J. Anim. Breed. Genet. ISSN 0931-2668

REVIEW ARTICLE

Reliable computing in estimation of variance components

I. Misztal

University of Georgia, Athens, USA

Keywords
computing algorithms; MCMC; REML; variance components.

Correspondence
Ignacy Misztal, Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA.
Tel: +1-706-548-5961;
Fax: +1-706-583-0274;
E-mail: ignacy@uga.edu

Received: 17 March 2008;
accepted: 23 July 2008

Summary

The purpose of this study is to present guidelines in selection of statistical and computing algorithms for variance components estimation when computing involves software packages. For this purpose two major methods are to be considered: residual maximal likelihood (REML) and Bayesian via Gibbs sampling. Expectation-Maximization (EM) REML is regarded as a very stable algorithm that is able to converge when covariance matrices are close to singular, however it is slow. However, convergence problems can occur with random regression models, especially if the starting values are much lower than those at convergence. Average Information (AI) REML is much faster for common problems but it relies on heuristics for convergence, and it may be very slow or even diverge for complex models. REML algorithms for general models become unstable with larger number of traits. REML by canonical transformation is stable in such cases but can support only a limited class of models. In general, REML algorithms are difficult to program. Bayesian methods via Gibbs sampling are much easier to program than REML, especially for complex models, and they can support much larger datasets; however, the termination criterion can be hard to determine, and the quality of estimates depends on a number of details. Computing speed varies with computing optimizations, with which some large data sets and complex models can be supported in a reasonable time; however, optimizations increase complexity of programming and restrict the types of models applicable. Several examples from past research are discussed to illustrate the fact that different problems required different methods.

Introduction

A common task in any animal breeding project is variance component estimation, for which there are several desirable characteristics. If multiple traits are involved, it is desirable to analyse them jointly in a multitrait analysis. One also would like to account for the nature of these traits, e.g. continuous or categorical, quasi-normally distributed or not, censored, longitudinal, etc. Models must incorporate important effects; effects can be considered important when their absence causes important changes in parameters of interest in the analyses, and not necessarily just because they are statistically significant. Finally, it is desirable for the methodology, implementing the model of interest, to have a reasonable computational cost, with high reliability.

In reality, initially one often selects software based on availability and ease of use, then edits data and selects the model according to software limitations. If results are as expected, the analysis is considered successful. If results seem unreasonable, there are efforts, including model refinements and additional edits, to obtain 'reasonable' results. If time runs out, the efforts to refine analyses are usually replaced by efforts to justify the results. 'Unreasonable' results

can be because of data limitations, model limitations or problems with the estimation methodology.

Software currently used in parameter estimation in animal breeding uses predominantly two methodologies: residual maximal likelihood (REML) and Bayesian (Thompson *et al.* 2005), the last one usually through Gibbs sampling (BaGS). There are many algorithms and implementations of both REML and BaGS. The purpose of this paper is to examine properties of various implementations of REML and BaGS such that can easily be programmed in software packages, and to provide examples of their use from studies mainly conducted at the University of Georgia.

Methods

Restricted maximum likelihood

In the method of REML (see review by Druet & Ducrocq 2006), one maximizes the likelihood with respect to parameters, yielding estimates of parameters corresponding to the maximum of that likelihood. The maximization can be implemented in many ways, depending on the order of derivatives available (Thompson & Mantysaari 1999). The simplest class of algorithms is derivative-free (DF), where no derivatives are used. EM REML is the most popular methodology involving the first derivatives, and AI REML is the most popular algorithm involving the second derivatives. All methods are iterative. In animal model problems, where the mixed model equations (MME) are sparse, the left-hand side of the MME is stored as a sparse matrix, and computations involve sparse matrix factorization and often sparse matrix inversion (Misztal & Perez-Enciso 1993).

Assuming a general-model REML applied to normally distributed traits, computing properties of some REML algorithms were described by Misztal (1994) and Thompson & Mantysaari (1999). DF is the simplest to program and is fast for simple models; however, it is very expensive and potentially unreliable for complicated models such as multiple trait models. For these reasons, DF is rarely used nowadays. The EM algorithm was traditionally regarded as the most reliable, though slow in convergence, potentially taking a few hundred rounds. However, Foulley & Van Dyk (2000) found that the EM algorithm fails with random regression models, converging to a different set of estimates dependent on starting values. According to Tsuruta *et al.* (2004), this happens if the starting parameters are smaller than the converged estimates, but not if they are bigger. Additionally, EM does not generate standard errors of the estimates (SE) directly. While EM is slow, it can be accelerated, and while it is simple to implement if there are no missing traits; with missing traits, the formulae are much more complex (Mäntysaari & Van Vleck 1989).

The AI algorithm is quite complex to program; however, SE are obtained directly, and for many models, convergence is achieved in just a few rounds. However, for complex models and poor starting values, the AI algorithm can 'overshoot', resulting in either very slow convergence or parameter estimates out of the parameter space. Therefore, heuristics, such as the initial use of the EM algorithm, a PX (parameter-expanded EM) algorithm (Meyer 2006), or a reparameterization (as in Meyer & Smith 1996), are needed to ensure that next round of estimates are 'better' than the previous round, and are a crucial part of any AI REML software. Slow convergence of AI algorithms indicates that there is little available information on some parameters. One of the best solutions is to simplify the model.

Some new REML algorithms were discussed by Thompson & Mantysaari (1999). For example, a PX-REML algorithm is computationally similar to EM REML but seems to converge faster and is reliable with random regression models (Foulley & Van Dyk 2000). However, PX-REML is more difficult to implement for general models than EM REML and is not as fast as AI REML for well-behaving analyses (Meyer 2006).

Assuming estimation with the animal model and typical software, where MME are stored and solved by sparse methods, the cost per round for EM or AI increases proportionally by about $n^2\ t^3$, where n is the number of animals and t is the number of traits. Memory requirements increase by $n^{1.5}\ t^2$, respectively (Misztal 1994). The increases are substantial enough to be a limiting cost factor for large models analysed using REML methodologies. In practical studies these limits could be around 200 000 animals, 2 million equations and five traits, although specific models and patterns of data can affect these limits.

Another issue with the accuracy of REML is its stability with a large number of traits. As the number of traits increase, the variance–covariance matrices tend to become almost semi-positive definite. Subsequently, MME become numerically poorly conditioned resulting in poor estimates. Thus, a general-model REML implementation where the

equations are set up explicitly is unsuitable for more than 5–10 traits.

When the number of traits is high, one option for parameter estimation is the canonical transformation (CT), where a *t* trait model is transformed to *t* single-trait models (Hayes & Hill 1980). The cost of CT increases proportionally to *t*, with little increase of memory and no problem with stability; >30 trait models have been successfully run with CT. However, CT is restricted to models with no missing traits, having the same design matrix for all traits, and a single random effect; approximations exist that lift the last restriction (Lin & Smith 1990).

REML formulae are well known in linear models when traits are distributed normally. When trait(s) are not distributed normally, the formulae can be very complicated, an approximation of the true formulae or unknown (Tempelman 1998). For example, an approximation of REML applied to threshold models results in increasingly inflated estimates of heritabilities with increasing number of levels of fixed effects.

BaGS

Another popular methodology for variance estimation is BaGS; see extensive coverage by Gianola & Sorensen (2002). In the simplest implementation, it involves repeated solving of MME by Gauss–Seidel while adding noise to solutions. Sampling of variance components is from chi-square or Wishart distributions. Samples of variances from initial rounds of sampling are discarded, and samples from later rounds for each (co)variance are assumed to be from the marginal distribution of that (co)variance. The total number of rounds needs to be large enough to collect a sufficient number of effective samples. By analysing samples, one can estimate modes and means, and their SE. For linear models, typically the estimates of the modes are very similar to the corresponding estimates from REML although estimates from REML are joint, not marginal modes. In practice, differences between estimates from REML and BaGS are often traced to problems with BaGS, e.g. insufficient length of chain or implementation problems.

One advantage of BaGS is simplicity in theory and programming. A mixed model program can be converted to BaGS by adding procedures for random number generations and a few statements for sampling. Thus, memory requirements are very small as no factorization or sparse inversion is required. In addition, modifications to more complex models can be performed relatively easily. For example, if some traits are missing, one predicts missing traits and then continues as if all traits were recorded. In threshold models, one predicts liabilities and then proceeds as if the model was linear. If a trait is censored, the censored values are predicted and the analysis continues as if traits were all recorded. BaGS allows for analyses of models that seem to be intractable under REML including highly hierarchical models, however, many such models require expert knowledge to run and cannot easily be implemented in a package.

The speed of BaGS depends on the optimization. The number of samples required or chain size may vary from 1000 to millions; a number around 20 000 is often used. If mixed model equations are recreated every round, the run may take a very long time. However, optimizations can greatly decrease the time per round, and many types of optimizations are possible. For example, with iteration on data, there is no need to create the system of equations explicitly. For some types of models, such as test-day random-regression models in dairy cattle, coefficients of random regressions can be recreated every round, decreasing the amount of data to be read (Jamrozik 2004). Additionally, all animal effects can be sampled at once, thus increasing the mixing rate. Although the iteration on data suitable for BaGS is easy with some models, it is complicated for general models. Another optimization is storing parts of the left-hand side of MME separately as contributions to data, relationship matrices and variance components. All of these parts are then combined during the iteration without the need to recreate the system of equations.

Another way of decreasing the run time of BaGS is via a reparameterization that decreases the number of unknowns. A mixing rate several times faster was reported when the regular animal model was converted to a reduced animal model (Quaas, personal communications). An extra benefit of this reparameterization was a smaller cost per round as the number of non-zero coefficients with the reduced animal model decreased. However, not all reparameterizations are easy to program or reduce the computing time per sample.

The quality of estimates from BaGS depends mainly on quality of pseudo-random number generators and the choice of burn in and chain length. There are many methods to determine the last two parameters (e.g. Garcia-Cortes *et al.* 1998); however, it is not an exact science. Aside from statistical criteria, a common way of determining the burn in is by visual inspection of samples.

BaGS can be used with flat or informative priors on parameters. In practical use, the main advantage of informative priors is ensuring convergence that otherwise may not be achievable with flat priors. The lack of convergence with flat priors may be because of many factors including the impropriety of the posterior distribution (Hobert & Casella 1996) or owing to numerical problems, e.g. imperfect random number generators, imperfect correction of non-positive matrices, etc. A disadvantage of informative priors is the extra work needed to specify them and the potentially undesirable effect of 'biasing' the estimates. In common analyses involving models that can also be analysed by REML, there are few problems caused by flat priors, however, they are more likely to occur in small data sets and with non-linear models (Stock *et al.* 2007). Proper priors can be obtained by such simple terms as imposing a large variances or large bounds on parameters of the model. It is possible that such bounds are implicitly provided by pseudo-random number generators.

In general, estimates from BaGS will be different each time an analysis is run because random sampling is involved. Therefore, estimates will be similar for parameters with more information (and small SE), but may be very different for parameters with little information. For the same reasons, a large number of samples is needed with small data sets; repeated runs are advised for validation. Results with small data sets stress the quality of pseudo-random number generators. For comparison, repeated applications of the REML methodology result in identical estimates, even with small data sets.

Nearly always, BaGS results in smaller, and sometimes much smaller memory requirements than REML. With various optimizations, the running time can be proportional to nt, where n is the number of animals and t is the number of traits. However, the total time may still be very large as the number of rounds to obtain sufficient number of samples can be very high. While running REML is fairly automatic, as the convergence point is obvious, running BaGS may require much more involvement. In particular, one needs to determine the length of burn in, the number of rounds for a sufficient number of effective samples and also to make sure that no unexpected events occurred, e.g. resets of chain or divergence.

General points

Many problems in the estimation process are unrelated to methodologies. For any estimation procedure, Karin Meyer (personal communication) recommends that any estimation effort can be performed in stages. They would include simple statistics, single-trait analysis with simple models and then increasingly more complex analysis. Subsequently, potential problems can be localized more easily.

Problems with variance component estimation often can be traced to unnecessarily high model complexity with too many parameters to estimate. As Box and Draper (1987) stated, 'Essentially, all models are wrong, but some are useful', thus the search for a perfect model is futile. While more complex models may be needed to reveal the biology of traits, simpler models may suffice for genetic evaluation. For example, Lopez-Romero & Carabano (2003) compared random regression models using Legendre polynomials of orders 2–6. While more complex models fit the data better, the predictive ability of all the models was almost identical, indicating almost identical rankings of sires. Good arguments for following productivity in model comparisons were made by Blasco (2006). Reports from literature that simple and complicated models provide similar estimated breeding values are abound, e.g. Piles *et al.* (2006).

Recommendations

Assume that a scientist has access to one or more software packages and that limited modifications to that package(s) are possible though not desirable. The strategy for a successful estimation effort could be as follows.

1. Attempt to run single-trait analyses by AI REML. Verify that estimates are realistic.
2. Move to multiple traits analyses, first through a combination of traits, then all traits if the number of traits is smaller than 5–7. Large discrepancies between single and multiple trait estimates indicate problems.
3. If estimates are unrealistic, check the data for errors. In addition, refine the model by adding or removing additional non-genetic effects. Small changes in estimates after removing or adding an effect suggest that such an effect is unimportant. Large changes mean the opposite.
4. If estimates by AI REML seem to strongly fluctuate or be out of the parameter space, try EM REML.
5. If time per round in either AI or EM REML seems long or the analyses do not finish owing to excessive computing requirements, move to CT REML (if applicable and number of traits high) or to BaGS. When the number of traits is large, these are the only applicable methodologies.

6. When the number of records is very large, the only applicable strategy would be optimized BaGS.
7. With BaGS, choose appropriate optimization.
8. With BaGS, always analyse samples and ensure that the available number of samples exceeds burn in and that the number of effective samples is sufficient. If estimates by any method seem unstable or SE are very high, consider simplifying the model.
9. If chain in BaGS seems erratic or has loops, verify whether adding priors (if initially flat) or making them stronger (but not too strong) eliminates problems.
10. If analyses by AI seem to finish complete but the results seem doubtful, repeat computations by BaGS. For most models, posterior modes by BaGS and estimates by REML should be similar.

Case studies

Case studies illustrate the use of the aforementioned rules. These studies almost always used software from the BLUPF90 suite of programs (Misztal 1999); major programs in this suite are listed in the Appendix. The purpose of having different programs instead of one 'superprogram' is to keep those programs simple. Nowadays, few animal breeders have substantial programming skills, and keeping the programs simple makes modifications of these programs possible, if needed, for those moderately skilled in programming. Moreover, simpler programs are not only likely to contain fewer bugs but fixing those bugs is likely to be faster.

Relationships between stillbirth, mortality and birth weight in pigs

In a three-trait study by Arango *et al.* (2006), stillbirth status was a categorical trait, birth weight was available only for piglets born alive and preweaning mortality was a binary-longitudinal trait of days at death. Stillbirth could be treated as a trait of the piglet, useful for estimating correlations with birth weight, or (as a count or categorical variable) a trait of the sow. Computations used THRGIBBSF90 as the only program that could support multitrait threshold-linear model. No convergence was achieved with the random regression model for mortality as this trait was binary, the incidence of death was relatively low and most of the deaths occurred at the beginning of the trajectory. Subsequently, this trait was split into early and late mortality. Still, the attempts to have one multitrait analysis were unsuccessful, requiring four different models to be run. Some runs needed a chain size as long as 250 000 and took a few days for computing. In general, one should expect estimation problems with multiple trait threshold models especially involving the animal model.

Genetic correlations between performance of pure- and cross-breds in pigs

In a study by Zumbach *et al.* (2007), data on both pure- and cross-breds were available. Purebreds had fairly complete pedigrees, while cross-bred had pedigrees mostly on the paternal side. While the maternal side, which was cross-bred, was identified, it had no recorded pedigrees. In a model, the effects of pure- and cross-breds were fit as separate traits. The genetic effect was animal for purebreds and sire for cross-breds; both effects were assumed correlated.

Computations used AIREMLF90. When the maternal effect was included for purebreds, the number of rounds to convergence increased from 5 to about 200. When the maternal effect was removed, the variance of the litter effect increased, and the remaining variances were unaffected. Whenever AI REML requires a large number of rounds, this is an indication of either an over-parameterized model or of little information available to estimate a particular variance. It was the second case in this study as litters and dams were nearly confounded.

Components of sow disposal

In a study by Arango *et al.* (2005b) there were three traits: age at disposal for fertility, age at disposal for reproductive reasons and age at disposal for remaining reasons. The ages of disposal from 2–5 could be treated as categorical. As the sow was disposed only for one reason, there was always one trait recorded and two traits censored. Computing focused on modifying THRGIBBS1F90 for censoring. No convergence was achieved with the threshold animal model. The estimation was successful when traits were regarded as linear. Heringstad *et al.* (2004) were able to achieve convergence using the multiple trait threshold sire model. It is not clear whether gains achieved with the threshold model are greater than losses incurred by neglecting some relationships with the sire model; for discussions, see Ramirez-Valverde *et al.* (2001).

Competition model

A model for social competition by Muir (2005) contains one direct effect and as many associated effects

as the number of animals in a pen minus one. All associated effects contribute to the same block of MME, and the direct and associated effects are correlated. Additionally, as the number of animals per pen varies, the number of effects in the model also varies by observation. The BLUPF90 parameter file supported the competition model directly. While more optimized programs did not work, the simplest versions: REMLF90 and GIBBSF90 worked correctly as verified by simulation.

The estimation with the commercial data set was unsuccessful (Arango *et al.* 2005a). Consequently, the model was simplified by setting the correlation between the direct and associative effects to 0. Then, REMLF90 was used to calculate likelihoods for a number of combinations of variances, mimicking a derivative-free REML. While the maximum of the likelihood for the variance of the associative effect was >0, the likelihood was very flat, indicating little information in the data about that effect. This was related to large pen sizes and incomplete pedigrees.

As noted by Cantet & Cappa (2008), the Muir (2005) model has the pen and competitive effects confounded; Arango *et al.* (2005a) treated the effects of pens as random. Moreover, the model by Muir (2005) makes a strong assumption that expression of the competitive effects is linear; better estimates may be obtained in a more realistic model, e.g. as proposed by Misztal & Rekaya (2007). Thus, problems to estimate specific variances could be not only because of numerical problems or lack of information but also because of unrealistic models.

Stochastic censoring in days from exposure to conception in beef

The traits analysed in a study by Urioste *et al.* (2007a) were days from exposure of bull to conception, for three consecutive seasons. These traits were censored as some animals did not conceive within the season. Analyses were by several multitrait models: (i) penalized, where non-calving cows had their record set to maximum days of exposure + penalty; (ii) regular censored, where missing traits were predicted from days past the exposure; and (iii) threshold-linear, where missing traits were treated as missing, but binary traits of calving success were added. The last model accounted for stochastic, rather than truncation, censoring (Foulley 2004). The regular censored procedure used the program TM by Andres Legarra, which uses informative priors, and the remaining analyses used THRGIBBS1F90.

Initially inconsistent results were caused by inadequate systematic effects. Mixing in all models was slow and 200 000 samples were needed for convergence. The six-trait model was by far the most accurate (Urioste *et al.* 2007b), probably because the assumption of stochastic censoring was more realistic than of truncation censoring. For example, an inference that long days to calving in the first parity greatly decreased chances of calving success in the second parity could not be reached with simpler models.

Changes in genetic correlations over time

Tsuruta *et al.* (2005) looked into changes in genetic correlations between productive life and production and type traits over time, for a total of 24 traits. Because substantial records on productive life are available only when bulls are old, there is an interest in predicting productive life indirectly from other traits. However, such prediction is strongly dependent on correctness of the genetic correlations among the traits, and therefore there was an interest in finding whether these correlations change much over time. A model to look at such changes included linear or quadratic random regression on year of birth. Not all traits were recorded, and the traits formed several clusters of fixed effects. The variance–covariance matrix had an order up to 72. Only the optimized Gibbs sampler could support models with such a large number of traits especially if the estimated matrix was not full rank. Computing was by GIBBS2F90 and took about 1 month.

Many correlations between the productive life and the remaining traits decreased or even changed signs over time. For example, for animals born in 1980 the milk yield and stature had correlations of 0.4 with PL. For cows born in 1998, the correlation with milk yield reduced to 0.1 and became negative with stature. To validate the changes estimated by the random regression model, sire model analyses (using GIBBS1F90) were run with data restricted to a few years first around 1980 and then around 1998; changes in genetic correlations predicted by the both models were similar.

Conclusions

Different problems require different methodologies. AI REML is the method of choice if it is converging and its cost is reasonable. If AI REML does not converge reliably, EM REML may, although it is much slower and caution with starting values is needed

with random regression models. Optimized Bayesian methods via Gibbs sampling are the methods of choice when the data sizes are large and the number of traits is high. However, running the Bayesian analyses is not automatic. Bayesian analyses are also the methods of choice for models that are too hard or impossible to program for REML. Additional issues affect the estimation process regardless of the methodology. Modelling fixed or systematic effects is often more important than modelling random animal effects. Model complexity needs to be balanced with the amount of data. Some models may seem far from 'true' but they may be very good approximations that are much easier to compute. Finally, caution is required with the use of statistical criteria for model comparisons, where more complex models are preferred despite little practical improvement over much simpler models.

Acknowledgements

The invitation to prepare this study by Katja Grandinson, the encouragement to write this paper by Asko Maki-Tanila, comments and editing by Kelly Robbins and discussions with Romdhane Rekaya are gratefully acknowledged.

References

Arango J., Misztal I., Tsuruta S., Culbertson M., Herring W. (2005a) Estimation of variance components including competitive effects of Large White growing gilts. *J. Anim. Sci.*, **83**, 1241–1246.

Arango J., Misztal I., Tsuruta S., Culbertson M., Herring W. (2005b) Study of codes of disposal at different parities of Large White sows using a linear censored model. *J. Anim. Sci.*, **83**, 2052–2057.

Arango J., Misztal I., Tsuruta S., Culbertson M., Holl J.W., Herring W. (2006) Genetic study of individual preweaning mortality and birth weight in Large White piglets using threshold-linear models. *Livest. Sci.*, **101**, 208–218.

Blasco A. (2006) Philosophy of science and animal breeding research. In: 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, 13–18 August 2006.

Box G.E.P, Draper N.R. (1987) . *Empirical Model-Building and Response Surfaces*, Wiley.

Cantet R.J.C., Cappa E.P. (2008) On identifiability of (co)variance components in animal models with competition effects. *J. Anim. Breed Genet.*, **125**, 371–381.

Druet T., Ducrocq V. (2006) Innovations in software packages in quantitative genetics. In: 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, 13–18 August 2006.

Foulley J.L. (2004) Prediction of random effects in linear mixed models under stochastic censoring. Proceedings of the International Conference on Statistics and Health Science, Nantes, France, 23–25 June. University of Nantes, Nantes, France, pp. 129–132.

Foulley J.L., Van Dyk D.A. (2000) The PX-EM algorithm for fast stable fitting of Henderson's mixed model. *Genet. Sel. Evol.*, **32**, 143–163.

Garcia-Cortes L.A., Rico M., Groeneveld E. (1998) Using coupling with the Gibbs sampler to assess convergence in animal models. *J. Anim. Sci.*, **76**, 441–447.

Gianola D., Sorensen D. (2002) Likelihood, Bayesian and MCMC Methods in Quantitative Genetics. Springer-Verlag Inc., New York.

Hayes J.F., Hill W.G. (1980) A reparameterisation of a genetic selection index to locate its sampling properties. *Biometrics*, **36**, 237–248.

Heringstad B., Chang Y.M., Gianola D., Klemetsdal G. (2004) Multivariate threshold model analysis of clinical mastitis in multiparous Norwegian dairy cattle. *J. Dairy Sci.*, **87**, 3038–3046.

Hobert J.P., Casella G. (1996) The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *J. Am. Stat. Assoc.*, **91**, 1461–1473.

Jamrozik J. (2004) Implementation issues for Markov Chain Monte Carlo methods in random regression test-day models. *J. Anim. Breed. Genet.*, **121**, 1–13.

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

Lopez-Romero P., Carabano M.J. (2003) Comparing alternative random regression models to analyse first lactation daily milk yield data in Holstein–Friesian cattle. *Livest. Prod. Sci.*, **82**, 81–96.

Mäntysaari E., Van Vleck L.D. (1989) Restricted maximum likelihood estimates of variance components from multitrait sire models with large number of fixed effects. *J. Anim. Breed. Genet.*, **106**, 409–422.

Meyer K. (2006) PX x AI: algorithmic for better convergence in restricted maximum likelihood estimation. Proceedings of the Eighth World Congress on Genetic Applied to Livestock Production, communication no. 24-15.

Meyer K., Smith S.P. (1996) Restricted maximum likelihood estimation for animal models using derivatives of the likelihood. *Genet., Sel., Evol.*, **28**, 23–49.

Misztal I. (1994) Comparison of computing properties of derivative and derivative-free algorithms in variance component estimation by REML. *J. Anim. Breed. Genet.*, **111**, 346–355.

Misztal I. (1999) Complex models, more data: simpler programming. Proceedings of the International

Workshop on Computation and Cattle Breeding, 18–20 March, Tuusala, Finland. *Interbull Bull.*, **20**, 33–42.
Misztal I., Perez-Enciso M. (1993) Sparse matrix inversion in restricted maximum likelihood estimation of variance components by expectation-maximization. *J. Dairy Sci.*, **76**, 1479.
Misztal I., Rekaya R. (2007) A social competitive model with the categorical expression. *J. Anim. Sci.*, **84**(Suppl. 1), 416.
Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., Lee D.H. (2002) BLUPF90 and related programs (BGF90). In: 7th World Congress on Genetics Applied to Livestock Production, Montpelier, 19–23 August 2002.
Muir W.M. (2005) Incorporation of competitive effects in forest tree or animal breeding programs. *Genetics*, **170**, 1247–1259.
Piles M., Garcia M.L., Rafel O., Ramon J., Baselga M. (2006) Genetics of litter size in three maternal lines of rabbits: repeatability versus multiple-trait models. *J. Anim. Sci.*, **84**, 2309–2315.
Ramirez-Valverde R., Misztal I., Bertrand J.K. (2001) Comparison of threshold and linear models for calving difficulty in beef cattle. *J. Anim. Sci.*, **79**, 333–338.
Stock K., Distl O., Hoeschele I. (2007) Influence of priors in Bayesian estimation of genetic parameters for multivariate threshold models using Gibbs sampling. *Genet. Sel. Evol.*, **39**, 123–137.
Tempelman R.J. (1998) Generalized linear mixed models in dairy cattle breeding. *J. Dairy Sci.*, **81**, 1428–1444.
Thompson R., Mantysaari E.A. (1999) Prospects for statistical methods in dairy cattle breeding. *Interbull Bull.*, **20**, 70–79.
Thompson R., Brotherstone S., White I.M.S. (2005) Estimation of quantitative genetic parameters. *Philos. Trans. R. Soc. B*, **360**, 1469–1477.
Tsuruta S., Misztal I., Lawlor T.J., Klei L. (2004) Modeling final scores in US Holsteins as a function of year of classification using a random regression model. *Livest. Prod. Sci.*, **91**, 199–299.
Tsuruta S., Misztal I., Lawlor T.J. (2005) Current day's prediction of a changing trait. Productive life of US Holsteins. *J. Dairy Sci.*, **88**, 1156–1165.
Urioste J.I., Misztal I., Bertrand J.K. (2007a) Fertility traits in spring-calving Aberdeen Angus cattle. 1. Model development and genetic parameters. *J. Anim. Sci.*, **85**, 2854–2860.
Urioste J.I., Misztal I., Bertrand J.K. (2007b) Fertility traits in spring-calving Aberdeen Angus cattle. 2. Model comparison. *J. Anim. Sci.*, **85**, 2861–2865.
Zumbach B., Misztal I., Tsuruta S., Holl J., Herring W., Long T. (2007) Genetic correlations between two strains of Durocs and crossbreds from differing production environments for slaughter traits. *J. Anim. Sci.*, **85**, 901–908.

Appendix

Table 1 lists selected computer programs for variance component estimation mostly in the BLUPF90 suite of programs (Misztal *et al.* 2002). The programs as well as course notes that describe the design of those programs including some optimizations are available at: http://nce.ads.uga.edu/~ignacy.

Table 1 List of selected programs in the BLUPF90 suite.

Program	Description	Comments
REMLF90	EM REML with Aitken acceleration	
AIREMLF90	Average information REML	Initially a mix of EM and AI
AIREMLRES	Average information REML	As above with support for heterogeneous residuals
GIBBSF90	Simple BaGS	Coefficient matrix recreated every round
GiBBS1F90	Optimized BaGS	Left-hand side dynamically recreated
GIBBS2F90	Optimized BaGS	As above but all correlated effects sampled jointly
GIBBS3F90	Optimized BaGS	As above but with support for heterogeneous residuals
THREGIBB1F90	Optimized BaGS for threshold-linear models	Modification of GIBBS2F90
PPOSTGIBBSF90	Graphical analyser of BaGS samples	
CBLUP90REML	Quasi REML for single-trait threshold models	Only stable for sire models
MTC	Canonical transformation REML	Older design; different parameter file

REML, residual maximal likelihood; BaGS, Bayesian Gibbs sampling; MTC, Multiple Trait Canonical.

