Original article

Full conjugate analysis of normal multiple traits with missing records using a generalized inverted Wishart distribution

Rodolfo Juan Carlos CANTET[a,b*], Ana Nélida BIRCHMEIER[a],
Juan Pedro STEIBEL[a]

[a] Departamento de Producción Animal, Universidad de Buenos Aires,
Avenida San Martín 4453, 1417 Buenos Aires, Argentina
[b] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Abstract – A Markov chain Monte Carlo (MCMC) algorithm to sample an exchangeable covariance matrix, such as the one of the error terms (R_0) in a multiple trait animal model with missing records under normal-inverted Wishart priors is presented. The algorithm (FCG) is based on a conjugate form of the inverted Wishart density that avoids sampling the missing error terms. Normal prior densities are assumed for the 'fixed' effects and breeding values, whereas the covariance matrices are assumed to follow inverted Wishart distributions. The inverted Wishart prior for the environmental covariance matrix is a product density of all patterns of missing data. The resulting MCMC scheme eliminates the correlation between the sampled missing residuals and the sampled R_0, which in turn has the effect of decreasing the total amount of samples needed to reach convergence. The use of the FCG algorithm in a multiple trait data set with an extreme pattern of missing records produced a dramatic reduction in the size of the autocorrelations among samples for all lags from 1 to 50, and this increased the effective sample size from 2.5 to 7 times and reduced the number of samples needed to attain convergence, when compared with the 'data augmentation' algorithm.

FCG algorithm / multiple traits / missing data / conjugate priors / normal-inverted Wishart

1. INTRODUCTION

Most data sets used to estimate genetic and environmental covariance components, from multiple trait animal models, have missing records for some traits. Missing data affect the precision of estimates, and reduce the convergence rates of the algorithms used for either restricted maximum likelihood (REML, [13]) or Bayesian estimation. The common Bayesian approach for

---

estimating (co)variance components in multiple trait animal models or test-day models with normal missing records, is to use prior distributions that would be conjugate if the data were complete. Using this approach Van Tassell and Van Vleck [17] and Sorensen [15] implemented 'data augmentation' algorithms (DA, [16]) by sampling from multivariate normal densities for the fixed effects and breeding values, and from inverted Wishart densities for the covariance matrices of additive genetic and environmental effects. The procedure needs sampling of *all* breeding values and errors for all individuals with at least one observed trait, conditional on the additive genetic and environmental covariance matrices sampled in the previous iteration of the Markov chain Monte Carlo (MCMC) augmentation scheme. The next step is to sample the additive genetic and environmental covariance matrices conditional on the breeding values and error terms just sampled. Therefore, the MCMC samples of breeding values tend to be correlated with the sampled covariance matrix of additive (co)variances and the sampled errors tend to be correlated with the sampled environmental covariance matrix, due to the dependence between one set of random variables to the other. Also, samples of the covariance matrices tend to be highly autocorrelated, indicating problems with slow mixing and hence inefficiency in posterior computations. The consequence of the correlation between parameter samples and autocorrelations among sampled covariance matrices is that the number of iterations needed to attain convergence tend to increase [10] in missing data problems, a situation similar to that which happens when the expectation-maximization algorithm of Dempster *et al.* [4] is used for REML estimation of multiple trait (co)variance components.

It would then be useful to find alternative MCMC algorithms for estimating covariance matrices in normal-inverted Wishart settings with smaller autocorrelations among samples and faster convergence rates than regular DA. Liu *et al.* [11] suggested that MCMC sampling schemes for missing data that collapse the parameter space (for example, by avoiding the sampling of missing errors) tend to substantially increase the effective sampling size and to consequently accelerate convergence. Dominici *et al.* [5] proposed collapsing the parameter space by integrating the missing values rather than imputing them. In the normal-inverted Wishart setting for multiple traits of Van Tassell and Van Vleck [17] or Sorensen [15], this would lead to the direct sampling of the covariance matrices of additive breeding values and/or error terms, without having to sample the associated missing linear effects. The scheme is possible for the environmental covariance matrix only, since error terms are exchangeable among animals while breeding values are not. The objective of this research is to show how to estimate environmental covariance components with missing data, using a Bayesian MCMC scheme that does not involve

sampling missing error terms, under a normal-inverted Wishart conjugate approach. The procedure (from now on called FCG for 'full conjugate' Gibbs) is illustrated with a data set involving growth and carcass traits of Angus animals, and compared with the DA sampler for multiple normal traits.

2. THE FCG SAMPLER

2.1. Multiple normal trait model with missing data

The model for the data on a normally distributed trait j ($j = 1, ..., t$) taken in animal i ($i = 1, ..., q$), is as follows:

$$y_{ij} = X_{ij}'\beta_j + a_{ij} + e_{ij} \tag{1}$$

where y_{ij}, a_{ij}, and e_{ij} are the observation, the breeding value and the error term, respectively . The vector β_j of 'fixed' effects for trait j is related to the observations by a vector X_{ij}' of known constants. As usual, y_{ij} is potentially observed, whereas a_{ij} and e_{ij} are not, $i.e.$, some animals may not have records for some or all of the traits. All individuals that have observations on the same traits (say $t_g \leq t$), share the same 'pattern' of observed and missing records. A pattern of observed data can be represented by a matrix (M_g) having t_g rows and t columns [5], where $g = 1, ...$ G, and G is the number of patterns in a given data set. Let n be the total number of animals with records in all traits. All elements in any row of M_g are 0's except for a 1 in the column where trait j is observed. Thus, $M_g = I_t$ whenever $t_g = 0$. For example, suppose $t = 6$ and traits 1, 2 and 5 are observed. Then:

$$M_g = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

There are 2^t-1 different matrices M_g related to patterns with at least one trait observed. We will set the complete pattern to be $g = 1$, so that $M_1 = I_t$. We will also assume that the complete pattern is observed in at least t animals, and denote with n_g the number of animals with records in pattern g. Then, $n = \sum_{g=1}^{G} n_g$ so that n is the total number of animals with at least one trait recorded.

Let y be the vector with the observed data ordered by traits within animal within pattern. Stacking the fixed effects by trait and the breeding values by animal within trait, results in the following matrix formulation of (1):

$$y = X\beta + Za + e. \tag{2}$$

The matrix Z relates records to the breeding values in a. In order to allow for maternal effects, we will take the order of a to be $rq \times 1(r \geq t)$, rather than $rt \times 1$. The covariance matrix of breeding values can then be written as:

$$\text{Var}(\boldsymbol{a}) = \begin{bmatrix} g_{1,1}\boldsymbol{A} & g_{1,2}\boldsymbol{A} & . & . & . & g_{1,r}\boldsymbol{A} \\ g_{2,1}\boldsymbol{A} & g_{2,2}\boldsymbol{A} & . & . & . & g_{2,r}\boldsymbol{A} \\ . & . & . & . & . & . \\ . & . & . & g_{i,j}\boldsymbol{A} & . & . \\ . & . & . & . & . & . \\ g_{r,1}\boldsymbol{A} & g_{r,2}\boldsymbol{A} & . & . & . & g_{r,r}\boldsymbol{A} \end{bmatrix} = \boldsymbol{G}_0 \otimes \boldsymbol{A} \qquad (3)$$

where $g_{jj'}$ is the additive genetic covariance between traits j and j' if $j \neq j'$, and equal to the variance of trait j otherwise. As usual [7], the matrix A contains the additive numerator relationships among all q animals.

In (2), the vector e contains the errors ordered by trait within animal within pattern. Thus, the vectors $\boldsymbol{e}_{(g1)}, \boldsymbol{e}_{(g2)},..., \boldsymbol{e}_{(g t_g)}$ denote the $t_g \times 1$ vectors of errors for the different animals with t_g observed traits in pattern g. Error terms have zero expectation and, for animal i with complete records, the variance-covariance matrix is equal to $\text{Var}(\boldsymbol{e}_{(1i)}) = \boldsymbol{R}_0 = [r_{jj'}]$, with $r_{jj'}$ the environmental (co)variance between traits j and j'. If animal i' has incomplete records in pattern g, the variance is then $\text{Var}(\boldsymbol{e}_{(i'g)}) = \boldsymbol{M}_g \boldsymbol{R}_0 \boldsymbol{M}'_g$. For the entire vector e with errors stacked traits within animal within pattern, the variance-covariance matrix is equal to:

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{I}_{n_1} \otimes \boldsymbol{M}_1 \boldsymbol{R}_0 \boldsymbol{M}'_1 & 0 & . & 0 \\ 0 & \boldsymbol{I}_{n_2} \otimes \boldsymbol{M}_2 \boldsymbol{R}_0 \boldsymbol{M}'_2 & . & 0 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 0 & 0 & ... & \boldsymbol{I}_{n_g} \otimes \boldsymbol{M}_g \boldsymbol{R}_0 \boldsymbol{M}'_g \end{bmatrix}$$
$$= \bigoplus_{g=1}^{G} \boldsymbol{I}_{n_g} \otimes \boldsymbol{M}_g \boldsymbol{R}_0 \boldsymbol{M}'_g. \qquad (4)$$

Under normality of breeding values and errors, the distribution of the observed data can be written as being proportional to:

$$p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{R}_0, \boldsymbol{M}_1, ..., \boldsymbol{M}_G) \propto |\boldsymbol{R}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{a})' \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{a})\right].$$
$$(5)$$

2.2. Prior densities

In order to avoid improper posterior densities, we take the prior distribution of the $p \times 1$ vector $\boldsymbol{\beta}$ to be multivariate normal: $\boldsymbol{\beta} \sim N_p (\mathbf{0}, \ \boldsymbol{K})$. If vague prior knowledge is to be reflected, the covariance matrix $\boldsymbol{K}$ will have very large diagonal elements (say $k_{ii} > 10^8$). This specification avoids having improper posterior distributions in the mixed model [8]. The prior density of $\boldsymbol{\beta}$ will then be proportional to:

$$p(\boldsymbol{\beta}|\boldsymbol{K}) \propto \left| \prod_{i=1}^{p} k_{ii} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{p} \frac{\beta_i^2}{k_{ii}} \right\}. \tag{6}$$

The breeding values of the t traits for the q animals are distributed *a priori* as $\boldsymbol{a} \sim N_{rq}(\mathbf{0}, \boldsymbol{G}_0 \otimes \boldsymbol{A})$, so that:

$$p(\boldsymbol{a}|\boldsymbol{G}_0, \boldsymbol{A}) \propto |\boldsymbol{G}_0|^{-\frac{q}{2}}|\boldsymbol{A}|^{-\frac{t}{2}} \exp\left\{ -\frac{1}{2}\boldsymbol{a}' \left(\boldsymbol{G}_0^{-1} \otimes \boldsymbol{A}^{-1}\right) \boldsymbol{a} \right\}. \tag{7}$$

The matrix of the additive (co)variance components $\boldsymbol{G}_0$ follows *a priori* an inverted Wishart (IW) density: $\boldsymbol{G}_0 \sim \text{IW}(\boldsymbol{G}_0^*, n_A)$, where $\boldsymbol{G}_0^*$ is the prior covariance matrix and n_A are the degree of belief. Thus:

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

We now discuss the prior density for $\boldsymbol{R}_0$. Had it been the data complete (no missing records for all animals and for all traits), the prior density for $\boldsymbol{R}_0$ would have been IW $(\boldsymbol{R}_0^*, \nu)$, the hyperparameters being the prior covariance matrix $\boldsymbol{R}_0^*$ and the degrees of belief ν. In order to allow for all patterns of missing data, and after the work of Kadane and Trader [9] and Dominici *et al.* [5], we take the following conjugate prior for $\boldsymbol{R}_0$:

$$p(\boldsymbol{R}|\boldsymbol{R}_0^*, \boldsymbol{M}_1, ..., \boldsymbol{M}_G, \nu_g) \propto \prod_{g=1}^{G} |\boldsymbol{M}_g\boldsymbol{R}_0\boldsymbol{M}_g'|^{-\frac{(\nu_g+2t_g+1)}{2}}$$
$$\times \exp\left\{ -\frac{1}{2}\text{tr}\left[\boldsymbol{M}_g\boldsymbol{R}_0^*\boldsymbol{M}_g'(\boldsymbol{M}_g\boldsymbol{R}_0\boldsymbol{M}_g')^{-1}\right] \right\}. \tag{9}$$

In words of Dominici *et al.* [5], the specification (9) mimics the natural conjugate prior for the t_g-dimensional problem of inference on the variables of pattern g.

2.3. Joint posterior distribution

Multiplying (5) with (6), (7), (8) and (9), produces the joint posterior density for all parameters, and this is proportional to:

$$p(\boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{G}_0, \boldsymbol{R}_0 | \boldsymbol{y}, \boldsymbol{M}_1, ..., \boldsymbol{M}_G) \propto$$

$$|\boldsymbol{R}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{a})' \boldsymbol{R}^{-1}(\boldsymbol{y}, \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{a})\right] \exp\left\{-\frac{1}{2}\sum_{i=1}^{p}\frac{\beta_i^2}{k_{ii}}\right\}$$

$$\times \exp\left\{-\frac{1}{2}\boldsymbol{a}'\left(\boldsymbol{G}_0^{-1} \otimes \boldsymbol{A}^{-1}\right)\boldsymbol{a}\right\} |\boldsymbol{G}_0|^{-\frac{(n_A+r+q+1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{G}_0^*\boldsymbol{G}_0^{-1}\right)\right\}$$

$$\times \prod_{g=1}^{G} |\boldsymbol{M}_g\boldsymbol{R}_0\boldsymbol{M}_g'|^{-\frac{(v_g+2t_g+1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{M}_g\boldsymbol{R}_0^*\boldsymbol{M}_g'(\boldsymbol{M}_g\boldsymbol{R}_0\boldsymbol{M}_g')^{-1}\right]\right\}. \quad (10)$$

A hybrid MCMC method may be used to sample from (10), by combining the classic DA algorithm for normal multiple traits of Van Tassell and Van Vleck [17] and Sorensen [15], for $\boldsymbol{\beta}$, $\boldsymbol{a}$ and $\boldsymbol{G}_0$, with the conjugate specification (9) for $\boldsymbol{R}_0$ proposed by Dominici *et al.* [5].

2.4. Conditional posterior distributions of $\boldsymbol{\beta}$, $\boldsymbol{a}$ and $\boldsymbol{G}_0$

The algorithm employed by Van Tassell and Van Vleck [17] and Sorensen [15] involves the sampling of the fixed effects and the breeding values first. In doing so, consider the following linear system:

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} + \boldsymbol{K}^{-1} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}_0 \otimes \boldsymbol{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{a}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix}. \quad (11)$$

The system (11) is a function of $\boldsymbol{K}$, $\boldsymbol{G}_0$ and $\boldsymbol{R}_0$, and allows writing the joint conditional posterior density of $\boldsymbol{\beta}$ and $\boldsymbol{a}$ as:

$$\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{a} \end{bmatrix} 1/2\boldsymbol{G}_0, \boldsymbol{R}_0, \boldsymbol{y} \sim N_{p+qr}\left(\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{a}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} + \boldsymbol{K}^{-1} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}_0 \otimes \boldsymbol{A}^{-1} \end{bmatrix}^{-1}\right).$$
$$(12)$$

The normal density (12) can be sampled either by a parameter or by a block of parameters [17].

To sample from the posterior conditional density of G_0, let S be

$$
S = \begin{bmatrix}
a'_1 A^{-1} a_1 & a'_1 A^{-1} a_2 & . & . & a'_1 A^{-1} a_r \\
a'_2 A^{-1} a_1 & a'_2 A^{-1} a_2 & . & . & a'_2 A^{-1} a_r \\
. & & a'_i A^{-1} a_j & . & . \\
. & . & . & . & . \\
a'_r A^{-1} a_1 & a'_r A^{-1} a_2 & . & . & a'_r A^{-1} a_r
\end{bmatrix}. \quad (13)
$$

Then, Van Tassell and Van Vleck [17] observed that the conditional posterior distribution of G_0 is inverted Wishart with the scale matrix $G_0^* + S$ and degrees of freedom equal to $n_A + q$, so that:

$$
p(G_0 | y, \beta, a, R_0) \propto |G_0|^{-\frac{(n_A + r + q + 1)}{2}} \exp\left\{ -\frac{1}{2} \mathrm{tr}\left((G_0^* + S) G_0^{-1} \right) \right\}. \quad (14)
$$

2.5. Conditional posterior distribution of R_0

In this section the sampling of the environmental covariance matrix R_0 is implemented. The procedure is different from the regular DA algorithm proposed by Van Tassell and Van Vleck [17] or Sorensen [15], since no missing errors are sampled: only those variances and covariances of R_0 that are missing from any particular pattern are sampled. In the Appendix it is shown that the conditional posterior distribution of R_0 has the following density:

$$
p(R_0 | y, \beta, a, G_0) \propto \prod_{g=1}^{G} |M_g R_0 M'_g|^{-\frac{(v_g + n_g + 2t_g + 1)}{2}}
$$

$$
\times \exp\left\{ -\frac{1}{2} \mathrm{tr}\left[(M_g R_0^* M'_g + E_g)(M_g R_0 M'_g)^{-1} \right] \right\}. \quad (15)
$$

To sample the covariance matrix of a normal distribution, Dominici *et al.* [5] proposed an MCMC sampling scheme based on a recursive analysis of the inverted Wishart density described by Bauwens *et al.* [1] (th. A.17 p. 305). This approach is equivalent to characterizing (15) as a generalized inverted Wishart density [2]. We now present in detail the algorithm to sample R_0. The first step of the algorithm requires calculating the "completed" matrices of hyperparameters for each pattern. We denote this matrix for pattern g as R_g^*, being composed by submatrices R_{oo}^{*g}, R_{om}^{*g} and R_{mm}^{*g} for the observed, observed by missing and missing traits, respectively. Once the R_g^*'s are computed, R_0 is sampled from an inverted Wishart distribution with the parameter covariance matrix equal to the sum of the "completed" matrices obtained in the previous step. The sampling of R_0 then involves the following steps:

2.5.1. Sampling of the hypercovariances between observed and missing traits in the pattern

This step requires a matrix multiplication to a sampled random matrix. Formally:

$$\left(R_{oo}^{*g}\right)^{-1} R_{om}^{*g} | R_0, R_{mm.o} \sim R_{oo}^{*g} \times N_{t_g \times (t - t_g)} \left[R_{oo}^{*g} R_{om}, R_{oo}^{-1} \otimes R_{mm.o} \right]. \qquad (16)$$

In practice, the sampling from the matricvariate normal matrix (see [1] p. 305) in (16) can be achieved by performing the following operations. First, multiply the Cholesky decomposition of the covariance matrix $(R_{oo}^{*g})^{-1} \otimes R_{mm.o}$ times a $[t_g(t - t_g)]$ vector of standard normal random variables, with $R_{mm.o} = R_{mm} - R_{mo} R_{oo}^{-1} R_{om}$. The resulting random vector is then set to a matrix of order a $t_g \times (t - t_g)$ by reversing the vec operation: the first column of the matrix is formed with the first t_g elements of the vector, the second column with the next t_g elements, and so on to end up with column $(t - t_g)$. Note that R_{oo}, R_{mm} and R_{mo} are submatrices of R_0 from the previous iteration of the FCG sampling. Then, the mean $R_{oo}^{-1} R_{om}$ should be added to the random $t_g \times (t - t_g)$ matrix. Finally, the resulting matrix should be premultiplied by the $t_g \times t_g$ matrix associated with the observed traits in pattern g equal to $R_{oo}^* = R_{0oo}^* + E_g$. The matrix R_{0oo}^* contains the hypervariances and covariances in R_0^*. The matrix with the sum of squared errors in pattern g (E_g) is defined in the Appendix.

2.5.2. Sampling of the hypervariance matrix among missing traits in pattern g

This step is performed by sampling from:

$$R_{mm.o}^{*g} | R_0 \sim W_m \left(v_g + 2t, \ R_{mm.o} \right) \qquad (17)$$

where W_m denotes the Wishart density. The covariance matrix is $R_{mm.o}$ and the degrees of freedom are v_g plus two times the number of traits.

2.5.3. Computation of the unconditional variance matrix among the missing traits in pattern g

$$R_{mm}^{*g} = R_{mm.o}^{*g} + R_{mo}^{*g} (R_{oo}^{*g})^{-1} R_{om}^{*g} \qquad (18)$$

2.5.4. Obtaining the hypercovariance matrix

Let P_g be a permutation matrix such that the observed and missing traits in pattern g recover their original positions in the complete covariance matrix. Then, the contribution of pattern g to the hypercovariance matrix for

sampling $\boldsymbol{R}_0$ is equal to:

$$\boldsymbol{P}_g \left[\begin{array}{cc} \boldsymbol{R}_{oo}^{*g} & \boldsymbol{R}_{om}^{*g} \\ \boldsymbol{R}_{mo}^{*g} & \boldsymbol{R}_{mm}^{*g} \end{array} \right] \boldsymbol{P}'_g = \boldsymbol{R}_g^* \tag{19}$$

so that the complete hypercovariance matrix is equal to $\sum_{g=1}^{G} \boldsymbol{R}_g^*$.

2.5.5. Sampling of R_0

The matrix $\boldsymbol{R}_0$ is sampled from:

$$p(\boldsymbol{R}_0|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{G}_0) \sim \mathrm{IW}\left(\sum_{g=1}^{G} v_g + n + (G-1)(t+1), \left(\sum_{g=1}^{G} \boldsymbol{R}_g^* \right) \right). \tag{20}$$

2.6. A summary of the FCG algorithm

1. Build and solve equations (11).
2. Sample $\boldsymbol{\beta}$ and $\boldsymbol{a}$ from (12).
3. Calculate the residuals: $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Za}$.
4. For every pattern do the following:
   4.1. Sample the hypercovariances between observed and missing traits in the pattern using (16).
   4.2. Sample the hypervariance matrix among missing traits in the pattern using the inverted Wishart density in (17).
   4.3. Calculate the unconditional variance matrix among the missing traits in the pattern using (18).
5. Calculate the hypercovariance matrix for $\boldsymbol{R}_0$ by adding all $\boldsymbol{R}_g^*$ as $\sum_{g=1}^{G} \boldsymbol{R}_g^*$.
6. Sample $\boldsymbol{R}_0$ from (20).
7. Calculate $\boldsymbol{S}$.
8. Sample $\boldsymbol{G}_0$ from (8), and go back to 1.

2.7. Analysis of a beef cattle data set

A data set with growth and carcass records from a designed progeny test of Angus cattle collected from 1981 to 1990, was employed to compare the autocorrelations among sampled $\boldsymbol{R}_0$ and the 'effective sample sizes' of the FCG and DA algorithms. Calves were born and maintained up to weaning at a property of Universidad de Buenos Aires (UBA), in Laprida, south central Buenos Aires province, Argentina. There were 1367 animals with at least one trait recorded, which were sired by 31 purebred Angus bulls and by commercial heifers. Almost half of the recorded animals were females that did not have observations in all traits but birth weight. Every year 6 to 7 bulls were

Table I. Descriptive statistics for the traits measured.

Trait	N	Mean (kg)	SD (%)	CV	Minimum (kg)	Maximum (kg)
Birth weight	1367	30.9	4.6	15	20	45
Weaning weight	561	194.5	31.2	16	99	305
18 months weight	405	332.1	53.4	16	180	472
Weight of three retail cuts	474	13.4	15.0	11	9.3	18.6
Hind *pistola* weight	474	50.7	51.0	10	32	66
Half carcass weight	466	121.6	12.7	10	75	154

tested, and either 1 or 2 sires were repeated the next year to keep the data structure connected. Every year different heifers were artificially inseminated under a completely random mating scheme. After weaning (average age = 252 days), all males were castrated and taken to another property for the fattening phase. The steers were kept on cultivated pastures until they had at least 5 mm of fat over the ribs, based on visual appraisal. The mean age at slaughter was 28 months and the mean weight was 447 kg. Retail cuts had the external fat completely trimmed. The six traits measured were: (1) birth weight (BW); (2) weaning weight (WW); (3) weight at 18 months of age (W18); (4) weight of three retail cuts (ECW); (5) weight of the hind pistola cut (WHP) and (6) half-carcass weight (HCW). Descriptive statistics for all traits, are shown in Table I.

The year of measure was included as a fixed classification variable in the models for all six traits, and sex was included in the model for BW. Fixed covariates were age at weaning for WW, age at measure for W18, and age at slaughter for ECW, WHP and HCW. Random effects were the breeding values and the error terms. The parameters of interest were the 6×6 covariance matrices G_0 and R_0. An empirical Bayes procedure was used to estimate the hyperparameters. The prior covariance matrices G_0^* and R_0^* were taken to be equal to the Expectation-Maximization [4] REML estimates of G_0 and R_0, respectively, as discussed by Henderson [7]. The degrees of belief were set equal to 10. After running 200 000 cycles of both FCG and DA, autocorrelations from lag 1 to 50 for all 21 parameters of R_0 were calculated using the BOA program [14]. Once the autocorrelations were calculated, the 'effective sample size' (Neal, in Kass *et al.* [10]) were computed as:
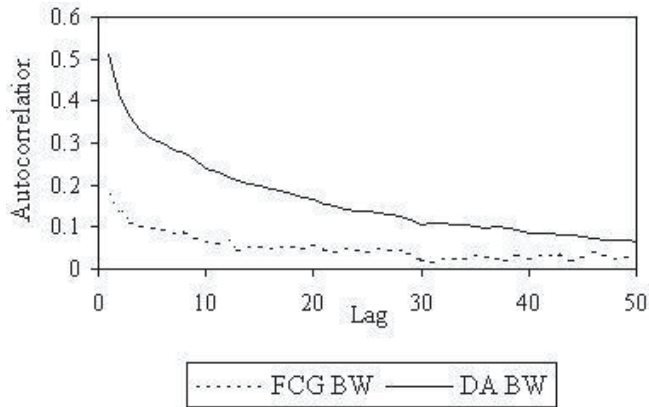
$$ESS = \frac{T}{1 + 2 \sum_{i=1}^{k} \rho(i)},$$

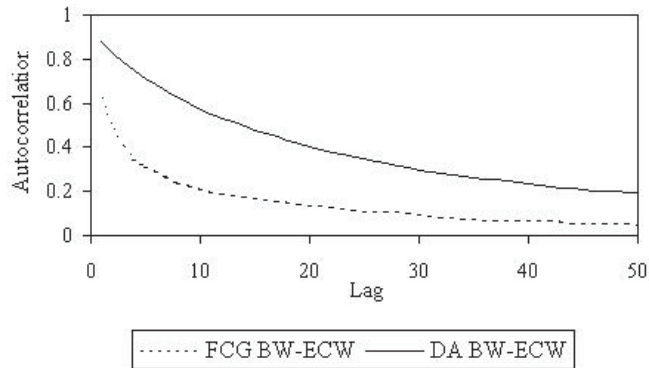**Figure 1.** Lag correlations of the environmental variance of BW for FCG and DA.

Figure 2. Lag correlations of the environmental covariance between BW and ECW for FCG and DA.

where $\rho(i)$ is the autocorrelation calculated at lag i, K is the lag at which ρ seems to have fallen to near zero and T is the total number of samples drawn. We took $K = 50$ and $T = 200\,000$.

3. RESULTS

To summarize the results, the environmental variances of a growth trait (BW, r_{BW}), and a carcass trait (ECW, r_{ECW}), as well as the covariance between them (BW and ECW, r_{BW-ECW}) are reported. The autocorrelation functions of the MCMC draws for r_{BW}, r_{BW-ECW}, and r_{ECW} for the FCG and DA samplers are displayed in Figures 1, 2 and 3, respectively.

For all environmental dispersion parameters the autocorrelations of the FCG sampler were in between 0.5 to 4 times smaller than those obtained with
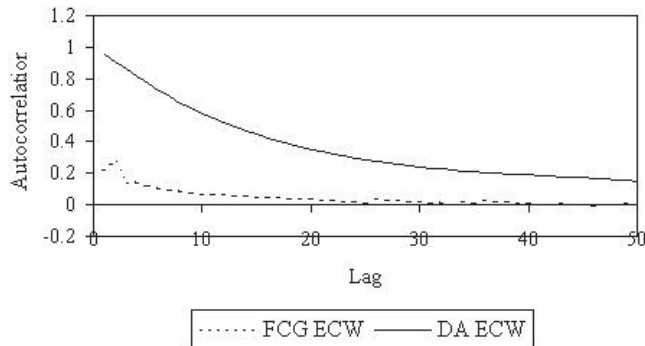
Figure 3. Lag correlations of the environmental variance of ECW for FCG and DA.

Table II. Posterior means (above main diagonal) and posterior standard deviations (below main diagonal) of the environmental parameters expressed as correlations.

	BW	WW	W18	ECW	HPW	HCW
BW	-	0.303	0.291	0.112	0.147	0.240
WW	0.396	-	0.640	0.247	0.468	0.577
W18	0.438	0.533	-	0.308	0.562	0.695
ECW	0.345	0.345	0.375	-	0.363	0.361
HPW	0.452	0.480	0.544	0.387	-	0.714
HCW	0.471	0.534	0.615	0.402	0.615	-

the DA procedure. The 'effective sample sizes' were ESS r_{BW} = 32 560, ESS r_{BW-ECW} = 12 480, and ESS r_{ECW} = 36 480, for FCG. The figures for DA were ESS r_{BW} = 11 320, ESS r_{BW-ECW} = 4880 and ESS r_{ECW} = 5240. As a result, the ESS were 2.87, 2.54, and 6.96 greater for r_{BW}, r_{BW-ECW}, and r_{ECW}, respectively. In other words, FCG required in between 2.5 to 7 times less samples to reach convergence than DA, a substantial decrease. Table II shows the posterior means and the posterior standard deviations, for all environmental (co)variance components, expressed as correlations. A graph of the posterior marginal density of r_{BW-ECW} is displayed in Figure 4.

4. DISCUSSION

The original contribution of this research was to implement a Bayesian conjugate normal-inverted Wishart approach for sampling the environmental covariance matrix R_0, in an animal model for multiple normal traits with missing records. This conjugate setting for normal missing data was first proposed by Kadane and Trader [9], and later implemented by Dominici *et al.* [5] to an exchangeable normal data vector with missing records. The procedure was
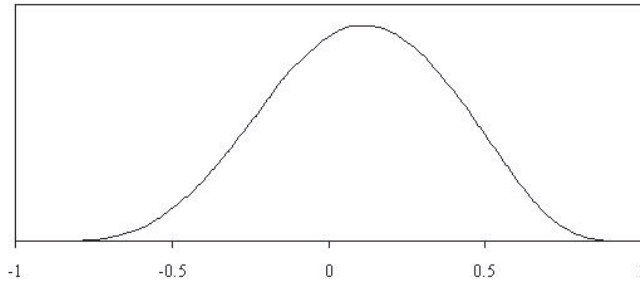
Figure 4. Posterior marginal density of the environmental correlation between BW and ECW.

different from the regular DA algorithm implemented by Van Tassell and Van Vleck [17] for multiple traits, or by Sorensen [15] for multivariate models, in the sense that there is no need to sample the residuals of missing records: only those variances and covariances of R_0 that are missing from any particular pattern are sampled. By avoiding the sampling of missing residuals from R_0, the effect that the correlation between these latter terms and the sampled R_0 has on the convergence properties of the algorithm is eliminated [11], which in turn has the effect of decreasing the total amount of samples needed to reach convergence. The use of the algorithm in a data set with an extreme pattern of missing records produced a dramatic reduction in the size of the autocorrelations among samples for all the lags from 1 to 50. In turn, this increased the effective sample size from 2.5 to 7 times, and reduced the number of samples needed to attain convergence compared with the DA algorithm of Van Tassell and Van Vleck [17] or Sorensen [15]. Additionally, the time spent on each iteration of FCG was approximately 80% of the time spent on each iteration of DA. The total computing time will depend on the fraction of missing to total residuals to sample and the total number of patterns, for the particular data set.

The sampling of the environmental covariance matrix without having to sample the residuals presented here was facilitated by the fact that error terms are exchangeable among animals. This allowed collapsing the parameter space by integrating out the missing residuals, rather than imputing them. Unfortunately, breeding values are not exchangeable (for example, the breeding values of a sire and its son can not be exchanged) so that the sampling can not be performed with the additive genetic covariance matrix. A possibility is to parametrize the model in terms of Mendelian residuals (φ) rather than breeding values, by expressing $a = [I_t \otimes (I_q - P)^{-1}]\varphi$ in (2) [3]. The matrix P has row i with all elements equal to zero except for two 0.5 in the columns corresponding to the sire and dam of animal i. However, this reparametrization does not reduce the number of random genetic effects and the programming is more

involved, so there is not much to gain with this formulation. The alternative parametrization of **a** leaving only the breeding values of those animals with records requires direct inversion of the resulting additive relationship matrices by trait one by one, since the rules of Henderson [7] can not be used, a rather impractical result for most data sets.

Although the algorithm FCG was presented for a multiple trait animal model, it can be equally used for the environmental covariance matrix of longitudinal data with a covariance function, such as the one discussed by Meyer and Hill [12]. In that case let $\mathbf{R}_0 = \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}'$ in (9) and onwards, where the parametric matrix of (co)variance components is $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ is a known matrix that contains orthogonal polynomial functions evaluated at given ages. Also, the prior covariance matrix is such that $\mathbf{R}_0^* = \mathbf{\Phi}\mathbf{\Gamma}^*\mathbf{\Phi}'$.

ACKNOWLEDGEMENTS

REFERENCES

[1] Bauwens L., Lubrano M., Richard J.F., Bayesian inference in dynamic econometric models, Oxford University Press, 1999.

[2] Brown P.J., Le N.D., Zidek J.V., Inference for a covariance matrix, in: Freeman P.R., Smith A.F.M. (Eds.), Aspects of uncertainty: a tribute to D.V. Lindley, Wiley, New York, 1994, pp. 77–92.

[3] Cantet R.J.C., Fernando R.L., Gianola D., Misztal I., Genetic grouping for direct and maternal effects with differential assignment of groups, Genet. Sel. Evol. 24 (1992) 211–223.

[4] Dempster A.P., Laird N.M., Rubin D.B., Maximum likelihood from incomplete data *via* the EM algorithm (with discussion), J. Royal Stat. Soc. Ser. B 39 (1977) 1–38.

[5] Dominici F., Parmigiani G., Clyde M., Conjugate analysis of multivariate normal data with incomplete observations, Can. J. Stat. 28 (2000) 533–550.

[6] Harville D.A., Matrix algebra from a statistician's perspective, Springer-Verlag, NY, 1997.

[7] Henderson C.R., Application of linear models in animal breeding, University of Guelph, Guelph, 1984.

[8] Hobert J.P., Casella G., The effects of improper priors on Gibbs sampling in hierarchical linear models, J. Amer. Statist. Assoc. 91 (1996) 1461–1473.

[9] Kadane J.B., Trader R.L., A Bayesian treatment of multivariate normal data with observations missing at random, Statistical decision theory and related topics IV, 1 (1988) 225–234.

[10] Kass R.E., Carlin B.P., Gelman A., Neal R.M., Markov chain Monte Carlo in practice: a roundtable discussion, Amer. Stat. 52 (1998) 93–100.

[11] Liu J.S., Wong W.H., Kong A., Covariance structure and convergence of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes, Biometrika 81 (1994) 27–40.

[12] Meyer K., Hill W.G., Estimation of genetic and phenotypic covariance functions for longitudinal or repeated records by restricted maximum likelihood, Livest. Prod. Sci. 47 (1997) 185–200.

[13] Patterson H.D., Thompson R., Recovery of inter-block information when block sizes are unequal, Biometrika 58 (1971) 545–554.

[14] Smith B.J., Bayesian Output Analysis Program (BOA) version 1.0.0 user's manual (2001), Available from http://www.public-health.uiowa.edu/boa/Home.html.

[15] Sorensen D.A., Gibbs sampling in quantitative genetics, Internal report 82, Danish Institute of Animal Science, Department of Breeding and Genetics (1996).

[16] Tanner M.A., Wong W.H., The calculation of posterior distributions by data augmentation, J. Amer. Stat. Assoc. 82 (1987) 528–540.

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

APPENDIX: DERIVATION OF THE FULL CONDITIONAL DENSITY OF R_0 IN (15)

Collecting all terms in (10) that are a function of $\boldsymbol{R}_0$ results in:

$$p(\boldsymbol{R}_0|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{G}_0) \propto |\boldsymbol{R}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{a})'\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{a})\right]$$

$$\times \prod_{g=1}^{G} |\boldsymbol{M}_g\boldsymbol{R}_0\boldsymbol{M}_g'|^{-\frac{(v_g+2t_g+1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{M}_g\boldsymbol{R}_0^*\boldsymbol{M}_g'(\boldsymbol{M}_g\boldsymbol{R}_0\boldsymbol{M}_g')^{-1}\right]\right\}. \quad (A.1)$$

Let $\boldsymbol{e}_{(g1)}$, $\boldsymbol{e}_{(g2)}$,..., $\boldsymbol{e}_{(go)}$ be the $n_g \times 1$ vectors of errors for the different observed traits in pattern g. The notation in parenthesis means, for example, that trait $(g1)$ is the first trait observed in the pattern and not necessarily trait 1. Using this notation, let the matrix $\boldsymbol{E}_g$ (of order $t_g \times t_g$) containing the sum of squared errors for pattern g be:

$$\boldsymbol{E}_g = \begin{bmatrix} \boldsymbol{e}'_{(g1)}\boldsymbol{e}_{(g1)} & \boldsymbol{e}'_{(g1)}\boldsymbol{e}_{(g2)} & \cdot & \boldsymbol{e}'_{(g1)}\boldsymbol{e}_{(g5)} & \boldsymbol{e}'_{(g1)}\boldsymbol{e}_{(t_g)} \\ \boldsymbol{e}'_{(g2)}\boldsymbol{e}_{(g1)} & \boldsymbol{e}'_{(g2)}\boldsymbol{e}_{(g2)} & \cdot & \boldsymbol{e}'_{(g2)}\boldsymbol{e}_{(g5)} & \boldsymbol{e}'_{(g2)}\boldsymbol{e}_{(t_g)} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \boldsymbol{e}'_{(i)}\boldsymbol{e}_{(j)} & \cdot \\ \boldsymbol{e}'_{(t_g)}\boldsymbol{e}_{(g1)} & \boldsymbol{e}'_{(t_g)}\boldsymbol{e}_{(g2)} & \cdot & & \boldsymbol{e}'_{(t_g)}\boldsymbol{e}_{(t_g)} \end{bmatrix}. \quad (A.2)$$

Using this specification we can write:

$$(y - X\beta - Za)'R^{-1}(y - X\beta - Za) = \text{tr}\left[ee'R^{-1}\right] = \text{tr}\left[ee'\left(\bigoplus_{g=1}^{G}\left(M_gR_0M_g'\right)^{-1}\right)\right]$$

$$= \sum_{g=1}^{G}\text{tr}\left[\sum_{i,j}e_{(gi)}e_{(gj')}'\left(M_gR_0M_g'\right)^{-1}\right] = \sum_{g=1}^{G}\text{tr}\left[E_g\left(M_gR_0M_g'\right)^{-1}\right]. \quad (A.3)$$

Now, by replacing on the exponential term in (A.1) with (A.3) produces:

$$\exp\left\{-\frac{1}{2}\left(e'R^{-1}e + \text{tr}\left[\bigoplus_{g=1}^{G}M_gR_0^*M_g'(M_gR_0M_g')^{-1}\right]\right)\right\} =$$

$$\exp\left\{-\frac{1}{2}\left(\sum_{g=1}^{G}\text{tr}\left(E_g\left(M_gR_0M_g'\right)^{-1}\right)\sum_{g=1}^{G}\text{tr}\left(M_gR_0^*M_g'(M_gR_0M_g')^{-1}\right)\right)\right\} =$$

$$\exp\left\{-\frac{1}{2}\sum_{g=1}^{G}\text{tr}\left(\left(E_g + M_gR_0^*M_g'\right)\left(M_gR_0M_g'\right)^{-1}\right)\right\}. \quad (A.4)$$

Operating on the determinants of (A.1) allows us to write down:

$$|R|^{-\frac{1}{2}}\prod_{g=1}^{G}|M_gR_0M_g'|^{-\frac{(v_g+2t_g+1)}{2}} = \prod_{g=1}^{G}|I_{n_g}\otimes M_gR_0M_g'|^{-\frac{1}{2}}|M_gR_0M_g'|^{-\frac{(v_g+2t_g+1)}{2}}$$

which, after using the formula for the determinant of a Kronecker product [6] is equal to:

$$\prod_{g=1}^{G}|M_gR_0M_g'|^{-\frac{n_g}{2}}|M_gR_0M_g'|^{-\frac{(v_g+2t_g+1)}{2}} = \prod_{g=1}^{G}|M_gR_0M_g'|^{-\frac{(v_g+n_g+2t_g+1)}{2}}. \quad (A.5)$$

Finally, multiplying (A.4) to (A.5) gives the density of R_0 as in (15):

$$\prod_{g=1}^{G}|M_gR_0M_g'|^{-\frac{(v_g+n_g+2t_g+1)}{2}}\exp\left\{-\frac{1}{2}\text{tr}\left(\left(E_g + M_gR_0^*M_g'\right)\left(M_gR_0M_g'\right)^{-1}\right)\right\}.$$