Validation of an approximate approach to compute genetic correlations between longevity and linear traits

J. Tarres[1]*, J. Piedrafita[1] and V. Ducrocq[2],
[1]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain. Email : jtarres@quiro.uab.es
[2]Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique, 78352 Jouy en Josas, France. Email : ducrocq@dga.jouy.inra.fr

1. INTRODUCTION

A common objective in animal breeding is to increase the economic efficiency of animal production by directional selection. In this context, economically important traits are often classified into production and functional traits. Functional traits refer to traits that increase the economic efficiency by reducing costs instead of increasing the quantity of saleable products. Their importance results from the fact that obviously, domestic animals have to be alive and to reproduce normally to be profitable for the breeder. Functional traits have increased significantly their economical importance, in particular in dairy cattle (Essl, 1998).

In general, with the exception of some type traits, functional traits exhibit rather low heritabilities, leading to genetic evaluations with low reliabilities for young sires (Ducrocq, 2001). Fortunately, more heritable traits can be used as early predictors of these functional traits. For example, early predictors of, e.g., somatic cells count or functional longevity (e.g., Visscher and Goddard, 1995, Weigel, 1996; Weigel et al., 1998, Druet et al., 1999, Larroque and Ducrocq, 2001, Buenger et al., 2001) can be found in the long list of type traits recorded in each breed. A technique to properly combine these pieces of information is needed, so they can be included in a more balanced total merit index combining production and functional traits

Ducrocq et al. (2001) proposed a two-step approach for multiple trait evaluation of functional and production traits: first, univariate analyses are performed for each trait to get genetic variance estimates and to compute for each recorded individual "pseudo-records" and their associated weights. Combining these "pseudo-records" in a multiple trait animal model while fixing the genetic and residual variances, one can get correlation estimates by AI-REML and approximate MT-BLUP breeding values that blend direct and indirect information.

The aim of this paper is to check via simulation the two-step approach for multiple trait genetic evaluation of longevity and two linear traits. After the analysis of a reference situation, a sensitivity analysis was performed to check the suitability of this approach in a wide range of situations.

2. MATERIAL AND METHODS

2.1. General strategy

Observations of longevity (t) and of two linear traits (y_1, y_2) with known genetic and residual correlations were simulated. For these traits, the current models of analysis used are so different (the longevity model has to deal with non linearity, censoring and non normal residuals) that an exact multiple trait approach is usually not feasible, at least on moderate or large size data sets. The proposed approach aims at summarising the data at the individual level in such a way that the simplest linear animal model can be used for each trait. This first step requires the calculation of a "pseudo-record" $y^*_{i,m}$ for each animal m and trait i corrected for all non genetic effects and an associated weight $w_{i,m}$ indicating the amount of information for that animal. These "pseudo-records" are obtained from a univariate (or a simpler multivariate) analysis, after estimation of the relevant dispersion parameters. Then, all "pseudo-records" are analysed together using a classical MT-BLUP framework assuming an animal model:

$$y^*_{i,m} = \mu_i + a_{i,m} + e_{i,m} \qquad [1]$$

where μ_i is the overall mean for trait *i*, $a_{i,m}$ is the additive genetic value of animal *m* for trait *i* and $e_{i,m}$ is the residual. To account for the variable amount of information summarised in $y^*_{i,m}$, its residual variance is assumed to be heterogeneous: $\text{var}(e_{i,m}) = \sigma^2_{e,i} / w_{i,m}$ where $\sigma^2_{e,i}$ is the residual variance for trait *i*.

The derivation of "pseudo-records" and their weights is based on the following principle: when analysed using the simplistic univariate BLUP animal model [1], these records should lead to EBVs equal or as close as possible to the EBVs obtained with the complete model and in the case of nonlinear traits, with the adequate methodology. Then, the application of a MT-BLUP animal model based on equation [1] is straightforward. It provides the appropriate EBVs for all traits and all animals. However, MT-BLUP requires an adequate knowledge of the correlations between traits. The variances can be supposed to be the ones estimated for the simpler univariate analysis used to compute "pseudo-records". Here, the objective is to check via

simulation whether these "pseudo-records" with their associated weights can be used to estimate appropriate genetic and residual correlations using a sire or an animal model and to provide the appropriate EBVs for all traits and all animals.

2.2. Simulation of the dataset

Records of 5000 animals roughly resembling length of productive life of dairy cows were simulated using the following Weibull log normal frailty model:

$$h(t) = \lambda\rho(\lambda t)^{\rho-1} \exp[X\beta + Za] \qquad [2]$$

where $h(t)$ is the hazard function at time $t > 0$. The Weibull parameters ρ and λ are strictly positive. A ρ value of 2.0 (increasing hazard) and a λ value such that the median time t_{med} was 1000 days were used for the simulation. In the case of the two linear traits simulated based on the model $y_i = \mu_i + X_i\beta_i + Z_i a_i + e_i$, two means $\mu_1 = 100$ and $\mu_2 = 200$ were also arbitrarily added without lack of generality. For each trait i, two fixed effects $\beta_i = (\beta_{i1}, \beta_{i2})$, with 10 and 100 levels respectively, were generated from a uniform distribution and an appropriate change of scale. These levels were attributed to each animal at random, in a relatively unbalanced way.

The true breeding values of the 5000 animals $a_{i,m}$, that were progeny of 50 unrelated sires, were obtained by adding half the breeding value of their sire $s_{i,m}$ to a value $u_{i,m}$ covering the dam contribution and mendelian sampling, i.e., representing three quarters of the total genetic variance. Values were drawn from a MVN(0,G) distribution, where G is the genetic covariance matrix. In the reference situation, genetic variances for longevity and linear characters were respectively 0.20, 400 and 600. Genetic correlations between all pairs of traits were 0.4.

Residual values $e_{i,m}$ for the two linear traits were also generated from values drawn from a bivariate normal distribution with 0 mean and residual (co)variance matrix R. Residual variances were chosen to lead to heritabilities of 0.25 for the first linear character and of 0.10 for the second, representing medium and low genetic variation. The distribution of the residual component for the longevity measure is proportional to an extreme value distribution. This residual was not explicitly simulated but resulted from the simulation process based on the hazard function (Kalbfleisch and Prentice, 1980). The residual correlation between the two linear traits was 0.4. In the reference situation, a residual correlation of 0 was assumed between longevity and the two linear traits.

2.3. First step: Calculation of "pseudo-records" and their associated weight.

Longevity was analysed using a Weibull frailty model (see Ducrocq and Sölkner, 1998; Ducrocq et al., 2001). First, the genetic variance was estimated using a sire model. Then, assuming that this estimated variance is the correct one, longevity "pseudo-records" and their weights were obtained using two procedures. The first one was a two-step procedure where the first step involved the estimation of fixed effects and sire EBVs using a sire model and the second step consisted in calculating the animals' EBVs, considering fixed effects and sire EBVs as known. This leads to an approximation of the EBV solutions $\hat{a}_{j,m}$ from a Weibull animal model, solving a non-linear equation for each animal *m* (see Ducrocq, 2001 for details). The second procedure consisted in directly estimating the animals' EBVs $\hat{a}_{j,m}$ from a Weibull animal model. The resulting "pseudo-record" for longevity for animal m in both procedures was (for details, see Ducrocq (2001)) :

$$y^*_{i,m} = \frac{\delta_m}{w_{i,m}} + \hat{a}_{i,m} - 1 \qquad [3]$$

where $\delta_m = 0/1$ is the censoring code. The associated weight $w_{i,m} = \frac{H_{i,m}}{\exp(\hat{a}_{i,m})}$ is a function of the cumulative risk $H_{i,m}$ of animal m from time 0 to culling or censoring time. Calculations were done using a modified version (version 5.0) of the Survival Kit (Ducrocq and Sölkner, 1998).

For linear traits, instead of a univariate analysis for each trait, it was decided to use a bivariate animal model, which is simple to implement here. First, genetic and residual dispersion parameters were estimated via a REML procedure. The "pseudo-record" for trait i and animal m was simply the record adjusted for fixed effects:

$$y^*_{i,m} = y_{i,m} - X\hat{\beta}_{i,j} \qquad [4]$$

The associated weight $w_{i,m}$ for the approximate MT-BLUP evaluation was the animal m diagonal element of the univariate mixed model coefficient matrix after absorption of fixed effects, but before adding the relationship matrix .

2.4. Second step: Joint analysis

Genetic and residual correlations were estimated via a REML procedure applied to "pseudo-records" assuming that the variances are known and equal to the estimates obtained in the first step. Data were analysed using either a sire or an animal model to compare the performance of both models. Several REML packages exist but none is really adapted to model equations such as [1] with heterogeneous residual variances. A simple trick to avoid this problem is to multiply both sides of the model equation [1] by $v_{i,m} = \sqrt{w_{i,m}}$. Then, one gets:

$$y^{\#}_{i,m} = v_{i,m}\ y^{*}_{i,m} = v_{i,m}\mu_i + v_{i,m}a_{i,m} + \varepsilon_{i,m} \qquad [5]$$

Now, the residual part $\varepsilon_{i,m}$ has homogeneous variance: $Var[\varepsilon_{i,m}] = v^2_{i,m}\ Var[e_{i,m}] = \sigma^2_{e,i}$. The REML estimation of the dispersion parameters of model [5] considering $y^{\#}_{i,m}$ as the data and $v_{i,m}$ as a continuous covariable gives results identical to the analysis of model [1] (Ducrocq et al., 2001). I. Misztal's AI-REMLsoftware was modified to impose constraints (Druet et al., 2003). Here, the genetic and residual variances were fixed to their estimates obtained in the first step.. Finally, a MT-BLUP evaluation based on "pseudo-records" (and their weights) was performed using the estimated genetic and residual (co)variance matrices.

2.5. Reliabilities

Two different longevity EBVs were obtained for each animal: one from direct evaluation (via the Weibull model) and the other taking into account indirect information from correlated traits (MT-BLUP approach). Average reliabilities for longevity of progeny animal and their sires were also computed in two ways. First, asymptotic mean reliability was obtained as the mean of the diagonal elements of the inverse of the Weibull Hessian matrix at convergence of the maximisation process. The second way was to compute the correlation between the simulated (true) BVs and the estimated ones (via Weibull model or via the MT-BLUP approach). The average reliability is the square of this correlation.

2.6. Sensitivity analysis

2.6.1. Batches of progeny with different censoring rates

To check how the reliability of the younger sires' proofs would increase with the multiple trait approach as a function of censoring percentage, in this section sires were simulated with different percentage of censoring among their progeny. The actual longevity measure for the first 1000 animals (progeny of the first 10 bulls) was set to min(400,T_m), i.e., censored at 400 d if the actual failure time T_m was larger than 400 d. Longevity for the next 1000 animals was set to min(800,T_m) days and the records of the following 1000 animals were set to min(1200,T_m) days. Finally, the last 2000 animals were not censored (progeny of old sires). The first 1000 animals would represent progeny of young sires (approximately 90% censoring) and the last animals, progeny of older sires (uncensored).

2.6.2. Non zero residual correlation between longevity and linear traits

Above, a zero residual correlation was assumed between longevity and linear traits. Assuming that longevity data follow a Weibull distribution, equation [2] is equivalent to:

$$\log t = -\frac{(\rho \log \lambda + X\beta + Za + e)}{\rho} \qquad [6]$$

where $e = \rho^{-1}\omega$ and ω follows an extreme value distribution (Kalbfleisch and Prentice, 1980). Then,

$$Var(e) = Var(\rho^{-1}\omega) = \rho^{-2}(\pi^2 / 6)$$

To generate a nonzero correlation, records for linear traits were simulated as:

$$y_1 = \mu_1 + X_1\beta_1 + Z_1a_1 + \varepsilon_1 + \omega_1 \qquad [7]$$

$$y_2 = \mu_2 + X_2\beta_2 + Z_2a_2 + \varepsilon_2 + \omega_2 \qquad [8]$$

where the complete residual values (e_1, e_2) were partitioned into two components: one correlated with longevity (ω_1, ω_2) with correlations r_{long1} and r_{long2} respectively, and the other correlated with the other linear character $(\varepsilon_1, \varepsilon_2)$, that is to say:

$$\text{cov}(e, e_1) = \text{cov}(e, \omega_1) + \text{cov}(e, \varepsilon_1) = r_{long1}\sigma_e\sigma_{e_1} + 0 = r_{long1}\sigma_e\sigma_{e_1}$$
$$\text{cov}(e, e_2) = \text{cov}(e, \omega_2) + \text{cov}(e, \varepsilon_2) = r_{long2}\sigma_e\sigma_{e_2} + 0 = r_{long2}\sigma_e\sigma_{e_2}$$
$$\text{cov}(e_1, e_2) = \text{cov}(\varepsilon_1, \varepsilon_2) = r_{12}\sigma_{e_1}\sigma_{e_2}.$$

To obtain ω_1 and ω_2, first the failure time T_m in [6] was simulated . Then, the residual e was obtained as $e = -\rho \log T_m - \rho \log \lambda - X\beta - Za$. The component correlated with longevity (ω_1, ω_2) was generated as $\omega_i = b_{longi} e_i$, i.e., the regression of e_i on e_{long} with $b_{longi} = r_{longi}(\sigma_{e_i} / \sigma_e)$.

Finally, the value pairs $(\varepsilon_1, \varepsilon_2)$ were generated from a bivariate normal distribution with the adequate covariance matrix:

$$Var\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\end{pmatrix} = Var\begin{pmatrix}e_1\\ e_2\end{pmatrix} - Var\begin{pmatrix}w_1\\ w_2\end{pmatrix} = \begin{pmatrix}\sigma^2_{e1} - b^2_{long1}\sigma^2_e & r_{12}\sigma_{e1}\sigma_{e2} - b_{long1}b_{long2}\sigma^2_e\\ r_{21}\sigma_{e2}\sigma_{e1} - b_{long2}b_{long1}\sigma^2_e & \sigma^2_{e2} - b^2_{long2}\sigma^2_e\end{pmatrix}$$

Note that the non zero residual correlation between longevity and the linear traits was obtained using equation (6) to model $\log t$ while the final model of analysis with "pseudo-records" is on a different scale, with heterogeneous residual variance. This will have implications on the interpretation of the results.

2.6.3. Biases generated by incorrect univariate analyses

So far, the model of analysis of simulated records has been assumed to be correct. Here, the hypothesis is that a genetic trend on all traits existed (progeny born in years 0, 1, …,10 do not have the same average genetic level) but was incorrectly estimated (biased) in the univariate analyses. To generate such a situation, the simulation models for longevity and linear traits were modified to:

$$h(t) = \lambda\rho(\lambda t)^{\rho-1}\exp\left[X\beta + \delta_j + Za\right] \quad [9]$$
$$y_1 = \mu_1 + X_1\beta_1 + \delta_{1j} + Z_1a_1 + e_1 \quad [10]$$
$$y_2 = \mu_2 + X_2\beta_2 + \delta_{2j} + Z_2a_2 + e_2 \quad [11]$$

where, for an animal born in year $j = 0, 1, ..., 10$ an effect $\delta_j = 0.05\ j\ \sigma_a$ was added for each trait (5% of a genetic standard deviation per year). To create an unbalanced but connected design, a sire with 100 progeny was assumed to have half of them born in year j and the other half born in year $j+1$ (where $j = 0$ for 10% of the sire, $j = 1$ for the second 10%, etc.. So, sires had their progeny in different years.

The analysis of such a dataset was done ignoring the year effect in the first step (univariate analysis). Therefore, this resulted in a situation where estimates of variances, "pseudo-records" and weights were biased. However, to see how the joint analysis of the data can cope with such a bias, a year effect was included in the last steps, i.e., for the estimation of the genetic and residual correlations and for the MT-BLUP evaluation of the three traits together.

3. RESULTS

3.1. Reference situation

Genetic and residual variances were estimated assuming a Weibull sire model for longevity and a bivariate linear animal model for linear traits. For linear traits, true genetic and residual variances were always within the 95% confidence interval of the 200 replicates mean (results not shown). In the case of longevity, the sire variance was slightly underestimated for the reference situation. However, this bias was negligible.

The calculation of the longevity "pseudo-records" and their weights was obtained according to two alternative procedures: an approximate two-step procedure based on a sire model and a direct procedure based on the exact animal model. The correlations between the results of the two procedures were 0.9985 for "pseudo-records" and 0.9910 for their weights. The approximate two-step procedure gave results equivalent to the more demanding exact one. Only results from the exact procedure are reported hereafter because the size of our datasets allowed the use of the animal model.

Next, univariate BLUP analyses were performed using "pseudo-records" for longevity.. The correlations between BLUP EBVs from "pseudo-records" and approximate Weibull animal model EBVs were

0.9980 for sires and 0.9999 for progeny. Their standard deviations were also nearly identical. Therefore, as desired, the use of "pseudo-records" in univariate BLUP evaluation based on an animal model led to sire and progreny EBVs similar to the approximate EBVs obtained in the Weibull analysis.

The "pseudo-records" allowed the estimation of genetic and residual correlations under a multiple trait sire or animal model. All mean genetic correlations estimates were similar to the simulated ones whatever the estimation models (Table 1). However, this was obtained only when a constraint was imposed stating that genetic and residual variances are known and equal to the ones estimated during the first step. Without this constraint,convergence was rarely obtained. The standard deviations (over replicates) of genetic correlations were between 0.139 and 0.160 for the sire model and were slightly smaller for the animal model (Table 1). The average asymptotic standard errors provided by the AI-REML algorithm were lower than these standard deviations (between 0.100 and 0.113 for the sire model), mainly because it was assumed that the variances were known without error. Nevertheless, standard errors provided by the AI-REML gave an idea of the order of magnitude of the correlations accuracy.

Table 1. Estimates of genetic and residual correlations between longevity (Long) and the two linear traits (L1, L2) using the AI-REML approach under a sire and an animal model for two levels of animal genetic variation for longevity: low (0.05) and high (or reference) (0.20).

Model of Analysis			Sire				Animal			
Longevity heritability			Low		High		Low		High	
Correlations [a]		True	Mean[a]	STD [a]	Mean[a]	STD [a]	Mean[a]	STD [a]	Mean[a]	STD[a]
Genetic	Long-L1	0.4	0.382	0.182	0.371	0.139	0.388	0.128	0.374	0.079
	Long-L2	0.4	0.410	0.209	0.389	0.160	0.402	0.179	0.380	0.121
	L1-L2	0.4	0.377	0.158	0.373	0.160	0.379	0.156	0.374	0.156
Residual	Long-L1	0.0	0.004	0.005	0.010	0.006	-0.000	0.006	-0.000	0.006
	Long-L2	0.0	0.003	0.005	0.007	0.006	-0.000	0.005	-0.000	0.006
	L1-L2	0.4	0.406	0.011	0.406	0.011	0.415	0.024	0.416	0.024

[a] Mean and standard deviations (STD) over 200 replicates.

Finally, average reliabilities for longevity computed as the squared correlations between true and EBVs were 0.890 for sires and 0.538 for progeny when just direct information was used in a Weibull model. These reliabilities were similar to the asymptotic ones obtained from EBV standard errors. When information on correlated linear traits was taken into account under a MT-BLUP approach, the gain in reliability was very limited for sires (0.2 %) and slightly higher (1.7 %) for progeny (Table 2). This modest increase may be related to the uncertainty on dispersion parameters and the initial level of reliability, which was already high for sires.

Table 2. Mean and standard deviation of reliabilities (squared correlations between true and estimated breeding values) of longevity obtained for sires and progeny with a Weibull model and the MT-BLUP approach under a sire and an animal model.

			Weibull model		MT-BLUP	
Genetic variance	Model of analysis	Animals	Mean [a]	STD [a]	Mean [a]	STD [a]
Low (0.05)	Sire	Sires	0.731	0.066	0.743	0.064
	Animal	Sires	0.732	0.066	0.748	0.063
		Progeny	0.403	0.043	0.434	0.040
High (0.20)	Sire	Sires	0.890	0.030	0.892	0.030
	Animal	Sires	0.892	0.030	0.895	0.029
		Progeny	0.538	0.032	0.555	0.029

[a] Mean and standard deviations (STD) over 200 replicates.

3.2. Sensitivity analysis

3.2.1. Progeny batches with different censoring rates

The previous results were obtained in the situation of no censoring. When sires were simulated with different percentages of censoring among their progeny, the average reliabilities for longevity decreased (and

their standard deviations increased) as the percentage of censoring increased. The average reliability of young sires (with 90 % of their daughters censored) was 0.36 and 0.60 for low and high genetic variation, respectively. When indirect information was included (Figure 1), there was no gain in reliability for old sires (0% or 30% progeny censored). However, the gain for young sires was important (4-5 %) when genetic variation was high, and even more when it was low (up to 10 %). This can be attributed to the fact that the information of older sires allowed a more accurate estimation of variances and genetic correlations.

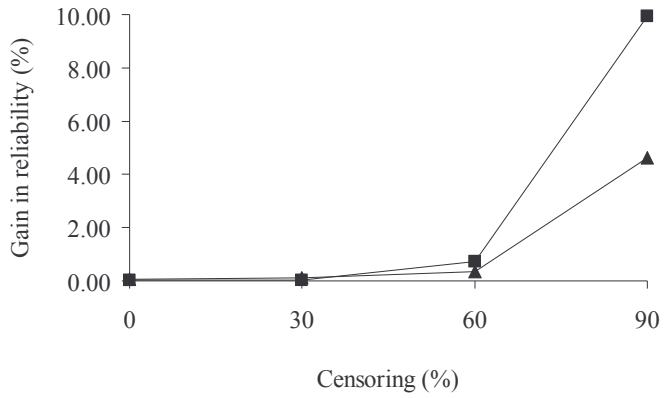

Figure 1. Increase in reliability for longevity of sires obtained by adding indirect information with the MT-BLUP approach as a function of censoring percentage. Different levels of genetic variance for longevity were simulated: low (■) and high (▲).

3.2.2. Effect of non zero residual correlations

When non zero residual correlations between longevity and linear traits existed, estimates of genetic and residual variances were again similar to the simulated values (results not shown). In the joint analysis, the sire model gave again virtually unbiased estimates of genetic correlations. However, the estimates from the animal model were clearly biased upwards (Table 3). The standard deviations were similar to those obtained with zero residual correlation. It should be reminded that the residual correlation estimates with the longevity trait are not comparable to the simulated ones here (Table 3): the latter ones corresponded to the modelling of log t with a residual variance of $\rho^{-2}\pi^2/6$. They are not directly transposable to the "pseudo-record" scale which assumes a heterogeneous residual variance. Average reliabilities for sires and progeny were similar to those obtained with zero residual correlation (results not shown). These facts again confirm the robustness of the REML approach on "pseudo-records" for the estimation of genetic correlations under a sire model. However, the animal model clearly appears unsuitable in the case of non zero residual correlations.

Table 3. Estimates of genetic and residual correlations between longevity (Long) and the two linear traits (L1, L2) using the AI-REML approach under a sire and an animal model. A high genetic variation for longevity (0.20) simulates assumed.

		Non zero residual			Incorrect univariate		
Correlations [a]		True	Sire	Animal	True	Sire	Animal
Genetic	Long-L1	0.4	0.398	0.773	0.4	0.383	0.359
	Long-L2	0.4	0.436	0.852	0.4	0.398	0.366
	L1-L2	0.4	0.377	0.674	0.4	0.400	0.394
Residual	Long-L1	0.4 [b]	0.108 [c]	0.090	0	0.010	-0.001
	Long-L2	0.4 [b]	0.111 [c]	0.093	0	0.007	-0.000
	L1-L2	0.4	0.378	0.331	0.4	0.406	0.414

[a] Mean over 200 replicates.
[b,c] Scale change; not comparable to the values on the same line

3.2.3. Effect of incorrect univariate analysis

When the univariate analyses were incorrect (for example, because of a hidden bias in the estimated genetic trend and/or an incorrect modelling of fixed effects), estimated genetic variances were slightly increased compared to the reference situation and residual variances were underestimated (Table 3). As a consequence, the "pseudo-records" are also biased. A year effect was included in the joint analysis of the data, to try to capture this bias. For longevity, the slope of the regression of the year effects estimated with the MT-BLUP approach on year was 0.0221 (sire model) and 0.0223 (animal model) and was very close to the simulated value (0.0224). Similar results were obtained for the year effects of the two linear traits. Given the correction for the annual bias with the inclusion of this year effect, and in spite of the use of biased genetic and residual variances, all genetic correlations were on average similar to the simulated ones for both estimation models (Table 3). The standard deviations of these correlations were also similar to those of the reference situation, between 0.141 and 0.171. On the other hand, the average reliabilities for sires and progeny were slightly lower (1 to 2 %) than the reference ones (results not shown).

4. DISCUSSION

The results obtained under a sire and an animal model confirmed the suitability of the approach proposed by Ducrocq et al. (2001), in a wide range of situations (different genetic correlations, heritabilities and levels of censoring) not all presented here. The two step approach starts with the estimation of dispersion parameters via univariate analyses. Using these estimates, "pseudo-records" are created. These "pseudo-records" are performances free of all environmental effects that can be used in a BLUP animal model to get the same breeding values as in the Weibull animal model. They can also be viewed as deregressed proofs. The longevity "pseudo-records" and their associated weights can be obtained according to two procedures. If the dataset is small, they can be computed based on an animal model. However, if there are computational constraints to implement it (e.g. for large national applications) the two step procedure based on a sire model is a less demanding alternative. The correlations between both procedures were very high for "pseudo-records" as well as for weights and, therefore, both of them could be used indistinctly.

Combining these "pseudo-records" into a multiple trait sire model and fixing genetic and residual variances to the previously estimated values, AI-REML estimates of genetic and residual correlations were virtually unbiased. The use of multiple trait linear approach under a sire model for analysing pseudo records appears justified, as long as the constraint that genetic and residual variances are known is imposed. This constraint led to standard errors of genetic and residual correlations provided by the AI-REML algorithm that were smaller than the actual standard deviations. Nevertheless, these standard errors provided a rough order of magnitude of the accuracy of the estimates. Another concern related to this constraint is that potentially biased genetic and residual variances might lead to biased estimates of correlations. In fact, it was found that the multiple trait approach under a sire model is rather robust: if the genetic trend is wrongly estimated in univariate analyses resulting in biased estimates of variances, "pseudo-records" and weights, the joint analysis of the data can correct and even estimate this bias if a time (year) effect is included in the model. Again, this leads to nearly unbiased estimates of genetic correlations.

On the other hand, the adequacy of the estimation of genetic correlations with the multitrait animal model was found to rely on the assumption of a zero residual correlation between longevity and linear traits. This assumption is natural when different traits are recorded on different animals, but is no longer satisfying when the traits are observed on the same animals. Then residual correlations can differ substantially from 0 for some pairs of traits (Ducrocq et al., 2001). In such a situation, one should estimate the genetic and residual correlations under a sire model and then evaluate the progeny under an animal model given these correlations. The correlation between individual "pseudo-records" residuals clearly deviated from the true residual correlation. A sire model somewhat averages residuals over progeny of a same sire and is more robust.

After the MT-BLUP evaluation accounting for the correlation between longevity and linear traits, a gain in true reliability is observed with respect to the situation when just direct information is used in a Weibull model. This gain is greater when the individual proof reliability in the initial Weibull model is lower. However, this gain was very limited in uncensored datasets, due to the use of possibly inaccurate dispersion parameters. The increase in reliability is more noticeable when progeny batches with different censoring rates are simulated. Then, the information from older sires allows an accurate estimation of variances and genetic correlations and the multiple trait evaluation uses this information to significantly improve the reliability for young sires with a high proportion of censored daughters.

Finally, one can note that the two step approach is an operational tool that can be implemented in many situations where a multiple trait approach is desirable but not applicable, either because of the huge size of the datasets analyzed or the complexity and heterogeneity of the models to be implemented. Recent applications include total merit index constructions (Ducrocq et al., 2001), joint analyses of longevity, discrete and linear traits (Besbes et al, 2002), joint cow and bull international evaluation (Canavesi et al., 2002).

REFERENCES

Besbes B., Ducrocq V., Protais M. An approximate total merit index combining linear traits, a survival trait and a categorical trait in laying hens., pp. in 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, August 19-23 (2002) Communication n° 20-05

Buenger A., Ducrocq V., Swalve H.H. Analysis of survival in dairy cows with supplementary data on type scores and housing systems from a region of northwest Germany. Journal of Dairy Science, 84 (2001) 1531-1541.

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). pp. in 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, August 19-23 (2002) Communication n° 01-59.

Colleau J.J., Regaldo D., Etablisement de l'objectif de selection dans les races bovines laitières. Renc. Rech. Ruminants 8 (2001)

Druet T., Sölkner J., Gengler N., Use of multitrait evaluation procedures to improve reliability of early prediction of survival, J. Dairy Sci. 82 (1999) 2054-2068.

Druet T., Jaffrezic F., Boichard D., Ducrocq V., Modelling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows, J. Dairy Sci. 86 (2003) 2480-2490.

Ducrocq, V. P., Sölkner J., The Survival Kit v3.12", a FORTRAN package for large analysis of survival data, in: Proceedings of 6th World Congress on Genetics Applied to Livestock Production, University of New-England, Armidale, Australia, 27 (1998) 447-450.

Ducrocq V., A two-step procedure to get animal model solutions in Weibull survival models used for genetic evaluations on length of productive life., pp. 147-152 in Proceedings of the 2001 Interbull Meeting, Budapest, Hungary, August 30-31, 2001, Interbull Bulletin n° 27 (2001).

Ducrocq V., Boichard D., Barbat A., Larroque H., Implementation of an approximate multitrait BLUP evaluation to combine production traits and functional traits into a total merit index. 52nd Annual Meeting of the European Association for Animal Production, Paper GI.4, Budapest, Hungary, August 26-29, 2001.

Essl, A. Length of productive life in dairy cattle breeding: a review. Livestock Production Science 57 (1998): 79-89.

Kalbsfleisch J.D., Prentice R.L. The statistical analysis of failure time data. Wiley, New York, 1980.

Larroque H., Ducrocq V., Relationships between type and longevity in the Holstein breed. Genetics Selection Evolution, 33 (2001) 39-59.

Visscher P.M., Goddard M.E., Genetic parameters for milk yield, survival, workability and type traits for Australian dairy cattle, J. Dairy Sci. 78 (1995) 205-220.

Weigel K.A., Use of correlated trait information to improve the accuracy of early prediction of breeding value for length of productive life, in: Preliminary of the International Workshop on EU Concerted Action Genetic Improvement of Functional Traits in Cattle (GIFT), Inter. Bull. 12 (1996) 125-135.

Weigel K.A., Lawlor T.J., Van Raden P.M., Wiggans G.R., Use of linear type and production data to supplement early predicted transmitting abilities for productive life, J. Dairy Sci. 81 (1998) 2040-2044.