J. Dairy Sci. 106:1518–1532
https://doi.org/10.3168/jds.2022-22629

Invited review: Reliability computation from the animal model era to the single-step genomic model era

Hafedh Ben Zaabza,[1,2*] Curtis P. Van Tassell,[2] Jeremie Vandenplas,[3] Paul VanRaden,[2] Zengting Liu,[4] Herwin Eding,[5] Stephanie McKay,[1] Katrine Haugaard,[6] Martin H. Lidauer,[7] Esa A. Mäntysaari,[7] and Ismo Strandén[7]

[1]Department of Animal and Veterinary Sciences, University of Vermont, Burlington 05405
[2]Animal Improvement Programs Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD 20705-2350
[3]Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH, Wageningen, the Netherlands
[4]IT Solutions for Animal Production (vit), Heinrich-Schröder-Weg 1, D-27283 Verden, Germany
[5]CRV BV, Wassenaarweg, 20, 6843 NW, Arnhem, the Netherlands
[6]Interbull Centre, SLU, Box 7023, S-75007 Uppsala, Sweden
[7]Natural Resources Institute Finland (Luke), FI-31600 Jokioinen, Finland

ABSTRACT

The calculation of exact reliabilities involving the inversion of mixed model equations poses a heavy computational challenge when the system of equations is large. This has prompted the development of different approximation methods. We give an overview of the various methods and computational approaches in calculating reliability from the era before the animal model to the era of single-step genomic models. The different methods are discussed in terms of modeling, development, and applicability in large dairy cattle populations. The paper also describes the problems faced in reliability computation. Many details dispersed throughout the literature are presented in this paper. It is clear that a universal solution applicable to every model and input data may not be possible, but we point out several efficient and accurate algorithms developed recently for a variety of very large genomic evaluations.
Key words: estimated breeding values, single-step genomic models, dairy cattle, reliability computation

INTRODUCTION

The computation of reliability for EBV is challenging, and this challenge has been further exacerbated by the increase in genomic information in the recent past. Reliability is defined as the squared correlation between the true breeding value and EBV (Wilmink and Dommerholt, 1985), and has measured the agreement of the true with the estimated genetic merit for almost a century (Goodale, 1928). When comparing EBV, breeders are interested not only in the genetic predictions but also in the reliability of these predictions (Meyer and Tier, 2003). The reliability of an EBV can also be considered as a measure of information content that contributed to that prediction. In practice, reliability is computed using the prediction error variance (**PEV**) of EBV and the genetic variance of true breeding value (VanVleck, 1993). The PEV can be obtained from elements of the inverse of the coefficient matrix of the mixed model equations (**MME**) (Henderson, 1984). However, usually, the MME inverse matrix cannot be computed due to its size or to the loss of numerical precision (Hickey et al., 2009), even when using sparse matrix techniques (Meyer and Tier, 2003). Traditional pedigree-based genetic evaluations use an animal model where the size of MME increases with the number of animals in the pedigree. A variety of methods have been proposed for approximating PEV for an animal model (Misztal and Wiggans, 1988; Harris and Johnson, 1998; Jamrozik et al., 2000; Liu et al., 2004; Tier and Meyer, 2004).

Inexpensive genotyping has led to the use of genomic data in dairy cattle evaluations (Van Tassell et al., 2011; García-Ruiz et al., 2016; VanRaden, 2020). Genomic data presents a different challenge than the use of only pedigree information. Genomic predictions can be computed by 2 equivalent models, typically referred to as genomic relationship-based BLUP (**GBLUP**) and SNP-based BLUP (**SNPBLUP**) (VanRaden, 2008; Strandén and Garrick, 2009). In GBLUP, the size of MME increases with the number of genotyped animals, whereas in SNPBLUP the dimension of this set of equations is bounded by the number of SNP markers. The pedigree and genomic information can be combined into a blended model, the single-step genomic BLUP (**ssGBLUP**) model (Aguilar et al., 2010; Christensen and Lund, 2010) that permits simultaneous analysis of phenotypes for genotyped and nongenotyped animals.

Received August 5, 2022.
Accepted November 7, 2022.
*Corresponding author: Hafedh.Ben-Zaabza@uvm.edu

There is a further class of single-step models that are otherwise equivalent to ssGBLUP but based on marker effect models (**ssSNPBLUP**) and with SNP markers treated as covariates (Fernando et al., 2014; Liu et al., 2014; Taskinen et al., 2017). We will refer to ssGBLUP and ssSNPBLUP as single-step models when both are considered.

Experience with single-step models have shown them to have at least as high or often higher prediction accuracy than pedigree-based animal models (Legarra et al., 2014; Misztal et al., 2013a, 2020). The striking feature of ssGBLUP is its ability to account for bias when selection is based on genotypes only (VanRaden, 2012; Mäntysaari et al., 2020; Misztal et al., 2020). However, the calculation of PEV by inverting the MME to compute reliabilities for ssGBLUP models often becomes either impossible or prohibitively expensive in the case of large numbers of animals. This limitation results from the genomic information producing large and dense matrix blocks in the MME that preclude the use of sparse matrix techniques.

Efficient computation of genomic accuracies for any model and data set continues to be an important research topic (Misztal et al., 2020). In this paper, we consider the computation of reliabilities for various models used in dairy cattle breeding, as a supplement to the review papers on EBV computations (Mäntysaari et al., 2020; Misztal et al., 2020). The objectives of this review are (1) to provide an overview of the different reliability methods in terms of modeling, development, and application in dairy cattle, with a special focus on genomic reliabilities, and (2) to discuss the problems faced in calculating reliabilities.

PEDIGREE-BASED MODELS

Sire Model

Sire models associate daughter phenotypes to sires genetic effects through a random sampling of genetic effects of each bull (Henderson, 1975). In the simplest sire model, each bull is assumed to be unrelated and the dams of the daughters represent a random sample of cows. The complexity of the MME in a sire model increases as the simplifying assumptions are relaxed by, for example, identifying pedigree relationships among sires and maternal grand sires, the addition of other random effects such as permanent environment or herd-year, or the addition of fixed effects such as parity.

Reliability of a sire EBV is often based on the reciprocal of the diagonal of sire model MME for EBV after the absorption of all fixed effects. For a simple sire model, ignoring the off-diagonal elements in the coefficient matrix of the MME has produced good reliability approximations (Meyer, 1989). When a sire relationship matrix is included in the MME, ignoring the off-diagonal elements of the coefficient matrix will lead to inaccurate approximations, due to exclusion of the information contained in those off-diagonals (Meyer, 1989).

Several methods have been proposed to account for the off-diagonal elements to give better approximations from the diagonal elements of the MME coefficient matrix (VanRaden and Freeman, 1985; Robinson and Jones, 1987). VanRaden and Freeman (1985) calculated bounds on the diagonal elements of the MME of the form $\mathbf{Z}'\mathbf{M}\mathbf{Z} + k\mathbf{I}$, where $\mathbf{I}$ is an identity matrix, $\mathbf{Z}$ is the incidence matrix of the random genetic effects; $\mathbf{M} = \mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'$, where $\mathbf{X}$ is the incidence matrix of the fixed effects; $(\mathbf{X}'\mathbf{X})^{-}$ denotes a generalized inverse of $\mathbf{X}'\mathbf{X}$; and $k = \sigma_e^2 / \sigma_s^2$ with the residual variance σ_e^2 and the sire genetic variance σ_s^2. The bounds were determined by using a partitioned matrix and positive definite forms. Two formulas for computing upper and lower bounds on accuracies, respectively, were D/(D + k) and D/(D + k + $\mathbf{d}'\mathbf{d}$/D), where D is the diagonal element of $\mathbf{Z}'\mathbf{M}\mathbf{Z}$ and $\mathbf{d}'\mathbf{d}$ is the sum of the squares of off-diagonal elements. Later, Robinson and Jones (1987) generalized this approach to a model with related sires.

Animal Model

The animal model enables simultaneous evaluation of sires, dams, and offspring, which prevents selection bias and improves the accuracy of prediction (Misztal and Wiggans, 1988). Unlike in a sire model where the observations on a daughter provide the information on her sire, even if relationships are ignored among those bulls, in an animal model, the relationships among animals (i.e., the covariances among random effects) are essential in the calculation of EBV and PEV. For animals without phenotypes, this is the only information contained in these EBV. For example, a bull will never have milk yield records, but the correlations imposed by the relationships with relatives could provide the highest genetic merit prediction in a population based on the performance of those relatives. Although the theory of animal model evaluation was developed about 60 years ago (Henderson, 1963), large-scale EBV prediction using an animal model only evolved in the early 1990s, leveraging the dramatic increase in the capacity of computers at that time. In routine animal model evaluations, the absorption technique is frequently used to reduce the size of the system of equations to be solved. Despite the use of this strategy, the MME coefficient matrix was still too

large to invert when millions of animals were involved in complex models. Schaeffer and Kennedy (1986) developed a matrix-free method, usually called "iteration on data," that enables obtaining solutions without explicitly setting up the MME. The iteration on data is computationally efficient in solving an animal model having millions of animals because it takes advantage of the sparsity of the MME.

Generally, the calculation of PEV in an animal model involves inverting MME of a size equal to the number of levels in fixed and in random effects in the model, and that constraint continued to limit applications of animal models for large data sets. Several approaches have been proposed to solve this problem and to approximate formulas for PEV and accuracy without the need to invert the MME coefficient matrix (Misztal and Wiggans, 1988; Liu et al., 2004; Tier and Meyer, 2004; Ducrocq and Schneider, 2007; Liu et al., 2010). Misztal and Wiggans (1988) pointed out that to achieve computing efficiency, the cost of PEV approximation should not exceed that of EBV evaluations.

Approximation of PEV for an Animal Model

Methods to approximate reliabilities for sire models are not suitable for animal models, because many important genetic relationships in an animal model are not considered in a sire model (Misztal and Wiggans, 1988). Therefore, several methods have been developed to approximate reliability in an animal model (Misztal and Wiggans, 1988; Boichard and Lee, 1992; Thompson et al., 1994). Misztal and Wiggans (1988) developed an iterative algorithm to approximate PEV in an animal model suitable for large data sets. They assumed that the loss of information in the estimation due to the nongenetic effects has been accounted for in a diagonal matrix $\mathbf{D}$ for every animal. A single observation for an animal having no loss of information due to estimation of nongenetic effects has a value of 1 on the diagonal. The diagonal elements of the matrix $\left(\mathbf{D}+\alpha\mathbf{A}^{-1}\right)^{-1}$, where $\alpha=\sigma_e^2/\sigma_a^2$, are assumed to be the same as in the submatrix of the inverse of the MME corresponding to the genetic effects of the animal along with its sire and dam, and $\mathbf{A}$ is the pedigree-based relationship matrix. Therefore, the total genetic information of an animal i, say b_i, can be decomposed in contributions from records and from relationships. They also demonstrated that the reliability of EBV for animal i can be calculated as $d_i/(d_i+\alpha)$, where d_i is the effective number of records for animal i. The authors approximated the contributions due to relationships for each relationship using the following iterative algorithm:

$$\begin{bmatrix} 1.5\alpha+q_s & 0.5\alpha & -\alpha \\ 0.5\alpha & 1.5\alpha+q_d & -\alpha \\ -\alpha & -\alpha & 2\alpha+q_i \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{(\alpha+b_s)} & * & * \\ * & \frac{1}{(\alpha+b_d)} & * \\ * & * & \frac{1}{(\alpha+b_i)} \end{bmatrix},$$

where q_i, q_s, and q_d are information on animal i, its sire, and its dam, respectively, without relationships considered, and b_i, b_s, and b_d are the total information corresponding to the same animals. Initially, the q vector is equal to the effective number of records, but q is updated by accumulating information from every animal using the relationship matrix information.

Misztal and Wiggans (1988) assessed their method with a simulated data set and compared the exact reliabilities obtained by inversion of the coefficient matrix and by approximation for both a full model and a model without fixed effects (herd-year-season). They reported correlation estimates of 0.996 with a single-trait repeatability model for a simulated data, which had 2,315 records from 1,000 daughters of 40 sires, and only 100 management groups over 3 generations. A major advantage of this method is that the memory requirement is restricted to 3 variables per animal. The iterative process only involves reading the data file once and the pedigree file once at each iteration. Later, VanRaden and Wiggans (1991) developed a similar method using the same contributions as in Misztal and Wiggans (1988), but expressed in daughter equivalents.

Meyer (1989) proposed an approach for approximating the reliability for a single-trait animal model. Let $\mathbf{C}$ be the usual MME and $\mathbf{C}^{-1}$ its inverse. A reduced set of elements in $\mathbf{C}^{-1}$ can be approximated by inverting the 4×4 submatrix, where the dimension is from an individual, its parents, and the fixed effect class containing the observation. The accuracy of this approximation depends on the number and magnitude of nonzero off-diagonal elements of $\mathbf{C}$ relevant for this animal that were ignored. The number of these elements could be substantial for an animal with many progenies. The author suggested the following 3-step procedure: (1) adjust the diagonal of each animal with records for limited subclass sizes, (2) accumulate adjustments to the diagonals of the parents for limited information on progeny, and (3) adjust diagonals of progeny for the adjusted diagonals of their parents, to avoid double counting the animal considered.

Multi-Trait Animal Models

Multi-trait animal models can incorporate valuable information from correlated traits on an animal and its relatives for the computation of EBV (Meyer, 1989). When genetic correlations between traits are moderate or high, reliability approximations from single-trait models may give poor estimates for reliabilities using analogous multi-trait model evaluations. Numbers of daughters that may be suitable for use in a single-trait sire model can be inappropriate because observations from correlated traits on a daughter do not provide the same amount of information on a sire as 2 measurements from different daughters. On the other hand, information of a second correlated trait on the same animal can generate more information on the EBV for that animal for the primary trait than if it were observed on its ancestors or its progeny (Meyer, 1989).

In the method suggested by Tier and Meyer (2004), the PEV and prediction error covariances of EBV were approximated by extending the 3 primary steps of Meyer (1989) as follows: (1) determine the record information for every animal, (2) accumulate information on the progeny and further descendants of the animal from the youngest to the oldest, and (3) gather the accumulated record information from parents, ancestors, and collateral relatives by traversing the pedigree from the oldest to the youngest animal. The following 2 additional steps are required: (1) removing the animal's own contribution to avoid double counting, and (2) combining the contributions from all information sources (parents, progeny, and its own records). The final block for each animal, say $\mathbf{F}_i$, can be written as $\mathbf{F}_i = \sum_{j=1}^{t_i}\mathbf{E}_j^* + \mathbf{D}_i + \sum_{l=1}^{p_i}\mathbf{E}_l$, where $\mathbf{E}_j^*$ is the contribution of the parents adjusted for animal i from step 3; $\mathbf{D}_i$ has the amount of information from its own records; $\mathbf{E}_l$ has the progeny contributions of animal i from step 2; t_i is the number of known parents for animal i; and p_i is the number of progeny of animal i.

The PEV of animal i can be expressed as $\mathbf{PEV}_i = diag\left\{\left(\mathbf{F}_i + \mathbf{G}_0^{-1}\right)^{-1}\right\}$, where $diag\{\}$ is a square matrix with the diagonal elements the same as the diagonal elements of the original matrix and all off-diagonals elements are zero, and $\mathbf{G}_0^{-1}$ is the variance-covariance matrix of additive genetic effects. See Tier and Meyer (2004) for further details. The method provides good approximations of PEV in a multi-trait model and has been shown to be efficient even for large populations. However, the method tends to overestimate reliabilities for animals with high reliabilities.

Exploiting Sparsity in the Animal Model

In addition to the calculation of PEV, the diagonal elements of the inverse of the coefficient matrix of the MME are often needed in the estimation of variance components by REML (VanRaden and Freeman, 1985). For example, PEV are needed in the first derivatives of the REML log-likelihood used by expectation maximization and average information REML approaches. These computational approaches are regarded as expensive when large data sets are used. Various sparse matrix techniques have been proposed to calculate PEV values required for the first derivative (Misztal and Perez-Enciso, 1993; Johnson and Thompson, 1995). Yet, it is worth noting that the sparse inverse generally includes a small fraction of the nonzero elements of the full inverse.

Exploiting sparse Cholesky factorization, the coefficient matrix of MME can be presented as $\mathbf{C} = \mathbf{U}'\mathbf{D}\mathbf{U}$, where the MME coefficient matrix $\mathbf{C}$ is expressed using the upper triangle matrix $\mathbf{U}$ and the diagonal matrix $\mathbf{D}$. Indeed, based on the method of Takahashi et al. (1973), Misztal and Perez-Enciso (1993) showed that the inverse of $\mathbf{C}$ noted $\mathbf{C}^{-1}$ can be written as $\mathbf{C}^{-1} = \mathbf{D}^{-1}\mathbf{U}'^{-1} + (\mathbf{I} - \mathbf{U})\mathbf{C}^{-1}$. They used the following equations to obtain the sparse matrix $\mathbf{C}^{-1}$: $c_{ii} = d_{ii}^{-1} - \sum_{k=i+1}^{n} u_{ik}c_{ik}$ and $c_{ij} = c_{ji} = -\sum_{k=j+1}^{n} u_{jk}c_{ik}$, where the elements of the $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{U}$ matrices are referred by indices $i = n, n-1, \ldots, 1$, and $j = i-1, i-2, \ldots, 1$. The advantage of this method is that the elements of $\mathbf{C}^{-1}$ are calculated only for corresponding nonzero elements of $\mathbf{U}$. For easy understanding see the numerical example given in Misztal and Perez-Enciso (1993). It is also worth noting that this type of method, based on the sparse inverse, can be used to compute PEV in an animal model. However, the total computing time heavily depends on the memory requirement that continues to present a major limitation in the case of large populations.

An alternative approach is to approximate the needed functions or elements of PEV in expectation maximization and average information REML by Monte Carlo (**MC**) simulation as in Matilainen et al. (2013), which can be applied in computation of PEV for reliability as well. However, although the memory limitations have been lifted by the use of the MC approach, each MC sample requires solving MME, which may be infeasible in practice as the required number of MC samples may be too large. Monte Carlo sampling procedures for calculating PEV have also been described in Hickey et al. (2008) and Hickey et al. (2009). Hickey et al. (2009)

found that PEV approximations using MC sampling are affected by the formulation used to calculate PEV and the level of exact PEV. However, they showed that the difference between the formulations tends to be small when the number of MC samples increases. See Hickey et al. (2009) for further details.

GENOMIC MODELS

Computation of Reliabilities in the Era of Genomic Prediction

Genomic selection uses genomic information, which needs to be considered in the evaluation models as well as in the calculation of reliabilities. Initially the number of genotyped animals was quite small. However, the rapid increase in the number of genotyped animals due to the decreasing cost of genotyping and the growing use of genomic selection, has led to new computational challenges to handle the vast amount of genomic information. In this section, we will review the different approaches for approximating reliabilities for models that use genomic information only.

Equivalence Between GBLUP and SNPBLUP Models

Genomic relationship-based BLUP or SNPBLUP models have been widely used for routine genomic evaluations in animal breeding. Both models generate equal EBV and PEV. Based on the paper of Meuwissen et al. (2001), a marker effect or SNPBLUP model can be written as $\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{e}$, where $\mathbf{y}$ is the vector of trait phenotypes, $\mathbf{Z}$ is the matrix of gene content, $\mathbf{b}$ is the vector of fixed effects, $\mathbf{g}$ is the vector of marker effects, and $\mathbf{e}$ is the vector of random residual effects. It is assumed that there are n genotyped animals with m SNP markers (i.e., $\mathbf{Z}$ has size n by m). The marker effects are assumed to be uncorrelated with variance $\mathrm{Var}(\mathbf{g}) = \mathbf{I}\sigma_{\mathrm{g}}^2$ and $\boldsymbol{Var}(\mathbf{e}) = \mathbf{I}\sigma_{\mathrm{e}}^2 = \mathbf{R}$.

Breeding values can be calculated from SNPBLUP model as $\mathbf{u} = \mathbf{Zg}$. Using selection index equations, VanRaden (2007, 2008) showed that the SNPBLUP model is equivalent to the GBLUP model, which directly models the breeding values $\mathbf{u}$ but not marker effects $\mathbf{g}$. The breeding values are assumed to have variance $\mathrm{Var}(\mathbf{u}) = \mathbf{G}\sigma_{\mathrm{u}}^2$, where $\mathbf{G}$ is the genomic relationship matrix and σ_{u}^2 is the genetic variance. The $\mathbf{G}$ matrix is $\mathbf{ZZ}'\Big/\sum_{i=1}^{m} p_i\left(1-p_i\right)$, with p_i as the frequency of the ith SNP marker, and $\sigma_{\mathrm{u}}^2 = \sigma_{\mathrm{g}}^2\sum_{i=1}^{m} p_i\left(1-p_i\right)$. Using the MME (Henderson, 1984) and the Woodbury matrix identity, Strandén and Garrick (2009) proved the equivalence between genomic effect and marker effect models for the genotyped reference animals. As shown by Henderson (1984), 2 models are linearly equivalent if and only if their fixed effect estimates are equal and their phenotypic variance matrices are equal. Fernando et al. (2014) showed that SNPBLUP and GBLUP provide equivalent predictions. The MME size increases with the number of genotyped animals n in GBLUP but is bounded by the number of SNP markers m in SNPBLUP.

Typically, SNP markers are unable to completely capture the genetic variation due to the incomplete linkage disequilibrium between the QTL and the genome-wide markers (Hayes et al., 2009; Liu et al., 2011; Ben Zaabza et al., 2020a). Inclusion of the residual polygenic ($\mathbf{RPG}$) effect in the aforementioned models would generally result in higher accuracy (Liu et al., 2011), but would also increase the size of the MME by the number of genotyped animals. The equivalence between GBLUP and SNPBLUP models with a fitted RPG has been demonstrated with RPG as a separate effect in SNPBLUP (Liu et al., 2016), and with RPG integrated in the marker genotype matrix (Ben Zaabza et al., 2020a).

Computational Cost of Reliability Calculation in Multistep Genomic Models Using Either GBLUP or SNPBLUP

Calculation of EBV reliability requires the elements from the inverse of the coefficient of the MME. In SNPBLUP model, reliability for an animal i can be calculated as $1-\mathbf{z}_i\mathbf{C}^{gg}\mathbf{z}_i^{'} \Big/ \left(\mathbf{G}_{ii}\sigma_u^2\right)$, where $\mathbf{z}_i$ is the row i in the marker matrix $\mathbf{Z}$, and $\mathbf{C}^{gg}$ is the submatrix corresponding to the marker effects in the inverse of SNPBLUP MME. The computationally most expensive operation with SNPBLUP is the inversion of the coefficient matrix of the MME, which has an approximately cubic cost with the number of SNP markers [i.e., $\mathrm{O}(m^3)$]. In GBLUP, reliability for an animal i can be calculated as $1-\mathbf{PEV}_i \Big/ \left(\mathbf{G}_{ii}\sigma_u^2\right)$, where $\mathbf{PEV}_i$ is the diagonal element i in the PEV matrix $\mathbf{C}^{\mathrm{uu}}$ obtained from the inverse of GBLUP MME. The inversion of GBLUP MME has a cubic cost (and quadratic memory) for genotyped animals [i.e., $\mathrm{O}(n^3)$], which is prohibitively expensive for a large population of genotyped animals. Further, the inversion of the $\mathbf{G}$ matrix is computed beforehand and requires almost (n^2) memory and (n^3) computations, which is not computationally feasible due to the density and large size of the genomic relationship matrix $\mathbf{G}$ in case of a high number of genotyped animals. Thus, when the number of genotyped animals exceeds the

number of SNP markers (i.e., $n > m$), SNPBLUP is more efficient because it does not involve calculation of the $\mathbf{G}$ matrix or its inverse (Strandén and Garrick, 2009; Fernando et al., 2014). For example, when the number of genotyped animals (n) is higher than the number of marker covariates (m), the computing cost required to invert the MME for GBLUP over SNPBLUP is approximately proportional to $(n/m)^3$ (Ben Zaabza et al., 2020a).

Computational Cost of GBLUP/SNPBLUP with Fitted RPG Effect

Including the RPG effect in multistep genomic BLUP is known to improve the reliability of the prediction and reduce bias (Guarini et al., 2018). In case the SNP markers fail to describe all the additive genetic variance of the trait, widely used sires with high genomic enhanced breeding values (**GEBV**) reliabilities will have slightly different prediction estimates compared with those from traditional pedigree analysis, unless RPG effect is included in the applied genomic model (Golden et al., 2016). In the GBLUP model, the RPG effect can be accounted for by an additional RPG or by combining the genomic and polygenic breeding value effects in the genomic relationship matrix, which can be written as

$$\mathbf{G}_w = (1-w)\mathbf{Z}_c\mathbf{Z}_c^{'} + w\mathbf{A}_{22},$$

where $\mathbf{Z}_c$ is the VanRaden (2008) method 1 centered and scaled genotype matrix and w is the RPG proportion. The RPG proportion w is the weight of the RPG effect and can be seen as the proportion of genetic variance not captured by SNP markers. If we assume a GBLUP model as $\mathbf{y} = 1_n\mu + \mathbf{u} + \mathbf{e}$, the computation of **PEV** requires the following 2 inversions: (1) inversion of the genomic relationship matrix $\mathbf{G}_w$ of size n, and (2) inversion of the MME of size $n + 1$. This is not feasible for a large population of genotyped animals, because the computing time increases cubically and the storage (memory) quadratically with the number of genotyped animals. Typically, the calculation of reliability in GBLUP models is dominated by the calculation of the inverse genomic relationship matrix, which includes making and inverting the matrix. This becomes clearer with the higher the number of genotyped animals.

In SNPBLUP with a fitted RPG effect, the marker effect coefficients can be augmented with RPG coefficients for the genotyped animals. In other words, if we consider an SNPBLUP model equivalent to the GBLUP model as $\mathbf{y} = 1_n\mu + \mathbf{Zg} + \mathbf{e}$, where $\mathbf{Z} = \left[\sqrt{(1-w)}\,\mathbf{Z}_c \quad \sqrt{w}\mathbf{L}\right]$, the square matrix $\mathbf{L}$ is the Cholesky decomposition of $\mathbf{A}_{22}$ (i.e., $\mathbf{A}_{22} = \mathbf{LL}'$), and $\mathbf{g}$ is m_n by 1 vector of pseudomarker effects, $m_n = m + n$. The computation of the PEV of SNP markers requires the inversion of the MME of size $m + n$, which is infeasible for large-scale genetic evaluation schemes involving millions of animals. Ben Zaabza et al. (2020a) reported a computing time of 18.6 h to invert the $\mathbf{G}$ matrix for 222,619 genotyped animals using 10 CPU threads. Extrapolating for 1 million genotyped animals, the equivalent computing time would be almost 70 d. The authors proposed several strategies to overcome the computational burden in multistep genomic model reliability estimation.

Approximating Reliabilities in SNPBLUP/GBLUP Models

MC-SNPBLUP. Heavy computation is inevitable in both large-scale GBLUP and SNPBLUP models when the RPG effect is included. In large-scale genomic evaluations, an approximation method based on reducing the size of the MME can be a practical solution. To this end, Ben Zaabza et al. (2020a,b) proposed a MC approach to estimate reliability for the SNPBLUP model with an RPG effect. The method allows reduction of the MME size, which determines the computational cost in the case of a large number of genotyped animals. The authors approximated the $\mathbf{A}_{22}$ matrix using MC sampling; that is, they applied MC samples instead of the Cholesky decomposition $\mathbf{L}$ for the RPG effect, as follows:

$$\widehat{\mathbf{A}_{22}} = \frac{1}{n_{\mathrm{MC}}}\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_{n_{\mathrm{MC}}}\end{bmatrix}\begin{bmatrix}\mathbf{a}_1^{'} \\ \vdots \\ \mathbf{a}_{n_{\mathrm{MC}}}^{'}\end{bmatrix} = \mathbf{U}_{22}\mathbf{U}_{22}^{'},$$

where $\boldsymbol{a}_i \sim \mathrm{N}(0, \mathbf{A}_{22}), \quad i = 1, \ldots, n_{\mathrm{MC}},$ and $\mathbf{U}_{22} = 1/\sqrt{n_{\mathrm{MC}}}\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_{n_{\mathrm{MC}}}\end{bmatrix}$.

Thus, the $\mathbf{G}_w$ matrix can be reformulated as $\mathbf{G}_w^* = (1-w)\mathbf{Z}_c\mathbf{Z}_c^{'} + w\mathbf{U}_{22}\mathbf{U}_{22}^{'}$. Subsequently, the $\mathbf{G}_w^*$ matrix can be written as $\mathbf{G}_w^* = \mathbf{SS}'$, where $\mathbf{S} = \begin{bmatrix}\mathbf{S}_{\mathrm{Z}} & \mathbf{S}_{\mathrm{U}}\end{bmatrix}$, with $\mathbf{S}_{\mathbf{Z}} = \sqrt{1-w}\mathbf{Z}_c$ and $\mathbf{S}_{\mathrm{U}} = \sqrt{w}\mathbf{U}_{22}$. Ben Zaabza et al. (2020a) used $\mathbf{G}_w^*$ to create the following equivalent MC-based SNPBLUP:

$$\mathbf{y} = 1_n\mu + \mathbf{Sg}_s + \mathbf{e},$$

where $\mathbf{g}_s$ is the vector of ($m + n_{MC}$) pseudomarker effects, of which m are due to SNP markers and n_{MC} are

dummy effects for the RPG breeding values. Thus, the MME can be written as

$$\begin{bmatrix} 1_n'\mathbf{R}^{-1}1_n & 1_n'\mathbf{R}^{-1}\mathbf{S} \\ \mathbf{S}'\mathbf{R}^{-1}1_n & \mathbf{S}'\mathbf{R}^{-1}\mathbf{S}+\mathbf{I}\sigma_u^{-2} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \widehat{\mathbf{g}_s} \end{bmatrix} = \begin{bmatrix} 1_n'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{S}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}.$$

The inversion of MME for MC-SNPBLUP incurs a cubic cost and requires quadratic storage space for $(m + n_{\text{MC}})$. Savings in memory requirements and computing time are due to replacing the number of genotyped animals by a smaller number of MC samples. The authors have shown that the MC-SNPBLUP model has computational advantages over exact SNPBLUP, with a fitted RPG effect, in that it gives good reliability approximations with several MC samples much smaller than the number of genotyped animals, when the number of genotyped animals is large. Consequently, the total computing time is substantially decreased.

The MC-SNPBLUP method was tested in a beef cattle carcass conformation evaluation involving a population of almost 223,000 genotyped animals and 13.35 million pedigree animals, and compared with the exact GBLUP model reliabilities under different MC sample sizes and RPG proportions. The authors reported correlations of 99% between the 2 model reliabilities with 20,000 MC samples. The corresponding computing time was around 250 min, which is less than 1/10th of that for GBLUP reliabilities. For more details, see Ben Zaabza et al. (2020a). The computational cost needed to invert the MME was approximately proportional to $\left(\frac{n}{m+n_{\text{MC}}}\right)^3$ for GBLUP over MC-SNPBLUP, indicating that the latter is computationally less demanding than GBLUP model when $n \gg (n + n_{MC})$.

Full MC Sampling for SNPBLUP. Expanding the SNPBLUP MME to include the RPG effect by using the MC approach offers computational advantages over the exact SNPBLUP, by reducing the MME size to $(m + n_{\text{MC}})$ instead of $(m + n)$, with n_{MC} always smaller than n. However, the computational cost using a high-density SNP marker chip is still expensive in the case of many SNP markers and genotyped animals. To overcome this issue, Ben Zaabza et al. (2021) extended the MC-SNPBLUP model to use MC samples for both SNP markers and the RPG effect. Because all genetic effects were now replaced by MC samples, the method was putatively called full-MC-SNPBLUP. The full-MC-SNPBLUP model equation is

$$\mathbf{y} = 1_n\mu + \mathbf{U}\mathbf{s} + \mathbf{e},$$

where $\hat{\boldsymbol{s}}$ are solutions for n_{MC} random pseudogenetic effects, with n_{MC} as the number of MC samples and $\mathbf{U}$ as an n by n_{MC} matrix of the MC samples. It is assumed that $\mathbf{s} \sim \text{N}\left(0, \mathbf{I}\sigma_u^2\right)$ and $\mathbf{e} \sim \text{N}\left(0, \mathbf{R}\sigma_e^2\right)$, where σ_u^2 and σ_e^2 are the genetic and residual variance, respectively. Each column i in the $\mathbf{U}$ matrix is an MC-simulated sample $\mathbf{u}_i = \mathbf{Z}_c\mathbf{g}_i + \mathbf{a}_i$, where it is assumed that $\mathbf{g} \sim \text{N}\left[0, (1-w)\mathbf{I}\right]$ and $\mathbf{a} \sim \text{N}\left(0, w\mathbf{A}_{22}\right)$. Thus, the MME for full-MC-SNPBLUP is

$$\begin{bmatrix} 1_n'\mathbf{R}^{-1}1_n & 1_n'\mathbf{R}^{-1}\mathbf{U} \\ \mathbf{U}'\mathbf{R}^{-1}1_n & \mathbf{U}'\mathbf{R}^{-1}\mathbf{U}+\mathbf{I}\sigma_u^{-2} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{s}} \end{bmatrix} = \begin{bmatrix} 1_n'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{U}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}.$$

The full-MC-SNPBLUP is a sequential procedure, which comprises the following 4 main steps: (1) making MC samples for SNP markers, (2) making MC samples for the RPG effect, (3) making the MME matrix, and (4) inverting the MME matrix. Even though the computing time for the first 2 steps increases linearly along with the number of MC samples, generating the MC samples for the RPG effect can be computationally more expensive than for the SNP marker effects. The cost of making the MME is $n(n_{\text{MC}})^2$. However, the cost of inverting the MME is a cubic function $\text{O}[(n_{\text{MC}})^3)]$ of the number of MC samples, and hence, the computing cost for making the MME becomes more expensive than that for MME inversion when the number of genotyped animals surpasses the number of MC samples. Considering the required computing time as well as the unbiasedness and accuracy of full-MC-SNPBLUP, this model outperformed the MC-SNPBLUP method, especially when high RPG proportion was used. However, the computing time is largely determined by the number of MC samples needed to obtain a high accuracy. Further, the application of the MC approach in a multi-trait model remains challenging. For more details, see Ben Zaabza et al. (2021).

SINGLE-STEP GENOMIC MODELS

Computation of reliabilities for the single-step models inherits the challenges from both the pedigree-based and genomic models. The following sections give the current status in the computation of approximate reliabilities for large data sets with many genotyped individuals.

Reliabilities for Single-Step GBLUP Models

When a part of the population is not genotyped, the genomic relationship matrix $\mathbf{G}$ can be combined with the pedigree-based additive relationship matrix $\mathbf{A}$ into

a unified relationship matrix $\mathbf{H}$ (Legarra et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). Let subscripts 1 and 2 refer to nongenotyped and genotyped animals, respectively. Then, the pedigree-based relationship matrix $\mathbf{A}$ can be written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},$$

where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ represent the pedigree relationship between nongenotyped and between genotyped animals, respectively, and the submatrix $\mathbf{A}_{12}$ is the pedigree relationship between nongenotyped and genotyped animals. The $\mathbf{H}$ matrix can be interpreted as an extended matrix of $\mathbf{A}$ to accommodate $\mathbf{G}$, and can be written as

$$\mathbf{H} = \begin{pmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{pmatrix},$$

and its inverse as

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{pmatrix}.$$

Clearly the augmented relationship matrix $\mathbf{H}$ is a dense matrix, but its inverse $\mathbf{H}^{-1}$ appears to have a simple form with dense blocks within otherwise sparse matrix. The $\mathbf{H}^{-1}$ matrix includes the inverse of the genomic relationship $\mathbf{G}$, which is often singular and needs to be blended to $\mathbf{A}_{22}$ to be of full rank. A modified genomic relationship matrix can be written (obtained) as $\mathbf{G}_w = (1 - w)\mathbf{G} + w\mathbf{A}_{22}$. The most expensive operation in ssGBLUP model reliabilities, as presented by Aguilar et al. (2010) and Christensen and Lund (2010), is making and then inverting the MME. Both operations have an approximately cubic cost with the number of equations (animal plus fixed effects). Moreover, ssGBLUP requires the inversion of $\mathbf{G}$ and $\mathbf{A}_{22}$ matrices. The corresponding computations increase cubically and the storage requirement quadratically with the number of genotyped animals.

Reliability Approximation Using the Direct Inversion of G Matrix

Misztal et al. (2013b) proposed an approximation method to calculate reliabilities in ssGBLUP models. Their approach is an extension of the approximation algorithm of Misztal and Wiggans (1988). Misztal et al. (2013b) showed that PEV for an animal i can be written as $1/\left(\alpha + d_i^r + d_i^p + d_i^g\right)$, where d_i^r, d_i^p, and d_i^g are contributions due to pedigree, phenotype, and genomic information, respectively, in terms of effective daughters or observations. The authors demonstrated that PEV for all animals can also be written as $diag\left\{\left[\mathbf{D}^r + \mathbf{D}^P + \left(\mathbf{I} + \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}\right)\alpha\right]^{-1}\right\}$. The $\mathbf{G}$ matrix accounts for genomic information and the $\mathbf{A}_{22}$ matrix provides an adjustment to ensure that the relationship information contained in $\mathbf{G}$ and $\mathbf{A}_{22}$ will not be double counted.

The algorithm developed by Misztal et al. (2013b) for reliability approximation with genomic information includes the following 3 steps: (1) approximating reliabilities in an animal model (i.e., no genomic information used), (2) transforming the already calculated reliabilities for genotyped animals into the effective number of records, and (3) calculating LHS_{uu}, where uu denotes the block of the left-hand side (**LHS**) of the MME, as $\left\{\left[\mathbf{D}^r + \mathbf{D}^P + \left(\mathbf{I} + \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}\right)\alpha\right]^{-1}\right\}$, and then genomic reliability for animal i as $1 - \alpha\text{LHS}_{uu}^{ii}$, where LHS_{uu}^{ii} is the diagonal value i in LHS_{uu}. An additional step, involving the adjustment of reliabilities for nongenotyped animals in case these are functions of the reliabilities of genotyped animals, is optional. Misztal et al. (2013b) also suggested a simpler form of the above algorithm by ignoring the off-diagonal elements of $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$, so that $\text{LHS}_{uu} \approx \left\{\left[\mathbf{D}^r + \mathbf{D}^P + (\mathbf{I} + \text{diag}\{\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}\})\alpha\right]^{-1}\right\}$. The second approximation enjoys a significant computational advantage over the first one, whereas the first approximation performs better. Its drawback is that it requires an explicit calculation of the inverse of $\mathbf{G}$ and $\mathbf{A}_{22}$ matrices, which is not computationally feasible for a large number of genotyped animals. Both methods were found to be viable for populations of up to 100,000 genotyped animals.

Reliability Approximation Based on a Multistep Approach

To make genomic reliabilities comparable across countries and consistent with pedigree-based animal model reliabilities, Liu et al. (2017) developed a standardized statistical procedure for approximating genomic reliabilities. The authors defined the following features which an accurate method for calculating genomic reliabilities has to satisfy: (1) ability to account for the RPG effect, (2) feasibility irrespective of the number of genotyped animals, (3) applicability to a single-step genomic model, (4) feasibility for routine genomic evaluations, and (5) compatibility of the approximated genomic reliability with the genomic vali-

dation PEV. The authors based their method on a pure multistep SNPBLUP model, which has the advantage of a manageable and stable MME size, as the model does not rely on the number of genotyped animals.

The Interbull method by Liu et al. (2017) involves a total of 6 steps, which are as follows: (1) calculating SNP reliability with the assumption that all the genetic variability is captured by the SNP markers (i.e., no RPG effect was included into the model), (2) approximating the reliability of genomic EBV or direct genomic value as $R_{DGV} = (1 - k)R_{imp}R_{SNP}$, where k is the proportion of RPG variance, R_{imp} is the reliability of genotype imputation, and R_{SNP} is the reliability of SNP markers, (3) adjusting the theoretical reliabilities to the realized reliabilities using genomic validation results. It is worth noting that the theoretical reliabilities tend to be inflated for the young genotyped candidate animals, (4) calculating the genomic effective daughter contribution (**EDC**) gain, namely the difference between the realized genomic EDC and the conventional EDC, (5) transferring the additional information of genomic origin to nongenotyped animals, and (6) calculating final reliabilities enhanced with genomic information. For more details, see Liu et al. (2017).

Reliability Approximation Using Algorithm of Proven and Young Animals

Bermann et al. (2021) presented a method for calculating PEV in the GBLUP model using an approximate inverse of $\mathbf{G}$ computed a recursive algorithm developed by Misztal et al. (2014), called the algorithm of proven and young animals (**APY**). The authors extended this method for approximated PEV of GBLUP to the ssGBLUP model to approximate reliabilities for single-trait and multi-trait ssGBLUP by using a procedure based on effective record contributions (**ERC**), as shown in Liu et al. (2017). It is worth noting that the APY method calculates the approximated $\mathbf{G}^{-1}$, hereinafter called $\mathbf{G}^{-1}_{APY}$, for a large number of noncore genotyped animals based on the direct $\mathbf{G}^{-1}$ for a minimum number of core genotyped animals, assuming that these core animals represent most of the chromosome segments in the genome (Tsuruta et al., 2021). The proposed $\mathbf{G}^{-1}_{APY}$ can be written as

$$\mathbf{G}^{-1}_{APY} = \begin{bmatrix} \mathbf{I} & -\mathbf{P}_{cn} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{G}^{-1}_{cc} & 0 \\ 0 & \mathbf{M}^{-1}_{nn} \end{bmatrix} \begin{bmatrix} \mathbf{I} & 0 \\ -\mathbf{P}_{nc} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{G}^{cc} & \mathbf{G}^{cn} \\ \mathbf{G}^{nc} & \mathbf{M}^{-1}_{nn} \end{bmatrix}.$$

Let subscripts c and n refer to the core and noncore animals, respectively; $\mathbf{P}_{cn} = \mathbf{G}_{nc}\mathbf{G}^{-1}_{cc}$ and $\mathbf{M}_{nn} = diag\left\{\mathbf{G}_{nn} - \mathbf{G}_{nc}\mathbf{G}^{-1}_{cc}\mathbf{G}_{cn}\right\}$.

The $\mathbf{G}^{-1}_{APY}$ has a cubic cost and a quadratic memory requirement for core genotyped animals and a linear cost and memory requirement for noncore genotyped animals. In other words, when the number of genotyped animals increases, the number of computations used in $\mathbf{G}^{-1}_{APY}$ increases linearly.

In their study, Bermann et al. (2021) exploited the sparse presentation of $\mathbf{G}^{-1}_{APY}$ to calculate PEV in GBLUP as the diagonal of $\left(\mathbf{D}+\mathbf{G}^{-1}_{APY}\right)^{-1}$, where $\mathbf{D}$ is a diagonal matrix having weights. The $\mathbf{D}$ matrix can be written as $diag\left\{\mathbf{W}'\left[\mathbf{I}-\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\right]\mathbf{W}\right\}$, where $\mathbf{X}$ and $\mathbf{W}$ are design matrices for the fixed and random effect, respectively. Their algorithm clearly does not require setting up the MME, but only approximating weights to be added to the diagonal elements of the $\mathbf{G}^{-1}_{APY}$ matrix. The authors used the APY-GBLUP to approximate reliabilities in a single-trait ssGBLUP using a multistep procedure similar to those developed by Liu et al. (2017) and Edel et al. (2019). Further, they extended both the single-trait APY-GBLUP and APY-ssGBLUP to multi-trait APY-GBLUP and APY-ssGBLUP models, respectively, by adjusting the single-trait reliabilities using the genetic and residual covariance matrices across traits as in Strabel et al. (2001).

In their multistep procedure, Bermann et al. (2021) showed that steps 4 to 6 are the most time-consuming steps of their algorithm, with a quadratic cost for core genotyped animals and a linear cost for noncore genotyped animals both for step 4, which involves calculating $\mathbf{G}^{cc}$ as $\mathbf{G}^{cc} - \mathbf{G}^{cn}\left(\mathbf{G}^{cn}\mathbf{M}_{nn}\right)'$, and for step 6, which involves calculating the matrix product $-\mathbf{G}^{cc}\left(\mathbf{G}^{cn}\mathbf{M}_{nn}\right)$. Step 5, however, which requires inverting the matrix $\mathbf{G}^{cc}$, has a cubic cost for core genotyped animals. The gains in computing cost and memory requirement are due to ignoring the computation and storage for both the blocks of noncore × noncore (except diagonal elements) for both $\mathbf{G}$ and $\mathbf{G}^{-1}_{APY}$ (Misztal, 2016).

The computations by the APY-based ssGBLUP reliabilities seem to be efficient in terms of computing times, but the success of this method can be dependent on the choice of appropriate core animals. Tsuruta et al. (2021) reported that some of the chosen core animals might give redundant information (i.e., collinearity between animals), which would require increasing the number of core animals representing all independent chromosome segments in the genome. The authors stated that even with the APY method, the computational cost for large-scale genomic evaluations is still high when all genotyped animals are included. They concluded that this could prove to be a bottleneck in conducting frequent evaluations.

Reliability Approximation Based on the Single-Step Bayesian Regression Method

The single-step Bayesian regression (**ssBR**) model for ssSNPBLUP was presented by Fernando et al. (2014). The MME for the ssBR is

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{M} & \mathbf{X}_1'\mathbf{R}_1^{-1}\mathbf{Z}_1 \\ \mathbf{M}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{M}'\mathbf{R}^{-1}\mathbf{M}+\mathbf{I}\frac{1}{\sigma_g^2} & \mathbf{M}_1'\mathbf{R}_1^{-1}\mathbf{Z}_1 \\ \mathbf{Z}_1'\mathbf{R}_1^{-1}\mathbf{X}_1 & \mathbf{Z}_1'\mathbf{R}_1^{-1}\mathbf{M}_1 & \mathbf{Z}_1'\mathbf{R}_1^{-1}\mathbf{Z}_1+\mathbf{A}^{11}\frac{1}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \\ \hat{\boldsymbol{\varepsilon}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{M}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_1'\mathbf{R}_1^{-1}\mathbf{y}_1 \end{bmatrix},$$

where $\mathbf{M}$ is the matrix of imputed and observed genotypes of all animals in the pedigree, $\mathbf{M}_1$ is a submatrix of imputed genotypes of nongenotyped animals with phenotype, and $\hat{\mathbf{b}}$, $\hat{\mathbf{g}}$, and $\hat{\boldsymbol{\varepsilon}}$ are solutions for fixed effects, marker effects, and imputation residual effects, respectively.

The ssBR model was originally implemented using Markov chain MC (**MCMC**), which allows computation of PEV from the MCMC chain. Gao et al. (2018) illustrated that the PEV from the MCMC chain was close to the correct value obtained by direct inversion. However, MCMC was considered to be computationally demanding for large MME.

Edel et al. (2019) described the theoretical aspects related to calculations of reliability of single-step predictions. Their derivation consists of calculating the PEV for marker effects, and then approximating the PEV for the imputation residuals. The authors demonstrated that the PEV for SNP marker effects, taking into account the contribution of the imputation residual, can be written as

$$\left[\left(\mathbf{M}'\mathbf{R}^{-1}\mathbf{M}+\mathbf{I}\frac{1}{\sigma_g^2}\right)-\mathbf{M}_1'\mathbf{R}_1^{-1}\mathbf{Z}_1\left(\mathbf{Z}_1'\mathbf{R}_1^{-1}\mathbf{Z}_1+\mathbf{A}^{11}\frac{1}{\sigma_a^2}\right)^{-1}\mathbf{Z}_1'\mathbf{R}_1^{-1}\mathbf{M}_1\right]^{-1}.$$

They also showed that the PEV of the imputation term of $\hat{\varepsilon}$ can be approximated by a diagonal matrix (i.e., by dividing the diagonal of $\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$ by the diagonal of $\mathbf{A}_{11}$). For more details, see Edel et al. (2019).

Edel et al. (2019) further presented a second approximation based on limiting the number of nongenotyped animals to those with a substantial contribution to the total amount of information. The consequent reduction in animal number had the following 3 benefits: (1) reducing the dimensionality of $\mathbf{Z}_1'\mathbf{R}_1^{-1}\mathbf{Z}_1+\mathbf{A}^{11}\frac{1}{\sigma_a^2}$, (2) lowering the cost of imputing and storing of genotypes of nongenotyped animals, and (3) reducing the size of the $\mathbf{M}$ matrix to equal the sum of genotyped animals plus those nongenotyped animals with substantial information contributions, instead of genotyped and all nongenotyped animals. The authors also proposed to combine the 2 approximation approaches to calculate reliability of genotyped animals for large populations of dairy cattle. Similar to Liu et al. (2017), they suggested a framework to approximate reliability for nongenotyped animals by propagating the information of genotyped animals by reverse-reliability estimation to nongenotyped animals. It is noteworthy that the Liu et al. (2017) method does not use the imputed genotypes but the observed genotypes only. One of the main advantages of the approach of Edel et al. (2019) is that there is no need to invert the $\mathbf{G}$ matrix, which becomes infeasible as the number of genotyped animals increases. Also, the calculation of imputed genotypes for all nongenotyped animals can easily be parallelized, which significantly decreases the total computing time. Even though the Edel et al. (2019) approach can be an efficient method to overcome the computational problems encountered with calculation of the exact ssBVBLUP (or regular ssGBLUP) reliabilities, it is worth mentioning that the diagonal approximation approach was derived from practical examples and holds for dairy cattle only. In contrast, the original ssBR approach with MCMC can be used for any model, data, and population structure.

Multi-Trait Single-Step Reliability Approximation

Ben Zaabza et al. (2022a) developed an ssGBLUP reliability approximation method for a multi-trait model. The method is based on a separate calculation of reliabilities for genotyped and nongenotyped animals, and comprises 3 steps. Firstly, the amount of nongenomic information for genotyped animals is calculated using approximate reliability in a traditional pedigree-based animal model. Secondly, the genomic information is added to estimate the total amount of information in ssGBLUP model reliability. This involves the use of a technique called reverse-reliability estimation to calculate ERC, instead of the simple formula $\text{ERC}=\frac{r^2}{\left(1-r^2\right)}\frac{\left(1-h^2\right)}{h^2}$. The reverse-reliability

technique consists of approximating an ERC value from a given reliability by reversing the method of Tier and Meyer (2004). The ERC for each genotyped animal were solved iteratively using a Newton-Raphson algorithm (Ben Zaabza et al., 2022a,b). The ERC based on the reverse-reliability estimation are then used as weights in calculating multi-trait GBLUP model reliability. Thirdly, reliabilities for the nongenotyped animals are calculated by adding the increase in information due to genomics to the genotyped animals through pseudo-observations in a conventional pedigree-based animal model.

Ben Zaabza et al. (2022a) showed that the computational cost of the approximation method was only about 12% of that of the exact ssGBLUP reliability approach which was based on matrix inversion. Calculation of genomic reliabilities using the multi-trait GBLUP model proved to be the most time-consuming step, with $O[(tn)^3]$ computations and $O[n(m + t^2n)]$ memory, where t is the number of traits and n is the number of genotyped animals. This is not tenable in case of many genotyped animals. One way to overcome issues with computing cost related to the multi-trait GBLUP model reliability is to use equivalent multi-trait SNPBLUP. Indeed, reliabilities for genotyped animals can be efficiently computed by an equivalent multi-trait SNPBLUP model with $O[(tm)^3]$ computations and $O[m(n + t^2m)]$ memory, which can be a practical choice to reduce the computational cost when the number of genotyped animals exceeds the number of markers. However, this advantage is lost if an RPG effect is fitted into the SNPBLUP model, which incurs a cost of $O[m(n + t^2m)]$ (Liu et al., 2014; Bermann et al., 2021). One way to overcome the problem is to use either partial MC-based sampling, where the size of the MME depends on the number of $m + n_{MC}$, or full MC-based sampling, where the MME depends on n_{MC}.

DISCUSSION

The purpose of this paper was to give an overview of the various methods for calculating model reliabilities, with special focus on genomic reliabilities. The computation of pedigree-based animal model reliabilities no longer presents as great of a challenge as in previous decades, even for large data sets. However, the calculation of exact genomic reliabilities involving MME inversion has proved infeasible if the number of genotyped animals is very high. Different approximation methods have been developed to overcome the heavy computational challenge. The rapid pace at which the new literature on model reliabilities is increasing makes it practically impossible to summarize it in such a short space, and so we will limit our discussion to only a few recent developments.

Among the reliability calculation methods proposed for GBLUP/SNPBLUP models, the MC approach by Ben Zaabza et al. (2020a, 2021) allows reduction of MME from $(m + n)$ to $(m + n_{MC})$ with partial MC or to only n_{MC} with full MC. Both methods were found to be effective when the number of genotyped animals n is larger than the number of SNP markers m. However, a compromise needs to be made between the number of MC samples, n_{MC}, and accuracy. Computation of exact reliabilities in ssGBLUP models is even more challenging when many more genotyped animals are used, which has prompted the development of different approximation methods to overcome the heavy computational challenge.

Liu et al. (2017) developed a method to harmonize the computation of genomic reliabilities in different countries. The authors proposed a multistep procedure for reliability calculation using a single-trait model. They quantified the nongenomic information for genotyped animals in a conventional animal model and then used the nongenomic information to obtain the total information in a genomic model. The method is simple and efficient both in terms of computation and accuracy, but its extension to a multi-trait model still requires the development of the necessary software.

The approach by Liu et al. (2017) has motivated the introduction of other approximation methods. For example, the ssBR method proposed by Edel et al. (2019) has proved effective. The authors examined the relative merit of their method in terms of accuracy by comparing it to the exact ssGBLUP reliabilities. The ssBR was found to do slightly better than the Misztal et al. (2013b) and Liu et al. (2017) methods, but its drawback is that the diagonal approximation was initially derived from practical examples and holds for dairy cattle only. The method developed by Bermann et al. (2022), which is based on an APY, enjoys a significant computational advantage over the others, but loses in accuracy compared with them. The quality of the method may also be dependent on the chosen set of core animals, and further study is needed to assess the effect of the chosen core animals on the goodness of the approximation.

Ben Zaabza et al. (2022a) developed a reliability approximation method similar to those by Liu et al. (2017) and Edel et al. (2019). The encouraging results reported in Ben Zaabza et al. (2022a) suggest that the method might be useful for large data sets. However, their procedure involves an intermediate step of calculating reliabilities in GBLUP or SNPBLUP, and this continues to be computationally expensive when

an RPG effect is included. Even though the MC approach can be applied for reliability approximation in SNPBLUP models when RPG is fitted, the method is sensitive to the MC sample size and the size of the RPG weight. The encouraging results of these recent approximation methods lend some support to their potential applicability for large data sets. However, all the methods produce some inflation or deflation in the reliabilities without a compelling explanation. One of the main concerns has to do with the double counting of information in the genomic reliability calculation. Solving this problem poses an interesting topic for future research.

It is worthwhile noting that all these proposed approaches for estimating genomic reliabilities center on calculating PEV for each SNP and the RPG effects separately, and finally summing these to arrive at a GEBV reliability. The EDC reliability approximation by VanRaden and Wiggans (1991) is fundamentally different in that it estimates the reliability contributed by information sources. In conventional pedigree-based BLUP (**PBLUP**), these are the contributions from own observation, from the parents, and from the offspring. This method could conceivably be extended by adding an animal's (imputed) genotype as a fourth source of information. The latter could be decomposed in a reliability of a fully known genotype and a reliability with which the (imputed) genotype is known (obtained from propagation). If an efficient method could be conceived to estimate genotype reliability, this could potentially greatly simplify GEBV reliability calculations, particularly if calculation of the reliability of each SNP can be avoided.

An interesting question concerns changes in the reliability of EBV due to genomic information. There are at least 2 aspects to be considered: model reliability versus observed reliability, and what base is used for the genomic and pedigree data.

The first is that the computed model reliabilities depend on the assumptions made in the statistical model. Thus, the reliability of an EBV/GEBV by a model is restricted by the assumptions of the prediction model. For example, consider a GBLUP/SNPBLUP model to analyze data having 50,000 markers versus 10,000 markers. The model with 50,000 markers is likely to show lower reliabilities than the one with 10,000 markers. However, the observed reliabilities will tend to be the opposite: that is, a model having more markers give higher prediction reliabilities than a model with fewer markers. Thus, a comparison of reliabilities between different models having different assumptions is often not meaningful. Another example is that a PBLUP model which uses an incomplete pedigree that has many unrelated animals will give higher EBV reliabilities than a PBLUP model with complete pedigree information where the animals will be more related and possibly inbred. A complete pedigree will have more closely related individuals than an incomplete pedigree and, thus, will have less uncorrelated information available to predict the breeding values. Similarly, genomic information may indicate a higher relationship between some animals than pedigree information, which can translate to lower model reliability for some individuals.

Any pedigree or genomic based model reliabilities depend on the used genetic base. The centering of markers, or so-called allele coding, affects EBV reliabilities by SNPBLUP/GBLUP. In theory, centering does not affect the PEV of marker effects in SNPBLUP such that the between-marker PEV in the inverse of MME (SNPBLUP) will be the same irrespective of the allele coding (Strandén and Christensen, 2011). Thus, the use of either −1,0,1 or 0,1,2 or centering by average genotype will not affect the PEV marker matrix. However, EBV reliabilities for animal i can be calculated as $1-\mathbf{z}_i\mathbf{C}^{gg}\mathbf{z}_i^{'}/\left(\mathbf{G}_{ii}\sigma_u^2\right)$ from SNPBLUP MME, as already explained. In the extreme, an individual i may have zero for all values in the genotype vector $\mathbf{z}_i$ for some centering but a vector of ones for another. The PEV matrix between the markers in SNPBLUP (i.e., $\mathbf{C}^{gg}$), is the same in both cases, but $\mathbf{z}_i\mathbf{C}^{gg}\mathbf{z}_i^{'}$ for an individual differs. In the extreme case of $\mathbf{z}_i$ being the zero vector, the genomic relationship matrix diagonal value $\mathbf{G}_{ii}$ would be zero as well, which illustrates that allele coding affects the reliability. To resolve the problem of inconsistent PEV due to different allele coding, Tier et al. (2018) proposed including into the $\mathbf{G}$ matrix an additional founder using base population allele frequencies, which model the mean of the founder population. The authors showed that the use of an implied founder resulted in a single set of PEV irrespective of the allele coding scheme. For more details, we refer the reader to Tier et al. (2018).

Interbull has the difficult task of trying to come up with general guidelines to compute genomic reliabilities in different countries. Allele frequencies can differ by population, and consequently, the base differs as well if it relies on allele frequencies, as is common. Use of a common allele frequency base for all populations would allow a unified approach. However, this would often differ from the used approach for a local population. Thus, genomic EBV reliabilities may not be as easily transformed from one population to another as pedigree-based EBV reliabilities by measures such as effective daughter contributions. Furthermore, different populations can use different SNP markers in their genomic evaluations. Perhaps, if all participating populations used the meta-founder approach (Legarra

et al., 2015) to model the relationships among pedigree founders, such that they are all related, this problem would diminish, as the same allele frequency of 0.5 is used for all markers. It is important to note that the meta-founder approach developed by Christensen et al. (2012) and Legarra et al. (2015) can be seen as a generalization of unknown parent groups or genetic groups, and consists of a coherent method to make the **G** and **A** matrices relative to the same base population and thus compatible. Properties, similarities, and differences between meta-founders and unknown parent groups are well discussed in Masuda et al. (2021).

CONCLUSIONS

The computation of reliability of EBV is challenging, and the recent increase in the number of genotyped animals has further exacerbated this challenge. Current single-step techniques for EBV computation have succeeded in making the computational cost of solving the MME linearly proportional to the number of genotyped animals. Thus, the solving algorithms are capable of handling models with large data and many genotyped animals. In practice, the calculation of EBV can be achieved by using a fast-iterative method rather than a direct inversion of the MME, whereas the calculation of reliability requires results from the direct inversion of the MME. Therefore, most of algorithms used to calculate EBV are unsuitable for use in genomic reliability models because the calculations of EBV and reliabilities do not share the same math. In fact, direct inversion of the MME may be the only case where the same strategies are used for solutions to both EBV and reliabilities. On the other hand, when approximations are used in the reliability computations, the choice of the reliability solver and EBV solver can be independent and share very little code. Consequently, several studies on reliability approximations have been conducted lately to handle any size of dairy cattle populations, with some promising results. A universal solution applicable to every model and input data may not be possible but recent intensive efforts to approximate reliability have developed efficient and accurate algorithms for a variety of very large genomic evaluations.

ACKNOWLEDGMENTS

This project was supported by the USDA-ARS Food Systems Research Center at the University of Vermont (Burlington). The authors have not stated any conflicts of interest.

REFERENCES

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 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. https://doi.org/10.3168/jds.2009-2730.

Ben Zaabza, H., E. A. Mäntysaari, and I. Strandén. 2020a. Using Monte Carlo method to include polygenic effects in calculation of SNP-BLUP model reliability. J. Dairy Sci. 103:5170–5182. https://doi.org/10.3168/jds.2019-17255.

Ben Zaabza, H., E. A. Mäntysaari, and I. Strandén. 2020b. Snp_blup_rel: Software for calculating individual animal SNP-BLUP model reliabilities. Agric. Food Sci. 29:297–306. https://doi.org/10.23986/afsci.95617.

Ben Zaabza, H., E. A. Mäntysaari, and I. Strandén. 2021. Estimation of individual animal SNP-BLUP reliability using full Monte Carlo sampling. JDS Commun. 2:137–141. https://doi.org/10.3168/jdsc.2020-0058.

Ben Zaabza, H., M. Taskinen, E. A. Mäntysaari, T. Pitkänen, G. P. Aamand, and I. Strandén. 2022a. Breeding value reliabilities for multiple-trait single-step genomic best linear unbiased predictor. J. Dairy Sci. 105:5221–5237. https://doi.org/10.3168/jds.2021-21016.

Ben Zaabza, H., M. Taskinen, E. A. Mäntysaari, T. Pitkänen, G. P. Aamand, and I. Strandén. 2022b. Avoiding double counting in genomic reliability calculations. Page 4 in Proceedings of the World Congress of Genetics Applied in Livestock Production, Rotterdam, the Netherlands.

Bermann, M., D. A. L. Lourenco, and I. Misztal. 2022. Efficient approximation of reliabilities for single-step genomic best linear unbiased predictor models with the algorithm for proven and young. J. Anim. Sci. 100:1–7.

Boichard, D., and A. J. Lee. 1992. Approximate accuracy of genetic evaluation under a single-trait animal model. J. Dairy Sci. 75:868–877. https://doi.org/10.3168/jds.S0022-0302(92)77827-8.

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

Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. Animal 6:1565–1571. https://doi.org/10.1017/S1751731112000742.

Ducrocq, V., and M. d. P. Schneider. 2007. Generalization of the information source method to compute reliabilities in test models. Interbull Bull. 37:82–87.

Edel, C., E. C. G. Pimentel, M. Erbe, R. Emmerling, and K.-U. Götz. 2019. Short communication: Calculating analytical reliabilities for single-step predictions. J. Dairy Sci. 102:3259–3265. https://doi.org/10.3168/jds.2018-15707.

Fernando, R. L., J. C. Dekkers, and D. J. Garrick. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genet. Sel. Evol. 46:50. https://doi.org/10.1186/1297-9686-46-50.

Gao, H., M. Koivula, J. Jensen, I. Strandén, P. Madsen, T. Pitkänen, G. P. Aamand, and E. A. Mäntysaari. 2018. Short communication: Genomic prediction using different single-step methods in the Finnish red dairy cattle population. J. Dairy Sci. 101:10082–10088. https://doi.org/10.3168/jds.2018-14913.

García-Ruiz, A., J. B. Cole, P. M. VanRaden, G. R. Wiggans, F. J. Ruiz-López, and C. P. Van Tassell. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genome selection. Proc. Natl. Acad. Sci. USA 113:E3995–4004.

Golden, B. L., R. L. Fernando, and D. J. Garrick. 2016. Bolt and an alternative approach to genomic EPDs. Pages 102–106 in Proceedings of the Beef Improvement Federation.

Goodale, H. D. 1928. Selecting a Herd Sire: The Mount Hope Bull Index. M.B. Brown Printing and Binding.

Guarini, A. R., D. A. L. Lourenco, L. F. Brito, M. Sargolzaei, C. F. Baes, F. Miglior, I. Misztal, and F. S. Schenkel. 2018. Comparison of genomic predictions for lowly heritable traits using multi-step

and single-step genomic best linear unbiased predictor in Holstein cattle. J. Dairy Sci. 101:8076–8086. https://doi.org/10.3168/jds.2017-14193.

Harris, B., and D. Johnson. 1998. Approximate reliability of genetic evaluations under an animal model. J. Dairy Sci. 81:2723–2728. https://doi.org/10.3168/jds.S0022-0302(98)75829-1.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92:433–443. https://doi.org/10.3168/jds.2008-1646.

Henderson, C. R. 1963. Selection index and expected genetic advance. In Statistical Genetics and Plant Breeding. W. D. Hanson and H. F. Robinson, ed. National Academy of Sciences-National Research Council. Pub. 982.

Henderson, C. R. 1975. Comparison of alternative sire evaluation methods. J. Anim. Sci. 41:760–770. https://doi.org/10.2527/jas1975.413760x.

Henderson, C. R. 1984. Applications of linear models in animal breeding. University of Guelph, Guelph, Ontario, Canada.

Hickey, J. M., M. G. Keane, D. A. Kenny, A. R. Cromie, H. A. Mulder, and R. F. Veerkamp. 2008. Estimation of accuracy and bias in genetic evaluations with genetic groups using sampling. J. Anim. Sci. 86:1047–1056. https://doi.org/10.2527/jas.2007-0653.

Hickey, J. M., R. F. Veerkamp, M. P. L. Calus, H. A. Mulder, and R. Thompson. 2009. Estimation of prediction error variances via Monte Carlo sampling methods using different formulations of the prediction error variance. Genet. Sel. Evol. 41:23. https://doi.org/10.1186/1297-9686-41-23.

Jamrozik, J., L. R. Schaeffer, and G. B. Jansen. 2000. Approximate accuracies of prediction from random regression models. Livest. Prod. Sci. 66:85–92. https://doi.org/10.1016/S0301-6226(00)00158-5.

Johnson, D. L., and R. Thompson. 1995. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. J. Dairy Sci. 78:449–456.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663. https://doi.org/10.3168/jds.2009-2061.

Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single Step, a general approach for genomic selection. Livest. Sci. 166:54–65. https://doi.org/10.1016/j.livsci.2014.04.029.

Legarra, A., O. F. Christensen, Z. G. Vitezica, I. Aguilar, and I. Misztal. 2015. Ancestral relationships using metafounders: Finite ancestral populations and across population relationships. Genetics 200:455–468. https://doi.org/10.1534/genetics.115.177014.

Liu, Z., M. E. Goddard, B. J. Hayes, F. Reinhardt, and R. Reents. 2016. Technical note: Equivalent genomic models with a residual polygenic effect. J. Dairy Sci. 99:2016–2025. https://doi.org/10.3168/jds.2015-10394.

Liu, Z., M. E. Goddard, F. Reinhardt, and R. Reents. 2014. A single-step genomic model with direct estimation of marker effects. J. Dairy Sci. 97:5833–5850. https://doi.org/10.3168/jds.2014-7924.

Liu, Z., F. Reinhardt, A. Bünger, and R. Reents. 2004. Derivation and calculation of approximate reliabilities and daughter yield-deviations of a random regression test-day model for genetic evaluation of dairy cattle. J. Dairy Sci. 87:1896–1907. https://doi.org/10.3168/jds.S0022-0302(04)73348-2.

Liu, Z., F. Seefried, F. Reinhardt, and R. Reents. 2010. Approximating reliabilities of estimated direct genomic values. Interbull Bull. 41:29–32.

Liu, Z., F. R. Seefried, R. Reinhardt, S. Rensing, G. Thaller, and R. Reents. 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genet. Sel. Evol. 43:19. https://doi.org/10.1186/1297-9686-43-19.

Liu, Z., P. M. VanRaden, M. H. Lidauer, M. P. Calus, H. Benhajali, H. Jorjani, and V. Ducrocq. 2017. Approximating genomic reliabilities for national genomic evaluation. Interbull Bull. 51:75–85.

Mäntysaari, E. A., M. Koivula, and I. Strandén. 2020. Symposium review: Single-step genomic evaluations in dairy cattle. J. Dairy Sci. 103:5314–5326. https://doi.org/10.3168/jds.2019-17754.

Masuda, Y., P. M. VanRaden, S. Tsuruta, D. A. L. Lourenco, and I. Misztal. 2021. Invited review: Unknown-parent groups and metafounders in single-step genomic BLUP. J. Dairy Sci. 105:923–939.

Matilainen, K., E. A. Mäntysaari, M. H. Lidauer, I. Strandén, and R. Thompson. 2013. Employing a Monte Carlo algorithm in Newton-type methods for restricted maximum likelihood estimation of genetic parameters. PLoS One 8:e80821. https://doi.org/10.1371/journal.pone.0080821.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829. https://doi.org/10.1093/genetics/157.4.1819.

Meyer, K. 1989. Approximate accuracy of genetic evaluation under an animal model. Livest. Prod. Sci. 21:87–100.

Meyer, K., and B. Tier. 2003. Approximate prediction covariances among multiple estimated breeding values for individuals. Interbull Bull. 31:133–136.

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

Misztal, I., S. E. Aggrey, and W. M. Muir. 2013a. Experiences with a single-step genome evaluation. Poult. Sci. 92:2530–2534. https://doi.org/10.3382/ps.2012-02739.

Misztal, I., A. Legarra, and I. Aguilar. 2014. Using recursion to compute the inverse of the genomic relationship matrix. J. Dairy Sci. 97:3943–3952. https://doi.org/10.3168/jds.2013-7752.

Misztal, I., D. A. L. Lourenco, and A. Legarra. 2020. Current status of genomic evaluation. J. Anim. Sci. 98:skaa101. https://doi.org/10.1093/jas/skaa101.

Misztal, I., and M. Perez-Enciso. 1993. Sparse matrix inversion for restricted maximum likelihood estimation of variance components by expectation-maximization. J. Dairy Sci. 76:1479–1483. https://doi.org/10.3168/jds.S0022-0302(93)77478-0.

Misztal, I., S. Tsuruta, I. Aguilar, A. Legarra, P. M. VanRaden, and T. J. Lawlor. 2013b. Methods to approximate reliabilities in single-step genomic evaluation. J. Dairy Sci. 96:647–654. https://doi.org/10.3168/jds.2012-5656.

Misztal, I., and G. R. Wiggans. 1988. Approximation of prediction error variance in large-scale animal models. J. Dairy Sci. 71:27–32. https://doi.org/10.1016/S0022-0302(88)79976-2.

Robinson, G. K., and L. P. Jones. 1987. Approximations for prediction errors. J. Dairy Sci. 70:1623–1632. https://doi.org/10.3168/jds.S0022-0302(87)80190-X.

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

Strabel, T., I. Misztal, and J. K. Bertrand. 2001. Approximation of reliabilities for multiple-trait models with maternal effects. J. Anim. Sci. 79:833–839. https://doi.org/10.2527/2001.794833x.

Strandén, I., and O. F. Christensen. 2011. Allele coding in genomic evaluation. Genet. Sel. Evol. 43:25. https://doi.org/10.1186/1297-9686-43-25.

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

Takahashi, K., J. Fagan, and M. S. Chen. 1973. Formation of a sparse bus impedance matrix and its application to short circuit study. Page 63 in Proc. 8th Inst. PICA Conf., Minneapolis, MN.

Taskinen, M., E. A. Mäntysaari, and I. Strandén. 2017. Single-step SNP-BLUP with on-the-fly imputed genotypes and residual polygenic effects. Genet. Sel. Evol. 49:36. https://doi.org/10.1186/s12711-017-0310-9.

Thompson, R., N. R. Wray, and R. E. Crump. 1994. Calculation of prediction error variances using sparse matrix methods. J. Anim.

Breed. Genet. 111:102–109. https://doi.org/10.1111/j.1439-0388.1994.tb00443.x.
Tier, B., K. Meyer, and A. Swan. 2018. On implied genetic effects, relationships and alternate coding. Page 5 in Proceedings of the 11th World Congress on Genetics Applied to Livestock Production, Auckland.
Tier, B., and K. Meyer. 2004. Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models. J. Anim. Breed. Genet. 121:77–89. https://doi.org/10.1111/j.1439-0388.2003.00444.x.
Tsuruta, S., D. A. L. Lourenco, Y. Masuda, T. J. Lawlor, and I. Misztal. 2021. Reducing computational cost of large-scale genomic evaluation by using indirect genomic prediction. JDS Commun. 2:356–360. https://doi.org/10.3168/jdsc.2021-0097.
Van Tassell, C. P., G. R. Wiggans, P. M. VanRaden, T. A. Cooper, and T. S. Sonstegard. 2011. Applications of genomics to genetic improvement of dairy cattle. Pages 57–68 in Proc. Cornell Nutr. Conf., Syracuse, NY.
Van Vleck, L. D. 1993. Variance of prediction error with mixed model equations when relationships are ignored. Theor. Appl. Genet. 85:545–549. https://doi.org/10.1007/BF00220912.
VanRaden, P. M. 2007. Genomic measures of relationship and inbreeding. Interbull Bull. 37:33–36.
VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980.
VanRaden, P. M. 2012. Avoiding bias from genomic pre-selection in converting daughter information across countries. Interbull Bull. 45.
VanRaden, P. M. 2020. Symposium review: How to implement genomic selection. J. Dairy Sci. 103:5291–5301. https://doi.org/10.3168/jds.2019-17684.
VanRaden, P. M., and A. E. Freeman. 1985. Rapid method to obtain bounds on accuracies and prediction error variances in mixed models. J. Dairy Sci. 68:2123–2133. https://doi.org/10.3168/jds.S0022-0302(85)81078-X.
VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737–2746. https://doi.org/10.3168/jds.S0022-0302(91)78453-1.
Wilmink, L. B. M., and J. Dommerholt. 1985. Approximate reliability of Best Linear Unbiased Prediction in models with and without relationships. J. Dairy Sci. 68:946–952. https://doi.org/10.3168/jds.S0022-0302(85)80913-9.

ORCIDS

Hafedh Ben Zaabza https://orcid.org/0000-0003-4648-3813
Curtis P. Van Tassell https://orcid.org/0000-0002-8416-2087
Jeremie Vandenplas https://orcid.org/0000-0002-2554-072X
Paul VanRaden https://orcid.org/0000-0002-9123-7278
Zengting Liu https://orcid.org/0000-0002-7112-0320
Herwin Eding https://orcid.org/0000-0002-6763-5110
Stephanie McKay https://orcid.org/0000-0003-1434-3111
Katrine Haugaard https://orcid.org/0000-0003-3793-6075
Martin H. Lidauer https://orcid.org/0000-0003-0508-9991
Esa A. Mäntysaari https://orcid.org/0000-0003-0044-8473
Ismo Strandén https://orcid.org/0000-0003-0161-2618