Restricted Maximum Likelihood

Stefan McKinnon Høj-Edwards & Peter Sørensen

2022-03-28

Contents

1 Introduction

Genetic parameters (heritability and genetic correlations) can be estimated using a restricted maximum likelihood (REML) method. This method allow for estimation of genetic parameters using phenotypic information and genetic relationships for individuals in a study population. Genetic relationships are inferred from a general pedigree or from genetic marker data. REML allow for unbalanced data and account for genetic relationships within and between families. REML is based on linear mixed model methodology and use a restricted likelihood approach for estimating genetic parameters.

2 Estimation of Genetic Parameters using REML

The REML method was developed by Patterson and Thompson [1971] as an improvement of the standard Maximum Likelihood (ML). The ML method was originally proposed by Fisher [1922] but was introduced to variance components estimation by Hartley and Rao [1967]. ML assumes that fixed effects are known without error which is in most cases false and, as consequence, it produces biased estimates of variance components (usually, the residual variance is biased downward). To solve this problem, REML estimators maximize only the part of the likelihood not depending on the fixed effects, by assuming that the fixed effects have been, so to speak, fixed. This entails that when comparing multiple models by their REML likelihoods, they must contain the same fixed effects, and that REML, by itself, does not estimate the fixed effects. A benefit of this is that the number of parameters in the model is restricted[1].

REML does not produce unbiased estimates owing to the inability to return negative values of variance components of many methods to obtain REML estimators, but it is still the method of choice due to the fact that this source of bias is also present in ML estimates [Lynch and Walsh, 1998].

[1] Whether this is the origin of the name may be debatable, as some claim REML is an abbreviation for *REsidual* Maximum Likelihood [Searle et al., 1992, p. 250].

There are no simple one-step solutions for estimating the variance components based on ML and REML [Lynch and Walsh, 1998]. Instead, we infer the partial derivatives of the likelihoods with respect to the variance components. The solutions to these involve the inverse of the variance-covariance matrix, which themselves includes the variance components, so the variance components estimates are non-linear functions of the variance components. It is therefore necessary to apply iterative methods to obtain the estimates.

In order to better understand the following derivation of ML and REML, it is useful to recall that the likelihood ($L(\theta|y)$) is any function of the parameter (θ) that is proportional to $p(y|\theta)$. Maximizing $L(\theta|y)$ leads to obtaining the most likely value of θ ($\hat{\theta}$) given the data y. Usually the likelihood is expressed in terms of its logarithm ($l(\theta|\mathbf{y})$) as it makes the algebra easier.

We will start by first getting the ML estimators and, then, will move on to REML as it is a modification of the first, as explained above. For both, this involves writing the likelihood function, taking the partial derivatives with respect to the parameters and equating these to zero. Once the likelihood function is found, the remaining work is purely algebraic and we will not immerse in these details, but instead simply show the important intermediate steps.

2.1 Maximum Likelihood

The likelihood for the general G-BLUP model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zg} + \mathbf{e}$ (**??**) is simply the probability density function of a multivariate normal distribution, conditional on the known elements. The log-transformed likelihood can therefore be written as:

$$l(\boldsymbol{\beta}, \mathbf{V}|\mathbf{X}, \mathbf{y}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{V}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{1}$$

where $\mathbf{V}$ is the variance-covariance matrix, $\mathbf{G}\sigma_g^2 + \mathbf{D}\sigma_e^2$ (**??**). The first term is just a constant that does not involve any parameter estimation and therefore it is usually omitted from computation.

We then derive the first derivatives with respect to the parameters $\boldsymbol{\beta}$, σ_g^2 and σ_e^2. For the variance components, however, we will make use of a general expression using σ_i^2 and the derivatives can later be adjusted to fit the specific variance components. To help with the derivation, appendix **??** contains some useful properties for derivatives of matrices.

$$\begin{aligned}\frac{\partial l(\boldsymbol{\beta}, \mathbf{V}|\mathbf{X}, \mathbf{y})}{\partial \boldsymbol{\beta}} &= \frac{\partial[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]}{\partial \boldsymbol{\beta}} \\ &= \mathbf{X}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\end{aligned} \tag{2}$$

For the derivatives for the variance components, we introduce the short-hand notation V_i as

$$\mathbf{V}_i = \frac{\partial \mathbf{V}}{\partial \sigma_i^2} = \begin{cases}\mathbf{D} & \text{when } \sigma_i^2 = \sigma_e^2 \\ \mathbf{ZGZ}' & \text{when } \sigma_i^2 = \sigma_g^2\end{cases} \tag{3}$$

and so we can write

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{V}|\mathbf{X}, \mathbf{y})}{\partial \sigma_i^2} = -\frac{1}{2}\mathrm{Tr}(\mathbf{V}^{-1}\mathbf{V}_i) + \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{4}$$

To get ML estimates of the parameters, the differential equations in (2) and (4) are set equal to zero and we would attempt to solve for the variable. This is possible for $\boldsymbol{\beta}$, but not for σ_i^2 as both terms in RHS of (4) contains $\mathbf{V}^{-1}$, which is a function of the variance components. For $\boldsymbol{\beta}$, we have the estimate:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y} \tag{5}$$

which, incidentally, is the BLUE (Best Linear Unbiased Estimator) for $\boldsymbol{\beta}$. Notice that the $\hat{\boldsymbol{\beta}}$ estimate requires an estimate on the variance components, hence the $\hat{\mathbf{V}}$.

We now turn our attention to the ML estimators for the variance components. If given an estimate of the fixed effects $\hat{\boldsymbol{\beta}}$, we can rewrite the last term in (4), to take into account the difference between our estimate and the true fixed effects (by adding and subtracting $\mathbf{X}\hat{\boldsymbol{\beta}}$ into the parenthesis):

$$\begin{aligned}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) \\ +(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\end{aligned} \tag{6}$$

The ML assumes that the fixed effects are known 'without error', i.e. we assume we have the true fixed effects. We can then set $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ in (6), cancelling the last term, and (4) is altered to

$$\frac{\partial l(\boldsymbol{\beta}, \mathbf{V}|\mathbf{X}, \mathbf{y})}{\partial \sigma_i^2} = -\frac{1}{2}\operatorname{Tr}(\mathbf{V}^{-1}\mathbf{V}_i) + \frac{1}{2}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}(\mathbf{y}-X\hat{\boldsymbol{\beta}}) \tag{7}$$

The change can easily be overlooked, but it is the core of the bias of the ML: it ignores the deviation between our estimate of the fixed effects and the true fixed effect.

As noted above, the ML estimators of the variance components are non-linear functions of the variance components. To get an estimate of the variance components, an iterative approach is needed, but before looking into this, we will cover the restricted maximum likelihood.

2.2 Restricted Maximum Likelihood

The fallacy of ML towards the use of 'true' fixed effects has in Restricted Maximum Likelihood (REML) been countered. The trick that REML uses in order to obtain unbiased estimates of variance components is a linear transformation of the observations, $\mathbf{y}$, that removes fixed effects from the model. For this purpose, a matrix $\mathbf{K}$ is used, such that $\mathbf{KX} = 0$. $\mathbf{K}$ does not need to be computed, as it humbly leaves the equations before we reach the results.

The REML model is thus:

$$\mathbf{y}* = \mathbf{Ky} = \mathbf{K}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Zg} + \mathbf{e}) = \mathbf{KZg} + \mathbf{Ke} \tag{8}$$

By substituting the following into the likelihood in (1)

$$\mathbf{Ky} \quad \text{for} \quad \mathbf{y}; \quad \mathbf{KX} = 0 \quad \text{for} \quad \mathbf{X}; \quad \mathbf{KZ} \quad \text{for} \quad \mathbf{Z}; \quad \mathbf{KVK}' \quad \text{for} \quad \mathbf{V}$$

we can obtain restricted log-likelihood (the constant term and terms that do not include variance components are not included for brevity):

$$l(\mathbf{V}|\mathbf{y}) \propto -\frac{1}{2}\ln|\mathbf{K}'\mathbf{VK}| - \frac{1}{2}(\mathbf{Ky})'(\mathbf{K}'\mathbf{VK})^{-1}(\mathbf{Ky}) \tag{9}$$

However Searle et al. [1992] proved that:

$$\ln|\mathbf{K}'\mathbf{VK}| = \ln|\mathbf{V}| + \ln|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|$$

and

$$(\mathbf{Ky})'(\mathbf{K}'\mathbf{VK})^{-1}(\mathbf{Ky}) = \mathbf{y}'\mathbf{Py} = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$$

leading to the Restricted likelihood:

$$l(\mathbf{V}|\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}) \propto -\frac{1}{2}\ln|\mathbf{V}| - \frac{1}{2}\ln|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \tag{10}$$

Comparing this to ML,

$$l(\beta, \mathbf{V}|\mathbf{X}, \mathbf{y}) \propto -\frac{1}{2}\ln|\mathbf{V}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{1}$$

we find the only difference is that the term with second quadratic form is included in REML, because they are assumed known (without error) in ML.

Again, an iterative approach is needed to get the estimates of the variance components, and further derivation of the variance component estimators depends on the iterative approach. The approaches used here are based on finding the set of parameters that maximises the likelihood. The problem can now be treated as a general problem in the sense that we have to search the parameter space for the combination of parameters that result in the largest likelihood and/or improve the likelihood from a given set of parameters.

We mention briefly the derivative free methods, as they circumvented inverting the coefficient matrix, which is required in the derivative-based algorithms [Hofer, 1998]. The Simplex method only uses the REML likelihood, and relies on other means for guessing the updated variance components [Nelder and Mead, 1965]. Another approach by Smith and Graser [1986] and Graser et al. [1987] rewrites the restricted likelihood, so a computational demanding operation can be solved by Gaussian elimination. These are simple, but may be plagued by numerical problems, especially if (...) many parameters (are) to be estimated [Jensen et al., 1997] and can be less efficient when used with increasing number of traits [Misztal, 1994, Jensen et al., 1997].

An alternate heuristic is the Expectation-Maximization (EM) algorithm [Dempster et al., 1977], which is based on the first derivatives. The idea here is that the model comprises observed data, unobserved, latent variables, and parameters. If any two of the three are known, the third is quite tractable. In terms of G-BLUP, the genetic values g are the unobserved, latent variables, and the parameters are the variance components. If the variance components and observations (and covariance matrices) are known, genetic values can be calculated from (**??**). If it was the genetic values, instead of variance components, that was known, the latter can be deduced by reflecting on the assumptions placed on the random variables. However, we rarely know both the genetic values *and* the variance components.

The EM algorithm consists of two steps, the E step which expresses the expectation of the unobserved variables conditional on the observed data and estimates of the parameters; and the M step which maximises the parameters, based on the observed data and expectation of the unobserved variables. This is still an iterative approach, and it is started with some initial guesses for the parameters, but by alternating between the E- and M-step, the algorithm approximates the (restricted) maximum likelihood [Knight, 2008]. There are however computational issues with the EM algorithm, as it a) requires inverting the coefficient matrix (although workarounds exists), and b) may require an extensive number of iterations to converge [Hofer, 1998].

2.3 Average-Information REML

We have now described the progress of estimating variance components from Maximum Likelihood, to Restricted Likelihood, with means of using the REML estimators of derivative free and first derivatives. The time has now come to the second derivatives, starting with the Newton-Raphson approach, and the Fisher Scoring Method, which naturally leads to the Average Information REML.

The section will conclude on commenting on convergence criteria and a short discussion on the interpretation of the estimated variances.

The Average-Information REML (AI-REML) algorithm is based on the Newton-Raphson (NR) approach to approximate a function's root, the function here being the first derivative of REML. It uses the first and second derivatives of the likelihood to estimate in which direction and distance (in the parameter space) an update to the parameters that increases the likelihood might be found.

Following the NR approach, the parameters θ at the t^{th}, step can be updated by

$$\theta^{(t+1)} = \theta^t - \left(\frac{\partial^2 l(\theta^t)}{\partial\theta^t\partial\theta^t}\right)^{-1}\frac{\partial l(\theta^t)}{\partial\theta^t} \tag{11}$$

where θ is the vector of variance components, i.e. $\theta = (\sigma_e^2,\ \sigma_g^2)$, and $l(\theta^t)$ is the restricted likelihood in (10), but written with emphasis on the vector of variance components at the t^{th}, step. This is similar to Euler's method for approximating a differential equation, but instead of a fixed step size, the step size is determined by the second derivative. Note: The above is *not* the AI-REML algorithm.

The first derivative $(\partial l(\theta^t)/\partial\theta^t)$ is a r-length vector, where r is the number of variance components, including the residual, to be estimated.

The second derivative is an $r \times r$ matrix. Normally referred to as a *Hessian* matrix, in this context it is the *observed information matrix*. Skipping the algebra of derivation, it can be expressed, cf. Lynch and Walsh [1998], as

$$\frac{\partial^2 l(\theta)}{\partial\sigma_i^2\partial\sigma_j^2} = \frac{1}{2}\operatorname{Tr}(\mathbf{PV}_i\mathbf{PV}_j) - \mathbf{y}'\mathbf{PV}_i\mathbf{PV}_j\mathbf{Py} \tag{12}$$

Note however that entries relating to the residual can be reduced to simpler expressions. The NR approach is not necessarily stable, as initial guesses that lie far from the maxima will lead to large steps that step past the maxima, resulting in an oscillating iteration that only – if it does – slowly converges. Furthermore, the calculation of the observed information matrix in (12) can be computational straining.

An alternate approach is the Fisher Method of Scoring. The basic principle is the same, except instead of the Hessian matrix, the negative expected Hessian matrix is used. This is said to always be positive and stabilizes the algorithm [Lynch and Walsh, 1998]. The *expected information matrix* can be calculated as

$$-E_y\left[\frac{\partial^2 l(\theta)}{\partial\sigma_i^2\partial\sigma_j^2}\right] = -\frac{1}{2}\operatorname{Tr}(\mathbf{PV}_i\mathbf{PV}_j) \tag{13}$$

and as before, entries including the residual variance component can be simplified. D.L. Johnson and Robin Thompson, Per Madsen, Just Jensen, and Esa A. Mantysaari [Johnson and Thompson, 1995, Madsen et al., 1994, Jensen et al., 1997] showed that using the average of the observed and expected information matrix was easier to compute [Jensen et al., 1997], hence *Average-Information REML*.

It was shown that the average-information matrix, $\mathbf{I}_A$, could be calculated by

$$\mathbf{I}_A(\theta) = \mathbf{F}'\mathbf{PF} = \mathbf{F}'\mathbf{R}^{-1}\mathbf{F} - \mathbf{T}'\mathbf{W}'\mathbf{R}^{-1}\mathbf{F} \tag{14}$$

where $\mathbf{F}$ is a $n \times r$ matrix and the j^{th}, column $(\mathbf{f}_j)$ corresponds to $\frac{\partial\mathbf{V}}{\partial\theta_j}\mathbf{Py}$. $\mathbf{W}$ is the total design matrix, i.e. $\mathbf{W} = (\mathbf{X}\ \mathbf{Z})$, and $\mathbf{T}$ is a $n \times r$ matrix whose columns are the solutions to the MME using $\mathbf{f}_j$ instead of $\mathbf{y}$. Jensen et al. [1997] note that once $\mathbf{F}$ is known the average information can be computed easily by solving the MME once for each parameter in θ using efficient techniques for solving large and sparse linear systems, such that the solutions can be found without computing the full inverse of the MME coefficient matrix.

For G-BLUP, as displayed in (**??**), the two columns of $\mathbf{F}$ can be calculated as:

$$\mathbf{f}_g = \frac{\mathbf{Z}\hat{\mathbf{g}}}{\sigma_g^2} \tag{15}$$

$$\mathbf{f}_e = \frac{\partial\mathbf{V}}{\partial\sigma_e^2}(\mathbf{D}\sigma_e^2)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{g}}) = \frac{\hat{\mathbf{e}}}{\sigma_e^2} \tag{16}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ are the MME solutions.

It should therefore be seen, that the AI-REML approach is easier to compute than the respective approaches mentioned prior to this. We briefly mention that the inverse of $\mathbf{G}$ can be calculated while constructing $\mathbf{G}$, and the inverse of $\mathbf{D}\sigma_e^2$ is fool proof. The algorithm has additional advantages, as the elements of the first derivatives can be calculated at the same time as calculating the second derivatives, without processing the data again.

However, the parameter update might result in negative estimates of the variance components, and the implementation must be safeguarded against this by using a weighted average of the AI update and an EM update, that is modified to use the average information matrix. If the parameter update is still outside the parameter space, the update is re-attempted while gradually increasing the weight on the EM update.

Finally, when the algorithm has converged, the observed information matrix contains estimates of the uncertainty of the parameter estimate where the inverse of the expected information matrix contains the standard errors of the parameter estimates [Lynch and Walsh, 1998, p. 796].

2.3.1 Convergence

The question now goes towards when to stop updating the parameter estimates. We do this when we believe they have converged. One example is when the change in values of the parameter estimates are sufficient small, i.e. $\|\theta^{t+1} - \theta^t\| < \epsilon_1$, where ϵ_1 is a very small value such as 10^{-5} or 10^{-8}. However, Jensen et al. [1997] notes that this criterion might be fulfilled under the EM algorithm, before the EM algorithm has found a maximum.

Another criterion would be to see if the first derivatives are small enough, in line with the idea behind the (RE)ML approaches. Jensen et al. [1997] emphasises that parameters estimated with low accuracy should be weighted heavier, thus another criterion could be

$$\left\| \frac{diag(\mathbf{I}_A^{-1})}{\sqrt{r}} \cdot \frac{\partial l(\theta)}{\partial \theta} \right\| < \epsilon_2 \tag{17}$$

but this criterion suffers from being very large when estimates are at the boundary of the parameter space.

References

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. R. Stat. Soc. Ser. B*, 39(1):pp. 1–38, 1977. ISSN 00359246. URL http://www.jstor.org/stable/2984875.

R. A. Fisher. On the Mathematical Foundations of Theoretical Statistics. *Philos. Trans. R. Soc. London*, 222(594-604):309–368, 1922. ISSN 1364-503X. doi: 10.1098/rsta.1922.0009. URL http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.1922.0009.

H.-U. Graser, S P Smith, and B. Tier. A Derivative-Free Approach for Estimating Variance Components in Animal Models by Restricted Maximum Likelihood. *J. Anim. Sci.*, 64(5):1362–1370, 1987. URL http://www.journalofanimalscience.org/content/64/5/1362.short.

H. O. Hartley and J. N. K. Rao. Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, 54(1):93–108, 1967. URL http://biomet.oxfordjournals.org/content/54/1-2/93.short.

A. Hofer. Variance component estimation in animal breeding: a review. *J. Anim. Breed. Genet.*, 115(1-6): 247–265, January 1998. ISSN 09312668. doi: 10.1111/j.1439-0388.1998.tb00347.x. URL http://doi.wiley.com/10.1111/j.1439-0388.1998.tb00347.x.

Just Jensen, Esa A. Mantysaari, Per Madsen, and Robin Thompson. Residual Maximum likelihood Estimation of (Co) Variance Components in Multivariate Mixed Linear Models Using Average Information. *J. Indian Soc. Agr. Stat.*, 49:215–236, 1997. URL http://isas.org.in/jisas/jsp/abstract.jsp?title=Residual.

D.L. Johnson and Robin Thompson. Restricted Maximum Likelihood Estimation of Variance Components for Univariate Animal Models Using Sparse Matrix Techniques and Average Information. *J. Dairy Sci.*, 78(2):449–456, February 1995. ISSN 00220302. doi: 10.3168/jds.S0022-0302(95)76654-1. URL http://www.journalofdairyscience.org/article/S0022-0302(95)76654-1/abstract.

Emma Knight. *Improved Iterative Schemes for REML Estimation of Variance Parameters in Linear Mixed Models*. Phd thesis, School of Agriculture, Food and Wine, University of Adelaide, 2008.

Michael Lynch and Bruce Walsh. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc., Sunderland, USA, 1998. ISBN 0-87893-481-2.

Per Madsen, Just Jensen, and Robin Thompson. Estimation of (co)variance components by REML in multivariate mixed linear models using average of observed and expected information. In *5th WCGALP*, pages 455–462, Guelph, Canada, 1994.

I. Misztal. Comparison of computing properties of derivative and derivative-free algorithms in variance-component estimation by REML. *J. Anim. Breed. Genet.*, 111(1-6):346–55, January 1994. ISSN 0931-2668. doi: 10.1111/j.1439-0388.1994.tb00473.x. URL http://www.ncbi.nlm.nih.gov/pubmed/21395785.

J. A. Nelder and R. Mead. A simplex method for function minimization. *Comput. J.*, 7(4):308–313, 1965. doi: 10.1093/comjnl/7.4.308. URL http://comjnl.oxfordjournals.org/content/7/4/308.abstract.

H. D. Patterson and R. Thompson. Recovery of Inter-Block Information when Block Sizes are Unequal. *Biometrika*, 58(3):545–554, December 1971. ISSN 00063444. doi: 10.2307/2334389. URL http://www.jstor.org/stable/2334389.

Shayle R. Searle, George Casella, and Charles E. McCulloch. *Variance Components*. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, New Jersey, 1992. ISBN 978-0-470-00959-8. Reprinted in 2006.

S.P. Smith and H.-U. Graser. Estimating Variance Components in a Class of Mixed Models by Restricted Maximum Likelihood. *J. Dairy Sci.*, 69(4):1156–1165, April 1986. ISSN 00220302. doi: 10.3168/jds.S0022-0302(86)80516-1. URL http://www.journalofdairyscience.org/article/S0022-0302(86)80516-1/abstract.