Managing genetic groups in single-step genomic evaluations applied on female fertility traits in Nordic Red Dairy cattle

K. Matilainen[1], M. Koivula[1], I. Strandén[1], G.P. Aamand[2] and E.A. Mäntysaari[1]

[1] Natural Resources Institute Finland (Luke), Myllytie 1, FI-31600 Jokioinen, Finland
[2] Nordic Cattle Genetic Evaluation, Agro Food Park 15, Skejby, DK-8200 Århus N, Denmark
e-mail: kaarina.matilainen@luke.fi

Abstract

Joint Nordic (Denmark, Finland, Sweden) genetic evaluations of female fertility traits are currently based on a multi-trait multi-lactation animal model for two trait groups. This enables straightforward re-evaluation of the fertility model with genomic information by the single-step (ssGBLUP). ssGBLUP was applied for the first trait group and Nordic Red Dairy cattle data. The ssGBLUP used the same model and variance components as the routine animal model evaluation (BLUP). In addition to BLUP, four genomic evaluations were performed. The first two evaluations were $ssGBLUP_0$ and $ssGBLUP_{QP}$ where either the pedigree relationships, or pedigree and genomic relationships, were accounted in the phantom parent group equations, respectively. Further development of the $ssGBLUP_{QP}$ model was in the third model to include inbreeding coefficients into the pedigree relationship matrix also, and in the fourth model to approximate genomic relationship matrix with APY algorithm. The performance of BLUP and ssGBLUP were studied using Interbull GEBV validation test tailored to multi-trait single step evaluations. Convergence of iterative solver was slow in the BLUP evaluation and extremely slow in $ssGBLUP_0$ evaluation. Convergence of the $ssGBLUP_0$ evaluation was significantly improved by considering effect of genomic relationships in PPG equations, including inbreeding coefficients into the pedigree relationship matrix and applying APY. With these modifications, the number of iterations with ssGBLUP was comparable to animal model, although each iteration round took much longer time. Increase in validation reliability due to genomic information was moderate or high depending on the trait. Thus, the routine Nordic fertility evaluation using ssGBLUP was found feasible after the inbreeding coefficients and PPG had been correctly accounted.

Key words: genetic groups, inbreeding coefficients, APY

Introduction

Nordic Cattle Genetic Evaluations NAV has estimated breeding values using joint Nordic (Denmark, Finland, Sweden) fertility model since 2005 (Fogh et al., 2003). The model was upgraded in 2015 from sire to animal model and from repeatability to multi-trait model for lactations (Muuttoranta et al., 2015). This enables straightforward upgrade of the fertility model to include also genomic information.

Single step genomic evaluation (ssGBLUP) is a method that takes into account phenotypic, pedigree and genomic data simultaneously (Aguilar et al., 2010; Christensen and Lund, 2010). Although ssGBLUP is regarded as a simple and accurate approach, numerical difficulties have been reported in many studies. In the joint Nordic fertility model, correct accounting of genetic groups was necessary for the convergence of the genomic model by preconditioned conjugate gradient (PCG) iteration (Matilainen et al., 2015). Strandén et al. (2016) noticed that convergence of the used iterative solving method can be impaired if inverse of the pedigree-based relationship matrix ($\mathbf{A}^{-1}$) is constructed without taking into

account the inbreeding coefficients. Even with good convergence, the solving of ssGBLUP remains computationally demanding. Computing times per round of iteration increase when the number of genotyped animals increases. To overcome the computational challenges in using and inverting the genomic relationship matrix, we investigated the use of Algorithm for Proven and Young (APY) (Misztal et al., 2014).

Our objective was to study the feasibility and validity of ssGBLUP with and without APY for female fertility traits of Nordic Red Dairy Cattle (RDC).

Materials and Methods

Data

Joint Nordic fertility evaluations involve two different trait groups of which both have 11 correlated traits. We studied trait group I which has 2 heifer traits, and 9 cow traits. The heifer traits were the non-return rate (NRR0) and the length of service period (IFL0). The cow traits in the first, second and third parities were the non-return rate (NRR1, NRR2, NRR3), the length of service period (IFL1, IFL2, IFL3), and days from calving to the first insemination (ICF1, ICF2, ICF3). For the ssGBLUP we used the same model and variance components as is used in the routine evaluations (AM-BLUP). Heritabilities were low for all traits (0.015-0.04) and genetic correlations among traits were high between the lactations (0.60-0.88).

The RDC May 2016 data contained 4.2 million animals with records, and pedigree consisted of 5.4 million animals of which 33 969 had genotypes. There were 332 phantom parent groups (PPG) which were regarded as random in the evaluations. The number of markers used in the study was 46 914. To attain a reduced data for validation test, observations from the last six years were removed. The six year cut-off gave enough third parity observations for the daughters of validation bulls in a validation test.

Analyses

Relationship matrix **H** used in ssGBLUP is comprised of relationship matrices based on both the pedigree (**A**) and the genomic (**G**) information. Also $\mathbf{A}_{22}$, a submatrix of **A** including only genotyped animals, is needed. Problems in convergence can occur if the information from matrices **A**, $\mathbf{A}_{22}$ and **G** contradict. Usually in ssGBLUP implementations with genetic groups, only **A** is augmented to include PPG. This means that so-called QP-transformation is carried out for the inverse of pedigree-based relationship matrix $\mathbf{A}^{-1}$. However, the QP-transformation can be carried out for the inverse of full unified relationship matrix $\mathbf{H}^{-1}$, not only on $\mathbf{A}^{-1}$ (Misztal et al., 2013). This accounts the contributions of genomic relationships to PPG and removes the conflict between expected values of breeding values according to **A** and $\mathbf{A}_{22}$ matrices. On the other hand, because both **G** and $\mathbf{A}_{22}$ naturally take inbreeding into account, there might be worth of including inbreeding coefficients to pedigree-based relationship matrix.

Further improvement in computing time can be achieved by using sparse presentations of $\mathbf{G}^{-1}$ such as APY. Here the QP-transformation was made to the full relationship matrix $\mathbf{H}_{APY}^{-1}$ which now contained approximated genomic information relationship matrix $\mathbf{G}_{APY}^{-1}$ instead of original $\mathbf{G}^{-1}$. In APY, **G** is partitioned to core and noncore animals. The inverse matrix $\mathbf{G}^{-1}$ is approximated so that submatrix pertaining to noncore animals is diagonal. Here, the core was 12 741 animals which had descendant(s) in the pedigree, i.e., none of the non-core animals had any progeny.

As summary, four ssGBLUP evaluations were performed for the full data: 1) ssGBLUP_0 where the PPG were accounted only in **A**, 2)

ssGBLUP$_{QP}$ where the PPG equations accounted both the pedigree and the genomic relationships, 3) ssGBLUP$_{QP_Inb}$ where the inbreeding coefficients were included also in $\mathbf{A}^{-1}$ and 4) ssGBLUP$_{QP_Inb_APY}$ which was as 3 but where APY was applied. To overcome problems with singular **G**, 10% weight for polygenic information was used in all the analyses. In addition to the ssGBLUP models, EBVs were calculated using AM-BLUP. Models were solved by MiX99 using iterative PCG algorithm. PCG iteration was assumed to be converged when the square root of relative difference between consecutive solutions was smaller than 1.0^{-5}.

Validation

Performance of AM-BLUP and ssGBLUP were studied by the Interbull GEBV validation test approach (Mäntysaari et al., 2010):

$$\mathbf{DRP} = b_0 + b_1(\mathbf{G})\mathbf{EBV} + \boldsymbol{e}$$

where DRP are deregressed proofs from the full data and EBV (or GEBV) are estimated breeding values (or genomic breeding values) from the reduced data. The validation reliability R^2 was the coefficient of determination of the above model divided by the reliability of DRP of the trait. Validation group contained 750 genotyped bulls for which the effective record contribution was over 10 based on full data but zero based on reduced data.

Results and Discussion

Convergence

A number of PCG rounds and computing times for the five analyses are in Table 1, and log10 of the convergence statistic can be seen in Figure 1. AM-BLUP model converged slowly and ssGBLUP model extremely slowly. Convergence was greatly improved by QP-transformation and it was improved even more by including inbreeding coefficients into $\mathbf{A}^{-1}$. Although GEBVs can be estimated using current data without APY, solving time was 8 to 9 times slower compared to AM-BLUP. By applying APY, computing time reduced to approximately 7 times slower than AM-BLUP.

Table 1. Number of PCG rounds, wall clock time in hours and time per round in seconds for the four analyses.

Model	PCG rounds	Time (h)	Time/ round (s)
AM-BLUP	2 420	5	7
ssGBLUP$_0$	16 282	220	49
ssGBLUP$_{QP}$	2 941	45	55
ssGBLUP$_{QP_Inb}$	2 373	41	62
ssGBLUP$_{QP_Inb_APY}$	2 573	34	47

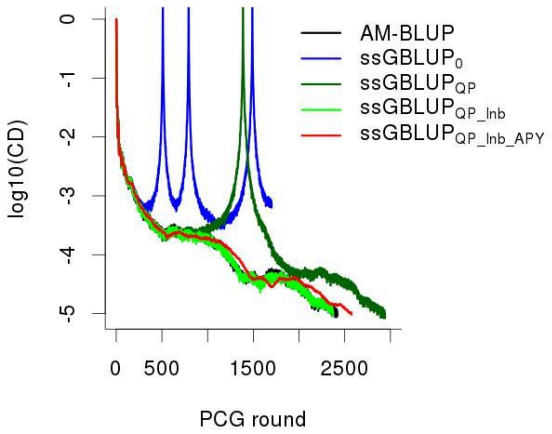

Figure 1. Convergence values plotted on the logarithmic scale during the four analyses. For ssGBLUP$_0$ model, first 1700 PCG rounds only.

Solutions

After QP-transformation, annual EBV and GEBV averages follow nicely each other (ICF2 as an example in Figure 2) and annual EBV and GEBV correlations were close to one for old animals, although decreased somewhat for young animals (ICF2 as an example in Figure 3).

Results from the APY approximated analyses corresponded well with the results from the original analyses. Correlations between GEBVs with and without APY were 1.00 among the core animals for all traits. Among non-core animals, correlations were between

0.998-0.999. Consequently, only validations results for ssGBLUP$_{QP_Inb_APY}$ are presented below.

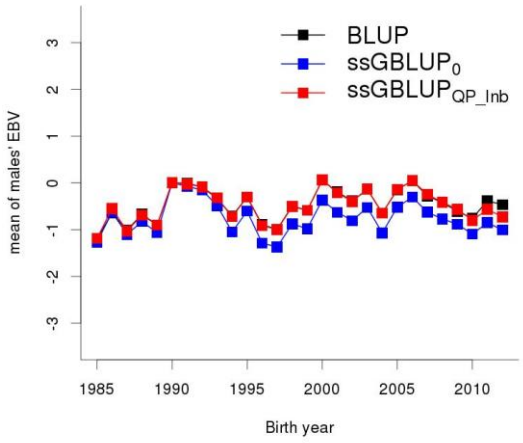

Figure 2. Males' annual EBV and GEBV averages for ICF2.

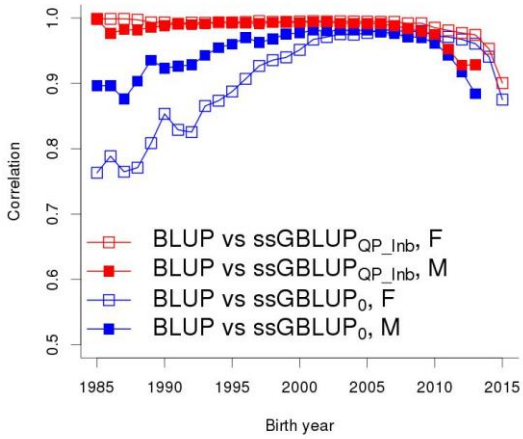

Figure 3. For ICF2, annual EBV and GEBV correlations for both females (F) and males (M).

Validation

Validation reliabilities (R^2) of DRPs of full data on the parent averages (EBV) and on the genomic enhanced breeding values (GEBV) based on reduced data are in Table 2 for all the traits. R^2 were low or moderate: 0.10-0.27 for EBVs and 0.22-0.31 for GEBVs. Increase in R^2 due to genomic information was moderate for the heifer traits, but clear for the cow traits. The increase was on average from 0.13 to 0.24 for NRR cow traits, from 0.18 to 0.29 for ICF cow traits, and from 0.18 to 0.30 for IFL cow traits.

Regression coefficients (b_1) of EBV and GEBV for all traits are in Table 2. The largest difference is in b_1 of the heifer traits, which were clearly higher for the EBV solutions than for the GEBV solutions (on average 1.03 v. 0.84). For the cow traits, b_1 were more similar for the two models. Especially for the third parity traits, b_1 were on average higher for GEBVs than EBVs.

Conclusions

The routine Nordic fertility evaluation using ssGBLUP was found feasible after the inbreeding coefficients and PPG had been correctly accounted. With these modifications, the number of iterations with ssGBLUP was comparable to animal model, but each iteration round took much longer computing time. APY-algorithm reduced the solving time with no effect on solutions. Model validation showed that ssGBLUP improved the fertility evaluations, especially for the cow traits.

Acknowledgements

The data for the analyses was kindly provided by Nordic Cattle Genetic Evaluation (NAV, Aarhus, Denmark).

References

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S. & Lawlor, T. J. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci. 93*, 743-752.

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

Fogh, A., Roth, A., Pedersen, O. M., Eriksson, J.-Å., Juga, J., Toivonen, M., Ranberg, I.

M. A., Steine, T., Nielsen, U. S. & Aamand, G. P. 2003. A joint Nordic model for fertility traits. *Interbull Bulletin 31*, 52-55.

Matilainen, K., Strandén, I., Aamand, G. P. & Mäntysaari, E. A. 2015. Single-step genomic evaluation for fertility in Nordic Red dairy cattle. In: 67th Annual Meeting of the European Federation of Animal Science, 29.8.-2.9.2016, Belfast, UK.

Misztal, I., Legarra, A., & Aguilar, I. 2014. Using recursion to compute the inverse of the genomic relationship matrix. *J. Dairy Sci. 97*, 3943-3952.

Misztal, I., Vitezica, Z. G., Legarra, A., Aguilar, I. & Swan, A. A. 2013. Unknown-parent groups in single-step genomic evaluation. *J. Anim. Breed. Genet. 130*, 252-258.

Muuttoranta, K., Tyrisevä, A.-M., Mäntysaari, E. A., Pösö, J., Aamand, G. P., Eriksson, J.-Å., Nielsen, U. S. & Lidauer, M. H. 2015. Genetic Parameters for Female Fertility in Nordic Dairy Cattle. *Interbull Bulletin 49*, 32-35.

Mäntysaari, E. A., Liu, Z. & VanRaden, P. 2010. Interbull validation test for Genomic evaluations. *Interbull bulletin 41*, 17-22.

Strandén, I., Matilainen, K., Aamand, G. P. & Mäntysaari, E. A. 2016. Solving efficiently large single-step genomic best linear unbiased prediction models. Submitted.

Table 2. Validation reliabilities (R^2) and regression coefficients (b_1) based on AM-BLUP (EBV) and ssGBLUP (GEBV) solutions, and change in reliabilities and in coefficients (Δ), for all traits.

Parity	Trait	R^2_{EBV}	R^2_{GEBV}	ΔR^2	b_{1EBV}	b_{1GEBV}	Δb_1
0	NRR	0.19	0.23	+0.04	1.00	0.81	-0.19
	IFL	0.27	0.29	+0.02	1.06	0.87	-0.19
1	NRR	0.16	0.27	+0.11	0.96	0.86	-0.10
	ICF	0.16	0.28	+0.12	0.99	0.90	-0.09
	IFL	0.17	0.31	+0.14	0.92	0.89	-0.03
2	NRR	0.12	0.24	+0.12	0.98	0.95	-0.03
	ICF	0.17	0.29	+0.12	0.88	0.86	-0.02
	IFL	0.16	0.29	+0.13	0.85	0.89	+0.04
3	NRR	0.10	0.22	+0.12	0.83	0.92	+0.09
	ICF	0.20	0.31	+0.11	0.92	0.90	-0.02
	IFL	0.20	0.31	+0.11	0.88	0.91	+0.03