# Validation of an Approximate REML Algorithm for Parameter Estimation in a Multitrait, Multiple Across-Country Evaluation Model: A Simulation Study

**J. Tarrés,*†[1] Z. Liu,* V. Ducrocq,† F. Reinhardt,* and R. Reents***
*Vereinigte Informationssysteme Tierhaltung w.v., Heideweg 1, 27283 Verden, Germany
†UR337, Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique, 78352 Jouy-en-Josas Cedex, France

## ABSTRACT

A multitrait, multiple across-country evaluation (MT-MACE) model permitting a variable number of correlated traits per country allows international genetic evaluation models to more closely match national models. Before the MT-MACE evaluation can be applied, genetic (co)variance components within and across country must be estimated. An approximate REML algorithm for parameter estimation was developed and was validated via simulation. This method is based on the expectation maximization REML (EM-REML) algorithm. Because obtaining the inverse of coefficient matrix is not usually feasible for large amounts of data, an algorithm using the multiple-trait effective daughter contribution (EDC) is proposed to provide approximate diagonal elements of the inverse matrix. The accuracy of the approximate EM-REML was tested with simulated data and compared with an average information REML (AI-REML) from available software. Two simulation studies were performed. First, data of 2 countries were simulated using a single-trait model. Estimates of across-country genetic correlations with the developed algorithm were unbiased and very precise. The precision, however, depended on the percentage of bulls with data in both countries. The results obtained with the approximate EM-REML software were very close to those obtained with the AI-REML software regarding estimated genetic correlations and bulls' estimated breeding values. The second simulation assumed a multiple trait model and the same number of traits, pedigree structure, EDC, and pattern of missing records as for actual observations for milk yield obtained from French and German national Holstein evaluations. As with the single-trait scenarios, the approximate EM-REML gave nearly unbiased and very precise estimates of within- and across-country genetic correla-

tions. The results obtained in both simulation studies confirmed the suitability of the MT-MACE model and approximate EM-REML software in a wide range of situations. Even when the genetic trend was incorrectly estimated by the national evaluations, a joint analysis including a time effect in the MT-MACE model adequately corrected for this bias.

**Key words:** multiple across-country evaluation, restricted maximum likelihood, daughter yield deviation, effective daughter contribution

## INTRODUCTION

The multiple across-country evaluation (**MACE**; Schaeffer, 1994) methodology is used for international dairy bull comparisons. Estimated breeding values in national genetic evaluations are deregressed within each country to obtain the values of the dependent variable for bulls that have daughters with records. Currently, a single EBV per bull is permitted for each country in international genetic evaluations by Interbull. As more and more countries have upgraded their national genetic evaluation system to a multiple-trait model or a multiple-lactation random regression test-day model (**RRTDM**), differences among models for national and international evaluations have become increasingly evident. To optimize genetic evaluation models for both national and international evaluations, Sullivan and Wilton (2001) proposed a multiple-trait MACE (**MT-MACE**) with a variable number of traits per country. This model extended the current single-trait MACE (**ST**-MACE) to multiple lactations, or traits for countries using a multiple trait model in national genetic evaluation. Schaeffer (2001) developed a multiple trait de-regression method for MACE evaluation. More recently, the simulation study by Sullivan et al. (2005) confirmed the theoretical expectation that MT-MACE methods should be preferred over methods that allow only one trait per country when assuming true genetic parameters. Mark and Sullivan (2006) applied MT-MACE to field data for udder health to quantify the benefits in terms of reliability and ability to predict

international genetic merit of MT-MACE compared with ST-MACE, and to discuss the practical implementation of MT-MACE. They concluded that the MT-MACE method is recommended for international genetic udder health evaluations.

In parallel to the development involving de-regression, Ducrocq et al. (2001) suggested using corrected records for a 2-step multitrait genetic evaluation. Indeed, this approach can also be applied to international genetic evaluations. Yield deviations (**YD**) of cows and daughter yield deviations (**DYD**) of bulls are, in fact, corrected records. For single-trait models, VanRaden and Wiggans (1991) developed formulas for the calculation of YD and DYD and its daughter equivalent, which were later called effective daughter contribution (**EDC**). Liu et al. (2004a) extended the YD and DYD derivation to multiple trait models, including RRTDM. Also, Liu et al. (2004a) developed a method for approximating reliabilities of EBV under general multiple trait models, which can also be used to approximate EDC associated with bulls' DYD.

Based on DYD, Liu et al. (2004b) presented an MT-MACE model for international bull comparison. This model allowed a variable number of correlated traits per country. For countries using multiple trait models in national genetic evaluations, a vector of DYD and its corresponding EDC matrix are needed for each bull with daughter performance information. An approximate REML algorithm was developed to estimate across country genetic correlations based on multitrait EDC (**MTEDC**, Liu et al., 2004a). For solving the large equation system of the MT-MACE model, a preconditioned conjugate gradients algorithm (**PCG**, Strandén and Lidauer, 1999) was applied together with the iteration on data technique. This MACE model can also be used to analyze YD of cows or de-regressed estimates of breeding values.

The aim of this paper was to validate the approximate REML method for estimating across-country genetic correlations via simulation. In addition to the analysis with a ST-MACE model, a MT-MACE analysis was performed to check the suitability of this approximate REML algorithm in a wide range of situations.

## MATERIALS AND METHODS

### *The MT-MACE Model*

For a country $j$ using a multitrait model in national genetic evaluation, the following statistical model was applied to the data of a bull $i$ from the country j:

$$\mathbf{y}_{ij} = \mathbf{f}_j + \mathbf{a}_{ij} + \mathbf{e}_{ij} \qquad [1]$$

where $\mathbf{y}_{ij}$ is a vector of data of the $i$th bull in country j, $\mathbf{f}_j$ is a vector of overall means for traits of the $j$th country, $\mathbf{a}_{ij}$ is a vector of additive genetic effects of bull $i$ in country $j$, and $\mathbf{e}_{ij}$ is a vector of residual effects. Model [1] is also valid for data from countries with a single-trait model in national genetic evaluations, with all terms above becoming scalars. The (co)variance matrix of genetic effects of the $m$ countries and its inverse are denoted as

$$\mathbf{G}_0 = \begin{bmatrix} \mathbf{G}_{0_{11}} & \mathbf{G}_{0_{12}} & \dots & \mathbf{G}_{0_{1m}} \\ & \mathbf{G}_{0_{22}} & \dots & \mathbf{G}_{0_{2m}} \\ & & \ddots & \vdots \\ symm. & & & \mathbf{G}_{0_{mm}} \end{bmatrix}, \qquad [2]$$

$$\text{and } \mathbf{G}_0^{-1} = \begin{bmatrix} \mathbf{G}_0^{11} & \mathbf{G}_0^{12} & \dots & \mathbf{G}_0^{1m} \\ & \mathbf{G}_0^{22} & \dots & \mathbf{G}_0^{2m} \\ & & \ddots & \vdots \\ symm. & & & \mathbf{G}_0^{mm} \end{bmatrix}$$

where $m$ is the number of countries, $\mathbf{G}_{0_{jj}}$ is the original genetic (co)variance matrix of country $j$, and $\mathbf{G}_{0_{jk}}$ is the genetic covariance matrix between countries $j$ and $k$. When country $j$ uses a single-trait model for national evaluation, $\mathbf{G}_{0_{jj}}$ is a scalar and all corresponding off-diagonal blocks, $\mathbf{G}_{0_{jk}}$ ($j \neq k$), become vectors or scalars, depending on the number of traits in other countries. As usually considered in MACE, data of a bull from different countries are assumed to be residually uncorrelated. The inverse of error (co)variance matrix of bull $i$ in country $j$ is

$$[\text{Var}(\mathbf{e}_{ij})]^{-1} = \boldsymbol{\Psi}_{ij} \qquad [3]$$

where $\boldsymbol{\Psi}_{ij}$ is the EDC matrix for bull $i$ in country $j$, converted from the reliability matrix contributed by his daughters' records in the $j$th country. The MTEDC procedure (Liu et al., 2004a) can be used to approximate matrix $\boldsymbol{\Psi}$ for each bull.

### *The Mixed Model Equations*

The mixed model equations (**MME**) of model [1] consist of equations for additive genetic effects of bulls and fixed effects of country means. Ignoring pedigree contributions, the equations corresponding to bull $i$ are