Using Pseudo-observations to combine Genomic and Conventional Data in the Dutch-Flemish National Evaluation

W.M. Stoop, H. Eding, M.L. van Pelt, L.C.M. de Haer, G. de Jong
CRV, P.O. Box 454, 6800 AL Arnhem, The Netherlands
E-mail: Marianne.Stoop@crv4all.com

Abstract
Since 2010, the Dutch-Flemish evaluation use a post-processing step to estimate Genomically Enhanced Breeding Values in their national evaluation. To optimize the use of genomic data, this paper describes a method where Direct Genomic Values (DGV) are transformed into equivalent daughter performances (pseudo-records), and analyzed in a multi-trait animal model as pseudo-trait to the conventional trait-of-interest with a heritability close to unity, and a genetic correlation set to the predictive value of the DGV. This method can account for pre-selection of data, allows for an animal model in which data can flow to relatives, is easy to incorporate technically, and avoids problems of genotype ownership, by directly using DGV. Results for overall conformation traits are shown as application of this method.

1. Introduction

Since 2010, the Dutch-Flemish evaluation use a post-processing step (Van Raden et al., 2009) to estimate Genomically Enhanced Breeding Values (GEBV) in their national evaluation. This method blends Direct Genomic Values (DGV), derived from deregressed proofs and SNP information, and conventional EBV. An obvious drawback of this method is that genomic information of an individual does not influence the breeding value of relatives.

Several methods have been described to use genomic information in a more appropriate way (e.g. Misztal *et al.*, 2009, Forni *et al.*, 2011, and Ducrocq and Liu, 2009), of which especially the method where DGV are transformed into equivalent daughter performances (pseudo-records) is appealing, because 1) it corrects for pre-selection of young bulls (including all DGV avoid bull dams to be severely over-estimated due to selected testing of their best sons; Patry et al. 2013), 2) it effectively drops genomic information down the pedigrees of the trait of interest, influencing relatives, 3) existing BLUP software is easily extended with an extra pseudo-trait, and is computationally feasible, 4) pseudo-records allow for the inclusion of DGV, rather than genotypes, and is therefore compatible with the current structure of reference populations and ownership of data, where genotypes are not freely exchanged.

The genomic part of the DGV is a more or less simple summation of SNP effects, and the observed variation in these SNP between animals is fully explained by their genetic makeup. Hence the genomic part of the DGV has a heritability close to unity. Thus, pseudo-records for a certain trait of interest can be analyzed as a correlated trait – the pseudo-trait – to the original trait of interest.

This paper describes a method to combine genomic and conventional data, and shows – as an example - the results and validation for overall conformation traits.

2. Material and Methods
2.1 GEBV
The used method is based on Mantysaari and Strandén (2010). In this method, a two-step approach is used in which: 1) a DGV is calculated in a genomic evaluation (de Roos *et al.*, 2009), and 2) the resulting DGV of an animal is transformed into a pseudo-observation (PSR) on absolute scale for a pseudo-trait with h^2 of 0.999, which is subsequently included in the national evaluation. The statistical model for analysing the pseudo-trait is:

$$y_i = \mu + anim_i + e_{ij}$$

Where y_i is the record of animal i. The μ fits the general level of the pseudo-observations and $anim_i$ is the genetic animal effect.

Data is analyzed in a multi-trait setting in which information from the pseudo-trait flows to the original conventional trait through the estimated genetic correlation, resulting directly in a genomically enhanced breeding value (GEBV). In multitrait systems with more than one pseudotrait, covariances between pseudo-traits are equal to those between conventional traits. Genetic variance of the pseudo-trait is assumed equal to the conventional trait.

Table 1. *Estimated reliabilities for DGV (R^2_{DGV}) and Pedigree Index (R^2_{PI}) from a DGV validation study, average reliability of the cohort validation bulls (R^2_{val}), and resulting genetic correlation between conventional trait and pseudo trait. Conventional heritability can be used to convert R^2_{DGV} and R^2_{PI} to EDC.*

Trait	h^2	R^2_{DGV}	R^2_{PI}	R^2_{val}	$r_{ebv,psr}$
Frame	0.350	0.420	0.186	0.882	0.613
Dairy Strength	0.110	0.490	0.331	0.802	0.630
Udder	0.336	0.478	0.201	0.889	0.670
Feet&Legs	0.167	0.454	0.279	0.797	0.621

The genetic correlation between pseudo-trait and original conventional trait is estimated in a DGV validation study (Table 1; Schrooten, 2012). In this DGV validation 4 years of data is deleted and the resulting Male Pedigree Index (PI) and DGV (of young validation bulls which no longer have daughter data) are both correlated to their realized daughter EBV (based on the 4 years of deleted data): r_{pi} and r_{dgv}.

The squared correlations give a measure for the reliabilities. Using the EDC equivalents an added reliability can be calculated for the DGV, which is then adjusted for the mean reliability of the validation cohort (R^2_{val}):

$$R^2_{dgv} => EDC_{dgv}, \text{ and } R^2_{pi} => EDC_{pi}$$

$$EDC_{add} = EDC_{dgv} - EDC_{pi}$$

$$EDC_{add} \sim R^2_{add}$$

$$r_{ebv,psr} = (R^2_{add}/R^2_{val})^{1/2}$$

For step 1, the genomic evaluation, data of the Eurogenomics consortium is used, consisting of approximately 24,000 bulls in the reference population for conformation traits. For the DGV validation study, 836 bulls are used as validation cohort (and thus no longer part of the reference population).

For step 2, the national evaluation, both conventional data (conformation scores of daughters) and DGV transformed to pseudo-observations, are included. Edited data comprised 5,411,754 records, including 27,373 pseudo-observations, of 5,407,253 animals with pedigree.

Results show the changes in (G)EBV compared to the EBV, and compared to the former GEBV, estimated from a blending procedure.

2.2 Validation of GEBV
To validate the GEBV and estimated GEBV reliabilities, a comparison is made between a run with only conventional data (only step 2: a classic EBV run), and a run where 4 years of data is deleted (both in step 1: the genomic evaluation, as in step 2: the national evaluation with conventional and DGV data, resulting for the 836 validation bulls in a GEBV based on a male pedigree index and DGV data, but no daughter performances; a full run).

The realized reliability of the estimated GEBV for the validation cohort can be derived from the correlation between EBV (classic run) and GEBV (full run) ($r_{ebv,gebv}$), and the reliability of the EBV (R^2_{gebv}):

$$R^2_{gebv} = (r_{ebv,gebv} /\sqrt{ R^2_{ebv} })^2$$

Table 2. *Comparison between GEBV and EBV: mean difference in breeding value GEBV-EBV (dBV) and reliability (dREL), and the correlation between EBV and GEBV (corBV). Differences in breeding values are absolute differences in relative breeding values ~N (100, 4) for bulls with at least 10 daughters and/or own pseudo-observation.*

Trait	No PSR, > 10 dau (N=15.614)			PSR, >10 dau (N=4.720)			PSR, no dau (N=2.692)		
	dBV	corBV	dREL	dBV	corBV	dREL	dBV	corBV	dREL
Frame	-0.24	0.998	0.16	0.09	0.992	8.06	0.04	0.922	28.42
Dairy Strength	0.52	0.997	0.35	0.01	0.986	8.87	0.00	0.901	23.45
Udder	0.02	1.000	0.04	0.11	0.988	5.77	0.48	0.878	28.88
Feet&Legs	-0.08	1.000	0.05	0.06	0.985	5.59	0.24	0.882	20.19

Table 3a. *Comparison between GEBV from blending method and from PSR method for young bulls without daughters (N=4.729): mean GEBV$_{PSR}$, mean GEBV$_{BLEND}$, difference and correlation. Breeding values are relative breeding values ~N (100, 4).*

Trait	PSR	BLEND	diff	corr
Frame	104.0	103.6	0.4	0.861
Dairy Strength	103.8	103.1	0.7	0.920
Udder	105.7	105.3	0.4	0.951
Feet&Legs	103.3	102.1	0.2	0.922

Table 3b. *Comparison between GEBV reliabilities from blending method and from PSR method for young bulls without daughters (N=4.729): mean reliabilities for GEBV$_{PSR}$ and GEBV$_{BLEND}$, and difference.*

Trait	PSR	BLEND	diff
Frame	64.5	53.8	10.7
Dairy Strength	57.7	53.8	3.9
Udder	66.2	63.3	2.9
Feet&Legs	55.3	52.1	3.2

This realized reliability should be in line with the estimated reliability from the national evaluation.

3. Results
3.1 GEBV
Table 2 shows the changes from EBV to GEBV for 3 groups of bulls: those with daughters but no PSR, those with daughters and PSR, and those with PSR but no daughters (Table 2). For bulls without own PSR only minor differences between EBV and GEBV are observed, caused by the influence of genotyped relatives. For bulls with own PSR and daughters, differences are small. Depending on the amount of daughter data available, the PSR still adds some information to the system, resulting in an increase in reliability of 3 to 6 percent. Young bulls without daughters obviously benefit the most from their PSR, resulting in considerable re-ranking in breeding values (correlations 0.88-0.92), and in a major increase in reliability of 20 to 29 percent.

Tables 3a and 3b show the changes in young bulls (with PSR but no daughters) between the blending method GEBV and the new PSR method GEBV. For bulls with daughters hardly any differences are observed (results not shown). For young bulls there are small differences in breeding values, caused by information from relatives that can now (slightly) influence their own observation (Table 3a). Reliabilities increase considerably with the PSR method, due to the multi-trait setting (Table 3b). The GEBV validation (paragraph 3.2) shows this increase is genuine.

3.2 Validation of GEBV
Figure 1 shows the result of the GEBV validation. The reliabilities calculated in the old blending method seem underestimated. Reliabilities from the new PSR method are more in line with the realized reliability (calculated from the correlation between EBV and GEBV and the EBV reliability, based on 836 bulls).

4. Discussion
The increase in GEBV reliability from blending to new PSR method is mainly caused by the multitrait setting in the latter method. The validation shows that this increase in reliability is mostly genuine, and that it seems no artefact of double counting; for both linear and overall conformation traits together, the average difference between PSR reliability and realized reliability was -0.0026 on a scale from 0.0 to 1.0 (results not shown). For the blending method, reliability was underestimated for all traits.

One of the advantages of the currently used blending method is that it avoids bias by using a male pedigree index and DGV data, and no data on bull dams. Although the PSR method partly accounts for this pre-selection, it fails to fully counter bias (based on empiric data). To avoid bias in conformation and production traits, young bulls with a PSR and no daughters will have their parent average replaced by a male pedigree index, and their reliabilities adapted to this change.

5. Conclusion
The PSR method is a feasible method to more effectively include genomic information in the national evaluation. It can account for pre-selection, it allows for an animal model in which data can flow to relatives, it is easy to incorporate the genomic data in the national evaluation, and avoids problems of genotype ownership, by directly using DGV. In the Netherlands, PSR are available for all traits analyzed in a multi-trait animal model. In conclusion, The PSR method is a good interim solution before a large-scale one-step approach is technically and politically feasible.

6. References

Schrooten, C., 2012, Method for computation of GEBV reliability in the Netherlands, presented at Interbull Technical Workshop on International Genomic Evaluations in Verona, Italy, 2012

De Roos, A.P.W., C. Schrooten, E. Mullaart, S. van der Beek, G. de Jong, and W. Voskamp, 2009, Genomic Selection at CRV, Interbulletin 39, 47-50

Ducrocq V. and Z. Liu, 2009, 'Combining genomic and classical information in national BLUP evaluations', Interbull bulletin 40, pp 172-177

Forni S., Aguilar I., and Misztal I., 2011, 'Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information' Genetics Selection Evolution 2011, 43:1

Mäntysaari E.A., and Strandén I., 2010, 'Use of bivariate EBV-DGV model to combine genomic and conventional breeding value evaluations', Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany

Misztal I., Aguilar I., Johnson D., Legarra A., Tsuruta S., and Lawlor T.J., 2009, 'Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information', Interbull bulletin 40, pp 240-244

Patry, C, H, Jorjani, V. Ducrocq, 2013, 'Effects of a national genomic preselection on the international genetic evaluations', J. Dairy Sci. 2013. 96(5):3272-3284

VanRaden P.M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor and F. S. Schenkel, 2009, '*Invited Review:* Reliability of genomic predictions for North American Holstein bulls', J. Dairy Sci. 2009. 92:16-24.

Figure 1. *Mean reliabilities of GEBV for overall conformation traits for validation cohort; for blending method, PSR method, and realized reliability of $GEBV_{PSR}$ from GEBV validation study (based on bulls in validation cohort with > 25 daughters)*