



HAL
open science

An approximate multitrait model for genetic evaluation in dairy cattle with a robust estimation of genetic trends

Jan Lassen, Morten Sorensen, Per Madsen, Vincent Ducrocq

► To cite this version:

Jan Lassen, Morten Sorensen, Per Madsen, Vincent Ducrocq. An approximate multitrait model for genetic evaluation in dairy cattle with a robust estimation of genetic trends. *Genetics Selection Evolution*, 2007, 39, pp.353-367. hal-02656502

HAL Id: hal-02656502

<https://hal.inrae.fr/hal-02656502>

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

An approximate multitrait model for genetic evaluation in dairy cattle with a robust estimation of genetic trends

(Open Access publication)

Jan LASSEN^{a,b*}, Morten Kargo SØRENSEN^a, Per MADSEN^a,
Vincent DUCROCQ^c

^a Department of Genetics and Biotechnology, Danish Institute of Agricultural Sciences,
P.O. Box 50, DK-8830 Tjele, Denmark

^b Department of Large Animal Sciences, The Royal Veterinary and Agricultural University,
Ridebanevej 12, DK-1870 Frederiksberg C, Denmark

^c Station de génétique quantitative et appliquée, UR 337, INRA, 78352 Jouy-en-Josas, France

(Received 6 November 2006; accepted 21 January 2007)

Abstract – In a stochastic simulation study of a dairy cattle population three multitrait models for estimation of genetic parameters and prediction of breeding values were compared. The first model was an approximate multitrait model using a two-step procedure. The first step was a single trait model for all traits. The solutions for fixed effects from these analyses were subtracted from the phenotypes. A multitrait model only containing an overall mean, an additive genetic and a residual term was applied on these preadjusted data. The second model was similar to the first model, but the multitrait model also contained a year effect. The third model was a full multitrait model. Genetic trends for total merit and for the individual traits in the breeding goal were compared for the three scenarios to rank the models. The full multitrait model gave the highest genetic response, but was not significantly better than the approximate multitrait model including a year effect. The inclusion of a year effect into the second step of the approximate multitrait model significantly improved the genetic trend for total merit. In this study, estimation of genetic parameters for breeding value estimation using models corresponding to the ones used for prediction of breeding values increased the accuracy on the breeding values and thereby the genetic progress.

stochastic simulation / multitrait model / genetic evaluation

1. INTRODUCTION

Most dairy cattle breeding programmes rely on multitrait selection, where predicted breeding values (EBV) for individual traits in the breeding goal are

* Corresponding author: jan.lassen@agrsci.dk

combined according to their economic values. EBV for the traits are frequently obtained from single trait models or in some cases from multitrait models for groups of traits. Over recent years, more and more focus has been put on lowly heritable traits leading to breeding values with low accuracies and thereby instability in breeding values for these traits [13]. A way to increase the accuracy of predicted breeding values for all traits in the breeding goal is to use multivariate methods for prediction of breeding values in order to fully exploit the data and to combine direct and indirect information on correlated traits in the breeding goal [21, 22]. For most breeding programmes with dairy cattle, a full multivariate prediction of breeding values is not computationally feasible due to the number of animals and number of traits involved [6]. An approximate method has been proposed by [8] analysing records adjusted for fixed effects so that each trait has one observation containing an overall mean, a genetic term and a residual term for all animals in a multitrait setting. This method has been shown to work with both longevity [18] and repeated observations [8]. The inclusion of breeding values from test day models into this approximate multitrait model is possible using MACE (multiple-trait across country genetic evaluation) methodology [19]. The reliability for longevity on newly proven bulls increased from 0.40 to 0.58 using this method. Also the selection differential for the 10% best bulls increased for SCC (somatic cell count), functional longevity and female fertility when comparing the approximate model with the model used previously. In this approach single trait models were used for each trait in the breeding goal and EBV were weighted together afterwards. This result though was a mixture of both changing the model and the breeding goal at the same time. Applying this approximate multitrait method in laying hens gave similar results as shown by [8]. It increased the selection differential for the 10% best animals for all traits in the breeding goal except precocity as compared to using a single trait model for all traits [1]. The method has also been shown to be superior in terms of genetic progress compared to a single trait approach [11].

Another benefit from multitrait evaluation including all traits in the breeding goal is that selection bias is considered [10]. The fact that the main selection in dairy cattle for many years has been on production, will lead to biased predicted breeding values when single trait analysis is used since genetic correlations are not accounted for. This will show up as bias in estimated genetic trends [14].

In the PROTEJE project [2] it has been shown that including an effect of time (year) of performance in the MACE model used for international

prediction of breeding values for bulls could account for selection bias in the national evaluations and thereby lead to more robust evaluations [9].

A drawback of multitrait evaluation is that the correlations used are only estimates, and they are often estimated with high uncertainty. Studies have shown that response to selection depends highly on the precision of the estimated and applied variance components [23]. Therefore, sometimes univariate models can provide more precise estimates than multitrait models.

The aim of this study was to augment the approximate multitrait method proposed by [4, 8] with a time effect to account for selection bias in predicted breeding values and variance components for the individual traits in national breeding value estimation. The augmented method was tested using stochastic simulation of a dairy cattle population.

2. MATERIALS AND METHODS

2.1. Outline of the simulation

The simulation was split into two steps each covering 15 years of selection. The idea behind this setup was in the first step to create a population that had undergone selection for a number of years before entering an experimental stage. In the second step, three methods of interest for prediction of breeding values were compared. Between the two steps, (co)-variance components for the traits of interest were estimated and used for prediction of EBV in the last step. This corresponds to a situation that often occurs in practical breeding programmes, where (co)-variance components are re-estimated, and new methods or models for prediction of EBV are introduced.

2.2. Structure of the simulated population

A population of approximately 100 000 animals was simulated, using a modified version of the stochastic simulation programme DairySim [17]. The simulation followed a yearly cycle as shown in Figure 1.

2.2.1. Traits

Seven traits were included in a total merit breeding goal: production (P), udder depth (UD), mastitis occurrence (MO), non return rate (NRR), dairy character (DC) and somatic cell score (SCS) (Tab. I). These traits were five normally distributed traits and two binary traits. At generation 0, the genetic

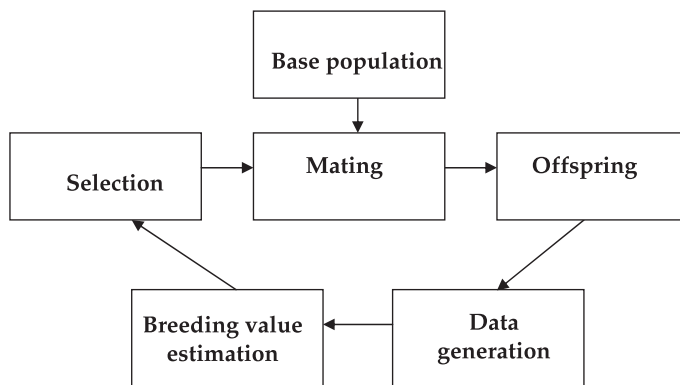


Figure 1. One year in the simulation study.

Table I. True heritabilities (diagonal), genetic (below diagonal) and phenotypic (above diagonal) correlations for traits in the simulation study.

Trait	P	UD	MO	NR	DF	DO	SCS
P	0.30	-0.20	0	-0.10	0.25	0.20	-0.15
UD	-0.35	0.30	-0.05	0	0	-0.10	-0.05
MO	0.35	-0.60	0.04	0	0	0	0.20
NR	-0.35	0.30	-0.10	0.03	0	0.05	0
DF	0.45	-0.10	0.25	0	0.25	0.10	0
DO	0.55	-0.10	0.05	-0.10	0.45	0.04	0
SCS	0.15	-0.30	0.75	-0.20	0.25	-0.25	0.10

Production (P), udder depth (UD), mastitis occurrence (MO), non return rate (NRR), dairy character (DC) and somatic cell score (SCS).

and phenotypic parameters for these traits were known and considered to be the true ones (Tab. I).

These traits were chosen to represent traits with substantial economic value in practical breeding and because they represented different types of traits with a wide range of genetic variance and correlation with other traits. For all traits, true breeding values and phenotypic observations were simulated multivariately as Gaussian traits. MO and NR were however, converted to binary traits with thresholds corresponding to frequencies of 0.18 and 0.5.

2.2.2. The first 15 years

Breeding values and phenotypes for the seven traits were simulated for base population animals, and they were allocated to age classes, gender and

Table II. Parameters used in the simulation study.

Information	Parameter
Number of years simulated	15
Number of replicates	20
Number of herds	1000
Herd size in number of females	100 cows
Young sires used for mating in percentage	30
Semen doses used from each young sire	750
Semen doses from each proven sire per year	24 000
Maximum age of bull at mating	7
Age when observations from females become available	33
Maximum number of offspring per dam	6
Pregnancy rate	0.9

herd-year-season groups. The observations were simulated using the parameters in Table II.

In the base population, the animals were mated randomly. In later years, the best males and females were mated to produce offspring based on the predicted total merit index (TM). The TM used was based on EBV for P and MO with economic weights of 19.4 and -50.0 per phenotypic standard deviation unit, respectively, corresponding to the economic values used at present in the Danish Holstein population for these two traits [20]. The EBV for P were obtained from a single trait animal model, while the EBV for MO were from a trivariate animal model including MO, UD and SCS. The binary trait MO was treated as a normally distributed trait. Models included a fixed herd-year-season effect, an animal effect and a residual term. Breeding values were estimated using genetic parameters corresponding to the ones presented in Table II. Selected bulls could be used in all herds, but in order to avoid strong inbreeding, mating between full- and halfsibs was avoided. Bulls were allocated to one of two groups representing young bulls and proven bulls and 30% of the bulls used in the simulation were unproven. Given the population size and the amount of semen produced by each young bull the number of young bulls tested per year were 40. An elite bull could not be used more than 24 000 times each year. Offspring were simulated individually and randomly distributed according to gender, and for each individual trait allocated to a herd-year-season. True breeding values were simulated as their parents average breeding values plus a Mendelian sampling term. Simulated observations were only realised for females and first when an appropriate age was reached. Observations from animals that did not reach this age were discarded. Cows were kept in the herds until better replacement heifers were available.

A proportion of the cows were involuntarily culled and all cows were culled when reaching a maximum age of 8 years. The replacement rate in the simulations was 30% in all scenarios. Also, the percentage of animals that died involuntarily per year was 5%, and these animals were picked randomly.

2.2.3. *The last 15 years*

A broader breeding goal was defined with weights on production (19.4), mastitis occurrence (-50.0), udder depth (4.2), non-return rate (13.0) and days open (-16.75). Dairy form and somatic cell score were both used as indicator traits.

Three different scenarios for prediction of EBV were simulated. In all scenarios, breeding values were predicted using different multitrait settings and the binary traits were analysed as linear normally distributed traits. The following general (co)-variance structure was used:

$$\text{var} \begin{pmatrix} a \\ e \end{pmatrix} \sim N \left(0; \begin{bmatrix} G_0 \otimes A & 0 \\ 0 & R_0 \otimes I \end{bmatrix} \right). \quad (1)$$

(Co)-variance components for use in prediction of breeding values in the three scenarios were estimated on data simulated for the first 15 years as described in the section on estimation of (co)-variance components. Even though different (co)-variance components were used for prediction of EBV in the three scenarios, the simulation was still based on the original values (Tab. I).

2.2.4. *Scenarios*

In scenario *A* an approximate multi-trait model as proposed by [4, 8] was applied. This is a two-step procedure, where the first step was applying models to each trait corresponding to the model used in the first 15 years. The observations were adjusted for the herd-year-season effect based on estimates from these models. Using these preadjusted data, breeding values for all traits were predicted in a multitrait setting using a model with a mean, a genetic effect and a residual:

$$y_{i,m} = hys_i + a_{i,m} + e_{i,m} \rightarrow y_{i,m}^* = \mu_i + a_{i,m}^* + e_{i,m}^* \quad (\text{model I})$$

where $y_{i,m}$ represents the record for animal m for trait i , hys_i the herd-year season effect for each trait and $a_{i,m}$ and $e_{i,m}$ the additive genetic and residual term for trait i of animal m . $y_{i,m}^*$ represents the preadjusted record for animal m

for trait i , μ_i the overall mean for each trait and $a_{i,m}^*$ and $e_{i,m}^*$ the additive genetic and residual term for the preadjusted record i of animal m .

The second step in (model I) was a complete multitrait setting, where G_0 and R_0 of (1) were 7×7 matrices estimated using model I (on precorrected data) as described in the section on estimation of co-variance components.

In scenario **B** a model similar to model I was used. The only difference was the inclusion of a time (year) effect in the second step where the multitrait evaluation was performed (model II):

$$y_{i,m} = hys_i + a_{i,m} + e_{i,m} \rightarrow y_{i,m}^* = \mu_i + year + a_{i,m}^* + e_{i,m}^*. \quad (\text{model II})$$

The purpose of the time effect is to correct for potential selection bias from ignoring correlations between traits when doing single trait evaluations as is done in the first step of the approximate multitrait model [9]. The effect of the improved model should show up both in terms of more accurate breeding values and in terms of more precise variance components. G_0 and R_0 of (1) were 7×7 matrices estimated using model II (on precorrected data) as described in the section on estimation of co-variance components. Using methodology like this one deletes some of the problematic aspects of multitrait evaluations. The model allows different models for different traits and it can handle repeated records. Repeated records were not simulated in this study because it makes computation even more complex but the approach is the same as for traits with one observation: the adjusted records are summarised to one per cow [8] and afterwards all record contributions are summed to adjusted records and appropriate weights. Doing the evaluation in two steps where fixed effects are subtracted from the observations after the first step makes the model capable of handling more traits since fewer equations need to be solved in the second multitrait step.

In scenario **C**, a full multitrait animal model on raw data for prediction of EBV (model III) was used, where fixed and random effects were estimated at the same time and where G_0 and R_0 of (model I) were 7×7 matrices, estimated using model III:

$$y_{i,m} = hys_i + a_{i,m} + e_{i,m}. \quad (\text{model III})$$

2.2.5. Estimation of co-variance components

After the first 15 years, (co)-variance components to be used in prediction of EBV for the last 15 years were estimated. For scenarios **A** and **B** (co)-variance components were estimated in two steps. In the first step four single trait animal models for P, NR, DF and DO, and a three trivariate animal model for

MO and UD with SCS were used as for breeding value estimation in the first 15 years. For the last step of scenarios **A** and **B**, covariance components were estimated in a full multivariate setting on preadjusted data and keeping the variances fixed at the values estimated for the first step. The only difference was that the model used in scenario **B** included a time (year) effect to account for potential selection bias. For scenario **C**, a full multitrait animal model was used. Estimation of the (co)-variance components for each of the three scenarios were performed on 10 sample datasets each with ~8000 animals with data. The average over subsets was used as parameters for prediction of EBV for the last 15 years. The estimations were conducted with the AI-REML module of the DMU-package [12].

2.2.6. Simulations of the last 15 years

All replicates of all scenarios were started from the same population simulated for the first 15 years. The simulations were as described for the first 15 years, but with the new (broader) breeding goal and the improved models/methods for prediction of EBV depending on scenario.

The efficiency of the three scenarios was compared based on the obtained progress in true total merit for the new and broader breeding goal.

3. RESULTS

(Co)-variance components estimated after the first 15 years of the simulation study are shown in Tables III, IV and V. The estimated heritabilities were not significantly different from the ones used to simulate the data using any of the models. For the approximate multitrait model without a time (year) effect (model I) 15 out of the 21 genetic correlations differed more than ± 0.10 from the ones used to simulate the data (Tab. III).

Only 5 of the 21 genetic correlations deviated more than ± 0.10 from the genetic correlations used to simulate the data both when using an approximate multitrait model with a time (year) effect (model II) and a full multitrait model (model III) (Tabs. IV and V).

In general, most of the estimated correlations were closer to 0 than the ones used to simulate the data. The residual correlations did not change as much as the genetic correlations. When using models I and II, only two of the residual correlations were different by more than ± 0.10 from the residual correlations used to simulate the data (Tabs. III and IV), and when using model III only a single residual correlation deviated more than 0.10 (Tab. V).

Table III. Heritabilities (on diagonal), genetic (below diagonal) and residual (above diagonal) correlations between the seven traits in the simulation estimated for scenario **A** on 10 subsets of the data simulated in the first 15 years. Heritabilities were estimated with univariate models and kept fixed in the multitrait setting. Correlations deviating ± 0.10 are underlined.

	P	UD	MO	NRR	DC	DO	SCS
P	0.27	-0.23	0.01	-0.08	0.19	0.21	-0.17
UD	<u>-0.16</u>	0.29	0.03	-0.05	<u>-0.24</u>	-0.15	0.03
MO	<u>0.17</u>	<u>-0.46</u>	0.05	-0.02	<u>-0.03</u>	0.00	<u>0.08</u>
NRR	<u>-0.22</u>	<u>0.10</u>	-0.06	0.03	0.01	0.04	0.01
DC	<u>0.31</u>	-0.13	<u>0.13</u>	0.10	0.25	0.06	0.02
DO	<u>0.24</u>	-0.08	<u>-0.12</u>	-0.17	<u>0.30</u>	0.06	0.03
SCS	-0.12	-0.32	<u>0.57</u>	<u>-0.07</u>	<u>-0.11</u>	<u>-0.46</u>	0.12

Production (P), udder depth (UD), mastitis occurrence (MO), non return rate (NRR), dairy character (DC) and somatic cell score (SCS).

Empirical standard errors from the 10 subsets on the heritabilities were below 0.01 and on correlations standard errors were all below 0.05.

Table IV. Heritabilities (on diagonal), genetic (below diagonal) and residual (above diagonal) correlations between the seven traits in the simulation estimated for scenario **B** on 10 subsets of the data simulated in the first 15 years. Heritabilities were estimated with univariate models and kept fixed in the multitrait setting. Correlations deviating ± 0.10 are underlined.

	P	UD	MO	NRR	DC	DO	SCS
P	0.27	-0.17	-0.05	-0.08	0.17	0.21	-0.21
UD	-0.38	0.29	0.03	-0.05	<u>-0.24</u>	-0.14	0.02
MO	0.42	<u>-0.39</u>	0.05	-0.01	<u>-0.04</u>	-0.00	<u>0.09</u>
NRR	-0.27	<u>0.20</u>	-0.15	0.03	0.00	0.04	0.02
DC	0.48	-0.16	0.24	<u>0.12</u>	0.25	0.06	0.02
DO	<u>0.32</u>	-0.17	0.11	-0.16	<u>0.34</u>	0.06	0.02
SCS	<u>0.16</u>	-0.24	0.68	-0.19	<u>-0.07</u>	-0.33	0.12

Production (P), udder depth (UD), mastitis occurrence (MO), non return rate (NRR), dairy character (DC) and somatic cell score (SCS).

Empirical standard errors from the 10 subsets on the heritabilities were below 0.01 and on correlations standard errors were all below 0.06.

In Table VI, the genetic trend for total merit in each of the three scenarios is shown. Adding a time (year) effect to the model improved genetic gain for total merit significantly.

In Tables VII and VIII trends in the true breeding values and in predicted breeding values for the individual traits for cows in each of the three scenarios are presented.

Table V. Heritabilities (on diagonal), genetic (below diagonal) and residual (above diagonal) correlations between the seven traits in the simulation estimated for scenario **C** on 10 subsets of the data simulated in the first 15 years. Correlations deviating ± 0.10 are underlined.

	P	UD	MO	NRR	DC	DO	SCS
P	0.27	-0.13	-0.06	-0.08	0.15	0.19	-0.20
UD	-0.41	0.33	0.03	-0.05	<u>-0.20</u>	-0.12	0.01
MO	0.39	<u>-0.38</u>	0.05	-0.01	<u>-0.05</u>	-0.01	0.11
NRR	-0.36	<u>0.16</u>	-0.12	0.04	0.00	0.04	0.02
DC	0.47	-0.21	0.24	0.08	0.24	0.05	0.01
DO	<u>0.34</u>	-0.19	<u>0.16</u>	-0.17	<u>0.32</u>	0.05	0.02
SCS	0.18	-0.23	0.75	-0.16	<u>-0.02</u>	-0.31	0.11

Production (P), udder depth (UD), mastitis occurrence (MO), non return rate (NRR), dairy character (DC) and somatic cell score (SCS)

Empirical standard errors from the 10 subsets on the heritabilities were below 0.02 and on correlations standard errors were all below 0.05.

Table VI. Regression of true and predicted genetic trends on year for total merit (TM) in simulation with standard errors (SE).

Scenario	Predicted		True	
	TM	SE	TM	SE
A	2.566	0.077	2.819	0.047
B	2.946	0.073	2.975	0.056
C	3.021	0.046	3.063	0.073

Table VII. Regression of predicted genetic trends on year for individual traits in simulation.

Scenario	P	UD	MO	NRR	DC	DO	SCC
A	0.113	0.027	-0.009	-0.001	0.032	0.016	-0.023
B	0.122	0.038	-0.011	0.001	0.024	0.009	-0.029
C	0.118	0.040	-0.014	0.003	0.032	0.011	-0.030

Production (P), udder depth (UD), mastitis occurrence (MO), non return rate (NRR), dairy character (DC) and somatic cell score (SCS).

Standard errors were below 0.012.

The breeding value estimation in the last round of the simulation took 12% more CPU-time using the full multitrait model used in scenario **C** compared to the approximate multitrait model used in scenario **B**. In each of the three scenarios, more than 2 GB of data was generated from one replicate given the setup of the simulations, with 20 replicates that lead to more than 120 GB of data.

Table VIII. Regression of true genetic trends on year for individual traits in simulation.

Scenario	P	UD	MO	NRR	DC	DO	SCC
<i>A</i>	0.097	0.043	-0.021	-0.002	0.035	0.017	-0.022
<i>B</i>	0.108	0.024	-0.022	-0.002	0.037	0.018	-0.017
<i>C</i>	0.107	0.030	-0.023	0.001	0.032	0.018	-0.018

Production (P), udder depth (UD), mastitis occurrence (MO), non return rate (NRR), dairy character (DC) and somatic cell score (SCS).

Standard errors were below 0.011.

4. DISCUSSION

The full multitrait model led to the highest genetic progress for total merit, but from the result of 20 replicates this model was not significantly better than an approximate multitrait model with a time (year) effect to correct for selection bias. Including a time (year) effect in the approximate multitrait model led to more robust breeding values and thereby the possibility to obtain higher genetic progress for total merit. Most countries are providing breeding values from single trait analysis, and thereby also calculating and publishing total merit indices from these breeding values. In situations where correlations between traits are not taken into account, these breeding values will often be biased [10, 15].

Survival is an economical important trait in a dairy cattle population. By nature the trait is measured late in life. Given the relatively long generation interval, the EBV for longevity of bulls based on direct observation will be very uncertain when the bull is interesting as a proven sire. Using information from correlated traits, the accuracy of predicted breeding values for traits with low heritability will increase. A MACE approach [5] as well as the model proposed in this study [18] can use indirect, direct and combined information from yield and functional traits to improve the genetic evaluation for longevity. When using direct, indirect and combined information, the mean reliability for all bulls was 0.50, 0.27 and 0.56 respectively. For the youngest proven bulls, the mean reliability was 0.26, 0.23 and 0.39 respectively. Using information from linear type traits and production traits [4] gave similar results in a linear multitrait BLUP setting. In the MACE setting as described by [5], residual correlations were assumed to be 0. This may not be true in multivariate prediction of breeding value for cows, where most of the traits are measured on each cow. The MACE methodology has been enhanced to be able to handle residual correlations different from 0 [16].

In general, the trend in true genetic merit for the individual traits did not change much when applying different models for prediction of breeding values in the simulation. The trends in general become more favourable when applying a more advanced model for prediction of breeding values. In total these changes sum up to a significant difference in true total merit. Models/methods similar to the ones used here have improved genetic response when applied on data from dairy cattle [8] and laying hens [1]. The selection differential for total merit of the 10% best sires changed in a favourable direction for all traits in the breeding goal when applying this methodology. The estimates changed differently depending on which model was used to estimate genetic correlations. When a time (year) effect was not included in the model, correlations between production and other traits changed. This was not observed to the same extent when a time (year) effect was included in the approximate multitrait model or if a full multitrait model was used. These correlations were estimated on data where there had been a strong selection for production without taking correlations between production and the other traits into account. Including a year of performance effect in a similar approximate multitrait model gave genetic parameters and sire solutions more robust to over- or underestimation of genetic trends [9]. The estimated correlations were highly dependent on which model was used, and focus should still be on the fact that correlations are only estimates of the true ones [23].

The approximate method used in this study can handle missing observations [1, 8] also in a canonical setting [7]. Then the residual variances are heterogeneous, but the algorithm for missing values described can be extended to this situation. In the EM algorithm used for missing values, the complete observations are observations that all have the same weight, equal to the largest ones on the observed scale. A canonical transformation of data was not applied in this study since the traits were very simple in nature. There were no missing observations, which often would occur in real life situations.

The approximate multitrait model as described by [4, 8] can be extended to handle traits that are not normally distributed traits. This has been shown on binary and survival traits [1, 8, 18]. An extension to handle test day records has also been developed [19]. In the present study, the main focus was to include a time (year) effect in the model to account for potential selection bias. In this study, a dairy cattle population was simulated but the model proposed could be useful in other populations with diverse breeding goals and potential long generation intervals as well. This has already been shown in laying hens [1]. In this study, especially the genetic gain for traits with low heritability seems

to increase since accuracy increases by exploiting the available information in the data. This was also expected and has been shown in other studies [21,23].

The increase of CPU-time for the full multitrait model in scenario *C* compared to the approximate multitrait model in scenario *B* of 12% is a conservative estimate of the reduction of computing time since the model for each trait was simple and only contained a herd-year-season effect. In real life, each model would contain more fixed effects, which would increase the computing time relatively more for the full multitrait model. In this study, seven traits were simulated. Addition of more traits would clearly not result in a linear increase in computing time. In addition, the present limit in terms of number of traits one would be able to handle in a full multitrait model in a reasonable size population as described in this study, would sooner be reached than using an approximate multitrait model. The amount of data produced in the simulations with 20 replicates limited the number of scenarios that could be compared in this study. Several other situations using different models, amount of data available and types of traits could have been interesting to examine, but were not investigated in this study.

Comparing models for prediction of breeding values in a simulation of a selection scheme has been done before to quantify the effect of not taking account of the use of BST (bovine somatotropin) in dairy cattle production when predicting breeding values [3]. Models that included an additive effect of the treatment seem to be more robust to the non-random allocation of BST, and a multitrait BLUP model would therefore allow better evaluation of the fixed effects in the model. The present study also shows that with biased models, one can have a distorted image of the accuracy of breeding values and genetic trends. On real data, only predicted breeding values are known. Therefore, simulations for which the true breeding values are also known can reveal such biases.

On a national scale, a full multitrait model for prediction of breeding values is usually not feasible. This is primarily due to computational limitations but also because traits are described by different models. Until these full multitrait models are feasible, an approximate model as described in this study is a better alternative than ignoring the correlations between traits.

5. CONCLUSION

An approximate multitrait model with a time (year) effect using preadjusted data from a two-step procedure is a feasible tool for prediction of breeding values. In this study, a full multitrait model was not significantly better. Including

a time (year) effect in the approximate multitrait model led to evaluations more robust to over- or underestimation of genetic trends and improved genetic progress significantly.

ACKNOWLEDGEMENTS

The authors would like to acknowledge one anonymous referee for a thorough reading of the paper leading to a significant improvement of the paper.

REFERENCES

- [1] Besbes B., Ducrocq V., Protais M., An approximate total merit index combining linear traits, a survival trait and a categorical trait in laying hens, in: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, 19–23 August 2002, Montpellier, CD-ROM communication n° 20-05.
- [2] Canavesi F., Boichard D., Ducrocq V., Gengler N., de Jong G., Liu Z., An alternative procedure for international evaluations: Production traits European joint evaluation (PROTEJE), in: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, 19–23 August 2002, Montpellier, CD-ROM communication n° 01-14.
- [3] Colleau J.J., Impact of the use of bovine somatotropin (BST) on dairy cattle selection, *Genet. Sel. Evol.* 21 (1989) 479–491.
- [4] Colleau J.J., Ducrocq V., Boichard D., Larroque H., Approximate multi-trait BLUP evaluation to combine functional traits information, Final GIFT Workshop. Breeding goals and selection schemes, Ede-Wageningen, The Netherlands, 7–9 November, *Interbull Bull.* 23 (1999) 151–160.
- [5] Druet T., Sölkner J., Gengler N., Use of multitrait evaluation procedures to improve reliability of early prediction of survival, *J. Dairy Sci.* (2001) online, <http://www.12.24.208.140/adsa/jds/index.asp>
- [6] Ducrocq V., Multi trait prediction: principles and problems, in: Proceedings of 5th World Congress on Genetics Applied to Livestock Production, 6–11 August 1994, Guelph, Canada, Vol. 18, pp. 455–462.
- [7] Ducrocq V., Besbes B., Solution of multiple trait animal models with missing data on some traits, *J. Anim. Breed. Genet.* 110 (1993) 81–92.
- [8] Ducrocq V., Boichard D., Barbat A., Larroque H., Implementation of an approximate multitrait BLUP to combine production traits and functional traits into a total merit index, 52nd annual meeting EAAP, Budapest, Hungary (2001).
- [9] Ducrocq V., Delaunay I., Boichard D., Mattalia S., A general approach for international genetic evaluations robust to inconsistencies of genetic trends in national evaluations, *Interbull Technical workshop*, Beltsville, Maryland, USA, 2–3 March, *Interbull Bull.* 30 (2003) 101–111.
- [10] Henderson C.R., Best linear unbiased estimation and prediction under a selection model, *Biometrics* 31 (1975) 423–447.

- [11] Lassen J., Sørensen M.K., Madsen P., Ducrocq V., A stochastic simulation study on validation of an approximate multitrait model using preadjusted data for prediction of breeding values, *J. Dairy Sci.* 90 (2007) 3002–3011.
- [12] Madsen P., Jensen J., DMU: a user's guide. A package for analysing multivariate mixed models, Version 6, release 4, DJF, Foulum, Denmark (2000).
- [13] Miglior F., Muir B.L., van Doormaal B.J., Selection indices in Holstein cattle of various countries, *J. Dairy Sci.* 88 (2005) 1255–1263.
- [14] Pollak E.J., van der Werf J., Quaas R.L., Selection bias and multiple trait evaluation, *J. Dairy Sci.* 67 (1984) 1590–1595.
- [15] Schaeffer L.R., Schenkel F.S., Fries L.A., Selection bias on animal model evaluation, in: *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production*, 11–16 January 1998, Vol. 25, University of Armidale, Australia, pp. 501–509.
- [16] Sullivan P.G., Wilton J.W., Schaeffer L.R., Jansen G.J., Robinson J.A.B., Allen O.B., Genetic evaluation strategies for multiple traits and countries, *Livest. Prod. Sci.* 92 (2005) 195–205.
- [17] Sørensen M.K., Berg P., Jensen J., Christensen L.G., Stochastic simulation of breeding schemes for total merit in dairy cattle, *GIFT Seminar on Genetic Improvement of Functional Traits in Cattle*, Wageningen, The Netherlands, 7–9 November, *Interbull Bull.* 23 (1999) 183–192.
- [18] Tarrés J., Piedrafitá J., Ducrocq V., Validation of an approximate approach to compute genetic correlations between longevity and linear traits, *Genet. Sel. Evol.* 38 (2006a) 65–85.
- [19] Tarrés J., Liu Z., Ducrocq V., Reinhardt F., Reents R., Parameter estimation and genetic evaluation of milkproduction traits from France and Germany with a multi-trait MACE model, in: *Proceeding at the 2006 Interbull meeting*, Kuopio, Finland, 4–6 June, *Interbull Bull.* 35 (2006b) 76–86.
- [20] The Danish Agricultural Advisory Centre, 2006, <http://www.lr.dk/kvaeg/diverse/principles.pdf>
- [21] Thompson R., Meyer K., A review of theoretical aspects in the estimation of breeding values for multitrait selection, *Livest. Prod. Sci.* 15 (1986) 299–313.
- [22] van der Werf J.H.J., van Arendonk J.A.M., De Vries A.G., Improving selection of pigs using correlated characters, in: *Proceedings of the 43rd EAAP Annual Meeting*, Madrid, Spain, 14–17 September, 1992, p. 13.
- [23] Villanueva B., Wray N.R., Thompson R., Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors, *Anim. Prod.* 57 (1993) 1–13.
- [24] Weigel K.A., Use of correlated trait information to improve the accuracy of early predictions of breeding values for length of productive life, in: *Proc. Int. Workshop on Genetic Improvement of Functional Traits in Cattle*, Gembloux, Belgium, 1996, *Interbull Bull.* 12, *Int. Bull. Eval. Serv.*, Uppsala, Sweden.