DEA de génétique multifactorielle
Master of multifactorial genetics

Application of a structural model to estimate genetic correlations between countries

Stéphanie MINÉRY

Supervised by Freddy FIKSE
INTERBULL Centre, SWEDEN

Defended the 02/09/2003

Acknowledgements

I thank Freddy Fikse for supervising my work during these 6 months at the INTERBULL Centre. He is an excellent supervisor: clever, patient, helpful and very pedagogic. I learnt a lot with him: about genetics of course, but also about reporting and presentation of the results. I hope that if I become a supervisor in the future, I will be as good as he was with me.

I thank Ulf Emanuelson for accepting me as a student in the INTERBULL team. He has welcomed me warmly at the INTERBULL Centre and in Sweden. He has helped me since my arrival until my departure to feel very well in Uppsala. He taught me about INTERBULL procedures. I also thank his family, Margareta, Klas and Magnus, for their kindness and their support.

I thank Jan Philipsson for accepting me for 6 months at the INTERBULL Centre and at the Department of Animal Breeding and Genetics at SLU.

I thank l'Institut de l'Elevage, my employer, for allowing me to go for a practical period to Sweden, at the INTERBULL Centre.

I thank Vincent Ducrocq, Sophie Mattalia, Tom Druet and Isabelle Delaunay for their help before and during my training period.

I thank Susanne Eriksson, Jette Jacobsen, Thomas Mark and Hossein Jorjani for integrating me, and for the good times spent together during these 6 months.
I thank in particular:
Susanne, for her large generosity, for the riding on Icelandic horses, and for the support that she and Freddy gave to me.
Jette, for her kindness and for all the tourist visits of Sweden that we have done together.
Thomas, for his kindness and his sense of humor which makes things nice.
Hossein, for the interesting discussions.

I thank Lotta Rydhmer and all the other teachers of the "Genetics and Animal Breeding Theory" for their interesting lectures.

I thank the students that I have met during these 6 months, exchange and Swedish students. I enjoyed my evenings and week ends in Sweden with them. Life in Gälbo was nice!

I thank my family for their support.

I thank Franck, for his love.

Contents

Tables

Figures and Appendix

Figures

Appendix

Abbreviations

-2logL: Minus two times the logarithm of the likelihood function
AIC: Akaike's Information criteria
AUS: Australia
BIC: Schwarz' Bayesian criteria
BLUE: Best Linear Unbiased Estimation
BLUP: Best Linear Unbiased Predictor
CM: Classical Model (No. countries)
DEU: Germany
DNK: Denmark
DYD: Daughter Yield Deviation
EBV: Estimated Breeding Value
EDC: Effective Daughter Contribution
EST: Estonia
ETA: Estimated Transmission Ability
FAO: Food and Agriculture Organization
FIN: Finland
FRA: France
Gb: Giga bytes
HUN: Hungary
INTERBULL: INTERnational BULL evaluation service
MACE: Multiple-trait Across Country Evaluation
MGD: Maternal Grand Dam
MGS: Maternal Grand Sire
MME: Mixed Model Equations
MTMACE: Multiple-trait multiple-country international evaluation
No.: Number
NLD: the Netherlands
NZL: New Zealand
PROTEJE: PROduction Traits European Joint Evaluation
Rg: Genetic correlation
SE: Standard error
SM: Structural model (No. axes, No. countries)
USA: United States of America

Introduction

For 25 years, countries exchanged an increasing amount of genetic material, as frozen semen, embryos or live animals. Direct comparison of national estimated breeding value (EBV) is not possible, because every country has its own way to calculate and to express genetic merit, with different units and/or level for their base population. As an example, USA uses transmission abilities (ETA) for a bull (half of his EBV), and pounds, whereas France uses EBV and kilograms. National genetic evaluation systems (number of lactation included for instance) can be very different from one country to another one.

Methods to compare bulls across countries became necessary. This comparison allows people involved in dairy production to select best bulls not only among domestic bulls but also among foreign bulls, anywhere in the world.

The current method used by INTERBULL, the organisation in charge of the international evaluation, is a multiple-trait across country evaluation (MACE). Application of this method requires estimation of the genetic parameters. The increasing number of countries involved in the international evaluation and lack of connectedness between some of the countries make this estimation more and more difficult, especially the estimation of the covariances between countries. In this context, structural models have been proposed to improve the estimation, more particularly by reducing the number of parameters. One structural model, presented by Delaunay et al. (2002) has been tested on simulated data and on the current matrix of genetic correlations used for the international evaluations by INTERBULL. Aim of the present study was to test it on field data.

After a review on international evaluations, the context of the study will be described carefully. Then, the main results obtained will be presented and discussed.

I. General presentation of international evaluations

I.A. From conversion formulas to MACE

The idea of comparing animals across countries was old. For example, the FAO (Food and Agriculture Organization) initiated a project where cows from different countries were placed in the same experimental farm to compare their genetic level (Stolzmann et al., 1981). Such a method is expensive, time consuming and does not allow comparison of many bulls.

I.A.1. Conversion formulas

The first method used is based on formulas that convert a bull's estimated transmitting ability in the exporting country (A) to the base and unit equivalent in the importing country (B). These formulas can be obtained from a simple regression of country B on country A (Banos and Sigurdsson, 1996):

$$I_B = a + b * I_A,$$

where
I_A = original national evaluation of a bull in country A,
I_B = converted evaluation of the same bull in country B,
a = intercept representing the genetic base difference between A and B, and
b = slope representing the scale difference, and the genetic correlation between A and B.

The theoretical expectation for the slope is:

$b = \rho_g * \frac{\sigma_B}{\sigma_A}$, where

ρ_g = genetic correlation between countries A and B,
σ_A = genetic standard deviation in country A, and
σ_B = genetic standard deviation in country B.

This method is straightforward and easy to apply.

Coefficients 'a' and 'b' are computed from bulls evaluated in both countries A and B (tested in both countries or first tested in country A and then exported in country B) (Banos, 1994). Different methods have been developed by Goddard (1985) and Wilmink et al. (1986) to calculate these regression coefficients. These require having unbiased breeding values of bulls in both countries.

Breeding values of imported bulls are known to be often problematic because of several reasons including preferential treatment of their daughters, and non-random mating. To avoid this, Mattalia and Bonaïti (1993) and Powel and Wiggans (1995) have proposed procedure based on analysis of full sibs families. A 'family' should have at least one full sib progeny tested in the importing country, and one in the exporting country. In this method, the coefficient 'a' is estimated by comparing within families Daughter Yield Deviations (DYD) of the bulls tested in one country to the DYD of the bulls tested in the other country. The full sibs comparison can be used only when sufficient number of families exists.

The DYD of a bull is the weighted average of daughter yields adjusted for solutions for all fixed effects and genetic merit of mates (VanRaden and Wiggans, 1991):

$$\text{DYD} = \frac{\sum_{i=1}^{n} q_{prog} w_{prog} \left(c_i - \frac{1}{2} a_{mi} \right)}{\sum_{i=1}^{n} q_{prog} w_{prog}},$$

where
n = number of daughters of the bull,
q_{prog} = 1 if daughter's dam is known and 2/3 if not,
w_{prog} = weighting factor depending of heritability of the trait and lactation length's weight,
a_{mi} = genetic merit of mate of daughter i, and
c_i = yield deviation of daughter i from all fixed effects, i.e., Y_i – FE,
with Y_i = phenotypic record of the daughter i, and
FE = BLUE solutions for all fixed effect.

Conversion formulas have some limits. First, computation of conversion coefficients is based only on a small and selected number of bulls, which can lead to biased conversion equations. Moreover, it ignores genetic relationship between bulls, and countries can only be compared 2 by 2 at a time.

I.A.2. Mixed linear model methods

I.A.2.a. ***Linear model of Schaeffer (1985)***

This approach is based on analysis of DYD or deregressed national breeding values of bulls from different countries with a linear model. Contrary to conversion equations, it allows for the use of genetic relationship among bulls, which utilizes genetic links between countries. In theory, an unlimited number of countries can be considered simultaneously.

The model proposed was:

$$y = Xc + ZQg + Zs + e ,$$

where

y = vector of observations,
c = vector of country of evaluation fixed effect,
g = vector of genetic groups effect, based on country and year of birth of the bull,
s = vector of random sire transmitting abilities,
e = vector of random residuals,
X = matrix that relates observations to countries,
Z = matrix that relates observations to sires, and
Q = matrix that relates sires to genetic groups.

Estimates of c, g and s are solutions of the Mixed Model Equation (MME). This international BLUP allows for the use of the inverse of the additive relationship matrix based on sire and maternal grandsire (MGS) of the bull, A^{-1}. In this model, the residual variance-covariance matrix is $D\sigma_e^2$, where D is a diagonal matrix with elements equal to 1 over n_{ij}, with n_{ij} = number of daughters for the j^{th} sire in the i^{th} country and σ_e^2 is the residual variance. The sire variance-covariance matrix is $A\sigma_s^2$ where σ_s^2 is the sire variance.

From these results, international sire evaluations are obtained by adding genetic group solutions to the sire solutions.

Observations used in this model can either be DYD or deregressed national breeding values of the bulls. DYD are not calculated by every country. Consequently, the variables chosen for the linear model are the deregressed breeding values, which are analogous to DYD (Schaeffer, 1994). The current method used to deregress national evaluation results is described by Jairath et al. (1998). This procedure makes the observations independent of genetic group effect and relationship among animals, to avoid counting it several times (Sigurdsson and Banos, 1995), and it takes into account accuracy of the national breeding values. Before analysis, observations are expressed in ETA and standardized within country using the sire standard deviation (Banos, 1994).

This linear model assumes that there is no genotype – environment interaction, contrary to the method of conversion formulas. Thus, in this linear model traits are supposed to be the same in different countries. In other words, genetic correlations between daughter performances across countries are equal to unity. Moreover, it does not allow for different heritabilities of the trait for countries: sire and error variances are assumed to be common to all countries.

I.A.2.b. MACE, Multiple-trait Across Country Evaluation (Schaeffer, 1994)

MACE is a multiple-trait across countries sire evaluation, where traits in different countries are considered as different and genetically correlated traits. It allows different heritability for each country and requires estimates of genetic correlations among sire genetic effects in different countries. The linear model uses is:

$$y_i = \mu_i 1 + Z_i Q g_i + Z_i s_i + e_i ,$$

where

y_i = vector of deregressed breeding value of bulls for country i,
μ_i = mean for country i which reflect the definition of the genetic base for country i,
g_i = vector of genetic groups effects,
s_i = vector of random sire transmitting abilities for country i,
e_i = vector of random residuals,
Z_i = matrix that relates observations to sires, and
Q = matrix that relates sires to genetic groups.

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

$$V\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_t \end{bmatrix} = \begin{bmatrix} D_1\sigma_{e_1}^2 & 0 & \cdots & 0 \\ 0 & D_2\sigma_{e_2}^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & D_t\sigma_{e_t}^2 \end{bmatrix} \text{ and } V\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_t \end{bmatrix} = \begin{bmatrix} A\sigma_{s_1}^2 & A\sigma_{s_{12}} & \cdots & A\sigma_{s_{1t}} \\ A\sigma_{s_{21}} & A\sigma_{s_2}^2 & \cdots & A\sigma_{s_{2t}} \\ \vdots & \vdots & \ddots & \vdots \\ A\sigma_{s_{t1}} & A\sigma_{s_{t2}} & \cdots & A\sigma_{s_t}^2 \end{bmatrix},$$

where

D_i = diagonal matrix with elements equal to 1 over n_{ij}, with n_{ij} = number of daughters for the j^{th} sire in the i^{th} country,
σ_{ei}^2 = residual variance for country i,
A = additive relationship matrix based on sire, MGS and maternal granddam (MGD) of the bull, with MGD considered as a phantom parent group,
σ_{si}^2 = sire variance for country i, and
σ_{sij} = sire covariance between countries i and j.

Current calculation of σ_{si}^2, σ_{sij} and σ_{ei}^2 is based on EM-REML procedures (Sigurdsson et al., 1996; Sullivan, 1999).

This multiple-trait approach allows different international rankings of sires for different countries. Indeed, the best bull of a country with extensive system and warm climate is not necessary adapted to an intensive system and cold climate, and MACE takes this into account.

I.A.2.c. Improvements in MACE

Since its first description in 1993, MACE has been improved. For example, weighting factor to take into account accuracy of national sire breeding value that simply was based on number of daughters of the bull (D_i in the equations above) was not sufficient. Differences exist between sires in the distribution of progeny across herds, the number of lactation of the daughters, their stage of lactation and the number of contemporaries (Weigel, 2002). Fikse and Banos (2001) proposed several factors to consider these different aspects which have an influence on the accuracy of daughter information. Since November 2002, one of these studied weighting factors, based on effective daughter contribution (EDC) considering contemporary group size, correlation between repeated records and the reliability of the daughter dam evaluation, is used in current MACE procedure (INTERBULL, 2003).

Some other studies have been done on definition of genetic groups, or on amount of historical data to be used (time edit) (Weigel and Banos, 1997; De Jong, 2003; Fikse, 2003).

Another concern is the consideration of multiple breeding values per bull for each evaluated trait (e.g., first to third lactation for milk). Nowadays, countries provide only one national breeding value per bull, as MACE can not handle more. Schaeffer (2001) extended MACE to accommodate multiple-trait models within a country (MTMACE). This approach involves many genetic (co)variances estimations, both within and across countries.

I.B. Other alternatives for international evaluations

Weigel (2002) suggested to consider individual performance records of cows instead of sire national EBV for international evaluation. In his approach, he also proposes to group herds without regarding country borders but using management, climate and genetic variables (e.g., herd size, milk yield, temperature, percent of North American genes). This idea is based on the fact that herds can be more different within a country (example from North to South of USA) than between two neighboring countries (e.g., Belgium and Netherlands). Herds of all countries participating in international evaluation would be grouped into clusters according these descriptive variables. Genetic correlations are assumed to be unity within the clusters (Weigel and Rekaya, 2000; Zwald et al., 2003). For Holstein data from 17 countries, the number of clusters varied from 5 to 7, depending on the implementation of the cluster algorithm. Only genetic correlations between clusters needed to be estimated, which reduced the number of parameters considerably. Then a direct multiple-trait BLUP analysis, where each cluster was considered as a different trait, was computed based on individual animal performances, using a sire model (Appendix 1).

First results show that this method is feasible, for a large number of countries. Further research should be developed to identify the most appropriate variables to form clusters.

The main disadvantage of the suggestion of Weigel is that the international sire model used is much less sophisticated than the national models, which usually included the most appropriated fixed effects for the countries. Other problems are more political, because countries are maybe not ready to accept the herd clustering procedure, and to loose control of fixed effects included in the model.

As an alternative, the joint European project (PROduction Traits European Joint Evaluation, PROTEJE) presented by Canavesi et al. (2001, 2002) is also based on individual performance records, but maintains the modelling of environmental effects at the national level. It suggests to provide records adjusted for all fixed effects in the national evaluation. A simple model, including additive genetic and residual effects, could be used for international evaluation, based on these pre-adjusted records. In this approach, each country is still considered as a different trait. Both cows and bulls can receive an international evaluation.

I.C. Roles of INTERBULL, the INTERnational BULL evaluation service

I.C.1. History

INTERBULL is a non profit organization founded in 1983 by FEZ (Fédération Européenne de Zootechnie), IDF (International Dairy Federation) and ICRPMA (International Committee for Recording the Productivity of Milk Analysis, the current ICAR, International Committee for Animal Recording). FAO joined this group later. Since 1988, INTERBULL has been a permanent sub-committee of the ICAR. In 1991, the INTERBULL Centre, its operational unit, was established in Upssala, Sweden, under contract with the Swedish University of Agricultural Sciences. In 1996, the European Union (EU) appointed the INTERBULL Centre as its reference body for international evaluation of dairy cattle (INTERBULL, 2001).

The first objective of INTERBULL was to standardize methodology used in making comparisons between countries in order to reduce the political tensions caused by different comparison practices (Schaeffer, 1993). In the beginning, INTERBULL proposed minimum requirement and guidelines to compute regression coefficients in conversion formulas which was the responsibility of importing country (INTERBULL, 1990).

More recently, INTERBULL has conducted international evaluations of dairy bulls, using MACE. The first evaluation was in August 1994, for Ayrshire and Holstein bulls from Denmark, Finland, Norway and Sweden. The second evaluation was in February

1995 for five breeds (Ayrshire, Brown Swiss, Guernsey, Holstein and Jersey) and a total of 10 countries (Canada, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Sweden and US) (Banos and Sigurdsson, 1996).

I.C.2. Activities

Nowadays, INTERBULL has four major activities (www.interbull.org):

- Communication (seminars, workshops, Interbulletin, web site)
- Research and Development
- International Genetic Evaluation Service
- Technical support

The International Genetic Evaluation Service involves the computation of 4 routine evaluations and 2 test runs per year for production, conformation and udder heath traits (Tables 1 & 2). The purpose of the test runs is to test new data or models. Other traits are scheduled to be included in the future, as longevity, calving traits and female fertility. A description of the procedure for test runs and routine evaluations is in Appendix 2.

Table 1: Schedule of routine evaluations and test-runs.

Service	Schedule
Routine evaluations	February, May, August & November
Test-runs	March & September

Table 2: INTERBULL Evaluation, May 2003.

Traits	Description of traits	Number of countries	Number of breeds
Production	Fat, Protein & Milk yield	25	6
Conformation	Stature, Chest width, Body depth, Angularity, Rump angle, Rump width, Rear leg set, Rear leg rear view, Foot angle, Fore udder, Rear udder height, Udder support, Udder depth, Teat placement, Teat length, Overall conformation, Overall udder, Overall feet & legs	18	4
Udder health	Milk somatic cell & Clinical mastitis	17	5

Research and development projects conduced at the INTERBULL Centre mainly deal with parameter estimation (use of data subsets, parsimonious models to reduce the number of parameters, incorporation of prior information), on treatment of genetic groups, MTMACE for traits as fertility, and validation of national evaluations.

I.C.3. INTERBULL Centre team

The INTERBULL Centre team consists of 5 full time persons: 3 for service activities and 2 for research and development (Table 3). It is supported by the INTERBULL Secretary and 2 part time persons (for computer services and secretarial assistance).

Table 3: INTERBULL Centre team.

Service's activities	Research & Development	Secretarial & Computer services
• Ulf Emanuelson (INTERBULL Centre Director) • Thomas Mark • Jette Jakobsen	• Freddy Fikse • Hossein Jorjani	• Interbull Secretary: Jan Philipsson • Secretarial assistance: Siw Karlsson • Computer services: Dan Englund

II. Context of this study

One of the major problems of MACE is the computation of genetic covariances. The current genetic (co)variance matrix is very large. For the Holstein breed, for instance, 26 populations (country*breed) are evaluated, which involved the estimation of 325 genetic correlations. The EM-REML algorithm currently used is time consuming and prevents the simultaneous estimation of all these genetic correlations because it needs too much memory (Appendix 3). Moreover, genetic correlations are close to the border of the parameter space (all are between 0.76 and 0.96 for production traits), which can lead to poor convergence.

This difficulty is becoming prohibitive because the number of countries involved in the international evaluation increases, and many of these new countries have small population and weak genetic connections with the other countries. In the future, the replacement of MACE by MTMACE can not be considered without solving this problem of estimation of genetic covariances.

Structural models to reduce the number of parameters are one of the possible solutions to face this problem.

II.A. Structural model of Rekaya et al. (2001; 2003)

II.A.1. Definition of the structural model

In parallel with the suggestion of clustering herds across country borders, Rekaya et al. (2001; 2003) proposed to use external data information to predict genetic correlations between countries to face the problem of estimation of huge number of parameters. The main idea of their approach is that countries are more or less correlated according to their production system, which depends not only on geographical proximity, but also on

climate conditions, management practices and genetic composition of the cow population. Two different structural models were studied, where the covariance between two countries (σ_{ij}) is written as a linear function of different explanatory variables, which were measures of genetic similarity (GS), management similarity (MS) and climate similarity (CS):

$$\sigma_{ij} = \mu + a\,GS_{ij} + b\,MS_{ij} + c\,CS_{ij},$$

where

μ = intercept common to all off-diagonal element of the genetic covariance matrix,
GS_{ij} = number of daughters of common bulls used in the two countries over the total number of daughters of all bull of country i and j,
MS_{ij} = absolute value of the difference between the average milk yield in country i and j, over the sum of both averages, and
CS_{ij} = absolute value of the difference between heat indices for the month July in country i and j, over the sum of both averages (with heat indices = T-R/20, where T = average temperature and R = total rainfall in July).

II.A.2. Results

The first structural model (SM1) was based on GS and MS only, whereas the second one (SM2) also included CS. No significant differences were observed between these 2 models.

These 2 models were tested on data of 13 countries (or regions) (Rekaya et al., 2001). This procedure reduced the number of parameters to estimate, from 78 covariances with a standard multiple-trait model to 4 (or 3 depending on the structural model): the regression coefficients (a, b, c) and the intercept, μ. Results showed that the structural model was capable to explain the genetic covariance structure and gave similar estimates of genetic correlations to the unstructured model. A similar observation was made by Rekaya et al. (2003) for data from five regions of the United States. The disadvantage of this structural model is that the external information required (GS, MS and CS) is not always available and the choice of variables is arguable. The advantages of this structural model are its abilities to reduce the number of parameters and to explain correlations even when countries have poor genetic links with other countries.

II.B. Structural model of Delaunay et al. (2002)

The study of Delaunay et al. (2002) is a part of PROTEJE (section I.B.). One of the objectives of this project was to find a method to reduce rank of genetic correlations matrix.

II.B.1. Definition of the structural model

Contrary to the structural model of Rekaya et al. (2001), the model of Delaunay et al. (2002) proposed to use the countries themselves to characterize differences between countries, instead of climate, management or genetic variables. The idea was that a set of unobserved variables for each country condition the genetic correlations between countries. The countries could be represented in a space in k dimensions (k < number of countries), in which the coordinates of the countries are the unobserved characteristics. In this space, the genetic correlation between 2 countries i and j was defined by:

$$\rho_{ij} = \exp(-D_{ij}),$$

where D_{ij} = the Euclidian distance between countries i and j:

$$D_{ij} = \sqrt{\sum_k (X_{ik} - X_{jk})^2},$$

where X_{ik} = coordinate of country i for axis k, and
X_{jk} = coordinate of country j for axis k.

The covariances between two countries i and j was computed as:

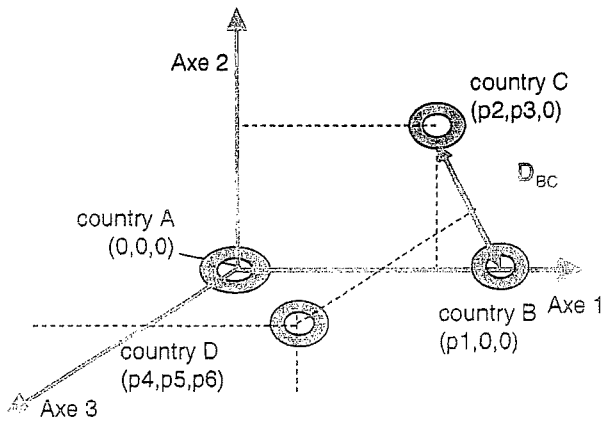
$$\sigma_{ij} = \sigma_i \sigma_j \exp(-D_{ij}),$$

where σ_i = genetic standard deviation in country i, and
σ_j = genetic standard deviation in country j.

This geometric representation can be illustrated by a simple example. Consider 4 countries A, B, C and D that define a 3-dimensional space. Country A is the center of the space, B defines the first axis, C the second axis and D the third axis. The distance between two countries determines the correlation between them. For example:

$$\rho_{BC} = \exp(-\mathbf{D_{BC}}) = \exp\left[-\sqrt{(p_2 - p_1)^2 + (p_3)^2}\right]$$

Figure 1: Representation of 4 countries in a 3-dimensional space.



In this parameterisation, the number of parameters to estimate is 6 (coordinates p_1 to p_6), used to the same number of genetic correlations (ρ_{AB}, ρ_{AC}, ρ_{AD}, ρ_{BC}, ρ_{BD}, ρ_{CD}; Figure 1). If a fifth country is added, the number of coordinates to estimate with the 3-dimensional structural model is 9 for 10 correlations (Figure 2). Table 4 shows the possible reduction of the number of parameters for different number of countries represented in a 3-dimensional space. For instance, the structural model could be used to explain 325 correlations among 36 countries, by the estimation of only 72 parameters.

Figure 2: Representation of 5 countries in a 3-dimensional space.

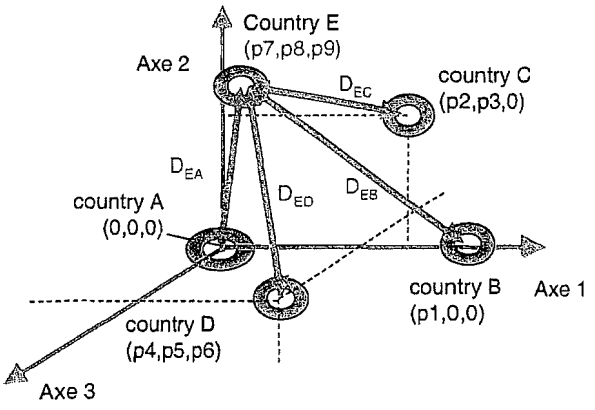
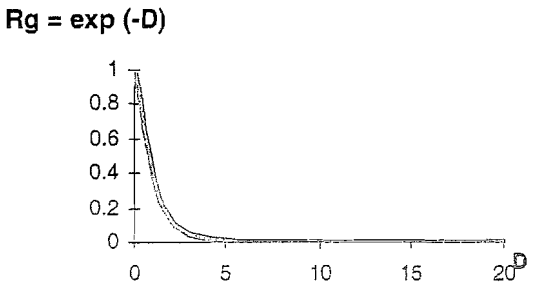


Table 4: Countries represented in a 3-dimensional space.

Number of countries	Number of parameters	Number of correlations
4	6	6
5	9	10
6	12	15
7	15	21
26	72	325

When two countries are located in the same place, the Euclidian distance between them is zero, which leads to a genetic correlation of unity. If they are far away, then distance is long and the genetic correlation is close to zero (Figure 3). Genetic correlations are always less or equal to 1. Inconvenience is that only positive correlations can be modelled. To allow for negative correlations with this structural model, observations should be multiplied by (-1).

Figure 3: Genetic correlation in relation to distance between 2 countries.



II.B.2. Results

The structural model has been tested by Delaunay et al. (2002) on simulated data. Simulated milk yields of 51200 cows from 5 generations and 4 countries were analyzed, assuming genetic correlations of either 0.90 or 0.99 between all the countries. An AI-REML algorithm (Gilmour et al., 1995), implemented in the ASREML software package (Gilmour et al., 2000) was used to estimate genetic (co)variances, either with an "unstructured" model or structured models in a 1, 2 or 3-dimensional space.

Results showed that the maximum log-likelihood value and the estimated genetic correlations were identical with the "unstructured" model and the structural model in a 3-dimensional space, where genetic correlations among countries were 0.90. It is interesting to notice that "unstructured" model failed when genetic correlations among countries were 0.99, because the correlations were too close to the border of parameter space. In this special case, the structural model gave good results. The structural model in a 2-dimensional space reduced the number of parameters to estimate from 6 to 5, and gave no results that were statistically different from the "unstructured" model. In a 1-dimensional space, the estimated genetic correlations for 0.90 were very different to the true ones, and there were less different for 0.99. The log-likelihood was not significantly lower than for a structural model with 2 or 3 dimensions (Table 5).

Table 5: Maximum log-likelihood value.

	Rg = 0.90	Rg = 0.99
Unstructured Model	-24887.7	Did not converge
Structural model – 3 dimensions	-24887.7	- 24543.5
Structural model – 2 dimensions	**-24888.2**	- 24543.7
Structural model – 1 dimension	-24914.1	**- 24544.9**

Source: Delaunay et al.. (2002)

The structural model was also tested with the genetic correlations matrix currently used by INTERBULL (Holstein breed, 26 countries, milk yield evaluation of May 2002; Delaunay et al., 2002). Genetic parameters were not estimated, but only a reparameterisation of the genetic correlation matrix with the structural model was done. They considered several numbers of axes, from 1 to 6. All the possible combinations of the countries were studied to define axes. For each combination and each dimension, the coordinates obtained with the structural model were used to compute a genetic correlation matrix, which was compared with the original one (Figure 4). Results are in Table 6.

Figure 4: Reparameterisation of the current genetic correlation matrix.

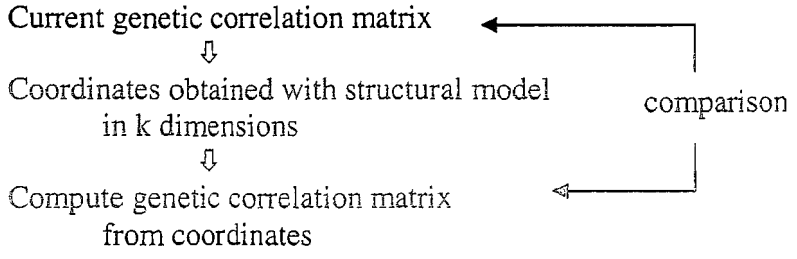


Table 6: Structural model with genetic correlation matrix used by INTERBULL.

Dim	Parameters	Best combination for axes countries	$\lvert Rg_{itb}-Rg_{str}\rvert$		
			Max	Mean	>0.03
1	25	AUS-GBR	0.13	0.04	37%
2	49	AUS-GBR -IRL	0.10	0.03	28%
3	72	NZL-ITA-CAN-IRL	0.08	0.02	21%
4	94	AUS-NDL-HUN-CSK-DEU	0.07	0.02	16%
5	115	NZL-ITA-CAN- CHE -CSK- DEU	0.07	0.01	10%
6	135	NZL-USA-NDL-DEU-CHE-CSK-HUN	0.08	0.01	10%
25	325	The 26 countries	0	0	0 %

With Rg_{itb} = genetic correlation matrix used by Interbull and Rg_{str} = genetic correlation matrix computed after parameterisation.
Source: Delaunay (personal communication)

In a structural model with 1 dimension, 37% of computed correlations differed from more than 0.03 in absolute value from the original correlations, with a maximum of 0.13. With 5 dimensions, this maximum was reduced to 0,07 and the average difference in absolute value was 0.01. Ten percent of the correlations still differed for more than 0.03 from the INTERBULL correlations: these were mainly for countries with poor links with other countries, like Slovakia, Estonia, Israel or South Africa. In this 5-dimensional space only 115 parameters needed to be estimates instead of 325 with an "unstructured" model.

Countries which defined axes that best explained the whole INTERBULL genetic correlation matrix were Australia (AUS), Great Britain (GBR), Ireland (IRL), New Zealand (NZL), Italy (ITA), Canada (CAN), The Netherlands (NLD), Hungary (HUN), Czech Republic (CSK), Germany (DEU), Switzerland (CHE) and The United-States (USA). It is important to notice that these "best combination of axes countries" were based on the comparison with the current genetic correlations matrix taken as a reference. These were not the best axes countries in an absolute sense for Holstein milk yield data, as they depended on the reliability on the estimates of the current genetic correlations.

III. Application of the structural model of Delaunay et al. (2002) on international data

The study of Delaunay et al. (2002) showed that the structural model was able to explain genetic covariances between countries with simulated data. The transformation of the genetic correlations matrix used by INTERBULL in routine evaluations with the structural model gave interesting results. The next step was to test this method on international data at INTERBULL Centre.

The first aim of this study was to compare results of the structural model with those obtained with an "unstructured" model, using AI-REML. This comparison was made for cases involving both well and poor-connected countries.

The second aim was to study the possible use of the coordinates to calculate directly genetic correlations, without having to estimate them. For example the country A (not an axis country, is in a k-dimensional space, and its coordinates are estimated with the

structural model. Another country B (not an axis country) is in the same k-dimensional space, and its coordinates are also estimated. By construction, the correlation between A and B could be obtained, from the coordinates of A and B relative to the same space.

The last objective of this work was to analyze how the choice of countries that define the axes of the space could influence the results.

III.A. Materials and Methods

III.A.1. Data

Data available were deregressed national breeding values of bulls and their EDC used for Holstein milk yield international genetic evaluation of February 2003 (26 countries*breeds). Data were edited to include only national evaluations for bulls born after 1984. All the observations were included in estimation of the genetic correlations, contrary to the current INTERBULL practice of creating subsets of well-connected bulls.

Different subsets of the11 following countries have been analyzed: Canada (CAN), Germany (DEU), Denmark (DNK), Finland (FIN), France (FRA), The Netherlands (NLD), The United States (USA), New Zealand (NZL), Australia (AUS), Estonia (EST) and Hungary (HUN). Number of bulls and common bulls are presented in Table 7, and a description of the different subsets of countries is in Table 8.

Table 7: Number of bulls with records in the country (on the diagonal) and number of common bulls (above the diagonal).

	CAN	DEU	DNK	FIN	FRA	NLD	USA	NZL	AUS	EST	HUN
CAN	5264	246	90	11	201	263	726	329	339	17	196
DEU		11197	134	23	306	556	467	200	197	31	178
DNK			4538	15	82	135	110	85	77	8	91
FIN				607	26	21	19	15	10	0	14
FRA					8037	253	420	146	180	8	120
NLD						6461	622	357	257	22	172
USA							17458	431	484	19	289
NZL								2997	412	8	147
AUS									3544	6	110
EST										265	10
HUN											1276

Table 8: Number of observations, bulls with records and genetic groups for each subset of countries analyzed.

Subset of countries	Number of observations	Number of bulls	Number of genetic groups
NLD-USA-NZL-HUN-DNK	32730	31009	33
NLD-USA-NZL-HUN-CAN-EST	33721	31208	32
NLD-USA-NZL-HUN-FRA-AUS	39773	36813	39
NLD-USA-NZL-HUN-DNK-FIN	33337	31586	34
NLD-USA-NZL-HUN-DEU-FRA	47426	44363	46
NLD-USA-NZL-HUN-DEU-CAN-EST	44918	41542	46
NLD-USA-NZL-HUN-FRA-AUS-DNK-FIN	44918	41760	47
NLD-USA-NZL-HUN-FRA-AUS-CAN-EST	45302	41263	44
NLD-USA-AUS-HUN-CAN-EST	34268	31804	33

III.A.2. Models

III.A.2.a. Structural model (SM)

The sire model described in Chapter I (section I.A.2.b.) was applied. The residual variance of the model considered the effective daughter contributions (EDC) as explained in section I.A.2.c. Genetic groups, considered as fixed effects, were based on selection path*year of birth*origin. Small groups were merged together first by origin, then by year of birth. Minimum group size was 500 bulls. Number of genetic groups formed for each subset of countries is in Table 8.

The structural model of Delaunay et al. (2002) described in Chapter II (section II.B.) was used. Choice of countries to define axes was based on results of Weigel and Zwald (2002) (Table 9). One country was selected in each cluster: The Netherlands defined the origin of the space, USA the first axis, New Zealand the second axis and Hungary the third axis, unless mentioned otherwise. These countries appeared also in the best combinations of axes countries determined by Delaunay (Table 6).

Table 9: Clusters of Weigel and Zwald (2002).

	Characterization	Herds from
Cluster 1	High average milk yield	Australia, Canada, Italy, **USA**
Cluster 2	Large herd size	Cesk Republic, Germany, **Hungary**, Italy, New Zealand, USA
Cluster 3	Low peak milk yield, low percentage of North American Holstein genes and low days to peak yield	Australia, Czech Republic, Germany, **New Zealand**
Cluster 4	Small herds with a high percentage of North American Holstein genes	Canada, Germany, **The Netherlands**

III.A.2.b. Classical Model (CM)

To compare results, an "unstructured" model, called classical model hereafter, was applied. To consider the EDC of each observation, it used the following linear model:

$$D_i^{-1/2}\, y_i = D_i^{-1/2}\, \mu_i\, 1 + D_i^{-1/2}\, Z_i\, (Qg_i + s_i) + D_i^{-1/2} e_i,$$

where D_i is a diagonal matrix with elements equal to 1 over the EDC_{ij}, effective daughter contribution of the j^{th} sire in the i^{th} country and y_i, μ_i, g_i, s_i, e_i, Z_i and Q are the same as in the structural model.

This model was obtained after pre-multiplying the left and the right side of MACE by $D_i^{-1/2}$. This was done to consider the weighing factor, as the program used to estimate variance components for the classical model could not handle weights in the R matrix. For more details, cf. Appendix 4 (F. Fikse).

Parameters to estimate

The total number of parameters for each model is given in Table 10. For both models, the residual variances and the genetic variances within countries need to be estimated. In the classical model the genetic covariances across countries are estimated. In the structural model the coordinates for each axis are estimated for each country.

Table 10: Number of parameters to estimate.

Classical model	Structural model
$2count + \frac{count(count-1)}{2}$	$2count + \frac{axes(axes-1)}{2} + (count - axes)axes$

count: No. countries; axes: No. axes

III.A.3. Algorithm

An AI-REML algorithm was used to estimate genetic (co)variances for the classical and the structural model.

The classical model used the program AIREMLF90 (Druet et al., 2003a), whereas the structural model used a modified version of the program AIREMLF90 that could weigh each observation by EDC (Druet T., personal communication).

The genetic (co)variance matrix contains genetic variances and covariances. For the structural model, the covariances were a nonlinear function of the coordinates. In this case, the AI-REML algorithm used a simplified average information matrix, ignoring non-zero terms of the second derivative of the genetic (co)variance matrix. (Gilmour et al., 1995).

Implementation of the AI-REML algorithm allowed for user-specified covariance structure for random effects (Druet et al., 2003b). This is achieved by defining a function "own". If the covariances are a function of the variables X, then the function "own" computes, from the genetic variances and these variables, the genetic (co)variance matrix and the first derivatives of the genetic (co)variances matrix with respect to the genetic variances and the variables X. Such an "own" function, where the variables X were the coordinates, has been programmed at the INTERBULL Centre for the structural model, for any number of countries and axes.

Asymptotic standard errors of genetic correlations were approximated from asymptotic standard errors of the parameters obtained as the inverse of the AI matrix. This procedure was based on Taylor Series expansion (Appendix 5, F. Fikse).

For the classical model, the AI-REML algorithm was based on:

$$P^{(t+1)} = P^{(t)} + \Delta ,$$

where $P^{(t+1)}$ were the new parameters at iteration (t+1), $P^{(t)}$ were the parameters at iteration t and Δ was the change between 2 iterations. The change Δ depended on the inverse of the information matrix $M^{(t)}$, and the first derivative of the log-likelihood with respect to the parameters P, called the gradient $d^{(t)}$. Each was evaluated at iteration t (Hofer, 1998).

$$\Delta = M^{(t)} d^{(t)} .$$

If the AI-REML update yielded a genetic covariance matrix that was not positive definite, a new update was computed that combined the AI with an EM update weighted such that the new parameters were in the parameter space (Jensen et al., 1997)

For the structural model, the algorithm was based on:

$$P^{(t+1)} = P^{(t)} + \gamma \Delta ,$$

where γ is the step size.

The algorithm started with $\gamma = 1$. Two situations were possible.
① -2logL computed from $(P^{(t)} + \gamma \Delta)$ decreased. Then new parameters $P^{(t+1)}$ were retained because they were better than previous ones $P^{(t)}$.
② -2logL increased. $P^{(t+1)}$ were not retained and the program continued with dividing γ repeatedly by 2 until -2logL decreased. Then parameters corresponding to the lower -2logL were kept as the new ones.
Next step was the computation of a new Δ from the parameters obtained from case ① or ②. A new cycle started with $\gamma = 1$ as described above.

III.A.4. Model comparison

Results of the estimation of parameters are presented as genetic correlations:

$$Rg_{ij} = \rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \ \sigma_j},$$

where ρ_{ij} is the genetic correlation between countries i and j (denoted Rg in the following), σ_{ij} is the genetic covariance between countries i and j, σ_i is the genetic standard deviation in country i.

Models were compared using:
- Differences in estimated genetic correlations (Rg).
- Minus two times the logarithm of the likelihood function: -2logL. For the classical model, -2logL was transformed to allow comparison with the structural model, as explained in Appendix 4 (F. Fikse).
- Two information criteria which take into account the number of parameters to estimate: Akaike's Information Criterion (AIC) and Schwarz' Bayesian Information Criterion (BIC). BIC puts more weight on number of parameter than AIC; it prefers parsimony (Wolfinger, 1993). Lower AIC or BIC values correspond to the better model.

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

$$BIC = -2\log L + q \log(n-p) \text{ ,}$$

where q is the number of parameters, n is the number of observations, p is the rank of fixed effects matrix computed as p = (No. genetic groups + 1)* No. countries. Number of observations and number of genetic groups are in Table 7.

III.B. Results and discussion

III.B.1. Comparison of structural and classical models (SM and CM)

III.B.1.a. Illustration with a 5 countries subset, including Denmark

Genetic correlations between 5 countries (the 4 axes countries and Denmark) have been estimated with 3 different models: the classical model (CM5), a structural model with 4 dimensions (SM45) and a parsimonious structural model, with 3 dimensions (SM35).

Table 11: Estimated genetic correlations from SM35 and their standard errors (above diagonal), deviations (SM35-CM5) in estimated genetic correlations and deviations in standard errors (below diagonal).

	NLD	USA	NZL	HUN	DNK
NLD		0.927	0.779	0.857	0.955
		(0.006)	*(0.016)*	*(0.015)*	*(0.007)*
USA	0.002		0.724	0.886	0.968
	-0.001		*(0.012)*	*(0.020)*	*(0.011)*
NZL	-0.005	0.017		0.685	**0.747**
	0.011	*0.006*		*(0.026)*	*(0.019)*
HUN	-0.001	-0.001	-0.005		0.872
	-0.004	*0.004*	*0.008*		*(0.020)*
DNK	-0.001	-0.006	**-0.056**	-0.006	
	-0.010	*-0.005*	*0.007*	*-0.007*	

In bold: more than 0.02 of difference in Rg

Table 12: Number of parameters and iterations, -2logL and information criteria for CM5, SM45 and SM35.

	CM5	SM45	SM35
No. parameters	20	20	19
No. iterations	4	2	34
-2log L	**424709.3**	424717.0	424717.0
AIC	**424749.3**	424757.0	424755.0
BIC	424917.1	424924.8	**424914.4**

Table 13: Coordinates of the countries estimated with SM35 and SM45, and their standard errors.

	SM35	*SE*	SM45	*SE*
NLD 1	0		0	
NLD 2	0		0	
NLD 3	0		0	
NLD 4			0	
USA 1	0.076	*(0.006)*	0.076	*(0.006)*
USA 2	0		0	
USA 3	0		0	
USA 4			0	
NZL 1	-0.240	*(0.081)*	-0.240	*(0.085)*
NZL 2	0.072	*(0.279)*	0.072	*(0.303)*
NZL 3	0		0	
NZL 4			0	
HUN 1	0.099	*(0.030)*	0.099	*(0.030)*
HUN 2	-0.065	*(0.285)*	-0.065	*(0.302)*
HUN 3	0.099	*(0.181)*	0.099	*(0.196)*
HUN 4			0	
DNK 1	0.045	*(0.006)*	0.045	*(0.006)*
DNK 2	0.009	*(0.026)*	0.009	*(0.116)*
DNK 3	-0.002	*(0.040)*	-0.002	*(0.092)*
DNK 4			-5.78E-05	*(1.732)*

Genetic correlations estimated from the classical model and the structural models varied from 0.685 to 0.974, and standard errors ranged from 0.006 to 0.027 (Table 11). The coordinates varied from -0.24 to +0.099 (Table 13). To understand how to compute the genetic correlations from the coordinates, let's illustrate it with Hungary and Denmark:

$$Rg_{HUN-DNK} = \exp(-D_{HUN-DNK}) = \exp\left(-\sqrt{(-0.002 - 0.099)^2 + (0.009 + 0.065)^2 + (0.045 - 0.099)^2}\right)$$
$$= \exp(-0.136) = 0.872.$$

SM45 and SM35 gave the same estimates of the genetic correlations and coordinates, and had the same -2logL (Tables 12 & 13). Information criteria were lower for SM35 which had one parameter less. The additional coordinate in SM45 was close to zero, which meant the 4th axis was not useful to estimate the genetic correlations. On the contrary, it led to higher standard errors because the same amount of information is used to estimate more parameters.

Genetic correlations estimated by SM35 were very close to those estimated with CM5, except the one between Denmark and New Zealand that differed for almost 0.06. This correlation was based on the lowest number of common bulls in this subset of countries (Table 7), which could explain why it was less stable than the others. Higher -2logL observed for SM35 was compensated by the reduction of parameters as shown by lower BIC criteria.

CM5 and SM45 gave different estimates of genetic correlations, although SM45 had the same number of parameters. CM5 has a better fit than SM45 as indicated by the lower -2logL. One explanation could be that SM45 imposed more constraints on the genetic correlations. Delaunay et al. (2002) had already mentioned this problem for a spatial representation of the correlations. For example, 3 countries A, B and C could be represented in a 2-dimensional space. Correlations between A and B could be transformed into a distance D_{AB}, using the definition of the correlation in the structural model. Similarly, the other correlations determine D_{AC} and D_{BC}. But if the sum of D_{AC} and D_{BC} is less than D_{AB}, then a spatial representation of these genetic correlations is impossible. The triangle (A,B,C) can not be formed. This was observed for the genetic correlations estimated with CM5: they could not be converted into coordinates in a 4-dimensional space.

III.B.1.b. Effect of connectness: illustration with a 6 countries subset including Canada and Estonia.

Two countries which differed in the amount of genetic ties were added to the four axes countries. Canada was chosen as a well connected country (number of common bulls ranged from 196 to 726 with the axes countries; Table 7), and Estonia as a poor connected one (number of common bulls vary from 8 to 22; Table 7). Canada and Estonia had 17 bulls in common (Table 7). The structural model and the classical model were used to estimate the genetic correlations.

Table 14: Estimated genetic correlations from SM36 and their standard errors (above diagonal), deviations (SM36-CM6) in estimated genetic correlations and deviations in standard errors (below diagonal).

	NLD	USA	NZL	HUN	CAN	EST
NLD		0.923	0.768	0.847	0.943	0.798
		(0,006)	*(0,016)*	*(0,016)*	*(0,006)*	*(0,042)*
USA	-0.001		0.715	0.883	0.959	0.841
	0.000		*(0,012)*	*(0,022)*	*(0,008)*	*(0,058)*
NZL	-0.012	0.008		0.680	**0.724**	**0.620**
	0.000	*-0.005*		*(0,020)*	*(0,015)*	*(0,020)*
HUN	0.002	0.000	0.006		0.872	**0.779**
	0.000	*0.010*	*-0.009*		*(0,026)*	*(0,053)*
CAN	-0.002	0.000	**0.024**	0.000		0.839
	0.000	*0.005*	*-0.005*	*0.011*		*(0,061)*
EST	0.018	-0.006	**0.098**	**0.029**	-0.015	
	-0.006	*0.013*	*-0.063*	*-0.001*	*0.016*	

In bold: more than 0.02 of difference in Rg

Table 15: Number of parameters and iterations, -2logL and information criteria for CM6 and SM36.

	CM6	SM36
nb parameters	27	24
nb iterations	7	22
-2log L	**440581.1**	440584.8
AIC	440635.1	**440632.8**
BIC	440862.4	**440834.9**

The approximated standard errors for SM36 were higher for Estonia than for Canada, which can be directly related to the number of common bulls (Tables 7 & 14).

Genetic correlations estimated with SM36 were not very different than those estimated with CM except the correlation between Estonia and New Zealand that was almost 0.10 higher with SM36. The same observation was made for the correlation between New Zealand and Finland that had also poor links with the other countries (Appendix 6). Similar -2logL were observed for CM6 and SM36, but information criteria were better for SM36, because of its parsimony (Table 15).

III.B.2. Use of the estimated coordinates to calculate correlations

Three combinations of 6 countries have been analyzed with the structural model (SM36). Pairs of countries that were added to the 4 axes countries were: France-Australia, Canada-Estonia and Denmark-Finland. Coordinates obtained with SM36 are in Table 16.

Table 16: Estimated coordinates and their standard errors for the 3 combinations with SM36.

	SM36 FRAAUS		SM36 DNKFIN		SM36 CANEST	
	Coordinates	*SE*	Coordinates	*SE*	Coordinates	*SE*
NLD1	0		0		0	
NLD2	0		0		0	
NLD3	0		0		0	
USA1	0.077	*(0.007)*	0.075	*(0.006)*	0.080	*(0.007)*
USA2	0		0		0	
USA3	0		0		0	
NZL1	-0.222	*(0.038)*	-0.222	*(0.074)*	-0.225	*(0.072)*
NZL2	0.125	*(0.075)*	0.118	*(0.147)*	0.138	*(0.123)*
NZL3	0		0		0	
HUN1	0.103	*(0.030)*	0.096	*(0.030)*	0.116	*(0.032)*
HUN2	0.054	*(0.112)*	-0.053	*(0.118)*	0.005	*(0.114)*
HUN3	0.103	*(0.064)*	0.108	*(0.055)*	0.119	*(0.019)*
FRA1	0.010	*(0.011)*				
FRA2	0.068	*(0.020)*				
FRA3	0.017	*(0.071)*				
AUS1	-0.113	*(0.043)*				
AUS2	0.148	*(0.043)*				
AUS3	-0.002	*(0.112)*				
DNK1			*0.046*	*(0.007)*		
DNK2			*0.010*	*(0.022)*		
DNK3			*-0.007*	*(0.023)*		
FIN1			*0.083*	*(0.059)*		
FIN2			*-0.095*	*(0.075)*		
FIN3			*-0.053*	*(0.113)*		
CAN1					*0.051*	*(0.006)*
CAN2					*-0.029*	*(0.007)*
CAN3					*0.004*	*(0.033)*
EST1					*0.169*	*(0.095)*
EST2					*-0.116*	*(0.172)*
EST3					*-0.093*	*(0.147)*

Coordinates of the axes countries appeared stable across combinations; differences were lower than the corresponding standard errors. The second axis defined by New Zealand had higher standard errors (from 0,075 to 0,147) than other axes countries, which was a little problematic and would have to be considered. France, Australia, Denmark and Canada were in the plane defined by the Netherlands, USA and New Zealand. Their coordinates for the 3rd axis were close to zero. For these countries, Hungary was not interesting as axis country. On the other hand, Finland and Estonia were not in the plane

NLD-USA-NZL, and moved away from it in the direction of the axis defined by Hungary. Supposing that the third axis was not correctly chosen as axes country, the positions of Finland and Estonia "imposed" by the third axis were not the optimum ones, which could have consequences for the genetic correlations.

The coordinates of France - Australia, Denmark - Finland and Canada - Estonia were relative to the same space defined by the 4 axes countries. That allowed to place France, Australia, Denmark and Finland in a common space as shown in Figure 5. This common space was defined as the average coordinates of the Netherlands, USA, New Zealand and Hungary obtained from SM36FRAAUS and SM36DNKFIN, as shown in Table 17 (CALC). Distances between countries in this common space was used to compute genetic correlations (CALC). Thus, correlations between France-Denmark, France-Finland, Australia-Denmark and Australia-Finland were obtained without having to estimate them.

Similarly, France, Australia, Canada and Estonia could be combined in a common space.

In both cases, genetic correlations, -2logL and information criteria were compared between CALC, SM38 and CM8 (Tables 18 &19). In addition, coordinates were compared between CALC and SM38 in the case of France, Australia, Denmark and Finland (Table 17). For the other example, coordinates are in Appendix 7.

Figure 5: Combination (CALC) of the coordinates obtained from SM36FRAAUS and SM36DNKFIN.

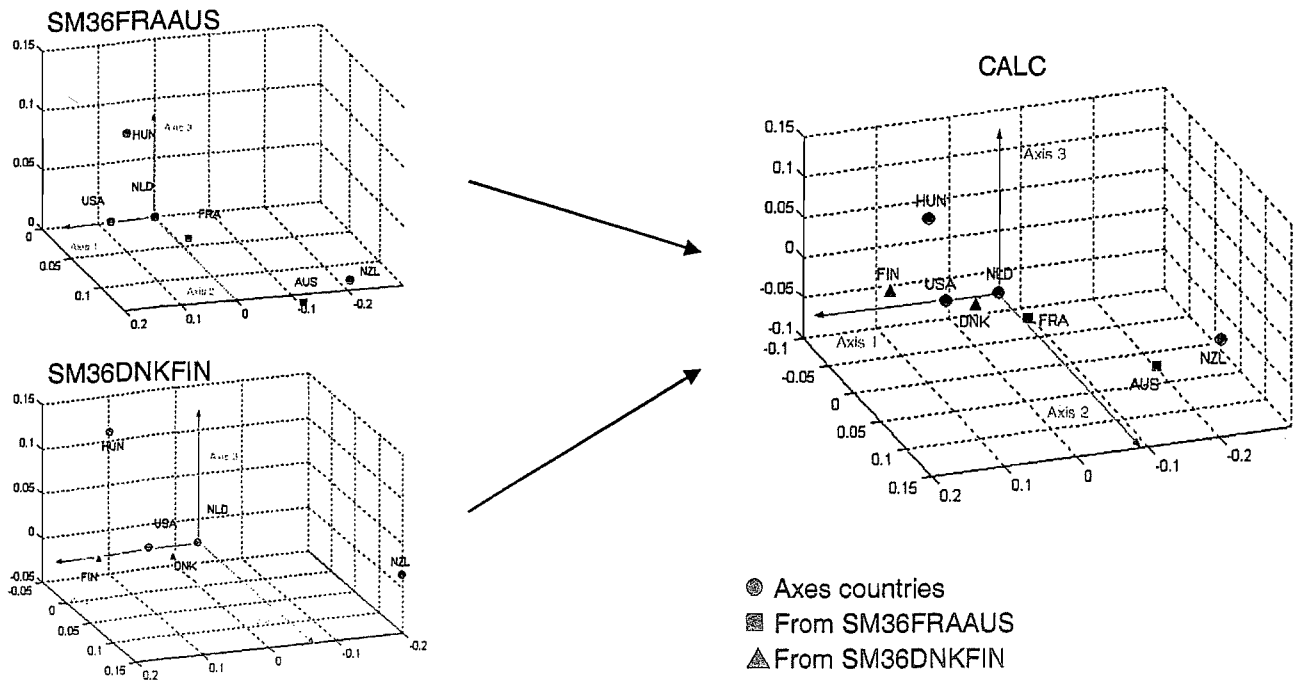


Table 17: Coordinates and standard errors obtained from combination of SM36FRAAUS and SM36DNKFIN (CALC) and those estimated with SM38.

	CALC	*SE*		SM38	*SE*
NLD1	0			0	
NLD2	0			0	
NLD3	0			0	
USA1	0.076			0.075	*0.006*
USA2	0			0	
USA3	0			0	
NZL1	-0.222		Average of SM36FRAAUS and SM36DNKFIN	-0.182	*0.069*
NZL2	0.121			0.175	*0.082*
NZL3	0			0	
HUN1	0.099			0.094	*0.029*
HUN2	0.001			-0.017	*0.055*
HUN3	0.106			0.116	*0.014*
FRA1	0.010	*0.011*		0.012	*0.011*
FRA2	0.068	*0.020*		0.052	*0.021*
FRA3	**0.017**	*0.071*	From SM36FRAAUS	**0.046**	*0.023*
AUS1	-0.113	*0.043*		-0.113	*0.039*
AUS2	0.148	*0.043*		0.128	*0.047*
AUS3	**-0.002**	*0.112*		**0.072**	*0.021*
DNK1	0.046	*0.007*		0.044	*0.007*
DNK2	0.010	*0.022*		0.013	*0.011*
DNK3	-0.007	*0.023*	From SM36DNKFIN	-0.005	*0.014*
FIN1	0.083	*0.059*		0.068	*0.054*
FIN2	-0.095	*0.075*		-0.099	*0.044*
FIN3	-0.053	*0.113*		-0.039	*0.077*

Table 18: Genetic correlations calculated from SM36 (CALC) or estimated with SM38 and CM8 (and standard errors) in the case "FRA-AUS-DNK-FIN" and in the case "FRA-AUS-CAN-EST".

		Case "FRA – AUS – DNK – FIN"		Case "FRA – AUS – CAN – EST	
		DNK	FIN	CAN	EST
FRA	CALC	0.930	0.825	0.899	0.766
	SM38	0.930 (*0.012*)	0.833 (*0.024*)	0.927 (*0.014*)	0.805 (*0.048*)
	CM8	0.941 (*0.003*)	0.798 (*0.009*)	0.930 (*0.007*)	0.862 (*0.045*)
AUS	CALC	0.810	0.729	0.786	0.672
	SM38	0.812 (*0.013*)	0.733 (*0.025*)	0.805 (*0.011*)	0.712 (*0.026*)
	CM8	0.823 (*0.007*)	0.643 (*0.013*)	0.780 (*0.014*)	0.639 (*0.068*)

Table 19: -2logL and information criteria for CALC, SM38 and CM8 in the case "FRA-AUS-DNK-FIN" and in the case "FRA-AUS-CAN-EST".

	Case "FRA – AUS – DNK – FIN"			Case "FRA – AUS – CAN – EST"		
	CALC	SM38	CM8	CALC	SM38	CM8
-2logL	572423.7a 572422.7b	572419.9	**572389.0**	580977.0a 580972.8b	580962.5	**580931.0**
AIC	-	572487.9	**572477.0**	-	581030.5	**581019.0**
BIC	-	**572783.8**	572859.9	-	**581326.7**	581402.4

a. calculated with residual and sire variances obtained from the two SM36
b. calculated with residual and sire variances obtained from SM38

Genetic correlations calculated (CALC) and estimated with SM38 were nearly the same for the case "FRA-AUS-DNK-FIN". In the other case, they differed by at most 0.039, but this difference was smaller than the standard error (0.048; Table 18). Differences between CALC and SM38 could be explained by the fact that the 8 countries were analyzed jointly in SM38. Additional information could be utilized, compared with the separate analyses of these countries with SM36. The number of ancestors was larger, which could create additional genetic links between countries. The higher differences between CALC and SM38 in the case "FRA-AUS-CAN-EST" could be explained by the strong links that could be created by Canada, a very well-connected country. Estimates of the coordinates were more precise with SM38 as shown by the difference in standard errors in Table 17.

Two options were considered for the computation of -2logL for CALC. Firstly, the parameters used were coordinates and residual and sire variances obtained from the two SM36 analyses. For the axes countries, for which there was a double set of estimates from the two SM36 runs, the average was considered (CALCa). Secondly, -2logL for CALC was calculated by using the same coordinates as before, but by considering the residual and sire variances obtained from SM38 (CALCb). This -2logL was lower than CALCa. For both country combinations considered, the -2logL computed for CALC were a little higher than with SM38 (Table 19).

The difference between CALCa and CALCb could be attributed to the different residual and sire variances. CALCb and SM38 had the same residual and sire variances but were still different due to the differences in the genetic correlations observed in Table 18.

Comparison of SM38 and CM8 results agreed with previous ones, in part 1: genetic correlations of poor connected countries (Finland and Estonia) varied the most; reduction of parameters from 44 with CM8 to 34 compensated the higher -2logL obtained with SM38, as shown by the lower BIC.

III.B.3.Influence of choice of the axes countries

III.B.3.a. Effect of replacing the Netherlands by Germany

Two subsets of countries were analyzed. For each of them, 2 different combinations of axes countries were used in the structural model: one with the Netherlands, USA, New Zealand and Hungary, and the other one replacing the Netherlands by Germany. Results are in Tables 20 and 21. Genetic correlations, -2logL and coordinates were compared. Volume defined by the axes countries were computed as explained in Figure 6.

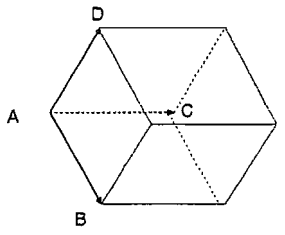
Table 20: Replacing the Netherlands by Germany in the axes countries.

Axes countries	Added countries	No. countries	-2logL	Deviation in Rg
NLD - USA - NZL - HUN	DEU - FRA	6	607687.2	0
DEU - USA - NZL - HUN	NLD - FRA	6	607687.2	0
NLD - USA - NZL - HUN	DEU - CAN - EST	7	577019.0	0
DEU - USA - NZL - HUN	NLD - CAN - EST	7	577019.0	0

Table 21: Coordinates and volume defined by the axes countries, for the combination of 7 countries.

	DEU as center	*SE*	NLD as center	*SE*
DEU or NLD 1	0		0	
DEU or NLD 2	0		0	
DEU or NLD 3	0		0	
USA 1	0.125	*0.008*	0.080	*0.006*
USA 2	0		0	
USA 3	0		0	
NZL 1	0.193	*0.084*	-0.185	*0.064*
NZL 2	0.329	*0.035*	0.208	*0.063*
NZL 3	0		0	
HUN 1	0.144	*0.027*	0.104	*0.029*
HUN 2	-0.017	*0.043*	0.010	*0.070*
HUN 3	0.115	*0.016*	0.115	*0.017*
NLD or DEU 1	0.090	*0.009*	0.024	*0.015*
NLD or DEU 2	0.072	*0.007*	-0.112	*0.007*
NLD or DEU 3	0.000	*0.031*	0.000	*0.076*
CAN1	0.080	*0.007*	0.050	*0.007*
CAN2	0.011	*0.007*	-0.036	*0.005*
CAN3	0.000	*0.015*	0.000	*0.027*
EST1	0.077	*0.073*	0.176	*0.066*
EST2	-0.131	*0.074*	-0.102	*0.095*
EST3	-0.082	*0.096*	-0.082	*0.102*
Mean SE		***0.035***		***0.043***
Volume	4.7602E-03		1.9131E-03	

Figure 6: Computation of the volume of the parallelepiped.



Coordinates:
A (x_A ; y_A ; z_A)
B (x_B ; y_B ; z_B)
C (x_C ; y_C ; z_C)
D (x_D ; y_D ; z_D)

Volume of parallelepiped defined by A, B, C, D:

$$V = | \det (\overrightarrow{AB}, \overrightarrow{AC}, \overrightarrow{AD}) |$$

$$V = | \det \begin{pmatrix} (x_B - x_A) & (x_C - x_A) & (x_D - x_A) \\ (y_B - y_A) & (y_C - y_A) & (y_D - y_A) \\ (z_B - z_A) & (z_C - z_A) & (z_D - z_A) \end{pmatrix}$$

Source: http://c.caignaert.free.fr/resume/node34.html

For both sets of countries, -2log likelihood and estimated genetic correlations were independent from whether Germany or the Netherlands defined the origin. The only difference was that space defined with Germany was more voluminous (4.7602E-03 cubic unit) than space defined with the Netherlands (1.9131E-03 cubic unit), and the standard errors were in average lower with Germany, as shown at the end of Table 21.

III.B.3.b. Effects of replacing New Zealand by Australia as axes country

Genetic correlations between The Netherlands, USA, Hungary, Canada and Estonia were compared when they were estimated with 2 different combinations of axes countries: Netherlands-USA-New Zealand-Hungary (SM36NZL) and Netherlands-USA-Australia-Hungary (SM36AUS) (Table 22).

Table 22: Deviations SM36NZL-SM36AUS for the genetic correlations and the standard errors.

	USA	HUN	CAN	EST
NLD	-0.003	0.000	-0.003	-0.004
	0.000	*0.000*	*0.000*	*0.000*
USA		0.000	0.000	-0.003
		0.000	*0.001*	*0.000*
HUN			0.000	-0.001
			-0.018	*-0.021*
CAN				0.000
				-0.002

As in the previous example, replacing New Zealand by Australia had small influence on the estimates of genetic correlations. The volume defined by axes countries including New Zealand was larger (1.32E-03 cubic unit) than with Australia (3.74E-04 cubic unit). Standard errors were equal or lower with New Zealand as axes-country.

III.B.4.General discussion

Selection of countries to define the axes, and number of dimension for the structural model are an important issue. Information criteria could be used to compare structural models with different combination of countries and number of axes, and select the best one. The main question is to know how much the differences between genetic correlations estimated with the structural model and with the classical model are acceptable compared to the benefit of reducing the number of parameters.

If the information criteria give the same results, the structural model with axes countries defining the most voluminous space seemed to be more interesting by reducing the standard errors. The further away the axes countries are, the more stable the genetic correlations are. Indeed, a small change of the distance affects the genetic correlations only slightly when the distance between two countries is large and an exponential function is used (Figure 3). In addition to volume, another criterion should be defined to

guarantee that the volume is large because all the countries are far away, and not because one country has extreme coordinates. This criterion could involve for example distances or areas.

For the current study, the total number of deregressed national breeding value was used to compute BIC. The reliability of the breeding value was not considered to weigh the number of observations. It would be interesting to verify if the BIC are also the lower with the structural model when the reliability is considered to compute BIC.

It seems reasonable to include at least one well-connected country among the axes countries, like the USA. Poor connected countries are not stable enough in their correlations to be chosen as axes countries.

The structural model might be more advantageous for poor connected countries than the classical model, because it uses indirect information. For example, 2 countries could have no links between each other but could be connected with some of the axes countries. These links with the axes countries would make it possible to determine the coordinates of these 2 unconnected countries in the same space, using the structural model. Then, the distance between the 2 unconnected countries could be calculated to determine their genetic correlation.

This study shows that the structural model would allow a decrease of the number of runs to determine the matrix of the genetic correlations between the countries. The coordinates of different sets of countries analyzed by a structural model with the same axes countries could be combined. Thus, if a new country wants to participate to INTERBULL evaluation, only correlations with the axes countries need to be estimated. This country would be placed in the same space as the other ones, and all the correlations could be computed from the distances between countries. This method would avoid a direct time-consuming estimation of all the genetic correlations with the current 26 breed*countries. Before to use it, it would be necessary to ensure that the new participating country is correctly represented in the space defined by the axes countries.

Most of the correlations estimated with SM or CM were similar or lower than those used by INTERBULL in international evaluation of February 2003. Differences could partly be attributed to difference in data selection to estimate genetic correlations. Moreover, when links are poor between countries, INTERBULL uses other procedures to determine the correlations, as explained in Appendix 3. Other difference is that INTERBULL uses the heritabilities provided by the countries to estimate the genetic correlations: the residual variances are not estimated contrary to the current study.

The residual and the sire variances estimated with the structural model and the classical model were in most cases similar, with differences less than 2%. However, in some cases, residual variances estimated with SM and CM differed by 5 to 10 %. In the classical and structural models, the estimates of residual and sire variances can be slightly different (from 1 to 10%) depending on the particular combination of countries included in the subset.

One of the advantages of the AI-REML algorithm is the asymptotic approximation of the standard errors of the estimates. Estimation of the parameters with the structural model needed 0.6 Gb of memory and took some hours for 5 countries, to 2.1 Gb and a half day for 8 countries. Current estimation of the correlations needs the same memory capacity, but takes longer time. The number of iterations with the structural model was usually higher than with the classical model. Some improvements of the algorithm can be done with respect to the rules of convergence and the way to change the step size. In the current algorithm the step size decreased by a factor two. Other more sophistical changes could be implemented for the step size. Such changes would allow reducing the time to estimate the correlations.

Finally, the structural model could be tested on other traits, like conformation traits which are less correlated than production traits. Best combinations of countries to define axes would not be necessarily the same than for milk yield.

Conclusion

The structural model of Delaunay et al. (2002) applied on Holstein milk yield international data and using an AI-REML was able to explain genetic covariance between countries. These examples show that reduction of the number of parameters is possible with the structural model. This reduction compensated the constraints of the spatial representation. These results are promising and much more drastic reduction of the number of parameters could be planed in the future. Moreover, the use of the coordinates of the countries to calculate the genetic correlations could reduce the number of runs (thus the time) needed to determine the all genetic (co)variance matrix.

Other advantages of the structural model are that it allowed estimations of correlations close to the border space. Also, countries themselves were used as explanatory variables, and no extra information is needed contrary to Rekaya's structural model.

Determination of the best axes countries and the optimal number of axes needs to be investigate further, using information criteria and regarding the volume of the space defined by the axes countries. An efficient procedure to select the best axes countries with the optimal number of axes countries should be defined and tested.

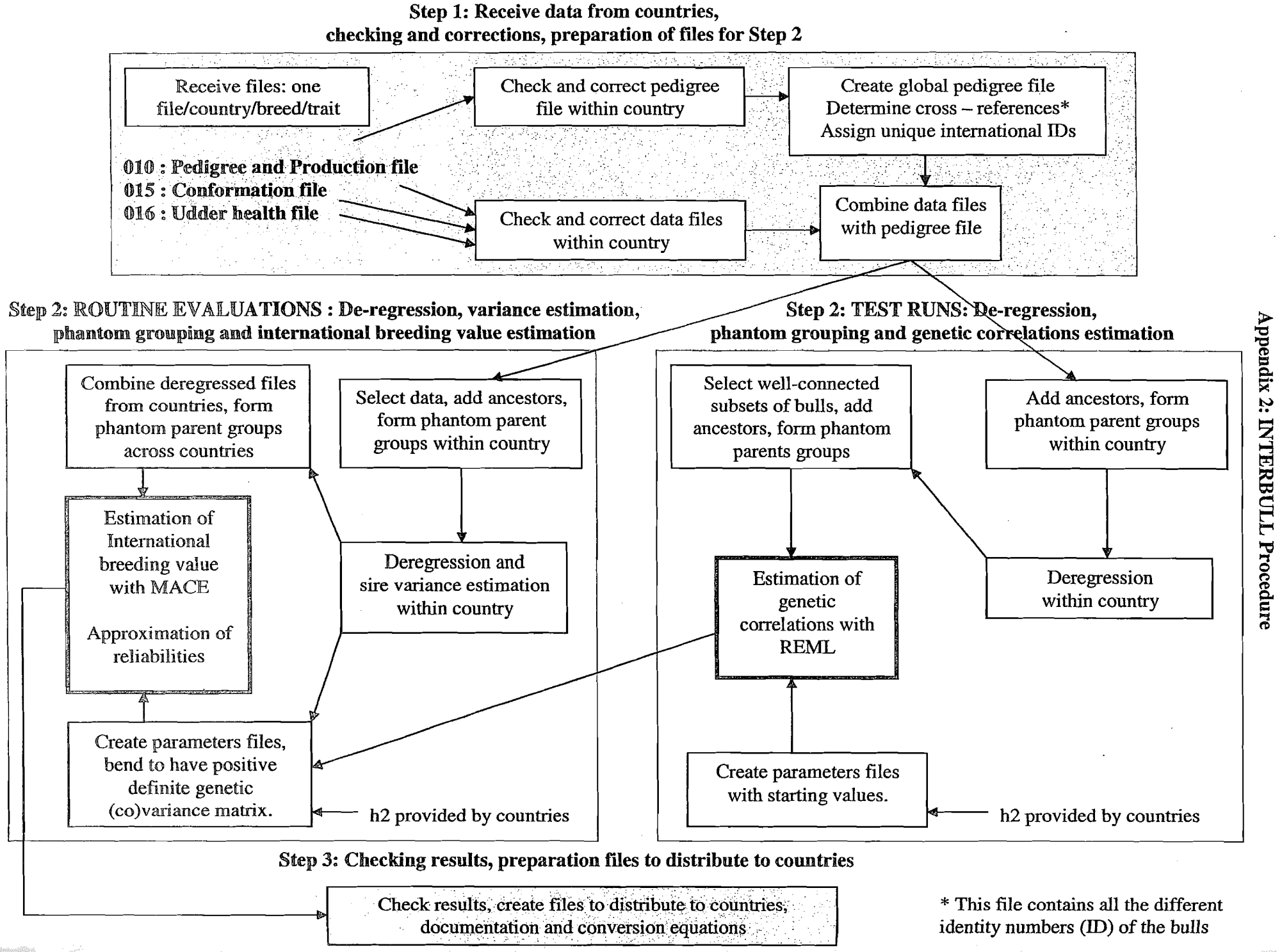
Appendix 1

The multiple-trait herd cluster model for international dairy sire evaluation.
(Weigel, K.A. and R. Rekaya. 2000)

Multiple-trait BLUP analysis, where each cluster is considered as a different trait, is computed based on individual animal performances with this sire model:

$$y_{ijklmno} = \text{herd-year}_i + \text{season}_j + \text{age}_k + \text{frequency}_l + \beta \times DIM_m + \text{sire}_n + \text{error}_{ijklmno},$$

where
$y_{ijklmno}$ = first lactation milk yield,
herd-year_i = interaction of herd and year of calving,
season_j = season of calving (3-months season were used),
age_k = age of cow at calving,
frequency_l = number of times milked per day,
β = regression coefficient,
DIM_m = days in milk,
sire_n = sire of cow, and
$\text{error}_{ijklmno}$= random residual.


Step 1: Receive data from countries, checking and corrections, preparation of files for Step 2
Receive files: one file/country/breed/trait
Check and correct pedigree file within country
Create global pedigree file
Determine cross – references*
Assign unique international IDs
010 : Pedigree and Production file
015 : Conformation file
016 : Udder health file
Check and correct data files within country
Combine data files with pedigree file
Step 2: ROUTINE EVALUATIONS : De-regression, variance estimation, phantom grouping and international breeding value estimation
Step 2: TEST RUNS: De-regression, phantom grouping and genetic correlations estimation
Combine deregressed files from countries, form phantom parent groups across countries
Select data, add ancestors, form phantom parent groups within country
Estimation of International breeding value with MACE
Approximation of reliabilities
Deregression and sire variance estimation within country
Create parameters files, bend to have positive definite genetic (co)variance matrix.
h2 provided by countries
Select well-connected subsets of bulls, add ancestors, form phantom parents groups
Add ancestors, form phantom parent groups within country
Estimation of genetic correlations with REML
Deregression within country
Create parameters files with starting values.
h2 provided by countries
Step 3: Checking results, preparation files to distribute to countries
Check results, create files to distribute to countries, documentation and conversion equations
* This file contains all the different identity numbers (ID) of the bulls

Appendix 3

Genetic correlation estimation procedure used by INTERBULL for production traits

From INTERBULL web site
www.interbull.org

" Estimation of genetic correlations among countries takes place in test-runs only, when new or modified data are submitted from a country, according to the following procedure (as per Interbull technical workshop of April 1995, Uppsala, Sweden):

Step 1:
Several subsets of countries are analysed. At the most 10 countries at a time are included in each subset. Countries that are major link contributors (judged from the number and origin of common bulls with multiple national evaluations) are always included in these subsets. If multiple genetic correlation estimates are computed for a country pair, the highest estimate is kept, as per Sigurdsson et al. (1996) showing that genetic correlations may be under-estimated but not over-estimated by the method used.

Step 2:
In some cases sufficient links between countries may be missing, resulting in close to zero genetic correlation estimates. If no reasonable correlations can be estimated via indirect links with third countries, one the the following procedures is followed:

a) Estimates from another breed for the country pair are used, if applicable
b) Product moment correlations of common bulls, adjusted for national evaluation accuracy are used (not a very frequent practice, since presence of common bulls will likely result in reasonable correlation estimates using the approximate REML method of Sigurdsson et al.)
c) Estimates from the low end of the correlation distribution are assigned; these would normally range from .86 to .89 between two North Hemisphere countries and from .75 to .78 between a North and a South Hemisphere country; the model of national evaluation is also taken into consideration (countries with similar national evaluation models are assigned higher genetic correlation estimates)

Step 3:
Since genetic correlation estimates are not derived simultaneously, the full covariance matrix need to be bent in order to ensure it's positive definite.

Efforts to improve the procedure are currently under way. The use of covariance structure of models that include genetic and non-genetic (eg, national evaluation model, management practice etc) components in determining correlation estimates between weakly or non-linked countries/populations is being studied. If a country is not linked to the other countries in the evaluation system, its data are not included in the international genetic evaluation."

Appendix 4

Two equivalent model specifications for MACE

The usual specification of a multiple-trait model is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \qquad (1)$$

with:

$$\begin{aligned}
\mathrm{var}(\mathbf{e}) &= \begin{bmatrix} \sigma_1^2 w_1^{-1} & & \varnothing \\ & \ddots & \\ \varnothing & & \sigma_n^2 w_n^{-1} \end{bmatrix} \\
&= \begin{bmatrix} w_1^{-1/2} & & \\ & \ddots & \\ \varnothing & & w_n^{-1/2} \end{bmatrix} \begin{bmatrix} \sigma_1^2 & & \varnothing \\ & \ddots & \\ \varnothing & & \sigma_n^2 \end{bmatrix} \begin{bmatrix} w_1^{-1/2} & & \varnothing \\ & \ddots & \\ \varnothing & & w_n^{-1/2} \end{bmatrix} \\
&= \mathbf{D}^{-1/2} \boldsymbol{\Lambda} \mathbf{D}^{-1/2} \\
&= \mathbf{R} \\
\mathrm{var}(\mathbf{a}) &= \mathbf{G}
\end{aligned}$$

and

$$\mathrm{var}(\mathbf{y}) = \mathbf{V} = \mathbf{R} + \mathbf{ZGZ}'$$

Pre-multiplying the vector of observations with $\mathbf{D}^{-1/2}$ and nested regression for fixed and random effects such that the design matrices for fixed and random effects become $\mathbf{D}^{-1/2}\mathbf{X}$ and $\mathbf{D}^{-1/2}\mathbf{Z}$, respectively, leads to the following model specification:

$$\mathbf{D}^{-1/2}\mathbf{y} = \mathbf{D}^{-1/2}\mathbf{Xb} + \mathbf{D}^{-1/2}\mathbf{Za} + \mathbf{D}^{-1/2}\mathbf{e}$$

or

$$\mathbf{y}_o = \mathbf{X}_o\mathbf{b} + \mathbf{Z}_o\mathbf{a} + \mathbf{e}_o \qquad (2)$$

Note that:

$$\begin{aligned}
\mathrm{var}(\mathbf{y}_o) = \mathbf{V}_o &= \mathrm{var}\left(\mathbf{D}^{-1/2}\mathbf{y}\right) = \mathbf{D}^{-1/2}\,\mathrm{var}(\mathbf{y})\,\mathbf{D}^{-1/2} = \mathbf{D}^{-1/2}\mathbf{V}\mathbf{D}^{-1/2} \\
\mathbb{V}_o^{-1} &= \left[\mathbb{D}^{-1/2}\mathbb{V}\mathbb{D}^{-1/2}\right]^{-1} = \mathbb{D}^{1/2}\mathbb{V}^{-1}\mathbb{D}^{1/2} \\
\mathrm{var}(\mathbf{e}_o) = \mathbb{R}_o &= \mathrm{var}\left(\mathbb{D}^{-1/2}\mathbf{e}\right) = \mathbb{D}^{-1/2}\,\mathrm{var}(\mathbf{e})\,\mathbb{D}^{-1/2} = \mathbb{D}^{-1/2}\mathbb{D}^{1/2}\Lambda\mathbb{D}^{1/2}\mathbb{D}^{-1/2} \\
&= \Lambda
\end{aligned}$$

Log likelihood for a mixed linear model

Assuming normality of the random effects, -2 times the logarithm of the restricted likelihood for the model (**1**) is:

$$-2\log\ell(\boldsymbol{\theta}\,|\,\mathbf{K}'\mathbf{y}) = (n-p)\log(2\pi) + \log|\mathbf{K}'\mathbf{V}\mathbf{K}| + \mathbf{y}'\mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}\mathbf{y}$$

Here n and p are the number of observations and rank of the design matrix for fixed effects, respectively. $\mathbf{K}$ is a matrix for which the following condition holds: $\mathbf{K}'\mathbf{X}=0$.

The following equalities can be shown to hold:

$$\mathbf{y}'\mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}'\mathbf{y} = (\mathbf{y}-\mathbf{X}\mathbf{b})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\mathbf{b})$$

$$\log|\mathbf{K}'\mathbf{V}\mathbf{K}| = \log|\mathbf{V}| + \log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|$$

$$\log|\mathbf{V}| + \log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| = \log|\mathbf{R}| + \log|\mathbf{G}| + \log|\mathbf{C}|$$

where $\mathbf{C}$ denotes the Henderson mixed model equation.

Hence, the logarithm of the restricted likelihood can be rewritten as:

$$-2\log\ell(\boldsymbol{\theta}\,|\,\mathbf{K}'\mathbf{y}) = (n-p)\log(2\pi) + \log|\mathbf{R}| + \log|\mathbf{G}| + \log|\mathbf{C}| + (\mathbf{y}-\mathbf{X}\mathbf{b})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\mathbf{b})$$

Similarly, the logarithm of the restricted likelihood for model (**2**) can be written as:

$$-2\log\ell(\boldsymbol{\theta}_o\,|\,\mathbf{K}'\mathbf{y}_o) = (n-p)\log(2\pi) + \log|\mathbf{R}_o| + \log|\mathbf{G}| + \log|\mathbf{C}_o| + (\mathbf{y}_o-\mathbf{X}_o\mathbf{b})'\mathbf{V}^{-1}(\mathbf{y}_o-\mathbf{X}_o\mathbf{b})$$

Next the difference between both log restricted likelihoods is determined. The first term is equal, as the number of observations and fixed effects is the same in both analyses. The genetic variance-covariance matrix - and therefor the third term - is the same for both models. With respect to the fourth term (determinant of the mixed model equations), note that:

$$\begin{aligned}
\mathbf{C}_o &= \begin{bmatrix} \mathbf{X}_o'\mathbf{R}_o^{-1}\mathbf{X}_o & \mathbf{X}_o'\mathbf{R}_o^{-1}\mathbf{Z}_o \\ \mathbf{Z}_o'\mathbf{R}_o^{-1}\mathbf{X}_o & \mathbf{Z}_o'\mathbf{R}_o^{-1}\mathbf{Z}_o + \mathbf{G}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{X}'\mathbf{D}^{-1/2}\boldsymbol{\Lambda}\mathbf{D}^{-1/2}\mathbf{X} & \mathbf{X}'\mathbf{D}^{-1/2}\boldsymbol{\Lambda}\mathbf{D}^{-1/2}\mathbf{Z} \\ \mathbf{Z}'\mathbf{D}^{-1/2}\boldsymbol{\Lambda}\mathbf{D}^{-1/2}\mathbf{X} & \mathbf{Z}'\mathbf{D}^{-1/2}\boldsymbol{\Lambda}\mathbf{D}^{-1/2}\mathbf{Z} + \mathbb{G}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{X}'\mathbb{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbb{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbb{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbb{R}^{-1}\mathbf{Z} + \mathbb{G}^{-1} \end{bmatrix} \\
&= \mathbb{C}
\end{aligned}$$

Thus the fourth term is the same for both models.

The last term is also the same in both log restricted likelihoods:

$$\mathbf{y}_o - \mathbf{X}_o\mathbf{b} = \mathbf{D}^{-1/2}\mathbf{y} - \mathbf{D}^{-1/2}\mathbf{X}\mathbf{b} = \mathbf{D}^{-1/2}\left(\mathbf{y} - \mathbf{X}\mathbf{b}\right)$$

$$\left(\mathbf{y}_o - \mathbf{X}_o\mathbf{b}\right)' \mathbf{V}_o^{-1}\left(\mathbf{y}_o - \mathbf{X}_o\mathbf{b}\right) = \left(\mathbf{y} - \mathbf{X}\mathbf{b}\right)' \mathbf{D}^{-1/2}\mathbf{D}^{1/2}\mathbf{V}^{-1}\mathbf{D}^{1/2}\mathbf{D}^{-1/2}\left(\mathbf{y} - \mathbf{X}\mathbf{b}\right)$$

$$= \left(\mathbf{y} - \mathbf{X}\mathbf{b}\right)' \mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\mathbf{b}\right)$$

The only remaining term is the second one:

$$\log|\mathbf{R}| = \log\left|\mathbf{D}^{-1/2}\Lambda\mathbf{D}^{-1/2}\right|$$

$$= \log\left|\mathbf{D}^{-1/2}\right| + \log|\Lambda| + \log\left|\mathbf{D}^{-1/2}\right|$$

$$= \log\left|\mathbf{D}^{-1/2}\right| + \log|\mathbf{R}_o| + \log\left|\mathbf{D}^{-1/2}\right|$$

Thus, the difference between logarithm of the restricted likelihood is twice the determinant of $\mathbf{D}^{-1/2}$. $\mathbf{D}^{-1/2}$ is a diagonal matrix with elements equal to the inverse of the square root of weights given to each observation. Note that the determinant of a diagonal matrix equals the product of all elements. Then:

$$2\log\left|\mathbf{D}^{-1/2}\right| = -2\log\left(\sqrt{w_1}\cdot\sqrt{w_2}\cdot\ldots\cdot\sqrt{w_n}\right)$$

$$= -\log\left(\sqrt{w_1}\cdot\sqrt{w_1}\cdot\sqrt{w_2}\cdot\sqrt{w_2}\cdot\ldots\cdot\sqrt{w_n}\cdot\sqrt{w_n}\right)$$

$$= -\log\left(w_1\cdot w_2\cdot\ldots\cdot w_n\right)$$

$$= -\log\left(w_1\right) - \log\left(w_2\right) - \ldots - \log\left(w_n\right)$$

Freddy Fikse
Interbull Centre
July 2003

Appendix 5

Approximation of SE of r_G among countries

Unstructured genetic variance-covariance matrix

The genetic correlation between two countries is computed as:

$$r_G = \frac{\sigma_{aa'}}{\sqrt{\sigma_a^2 \cdot \sigma_{a'}^2}},$$

where:

$\sigma_{aa'}$ = genetic covariance between both countries

σ_a^2 = genetic variance in a country

The estimate of the genetic correlation, $\hat{r}_G$, can be approximated with a first order Taylor series expansion around the true variance components:

$$\hat{r}_G \approx r_G + \left(\hat{\sigma}_{aa'} - \sigma_{aa'}\right)\frac{\partial r_G}{\partial \sigma_{aa'}} + \left(\hat{\sigma}_a^2 - \sigma_a^2\right)\frac{\partial r_G}{\partial \sigma_a^2} + \left(\hat{\sigma}_{a'}^2 - \sigma_{a'}^2\right)\frac{\partial r_G}{\partial \sigma_a^2}$$

The variance of $\hat{r}_G$ can then be approximated as:

$$
\begin{aligned}
Var(\hat{r}_G) \approx Var(r_G) &+ Var\left((\hat{\sigma}_{aa'} - \sigma_{aa'}) \frac{\partial r_G}{\partial \sigma_{aa'}} \right) \\
&+ Var\left((\hat{\sigma}_a^2 - \sigma_a^2) \frac{\partial r_G}{\partial \sigma_a^2} \right) \\
&+ Var\left((\hat{\sigma}_a^2 - \sigma_a^2) \frac{\partial r_G}{\partial \sigma_a^2} \right) \\
&+ 2 \cdot Cov\left(r_G, (\hat{\sigma}_{aa'} - \sigma_{aa'}) \frac{\partial r_G}{\partial \sigma_{aa'}} \right) \\
&+ 2 \cdot Cov\left(r_G, (\hat{\sigma}_a^2 - \sigma_a^2) \frac{\partial r_G}{\partial \sigma_a^2} \right) \\
&+ 2 \cdot Cov\left(r_G, (\hat{\sigma}_{a'}^2 - \sigma_{a'}^2) \frac{\partial r_G}{\partial \sigma_{a'}^2} \right) \\
&+ 2 \cdot Cov\left((\hat{\sigma}_{aa'} - \sigma_{aa'}) \frac{\partial r_G}{\partial \sigma_{aa'}}, (\hat{\sigma}_a^2 - \sigma_a^2) \frac{\partial r_G}{\partial \sigma_a^2} \right) \\
&+ 2 \cdot Cov\left((\hat{\sigma}_{aa'} - \sigma_{aa'}) \frac{\partial r_G}{\partial \sigma_{aa'}}, (\hat{\sigma}_{a'}^2 - \sigma_{a'}^2) \frac{\partial r_G}{\partial \sigma_{a'}^2} \right) \\
&+ 2 \cdot Cov\left((\hat{\sigma}_a^2 - \sigma_a^2) \frac{\partial r_G}{\partial \sigma_a^2}, (\hat{\sigma}_{a'}^2 - \sigma_{a'}^2) \frac{\partial r_G}{\partial \sigma_{a'}^2} \right)
\end{aligned}
$$

The parameters (r_G , $\sigma_{aa'}$, σ_a^2) are fixed variables, but the estimators ($\hat{r}_G$, $\hat{\sigma}_{aa'}$, $\hat{\sigma}_a^2$) are random variables. Thus all variance and covariance terms involving r_G in the previous expression are zero. Likewise, the other variance and covariance terms can be simplified, for example:

$$
Cov\left((\hat{\sigma}_{aa'} - \sigma_{aa'}) \frac{\partial r_G}{\partial \sigma_{aa'}}, (\hat{\sigma}_a^2 - \sigma_a^2) \frac{\partial r_G}{\partial \sigma_a^2} \right) = Cov(\hat{\sigma}_{aa'}, \hat{\sigma}_a^2) \cdot \frac{\partial r_G}{\partial \sigma_{aa'}} \cdot \frac{\partial r_G}{\partial \sigma_a^2}
$$

The expression for the variance of $\hat{r}_G$ can then be simplified to:

$$Var(\hat{r}_G) \approx Var(\hat{\sigma}_{aa'})\left(\frac{\partial r_G}{\partial \sigma_{aa'}}\right)^2 + Var(\hat{\sigma}_a^2)\left(\frac{\partial r_G}{\partial \sigma_a^2}\right)^2 + Var(\hat{\sigma}_{a'}^2)\left(\frac{\partial r_G}{\partial \sigma_a^2}\right)^2$$
$$+2 \cdot Cov(\hat{\sigma}_{aa'}, \hat{\sigma}_a^2) \cdot \frac{\partial r_G}{\partial \sigma_{aa'}} \cdot \frac{\partial r_G}{\partial \sigma_a^2}$$
$$+2 \cdot Cov(\hat{\sigma}_{aa'}, \hat{\sigma}_{a'}^2) \cdot \frac{\partial r_G}{\partial \sigma_{aa'}} \cdot \frac{\partial r_G}{\partial \sigma_{a'}^2}$$
$$+2 \cdot Cov(\hat{\sigma}_a^2, \hat{\sigma}_{a'}^2) \cdot \frac{\partial r_G}{\partial \sigma_a^2} \cdot \frac{\partial r_G}{\partial \sigma_{a'}^2}$$

The variances and covariances in this expression are for example obtained from the inverse of the average-information matrix of an AI-Reml program.

The partial derivatives of the genetic correlation are:

$$\frac{\partial r_G}{\partial \sigma_{aa'}} = \frac{\partial}{\partial \sigma_{aa'}} \frac{\sigma_{aa'}}{\sqrt{\sigma_a^2 \cdot \sigma_{a'}^2}} = \frac{1}{\sqrt{\sigma_a^2 \cdot \sigma_{a'}^2}}$$

$$\frac{\partial r_G}{\partial \sigma_a^2} = \frac{\partial}{\partial \sigma_{aa'}} \frac{\sigma_{aa'}}{\sqrt{\sigma_a^2 \cdot \sigma_{a'}^2}} = -\frac{\sigma_{aa'}}{2 \cdot \sigma_a^2 \cdot \sqrt{\sigma_a^2 \cdot \sigma_{a'}^2}} = -\frac{r_G}{2 \cdot \sigma_a^2}$$

Since the true parameters are not known, the (restricted) maximum likelihood estimator is used instead.

References

Árnason, T. 1998. Varians i skattningar av kvotfunktioner av varians- och kovarianskomponenter med hjälp av Taylorutveckling av skattningsvarians kring varianskomponenterna. Mimeo.

Gerstmayr, S. 1992. Impact of the data structure on the reliability of the estimated genetic parameters in an animal model with maternal effects. J. Anim. Breed. Genet. 109, 321-336.

Freddy Fikse
Interbull Centre
May 2003

Approximation of SE of r_G among countries

Structural model

The genetic correlation between two countries is computed as:

$$r_G = \exp\left(-d_{ij}\right),$$

where:

$d_{ij} = \sqrt{\sum_k \left(x_{ik} - x_{jk}\right)^2}$, the distance between both countries as a function of coordinates x_{ik} and x_{jk}.

For example, four countries can be represented in a three-dimensional space. Let the coordinates for the second and third country be $(p_1,0,0)$ and $(p_2,p_3,0)$, respectively. The genetic correlation between the second and third country is equal to:

$$r_G = \exp\left(-\sqrt{\left(p_2 - p_1\right)^2 + \left(p_3\right)^2}\right)$$

The Taylor series expansion for $\hat{r}_G$ develops into:

$$\hat{r}_G \approx r_G + \left(\hat{p}_1 - p_1\right)\frac{\partial r_G}{\partial p_1} + \left(\hat{p}_2 - p_2\right)\frac{\partial r_G}{\partial p_2} + \left(\hat{p}_3 - p_3\right)\frac{\partial r_G}{\partial p_3}$$

In matrix notation this can be written as:

$$\hat{r}_G \approx r_G + \left[\frac{\partial r_G}{\partial p_1} \quad \frac{\partial r_G}{\partial p_2} \quad \frac{\partial r_G}{\partial p_3}\right]\begin{bmatrix}\hat{p}_1 - p_1 \\ \hat{p}_2 - p_2 \\ \hat{p}_3 - p_3\end{bmatrix} = r_G + \mathbf{D}' \cdot \left[\hat{\mathbf{P}} - \mathbf{P}\right]$$

Hence the variance of $\hat{r}_G$ is computed as:

$$\begin{aligned} Var\left(\hat{r}_G\right) &\approx Var\left(r_G\right) + Var\left(\mathbf{D}' \cdot \left[\hat{\mathbf{P}} - \mathbf{P}\right]\right) \\ &= \mathbb{D}' \cdot Var\left(\left[\hat{\mathbb{P}} - \mathbb{P}\right]\right) \cdot \mathbb{D} \\ &= \mathbb{D}' \cdot Var\left(\hat{\mathbb{P}}\right) \cdot \mathbb{D} \end{aligned}$$

since both r_G and $\mathbb{P}$ are fixed variables. Note that $Var\left(\hat{\mathbb{P}}\right)$ is (a submatrix of) the inverse

of the AI matrix!

The elements of **D** are:

$$\frac{\partial r_G}{\partial p_1} = \exp\left(-\sqrt{(p_2-p_1)^2+(p_3)^2}\right)\cdot\frac{p_2-p_1}{\sqrt{(p_2-p_1)^2+(p_3)^2}}$$

$$\frac{\partial r_G}{\partial p_2} = -\exp\left(-\sqrt{(p_2-p_1)^2+(p_3)^2}\right)\cdot\frac{p_2-p_1}{\sqrt{(p_2-p_1)^2+(p_3)^2}}$$

$$\frac{\partial r_G}{\partial p_3} = -\exp\left(-\sqrt{(p_2-p_1)^2+(p_3)^2}\right)\cdot\frac{p_3}{\sqrt{(p_2-p_1)^2+(p_3)^2}}$$

Freddy Fikse
Interbull Centre
May 2003

Appendix 6

Comparison CM6-SM36
Denmark and Finland added to the axes countries.

Estimated genetic correlations from SM36 and their standard errors (above diagonal), deviations (SM36-CM6) in estimated genetic correlations and deviations in standard errors (below diagonal).

	NLD	USA	NZL	HUN	DNK	FIN
NLD		0.928	0.778	0.857	0.954	0.872
		(0.006)	*(0.016)*	*(0.016)*	*(0.007)*	*(0.030)*
USA	0.002		0.726	0.885	0.969	0.897
	0.002		*(0.012)*	*(0.020)*	*(0.010)*	*(0.039)*
NZL	-0.007	0.016		0.686	0.749	0.687
	0.007	*0.001*		*(0.014)*	*(0.017)*	*(0.023)*
HUN	0.000	0.000	-0.001		0.869	0.846
	0.009	*0.015*	*0.001*		*(0.013)*	*(0.075)*
DNK	0.000	-0.004	**-0.062**	-0.008		0.887
	0.005	*0.009*	*0.009*	*0.007*		*(0.044)*
FIN	0.009	**0.030**	**0.078**	0.014	**-0.029**	
	0.023	*0.032*	*0.007*	*0.067*	*0.039*	

In bold: more than 0.02 of difference in Rg

Appendix 7

Coordinates and standard errors obtained from combination of SM36FRAAUS and SM36CANEST (CALC) and those estimated with SM38

	CALC	SE		SM38	SE
NLD1	0			0	
NLD2	0			0	
NLD3	0			0	
USA1	0.079			0.080	*(0.006)*
USA2	0			0	
USA3	0			0	
NZL1	-0.224		Average of SM36FRAAUS and SM36CANEST	-0.218	*(0.050)*
NZL2	0.132			0.154	*(0.077)*
NZL3	0			0	
HUN1	0.109			0.133	*(0.026)*
HUN2	**0.030**			**0.104**	*(0.025)*
HUN3	**0.111**			**0.017**	*(0.084)*
FRA1	0.010	*(0.011)*	From SM36FRAAUS	0.003	*(0.012)*
FRA2	0.068	*(0.020)*		0.057	*(0.029)*
FRA3	0.017	*(0.071)*		0.044	*(0.037)*
AUS1	-0.113	*(0.043)*		-0.132	*(0.029)*
AUS2	0.148	*(0.043)*		0.114	*(0.046)*
AUS3	-0.002	*(0.112)*		0.061	*(0.027)*
CAN1	0.051	*(0.006)*	From SM36CANEST	0.049	*(0.006)*
CAN2	-0.029	*(0.007)*		-0.002	*(0.020)*
CAN3	**0.004**	*(0.033)*		**0.029**	*(0.006)*
EST1	0.169	*(0.095)*		0.129	*(0.083)*
EST2	-0.116	*(0.172)*		-0.086	*(0.136)*
EST3	-0.093	*(0.147)*		0.146	*(0.090)*

References

Banos, G. 1994. International genetic evaluation of dairy cattle. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production. **17**: 3-10.

Banos, G. and A. Sigurdsson. 1996. Application of contemporary methods for the use of international data in national genetic evaluations. J. Dairy Sci. **79**: 1117-1125.

Canavesi, F., D. Boichard, V. Ducrocq, N. Gengler, G. De Jong and Z. Liu. 2001. PROduction Traits European Joint Evaluation (PROTEJE). INTERBULL Bulletin **27**:32-34.

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 evaluation (PROTEJE). Proceedings of the 7th World Congress on Genetics Applied to Livestock Production. CD-ROM communication n° 01-59.

De Jong, G. 2003. MACE – Options for improvement. INTERBULL Bulletin **30**: 112-116.

Delaunay, I., V. Ducrocq and D. Boichard. 2002. A structural model for the matrix of genetic correlations between countries in international evaluations. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production. CD-ROM communication n° 01-14.

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 first three lactations of French Holstein cows. Proceedings of the 54th annual meeting of the European Association for Animal Production, Roma, Italy.

Fikse, W.F. and G. Banos. 2001. Weighting factors of sire daughter information in international genetic evaluations. J. Dairy Sci. **84**: 1759-1767.

Fikse, W.F. 2003. Approaches to incorporate genetic groups in Mace. Paper presented at the INTERBULL Workshop, March 2-3, 2003, Beltsville, MD, USA.

Gilmour, A. R., R. Thomson 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., B. R. Cullis, S.J. Welham and R. Thomson. 2000. ASREML Reference Manual.

Goddard, M. 1985. A method of comparing sires evaluated in different countries. Livst. Prod. Sci. **13:** 321-331.

Hofer, A. 1998. Variance component estimation in animal breeding: a review. J. Anim. Breed. Genet. **115**: 247-265.

INTERBULL. 1990. Recommended procedures for international use of sire proofs. INTERBULL Bulletin **4**: 17 pages.

INTERBULL. 2001. INTERBULL Guidlines for National & International Genetic Evaluation Systems in Dairy Cattle with Focus on Production Traits. INTERBULL Bulletin **28**: 27 pages.

INTERBULL. 2003. Effective daughter contribution (EDC) definition (pdf). http://www-interbull.slu.se/service_documentation/General/framesida-general.htm. Accessed April 1, 2003.

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. Mantysaari, P. Madsen and R. Thompson. 1997. Residual maximum likelihood estimation of (co) variance components in multivariate mixed linear models using average information. Jour. Ind. Soc. Ag. Statistics **49**: 215-236.

Mattalia, S. and B. Bonaïti. 1993. Use of full sibs families to estimate the 'a' coefficients of conversion formulas between countries. INTERBULL Bulletin **8**: 4 pages.

Powell, R. L. and G. R. Wiggans. 1995. Holstein conversion equations based on population variance and full-brother model. J. Dairy Sci. **78**: 1362-1368.

Rekaya R., K. Weigel and D. Gianola. 2001. Application of a structural model for genetic covariances in international dairy sire evaluations. J. Dairy Sci. **84**: 1525-1530.

Rekaya R., K. Weigel and D. Gianola. 2003. Bayesian estimation of parameters of a structural model for genetic covariances between milk yield in five regions of the United States. J. Dairy Sci. **86**: 1837-1844.

Schaeffer, L.R. 1985. Model for international evaluation of dairy sires. Livest. Prod. Sci. 12: 105-115.

Schaeffer, L.R. 1993. Multiple-trait, Across Country evaluation of dairy sires. INTERBULL Bulletin 8: 21 pages.

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

Sigurdsson, A. and G. Banos. 1995. Dependent variables in international sire evaluations. Acta Agric. Scand., Sect A, Animal Sci. **45:** 207-219.

Sigurdsson, A., G. Banos and J. Philipsson. 1996. Estimation of genetic (co)variance components for international evaluation of dairy bulls. Acta Agric. Scand., Sect A, Animal Sci. **46:** 129-136.

Stolzmann, M., H. Jasiorowski, Z. Reklexski, A. Zarnecki and G. Kalinowska. 1981. Friesian cattle in Poland. Preliminary results of testing different strains. World Animal Review **38**: 9.

Sullivan, P. G. 1999. REML estimation of heterogeneous sire (co)variances for MACE. INTERBULL Bulletin **22**: 146-148.

VanRaden, P. M. and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. **74**: 2737-2746.

Weigel, K. A. and G. Banos. 1997. Effect of time period of data used in international dairy sire evaluations. J. Dairy Sci. **80**: 3425-3430.

Weigel, K.A. and R. Rekaya. 2000. A multiple-trait herd cluster model for international dairy sire evaluation. J. Dairy Sci. **83**: 815-821.

Weigel, K.A. and N.R. Zwald. 2002. Selection and grouping of herds in international genetic evaluation of daughter performance records. INTERBULL Bulletin **29**: 7-11.

Westell, R.A., R.L. Quaas and L.D. Van Vleck. 1988. Genetics groups in an animal model. J. Dairy Sci. **71**: 1310-1318.

Wilmink, J.B.M., A. Meijering and B. Engel. 1986. Conversion of breeding values for milk from foreign populations. Livest. Prod. Sci. **14**: 223-229.

Wolfinger, R. 1993. Covariance structure selection in general mixed models. Comm. Statist.-Simula. **22(4)**: 1079-1106.

Zwald N.R., K.A. Weigel, W.F. Fikse and R. Rekaya. 2003. Application of a multiple-trait herd cluster model for genetic evaluation of dairy sires from seventeen countries. J. Dairy Sci. **86**: 376-382.

Abstract

One of the major problems of MACE is the computation of the genetic correlations between countries. An advantage of structural models is the option to reduce the number of parameters to estimate. The structural model used here defines the genetic correlations between 2 countries as an exponential function of the Euclidian distance between them. In this structural model, (k+1) countries can be represented in a k-dimensional space. The reduction of the number of dimensions of the space allows to reduce the number of parameters. For example, in a 3-dimensional space only 72 coordinates need to be estimated to compute 325 genetic correlations between 26 countries.

This structural model was successfully tested on simulated data, and on the current genetic correlations matrix used by INTERBULL. The first aim of the present study was to compare results of the structural model used on international data with results of an "unstructured" model. The second aim was to study the possible use of the coordinates of different structural models related to the same space, to calculate directly genetic correlations, without having to estimate them. The third aim was to analyze the influence of the choice of the axes countries.

Deregressed national breeding values used for Holstein milk yield international evaluation of February 2003 were analyzed. Several subsets of countries were considered. The structural model and a classical model were applied to estimate the genetic correlations between countries, using an AI-REML algorithm implemented in the AIREMLF90 program. Countries chosen to define axes in the structural model were based on the results from a previous study that applied a cluster analysis: The Netherlands as centre of the space, USA to define the 1st axis, New Zealand for the 2nd axis and Hungary for the 3rd axis.

Genetic correlations estimated with the structural model were very close to those estimated with the classical model. Larger differences (e.g., 0.06 for the correlation between Denmark and New Zealand) concerned the least connected of the countries considered. The standard errors of genetic correlations ranged from 0.006 to 0.058 depending on the amount of genetic ties. The -2logL was slightly higher for the structural model than for the classical model, but Bayesian information criterion favoured the structural model because of the lower number of parameters. Combining the coordinates obtained from different structural models related to the same space gave similar genetic correlations (differences were lower than the standard errors) and similar -2logL compared to the joint analyses of the countries. By this way, the number of runs needed to estimate all the genetic correlations can be reduced drastically. The change of some of the axes countries shows that the estimated genetic correlations were more precise when the volume defined by the axes countries was large.

Determination of the optimal number of countries to define the axes and the choice of the axes countries needs to be investigated further, using e.g. information criteria and volume of the space defined by axes countries.