Approximation of Reliability in Single Step Models using the Interbull Standardized Genomic Reliability Method

M. Erbe, C. Edel, E.C.G. Pimentel, J. Dodenhoff and K.-U. Götz
Bavarian State Research Centre for Agriculture, Institute for Animal Breeding, 85586 Grub, Germany

Abstract

Software for estimating breeding values within a single step system has become available in the last years. However, approaches to approximate reliabilities for those breeding values in routine dimensions have been rare. Last year, the Genomic Reliability Working Group of Interbull presented a general step-wise framework (Interbull Standardized Genomic Reliability Method, ISGRM) for approximating reliabilities which is applicable to two-step as well as to single-step models. For assessing the accuracy of the approach and its performance in the different steps, a small test data set (16.5k individuals in the pedigree, 4.3k of them with phenotypes, 5.8k of them genotyped) was created. Exact theoretical single step reliabilities could be obtained for this set via numerical inversion of the total system. These reliabilities were compared with values obtained with ISGRM. Results looked very promising for the genotyped individuals, while they were not completely satisfying for non-genotyped individuals in this data set. The lines of action of calculating effective record contributions for the genomic reference set and of considering the residual polygenic contribution were identified to have an influence on the performance. For larger routine data sets, however, not only the quality of the results, but also the possibility that all necessary calculations can be performed in a reasonable time frame with given hardware and software configurations is important. We thus assessed approximation options for different steps of ISGRM with the software ApaX99 and options to calculate reliabilities of direct genomic values via SNP reliabilities with the snp_blup_rel program. Performance testing in a routine data set for conformation traits in Fleckvieh cattle (~ 3.3M individuals in pedigree, ~ 1.4M of them with phenotypes, 78k of them genotyped) revealed that only the first step, namely the numerical inversion of a system with dimension (n_{SNPs}+1) x (n_{SNPs}+1), is computationally demanding (took ~ half an hour time and 38 GB RAM in the given data set). All other steps could be performed without any larger memory or CPU requirements in very short time.

Key words: single step model, reliability, genomic evaluation, Interbull Standardized Genomic Reliability Method

Introduction

Using single step models (e.g. Legarra et al., 2014) to obtain genomically enhanced breeding values (GEBV) for all individuals in the pedigree with some of them being genotyped has become popular in the last years. Different software (e.g. Aguilar et al., 2018; Strandén et al., 2018) for applying such models has become available so that GEBV can be obtained even for large(r) routine data sets. Calculating reliabilities for such GEBV is possible via numerical inversion of the left hand side only in very small data sets. However, only a few methods for approximation of reliabilities have been described in literature so far.

Misztal et al. (2013) proposed a method that allows approximating reliabilities for all genotyped individuals in a single step system by solving $\boldsymbol{Q}^{-1} = [\boldsymbol{D} + (\boldsymbol{I} + \boldsymbol{G}^{-1} - \boldsymbol{A}_{22}^{-1})\alpha]^{-1}$ where $\boldsymbol{D}$ is a diagonal matrix with weights derived from conventional reliabilities, $\boldsymbol{I}$ is an identity matrix, $\boldsymbol{G}$ is a genomic relationship matrix, $\boldsymbol{A}_{22}$ is the numerator relationship matrix between all genotyped individuals and α is the variance ratio of error and genetic variance. The final reliability for a genotyped individual i is then calculated as $r_i^2 = 1 - \alpha q^{ii}$.

With this approach, all genotyped individuals obtain a reliability gain due to

genomics, but non-genotyped individuals are not considered. The necessity to invert $\boldsymbol{G}$ and $\boldsymbol{A}_{22}$ might be a limitation of this approach in routine data sets with a large number of individuals genotyped.

Based on the idea to develop an Interbull standard method for approximating genomic(ally enhanced) reliabilities, Liu et al. (2017) proposed a new framework which is termed Interbull Standardized Genomic Reliability Method (ISGRM). ISGRM is thought to be used by Interbull members in routine applications in order to harmonize the way genomic reliabilities are calculated. It is a multi-step procedure which was not specially designed for single step models. It may, however, be used as a close approximation even in that context, since it proposes an additional step of propagating reliability gains from genomics to non-genotyped animals. The maximum system size to be solved is set fixed by the number of markers used and thus ISGRM seems to be applicable also to very large routine data sets.

The aim of this study was to assess ISGRM using two different data sets. First, with a very small data set approximated reliabilities from this approach were compared with true model based reliabilities. Since not only the quality of results, but also the feasibility of the necessary computations is an issue for routine applications, we further tested applicability and computational demands in a routine-like data set.

Materials and Methods

Single Step Model

The underlying Single Step Model was assumed to be

$$\begin{bmatrix} \boldsymbol{X'X} & \boldsymbol{X'U} \\ \boldsymbol{U'X} & \boldsymbol{U'U} + \boldsymbol{H^{-1}}\lambda \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{b}} \\ \widehat{\boldsymbol{a}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'y} \\ \boldsymbol{U'y} \end{bmatrix} \qquad [1]$$

where $\boldsymbol{X}$ and $\boldsymbol{U}$ were design matrices relating phenotypes to fixed and random effects, respectively, $\boldsymbol{y}$ was the vector with phenotypic observations, λ the variance ratio of error and genetic variance and $\boldsymbol{H}$ the combined pedigree and genomic relationship matrix as defined in e.g. Aguilar et al. (2010) or Christensen & Lund (2010) so that

$$\boldsymbol{H^{-1}} = \boldsymbol{A^{-1}} + \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{G^{-1}} - \boldsymbol{A_{22}^{-1}} \end{bmatrix}$$

and

$$Var(\boldsymbol{a}) = \boldsymbol{H}\sigma_a^2$$

$\boldsymbol{G}$ was calculated as

$$\boldsymbol{G} = (1-k)\boldsymbol{G}^* + k\boldsymbol{A}_{22}$$

with $\boldsymbol{G}^*$ being a genomic relationship matrix (Method 1 of VanRaden (2008) with basis allele frequencies (Gengler et al., 2007)) and k being the proportion of genetic variance not explained by markers.

Small test data set

The small test data set was a subset of pig routine evaluation data. In order to be able to compare approximated R^2 values with theoretical model based ones, the size of the routine data set was reduced to a pedigree of 16 500 individuals. 4300 of them had a phenotype for a conformation trait with a heritability of 0.33. 5800 individuals were genotyped with a 60k SNP chip. 180 of the 5800 individuals had an own phenotype and 600 of the 5 800 were parent of at least one phenotyped, but non-genotyped offspring.

Description of ISGRM as used in this study

ISGRM is a multi-step procedure. First prediction error co(variances) of direct genomic values (DGVs) were obtained via SNP effect reliabilities assuming

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{1'W^{-1}1} & \boldsymbol{1'W^{-1}Z_r} \\ \boldsymbol{Z_r'W^{-1}1} & \boldsymbol{Z_r'W^{-1}Z_r} + \boldsymbol{I}\dfrac{\sigma_e^2}{\sigma_{SNP}^2} \end{bmatrix}$$

and

$$\boldsymbol{PEV_g} = \boldsymbol{ZC^{22}Z'}\sigma_e^2 \qquad [2]$$

where $\boldsymbol{Z}$ was a (n_{geno} x n_{SNPs}) matrix with genotypes 0,1,2 corrected for two times the base allele frequencies, n_{geno} was the number of genotyped individuals and n_{SNPs} was the number of SNPs. Subscript r indicates that the matrix only contained reference individuals, i.e. $\boldsymbol{Z}_r$ was of dimension (n_{ref} x n_{SNPs}). $\boldsymbol{W}$ was a matrix of weighting factors used for all genotyped reference individuals. Detailed description of reference individuals and

corresponding weighting factors can be found in the following section.

The corresponding reliabilities $\boldsymbol{R_g^2}$ might be corrected (in a multiplicative way) for residual polygenic variance used in the model, for the imputation quality in case of imputed genotypes and can be adapted to fit to validation reliabilities. As we did not have imputed genotypes in the data set and we were interested in comparing true model based and approximated reliabilities, we only applied a correction for the polygenic component which is described in detail in the next section. The resulting reliabilities from this step will be termed $\boldsymbol{R_{DGV}^2}$ in the following.

Next, reliabilities ($\boldsymbol{R_{A_{22}}^2}$) for a subset EBV model with $\boldsymbol{A_{22}}$ as covariance structure were calculated using the same reference population and weights as in the first step. In the following, a genomic gain for each genotyped individual was calculated as

$$\varphi_{gain_i} = \frac{R^2_{DGV_i}}{1-R^2_{DGV_i}}\lambda - \frac{R^2_{A22_i}}{1-R^2_{A22_i}}\lambda \qquad [3]$$

and then added to the equivalent record contributions ($\boldsymbol{\varphi_{conv}}$) of the conventional full system (as [1], but with $\boldsymbol{A}$ instead of $\boldsymbol{H}$) so that the final reliability of a genotyped individual *i* was

$$R^2_{final_i} = \frac{\varphi_{final_i}}{\varphi_{final_i} + \lambda}$$

with $\varphi_{final_i} = \varphi_{conv_i} + \varphi_{gain_i}$.

Propagation gains for non-genotyped individuals were obtained by solving

$$\boldsymbol{PEV}_{prop} = \begin{bmatrix} \boldsymbol{1'D^{-1}1} & \boldsymbol{1'D^{-1}K} \\ \boldsymbol{K'D^{-1}1} & \boldsymbol{K'D^{-1}K + A^{-1}\lambda} \end{bmatrix}^{-1} \sigma_e^2$$

where $\boldsymbol{D^{-1}}$ was a diagonal matrix with genomic gains for all reference individuals and $\boldsymbol{K}$ was a ($n_{ref}\ x\ n_{all}$) design matrix and n_{all} was the number of animals in the pedigree. From $\boldsymbol{PEV}_{prop}$ reliabilities and corresponding contributions ($\boldsymbol{\varphi_{prop}}$) were calculated for all non-genotyped individuals. Final reliabilities for non-genotyped individuals were then

$$R^2_{final\_ng_i} = \frac{\varphi_{final\_ng_i}}{\varphi_{final\_ng_i} + \lambda}$$

with $\varphi_{final\_ng_i} = \varphi_{conv_i} + \varphi_{prop_i}$.

Reference individuals and weighting factors

While it is relatively straightforward to categorize an individual as a reference animal in the two-step approach it is not as obvious in the single-step context. Due to the way ISGRM is performed the set of reference individuals at maximum includes all genotyped individuals; non-genotyped (but implicitly imputed) individuals cannot be considered. In a supplementary document Liu et al. (2018) presented guidelines on how to define the reference population for ISGRM.

For strategy $\mathbf{w_1}$ we followed these ideas and defined a reference population with a) all genotyped sires with at least one phenotyped, but non-genotyped offspring, and b) genotyped females with phenotypes. The weights used in the $\boldsymbol{W^{-1}}$ matrix are equivalent daughter contributions (EDC) for the sires which only contained information of phenotyped daughters that are not themselves genotyped. All genotyped and phenotyped females obtained equivalent record contributions (ERC). EDCs as well as ERCs were calculated using variance ratios referring to an animal model.

Strategy $\mathbf{w_1}$ ignores some potential reference individuals, namely genotyped, but non-phenotyped dams with phenotyped, but non-genotyped progeny. For strategy $\mathbf{w_2}$ we additionally defined such dams as reference animals with EDC weights calculated equivalently to the sires' EDCs.

Assuming that the reliability of an individual's breeding value is predominantly determined by direct progeny information (e.g. like in milk production traits), all relevant information will be captured by the respective EDCs of genotyped parents. However, if information of grand progeny and other relatives is considered to contribute significantly to the genotyped individual, another approach might be better for obtaining the weights. Strategy $\mathbf{w_3}$ was based on ideas of Harris & Johnson (1998, 2010). Contributions of all other genotyped individuals from each genotyped individual's conventional reliability estimate was removed, thus keeping only information from itself and all other phenotyped, but non-genotyped relatives. These de-regressed reliabilities were converted to ERCs and used as weights. The reference

population then consisted of all individuals with a non-zero weight (0.2 used as limit in this study).

Polygenic contribution

Assuming that not all genetic variance can be captured by the given marker set, the genomic relationship matrix used to build $\boldsymbol{H}$ is a combination of the raw genomic relationship matrix $\boldsymbol{G}^*$ and $\boldsymbol{A_{22}}$. When performing step 1 of ISGRM via a SNP BLUP model as in [2], it is not possible to account for the missing polygenic part exactly. Liu et al. (2017) thus proposed to use (strategy $\mathbf{p_1}$)

$$R^2_{DGV_i} = \begin{cases} (1-k)R^2_{g_i} & \text{for any candidate} \\ R^2_{g_i} & \text{for any reference individual} \end{cases}$$

where k is the proportion of residual polygenic contribution (0.1 assumed in this study).

We further tested two other approaches to approximate $R^2_{DGV_i}$, namely

Strategy $\mathbf{p_2}$: $(1-k)R^2_{g_i} + kR^2_{\boldsymbol{A}_{22}i}$

Strategy $\mathbf{p_3}$:
$(1-k)^2R^2_{g_i} + (1-(1-k)^2)R^2_{\boldsymbol{A}_{22}i}$

Both these additional approaches were applied to all genotyped individuals after we had seen that the differentiation into reference individuals and candidates led to worse results (results not shown).

Comparison of results

To assess the quality and the exactness of ISGRM itself and the optimization potential of the different options regarding weighting factors and polygenic contribution, we numerically inverted the whole left hand side of the single-step system (see equation [1]) for the small test data set. This true model based reliability was used for comparisons to results from ISGRM. We also calculated the prediction error variance for conventional and A_{22} systems by numerical inversion and solved the model used for propagation of the gain to the non-genotyped individuals exactly. This was done in order to avoid inaccuracies due to different approximation steps. For assessing the different approaches of modelling the polygenic contributions, $\boldsymbol{R^2_{DGV}}$ were compared to true model based reliabilities obtained via the inverse of the following left hand side:

$$\begin{bmatrix} \mathbf{1}'\boldsymbol{W}^{-1}\mathbf{1} & \mathbf{1}'\boldsymbol{W}^{-1}\boldsymbol{V} \\ \boldsymbol{V}'\boldsymbol{W}^{-1}\mathbf{1} & \boldsymbol{V}'\boldsymbol{W}^{-1}\boldsymbol{V} + \boldsymbol{G}^{-1}\lambda \end{bmatrix}$$

where $\boldsymbol{V}$ is the $(n_{ref}\, x\, n_{geno})$ design matrix relating observations to random effects and $\boldsymbol{G} = (1-k)\boldsymbol{G}^* + k\boldsymbol{A}_{22}$.

Feasibility and computation time in a routine data set

In order to study if and how ISGRM can be applied in large data set, we made some trials with a routine data set of Fleckvieh cattle. This data set consisted of a pedigree of around 3.3 million individuals with 1.4 million of them having an observation for the conformation trait "udder". 78 000 individuals were genotyped with a 50K SNP chip. 5500 of the genotyped individuals had phenotypic information themselves and 12 000 of the genotyped individuals had at least one phenotyped, but non-genotyped offspring. We developed a pipeline to handle this larger data set with available software and perform all steps of ISGRM efficiently.

Results & Discussion

Figures 1a) and b) show the results for all genotyped and non-genotyped individuals in the small test data set.

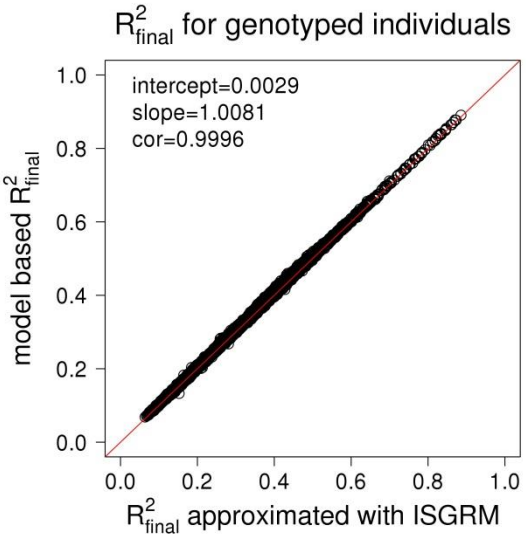

Figure 1a. Model based vs approximated R^2_{final} for all genotyped individuals in the small test data set.

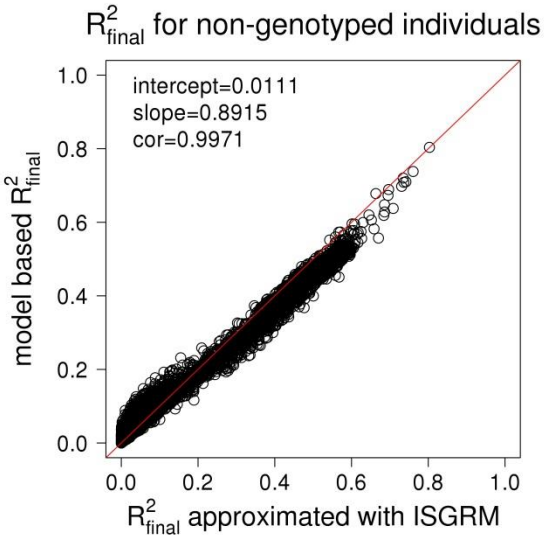

Figure 1b. Model based vs approximated R^2_{final} for all non-genotyped individuals in the small test data set.

The results were obtained with weighting strategy **w**$_3$ and polygenic strategy **p**$_3$ which were found to be optimal in this data set (see below). For genotyped individuals, the approximation of ISGRM was very accurate with a correlation of >0.999 and no bias.

For non-genotyped individuals the correlation of model based and approximated reliabilities was less than for the genotyped individuals and results were unsatisfying. It has to be studied further if this result is only due to the specific structure of the small data set or if there is a methodological error or inaccuracy in the way the propagation to the non-genotyped individuals is performed.

As the aim of this part of the study was to assess how well the multistep ISGRM approximations can match the "true" reliabilities, all inverse matrices needed in the calculations were numerically inverted in this small test data set. In larger real data approximations would replace these inversions except for the one in equation [2]. Cumulative approximations might lessen the quality of the final reliabilities, but test runs have shown that approximation of $R^2_{\boldsymbol{A}_{22}}$ and R^2_{prop} can be performed quite accurately (results not shown) and are not assumed to influence the quality of the final reliabilities much.

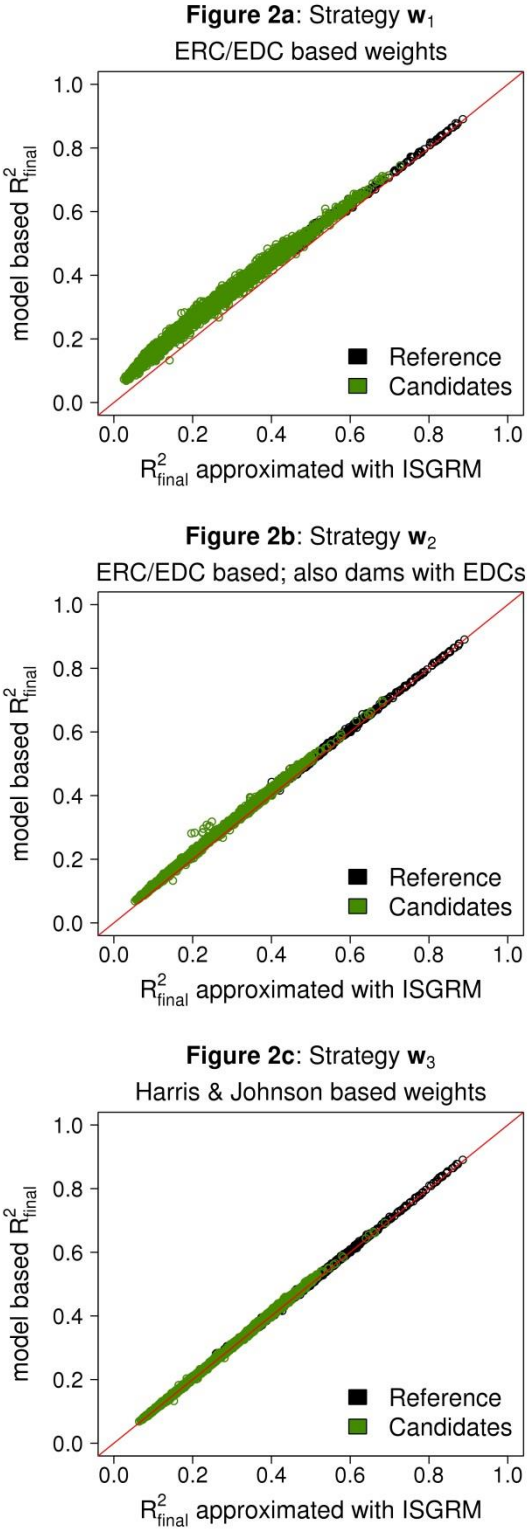

Figure 2. Model based vs approximated R^2_{final} for all genotyped individuals in the small test data set using ERC/EDC based weights as proposed by Liu et al. (2018) in 2a), also including dams' EDCs in 2b), or weights based on de-regressing ideas following Harris & Johnson (1998, 2010) in 2c).

Reference individuals and weighting factors

Figures 2a) to c) show that the quality of the final reliabilities for genotyped individuals obtained with ISGRM can vary depending on the definition of the reference set and the weights given to the reference individuals. If only sires' EDCs and females' ERCs were used, ISGRM R^2 underestimated the model based R^2, arguably because some phenotypic information relevant for the single-step system was lost. Figure 2c shows that, in the small test data set, results were best when we used a weight based on a full model de-regressing approach. Especially for data sets that are different from the classic milk production data structure (almost all females with phenotypes and sires with large numbers of daughters), defining weights not via EDC/ERC but with an approach considering all pathways of information sources seems to lead to favorable results (Figures 2a-c).

In single step models non-genotyped individuals are imputed implicitly. There can be an apparent number (depending of data structure) of individuals that are very accurately imputed via this step (e.g. because they have a larger number of genotyped offspring) and build links to phenotypes. Even the imputed genotypes have a non-ignorable imputation error (Edel et al., 2018), these individuals might valuably contribute to the SNP effect estimation in [1]. In ISGRM, however, reference individuals can only be individuals that are genotyped. It might be worth to study further on how such imputed individuals could be considered as reference individuals.

Residual polygenic contribution

We tested three different approaches regarding the consideration of the residual polygenic contribution. Results are shown in Figures 3a) to c) for $k = 0.1$. For candidates, R^2_{DGV} seems to be somewhat underestimated and biased with strategy $\mathbf{p_1}$. This is the more pronounced, the larger k is chosen (results not shown). Both strategies $\mathbf{p_2}$ and $\mathbf{p_3}$ worked well with a small advantage for $\mathbf{p_3}$, which also persists in more extreme scenarios with larger k values (results not shown). While strategy $\mathbf{p_1}$ only uses R^2_g, for strategies $\mathbf{p_2}$ and $\mathbf{p_3}$ reliabilities of the subset EBV ($R^2_{\mathbf{A}_{22}}$) for all individuals are necessary. However, these values have to be calculated anyway for [3] and approximation is straightforward.

Figure 3a

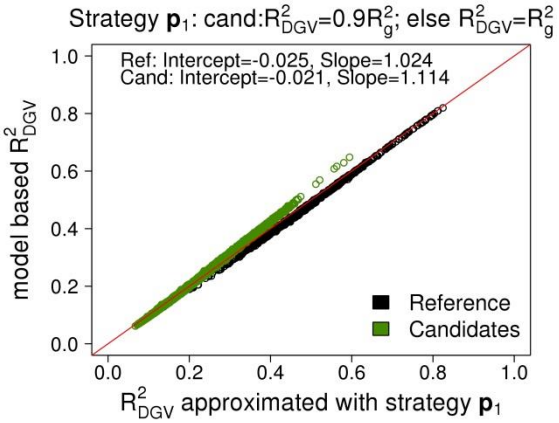

Figure 3b

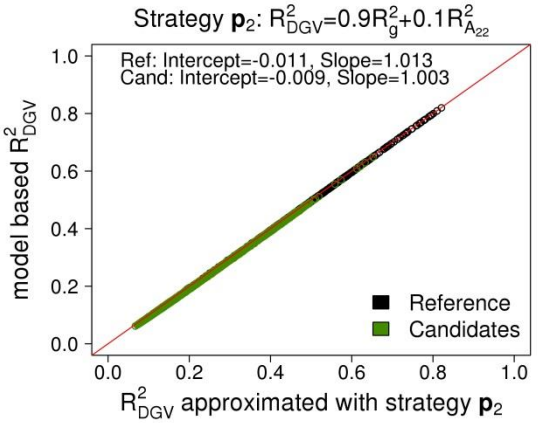

Figure 3c

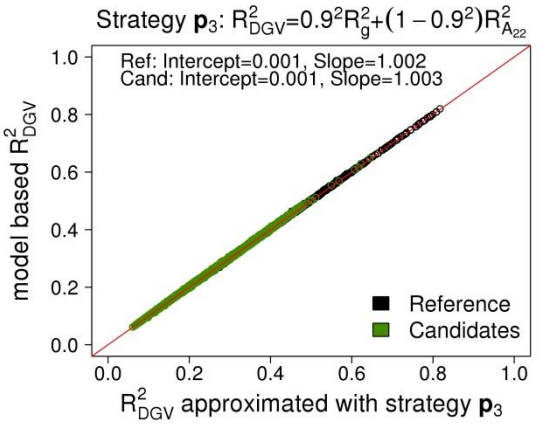

Figure 3. Model based vs approximated R^2_{DGV} in the small test data set for reference individuals and candidates using different approaches to model the residual polygenic contribution.

Application to a large data set

We further established a pipeline that made it possible to perform ISGRM in a routine-like setting. Table 1 shows which programs were used for which step. Building on Harris & Johnson algorithms, the reliability approximation program ApaX99 (included in the MiX99 software package, MiX Development Team (2018)) allows, for a given set of individuals, to obtain ERC with information from all other individuals in the set removed. ApaX99 can also be used for all other steps that require reliability approximations for models with covariance between individuals assumed to be pedigree-based. Apart from R^2_{conv} (which might also be already available from regular routine conventional breeding value estimation) it therefore also allows to approximate $R^2_{\boldsymbol{A}_{22}}$ and R^2_{prop} in a few minutes with very little memory requirements.

Table 1. Programs used to perform the different steps of ISGRM.

Step as described in Liu et al. (2017)	**ApaX99** (MiX99 Development Team, 2018)	Basic arithmetic in **R** (R Core Team, 2016)	**snp_blup_rel** (Mäntysaari et al., 2018)
Reliabilities of SNP genotypes	X	X	X
Reliabilities of DGV		X	
Adjusting the theoretical reliabilities		X	
Calculating the genomic EDC gain	X	X	
Propagation of genomic information to non-genotyped relatives	X	X	
Final reliabilities enhanced with genomic information		X	

The software snp_blup_rel (Mäntysaari et al., 2018) uses genotypes and weights for reference individuals as input and returns raw DGV reliabilities R^2_g. These DGV reliabilities are calculated via numerical inversion of the left hand side of a SNP BLUP model. Running times and memory requirements for tests with different reference set sizes and different SNP set sizes are presented in Table 2.

Running this program once for one trait is easily feasible. However, one should keep in mind that genomic evaluations in dairy cattle usually involve several traits. Obtaining approximations for several traits also means that all steps of ISGRM would have to be run repeatedly. Alternative strategies that make it possible to invert the SNP BLUP model only once (or only a few times) by grouping traits with similar heritability and/or similar groups of phenotyped individuals and approximate R^2_g of one trait from the solution of the other trait might be very helpful and will be subject of further study.

Table 2. Time and memory requirement to perform the calculation of raw DGV reliabilities via SNP effect reliabilities with snp_blup_rel[1] (Mäntysaari et al., 2018).

No of reference animals	No of SNPs	Peak virtual memory	Approx. time in total (building MME/ inversion/reliabilities)
78k	41k	48 GB	35 (7.5/ 9.5/ 11) min
16k	41k	38 GB	26 (1.5/ 10/ 8.5) min
16k	20k	19 GB	8 (<1/ <1/ 1.5) min
16k	8k	10 GB	5.5 (< 1/ < 1/ <1) min
11k	41k	38 GB	25 (1/ 10.5/ 8) min

[1] applied with default options on a Linux-Server with 96 threads and 512 GB RAM

Conclusions

The quality of the approximated reliabilities with ISGRM was very good for genotyped individuals while it was not completely satisfying for the non-genotyped individuals in this data set. Weights used for reference individuals and the way of considering of the polygenic contribution have an influence on the results. The only step that is computationally demanding is the numerical inversion of the SNP BLUP model while all other steps can be performed quickly also in larger data sets.

Acknowledgements

We gratefully acknowledge the Arbeitsgemeinschaft Süddeutscher Rinder-zucht- und Besamungsorganisationen e.V. for their financial support within the research cooperation „Zukunftswege" and the Association for Bioeconomy Research (FBF) for their support in the continuous improvement of genomic routine applications.

References

Aguilar et al., 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.

Aguilar et al., 2018. BLUPF90 suite of programs for animal breeding with focus on genomics. *Proc. 11th WCGALP*, 751.

Christensen & Lund, 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol. 42*: 2.

Edel et al., 2018. Short communication: Calculating Analytical Reliabilities for Single-Step Predictions. *J. Dairy Sci.*, submitted.

Gengler et al., 2007. A simple method to approximate gene content in large pedigree populations: application to the myostatin gene in dual-purpose Belgian Blue cattle. *Animal 1*, 21-28.

Harris & Johnson, 1998. Approximate Reliability of Genetic Evaluations Under an Animal Model. *J. Dairy Sci. 81*, 2723-2728.

Harris & Johnson, 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. *J. Dairy Sci. 93*, 1243-1252.

Legarra et al., 2014. Single Step, a general approach for genomic selection. *Livest. Sci. 166*, 54-65.

Liu et al., 2017. Approximating Genomic Reliabilities for National Genomic Evaluation. *Interbull Bulletin 51*, 75-85.

Liu et al., 2018. A supplementary document to the Interbull genomic reliability method. Interbull Center, Uppsala, Sweden.

Mäntysaari et al., 2018. snp_blup_rel – A tool to calculate reliabilities of individual genomic evaluations. *Presentation at 2018 Interbull Technical Workshop*, Aug 26th 2018, Dubrovnik, Croatia. http://www.interbull.org/static/web/SNP_BLUP_REL_M%C3%A4ntysaari_etal.pdf

Misztal et al., 2013. Methods to approximate reliabilities in single-step genomic evaluation. *J. Dairy Sci. 96*, 647-654.

MiX99 Development Team, 2018. MiX99 software package. http://www.luke.fi/mix99.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Strandén et al., 2018. Efficient single-step BLUP computations with MiX99 software. *Proc. 11th WCGALP*, 566.

VanRaden, 2008. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414-4423.