

**AN ALTERNATIVE PROCEDURE FOR INTERNATIONAL EVALUATIONS :  
PRODUCTION TRAITS EUROPEAN JOINT EVALUATION (PROTEJE)**

**F. Canavesi<sup>1</sup>, D. Boichard<sup>2</sup>, V. Ducrocq<sup>2</sup>, N. Gengler<sup>3</sup>, G. De Jong<sup>4</sup> and Z. Liu<sup>5</sup>**

<sup>1</sup> ANAFI, Via Bergamo 292, 26100, Cremona, Italy

<sup>2</sup> INRA, SGQA, 78352 Jouy-en-Josas Cédex, France

<sup>3</sup> Gembloux University, Gembloux, Belgium

<sup>4</sup> CR-DELTA, Arnhem, The Netherlands

<sup>5</sup> VIT, Verden, Germany

**INTRODUCTION**

The November 2001 official international genetic evaluation of dairy bulls for production traits has seen 25 countries sending and receiving data from Interbull. The current method used for international comparisons of dairy bulls was developed by Schaeffer in 1993 and it is known as Multi-trait Across Country Evaluation (MACE). Bull proofs computed at the national level are provided to Interbull every four months. They are de-regressed and used as elementary observations in a multiple trait BLUP sire model considering production in different countries as different traits. At the end of the process, each country receives a list of EBVs on its own scale for all bulls evaluated in any of the participating countries. Over time, the method has been improved but has been found highly sensitive to changes in sire variance (within country as well as over time) to quality of national proofs and to changes in genetic correlations. As the number of Interbull countries is growing, the precise estimation of the about 350 genetic correlations for each trait between countries for Holsteins is becoming a formidable task. The present approach only evaluates bulls and with the increasing international exchange of female embryos and live cows, an international genetic evaluation of elite females is highly desirable.

In order to overcome potential biases resulting from the partial use of basic information, (summarised in the national sire proofs), several attempts to run a global animal model (GAM) using raw data (i.e. records and pedigree on all cows within each population and their related sires) have been made (Weigel *et al.*, 1999, 2001). While this alternative is very appealing, it has several limitations. Production systems, management and data collection conditions widely vary across countries. National evaluations have been developed to account for the most appropriate factors affecting production (Jorjani, 2000). It is unrealistic to include in GAM the exhaustive list of fixed effects considered in at least one country. But any selection of a subset of these fixed effects may lead to biased comparisons. Moreover, putting together raw data from many different countries raises questions about the feasibility to properly estimate genetic parameters for all of them, and the use of multiple lactations or test-day records will increase the difficulties to be faced. Following the theory behind the concept of borderless genetic evaluations (Lohuis, 1997), it has been suggested that differences among countries in MACE are sometimes imposed by national borders, and not by real differences at the genetic level. Clustering of countries (or herds across countries) with similar characteristics has thus been advocated as a tool to simplify the computations in order to make feasible a GAM evaluation across countries (Zwald *et al.*, 2001): the number of traits to evaluate would be equal to the number of distinct clusters. This approach brings up new questions: which are the most

appropriate variables to include in order to define clusters ? Do they exhaustively describe differences between countries ? Are they available in all participating countries ? How adequate is the final model in estimating effects found in particular areas of each of the countries analysed? How to deal in practice with several lists within a country ?

The objective of the PROTEJE project is to develop an alternative methodology for international evaluation, applicable to both bulls and cows, maintaining the modelling of environmental effects at the national level and reducing the computational burden of the estimation of a full rank, huge matrix of genetic correlations.

## MATERIAL AND METHODS

National genetic evaluation models have been built, developed and improved over time gathering expertise from all over the world, and more and more precisely accounting for local specificities (i.e., population structure, trait mean and variability, breeding schemes, data recording etc...). The approach envisioned in the PROTEJE project for the international genetic evaluation for bulls and cows assumes that this know-how can be entirely preserved. It is suggested to base the evaluation on *pre-corrected* records defined as lactations or, in the future, test-day records adjusted for all fixed effects estimated in *national* evaluations: the precise modelling at the local level should ensure that all appropriate effects have been accounted for. Pre-corrected records or yield deviations sent to international level could then be analyzed assuming a very simple model including only additive genetic and residual effects. As a result, a substantial part of the computational difficulty associated with a BLUP multiple trait animal model disappears. The use in a huge multiple trait evaluation of pre-corrected records obtained in a univariate context has already proved to be successful in other situations (Ducroq *et al.*, 2001; Besbes *et al.*, 2002). In order to clarify the goal of the project, Figure 1 displays the current system and the present proposal (PROTEJE) is illustrated in Figure 2.

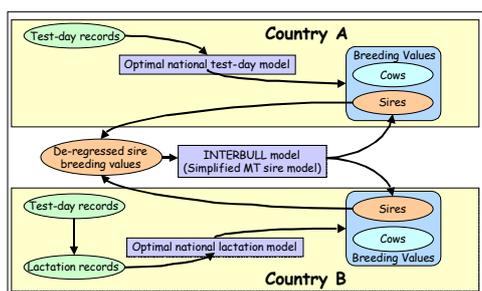


Figure 1. Current MACE procedures

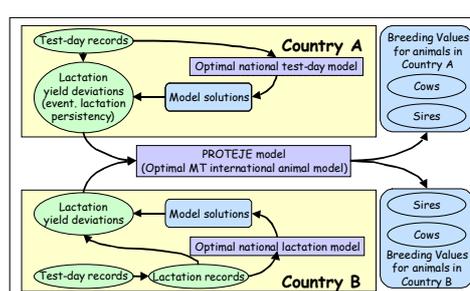


Figure 2. Proposed PROTEJE procedure

To alleviate the problem of estimating a huge matrix of genetic correlation, a reduction of its dimension seems a natural solution. Clustering is just one of the options to achieve this. An alternative procedure is to still consider distinct traits for different countries, while recognizing that globally these scales are related, imposing a reduced rank genetic correlation matrix. Simultaneously, a reparameterisation of the correlation matrix is proposed in order to avoid the difficulty associated with the estimation of high genetic correlations, possibly located on the border of the parameter space (see Delaunay *et al.*, 2002).

In practice the project implies:

1. the design and study of pre-correction steps for all fixed effects at the national level;
2. the design and study of the use of pre-corrected data into international genetic evaluations;
3. the estimation of a reparameterised, reduced rank matrix of genetic correlations based on pre-corrected records;
4. the comparison of this approach with national results and other international evaluation approaches.

The PROTEJE research group was set up within Europe following a request from the European Holstein Friesian Confederation. It includes France, Germany, Italy, The Netherlands and the University of Gembloux. Pre-corrected records were computed by each institute responsible for official national genetic evaluation and sent to all partners along with the original lactation data and the corresponding pedigree information. The total pedigree file consists of 21 millions animals and the total production file of more than 30 million records. Four countries are providing data and five research units are developing the project. The different tasks of the project are distributed among the different research groups according to computing resources and specific skills and interests of each research unit. One group looks at the pedigree file trying to identify problems arising from mishandling of original identifications for cows moving from one country to another one: very often cows sold to a different country were re-identified and the original ID lost as well as the true origin of the animal. A cross reference file for females is being built in order to keep track of imported and exported animal with multiple identifications and guidelines are developed in order to avoid similar problems in the future.

Another group verifies the theoretical properties of the proposed approach through a simulation study aiming at:

- assessing the quality of the use of pre-corrected records for the estimation of the genetic variance covariance matrix;
- studying the possibility to test for the proper rank of this matrix;
- illustrating the use of a reparameterised version of the reduced rank matrix for the REML estimation of the correlation matrix, using a second order algorithm (AI-REML).

Another research unit works on the international evaluation model itself and on the methodology required to provide pre-adjusted lactation records in case of a test day model used as the official national evaluation model. Proper weighting of pre-corrected records needs to be defined in order to account for every adjustment applied at the national level: contemporary group definition, adjustment for heterogeneity etc..

Finally, the issue of the origin of genetic correlations less than 1 between countries will be further investigated: what is true GxE and what is related to differences in evaluation models ?

## **DISCUSSION AND PERSPECTIVE**

The success of the PROTEJE project is conditioned by two major conjectures. It is assumed that a) little information is lost when international evaluations are based on pre-corrected records rather than on real data, b) a reduced rank, reparameterised genetic correlation matrix would stop the increase in complexity of the international evaluation itself and of the prior estimation of its required genetic parameters each time a new country is added to the system.

Many different aspects need to be tested in order to ensure that the methodology, -especially the use of pre-corrected data - is working nearly as well as a strict multitrait multicountry

animal model based on original performances. The use of simulated data will help to sequentially test how good the method is and how it compares with the current approach used by Interbull: first, the correlation matrix (G) will be estimated based on complete records, assuming a complete multitrait BLUP animal model. This estimate will be compared to situations based on pre-corrected records, in which fixed effects are obtained from national evaluations based either on true genetic parameters or computed estimates. Then, if the use of pre-corrected records seems satisfactory, G will be estimated using a sire model: in practice, at the international level with a large number of countries, only a sire model seems compatible with the estimation of genetic correlations. The benefits of reparameterisation and the quality of reduced rank estimates will be assessed. Finally, sire and cow proofs obtained from multiple-trait animal models based on complete or pre-corrected records and on different estimates of G will be compared.

### CONCLUSION

The PROTEJE project is a three year project and is only at its beginning. The team work of four European countries and five institutes will hopefully lead to results beneficial to all countries interested in international comparisons. Results are often discussed between the participating scientists and will be regularly presented at Interbull meetings. It is hoped that they will pave the way to a more flexible and precise international evaluation of both bulls and cows by Interbull, while maintaining the existing expertise at the national level.

### REFERENCES

- Besbes *et al.*, (2002) These proceedings  
Delaunay *et al.*, (2002) These proceedings  
Ducrocq, V., Boichard, D., Barbat, A. and Larroque, H. (2001) *J. Dairy Sci.*, **84** : suppl. 1: 25.  
Jorjani, H., (2000) *Interbull bulletin* **24**.  
Lohuis, M.M. and Dekkers, J.C.M. (1997) *Proc. 6<sup>th</sup> WCGALP* **26** : 169-172.  
Schaeffer, L.R. (1994) *J. Dairy Sci* **77** : 2671-2678.  
Weigel, K.A. and Rekaya, R. (2000) *J. Dairy Sci* **83** : 815-821.  
Weigel, K.A., Zwald, N.R. and Rekaya, R. (2001) *Interbull bulletin* **27** : 19-25.  
Zwald, N.R., Weigel, K.A. and Rekaya, R. (2001) *Interbull Bulletin* **27** : 13-18.