the final order is traits within tree within pattern of missing data. Then, the expected value of $\boldsymbol{e}$ is zero and its covariance matrix is equal to

$$[5]\qquad \text{Var}(\boldsymbol{e}) = \mathbf{R} = \begin{bmatrix} \mathbf{I}_{n_1}\otimes\mathbf{M}_1\mathbf{R}_0\mathbf{M}_1' & 0 & \cdot & 0\\ 0 & \mathbf{I}_{n_2}\otimes\mathbf{M}_2\mathbf{R}_0\mathbf{M}_2' & \cdot & 0\\ \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\\ 0 & 0 & \cdot & \mathbf{I}_{n_K}\otimes\mathbf{M}_K\mathbf{R}_0\mathbf{M}_K'\end{bmatrix}$$

In [5], $\mathbf{R}_0 = [\boldsymbol{r}_{jj'}]$, where $\boldsymbol{r}_{jj'}$ is the environmental (co)variance between traits j and j'. Vectors $\boldsymbol{a}$ and $\boldsymbol{e}$ are independent and normally distributed.

The parameters of the covariance matrix of $\boldsymbol{y}$ are the elements of the covariance matrices $\mathbf{G}_0$ and $\mathbf{R}_0$ of additive genetic and error effects, respectively. These dispersion parameters are estimated by a Bayesian approach using the FCG algorithm proposed by Dominici et al. (2000), which was adapted to a multiple-trait setting by Cantet et al. (2004). The algorithm can handle different patterns of observed and "missing" traits more efficiently than can "data augmentation" (Van Tassell and Van Vleck 1996). We now describe the priors for all parameters and the likelihood of the data, and then explain the FCG algorithm for multiple traits. In doing so, we follow closely Cantet et al. (2004).

Specification of prior distributions and likelihood

In a conjugate approach the prior densities for all parameters are chosen to be closed under sampling (Robert and Casella 1999, p. 31), which means that both the prior and posterior densities belong to the same family of distributions. To reflect a prior state of uncertainty for the fixed effects in a mixed linear model, while keeping the posterior distribution proper (Hobert and Casella 1996), β is taken to be $\boldsymbol{\beta} \sim N_p(\boldsymbol{0}, \mathbf{K})$, where p is the number of fixed effects. The matrix $\mathbf{K}$ is diagonal, with large elements ($k_{ii} > 10^8$; Cantet et al. 2004), and the prior density of $\boldsymbol{\beta}$ is then proportional to

$$[6]\qquad p(\boldsymbol{\beta}|\mathbf{K}) \propto \left|\prod_{i=1}^{p} k_{ii}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{p}\frac{\boldsymbol{\beta}_i^2}{k_{ii}}\right)$$

The vector of breeding values is distributed a priori as $\boldsymbol{a} \sim N_{rq}(\boldsymbol{0}, \mathbf{G}_0 \otimes \mathbf{A})$ (see [13.38] in Sorensen and Gianola 2002, p. 578), so that

$$[7]\qquad p(\boldsymbol{a}|\mathbf{G}_0, \mathbf{A}) \propto \left|\mathbf{G}_0\right|^{-\frac{q}{2}}\left|\mathbf{A}\right|^{-\frac{r}{2}} \exp\left[-\frac{1}{2}\boldsymbol{a}'(\mathbf{G}_0^{-1}\otimes\mathbf{A}^{-1})\boldsymbol{a}\right]$$

A priori the additive (co)variance matrix $\mathbf{G}_0$ follows an inverted Wishart density $\mathbf{G}_0 \sim \text{IW}(\mathbf{G}_0^*, n_A)$, where $\mathbf{G}_0^*$ is the hypercovariance and n_A is the degrees of belief (Sorensen and Gianola 2002, page 57) so that

$$[8]\qquad p(\mathbf{G}_0|\mathbf{G}_0^*, n_A) \propto \left|\mathbf{G}_0^*\right|^{\frac{n_A}{2}}\left|\mathbf{A}\right|^{-\frac{(n_A+r+1)}{2}} \exp\left[-\frac{1}{2}\text{tr}(\mathbf{G}_0^*\mathbf{G}_0^{-1})\right]$$

The covariance matrix for the error terms also follows an inverted Wishart prior: $\mathbf{R}_0 \sim \text{IW}(\mathbf{R}_0^*, v_k)$. The hyperparameters are the hypercovariance matrix $\mathbf{R}_0^*$ and the degrees of belief for the kth pattern v_k. To account for the patterns of missing data, the prior density depends on $\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_K$ such that it is written as

$$[9]\qquad p(\mathbf{R}_0|\mathbf{R}_0^*,\mathbf{M}_1,\ldots,\mathbf{M}_K,\nu_k)\propto\prod_{k=1}^{K}|\mathbf{M}_k\mathbf{R}_0\mathbf{M}_k'|^{-\frac{(\nu_k+2r_k+1)}{2}}\times\exp\left\{-\frac{1}{2}\operatorname{tr}[\mathbf{M}_k\mathbf{R}_0^*\mathbf{M}_k'\,(\mathbf{M}_k\mathbf{R}_0\mathbf{M}_k')^{-1}]\right\}$$

In the Bayesian view of the mixed linear model (Sorensen and Gianola 2002) the likelihood of the data is proportional to

$$[10]\qquad p(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{a},\mathbf{R})\propto|\mathbf{R}|^{-\frac{1}{2}}\times\exp\left[-\frac{1}{2}(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{a})'\mathbf{R}^{-1}(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{a})\right]$$

Joint and conditional posterior densities

The joint posterior density, that is, the density of the parameters given the data and the prior information, is written as the product of the likelihood function [10] and the prior distributions [6], [7], [8], and [9], which results in:

$$\begin{aligned}[11]\qquad & p(\boldsymbol{\beta},\boldsymbol{a},\mathbf{G}_0,\mathbf{R}_0|\,\boldsymbol{y},\mathbf{M}_1,\ldots,\mathbf{M}_K)\propto\\ & |\mathbf{R}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{a})'\mathbf{R}^{-1}(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{a})\right]\\ & \times\exp\left(-\frac{1}{2}\sum_{i=1}^{p}\frac{\boldsymbol{\beta}_i^2}{k_{ij}}\right)\exp\left[-\frac{1}{2}\boldsymbol{a}'(\mathbf{G}_0^{-1}\otimes\mathbf{A}^{-1})\boldsymbol{a}\right]\\ & \times|\mathbf{G}_0|^{-\frac{(n_A+r+q+1)}{2}}\exp\left[-\frac{1}{2}\operatorname{tr}(\mathbf{G}_0^*\mathbf{G}_0^{-1})\right]\\ & \times\prod_{k=1}^{K}|\mathbf{M}_k\mathbf{R}_0\mathbf{M}_k'|^{-\frac{(\nu_k+2r_k+1)}{2}}\\ & \times\exp\left\{-\frac{1}{2}\operatorname{tr}[\mathbf{M}_k\mathbf{R}_0^*\mathbf{M}_k'(\mathbf{M}_k\mathbf{R}_0\mathbf{M}_k')^{-1}]\right\}\end{aligned}$$

To take advantage of all information in the data about any parameter, inferences about $\boldsymbol{\beta}$, $\boldsymbol{a}$, $\mathbf{G}_0$, and $\mathbf{R}_0$ are based on their respective marginal posterior densities. A useful property of these marginal distributions is their lack of dependence upon any particular value of the other parameters. Thus, each marginal density is obtained by integrating out the joint distribution [11] with respect to the parameters other than the one of interest. This is accomplished by using the MCMC procedure known as Gibbs sampling while taking advantage of the marginal conditional densities resulting from [10] that are feasible for sampling. For the multiple-trait model with individual additive effects, these conditional densities were obtained by Van Tassell and Van Vleck (1996), Sorensen and Gianola (2002), and Cantet et al. (2004). Thus, for the linear parameters in $\boldsymbol{\beta}$ and $\boldsymbol{a}$ the posterior conditional density is equal to

$$[12]\qquad \begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{a}\end{bmatrix}\frac{1}{2}\,\boldsymbol{y},\mathbf{G}_0,\mathbf{R}_0\sim N_{p+rq}\times\left(\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\boldsymbol{a}}\end{bmatrix},\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}+\mathbf{K}^{-1} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}_0^{-1}\otimes\mathbf{A}^{-1}\end{bmatrix}^{-1}\right)$$

The vectors $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{a}}$ in [12] are the solutions to the following system of equations:

$$[13]\qquad \begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}+\mathbf{K}^{-1} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}_0^{-1}\otimes\mathbf{A}^{-1}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\boldsymbol{a}}\end{bmatrix}=\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\boldsymbol{y}\\ \mathbf{Z}'\mathbf{R}^{-1}\boldsymbol{y}\end{bmatrix}$$

The coefficient matrix in [13] is symmetric, so that the ***L–D–L′*** decomposition (Golub and Van Loan 1983; p. 82) for sparse systems with the aid of special interface subroutines (Kincaid et al. 1982) was used to obtain solutions.

The posterior conditional density of the covariance matrix $\mathbf{G}_0$ is an inverted Wishart distribution, with scaling matrix $\mathbf{G}_0^*+\mathbf{S}$ and degrees of belief equal to $n_A+q+r+1$:

$$[14]\qquad p(\mathbf{G}_0|\,\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{a},\mathbf{R}_0\propto|\mathbf{G}_0|^{-\frac{(n_A+q+r+1)}{2}}\times\exp\left\{-\frac{1}{2}\operatorname{tr}[(\mathbf{G}_0^*+\mathbf{S})\mathbf{G}_0^{-1}]\right\}$$

where $\mathbf{S}$ is defined as

$$[15]\qquad \mathbf{S}=\begin{bmatrix}\boldsymbol{a}_1'\mathbf{A}^{-1}\boldsymbol{a}_1 & \boldsymbol{a}_1'\mathbf{A}^{-1}\boldsymbol{a}_2 & \cdot & \cdot & \boldsymbol{a}_1'\mathbf{A}^{-1}\boldsymbol{a}_r\\ \boldsymbol{a}_2'\mathbf{A}^{-1}\boldsymbol{a}_1 & \boldsymbol{a}_2'\mathbf{A}^{-1}\boldsymbol{a}_2 & \cdot & \cdot & \boldsymbol{a}_2'\mathbf{A}^{-1}\boldsymbol{a}_r\\ \cdot & & \boldsymbol{a}_i'\mathbf{A}^{-1}\boldsymbol{a}_j & \cdot & \cdot\\ \cdot & & \cdot & \cdot & \cdot\\ \boldsymbol{a}_r'\mathbf{A}^{-1}\boldsymbol{a}_1 & \boldsymbol{a}_r'\mathbf{A}^{-1}\boldsymbol{a}_2 & \cdot & \cdot & \boldsymbol{a}_r'\mathbf{A}^{-1}\boldsymbol{a}_{\mathrm{r}}\end{bmatrix}$$

Finally, the marginal posterior conditional density of $\mathbf{R}_0$ was obtained by Cantet et al. (2004), and it is also an inverted Wishart distribution:

$$[16]\qquad p(\mathbf{R}_0|\,\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{a},\mathbf{G}_0)\sim\mathrm{IW}\left[\sum_{k=1}^{K}\nu_k+n+(K-1)(r+1),\left(\sum_{k=1}^{K}\mathbf{R}_k^*\right)\right]$$

Its kernel is equal to

$$[17]\qquad p(\mathbf{R}_0|\,\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{a},\mathbf{G}_0)\propto\prod_{k=1}^{K}\left|\mathbf{M}_k\mathbf{R}_0\mathbf{M}_k'\right|^{-\frac{(\nu_k+n_k+2r_k+1)}{2}}\times\exp\left[-\frac{1}{2}\operatorname{tr}(\mathbf{M}_k\mathbf{R}_0\mathbf{M}_k'+\mathbf{E}_k)(\mathbf{M}_k\mathbf{R}_0\mathbf{M}_k')^{-1}\right]$$

Matrix $\mathbf{E}_k$ for pattern k is $r\times r$, and its elements are equal to $\boldsymbol{e}_{(k)i}'\boldsymbol{e}_{(k)j}$, that is, inner products between the error vectors for traits i and j in pattern k. As an example, suppose three traits are studied. Let pattern $k=1$ be the complete one,

Table 1. Distribution of observations on each trial, total number of trees with records, and means and standard deviations (SD) for each genotype and for both traits (DBH and height).

	DBH					Height				
	No. of observations					No. of observations				
Genotype[a]	Trial 1	Trial 2	Total	Mean	SD	Trial 1	Trial 2	Total	Mean	SD
E	15	48	63	18.51	3.45	15	47	62	11.04	2.30
F1: E × H	667	631	1298	22.88	4.60	666	615	1281	12.55	2.12
F1: H × E	15	0	15	22.65	4.15	14	0	14	11.08	3.30
$BC_{(E)}$	13	46	59	21.70	3.44	13	42	55	11.24	3.23
$BC_{(H)}$	37	43	80	22.01	4.33	37	38	75	11.70	3.67
F2	39	44	83	20.39	4.38	39	44	83	12.37	1.59
Total	786	812	1598	22.49	4.60	784	786	1570	12.57	2.31

[a]E, *Pinus elliottii* var. *elliottii*; H, *Pinus caribaea* var. *hondurensis*.

whereas in pattern $k = 2$ only traits 1 and 3 are recorded. Then, matrices $\mathbf{E}_1$ and $\mathbf{E}_2$ are, respectively,

$$\mathbf{E}_1 = \begin{bmatrix} \boldsymbol{e}'_{(1)1}\boldsymbol{e}_{(1)1} & \boldsymbol{e}'_{(1)1}\boldsymbol{e}_{(1)2} & \boldsymbol{e}'_{(1)1}\boldsymbol{e}_{(1)3} \\ \boldsymbol{e}'_{(1)2}\boldsymbol{e}_{(1)1} & \boldsymbol{e}'_{(1)2}\boldsymbol{e}_{(1)2} & \boldsymbol{e}'_{(1)2}\boldsymbol{e}_{(1)3} \\ \boldsymbol{e}'_{(1)3}\boldsymbol{e}_{(1)1} & \boldsymbol{e}'_{(1)3}\boldsymbol{e}_{(1)1} & \boldsymbol{e}'_{(1)3}\boldsymbol{e}_{(1)3} \end{bmatrix}$$

$$\mathbf{E}_2 = \begin{bmatrix} \boldsymbol{e}'_{(2)1}\boldsymbol{e}_{(2)1} & 0 & \boldsymbol{e}'_{(2)1}\boldsymbol{e}_{(2)3} \\ 0 & 0 & 0 \\ \boldsymbol{e}'_{(2)3}\boldsymbol{e}_{(2)1} & 0 & \boldsymbol{e}'_{(2)3}\boldsymbol{e}_{(2)3} \end{bmatrix}$$

The missing elements within each pattern (for example the zeros in $\mathbf{E}_2$) are sampled using properties of the inverted Wishart distribution (Cantet et al. 2004).

Sampling scheme from the FCG algorithm

Cantet at al. (2004) implemented the sampling scheme for the FCG algorithm in a multiple-trait model as follows:

(1) Build and solve [13];
(2) Sample $\boldsymbol{\beta}$ and $\boldsymbol{a}$ from [12];
(3) Calculate the residuals: $\boldsymbol{e} = \boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{a}$;
(4) Sample the unconditional variance matrices among the missing traits in each pattern (zero elements of the $\mathbf{E}_k$ matrices) and add them up to calculate the hyper-covariance matrix for $\mathbf{R}_0$, as described by Cantet et al. (2004) in detail;
(5) Sample $\mathbf{R}_0$ from [16];
(6) Calculate $\mathbf{S}$ in [15];
(7) Sample $\mathbf{G}_0$ from [14], and go back to step 1.

Model comparison using the DIC

In the previous section, Bayesian estimation in a multiple-trait individual-tree model via the FCG algorithm is presented conditional on a particular model M. However, there may be several competitive mixed models to be used in the analysis. The issue is especially relevant when analyzing data from different genotypes and crosses and the researcher is interested in testing the inclusion of mean dominance, or mean additive × additive effects, or both. In this section, we adapt the Bayesian method of DIC comparison (Spiegelhalter et al. 2002), to multiple-trait genetic-tree models with missing data. All calculations are obtained with little numerical effort beyond that required to implement the FCG algorithm. The DIC combines a measure of model fit (the posterior mean deviance, $\overline{D}(\boldsymbol{\theta}_M)$) with a measure of model complexity (the "effective number of parameters", p_D). Models with more parameters display better fit, but at the expense of adding complexity to the model. Similar to the Akaike information criterion, the DIC penalizes the additional parameters that improve the fit while searching for a more parsimonious model. Therefore, models having a smaller DIC should be favored, as this indicates a better fit and a lower degree of model complexity. On adding both terms, DIC results in

$$[18] \quad \text{DIC} = \overline{D}(\boldsymbol{\theta}_M) + p_D$$

To gain insight on the term related to model fit, let $D(\boldsymbol{\theta}_M)$ or the Bayesian deviance for model M be indexed by the parametric vector $\boldsymbol{\theta}_M$

$$[19] \quad D(\boldsymbol{\theta}_M) = -2\log p(\boldsymbol{y}|\boldsymbol{\theta}_M) + 2\log f(\boldsymbol{y})$$

The expression $-2\log p(\boldsymbol{y}|\boldsymbol{\theta}_M)$ is the residual information in the data vector $\boldsymbol{y}$ conditional on $\boldsymbol{\theta}$ and is interpreted as a measure of uncertainty. The standardizing term $f(\boldsymbol{y})$ does not depend on $\boldsymbol{\theta}$, so it does not affect model comparison. Therefore, $\overline{D}(\boldsymbol{\theta}_M)$, or the posterior expectation of the deviance, is equal to

$$[20] \quad \overline{D}(\boldsymbol{\theta}_M) = -2\int \log p(\boldsymbol{y}|\boldsymbol{\theta}_M)p(\boldsymbol{\theta}_M|\boldsymbol{y}, \mathbf{M})\mathrm{d}\boldsymbol{\theta}_M$$

The effective number of parameters is defined as

$$[21] \quad p_D = \overline{D}(\boldsymbol{\theta}_M) - D(\overline{\boldsymbol{\theta}}_M)$$

where $\overline{\boldsymbol{\theta}}_M$ is an estimate of $\boldsymbol{\theta}_M$ that depends on $\boldsymbol{y}$, so that $D(\overline{\boldsymbol{\theta}}_M)$ is the deviance evaluated at the posterior mean of $\boldsymbol{\theta}_M$. It can also be viewed as the "degrees of freedom" of model M (Spiegelhalter et al. 2002, p. 592). Using [21], the DIC in [18] can be rewritten as

$$[22] \quad \text{DIC} = \overline{D}(\boldsymbol{\theta}_M) + \overline{D}(\boldsymbol{\theta}_M) - D(\overline{\boldsymbol{\theta}}_M)$$

so that

$$[23] \quad \text{DIC} = 2\overline{D}(\boldsymbol{\theta}_M) - D(\overline{\boldsymbol{\theta}}_M)$$

Expression [23] is used to calculate DIC in practice from the chains of an MCMC algorithm such as FCG. First com-

pute $D(\boldsymbol{\theta}_M)$ at each FCG iteration, and then calculate both $\overline{D}(\boldsymbol{\theta}_M)$ and $D(\overline{\boldsymbol{\theta}}_M)$ at the end of the chain. Let

$$\boldsymbol{\theta}' = \left[\boldsymbol{\beta}'_M \;\vdots\; \boldsymbol{a}'_M\right]$$

then the deviance is

$$\begin{aligned} D(\boldsymbol{\theta}) &= -2\log p(\boldsymbol{y}\,|\,\boldsymbol{\theta},\mathbf{R}) \\ &= N\log(2\pi) + \log|\mathbf{R}| + (\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta} \\ &\qquad -\mathbf{Z}\boldsymbol{a})'\mathbf{R}^{-1}(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{a}) \end{aligned}$$

where N is the total number of trees with at least one trait recorded. The quantity $(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{a})'\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{a})$ is a weighted sum of squares for error terms, usually taken as a measure of fit. Observe that the deviance decreases as the error sum of squares decreases. Equation A3 in the appendix of Cantet et al. (2004) shows that

$$[24]\quad (\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{a})'\mathbf{R}^{-1}(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\boldsymbol{a}) = \sum_{k=1}^{K}\mathrm{tr}\left[(\mathbf{M}_k\mathbf{R}_0\mathbf{M}'_k)^{-1}\mathbf{E}_{(\boldsymbol{\theta})k}\right]$$

with matrix $\mathbf{E}_{(\boldsymbol{\theta})k}$ as defined in [17]. The notation stresses the dependence of the $\mathbf{E}_k$'s on $\boldsymbol{\theta}$. To obtain a workable expression for $|\mathbf{R}|$ in the deviance, take determinants in [5] to obtain

$$[25]\quad |\mathbf{R}| = \prod_{k=1}^{K}|\mathbf{I}_{mk}\otimes\mathbf{M}_k\mathbf{R}_0\mathbf{M}'_k| = \prod_{k=1}^{K}|\,\mathbf{M}_k\mathbf{R}_0\mathbf{M}'_k|^{n_k}$$

Now take the logarithm in [25]

$$[26]\quad \log|\mathbf{R}| = \sum_{k=1}^{K} n_k\log|\,\mathbf{M}_k\mathbf{R}_0\mathbf{M}'_k|$$

All in all, the deviance is

$$[27]\quad D(\boldsymbol{\theta}) = N\log(2\pi) + \sum_{k=1}^{K} n_k\log|\,\mathbf{M}_k\mathbf{R}_0\mathbf{M}'_k| + \sum_{k=1}^{K}\mathrm{tr}\left[(\mathbf{M}_k\mathbf{R}_0\mathbf{M}'_k)^{-1}\,\mathbf{E}_{(\boldsymbol{\theta})k}\right]$$

$D(\boldsymbol{\theta})$ is calculated and accumulated at each iteration and averaged at the end of the MCMC chain to obtain $\overline{D}(\boldsymbol{\theta})$. Also at the end, $D(\overline{\boldsymbol{\theta}})$ is calculated as follows

$$[28]\quad D(\overline{\boldsymbol{\theta}}) = N\log(2\pi) + \sum_{k=1}^{K} n_k\log|\,\mathbf{M}_k\overline{\mathbf{R}}_0\mathbf{M}'_k| + \sum_{k=1}^{K}\mathrm{tr}\left[(\mathbf{M}_k\overline{\mathbf{R}}_0\mathbf{M}'_k)^{-1}\overline{\mathbf{E}}_{(\boldsymbol{\theta})k}\right]$$

with the bars on top of $\mathbf{R}_0$ and $\mathbf{E}_{(\boldsymbol{\theta})k}$ indicating the means.

A working example

Data

Data used in the study belong to the Forestry Research and Experimentation Centre (Centro de Investigaciones y Experiencias Forestales, CIEF), Buenos Aires, Argentina, and were collected in the northeastern province of Corrientes. The 1992 plantations reported here are part of a series of collaborative trials between the Queensland Forest Service and private firms in Argentina. Traits evaluated were diameter at breast height (1.3 m, DBH, in cm) and total height (Height, in m). The genotypes included in the study were *Pinus elliottii* var. *elliottii* (E) and *Pinus caribaea* var. *hondurensis* (H): (1) open-pollinated bulk of *Pinus elliottii* var. *elliottii* (E), (2) F1 reciprocal crosses: E × H and H × E; (3) the backcrosses E × F1 ($BC_{(E)}$) and H × F1 ($BC_{(H)}$); (4) F2 hybrid bulk (E × H) × (E × H). The design was a randomized complete block. Table 1 displays the distribution of records per trial, total number of trees with records, and the means and the standard deviations for each genotype and for both traits. The remaining families had unknown relationships with other genotypes. The number of E, F2, $BC_{(E)}$, and $BC_{(H)}$ families, taxa, and the relationships between the F1 and those genotypes were unknown. After including all known genetic relationships, 1636 individual trees were used in the analysis.

Table 2. Coefficients of the parameters of Hill (1982) for the genotypes present in the *Pinus* data set.

Genotype	Mean	a_k	d_k	a_ka_k	a_kd_k	d_kd_k
Parental						
H	1	1	–1	1	–1	1
E	1	–1	–1	1	1	1
F1						
E × H	1	0	1	0	0	1
H × E	1	0	1	0	0	1
Backcross						
H × (E × H)	1	0.5	0	0.25	0	0
E × (E × H)	1	–0.5	0	0.25	0	0
F2						
(E × H) × (E × H)	1	0	0	0	0	0

A previous analysis found no significant differences among blocks in both trials and for the two traits. Therefore, no effects of blocks were used in any analysis.

We had two patterns of observed and missing data: the complete pattern ($\mathbf{M}_1 = \mathbf{I}_2$) and observations on DBH alone, that is, $\mathbf{M}_2 = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$.

Candidate genetic model with different mean genetic effects

As different genotypes and crosses were involved in the analysis, it was necessary to test which genetic effects affected DBH and Height, and the parameters defined by Hill (1982) were used for that purpose. This parameterization does not require that the materials be homozygous (Lynch and Walsh 1998; page 208) and remains valid whatever the values of the gene frequencies in the parental populations. Moreover, the parameters defined by Hill (1982) are defined even when the data structure is unbalanced (Cardoso and Tempelman 2004). The parameters explain the performance of different crosses as linear combinations of additive, domi-

Table 3. Expectation of the Bayesian deviance $\bar{D}(\boldsymbol{\theta})$, Bayesian deviance evaluated at the posterior mean of model parameters $D(\bar{\boldsymbol{\theta}})$, effective number of parameters (p_D), and deviance information criterion (DIC) for the eight models analyzed.

	Model							
	1	2	3	4	5	6	7	8
$\bar{D}(\boldsymbol{\theta})$	3943.355	3944.455	3945.583	3946.168	3944.593	3944.913	3944.908	3946.408
$D(\bar{\boldsymbol{\theta}})$	3940.733	3941.806	3942.928	3943.502	3941.956	3942.256	3942.262	3943.767
p_D	2.622	2.649	2.655	2.666	2.637	2.657	2.646	2.641
DIC	3945.977	3947.104	3948.238	3948.834	3947.230	3947.570	3947.554	3949.049

Table 4. Posterior statistics for heritability of diameter (h^2_{DBH}), heritability of height (h^2_{Height}), additive correlation (r_{A}), and environmental correlation (r_{E}).

	Mean	Median	Mode	Effective SE	95% HPD	ESS
h^2_{DBH}	0.116	0.115	0.114	0.019	0.080–0.156	5207
h^2_{Height}	0.139	0.138	0.121	0.022	0.098–0.184	5631
r_{A}	0.524	0.524	0.515	0.057	0.411–0.636	5306
r_{E}	0.394	0.394	0.387	0.042	0.323–0.465	56744

Note: HPD, high posterior density interval; ESS, effective sample size.

nance, and all two-way epistatic interactions, weighted by appropriate coefficients that depend on the cross. Parameters are expressed as deviations from the F2, which is taken to be the reference population. More formally, let p_{fk} be the fraction of the genes that descend from H in the father of genotype k, and let p_{mk} be the fraction of H genes in the mother. Using p_{fk} and p_{mk} Lynch (1991) calculated the coefficient (a_k) for the additive effect A of Hill (1982) as $a_k = p_{fk} + p_{mk} - 1$. Similarly, the coefficient for the dominance effect D was obtained as $d_k = 2[p_{fk}(1 - p_{mk}) + p_{mk}(1 - p_{fk})] - 1$. To exemplify, for the backcross that results from crossing H (father) to an F1 (mother), the fractions are $p_{fB} = 1$ and $p_{mB} = 0.5$. Then, $a_B = 0.5$ and $d_B = 2[1(1 - 0.5) + 0.5(1 - 1)] - 1 = 0$. The coefficients for the different parameters of the *Pinus* genotypes are shown in Table 2.

Data were analyzed using a two-trait model (Henderson and Quaas 1976) with the same model equation for both traits:

$$[29] \quad y_{ijbl} = \boldsymbol{X}'_{jl}\boldsymbol{\beta}_{jl} + \boldsymbol{X}'_{jb}\,\boldsymbol{\omega}_{jb} + a_{ijbl} + e_{ijbl}$$

where y_{ijbl} is a record on tree i (i = 1, 2, …, 1598) for trait j (where j is DBH or Height), with genotype b (where b is E, F1: E × H, F1: H × E, $BC_{(E)}$, $BC_{(H)}$, or F2), from trial l. The vector $\boldsymbol{\beta}_{jl}$ contains the systematic environmental effects (means of the groups) of a tree in trial l with respect to trait j; $\boldsymbol{X}_{jl}$ is a column vector that relates y_{ijbl} to the elements of $\boldsymbol{\beta}_{jl}$. Also, $\boldsymbol{\omega}_{jb}$ is a parametric vector including different combinations of the following genetic effects: additive (A), dominance (D), and epistatic (A × A, A × D, D × D). These are represented by $\boldsymbol{\omega}_{j\mathrm{Ab}}$, $\boldsymbol{\omega}_{j\mathrm{Dbb}'}$, $\boldsymbol{\omega}_{j\mathrm{AAbb}'}$, $\boldsymbol{\omega}_{j\mathrm{ADbb}'}$, $\boldsymbol{\omega}_{j\mathrm{DDbb}}$, with b and b' denoting the parental genotypes. Each row vector $\boldsymbol{X}_{jb}$ contains the coefficients of genotype b displayed in Table 2. Eight models involving different mean genetic parameters were compared: (1) A, (2) A + D, (3) A + A × A, (4) A + D × D, (5) A + D + A × A, (6) A + D + D × D, (7) A + D + A × D, and (8) A + A × A + A × D.

Posterior inference

The DIC was computed for each model using the MCMC output from [23]. At each iteration of the FCG sampler, g_{11}, g_{12}, g_{22}, r_{11}, r_{12}, and r_{22} were reparameterized to heritabilities (h^2_{DBH} and h^2_{Height}), the additive correlation (r_{A}), and the environmental correlation (r_{E}) as follows:

$$[30] \quad h^2_{\mathrm{DBH}} = \frac{g_{11}}{g_{11} + r_{11}} \qquad h^2_{\mathrm{Height}} = \frac{g_{22}}{g_{22} + r_{22}}$$

$$r_{\mathrm{A}} = \frac{g_{12}}{\sqrt{g_{11}g_{22}}} \qquad r_{\mathrm{E}} = \frac{r_{12}}{\sqrt{r_{11}r_{22}}}$$

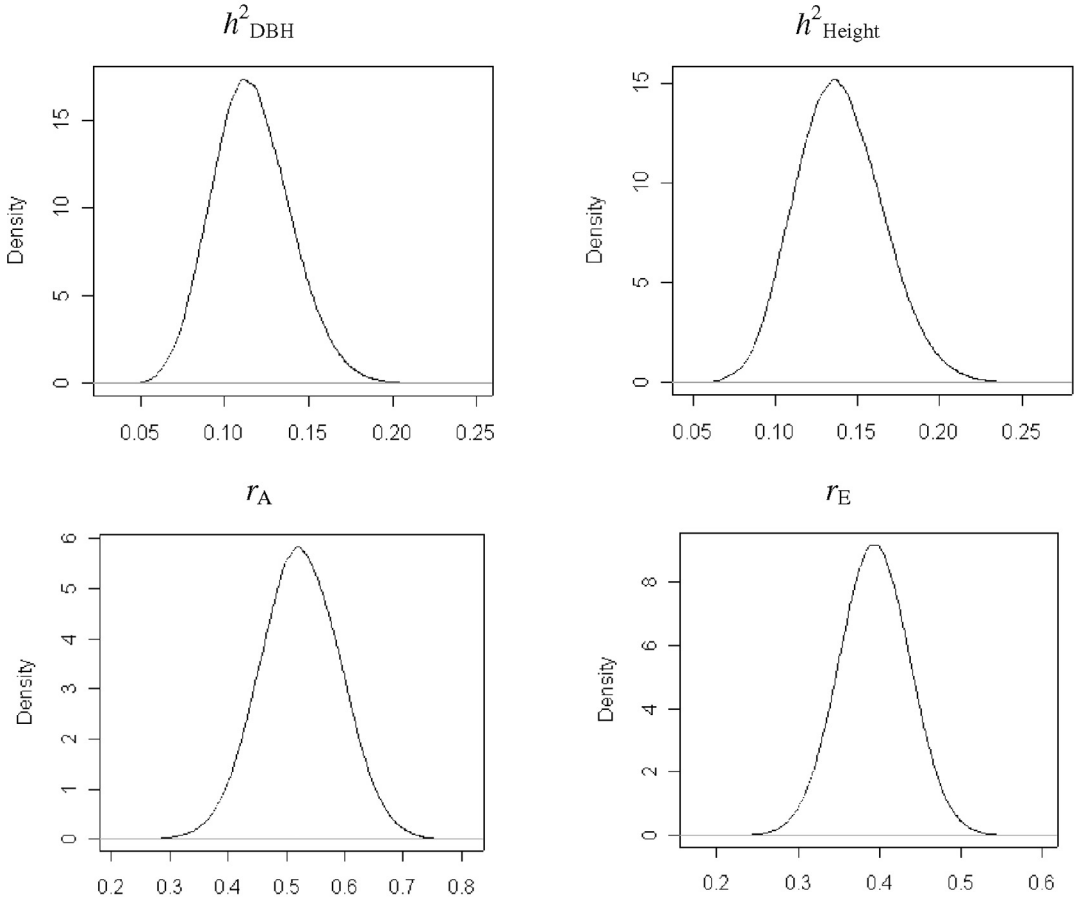
The values of the hypervariances (in $\mathbf{G}^*_0$ and $\mathbf{R}^*_0$) for g_{11}, g_{22}, r_{11}, and r_{22} were obtained using a single-trait Gibbs sampler from the same data set, whereas prior covariances for g_{12} and r_{12} were chosen to be small (but not zero) and positive, so that both $\mathbf{G}^*_0$ and $\mathbf{R}^*_0$ were positive definitive. The degrees of belief were then set to 10 (i.e., $n_A = v_k = 10$) to reflect a relatively high degree of uncertainty.

Using the FCG algorithm, 1 010 000 samples were drawn in a single chain, and the first 10 000 iterates were discarded because of burn-in. Autocorrelations were calculated with Bayesian Output Analysis (BOA version 1.0.1, Smith 2003) for all lags from 1 to 200. Posterior standard errors of each parameter were corrected for an "effective sample size" (ESS; Neal in Kass et al. 1998), to account for the impact of autocorrelations in the chain on measures of variability. The ESS of each parameter was calculated as

$$\mathrm{ESS} = \frac{1\,000\,000}{1 + 2\sum_{i=1}^{200}\rho(i)}$$

where $\rho(i)$ is the autocorrelation measured at lag i. The posterior effective standard error (effective SE) for parameter i (s_i) was calculated as $\sqrt{s_i^2/\mathrm{ESS}}$. Marginal posterior densities for all parameters were estimated by the Gaussian kernel method (Silverman 1986; chapter 2):

Fig. 1. Marginal posterior distributions of heritabilities of diameter (h^2_{DBH}) and height (h^2_{Height}) and genetic (r_A) and environmental (r_E) correlation.



$$[31] \quad f(\theta) = \frac{1}{1\,000\,000} \sum_{i=1}^{1000000} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{(z-\theta_i)^2}{h}\right]\right\}$$

In [31], $f(\theta)$ is the estimated posterior density, θ_i (i = 1, 2, …, 1 000 000) is a sampled value, and h is the window width estimated by unbiased cross validation. Mean, mode, median, standard deviation, and 95% high posterior density interval (95% HPD), were then calculated with BOA for all parameters from the individual marginal posteriors, under the free software R (http://www.r-project.org/).

Results

The values of DIC, and its two constituent elements, for all models are displayed in Table 3. The additive model 1 gave the smallest values of DIC, $\overline{D}(\boldsymbol{\theta})$, and p_D followed by model 2 (A + D). The deviance term was responsible for 99% of the DIC in all eight models analyzed.

Posterior statistics for h^2_{DBH}, h^2_{Height}, r_A, and r_E are shown in Table 4. For all parameters, posterior means, medians, and modes were quite similar, indicating that the marginal posterior distributions are symmetrical. The marginal posterior means of h^2_{DBH} and h^2_{Height} were, respectively, 0.116 and 0.139, whereas the effective SE was 0.019 for h^2_{DBH} and 0.022 for h^2_{Height}. The marginal posterior mean of r_A was 0.524 and the effective SE was 0.057. Corresponding values for r_E were 0.394 and 0.042. The effective SE as calculated here takes into account ESS. None of the 95% HPD for h^2_{DBH}, h^2_{Height}, r_A, or r_E included zero, which suggests that none of the parameters is zero. Looking at the SE indicates that all estimates were quite precise, though a large number of samples was drawn to attain reasonable ESS. Figure 1 shows that the graphs of the posterior densities for h^2_{DBH}, h^2_{Height}, r_A, and r_E tended to be symmetrical.

Discussion

The Bayesian multiple-trait analysis with the FCG algorithm used in the present study is a precise tool for estimating heritabilities and genetic and environmental correlations in forestry data. It is expected that the advantage of using the FCG will be greater for data sets with more traits and with more patterns of unbalance than in the one used here. The algorithm would also be advantageous when estimating covariance matrices between traits that are easy to measure and traits that are either expensive or difficult to score and only a small number of observations are available. The FCG algorithm is also useful in models with repeated measures, especially when there are missing data due to tree mortality with time. In this case, it is customary to analyze data from trees that have measures at all times only while discarding records from dead trees (for example, Balocchi et al. 1993).

Mechanisms that induce missing data will affect the form of the distribution of REML estimates of genetic parameters and of statistics that are functions of those estimates. For example, Apiolaza et al. (1999) simulated random subsamples of a trait and found a skewed distribution of the REML estimates of heritability. The symmetry of the posterior distributions of heritabilities and correlations estimated here suggests that similar results would have been obtained had REML been the method of estimation. There is no small sample distribution for REML, but asymptotic normality under certain conditions on the eigenvalues of the information matrix (see Cressie and Lahiri 1993). Even for two-trait models such as [29], the calculation of the information matrix may be unfeasible with large data sets. Moreover, whereas REML provides only point estimates of the parameters and the asymptotic approximation of their variances, the Bayesian approach allows more general inferences, as the exact posterior distribution is available. Therefore, variance, standard errors, posterior HPD intervals, or the probability of a parameter being less than a given value can be reported.

Although the use of DIC outside of data originating from exponential families is a topic of ongoing discussion, its performance in normal linear models, as in the one described in this research, is sound and well defined (see Spiegelhalter et al. 2002, and subsequent discussion on the same paper). The argument that DIC may be sensitive to the choice of prior distribution of the parameters to estimate becomes inconsequential if the researcher has a strong belief in its prior. Notwithstanding this, the noninformative prior density for the linear genetic parameters in [29] induces DIC to behave in a similar way to the Akaike information criterion.

In the current research, the genetic model used for all analyses included linear combinations of generation means for the different crosses plus random additive effects fitted at the individual (tree) level. Different genetic models responsible for the variation in mean DBH and Height were tested using the DIC. All of these employ the parameterization defined by Hill (1982), which is equivalent to a regression version of generation-mean analysis. For the different genotypes that result from crossing two species, the chosen parameterization would take into account mean general effects of A, D, A × A, D × D, and A × D. As such, it cannot detect genetic effects at the individual genetic locus level, as discussed by Wu and Li (1999). More formally, the absence of linear components attributable to A × A, D × D, and A × D effects does not imply that epistasis effects are null, since linear functions of the means may be zero even though individual effects are not. Thus, the model used here is not useful to search for epistatic values at the level of two or more specific loci. However, mean generation analysis plus random additive effects may not be restrictive for a tree breeder in search of a workable genetic model that fits the data in order to perform selection.

Acknowledgements

The authors would like to thank to the Forestry Research and Experimentation Centre (CIEF, Buenos Aires, Argentina) for kindly providing the data used for the study and three anonymous referees for constructive comments on an earlier version of this manuscript. This research was supported by grants from Agencia Nacional de Ciencia y Tecnología (FONCyT PICT 08009502) and Universidad de Buenos Aires (UBACyT G018, 2004-2007), of Argentina.

References

Apiolaza, L.A., Burdon, R.D., and Garrick, D.J. 1999. Effect of univariate subsampling on the efficiency of bivariate parameter estimation and selection using half-sibs progeny tests. For. Genet. **9**: 79–87.

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.

Borralho, N.M.G. 1995. The impact of individual tree mixed models (BLUP) in tree breeding strategies. *In* Eucalypts Plantations: Improving Fibre Yield and Quality. Proceedings of a CRC-IUFRO Conference. *Edited by* B.M. Potts, N.M.G. Borralho, J.B. Reid, R.N. Cromer, W.N. Tibbits, and C.A. Raymond. CRC Temperate Hardwood Forestry, Hobart, Tasmania, Australia. pp. 141–145.

Cantet, R.J.C., Birchmeier, A.N., and Steibel, J.P. 2004. Full conjugate analysis of normal multiple traits with missing records using a generalized inverted Wishart distribution. Genet. Sel. Evol. **36**: 49–64.

Cardoso, F.F., and Tempelman, R.J. 2004. Hierarchical Bayes multiple-breed inference with an application to genetic evaluation of a Nelore–Hereford population. J. Anim. Sci. **82**: 1589–1601.

Cressie, N., and Lahiri, S. N. 1993. The asymptotic distribution of REML estimators. J. Multivar. Anal. **45**: 217–233.

Dieters, M.J., White, T.L., Litell, R.C., and Hodge, G.R. 1995. Application of approximate variances of variance components and their ratios in genetic tests. Theor. Appl. Genet. **91**: 15–24.

Dominici, F., Parmigiani, G., and Clyde, M. 2000. Conjugate analysis of multivariate normal data with incomplete observations. Can. J. Stat. **28**: 533–550.

Dungey, H.S., Dieters, M.J., Gwaze, D.P., Toon, P.G., and Nikles, D.G. 2000. Interspecific pine hybrids II. genotype by environment interactions across Australia, Swaziland and Zimbabwe. For. Genet. **7**: 21–30

Golub, G.H., and Van Loan, C.F. 1983. Matrix computations. The Johns Hopkins University Press, Baltimore, Md.

Gwaze, D.P., and Woolliams, J.A. 2001. Making decisions about the optimal selection environment using Gibbs sampling. Theor. Appl. Genet. **103**: 63–69.

Harville, D.A. 1974. Bayesian inference for variance components using only error contrasts. Biometrika, **61**: 383–384.

Henderson, C.R. 1984. Applications of linear models in animal breeding. University of Guelph, Guelph, Ont.

Henderson, C.R., and Quaas, R.L. 1976. Multiple trait evaluation using relatives' records. J. Anim. Sci. **43**: 1188–1197.

Hill, W.G. 1982. Dominance and epistasis as components of heterosis. J. Anim. Breed. Genet. **99**: 161–168.

Hobert, J.P., and Casella, G. 1996. The effects of improper priors on Gibbs sampling in hierarchical linear models. J. Am. Stat. Assoc. **91**: 1461–1473.

Huber, D.A., White, T.L., and Hodge, G.R. 1994. Variance component estimation techniques compared for two mating designs with forest genetic architecture through computer simulation. Theor. Appl. Genet. **88**: 236–242.

Kass, R.E., Carlin, B.P., Gelman, A., and Neal, R.M. 1998. Markov chain Monte Carlo in practice: a roundtable discussion. Am. Stat. **52**: 93–100.

Kincaid, D.R., Respess, J.R., and Young, D.M. 1982. Algorithm 586 ITPACK 2C: A FORTRAN package for solving large sparse

linear systems by adaptive accelerated iterative methods. ACM Trans. Math. Softw. **8**: 302–322.

Lynch, M. 1991. The genetic interpretation of inbreeding depression and outbreeding depression. Evolution, **45**: 622–629.

Lynch, M., and Walsh, B. 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, Mass.

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

Robert, C.P., and Casella, G.C. 1999. Monte Carlo statistical methods. Springer-Verlag, New York.

Rubin, D.B. 1976. Inference and missing data. Biometrika, **63**: 581–592.

Smith, B.J. 2003. Bayesian Output Analysis Program (BOA) version 1.0 user's manual [online]. Available from http://www.public-health.uiowa.edu/boa/Home.html.

Shepherd, M., Cross, M., Dieters, M.J., and Henry, R. 2002. Branch architecture QTL for *Pinus elliottii* var. *elliottii* × *Pinus caribaea* var. *hondurensis* hybrids. Ann. For. Sci. **59**: 617–625.

Sorensen, D., and Gianola, D. 2002. Likelihood, Bayesian, and MCMC methods in quantitative genetics. Springer-Verlag, New York.

Sorensen, D.A., Wang, C.S., Jensen, J., and Gianola, D. 1994. Bayesian analysis of genetic change due to selection using Gibbs sampling. Genet. Sel. Evol. **26**: 333–360.

Soria, F., Basurco, F., Toval, G., Silió, L., Rodriguez, M.C., and Toro, M. 1998. An application of Bayesian techniques to the genetic evaluation of growth traits in *Eucalyptus globulus*. Can. J. For. Res. **28**: 1286–1294.

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. R. Stati. Soc. Ser. B, **64(4)**: 583–639.

Van Tassell, C.P., and Van Vleck, L.D. 1996. Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference. J. Anim. Sci. **74**: 2586–2597.

Wang, C.S., Rutledge, J.J., and Gianola, D. 1994. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. Genet. Sel. Evol. **26**: 91–115.

Wu, R., and Li, B. 1999. A multiplicative-epistatic model for analysing intespecific differences in outcrossing species. Biometrics, **55**: 355–365.

Zeng, W., Ghosh, S. K., and Li, B. 2004. A blocking Gibbs sampling method to detect major genes affecting a quantitative trait for diallel mating design. Genet. Res. **84**: 1–12.