

Bayesian Inclusion of External Evaluations into a National Evaluation System: Application to Milk Production Traits

N. Gengler^{1,2} and S. Vanderick¹

¹ Animal Science Unit, Gembloux Agricultural University, B-5030 Gembloux, Belgium

² National Fund for Scientific Research, B-1000 Brussels, Belgium

1. Introduction

For beef cattle Klei *et al.* (1996) proposed using a Bayesian approach to integrate estimated breeding values (EBV) from external genetic evaluations, in their case from other breeds, into genetic evaluations of a given breed. Surprisingly genetic evaluation systems in dairy cattle are not using currently this type of approach even if the context could be considered similar. Indeed, we have often rather small scale national or local evaluations and international evaluations provided by Interbull summarizing results for sires in dozens of populations. Similarly one can expect that in the future additional external sources of information, as genomic EBV, will be available for national or local genetic evaluation systems, therefore development of appropriated procedures is becoming even more important.

Recently Gengler and Verkenne (2007) presented a Bayesian approach to integrate molecular a priori known information into genetic evaluations. Their formulas were very similar to the one presented by Legarra *et al.* (2007) for a setting similar to that presented by Klei *et al.* (1996). The main difference was that Gengler and Verkenne considered that molecular information was known without error. Another issue that prevents immediate adaptation of the strategies used for beef evaluations is that external evaluations as Interbull evaluations are not independent from national or local evaluations. Finally for a certain number of traits the reported breeding value is not necessarily the evaluated one and therefore other adaptations need to be made.

The objective of this study was developing appropriated procedures to allow Bayesian inclusion of external evaluations in dairy cattle evaluations as a new and innovative way to make best use of external evaluations.

2. Material and Methods

2.1. Estimating external (international) contributions to Interbull EBV

Most current standard procedures to combined EBV of different sources are based on a meta-analyzes type of approach. VanRaden (2001) gave a comprehensive overview of several available methods based on selection index (SI) type of procedures. Current procedures used in the Walloon region of Belgium are based on the selection index procedures presented by VanRaden. Conceptually these SI procedures can also be modified for a slightly different purpose. Indeed for a given sire Interbull EBV (u_i) contain both local population (u_L) and other sources (u_E) of information, only this being external. The value u_E can be considered being an unknown breeding value on the same scale as u_L summarizing all the information from other populations. Therefore its theoretical correlation to u_L should be 1 and it is considered having theoretically the same heritability, variance and mean. However there is no estimate available for u_E and it has to be obtained by appropriate SI procedures where the information vector is based on u_i that contain both local population and u_L that contains only local information

$$\hat{u}_E = \mathbf{c}' \mathbf{V}^{-1} \begin{bmatrix} \hat{u}_L \\ \hat{u}_I \end{bmatrix} \text{ with}$$

$$\mathbf{c}' = [\text{cov}(\hat{u}_E, \hat{u}_L) \quad \text{cov}(\hat{u}_E, \hat{u}_I)]$$

$$\mathbf{V} = \begin{bmatrix} \text{var}(\hat{u}_L) & \text{cov}(\hat{u}_L, \hat{u}_I) \\ \text{cov}(\hat{u}_I, \hat{u}_L) & \text{var}(\hat{u}_I) \end{bmatrix}$$

Needed variances and covariance are based on daughter equivalents (DE) DE_L , DE_I and DE_E defined as $DE_I - DE_L$ and trait variances.

In our implementation DE were backsolved from reliabilities and EDC not directly used. This was also necessary as EDC for u_i are not provided. DE_E will be needed to compute associated reliabilities (REL_E). VanRaden (2001) showed appropriate formulas also stressing that subtracting parent average PA from EBV allows to remove different genetic trends. In our situation this is a somewhat a philosophical problem because theoretically the trends should be identically. Also it is no easy to choose the right PA especially for external EBV. The final formula used in this implementation was $u_E = w_L u_L + w_I u_I$. An alternative implementation that was not yet tested, would be to compute theoretical coefficients to predict u_i from u_L and u_E using $u_i = v_L u_L + v_E u_E$ and backsolve for unknown u_E $u_E = (u_i - v_E u_E) / v_L$.

2.2. Transforming external (international) contributions to locally evaluated EBV

Unfortunately evaluated traits are not necessarily those reported to Interbull and send back on that scale. Therefore in practice an additional step might be necessary converting external EBV. An example is milk, fat and protein yields were the trait reported to Interbull is for most countries the average of lactation yields for mostly three lactations. However mostly test-day records are evaluated (u_{lact}). Also genetic effects are random regressions. In such a situation, elements of the real local genetic solution vector (u_L) have to be chosen in an appropriate way to reflect $u_{lact,0}$. For the Walloon genetic evaluation system and in this implementation, for yields the constant effect can be considered affected only which means that we assume no *a priori* knowledge of lactation curve shape parameters. Differentiating between first second and third lactation can be based on expected differences between EBV for those three lactations. In our implementation we considered all lactations as being equally affected. Also any base and expression of results change have to be taken care of. In general these steps need to be fine tuned for any routine implementation of this method.

2.3. Mixed model equations for Bayesian integration

Modifications to standard mixed model equations for a local genetic evaluation with a generic mixed model ($y = Xb + Zu + e$) are small. We follow here the notation and set up as presented by Legarra *et al.* (2007):

$$\begin{bmatrix} X'_L R^{-1} X_L & X'_L R^{-1} Z_L \\ Z'_L R^{-1} X_L & Z'_L R^{-1} Z_L + G^{*-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'_L R^{-1} y_L \\ Z'_L R^{-1} y_L + \theta \end{bmatrix}$$

where y_L is a vector of locally available data, X_L and Z_L are incidence matrices in the local linking effects and y_L , b is a vector of fixed effects, u is a vector of random effects, external animal, that can also have local records, and exclusively locals animals will be considered sorted. Two basic changes are done where new elements can be defined as

$$G^{*-1} = \begin{bmatrix} D^{-1} - G_{EE}^{-1} + G^{EE} & G^{EL} \\ G^{LE} & G^{LL} \end{bmatrix} \quad \text{and} \\ \theta = \begin{bmatrix} D^{-1} \hat{u}_E \\ \mathbf{0} \end{bmatrix}. \quad \text{Also we can define}$$

$\text{Var}(u_E) = G_{EE}$ and $\text{PEV}(\hat{u}_E) = D$. Legarra *et al.* (2007) showed that conceptually $D^{-1} - G_{EE}^{-1}$ can be considered being the least square part of the left-hand side and $D^{-1} \hat{u}_E$ can be considered being the right-hand side of the theoretical BLUP equations that provided \hat{u}_E . In reality especially the complete D matrix is unknown as prediction error covariances are not known and D needs to be replaced by a diagonal matrix of PEV for every external evaluation. Similarly G_{EE} needs to be simplified ignoring all relationships. Even if basic equations are simple additional non-trivial assumptions have to be made. First this consists in the nature of the theoretical external evaluation. In fact the external EBV can have any origin, however in our implementation we tried to consider them coming from a completely similar theoretical evaluation using

the same variances as in the local genetic evaluation. This is not required; however it is the simplest and most logical assumption.

2.4. Application to milk production traits

Data provided by the Walloon Breeding Association (AWE) for the routine evaluation for milk production of January 2008 were used and included 14,358,255 test-day records for 814,596 dairy and dual-purpose cows in production. The pedigree file contained 1,453,396 animals (cows with production records and ancestors). The model used in the Walloon region of Belgium was presented by Auvray and Gengler (2002) and Croquet *et al.* (2006).

Interbull files from the January 2008 routine-run were used that contained a total of 100,156 sires. Interbull files were matched with local evaluation files and sires identified in both evaluations were extracted. Interbull evaluations were discarded if not considered official. The following rules were used to identify prior evaluations:

1. If there was no local evaluation send and used by Interbull, the Interbull evaluation was free of local information, therefore this evaluation was directly considered the external evaluation.
2. If a local evaluation contributed to the Interbull evaluation, the external evaluation was computed as described before

In order to speed up convergence, converged solutions from the pure local evaluations were used as starting values. Additionally for animals with external evaluations starting values were computed a weighted average between local and external evaluations. This was preferred instead of directly putting the Interbull evaluation as those evaluations were not taking care of differences across the lactations.

3. Results and Discussion

3.1. External evaluations

A total of 4041 bulls with Interbull evaluations were also present in our local evaluation. Of these bulls 638 had local evaluations that were also provided to Interbull. The remaining animal had only Interbull evaluations, but had also independent local evaluations. Table 1 shows the average coefficients used to compute external EBV from Interbull EBV and from local EBV. The coefficients depend on heritabilities, and reliabilities. As those are potentially different across the three traits slightly different weightings were used.

Table 1. Average coefficients used to compute external EBV from Interbull EBV and from local EBV.

Trait	Interbull	Local
Milk	2.37	-1.63
Fat	2.37	-1.56
Protein	2.49	-1.69

In Table 2 gives correlations for all three traits of external EBV with Interbull EBV and local EBV.

Table 2. Correlations for all three traits of external EBV with Interbull EBV and local EBV.

Trait	Interbull EBV	Local EBV
Milk	0.85	0.78
Fat	0.94	0.87
Protein	0.95	0.90

Table 3 summarizes average reliabilities of local EBV, Interbull EBV and external EBV.

Table 3. Average reliabilities of local, Interbull and external EBV

Trait	Interbull	Local	External
Milk	0.93	0.89	0.73
Fat	0.93	0.89	0.82
Protein	0.93	0.89	0.81

3.2. Setting up and solving of modified mixed model equation

As shown before modifications of regular mixed model equations were small and could be done in a few lines of code. Solving was done using the same regular precondition conjugate gradient solver as for current genetic evaluations. Convergence of the modified system of equations was sufficient to obtain after 300 rounds a relative squared error, estimated as the difference between observed and computed right hand side elements, of 0.20×10^{-9} while the largest relative change of a genetic effect from one round to the other was below 0.8%.

3.3. Modified mixed model equation solutions

Bayesian integration of external EBV affects the whole system of equations from fixed effects to random effects. By keeping the joint estimation, estimates of contemporary effects will take into account differences of genetic merit of each member of such a group. Also all related animals will be affected by external EBV. Despite this, overall correlations for all effects were over 0.99 when comparing solutions with and without integrated external EBV. More important is in this context if the EBV for sires with external EBV moved closer to the expected values. In our case we expect that Interbull EBV should be the expected values. Table 4 shows the correlation of EBV of the 4041 of strictly local and integrated external EBV with Interbull EBV. The improvement of correlations was very impressive.

Table 4. Correlations of strictly local and with integrated external EBV with Interbull EBV.

Trait	Local	Integrated
Milk	0.77	0.96
Fat	0.69	0.96
Protein	0.78	0.96

Despite these very promising results, we noticed also that solutions for animals after integrated external EBV had different means and variances than original EBV. The observed differences were higher than what would be expected.

4. Conclusions and Implications

Bayesian inclusion of external evaluations into a national evaluation system and its application to milk production traits proved to be feasible. Even if overall solutions had very high correlations to solutions without integrated EBV, EBV for animals with external EBV changed as expected and became more similar to Interbull EBV. However the presented method needs some important fine tuning as several non-trivial assumptions are made and results are potentially very sensitive to these assumptions. The most important topics of future research are to investigate improved methods to get external EBV free of local information, to translate external EBV of Interbull traits to EBV of locally evaluated traits and other issues that lead to changes in means and variances. Legarra *et al.* (2007) showed a strategy to address mean differences for external EBV, however also changes in implementation or even in the theory to address variance differences can be imagined.

Acknowledgements

Nicolas Gengler, who is Research Associate of the National Fund for Scientific Research (Brussels, Belgium), acknowledges his support. The authors gratefully acknowledge the financial support of the Ministry Agriculture of the Walloon Region of Belgium (MRW-DGA) (Namur, Belgium). Additional support was provided through grants 2.4507.02F (2) and F.4552.05 of the National Fund for Scientific Research.

References

- Auvray, B. & Gengler, N. 2002. Feasibility of a Walloon test-day model and study of its potential as tool for selection and management. *Interbull Bulletin* 29, 123-127.
- Croquet, C., Mayeres, P., Gillon, A., Vanderick, S. & Gengler, N. 2006. Inbreeding depression for global and partial economic indexes, production, type, and functional traits. *J. Dairy Sci.* 89, 2257-2267.
- Legarra, A., Bertrand, J.K., Strabel, T., Sapp, R.L., Sanchez, J.P. & Misztal, I. 2007. Multi-breed genetic evaluation in a Gelbvieh

- population. *J. Anim. Breed. Genet.* 124, 286-295.
- Klei, L., Quaas, R.L., Pollak, E.J. & Cunningham, B.E. 1996. Multiple breed evaluation. In: *Proceedings of Research Symposium and Annual Meeting*, Beef Improvement Federation. Birmingham, AL, USA. Pp. 106-115.
- Gengler, N. & Verkenne, C. 2007. Bayesian approach to integrate molecular data into genetic evaluations. *Interbull Bulletin* 37, 37-41.
- VanRaden, P.M. 2001. Methods to combine estimated breeding values obtained from separate sources. *J. Dairy Sci.* 84:E47-55E.