Bayesian MACE for Ayrshire Conformation Traits

Thomas Mark[1] and Per Madsen[2]
[1]*Interbull Centre, Uppsala, Sweden*
[2]*Danish Institute of Agricultural Sciences, Foulum, Denmark.*

Introduction

Routine international genetic evaluations for conformation traits are performed for Holstein and Jersey. Results from a pilot-run for Guernsey (Klei & Lawlor, 2002) and a test-run for Brown Swiss were promising, and these breeds will probably join Interbull conformation evaluations in the near future. There is also a strong wish from various national Ayrshire societies to have across country genetic evaluations for Ayrshire conformation traits. However, such evaluations are hampered by very poor genetic connectedness among Ayrshire populations. Poor connectedness makes estimation of genetic correlations and contrasts between country effects difficult. In a recent study, Klei and Lawlor (2001) estimated genetic correlations among conformation traits in nine different Ayrshire populations. Results from this study were in some cases unexplainable, except because of sampling due to weak connectedness.

Bayesian methodology incorporates prior information in the inference of dispersion parameters. With such methodology it is possible to use parameters obtained from another breed (*i.e.* Holstein) as prior information in the estimation of Ayrshire parameters. This is a viable approach when reliable estimates are available from another breed, little information is available in Ayrshire data, and traits are expected to be similar for the two breeds. Recently, the classification of Ayrshire conformation traits has been harmonised with Holstein, and traits are expected to be quite similar between Ayrshire and Holstein within countries. The uncertainty of all parameters is taken into account in Bayesian inferences. Thus if breeding values are predicted the uncertainty of genetic parameters and contemporary group effects (*e.g.* country effect) are accounted for and affect predictions and their reliabilities. This property seems desirable for especially poorly connected data.

A Gibbs Sampler applied to Multiple-trait Across Country Evaluations (**MACE**) was developed by Jensen and Madsen (2002). The aim of this study was to apply this method to simultaneously estimate genetic parameters, country effects and predict international breeding values for Ayrshire conformation traits and to study the impact of prior information on posterior parameters.

Material and Methods

Data

National genetic evaluation results for 18 Ayrshire conformation traits from nine countries were available. These countries were Australia (**AUS**), Canada (**CAN**), Denmark (**DNK**), Finland (**FIN**), the Unted Kingdom (**GBR**), Norway (**NOR**), New Zealand (**NZL**), Sweden (**SWE**) and the United States (**USA**). For AUS, GBR, NOR and NZL the same data as described by Klei and Lawlor (2001) were used whereas more recent evaluation results were submitted from the other countries. For simplicity only results for Fore Udder are shown here. Fore udder was chosen because all participating countries had data available for this trait unlike most other traits.

The amount of genetic ties and total number of bulls with data included in the analysis are shown in Table 1. The number of genetic ties has increased slightly for some country combinations, since data were collected by Klei and Lawlor (2001), but mostly for those country combinations that already had most genetic ties. The use of foreign bulls was still very unevenly distributed across countries for the Ayrshire data considered in this study (Table 1).

Table 1. Number of common bulls (below diagonal), common ¾ sib families (above diagonal) and total number of bulls with data (on diagonal) included in parameter estimation and prediction of breeding values.

	AUS	CAN	DNK	FIN	GBR	NOR	NZL	SWE	USA
AUS	148	8	0	1	2	0	3	1	6
CAN	8	478	2	6	21	0	17	5	70
DNK	0	2	864	4	0	1	2	17	0
FIN	1	6	4	1453	1	3	5	29	2
GBR	2	26	0	1	189	0	11	1	13
NOR	0	0	1	2	0	1732	0	10	0
NZL	3	15	2	5	12	0	288	6	11
SWE	1	5	14	23	1	7	6	586	1
USA	6	61	0	2	13	0	10	1	175

Methods and prior information

National evaluation results were deregressed within country (Jairath *et al.*, 2001). The deregressed national evaluation results were used as dependent variable in Bayesian MACE. The same heritabilities as was used in national evaluations were used in the deregression. Heritabilities ranged between 0.19 (CAN) and 0.30 (SWE), except for NOR where the heritability was 0.061.

A Gibbs Sampler applied to MACE (Jensen & Madsen, 2002) was used to simultaneously estimate (co)variance components and predict international breeding values. Prior (co)variance matrices were assumed to follow an inverse Wishart distribution with equal prior belief (*i.e.* degrees of freedom; **df**) for all elements. Estimated genetic correlations for Holstein fore udder (Interbull, 2002) measured in the same countries were used as prior information when available. Norway was the only country where Holstein estimates were not available. The prior correlations between Norway and other countries were set to the lowest estimated Holstein correlation for the other country. This resulted in the prior correlations shown in Table 2. For simplicity EM-REML estimates (from a well-connected subset of the same data) for sire and residual variances were used as prior for these components.

The same degree of belief (weight) was given to the prior for each (co)variance component in each analyses, but weights of 5, 10 and 100 df were used in different analyses to study the effect of weight on prior information on posterior estimates. One df corresponds to one extra animal with its true breeding value known.

Table 2. Prior genetic correlations (above diagonal) and sire standard deviations (on diagonal).

	AUS	CAN	DNK	FIN	GBR	NOR	NZL	SWE	USA
AUS	.231	.80	.79	.64	.85	.59	.80	.59	.77
CAN		2.129	.95	.70	.94	.70	.82	.73	.94
DNK			.751	.78	.92	.78	.89	.81	.92
FIN				.292	.67	.63	.78	.92	.63
GBR					.303	.67	.88	.70	.92
NOR						.849	.72	.59	.63
NZL							.184	.72	.82
SWE								.293	.68
USA									1.633

(Co)variances were also estimated with EM-REML applied to a reduced set of MACE equations (Klei & Weigel, 1998) and country effects and breeding values were obtained with traditional MACE (Schaeffer, 1994) for comparison.

Genetic groups were created based on selection path, birth year of the bull and country of first registration. A relaxed grouping strategy was used to avoid initial problems with many groups having very little information, *i.e.* only 12 groups consisting of 398 to 1042 phantom parents were formed.

210,000 rounds of iteration were used as burn-in and were followed by 100,000 rounds with an interleave of 10 rounds, *i.e.* a total of 10,000 samples were available for inferences for each of the three analysis.

Results and Discussion

The 10,000 samples used for inferences corresponded to 48-199, 66-275 and 502-2254 effective samples for sire (co)variances for 5, 10 and 100 df, respectively. Effective sample sizes were higher for residual variances and ranged between 114-465, 127-298 and 567-2881 for 5, 10 and 100 df, respectively. Longer chains could be run to increase effective sample sizes. The burin period was unusually high, which could be due to the many parameters which needed to be estimated based on little information. There were very high auto-correlations between samples (lag10 correlations >0.9), which resulted in poor mixing and made it necessary to use many rounds as burnin.

Estimated genetic correlations for different methods and weights on prior information are shown in Table 3 for three different country combinations representing poor, medium and well-connected populations, respectively.

Gibbs sampling estimates were expected to be similar to REML estimates when a flat prior was used, but were expected to move towards the prior correlation as the prior weight increased and at a higher rate for poorly connected estimates. Gibbs sampling estimates were different from EM-REML estimates. Differences between Gibbs sampling estimates (df=5) and EM-REML estimates ranged from -0.04 to 1.25 for all countries (average difference was 0.44). But as the prior weight (df) increased the estimates did tend to move towards the prior correlations, although relatively little except for estimates based on little information in data. The largest change in estimated genetic correlation when increasing the prior weight from 5 to 100 df was 0.32 for the correlation between NOR and SWE. The largest change in the opposite direction of the prior correlation was for the correlation between FIN and SWE (-0.02) and the average change for all correlations was 0.12.

Posterior standard deviations of Gibbs sampling estimates were expected to decrease as the prior weight increased. On average this was also the case $(i.e. 100\%(SE_{DF=5} - SE_{DF=100}) / SE_{DF=5}$ was on average 10.2 percent for genetic correlations), but single posterior standard deviations on estimated genetic correlations were higher for 100 df compared with 5 df. This was because the estimated correlations in some cases changed substantially, and posterior standard deviations on (co)variance components more consistently moved in the expected direction, *i.e.* the relative decrease in posterior standard deviations on sire (co)variances ranged from 0.0 to 40.4 percent when df increased from 5 to 100. Similarly the relative decrease in residual (co)variances ranged from 6.5 to 54.2 percent. Posterior standard deviations tended to be largest for parameters for poorly connected countries and countries with low heritability.

Table 3. Genetic correlations estimated with EM-REML and mean posterior genetic correlation estimates (± posterior standard deviation) for Gibbs sampler (Gibbs) for different weights on prior information (df) and three selected country combinations.

Method	df	USA-NOR	USA-NZL	USA-CAN
REML	-	.948	.772	.908
Gibbs	5	.913 ± .057	.906 ± .049	.996 ± .003
Gibbs	10	.733 ± .131	.906 ± .045	.989 ± .005
Gibbs	100	.663 ± .056	.917 ± .014	.971 ± .005
Prior	-	.63	.82	.94

Estimated genetic parameters for 10 df are shown in Table 4. Estimated genetic correlations ranged between 0.66 (FIN-NOR) and 0.99 (*e.g.* CAN-USA). Estimated genetic correlation were in most cases higher compared with EM-REML estimates from the same data (results not shown) and showed less variability across countries. Posterior standard deviations of estimated genetic correlations ranged between 0.005 (CAN-USA) and 0.15 (FIN-NOR). Residual correlations were also estimated and were close to zero (0.00 to 0.02) among countries as expected.

Table 4. Estimated (± posterior standard deviation) genetic correlations (above diagonal), sire variances (on diagonal) and residual correlations (below diagonal) among fore udder in 9 Ayrshire countries (df=10).

	AUS	CAN	DNK	FIN	GBR	NOR	NZL	SWE	USA
AUS	.054 ± .009	.90 ± .05	.91 ± .05	.88 ± .05	.91 ± .05	.73 ± .13	.95 ± .03	.90 ± .05	.88 ± .06
CAN	.00 ± .03	5.99 ± .62	.99 ± .01	.74 ± .10	.96 ± .02	.75 ± .13	.93 ± .04	.89 ± .05	.99 ± .00
DNK	.00 ± .03	.00 ± .02	.664 ± .068	.79 ± .09	.98 ± .01	.81 ± .10	.96 ± .02	.91 ± .04	.97 ± .01
FIN	.00 ± .02	.00 ± .02	.00 ± .02	.051 ± .007	.80 ± .09	.66 ± .15	.85 ± .06	.91 ± .04	.68 ± .12
GBR	.02 ± .05	.00 ± .04	.00 ± .03	.00 ± .02	.117 ± .016	.80 ± .10	.94 ± .03	.93 ± .03	.94 ± .03
NOR	.00 ± .02	.00 ± .02	.00 ± .02	.00 ± .02	.00 ± .02	.278 ± .079	.83 ± .09	.69 ± .13	.73 ± .13
NZL	.00 ± .04	.00 ± .03	.00 ± .03	.00 ± .02	.00 ± .04	.00 ± .02	.025 ± .004	.90 ± .04	.91 ± .04
SWE	.00 ± .03	.00 ± .03	.00 ± .02	.01 ± .02	.00 ± .03	.00 ± .02	.00 ± .03	.071 ± .011	.86 ± .06
USA	.01 ± .05	.01 ± .04	.00 ± .03	.00 ± .02	.00 ± .05	.00 ± .02	.00 ± .04	.00 ± .03	4.29 ± .51

Estimated country effects changed substantially for some countries when Gibbs Sampling were used to simultaneously estimate parameters and predict breeding values compared with traditional MACE (Table 5). The prior weights and correlations used had a smaller effect on estimated country effects. Posterior standard deviations on country effects tended to decrease with increased prior weights as expected. Further studies should be done to determine the impact of changes in estimated parameters on breeding values. It is likely that breeding values will be affected, but changes in some parameters may also to some extend be counteracted, *e.g.* in genetic group effects.

Table 5. Estimated country effects (± standard error or posterior standard deviation) for different methods and prior weights (df).

Method	AUS	CAN	DNK	FIN	GBR	NOR	NZL	SWE	USA
I	-0.23±0.14	-2.71±0.95	0.75±0.53	0.40±0.22	-0.26±0.17	-2.15±1.54	0.002±.076	0.34±0.32	-1.97±0.70
II	-0.25±0.20	-2.25±1.05	0.84±0.81	0.48±0.29	-0.51±0.18	-2.13±1.83	0.029±.090	0.36±0.37	-1.77±0.92
III	-0.40±0.17	-0.92±0.56	-0.49±0.35	0.29±0.19	-0.29±0.13	-2.03±1.19	0.068±.056	-0.03±0.21	-0.58±0.59
VI	-0.38±0.16	-0.88±0.57	-0.54±0.34	0.29±0.19	-0.31±0.13	-2.25±0.99	0.070±.057	-0.05±0.21	-0.53±0.62
V	-0.37±0.14	-0.97±0.54	-0.55±0.35	0.28±0.18	-0.33±0.13	-2.66±1.35	0.064±.060	-0.03±0.22	-0.58±0.54

I) Normal MACE and EM-REML; II) Normal MACE, but correlations estimated with Gibbs (df = 10); III) Gibbs (df = 5); VI) Gibbs (df = 10); V) Gibbs (df = 100).

General discussion: Implications and future research

For the current Interbull practice, data on older bulls are excluded in breeding value prediction and estimation of sire variances to avoid problems with heterogeneous variances over time. A well connected subset of all historical data is, however, used for estimation of genetic correlations. The same data set is required for correlation estimation and prediction of breeding values when these two steps are carried out simultaneously. In this study data on all bulls regardless of birth year and connectedness were included in the analysis. Bayesian Mace can account for heterogeneous variances over time. Specific variance components can be assigned to groups of bulls, *e.g.* based on birth year. Including all data and accounting for heterogeneity in the model have theoretical advantages over the current ad hoc procedure of employing different data edits for different tasks. This was, however, not done in this study, since the main aim was to study the impact of prior information. Further research regarding heterogeneous variance and which data to include is needed before any implementation can be envisioned.

The beliefs on each prior correlation may differ, but this was not taken into account in this study. Each prior (co)variance had the same weight. However, it seems more pertinent to use available resources on other research questions given the relative small influence of different prior weights on most parameters except correlations between poorly connected countries.

Computing time for Bayesian MACE was higher compared with the current EM-REML procedure (Klei & Weigel, 1998). 500,000 rounds of iteration took approximately 7 CPU days per trait (on an IBM RS6000 Power3-II Thin Node, 375 MHz computer) for the current implementation and for the nine countries considered in this study. Therefore, an implementation of Bayesian MACE for Holstein (at least production and conformation traits) is currently not realistic. However, each trait can be run on separate computers and no efforts have been made to optimize the program for computing speed. Also, Bayesian MACE will be most advantageous for breeds other than Holstein. This is because Holstein data usually contain more information compared with other breeds. Thus the benefits in terms of being able to incorporate prior information are limited for Holstein.

If a simultaneous estimation of genetic parameters and prediction of breeding values are implemented, it will affect the Interbull service structure. The turn over time will increase substantially for routine runs, but the need for test-runs will diminish. The only purpose of a test-run would be to prepare members for future changes and this would even be made more difficult when correlations are re-estimated during routine runs. The use of Bayesian techniques to estimate genetic parameters for genetic evaluation is becoming common practise, but a full scale Bayesian genetic evaluation that simultaneously estimate genetic parameters and predict breeding values has not yet been implemented anywhere to the knowledge of the authors. Thus little experience have been accumulated to support the implementation of Bayesian MACE. The results on correlations and country effects obtained in this study were substantially different compared with results from traditional MACE and the dairy industry will have to get used to and accept this before Bayesian MACE can be implemented.

Further research could address heterogeneous variances, computing strategies, the impact of estimating residual (co)variances instead of keeping them fixed and strategies for handling genetic groups in Bayesian MACE.

Conclusion

Bayesian MACE was applied for Ayrshire conformation (fore udder). Results were generally within the range of expectations. The influence of different weights on prior information was largest for estimated correlations between poorly connected countries and small for country effects. Estimated genetic correlations were high (0.66 to 0.99) and the method seemed appropriate for data characterized by poor connectedness. However, further research on *e.g.* heterogeneous variance and which data to include is needed before we are ready for a pilot run. Implementation of Bayesian MACE in full scale will reduce the need and usefulness of test-runs and will increase the turn-over time for routine runs.

Acknowledgements

We are grateful to the countries that provided data and to Bert Klei for making older data available. We also thank Just Jensen, Daniel Sorensen and Freddy Fikse for valuable discussions and technical assistance.

References

Interbull, 2002. Results from Interbull February 2002 test-run (not published).

Jairath, L., Dekkers, J.M.C., Schaeffer, L.R., Liu, Z., Burnside, E.B. & Kolstad, B. 1998. Genetic evaluation for herd life in Canada. *J. Dairy Sci. 81,* 550-562.

Jensen, J. & Madsen, P. 2002. Bayesian estimation of across country genetic parameters for MACE. *Paper presented at Interbull open meeting,* Interlaken, May 2002.

Klei, B. & Lawlor, T. 2001. Mace for conformation – Ayrshires. *Interbull Bulletin 27,* 51-55.

Klei, B. & Lawlor, T. 2002. A pilot study on MACE for conformation for Guernseys. *Paper presented at Interbull open meeting,* Interlaken, May 2002.

Klei, B. & Weigel, K.A. 1998. A method to estimate correlations among traits in different countries using data on all bulls. *Interbull Bulletin 17,* 8-14.

Schaeffer, L.R. 1994. Multiple-country comparison of dairy sires. *J. Dairy Sci. 77,* 2671-2678.