Received: 17 December 2018 | Revised: 1 April 2019 | Accepted: 3 April 2019

DOI: 10.1111/jbg.12400

ORIGINAL ARTICLE

Journal of Animal Breeding and Genetics WILEY

Desert island papers—A life in variance parameter and quantitative genetic parameter estimation reviewed using 16 papers

Robin Thompson

Rothamsted Research, Harpenden, UK

Correspondence
Robin Thompson, Rothamsted Research, Harpenden, Herts, AL5 2JQ, UK.
Email: robin.thompson@rothamsted.ac.uk

Abstract

I review my scientific career in terms of eight areas and 16 papers. The first two areas are associated with childhood. The other six are associated with residual maximum likelihood (REML), canonical transformation, inbreeding in selected populations, average information residual maximum likelihood (AIREML), the computer program ASReml and sampling-based estimation.

KEYWORDS

mixed models, quantitative genetic parameter estimation, REML

1 | INTRODUCTION

When I was recently asked to review my scientific career, I thought of adapting the device used in the radio programme Desert Island Discs when "castaways" are invited to discuss their life and suggest eight record tracks they would like if stranded on a desert island. I thought of replacing eight discs by eight papers but quickly realized eight papers were not self-contained and so I consider eight areas and I arbitrarily increase the number of papers to 16 to make a more coherent story.

The first two areas are associated with my early life.

- Childhood (2)
- University education (3)

And the other six are associated with statistical and genetic topics (the numbers in brackets refer to the article sections).

- Residual maximum likelihood (REML)(4)
- Canonical transformation(5)
- Inbreeding in selected populations(6)
- Average information residual maximum likelihood (AIREML)(7)
- The computer program ASReml(8)
- Sampling-based estimation(9)

2 | CHILDHOOD (1945–1963)

I was reared on two small dairy farms (circa 12 hectares) in Lancashire, UK. This gave me a lifelong interest in agriculture, but I had no practical skills and quickly realized a farming career was not sensible. The only academic subject I was good at was mathematics, and in my early teens, I used to say I wanted to be an aeronautical engineer to stop the conversation. Then, at the age of 13 I read an article in a farming periodical, the Farmer and Stockbreeder, about a geneticist from Edinburgh Alan Robertson. The periodical explained the work he had done on developing a method of evaluating bulls called the contemporary comparison (CC). I had seen some of these values in descriptions of A.I. bulls my father was using. Alan Robertson was described as an agricultural statistician and from that day forward I wanted to be an agricultural statistician to combine my interests in agriculture and mathematics.

To acknowledge Alan Robertson's contribution to my career, my first desert island paper is Robertson and Rendel (1954) [D] (I denote desert island papers by [D]), one of a series of papers on CC, a method devised by Alan that allowed A.I. to be used to increase rates of improvement in dairy cattle. An ingenious method illustrated Alan's mathematical skill and intuition that had worldwide effect. Something that

inspired me to try do something as useful in my career. This work on CC also helped in understanding and clarifying other methods of sire evaluation (Thompson 1976b and Thompson, 1979).

3 | UNIVERSITY (1963–1967)

I studied at Newcastle University, obtaining a B.Sc. in mathematics in 1996 and a M.Sc. in statistics in 1967. Two reasons for going to Newcastle were their low requirements for English and French and that they had a very good agricultural department. I took advantage of this to do a M.Sc. thesis on variance component estimation in poultry populations. I was supervised by Maurice Bichard and he introduced me to my second desert island paper Pig Industry Development Authority (PIDA) (1966) [D], which was a mimeograph publication (DA188) on combined testing in pig populations. The scientific ideas became more accessible in the paper by Fowler, Bichard, and Pease (1976), but the idea I have used throughout my scientific life was a financial analogy that you in contrast to biological data where one has variation one should not make mistakes with financial data as income minus expenses equals what is left in the bank. I adopted that so when I was not sure about something I would do calculations two ways to identify the flaw.

4 | RESIDUAL MAXIMUM LIKELIHOOD

After completing my M.Sc., my first job was with the Agricultural Research Council Unit of Statistics (ARCUS), Edinburgh, and I was employed there from 1967 to 1983. ARCUS had the dual role of trying to meet the statistical needs of agricultural research workers in institutes and colleges in Scotland and developing statistical theory with reference to agriculture. Early in my career, I read a paper by Cunningham and Henderson (1968) on iterative estimation of variance components and noticed a flaw in the argument so, with diffidence, as Henderson's methods were ubiquitous for estimation of variance components. I constructed a correction note and sent it to my director David Finney on a Friday afternoon. It came back on Monday as a draft paper, parts in blue ink that David was certain about and parts in red ink he wanted a colleague of ours, Desmond Patterson, to check, and finally in black ink 'ROBIN CANNOT WRITE'. In my defence, I was just trying to write a correction note and did not have the vision to turn the material into a paper. David's comments had a tremendous effect on me. I realized I would need to work as a team member and co-opt good writers if I was going to get work published (approximately only 10% of my publications are singly authored and approximately half of those are conference papers). I was also impressed with his generosity in helping a younger colleague. This is something I have tried to do, and found extremely rewarding, in my career. To acknowledge David's influence, I include the resulting paper Thompson (1969) [D] in my desert island papers.

This work led to a discussion with Desmond Patterson on whether we could develop better methods. Both of us were interested in mixed models but, coming from different areas of agriculture our terminology was somewhat different. I thought of animal genetic data having contemporary groups as fixed effects, sire effects as random effects and the sire variance as a variance component to be estimated, while Desmond, with an experimental design background, thought of treatment effects as fixed effects, block effects as random effects and the block variance as a variance component to be estimated. In the animal context, the sire variance enabled the heritability to be estimated and was used to downweight the sire effects to give predictions. By contrast, in the experimental design context the treatment effects were the main emphasis. There were two possible estimates: (a) an *intrablock* estimate essentially adjusting for blocks and (b) an *interblock* estimate based on block totals. These estimates can be combined to give an efficient estimate, and this is called the recovery of *interblock* information, but (b) needs estimation of the block variance. With blocks of equal size, methods were available to estimate efficiently the block variance. However, no such method was available for unequal block sizes.

We were interested in whether we could do better than the corrected Cunningham and Henderson method. Desmond suggested that maximum likelihood (ML) was (a) difficult to calculate and (b) "too clever." I thought I could suggest feasible computational schemes for some of the terms in a ML estimation scheme based on sums of squares of terms in the mixed model equations and some were difficult and based on differentials of determinants of variance matrices or traces of terms in the mixed model equations. So, I thought why not equate sums of squares to expectation as was done in Henderson's methods? I also noted that Patterson (1964) had given asymptotic variance estimates for ML estimates. I did not understand the formulae as they were based on eigenvalues, so I decided to check numerically for the design where Desmond used his asymptotic variances with mine. They agreed to four figures. I showed my results to Desmond and he said he had used not the likelihood of all the data but that of error contrasts, that is comparisons that did not provide information on treatments. He explained this is what his previous head of department while at Rothamsted, Frank Yates, would do.

The results were published in Patterson and Thompson (1971). The rest is history. Why was it a success? I suppose the ubiquity of correlated data, the flexibility of mixed models and the reduction to analysis of variance methods in balanced cases help. The method uses the three parts of

the mixed model: fixed effects, random effects and variance components in an integrated way. Fixed effects are adjusted for random effects, random effects are adjusted for fixed effects and variance components can be thought of as equating sums of squares to their expectation.

There were alternative ways of justifying the estimators. Rao (1972) and LaMotte (1973) introduced minimum norm quadratic unbiased and minimum variance unbiased estimators, respectively, that when iterated lead to our estimators. Harville (1974) introduced a Bayesian argument for justifying the use of error contrasts. Verbyla (1990) used a conditional likelihood argument to provide the same estimators. Later Lee and Nelder (1996) suggested a penalty function argument to allow for the uncertainty in the effects. Further, Robertson (1962) suggested a weighting for random effects in single classification that essentially was equivalent to equating sums of squares of predicted values to their expectation.

The method had several important champions. There was an influential review by Harville (1977). Dorothy Robinson wrote a computer program for British variety trials. Sue Welham extended this program into Genstat. Karin Meyer wrote several programs (Meyer, 1988) [D], suitable for animal breeding problems. Terry Speed suggested the name residual maximum likelihood (REML) as residuals were used in the analysis (a name much more appropriate than the limited and little understood description recovery of *interblock* information !!).

One useful development in Patterson and Thompson's (1971) paper to justify using the error contrast analysis was to transform the data into independent linear combinations with zero expectation.

If we have a mixed model

$$\mathbf{y}=\mathbf{Xa}+\mathbf{Zu}+\mathbf{e} \tag{1}$$

with **y,** a vector of length n, representing data**, X** representing a full rank fixed effect design matrix of size (n ×t) **Z** a random effect design matrix of size (n×q), respectively, and with **a** and **u** representing fixed and random effects.

$$\mathbf{u}\sim\mathcal{N}(0,\mathbf{G}) \tag{2}$$

$$\mathbf{e}\sim\mathcal{N}(0,\mathbf{R}) \tag{3}$$

so var $(\mathbf{y})=\mathbf{V}=\mathbf{R}+\mathbf{ZGZ}^{T}$

Patterson and Thompson (1971) considered the case with uncorrelated random and residual effects with $\mathbf{G}=\mathbf{I}\sigma^2_{\mathbf{u}}$,$\mathbf{R}=\mathbf{I}\sigma^2_{\mathbf{e}}$ and $\gamma=\sigma^2_{\mathbf{u}}/\sigma^2_{\mathbf{e}}$.

Using $\mathbf{S}=\mathbf{I}\text{-}\mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}$, $\mathbf{QQ}^{\mathbf{T}}=\mathbf{I}$ and $\mathbf{Z}^{\mathbf{T}}\mathbf{SZ}=\mathbf{Q}\Lambda\mathbf{Q}^{\mathbf{T}}$ we can form $\boldsymbol{y}_s$ a scaled version of $\mathbf{Z}^{\mathbf{T}}\mathbf{Sy}$, with var $(\boldsymbol{y}_s)=\mathbf{I}\sigma^2_{\mathbf{e}}+\Lambda\sigma^2_{\mathbf{u}}$ and sum of squares $\mathrm{R}=\mathbf{y}^{\mathbf{T}}\mathbf{Sy}\text{-}\mathbf{y}^{\mathbf{T}}\mathbf{SZ}(\mathbf{Z}^{\mathbf{T}}\mathbf{SZ})^{-}\mathbf{Z}^{\mathbf{T}}\mathbf{Sy}=\mathbf{y}^{\mathbf{T}}\mathbf{Py}$ with expectation $(\mathrm{n}-\mathrm{t}-\mathrm{q}+1)\,\sigma^2_e$.

Soon afterwards, Nelder and Wedderburn (1972) [D] published a paper that I would also like on my desert island. They introduced generalized linear models allowing variables distributed in the exponential family (normal, gamma, Poisson, binomial) and allow a non-linear link between the observation and the linear predictor. All fit into a weighted least squares algorithm, very useful generalization in many areas. The REML likelihood, based on $\boldsymbol{y}_s$ and R, is a special case of a generalized linear model with gamma-distributed "data" $(y_{si})^2$ with expectation $\sigma^2_e+\lambda_i\sigma^2_u$ and mean square R/(n -t-q + 1) with expectation $\sigma^2_e+\lambda_0\sigma^2_u$ with $\lambda_0=0$. This suggests an obvious diagnostic plot for the goodness of fit of the model by plotting the "data" against $\lambda_i(i=0,1,\ldots\ \mathrm{q}\text{-}1)$ with a slope σ^2_u and intercept σ^2_e. Thompson (1976a,1977b) discussed extension to the multivariate case of sums of squares and cross-products distributed as Wishart distributions used to model data on parents and offspring and estimate maternal genetic variances in *Tribolium*. In practice, this was not very useful apart from discussion of iterative schemes and designs, because data usually needed correction in a non-orthogonal way for fixed effects. Cullis, Smith, and Thompson (2004) [D] extended this idea to make more explicit the relationship between REML and analysis of variance, especially for generally balanced designs, and discussed a Cholesky transformation to give a simple quantitative interpretation of the information on the variance parameters.

5 | CANONICAL VARIATES

Thompson (1977a) gave a brief introduction to the idea of using canonical transformation to simplify BLUP calculations. The data were assumed to represent t variates and similar models held for all t variates, the so-called equal design matrices model with Equations (1), (2) and (3) replaced by

$$\mathbf{y}=\left(\mathbf{I}_{\mathbf{t}}\otimes\mathbf{X}\right)\mathbf{a}+\left(\mathbf{I}_{\mathbf{t}}\otimes\mathbf{Z}_{\mathbf{u}}\right)\mathbf{u}+\mathbf{e}$$

$$\mathbf{u}\sim\mathcal{N}\left(0,\mathbf{G}_{\mathbf{o}}\otimes\mathbf{A}_{\mathbf{q}}\right)\ \ \mathbf{e}\sim\mathcal{N}(0,\mathbf{R}_{\mathbf{o}}\otimes\mathbf{I}_{\mathbf{n}})$$

The random effect vector, **u**, is of size t × q representing effects on t variates on q animals with relationship matrix $\mathbf{A}_{\mathbf{q}}$. The symmetric variance matrices $\mathbf{R}_{\mathbf{o}}$ and $\mathbf{G}_{\mathbf{o}}$ can be factorized using a canonical transformation as $\mathbf{R}_{\mathbf{o}}=\mathbf{PP}^{\mathbf{T}}$ and $\mathbf{G}_{\mathbf{o}}=\mathbf{P}\Lambda\mathbf{P}^{\mathbf{T}}$. This allows the data to be transformed into *t* independent parts using $\mathbf{y}^{*}=(\mathbf{P}\otimes\mathbf{I}_{\mathbf{n}})\mathbf{y}$.

The motivation was to simplify BLUP calculations, but the idea was used by Hill and Thompson (1978) to consider the probability of non-positive definite estimates of **G**. Canonical variates were also used by Hayes and Hill (1980, 1981) to simplify calculation of sampling properties of selection indices and suggest improved selection indices. The transformation was also used by Meyer (1985) and Furlotte and Eskiny (2015) to suggest simplified estimation of **R** and **G**. More recently, De Faveri, Verbyla, Cullis, Pitchford, and Thompson (2017) [D] considered a two-dimensional invariant multivariate autoregressive model sometimes used in plant breeding trials when there is need to take account of spatial variation in two dimensions. The residuals can be defined recursively

$$\boldsymbol{e}_{11} = \boldsymbol{\varepsilon}_{11}$$

$\boldsymbol{e}_{i1} = \boldsymbol{\Omega}_{\mathbf{r}}\boldsymbol{e}_{(i-1)1} + \boldsymbol{\varepsilon}_{i1}$ for i = 2, … , r
$\boldsymbol{e}_{1c} = \boldsymbol{\Omega}_{\boldsymbol{c}}\boldsymbol{e}_{1(j-1)} + \boldsymbol{\varepsilon}_{1j}$ for j = 2,… , c
$\boldsymbol{e}_{ij} = \boldsymbol{\Omega}_{\mathbf{r}}\boldsymbol{e}_{(i-1)j} + \boldsymbol{\Omega}_{\mathbf{c}}\boldsymbol{e}_{i(j-1)} - \boldsymbol{\Omega}_{\mathbf{r}}\boldsymbol{\Omega}_{\mathbf{c}}\boldsymbol{e}_{(i-1)(j-1)} + \boldsymbol{\varepsilon}_{ij}$ for i = 2,..., r and j = 2,..., c

$$\mathbf{var}\left(\boldsymbol{\varepsilon}_{ij}\right) = \boldsymbol{\Sigma}$$

In one-dimensional cases, we would sometimes assume invariance in the model with prediction errors not depending on which direction the model is fitted. This imposes constraints on the parameters. For row invariance, $\boldsymbol{\Omega}_{\mathbf{r}}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{\Omega}_{\mathbf{r}}^{\mathbf{T}}$, and similarly, for column invariance $\boldsymbol{\Omega}_{\mathbf{c}}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{\Omega}_{\mathbf{c}}^{\mathrm{T}}$. De Faveri (2013) showed that for row–column invariance, $\boldsymbol{\Omega}_{\mathbf{r}}\boldsymbol{\Omega}_{\mathbf{c}} = \boldsymbol{\Omega}_{\mathbf{c}}\boldsymbol{\Omega}_{\mathbf{r}}$. Because these three matrices $\boldsymbol{\Sigma}$, $\boldsymbol{\Omega}_{\mathbf{r}}$ and $\boldsymbol{\Omega}_{\mathbf{c}}$ commute, there is a transformation so that

$$\boldsymbol{\Sigma} = \mathbf{P}\mathbf{P}^{\mathbf{T}}\,\boldsymbol{\Omega}_{\mathbf{r}} = \mathbf{P}\mathbf{D}_{\mathbf{r}}\mathbf{P}^{\mathbf{T}}\,\boldsymbol{\Omega}_{\mathbf{c}} = \mathbf{P}\mathbf{D}_{\mathbf{c}}\mathbf{P}^{\mathbf{T}}$$

Thus, there are underlying *t* independent two-dimensional separable autoregressive processes, (AR1 ⊗ AR1), with autoregressive parameters $\mathbf{D}_{\mathbf{ri}}$ and $\mathbf{D}_{\mathbf{ci}}$(i = 1, t). This model is relatively easy to fit in well-constructed mixed model software.

6 | INBREEDING IN SELECTED POPULATIONS

I was transferred to the Animal Breeding Research Organization in Edinburgh in 1983. One of my initial interests there was the analysis of selection experiments. An almost classic example was an 18-year experiment on divergent selection for an index based on cannon bone length and body weight in Scottish Blackface sheep that had a linear response to index selection in two divergent lines (Atkins & Thompson, 1986a, 1986b). Selection experiments have a sophisticated autoregressive error structure comprehensively developed by Bill Hill in a series of papers, and I would like a sample paper of these, Hill (1972) [D], on my desert island. I had great difficulty in understanding these papers, so in frustration and tongue in cheek I used to ask why there were so many effective numbers—so many they all could not be effective. The sheep selection experiment gave an opportunity to investigate what was contributing to the heritability estimate from different parts of data, for example within lines and between lines, and Thompson and Atkins (1993) [D] decomposed the likelihood to show these contributions. As a by-product, we calculated a variance–covariance matrix of year estimates. This showed an autoregressive structure agreeing well with the Hill predictions.

This work eventually led to work with Naomi Wray on prediction of inbreeding in selected populations, for example by Wray and Thompson (1989) [D]. Only after discussions with Bill during the start of Naomi's thesis did I begin to understand (belatedly after 15 years) the underlying genetic arguments that are crucial to this work. Woolliams and Thompson (1994) have recently given a review of my contribution to this area. A related topic that intrigues me, but I do not have space for in this paper, is the use of relationships in statistical models to share information between relatives (Thompson, 1977a, 1977b and Thompson, 1979).

7 | AVERAGE INFORMATION RESIDUAL MAXIMUM LIKELIHOOD (AIREML)

This work started in 1992 with discussion with Arthur Gilmour and Brian Cullis on the analysis of plant breeding trials and with David Johnson on the analysis of animal breeding data. At that time, derivative-free methods had become popular because of their computational feasibility because they only required the computation of a determinant each iteration and used numerical search techniques to find a maximum. Methods that use first and second differentials were thought to have better numerical properties especially as the number of parameters increases. However, the first and second differentials of the likelihood have terms that depend on the inverse of the coefficient matrix and the inversion is computationally expensive especially for large matrices. Two ideas, firstly only calculating terms in the inverse needed for the first differentials and secondly using a simple approximate second differential or information matrix that avoids calculating the whole inverse will be briefly discussed.

7.1 | First differentials

Using model Equations (1)–(3), if the random and residual matrices $\mathbf{R}$ and $\mathbf{G}$ are functions of v_g and v_r, with variance parameters $\boldsymbol{\theta}_g$ and $\boldsymbol{\theta}_r$, and the variance matrix $\mathbf{V}$ of $\mathbf{y}$ is a function of the combined set of $v = v_g + v_r$ variance parameters, $\boldsymbol{\theta}$, then the residual log likelihood function, Log $(\boldsymbol{\theta})$, has a gradient vector $\mathbf{J}(\boldsymbol{\theta}) = -[\delta \text{Log}(\boldsymbol{\theta}) / \delta\theta_i]_{i=1,\ldots,v}$. Using the matrix $\mathbf{P}=\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^{T}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$, the component $[\mathbf{J}(\boldsymbol{\theta})]_i$ of $\mathbf{J}(\boldsymbol{\theta})$ is given by

$$2\mathbf{J}[(\boldsymbol{\theta})]_i = \text{tr}[(\delta\mathbf{V}/\delta\theta_i)\mathbf{P}] - \mathbf{y}^{T}\mathbf{P}(\delta\mathbf{V}/\delta\theta_i)\mathbf{P}\mathbf{y} \quad (5)$$

One natural interpretation of (5) is that maximum-likelihood estimates that make the gradients zero equate a set of quadratic forms to their expectation since

$$E(\mathbf{y}^{T}\mathbf{P}(\delta\mathbf{V}/\delta\theta_i)\mathbf{P}\mathbf{y}) = \text{tr}\left[(\delta\mathbf{V}/\delta\theta_i)\mathbf{P}\right] \quad (6)$$

Johnson and Thompson (1995) note that

$$\mathbf{P} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{W}\mathbf{C}^{-1}\mathbf{W}^{T}\mathbf{R}^{-1} \quad (7)$$

with $\mathbf{W} = (\mathbf{X}\mathbf{Z})$ and $\mathbf{C}$ the coefficient matrix on the left-hand side of the mixed model equations

$$\begin{bmatrix} \mathbf{X}^{T}\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^{T}\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^{T}\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^{T}\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} a \\ u \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{T}\mathbf{R}^{-1}y \\ \mathbf{Z}^{T}\mathbf{R}^{-1}y \end{bmatrix} \quad (8)$$

Alternatively, with $\mathbf{b}^{T} = [\mathbf{a}^{T}\mathbf{u}^{T}]^{T}$ and $\mathbf{c} = \mathbf{W}^{T}\mathbf{R}^{-1}\mathbf{y}$ (8) can be written concisely as

$$\mathbf{C}\mathbf{b} = \mathbf{c} \quad (9)$$

And the matrix transform $\mathbf{P}\mathbf{y}$ can be written as $\mathbf{R}^{-1}\tilde{\mathbf{e}} = \left(\mathbf{y} - \mathbf{X}\hat{\mathbf{a}} - \mathbf{Z}\tilde{\mathbf{u}}\right)$ where â and ũ are solutions to the mixed model Equation (8). Using (7) and (8) and letting $\mathbf{C}_{uu} = \mathbf{Z}^{T}\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}$ and $\mathbf{C}^{uu}$ the submatrix of $\mathbf{C}^{-1}$ pertaining to the random effects giving the prediction error variance matrix, then the gradient vectors for the $\boldsymbol{\theta}_r$ and $\boldsymbol{\theta}_g$ parameters can be written as

$$2\mathbf{J}[(\boldsymbol{\theta}_r)]_i = \text{tr}[\mathbf{R}^{-1}(\delta\mathbf{R}/\delta\theta_{ri})\mathbf{R}^{-1}(\mathbf{R} - \mathbf{W}\mathbf{C}^{-1}\mathbf{W}^{T})] - \tilde{\mathbf{e}}^{T}\mathbf{R}^{-1}(\delta\mathbf{R}/\delta\theta_{ri})\mathbf{R}^{-1}\tilde{\mathbf{e}} \quad (10)$$

$$2\mathbf{J}[(\boldsymbol{\theta}_g)]_i = \text{tr}[\mathbf{G}^{-1}(\delta\mathbf{G}/\delta\theta_{gi})\mathbf{G}^{-1}(\mathbf{G} - \mathbf{C}^{uu})] - \tilde{\mathbf{u}}^{T}\mathbf{G}^{-1}(\delta\mathbf{G}/\delta\theta_{gi})\mathbf{G}^{-1}\tilde{\mathbf{u}} \quad (11)$$

or alternatively as

$$2\mathbf{J}[(\boldsymbol{\theta}_r)]_i = \text{tr}[(\delta\mathbf{R}/\delta\theta_{ri})\mathbf{R}^{-1}] - \text{tr}[(\delta\mathbf{C}/\delta\theta_{ri})\mathbf{C}^{-1}] - \tilde{\mathbf{e}}^{T}\mathbf{R}^{-1}(\delta\mathbf{R}/\delta\theta_{ri})\mathbf{R}^{-1}\tilde{\mathbf{e}} \quad (12)$$

$$2\mathbf{J}[(\boldsymbol{\theta}_g)]_i = \text{tr}[(\delta\mathbf{G}/\delta\theta_{gi})\mathbf{G}^{-1}] - \text{tr}[(\delta\mathbf{C}_{uu}/\delta\theta_{gi})\mathbf{C}^{uu}] - \tilde{\mathbf{u}}^{T}\mathbf{G}^{-1}(\delta\mathbf{G}/\delta\theta_{gi})\mathbf{G}^{-1}\tilde{\mathbf{u}} \quad (13)$$

showing that to calculate the gradients one only needs to calculate the elements in the inverse $\mathbf{C}^{-1}$ corresponding to non-zero elements in $\mathbf{C}$, often called the sparsity pattern of $\mathbf{C}$.

A result of Takahashi, Fagan, and Chin (1973) showed that calculation of the elements of the sparse inverse subset, that is elements that correspond to elements in the sparsity pattern of the original matrix, only needs twice as many numerical operations as forming the determinant. Takahashi et al. (1973) used results on triangular matrices to show the result, but I prefer an alternative formulation used by Thompson, Wray, and Crump (1994) [D] when considering the calculation of prediction error variances for beef cattle populations. This is based on using partitioned matrices to absorb variables sequentially in a set of equations and then use a similar technique to calculate the sparse inverse.

The system of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$, with $\mathbf{A}$ a symmetric matrix, can be transformed to $\mathbf{L}\mathbf{x} = \mathbf{k}$ with $\mathbf{L}$ a lower triangular matrix by absorbing each variable in $\mathbf{x}$ in turn. Subscripts are used to indicate a matrix constructed from a subset of row and columns of a matrix, and superscripts are used to indicate the stage of the recursive procedure. So, if $\mathbf{A}$ and $\mathbf{b}$ are, respectively, an $(n \times n)$ matrix and a vector of length n, then starting at stage n $\mathbf{A}^{(n)}_{1:n,1:n}$ and $\mathbf{b}^{(n)}_{1:n}$ are equivalent to $\mathbf{A}$ and $\mathbf{b}$, respectively, and for the stages i =(n-1), …,1

$$\mathbf{A}^{(i)}_{1:i,1:i} = \mathbf{A}^{(i+1)}_{1:i,1:i} - \mathbf{A}^{(i+1)}_{1:i,i+1}[\mathbf{A}^{(i+1)}_{1+i,i+1}]^{-1}\mathbf{A}^{(i+1)}_{i+1,1:i} \text{ and } \mathbf{b}^{(i)}_{1:i} = \mathbf{b}^{(i+1)}_{1:i} - \mathbf{b}^{(i+1)}_{i+1}[\mathbf{A}^{(i+1)}_{1+i,i+1}]^{-1}\mathbf{b}^{(i+1)}_{1:i}$$

can be constructed, and then, $\mathbf{L}$ and $\mathbf{k}$ can be found using $\mathbf{L}_{i,1:i} = \mathbf{A}^{(i)}_{i,1:i}$ $\mathbf{k}_i = \mathbf{b}^{(i)}_i$

The sparse inverse subset can be found in a similar recursive way: the first diagonal element can be found using $\mathbf{A}^{-1}_{1,1} = [\mathbf{L}_{1,1}]^{-1}$. Then for i-1, …, (n-1), a scaled version of $\mathbf{L}_{i+1,1:i}$ can be constructed using

$$\mathbf{A}^{-1}_{i+1,i+1} = [\mathbf{L}_{i+1,i+1}]^{-1}\mathbf{L}_{i+1,1:i} = \mathbf{A}^{-1}_{i+1,i+1}\mathbf{L}_{i+1,1:i+1}$$

and then terms in the sparse inverse subset can be calculated using

$$\mathbf{A}^{-1}_{i+1,1:i} = \mathbf{L}_{i+1,1:i}\mathbf{A}^{-1}_{1:i,1,i} \quad \text{and} \quad \mathbf{A}^{-1}_{i+1,i+1} = \mathbf{A}^{-1}_{i+1,i+1} + \mathbf{A}^{-1}_{i+1,1:i}\mathbf{L}^{T}_{i+1,1:i}$$

We note that we only use non-zero elements of $\mathbf{L}_{i+1,1:i}$, $\mathbf{A}^{(i+1)}_{i+1.1:i}$ and $\mathbf{A}^{-1}_{i+1,1:i}$. The computational effort depends on the order of fitting. If we fit a model with mean and q groups, absorbing groups first, the numerical operations are proportional to q. However, if we fit q groups and mean, absorbing the mean first, the matrices become denser and operations are proportional to q^3.

7.2 | Information matrices

A paper that leads to my interest in this area was by Henderson, Kempthorne, Searle, and Von Krosigk (1959) [D] on the estimation of genetic and environmental trends from records subject to culling. They were interested in the situation of dairy cattle records on cows in a closed herd with culling on milk yield with fixed effects for year. An analysis with fixed effects for cows was known to lead to biased environmental trend. This paper integrated distinct contributions from three eminent statisticians. Henderson suggested using mixed models including random effects for cows, Kempthorne suggested a sequential approach: likelihood of first lactation record, likelihood of second lactation record given first lactation record, etc., while Searle showed that these methods were equivalent. Thompson (1973) considered the estimation of multivariate parent–offspring regression using Kempthorne's suggestion of considering parent records and offspring records given parents. In a p variate case, there were two symmetric p × p matrices **G** and **R** with parameters θ_g and θ_r to be estimated and two alternative information matrices were available. One (observed information, **OInf**) based on second differentials of the likelihood and another (expected information, **EInf**)) based on the expected value of the second differentials. The information matrices could be partitioned as

$$\mathbf{Inf} = \begin{bmatrix} \mathbf{Inf_{gg}} & \mathbf{Inf_{gr}} \\ \mathbf{Inf_{rg}} & \mathbf{Inf_{rr}} \end{bmatrix}$$

and the partitions $\mathbf{Inf_{gg}}$, $\mathbf{Inf_{rg}}$, $\mathbf{Inf_{gr}}$ and $\mathbf{Inf_{rr}}$ were all symmetric, but in the inverse matrix, the asymptotic variance inverse, **Var** , the off-diagonal $\mathbf{Var_{rg}}$ and $\mathbf{Var_{gr}}$ term were not. This worried me in case I had done the calculations incorrectly. Some more investigation showed that there was another information matrix, an average information matrix (**AInf**), based on the average of the other information matrices so **AInf**=(**OInf** + **EInf**)/2 that had the advantage of being based purely on data terms and in this example so much easier to calculate than **OInf** and **EInf**. I did nothing with this result for 20 years!!

When Arthur Gilmour, Brian Cullis, David Johnson and I discussed estimation in mixed models, I suggested that using sparse matrix methods to calculate the first differentials and that an approximate average information might be more feasible and efficient than the derivative-free methods.

If **V** is linear in the variance parameters, $\boldsymbol{\theta}$, then the component $[\mathbf{OInf}(\boldsymbol{\theta})]_{i,j}$ of the observed information matrix $[\mathbf{OInf}(\boldsymbol{\theta})]$ is given by

$$2[\mathbf{OInf}(\boldsymbol{\theta})]_{i,j} = -\mathrm{tr}[(\delta\mathbf{V}/\delta\theta_i)\mathbf{P}(\delta\mathbf{V}/\delta\theta_j)\mathbf{P}] + 2\mathbf{y}^{\mathbf{T}}\mathbf{P}(\delta\mathbf{V}/\delta\theta_i)\mathbf{P}(\delta\mathbf{V}/\delta\theta_j)\mathbf{Py}$$

and since

$$\mathbf{E}(\mathbf{y}^{\mathbf{T}}\mathbf{P}(\delta\mathbf{V}/\delta\theta_i)\mathbf{P}(\delta\mathbf{V}/\delta\theta_j)\mathbf{Py}) = \mathrm{tr}[(\delta\mathbf{V}/\delta\theta_i)\mathbf{P}(\delta\mathbf{V}/\delta\theta_j)\mathbf{P}]$$

the component $[\mathbf{EInf}(\boldsymbol{\theta})]_{i,j}$ of the expected information matrix $[\mathbf{EInf}(\boldsymbol{\theta})]$ is given by

$$2[\mathbf{EInf}(\boldsymbol{\theta})]_{i,j} = \mathrm{tr}[(\delta\mathbf{V}/\delta\theta_i)\mathbf{P}(\delta\mathbf{V}/\delta\theta_j)\mathbf{P}]$$

Hence, component $[\mathbf{AInf}(\boldsymbol{\theta})]_{i,j}$ of the average information matrix $[\mathbf{AInf}(\boldsymbol{\theta})]$ is given by

$$2[\mathbf{AInf}(\boldsymbol{\theta})]_{i,j} = \mathbf{y}^{\mathbf{T}}\mathbf{P}(\delta\mathbf{V}/\delta\theta_i)\mathbf{P}(\delta\mathbf{V}/\delta\theta_i)\mathbf{Py}$$

If $\mathbf{f}_i = (\delta\mathbf{V}/\delta\theta_i)\mathbf{Py}$ is the component $[\mathbf{F}]_i$ of **F** then $2[\mathbf{AInf}(\boldsymbol{\theta})]_{i,j} = \mathbf{f}_i^{\mathrm{T}}\mathbf{P}\mathbf{f}_j$ and

$$2\mathbf{AInf}(\boldsymbol{\theta}) = \mathbf{F}^{\mathbf{T}}\mathbf{PF} \tag{14}$$

Just as the term **Py** can be calculated from fitting a linear mixed model to **y** from fitting a mixed model to **y** so can $\mathbf{Pf}_i$ can be calculated from fitting a mixed model to $\mathbf{f}_i$. By analogy with the working variate in generalized linear models, we call $\mathbf{f}_i$ a working covariate. The term $[\mathbf{AInf}(\boldsymbol{\theta})]_{i,j}$ in the average information matrix can be calculated from the product of $\mathbf{Pf}_i$ with $\mathbf{f}_i^{\mathrm{T}}$.

Alternatively, if we form a matrix with the sum of squares $\mathbf{y}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{y}$, the right-hand side of the mixed model equations $\mathbf{W}^{\mathbf{T}}\mathbf{R}^{-1}\boldsymbol{y}$ and **C** of the form $\mathbf{A} = \begin{bmatrix} \mathbf{y}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{y} & \mathbf{y}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{y} & \mathbf{C} \end{bmatrix}$ then using the absorption scheme of section 5 leads to a triangular matrix with $\mathbf{L}_{1,1} = \mathbf{y}^{\mathbf{T}}\mathbf{Py}$. Gilmour, Thompson, and Cullis (1995) noted that if we replace **y** by **F**, a matrix with v columns, and perform a partial absorption terminating at stage v, then

$$\mathbf{A}^{(\mathrm{v})}_{1:\mathrm{v},1:\mathrm{v}} = \mathbf{F}^{\mathbf{T}}\mathbf{PF}.$$

Hence, using a quasi-Newton scheme an increment $\boldsymbol{\Delta}$ to the variance parameters can be found from

$$\mathbf{AInf}(\boldsymbol{\theta})\boldsymbol{\Delta} = -\mathbf{J}(\boldsymbol{\theta})$$

This iterative scheme can be put in a quasi-linear form by letting $\mathbf{J}_1(\boldsymbol{\theta})$ be a scaled component of $\mathbf{J}(\boldsymbol{\theta})$ with

$$[\mathbf{J}_1(\boldsymbol{\theta})]_i = -\mathrm{tr}[(\delta\mathbf{V}/\delta\theta_i)\mathbf{P}]$$

we find

$$\mathbf{F}^{\mathbf{T}}\mathbf{P}\mathbf{F}\boldsymbol{\Delta}=\mathbf{F}^{\mathbf{T}}\mathbf{P}\mathbf{y}+\mathbf{J}_1(\theta)$$

This suggests $\boldsymbol{\Delta}$ satisfies a mixed model equation of the form

$$\begin{bmatrix}\mathbf{F}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{F}+\mathbf{C}_{\boldsymbol{\Delta}} & \mathbf{F}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{W}\\ \mathbf{W}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{F} & \mathbf{C}\end{bmatrix}\begin{bmatrix}\Delta\\ b\end{bmatrix}=\begin{bmatrix}\mathbf{F}^{\mathbf{T}}\mathbf{R}^{-1}y+\mathbf{I}^{\mathbf{T}}\mathbf{R}_{\boldsymbol{\Delta}}^{-1}\mathbf{y}_{\boldsymbol{\Delta}}\\ \mathbf{W}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{y}\end{bmatrix} \quad (15)$$

with $\mathbf{y}_{\boldsymbol{\Delta}}=\mathbf{J}_1(\theta),\mathbf{b}^{\mathbf{T}}=[\mathbf{a}^{\mathbf{T}}\mathbf{u}^{\mathbf{T}}]^{\mathbf{T}},\mathbf{C}_{\boldsymbol{\Delta}}=\mathbf{R}_{\boldsymbol{\Delta}}^{-1}+\mathbf{G}_{\boldsymbol{\Delta}}^{-1}$ with $\mathbf{R}_{\boldsymbol{\Delta}}^{-1}=\mathbf{I}=-\mathbf{G}_{\boldsymbol{\Delta}}^{-1}$ corresponding an incremental model with random working covariate effects

$$\mathbf{y}=\mathbf{F}\boldsymbol{\Delta}+\mathbf{X}\mathbf{a}+\mathbf{Z}\mathbf{u}+\mathbf{e} \quad (16)$$

$$\mathbf{y}_{\boldsymbol{\Delta}}=\mathbf{I}\boldsymbol{\Delta}+\mathbf{e}_{\boldsymbol{\Delta}} \quad (17)$$

and

$$\boldsymbol{\Delta}\sim\mathcal{N}(0,\mathbf{G}_{\boldsymbol{\Delta}})\text{ and }\mathbf{e}_{\boldsymbol{\Delta}}\sim\mathcal{N}(0,\mathbf{R}_{\boldsymbol{\Delta}}) \quad (18)$$

Our development was motivated by models that generated sparse mixed model equations and taking advantage of this sparsity. In other cases other computing strategies might be appropriate. For example, for data that has dense genomic matrices, Lee and Van der Werf (2006) and Yang, Lee, Goddard, & Visscher (2011) [D] discuss methods and software that use weighted least squares equations rather than the mixed model equations in conjunction with the use of average information and has computational advantages in their situations.

8 | THE COMPUTER PROGRAM ASREML

The success of our early experience with the average information algorithm led to an agreement between Arthur Gilmour, Brian Cullis, Sue Welham and me to agree to pool our recourses to produce software based on the average information algorithm. We brought different expertise to the project: Arthur Gilmour had written programs for regression analysis, breeding value prediction and spatial analysis; Brian Cullis had experience of spatial analysis of field experiments which utilized the concept of separability which we found useful more generally, for example, for specifying multivariate and spatial model; and Sue Welham had introduced a REML program into GENSTAT and had experience of repeated measure and spline analysis. Later, we co-opted Beverley Gogel and Dave Butler to the team to improve the draft documentation and provide computing skills, respectively.

I had experience of fitting a variety of models in Karin Meyer's programs (DFREML and REMLPK) with a succession of Ph.D. students. Mrode, Smith, and Thompson (1990) had fitted, for the first time, an animal model in a selection experiment and we tested Karin's program to its limits, for example Crump, Thompson, Haley, and Mercer (1997) fitted genetic covariances between traits measured on males and females, Heath, Bulfield, Thompson, and Keightley (1995) tested for change in variance components over time in a selection experiment, Koerhuis and Thompson (1997) fitted maternal models based on ideas of Falconer while Visscher and Thompson (1992) compared genetic variances from the male and female sides. In all these cases, we were investigating what seemed reasonable genetic hypotheses. I doubt these hypotheses would be seen as important in the new animal and plant genomic era, but it has been suggested that this paradigm is relevant for analysis of human genomic data. This experience emphasized to me John Nelder's assertion that, because of scarce resources, statistical software should be carefully designed to be as useful as possible. This we attempted to do by surveying and reviewing the possible models we would wish to fit. These include various correlated variance structures including autoregressive, moving average, ante dependence, repeated measure and separable models (Gilmour & Thompson, 1998; Welham, Thompson, & Gilmour, 1998), genetic, spatial and temporal relationships and plant breeding models (Gilmour, Cullis, Frensham, & Thompson, 1998). Not everything was anticipated; so facilities were introduced to allow users to generate their own variance structures and basis functions. Later, facilities were added to deal with singular matrices (Thompson, 2009; Kelly, 2013).

Throughout, Arthur Gilmour has taken responsibility for programming. The initial agreement was not to write a self-contained program but provide a kernel for Genstat. However, Genstat could not keep up with Arthur Gilmour's speed of programming so a self-contained program was developed and released. ASReml has dominated Arthur's life driven by a desire to give colleagues more appropriate analyses. For instance, his Association for the Advancement of Animal Breeding and Genetics citation for fellowship states: "A discussion group has also been set up that is better described as ask Arthur a question. His generosity in time to individually answer and his resistance to describing perhaps 50% of the questions as stupid are exemplary."

Development of the program highlighted areas where improved computing algorithms and improved statistical estimation were needed. For instance, in the computing area, a sparse algorithm was developed for forming linear combinations of estimates (Gilmour, Cullis, Welham, Gogel, & Thompson, 2004) and concepts of linear model specification were extended to develop a succinct user-friendly syntax for specifying the required combinations (Welham, Cullis, Gogel, Gilmour, & Thompson, 2004). Also, a popular equation ordering algorithm was improved to deal with animal and plant breeding data (Gilmour & Thompson, 2006). In the statistical area,

estimation for multivariate data was developed in detail (Jensen, Mantysaari, Madsen, & Thompson, 1997 [D]) leading to a convenient PX(EM) algorithm that was introduced into ASReml. Factor analysis and reduced rank variance models methods that took advantage of sparsity were developed (Thompson, Cullis, Smith, & Gilmour, 2003, and Kelly, Cullis, Gilmour, Eccleston, & Thompson, 2009). A recent application of these methods is in Tolhurst, Matthews, Smith and Cullis (2019). Methods were also developed for spline models (Welham, Cullis, Kenward, & Thompson, 2006, 2007). In the genetic area, methods were introduced for QTL analysis (Gilmour, 2007 and Verbyla, Cullis, & Thompson, 2007) and a multi-environmental reduced tree model with useful computational savings was introduced (Cullis, Jefferson, Thompson, & Smith, 2014). Interaction with Brian Cullis and a series of students also led to a better ASReml program. This interaction included work on spatial modelling (Gogel, 1997, Haskard, 2006, Stringer, 2006 and De Faveri, 2013), multiplicative mixed models (Smith, 1999 and Ganesalingam, 2013), iterative schemes (Knight, 2008 and Diffey, 2012), singular variance matrices (Kelly, 2013) and generalized linear mixed models (Collins, 2008).

The focus in developing ASReml was initially on its kernel, and it was freely acknowledged that its user interface was not to the level of other packages. Other packages such as R and Genstat allow the data preparation, analysis, postprocessing to be carried out within a single framework. To take account of this need, Sue Welham ported the kernel of ASReml into Genstat (Payne et al., 1997) and Dave Butler and Brian Cullis built on the kernel to write ASReml-R (Butler, Cullis, Gilmour, Gogel, & Thompson, 2018) [D]. Castaways on the desert island are asked if they could only take one record which would it be, in my case I would take this paper, a product of synergistic team work, as it includes derivation of REML, the AI algorithm, REML made accessible for a wide range of models, a user friendly syntax, a skilfully constructed kernel and is a well-documented user guide. Simon Harding wrote ASReml-W, a graphical tool allowing the user to edit programs, run ASReml and then view the output, before saving results.

One important development has been the simplification in the way models are specified. Gilmour (2019) discusses this in this issue.

9 | SAMPLING-BASED ESTIMATION

Estimation based on using sparse matrix methods to calculate the terms involved in REML estimation has found wide application, but in some cases, for example large data sets, complex models including multivariate data and genomic relationship matrices, it is often impossible to calculate the terms required from the exact inverse of the coefficient matrix using direct methods. A natural suggestion, for example Garcıa-Cortes, Moreno, Varona, and Altarriba (1992), is to use sampling schemes to estimate the required terms. Two sampling methods, data augmentation, based on Gibbs sampling, and Monte Carlo AIREML based on a simple sample generation scheme will now be discussed.

9.1 | Data augmentation

This data augmentation method is based on work by Thompson (1994) and Clayton and Rasbash (1999) using ideas based on Gibbs sampling. To motivate the method, we note that if we were just interested in estimates of the linear effects from (9) we might avoid inversion of $\mathbf{C}$ by solving (9) iteratively using block Gauss–Seidel methods with the blocks associated with sub-models. For instance, writing the model (1) $\mathbf{y}=\mathbf{Xa}+\mathbf{Zu}+\mathbf{e}$ in the more homogeneous form $\mathbf{y}=\mathbf{Wb}+\mathbf{e}$ then partitioning the model into terms associated with m sub-models $\mathbf{y}=\sum_{j=1}^{m}\mathbf{W}_j\mathbf{b}_j+\mathbf{e}$ using $\mathbf{W}=[\mathbf{W}_1\mathbf{W}_2\ldots\mathbf{W}_m]$ and $\mathbf{b}^T=\left[\mathbf{b}_1^T\mathbf{b}_1^T\ \ldots\ \mathbf{b}_m^T\right]$ with effects in one sub-model independent of effects in the other sub-models. In a similar manner, we can partition $\mathbf{C}$ into component matrices $\mathbf{C}_{ij}$ associated with sub-models i and j and define $\mathbf{c}_i=\mathbf{W}_i^T\mathbf{R}^{-1}\mathbf{y}$. Then, block Gauss–Seidel iteration involves successively solving

$$\mathbf{C}_{ii}\mathbf{b}_i=\mathbf{c}_i-\sum_{j=1,j\neq i}^{m}\mathbf{C}_{ij}\mathbf{b}_j$$

for $\mathbf{b}_i$ for i = 1, …, m and iterating the process until the estimates converge. A simplified form of Gibbs block sampling follows a similar paradigm, first obtaining the expected value of $\mathbf{b}_i$, $\mathbf{b}_{ei}$, given sample values of the other estimates, (b_{sj}, j = 1, …, m, j ≠ i) from

$$\mathbf{C}_{ii}\mathbf{b}_{ei}=\mathbf{c}_i-\sum_{\mathbf{j=1,j\neq i}}^{m}\mathbf{C}_{ij}\mathbf{b}_{sj} \tag{19}$$

Then, a sample of the estimate of $\mathbf{b}_i$, $\mathbf{b}_{si}$ can be found using $\mathbf{b}_{si}=\mathbf{b}_{ei}+\mathbf{e}_i$, where $\mathbf{e}_i$ is sampled from a normal distribution with inverse variance $\mathbf{C}_{ii}$.

An alternative interpretation of (19) is that for the i-th sub-model, data are augmented by adjustments for fitted values in the other sub-models ($\mathbf{w}_j=\mathbf{W}_j\mathbf{b}_{sj}$) of the form $\mathbf{y}_i=\mathbf{y}-\sum_{j=1,j\neq i}^{m}\mathbf{w}_j$ and a model to the augmented data of the form $\mathbf{y}_i=\mathbf{W}_i\mathbf{b}_{ei}+\mathbf{e}$ is fitted, to find expected values $\mathbf{b}_{ei}$, updating the variance parameters in the i-th sub-model and forming sample values $\mathbf{b}_i$ from $\mathbf{b}_{ei}$ and $\mathbf{e}_i$. The procedure is carried out for each sub-model in turn and repeated for several iterations. Estimates of the variance parameters are based on the average over all iterations excluding burn-in iterations.

We can contrast the three methods. AIREML solves the full set of equations to get $\tilde{\mathbf{u}}$ and the sparse inverse subset

of $\mathbf{C}$, data augmentation uses the expected value of terms in the sub-models, $\mathbf{b}_{ei}$ given samples from the other sub-models and the sparse inverse subset of the $\mathbf{C}_{ii}$ from the sub-models and Gibb's sampling just uses samples $\mathbf{b}_{si}$. In cases where the sparse inverse subset is difficult to calculate methods using sampling have advantages. Data augmentation, because it uses the expected values $\mathbf{b}_{ei}$ rather than the sampled values $\mathbf{b}_{si}$ used in Gibb's sampling, has less sampling variation than Gibb's sampling. This procedure is often called Rao–Blackwellization (Gelfand & Smith, 1990). There are several ways of constructing the variance parameter updates in data augmentation. For instance,

- an EM update could be used, based on the Gibbs sampling analogy,
- as Clayton and Rasbash (1999) suggest, iteration could continue until the estimates converge,
- as Stewart et al. (2012) do, use AIREML updates based on the sub-model,
- if several sub-models are used, fit the sub-models then use updates to variance parameters based on solving an analogy of (15).

9.2 | Monte Carlo AIREML (MCAIREML)

An alternative approach making use of the AI algorithm based on the simple sampling paradigm of Garcıa-Cortes et al. (1992) was suggested by Matilainen, Mäntysaari, Lidauer, Strandén and Thompson (2013) and Loh et al. (2015). They suggest calculating the solutions to the MME model equations needed in the calculation of the gradients (8)–(12) using an iterative scheme based on preconditioned conjugate gradient schemes. They also suggest that the trace terms in (9) and (10) can be estimated using s samples of $\mathbf{y}_j$ (j = 1..., s) generated as $\mathbf{y}_j = \mathbf{Z}\mathbf{u}_j + \mathbf{e}_j$ with

$$\mathbf{u}_j \sim \mathcal{N}(0,\mathbf{G}) \text{ and } \mathbf{e}_j \sim \mathcal{N}(0,\mathbf{R}) \tag{20}$$

Then, the trace terms in (8) and (9) can be estimated as

$$\mathrm{tr}[\mathbf{R}^{-1}(\delta\mathbf{R}/\delta\theta_{ri})\mathbf{R}^{-1}(\mathbf{R}-\mathbf{W}\mathbf{C}^{-1}\mathbf{W}^{\mathbf{T}})] = 1/s(\sum_{j=1}^{s}(\hat{\mathbf{e}}_j^{\mathbf{T}}\mathbf{R}^{-1}(\delta\mathbf{R}/\delta\theta_{ri})\mathbf{R}^{-1}\hat{\mathbf{e}}_j)) \tag{21}$$

and

$$\mathrm{tr}[\mathbf{G}^{-1}(\delta\mathbf{G}/\delta\theta_{gi})\mathbf{G}^{-1}(\mathbf{G}-\mathbf{C}^{\mathbf{uu}})] = 1/s(\sum_{j=1}^{s}(\hat{\mathbf{u}}_j^{\mathbf{T}}\mathbf{G}^{-1}(\delta\mathbf{G}/\delta\theta_{gi})\mathbf{G}^{-1}\hat{\mathbf{u}}_j) \tag{22}$$

where $\hat{j}$ and $\hat{\mathbf{e}}_j$ are formed from the mixed model equations with data $\mathbf{y}_j$. There are alternative terms to estimate the trace terms in (11) and (12) in terms of prediction errors $(\mathbf{u}_j - \hat{\mathbf{u}}_j)$ and $(\mathbf{e}_j - \hat{\mathbf{e}}_j)$

$$\mathrm{tr}[(\delta\mathbf{C}/\delta\theta_{ei})\mathbf{C}^{-1}] = 1/s\,(\sum_{j=1}^{s}(\mathbf{e}_j - \hat{\mathbf{e}}_j)^{\mathbf{T}}\mathbf{R}^{-1}(\delta\mathbf{R}/\delta\theta_{ri})\mathbf{R}^{-1}(\mathbf{e}_j - \hat{\mathbf{e}}_j) \tag{23}$$

and

$$\mathrm{tr}[(\delta\mathbf{C}_{uu}/\delta\theta_{gi})\mathbf{C}^{\mathbf{uu}}] = 1/s\,(\sum_{j=1}^{s}(\mathbf{u}_j - \hat{\mathbf{u}}_j)^{\mathrm{T}}\mathbf{G}^{-1}(\delta\mathbf{G}/\delta\theta_{gi})\mathbf{G}^{-1}(\mathbf{u}_j - \hat{\mathbf{u}}_j) \tag{24}$$

Matilainen et al. (2013) note that the optimal convergence criterion is more difficult to ascertain because of the sampling at each iteration. Loh et al. (2015) suggest that these problems are reduced if instead of generating new samples in each iteration base samples with

$$\mathbf{u}_{bi} \sim \mathcal{N}(0,\mathbf{I}) \text{ and } \mathbf{e}_{bi} \sim \mathcal{N}(0,\mathbf{I}) \tag{25}$$

are first generated and then in each iteration scaled values of $\mathbf{u}_i$ and $\mathbf{e}_i$ are constructed pausing

$\mathbf{u}_i = \mathbf{G}^{1/2}\mathbf{u}_{bi}$ and $\mathbf{e}_i = \mathbf{R}^{1/2}\mathbf{e}_{bi}$ with $\mathbf{G} = \mathbf{G}^{1/2}(\mathbf{G}^{1/2})^{\mathrm{T}}$ and $\mathbf{R} = \mathbf{R}^{1/2}(\mathbf{R}^{1/2})^{\mathrm{T}}$.

Matilainen, Mantysarri and Stranden (2019) in this issue take a similar approach by reusing the same random numbers within each sampling to remove the fluctuations of convergence criteria. Both Matilainen et al. (2013) and Loh et al. (2015) advocate that the average information matrix be calculated from (13) with $\mathbf{Pf}_j$ being calculated from fitting a mixed model to $\mathbf{f}_j$ again using preconditioned conjugate gradient methods.

These require solving v sets of equations, one for each variance parameter. In some cases, this is computationally expensive, for example Stewart et al. (2012) considered models with over 200 parameters. Using the incremental model (14)-(16) allows the increments to be calculated with only solving one set of equations.

9.3 | Open questions

These methods raise several open questions. These include the following:

- Can the data augmentation method be easily adapted to give REML estimates?
- In the data augmentation method, when most of the computations are concerned with absorbing and forming the sparse inverse subset of the sub-model coefficient matrices, would fitting s samples based on independent noise (as is used in the MCAIREML method) and update estimates based on the s samples be computationally and statistically useful?
- For both methods, could the latter be improved by generating dependent noise based on the s quantiles of the noise distribution?
- Can the data augmentation method be adapted for cases when using different elements in a set of correlated random effects in different sub-models would reduce the computation, for example separating male and female genetic

effects based on a pedigree relationship matrix ? If the $\mathbf{G}$ matrix can be written as $\mathbf{G}=\mathbf{A}\sigma^2_{\mathbf{u}}$, one suggestion would be to augment the data with extra data ,$\mathbf{y_a}$, with values zero and model $\mathbf{y_a}=\mathbf{Lu}+\mathbf{e_a}, \mathbf{e_a}\sim\mathcal{N}\left(0,\mathbf{D}\sigma^2_{\mathbf{u}}\right)$ with $\mathbf{A}^{-1}=\mathbf{LDL}^{\mathbf{T}}$ and $\mathbf{L}$, a lower triangular matrix with diagonal elements 1. The extra data could be thought of as Mendelian sampling terms. With this parameterization, $\mathbf{u}$ and its associated design matrices can easily be apportioned into simpler submodels based on say male, female and perhaps ancestor effects.

- Can the terms in (21)–(24) be combined to give estimates of the trace terms with less sampling variation, just as Hickey, Veerkamp, Calus, Mulder, and Thompson (2009) [D] showed when estimating prediction error variances?
- Which method, data augmentation or MCAIREML, is the computationally more efficient? Is there a hybrid scheme that takes advantage of the strengths of the data augmentation method in reducing sampling variation and giving a relatively easy paradigm for estimating the likelihood of a model, and the strengths of the MCAIREML scheme in reducing the effort in calculating solutions?

In conclusion , my Roslin colleagues asked me, based on my experience, what advice I give to students. My advice would be

- Choose mentors who give wise counsel.
- If possible work in a happy team that has a team ethic, generates synergy, has mutual trust and respect, respect for sponsors and users and has a shared vision.
- Work with a supervisor who will give a rigorous grounding, as Bill Hill and Brian Cullis did, but I could only aspire to.
- Adapt to new methods quickly, I think I was slow to see the potential of sampling methods.
- Using two alternative methods to check results can aid understanding and give extra confidence to believe I have the to believe I have got correct results.

ACKNOWLEDGEMENTS

The device of using desert island papers has allowed me to acknowledge the influence of mentors and colleagues on my career. Others, who I would class as facilitators, do not naturally fit into this framework, and they include Roger Land who was instrumental in my transfer to what is now Roslin Institute and he had the vision to see my skills and interests fitted the needs of Roslin. This enabled me to play an active role in postgraduate education, including supervising over 20 Ph.D. students. I found it very rewarding helping people to grow and learn to fly. I am grateful for the support to ASReml by VSNi that has allowed better documentation, more rigorous quality control and probably lead to more impact of the program. Rothamsted Research's share of the ASReml royalties has supported further statistical research in the institute and at the University of Woolongong. I am pleased that more recently VSNi has agreed that ASReml be free for students and for users in over 100 less well-off countries. A triumvirate of women have had lasting influence on my career my mother, Mary Thompson, was insistent I was not going to be a farmer, Liz Archibald skilfully organized me when I was at Roslin, and Harriet Mackintosh has unstintingly supported me throughout my career. Finally, this paper is based on a talk John Hickey suggested I give and he and Gregor Gorjanc organized an international meeting that allowed me to present the ideas. Castaways on the desert island are allowed one book, it seems appropriate to take this issue of the journal as it includes some of the material presented at the international meeting.

ORCID

Robin Thompson https://orcid.org/0000-0003-3319-6938

REFERENCES

Atkins, K. D., & Thompson, R. (1986a). Predicted and realised responses to selection for an index of bone length and body weight in Scottish Blackface sheep. 1. Responses in the index and component traits. *Animal Production*, *43*, 421–436.

Atkins, K. D., & Thompson, R. (1986b). Predicted and realised responses to selection for an index of bone length and body weight in Scottish Blackface sheep. 2. Correlated responses in lifetime productivity. *Animal Production*, *43*, 437–446.

Butler, D., Cullis, B. C., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2018). *ASReml-R Reference Manual Version 4*. 176 pp VSN International, Hemel Hempstead.

Clayton, D., & Rasbash, J. (1999). Estimation in large cross random-effect models by data augmentation. *Journal of the Royal Statistical Society*, *162*,425–436.

Collins, D. (2008). *Performance of estimation methods for generalized linear mixed models*. PhD. Thesis, University of Wollongong.

Crump, R. E., Thompson, R., Haley, C. S., & Mercer, J. (1997). Individual animal model estimates of genetic correlations between performance test and reproduction traits of Landrace pigs performance tested in a commercial nucleus herd. *Animal Science*, *65*, 291–298. https://doi.org/10.1017/S135772980001660X

Cullis, B. R., Jefferson, P., Thompson, R., & Smith, A. B. (2014). Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a Pinus radiata breeding programme. *Theoretical and Applied Genetics*, *127*, 2193–2210. https://doi.org/10.1007/s00122-014-2373-0

Cullis, B. R., Smith, A. B., & Thompson, R. (2004). Perspectives of ANOVA, REML and a General linear mixed model. In N. M. Adams, M. J. Crowder, D. Hand & D. A. Stephens (Eds.), *Methods and models in statistics in honour of Professor John Nelder FRS* (pp. 53–94). London, UK: Imperial College Press.

Cunningham, E. P., & Henderson, C. R. (1968). An iterative procedure for estimating fixed effects and variance components in mixed model situations. *Biometrics*, *24*, 13–25. https://doi.org/10.2307/2528457

De Faveri, J. (2013). *Spatial and temporal modelling of perennial crop variety selection trials*. PhD thesis, The University of Adelaide.

De Faveri, J., Verbyla, A. P., Cullis, B. R., Pitchford, W. S., & Thompson, R. (2017). Residual variance-covariance modelling in analysis of multivariate data from variety selection trials. *Journal of Agricultural, Biological and Environmental Statistics*, *22*, 1–22. https://doi.org/10.1007/s13253-016-0267-0

Diffey, S. M. (2012). *A new REML (PX)EM algorithm for linear mixed models and factor analytic mixed models*. PhD. Thesis, Australian National University.

Fowler, V. R., Bichard, M., & Pease, A. H. R. (1976). Objectives in pig breeding. *Animal Production*, *23*, 365–387.

Furlotte, N. A., & Eskiny, E. (2015). Efficient multiple-trait association and estimation of genetic correlation using the matrix-variate linear mixed model. *Genetics*, *200*, 59–68. https://doi.org/10.1534/genetics.114.171447

Ganesalingam, A. (2013). *Improving the efficiency of selection in a plant breeding program using information on correlated traits, ancestry and environments*. PhD. thesis, University of Western Australia.

Garcıa-Cortes, L. A., Moreno, C., Varona, L., & Altarriba, J. (1992). Variance component estimation by resampling. *Journal of Animal Breeding and Genetics*, *109*, 358–363.

Gelfand, A., & Smith, A. F. M. (1990). Sampling approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398–409.

Gilmour, A. R. (2007). Mixed model regression mapping for QTL detection in experimental crosses. *Computational Statistics and Data Analysis*, *51*, 3749–3764. https://doi.org/10.1016/j.csda.2006.12.031

Gilmour, A. R. (2019) Average Information REML in practice. *Journal of Animal Breeding and Genetics*, *136*, 262–272.

Gilmour, A. R., Cullis, B. R., Frensham, A. B., & Thompson, R. (1998). (Co)Variance structures for linear models in the analysis of plant improvement data. *Compstat 1998 Proceedings*, 53–64.

Gilmour, A., Cullis, B., Welham, S. J., Gogel, B., & Thompson, R. (2004). An efficient computing strategy for prediction in mixed linear models. *Computational Statistics and Data Analysis*, *44*, 571–586. https://doi.org/10.1016/S0167-9473(02)00258-X

Gilmour, A. R., & Thompson, R.. (1998). Modelling variance parameters in ASREML for repeated measures data. *Proceedings of 6th World Congress on Genetics Applied to Livestock Production*, *27*, 453–454.

Gilmour, A. R., & Thompson, R.. (2006). Equation ordering for Average Information REML. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production*, *27*, 13–16.

Gilmour, A. R., Thompson, R., & Cullis, B. R. (1995). Average information REML, an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics 51*,1440–1450.

Gogel, B. J. (1997). *Spatial analysis of multi-environment variety trials*. PhD. Thesis, University of Adelaide.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, *61*, 383–385.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation. *Journal of the American Statistical Association*, *72*, 320–338.

Haskard, K. A. (2006). *Anisotropic Matern correlation and other issues in model-based geostatistics*. PhD thesis, University of Adelaide.

Hayes, J. F., & Hill, W. G. (1980). A reparameterization of a genetic selection index to locate its sampling properties. *Biometrics*, *36*, 237–248. https://doi.org/10.2307/2529975

Hayes, J. F., & Hill, W. G. (1981). Modification of estimates of parameters in the construction of genetic selection indices ('Bending'). *Biometrics*, *37*, 483–493.

Heath, S. C., Bulfield, G., Thompson, R., & Keightley, P. D. (1995). Rates of change of genetic parameters of body weight in selected mouse lines. *Genetical Research*, *66*, 19–25. https://doi.org/10.1017/S0016672300034352

Henderson, C. R., Kempthorne, O., Searle, S. R., & Von Krosigk, C. M. (1959). The estimation of genetic and environmental trends from records subject to culling. *Biometrics*, *15*, 192–218.

Hickey, J. M., Veerkamp, R. V., Calus, M. P. L., Mulder, H., & Thompson, R. (2009). Estimation of prediction error variances via Monte Carlo sampling methods using different formulations of the prediction error variance. *Genetics Selection Evolution*, *41*, 22–32. https://doi.org/10.1186/1297-9686-41-23

Hill, W. G. (1972). Estimation of genetic change. I. General theory and design of control populations. *Animal Breeding Abstracts*, *40*, 1–16.

Hill, W. G., & Thompson, R. (1978). Probabilities of non-positive definite between-group or genetic covariance matrices. *Biometrics*, *34*, 429–439. https://doi.org/10.2307/2530605

Jensen, J., Mantysaari, E. A., Madsen, P., & Thompson, R. (1997). Residual maximum likelihood estimation of (Co) variance components in multivariate mixed linear models using average information. *Indian Journal Agricultural Statistics*, *49*, 215–236.

Johnson, D. L., & Thompson, R. (1995). Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and a quasi-Newton procedure. *Journal of Dairy Science*, *78*, 449–456.

Kelly, A. M. (2013). *Estimation of a linear mixed model with a reduced rank variance matrix*. PhD. Thesis, University of Queensland.

Kelly, A. M., Cullis, B. R., Gilmour, A. R., Eccleston, J. A., & Thompson, R. (2009). Estimation in a multiplicative mixed model involving a genetic relationship matrix. *Genetics Selection Evolution*, *41*, 33–42. https://doi.org/10.1186/1297-9686-41-33

Knight, E. (2008). *Improved iterative schemes for REML estimation of variance parameters in linear mixed models*. PhD thesis, University of Adelaide.

Koerhuis, A. N. M., & Thompson, R. (1997). Models to estimate maternal effects for juvenile body weight in broiler chickens. *Genetics, Selection and Evolution*, *29*, 225–249.

LaMotte, L. R. (1973). Quadratic estimation of variance components. *Biometrics*, *29*, 311–330. https://doi.org/10.2307/2529395

Lee, Y., & Nelder, J. A. (1996). Hierarchical generalised linear models (with discussion). *Journal of the Royal Statistical Society B*, *58*, 619–656.

Lee, S. H., & Van Der Werf, J. H. J. (2006). An efficient variance component approach implementing an average information REML suitable for combined LD and linkage mapping with a general complex pedigree. *Genetics Selection Evolution*, *38*, 25–43.

Loh, P.-R., Bhatia, G., Gusev, A., Finucane, H. K., Bulik-Sullivan, B. K., Pollack, S. J.… Price, A. L. (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature Genetics*, *47*, 1385–1392.

Matilainen, K., Mäntysaari, E. A., Lidauer, M. H., Strandén, I., & Thompson, R. (2013). Employing a Monte Carlo algorithm in Newton-type methods for restricted maximum likelihood

estimation of genetic parameters. *PLoS ONE*, *8*, e80821. https://doi.org/10.1371/journal.pone.0080821

Matilainen, K., Mäntysaari, E. A., & Strandén, I. (2019). Efficient Monte Carlo algorithm for restricted maximum likelihood (REML) estimation of genetic parameters. *Journal of Animal Breeding and Genetics*, *136*, 252–261.

Meyer, K. (1985). Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics*, *41*, 163–165. https://doi.org/10.2307/2530651

Meyer, K. (1988). DFREML—A set of programs to estimate variance components under an individual animal model. *Journal of Dairy Science*, *71*, 33–34. https://doi.org/10.1016/S0022-0302(88)79977-4

Mrode, R., Smith, C., & Thompson, R. (1990). Selection for rate and efficiency of lean gain in hereford cattle. I. Selection pressure applied and direct responses. *Animal Production*, *51*, 23–34.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A*, *135*, 370–384. https://doi.org/10.2307/2344614

Patterson, H. D. (1964). Theory of cyclic rotation experiments. *Journal of the Royal Statistical Society B*, *26*, 1–45. https://doi.org/10.1111/j.2517-6161.1964.tb00535.x

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, *58*, 545–554. https://doi.org/10.1093/biomet/58.3.545

Payne, R. W., Lane, P. W., Baird, D. B., Gilmour, A. R., Harding, S. A., Morgan, G. W., … Welham, S. J. (1997). *Genstat 5 Release 4.1 reference manual supplement* (pp. 229). Oxford, UK: Numerical Algorithms Group.

Pig Industry Development Authority (PIDA) (1966). *Combined Testing. PIDA Internal Report (DA 188). Meat and Livestock Commission*, Bletchley, Bucks.

Rao, C. R. (1972). Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association*, *67*, 112–115. https://doi.org/10.1080/01621459.1972.10481212

Robertson, A. (1962). Weighting in the estimation of variance components in the unbalanced single classification. *Biometrics*, *18*, 413–417.

Robertson, A., & Rendel, J. M. (1954). The performance of heifers got by artificial insemination. *The Journal of Agricultural Science*. *44*,184–192.

Smith, A. B. (1999). *Multiplicative mixed models for the analysis of multi-environment trial data*. PhD thesis, University of Adelaide.

Stewart, I. D., White, I. M. S., Gilmour, A. R., Thompson, R., Woolliams, J. A., & Brotherstone, S. (2012). Estimating variance components and predicting breeding values for eventing disciplines and grades in sport horses. *Animal*, *6*, 1377–1388. https://doi.org/10.1017/S1751731112000596

Stringer, J. K. (2006). *Joint modelling of spatial variability and interplot competition to improve the efficiency of plant improvement*. PhD Thesis, University of Queensland.

Takahashi, K., Fagan, J., & Chin, M.-S. (1973). *Formation of a sparse bus impedance matrix and its application to short circuit study*. 8th PICA Conf. Proc., Minneapolis, MN.

Thompson, R. (1969). Iterative estimation of variance components for non-orthogonal data. *Biometrics*, *25*, 767–773. https://doi.org/10.2307/2528574

Thompson, R. (1973). The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics*, *29*, 527–550. https://doi.org/10.2307/2529174

Thompson, R. (1976a). The estimation of maternal genetic variances. *Biometrics*, *32*, 903–917. https://doi.org/10.2307/2529273

Thompson, R. (1976b). Relationship between the cumulative difference and best linear unbiased predictor methods of evaluating bulls. *Animal Production*, *23*, 15–24.

Thompson, R. (1977a). Estimation of quantitative genetic parameters. *Proceedings of the International Conference on Quantitative Genetics* (pp. 639-657), Ames, IO: Iowa State University Press.

Thompson, R. (1977b). The estimation of heritability with unbalanced data: II. Data available on more than two generations. *Biometrics*, *33*, 497–504.

Thompson, R. (1979). Sire evaluation. *Biometrics*, *35*, 339–353.

Thompson, R. (1994). Integrating best linear unbiased prediction and maximum likelihood estimation. *5th World Congress on Genetics Applied to Livestock Production*, *18*, 337–340.

Thompson, R. (2009). Latent mixed models. In *Proceedings 18th Conference of the Association for the Advancement of Animal Breeding and Genetics* (pp. 398-405).

Thompson, R., & Atkins, K. A. (1993). Sources of information for estimating heritability from selection experiments. *Genetical Research*, *63*, 49–55. https://doi.org/10.1017/S0016672300032079

Thompson, R., Cullis, B. R., Smith, A., & Gilmour, A. (2003). A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. *Australian and New Zealand Journal of Statistics*, *45*, 445–459. https://doi.org/10.1111/1467-842X.00297

Thompson, R., Wray, N. R., & Crump, R. E. (1994). Calculation of prediction error variances using sparse matrix methods. *Journal of Animal Breeding and Genetics*, *111*, 102–109. https://doi.org/10.1111/j.1439-0388.1994.tb00443.x

Tolhurst, D. J., Mathews, K. L., Smith, A. B., & Cullis, B. C. (2019). Genomic selection in multi-environment plant breeding trials using a factor analytic linear mixed model. *Journal of Animal Breeding and Genetics*, *136*, 279-300.

Verbyla, A. P. (1990). A conditional derivation of residual maximum likelihood. *Australian Journal of Statistics*, *32*, 227–230. https://doi.org/10.1111/j.1467-842X.1990.tb01015.x

Verbyla, A. P., Cullis, B. R., & Thompson, R. (2007). The analysis of QTL by simultaneous use of the full linkage map. *Theoretical and Applied Genetics*, *116*, 95–111.

Visscher, P. M., & Thompson, R. (1992). Comparisons between genetic variances estimated from different types of relatives in dairy cattle. *Animal Production*, *55*, 315–320. https://doi.org/10.1017/S0003356100021000

Welham, S. J., Cullis, B. R., Gogel, B., Gilmour, A., & Thompson, R. (2004). Prediction in linear mixed models. *Australian and New Zealand Journal of Statistics*, *46*, 325–347. https://doi.org/10.1111/j.1467-842X.2004.00334.x

Welham, S. J., Cullis, B. R., Kenward, M. G., & Thompson, R. (2006). The analysis of longitudinal data using mixed model L-Splines. *Biometrics*, *62*, 392–401.

Welham, S. J., Cullis, B. R., Kenward, M. G., & Thompson, R. (2007). A comparison of mixed model splines for curve fitting. *Australian and New Zealand Journal of Statistics*, *49*, 1–23. https://doi.org/10.1111/j.1467-842X.2006.00454.x

Welham, S. J., Thompson, R., & Gilmour, A. R. (1998). A general form for specification of correlated error models, with allowance for heterogeneity. *Compstat1998 Proceedings*, 479–484.

Woolliams, J. A., & Thompson, R. (1994). A theory of genetic contributions. *Proceedings of the World Congress on Genetics Applied to Livestock Production*, *19*, 127–134.

Wray, N. R., & Thompson, R. (1989). Prediction of rates of inbreeding in selected populations. *Genetical Research*, *55*, 41–54. https://doi.org/10.1017/S0016672300025180

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*, 76–82.

How to cite this article: Thompson R. Desert island papers—A life in variance parameter and quantitative genetic parameter estimation reviewed using 16 papers. *J Anim Breed Genet*. 2019;136:230–242. https://doi.org/10.1111/jbg.12400