

Rothamsted Repository Download

A - Papers appearing in refereed journals

Thompson, R. and Mantysaari, E. 2004. Prospects for statistical methods in animal breeding. *Journal of the Indian Society of Agricultural Statistics*. 57 (Special Volume), pp. 15-25.

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/895yw>.

© Indian Society of Agricultural Statistics.

Prospects for Statistical Methods in Animal Breeding

Robin Thompson¹ and Esa Mantysaari²
Rothamsted Research, Harpenden, Herts AL5 2JQ, England

SUMMARY

Accurate prediction of breeding values is of great importance for animal improvement programmes. The prediction of breeding values requires knowledge of the magnitude of the variances and covariances of random effects. This paper gives a short review of methods of estimation of genetic variance parameters, contrasting analytical estimates with iterative and sampling based methods.

Key words : Variance component estimation, Residual maximum likelihood, Derivative free methods.

1. Introduction

A recent GIFT workshop had two papers that discussed future dairy cattle research. The two papers (Goddard [10] and Hill *et al.* [16]) were in good agreement of the future statistical needs. These included methods for test day models, international comparisons, non additive variance, non-linear models and individual gene models. They also highlighted a trend to more sophisticated analysis leading to less biased predictions and more progress at the expense of greater variance or risk. Cost of analysis was suggested to be small compared to the cost of collection of data. There was a concern that uncertainties in parameters might erode possible gains. There was also a hope that prediction were robust to bad luck. Variance parameter estimation plays an integral role in several of these topics. We therefore intend to review this area hoping to identify themes that will lead to more rapid change. Prem Narain has made an outstanding contribution to statistical genetics, including the application of statistics to plant and animal breeding, so we think our small paper contribution is particularly apt to this volume in honour of Prem's 70th birthday.

¹ Roslin Institute(Edinburgh),Roslin, Midlothian EH25 9PS, Scotland

² Agricultural Research Centre, MTT, Institute of Animal Production, FIN-31600, Jokioinen, Finland

2. Variance Component Estimation

We consider a linear model

$$y = Xb + Zu + e$$

$$\text{with } \text{var}(y) = ZGZ' + R, \text{var}(u) = G, \text{var}(e) = R$$

The matrices G and R are often linear functions of unknown genetic parameters such as genetic and phenotype variance. Estimation of variance and covariances by Residual Maximum Likelihood (REML) (Patterson and Thompson [33]).

Here is often the method of choice. The log-likelihood is of the form

$$L \propto (y - X\hat{b})'V^{-1}(y - X\hat{b}) - \log \det(V) - \log \det(X'V^{-1}X)$$

This is different from the usual likelihood form in that it is a function of error contrasts – contrasts that do not tell us about fixed effects. This difference has two consequences, the use of the weighted least squares estimate of b , \hat{b} given by

$$X'V^{-1}X\hat{b} = X'V^{-1}y$$

The term in $\det X'V^{-1}X$ that is sometimes thought of as a penalty function because the fixed effects are not known. Mixed model equations (Henderson [15]) pay an important part in the analysis process. These are of the form

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

Terms derived from these include prediction error variances found from writing the mixed model equations as

$$Cs = R$$

so that

$$\text{var} \begin{bmatrix} \hat{b} \\ \hat{u} - u \end{bmatrix} = C^{-1}$$

It is often useful to express relevant quantities in terms of the projection matrix

$$P = N^{-1} - X(X'V^{-1}X)^{-1}X'V^{-1}$$

$$L \propto y'Py - \log \det(V) - \log(X'V^{-1}X)$$

Estimation of a variance parameter θ_i involves setting to zero the first derivatives

$$\partial L / \partial \theta_i = y'P(\partial V / \partial \theta_i)Py - \text{tr}[P(\partial V / \partial \theta_i)]$$

These could be thought of equating a function of the data to its expectation.

Normally finding a maximum of the likelihood requires an iterative scheme. One suggested by Patterson and Thompson [33] is based on the expected value of the second differential, i.e.

$$E(\partial^2 L / \partial \theta_i^2) = -(1/2) \text{tr}[P(\partial V / \partial \theta_i)P(\partial V / \partial \theta_i)P]$$

This is called the Expected Information. Using the first and second differentials we can update θ using the rate that all the terms from solution of MME and C^{-1} for example

$$\hat{\theta} = \theta + E \text{Inf}^{-1}(\partial L / \partial \theta)$$

Whilst this development is very direct, later developments tried to take account of the structure to reduce the computational effort. For example eliminating u from the mixed model equations gives weighted least squares equations for \hat{b} and \hat{u} calculated as $\hat{u} = (Z'R^{-1}Z - G^{-1})^{-1}Z'R^{-1}(y - X\hat{b})$ and similarly C^{-1} as $(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}X(X'V^{-1}X)^{-1}X'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}$ where the second term is a correction for the uncertainty in \hat{u} . The correction term used in the first differential can be written as

$$\text{trace} [(X'V^{-1}X)^{-1}X'R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}X]$$

Showing that not all the elements of C^{-1} need calculation in order to form the first differential (Thompson [40]).

An alternative algorithm was suggested by Dempster *et al.* [2].

This EM algorithm is based on thinking of the random effects 'missing'.

The estimation is based on using

$$s\hat{\sigma}_g^2 = u'u + \text{PEV}(u) \text{ writing this as}$$

$$s\hat{\sigma}_g^2 = y'V^{-1}(\partial G / \partial \theta_i)V^{-1}y + s\sigma_g^2 - \text{tr}[V^{-1}(\partial L / \partial \theta)]$$

we see this as a manipulation of equating first differential to zero. It can also be written as $\hat{\theta} = \theta + \text{Inf}^{-1}(\partial L / \partial \theta)$ with Inf representing the information on the complete data. One advantage of this method is σ_g^2 that stays in the parameter space $\sigma_g^2 \geq 0$.

Another advantage is that there is an increase in likelihood in each iteration. Disadvantages are that the method can be slow to converge (indeed this method is said to be the most widely used in terms of numbers of iterations) and it requires the inversion of C in each iteration.

An important development was the introduction by Smith and Graser [34] of an alternative form for the likelihood that naturally leads to sequential formation of the likelihood

$$L = \log \det(R) - \log \det(G) - \log \det(C) - y'Py$$

If we write equations for $n+1$ variables in the form

$$\begin{bmatrix} X_{nn} & x_{nn+1} \\ x_{n+1n} & x \end{bmatrix} \begin{bmatrix} u_n \\ u_{n+1} \end{bmatrix} = \begin{bmatrix} y_n \\ y \end{bmatrix}$$

then the contribution from the $(n+1)^{\text{th}}$ term to $y'Py = y^2/x$ and to $\log \det(C) = \log(x)$. Using regression coefficient $r_{n+1} = x_{n+1n}/x$ a correction to $X_{nn} = X_{nn} - x_{nn+1}r_{n+1n}$, and $y_n = y_n - y r_{n+1}$ and can be formed and the procedure repeated with $n = n - 1$. If x_{nn+1} is sparse then this can lead to a reduction in calculations, especially if it is taken to reorder equations to keep X_{nn} sparse.

To maximize the likelihood with one parameter Smith and Graser [35] suggested using a quadratic approximation. With more than one parameter simplex methods become a popular flexible alternative as they avoid calculating derivatives. The methods were used for Animal and Reduced Animal Models, both for univariate and multivariate data (Meyer [29]). After more biological appropriate models with genetic components naturally fitted into their framework including maternal models with both Wilham and Falconer terms (Koerhuis and Thompson [19]) and models with mutation terms (Wray [42]).

However it was realized that the computational effort for derivative-free methods increased dramatically as the number of variance parameters increased.

An important advance was the rediscovery (Misztal and Perez-Enriso [31]) of an algorithm (Takahashi *et al.*, [37]) that allowed the calculation of the 'relevant' terms in the inverse of C required for forming the first differentials without calculating all the elements of the inverse. Meyer and Smith introduced an alternative way of calculating these first differentials by performing the 'automatic' differentiation of the Choleski decomposition of C . These techniques both requiring twice the computational effort of forming the likelihood were derived using properties of Choleski decompositions. An alternative derivation in terms of sequential of C^{-1} parallels the sequential formation of the likelihood (Thompson *et al.* [39]). If X_{nn}^{-1} contains the partition of C^{-1} for the first n elements then terms in C^{-1} for the first $n + 1$ elements are given by

$$\begin{bmatrix} X_{nn}^{-1} & c_{nn+1} \\ c_{n+1n} & c \end{bmatrix}$$

with $c_{n+1n} = x_{n+1n} X_{nn}^{-1}$
 and $c = \frac{1}{x} + c_{n+1n} x_{nn+1}$

Terms in c_{n+1} only need to calculating if the terms x_{n+1n} are non-zero leading to a major reduction in computation.

This result allowed the implementation of EM algorithms to estimate variance parameters, (Misztal [30]). These were an improvement on derivative free methods but could still be slow to converge. It is possible to calculate second differentials using the automatic differentiation ideas of Smith [34] but the calculation of each second differential requires the computation of the order of six likelihood calculations (Smith [34]) and this becomes more costly as the number of parameters increase. There are various suggestions on approximating the second differential. Mantysaari and Van Vleck [25] suggest accelerating the EM algorithm based on the observed geometric rate of convergence. Neumaier and Groenevald [32] suggest quasi-Newton scheme using first differential values to build up an approximate second differential. A third suggestion by Thompson and co-workers (Johnson and Thompson [18], Gilmour *et al.* [9]) is based on manipulation of the alternative information matrices.

The second differential of C with respect to θ_i and θ_j is as

$$\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = \frac{1}{2} \text{tr} \left[P \frac{\partial V}{\partial \theta_i} P \frac{\partial V}{\partial \theta_j} \right] - y' P \frac{\partial V}{\partial \theta_i} P \frac{\partial V}{\partial \theta_j} P y$$

and

$$E \left[\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right] = -\frac{1}{2} \text{tr} \left[P \frac{\partial V}{\partial \theta_i} P \frac{\partial V}{\partial \theta_j} \right]$$

Both these terms often called observed and expected information are difficult to calculate but the average

$$AI \left[\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right] = -\frac{1}{2} \text{tr} \left[P \frac{\partial V}{\partial \theta_i} P \frac{\partial V}{\partial \theta_j} \right]$$

can be calculated by using $\frac{\partial V}{\partial \theta_i} P y$ and $\frac{\partial V}{\partial \theta_j} P y$ as working variables and obtaining the residual cross product between these working variables. This calculation is much simpler than calculating either the observed and expected information.

A synthesis of comparisons of these methods was carried out by Hofer [17] and is updated in Table 1. These show the expected improvement of EM methods over derivative free methods. They also show that most second differential methods converge in relatively small number of iterations. Rather

embarrassingly I think that theoretical calculations suggest that the Jensen *et al.* (1997) times for the AI method can be improved dramatically.

In some cases transformations can aid in estimation. If we have multivariate data with two ($p \times p$) variance matrices to estimate, say G and R, then a economical transformation (Thompson [41], Meyer [29]) can help in reducing one $p \times p$ estimation into p independent analyses. They are modifications using the EM algorithm that allow the same techniques to be used with missing values (Ducrocq [5]) and with different designs with different variates (Ducrocq and Chapuis [6]).

A related problem is that often we require G and R to be positive definite and schemes based on second differentials do not necessarily lead to positive definite matrices. One suggestion is to use transformed parameters for example σ or $\log \sigma$ instead of σ^2 , or multivariate analogues such as Choleski transformations (Lindstrom and Bates [21]).

Table 1. Results of empirical comparison of REML algorithms with regards to rounds of iteration (function evaluations for DF and total time (h) to convergence

Ref ^b	MME ^c	Par ^d	DF		EM		NR/AI	
			F.Eval	Time	Rounds	Time	Rounds	Time
1	4895	3	26	0.01	24	0.05		
	9790	9	238	0.31	33	0.26		
	14685	18	583	1.77	45	1.02		
2	6192	9	699	1.27			6	0.45
	10230	12	1236	2.33			8	0.90
	14274	18	4751	11.10			18	3.33
3	5731	5	169	0.34			6	0.07
4	8765	6	927	70.60	109	1.14	7	1.86
5	5073	2	39	0.02	23	4.97	5	0.02
	10146	6	472	0.52			9	0.09
6 ^e	235796	55	37021	2083			185	40.10
7	46581	12	1435	15.2	1006	88.60	6	0.58
	55410	19	5813	30.6			6	1.00

a Updated from Hofer [17].

b References (1) Misztal [30]; (2) Meyer and Smith (1996); (3) Johnson and Thompson [18] (4) Gilmour *et al.* [9]; (5) Madsen *et al.* [24]; (6) Neumaier and Groenevald [32] (7) Jansen *et al.* (1997)

c Dimension of mixed model equations (MME)

d Number of (co)variance components

e 'DF' = quasi Newton using finite differences

'NR/AI' = quasi-Newton using computed analytic differences

For example, Foulley and Quass [7] use a model $y = X\alpha + \sigma_G Z u^* + e$ and given σ predict u with natural mixed model equations. Regression of y on σ and Zu^* (taking into account uncertainty of u) gives a natural way of updating σ (keeping σ_G^2 within the parameter space). For a balanced sire model Foulley and

Quass [7] note that the rate of convergence depends on $(n/(n + \alpha))$ with $\alpha = \frac{\sigma_e^2}{\sigma_g^2}$. For $\frac{n}{n + \alpha} = 0.2, 0.5, 0.8$ the rates of convergence for σ_G using an EM algorithm are 0.27, 0.45, and 0.31, compared with 0.03, 0.25, and 0.63 for a scheme based on updating σ_G^2 , showing the advantage of the σ_G parameterization for small values of $(n/(n + \alpha))$.

A more recent development is the suggestion of Lui *et al.* [22] who suggest a parameter extension or PX-EM algorithm. In our case it involves estimating $\sigma_G^2 = (\sigma_{G1})^2 \sigma_{G2}^2$ and σ_{G1} estimated by the linear scheme and σ_{G2}^2 by the quadratic scheme. This scheme at first sight counter-initiative in that σ_{G2}^2 is confounded, has a rate of convergence that again depending on $n/(n + \alpha)$ but is faster than the two previous schemes. With rates of convergence of 0.30, 0.60, 0.80 for $(n/(n + \alpha)) = 0.2, 0.5$ and 0.8. In one sense missingness helps to avoid redundancy and σ_{G1} parameter is perhaps analogous to parameters in conjugate gradient methods that decide on the optimal distance to travel in a specified direction.

I have found the following argument in trying to understand some of these improved EM schemes. Consider the case when we have N moment matrices M_i , ($i = 1, \dots, N$) with expectation $G + R_i$. This might arise in considering a p multivariate problem with 'equal designs' with $2p \times p$ multivariate components and we use a spectral decomposition to construct N independent sets of sums of squares and cross products. We consider the case when R_i is known and we are interested in estimating G . We let $G = SUS^1$ that allows a wide range of possible models. If $S = I$ we have the most common model. If $U = I$ and S lower triangular we have a Choleski parameterization. The matrix S could be thought representing a set of factors and if S is of size $p \times f$ we have V factors, and so the parameterization can represent a reduced rank or latent regression parameterization. As G is a symmetric matrix, there are $p \times (p + 1)/2$ parameters. Obviously care needs to be taken with S and V as these have $p \times (2p^2 + p + 1)/2$ parameters. An estimation procedure based on differentiating the likelihood can be informally thought of as thinking of M_i as $y_i y_i^1$ with $E(y_i) = u_i = sf_i$ estimation of the terms of S can be thought of as predicting f_i from y_i and regressing y_i on the prediction of f_i taking into account the uncertainty in f_i . Estimation of U follows the recipe involving the prediction of f_i and the prediction error variance of f_i . Note that formally y_i does not need forming as all required terms can be constructed from M_i . A similar algorithm can be constructed from PX-EM arguments (B.C. Cullis and A. Smith *pers. comm.*). I have found this argument useful in (a) understanding the PX-BM methodology, (b) estimation in reduced rank or latent factor models and (c) as a way of constructing hybrid iterative schemes.

We think that AI iterative schemes are all attractive in that they usually only need a small number of iterations. The two drawbacks are that they do not always improve the likelihood but this difficulty reduces as the parameters get

nearer to a minimum value of the likelihood and can lead to estimates outside the parameter space. One suggestion is motivated by Lee and Nelder [20] who base estimation of variance parameters in hierarchical models of pseudo data based on sums of squares of predicted values and their prediction error variances. This suggests that AI algorithms are having problems using (PX)-EM schemes based on constructing pseudo moment matrices and expectations from relevant prediction error matrices for difficult parameters. A maximization of this likelihood of this pseudo-data could be perhaps used. This updating should get parameters nearer the maximum computationally faster than updating all parameters after each (PX)-EM iterate.

We have concentrated on exact method of analysis because a recent excellent book (Soensen and Gianola [36]) have discussed Bayesian and Markov Chain Monte Carlo (MCMC) methods. In a sense there is a direct analogy between direct and iterative estimation in linear estimation and exact and sampling based methods in quadratic estimation. I tend to think of Gibbs sampling methods as adding noise at every step of a simplified exact analysis. For instance estimate b and add noise, estimate u and add noise, form sums of squares for u and add noise to give an estimate of σ_G^2 . One does not need to Bayesian to use MCMC methods and Guo and Thompson [13] use the above paradigm with the estimation of σ_G^2 given by an EM step. In a sense the difficulties of calculating prediction error variances is replaced by sampling them. Thompson [38] and Groeneveld and Garcia-Cortes [8] have pointed out that the sampling error can be reduced when updating σ_G^2 taking account of the variance of the noise added to u although this is simpler to do for uncorrelated effects. One can also get nearer to exact methods by using block updating but this leads to more complicated variance correction formula. It is not always clear which computational scheme, exact, Gibbs sampling or intermediate will minimize computational effort.

A recent suggestion by Clayton and Rasbash [1] for imputation can also reduce the computational effort. In our model, their idea suggests fitting two models

$$y - Z\tilde{u} = Xb + e \quad (1)$$

and

$$y - X\tilde{b} = Zu + e \quad (2)$$

In (1) we fit \hat{b} and construct \tilde{b} as \hat{b} plus noise. In (2) we adjust y for \tilde{b} , estimate σ_G^2 and σ_e^2 and fit \hat{u} and add noise to \tilde{u} . Then y is adjusted for $Z\tilde{u}$ and the procedure repeated. After burn in averages of σ_G^2 and σ_e^2 provide estimates of σ_G^2 and σ_e^2 in the spirit of Gibbs sampling but avoiding some of

the noise in \tilde{u} when σ_G^2 and σ_e^2 are estimated. Garcia Cortes, and Sorensen [8] and Harville [14] have recently discussed related ideas.

Conclusions

We have shown that the area of genetic parameter estimation has advanced tremendously over the last thirty years allowing more appropriate models to be fitted to larger data sets. There are still challenging problems to be solved that we think will build on existing knowledge.

REFERENCES

- [1] Clayton, D. and Rasbash, G. (1999). Estimation in large crossed random effects models by data augmentation. *J. Roy. Statist. Soc.*, **A162**, 425-436.
- [2] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc.*, **B39**, 1-38.
- [3] Ducrocq, V. (1990). Estimation of genetic parameters arising in nonlinear models. In: *Proc. 4th World Congress on Genetics Applied to Livestock Production*, Edinburgh, **13**, 419-448.
- [4] Ducrocq, V. (1992). Solving animal model equations through an approximate incomplete Cholesky decomposition. *Genet. Sel. Evol.*, **24**, 193-209.
- [5] Ducrocq, V. (1993). Solution of multiple trait animal models with missing data on some traits. *J. Anim. Breed. Genet.*, **110**, 81-92.
- [6] Ducrocq, V. and Chapuis, H. (1997). Generalizing the use of the canonical transformation for the solution of multivariate mixed model equations. *Genet. Sel. Evol.*, **29**, 205-224.
- [7] Foulley, J.L. and Quaas, R.L. (1995). Heterogeneous variances in Gaussian linear mixed models. *Genet. Sel. Evol.*, **27**, 211-225.
- [8] Garcia-Cortes, L.A. and Sorensen, D. (2001). Alternative implementations of Monte Carlo EM algorithms for likelihood inferences. *Genet. Sel. Evol.*, **33**, 443-452.
- [9] Gilmour, A.R., Thompson, R. and Cullis, B.R. (1995). Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, **51**, 1440-1450.
- [10] Goddard, M. (1997). Advances in dairy cattle breeding research. *Interbull Bull.*, **19**, 37-49.
- [11] Groeneveld, E. (1994b). A reparameterization to improve numerical optimization in multivariate REML (co)variance component estimation. *Genet. Sel. Evol.*, **26**, 537-545.
- [12] Groeneveld, E. and García-Cortés, A. (1998). VCE4.0, a (co)variance component package for frequentists and bayesians. In: *Proc. 6th World Congress on Genetics Applied to Livestock Production*, Armidale, **27**, 455-456.

- [13] Guo, S.W. and Thompson, E.A. (1994). Monte-Carlo estimation of mixed models for large complex pedigrees. *Biometrics*, **50**, 417-432.
- [14] Harville, D.A. (1999). Use of the Gibbs sampler to invert large, possibly sparse, positive definite matrices. *Linear Alg. Appl.*, **289**, 203-224.
- [15] Henderson, C.R. (1973). Sire evaluation and genetic trends. In: *Proc. Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush, American Society of Animal Science and American Dairy Science Association, Champaign, Illinois*, 10-41.
- [16] Hill, W.G., Visscher, P.M. and Brotherstone, S. (1997). Black and white spots in the application of genetics to dairy cattle breeding. *Interbull Bull.*, **19**, 51-58.
- [17] Hofer, A. (1998). Variance component estimation in animal breeding: A review. *J. Anim. Breed. Genet.*, **115**, 247-265.
- [18] Johnson, D.L. and Thompson, R. (1995). Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *J. Dairy Sci.*, **78**, 449-456.
- [19] Koerhuis, A.N.M. and Thompson, R. (1997). Models to estimate maternal effects for juvenile body weight in broiler chickens. *Genet. Sel. Evol.*, **29**, 225-299.
- [20] Lee, Y. and Nelder, J. A. (2001). Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika*, **88**, 987-1006.
- [21] Lindstrom, M.J. and Bates, D.M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J. Amer. Statist. Assoc.*, **83**, 1014-1022.
- [22] Lui, C., Rubin, D.B. and Wu, Y.N. (1997). Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika*, **85**, 755-770.
- [23] Garcia-Cortes, L.A. and Sorensen, D. (2001). Alternative implementations of Monte Carlo EM algorithms for likelihood inferences. *Genet. Sel. Evol.*, **33**, 443-452.
- [24] Madsen, P., Jensen, J. and Thompson, R. (1994). Estimation of (co)variance components by REML in multivariate mixed linear models using average of observed and expected information. In: *Proc. 5th World Congress on Genetics Applied to Livestock Production, Guelph*, **22**, 19-22.
- [25] Mantysaari, E. and Van Kleck, 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.
- [26] Meng, X.L. and Van Dyk, D.A. (1998). Fast EM-type implementation for mixed effects models. *J. Roy. Statist. Soc.*, **B60**, 559-578.
- [27] Meyer, K. (1983). Maximum likelihood procedures for estimating genetic parameters for later lactations of dairy cattle. *J. Dairy Sci.*, **66**, 1988-1997.

- [28] Meyer, K. (1985). Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics*, **41**, 153-165.
- [29] Meyer, K. (1989). Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genet. Sel. Evol.*, **21**, 317-340.
- [30] Misztal, I. (1994). Comparison of computing properties of derivate and derivate-free algorithms in variance component estimation by REML. *J. Anim. Breed. Genet.*, **111**, 346-355.
- [31] Misztal, I. and Perez-Enciso, M. (1993). Sparse matrix inversion for restricted maximum likelihood estimation of variance components by expectation-maximization. *J. Dairy Sci.*, **76**, 1479-1483.
- [32] Neumaier, A. and Groeneveld, E. (1998). Restricted maximum likelihood estimation of covarances in sparse linear models. *Genet. Sel. Evol.*, **30**, 3-26.
- [33] Patterson, H.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545-554.
- [34] Smith, S.P. (1995). Differentiation of the Cholesky algorithm. *J. Comp. Graph Stat.*, **4**, 134-147.
- [35] Smith, S.P. and Graser, H.U. (1986). Estimating variance components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.*, **69**, 1156-1165.
- [36] Sorensen, D. and Gianola, D. (2002). *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics*. Springer-Verlag.
- [37] Takahashi, K., Fagan, J. and Chin, M.S. (1973). Formation of a sparse bus impedance matrix and its application to short circuit study. In: *Proc. 8th Inst. PICA Conf.*, Minneapolis, 63.
- [38] Thompson, R. (1994). Integrating best linear unbiased prediction and maximum likelihood estimation. *5th World Congress on Genetics Applied to Livestock Production*, **18**, 337-340.
- [39] Thompson, R., Wray, N.R. and Crump, R.E. (1994). Calculation of prediction error variances using sparse matrix methods. *J. Anim. Breed. Genet.*, **111**, 102-109.
- [40] Thompson, R. (1997a). The estimation of heritability with unbalanced data. *Biometrics*, **33**, 485-504.
- [41] Thompson, R. (1997b). Estimation of quantitative genetic parameters. *Proc. of the International Conference on Quantitative Genetics*. (Eds. O. Kempthorne, E. Pollock and T. Bailey). Iowa State University Press, Ames, Iowa, 639-657.
- [42] Wray, N.R. (1990). Accounting for mutation effects in the additive genetic variance covariance matrix and its inverse. *Biometrics*, **46**, 197-186.