CONSIDERATIONS ON THE PATH FROM TEST-DAY RECORDS TO NATIONAL AND INTERNATIONAL GENETIC EVALUATIONS AND ITS CONSEQUENCES

N. Gengler[1,2]

[1] Gembloux Agricultural University, B-5030 Gembloux, Belgium
[2] National Fund for Scientific Research, B-1000 Brussels, Belgium

INTRODUCTION

Direct use of test-day yields in genetic evaluations has been a major focus of research in recent years and their are numerous advantages associated with this approach as detailed by many authors (e.g., Swalve 2000). Despite considerable recent research there seems to be some confusion among people how to consider test-day records and their relationship to full lactation records. This is however a major issue if we have to consider the combination of both types of records in national or international evaluations. The objective of this report was to summarize considerations on the path from test-day records to genetic evaluations with and without the intermediate step of estimation lactation records in order to provide an underlying framework for all evaluation models.

GENERAL MODEL

With the apparent differences between genetic evaluation systems based either on test-day or full or total (mostly on a 305 day basis) lactation yields it is often very difficult to find common ground. Most methods can however be classified as being based on a general model where only the nature of the functions, effects, statistical distributions and the manner of solving are different. We assume here that the general model could be as follows modifying the two-stage notation used by Jamrozik *et al.* (2001) where every stage can be uni- or multivariate:

Stage 1: $$\mathbf{f}\left(\mathbf{y}_{TD_i}\right) = \mathbf{H}_i\mathbf{h}_i + \mathbf{g}(\boldsymbol{\theta}_i, \mathbf{t}_i) + \boldsymbol{\varepsilon}_i \quad [1a]$$

Stage 2: $$\begin{bmatrix} \boldsymbol{\theta}_i \\ \boldsymbol{\Phi}(\boldsymbol{\theta}_i) \end{bmatrix} = \mathbf{X}_i\mathbf{b}_i + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i \quad [1b]$$

where $\mathbf{f}\left(\mathbf{y}_{TD_i}\right)$ is a function of test day records for a cow i; $\mathbf{h}_i$ is a vector of test-day related effects; $\mathbf{g}(\boldsymbol{\theta}_i, \mathbf{t}_i)$ is function of cow specific parameters $\boldsymbol{\theta}_i$ and time (days in milk at test) $\mathbf{t}_i$; $\boldsymbol{\varepsilon}_i$ is a vector of residuals on a test-day level; $\mathbf{H}_i$ is an incidence matrix relating $\mathbf{f}\left(\mathbf{y}_{TD_i}\right)$ to $\mathbf{h}_i$; and $\left[\boldsymbol{\theta}_i' \quad \boldsymbol{\Phi}(\boldsymbol{\theta}_i)'\right]'$ is a vector of cow specific parameters $\boldsymbol{\theta}_i$ or functions of those parameters for cow i where a typical function would be the full lactation yield; $\mathbf{b}_i$ vector of cow related fixed effects for cow i; $\mathbf{u}_i$ is a vector of cow related random effects for cow i; $\mathbf{e}_i$ is a vector of

lactation related residuals for cow i; $\mathbf{X}_i$ and $\mathbf{Z}_i$ are incidence matrices linking $[\boldsymbol{\theta}_i' \quad \boldsymbol{\Phi}(\boldsymbol{\theta}_i)']'$ with $\mathbf{b}_i$ and $\mathbf{u}_i$.

The assumptions about the detailed nature of the functions, effects, statistical distributions and the manner of solving will allow us to show how out of this general model the different type of evaluation systems can be developed.

GENETIC EVALUATIONS BASED ON LACTATION YIELDS

Current lactation based genetic evaluation systems follow the formulas given in [1a] and [1b] very well. Their first stage consists of the computation; or better estimation; of full 305 day lactation records and the second stage consists of a genetic model on those records.

Test Interval and Centering Date Methods. Very few assumptions can lead directly to the test-interval (TIM) or centering date methods (CDM) where [1a] becomes a weighted linear model and the weights represent the contribution of each test-day yield has to full of total yield (305 x mean) for the given lactation :

Stage 1: $$\mathbf{y}_{TD\,i} = \mathbf{L}_i\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i \qquad [2a]$$

because: $\mathbf{f}(\mathbf{y}_{TD\,i}) = \mathbf{y}_{TD\,i}$, $\mathbf{H}_i\mathbf{h}_i = \mathbf{0}$, $\mathbf{g}(\boldsymbol{\theta}_i, \mathbf{t}_i) = \mathbf{L}_i\boldsymbol{\lambda}_i$, $\mathbf{L}_i = [1/305 \quad ... \quad 1/305]'$ and $\boldsymbol{\theta}_i = \lambda_i$ with $V(\mathbf{y}_{TD\,i}) = V(\boldsymbol{\varepsilon}_i) = \mathbf{W}_i^{-1}\sigma_\varepsilon^2$ where $\mathbf{W}_i$ is a diagonal matrix with the weights for every test-day of cow i (in a given lactation). The weights used are equal to the number of days in milk associated with this record for CDM or half of the days of the interval before and half of the days after for TIM. Then the following well-known genetic model is used on lactation records:

Stage 2: $$\boldsymbol{\lambda}_i = \mathbf{y}_{LY_i} = \mathbf{X}_i\mathbf{b}_i + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i \qquad [2b]$$

as $[\boldsymbol{\theta}_i' \quad \boldsymbol{\Phi}(\boldsymbol{\theta}_i)']'$ is replaced by the estimated lactation yields $\mathbf{y}_{LY_i}$ of a cow i. In most situations solving of [2b] is done adjusting equations for the precision of the estimates of lactation yields. Some systems also adjust for heterogeneous phenotypic variances, some expand the variances of the estimated lactation yields to account for the lost of variance due to the estimation process.

Multiple trait prediction. Alternative methods were developed to replace TIM and CDM. An interesting example is the multiple trait prediction (MTP) method by Schaeffer and Jamrozik (1996). They used a model where logarithms of test-day data were modeled for milk, fat and protein assuming Wood's model:

Stage 1: $$\ln(\mathbf{y}_{TD_i}) = \mathbf{K}_i\mathbf{c}_i + \boldsymbol{\varepsilon}_i \qquad [3a]$$

because: $\mathbf{f}(\mathbf{y}_{TD\,i}) = \mathbf{y}_{TD\,i}$, $\mathbf{H}_i\mathbf{h}_i = \mathbf{0}$, $\mathbf{g}(\boldsymbol{\theta}_i, \mathbf{t}_i) = \mathbf{K}_i\mathbf{c}_i$ .where for every test-day j the incidence matrix $\mathbf{K}_{ij}$ is written as $\mathbf{I}_3 \otimes [1 \quad \ln(t_j) \quad t_j]$ and $\otimes$ is the Kronecker product.

A specificity of the MTP method is that the equations are solved assuming $E(\mathbf{c}_i) = \mathbf{c}_0$ and $V(\mathbf{c}_i) = \mathbf{G}$ where $\mathbf{c}_0$ is vector of expected coefficients for all cow with the same production characteristics. Doing this regresses the solutions for every cows towards their expected values or in a Bayesian sense combines it with a priori information. After solving the equation full lactation (305 day) yields are computed from the estimated lactation curve coefficients for Wood's model.

Best prediction. VanRaden (1997) proposed another method as replacement of TIM or CDM in Stage 1. His idea was to use multiple regression techniques to obtain estimates for the yields at non-observed test-days and called it Best Prediction (BP). One characteristic is that the assumed variance structures regressed the unknown test-days and computed the lactation yield as the sum of unregressed known and regressed unknown test-days. In reality this method does therefore only interpolate the missing test-days. The question if observed test-days should also be regressed could be considered nearly a philosophical one. However it should be noticed that other methods such as MTP do assume this and regresses also known test-days towards the expected values.

Other genetic systems making indirect use of test-day yields. Several other evaluation systems exist that use similar two step strategies computing lactation parameters (generally 305 day) yields and then combine them. Examples are the AI Multiple Trait Cow and Sire Evaluations by Cornell University (e. g., ABC, 2002), the Australian (e.g., Jones and Goddard, 1990) and New Zealand genetic evaluation systems (e. g., Johnson, 1996).

GENETIC EVALUATIONS BASED DIRECTLY ON TEST-DAY YIELDS

Different types of test-day models (e. g., Swalve, 2000) can be defined, we consider here those based directly on test-day yields. All the current test-day model evaluation systems can be expressed under a general form that combines Stage 1 and Stage 2.

Stages 1 + 2: $$\mathbf{y}_{TD_i} = \mathbf{H}_i\mathbf{h}_i + (\mathbf{Q}_{\mathbf{b}_i}\mathbf{b}_i + \mathbf{Q}_{\mathbf{u}_i}\mathbf{u}_i + \mathbf{Q}_{\mathbf{p}_i}\mathbf{p}_i) + \boldsymbol{\varepsilon}_i \quad [5]$$

where the incidence matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ are replaced by matrices of regression coefficients $\mathbf{Q}_{\mathbf{b}_i}$ and $\mathbf{Q}_{\mathbf{u}_i}$ and the residual on lactation level is replaced by $\mathbf{Q}_{\mathbf{p}_i}\mathbf{p}_i$. This later point is important in the sense that it shows that the lactation residuals can in fact be considered linked to the permanent environmental effect on a test-day level.

Test-Day Models with and without cow specific regression effects. The nature of the different effects in [5] defines the type of test-day model. The German (VIT) incomplete test-day model (Reents *et al*., 1995) does not allow cow specific random regressions. The genetic and permanent environmental effects in a lactation are considered constant over the whole lactation. The Canadian (Schaeffer *et al*., 2000) and Finnish (Lidauer *et al*., 2000) models however allow the modeling of the cow specific lactation shape functions. This allows for genetic and permanent environmental differences in the shape of the lactation curves.

Alternative solving methods. Gengler *et al.* (2000) showed that iterative solving of a two stage model achieved rapidly nearly identical solutions to the equivalent full test-day models. Reciprocally this means that at least some types of test-day models that can be split up in two stages could be solved by iterative solving of both stages adding increased flexibility to modeling those stages and to allow use of larger data sets.

COMBINING TEST-DAY AND FULL LACTATION RECORDS

National evaluations. In a lot of countries there is a need to optimally use existing older full lactation records. The most logical way of doing this would be to define a two stage model which combine those two types of records in the second stage. Iterative solving of both stages could achieve near BLUE and BLUP properties for all effects.

International evaluations. Current international genetic evaluation is adding an additional third stage on an international level. This stage is using a very much simplified approach based on deregressed sire breeding values and an multivariate sire model (Schaeffer, 1994). Alternatives are under development that try to make more or less direct use of records. Those alternatives will need to address very quickly the problem of combining test-day and full lactation records and this for situations in very large populations.

CONCLUSION

Understanding the underlining framework linking all present genetic evaluation systems for milk traits is extremely useful as it gives us the possibility to solve issues linked to the use of optimal genetic evaluation system also in extremely large populations and complicated situations as in future national and international genetic evaluation systems.

REFERENCES

ABC (2002) http://www.ansci.cornell.edu/abc/siretabs/about.html
Gengler, N., Tijani, A. and Wiggans, G.R. (2000) *J. Dairy Sci.* **83** : february online.
Jamrozik, J., Gianola, D. and Schaeffer, L.R. (2001) *Livest. Prod. Sci.* **71** : 223-240.
Johnson, D.L. (1996) *Proc. N.Z. Soc. Anim. Prod*. **56** :16.
Jones, L.P. and Goddard, M.E. (1990) *Proc 4th WCGALP* **13** : 382.
Lidauer, M., Mäntysaari, E.A., .Strandén, I. and Pösö, J. (2000) *Interbull Bulletin* **25** : 81-86.
Reents, R., Dekkers, J.C.M. and Schaeffer, L.R. (1995) *J. Dairy Sci.* **78** : 2858-2870.
Schaeffer, L.R. (1994) *J. Dairy Sci*. **77** : 2671-2678.
Schaeffer, L.R. and Jamrozik, J. (1996) *J. Dairy Sci.* **79** : 2044-2055.
Schaeffer, L.R., Jamrozik, J., Kistemaker, G.J. and Van Doormaal, B.J. (2000) *J. Dairy Sci*. **83** : 1135-1144.
Swalve, H.H. (2000) *J. Dairy Sci.* **83** : 1115-1124.
VanRaden, P.M. (1997) *J. Dairy Sci.* **80** : 3015-3022.