Estimation of Genetic Correlations Using a Structural Model for Milk and Type Traits: Limits and Opportunities

H. Leclerc[1,3]*, S. Minéry*[2]*, W.F. Fikse*[3]*, V. Ducrocq*[4] *and T. Druet*[4]

[1] *Union Nationale des Coopératives agricoles d'Elevage et d'Insémination Animale, 75 595 Paris Cedex 12, France*
[2] *Département de Génétique, Institut de l'Elevage, INRA-SGQA, 78 352 Jouy-en-Josas Cedex, France*
[3] *Interbull Centre, Department of Animal Breeding & Genetics, SLU, Box 7023, S – 750 07, Uppsala, Sweden*
[4] *INRA, Station de Génétique Quantitative et Appliquée, 78 352 Jouy-en-Josas Cedex, France*

Introduction

The estimation of genetic correlations between countries needed for Mutiple-trait Across Country Evaluation (MACE) is one of the major problems of international genetic evaluation. The increase of the number of participating countries and the lack of genetic links between some of them lead to computational difficulties. To avoid these problems, the use of structural models has been suggested to exploit patterns in the genetic correlations and to reduce the number of parameters to estimate.

In the structural model proposed by Delaunay *et al.* (2002) as a part of PROTEJE project, genetic correlations are described as a simple function of unspecified country characteristics. The characteristics conceptually define a k-dimensional space whose axes are based on the characteristics of (k+1) "axis countries". The coordinates of the other countries in this base are enough to define all the other correlations. The link function used by Delaunay *et al.* (2002) to define the correlation between two countries i and j was the exponential of the Euclidian distance between the coordinates of these two countries (d_{ij}). Clearly, the smaller dimension k is, the smaller the number of parameters to estimate.

Such a structural model to estimate genetic correlation between countries was successfully tested on a small scale, using milk yield international evaluation data (Minéry *et al.*, 2003).

However, there was some concern on the fact that the use of distances imposed important constraints, since not all correlation matrices can be described appropriately (Delaunay *et al.*, 2002, Goddard, 2004, personal communication) As an example with three countries A, B and C, if rG_{AB}=0.9 ; rG_{AC}=0.7 ; rG_{BC}=0.8), it is not possible to find country coordinates such that correlation $rG_{ij}= \exp(-d_{ij})$. Indeed, on the "distance scale", it is impossible to have at the same time d_{AB}= 0.105, d_{AC}= 0.357, d_{BC}= 0.223, since $d_{AB} + d_{BC} < d_{AC}$

The aim of this study was to test the structural model on traits which are less correlated across countries like type traits, and to enlarge the study on milk field data to investigate the limits and the opportunities of the structural model for the estimation of genetic correlations.

Material & Methods

The data used were deregressed national breeding values of Holstein bulls and their effective daughter contributions (EDC) used in the Interbull routine evaluation of November 2003 for stature and foot angle (concerning respectively 22 and 21 populations, here after referred to as countries) and of August 2003 for milk yield (27 populations).

The sire model currently used in international genetic evaluations (Schaeffer, 1994) was applied (Minéry *et al.*, 2003 for more details).

Genetic correlations were estimated with three different models for genetic correlations: an "unstructured" model, called here classical model (CM), the structural model SM(d_{ij}), of Delaunay *et al.* (2002), in which correlations were calculated as $\exp(-d_{ij})$ and another structural model, SM(d_{ij}^2), with a different function: $rG_{ij} = \exp(-d_{ij}^2)$. This function avoids many of the cases where "triangle inconsistencies" occur. For the example above: d_{AB}= 0.325, d_{AC}= 0.597, d_{BC}= 0.472, and the reparametrisation is possible ($d_{AB} + d_{BC} > d_{AC}$).

In the rest of this paper, SMxy will represent a structural model for which the correlations between y countries are estimated based on the country coordinates in a space of dimension x.

The axis countries were chosen in order to maximize the volume of the space defined by their coordinates. However, at least one well-connected country of each hemisphere (most of time USA for north, and Australia or New Zealand for south) was included as axis countries.

An AI-REML algorithm implemented in a program of Druet *et al.*, (2003) was used for parameters estimation (see Minéry *et al.*, 2003, for more details). Parameters could be kept constant during the iteration process by setting to zero the first derivatives of the likelihood with respect to these parameters. By fixing coordinates for the axis countries, the coordinates for other countries estimated in different runs were relative to the exact same space.

The three models for genetic correlations were compared on the basis of minus two log likelihood (-2logL), Akaike's Information Criterion (AIC), Schwarz's Bayesian Information Criterion (BIC). The latter criterion give increasing importance to the reduction of the number of parameters to estimate. Estimated correlations were compared with each other.

Results and Discussion

1. Stature

The axis countries that defined the larger volume in a three dimensional space were the same as those chosen for the milk yield (Minéry *et al.*, 2003), i.e. Germany, Hungary, New Zealand and USA.

Table 1. Number of parameters, -2logL and information criteria for CM5 and SM35.

Stature	CM5	SM35(d_{ij})
No. parameters	20	19
-2log L	**42863.8**	42865.1
AIC	42903.8	**42903.1**
BIC	42953.5	**42950.3**

Axis countries SM35: USA, NZL, DEU, HUN

The classical model and the structural model in three dimensions (SM35(d_{ij})) gave similar genetic correlations (Appendix 1) and BIC favoured SM in comparison to CM (Table 1). These results were comparable with those obtained with milk yield (Minéry *et al.*, 2003).

2. Foot angle

With SM(d_{ij}), the axis countries determined by the volume of the space were different than those used for milk yield and stature (Australia, Poland, Switzerland and USA). This reflects that the definition of foot angle varies considerable across countries; for Switzerland the trait heel depth is submitted as measure for foot angle.

40% of the estimated genetic correlations differed from more than 0.03 between CM and SM(d_{ij}). Some estimates were extremely different: e.g., correlation between USA and Switzerland was 0.50 with SM35(d_{ij}) and 0.32 with CM5 (Appendix 2). The likelihood, AIC and BIC favoured the classical model (Table 2).

Table 2. Number of parameters, -2logL and information criteria for CM6 and SM36.

Foot angle	CM6	SM36(d_{ij})
No. parameters	27	24
-2log L	**135859.4**	135874.9
AIC	**135913.4**	135922.9
BIC	**135978.8**	135981.1

Axis countries SM36: USA, AUS, CHE, POL

The reason why the structural model does not perform as well as for milk yield and stature may be that it is less adapted for lowly correlated and poorly defined traits like foot angle. Low correlations lead to country coordinates that are farther apart, exacerbating the "triangle" problem. Another reason could be that the program to estimate genetic correlations stopped at different local maxima.

A structural model SM(d_{ij}) in a four dimensional space was used to analyze data from five well-connected countries (Australia, Canada, France, Germany and United Kingdom). From different sets of starting values, the AI-REML program stopped at different points, which were local maxima (Table 3). None of them was equivalent to the

CM probably because of the geometrical constraints imposed by the SM.

Table 3. -2logL from four sets of starting values (SV1 to SV4) with SM(d_{ij}) and CM.

Foot Angle	-2logL
CM5	**78331,4**
SM45, SV1	78406.3
SM45, SV2	78365.8
SM45, SV3	78361.0
SM45, SV4	78385.6

Axis countries SM45: AUS, CAN, DEU, FRA, GBR

The three models CM, SM(d_{ij}) and SM(d_{ij}^2) were then compared for a subset of eight countries (Australia, Canada, France, Germany, Italy, The Netherlands, United Kingdom and USA). When the dimension was high, SM(d_{ij}^2) gave better AIC and BIC than SM(d_{ij}) (Table 4a), and genetic correlations more similar to CM (Appendix 3). However, when the dimension was reduced (Table 4b), SM(d_{ij}) gave better AIC and BIC than SM(d_{ij}^2) and genetic correlations more similar to CM (Appendix 4).

Table 4a. Number of parameters, -2logL, AIC and BIC for CM8, SM78(d_{ij}) and SM78(d_{ij}^2).

Foot angle	CM8	SM78 (d_{ij})	SM78 (d_{ij}^2)
No. param.	44	44	44
-2log L	**201790.7**	202083.7	201843.1
AIC	**201878.7**	202171.7	201931.1
BIC	**201998.4**	202291.4	202050.8

Axis countries SM78: AUS, CAN, DEU, FRA, GBR, ITA, NLD, USA

Table 4b. Number of parameters, -2logL, AIC and BIC for CM8, SM38(d_{ij}) and SM38(d_{ij}^2).

Foot angle	CM8	SM38 (d_{ij})	SM38 (d_{ij}^2)
No. param.	44	34	34
-2log L	**201790.7**	201837.0	201929.2
AIC	**201878.7**	201905.0	201997.2
BIC	201998.4	**201997.5**	202089.7

Axis countries SM38: AUS, CAN, DEU, FRA

3. Milk yield

We started with nine well-connected countries (out of 27): Australia, Canada, France, Germany, Italy, New-Zealand, The Netherlands, United Kingdom and USA. Correlations between these countries are estimated with high precision and allowed us to investigate the severity of the restriction imposed by a structural model. Moreover, the estimation of correlations with the other 18 countries was made easier, since these nine countries provided a maximum amount of links with these other countries.

Table 5. Comparison of three SMs computed with SM(d_{ij}) and the corresponding CM.

Milk	CM9	SM49 (d_{ij})	SM39 (d_{ij})	SM29 (d_{ij})
No. param.	54	44	39	33
-2log L	**817180.2**	817231.3	817236.2	817253.3
AIC	**817288.2**	817319.3	817314.2	817319.3
BIC	817778.2	817718.6	817668.1	**817618.8**
Average deviation corr. SM-CM		0.008	0.008	0.006
Average abs. deviat. corr. SM-CM		0.015	0.015	0.015
Maximum deviation corr. SM-CM		0.061	0.058	0.061
No of abs. deviations corr. >0.01		15	15	17
No of abs. deviations corr. >0.03		8	8	6

Axis countries: SM29: 1-3 ; SM39: 1-4 ; SM49 : 1-5 1=DEU, 2=USA, 3=NZL, 4=FRA, 5=CAN

For SM(d_{ij}) and this subset of countries, only the models including at most four axes converged (Table 5). The best SM(d_{ij}) was the one with only two axes (SM29). It had the lowest BIC, and surprisingly, the correlations estimated with SM29 did not deviate more from CM9 than SMs with higher dimensions.

Table 6. Comparison of three SMs computed with SM(d_{ij}^2) and the corresponding CM.

Milk	CM9	SM89 (d_{ij}^2)	SM79 (d_{ij}^2)	SM69 (d_{ij}^2)
No. param.	54	54	53	51
-2log L	**817180.2**	817180.6	817192.4	817213.7
AIC	**817288.2**	817288.6	817298.4	817315.7
BIC	**817778.2**	817778.6	817779.3	817778.5
Average deviation corr. SM-CM		0.000	-0.005	-0.014
Average abs. deviat. corr. SM-CM		0.000	0.005	0.014
Maximum deviation corr. SM-CM		0.000	-0.045	-0.076
No of abs. deviations corr. >0.01		0	6	14
No of abs. deviations corr. >0.03		0	2	5

Axis countries: SM69: 1-7 ; SM79: 1-8 ; SM89 : 1-9 1=DEU, 2=USA, 3=NZL, 4=FRA, 5=AUS, 6=ITA, 7=GBR, 8=NLD, 9=CAN

For SM(d_{ij}^2), the results for the nine well-connected countries (Table 6) were reasonable only with models having at least six axes. In the other situations, the likelihoods obtained with SM(d_{ij}^2) were lower than the ones with SM(d_{ij}) with the same dimension: for example, -2logL was higher for SM49(d_{ij}^2) than for SM49(d_{ij}) (not shown).

Exactly the same results as CM9 were obtained for SM89(d_{ij}^2) that had the same number of parameters as CM9. It seems that the triangle restriction is much less severe for SM(d_{ij}^2). The best compromise for SM(d_{ij}^2) was SM79. Its BIC was close to the one of CM9.

Genetic correlations estimated by SM79 were not very different from those estimated with CM9, and only two deviated by more than 0.03 in absolute value.

The addition of a new axis in SM(d_{ij}) and SM(d_{ij}^2) improved the likelihood. For SM(d_{ij}^2), increasing the dimension also gave correlations that were closer to CM estimates, whereas we did not observe this for SM(d_{ij}).

Correlations among 22 countries computed with these three models were compared. Of the 231 genetic correlations, only 211 could be computed with CM on different subsets of countries (between four and ten countries per run). Most of missing correlations involved Israel, Estonia, Finland, Switzerland Red and South Africa.

To compute the SM correlations, we added different subsets of countries (added countries) to the axis countries. The coordinates for axis countries were fixed. The coordinates of added countries were thus estimated in the exact same base and were not influenced by small variations of axis countries' coordinates that we observed when the space was not fixed. With the coordinates of each country in the space defined by the axis countries, it was possible to compute the distance between all of them, and so their correlation.

Table 7. Comparison CM and SM (based on 211 common genetic correlations).

Milk		**SM27 & SM28 (d_{ij})**	**SM79 & SM710 (d_{ij}^2)**
Average deviation corr. SM-CM		-0.022	-0.005
Average abs. deviat. corr. SM-CM		0.044	0.024
Maximum deviation corr. SM-CM		-0.290	-0.241
Frequency of correlation deviations SM – CM (%)	0.05 < x	7.1	3.8
	0.03 < x ≤ 0.05	6.2	4.3
	0.01 < x ≤ 0.03	9.0	15.6
	-0.01 ≤ x ≤ 0.01	22.7	45.0
	-0.03 ≤ x < -0.01	16.1	16.1
	-0.05 ≤ x < -0.03	11.4	5.2
	-0.05 < x	27.5	10.0

Axis countries: SM2x: 1-3 ; SM7x: 1-8 ; 1=DEU, 2=USA, 3=NZL, 4=FRA, 5=AUS, 6=ITA, 7=GBR, 8=NLD

Correlations estimated with SM(d_{ij}^2) were closer to CM correlations than those estimated with SM(d_{ij}) (Table 7). 76.7% of correlations with SM(d_{ij}^2) deviated by less than 0.03 from CM estimates, whereas this was only 47.8% for SM(d_{ij}). The maximum deviations were large in the two cases, but were obtained for correlations between countries with poor links: Poland and Switzerland in the case of SM(d_{ij}) and Israel and Hungary with SM(d_{ij}^2).

Deviations between SM(d_{ij}^2) and CM estimations were largest for a few countries with poor genetic ties (e.g., Czech Republic, Estonia, Israel, Poland, Switzerland Red).

When CM has convergence problems because one population has poor ties with the other ones, SM(d_{ij}^2) may be used to bypass these difficulties (Delaunay *et al.,* 2002).

For instance, the CM estimation with nine well-connected countries and Israel did not converge after iterating ten times longer than it took for the SM to converge, despite of similar starting values were used.

Table 8. Comparison of CM / SMs (d_{ij}^2) runs on nine well-connected countries and Israel.

Milk	**CM10**	**SM810 (d_{ij}^2)**	**SM710 (d_{ij}^2)**
No. param.	65	64	62
-2log L	**766089.3**	766089.8	766094.1
AIC	766219.3	**766217.8**	766218.1
BIC	766804.9	766794.4	**766776.7**
Average deviation corr. SM-CM		-0.001	-0.005
Average abs. deviat. corr. SM-CM		0.002	0.005
Maximum deviation corr. SM-CM		-0.019	-0.027
No of abs. deviations corr. >0.01		3	7

Axes countries: SM710: 1-8 ; SM810: 1-9 ; 1=DEU, 2=USA, 3=NZL, 4=FRA, 5=AUS, 6=ITA, 7=GBR, 8=NLD, 9=ESP

The likelihood for SM810(d_{ij}^2) and SM710(d_{ij}^2) were not significantly different ($p<0.05$) from one of CM10 which did not reached convergence (Table 8). Furthermore, the estimated correlations with SMs, in particularly for SM810, were really close to the CM estimates: only three correlations estimated with SM810 deviated more than 0.01 from the CM estimates.

Conclusion

Genetic correlations between countries for stature and milk yield were estimated quite accurately with a structural model SM(d_{ij}) in examples involving a low number of axis countries. When a larger space is defined including more axis countries in order to add more flexibility to the model, serious convergence problems appear and/or local maxima are reached. Furthermore, for foot angle, a trait that is defined differently across countries, correlations estimated with a structural model deviated substantially from correlations estimated with the classical model. In these two situations (low correlations or choice of a large dimensional space), the geometrical restriction imposed by the use of the Euclidian distance seems much more severe than initially found.

The structural model that defined the correlation between two countries as the exponential of minus the square of the Euclidian distance seemed to be less affected by geometrical constraints, although these could not be ruled out completely.

In a large-scale application, the estimation of genetic correlations between 22 countries, revealed that, this time, the dimension of the structural model SM(d_{ij}^2) could not be reduced as much as was anticipated before. A reduction of the number of parameters from 231 correlations (with CM) to 126 coordinates (with SM) caused some 15% of the correlations to change by more than 0.05. Most of these cases concerned country pairs with a low number of common bulls.

It is concluded that the structural models are mainly interesting to use to deal with cases where correlation estimates are near the border of the parameter space, to get reasonable genetic correlations for countries with limited links with most the others.

References

Delaunay, I., Ducrocq, V. & Boichard, D. 2002. A structural model for the matrix of genetic correlations between countries in international evaluations. *Proc. 7th WCGALP*. CD-Rom comm. nº01-14.

Druet, T., Jaffrézic, F., Boichard, D. & Ducrocq, V. 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holsteins cows. *J. Dairy Sci. 86*, 2480-2490.

Minéry, S., Fikse, W.F. & Ducrocq, V. 2003. Application of a structural model to estimate genetic correlations between countries. *Interbull Bulletin 17*, 175-179.

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

APPENDIX

Appendix 1. Estimated genetic correlations for Stature with SM35(d_{ij}) (above the diagonal), and deviations from CM5 (SM35-CM5) (below the diagonal).

Stature	DEU	USA	NZL	HUN	SWE
DEU		0.942	0.879	0.941	0.995
USA	0.010		0.846	0.935	0.942
NZL	0.002	0.005		0.828	0.883
HUN	-0.005	-0.003	-0.001		0.937
SWE	0.005	0.024	-0.006	0.007	

Axis countries SM35: USA, NZL, DEU, HUN

Appendix 2. Estimated genetic correlations for Foot angle with SM36(d_{ij}) (above the diagonal), and deviations from CM6 (SM36-CM6) (below the diagonal).

Foot angle	USA	AUS	CHE	POL	BEL	CAN
USA		0.687	0.503	0.899	0.680	0.915
AUS	0.005		0.485	0.673	0.616	0.714
CHE	**0.179**	**0.054**		0.554	0.737	0.531
POL	-0.023	0.003	-0.016		0.749	0.889
BEL	-0.021	-0.020	**-0.048**	**-0.100**		0.712
CAN	-0.009	-0.003	**0.074**	0.001	**-0.037**	

Axis countries SM36: USA, AUS, CHE, POL

Appendix 3. Deviations (SM78-CM8) with SM78(d_{ij}) above the diagonal and with SM78(d_{ij}^2) below the diagonal.

Foot angle	AUS	CAN	DEU	FRA	GBR	ITA	NLD	USA
AUS		0.000	0.000	0.000	**-0.141**	**-0.125**	**-0.174**	**0.095**
CAN	0.001		0.000	0.000	**-0.135**	**-0.279**	**-0.065**	**-0.117**
DEU	-0.002	-0.005		0.000	**-0.071**	**-0.227**	**-0.046**	**-0.035**
FRA	0.002	-0.004	0.003		**-0.153**	**-0.275**	**-0.074**	**-0.039**
GBR	0.003	0.002	0.004	-0.002		**-0.236**	**-0.146**	**-0.063**
ITA	0.000	-0.003	0.004	-0.002	-0.005		**-0.102**	**-0.361**
NLD	0.012	0.006	0.010	-0.002	0.001	-0.011		**-0.091**
USA	**-0.047**	**-0.060**	**-0.053**	**-0.050**	**-0.049**	**-0.051**	**-0.067**	

Axis countries SM78: AUS, CAN, DEU, FRA, GBR, ITA, NLD, USA

Appendix 4. Deviations (SM38-CM8) with SM38(d_{ij}^2) above the diagonal and with SM38(d_{ij}^2) below the diagonal.

Foot angle	AUS	CAN	DEU	FRA	GBR	ITA	NLD	USA
AUS		0.027	0.016	0.020	-0.013	**0.112**	0.016	**0.052**
CAN	**-0.251**		0.000	0.018	-0.010	0.005	0.004	-0.001
DEU	-0.011	**-0.363**		-0.022	**0.049**	0.003	-0.004	0.020
FRA	**0.034**	**-0.179**	0.026		-0.014	-0.005	0.011	0.008
GBR	**-0.049**	-0.025	**-0.057**	-0.021		-0.005	-0.007	-0.006
ITA	**-0.097**	**-0.033**	**-0.070**	-0.020	**-0.050**		-0.012	-0.002
NLD	**-0.040**	**-0.039**	**-0.091**	**-0.052**	-0.006	**-0.112**		-0.002
USA	**-0.145**	-0.010	**-0.100**	**-0.082**	-0.029	-0.006	0.000	

Axis countries SM38: AUS, CAN, DEU, FRA