meBLUP = ssBLUP without relationship matrices

Arnold Neumaier

Fakultät für Mathematik
Universität Wien, Österreich

joint work with

Eildert Groeneveld

Institut für Nutztiergenetik
Bundesforschungsinstitut für Tiergesundheit
Mariensee, Deutschland

Lecture given on September 20, 2018, Göttingen
Tagung der Projektgruppe
"Genetisch-Statistische Methoden" der
Deutschen Gesellschaft für Züchtungskunde

These slides can be found at
`http://www.mat.univie.ac.at/~neum/ms/meBLUPSlides.pdf`

meBLUP is a new method for the genetic evaluation of joint pedigree and genomic data, totally avoiding relationship matrices or their inverse, which figure prominently in all previous approaches.

When either the pedigree information or the genomic information is absent, **meBLUP** reduces to equations that give exactly the same BLUPs and BLUEs as the standard formulation.

This is work in progress.
First numerical results indicate that **meBLUP** is likely to be competitive with **ssBLUP**, the current technique for the joint evaluation of pedigree and genomic data.

Contents

- Mixed models
- Animal breeding models
- Breeding values from genomic data
- The model equation (ME) approach
- Equivalent mixed models
- Numerical results and conclusions

Mixed models

- Linear stochastic models
- The Gauss–Markov theorem
- Mixed models
- Mixed model equations
- The ME formulation of a mixed model

Linear stochastic models

A **linear stochastic model** is a relation of the form

$$y = Ax + \text{noise}(C)$$

between random vectors x and y with constant matrices A and C.

This is interpreted as the statement that $y = Ax + \varepsilon$,
where the residual $\varepsilon := y - Ax$ is noise with zero mean
and covariance matrix C:

$$\langle \varepsilon \rangle = 0, \quad \langle \varepsilon \varepsilon^T \rangle = C.$$

As a conseqence, the means $\overline{x}$ of x and $\overline{y}$ of y are related by $\overline{y} = A\overline{x}$.

No other assumptions about the noise are needed; it does not need to be Gaussian.

The Gauss–Markov theorem

Theorem. Suppose that x and y are random vectors and $y = Ax + \text{noise}(C)$. If $\begin{pmatrix} C & A \\ A^T & 0 \end{pmatrix}$ is nonsingular, there is a unique **best linear unbiased estimator (BLUE)** $\widehat{x}$ in the sense of uniform optimality:

For every linear unbiased estimator $\widetilde{x}$ of x and all $a \in \mathbb{R}^n$,

$$\text{var}(a^T \widetilde{x}) \geq \text{var}(a^T \widehat{x}).$$

If C and $A^T C^{-1} A$ are nonsingular, the BLUE is given by the solution of the **normal equations**

$$A^T C^{-1} A x = A^T C^{-1} y.$$

Mixed models

A **mixed model** is a linear stochastic model of the form

$$y = X\beta + Zu + \text{noise}(D), \quad u = \text{noise}(G),$$

where D and G are covariance matrices,

• y is a data vector of **records** of observed **traits**,

• β is a (fixed but unknown) coefficient vector of **fixed effects** with corresponding model matrix X,

• $u = \text{noise}(G)$ is a (random, unknown) coefficient vector of **random effects** with corresponding model matrix Z, and

• $\eta = y - X\beta - Zu = \text{noise}(D)$ is noise from observation errors and modeling errors, uncorrelated with u.

Mixed model equations

HENDERSON 1949 discovered his **mixed model equations**

$$\begin{pmatrix} X^T D^{-1} X & X^T D^{-1} Z \\ Z^T D^{-1} X & Z^T D^{-1} Z + G^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ u \end{pmatrix} = \begin{pmatrix} X^T D^{-1} y \\ Z^T D^{-1} y \end{pmatrix}$$

for estimating β and u, which only involve the inverse G^{-1}.

This made estimation practical at large scale since in the traditional applications to animal breeding, G^{-1} is always sparse and (HENDERSON 1976) easily computable from the pedigree.

The vector $\widehat{u}$ (containing the estimated breeding values) is usually called the **best linear unbiased predictor (BLUP)** of the random effects.

The ME formulation of a mixed model

Treating the measurement relation and the random effect condition as two separate model equations (ME) gives the **ME formulation**

$$\begin{pmatrix} y \\ 0 \end{pmatrix} = \begin{pmatrix} X & Z \\ 0 & -I \end{pmatrix} \begin{pmatrix} \beta \\ u \end{pmatrix} + \text{noise} \begin{pmatrix} D & 0 \\ 0 & G \end{pmatrix}.$$

This is a stochastic linear model of the form to which the Gauss–Markov theorem applies.

As observed by Fellner 1986, Henderson's mixed model equations are just the normal equations for this model. Thus the BLUP vector $\widehat{u}$ is part of the best linear unbiased estimator $\begin{pmatrix} \widehat{\beta} \\ \widehat{u} \end{pmatrix}$ of the ME formulation of the mixed model.

Animal breeding models

- Animal breeding via mixed models
- Animal breeding models
- Equivalent breeding models without relationship matrix
- Pedigree equations
- Missing data patterns
- Solving the normal equations

Animal breeding via mixed models

Mixed model methodology provides the machinery for genetic evaluation in large-scale animal breeding.

The prediction of breeding values is generally done by Henderson's Mixed Model Equations (MME). These involve G^{-1}, which is proportional to the inverse of the relationship matrix.

If the information consists of phenotype, pedigree, and genomic data, the method of choice today is single-step BLUP (ssBLUP). In ssBLUP, the inverse relationship matrix is computed by a more or less heuristic combination of information from the pedigree-based (numerator) relationship matrix and a genomic relationship matrix.

The details differ from author to author and from application to application, and must be specially tuned to each extension of the structure of the estimation problem (multiple traits, dominance, unknown parent groups, etc.).

Simple animal breeding models

In the simplest animal breeding models:

• The entries of the vector of fixed effects β are coefficients measuring the influence of known factors on the traits (phenotypes).

• The entries of the vector of random effects u are the **breeding values** of the individual animals, measuring the influence of their genetic constitution on the traits.

• The covariance matrix G of the random effects are a multiple (or Kronecker factor) of the relationship matrix. This is a dense matrix, in large populations far too big to be stored or even generated.

Equivalent breeding models without relationship matrix

The relationship matrix can be completely avoided by modeling the pedigree directly in the form of model equations relating the breeding values of each animal $T(\nu)$ and of their parents $F(\nu)$ and $M(\nu)$.

In the conventional approach, these are just byproducts of the derivation of the mixed model. The ME approach promotes them to fundamental equations.

The normal equations for the resulting augmented system lead to an even sparser system of equations from which the BLUP can be calculated.

Pedigree equations

If both parents are represented in the pedigree, and assuming that each trait is genetically influenced by many independent genes, Mendel's laws lead to a relation of the form

$$\frac{1}{\sqrt{k}}(2\beta_T - \beta_V - \beta_M) = \varepsilon_T = \text{noise}(C_P)$$

where, for unknown parents, β_V or β_M are zero, and

$$k = 4 - \text{number of known parents},$$

with noise independent of the measurement noise.

The resulting mixed model equations are again equivalent to Henderson's when the relationship matrix is obtained by the recipe of Westell et al. 1987.

Missing data patterns

In practice it frequently happens that some data are missing for nonsystematic reasons.

In place of a group of model equations

$$y_\nu = A_\nu x + \text{noise}(C_\nu)$$

one then has to use the reduced model equations

$$Py_\nu = PA_\nu x + \text{noise}(PC_\nu P^T),$$

where P is a projection operator that removes the rows in which data are missing. (For groups of equations where no data are missing, P is the identity matrix.)

Solving the normal equations

The normal equations can be solved by preconditioned conjugate gradients.

This requires routine for the multiplication of the normal equation matrix with a vector.

This product can be computed very efficiently by iteration on data, without forming the normal equation matrix, making the solution process very efficient.

A few hundred iterations produce adequate accuracy, even for the largest problems.

Breeding values from genomic data

- Genomic regression models
- Multistep GBLUP
- ssBLUP (single step BLUP)
- Data compression in GBLUP

Genomic regression models

For some animals, SNP (single nucleotid polymorphism) marker information is available and exploited for BLUP estimation.

In genomic regression models one assumes an approximate relation between the breeding values u_ν and the SNP data $S_{\nu k}$, forming the **SNP matrix** S. Here ν runs over the genotyped animals and k over the different SNP indices.

Typically, this relation is linear, resulting in a relation

$$u \approx Sx.$$

This relation must be solved for the vector x of SNP weights, needed for predicting breeding values of new animals for which performance data are not yet available.

Multistep GBLUP

When SNP information first became available, the common practice was to proceed in two steps.

In step 1, one calculates breeding values in the conventional way from pedigrees.

In step 2, one solves the genomic regression model.

This is still a much used approach.

Alternatively, one could fit the traits directly to the SNP information, bypassing the breeding values.

ssBLUP (single step BLUP)

ssBLUP is a successful attempt by Misztal et al. 2009
to use both pedigree information and genomic information
simultaneously.

This is done by combining – in an appropriate way suggested by
Bayesian analysis – the pedigree relationship matrix and the
genomic relationship matrix into an effective relationship matrix
that is then used in the mixed model equations.

ssBLUP involves heavy linear algebra, since large matrices must be
(at least approximately) inverted.

Data compression in GBLUP

As a preparation for handling the case of very large amounts of genomic data, we made preliminary investigations of various dimension reduction techniques.

This can be studied in the simpler situation of GBLUP, i.e., without pedigree information.

When the number of SNPs is close to or larger than the number of genotyped animals, the SNP matrix is often extremely ill-conditioned or even rank deficient (always if there are more SNPs than animals).

Hence some form of **regularization** is needed that adds qualitative information which permits a sensible estimate of x. One can use this qualitative information also to reduce the number of genomic breeding values.

Günther Jedenastik 2018 investigated numerically six different regularization methods on real data – poultry and Nelore cattle.

The goal was to find out which one has the best generalization to unseen data in a 2-stage **cross validation** approach that allows one to optimize a regularization parameter on a training set.

• The assumption that a few QTL's (quantitative trait loci) are responsible for the bulk of the breeding value leads to L^1 regularization. This can be efficiently solved with the **LASSO** procedure, but was found to be not very predictive.

• Best results were obtained with **ridge regression** with a ridge parameter optimized on the training set. This suggests that a huge number of SNPs influence the traits, each one by a small amount only.

The model equation approach

- meBLUP (model equation BLUP)
- Phenotype model equations
- Pedigree model equations
- Genomic model equations
- Trivial model equations

meBLUP (model equation BLUP)

Model equation BLUP (meBLUP) is an genomic extension of the method described in our recent paper

E. Groeneveld and A. Neumaier,
BLUP without (inverse) relationship matrix,
Proc. World Congress Genetics Livestock Production,
vol. Theory to Application 3 (2018), 21.
`http://www.wcgalp.org/system/files/proceedings/2018/blup-without-inverse-relationship-matrix.pdf`

for the case of BLUP without genomic information. Its only free parameters are variance components, so they can in principle be estimated by REML procedures. meBLUP is implemented in the PEST/VCE package of the second author.

Phenotype model equations

For those animals for which measurements are available, the measurement vectors $y_\nu \in \mathbb{R}^{n_{\text{trait}}}$ are explained in terms of a linear combination of effects $\beta_i \in \mathbb{R}^{n_{\text{trait}}}$,

$$y_\nu \approx \sum_{l=1}^{n_{\text{eff}}} \mu_{\nu l} \beta_{i_{\nu l}},$$

leading to the model equations

$$\textbf{(PHE)} \qquad \sum_{l=1}^{n_{\text{eff}}} \mu_{\nu l} \beta_{i_{\nu l}} = y_\nu + \text{noise}(C_M). \tag{1}$$

Here the $i_{\nu l}$ form an $n_{\text{rec}} \times n_{\text{eff}}$ index matrix and the $\mu_{\nu l}$ form an $n_{\text{rec}} \times n_{\text{eff}}$ coefficient matrix.

Pedigree model equations

For some animals, identified by the index T of their additive genetic effect β_T, we may know the parents, with corresponding indices S (sire, father) and D (dam, mother). Their genetic dependence is modeled by a relation

$$\beta_{T(\nu)} \approx \frac{1}{2}\beta_{S(\nu)} + \frac{1}{2}\beta_{D(\nu)}$$

and similar formulas when one or both parents are missing. This leads to model equations of the following form:

$$\textbf{(PED)} \qquad \frac{1}{\sqrt{k}}(2\beta_T - \beta_S - \beta_D) = \varepsilon_T = \text{noise}(C_{P\gamma}) \qquad (2)$$

where, for unknown parents, β_S or β_D are zero, and

$$k = 4 - \text{number of known parents.}$$

for animals $T(\nu)$ of class γ, according to some classification. In the simplest case there is only one class. For missing parents, the corresponding β is zero.

Genomic model equations

For some animals, SNP information is assumed to be given by a **marker incidence matrix** K with entries $K_{\nu\ell} \in \{0, 1, 2\}$.

For those animals ν for which SNPs are available, the additive genetic effect $\beta_{T(\nu)} \in \mathbb{R}^{n_{\text{trait}}}$ is explained by

$$\beta_{T(\nu)} \approx \beta_S g_{T(\nu)} + \beta_0$$

in terms of **additive genomic effects** $\beta_S \in \mathbb{R}^{n_{\text{trait}} \times n_S}$ and a **genomic mean effect** $\beta_0 \in \mathbb{R}^{n_{\text{trait}}}$, leading to element equations

$$\textbf{(GEN)} \qquad \beta_S g_{T(\nu)} + \beta_0 - \beta_{T(\nu)} = 0 + \text{noise}(C_S). \qquad (3)$$

Here e is the all-one vector of size n_{trait}, the g_ν are feature vectors of size n_{SNP} containing **marker features**.

Trivial model equations

Certain effects $\beta_{R(\gamma)}$, namely all random environmental effects and all additive genomic effects β_S, are limited in size by including trivial model equations

$$\textbf{(TRI)} \qquad \beta_{R(\gamma)} = 0 + \text{noise}(C_\gamma). \tag{4}$$

This has a regularizing effect as it reduces the condition number of the normal equations.

Equivalent mixed models

- The abstract relationship matrix
- Model equations in matrix form
- Special cases

The abstract relationship matrix

To see what goes on, we reduce a number of instances of the general ME model to the mixed model form

$$y = X\beta + Zu + \text{noise}(D), \quad u = \text{noise}(G), \tag{5}$$

so that we can see how the resulting G looks like.

We call G the **abstract relationship matrix** since in simple cases it is proportional to the numerator relationship matrix.

In terms of a model equation interpretation, the mixed model equations contain only phenotype equations and trivial equations.

This implies that all other information must be packed into the abstract relationship matrix, and is the reason why G becomes complicated and dense.

G would figure in Henderson's mixed model equations if we would solve the ME problem by the standard method.

Instead we solve the ME problem directly via the normal equations, without any modeling of a relationship matrix or its inverse.

Model equations in matrix form

To relate our ME approach to the mixed model we rewrite the model equations in matrix form.

For definiteness, we assume the presence of (some or all of)

- fixed effects
- random environmental effects
- additive genomic effects
- a mean additive genomic effect
- trivial genomic and random environmental effects

We write the phenotype equations as

$$X\beta + Z_e\beta_e + Zu = y + \text{noise}(D), \quad D = C_M, \tag{6}$$

where β contains the fixed effects, β_e contains the random environmental effects, and u the additive genetic effects in the phenotype equations.

The pedigree equations are written as

$$Pu = \text{noise}(C_P), \tag{7}$$

where P is a nonsingular triangular matrix.

The genomic equations are written as

$$e\beta_0 + Sv - Ju = \text{noise}(C_S), \tag{8}$$

where β_0 is the mean genomic effect, v contains all additive genomic effects, e is an all-one vector, and J projects to the part for which genomic information is available.

Finally, the trivial element equations are written as

$$\beta_e = \text{noise}(C_E) \tag{9}$$

for the random environmental effects and

$$v = \text{noise}(C_A) \tag{10}$$

for the additive genomic effects.

This leads to the linear stochastic model $Ax = y + \text{noise}(C)$ with

$$A = \begin{pmatrix} X & Z & 0 \\ 0 & P & 0 \\ 0 & -J & S \\ 0 & 0 & I \end{pmatrix}, \quad x = \begin{pmatrix} \beta \\ u \\ v \end{pmatrix}, \quad b = \begin{pmatrix} y \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

and $C = \text{Diag}(D, C_P, C_S, C_A)$.

This model is easy to assemble and to solve by conjugate gradients, even for very large data.

This is the way meBLUP is implemented in PEST and VCE.

In the ME formulation, the resulting normal equation matrix equals that of the matrix of Henderson's MME for the equivalent mixed model formulation when there are no genotypes, and only then.

Special cases

To see the emaning of the equivalent mixed model form
we discuss several special cases:

1. No SNP genotypes, no random environmental effect
2. No SNP genotypes but random environmental effects
3. All animals genotyped, no genomic mean effect,
no pedigree, no random environmental effects
4. All animals genotyped, with genomic mean effect,
no pedigree, no random environmental effects
5. Full pedigree, only some animals genotyped,
no genomic mean effect, no random environmental effects

In addition, we discuss some special or limiting cases
of these five cases.

CASE 1: *No SNP genotypes, no random environmental effect.*

In this case, the phenotype equations are

$$X\beta + Zu = y + \text{noise}(D), \quad D = C_M,$$

where β contains the fixed effects and u the additive genetic effects in the measurement equations.

The pedigree equations $Pu = \text{noise}(C_P)$ must be solved for u. This gives $u = \text{noise}(P^{-1}C_P P^{-T})$, so that $G = P^{-1}C_P P^{-T}$, and the inverse abstract relationship matrix is

$$G^{-1} = H^{\text{ped}} := P^T C_P^{-1} P.$$

This is precisely the sparse factored form of the inverse relationship matrix discovered by HENDERSON 1976. We can infer that in this case, C_P is a diagonal matrix whose diagonal entries are determined by the inbreeding coefficients.

CASE 2: *No SNP genotypes but random environmental effects* β_e.

Now the phenotype equations have the form

$$X\beta + Zu + Z_e\beta_e = y + \text{noise}(D), \quad D = C_M,$$

and we have additional trivial equations $\beta_e = \text{noise}(C_E)$.

Thus the vector of random effects figuring as u in the standard mixed model form is in fact the larger vector

$$\begin{pmatrix} \beta_e \\ u \end{pmatrix} = \text{noise}\left(\begin{pmatrix} C_E & 0 \\ 0 & P^{-1}C_P P^{-T} \end{pmatrix} \right),$$

so that

$$G^{-1} = \begin{pmatrix} C_E & 0 \\ 0 & P^T C_P^{-1} P \end{pmatrix} = \begin{pmatrix} C_E & 0 \\ 0 & H^{\text{ped}} \end{pmatrix}.$$

CASE 3: *All animals genotyped, no genomic mean effect, no pedigree, no random environmental effects.*

Now we have genomic equations of the form

$$Sv - u = \text{noise}(C_S)$$

and trivial equations for the additive genomic effect vector, $v = \text{noise}(C_A)$. Hence

$$u = Sv + \text{noise}(C_S) = \text{noise}(F)$$

where

$$F := C_S + SC_AS^T.$$

Thus $G = F$ and

$$G^{-1} = H^{\text{gen}} := (C_S + SC_AS^T)^{-1}.$$

This defines the **genomic inverse relationship matrix** H^{gen} in the ME approach.

CASE 3A: *Limiting case of Case 3 where* $C_S = 0$.

Now $u = Sv$ is an exact relation, and we obtain $G = F = SC_AS^T$.

This has the form of the genomic relationship matrix appearing in the mixed model of MEUWISSEN et al. 2001 and identified as such by VANRADEN 2008.

CASE 4: *All animals genotyped, with genomic mean effect, but no pedigree, no random environmental effects.*

Now the genomic equations are

$$Sv + e\beta_0 - u = \text{noise}(C_S),$$

where e is an all-one vector. If we treat the genomic mean effect also as random, $\beta_0 = \text{noise}(\sigma^2)$, we find

$$u = \beta_0 e + Sv + \text{noise}(C_S) = \text{noise}(\sigma^2 ee^T + C_S + SC_A S^T) = \text{noise}(\sigma^2 ee^T + F),$$

Therefore $G = \sigma^2 ee^T + F$. Since $F^{-1} = H^{\text{gen}}$, we conclude from the matrix inversion formula of Sherman and Morrison that

$$G^{-1} = H^{\text{gen}} - \frac{H^{\text{gen}} ee^T H^{\text{gen}}}{\sigma^{-2} + e^T H^{\text{gen}} e}.$$

CASE 4A: *Special case of Case 4 where all animals have phenotypes.*

Now $Z = I$, and the model equations reduce to the mixed model

$$X\beta + Sv + e\beta_0 = y + \text{noise}(D + C_S)$$

used (after absorbing C_S into D) by MEUWISSEN 2001.

CASE 4B: β_0 becomes a fixed effect in the limit $\sigma^2 \to \infty$.

In this limit

$$G^{-1} = H^{\text{gen}} - \frac{H^{\text{gen}} e e^T H^{\text{gen}}}{e^T H^{\text{gen}} e}$$

satisfies $G^{-1}e = 0$, so that we have a singular inverse covariance matrix.

(Note that a true inverse matrix cannot be singular.)

CASE 5: *Full pedigree, only some animals genotyped, but no with genomic mean effect and no random environmental effects.*

Now the model equations are

$$X\beta + Zu = y + \text{noise}(D), \quad Pu = \text{noise}(C_P),$$

$$Sv - Ju = \text{noise}(C_S), \quad v = \text{noise}(C_A).$$

Proceeding as before, we find that $Ju = \text{noise}(C_S + SC_TS^T)$.

This is additional covariance information generally inconsistent with that in the pedigree equations (which already fully determines the covariance matrix of u), and leads to the overdetermined specification

$$\begin{pmatrix} P \\ J \end{pmatrix} u = \text{noise}\left(\begin{pmatrix} C_P & 0 \\ 0 & C_S + SC_TS^T \end{pmatrix} \right).$$

In meBLUP, this overspecification is automatically resolved. In analogy to Case 1, this leads after some calculation to

$$G^{-1} = \begin{pmatrix} P \\ J \end{pmatrix}^T \begin{pmatrix} C_P & 0 \\ 0 & C_S + SC_TS^T \end{pmatrix} \begin{pmatrix} P \\ J \end{pmatrix} = H^{\text{ped}} + J^T H^{\text{gen}} J.$$

If H^{ped} is partitioned into blocks corresponding to ungenotyped and genotyped animals then $J = (0 \;\; I)$, hence

$$G^{-1} = H^{\text{ped}} + J^T H^{\text{gen}} J = \begin{pmatrix} H^{\text{ped}}_{11} & H^{\text{ped}}_{12} \\ H^{\text{ped}}_{21} & H^{\text{ped}}_{22} + H^{\text{gen}} \end{pmatrix}.$$

Effectively, the breeding value information from the pedigree is treated as Bayesian prior information for the estimation of the breeding value from the genomic information.

This is different from the way ssBLUP reconciles the inconsistency between the pedigree covariance and the genomic covariance:

ssBLUP subtracts from the meBLUP matrix just derived some (expensive to compute) term in the bottom right corner.

This subtraction produces in ssBLUP a more ill-conditioned or even singular G^{-1} than in meBLUP.

That a singular G^{-1} is possible is a mathematical defect of ssBLUP that shows that something is problematic in the derivation of the ssBLUP.

Indeed, the working implementations of ssBLUP never use the theoretically formulas but heuristic modifications involving additional scaling and averaging steps justified not by theory but only by their need to make the method work in practice.

Numerical results and conclusions

- Preliminary numerical tests
- Correlations true/estimated breeding value
- Summary: Features of meBLUP

Preliminary numerical tests

We used **QMSim** by Sargolzaei & Schenkel 2009 to simulate

- 800 QTL on 30 chromosomes for 30000 or 60000 markers
- phenotypes with different heritabilities

$$h^2 = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$$

- 11 generations (840+10*1600) with a total of 16840 animals;

of these

- the last 3 generations (4800 animals) were genotyped
- the last generation (1600 animals) had no phenotype.

We compared meBLUP with suitable variances [x=13616] with the ssBLUP implementation in Misztal's blupf90.

Correlations true/estimated breeding value

for the last generation (without phenotypes)

h^2	number of SNPs used			method
	40000	50000	60000	
0.05	0.435+	0.446+	0.450-	meBLUP
0.05	0.426	0.445	0.455	ssBLUP
0.10	0.425+	0.429-	0.446-	meBLUP
0.10	0.418	0.436	0.459	ssBLUP
0.20	0.591+	0.593+	0.599+	meBLUP
0.20	0.581	0.586	0.597	ssBLUP
0.30	0.528-	0.537-	0.555-	meBLUP
0.30	0.530	0.551	0.584	ssBLUP
0.40	0.572-	0.591-	0.598-	meBLUP
0.40	0.577	0.615	0.635	ssBLUP
0.50	0.617+	0.640-	0.650-	meBLUP
0.50	0.616	0.655	0.679	ssBLUP
0.60	0.717+	0.735-	0.745-	meBLUP
0.60	0.704	0.745	0.767	ssBLUP

Summary: Features of meBLUP – I

Instead of a formulation as a mixed model based on some form of (inverse) relationship matrix meBLUP uses a linear model expressing all assumed numerical relations directly in the form of model equations.

• The ME approach is fully automated and very simple, both in theory and practice.

• No inverse or partial inverse is needed, no factorization, only trivial linear algebra. Hence the method is very easy to parallelize.

• Neither the relationship matrix nor the mixed model equations are formed.

• There are no problems with singularity or poor conditioning.

• The resulting linear system is always positive definite and can be solved by conjugate gradients, without slow convergence.

Summary: Features of meBLUP – II

• Other information is easy to integrate.

• General software is available for everything that is in PEST; in particular, for incomplete records, multiple traits, genetic groups, arbitrary fixed effects and random environmental effects.

• In principle, arbitrary expected, conjectured, or desired approximate linear relations and residual sizes can be imposed.

• The ME approach leads to a different way of combining pedigree data and genomic data. It is not equivalent to ssBLUP.

• First numerical results indicate that, for an appropriate choice of parameters, meBLUP has a similar quality as ssBLUP.

• Finding the best parameters needs further study. All free parameters are variance components and hence can in principle be estimated by REML.

Thank you for your attention!

Some references can be found on the following pages.

A copy of the slides is be available at

`http://www.mat.univie.ac.at/~neum/ms/meBLUPSlides.pdf`

References

W.H. Fellner, Robust estimation of variance components, Technometrics 28 (1986), 51–60.

E. Groeneveld and A. Neumaier,
BLUP without (inverse) relationship matrix,
Proc. World Congress Genetics Livestock Production,
vol. Theory to Application 3 (2018), 21.
`http://www.wcgalp.org/proceedings/2018/blup-without-inverse-relationship-matrix`

C.R. Henderson, Estimation of changes in herd environment, J. Dairy Sci, 32 (1949), 706–715.

G. Jedenastik, Determine breeding value of animal livestock using genetic data, Master's thesis, Dept. of Mathematics, University of Vienna, 2018.

T.H.E. Meuwissen, B.J. Hayes and M.E. Goddard, Prediction of total genetic value using genome-wide dense marker maps, Genetics 157 (2001), 1819–1829.

Y. Miar, G. Plastow and Z. Wang, Genomic selection, a new era for pork quality Improvement, Springer Science Reviews 3 (2015), 27–37.

I. Misztal, A. Legarra and I. Aguilar, Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information, J. Dairy Sci. 92 (2009), 4648–4655.

I. Misztal, S. Tsuruta, D. Lourenco, I. Aguilar, A. Legarra and Z. Vitezica, Manual for BLUPF90 family of programs. Athens: University of Georgia, 2014.

A. Neumaier and E. Groeneveld, Restricted maximum likelihood estimation of covariances in sparse linear models, Genetics Selection Evolution 30 (1998), 3–26.
`http://www.mat.univie.ac.at/~neum/ms/reml.pdf`

M. Sargolzaei and F.S. Schenkel, QMSim: a large-scale genome simulator for livestock, Bioinformatics 25 (2009), 680–681.

P.M. VanRaden, Efficient methods to compute genomic predictions, J. Dairy Science 91 (2008), 4414–4423.

R.A. Westell, R.L. Quaas and L.D. Van Vleck, Genetic groups in an animal model, J. Animal Sci. 71 (1987), 1310–1318.