Reduced Animal Models Fitting Only Equations for Phenotyped Animals

Mohammad Ali Nilforooshan[1]* and Dorian Garrick[2]

[1] *Livestock Improvement Corporation, Hamilton, New Zealand,* [2] *AL Rae Centre for Genetics and Breeding, School of Agriculture, Massey University, Palmerston North, New Zealand*

Reduced models are equivalent models to the full model that enable reduction in the computational demand for solving the problem, here, mixed model equations for estimating breeding values of selection candidates. Since phenotyped animals provide data to the model, the aim of this study was to reduce animal models to those equations corresponding to phenotyped animals. Non-phenotyped ancestral animals have normally been included in analyses as they facilitate formation of the inverse numerator relationship matrix. However, a reduced model can exclude those animals and obtain identical solutions for the breeding values of the animals of interest. Solutions corresponding to non-phenotyped animals can be back-solved from the solutions of phenotyped animals and specific blocks of the inverted relationship matrix. This idea was extended to other forms of animal model and the results from each reduced model (and back-solving) were identical to the results from the corresponding full model. Previous studies have been mainly focused on reduced animal models that absorb equations corresponding to non-parents and solve equations only for parents of phenotyped animals. These two types of reduced animal model can be combined to formulate only equations corresponding to phenotyped parents of phenotyped progeny.

Keywords: animal model, phenotyped, reduced model, back-solve, relationship matrix

1. INTRODUCTION

Computational limitations have been a major challenge facing animal breeders, especially for solving large mixed model equations, such as those including multiple traits and millions of animals. Computational speed, power and technology have advanced, making most of the computations that were impossible in the past, feasible. However, the amount of data is growing rapidly, and the complexity of evaluation models are increasing. Henderson (1974) introduced the concepts of equivalent models and reduced models to animal breeding. Any linear model can be written in various forms of equivalent models yielding the same first and second moments of the data (Henderson, 1985). A reduced model is an equivalent model to the full model, allowing computational simplifications by reduction in the number of equations to be solved.

Limited by available memory to store the entire set of equations, Henderson (1974) proposed an equivalent reduced model in which the equations for fixed herd effects were absorbed into the equations for sire effects. The full model explicitly fitted herd and sire effects. Generally, absorption is a model rank reduction (reduced number of equations to be solved), by which some effects get absorbed into other effects with no influence on the solutions of the remaining effects in the model.

Reduced models that absorbed multiple-trait equations for random effects were first introduced to animal models by Quaas and Pollak (1980), where they showed that the best linear unbiased

predictions (BLUP) of parents (with phenotyped progeny) can be obtained from a reduced model that only includes equations for breeding values for parents. This was achieved by absorbing equations of non-parent individuals (and parents with no phenotyped progeny). The BLUP of breeding values for non-parent individuals can then be back-solved given the BLUP of breeding values for their parents. The concept of reduced models was extended to animal models incorporating quantitative trait loci, introduced by Fernando and Grossman (1989), and was also based on absorbing equations corresponding to non-parent individuals (Cantet and Smith, 1991; Hoeschele, 1993; Arendonk et al., 1994; Saito and Iwaisaki, 1996, 1997).

Henderson (1986) proposed the estimation of variance components in reduced (based on the methodology of Quaas and Pollak, 1980) rather than full animal models for single traits and single records, based on minimum variance quadratic unbiased estimation (MIVQUE), restricted maximum likelihood (REML) by iterated MIVQUE, and REML by the expectation-maximization (EM, Dempster et al., 1977) algorithm. Providing numerical examples, he showed that the reduced model can be used for estimation of variances, converging to the same solutions obtained by the full model.

This study revisited the reduced animal model from a different perspective by only including equations corresponding to phenotyped animals in the model. In a full model, non-phenotyped parents do not contribute any phenotypic information to the model, but make it easier to account for similarity between animals with phenotypes when they are included in the construction of the inverse numerator relationship matrix. This form of reduced animal model that absorbs the effects of non-phenotyped breeding values is first explained in the context of a simple univariate animal model, and then extended to several other forms of animal model. Numerical examples and R scripts are provided to facilitate understanding.

2. METHODS

Animal models can be expressed and fitted in various forms and combinations. In this section, some equivalent forms of animal model, their size reduction (absorption) to phenotyped animals (denoted as p), and approach to back-solve solutions corresponding to non-phenotyped animals (denoted as n) are presented.

2.1. Basic Animal Model

In its simplest commonly-used form ($\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$), an animal model contains the vectors of phenotypes ($\mathbf{y}$), fixed effects ($\mathbf{b}$), random direct additive genetic effects [$\mathbf{u} \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$, representing breeding values] and random residuals ($\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$), with σ_u^2 and σ_e^2 being the corresponding known variances, and $\mathbf{A}$ being the numerator relationship matrix. Matrices $\mathbf{X}$ and $\mathbf{Z}$ relate phenotypes to fixed effects and animals, respectively. In matrix notation solutions to this single-trait model are obtained by solving the equation system:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \lambda\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}, \qquad (1)$$

where $\hat{\mathbf{b}}$ is a vector of solutions for fixed effects ($\mathbf{b}$) and $\hat{\mathbf{u}}$ is the vector of best linear unbiased predictions for $\mathbf{u}$ (known as estimated breeding values) and $\lambda = \sigma_e^2/\sigma_u^2$. Permuting rows and columns of the sparse matrix $\mathbf{A}^{-1}$ so that non-phenotyped animals appear before phenotyped animals, $\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}^{nn} & \mathbf{A}^{np} \\ \mathbf{A}^{pn} & \mathbf{A}^{pp} \end{bmatrix}$, $\hat{\mathbf{u}}' = \begin{bmatrix} \hat{\mathbf{u}}'_n & \hat{\mathbf{u}}'_p \end{bmatrix}$ and $\mathbf{Z} = \begin{bmatrix} \mathbf{0} & \mathbf{Z}_p \end{bmatrix}$, where $\mathbf{0}$ is a matrix with elements equal to zero. Here, $\mathbf{0}$ has number of rows equal to the number of phenotypes and number of columns equal to the number of non-phenotyped animals. Then, Equation (1) can be written as:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{0} & \mathbf{X'Z}_p \\ \mathbf{0} & \lambda\mathbf{A}^{nn} & \lambda\mathbf{A}^{np} \\ \mathbf{Z}'_p\mathbf{X} & \lambda\mathbf{A}^{pn} & \mathbf{Z}'_p\mathbf{Z}_p + \lambda\mathbf{A}^{pp} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_n \\ \hat{\mathbf{u}}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{0} \\ \mathbf{Z}'_p\mathbf{y} \end{bmatrix}. \qquad (2)$$

Permuting $\mathbf{y}$ so that animals are in the same order as $\mathbf{A}^{pp}$, $\mathbf{Z}_p = \mathbf{I}$. Then,

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{0} & \mathbf{X'} \\ \mathbf{0} & \lambda\mathbf{A}^{nn} & \lambda\mathbf{A}^{np} \\ \mathbf{X} & \lambda\mathbf{A}^{pn} & \mathbf{I} + \lambda\mathbf{A}^{pp} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_n \\ \hat{\mathbf{u}}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{0} \\ \mathbf{y} \end{bmatrix}. \qquad (3)$$

If we absorb the equations corresponding to breeding values for non-phenotyped animals by modifying the equations corresponding to breeding values for phenotyped animals, Equation (3) is transformed to Equation (4), which contains breeding values only for phenotyped animals:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'} \\ \mathbf{X} & \mathbf{I} + \lambda\Phi \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{y} \end{bmatrix}, \qquad (4)$$

where

$$\Phi = \mathbf{A}^{pp} - \mathbf{A}^{pn}(\mathbf{A}^{nn})^{-1}\mathbf{A}^{np}. \qquad (5)$$

Derivation of Φ is provided later in this subsection. Considering equations corresponding to breeding values for non-phenotyped animals in Equation (3), we have the identity:

$$\mathbf{A}^{nn}\hat{\mathbf{u}}_n = -\mathbf{A}^{np}\hat{\mathbf{u}}_p. \qquad (6)$$

Thus, $\hat{\mathbf{u}}_n$ can be back-solved from $\mathbf{A}^{nn}$, $\mathbf{A}^{np}$, and $\hat{\mathbf{u}}_p$. Comparing Equations (1) and (4), $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$ is reduced to $\mathbf{y} = \mathbf{Xb} + \mathbf{u}_p + \mathbf{e}$. Thus, phenotypes depend only on the genetic merit of phenotyped animals and are independent from the genetic merit of their non-phenotyped relatives. The reduced model has a different structure of additive genetic variance ($\mathbf{u}_p \sim N(\mathbf{0}, \Phi^{-1}\sigma_u^2)$ vs. $\mathbf{u} \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$).

Consider the typically dense matrix $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{nn} & \mathbf{A}_{np} \\ \mathbf{A}_{pn} & \mathbf{A}_{pp} \end{bmatrix}$. The following mixed model equations are equivalent to Equation (1),

but can only be used to directly predict $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}_p$, and it involves no changes in model assumptions (Henderson, 1986):

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' \\ \mathbf{X} & \mathbf{I} + \lambda\mathbf{A}_{pp}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{y} \end{bmatrix}. \qquad (7)$$

The above representation of the model is useful to prove the derivation of Φ, but not computationally efficient, because $\mathbf{A}_{pp}$ is generally not sparse and it cannot be inverted using the indirect inversion methods of Henderson (1976) or Quaas (1976). Equating row 3 of Equation (3) $(\mathbf{X}\hat{\mathbf{b}} + \lambda\mathbf{A}^{pn}\hat{\mathbf{u}}_n + (\mathbf{I} + \lambda\mathbf{A}^{pp})\hat{\mathbf{u}}_p = \mathbf{y})$ and row 2 of Equation (7) $(\mathbf{X}\hat{\mathbf{b}} + (\mathbf{I} + \lambda\mathbf{A}_{pp}^{-1})\hat{\mathbf{u}}_p = \mathbf{y})$ results in $\lambda\mathbf{A}^{pn}\hat{\mathbf{u}}_n + (\mathbf{I} + \lambda\mathbf{A}^{pp})\hat{\mathbf{u}}_p = (\mathbf{I} + \lambda\mathbf{A}_{pp}^{-1})\hat{\mathbf{u}}_p$, which is rearranged to $\mathbf{A}^{pn}\hat{\mathbf{u}}_n = (\mathbf{A}_{pp}^{-1} - \mathbf{A}^{pp})\hat{\mathbf{u}}_p$. Solving the obtained equation with Equation (6):

$$\begin{cases} \mathbf{A}^{pn}\hat{\mathbf{u}}_n = (\mathbf{A}_{pp}^{-1} - \mathbf{A}^{pp})\hat{\mathbf{u}}_p \\ \mathbf{A}^{nn}\hat{\mathbf{u}}_n = -\mathbf{A}^{np}\hat{\mathbf{u}}_p \quad \text{(Equation 6)} \end{cases}$$

$$\Rightarrow \begin{cases} \mathbf{A}^{pn}\hat{\mathbf{u}}_n = (\mathbf{A}_{pp}^{-1} - \mathbf{A}^{pp})\hat{\mathbf{u}}_p \\ \hat{\mathbf{u}}_n = -(\mathbf{A}^{nn})^{-1}\mathbf{A}^{np}\hat{\mathbf{u}}_p \end{cases}$$

$$\Rightarrow -\mathbf{A}^{pn}(\mathbf{A}^{nn})^{-1}\mathbf{A}^{np}\hat{\mathbf{u}}_p = (\mathbf{A}_{pp}^{-1} - \mathbf{A}^{pp})\hat{\mathbf{u}}_p$$

$$\Rightarrow \mathbf{A}^{pn}(\mathbf{A}^{nn})^{-1}\mathbf{A}^{np} = \mathbf{A}^{pp} - \mathbf{A}_{pp}^{-1}$$

$$\Rightarrow \mathbf{A}^{pp} - \mathbf{A}^{pn}(\mathbf{A}^{nn})^{-1}\mathbf{A}^{np} = \mathbf{A}_{pp}^{-1}$$

$$\Rightarrow \Phi = \mathbf{A}_{pp}^{-1}$$

Thus, forming and solving the mixed model equations using Φ will produce the same results as forming and solving the mixed model equations using $\mathbf{A}_{pp}^{-1}$, but without the need to form $\mathbf{A}$, invert or store $\mathbf{A}_{pp}$.

We carried out a test to compare the computational time between the full and the reduced animal model. Pedigree and phenotypes of 10,000 animals were simulated using R package pedSimulate (Nilforooshan, 2021). The pedigree was initiated with 100 founder animals (generation 0) of both sexes (50 each) and continued to eight generations. Except 2,000 randomly selected phenotypes from generations 6 to 8, other phenotypes were set to missing. Phenotypes were randomly assigned to five herds and the herd effect was added to phenotypes. The R statistical software (R Core Team, 2019) was used to analyse the data.

2.2. Multiple Trait Animal Model

The mixed model equations for a multiple trait animal model in its simplest form with no additional random effects other than for breeding values can be represented as:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}, \qquad (8)$$

where $\mathbf{G}$ is the $t \times t$ matrix of genetic covariances between the traits, $\mathbf{R}$ is the matrix of residual covariances between all the residuals on all the phenotyped animals, and $\otimes$ is the Kronecker product (Horn and Johnson, 1991). Permuting rows and columns

of $\mathbf{A}^{-1}$ so that animals that are non-phenotyped (for all the traits) appear before animals that are phenotyped for at least one of the traits, $\mathbf{Z} = \begin{bmatrix} \mathbf{0} & \mathbf{Z}_p \end{bmatrix}$, $\mathbf{Z}_p$ corresponds to animals phenotyped for at least one of the traits, and:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{0} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{0} & \mathbf{A}^{nn} \otimes \mathbf{G}^{-1} & \mathbf{A}^{np} \otimes \mathbf{G}^{-1} \\ \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{X} & \mathbf{A}^{pn} \otimes \mathbf{G}^{-1} & \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{Z}_p + \mathbf{A}^{pp} \otimes \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_n \\ \hat{\mathbf{u}}_p \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{0} \\ \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}. \qquad (9)$$

In the case of a 2-trait model, $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix}$ and $\mathbf{Z}_p = \begin{bmatrix} \mathbf{Z}_{p_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{p_2} \end{bmatrix}$. $\mathbf{Z}_{p_1}$ and $\mathbf{Z}_{p_2}$ contain columns for the same set of animals (those phenotyped for either or both of the two traits).

Following the same procedure as for the single-trait animal model, the model is reduced to fitting phenotyped animals for at least one of the traits in the model:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{Z}_p + \Phi \otimes \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}. \qquad (10)$$

Considering Equation (9), solutions for non-phenotyped animals ($\hat{\mathbf{u}}_n$) can be back-solved as:

$$(\mathbf{A}^{nn} \otimes \mathbf{G}^{-1})\hat{\mathbf{u}}_n = -(\mathbf{A}^{np} \otimes \mathbf{G}^{-1})\hat{\mathbf{u}}_p. \qquad (11)$$

2.3. Repeatability Animal Model

Repeatability animal models are employed for the analysis of repeated phenotypes of the same trait on an animal. In the simplest repeatability animal models, a genetic correlation of 1 and equal residual correlations are considered between repeated records from an animal, and the same variances are applied to all the phenotypes (Mrode, 2005). Repeatability animal models are typically fitted including an additional random term for permanent environmental effects ($\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wp} + \mathbf{e}$). In matrix notation:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{A}^{-1} & \mathbf{Z}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{Z} & \mathbf{W}'\mathbf{W} + \alpha\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}, \qquad (12)$$

where $\hat{\mathbf{p}}$ is the vector of solutions for random permanent environmental effects, $\mathbf{W}$ is the incidence matrix relating phenotypes to permanent environmental effects), $\alpha = \sigma_e^2/\sigma_p^2$, and σ_p^2 is the permanent environment variance. Considering $\mathbf{W}$ being already limited to phenotyped animals, $\mathbf{W} = \mathbf{Z}_p$, and:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{0} & \mathbf{X'Z}_p & \mathbf{X'W} \\ \mathbf{0} & \lambda\mathbf{A}^{nn} & \lambda\mathbf{A}^{np} & \mathbf{0} \\ \mathbf{Z}_p'\mathbf{X} & \lambda\mathbf{A}^{pn} & \mathbf{Z}_p'\mathbf{Z}_p + \lambda\mathbf{A}^{pp} & \mathbf{Z}_p'\mathbf{W} \\ \mathbf{W'X} & \mathbf{0} & \mathbf{W'Z}_p & \mathbf{W'W} + \alpha\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_n \\ \hat{\mathbf{u}}_p \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{0} \\ \mathbf{Z}_p'\mathbf{y} \\ \mathbf{W'y} \end{bmatrix}. \tag{13}$$

Equation (13) can be reduced to phenotyped animals (Equation 14) by absorbing the equations corresponding to breeding values for non-phenotyped animals into the equations corresponding to breeding values for phenotyped animals:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z}_p & \mathbf{X'W} \\ \mathbf{Z}_p'\mathbf{X} & \mathbf{Z}_p'\mathbf{Z}_p + \lambda\Phi & \mathbf{Z}_p'\mathbf{W} \\ \mathbf{W'X} & \mathbf{W'Z}_p & \mathbf{W'W} + \alpha\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_p \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z}_p'\mathbf{y} \\ \mathbf{W'y} \end{bmatrix}. \tag{14}$$

Back-solving solutions for non-phenotyped animals is done according to Equation (6). If permanent environmental effect solutions are not of interest, Equation (14) can be further reduced by absorbing the equations for random permanent environmental effects. Any random effect not correlated to other random effects remaining in the model can be confounded with the residual term (Quaas and Pollak, 1980).

2.4. Maternal Trait Animal Models

For some traits, dams not only influence the phenotypic expression of their progeny via directly passed genes, but also through maternal ability that influences the environment for their progeny, such as the milk volume they provide to the offspring (Mrode, 2005). Maternal ability is partly genetic and partly environmental. Accordingly, there can be animal models with maternal genetic effects, maternal permanent environment effects, or both. Animal models with maternal permanent environment effects have similar structure to repeatability animal models with the differences that there might be no repeated records per animal, $\mathbf{W}$ is the incidence matrix relating phenotypes to the dams of phenotyped animals, and a different variance than σ_p^2 is involved (σ_{mpe}^2, maternal permanent environment variance). Maternal permanent environmental effect is already limited to phenotyped animals as the incidence matrix used for this effect has rows corresponding to phenotypes and columns corresponding to dams of phenotyped animals.

Animal models with maternal genetic effects ($\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Sm} + \mathbf{e}$) uncorrelated to direct additive genetic effects have the variance structure:

$$var\begin{bmatrix} \mathbf{u} \\ \mathbf{m} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_u^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}\sigma_m^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2 \end{bmatrix},$$

where σ_m^2 is the maternal genetic variance. In matrix notation:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'S} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \lambda_1\mathbf{A}^{-1} & \mathbf{Z'S} \\ \mathbf{S'X} & \mathbf{S'Z} & \mathbf{S'S} + \lambda_2\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{S'y} \end{bmatrix}. \tag{15}$$

Given $\mathbf{Z} = [\mathbf{0}\ \mathbf{Z}_p]$, $\mathbf{S} = [\mathbf{S}_n\ \mathbf{S}_p]$ and $\mathbf{S'S} = \begin{bmatrix} \mathbf{S}_n'\mathbf{S}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_p'\mathbf{S}_p \end{bmatrix}$, Equation (15) can be re-written as:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{0} & \mathbf{X'Z}_p & \mathbf{X'S}_n & \mathbf{X'S}_p \\ \mathbf{0} & \lambda_1\mathbf{A}^{nn} & \lambda_1\mathbf{A}^{np} & \mathbf{0} & \mathbf{0} \\ \mathbf{Z}_p'\mathbf{X} & \lambda_1\mathbf{A}^{pn} & \mathbf{Z}_p'\mathbf{Z}_p + \lambda_1\mathbf{A}^{pp} & \mathbf{Z}_p'\mathbf{S}_n & \mathbf{Z}_p'\mathbf{S}_p \\ \mathbf{S}_n'\mathbf{X} & \mathbf{0} & \mathbf{S}_n'\mathbf{Z}_p & \mathbf{S}_n'\mathbf{S}_n + \lambda_2\mathbf{A}^{nn} & \lambda_2\mathbf{A}^{np} \\ \mathbf{S}_p'\mathbf{X} & \mathbf{0} & \mathbf{S}_p'\mathbf{Z}_p & \lambda_2\mathbf{A}^{pn} & \mathbf{S}_p'\mathbf{S}_p + \lambda_2\mathbf{A}^{pp} \end{bmatrix}$$
$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_n \\ \hat{\mathbf{u}}_p \\ \hat{\mathbf{m}}_n \\ \hat{\mathbf{m}}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{0} \\ \mathbf{Z}_p'\mathbf{y} \\ \mathbf{0} \\ \mathbf{S}_p'\mathbf{y} \end{bmatrix}, \tag{16}$$

where λ_1 and λ_2 are equal to σ_e^2/σ_u^2 and σ_e^2/σ_m^2, respectively. Considering equations corresponding to $\hat{\mathbf{u}}_n$ in Equation (16), Equation (6) is derived. Equation (16) can then be reduced to:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z}_p & \mathbf{X'S} \\ \mathbf{Z}_p'\mathbf{X} & \mathbf{Z}_p'\mathbf{Z}_p + \lambda_1\Phi & \mathbf{Z}_p'\mathbf{S} \\ \mathbf{S'X} & \mathbf{S'Z}_p & \mathbf{S'S} + \lambda_2\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_p \\ \hat{\mathbf{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z}_p'\mathbf{y} \\ \mathbf{S'y} \end{bmatrix}. \tag{17}$$

Equations corresponding to $\hat{\mathbf{m}}$ cannot be reduced to equations corresponding to $\hat{\mathbf{m}}_p$. The reason is the non-zero $\mathbf{S}_n$ matrix.

Animal models with maternal genetic effects correlated to direct additive genetic effects have the variance structure:

$$var\begin{bmatrix} \mathbf{u} \\ \mathbf{m} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_u^2 & \mathbf{A}\sigma_{um} & \mathbf{0} \\ \mathbf{A}\sigma_{mu} & \mathbf{A}\sigma_m^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2 \end{bmatrix}.$$

In matrix notation:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'S} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \lambda_{11}\mathbf{A}^{-1} & \mathbf{Z'S} + \lambda_{12}\mathbf{A}^{-1} \\ \mathbf{S'X} & \mathbf{S'Z} + \lambda_{21}\mathbf{A}^{-1} & \mathbf{S'S} + \lambda_{22}\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{S'y} \end{bmatrix}, \tag{18}$$

where $\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} = \sigma_e^2\mathbf{G}^{-1}$, and $\mathbf{G}^{-1} = \begin{bmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{bmatrix} = \begin{bmatrix} \sigma_u^2 & \sigma_{um} \\ \sigma_{mu} & \sigma_m^2 \end{bmatrix}^{-1}$.

For an efficient model reduction, it is important that all $\mathbf{A}^{-1}$ occurrences in Equation (18) are converted to the same matrix. In order to do that, the definition of n and p are changed to p being phenotyped animals or non-phenotyped dams, and n being other animals. That enables both $\mathbf{Z}_n$ and $\mathbf{S}_n$ being zero matrices. Then, permuting rows and columns of $\mathbf{A}^{-1}$ so that animals n appear before animals p, Equation (18) can be written as:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{0} & \mathbf{X'Z}_p & \mathbf{0} & \mathbf{X'S}_p \\ \mathbf{0} & \lambda_{11}\mathbf{A}^{nn} & \lambda_{11}\mathbf{A}^{np} & \lambda_{12}\mathbf{A}^{nn} & \lambda_{12}\mathbf{A}^{np} \\ \mathbf{Z}_p'\mathbf{X} & \lambda_{11}\mathbf{A}^{pn} & \mathbf{Z}_p'\mathbf{Z}_p + \lambda_{11}\mathbf{A}^{pp} & \lambda_{12}\mathbf{A}^{pn} & \mathbf{Z}_p'\mathbf{S}_p + \lambda_{12}\mathbf{A}^{pp} \\ \mathbf{0} & \lambda_{21}\mathbf{A}^{nn} & \lambda_{21}\mathbf{A}^{np} & \lambda_{22}\mathbf{A}^{nn} & \lambda_{22}\mathbf{A}^{np} \\ \mathbf{S}_p'\mathbf{X} & \lambda_{21}\mathbf{A}^{pn} & \mathbf{S}_p'\mathbf{Z}_p + \lambda_{21}\mathbf{A}^{pp} & \lambda_{22}\mathbf{A}^{pn} & \mathbf{S}_p'\mathbf{S}_p + \lambda_{22}\mathbf{A}^{pp} \end{bmatrix}$$
$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_n \\ \hat{\mathbf{u}}_p \\ \hat{\mathbf{m}}_n \\ \hat{\mathbf{m}}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{0} \\ \mathbf{Z}_p'\mathbf{y} \\ \mathbf{0} \\ \mathbf{S}_p'\mathbf{y} \end{bmatrix}. \tag{19}$$

Reducing the above equation to phenotyped animals, equations corresponding to $\hat{\mathbf{u}}_n$ are absorbed in equations corresponding to $\hat{\mathbf{u}}_p$, and equations corresponding to $\hat{\mathbf{m}}_n$ are absorbed in equations corresponding to $\hat{\mathbf{m}}_p$, resulting to:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z}_p & \mathbf{X'S}_p \\ \mathbf{Z}_p'\mathbf{X} & \mathbf{Z}_p'\mathbf{Z}_p + \lambda_{11}\Phi & \mathbf{Z}_p'\mathbf{S}_p + \lambda_{12}\Phi \\ \mathbf{S}_p'\mathbf{X} & \mathbf{S}_p'\mathbf{Z}_p + \lambda_{21}\Phi & \mathbf{S}_p'\mathbf{S}_p + \lambda_{22}\Phi \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_p \\ \hat{\mathbf{m}}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z}_p'\mathbf{y} \\ \mathbf{S}_p'\mathbf{y} \end{bmatrix}. \tag{20}$$

$\hat{\mathbf{u}}_n$ and $\hat{\mathbf{m}}_n$ are back-solved given $\hat{\mathbf{u}}_p$, $\hat{\mathbf{m}}_p$, $\mathbf{A}^{nn}$, $\mathbf{A}^{np}$, and $\mathbf{G}^{-1}$. For the proof, consider rows 2 and 4 in Equation (19):

$$\begin{cases} \lambda_{11}\mathbf{A}^{nn}\hat{\mathbf{u}}_n + \lambda_{11}\mathbf{A}^{np}\hat{\mathbf{u}}_p + \lambda_{12}\mathbf{A}^{nn}\hat{\mathbf{m}}_n + \lambda_{12}\mathbf{A}^{np}\hat{\mathbf{m}}_p = \mathbf{0} \\ \lambda_{21}\mathbf{A}^{nn}\hat{\mathbf{u}}_n + \lambda_{21}\mathbf{A}^{np}\hat{\mathbf{u}}_p + \lambda_{22}\mathbf{A}^{nn}\hat{\mathbf{m}}_n + \lambda_{22}\mathbf{A}^{np}\hat{\mathbf{m}}_p = \mathbf{0} \end{cases}$$
$$\Rightarrow \begin{cases} \mathbf{A}^{nn}(\lambda_{11}\hat{\mathbf{u}}_n + \lambda_{12}\hat{\mathbf{m}}_n) = -\mathbf{A}^{np}(\lambda_{11}\hat{\mathbf{u}}_p + \lambda_{12}\hat{\mathbf{m}}_p) \\ \mathbf{A}^{nn}(\lambda_{21}\hat{\mathbf{u}}_n + \lambda_{22}\hat{\mathbf{m}}_n) = -\mathbf{A}^{np}(\lambda_{21}\hat{\mathbf{u}}_p + \lambda_{22}\hat{\mathbf{m}}_p) \end{cases}$$
$$\Rightarrow \begin{cases} \mathbf{A}^{nn}(g^{11}\hat{\mathbf{u}}_n + g^{12}\hat{\mathbf{m}}_n) = -\mathbf{A}^{np}(g^{11}\hat{\mathbf{u}}_p + g^{12}\hat{\mathbf{m}}_p) \\ \mathbf{A}^{nn}(g^{21}\hat{\mathbf{u}}_n + g^{22}\hat{\mathbf{m}}_n) = -\mathbf{A}^{np}(g^{21}\hat{\mathbf{u}}_p + g^{22}\hat{\mathbf{m}}_p) \end{cases}$$

$$\Rightarrow \left(\mathbf{A}^{nn} \otimes \mathbf{G}^{-1}\right) \begin{bmatrix} \hat{\mathbf{u}}_n \\ \hat{\mathbf{m}}_n \end{bmatrix} = -\left(\mathbf{A}^{np} \otimes \mathbf{G}^{-1}\right) \begin{bmatrix} \hat{\mathbf{u}}_p \\ \hat{\mathbf{m}}_p \end{bmatrix} \tag{21}$$

2.5. Animal Models With Genetic Groups

The idea of sire models with genetic groups (Quaas and Pollak, 1981) was extended to animal models by Westell and Van Vleck (1987), where grouping stems from the fact that unknown parents belonging to different genetic backgrounds and periods of time have different genetic means. An animal model with genetic groups and no additional random effect other than breeding values is written as $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{ZQg} + \mathbf{e}$. In matrix notation:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'ZQ} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \lambda\mathbf{A}^{-1} & \mathbf{Z'ZQ} \\ \mathbf{Q'Z'X} & \mathbf{Q'Z'Z} & \mathbf{Q'Z'ZQ} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{Q'Z'y} \end{bmatrix}, \tag{22}$$

where $\hat{\mathbf{g}}$ is the vector of solutions for genetic groups and $\mathbf{Q}$ is the matrix relating genetic groups to animals. Permuting rows and columns of $\mathbf{A}^{-1}$ so that non-phenotyped animals appear before

phenotyped animals and phenotyped animals appear in the same order in $\mathbf{y}$ and $\mathbf{A}^{pp}$, $\mathbf{Z} = \begin{bmatrix} \mathbf{0} & \mathbf{Z}_p \end{bmatrix}$, $\mathbf{Q'} = \begin{bmatrix} \mathbf{Q}_n' & \mathbf{Q}_p' \end{bmatrix}$, $\mathbf{ZQ} = \mathbf{Z}_p\mathbf{Q}_p$, and $\mathbf{Z}_p = \mathbf{I}$. Thus, Equation (22) can be written as:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{0} & \mathbf{X'} & \mathbf{X'Q}_p \\ \mathbf{0} & \lambda\mathbf{A}^{nn} & \lambda\mathbf{A}^{np} & \mathbf{0} \\ \mathbf{X} & \lambda\mathbf{A}^{pn} & \mathbf{I} + \lambda\mathbf{A}^{pp} & \mathbf{Q}_p \\ \mathbf{Q}_p'\mathbf{X} & \mathbf{0} & \mathbf{Q}_p' & \mathbf{Q}_p'\mathbf{Q}_p \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_n \\ \hat{\mathbf{u}}_p \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{0} \\ \mathbf{y} \\ \mathbf{Q}_p'\mathbf{y} \end{bmatrix}. \tag{23}$$

Absorbing the equations corresponding to $\hat{\mathbf{u}}_n$ in the equations corresponding to $\hat{\mathbf{u}}_p$, the above model is reduced to:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'} & \mathbf{X'Q}_p \\ \mathbf{X} & \mathbf{I} + \lambda\Phi & \mathbf{Q}_p \\ \mathbf{Q}_p'\mathbf{X} & \mathbf{Q}_p' & \mathbf{Q}_p'\mathbf{Q}_p \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_p \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{y} \\ \mathbf{Q}_p'\mathbf{y} \end{bmatrix}. \tag{24}$$

Back-solving $\hat{\mathbf{u}}_n$ is done using Equation (6) and the breeding value of animals are calculated as $\hat{\mathbf{u}} + \mathbf{Q}\hat{\mathbf{g}}$. Forming $\mathbf{Q}$ to calculate breeding values outside of the animal model is an additional computational step. To avoid that step, Quaas and Pollak (1981) transformed the mixed model equations so that the solution for the breeding value includes the genetic group effect:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{0} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \lambda\mathbf{A}^{-1} & -\lambda\mathbf{A}^{-1}\mathbf{Q} \\ \mathbf{0} & -\lambda\mathbf{Q'A}^{-1} & \lambda\mathbf{Q'A}^{-1}\mathbf{Q} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} + \mathbf{Q}\hat{\mathbf{g}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{0} \end{bmatrix}. \tag{25}$$

Let's consider:

$$\Lambda^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{Q} \\ -\mathbf{Q'A}^{-1} & \mathbf{Q'A}^{-1}\mathbf{Q} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{A}^{npg} \\ \mathbf{A}^{gpn} & \mathbf{A}^{gg} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{A}^{nn} & \mathbf{A}^{np} & \mathbf{A}^{ng} \\ \mathbf{A}^{pn} & \mathbf{A}^{pp} & \mathbf{A}^{pg} \\ \mathbf{A}^{gn} & \mathbf{A}^{gp} & \mathbf{A}^{gg} \end{bmatrix},$$

where g corresponds to genetic groups in a larger numerator relationship matrix (i.e., Λ is larger than $\mathbf{A}$). Then, Equation (25) can be written as:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{0} & \mathbf{X'Z}_p & \mathbf{0} \\ \mathbf{0} & \lambda\mathbf{A}^{nn} & \lambda\mathbf{A}^{np} & \lambda\mathbf{A}^{ng} \\ \mathbf{Z}_p'\mathbf{X} & \lambda\mathbf{A}^{pn} & \mathbf{Z}_p'\mathbf{Z}_p + \lambda\mathbf{A}^{pp} & \lambda\mathbf{A}^{pg} \\ \mathbf{0} & \lambda\mathbf{A}^{gn} & \lambda\mathbf{A}^{gp} & \lambda\mathbf{A}^{gg} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_n + \mathbf{Q}_n\hat{\mathbf{g}} \\ \hat{\mathbf{u}}_p + \mathbf{Q}_p\hat{\mathbf{g}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{0} \\ \mathbf{Z}_p'\mathbf{y} \\ \mathbf{0} \end{bmatrix}.$$

Considering $\mathbf{Z}_{pg} = \begin{bmatrix} \mathbf{Z}_p & \mathbf{0} \end{bmatrix}$, in which $\mathbf{0}$ is a matrix of zeros with columns corresponding to genetic groups:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{0} & \mathbf{X'Z}_{pg} \\ \mathbf{0} & \lambda\mathbf{A}^{nn} & \lambda\begin{bmatrix}\mathbf{A}^{np} & \mathbf{A}^{ng}\end{bmatrix} \\ \mathbf{Z}_{pg}'\mathbf{X} & \lambda\begin{bmatrix}\mathbf{A}^{pn} \\ \mathbf{A}^{gn}\end{bmatrix} & \mathbf{Z}_{pg}'\mathbf{Z}_{pg} + \lambda\begin{bmatrix}\mathbf{A}^{pp} & \mathbf{A}^{pg} \\ \mathbf{A}^{gp} & \mathbf{A}^{gg}\end{bmatrix} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_n + \mathbf{Q}_n\hat{\mathbf{g}} \\ \hat{\mathbf{u}}_p + \mathbf{Q}_p\hat{\mathbf{g}} \\ \hat{\mathbf{g}} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{X'y} \\ \mathbf{0} \\ \mathbf{Z}_{pg}'\mathbf{y} \end{bmatrix}. \tag{26}$$

Applying the same technique used to reduce Equation (3) to Equation (4), Equation (26) is reduced to Equation (27):

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_{pg} \\ \mathbf{Z}'_{pg}\mathbf{X} & \mathbf{Z}'_{pg}\mathbf{Z}_{pg} + \lambda\Phi \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_p + \mathbf{Q}_p\hat{\mathbf{g}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'_{pg}\mathbf{y} \end{bmatrix}, \quad (27)$$

where extending Equation (5) to include genetic groups:

$$\Phi = \begin{bmatrix} \Phi_{pp} & \Phi_{pg} \\ \Phi_{gp} & \Phi_{gg} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{pp} & \mathbf{A}^{pg} \\ \mathbf{A}^{gp} & \mathbf{A}^{gg} \end{bmatrix} - \begin{bmatrix} \mathbf{A}^{pn} \\ \mathbf{A}^{gn} \end{bmatrix} (\mathbf{A}^{nn})^{-1} \begin{bmatrix} \mathbf{A}^{pn} \\ \mathbf{A}^{gn} \end{bmatrix}' \quad (28)$$

and extending Equation (6) to include genetic groups:

$$\mathbf{A}^{nn}(\hat{\mathbf{u}}_n + \mathbf{Q}_n\hat{\mathbf{g}}) = -\begin{bmatrix} \mathbf{A}^{np} & \mathbf{A}^{np} \end{bmatrix} \begin{pmatrix} \hat{\mathbf{u}}_p + \mathbf{Q}_p\hat{\mathbf{g}} \\ \hat{\mathbf{g}} \end{pmatrix}. \quad (29)$$

2.6. Animal Models With a Genomic Relationship Matrix (GBLUP)

In the previous sections, animal models with pedigree-based relationship matrices were discussed. With the availability of genomic data, various forms of genomic relationship matrix (e.g., VanRaden, 2008; Hickey et al., 2013) are used in animal models. Replacing $\mathbf{A}$ with the genomic relationship matrix $\mathbf{G}$ in Equation (1) results in:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}. \quad (30)$$

However, it should be noted that unlike Equation (1), Equation (30) is limited to genotyped animals. Due to efficient methods for deriving $\mathbf{A}^{-1}$ without calculating and inverting $\mathbf{A}$ (Henderson, 1976; Quaas, 1976), Equation (4) is preferred over Equation (7). However, those methods are not applicable to $\mathbf{G}$, and this form of the equations requires $\mathbf{G}$ to be calculated and directly inverted. As a result, the reduced method of choice is equivalent to Equation (7):

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' \\ \mathbf{X} & \mathbf{I} + \lambda\mathbf{G}_{pp}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{y} \end{bmatrix}. \quad (31)$$

Therefore, only $\mathbf{G}_{pp}$ needs to be formed and inverted. If the calculation of $\mathbf{G}_{pp}$ requires allele frequencies, estimation of allele frequencies should not be limited to phenotyped animals. Then, $\hat{\mathbf{u}}_n$ can be derived indirectly by, first back-solving marker effects ($\hat{\mathbf{a}}$) from $\hat{\mathbf{u}}_p$ (Strandén and Garrick, 2009), and then summing up marker effect solutions over all markers for non-phenotyped animals ($\hat{\mathbf{u}}_n = \mathbf{M}_n\hat{\mathbf{a}}$ (Meuwissen et al., 2001), where $\mathbf{M}_n$ contains rows of $\mathbf{M}$ for non-phenotyped animals, and $\mathbf{M}$ is the genotype matrix with coefficients -1, 0, and 1).

2.7. Animal Models With an Augmented Pedigree-Genomic Relationship Matrix (ssGBLUP)

Aguilar et al. (2010) and Christensen and Lund (2010) introduced an animal model for simultaneous genetic evaluation of

genotyped and non-genotyped animal, using a pedigree-genomic augmented inverse matrix ($\mathbf{H}^{-1}$) to be used instead of $\mathbf{A}^{-1}$ in the mixed model equations, where

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{H}^{11} & \mathbf{H}^{12} \\ \mathbf{H}^{21} & \mathbf{H}^{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} + \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}, \quad (32)$$

$\mathbf{H}^{11}$ corresponds to the block for non-genotyped animals, and $\mathbf{H}^{22}$ corresponds to the block for genotyped animals. Though, it is possible to partition $\mathbf{H}^{-1}$ to $\begin{bmatrix} \mathbf{H}^{nn} & \mathbf{H}^{np} \\ \mathbf{H}^{pn} & \mathbf{H}^{pp} \end{bmatrix}$ and reduce the model to phenotyped animals (similar to the procedures already explained), perhaps a more practical approach would be to split animals into three groups of "non-phenotyped and non-genotyped" animals (4), "phenotyped and non-genotyped" animals (3), and genotyped animals (2):

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{H}^{44} & \mathbf{H}^{43} & \mathbf{H}^{42} \\ \mathbf{H}^{34} & \mathbf{H}^{33} & \mathbf{H}^{32} \\ \mathbf{H}^{24} & \mathbf{H}^{23} & \mathbf{H}^{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{44} & \mathbf{A}^{43} & \mathbf{A}^{42} \\ \mathbf{A}^{34} & \mathbf{A}^{33} & \mathbf{A}^{32} \\ \mathbf{A}^{24} & \mathbf{A}^{23} & \mathbf{H}^{22} \end{bmatrix}.$$

Grouping phenotyped or genotyped animals together (5):

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{A}^{44} & \mathbf{A}^{45} \\ \mathbf{A}^{54} & \mathbf{H}^{55} \end{bmatrix}, \mathbf{H}^{55} = \begin{bmatrix} \mathbf{A}^{33} & \mathbf{A}^{32} \\ \mathbf{A}^{23} & \mathbf{H}^{22} \end{bmatrix}, \mathbf{A}^{54} = \begin{bmatrix} \mathbf{A}^{34} \\ \mathbf{A}^{24} \end{bmatrix}.$$

Then, an animal model like the following:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \quad (33)$$

can be reduced to:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_5 \\ \mathbf{Z}'_5\mathbf{X} & \mathbf{Z}'_5\mathbf{Z}_5 + \lambda\Theta \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_5 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'_5\mathbf{y} \end{bmatrix}, \quad (34)$$

where $\Theta = \mathbf{H}^{55} - \mathbf{A}^{54}(\mathbf{A}^{44})^{-1}\mathbf{A}^{45}$ and $\mathbf{Z}_5$ contains columns of $\mathbf{Z}$ corresponding to group 5 of animals. After obtaining $\hat{\mathbf{u}}_5$, $\hat{\mathbf{u}}_4$ can be back-solved by solving

$$\mathbf{A}^{44}\hat{\mathbf{u}}_4 = -\mathbf{A}^{45}\hat{\mathbf{u}}_5. \quad (35)$$

3. RESULTS

Results from numerical examples are provided. The numerical examples were chosen from small datasets used in Mrode (2005) for different animal models. R scripts for producing the data and the analyses are available in the data repository. Files for any given filename in this study are also available in the data repository. Back-solving were done by direct solving of Equations (6), (11), (21), and (29). Where solving these equations for large pedigree is computationally challenging, preconditioned conjugate gradient method (Strandén and Lidauer, 1999) can be adopted.

3.1. Basic Animal Model

Data from Example 3.1 of Mrode (2005) was used (**Supplementary Table 1**). The data can be produced using `e31data.R` in the data repository. Equation (1) is coded in function "BLUP" in `e31.R`. The following code produces the data and solutions.

```
# Create data
source("e31data.R")
# Call the solver function
source("e31.R")
# Solve the full model
sol1 = BLUP(X, y, Z, lambda, Ai)
```

Applying Equation (4) instead of Equation (1) requires data transformation. Matrices Φ, $\mathbf{A}^{nn}$ and $\mathbf{A}^{np}$ are obtained using function "getPhi" coded in `getPhi.R`. Solutions of the reduced model are obtained as:

```
# Call the function to create Phi matrix
source("getPhi.R")
# Define phenotyped animals
phe=4:8
# Create Phi and back-solving matrices
mats = getPhi(Ai,phe)
# Solve the reduced model
sol2 = BLUP(X, y, Z=Z[,phe], lambda,
 Ai=mats$Phi)
```

Following Equation (6), $\hat{\mathbf{u}}_n$ are back-solved using the following code:

```
u_n = solve(mats$Ann, mats$Anp %*%
 -sol2[-2:-1])
```

The following command confirms that solutions from the full (Equation 1) and the reduced models (Equation 4) are identical.

```
source("check.R") # Call the function
check(sol1, c(sol2[1:2], u_n, sol2[3:7]))
```

We run a test to compare the computational time for the full and the reduced animal model for a pedigree of 10,000 animals, of which 2,000 animals had been phenotyped. The runtime were 10 and 8.8 s for the full and the reduced animal model, respectively. Additionally, it took 2 s for the reduced animal model to calculate Φ. Matrix Φ was considerably denser than the $\mathbf{A}^{-1}$ matrix.

3.2. Multiple Trait Animal Model

Data from Example 5.3 of Mrode (2005) was used (**Supplementary Table 2**). The data can be produced using `e53data.R` in the data repository. Equation (8) is coded in function "BLUP" in `e53.R`. The following code produces the data and solutions.

```
# Create data
source("e53data.R")
# Call the solver function
source("e53.R")
# Solve the full model
sol3 = BLUP(X, y, Z, Ri, Gi, Ai)
```

Applying Equation (10) instead of Equation (8) requires data transformation. Matrices Φ, $\mathbf{A}^{nn}$ and $\mathbf{A}^{np}$ are obtained using function "getPhi" coded in `getPhi.R`. Solutions of the reduced model are obtained as:

```
# Call the function to create Phi matrix
source("getPhi.R")
# Define phenotyped animals
phe=4:9
# Create Phi and back-solving matrices
mats = getPhi(Ai,phe)
# Solve the reduced model
sol4 = BLUP(X, y, Z=Z[,c(phe,phe+n)],
 Ri, Gi, Ai=mats$Phi)
```

Following Equation (11), $\hat{\mathbf{u}}_n$ are back-solved using the following code:

```
u_n = solve(kronecker(Gi, mats$Ann),
    kronecker(Gi, mats$Anp) %*%
    -sol4[-4:-1])
```

The following command confirms that solutions from the full (Equation 8) and the reduced models (Equation 10) are identical.

```
source("check.R") # Call the function
check(sol3, c(sol4[1:4], u_n[1:3],
 sol4[5:10], u_n[4:6], sol4[11:16]))
```

3.3. Repeatability Animal Model

Data from Example 4.1 of Mrode (2005) was used (**Supplementary Table 3**). The data can be produced using `e41data.R` in the data repository. Equation (12) is coded in function "BLUP" in `e41.R`. The following code produces the data and solutions.

```
# Create data
source("e41data.R")
# Call the solver function
source("e41.R")
# Solve the full model
sol5 = BLUP(X, y, Z, W, Va, Vpe, Ve, Ai)
```

Applying Equation (14) instead of Equation (12) requires data transformation. Matrices Φ, $\mathbf{A}^{nn}$ and $\mathbf{A}^{np}$ are obtained using function "getPhi" coded in `getPhi.R`. Solutions of the reduced model are obtained as:

```
# Call the function to create Phi matrix
source("getPhi.R")
# Define phenotyped animals
phe=4:8
# Create Phi and back-solving matrices
mats = getPhi(Ai,phe)
# Solve the reduced model
sol6 = BLUP(X, y, Z=Z[,phe], W, Va,
 Vpe, Ve, Ai=mats$Phi)
```

Following Equation (6), $\hat{\mathbf{u}}_n$ are back-solved using the following code:

```
u_n = solve(mats$Ann, mats$Anp %*%
   -sol6[5:9])
```

The following command confirms that solutions from the full (Equation 12) and the reduced models (Equation 14) are identical.

```
source("check.R") # Call the function
check(sol5, c(sol6[1:4], u_n,
   sol6[5:14]))
```

3.4. Maternal Trait Animal Models

Data from Example 6.1 of Mrode (2005) was used (**Supplementary Table 4**). This example includes maternal genetic and maternal permanent environment effects. The data can be produced using e61data.R in the data repository. The equation system is coded in function "BLUP" in e61.R. The following code produces the data and solutions.

```
# Create data
source("e61data.R")
# Call the solver function
source("e61.R")
# Solve the full model
sol7 = BLUP(X, y, Z, W, S, G, Vmpe,
   Ve, Ai)
```

Applying the reduced model requires data transformation. Matrices Φ, $\mathbf{A}^{nn}$, and $\mathbf{A}^{np}$ are obtained using function "getPhi" coded in getPhi.R. Solutions of the reduced model are obtained as:

```
# Call the function to create Phi matrix
source("getPhi.R")
# Define phenotyped animals
phe=5:14
# Define dams
dams = c(2,5:7)
# Find being phenotyped or dam
pheORdam = sort(unique(c(phe, dams)))
# Create Phi and back-solving matrices
mats = getPhi(Ai,pheORdam)
# Solve the reduced model
sol8 = BLUP(X, y, Z=Z[,pheORdam],
   W=W[,pheORdam],
   S, G, Vmpe, Ve, Ai=mats$Phi)
```

Following Equation (21), $\hat{\mathbf{u}}_n$ and $\hat{\mathbf{u}}_n$ are back-solved using the following code:

```
Gi = solve(G)
sol_n = solve(kronecker(Gi, mats$Ann),
       kronecker(Gi, mats$Anp) %*%
          -sol8[5:26])
# Separate direct and maternal genetic
   back-solved solutions
u_n = sol_n[1:3]
m_n = sol_n[-3:-1]
```

The following command confirms that solutions from the full and the reduced models are identical.

```
source("check.R") # Call the function
check(sol7, c(sol8[1:4], u_n[1],
   sol8[5], u_n[2:3],
       sol8[6:15], m_n[1], sol8[16],
       m_n[2:3], sol8[17:30]))
```

3.5. Animal Models With Genetic Groups

Data from Example 3.4 of Mrode (2005) was used (**Supplementary Table 5**). The data can be produced using e34data.R in the data repository. Equation (25) is coded in function "BLUP" in e34.R. The following code produces the data and solutions.

```
# Create data
source("e34data.R")
# Call the solver function
source("e34.R")
# Solve the full model
sol9 = BLUP(X, y, Z, lambda, Ai, Q)
```

Applying Equation (27) instead of Equation (25) requires data transformation. Matrices Φ, $\mathbf{A}^{nn}$, and $\mathbf{A}^{np}$ are obtained using function "getPhi" coded in getPhi.R. Solutions of the reduced model are obtained using function "BLUP" coded in e31.R, and the following code.

```
# Call the function to create Phi matrix
source("getPhi.R")
# Call the solver function for the
   reduced model
source("e31.R")
# Define phenotyped animals
phe=4:8
# Create Z matrix for the reduced model
Z = Z[,phe]
Z = cbind(Z, matrix(0,nrow(Z),ncol(Q)))
# Make extended A^-1 including genetic
 groups
tmp = cbind(diag(nrow(Q)), -Q)
Ai2 = crossprod(tmp, Ai) %*% tmp
# Create Phi and back-solving matrices
# Reduction to phenotyped animals and
 genetic groups
mats = getPhi(Ai=Ai2, c(phe,9))
# Solve the reduced model
sol10 = BLUP(X, y, Z, lambda,
 Ai=mats$Phi)
```

Following Equation (29), $\hat{\mathbf{u}}_n + \mathbf{Q}_n\hat{\mathbf{g}}$ are back-solved using the following code:

```
u_n = solve(mats$Ann, mats$Anp %*%
   - sol10[-2:-1])
```

The following command confirms that solutions from the full (Equation 25) and the reduced models (Equation 27) are identical.

```
source("check.R") # Call the function
check(sol9, c(sol10[1:2], u_n,
  sol10[3:8]))
```

3.6. Animal Models With a Genomic Relationship Matrix (GBLUP)

Data from Example 3.1 of Mrode (2005) was used (**Supplementary Table 1**). The data can be produced using e31data.R in the data repository. Rather than using the $\mathbf{A}^{-1}$ matrix produced by e31data.R (data repository), a $\mathbf{G}^{-1}$ matrix was used, created by makeGi.R (data repository). The following code creates the data and solutions. The solutions are obtained using the function "BLUP" in e31.R.

```
# Create data
source("e31data.R")
# Call the solver function
source("e31.R")
# Make a G^-1 matrix
source("makeGi.R")
# Solve the full model
sol11 = BLUP(X, y, Z, lambda, Ai=Gi)
```

Applying Equation (31) instead of Equation (30) requires data transformation. Solutions of the reduced model are obtained as:

```
# Define phenotyped animals
phe=4:8
# Define non-phenotyped animals
nonphe = which(! 1:nrow(ped) %in% phe)
# Make G for phenotyped animals
 (reduced model)
G = tcrossprod(M[phe,]-2*P[phe,])/
(2*sum(p*(1-p))) +
    diag(length(phe))/10
# Invert G
Gi2 = solve(G)
# Solve the reduced model
sol12 = BLUP(X, y, Z=Z[,phe], lambda,
 Ai=Gi2)
```

Following Equation (6) (substituting $\mathbf{A}$ to $\mathbf{G}$), $\hat{\mathbf{u}}_n$ are back-solved using the following code:

```
u_n = solve(Gi[nonphe,nonphe],
  Gi[nonphe,phe] %*%  -sol12[-2:-1])
```

The following command confirms that solutions from the full (Equation 30) and the reduced (Equation 31) models are identical.

```
source("check.R") # Call the function
check(sol11, c(sol12[1:2], u_n,
  sol12[3:7]))
```

3.7. Animal Models With an Augmented Pedigree-Genomic Relationship Matrix (ssGBLUP)

Data from Example 3.1 of Mrode (2005) was used (**Supplementary Table 1**). The data can be produced using

e31data.R in the data repository. It was considered that animals 3, 7, and 8 are genotyped and a 3×3 arbitrary $(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})$ matrix was considered for genotyped animals. An $\mathbf{H}^{-1}$ matrix was created (Equation 32) to be used instead of $\mathbf{A}^{-1}$. The following code creates the data and solutions. The solutions are obtained using the function "BLUP" in e31.R.

```
# Create data
source("e31data.R")
# Call the solver function
source("e31.R")
# Define genotyped animals
geno = c(3,7,8)
# Make an arbitrary G^-1 - A_22^-1
GimA22i = matrix(c(50,2,-5,2,40,6,-5,6,
60)/100, nrow=3)
# Create H^-1
Hi = Ai
Hi[geno, geno] = Ai[geno, geno]
 + GimA22i
# Solve the full model
sol13 = BLUP(X, y, Z, lambda, Ai=Hi)
```

Applying Equation (34) instead of Equation (33) requires data transformation. Solutions of the reduced model are obtained as:

```
# Define phenotyped animals
phe=4:8
# Define phenotyped or genotyped
 animals
pheORgeno = sort(union(phe, geno))
# Define non-phenotyped and
 non-genotyped animals
nonpheANDnongeno = setdiff(1:nrow(Ai),
 pheORgeno)
# Create H^55
H55 = Hi[pheORgeno, pheORgeno]
# Create A^44
A44 = Ai[nonpheANDnongeno,
nonpheANDnongeno]
# Create A^45
A45 = Ai[nonpheANDnongeno, pheORgeno]
# Create Theta matrix (Eq. 34)
Theta = H55 - crossprod(A45,solve(A44))
 %*% A45
# Solve the reduced model
sol14 = BLUP(X, y, Z=Z[,pheORgeno],
lambda, Ai=Theta)
```

Following Equation (35), solutions for non-phenotyped and non-genotyped animals are back-solved using the following code:

```
u_n = solve(A44, A45 %*% -sol14[-2:-1])
```

The following command confirms that solutions from the full (Equation 33) and the reduced (Equation 34) models are identical.

```
source("check.R") # Call the function
check(sol13, c(sol14[1:2], u_n,
  sol14[3:8]))
```

4. DISCUSSION

Genetic evaluation centers around the world deal with tight deadlines for delivering breeding value predictions of animals to the industry. The amount of data to be dealt with usually contains millions of animals and that can make restraints around computational time and resources. Often, there are many historical or ancestral animals as well as inactive (culled) breeding animals, for which breeding values are not routinely needed. Computational demand can be reduced by applying reduced animal models that are equivalent to the full animal models.

Several forms of animal model were reduced to comprise equations for breeding values only for phenotyped animals, and extension to other forms of animal model can be done following the same logic. The efficiency of these models depends on several factors, such as the proportion of phenotyped animals in the population being evaluated, sparseness of Φ, and the number of effects in common between the full and the reduced model. There is also a trade-off between the number of phenotyped and non-phenotyped animals. The smaller the proportion of phenotyped animals the smaller the size of MME, but the greater the computational cost of calculating Φ as a result of a larger $\mathbf{A}^{nn}$. The sparsity of Φ, which itself depends on the sparsity of $\mathbf{A}^{nn}$ [the sparser the $\mathbf{A}^{nn}$, the sparser the $(\mathbf{A}^{nn})^{-1}$] can heavily reduce the efficiency of the reduced animal model. Depending on the size of Φ, it might have more non-zero elements than the $\mathbf{A}^{-1}$ matrix. This might erode any computational benefit from the reduced model compared to the full model. We suggest two solutions for this problem. One is discarding non-phenotyped animals that have little or no relatedness with phenotyped progeny (i.e., a few generations distant from any phenotyped descendant), and the other is re-ordering $\mathbf{A}^{nn}$ to obtain two diagonal blocks, one very sparse ($\mathbf{A}^{ss}$) and the other ($\mathbf{A}^{dd}$) relatively dense $\left(\mathbf{A}^{nn} = \begin{bmatrix} \mathbf{A}^{ss} & \mathbf{A}^{sd} \\ \mathbf{A}^{ds} & \mathbf{A}^{dd} \end{bmatrix}\right)$. Then, $\mathbf{A}^{dd}$, $\mathbf{A}^{sd}$, and $\mathbf{A}^{ds}$ are appended to $\mathbf{A}^{pp}$, $\mathbf{A}^{np}$, and $\mathbf{A}^{pn}$. This also requires appending rows of 0 to $\mathbf{X}$ and columns of 0 to $\mathbf{Z}$ (corresponding to the appended non-phenotyped animals). A pedigree-only-based back-solving (such as for BLUP and for ssGBLUP dividing animals to "phenotyped or genotyped" and "non-phenotyped and non-genotyped") can be done efficiently for millions of animals.

Solutions for non-phenotyped animals can be back-solved after the solutions for phenotyped animals are obtained. Generally, back-solving is more efficient and faster than solving the mixed model equations, and that is due to the known structure of $\mathbf{A}^{-1}$. Usually, breeding values of only a small proportion of non-phenotyped animals are of interest (e.g., active non-phenotyped breeding animals). Consequently, the back-solving procedure can be omitted by explicitly including those animals together with phenotyped animals in the reduced animal model.

The same reduction technique is applicable to BLUP using the genomic relationship matrix (GBLUP). That would reduce the computational demand to form and invert $\mathbf{G}$ by limiting it to phenotyped animals only, and the computational demand to solve a smaller set of mixed model equations. Note that the computational cost of inverting $\mathbf{G}$ is cubic to its dimension unless it is not full rank because it was formed with fewer markers than the number of genotyped animals, where an equivalent model (Fernando et al., 2016) can be used or it uses a reduced rank approximation, such as the algorithms introduced by Misztal (2016) and Mäntysaari et al. (2017). Though, back-solving is possible, it is discouraged as it involves the availability of the blocks of $\mathbf{G}^{-1}$ among non-phenotyped animals and between non-phenotyped and phenotyped animals. Because most of the computational complexity of GBLUP is due to forming and inverting $\mathbf{G}$, it is recommended to form and invert $\mathbf{G}$ for phenotyped animals together with non-phenotyped animals of interest. However, it is assumed that either the allele frequencies in the whole population are known or allele frequencies in the individuals present in the reduced model are not different from the allele frequencies in the whole population. Also, as it was mentioned earlier, another way of back-solving for non-phenotyped animals is through back-solving of marker effects. There are other forms of reduced GBLUP proposed by other studies (Cantet and Smith, 1991; Hoeschele, 1993; Arendonk et al., 1994; Saito and Iwaisaki, 1996, 1997). Comparing the efficiency of these methods and possibly combining them with the reduced GBLUP presented in this study are yet to be investigated. The computational efficiency of the reduced GBLUP presented in this study heavily depends on the proportion of animals not included in the reduced model (genotyped and non-phenotyped animals whose evaluation is not of interest).

There are two possibilities for ssGBLUP; dividing animals to phenotyped and non-phenotyped animals or dividing animals to "phenotyped or genotyped" and "non-phenotyped and non-genotyped" groups. The latter is preferred as it makes back-solving independent from $\mathbf{H}^{22}$ (block of $\mathbf{H}^{-1}$ corresponding to genotyped animals) and dependent to blocks of $\mathbf{A}^{-1}$ between the two groups of animals and among "non-phenotyped and non-genotyped" animals. Blocks of $\mathbf{A}^{-1}$ are much easier to calculate and keep in the memory, because of its sparsity.

Variance components can be estimated using reduced animal models (e.g., Henderson, 1986; Besbes et al., 1992; White et al., 2006; Guy et al., 2009). Restricted maximum likelihood (REML, Patterson and Thompson, 1971) has been a popular method for variance components estimation. Generalized inverses of coefficient matrices of mixed model equations are required under both the full and the reduced animal models (Henderson, 1986). Reduced animal models have considerably smaller order coefficient matrix and consequently less computational complexity of variance components estimation. Henderson (1986) estimated variances for the reduced model based on the method of Quaas and Pollak (1980). Parents with phenotyped progeny are evaluated in this reduced model. To account for Mendelian sampling and for the genetic contribution of unknown parents, there is an additional component (compared to the full model) to the variance of residuals (Henderson, 1986).

The introduced reduction method in this study evaluates all the phenotyped animals and no extra assumption or component is required to hold the assumption on $var(\mathbf{y})$. The matrices used in the reduced model do not make any deviation from the assumptions on variance components. Also, creating the $\mathbf{Z}$ matrix for the reduced model in this study is trivial (if the vector of phenotypes is ordered by the order of animals in the pedigree, $\mathbf{Z}$ is an identity matrix with the order of the number of phenotypes; for a repeatability model, it is still an incidence matrix with coefficients 0 and 1, relating phenotypes to animals). Creating the $\mathbf{Z}$ matrix for the reduced model of Quaas and Pollak (1980) requires pedigree information.

The presented reduced animal models can further be reduced. This can be achieved by a "2-fold" model reduction, which involves reducing the model to equations for phenotyped animals followed by reducing the model to parents of phenotyped animals (Quaas and Pollak, 1980). Such 2-fold reduced animal model contains equations corresponding to phenotyped parents of phenotyped animals. A 2-fold reduced animal model also requires a 2-fold back-solving procedure, in which the solutions of non-phenotyped parents are back-solved from the solutions of phenotyped parents, and then the solutions of non-parents (and parents of non-phenotyped progeny) are back-solved from the solutions of parents. This 2-fold back-solving can be omitted or reduced to a 1-fold back-solving by putting together a group of animals of interest to phenotyped parents with phenotyped progeny. Further research is required on this subject.

Finally, depending on the size of the data and the number of traits and effects in the model, the equation system of the reduced animal model may become small enough that direct solving of the mixed model equations might become faster and more efficient than iterative procedures for solving mixed model equations (Schaeffer and Kennedy, 1986; Berger et al., 1989; Strandén and Lidauer, 1999).

DATA AVAILABILITY STATEMENT

The R scripts to produce the data and perform the analyses for this study can be found in the figshare repository [https://doi.org/10.6084/m9.figshare.13369607].

AUTHOR CONTRIBUTIONS

DG initiated the idea of the study. MN extended the idea to other forms of animal model, drafted the manuscript, coded the examples, and validated the results. Both DG and MN took active part in the preparation and revision of the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.637626/full#supplementary-material

REFERENCES

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of holstein final score. *J. Dairy Sci.* 93, 743–752. doi: 10.3168/jds.2009-2730

Arendonk, J. A. V., Tier, B., and Kinghorn, B. P. (1994). Use of multiple genetic markers in prediction of breeding values. *Genetics* 137, 319–329. doi: 10.1093/genetics/137.1.319

Berger, P. J., Luecke, G. R., and Hoekstra, J. A. (1989). Iterative algorithms for solving mixed model equations. *J. Dairy Sci.* 72, 514–522. doi: 10.3168/jds.S0022-0302(89)79134-7

Besbes, B., Ducrocq, V., Foulley, J. L., Protais, M., Tavernier, A., Tixier-Boichard, M., et al. (1992). Estimation of genetic parameters of EGG production traits of laying hens by restricted maximum likelihood applied to a multiple-trait reduced animal model. *Genet. Select. Evol.* 24:539. doi: 10.1186/1297-9686-24-6-539

Cantet, R., and Smith, C. (1991). Reduced animal model for marker assisted selection using best linear unbiased prediction. *Genet. Select. Evol.* 23:221. doi: 10.1186/1297-9686-23-3-221

Christensen, O. F., and Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genet. Select. Evol.* 42:2. doi: 10.1186/1297-9686-42-2

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *J. R. Stat. Soc. B Methodol.* 39, 1–38. doi: 10.1111/j.2517-6161.1977.tb01600.x

Fernando, R. L., Cheng, H., and Garrick, D. J. (2016). An efficient exact method to obtain gblup and single-step gblup when the genomic relationship matrix is singular. *Genet. Select. Evol.* 48:80. doi: 10.1186/s12711-016-0260-7

Fernando, R. L., and Grossman, M. (1989). Marker assisted selection using best linear unbiased prediction. *Genet. Select. Evol.* 21:467. doi: 10.1186/1297-9686-21-4-467

Guy, D. R., Bishop, S. C., Woolliams, J. A., and Brotherstone, S. (2009). Genetic parameters for resistance to infectious pancreatic necrosis in pedigreed atlantic salmon (*Salmo salar*) post-smolts using a reduced animal model. *Aquaculture* 290, 229–235. doi: 10.1016/j.aquaculture.2009.02.015

Henderson, C. R. (1974). General flexibility of linear model techniques for sire evaluation. *J. Dairy Sci.* 57, 963–972. doi: 10.3168/jds.S0022-0302(74)84993-3

Henderson, C. R. (1976). A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32, 69–83. doi: 10.2307/2529339

Henderson, C. R. (1985). Equivalent linear models to reduce computations. *J. Dairy Sci.* 68, 2267–2277. doi: 10.3168/jds.S0022-0302(85)81099-7

Henderson, C. R. (1986). Estimation of variances in animal model and reduced animal model for single traits and single records. *J. Dairy Sci.* 69, 1394–1402. doi: 10.3168/jds.S0022-0302(86)80546-X

Hickey, J. M., Kinghorn, B. P., Tier, B., Clark, S. A., van der Werf, J. H., and Gorjanc, G. (2013). Genomic evaluations using similarity between haplotypes. *J. Anim. Breed. Genet.* 130, 259–269. doi: 10.1111/jbg.12020

Hoeschele, I. (1993). Elimination of quantitative trait loci equations in an animal model incorporating genetic marker data. *J. Dairy Sci.* 76, 1693–1713. doi: 10.3168/jds.S0022-0302(93)77503-7

Horn, R., and Johnson, C. (1991). *Topics in Matrix Analysis*. Cambridge: Cambridge University Press.

Mäntysaari, E. A., Evans, R. D., and Strandén, I. (2017). Efficient single-step genomic evaluation for a multibreed beef cattle population having many genotyped animals. *J. Anim. Sci.* 95, 4728–4737. doi: 10.2527/jas2017.1912

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

Misztal, I. (2016). Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. *Genetics* 202, 401–409. doi: 10.1534/genetics.115.182089

Mrode, R. A. (2005). *Linear Models for the Prediction of Animal Breeding Values*. Oxfordshire: CABI Publishing.

Nilforooshan, M. A. (2021). *pedSimulate: Pedigree, Genetic Merit and Phenotype Simulation*. Version 0.1.2. Available online at: https://cran.r-project.org/package=pedSimulate (accessed February 23, 2021).

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

Quaas, R. L. (1976). Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* 32, 949–953. doi: 10.2307/2529279

Quaas, R. L., and Pollak, E. J. (1980). Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.* 51, 1277–1287. doi: 10.2527/jas1981.5161277x

Quaas, R. L., and Pollak, E. J. (1981). Modified equations for sire models with groups. *J. Dairy Sci.* 64, 1868–1872. doi: 10.3168/jds.S0022-0302(81)82778-6

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Version 3.6.2. Vienna: R Foundation for Statistical Computing.

Saito, A., and Iwaisaki, H. (1996). A reduced animal model with elimination of quantitative trait loci equations for marker-assisted selection. *Genet. Select. Evol.* 28:465. doi: 10.1186/1297-9686-28-6-465

Saito, A., and Iwaisaki, H. (1997). A reduced animal model approach to predicting total additive genetic merit for marker-assisted selection. *Genet. Select. Evol.* 29:25. doi: 10.1186/1297-9686-29-1-25

Schaeffer, L. R., and Kennedy, B. W. (1986). Computing strategies for solving mixed model equations. *J. Dairy Sci.* 69, 575–579. doi: 10.3168/jds.S0022-0302(86)80441-6

Strandén, I., and Garrick, D. J. (2009). Technical note: derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J. Dairy Sci.* 92, 2971–2975. doi: 10.3168/jds.2008-1929

Strandén, I., and Lidauer, M. (1999). Solving large mixed linear models using preconditioned conjugate gradient iteration. *J. Dairy Sci.* 82, 2779–2787. doi: 10.3168/jds.S0022-0302(99)75535-9

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Westell, R. A., and Van Vleck, L. D. (1987). Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. *J. Dairy Sci.* 70, 1006–1017. doi: 10.3168/jds.S0022-0302(87)80106-6

White, I. M., Roehe, R., Knap, P. W., and Brotherstone, S. (2006). Variance components for survival of piglets at farrowing using a reduced animal model. *Genet. Select. Evol.* 38:359. doi: 10.1186/1297-9686-38-4-359