

Article

Accounting for Missing Pedigree Information with Single-Step Random Regression Test-Day Models

Minna Koivula ^{1,*}, Ismo Strandén ¹ , Gert P. Aamand ² and Esa A. Mäntysaari ¹

¹ Natural Resources Institute Finland (Luke), 31600 Jokioinen, Finland; ismo.stranden@luke.fi (I.S.); esa.mantysaari@luke.fi (E.A.M.)

² NAV Nordic Cattle Genetic Evaluation, 8200 Aarhus, Denmark; gap@seg.es.dk

* Correspondence: Minna.koivula@luke.fi

Abstract: Genomic selection is widely used in dairy cattle breeding, but still, single-step models are rarely used in national dairy cattle evaluations. New computing methods have allowed the utilization of very large genomic data sets. However, an unsolved model problem is how to build genomic- (G) and pedigree- (A_{22}) relationship matrices that satisfy the theoretical assumptions about the same scale and equal base populations. Incompatibility issues have also been observed in the manner in which the genetic groups are included in the model. In this study, we compared three approaches for accounting for missing pedigree information: (1) GT_H used the full Quaas and Pollak (QP) transformation for the genetic groups, including both the pedigree-based and the genomic-relationship matrices, (2) GT_A₂₂ used the partial QP transformation that omitted QP transformation in G^{-1} , and (3) GT_MF used the metafounder approach. In addition to the genomic models, (4) an official animal model with a unknown parent groups (UPG) from the QP transformation and (5) an animal model with the metafounder approach were used for comparison. These models were tested with Nordic Holstein test-day production data and models. The test-day data included 8.5 million cows with a total of 173.7 million records and 10.9 million animals in the pedigree, and there were 274,145 genotyped animals. All models used VanRaden method 1 in G and had a 30% residual polygenic proportion (RPG). The G matrices in GT_H and GT_A₂₂ were scaled to have an average diagonal equal to that of A_{22} . Comparisons between the models were based on Mendelian sampling terms and forward prediction validation using linear regression with solutions from the full- and reduced-data evaluations. Models GT_H and GT_A₂₂ gave very similar results in terms of overprediction. The MF approach showed the lowest bias.

Keywords: ssGBLUP; ssGTBLUP; genomic evaluation; single-step; Holstein; genetic groups; metafounder



Citation: Koivula, M.; Strandén, I.; Aamand, G.P.; Mäntysaari, E.A. Accounting for Missing Pedigree Information with Single-Step Random Regression Test-Day Models. *Agriculture* **2022**, *12*, 388. <https://doi.org/10.3390/agriculture12030388>

Academic Editors: Heather Burrow and Michael Goddard

Received: 19 January 2022

Accepted: 8 March 2022

Published: 10 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Meuwissen et al. [1] introduced the genome-wide marker-assisted prediction model called genomic selection. Many alternative prediction models have been developed to use genomic selection in dairy cattle breeding [2]. Theoretically, the best model is single-step genomic BLUP (ssGBLUP) when phenotypes are available from both genotyped and non-genotyped individuals. The single-step approach offers a unified method for the analysis of all animals simultaneously [3,4]. Even though a decade has passed since the introduction of ssGBLUP, it is still not widely implemented in national dairy cattle evaluations.

Practical implementation of ssGBLUP has encountered different computational and modeling challenges. Some of the computational challenges related to very large genomic data sets can be overcome by using an alternative expression of the inverse genomic-relationship matrix or a model equivalent to ssGBLUP [1,5,6]. However, an unsolved modeling problem is how to build genomic- (G) and pedigree- (A_{22}) relationship matrices that satisfy the theoretical assumptions about the same scale and equal base populations [7]. Several methods were proposed that make G and A_{22} compatible. For example, base-population allele frequencies (AF) are used [8], and elements of G are scaled and centered

to have, on average, the same diagonal and off-diagonal elements as in \mathbf{A}_{22} [7,9]. Similar incompatibility issues were observed when genetic groups were included in the model [10].

Genetic groups can be included in a model as regression coefficients when the group proportions of an individual are traced to groups of unknown parents in the pedigree [11]. When the number of genetic groups is small, it is easy to explicitly include such groups in the model. However, this may lead to significant memory and computational costs especially in complicated models having many genetic groups [10]. A computationally more efficient approach with such models is to re-express genetic groups as unknown parent groups (UPG) resulting from the Quaas and Pollak (QP) transformation [12,13]. After the QP transformation, the genetic groups, by regression, are replaced by UPGs that extend the inverse-relationship matrix of all animals \mathbf{A}^{-1} in the mixed-model equations (MME) of the pedigree-based animal model.

The ssGBLUP uses relationship matrix \mathbf{H} , which includes both pedigree-based and genomic-based relationship matrix information. UPGs can be accounted for in many ways in the MME of ssGBLUP. If UPGs are included in \mathbf{A}^{-1} but not accounted for in \mathbf{A}_{22}^{-1} and \mathbf{G}^{-1} , severe convergence problems may arise, suggesting an incorrect model [14]. In many cases, this problem can be solved by properly accounting for genetic groups. The full QP transformation as described in [14,15] includes UPGs in \mathbf{A}^{-1} , \mathbf{A}_{22}^{-1} , and \mathbf{G}^{-1} in the MMEs of the single-step models. However, there is an alternative approach where the contributions of \mathbf{G}^{-1} are ignored in the QP transformation, or the altered QP \mathbf{H} inverse is used [16–18]. Ignoring \mathbf{G}^{-1} in the QP transformation can be justified by considering that the \mathbf{G} matrix contains all the information, whereas the pedigree information to build the \mathbf{A} matrix is incomplete and requires a UPG.

An alternative to the genetic groups is the use of metafounders (MF). The MF approach proposed by Legarra et al. [19] attempts to make the \mathbf{A}^{-1} and \mathbf{A}_{22}^{-1} matrices compatible with the \mathbf{G} matrix. The MF approach is based on the idea of using an allele frequency (AF) equal to 0.5 for all the markers when calculating the \mathbf{G} matrix [7]. The unknown parents are replaced by MF or pseudo-individuals with relationships and self-relationships that augment the \mathbf{A} matrix. The MFs are like UPGs but allow a related base-population or populations with nonzero inbreeding coefficients, e.g., as in [19,20]. The relationships within and between the MFs are modeled by a $\mathbf{\Gamma}$ matrix, which is used in forming the pedigree-based relationship matrix (\mathbf{A}^{Γ}).

When MFs are defined as the same as the genetic groups, the large number of UPGs that are common in the breeding-value prediction models of dairy cattle can make the estimation of the $\mathbf{\Gamma}$ matrix challenging [21]. Because the estimation of the $\mathbf{\Gamma}$ matrix requires the base-population AF, the large number of UPGs increases the chances that a UPG is associated with some rare allele genotypes, which can make the base-population AF estimate very uncertain. Thus, instead of estimating the base-population AF for all UPGs, a well-estimated $\mathbf{\Gamma}$ submatrix can be extended to a full $\mathbf{\Gamma}$ matrix using a covariance function, [22] as suggested by Kudinov et al. [23].

The aim of this study was to test different options to handle genetic groups in single-step models with the Nordic Holstein test-day (TD) data. We applied three different single-step TD models: (1) UPG by full QP transformation (GT_H), (2) UPGs in \mathbf{A}^{-1} and \mathbf{A}_{22}^{-1} or partial QP (GT_ \mathbf{A}_{22}), and (3) the MF approach (GT_MF). In addition to the genomic models, (4) an official pedigree-based animal model with a UPG by QP transformation and (5) a pedigree-based animal model with the metafounder approach were run for comparison.

2. Materials and Methods

2.1. Data

We used the official Nordic Holstein (HOL) milk production evaluation data obtained from the Nordic Cattle Genetic Evaluation (NAV). The data included TD records of milk, fat, and protein production from Denmark (DNK), Finland (FIN), and Sweden (SWE). The TD data included 8.8 million cows with a total of 173.7 million test-day records and a total of 447.5 million observations. The pedigree file had 10.9 million animals.

There were 274,145 genotyped animals, of which 75,802 were genotyped bulls (also including Holstein bulls from the EuroGenomics genotype exchange (EuroGenomics, 2020) and young bull calves), and 198,343 genotyped cows and heifers. Until 2019, bulls were genotyped using the BovineLD Bead Chip (Illumina, San Diego, CA, USA) (25% of all genotyped) and cows with the Eurogenomics custom LD chip (51% of all genotyped). Since 2019, all animals have been genotyped using the Eurogenomics EG MD chip (24% of all genotyped) [24]. After applying editing criteria for a minor allele frequency of 0.01 and a locus average GenCall score of 0.60, a total of 46,342 SNP markers on the 29 bovine autosomes were chosen for the analyses, and all the genotypes were imputed to this density. Genotype imputation was carried out using a family-and-population-based approach implemented in the FImpute program [25].

2.2. Models

Instead of the original ssGBLUP model using the \mathbf{H} matrix, we used the ssGTBLUP approach in the computations. The ssGTBLUP approach allows the key computations involving the \mathbf{G}^{-1} matrix to be replaced by efficient multiplications with a Woodbury matrix identity [26]. Thus, a dense \mathbf{T} matrix of size m by n is used instead of the dense \mathbf{G}^{-1} matrix of size n by n , where n is the number of genotyped animals and m is the number of SNP markers. Three different single-step models named GT_H, GT_A₂₂, and MF were used in the comparisons. In all models, the genomic-relationship matrix was as in VanRaden method 1 and included 30% residual polygenic proportion (RPG) and an AF of 0.5 for all markers. Earlier studies indicated that 30% RPG reduces the inflation of genomic evaluations more than models with smaller or larger RPGs (unpublished data). In the GT_H model, full QP transformation of genetic groups for \mathbf{A}^{-1} and \mathbf{A}_{22}^{-1} and \mathbf{G}^{-1} was used. The computations by ssGTBLUP were described in Koivula et al. [15]. In GT_A₂₂, the QP transformation was applied to \mathbf{A}^{-1} and \mathbf{A}_{22}^{-1} . The QP part for \mathbf{A}_{22} can be completed with an equivalent sparse formulation by reading the pedigree and including UPGs in the \mathbf{A}_{22} as described in Koivula et al. [15]. In GT_MF, the metafounder approach was used. In addition to the single-step models, an animal model with UPGs by QP transformation (EBV) and an animal model with metafounders (EBV_MF) were used for comparison and to observe the changes in predictions due to genomic information only. The pedigree inbreeding coefficients were accounted for in \mathbf{A}^{-1} and \mathbf{A}_{22}^{-1} , and in GT_MF, inbreeding coefficients were accounted for in corresponding inverse matrices \mathbf{A}^{Γ} and \mathbf{A}_{22}^{Γ} . The \mathbf{G} matrices in GT_H and GT_A₂₂ were scaled to have an average diagonal equal to the pedigree-based relationship matrix of the genotyped animals (\mathbf{A}_{22}).

Genetic groups and MFs were defined by the same logic. First, we defined fewer genetic groups than in the original NAV evaluation. The new groups were based on 4 breed groups (Holstein, red dairy cattle, Jersey, and 'other breeds') and 5 country-of-origin groups within the Holstein group (HOLDNK, HOLSWE, HOLFAN, HOLother, and HOLred). Second, each of these nine sources was further grouped by birth year decade and by selection path when appropriate. Thus, the original 446 groups were reduced to 176. These groups were considered random UPGs with variances equal to the genetic (co)variance in GT_H, GT_A₂₂, and EBV. The 176 groups were used as metafounders in GT_MF and EBV_MF.

The MF approach needs a covariance matrix for the metafounders, i.e., the $\mathbf{\Gamma}$ matrix. The MF self-relationship matrix $\mathbf{\Gamma}$ was defined using a covariance function (CF) model [22]. The $\mathbf{\Gamma}_0$ matrix of nine base MFs used values from [21]. In the 176 groups, a breed-specific linear time trend of a decade was assumed in the self-relationships, which were estimated using the base $\mathbf{\Gamma}_0$ matrix in the CF model. The known regression coefficients and design matrices in the CF model allow describing $\mathbf{\Gamma}$ matrices of any size, such as for our 176 groups, $\mathbf{\Gamma}_{176} = \mathbf{\Phi}_{176} \mathbf{K} \mathbf{\Phi}_{176}'$. We chose heuristic values for \mathbf{K} based on expectations from numerous descriptive analyses. For more formal analyses, see Kudinov et al. [23]. After solving the $\mathbf{\Gamma}_{176}$ -matrix, we computed the $\mathbf{\Gamma}$ -matrix-compliant inbreeding coefficients needed for the computations involving the inverses \mathbf{A}^{Γ} and \mathbf{A}_{22}^{Γ} when solving MMEs. As the MF

approach changes the relationship structure, Legarra et al. [19] derived a correction factor to be used to change the unrelated base-population-variance components to related base-population components. For our derived Γ matrix, the correction factor was close to one. Therefore, the same genetic-variance components were used in all models.

The study used the Nordic multiple-trait, reduced-rank, random regression TD model from the routine genetic evaluations for milk, fat, and protein in Finland, Sweden, and Denmark (a detailed description of the model can be found in Lidauer et al. [27]). Production records from the first three lactations are considered as nine different traits, with each having its own lactation curve. Therefore, the model had 27 traits in total: 3 countries, 3 yield traits, and 3 lactations. In this-reduced rank random regression TD model, each animal received 15 solutions to the random regression effects in the same way as in the official NAV evaluations. The TD-model solutions of genetic random regression coefficients were used to compile the total yields for milk, protein, and fat for the 305 d lactation [27] where the first, second, and third lactations had weights of 0.3, 0.25, and 0.45, respectively.

Comparison of models was based on forward prediction validation with solutions from the full- and the reduced-data evaluations. The reduced data were extracted from the full data by removing the last four years of observations from the full data. The linear regression validation (LR) method [28] was used for validation. The LR method compares predictions based on reduced and full data, which results in estimates of accuracy and bias. Candidate animals in the validation were selected according to their effective daughter contributions (EDC). Danish, Finnish, and Swedish (DFS) Holstein bulls born between the years 2010 and 2016 and having their EBV based on an $EDC \geq 3.0$ (corresponding to roughly 20 daughters) in the full-data set but an EDC of zero in the reduced-data set were defined as candidate bulls. This resulted in 524 candidate bulls for validation. The EDCs were calculated using the Interbull-EDC method tailored for the animal model by the ApaX99-program [29] for all bulls in the pedigree using both the full and the reduced data sets.

All MMEs of the TD models were solved by MiX99 software [30], which uses preconditioned conjugate gradient (PCG) iteration. The main computational cost in every iteration of the PCG method was due to the MME coefficient matrix times a vector product, where the most time-consuming computations were due to genomic information. To save memory and computing time, the inverse of the A_{22} matrix was not computed in advance, but instead, the computations used the method by Strandén et al. [31], which is based on sparse submatrices of A^{-1} by pedigree information. The PCG method was assumed to be converged when $C_r < 10^{-7}$, where C_r is defined as a Euclidean norm of the difference between the right-hand side (RHS) of the MME and the one predicted by the current solutions relative to the norm of RHS.

3. Results and Discussion

Average diagonal elements of the A_{22} , $A_{22}^{\Gamma^{176}}$, and G by birth year of genotyped animals are presented in Figure 1. The use of the Γ_{176} matrix lifted the diagonal elements of the A_{22} matrix close to those of the G matrix. Average inbreeding coefficients in A_{22} and G were 0.05 and 0.34, respectively. The difference is close to those reported earlier by VanRaden et al. [32] and Kudinov et al. [21]. The average inbreeding coefficient increased to 0.29 in $A_{22}^{\Gamma^{176}}$.

Wall clock times for preprocessing and solving MMEs are given in Table 1. The preprocessing, i.e., the computing time and the peak memory used in the construction of the T matrix for the G^{-1} matrix computations, did not differ considerably between the single-step models. For the MF model, there was an additional step of building the self-relationship matrix Γ . However, this step only marginally increased the total time.

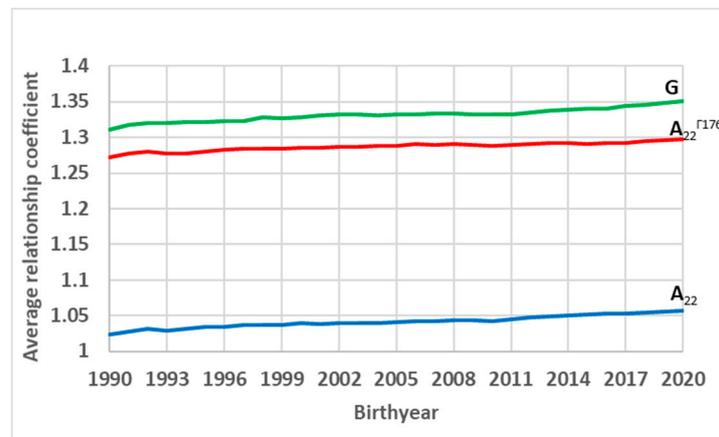


Figure 1. Average diagonal elements of the pedigree-relationship matrix of the genotyped animals (A_{22}), the genomic-relationship matrix constructed assuming that all allele frequencies were 0.5 (G), and the pedigree-relationship matrix of the genotyped animals augmented by Γ_{176} ($A_{22}^{\Gamma_{176}}$) presented by the birth year of the animal.

The number of PCG iterations to solve (G)EBV was 1227 for the animal model (EBV) and 1264 for the animal model with metafounders (EBV_MF). The total solving time to calculate EBV_MF was longer than for the animal model with UPGs by the QP transformation. Of the single-step models, GT_H with full QP needed 1019 iteration rounds for convergence, GT_A₂₂ with QP only in A^{-1} and A_{22} needed 1051 iteration rounds, and GT_MF with metafounders needed 1307 iteration rounds. Thus, the MF model required more iterations. However, because the time per iteration round in the GT_MF model was less than that in the GT_H and GT_A₂₂ models, the total computing time by GT_MF was lowest among the single-step models. Compared with the pedigree-based animal models, the single-step models required about 40–50% more time to calculate the solutions using the PCG iteration.

Table 1. Computing time (wall clock time in hours) and peak memory use (in GB) for an animal model with UPGs (EBV), an animal model with metafounders (EBV_MF), and different single-step models. The single-step models had unknown parent groups (UPG) with full QP (GT_H), UPGs with partial QP (GT_A₂₂), and metafounders (GT_MF). Convergence was assumed when $C_r < 10^{-7}$ (the norm of relative error in predicted RHS of MME).

	EBV	EBV_MF	GT_H	GT_A ₂₂	GT_MF
Building T matrix time (h)			23	20	20
Peak memory GB			208	207	207
Solving					
Iterations	1227	1264	1019	1051	1307
Seconds/PCG round	101	125	173	178	125
Time (h)	34	44	49	52	45
Peak memory GB	14.9	15.0	114.5	114.3	114.5
Total computing time (h)	34	44	72	72	65

Table 2 illustrates the LR-validation results from the different models for 524 DFS Holstein validation bulls. The level differences in the (G)EBV predictions were corrected by standardizing the (G)EBV so that the mean (G)EBV for cows born in 2007 was the same in all models. The b_0 column in Table 1 shows the mean difference (in kg) between the full- and reduced-data (G)EBV evaluations. This illustrates the realized bias. GT_MF showed a slightly smaller difference than the UPG models. The regression coefficients (b_1) showed the same trend as b_0 . The largest b_1 , i.e., the smallest overdispersion, was observed for GT_MF for all traits. Although there were no large differences in the coefficients of correlation (R^2) between the models, GT_MF had the highest R^2 values. The R^2 from

the LR validation can be interpreted as a reciprocal of the increase in reliability from the reduced-data evaluations to the full-data evaluations. Thus, the results indicate that, both in terms of bias and reliability, the MF model was slightly better than the other models. This conclusion is similar to that of other studies where the use of the MF model improved the single-step evaluations, e.g., [17,18]. Similarly, the MF approach seems to give better b_1 and R^2 values than the UPG model for the animal model without genomic information. Our results indicate that when the QP transformation is used, GT_H is as practical an alternative as GT_A22. This result is different from the results from other studies [16,18], where GT_A22, which was called altered H-inverse, had better predictive abilities than the full QP model GT_H.

Table 2. Bull linear regression validation (number of bulls = 524) results. Regression coefficients (b_1) and coefficients of correlation (R^2) from the models. The b_0 = mean (Full_(G)EBV—reduced_(G)EBV). The different models are an animal model with UPGs (EBV), an animal model with metafounders (EBV_MF), and different single-step models. The single-step models used UPGs with full QP (GT_H), UPGs with partial QP (GT_A22), and metafounders (GT_MF).

	Model	b_0	b_1	R^2
Milk	EBV	−101.7	0.84	0.32
	EBV_MF	−141.36	0.89	0.35
	GT_H	−319.8	0.87	0.67
	GT_A22	−315.2	0.87	0.67
	GT_MF	−272.3	0.89	0.68
Protein	EBV	0.80	0.74	0.24
	EBV_MF	0.58	0.82	0.27
	GT_H	−11.10	0.81	0.63
	GT_A22	−10.99	0.81	0.62
	GT_MF	−9.71	0.83	0.64
Fat	EBV	−2.18	0.73	0.23
	EBV_MF	−2.24	0.80	0.26
	GT_H	−16.16	0.82	0.64
	GT_A22	−15.61	0.81	0.63
	GT_MF	−14.67	0.85	0.65

In a genomic-selection program, the genomic pre-selection of bull calves [9,32] conflicts with the usual assumptions in the pedigree-based evaluations, which do not use genomic information. The pre-selected bulls are no longer a random sample of the progeny of their parents. This leads to an inflated mean for Mendelian sampling (MS) terms for these bulls and to a violation of normal assumptions in the pedigree-based animal model. Additionally, the MS term for the genotyped animals is likely to be different from zero with selective genotyping. In contrast, MS is expected to be zero when genotyping involves all young animals or is random.

Genomic selection allows selecting animals with superior MS. Consequently, the mean GEBV of all candidate animals is lower than for selected animals and their progeny [33]. The selected animals with many genotyped progenies are also more likely to have MS greater than zero [34]. The larger MS has an impact on genetic trends when animals are selected based on genomic information, especially if the selection happens before phenotypes are recorded [35].

Figure 2 shows the mean MS of genotyped bulls by birth year for protein using the reduced-data GEBV. Means are for DFS bulls and include all genotyped young bulls, those without daughters, and those that never entered AI service. In the reduced-data model, the bulls born after 2011 only have genomic information. There were no significant differences in the mean MS terms between the single-step models. The figure shows that for the youngest age classes, the difference is about 4 kg for different single-step models, whereas in both animal models, the MS term is zero, as expected. Before the start of genomic selection, the mean MS terms were quite stable. The mean MS was presumably below zero

because of the overprediction of bull dam EBVs. After genomic selection began to take effect, the mean MS also started to increase. In animal models, both approaches seem to give zero MS for the youngest age classes, but in the older bulls, the MS term in EBV_MF was not as negative as in the other models.

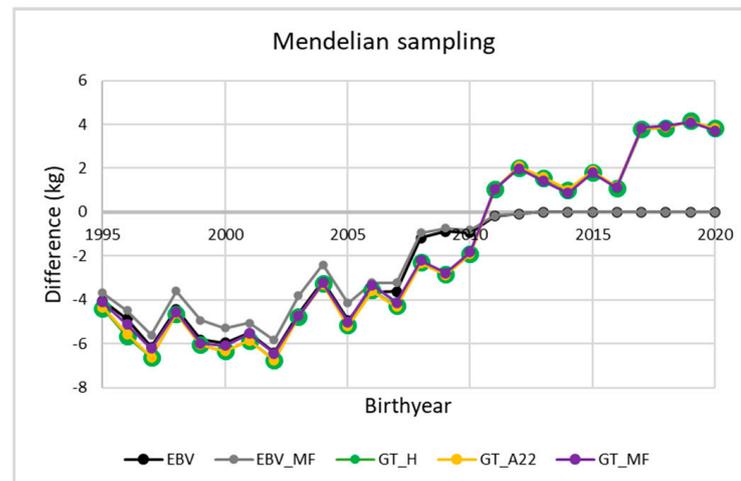


Figure 2. Mendelian sampling term means for protein for all genotyped DFS bulls by birth year calculated from (G)EBV from the reduced data. The different models are animal model (EBV), animal model with metafounders (EBV_MF), and three different single-step models by ssGTBLUP. The single-step models had unknown parent groups (UPG) with full QP (GT_H), UPGs with partial QP (GT_A22), and metafounders (GT_MF).

Figure 3 shows the genetic trend and yearly SD of protein (G)EBV for DFS Holstein bulls. Solid lines are from the full-data runs and dashed lines from the reduced-data runs. Except for the lower trend for animal-model EBVs, the trends from the single-step models were quite similar. However, for the MF model in the reduced-data set, the trend was slightly lower than in the other single-step models, indicating lower overprediction in the MF model compared with the UPG models. The same can be observed in estimates of b_0 in Table 2. The genetic trends in the official animal model (EBV) and the animal model using the MF approach (MF_EBV) were similar, as were their SD trends.

Based on all our comparisons, it seems that the traditional genetic group model and the MF model are both feasible options for handling genetic groups in single-step evaluations. The single-step MF model can be a more sophisticated way to combine pedigree and genomic information than the traditional single-step model with UPGs because genomic information affects both the genomic- and the pedigree-based relationship matrices in the MF model. Moreover, it seems that the MF model does not increase the trend of young, genotyped animals as much as the UPG-based single-step methods. The MF model also gives marginally better validation results compared with the other models. However, the current MF approach might still require some further development in building the Γ matrix.

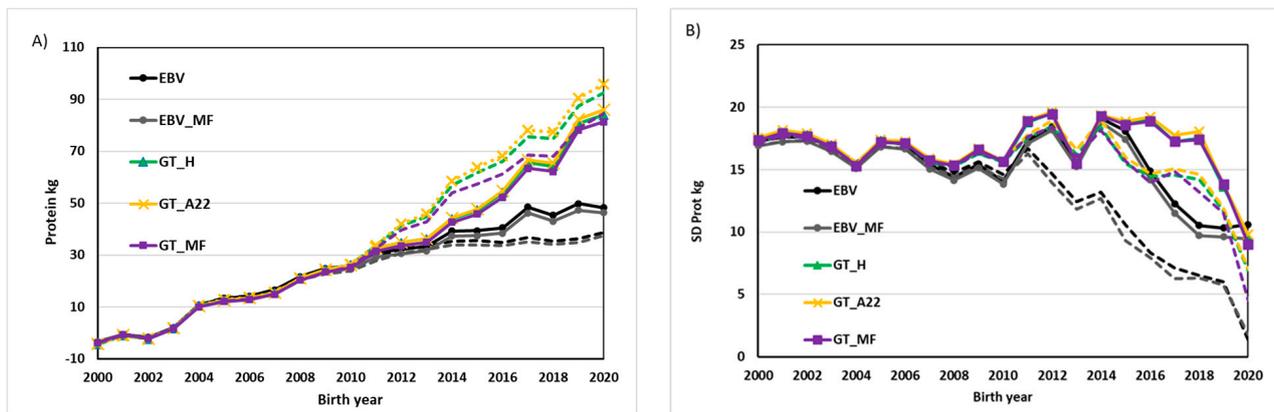


Figure 3. (A) Genetic trends for protein (G)EBV (kg) for the bulls presented by birth year averages. (B) SD for protein (G)EBVs (kg) by birth year. The different models are animal model (EBV), animal model with metafounders (EBV_MF), and three different single-step models by ssGTBLUP. The single-step model had UPGs with full QP (GT_H), UPGs with partial QP (GT_A22), and metafounders (GT_MF). Solid lines are means and SDs for full-data trends, and dashed lines are for reduced-data trends.

4. Conclusions

Both the traditional UPG models and the MF approach can be implemented efficiently in single-step models with large genomic data sets. In our study, the MF approach had a lower bias than the UPG models. Moreover, when the QP transformation was used to arrive at a UPG model, the results from the full QP transformation were similar to the partial QP transformation in which the G^{-1} contributions were not included in the UPG computations. The mean MS by birth year was positive for the genotyped bulls during the last decade according to the single-step models but close to zero in the pedigree-based animal model. Selective genotyping can explain some of the positive mean MS values, but the most recent years need further investigation.

Author Contributions: Conceptualization and original draft preparation, M.K.; writing, review, and editing, E.A.M., I.S. and G.P.A. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the following Nordic cattle-breeding organizations: Viking Genetics and Nordic Cattle Genetic Evaluation.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work was a part of the Genomics in BLUP project. The Nordic cattle-breeding organizations: Viking Genetics (Randers, Denmark), Nordic Cattle Genetic Evaluation (NAV, Aarhus, Denmark), and Faba (Hollola, Finland) are acknowledged for providing the genotype data and the test-day data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829. [[CrossRef](#)] [[PubMed](#)]
2. VanRaden, P.M. Symposium review: How to implement genomic selection. *J. Dairy Sci.* **2020**, *103*, 5291–5301. [[CrossRef](#)] [[PubMed](#)]
3. Aguilar, I.; Misztal, I.; Johnson, D.-L.; Legarra, A.; Tsuruta, S.; Lawlor, T.J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* **2010**, *93*, 743–752. [[CrossRef](#)]
4. Christensen, O.F.; Lund, M.S. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* **2010**, *42*, 2. [[CrossRef](#)] [[PubMed](#)]

5. Mäntysaari, E.A.; Koivula, M.; Strandén, I. Symposium review: Single-step genomic evaluations in dairy cattle. *J. Dairy Sci.* **2020**, *103*, 5314–5326. [[CrossRef](#)]
6. Misztal, I.; Lourenco, D.; Legarra, A. Current status of genomic evaluation. *J. Anim. Sci.* **2020**, *98*, 1–14. [[CrossRef](#)]
7. Christensen, O.F. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genet. Sel. Evol.* **2012**, *44*, 37. [[CrossRef](#)]
8. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [[CrossRef](#)]
9. Vitezica, Z.G.; Aguilar, I.; Misztal, I.; Legarra, A. Bias in genomic predictions for populations under selection. *Genet. Res.* **2011**, *93*, 357–366. [[CrossRef](#)]
10. Misztal, I.; Vitezica, Z.G.; Legarra, A.; Aguilar, I.; Swan, A.A. Unknown-parent groups in single-step genomic evaluation. *J. Anim. Breed. Genet.* **2013**, *130*, 252–258. [[CrossRef](#)]
11. Thompson, R. The estimation of heritability with unbalanced data: II. Data available on more than two generations. *Biometrics* **1979**, *33*, 497–504. [[CrossRef](#)]
12. Westell, R.A.; Quaas, R.L.; Van Vleck, L.D. Genetic groups in an animal model. *J. Dairy Sci.* **1988**, *71*, 1310–1318. [[CrossRef](#)]
13. Quaas, R.L.; Pollak, E.J. Modified equations for sire models with groups. *J. Dairy Sci.* **1981**, *64*, 1868–1872. [[CrossRef](#)]
14. Matilainen, K.; Strandén, I.; Aamand, G.P.; Mäntysaari, E.A. Single step genomic evaluation for female fertility in Nordic Red dairy cattle. *J. Anim. Breed. Genet.* **2018**, *135*, 337–348. [[CrossRef](#)] [[PubMed](#)]
15. Koivula, M.; Strandén, I.; Aamand, G.P.; Mäntysaari, E.A. Practical implementation of genetic groups in single-step genomic evaluations with Woodbury matrix identity based genomic relationship inverse. *J. Dairy Sci.* **2021**, *104*, 10049–10058. [[CrossRef](#)]
16. Tsuruta, S.; Lourenco, D.A.L.; Masuda, Y.; Misztal, I.; Lawlor, T.J. Controlling bias in genomic breeding values for young genotyped bulls. *J. Dairy Sci.* **2019**, *102*, 9956–9970. [[CrossRef](#)] [[PubMed](#)]
17. Cesarani, A.; Masuda, Y.; Tsuruta, S.; Nicolazzi, E.K.; VanRaden, P.M.; Lourenco, D.; Misztal, I. Genomic predictions for yield traits in US Holsteins with unknown parent groups. *J. Dairy Sci.* **2021**, *104*, 5843–5853. [[CrossRef](#)]
18. Masuda, Y.; Tsuruta, S.; Bermann, M.; Bradford, H.L.; Misztal, I. Comparison of models for missing pedigree in single-step genomic prediction. *J. Anim. Sci.* **2021**, *99*, 1–10. [[CrossRef](#)]
19. Legarra, A.; Christensen, O.F.; Vitezica, Z.G.; Aguilar, I.; Misztal, I. Ancestral relationships using metafounders: Finite ancestral populations and across population relationships. *Genetics* **2015**, *200*, 455–468. [[CrossRef](#)]
20. Garcia-Baccino, C.; Legarra, A.A.; Christensen, O.F.; Misztal, I.; Pocrnic, I.; Vitezica, Z.G.; Cantet, R.J. Metafounders are related to Fst fixation indices and reduce bias in single-step genomic evaluations. *Genet. Sel. Evol.* **2017**, *49*, 34. [[CrossRef](#)]
21. Kudinov, A.A.; Mäntysaari, E.A.; Aamand, G.P.; Uimari, P.; Strandén, I. Metafounder approach for single-step genomic evaluations of Red Dairy cattle. *J. Dairy Sci.* **2020**, *103*, 6299–6310. [[CrossRef](#)] [[PubMed](#)]
22. Kirkpatrick, M.; Hill, W.G.; Thompson, R. Estimating the covariance structure of traits during growth and ageing, illustrated with lactation in dairy cattle. *Genet. Res.* **1994**, *64*, 57–69. [[CrossRef](#)] [[PubMed](#)]
23. Kudinov, A.A.; Koivula, M.; Strandén, I.; Aamand, G.P.; Mäntysaari, E.A. Single-step genomic predictions of a minor breed concurrently with a main breed large national genomic evaluation. *Interbull Bull.* **2021**, *56*, 174–179.
24. EuroGenomics. A European Network for a Reliable Cattle Breeding. 2022. Available online: https://www.eurogenomics.com/?rub=88&unce_contentus_webclient=0&view=afficher_elasticsearch_results&submitted=1&q=A+European+Network+for+a+Reliable+Cattle+Breeding.+2022 (accessed on 10 January 2022).
25. Sargolzaei, M.; Chesnais, I.P.; Schenkel, F.S. A new approach for efficient genotype imputation using information from relatives. *BMC Genom.* **2014**, *15*, 478. [[CrossRef](#)]
26. Mäntysaari, E.A.; Evans, R.D.; Strandén, I. Efficient single-step genomic evaluation for a multibreed beef cattle population having many genotyped animals. *J. Anim. Sci.* **2017**, *95*, 4728–4737. [[CrossRef](#)]
27. Lidauer, M.; Pösö, J.; Pederson, J.; Lassen, J.; Madsen, P.; Mäntysaari, E.A.; Nielsen, U.; Eriksson, K.-Å.; Johansson, K.; Pitkänen, T.; et al. Across-country test-day model evaluations for Nordic Holstein, Red Cattle and Jersey. *J. Dairy Sci.* **2015**, *98*, 1296–1309. [[CrossRef](#)]
28. Legarra, A.; Reverter, A. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel. Evol.* **2018**, *50*, 53. [[CrossRef](#)]
29. Strandén, I.; Lidauer, M.; Mäntysaari, E.A.; Pösö, J. Calculation of Interbull weighting factors for the Finnish test day model. *Interbull Bull.* **2001**, *26*, 78–79.
30. Strandén, I.; Lidauer, M. Solving large mixed models using preconditioned conjugate gradient iteration. *J. Dairy Sci.* **1999**, *82*, 2779–2787. [[CrossRef](#)]
31. Strandén, I.; Matilainen, K.; Aamand, G.P.; Mäntysaari, E.A. Solving efficiently large single-step genomic best linear unbiased prediction models. *J. Anim. Breed. Genet.* **2017**, *134*, 264–274. [[CrossRef](#)]
32. VanRaden, P.M.; Olson, K.M.; Wiggans, G.R.; Cole, J.B.; Tooker, M.E. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. *J. Dairy Sci.* **2011**, *94*, 5673–5682. [[CrossRef](#)] [[PubMed](#)]
33. Tyrisevä, A.-M.; Mäntysaari, E.A.; Jakobsen, J.; Aamand, G.P.; Dürr, J.; Fikse, W.F.; Lidauer, M.H. Detection of evaluation bias caused by genomic preselection. *J. Dairy Sci.* **2018**, *101*, 3155–3163. [[CrossRef](#)]

34. Masuda, Y.; VanRaden, P.M.; Misztal, I.; Lawlor, T.J. Differing genetic trend estimates from traditional and genomic evaluations of genotyped animals as evidence of preselection bias in US Holsteins. *J. Dairy Sci.* **2018**, *101*, 5194–5206. [[CrossRef](#)] [[PubMed](#)]
35. Abdollahi-Arpanahi, R.; Lourenco, D.; Misztal, I. Detecting effective starting point of genomic selection by divergent trends from best linear unbiased prediction and single-step genomic best linear unbiased prediction in pigs, beef cattle, and broilers. *J. Anim. Sci.* **2021**, *99*, 1–11. [[CrossRef](#)] [[PubMed](#)]