Data transformation for rank reduction in Multi-Trait MACE Model for International Bull Comparison.

J. Tarrès, Z. Liu, Vincent Ducrocq, F. Reinhardt, R. Reents

HAL Id: hal-02655105

https://hal.inrae.fr/hal-02655105v1

Submitted on 29 May 2020

Original article

Data transformation for rank reduction in multi-trait MACE model for international bull comparison

Joaquim Tarres[1,2*], Zengting Liu[1], Vincent Ducrocq[2], Friedrich Reinhardt[1], Reinhard Reents[1]

[1] VIT, Heideweg 1, 29283 Verden, Germany
[2] UR337, Station de génétique quantitative et appliquée, INRA, 78352 Jouy-en-Josas Cedex, France

Abstract – Since many countries use multiple lactation random regression test day models in national evaluations for milk production traits, a random regression multiple across-country evaluation (MACE) model permitting a variable number of correlated traits per country should be used in international dairy evaluations. In order to reduce the number of within country traits for international comparison, three different MACE models were implemented based on German daughter yield deviation data and compared to the random regression MACE. The multiple lactation MACE model analysed daughter yield deviations on a lactation basis reducing the rank from nine random regression coefficients to three lactations. The lactation breeding values were very accurate for old bulls, but not for the youngest bulls with daughters with short lactations. The other two models applied principal component analysis as the dimension reduction technique: one based on eigenvalues of a genetic correlation matrix and the other on eigenvalues of a combined lactation matrix. The first one showed that German data can be transformed from nine traits to five eigenfunctions without losing much accuracy in any of the estimated random regression coefficients. The second one allowed performing rank reductions to three eigenfunctions without having the problem of young bulls with daughters with short lactations.

rank reduction / principal components / genetic correlation matrix / multiple across country evaluation / dairy cattle

1. INTRODUCTION

The multiple across country evaluation (MACE) [17] methodology is currently used for international dairy bull comparisons. Estimated breeding values from each country are deregressed to obtain a value analogous to daughter yield deviations (DYD) for bulls that have daughters with records. Despite the

---

fact that only a single EBV per bull is permitted for each country in inter-
national genetic evaluation, the current MACE has a large number of equa-
tions, since each evaluated sire will, conceptually, have a breeding value for all
traits, *i.e.* for all countries, although it might have daughters only in one. Tradi-
tionally, the corresponding genetic covariance matrices have been considered
unstructured, *i.e.* for k countries there were k(k + 1)/2 distinct (co)variance
components to be estimated. For example, when the current Interbull Holstein
evaluation run for production traits includes 26 countries, the (co)variance ma-
trix of genetic effects involves 325 correlations. By and large, restrictions on
estimates are imposed only to ensure that estimates were within the parameter
space, *i.e.* that all variances and conditional variances are positive, that all cor-
relation estimates are in the range of −1 to +1, and that all partial correlations
are consistent with each other [13]. In statistical terms, this is equivalent to
the requirement that the estimated covariance matrix is positive semidefinite,
i.e. that none of its eigenvalues is negative.

In contrast, other areas of statistics have long since assumed and estimated
structured covariance matrices. A well-designed structural model for genetic
covariances uses information, external to the data, to explain genetic covari-
ability in terms of few parameters, leading to more precise estimates of ge-
netic correlations between countries [16]. For example, in international dairy
sire evaluations, traits are currently defined according to country borders. How-
ever, similarity in production systems between herds in different countries de-
pends not only on geographical proximity but also on climatic conditions, on
management practices, and on the genetic composition of the cow popula-
tion. If information is available about these variables, it can be used to ex-
plain the genetic covariance structure between countries. Rekaya *et al.* [16],
Minéry *et al.* [14] and Leclerc *et al.* [6] have proposed structural models in
order to reduce the number of parameters to be estimated across countries.

On the contrary, principal component (PC) analysis is widely used as a
dimension reduction technique, but so far has had only limited applications
in quantitative genetic analyses. PC analysis requires estimating eigenvectors
and -values of covariance matrices [3]. Eigenvectors define independent lin-
ear functions of the variables considered, the so-called principal components,
that successively explain the maximum amount of variation, measured by the
corresponding eigenvalues. This implies that for a given number of compo-
nents considered, PC approximate the multivariate data most accurately. It fol-
lows that PC with variances (eigenvalues) close to zero contribute virtually no
information to the analysis that is not already contained in the PC with larger
eigenvalues. Hence, these components can be ignored resulting in equivalent

analyses involving fewer variables, *i.e.* traits or countries, and often reduced sampling variation. Reducing the dimension of analyses by considering fewer variables can considerably decrease computational requirements. Different authors have proposed the reduction of the number of parameters to be estimated across countries using PC analysis, and the closely related factor analysis [5, 10–13].

Since more and more countries have upgraded their national genetic evaluation system to a multiple trait model or a multiple lactation random regression test day model (RRTDM) [7], differences between models for national and international evaluations have become increasingly evident. In order to optimise genetic evaluation models for both national and international evaluations, Sullivan and Wilton [18] and Liu *et al.* [9] proposed a multiple trait MACE (MT-MACE) model for international bull comparison. This model extended the current single trait MACE (ST-MACE) allowing a variable number of correlated traits for countries using a multiple trait model in national genetic evaluation. Although the MT-MACE model can better utilise the information derived from the RRTDM in national genetic evaluation than the ST-MACE model, the huge size of the MT-MACE system can be a limiting factor for international genetic evaluation involving all dairy populations. In order to reduce the size of the MT-MACE system, it is even more necessary than for the ST-MACE to apply rank reduction techniques to find intermediate MACE models that can be a reasonable compromise between feasibility and accuracy.

In parallel to international bull comparison, France and Germany strengthened their collaboration at the end of 2005 in a joint project. One of the main goals of this project was to perform joint French-German bull and cow evaluations using pre-corrected records (*i.e.* yield deviations) in a two-step approach [1] following the multiple trait MACE model proposed by Liu *et al.* [9]. Since Germany uses a multiple lactation random regression test day model in national evaluation, a random regression MACE model (RR-MACE) was performed and was feasible for the parameter estimation and the joint bull evaluation of milk production traits from France and Germany [19, 20]. In the near future, the RR-MACE model will be applied for the joint bull and cow evaluation for milk production traits from France and Germany. In this case, the number of traits per country can be a limiting factor due to the much larger equation system (*i.e.* millions of equations).

The aim of this paper was to explore different methods to reduce the rank of the German RRTDM in order to make MT-MACE applicable for joint French-German evaluation and/or international genetic evaluation involving a higher number of countries. After the analysis of a multiple lactation MACE

model (ML-MACE), two rank reduction alternatives based on principal component analysis were performed to investigate their accuracy and suitability.

2. MATERIALS AND METHODS

2.1. Data

Once February 2006 German national genetic evaluations were run, DYD of bulls were obtained as the average of daughter performance adjusted for fixed effects and non-genetic random effects of daughters and genetic effects of bull's mates [8]. There were 14 887 Black and White Holstein bulls with DYD available from Germany.

Full pedigree information of bulls with sire and dam relationship was used for MACE evaluation. There were 67 541 animals in the pedigree file and 32 genetic groups for unknown parents. Genetic groups were defined according to the breed, country of origin, selection path (son to sire, son to dam, daughter to sire and daughter to dam) and birth year of the animal. Small phantom groups were merged automatically given a predefined minimum number of animals per group. The following rules were applied to combine or merge small groups: selection paths were merged based on the sex of the parent (son to sire with daughter to sire, and son to dam with daughter to dam), countries were merged accordingly (North America, Western Europe, and the rest), minor breeds within Holstein breed were merged.

2.2. The MT-MACE model

For a country using a multi-trait model in national genetic evaluation, the following statistical model was applied to DYD of a bull i:

$$\mathbf{q}_i \;=\; \mathbf{f} \;+\; \mathbf{u}_i \;+\; \boldsymbol{\varepsilon}_i \tag{1}$$

where $\mathbf{q}_i$ is a vector of DYD of the i-th bull, $\mathbf{f}$ is a vector of general means for all traits, $\mathbf{u}_i$ is the vector of additive genetic effects of bull i, and $\boldsymbol{\varepsilon}_i$ is a vector of residual effects. (Co)variance matrices for the random effects are the following:

$$\text{Var}\,(\mathbf{u}_i) = \mathbf{G}, \quad \text{and} \quad [\text{Var}\,(\boldsymbol{\varepsilon}_i)]^{-1} = \boldsymbol{\Psi}_i$$

where $\mathbf{G}$ is genetic (co)variance matrix, and $\boldsymbol{\Psi}_i$ is a multitrait equivalent daughter contribution (MTEDC) matrix associated with the DYD vector $\mathbf{q}_i$.

Four sub-models of the MT-MACE model that differed with respect to the data analysed were considered in this investigation. The subindex i will be omitted from the formulae to simplify nomenclature.

2.2.1. The random regression MACE model

The random regression MACE model (RR-MACE) analysed $\mathbf{q}_i$ on a random regression coefficient (RRC) basis. Ignoring pedigree information, the mixed-model equation (MME) for estimating the breeding value $\hat{\mathbf{u}}_{RRC}$ for bull i is:

$$\left(\boldsymbol{\Psi}_{RRC} + a^{ii}\mathbf{G}_{RRC}^{-1}\right) \hat{\mathbf{u}}_{RRC} = \boldsymbol{\Delta}_{RRC}$$

where a^{ii} is the diagonal element of the inverse of the relationship matrix for bull i. The matrix $\boldsymbol{\Psi}_{RRC}$ and the vector $\boldsymbol{\Delta}_{RRC}$ were generated in national evaluation using the MTEDC procedure [8]. The equation system was solved using a pre-conditioned conjugate gradient algorithm (PCG) with an iteration-on-data technique [9]. The convergence criterion, defined as the logarithm of the sum of squares of differences in solutions between two consecutive rounds of iteration divided by the sum of squares of solutions in the last round, was set to -10.

The RRTDM in Germany modelled the additive genetic effects of an animal as a normalised orthogonal third-order Legendre polynomial per each of the first three lactations [7]. Thus, in order to model the additive genetic effects more closely to national evaluation, the RR-MACE evaluation estimated nine breeding values for each animal in the pedigree file. These breeding values on a RRC basis can be easily summarised to a combined lactation basis [7].

2.2.2. The multiple lactation MACE

In order to reduce the number of traits per bull, the ML-MACE model was designed for evaluating DYD on a 305-day lactation basis. Now, each bull with data in Germany had one DYD for each of the three lactations. In order to conduct the evaluation with the ML-MACE model, the following conversions need to be performed to equation system $\left(\boldsymbol{\Psi}_L + a^{ii}\mathbf{G}_L^{-1}\right) \hat{\mathbf{u}}_L = \boldsymbol{\Delta}_L$ prior to evaluation:

$$\boldsymbol{\Psi}_L = \left(\mathbf{L}\left(\boldsymbol{\Psi}_{RRC}\right)^{-1}\mathbf{L}'\right)^{-1} \tag{2}$$

$$\boldsymbol{\Delta}_L = \boldsymbol{\Psi}_L\left(\mathbf{L}\left(\boldsymbol{\Psi}_{RRC}\right)^{-1}\boldsymbol{\Delta}_{RRC}\right) \tag{3}$$

$$\mathbf{G}_L = \mathbf{L}\,\mathbf{G}_{RRC}\,\mathbf{L}' \tag{4}$$

where matrix $\mathbf{L}$ converts the information from RRC to a 305-day lactation basis. Note that the EDC matrix $\boldsymbol{\Psi}_{RRC}$ and the matrix product $\mathbf{L}(\boldsymbol{\Psi}_{RRC})^{-1}\mathbf{L}'$ do not have a regular inverse when bull i has daughters with some missing lactations. In this case, only a full rank submatrix corresponding to lactations with data need to be inverted.

Table I. Genetic variances (on the diagonal) and correlations (above the diagonal) used in the German national evaluation with a random regression test day model.

RRC of 1st lactation			RRC of 2nd lactation			RRC of 3rd lactation		
First	Second	Third	First	Second	Third	First	Second	Third
3.715	0.147	–0.479	0.835	–0.026	–0.409	0.838	–0.037	–0.400
	0.397	–0.263	0.216	0.697	–0.295	0.193	0.642	–0.271
		0.184	–0.444	0.184	0.909	–0.450	0.098	0.909
			3.362	0.066	–0.435	0.970	0.082	–0.429
				0.801	0.156	0.043	0.961	0.163
					0.252	–0.435	0.101	0.977
						3.697	0.075	–0.427
							0.849	0.097
								0.293

Table II. Eigenvalues of genetic correlation matrix (GC) and combined lactation weight matrix (CW).

rG	CW
4.218	9.914
2.574	0.755
1.519	0.106
0.304	0.0000306
0.214	0.0000012
0.102	0.0000003
0.032	1×10^{-10}
0.024	1×10^{-14}
0.015	1×10^{-12}

2.2.3. *The rank reduced random regression MACE model based on genetic correlations*

Because the genetic (co)variance matrix from Germany (Tab. I) has some relatively low eigenvalues, principal component analysis could be applied in order to reduce the number of equations per bull from nine traits to a lower number of eigenfunctions. The German genetic (co)variance matrix $\mathbf{G} = \mathbf{S}_{rG}\mathbf{U}_{rG}\mathbf{D}_{rG}\mathbf{U}'_{rG}\mathbf{S}_{rG}$ is decomposed as a product of the eigenvectors $\mathbf{U}_{rG}$ and the eigenvalues $\mathbf{D}_{rG}$ of the genetic correlation matrix (Tab. II). The matrix $\mathbf{S}_{rG}$ is the diagonal matrix of genetic standard deviations. The data transformation matrix is defined as $\mathbf{T} = \mathbf{S}_{rG}\mathbf{U}_{rG}\mathbf{D}_{rG}^{\frac{1}{2}}$. For rank reduction, the smallest eigenvalues in $\mathbf{D}_{rG}$ can be set to zero and deleted, and the corresponding columns in $\mathbf{U}_{rG}$ removed.

In order to conduct the evaluation with the rank reduced random regression MACE model based on genetic correlations (rG-MACE), the following conversions need to be performed to the equation system $\left(\mathbf{\Psi}_{rG} + a^{ii}\mathbf{G}_{rG}^{-1}\right)\hat{\mathbf{u}}_{rG} = \Delta_{rG}$ prior to evaluation:

$$\mathbf{\Psi}_{rG} = \mathbf{T}'\mathbf{\Psi}_{RRC}\mathbf{T} \tag{5}$$

$$\Delta_{rG} = \mathbf{T}'\Delta_{RRC} \tag{6}$$

$$\mathbf{G}_{rG} = \mathbf{I}. \tag{7}$$

Note that since no matrix has to be inverted, there was no problem to calculate $\mathbf{\Psi}_{rG}$ for bulls with missing values in second and/or third lactation (see Appendix).

Once the breeding values $\hat{\mathbf{u}}_{rG}$ were estimated, they were back transformed to the RRC basis as $\hat{\mathbf{u}} = \mathbf{T}\,\hat{\mathbf{u}}_{rG}$ (10). These back transformed EBV are approximate solutions of RRC breeding values.

2.2.4. The rank reduced random regression MACE model based on combined lactation weights

The previous rank reduction based on the genetic correlation matrix gave the same relative importance to all random regression coefficients. However, the German combined lactation EBV depend mainly on the first coefficient of each lactation [7]. In order to give more weight to the coefficients that influence lactation production the most, the genetic (co)variance matrix was decomposed $\mathbf{G} = \mathbf{S}_{CW}^{-1}\mathbf{U}_{CW}\mathbf{D}_{CW}\mathbf{U}_{CW}'\mathbf{S}_{CW}^{-1}$ as a product of a matrix $\mathbf{S}_{CW}^{-1}$ and the eigenvectors $\mathbf{U}_{CW}$ and the eigenvalues $\mathbf{D}_{CW}$ of a matrix $\mathbf{C}_{CW}$. The matrix $\mathbf{C}_{CW} = \mathbf{S}_{CW}\mathbf{G}\mathbf{S}_{CW}$ is the product of the genetic (co)variance matrix and a diagonal matrix $\mathbf{S}_{CW}$ that contains the weight of each coefficient for calculating the combined lactation EBV [7]. In order to conduct the evaluation with the rank reduced random regression MACE model based on combined lactation weights (CW-MACE), the transformation matrix was defined as $\mathbf{T} = \mathbf{S}_{CW}^{-1}\mathbf{U}_{CW}\mathbf{D}_{CW}^{\frac{1}{2}}$ and used in equation (5) and (6) as for rG-MACE.

3. RESULTS

3.1. Multiple lactation MACE model

Pearson correlations between the ML-MACE and the RR-MACE were 0.994, 0.995 and 0.995 for the first, second and third lactation EBV, respectively and 0.995 for the combined lactation EBV (Tab. III). The correlations

Table III. EBV correlations on a combined lactation basis by birth year of Holstein bulls of the multiple lactation MACE (ML-MACE) and the reduced rank random regression MACE models with the random regression MACE model. Rank reductions to 5 and 3 eigenfunctions were performed based on genetic correlations (rG-MACE) or combined lactation weights (CW-MACE).

Birth year	No. of bulls	rG-MACE		CW-MACE	ML-MACE
		Rank 5	Rank 3	Rank 3	Rank 3
1985–1989	3955	1.000	0.988	0.996	1.000
1990–1994	4088	1.000	0.985	0.995	1.000
1995–1999	5194	0.999	0.979	0.992	0.999
2000	867	0.999	0.963	0.992	0.980
2001	214	1.000	0.983	0.993	0.924
§	172	1.000	0.985	0.989	0.840
Overall	14 487	1.000	0.987	0.995	0.995

§ Bulls did not fulfil the criterion, *i.e.* a bull must have at least 10 daughters with lactation passed 120 DIM in the German national evaluation.

by birth year were over 0.999 for old bulls although the values dropped for the very last years (Tab. III). The correlations also dropped for bulls with low average number of test day records of daughters (results not shown). However, the number of daughters had a very limited impact on correlations. In general, very similar combined lactation EBV were obtained using RR-MACE and ML-MACE, but significant differences existed for the youngest bulls with short lactations of daughters. These differences led to a correlation of 0.84 within the top 300 Holstein bulls (Tab. V).

3.2. Rank reduction based on genetic correlations

The rank reduction from nine traits to five eigenfunctions based on the genetic correlation matrix was done discarding the smallest four eigenvalues in Table II. EBV correlations between the rG-MACE of rank 5 and the RR-MACE were over 0.99 for the first and second random regression coefficients (Tab. IV), but slightly lower for the third RRC. On a 305-day lactation basis, EBV correlations were 1.000, 0.997 and 0.991 for the first, second and third lactation. When the three lactations were summarised to a combined lactation basis, the EBV correlations were over 0.999 independently from the birth year (Tab. III), the number of daughters and the average number of test day records of daughters of the bulls. The ranking correlations within the top 300 Holstein bulls were around 0.98 between the rG-MACE of rank 5 and the RR-MACE (Tab. V).

Table IV. EBV correlations on a random regression coefficient (RRC) basis of the reduced rank random regression MACE models with the random regression MACE model. Rank reductions to 5 and 3 eigenfunctions were performed based on genetic correlations (rG-MACE) or combined lactation weights (CW-MACE).

		rG-MACE		CW-MACE
	Lactation	Rank 5	Rank 3	Rank 3
1st RRC	1	1.000	0.992	0.998
	2	0.997	0.963	0.984
	3	0.991	0.952	0.984
2nd RRC	1	1.000	0.948	0.783
	2	0.994	0.946	0.645
	3	0.988	0.899	0.531
3rd RRC	1	0.981	0.975	0.614
	2	0.975	0.960	0.571
	3	0.969	0.950	0.532

Table V. Ranking correlations for Holstein bulls of the multiple lactation MACE (ML-MACE) and the reduced rank random regression MACE models with the random regression MACE model. Rank reductions to 5 and 3 eigenfunctions were performed based on genetic correlations (rG-MACE) or combined lactation weights (CW-MACE).

	rG-MACE		CW-MACE	ML-MACE
	Rank 5	Rank 3	Rank 3	Rank 3
Top 300 bulls	0.980	0.674	0.762	0.847
All 13 718 bulls	1.000	0.987	0.995	0.998

Only official bulls were included in the rankings.

Further rank reduction to three eigenfunctions had lower performance. In this case, almost all EBV correlations with RR-MACE were under 0.99 both on a RRC basis (Tab. IV) and a combined lactation basis (Tab. III).

3.3. Rank reduction based on combined lactation weights

Rank reduction to three eigenfunctions performed better when it was based on combined lactation weights. In this case, only three eigenvalues and their associated eigenvectors explain most of the total genetic variance (Tab. II). After discarding the lowest eigenvalues, it was possible to perform CW-MACE with three eigenfunctions that kept EBV correlations with RR-MACE around 0.99 for the first random regression coefficients (Tab. IV), although the EBV correlations for the second and third coefficients were poor. The EBV

correlations on a combined lactation basis were 0.995 and were quite independent from the birth year (Tab. III), the number of daughters and the average number of test day records of daughters of the bulls. Then, using this data transformation, it is possible to reduce the rank to three eigenfunctions without having the short lactation problem encountered with ML-MACE. In spite of the high EBV correlations, the correlations within the top 300 Holstein bulls was around 0.762 with the RR-MACE (Tab. V).

4. DISCUSSION

Principal component analysis can be applied to countries having a genetic (co)variance matrix close to singular in order to reduce the number of traits. Ignoring the lowest eigenvalues of the genetic correlation matrix, one can perform data transformation to reduce the number of traits to a lower number of eigenfunctions. Generally, the eigenvalues should be obtained from correlation rather than covariance matrices, especially if traits with greatly differing variation are included. In a covariance matrix, functions of traits with high variances have high eigenvalues and are selected first whereas functions of the other traits with less variances may be discarded. Data transformation using eigenvalues of the genetic correlation matrix allowed reducing the German dataset from nine traits to five eigenfunctions without losing accuracy in any of the random regression coefficients. Back transformation of breeding values allowed getting back the same traits that are presented to breeders so that these traits will be more closely inspected and easier to understand for the industry. The impact of rank reduction to five eigenfunctions in the top lists can be considered negligible.

In the near future, the previous rank reduction can be applied in the joint French-German bull and cow evaluation for milk production traits. However, if the number of traits is still a limiting factor, the German dataset can be reduced to three eigenfunctions. In this case, PCA could not be based on genetic correlations because there was a loss of accuracy in all RRC, but it can be applied based on combined lactation weights. This approach concentrated the loss of accuracy on the less important random regression coefficients but keeping the high EBV correlations on a combined lactation basis. This higher rank reduction led to higher differences in the top lists with respect to the RR-MACE and to a lower accuracy for second and third coefficients that would not allow selection for lactation persistency. However, it can be a compromise between feasibility and accuracy.

In the future, the multi-trait MACE model can be applicable for international bull comparison involving a higher number of countries. In such a model, the number of traits would be enormous (Germany 9 RRC, Canada, Netherlands and Italy 15 RRC and so on) and rank reduction would be necessary. Rank reductions could be performed within country and principal components could be used as exchange tools across countries. The problem here is that genetic parameters differ a lot across countries. For example, the genetic correlation of second with third lactation was estimated to be 0.97 in Germany but only 0.83 in Canada [15]. This could lead to much different eigenvectors and eigenvalues across countries. If this is a problem, the within country data can be reduced to three lactations. This ML-MACE model performed very well except for the youngest bulls, which can be traced back to the issue of short lactations. Instead of first, second and third lactations, Guo *et al.* [2] proposed to exchange average yield and a regression on maturity. This uses two traits instead of three to provide nearly the same information and allows lactations higher than three to be easily included, thereby increasing reliability especially for cows. All these options still should be analysed in order to find a compromise between feasibility and accuracy. The within country rank reductions can be combined with across country ones, *e.g.* Leclerc *et al.* [5,6], in order to make MT-MACE applicable for international genetic evaluation involving a higher number of countries.

Another way to reduce the rank of MT-MACE models is to exchange the same traits across countries as in national evaluations and perform the principal component analysis at the Interbull level. Here the estimation of genetic (co)variance matrix for so many traits and countries prior to rank reduction would be difficult [4].

Other applications of data transformation are in total merit index construction within country using a multiple trait animal model [1]. In Germany, production traits (milk, fat and protein) and somatic cell counts are analysed with RRTDM. In this case, data reduction within and across traits could be performed before the joint analysis of all traits.

5. CONCLUSIONS

Different sub-models of the multiple trait MACE model were implemented for the evaluation of German DYD for production traits. If computing requirements of the random regression MACE model are a limiting factor for international genetic evaluations involving different dairy populations, then the data transformation for rank reduction to five eigenfunctions based on eigenvalues

of genetic correlation matrix would be a reasonable compromise between feasibility and accuracy while keeping the possibility to select for lactation persistency. Higher rank reductions lead to more changes on the rankings of top bulls and lower accuracies for coefficients related to lactation persistency, but appear satisfactory for international 305-day combined lactation EBV.

ACKNOWLEDGEMENTS

REFERENCES

[1] Ducrocq V., Boichard D., Barbat A., Larroque H., Implementation of an approximate multi-trait BLUP evaluation to combine production traits and functional traits into a total merit index, in: Proceedings of the 52nd Annual Meeting of the European Association for Animal Production, 26–29 August 2001, Budapest, Hungary, Paper GI.4.

[2] Guo Z., Lund M.S., Madsen P., Korsgaard I., Jensen J., Genetic parameter estimation for milk yield over multiple parities and various lengths of lactation in Danish Jerseys by random regression models, J. Dairy Sci. 85 (2002) 1596–1606.

[3] Joliffe I.T., Principal Component Analysis, Springer Series in Statistics, Springer-Verlag, New York, 1986.

[4] Jorjani H., International genetic evaluation for female fertility traits, Interbull Bull. 34 (2006) 57–60.

[5] Leclerc H., Fikse W.F., Ducrocq V., Principal components and factorial approaches for estimating genetic correlations in international sire evaluation, J. Dairy Sci. 88 (2005) 3306–3315.

[6] Leclerc H., Minéry S., Delaunay I., Druet T., Fikse W.F., Ducrocq V., Estimation of genetic correlations among countries in international dairy sire evaluations with structural models, J. Dairy Sci. 89 (2006) 1792–1803.

[7] Liu Z., Reinhardt F., Bünger A., Dopp L., Reents R., Application of a random regression model to genetic evaluations of test day yields and somatic cell scores in dairy cattle, Interbull Bull. 27 (2001) 159–166.

[8] Liu Z., Reinhardt F., Bünger A., Reents R., Derivation and calculation of approximate reliabilities and daughter yield deviations of a random regression test-day model for genetic evaluation of dairy cattle, J. Dairy Sci. 87 (2004) 1896–1907.

[9] Liu Z., Reinhardt F., Reents R., A multi-trait MACE model for international bull comparison based on daughter yield deviations, Interbull Bull. 32 (2004) 46–52.

[10] Madsen P., Jensen J., Mark T., Reduced rank estimation of (co)variance components for international evaluation using AI-REML, Interbull Bull. 25 (2000) 46–50.

[11] Mantysaari E.A., Multiple-trait across-country evaluation using singular (co)variance matrix and random regression model, Interbull Bull. 32 (2004) 70–74.

[12] Meyer K., To have your steak and eat: genetic principal component analyses of beef cattle data, in: Proceedings of the 8th World Congress on Genetics Applied to Livestock, 13–18 August 2006, Belo Horizonte, Brasil, Paper 25-03.

[13] Meyer K., Kirkpatrick M., Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices, Genet. Sel. Evol. 37 (2005) 1–30.

[14] Minéry S., Fikse W.F., Ducrocq V., Application of structural model to estimate genetic correlations between countries, Interbull Bull. 31 (2003) 175–179.

[15] Norman H.D., Wright J.R., Powell R.L., VanRaden P.M., Impact of maturity rate of daughters on genetic ranking of Holstein bulls, J. Dairy Sci. 88 (2005) 3337–3345.

[16] Rekaya R., Weigel K.A., Gianola D., Application of a structural model for genetic covariances in international dairy sire evaluations, J. Dairy Sci. 84 (2001) 1525–1530.

[17] Schaeffer L.R., Multiple-country comparison of dairy sires, J. Dairy Sci. 77 (1994) 2671–2678.

[18] Sullivan P., Wilton J.W., 2001. Multiple-trait MACE with a variable number of traits per country, Interbull Bull. 27 (2001) 68–72.

[19] Tarres J., Liu Z., Ducrocq V., Reinhardt F., Reents R., Parameter estimation and genetic evaluation of milk production traits from France and Germany with a multi-trait MACE model, Interbull Bull. 35 (2006) 76–87.

[20] Tarres J., Liu Z., Ducrocq V., Reinhardt F., Reents R., Validation of an approximate REML algorithm for parameter estimation in multi-trait MACE model. A simulation study, J. Dairy Sci. 90 (2007) 4846–4855.

APPENDIX

Assuming the following simple multi-trait animal model for data of a bull i:

$$\mathbf{y}_i \;=\; \mathbf{Z}_i\mathbf{u}_i \;+\; \mathbf{e}_i$$

where $\mathbf{y}_i$ is a vector of data of the i-th bull, $\mathbf{u}_i$ is the vector of additive genetic effects of bull i, and $\mathbf{e}_i$ is a vector of residual effects, the (co)variance matrices for the random effects are:

$$\text{Var}\,(\mathbf{u}_i) \;=\; \mathbf{G}, \;\; \text{and} \;\; \text{Var}\,(\mathbf{e}_i) \;=\; \mathbf{R}_i$$

where $\mathbf{G}$ is the genetic (co)variance matrix, and $\mathbf{R}_i$ is the residual (co)variance matrix. Ignoring pedigree information, the mixed-model equations (MME) for estimating the breeding value $\hat{\mathbf{u}}_i$ for bull i are:

$$\left(\mathbf{Z}'_i\mathbf{R}^{-1}_i\mathbf{Z}_i + a^{ii}\mathbf{G}^{-1}\right)\hat{\mathbf{u}}_i = \mathbf{Z}'_i\mathbf{R}^{-1}_i\mathbf{y}_i$$

where a^{ii} is the diagonal element of the inverse of the relationship matrix for bull i.

The genetic (co)variance matrix can be decomposed as $\mathbf{G} = \mathbf{TT}'$ and the vector of additive genetic effects of bull i approximated by $\mathbf{u}_i = \mathbf{T}\mathbf{u}^*_i$. The model can be rewritten as:

$$\mathbf{y}_i = \mathbf{Z}_i\mathbf{T}\mathbf{u}^*_i + \mathbf{e}_i$$

where $\mathbf{y}_i$ is a vector of data of the i-th bull, $\mathbf{u}^*_i$ is the vector of transformed additive genetic effects of bull i, and $\mathbf{e}_i$ is a vector of residual effects. (Co)variance matrices for the random effects are:

$$\text{Var}\left(\mathbf{u}^*_i\right) = \mathbf{G}^*, \quad \text{and} \quad \text{Var}\left(\mathbf{e}_i\right) = \mathbf{R}_i$$

where the genetic (co)variance matrix $\mathbf{G}^* = \mathbf{T}^{-1}\mathbf{G}(\mathbf{T}^{-1})' = \mathbf{I}$ is an identity matrix and $\mathbf{R}_i$ is the residual (co)variance matrix. Ignoring pedigree information, the mixed-model equation for estimating the transformed breeding value $\hat{\mathbf{u}}^*_i$ for bull i is:

$$\left(\mathbf{T}'\mathbf{Z}'_i\mathbf{R}^{-1}_i\mathbf{Z}_i\mathbf{T} + a^{ii}\mathbf{G}^{*-1}\right)\hat{\mathbf{u}}^*_i = \mathbf{T}'\mathbf{Z}'_i\mathbf{R}^{-1}_i\mathbf{y}_i.$$

The transformed breeding values can be back transformed to original basis as $\hat{\mathbf{u}}_i = \mathbf{T}\hat{\mathbf{u}}^*_i$. Here it is important to note that $\mathbf{T}'\mathbf{Z}'_i\mathbf{R}^{-1}_i\mathbf{Z}_i\mathbf{T}$ has no inverse if bull i has missing values for some traits. However, $\mathbf{T}'\mathbf{Z}'_i\mathbf{R}^{-1}_i\mathbf{Z}_i\mathbf{T} + a^{ii}\mathbf{I}$ is always invertible also for bulls with missing values.