Estimation of Genetic Correlations Among Countries in International Dairy Sire Evaluations with Structural Models

H. Leclerc,*†¹ S. Minéry,‡ I. Delaunay,†§ T. Druet,§ W. F. Fikse,* and V. Ducrocq§

*Interbull Centre, Department of Animal Breeding & Genetics, SLU, Box 7023, Uppsala 75007, Sweden

†Union Nationale des Coopératives agricoles d'Élevage et d'Insémination Animale, Paris 75595, France

‡Institut de l'Élevage, Département de Génétique, INRA-SGQA, Jouy-en-Josas 78352, France

§Station de Génétique Quantitative et Appliquée, INRA, Jouy-en-Josas 78352, France

ABSTRACT

The increase in the number of participating countries and the lack of genetic ties between some countries has lead to statistical and computational difficulties in estimating the genetic (co)variance matrix needed for international sire evaluation of milk yield and other traits. Structural models have been proposed to reduce the number of parameters to estimate by exploiting patterns in the genetic correlation matrix. Genetic correlations between countries are described as a simple function of unspecified country characteristics that can be mapped in a space of limited dimensions. Two link functions equal to the exponential of minus the Euclidian distance between the coordinates of two countries and the exponential of minus the square of this Euclidian distance were used for the study on international simulated and field data. On simulated data, it was shown that structural models might allow an easier estimation of genetic correlations close to the border of the parameter space. This is not always possible with an unstructured model. On milk yield data, genetic correlations obtained from 22 countries for structural models based on 2 and 7 dimensions, respectively, were analyzed. Only a structural model with a large number of axes gave reasonable estimates of genetic correlations compared with correlations obtained for an unstructured model: 76.7% of correlations deviated by less than 0.030. Such a model reduces the number of parameters from 231 genetic correlations to 126 coordinates. On foot angle data, large deviations were observed between genetic correlations estimated with an unstructured model and correlations estimated with a structural model, regardless of the number of axes taken into account.

Key words: genetic correlation matrix, structural model, international evaluation

INTRODUCTION

The increasing international trade of genetic material from dairy cattle prompted the development of across-country bull comparisons, which led to the establishment of the International Bull Evaluation Service (**Interbull**, Uppsala, Sweden). Interbull provides international predicted breeding values for each bull on the scale of each participating country using a multiple-trait across country evaluation (**MACE**) approach (Schaeffer, 1994) with national genetic evaluation results as input data. With this approach, performance in each country is considered a different trait, allowing for different genetic parameters in different countries and genetic correlations among countries less than unity. The estimation of genetic correlations has become a challenge with the increase in the number of participating countries and the lack of genetic links between some of them. For instance, in May 2002, among the 351 genetic correlations across 27 populations (hereafter referred to as countries) for milk yield in the Holstein breed, 124 could not be directly estimated (Interbull, 2005a) and were assigned according to similarities between production systems. These correlations ranged from 0.86 to 0.89 between two northern hemisphere countries, and from 0.75 to 0.78 between a northern and a southern hemisphere country (Interbull, 2005b).

In international dairy sire evaluation, traits are currently defined according to country borders, even though the underlying trait (e.g., milk yield) is quite similar for all countries. Thus, the expression of this trait in different countries tends to be highly correlated. To avoid computational difficulties, structural models are often proposed as an alternative to the classical approach (Rekaya et al., 2001; Delaunay et al., 2002). Here, the basic idea behind structural models is to describe the full genetic covariance matrix as a function of fewer parameters. Rekaya et al. (2001) suggested the use of external information to characterize production systems in different regions or countries to describe genetic covariances. However, the use of external information on climate conditions, management practices,

Received June 28, 2005.

Accepted November 23, 2005.

¹Corresponding author: helene.leclerc@jouy.inra.fr

Table 1. Number of countries, generations, animals per generation, and total animals in 2 simulated populations

	Population	
	1	2
No. of countries	4	8
No. of generation	5	4
No. of cows/country/generation	2,560	1,520
No. of cows in the population	51,200	48,640
No. of progeny-tested bulls/generation	40	24
No. of proven bulls/generation	16	8
No. of bull sires/generation	8	3

and genetic composition of the cow population to measure similarities across regions or countries is ambiguous due to lack of uniformity in recording such information across countries. In the structural model proposed by Delaunay et al. (2002) as a part of the Production Traits European Joint Evaluation project (Canavesi et al., 2002), genetic correlations between countries are described as a simple function of unspecified country characteristics that can be mapped in a space of limited dimensions. The link function used by Delaunay et al. (2002) to define the correlation between two countries was the exponential of minus the Euclidian distance between the coordinates of two countries. However, there was some concern about the fact that the use of distances imposed important constraints, because not all correlation matrices can be described with such an approach (Delaunay et al., 2002; Minéry et al., 2003; M. E. Goddard, Univ. Melbourne, Australia, personal communication).

The objective of this study was to present and test the structural model of Delaunay et al. (2002) and a variant of it in the context of international dairy sire evaluations on simulated data and field data for different levels of correlations between countries.

MATERIALS AND METHODS

Data

Simulated Data. Two hypothetical international dairy cattle populations were created. The first was simulated assuming equal genetic correlations among 4 countries, whereas the second one assumed varying genetic correlations among 8 countries (representing a more realistic situation). Designs of populations were made to obtain reasonable computation time and approximately the same number of cows in both populations: 51,200 and 48,640 in the first and second populations, respectively. The number of nonoverlapping generations was reduced from 5 in the first population to 4 in the second one. The characteristics of these populations are described in Table 1. The genetic correlations

used will be presented later. Populations were of equal size per country and per generation. Parents of the next generation were selected across countries based on the results of a MACE at each generation, for which the true genetic parameters were used. To ensure a strong connectedness among countries, each young or proven bull was used for mating in each country. At each generation, 40 and 24 young bulls were progeny-tested based on 24 and 30 female offspring in each country, respectively (i.e., 96 and 240 daughters per young bull). Of these young bulls, the 16 and 8 highest ranking bulls were used for all countries as proven bulls in the next generation and the 8 and 3 highest ranking bulls among the proven bulls were selected as bull sires, respectively. Proven bulls that passed the progeny test had 100 additional female offspring in each country in the next generation. Within each generation, 40 and 24 elite dams were selected across country as bull dams and each one was mated to one bull sire to produce a progeny-tested bull for the next generation. Each cow was mated to either a proven or a young bull to produce one female offspring for the next generation. Mating of selected animals was at random. The cow performance was simulated according to the model:

$$y_{ijk} = b_k + a_{ij} + e_{ijk} \quad [1]$$

where y_{ijk} is the simulated performance of cow i in herd k of country j ; b_k is the herd effect (25 herds per country and per generation) with $b_k \sim N(0,1)$; a_{ij} is the true additive genetic value of cow i in country j with $a_{ij} = \frac{1}{2}a_{sirej} + \frac{1}{2}a_{damj} + \phi_{ij}$, ϕ_{ij} being the cow's Mendelian sampling, $\mathbf{a} \sim N(\mathbf{0}, \mathbf{G} \otimes \mathbf{A})$ where $\mathbf{G}$ is the across-country genetic (co)variance matrix and $\mathbf{A}$ is the additive genetic relationship matrix between animals; and e_{ijk} is the residual with $\mathbf{e} \sim (\mathbf{0}, \mathbf{I}\sigma_e^2)$ where σ_e^2 is the residual variance for country j . For all countries, genetic and residual variance were fixed such as $\sigma_a^2 = 0.25$ and $\sigma_e^2 = 0.75$, that is, a resulting heritability value of 0.25.

Field Data. Field data were used to assess the generality of the results obtained on simulated data for which the populations' structure was not realistic. Data available were the deregressed national breeding values of the bulls used for the Holstein international genetic evaluation. Milk yield analyses were based on data of August 2003, and type trait analyses (foot angle) were based on data of November 2003. The deregression procedure was done by Interbull from the national breeding values sent by the participating countries. It removed the double counting of effects that subsequently were included in the prediction of international breeding values (Jairath et al., 1998). Each deregressed

breeding value was weighed in the analyses by their effective daughter contribution (**EDC**), which considers contemporary group size, correlations between repeated records, and the reliability of the daughter's dam evaluation (Fikse and Banos, 2001). These EDC were sent by each country to Interbull.

Data were edited to include only national evaluations for bulls born after 1984. All the observations were included in the estimation of genetic correlations, in contrast with the current Interbull practice based on subsets of well-connected bulls.

Data from 22 Interbull member countries (Australia, Belgium, Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Israel, Italy, New Zealand, Poland, Spain, South Africa, Switzerland, Switzerland Red, The Netherlands, United Kingdom, and United States) were used for the milk production study. These data sets are characterized in Table 2. The selected countries represented a wide range of production systems and environments. Links between countries were variable, with a number of bulls with daughters in 2 countries (hereafter referred to as common bulls) ranging from 0 (e.g., Estonia–Finland) to 772 (Canada–United States). For foot angle, data from 8 countries were selected (Australia, Canada, France, Germany, Italy, The Netherlands, United Kingdom, and United States). These data sets are characterized in Table 3.

International Evaluation

Model for Simulated Data. Variance components were estimated applying the same model as for simulation [1], and cow performance records were used as input data.

Model for Field Observations. For the variance components estimation, the sire model used in the international genetic evaluations was applied (Schaeffer, 1994). In this MACE, traits in different countries are considered different traits that are genetically correlated. The linear model used was

$$\mathbf{y}_i = \mu_i \mathbf{1} + \mathbf{Z}_i \mathbf{Q} \mathbf{g}_i + \mathbf{Z}_i \mathbf{s}_i + \mathbf{e}_i \quad [2]$$

where $\mathbf{y}_i$ is the vector of deregressed breeding values of bulls for country i , μ_i is the mean for country i , $\mathbf{g}_i$ is the vector of genetic groups effects treated as fixed for the estimation of genetic correlations between countries, $\mathbf{s}_i$ is the vector of random sire transmitting abilities for country i , $\mathbf{e}_i$ is the vector of random residuals, $\mathbf{Z}_i$ is the sire incidence matrix, and $\mathbf{Q}$ is the matrix assigning sires to genetic groups.

For t countries, the variance-covariance matrices of the random effects are

$$\text{Var} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_t \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 \sigma_{e_1}^2 & 0 & \cdots & 0 \\ 0 & \mathbf{D}_2 \sigma_{e_2}^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \mathbf{D}_t \sigma_{e_t}^2 \end{bmatrix} \text{ and}$$

$$\text{Var} \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_t \end{bmatrix} = \begin{bmatrix} \mathbf{A} \sigma_{s_1}^2 & \mathbf{A} \sigma_{s_{12}} & \cdots & \mathbf{A} \sigma_{s_{1t}} \\ \mathbf{A} \sigma_{s_{21}} & \mathbf{A} \sigma_{s_2}^2 & \cdots & \mathbf{A} \sigma_{s_{2t}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A} \sigma_{s_{t1}} & \mathbf{A} \sigma_{s_{t2}} & \cdots & \mathbf{A} \sigma_{s_{tt}}^2 \end{bmatrix},$$

where $\mathbf{D}_i$ is a diagonal matrix with EDC weighting factors (Fikse and Banos, 2001) as elements, $\sigma_{e_i}^2$ is the residual variance for country i , $\mathbf{A}$ is the additive relationship matrix based on sire, maternal grand sire, and maternal grand-dam of the bull, with maternal grand-dam treated as missing and assigned to phantom parent groups, $\sigma_{s_i}^2$ is the sire variance for country i , and $\sigma_{s_{ij}}$ is the sire covariance between countries i and j .

Genetic groups for missing ancestors were formed, based on selection path, year of birth, and origin (defined following country borders). Small groups were merged together first by year of birth and then by origin to achieve a minimum group size of 500 bulls with unknown parents for the estimation of genetic correlations across countries.

Models for the Genetic Covariances. Three different models were used for the genetic variance-covariance matrix: a classical model (**CM**), the structural model of Delaunay et al. (2002) referred to as **SM(d_{XY})**, and a structural model derived from that of Delaunay et al., referred to as **SM(d_{XY}²)**.

In the classical model, the variance-covariance matrix was assumed unstructured. In the structural model of Delaunay et al. (2002), the covariances between countries were defined as a function of a set of unobserved variables (characteristics) for each country that condition the genetic correlations between countries. The country characteristics are represented in a space of k dimensions ($k < \text{number of countries}$), in which the coordinates of the countries are the unobserved characteristics. In this space, the genetic correlation between 2 countries, X and Y , ($r_{G_{XY}}$) was defined as

$$r_{G_{XY}} = \exp(-d_{XY}) \quad [3]$$

where d_{XY} is the Euclidian distance between countries X and Y , computed as $d_{XY} = \sqrt{\sum_k (P_{Xi} - P_{Yi})^2}$, with P_{Xi} and P_{Yi} being the coordinates of countries X and Y ,

Table 2. Number of bulls per country¹ for milk production traits with a national evaluation (diagonal) and common bulls that had progeny in each country pair (above the diagonal)

	AUS	BEL	CAN	CHE	CHR	CZE	DEU	DNK	ESP	EST	FIN	FRA	GBR	HUN	IRL	ISR	ITA	NLD	NZL	POL	USA	ZAF
AUS	3,596	73	346	60	29	119	214	82	139	6	10	193	235	116	105	19	154	277	417	44	501	116
BEL		424	96	61	47	106	184	50	104	9	9	152	138	68	84	9	117	286	76	48	109	62
CAN			5,446	83	61	205	282	98	201	19	12	215	665	213	108	13	225	286	334	70	772	158
CHE				519	47	65	104	42	83	3	5	96	111	48	62	6	83	101	59	30	91	60
CHR					648	42	95	17	31	3	0	36	39	23	22	0	43	104	23	17	93	22
CZE						1,653	419	89	184	21	19	247	228	211	103	16	220	298	150	98	371	90
DEU							11,782	162	229	32	23	331	329	197	156	30	350	616	214	134	521	133
DNK								4,737	92	8	16	89	129	97	72	16	140	148	89	50	124	72
ESP									905	10	21	172	252	153	122	17	230	222	134	80	267	134
EST										276	0	8	15	10	4	1	8	23	9	11	21	2
FIN											615	26	27	14	15	5	24	22	15	6	20	7
FRA												8,371	252	127	139	12	241	265	152	74	443	103
GBR													3,118	207	211	27	284	372	285	81	445	155
HUN														1,329	97	17	187	185	152	75	306	106
IRL															878	20	126	163	125	46	135	90
ISR																570	25	32	31	7	27	15
ITA																	4,590	298	183	70	348	117
NLD																		6,819	376	89	657	136
NZL																			3,009	50	437	116
POL																				3,765	91	39
USA																					18,121	209
ZAF																						611

¹Country codes: AUS = Australia, BEL = Belgium, CAN = Canada, CHE = Switzerland, CHR = Switzerland Red, CZE = Czech Republic, DEU = Germany, DNK = Denmark, ESP = Spain, EST = Estonia, FIN = Finland, FRA = France, GBR = United Kingdom, HUN = Hungary, IRL = Ireland, ISR = Israel, ITA = Italy, NLD = The Netherlands, NZL = New Zealand, POL = Poland, USA = United States, ZAF = South Africa.

Table 3. Number of bulls per country¹ for foot angle trait with a national evaluation (diagonal) and common bulls that had progeny in each country pair (above the diagonal)

	AUS	CAN	DEU	FRA	GBR	ITA	NLD	USA
AUS	1,762	244	107	126	177	124	167	269
CAN		5,406	151	182	373	196	189	601
DEU			10,494	219	208	268	297	305
FRA				7,394	239	215	173	358
GBR					3,008	330	300	442
ITA						4,556	251	356
NLD							4,308	470
USA								16,065

¹Country codes: AUS = Australia, CAN = Canada, DEU = Germany, FRA = France, GBR = United Kingdom, ITA = Italy, NLD = The Netherlands, USA = United States.

respectively, for axis i . According to this definition, the covariance between 2 countries X and Y is $\sigma_{XY} = \sigma_X \cdot \sigma_Y \cdot \exp(-d_{XY})$ with σ_X and σ_Y being the genetic standard deviations in countries X and Y , respectively.

To illustrate, consider 4 countries (A, B, C, and D). Their characteristics can be used to conceptually define the axes of a 3-dimensional space (Figure 1). They are referred to as axis countries. Country A defines the center of the space. Adding country B determines the first axis. The inclusion of countries C and D position the second and third axes, respectively. For example, the correlation between countries B and C (r_{BC}) computed from their coordinates is $r_{BC} = \exp(-d_{BC}) = \exp[-\sqrt{(P_{C1} - P_{B1})^2 + (P_{C2} - 0)^2}]$. In this space, a fifth

country, E, could be added without contributing to the definition of the space. Country E is referred to as an added country. For country E, only 3 coordinates need to be estimated to determine the 4 genetic correlations

with the axis countries. Similarly, 3 additional coordinates need to be estimated for an added country F to determine the 5 additional genetic correlations (4 with the axis countries and one with country E). With this structural model, even if countries E and F do not have any direct link between them, the genetic correlation between them can be computed from the Euclidian distance d_{EF} obtained from their coordinates in the 3-dimensional space by $r_{EF} = \exp(-d_{EF}) = \exp[-\sqrt{(P_{F1} - P_{E1})^2 + (P_{F2} - P_{E2})^2 + (P_{F3} - P_{E3})^2}]$.

In a second structural model derived from that of Delaunay et al. (2002), the genetic correlation between countries X and Y was defined as

$$r_{XY} = \exp(-d_{XY}^2). \quad [4]$$

The use of the square of the Euclidian distance (d_{XY}^2) allowed more flexibility in the model. It solved many of the cases where “triangular inconsistencies” occurred (Delaunay et al., 2002; Minéry et al., 2003; M. E. Goddard, Univ. Melbourne, Australia, personal communication). As an example, take 3 countries, A, B, and C. If $r_{AB} = 0.90$, $r_{AC} = 0.70$, and $r_{BC} = 0.80$, it is not possible to find country coordinates such that $r_{XY} = \exp(-d_{XY})$. Indeed, on the “distance scale”, it is impossible to have, at the same time, $d_{AB} = 0.105$, $d_{AC} = 0.357$, $d_{BC} = 0.223$, because $d_{AB} + d_{BC} < d_{AC}$. With $r_{XY} = \exp(-d_{XY}^2)$, $d_{AB} = 0.325$, $d_{AC} = 0.597$, $d_{BC} = 0.472$, and the reparameterization is possible ($d_{AB} + d_{BC} > d_{AC}$). Another obvious restriction is that correlations must be between 0 and 1 in both cases.

The use of these structural models makes it possible to reduce the number of parameters to estimate for the genetic covariance matrix from $\frac{m(m+1)}{2}$ to $m(k+1) - \frac{k(k+1)}{2}$, where m is the number of countries and k the number of axes.

Figure 1. Geometrical representation of a structural model for 6 countries in a 3-dimensional space. The genetic correlation between 2 countries; e.g., E and F is assumed to be $r_{EF} = \exp(-d_{EF})$ or $r_{EF} = \exp(-d_{EF}^2)$.

In the rest of this paper, CM m will represent a classical model used to estimate the genetic correlations among m countries. The terms SM(d_{XY}) km and SM(d_{XY}^2) km will represent structural models for which the correlations among m countries are estimated based on the country coordinates in a space of dimension k , with the genetic correlations defined as in [3] and [4] respectively.

Algorithm. An average information-REML (AI-REML) algorithm was used for parameter estimation (Johnson and Thompson, 1995). The main advantages of AI-REML are that it converges faster than the expectation maximization-REML algorithm used by Interbull and it provides asymptotic standard errors of the estimates, obtained as the inverse of the AI matrix. For the first population of simulated data, the ASREML software (Gilmour et al., 2002) was used for the estimation of parameters of both the classical model and the structural model SM(d_{XY}). For the second population of simulated data and field data analyses, the AI-REML algorithm implemented by Druet et al. (2003a,b), which allows the user to define parametric structures for the random effects, was used for the estimation of the parameters. For the structural models, the genetic variance-covariance matrix was a nonlinear function of parameters (i.e., the coordinates), so the AI-REML algorithm used a simplified AI matrix, ignoring nonzero terms of the second derivative of the genetic (co)variance matrix (Gilmour et al., 1995).

In the ASREML software and in the AI-REML software used to analyze the field data, the update of the parameters was based on a line search procedure in which the step size was repeatedly divided by 2 until the likelihood increased (Dennis and Schnabel, 1983). For the classical model with field data, the update of the genetic covariance matrix was a combined AI-EM update if the AI-REML update alone lead to a nonpositive definite genetic covariance matrix (Jensen et al., 1996).

Some parameters could be forced to remain 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.

Model Comparison. For the simulated data, the structural and the classical models were compared with the true genetic correlations and on the basis of minus twice the logarithm of the likelihood ($-2\log L$). For the field data, the structural and the classical models were compared on the basis of the estimated genetic correlations, of minus twice the logarithm of the likelihood ($-2\log L$), and of 2 information criteria that take the number of parameters to estimate into account: the

Table 4. Genetic correlations among 8 countries (Co 1 to Co 8) considered for the second simulated population

	Co 1	Co 2	Co 3	Co 4	Co 5	Co 6	Co 7	Co 8
Co 1		0.96	0.97	0.92	0.89	0.92	0.75	0.70
Co 2			0.95	0.94	0.92	0.93	0.78	0.68
Co 3				0.91	0.86	0.92	0.74	0.69
Co 4					0.94	0.94	0.83	0.75
Co 5						0.88	0.77	0.66
Co 6							0.88	0.81
Co 7								0.89
Co 8								

Akaike's information criterion (AIC; Akaike, 1974) and the Schwarz's Bayesian information criterion (BIC; Schwarz, 1978):

$$AIC = -2\log L + 2q, \text{ and}$$

$$BIC = -2\log L + q\log(n - p)$$

where q is the number of parameters, n is the number of observations, and p is the rank of fixed effects matrix computed as $p = (\text{number of genetic groups} + 1) \times \text{number of countries}$. Although not necessarily the most accurate ones, the results obtained with the classical model were used as reference.

Analyses

Simulated Data. For the first population, genetic correlations between countries estimated for the structural model SM(d_{XY}) taking between 1 and 3 axes into account and a classical model were compared with the true values of genetic correlations. A first data set was simulated assuming a genetic correlation between all countries of 0.90; whereas in a second set, a genetic correlation between countries of 0.99 was simulated. This extreme correlation value makes it possible to check the ability of the structural models to deal with the estimation of genetic parameters very close to the border of the parameter space. For the second population, the data were simulated in such a way that genetic correlations between countries ranged from 0.66 to 0.97 (Table 4). Genetic correlations between countries estimated for the structural models SM(d_{XY}) and SM(d_{XY}^2) taking between 1 and 7 axes into account and a classical model were compared with the true values of genetic correlations. No replications of either scenario were made.

Field Data: Determination of Axis Countries. Two traits with various degrees of resemblance between countries were studied: milk yield, as an example of traits that are very similar across countries (i.e., with an average genetic correlation estimated at 0.88

Table 5. Number of parameters to estimate the genetic covariance matrix, the log likelihood, and average absolute deviations of genetic correlations (rG) estimated based on simulated data for a classical model (CM4) and 3 structural models $SM(d_{XY})k4$ for 4 countries ($k = 1$ to 3 axes) from the true genetic correlations

		Model			
		CM4	$SM(d_{XY})14$	$SM(d_{XY})24$	$SM(d_{XY})34$
Number of genetic parameters:		10	7	9	10
$rG = 0.90$	Log likelihood	-24,887.7	Did not converge	-24,888.2	-24,887.7
	Average abs. deviation. rG model-true correlation	0.011		0.029	0.011
$rG = 0.99$	Log likelihood	Did not converge	-24,544.9	-24,543.7	-24,543.5
	Average abs. deviation. rG model-true correlation		0.005	0.005	0.004

in the international evaluation of August 2003; Interbull, 2005a), and foot angle, to represent a more variable trait; that is, with an average genetic correlation between 2 countries estimated at 0.65 (range 0.06 to 0.91) in the international evaluation of November 2003 (Interbull, 2005c). Firstly, the different models were compared based on a subset of well-connected countries (i.e., 8 or 9 countries having a large number of common bulls). This preliminary step made it possible to determine how many axes were needed to obtain accurate estimates of genetic correlations while reducing the number of parameters. Only countries with strong genetic links, measured by the large number of common bulls (Tables 2 and 3) between each other and with the remaining countries, were chosen as axis countries. Such links facilitate the estimation of genetic correlations between countries with weak ties, utilizing indirect links provided by the axis countries. Moreover, the axis countries should be selected to represent a wide range of production environments. In practice, at least one well-connected country of each hemisphere was included as an axis country (United States for the northern hemisphere and Australia or New Zealand for the southern hemisphere). For milk yield, the first 2 axes were always defined by Germany–United States–New Zealand based on the results of Minéry et al. (2003). For milk yield, the axis countries for dimension 3 and higher were chosen to maximize the volume V of the space defined by their coordinates, where $V = |\det(\vec{OK}, \vec{OL}, \vec{OM})|$, if O is the center of the space, K , L , and M define the first, second, and third axis, respectively, as suggested by Minéry (2003). The increase of the number of axes was expected to give more flexibility to the model.

Field Data: Complete Data. For milk yield, 2 specific structural models with a fixed number of axes were selected based on the results from the 9 well-connected countries: one for $SM(d_{XY})$ and one for $SM(d_{XY}^2)$. Genetic correlations estimated with the structural models were compared with estimates obtained for a classical model when all 22 countries described in Table 2 were in-

cluded. Due to memory constraints, it was not possible to estimate correlations between more than 10 countries per run. Therefore, subsets including between 4 and 10 countries per run were created to estimate all correlations with CM. At least 1 or 2 countries providing many links, such as France, Germany, The Netherlands, and United States, were used in each subset. Only 211 of 231 genetic correlations could be computed with CM from the different subsets. For the 211 estimated country pairs, the number of estimates per country pair ranged from 1 (e.g., Czech Republic–Finland, Israel–Spain) to 28 (New Zealand–United States) with, on average, 3.4 estimates per country pair. For country pairs with several estimates, we used the average genetic correlation. For missing genetic correlations (20 country pairs out of 231), which were not considered in this analysis, most of them involved countries with weak links; that is, Estonia, Finland, Israel, South Africa, and Switzerland Red.

To estimate the coordinates for the structural models, different subsets of countries were considered in addition to the axis countries. Here, the coordinates for axis countries were fixed. Therefore, the coordinates of other countries were estimated in the exact same space and were not influenced by small variations of coordinates for axis countries that were observed when the space was not fixed (Minéry, 2003). With the coordinates of all countries in the space defined by the axis countries, it was possible to compute the distance between all pairs of countries, and thus their genetic correlations.

RESULTS

Simulated Data

As expected for the first population (Table 5), when all simulated genetic correlations between countries were 0.90, CM4 and $SM(d_{XY})34$, which presented the same number of parameters, gave equal likelihood and average absolute deviations from the true values. When the true genetic correlations were 0.90, $SM(d_{XY})14$ gave correlations outside the parameter space. This struc-

Table 6. Number of parameters to estimate the genetic covariance matrix, minus twice the log likelihood ($-2\log L$) Akaike's information criterion (AIC), and average absolute and distribution of deviations (rG) of genetic correlations estimated based on simulated data for a classical model (CM8) and structural models $SM(d_{XY})k8$ and $SM(d_{XY}^2)k8$ for 8 countries ($k = 1$ to 7 axes) from the true genetic correlations

	$SM(d_{XY})k8$					$SM(d_{XY}^2)k8$				
	CM8	SM18	SM28	SM38	SM48	SM38	SM48	SM58	SM68	SM78
Number of genetic parameters:	36	15	21	26	30	26	30	33	35	36
$-2\log L$ model – CM8	0.0	+21.0	+12.3	+12.1	+12.9	+1.4	–0.4	–1.8	+14.9	+15.5
AIC model – CM8	0.0	–21.0	–17.7	–7.8	+0.9	–18.6	–12.4	–7.8	+12.9	+15.5
Av. abs. dev. rG model-true rG	0.018	0.031	0.022	0.022	0.024	0.024	0.024	0.020	0.020	0.020
No. of abs. deviation $rG > 0.030$	6	13	9	8	6	10	8	6	5	6
No. of abs. deviation $rG > 0.050$	3	3	5	4	4	5	4	2	2	2

tural model has only one dimension, and it was not possible to obtain a spatial representation of such genetic correlations on a straight line, because all countries should be equidistant to each other. When genetic correlations of 0.99 were simulated, the unstructured model (CM4) failed to converge. The AI-REML algorithm converged to a nonpositive definite genetic covariance matrix and the results were discarded. All structural models tested gave consistent results with an average absolute deviation to the true value equal or lower than 0.005.

For the second population (Table 6), the iteration process of the unstructured model took 5 times longer than it took for $SM(d_{XY}^2)48$ to obtain similar correlations, and the likelihood was lower than the one obtained for some structural models. Average absolute deviations obtained for the structural models to the true correlations were around 0.020, as for the unstructured model. Surprisingly, $SM(d_{XY}^2)58$ presented the highest likelihood obtained for this population, indicating that CM may not have reached convergence and that $SM(d_{XY}^2)68$ and $SM(d_{XY}^2)78$ may have converged to a local maximum.

Milk Yield Field Data

Determination of Axis Countries. The 9 well-connected countries selected were Australia, Canada,

France, Germany, Italy, New Zealand, The Netherlands, United Kingdom, and United States. Strong links allowed an accurate and precise estimation of the genetic correlations using a CM. The lowest number of common bulls was 152 between France and New Zealand (Table 2). These countries were expected to define a stable space into which the remaining countries could be mapped.

For $SM(d_{XY})$, only the models including at most 4 axes gave consistent results; that is, where the increase of the space dimension led to an increase of the likelihood (Table 7). Model $SM(d_{XY})29$ appeared to be the most interesting compromise between the accuracy of estimated genetic correlations and the reduction of the number of parameters compared with the CM9 results (Table 8). The lower likelihood observed for $SM(d_{XY})29$ was compensated by the reduction of parameters and led to the lowest BIC. Interestingly, the correlations estimated with this model did not deviate substantially more from the CM9 estimates than $SM(d_{XY})$ with 3 or 4 dimensions (Table 8). All deviations of correlations larger than 0.030 were for pairs of countries from the southern and northern hemispheres.

For $SM(d_{XY}^2)$, the results with the 9 well-connected countries (Table 7) were only reasonable, in terms of likelihood, for models having at least 6 axes. For the other situations, the likelihoods obtained for $SM(d_{XY}^2)$ were lower than for $SM(d_{XY})$ with the same dimension;

Table 7. Number of parameters to estimate the genetic covariance matrix for milk yield trait, minus twice the log likelihood ($-2\log L$) obtained for the structural models $SM(d_{XY})k9$ and $SM(d_{XY}^2)k9$ for 9 countries (Australia, Canada, France, Germany, Italy, New Zealand, The Netherlands, United Kingdom, and United States) with models taking from 2 to 8 axes (k) into account, in comparison with the classical model (CM9).

	Model						
	SM29	SM39	SM49	SM59	SM69	SM79	SM89
Number of parameter SM-CM9:	–21	–15	–10	–6	–3	–1	0
$-2\log L$ $SM(d_{XY})$ -CM9	+73.1	+56.0	+51.1	+52.7	+72.1	Did not converge ¹	Did not converge ¹
$-2\log L$ $SM(d_{XY}^2)$ -CM9	+994.2	+468.2	+226.8	+81.6	+33.5	+12.2	+0.4

¹Average information-REML algorithm did not converge.

Table 8. Number of parameters to estimate the genetic covariance matrix for milk yield trait, minus twice the log likelihood, ($-2\log L$), Akaike's and Schwarz's Bayesian information criteria, and average, maximum and distribution of deviations of genetic correlations for three $SM(d_{XY})k9$ and two $SM(d_{XY}^2)k9$ with models taking from 2 to 4 axes (k) and from 6 to 7 axes, respectively, into account for 9 countries (Australia, Canada, France, Germany, Italy, New Zealand, The Netherlands, United Kingdom and United States) in comparison with the classical model (CM9)

No. of parameter SM-CM9:	Model				
	$SM(d_{XY})29^1$	$SM(d_{XY})39^1$	$SM(d_{XY})49^1$	$SM(d_{XY}^2)69^2$	$SM(d_{XY}^2)79^2$
	−21	−15	−10	−3	−1
$-2\log L$ SM-CM9	+73.1	+56.0	+51.1	+33.5	+12.2
AIC SM-CM9	+31.1	+26.0	+31.1	+27.5	+10.2
BIC SM-CM9	−159.6	−110.1	−59.4	+0.3	+1.1
Average deviation rG SM-CM9	0.006	0.008	0.008	−0.014	−0.005
Average abs. deviation rG SM-CM9	0.015	0.015	0.015	0.014	0.005
Maximum deviation rG SM-CM9	0.061	0.058	0.061	−0.076	−0.045
No. of abs. deviations $rG > 0.010$	17	15	15	14	6
No. of abs. deviations $rG > 0.030$	6	8	8	5	2

¹Axis countries: $SM(d_{XY})29$ = countries 1 to 3; $SM(d_{XY})39$ = countries 1 to 4; $SM(d_{XY})49$ = countries 1 to 5; with 1 = Germany, 2 = United States, 3 = New Zealand, 4 = France, 5 = Canada.

²Axis countries: $SM(d_{XY}^2)69$ = countries 1 to 7; $SM(d_{XY}^2)79$ = countries 1 to 8; with 1 = Germany, 2 = United States, 3 = New Zealand, 4 = France, 5 = Australia, 6 = Italy, 7 = United Kingdom, 8 = The Netherlands.

for instance, $-2\log L$ was higher for $SM(d_{XY}^2)49$ than for $SM(d_{XY})49$ (+175.7). The likelihood and genetic correlations obtained for $SM(d_{XY}^2)89$ were the same as for CM9. In contrast with what Minéry (2003) obtained with $SM(d_{XY})$, it seems that the triangular restriction is less severe for $SM(d_{XY}^2)$ than for $SM(d_{XY})$. The most interesting model appeared to be $SM(d_{XY}^2)79$. Genetic correlations estimated by $SM(d_{XY}^2)79$ were not very different from those estimated with CM9 (Table 8); only 2 deviated by more than 0.030 in absolute value (0.042 for Germany–New Zealand and 0.045 for Canada–New Zealand). Although the reduction of the number of parameters to estimate is negligible in that case, it should be remembered that the final goal is to use these 8 axis countries to find the position in the space spanned of the 14 other countries. Therefore, the overall reduction of the number of parameters is significant.

Complete Data. Genetic correlations among all countries estimated with $SM(d_{XY})7m$ were on average closer to the CM correlations than those estimated with $SM(d_{XY})2m$ (Figure 2A). With $SM(d_{XY}^2)7m$, 76.7% of the correlations deviated by less than 0.030 from the CM estimates, whereas this proportion was only 47.8% for $SM(d_{XY})2m$. Both models presented similar extreme deviations: -0.290 for $SM(d_{XY})2m$ and -0.241 for $SM(d_{XY}^2)7m$, which were for pairs of countries with weak links to each other: Poland–Switzerland (30 common bulls) and Hungary–Israel (17 common bulls), respectively.

When a distinction was made between correlations estimated among axis countries or between axis and nonaxis countries and correlations computed from coor-

dinates estimated in different runs, the pattern of deviations differed from above (Figure 2B and 2C). For correlations estimated among axis countries or between axis and nonaxis countries, 93.1% deviated by less than 0.030 from the CM correlations for $SM(d_{XY}^2)7m$, compared with only 66.7% for $SM(d_{XY})2m$. However, correlations among nonaxis countries computed from the coordinates estimated in different runs showed large deviations for both structural models. The average absolute deviations were 0.050 and 0.055 for $SM(d_{XY}^2)7m$ and $SM(d_{XY})2m$, respectively, and 60.0% and 66.9% of correlations deviated by more than 0.030 from CM estimates.

Foot Angle Field Data. The 8 well-connected countries selected were Australia, Canada, France, Germany, Italy, The Netherlands, United Kingdom, and United States. As for milk yield results, both structural models were compared using a model including a limited number of axes [3] and a model including a large number of axes (7). Model $SM(d_{XY}^2)78$ had the same number of parameters to estimate as CM8 but it did not give values of $-2\log L$, AIC, and BIC close to those of CM8 (+52.4). It was also noticed that the likelihood of $SM(d_{XY}^2)78$ was even lower than the one of $SM(d_{XY})38$, indicating computational inconsistency. For both structural models, it was found that different sets of starting values could lead to very different likelihood values, corresponding to different local maxima. For instance, with 4 sets of starting values, we obtained 4 different values of $-2\log L$ at convergence with $SM(d_{XY})45$ (ranging from +29.6 to +74.9 relative to $-2\log L$ for the classical model CM5).

Figure 2. Percentage of deviations for milk yield trait between genetic correlations computed based on coordinates estimated for structural model $SM(d_{XY})2m$ (i.e., model taking 2 axes into account; gray bars) and $SM(d_{XY}^2)7m$ (i.e., model taking 7 axes into account; black bars) compared with the corresponding correlations estimated with classical model (CM) for 22 countries (panel A). A distinction was made between correlations estimated among axis countries or between axis and nonaxis countries (panel B) and correlations among nonaxis countries computed from coordinates estimated in different runs (panel C).

DISCUSSION

An unstructured model (CM) was used as reference to assess the quality of genetic correlation estimates using two structural models. However, CM estimates do not represent the true values, the most accurate ones. Unstructured models also present some limitations. The number of parameters to estimate is large, so the precision of each parameter estimated is lower on average than with a parsimonious model such as the two structural models used here. Moreover, it is necessary to create subsets of countries to estimate CM genetic correlations with a large number of countries, because of memory constraints. This is why estimation of genetic correlations using an unstructured model is a time-consuming step in international genetic evaluation. The use of the structural models proposed here for the estimation of genetic correlations seems interesting in terms of reduction of number of runs, when dealing with a large number of countries. For instance, in the application to 22 countries, the 231 genetic correlations estimated for a classical model can be derived from the 126 coordinates estimated with a structural model defined with 7 axes [$SM(d_{XY}^2)7m$] and from only 41 coordinates with a structural model defined with 2 axes [$SM(d_{XY})2m$].

The agreement between SM and CM correlation estimates was rather disappointing. The use of $SM(d_{XY}^2)7m$ or $SM(d_{XY})2m$ led to deviations with respect to CM of more than 0.030 for 23 and nearly 50% of the correlations, respectively. A more detailed examination showed that between 70 and 80% of the large deviations of these estimates (larger than 0.030) were for correlations computed among nonaxis countries. The accuracy of the calculated genetic correlations among nonaxis countries were much lower than the ones estimated among axis countries or between axis and nonaxis countries, and were lower than expected based on a previous study (Minéry et al., 2003).

According to the analysis of simulated data, the structural model allows the estimation of genetic correlations that are very close to the border of the parameter space, which was not possible with an unstructured model. The structural model intrinsically involves a restriction on correlations that is not included in the unstructured model. This study only considers the comparison of the results obtained from structural models against methods used by Interbull (AI-REML or expectation maximization-REML algorithm). Very high genetic correlations (e.g., larger than 0.95) are common in international evaluations, especially with traits very similar across countries such as milk, fat, and protein yield. Unfortunately, the structural models used in this study did not perform well for traits moderately corre-

lated such as foot angle, whose definition varies considerably across countries (e.g., in Switzerland, another trait, heel depth, is used as a measure for foot angle). Low correlations lead to country coordinates that are further apart, exacerbating problems of triangular inconsistency.

The accuracy of genetic correlation estimates largely depended on the dimension of the parameter space defined by the axis countries. For milk yield field data, genetic correlations between countries were estimated quite accurately with a structural model $SM(d_{XY})$ in examples involving a low number of axis countries. However, when a larger space was considered, including more axis countries to add more flexibility to the model, serious convergence problems appeared and correlations estimated with $SM(d_{XY})$ deviated substantially from correlations estimated with the classical model. At least some of these problems are related to the evidence of local maxima, due to strong geometrical constraints or flat likelihood profiles. Such problems exist with $SM(d_{XY}^2)$ even though geometrical constraints are lower. These problems indicate that the triangular inconsistency should not be overlooked. The constraints have consequences on the maximization procedure, with different maxima reached depending on starting values. In contrast, for $SM(d_{XY}^2)$, the estimates were mostly accurate only with models involving a large number of axis countries. In this case, the reduction of the number of parameters to estimate is not as large as hoped (Minéry et al., 2003).

The choice of axis countries is another important issue for our structural models. The axis countries should represent most of the production systems that exist in the participating countries. The choice of axis countries was based on maximization of the volume of space defined by the coordinates of the axis countries. However, in view of the disappointing results obtained for some countries (e.g., Czech Republic, Estonia, Poland), it seems that some of the 22 countries were not correctly represented in the space defined by the 3 [for $SM(d_{XY})2m$] or 8 [for $SM(d_{XY}^2)7m$] well-connected countries that were chosen as axis countries. The motivation for working with a preliminary subset of 9 well-connected countries was to provide, through these countries, indirect links to the other countries, but it seems that this is not sufficient. Therefore, a compromise should be found between the amount of links between countries and their representativeness.

One of the expected advantages of our structural models was that only the coordinates of the country in the space defined by axis countries were needed to estimate correlations with all the other countries. This is an attractive property when a new country wants to

join international evaluation or one of the other countries changes something in its genetic evaluation, because structural models avoid time-consuming estimation. From this perspective, structural models with a large number of axes (e.g., 8 countries), although preferred for their higher accuracy, are not appealing from a practical point of view because the probability that one of these countries modifies its evaluation is high, which means that the considered space is modified and all coordinates need to be reestimated.

The use of Euclidian distance to define our structural models imposes obvious restrictions because only positive correlations can be estimated. Alternative link functions exist to encompass the correlation range from -1 to 1 [e.g., $rG_{XY} = -\left(\frac{d_{XY} - 1}{d_{XY} + 1}\right)$ or $rG_{XY} = 2\exp(-d_{XY}) - 1$]. The geometrical restrictions imposed by the use of

the Euclidian distance are strong. It can be shown that a similar structural model could be implemented to completely remove these constraints, enlarging the space to complex numbers. Unfortunately, in that case, the genetic correlation matrix is no longer ensured to be positive definite. However, as the number of participating countries to international evaluation is likely to keep growing, it seems obvious that unstructured correlation matrices are far from optimal. There are other structural models, such as models based on principal components or factor analysis, for which promising results have been obtained (Leclerc et al., 2005).

CONCLUSIONS

The structural model $SM(d_{XY}^2)$ seems, as expected, to be less affected by geometrical restrictions than $SM(d_{XY})$, but such restrictions still exist. The model $SM(d_{XY}^2)$ was able to explain genetic correlations between countries better than $SM(d_{XY})$ but required a larger number of axes to obtain estimates of correlations close to those obtained for a classical model. The benefits in terms of reduction of the number of parameters were thus lower than expected and numerical problems to find a global maximum were encountered.

It is concluded that the structural models envisioned here are mainly interesting 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 of the others.

REFERENCES

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automatic Control* 19:716–723.
- Canavesi, F., D. Boichard, V. Ducrocq, N. Gengler, G. De Jong, and Z. Liu. 2002. An alternative procedure for international evaluations:

- PROduction Traits European Joint Evalauton (PROTEJE). Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France. Communication 01-59.
- Delaunay, I., V. Ducrocq, and D. Boichard. 2002. A structural model for the matrix of genetic correlations between countries in international evaluation. Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France. Communication 01-14.
- Dennis, J. E., and R. B. Schnabel. 1983. Numerical methods for unconstrained optimization and nonlinear equations. Prentice-Hall, Englewood Cliffs, NJ.
- Druet, T., F. Jaffrézic, D. Boichard, and V. Ducrocq. 2003a. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. *J. Dairy Sci.* 86:2480–2490.
- Druet, T., F. Jaffrézic, and V. Ducrocq. 2003b. Estimation of genetic parameters of test day records for milk yield for the three first lactations of French Holstein cows. Proc. 54th Annu. Mtg. Eur. Assoc. Anim. Prod., Roma, Italy. Communication G5-08.
- Fikse, W. F., and G. Banos. 2001. Weighting factors of sire daughter information in international genetic evaluations. *J. Dairy Sci.* 84:1759–1767.
- Gilmour, A. R., R. Thompson, and B. R. Cullis. 1995. Average Information REML: An efficient algorithm for variance parameters estimation in linear mixed models. *Biometrics* 51:1440–1450.
- Gilmour, A. R., R. Thompson, B. R. Cullis, and S. J. Welham. 2002. ASREML estimates variance matrices from multivariate data using the animal model. Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France. Communication 28-05.
- Interbull. 2005a. Genetic evaluations—Production—August 2003—Appendix I – Holstein Milk. <http://www-interbull.slu.se/eval/framesida-prod.htm> Accessed Jan. 18, 2005.
- Interbull. 2005b. Genetic evaluations—Production—May 2002. <http://www-interbull.slu.se/eval/framesida-prod.htm> Accessed Jan. 18, 2005.
- Interbull. 2005c. Genetic evaluations—Conformation—November 2003—Appendix IV – Foot Angle. <http://www-interbull.slu.se/conform/framesida-conf.htm> Accessed Jan. 18, 2005.
- Jairath, L., J. C. M. Dekkers, L. R. Schaeffer, Z. Liu, E. B. Burnside, and B. Kolstad. 1998. Genetic evaluation for herd life in Canada. *J. Dairy Sci.* 81:550–562.
- Jensen, J., E. A. Mäntysaari, P. Madsen, and R. Thompson. 1996. Residual maximum likelihood estimation of (co)variance components in multivariate mixed linear models using average information. *J. Ind. Soc. Agric. Stat.* 49:215–236.
- Johnson, D. L., and R. Thompson. 1995. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and an average information. *J. Dairy Sci.* 78:449–456.
- Leclerc, H., W. F. Fikse, and V. Ducrocq. 2005. Principal components and approximate factorial approaches for estimating genetic correlations among countries in International dairy sire evaluation. *J. Dairy Sci.* 88:3306–3315.
- Minéry, S. 2003. Application of a structural model to estimate genetic correlations between countries. M.Sc. Thesis, Inst. Nat. Agron. Paris-Grignon, Paris, France.
- Minéry, S., W. F. Fikse, and V. Ducrocq. 2003. Application of a structural model to estimate genetic correlations between countries. *Interbull Bull.* 31:175–179.
- Rekaya, R., K. A. Weigel, and D. Gianola. 2001. Application of a structural model for genetic covariances in international dairy sire evaluations. *J. Dairy Sci.* 84:1525–1530.
- Schaeffer, L. R. 1994. Multiple-country comparison of dairy sires. *J. Dairy Sci.* 77:2671–2678.
- Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat.* 6:461–464.