Prospects for Statistical Methods in Dairy Cattle Breeding

Robin Thompson[1] and Esa A. Mäntysaari[2]
[1]IACR-Rothamsted Experimental Station, Harpenden, Herts AL5 2JQ, ENGLAND
[2]Agricultural Research Centre, MTT, Institute of Animal Production, FIN-31600, Jokioinen, FINLAND

Abstract

Accurate prediction of breeding values is of great importance for cattle 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.

1. Introduction

A recent GIFT workshop had two papers that discussed future dairy cattle research. The two papers (Goddard, 1998 and Hill, Visscher and Brotherstone (1998)) 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.

2. Variance component estimation

We consider a linear model

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

with $\mathrm{var}(\mathbf{y}) = \mathbf{ZGZ}' + \mathbf{R}$, $\mathrm{var}(\mathbf{u}) = \mathbf{G}$, $\mathrm{var}(\mathbf{e}) = \mathbf{R}$. 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, 1971) is often the method of choice. The log-likelihood is of the form

$$L\,a\,(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) - \log|\mathbf{V}| - \log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|$$

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,** $\hat{\mathbf{b}}$ given by

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

The term in $|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|$ that is sometimes thought of as a penalty function because the fixed effects are not known. Mixed model equations (Henderson, 1973) pay an important part in the analysis process. These are of the form

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

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

$\mathbf{Cs} = \mathbf{r}$

so that

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

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

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

This lets us to rewrite the log likelihood

$$L \;\; a \;\; \mathbf{y'Py} - \log|\mathbf{V}| - \log|\mathbf{X'V}^{-1}\mathbf{X}|$$

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

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

These could be thought of as 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 (1971) is based on the expected value of the second differential (*Einf*) that has terms such as

$$E(\partial^2 L / \partial q_i^{\,2}) = -\tfrac{1}{2} tr[\mathbf{P}(\partial \mathbf{V} / \partial q_i)\mathbf{P}(\partial \mathbf{V} / \partial q_i)]$$

This is called the Expected Information. Then we can update θ using

$$\hat{q} = q + EInf^{-1}(\partial L / \partial q)$$

All the terms in the update can be found from the solution of MME and $\mathbf{C}^{-1}$. Whilst this development is very direct, later developments have tried to take account of the structure of the model to reduce the computational effort. For example eliminating $\mathbf{u}$ from the mixed model equations gives weighted least squares equations for $\hat{\mathbf{b}}$, and $\hat{\mathbf{u}}$ can be calculated as $\hat{\mathbf{u}} = (\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}})^{-1}\,\mathbf{Z'R^{-1}}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$ and similarly the required part of $\mathbf{C}^{-1}$ can be found from

$$\text{var}(\hat{u} - u) = (\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}})^{-1} + (\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}})^{-1}\mathbf{Z'R^{-1}X}(\mathbf{X'V^{-1}X})^{-1}\mathbf{X'R^{-1}Z}(\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}})^{-1}$$

where the second term is a correction for the uncertainty in $\hat{\mathbf{u}}$. The trace of this correction term contributes to the first differential. By applying the rotation rule it can be written as

$$tr\left[(\mathbf{X'V^{-1}X})^{-1}\mathbf{X'R^{-1}Z}(\mathbf{Z'R^{-1}Z} + \mathbf{G^{-1}})^{-2}\mathbf{Z'R^{-1}X}\right]$$

This shows that not all the elements of $\mathbf{C}^{-1}$ need calculation in order to form the first differential. (Thompson, 1977a).

An alternative algorithm was suggested by Dempster, Laird and Rubin (1977). This Expectation maximization (EM) algorithm is based on thinking of the random effects `missing'.

The estimation is based on

$$s\hat{s}_g^{\,2} = \mathbf{u'u} + Var(\mathbf{u} - \hat{\mathbf{u}})$$

by writing this as

$$s\hat{s}_g^{\,2} = \mathbf{y'V^{-1}}(\partial \mathbf{G} / \partial q_i)\mathbf{V^{-1}y} + s s_g^{\,2} - tr[\mathbf{V}^{-1}(\partial \mathbf{G} / \partial q_i)],$$

we see this as a manipulation of equating first differential to zero. It can be also written as

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

with *Inf* representing the information on the complete data. An advantage of this method is that s^2_g stays in the parameter space, i.e., $s_g^2 \geq 0$.

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 to attain Var($\mathbf{u} - \hat{\mathbf{u}}$) it requires the inversion of **C** in each iteration.

An important development was the introduction by Smith and Graser (1985) of an alternative form for the likelihood that naturally leads to sequential formation of the likelihood:

$$L = \log|\mathbf{R}| - \log|\mathbf{G}| - \log|\mathbf{C}| - \mathbf{y}'\mathbf{P}\mathbf{y}$$

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

$$\begin{bmatrix} \mathbf{X}_{nn} & \mathbf{x}_{nn+1} \\ \mathbf{x}_{n+1n} & x \end{bmatrix} \begin{bmatrix} \mathbf{u}_n \\ u_{n+1} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_n \\ y \end{bmatrix}$$

then the contribution from the $(n+1)^{th}$ term to $\mathbf{y}'\mathbf{P}\mathbf{y} = y^2 / x$ and to log $|\mathbf{C}|$ = log(x). Using regression coefficients $r_{n+1} = \mathbf{x}_{n+1n} / x$ a correction to $\mathbf{X}_{nn} = \mathbf{X}_{nn} - \mathbf{x}_{nn+1} * \mathbf{r}_{n+1}$, and to $\mathbf{y}_n = \mathbf{y}_n - \mathbf{y} * \mathbf{r}_{n+1}$ can be formed, and the procedure is repeated with $n = n - 1$. If $\mathbf{x}_{nn+1}$ is sparse then this can lead to a reduction in calculations, especially if it is taken to reorder equations to keep $\mathbf{X}_{nn}$ sparse.

To maximize the likelihood with one parameter Smith and Graser (1986) suggested to use a quadratic approximation. With more than one parameter the 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, 1989). Biologically more appropriate models with genetic components naturally fitted into their framework including, maternal models with both Willham and Falconer terms (Koerhuis and Thompson, 1997) and models with mutation terms (Wray, 1990).

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-Enciso, 1993) of an algorithm (Takahashi, *et al*. 1973) that allowed the calculation of the `relevant' terms in the inverse of **C** required forming the first differentials without calculating all the elements of the inverse. Meyer and Smith (1996) introduced an alternative way of calculating these first differentials by performing the `automatic' differentiation of the Cholesky decomposition of **C**. These techniques requiring twice the computational effort of forming the likelihood were derived using properties of Cholesky decompositions. An alternative derivation in terms of sequential of $\mathbf{C}^{-1}$ parallels the sequential formation of the likelihood (Thompson *et al*., 1994). If $\mathbf{X}_{nn}^{-1}$ contains the partition of $\mathbf{C}^{-1}$ for the first *n* elements, then the terms in $\mathbf{C}^{-1}$ for the first *n+1* elements are given by

$$\begin{bmatrix} \mathbf{X}_{nn}^{-1} & \mathbf{c}_{nn+1} \\ \mathbf{c}_{n+1n} & c \end{bmatrix}$$

with $\mathbf{c}_{n+1n} = \mathbf{x}_{n+1\,n}\ \mathbf{X}_{nn}^{-1}$
and $\mathbf{c} = 1/x + \mathbf{c}_{n+1n}\ \mathbf{x}_{nn+1}$
Terms in $\mathbf{c}_{n+1}$ only need 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 for more complicated models (Misztal, 1994). 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 differentation ideas of Smith (1995) but the calculation of each second differential requires the computation of the order of six likelihood calculations (Smith 1995) and this becomes more costly as the number of parameters increase. There are various suggestions on approximating the second differential. Mäntysaari and Van Vleck (1989) suggest accelerating the EM algorithm based on the observed geometric rate of convergence. Neumaier and Groeneveld (1998) 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, 1995, Gilmour *et al*., 1995, and Jensen, *et al*., 1997) is based on manipulation of the alternative information matrices.

The second differential of L with respect to θ_i and θ_j.

$$\frac{\partial^2 L}{\partial q_i \partial q_j} = \frac{1}{2} tr\left[\mathbf{P}\frac{\partial \mathbf{V}}{\partial q_i}\mathbf{P}\frac{\partial \mathbf{V}}{\partial q_j}\right] - \mathbf{y}'\mathbf{P}\frac{\partial \mathbf{V}}{\partial q_i}\mathbf{P}\frac{\partial \mathbf{V}}{\partial q_j}\mathbf{P}\mathbf{y}$$

and

$$E\left[\frac{\partial^2 L}{\partial q_i \partial q_j}\right] = -\frac{1}{2} tr\left[\mathbf{P}\frac{\partial \mathbf{V}}{\partial q_i}\mathbf{P}\frac{\partial \mathbf{V}}{\partial q_j}\right]$$

Both these terms often called observed and expected information are difficult to calculate but the average, which we can call average information (AI)

$$AI\left[\frac{\partial^2 L}{\partial q_i \partial q_j}\right] = -\frac{1}{2}\mathbf{y}'\mathbf{P}\frac{\partial \mathbf{V}}{\partial q_i}\mathbf{P}\frac{\partial \mathbf{V}}{\partial q_j}\mathbf{P}\mathbf{y}$$

can be calculated by using $\frac{\partial \mathbf{V}}{\partial q_i}\mathbf{Py}$ and $\frac{\partial \mathbf{V}}{\partial q_j}\mathbf{Py}$ 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 (1998) and is updated in Table 1. These show the expected improvement of EM methods over the 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) execution 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 canonical transformation (Thompson, 1977b, Meyer, 1985) can help in reducing one $p \times p$ estimation into p independent analyses. There are modifications using the EM algorithm that allow the same techniques to be used with missing values (Ducrocq, 1993) and with different designs with different variates. (Ducrocq and Chapuis, 1997).

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 σ instead of σ^2, or multivariate analogues such as Cholesky transformations (Lindstrom and Bates, 1988, and Groeneveld, 1994). Recent work on EM algorithms (Foulley and Quass, 1995, Meng and Van Dyk, 1998) have suggested that this Cholesky or linear parameterization has a natural interpretation and can lead to faster convergence.

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 [a].

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	70.6	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]	233796	55	37021	2083			185	40.1
7	46581	12	1435	15.2	1006	88.6	6	0.58
	55410	19	5813	30.6			6	1.00

a Updated from Hofer (1998).

b References 1 Misztal (1999); 2 Meyer and Smith (1996); 3 Johnson and Thompson (1995); 4 Gilmour *et al.* (1995); 5 Madsen *et al.* (1994); 6 Neumaier and Groenevald (1998); 7 Jensen *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 (1995) use a model $y = X\beta + \sigma_G Zu^* + e$ 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^2_G within the parameter space). For a balanced sire model Foulley and Quass (1995) note that the rate of convergence depends on $(n/(n+\alpha))$ with $\alpha = \sigma_e^2/\sigma_G^2$. For $(n/(n+\alpha)) = 0.2, 0.5, 0.8$ 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*. (1998) who suggest a parameter extension or PX-EM algorithm. In our case it involves estimating $\sigma^2_G = (\sigma_{G1})^2$ σ^2_{G2} and σ_{G1} estimated by the linear scheme and σ^2_{G2} by the quadratic scheme. At first sight this scheme seems counter-initiative in that σ^2_{G2} are confounded, but it has a rate of convergence that again depends on $n/(n+\alpha)$ and 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 the σ_{G1} parameter is perhaps analogous to parameters in conjugate gradient methods that

decide the optimal distance to travel in a specified direction.

We have concentrated on exact method of analysis because other papers (Korsgaard *et al.*, 1999; Janss and de Jong, 1999) have discussed Bayesian and Markov Chain Monte Carlo (MCMC) methods. In a sense there is a direct analogy between direct and iterative methods 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 be Bayesian to use MCMC methods and so Guo and Thompson (1994) use the above paradigm with the estimation of σ^2_G given by an EM step. In a sense the difficulties of calculating prediction error variances is replaced by sampling them, Thompson (1994) and Groeneveld and Garcia-Cortes (1998) have pointed out that the sampling error can be reduced when σ^2_G is updated 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 (1998) for data augmentation can also reduce the computational effort. In our model, their idea suggests fitting two models.

$$\mathbf{y} - \mathbf{Z}\tilde{\mathbf{u}} = \mathbf{X}\mathbf{b} + \mathbf{e} \qquad (1)$$

and

$$\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}} = \mathbf{Z}\mathbf{u} + \mathbf{e} \qquad (2)$$

In (1) we fit $\hat{\mathbf{b}}$ and construct $\tilde{\mathbf{b}}$ as $\hat{\mathbf{b}}$ plus noise. In (2) we adjust **y** for $\tilde{\mathbf{b}}$ **,** estimate σ^2_G and σ^2_e, fit $\hat{\mathbf{u}}$ and add noise to get $\tilde{\mathbf{u}}$. Then **y** is adjusted for $\mathbf{Z}\tilde{\mathbf{u}}$ and the procedure repeated. After burn in the averages $\bar{s}^2_G$ and $\bar{s}^2_e$ provide estimates of σ^2_G and σ^2_e in the spirit of Gibbs sampling but avoiding some of the noise in $\tilde{\mathbf{u}}$ when σ^2_G and σ^2_e are estimated.

3. 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

Clayton, D., and G. Rasbash. 1998. Estimation in large crossed random effects models by data augmentation. *Presented at Royal Statistical Society Meeting on Multilevel Modelling.*

Dempster, A.P., N.M. Laird and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc.*, Series B **39**: 1-38.

Ducrocq, V. 1993. Solution of multiple trait animal models with missing data on some traits. *J. Anim. Breed. Genet.*, **110**, 81-92.

Ducrocq, V. and H. Chapuis. 1997. Generalizing the use of the canonical transformation for the solution of multivariate mixed model equations. *Genet. Sel. Evol.*, **29**, 205-224.

Foulley, J.L. and R.L. Quaas. 1995. Heterogeneous variances in Gaussian linear mixed models. *Genet. Sel. Evol.*, **27**, 211-225.

Gilmour, A.R., R. Thompson and B.R. Cullis. 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, **51**, 1440-1450.

Goddard, M. 1998. Advances in dairy cattle breeding research. *Interbull Bull.*, **19**, 37-49.

Groeneveld, E. 1994. A reparameterization to improve numerical optimization in multivariate REML (co)variance component estimation. *Genet. Sel. Evol.*, **26**, 537-545.

Groeneveld, E. and A. García-Cortés. 1998. VCE4.0, a (co)variance component package for frequentists and bayesians. In: *Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale,* Vol. **27**, pp. 455-456.

Guo, S.W. and E.A. Thompson. 1994. Monte-Carlo estimation of mixed models for large complex pedigrees. *Biometrics*, **50**, 417-432.

Henderson, C.R. 1973. Sire evaluation and genetic trends. In: *Proc. Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush, Am. Soc. Animal Sci. and Am. Dairy Sci. Assoc., Champaign, Illinois*, pp. 10-41.

Hill, W.G., P.M. Visscher and S. Brotherstone. 1998. Black and white spots in the application of genetics to dairy cattle breeding. *Interbull Bull.*, **19**, 51-58.

Hofer, A. 1998. Variance component estimation in animal breeding: a review. *J. Anim. Breed. Genet.*, **115**, 247-265.

Janss, L. and G. de Jong. 1999. MCML based estimation of variance components in a very large dairy cattle data set. *Interbull Bull.*, **20**.

Jensen, J., E.A. Mäntysaari, P. Madsen and R. Thompson. 1997. Residual maximum likelihood estimation of (co)variance components in multivariate mixed linear models using averge information. *J. Indian Soc. Agr. Stat.* 49:215-236.

Johnson, D.L. and R. Thompson. 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.

Korsgaard, I.R., M.S. Lund, D. Sorensen, D. Ganda, P. Madsen and J. Jensen. 1999. Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits – using Gibbs sampling. *Interbull Bull.*, **20**.

Koerhuis, A.N.M. and R. Thompson. 1997. Models to estimate maternal effects for juvenile body weight in broiler chickens. *Genet. Sel. Evol.*, **29**, 225-299.

Lindstrom, M.J. and D.M. Bates. 1988. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J. Am. Stat. Assoc.***, 83**, 1014-1022.

Lui, C., D.B. Rubin and Y.N. Wu. 1997. Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika*, **85**, 755-770.

Madsen, P., J. Jensen and R. Thompson. 1994. Estimation of (co)variance components by REML in multivariate mixed linear models using average of observed and expected information. In: *Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph,* Vol. **22**, pp. 19-22.

Mäntysaari, E.A. and L.D. Van Vleck. 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.

Meng, X.L. and D.A. Van Dyk. 1998. Fast EM-type implementation for mixed effects models. *J.R. Stat. Soc. B.*, **60**, 559-578.

Meyer, K. 1983. Maximum likelihood procedures for estimating genetic parameters for later lactations of dairy cattle. *J. Dairy Sci.*, **66**, 1988-1997.

Meyer, K. 1985. Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics*, **41**, 153-165.

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.

Meyer, K. and S. Smith. 1996. Restricted maximum likelihood estimation for animal models using derivatives of the likelihood. *Genet. Sel. Evol.* 28:23-50.

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.

Misztal, I. and M. Perez-Enciso. 1993. Sparse matrix inversion for restricted maximum likelihood estimation of variance components by expectation-maximization. *J. Dairy Sci.*, **76**, 1479-1483.

Neumaier, A. and E. Groeneveld. 1998. Restricted maximum likelihood estimation of covariances in sparse linear models. *Genet. Sel. Evol.*, **30**, 3-26.

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

Smith, S.P. 1995. Differentiation of the Cholesky algorithm. *J. Comp. Graph Stat.*, **4**, 134-147.

Smith, S.P. and H.U. Graser, H.-U. 1986. Estimating variance components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.*, **69**, 1156-1165.

Takahashi, K., J. Fagan and M.S. Chin. 1973. Formation of a sparse bus impedance matrix and its application to short circuit study. In: *Proc. 8th Inst. PICA Conf., Minneapolis,* p. 63.

Thompson, R. 1977a. The estimation of heritability with unbalanced data. *Biometrics*, **33**, 485-504.

Thompson, R. 1977b. Estimation of quantitative genetic parameters. *Proc. International Conf. Quantitative Genet.* Eds. O. Kempthorne, E. Pollock and T. Bailey. *Iowa State University Press, Ames, Iowa*, 639-657.

Thompson, R. 1994. Integrating best linear unbiased prediction and maximum likelihood estimation. *Proc. 5th World Congr. Genet. Appl. Livest. Prod.*, **18**, 337-340.

Thompson, R., N.R. Wray and R.E. Crump. 1994. Calculation of prediction error variances using sparse matrix methods. *J. Anim. Breed. Genet.*, **111**, 102-109.

Wray, N.R. 1990. Accounting for mutation effects in the additive genetic variance covariance matrix and its inverse. *Biometrics*, **46**, 197-186.