

Joint Nordic Test Day Model: Evaluation Model

M. Lidauer¹, J. Pedersen², J. Pösö³, E. A. Mäntysaari¹, I. Strandén¹, P. Madsen⁴, U.S. Nielsen², J.-Å. Eriksson⁵, K. Johansson⁵, G.P. Aamand⁶

¹MTT Agrifood Research Finland, Biotechnology and Food Science, Biometrical Genetics, Jokioinen, Finland;

²The Danish Agricultural Advisory, Aarhus, Denmark; ³Faba Breeding, Vantaa, Finland; ⁴Danish Institute of Agricultural Sciences, Genetics and Biotechnology, Foulum, Denmark; ⁵Swedish Dairy Association, Stockholm, Sweden; ⁶NAV Nordic Cattle Genetic Evaluation, Aarhus, Denmark.

Abstract

This paper is one of three papers describing the joint Nordic test day model for yield traits. The focus of this paper is on the evaluation models for Red Breeds, Holstein, and Jersey. Environmental effects and adjustment of heterogeneous variance are illustrated.

1. Introduction

In April 2006, the joint evaluation for yield traits of Danish, Finnish and Swedish dairy cattle was taken into official use. The new evaluation is carried out by Nordisk Avelsværdivurdering (NAV), a company responsible for the genetic evaluation across the three countries. The new NAV random regression model (RRM) replaces the national evaluation models for yield traits.

In the new NAV evaluation Denmark and Finland use test-day (TD) yield records, whereas Sweden continues with 305-day (305d) yield records. Mäntysaari (2002) presented a model that incorporates 305d yield records into a RRM. Based on this meta-model idea (Mäntysaari, 2006b), the NAV RRM was designed to allow different variance components and heritabilities across countries for the same biological traits. This is different from across-country evaluations where the traits are assumed to be the same across countries (i.e. Emmerling et al, 2002).

The Nordic *meta-model* for the random animal effects is outlined by Mäntysaari et al. (2006a). In this work we describe the environmental effects of the NAV RRM and the method for adjustment of heterogeneous variance. An analysis of the NAV breeding values is given in Pösö et al. (2006).

2. Material and methods

2.1 Data

TD data and 305d data from all dairy cattle of the three countries are merged into three data sets, one for Red Breeds, one for Holstein and one for Jersey. Holstein is the largest breed (Table 1), whereas the Red Breeds are the most heterogeneous, originating from eight different breeds. The Jersey breed is only located in Denmark with few herds in Sweden. Finnish Ayrshire and Holstein data are included into both, Red Breeds and Holstein data, due to many small mixed herds. For the same reason, all available lactations are included for the Finnish data. Data of the Finncattle is included in the evaluation of Red Breeds.

Table 1. Size of the evaluation data (in million).

	Red Breeds	Holstein	Jersey
Animals	4.05	6.64	0.59
TD yields	45.66	81.61	7.21
305d-yields	1.93	1.60	-

2.2 Evaluation model for Red Breeds

The NAV RRM is a multiple trait model that describes the biological traits milk, protein and fat yield. A multiple trait model was preferred because in Finland protein and fat yield are measured only every second

TD. The 1st, 2nd, and 3rd plus later parities as well as the countries Finland, Denmark, and Sweden are modeled as different traits. Hence, the model includes 27 traits. For simultaneous accounting of heterogeneous variance a multiplicative mixed model (Meuwissen et al., 1996) is used, and has the form:

$$\Lambda \mathbf{y} = \mathbf{Xb} + \mathbf{Kh} + \mathbf{Ls} + \mathbf{Md} + \mathbf{Rf} + \mathbf{Qg} + \mathbf{Z_a a} + \mathbf{Z_p p} + \mathbf{Z_w w} + \mathbf{e} \quad [1]$$

where Λ is a diagonal matrix with multiplicative adjustment factors and $\mathbf{y}$ is the

$$\begin{bmatrix} \Lambda_{Tc1t} \mathbf{y}_{Tc1t} \\ \Lambda_{L1t} \mathbf{y}_{L1t} \\ \Lambda_{Tc2t} \mathbf{y}_{Tc2t} \\ \Lambda_{L2t} \mathbf{y}_{L2t} \\ \Lambda_{Tc3t} \mathbf{y}_{Tc3t} \\ \Lambda_{L3t} \mathbf{y}_{L3t} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{Tc1t} \mathbf{b}_{Tc1t} \\ \mathbf{X}_{L1t} \mathbf{b}_{L1t} \\ \mathbf{X}_{Tc2t} \mathbf{b}_{Tc2t} \\ \mathbf{X}_{L2t} \mathbf{b}_{L2t} \\ \mathbf{X}_{Tc3t} \mathbf{b}_{Tc3t} \\ \mathbf{X}_{L3t} \mathbf{b}_{L3t} \end{bmatrix} + \begin{bmatrix} \mathbf{K}_{Tc1t} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{L1t} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{K}_{Tc2t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{K}_{L2t} \\ \mathbf{0} & \mathbf{0} & \mathbf{K}_{Tc3t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{K}_{L3t} \end{bmatrix} \begin{bmatrix} \mathbf{h}_{Tc1t} \\ \mathbf{h}_{L1t} \\ \mathbf{h}_{Tc2t} \\ \mathbf{h}_{L2t} \end{bmatrix} + \begin{bmatrix} \mathbf{L}_{Tc1t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{Tc1t} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{Tc3t} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{s}_{Tc1t} \\ \mathbf{s}_{Tc2t} \end{bmatrix} + \begin{bmatrix} \mathbf{M}_{Tc1t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{Tc2t} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{Tc3t} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{d}_{Tc1t} \\ \mathbf{d}_{Tc2t} \end{bmatrix} + \begin{bmatrix} \mathbf{R}_{Tc1t} & \mathbf{0} & \mathbf{0} \\ \mathbf{R}_{L1t} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{Tc2t} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{L2t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{Tc3t} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{L3t} \end{bmatrix} \begin{bmatrix} \mathbf{f}_{1t} \\ \mathbf{f}_{2t} \\ \mathbf{f}_{3t} \end{bmatrix} + \begin{bmatrix} \mathbf{Q}_{Tc1t} \mathbf{g}_{Tc1t} \\ \mathbf{Q}_{L1t} \mathbf{g}_{L1t} \\ \mathbf{Q}_{Tc2t} \mathbf{g}_{Tc2t} \\ \mathbf{Q}_{L2t} \mathbf{g}_{L2t} \\ \mathbf{Q}_{Tc3t} \mathbf{g}_{Tc3t} \\ \mathbf{Q}_{L3t} \mathbf{g}_{L3t} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{aTc1t} & \mathbf{0} \\ \mathbf{Z}_{aL1t} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{aTc2t} \\ \mathbf{0} & \mathbf{Z}_{aL2t} \\ \mathbf{0} & \mathbf{Z}_{aTc3t} \\ \mathbf{0} & \mathbf{Z}_{aL3t} \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_{23} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{pTc1t} \mathbf{p}_{Tc1} \\ \mathbf{Z}_{pL1t} \mathbf{p}_{L1} \\ \mathbf{Z}_{pTc2t} \mathbf{p}_{Tc2} \\ \mathbf{Z}_{pL2t} \mathbf{p}_{L2} \\ \mathbf{Z}_{pTc3t} \mathbf{p}_{Tc3} \\ \mathbf{Z}_{pL3t} \mathbf{p}_{L2} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{Z}_{wTfi3t} \mathbf{w}_{Tfi3} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{Tc1t} \\ \mathbf{e}_{L1t} \\ \mathbf{e}_{Tc2t} \\ \mathbf{e}_{L2t} \\ \mathbf{e}_{Tc3t} \\ \mathbf{e}_{L3t} \end{bmatrix}, \quad [2]$$

where subscript T stands for TD yields, subscript c for the TD yield country (either Denmark or Finland), subscript L stands for 305d-yields from Sweden, subscript *fi* for Finland, subscript 1, 2, and 3 stands for the parities, and subscript t for the biological trait milk, protein, and fat yield. TD yields from Denmark and Finland are modeled as different traits, having same effects in the model but different variance components. Vector $\mathbf{b}$ in [1] and [2] includes fixed effects that are specific to each country and parity, where Finnish 5th and later parities form one parity class. The effects in $\mathbf{b}$ are:

Season effect is modeled as production year×month for TD traits. This effect is necessary because of the small herd size, which requires a random herd-test-day

vector with observations. Vector $\mathbf{b}$, $\mathbf{h}$, $\mathbf{s}$, $\mathbf{d}$, $\mathbf{f}$ and $\mathbf{g}$ contain the environmental effects, and vector $\mathbf{a}$, $\mathbf{p}$ and $\mathbf{w}$ the random animal effects. Matrices $\mathbf{X}$, $\mathbf{K}$, $\mathbf{L}$, $\mathbf{M}$, $\mathbf{R}$, $\mathbf{Q}$, $\mathbf{Z}_a$, $\mathbf{Z}_p$, and $\mathbf{Z}_w$ are the corresponding design matrices and vector $\mathbf{e}$ contains the residuals. The model has the same structure for each biological trait but differs across parities and countries. Model [1] can be expanded to present the structure for one biological trait (t) as well as differences in modeling TD yields and 305d yields:

effect. For 305d traits a calving year×month is defined.

Calving age×breed is modeled by a linear and quadratic regression on calving age. The calving age is centered on mean calving age and positive and negative extreme values are discarded. To account for breed differences, additional calving age curves are defined. These are modeled as linear and quadratic regression on calving age times breed proportion. Such curves are modeled for the breeds Brown Swiss and Holstein in Danish traits, and for the breeds Finncattle and Holstein in Finnish traits.

Days open and *calving interval* are modeled for 305d traits as a linear regression on deviation from average days open, and as preceding calving interval from 2nd parity

onwards. Both effects are nested within 5-year time periods.

Days carried calf and *days dry* are defined for TD traits, where days dry is modeled as previous days dry from 2nd parity onwards. Both effects are nested within 5-year time periods.

Stage of lactation is modeled with four regressions on days in milk (DIM) nested within calving age, production month, and 4-year production year period. Three calving age classes were used. The corresponding covariables c , c^2 , w_1 , and w_2 represent the linear and quadratic terms of a Legendre polynomial and two Wilmink exponential terms, where $w_1 = \exp(p_1 \text{DIM})$ and $w_2 = \exp(p_2 \text{DIM})$. Parameter p_1 is -0.05 for milk and -0.04 for protein and fat, as was found optimal for modeling the random animal effects (Mäntysaari et al., 2006a). Parameter p_2 was estimated for each trait to ensure the best function fit. Values for p_2 range from -0.22 to -0.01. Note, that the covariables do not include an intercept since it is modeled by the effect of season.

Effects modeled by the remaining vectors in [1] and [2] were:

Herd effects are included in the vectors **h**, **s** and **d**, where **h** contains the fixed herd-year effects, vector **s** contains a fixed linear regression effect on DIM nested within herd and 5-year time period, and vector **d** contains a random herd-test-day effect. For all three effects, two parity classes are considered, one for first parity and another one for all other parities.

Heterosis and *recombination loss* are modeled the same way and therefore, only heterosis effects are explained. The across-country evaluation requires a careful modeling of crossbreeding effects. Within the Red Breeds twelve different breeds, with significant breed contributions, are considered. The model includes a fixed regression on the total sum of heterosis to account for all heterosis affecting the yield of a cow. This effect is modeled across countries (see vector **f** in [2]). Additionally,

within each country, five random heterosis effects (vector **g**) are included, one for each of the five most important crosses. A correlation structure is imposed between same random heterosis effects across countries to ensure that genetic levels of countries are not impaired by unreliable estimates of heterosis. A detailed description of modeling the cross breeding effects is given in Lidauer et al. (2006).

Random animal effects are modeled by a *meta-model* (Mäntysaari, 2006b). It exploits different variance components and heritabilities across countries. Even countries are considered as different traits, the genetic correlation across countries is set to unity to ensure for each animal a unique set of 14 breeding values coefficients (in vector **a**). For TD traits, vector **p** models non-genetic animal effects across parities, and vector **w** models non-genetic animal effects within later parities for Finnish 3rd and later parity observations. A cow with observations enters **p** with 18 coefficients and **w** with 6 coefficients for each later parity. A closer look of the random animal effects and applied variance components is given in Mäntysaari et al. (2006a).

2.3 Evaluation models for Holstein and Jersey

The Holstein evaluation model has the same setup as the Red Breeds except some differences in the environmental effects. These differences are: a) the calving age effect in vector **b** does not account for breed differences. b) for Danish traits, the fixed herd-year effect in vector **h** is removed and the herd-test-day effect in vector **d** is fixed. c) the fixed regressions on the total sum of heterosis (in vector **f**) are nested within countries. There are neither random heterosis nor recombination loss effects defined in the Holstein model. The model for the Jersey evaluation includes only nine traits and these traits are modeled the same way as the Danish traits in the Holstein model.

2.4 Accounting for heterogeneous variance

The simultaneous solving of the evaluation model and of the variance-model is carried out using a similar procedure as described by Lidauer et al. (2002). The model for the dispersion parameters includes a fixed country×year×month×parity effect and a random herd-year×parity group effect. An autoregressive variance structure is assumed between herd-years. There are two parity groups, one for 1st parity and another for later parities. Variance components for the variance-model were estimated for all 27 traits and for all breeds. For simplicity, traits are assumed to be uncorrelated in the variance-model. Applied parameters for the variance ratio (herd-year variance) / (residual variance) ranges between 0.08 and 0.63. The applied autoregressive correlation between herd-years is 0.90 for 1st parity traits and 0.92 for later parity traits.

While setting up the model, an across-country standardization procedure was applied. This iterative procedure calibrates the levels of the adjustment factors for each country until the approximated re-estimates of the genetic variances are the same in each country. Re-estimation of genetic variances was based on Mendelian sampling terms and approximated reliabilities for Mendelian sampling terms (Fikse et al., 2003). The standardization procedure was necessary because of the differences in variance components applied across countries.

3. Results and discussion

Solving the models is a computational challenge. Models are largest for Holstein with currently 228 million unknowns for the evaluation model and 9.7 million unknowns for the variance-model. With parallel computing (Strandén and Lidauer 2001) solving the multiplicative NAV RRM on 6 Intel Xeon 2.8 GHz cpu's requires 4.2 days of computations for Holstein.

The chosen strategy for modeling cross breeding effects gave good results. The

common across-country heterosis estimates for milk yield were 2.64, 2.76, and 2.75 % of the phenotypic mean for 1st, 2nd, and 3rd parity, respectively. In Table 2 the random estimates for two important breed crosses are summed to the common estimates. It shows, that estimates are different for different crosses but very similar within same crosses across countries.

Table 2. Heterosis estimates (in % of phenotypic mean) for crosses between Red Danes (RDM), Swedish Red (SRB) and Finnish Ayrshire (FAY) for milk yield given for Danish (*dk*), Swedish (*se*) and Finnish (*fi*) traits.

Traits	Cross	Lactation		
		1.	2.	3.
<i>dk</i>	RDM×SRB	4.40	3.86	2.90
<i>se</i>	SRB×RDM	4.60	3.72	2.93
<i>se</i>	SRB×FAY	2.60	2.88	2.18
<i>fi</i>	FAY×SRB	2.11	2.47	1.68

The heterogeneous variance adjustment ensures that re-estimated genetic variances for the NAV indices are the same for each country (Table 3, Figure 1). A NAV index is formed from the 305d breeding values of 1st, 2nd, and 3rd parity weighted by 0.5, 0.3, and 0.2, respectively. This guarantees that the standard deviation of sire breeding values in each country is independent from phenotypic variation or model variance parameters and is determined by the accuracy only. This was one important requirement set by breeders before accepting the model. However, due to differences in heritabilities and reliabilities standard deviations of cow breeding values vary across countries, as shown in Figure 2.

Table 3. Re-estimated genetic standard deviation for NAV index (in kg), obtained from Red Breed sires with at least 20 daughters.

	Denmark	Finland	Sweden
Milk	568.6	568.9	565.4
Protein	15.0	14.9	14.9
Fat	24.0	24.0	23.9

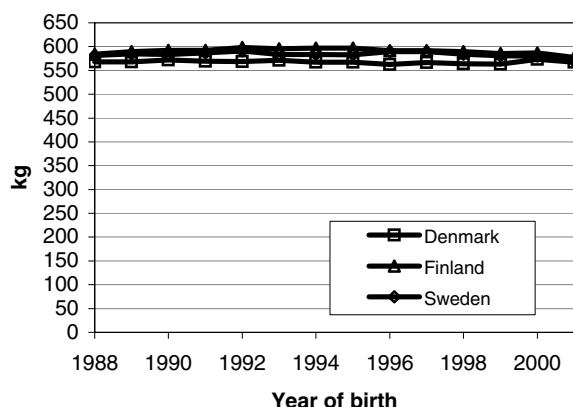


Figure 1. Re-estimated genetic standard deviation for NAV milk index, obtained from Red Breed cows.

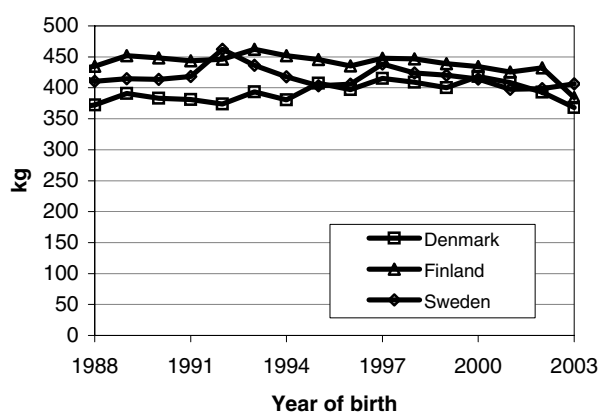


Figure 2. Standard deviation of NAV milk index for Red Breed cows.

4. Conclusions

Ideally an across-country evaluation based on raw data and animal model is preferable compared to MACE (Schaeffer, 1994). However, complexity of such a model bears its own problems. Along model developments unforeseen problems were aroused. One was the sensitivity of across-country evaluation to modeling of cross breeding effects. Another was to find a correct across-country adjustment procedure for heterogeneous variance. Solving of both problems paved the road towards official implementation of the Nordic yield model.

References

- Emmerling, R., Lidauer, M. & Mäntysaari, E.A. 2002. Multiple lactation random regression test-day model for Simmental and Brown Swiss in Germany and Austria. *Interbull Bulletin*. 29, 111-117.
- Fikse, W.F., Klei, L., Liu, Z. & Sullivan, P.G. 2003. Procedure for validation of trends in genetic variances. *Interbull Bulletin*. 31, 30-36.
- Lidauer, M., Mäntysaari, E.A., Strandén, I., Pösö, J., Pedersen, J., Nielsen, U.S., Johansson, K., Eriksson, J.-Å., Madsen, P. & Aamand, G.P. (2006). Random heterosis and recombination loss effects in a multibreed evaluation for Nordic red dairy cattle. *Proc. 8th WCGALP*. (accepted)
- Lidauer, M., Emmerling, R. & Mäntysaari, E.A. 2002. Accounting for heterogeneous variance in a test-day model for joint genetic evaluation of Austrian and German Simmental cattle. *Proc. 7th WCGALP*, Montpellier, Aug 19-23, 2002, p.529-532.
- Meuwissen, T.H.E., De Jong, G., Engel, B., 1996. Joint estimation of breeding values and heterogeneous variance for large data files. *J. Dairy Sci.* 79, 310-316.
- Mäntysaari, E.A., 2002. Combining test day and full lactation records in prediction of breeding values. *Proc. 7th WCGALP*. Montpellier, Aug 19-23, 2002, p.35-50.
- Mäntysaari, E.A., Lidauer, M., Pösö, J., Madsen, P., Strandén, I., Pedersen, J., Nielsen, U.S., Johansson, K., Eriksson, J.-Å. & Aamand, G.P. 2006a. Joint Nordic test day model: variance components. *Interbull Bulletin*.
- Mäntysaari, E.A., 2006b. Meta-model combines traits with different genetic models and data structures. *Proc. 8th WCGALP*. (accepted)
- Pösö, J., Pedersen, J., Lidauer, M., Mäntysaari, E.A., Strandén, I., Madsen, P., Nielsen, U.S., Eriksson, J.-Å., Johansson, K. & Aamand, G.P. 2006. Joint Nordic test day model: experience with the new model. *Interbull Bulletin*.
- Schaeffer, L.R. 1994. Multiple-country comparison of dairy sires. *J. Dairy Sci.* 77, 2671-2678.
- Strandén, I. & Lidauer, M., 2001. Parallel computing applied to breeding value estimation in dairy cattle. *J. Dairy Sci.* 84, 276-285.