Estimation of Residual Dam-Offspring Correlation for a Maternal Animal Model through a Griddy Gibbs Sampler

S. Munilla Leguizamón[*] and R.J.C. Cantet[*]

Introduction

Mixed linear models used to fit records on maternally influenced traits are termed maternal animal models (MAM). The additive covariance structure in these models is based on the theory of the covariance among relatives of Willham (1963), and its formulation within the mixed linear models theory is indebted to Quaas and Pollak (1980). Under these models, strong, negative estimates for the direct-maternal genetic correlation lead to predictions of breeding values for the direct effects that tend to be opposite in sign than the predictions of maternal breeding values, which raises discredit among breeders who uses those predictions as selection criteria. Now, while a weak adverse association is considered plausible, such strong, negative estimates have been met with skepticism among researchers (Meyer (1997)). It is generally accepted that estimates of the direct-maternal genetic correlation are biased by a negative environmental association between maternal effects in adjacent generations (Koch, (1972); Baker (1980)). To avoid this bias, Cantet (1990) suggested including a covariance between the maternal permanent environmental effects and the residuals. This approach, however, induces a covariance structure that is non-linear in the correlation between dam-offspring pairs, which in turn hampers the parameter estimation. To overcome this problem, Bijma (2006) fitted a moving-average time-series, and obtained unbiased estimates using simulated data. However, this approach is only valid when the recorded dams have a single progeny. The goal of the present research is to develop an estimation procedure in the MAM with correlation between dam-offspring pairs in a more general setting. In particular, based on a Bayesian approach, we describe the estimation of the dam-offspring correlation parameter through a Griddy Gibbs sampler. The algorithm was used to fit the model on weaning weight records of beef calves, and preliminary results are shown.

Material and methods

Maternal animal model with residual dam-offspring correlation. Consider a maternally influenced trait, and assume therefore the covariance structure described by Willham (1953). As usual, we will use subscripts “*o*” and “*m*” to differentiate between direct and maternal effects, respectively. The model equation is then

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}_o\boldsymbol{a}_o + \boldsymbol{Z}_m\boldsymbol{a}_m + \boldsymbol{Z}_p\boldsymbol{e}_p + \boldsymbol{e}_o$$

---

[*] Departamento de Producción Animal, Facultad de Agronomía, UBA, Av. San Martín 4453, Buenos Aires, Argentina, y CONICET, Av. Rivadavia 1917, Argentina.

where $\boldsymbol{y}$ ($n \times 1$) is a data vector, and $\boldsymbol{X}$ ($n \times p$) is the incidence matrix for the fixed effects vector $\boldsymbol{b}$ ($p \times 1$). Additionally, $\boldsymbol{a}_o$ and $\boldsymbol{a}_m$ are ($q \times 1$) random vectors with entries corresponding to the direct and maternal breeding values, respectively, and $\boldsymbol{e}_p$ ($d \times 1$) is a random vector accounting for maternal permanent environmental effects. Accordingly, $\boldsymbol{Z}_o$, $\boldsymbol{Z}_m$ and $\boldsymbol{Z}_p$ are the corresponding incidence matrices. Finally, $\boldsymbol{e}_o$ ($n \times 1$) represents the vector of errors. To simplify the notation, let $\boldsymbol{Z} = [\boldsymbol{Z}_o \mid \boldsymbol{Z}_m]$. Following Bijma (2006), we assume further that

$$\boldsymbol{e}_o \mid \rho, \sigma^2_{e_o} \sim N\left(\boldsymbol{0}, \boldsymbol{R}\sigma^2_{e_o}\right), \quad \boldsymbol{R}_{(\mathrm{N}\times\mathrm{N})} = \left\{r_{ij}\right\}, \quad r_{ij} = \begin{cases} 1 & \text{if } i = j \\ \rho & \text{if } (i, j) \text{ is a dam - offspring pair} \\ 0 & \text{otherwise} \end{cases}$$

The mixed model equations could be succinctly written as $\left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{W} + \Sigma^{-1}\right)\hat{\boldsymbol{\theta}} = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{y}$, where, $\boldsymbol{W} = \left[\boldsymbol{X} \mid \boldsymbol{Z} \mid \boldsymbol{Z}_p\right]$ and $\Sigma^{-1} = \left[\begin{array}{c:c:c} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \left(\boldsymbol{G}_0^{-1} \otimes \boldsymbol{A}^{-1}\right)\sigma^2_{e_o} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_d\sigma^2_{e_o}\sigma^{-2}_{e_p} \end{array}\right]$. Note that $\sigma^2_{e_o}$ has been factored out and cancelled on both sides. From this setting, it is apparent that in order to estimate (co)variance components through a MCMC scheme, the computing effort entailed in setting-up the system in each cycle will be enormous, especially for large data sets. However, this problem is somehow alleviated given the particular structure of the matrix $\boldsymbol{R}^{-1}$. It is verifiable that if records are properly ordered, $\boldsymbol{R}^{-1}$ has a block diagonal structure, that is, $\boldsymbol{R}^{-1} = \oplus_{k=0}^{mf}\boldsymbol{R}_k^{-1}$. Each block $\boldsymbol{R}_k^{-1}$ ($k = 1, \ldots, mf$) is associated with the batch of individuals that belongs to the k^{th} maternal family. Each maternal family, in turn, is defined by the older recorded dam within a female pathway, and all her descendants. The block $\boldsymbol{R}_0^{-1}$, on the other hand, includes all the recorded animals that do not belong to any maternal family, and, of course, $\boldsymbol{R}_0^{-1} = \boldsymbol{I}$. However, for any maternal family, $\boldsymbol{R}_k^{-1}$ is dense. As females usually have less offspring than males, maternal families tend not to be large and, as a consequence, the direct inversion of the block will not become an insurmountable task.

Bayesian analysis. Consider a hierarchical Bayes construction for the model, where the objective is to make inferences about parameters of interest, typically the (co)variance components. Except for the correlation parameter, it is verifiable that all other full conditional posterior densities are analytically recognizable and thus can be sampled using standard procedures as those described by Jensen, J., Wang C., Sorensen, D. *et al.* (1994) for MAMs. On the other hand, it can be shown that

$$p\left(\rho \mid \boldsymbol{\theta}, \boldsymbol{G}_0, \sigma^2_{e_p}, \sigma^2_{e_o}, \boldsymbol{y}\right) \sim \left|\boldsymbol{R}\right|^{-1/2} \exp\left\{-\frac{\boldsymbol{e}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{e}}{2\sigma^2_{e_o}}\right\}$$

Note that ρ is within $\boldsymbol{R}$ and $\boldsymbol{R}^{-1}$ matrices, and that, as a function of ρ, the density is not analytically recognizable. At this point, we make use of the Griddy Gibbs sampler algorithm (Ritter and Tanner (1992)) to sample from the full conditional distribution of the parameter.

Griddy Gibbs sampler. As originally proposed, the Griddy Gibbs sampler is based on forming an approximation to the inverse cdf through the evaluation of the full conditional function on a grid of points. The function only needs to be known up to proportionality, because the normalizing constant can be obtained using the evaluated grid points. Once the cdf is approximated, the algorithm proceeds by sampling a uniform deviate and, afterwards, by interpolating the value between the closest grid points. In our case, we used a linear interpolation as suggested by Dr. Luc Bauwens (*pers. comm.*). A Griddy Gibbs sampler was specifically coded in FORTRAN 90.

Weaning weight data. Records of 7229 weaning weight of Angus calves, involving 533 maternal families, were analyzed using this approach. Data belongs to the firm “Estancias y Cabañas Las Lilas”, from Argentina. The model included the direct and maternal breeding values, a maternal permanent environmental effect, and fixed effects for *sex*, *age of dam*, *contemporary group*, and *age at weaning* (fitted as a covariate). Preliminary results are shown here.

Results and discussion

Marginal posterior summaries for the (co)variance components corresponding to one thousand MCMC cycles are shown in table 1. Two hundred cycles were discarded as burn-in. For some of the parameters, convergence has not yet been attained. At present, the program is under execution to obtain more samples. The execution time is about a minute per cycle in a personal computer with a Pentium® 4 (CPU 3.6GHz, 3.11 GB of RAM) processor, what posits a strong restriction to the usefulness of the approach. The main drawback is the computing time consumed in setting-up the mixed model equations in each cycle. Although the residual covariance matrix has a block diagonal structure, as we have pointed out, the fact that the inverse covariance matrix of each family is dense induces too many contributions to the coefficient matrix when there are large families in the data file. For instance, the larger maternal family in the Angus data set has more than one hundred individuals, which implies almost 50 million contributions to the coefficient matrix.

One of the drawbacks when implementing the Griddy Gibbs sampler was the evaluation of the full conditional function. This involves the exponential of the quadratic form in the residuals which renders a number so small that has no machine representation. To overcome this problem, we evaluated the log function instead, which requires summing up the logarithm of the determinant of each block matrix, and multiplying the result to the quadratic form. Now, to compute the normalizing constant, the sum of the logarithms for each grid point is needed. This was accomplished through the algorithm suggested by Primaeux (2005). However, when searching within the entire parametric space, some grid values has such a negligible probability that the algorithm failed. To overcome this problem, we then used an adaptive grid. First, a global grid covering the entire parametric space was evaluated, with grid values spaced at 0.05. Further, the point value that maximizes the log function was selected and, centered on it, a local grid, spaced at 0.01 was used to approximate the inverse cdf. As a measure of the performance of the procedure, 92% of the local grid points used engulfed the .95 cdf density mass, thus ensuring a quite precise interpolation.

Table 1: Posterior summaries for each variance component marginal density*

CVC[1]	$\sigma^2_{e_o}$	ρ	$\sigma^2_{e_p}$	$\sigma^2_{a_o}$	$\sigma_{a_o a_m}$	$\sigma^2_{a_m}$
Mean	463.66	0.07	96.56	170.65	-113.30	121.08
Mode	462.84	0.07	99.74	187.38	-129.51	125.86
SD	15.46	0.03	8.27	24.89	18.49	11.56
HPD95	(434.77, 494.37)	(0.01, 0.13)	(81.56, 112.34)	(123.05, 203.50)	(-138.59, -78.69)	(99.98, 143.30)

*1000 cycles. 200 cycles discarded as burn-in. [1](Co)variance components: $\sigma^2_{e_o}$ = residual variance; $\sigma^2_{e_p}$ = maternal permanent environmental variance; $\sigma^2_{a_o}$ = direct additive variance; $\sigma_{a_o a_m}$ = direct-maternal additive covariance, $\sigma^2_{a_m}$ = maternal additive variance; ρ = residual dam-offspring correlation.

Conclusion

In this study we showed that fitting a maternal animal model with residual dam-offspring correlation is computationally feasible, given the block diagonal structure of the residual covariance matrix. As we have explained, each block is associated with a maternal family present in the data file. Further, we described the estimation of the dam-offspring correlation parameter through a Griddy Gibbs sampler algorithm. As far as we know, we introduced the use of this algorithm for the first time in a livestock parameter estimation problem.

References

Baker, R (1980). In *Proc. N. Z. Soc. Anim. Prod.*, volume 40, pages 285–303.

Bijma, P. (2006). *J. Anim. Sci.*, 84:800–806.

Cantet, R. (1990). *PhD. Thesis.*

Jensen, J., Wang C., Sorensen, D. *et al.* (1994). *Acta Agric. Scand.*, 44:193–201.

Koch, R. (1972). *J. Anim. Sci.*, 35:1316–1323.

Meyer, K. (1997). *Livest. Prod. Sci.*, 52:187–199.

Primeaux, D. (2005). In *Proc. 6th ICSEAINP/DC*, pages 129–132.

Quaas, R. and Pollak, E. (1980). *J. Anim. Sci.*, 51:1277–1287.

Ritter, C., and Tanner, M. (1992). *J. Amer. Stat. Asoc.*, 87:861–868.

Willham, R. (1963). *Biometrics*, 19:18–27.