Genet. Sel. Evol. 38 (2006) 45–64

© INRA, EDP Sciences, 2005
DOI: 10.1051/gse:200502

Original article

A bivariate quantitative genetic model for a linear Gaussian trait and a survival trait

Lars Holm DAMGAARD[a,b*], Inge Riis KORSGAARD[b]

[a] Department of Large Animal Sciences, Royal Veterinary and Agricultural University, Grønnegårdsvej 2, 1870 Frederiksberg C, Denmark
[b] Department of Genetics and Biotechnology, Danish Institute of Agricultural Sciences, P.O. Box 50, 8830 Tjele, Denmark

(Received 20 April 2005; accepted 16 September 2005)

Abstract – With the increasing use of survival models in animal breeding to address the genetic aspects of mainly longevity of livestock but also disease traits, the need for methods to infer genetic correlations and to do multivariate evaluations of survival traits and other types of traits has become increasingly important. In this study we derived and implemented a bivariate quantitative genetic model for a linear Gaussian and a survival trait that are genetically and environmentally correlated. For the survival trait, we considered the Weibull log-normal animal frailty model. A Bayesian approach using Gibbs sampling was adopted. Model parameters were inferred from their marginal posterior distributions. The required fully conditional posterior distributions were derived and issues on implementation are discussed. The two Weibull baseline parameters were updated jointly using a Metropolis-Hasting step. The remaining model parameters with non-normalized fully conditional distributions were updated univariately using adaptive rejection sampling. Simulation results showed that the estimated marginal posterior distributions covered well and placed high density to the true parameter values used in the simulation of data. In conclusion, the proposed method allows inferring additive genetic and environmental correlations, and doing multivariate genetic evaluation of a linear Gaussian trait and a survival trait.

survival / Gaussian / bivariate / genetic / Bayesian

1. INTRODUCTION

In recent years, several breeding organizations have implemented a routine genetic evaluation of sires for longevity of dairy cows. The evaluations are mainly based on univariate sire frailty models for survival data, as described in [7], and implemented in the Survival Kit [8]. However, for genetic evaluation of animals based on several traits, a multivariate analysis is advantageous both

* Corresponding author: lars.damgaard@agrsci.dk

Article published by EDP Sciences and available at http://www.edpsciences.org/gse or http://dx.doi.org/10.1051/gse:200502

in increasing the efficiency with which animals are ranked for selection, and in providing information about the genetic correlation between traits [10, 23, 25]. The latter measures the extent to which different traits are controlled by the same genes and provides important information about how selective breeding is expected to lead to correlated and not necessarily favorable responses in different traits. Furthermore, the bias introduced by artificial selection for one or more of the traits considered will be avoided in a multivariate analysis as opposed to the corresponding univariate analyses.

If the traits considered are multivariate normally distributed, then a multivariate quantitative genetic analysis using REML is common practise [27]. Recent contributions also include Bayesian methods for drawing inferences in multivariate quantitative genetic models of linear Gaussian and ordered categorical traits [18, 32]. So far, no methods exist for analyzing a survival trait jointly with linear Gaussian traits, which are genetically and environmentally correlated. Alternatively, the genetic correlation has been approximated by the product moment correlation between estimated sire effects from two univariate analyses (*e.g.* [16, 26, 28]). In other cases, and again based on estimated sire effects obtained from univariate analysis, the genetic correlation has been approximated by the principles applied in Multiple Trait Across Country Evaluations [19, 30]. A central assumption in both approaches is that the residual effects of the two traits are uncorrelated. Data for joint analysis of several traits are often based on recordings on the same animal. Therefore, a zero environmental correlation can generally not be assumed.

In this paper, we suggest a bivariate model for a linear Gaussian and a survival trait, which are genetically and environmentally correlated. We assumed that the linear Gaussian trait and the unobserved log-frailty of the survival trait followed a bivariate normal distribution. Model parameters were inferred from a Bayesian analysis using Gibbs sampling. The bivariate model was illustrated in two simulation studies. In the first simulation study, data with a relatively simple covariate structure was simulated according to a half-sib design and analyzed with a sire model and its equivalent animal model. In the second simulation study, we extended the simulation of the survival data to also involve time-dependent covariates.

2. MATERIALS AND METHODS

2.1. Model without missing data

Let Y_{1i} be a random variable of a linear Gaussian trait, and for the survival trait let T_i and C_i be random variables representing survival time and censoring

time of animal i for $i = 1, ..., n$, where n is the total number of animals with records. In field data, there will often be at least some animals on which records are available for one of the two traits only. In order to keep notation simple we define the bivariate model and implement the Gibbs sampler under the assumption that measurements are available for both traits. Later we consider the case where data on one of the two traits are missing at random. In the case of no missing observations, data of animal i are given by $(y_{1i}, y_{2i}, \delta_{2i})$, where y_{1i} is the observed value of Y_{1i}, y_{2i} is the observed value of $Y_{2i} = \min\{T_i, C_i\}$, and δ_{2i} is the outcome of the censoring indicator variable, such that δ_{2i} is equal to 1 if $T_i \leq C_i$ and 0 otherwise. For the survival trait, we considered the Weibull log-normal animal frailty model including a normally distributed residual effect on the log-frailty scale [1].

The bivariate model will be represented by the conditional hazard function of T_i and the joint distribution of $\mathbf{Y}_1$ and $\mathbf{e}_2$

$$\lambda_i(t|\boldsymbol{\theta}, \mathbf{e}_2) = \rho t^{(\rho-1)} \exp\left\{\mathbf{x}'_{2i}(t)\boldsymbol{\beta}_2 + \sum_{j\in Q(2)} \mathbf{z}'^{j}_{2i}(t)\mathbf{u}^j_2 + \sum_{j\in Q(1,2)} \mathbf{z}'^{j}_{2i}\mathbf{u}^j_2 + \mathbf{z}'^{g}_{2i}\mathbf{a}_2 + e_{2i}\right\}$$

$$\begin{matrix}\mathbf{Y}_1\\ \mathbf{e}_2\end{matrix}\Bigg|\,\boldsymbol{\theta} \sim N\left(\begin{pmatrix}\mathbf{X}_1\boldsymbol{\beta}_1 + \sum\limits_{j\in Q(1)\cup Q(1,2)} \mathbf{Z}^j_1\mathbf{u}^j_1 + \mathbf{Z}^g_1\mathbf{a}_1\\ \mathbf{0}\end{pmatrix}, \mathbf{R}_e \otimes \mathbf{I}_n\right) \quad (1)$$

where $\mathbf{e}_2$ with elements $(e_{2i})_{i=1,...,n}$ is a vector of residual effects of the survival trait on the log-frailty scale, which is assumed to account for the variation between individuals not otherwise accounted for by the specification of the log-frailty. Here $\lambda_i(t|\boldsymbol{\theta}, \mathbf{e}_2)$ is the hazard function of T_i conditional on model parameters $(\boldsymbol{\theta}', \mathbf{e}_2)$, where $\boldsymbol{\theta}' = (\rho, \boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, (\mathbf{u}'^j_1)_{j\in Q(1)\cup Q(1,2)}, (\mathbf{u}'^j_2)_{j\in Q(2)\cup Q(1,2)}, \mathbf{a}'_1, \mathbf{a}'_2, (\sigma^2_{1j})_{j\in Q(1)}, (\sigma^2_{2j})_{j\in Q(2)}, (\mathbf{R}_j)_{j\in Q(1,2)}, \mathbf{G}, \mathbf{R}_e)$. The set of indices $Q(1)$ define random environmental effects specific for the linear Gaussian trait, and $Q(2)$ define the set indices for time-independent and time-dependent random environmental effects specific for the survival trait. The set of indices $Q(1, 2)$ defines random environmental effects (possibly time-dependent), that are correlated between traits. The Weibull baseline hazard function is generally given as $\lambda^\rho \rho t^{(\rho-1)}$ with parameters ρ and λ. Here the term λ^ρ is the first element of the vector $\boldsymbol{\beta}_2$ (*i.e.* $\boldsymbol{\beta}_{21} = \rho \log(\lambda)$). The p_1 dimensional vector $\boldsymbol{\beta}_1$ and the p_2 dimensional vector $\boldsymbol{\beta}_2$ are regression parameters of the linear Gaussian trait and the survival trait measuring the effect of fitted covariates. The vectors $\mathbf{u}'^j_1$ of dimension q_{1j} for $j \in Q(1)$ and $\mathbf{u}'^j_2$ of dimension q_{2j} for $j \in Q(2)$ represent subject specific environmental effects common to individuals within traits. The vectors $\mathbf{u}^j_1$ and $\mathbf{u}^j_2$ of dimension q_{cj} for $j \in Q(1, 2)$ represent correlated random environmental effects for the two traits. The latter can be used to model common

environment affecting individuals that are grouped together in, for example, a pen, cage, stall etc. $\mathbf{a}_1$ and $\mathbf{a}_2$ of dimension q_g are the vectors of additive genetic effects for the two traits, where q_g is the total number of animals in the pedigree. Vectors $\mathbf{x}'_{1i}$, $\mathbf{x}'_{2i}(t)$, $(\mathbf{z}'^{j}_{1i}(t))_{j\in Q(1)}$, $(\mathbf{z}'^{j}_{2i})_{j\in Q(2)}$, $(\mathbf{z}'^{j}_{1i}, \mathbf{z}'^{j}_{2i})_{j\in Q(1,2)}$, $\mathbf{z}'^{g}_{1i}$, $\mathbf{z}'^{g}_{2i}$ are incidence arrays (possibly time-dependent) relating parameter effects to observations. Finally, $\mathbf{R}_j = \begin{bmatrix} R_{j11} & R_{j12} \\ R_{j21} & R_{j22} \end{bmatrix}$ for $j \in Q(1,2)$ are the covariance matrices of random environmental effects with the exception of residual effects, $\mathbf{G} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$ is the genetic covariance matrix, and $\mathbf{R}_e = \begin{bmatrix} R_{e11} & R_{e12} \\ R_{e21} & R_{e22} \end{bmatrix}$ is the residual covariance matrix.

The time-dependent covariates of animal i are assumed to be left-continuous and piecewise constant on the intervals $h_{i(m-1)} < t \leq h_{i(m)}$ for $m = 1, ..., M_i$, where $h_{i(0)} = 0$ and $h_{i(M_i)} = y_{2i}$, and $h_{i(m)}$ for $m = 1, ..., (M_i - 1)$ are the ordered time points at which one or more of the time-dependent covariates of animal i changes. $M_i - 1$ is the number of different time points with changes in one or more of the time-dependent covariates associated with animal i.

Prior specification. *A priori* model parameters $(\beta_{1b})_{b=1,...,p_1}$, $(\beta_{2b})_{b=1,...,p_2}$, ρ, $(\mathbf{u}^j_1, \sigma^2_{1j})_{j\in Q(1)}$, $(\mathbf{u}^j_2, \sigma^2_{2j})_{j\in Q(2)}$, $(\mathbf{u}^j_1, \mathbf{u}^j_2, \mathbf{R}_j)_{j\in Q(1,2)}$, and $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{G})$ are assumed to be mutually independent. Improper uniform priors are assigned to $(\beta_{1b})_{b=1,...,p_1}$, $(\beta_{2b})_{b=1,...,p_2}$ and ρ over their range of positive support. *A priori* the following multivariate normal distributions are assumed for environmental effects: $\mathbf{u}^j_1|\sigma^2_{1j} \sim N\left(\mathbf{0}, \sigma^2_{1j}\mathbf{I}_{q_{1j}}\right)$ for $j \in Q(1)$, $\mathbf{u}^j_2|\sigma^2_{2j} \sim N\left(\mathbf{0}, \sigma^2_{2j}\mathbf{I}_{q_{2j}}\right)$ for $j \in Q(2)$, and $(\mathbf{u}'^j_1, \mathbf{u}'^j_2)'|\mathbf{R}_j \sim N\left(\mathbf{0}, \mathbf{R}_j \otimes \mathbf{I}_{q_{cj}}\right)$ for $j \in Q(1,2)$.

The prior distribution of additive genetic effects is by assumption of the additive genetic infinitesimal model [3] assumed to be multivariate normally distributed $\left(\mathbf{a}'_1, \mathbf{a}'_2\right)'|\mathbf{G} \sim N\left(\mathbf{0}, \mathbf{G} \otimes \mathbf{A}_{q_g}\right)$, where $\mathbf{A}_{q_g}$ is the additive genetic relationship matrix.

The hyperparameters σ^2_{1j} and σ^2_{2j} are assumed *a priori* to be inverse Gamma distributed according to $\sigma^2_{1j} \sim IG(v_{1j}, f_{1j})$ for $j \in Q(1)$ and $\sigma^2_{2j} \sim IG(v_{2j}, f_{2j})$ for $j \in Q(2)$. The covariance matrices $(\mathbf{R}_j)_{j\in Q(1,2)}$, $\mathbf{G}$ and $\mathbf{R}_e$ are *a priori* assumed to be inverse Whishart distributed according to $\mathbf{R}_j \sim IW(\mathbf{F}_j, f_j)$ for $j \in Q(1,2)$, $\mathbf{G} \sim IW(\mathbf{F}_g, f_g)$ and $\mathbf{R}_e \sim IW(\mathbf{F}_e, f_e)$.

2.1.1. Joint posterior distribution

Using Bayes' theorem, the joint posterior distribution of $\boldsymbol{\theta}$ and $\mathbf{e}_2$ is obtained up to proportionality by multiplying the conditional likelihood with the prior

distribution of $\boldsymbol{\theta}$ and $\mathbf{e}_2$

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{e_2}|\mathbf{y_1}, \mathbf{y_2}, \boldsymbol{\delta_2}) &\propto p(\mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\delta}_2|\boldsymbol{\theta}, \mathbf{e}_2)p(\boldsymbol{\theta}, \mathbf{e}_2) \\ &= p(\mathbf{y}_2, \boldsymbol{\delta}_2|\boldsymbol{\theta}, \mathbf{e}_2)p(\mathbf{y}_1, \boldsymbol{\theta}, \mathbf{e}_2)p(\boldsymbol{\theta}, \mathbf{e}_2) \\ &= p(\mathbf{y}_2, \boldsymbol{\delta}_2|\boldsymbol{\theta}, \mathbf{e}_2)p(\mathbf{y}_1, \mathbf{e}_2|\boldsymbol{\theta})p(\boldsymbol{\theta}) \end{aligned} \tag{2}$$

where we in the second step use that $(\mathbf{Y}_2, \boldsymbol{\delta}_2)$ and $\mathbf{Y}_1$ are conditionally independent given $\boldsymbol{\theta}$ and $\mathbf{e}_2$. Conditional on model parameters censoring is assumed to be independent and non-informative [2] implying that $p(\mathbf{y}_2, \boldsymbol{\delta}_2|\boldsymbol{\theta}, \mathbf{e}_2) \propto \prod_i S_i(y_{2i}|\boldsymbol{\theta}, \mathbf{e}_2)\left[\lambda_i(y_{2i}|\boldsymbol{\theta}, \mathbf{e}_2)\right]^{\delta_{2i}}$, where $S_i(y_{2i}|\boldsymbol{\theta}, \mathbf{e}_2) = \exp\left\{-\int_0^{y_{2i}} \lambda_i(s|\boldsymbol{\theta}, \mathbf{e}_2)\mathrm{d}s\right\}$ is the conditional survival function. Under the assumption of a proportional Weibull log-normal animal frailty model, $p(\mathbf{y}_2, \boldsymbol{\delta}_2|\mathbf{e}_2, \boldsymbol{\theta})$ is up to proportionality given by

$$\begin{aligned} &\rho^{\sum_{i=1}^n \delta_{2i}} \left(\prod_{i=1}^n y_{2i}^{\delta_{2i}}\right)^{(\rho-1)} \exp\left\{\sum_{i=1}^n \delta_{2i}\left(\mathbf{x}'_{2i}(y_{2i})\boldsymbol{\beta}_2 + \mathbf{z}'_{2i}(y_{2i})\mathbf{u}_2 + \mathbf{z}'^{g}_{2i}\mathbf{a}_2 + e_{2i}\right)\right\} \\ &\times \exp\left\{-\sum_{i=1}^n \sum_{m=1}^{M_i} \exp\left(\mathbf{x}'_{2i}(h_{i(m)})\boldsymbol{\beta}_2 + \mathbf{z}'_{2i}(h_{i(m)})\mathbf{u}_2 + \mathbf{z}'^{g}_{2i}\mathbf{a}_2 + e_{2i}\right)\left(h^{\rho}_{i(m)} - h^{\rho}_{i(m-1)}\right)\right\} \end{aligned} \tag{3}$$

where $\mathbf{z}'_{2i}$ is a row vector with vector elements $(\mathbf{z}'^{j}_{2i})_{j\in Q(2)\cup Q(1,2)}$, and $\mathbf{u}_2$ is a column vector with vector elements $(\mathbf{u}^{j}_{2i})_{j\in Q(2)\cup Q(1,2)}$.

The conditional distribution of $\mathbf{Y}_1, \mathbf{e}_2|\boldsymbol{\theta}$ is multivariate normally distributed according to (1).

2.1.2. Conditional posterior distributions

Implementation of a Gibbs sampler involves generating samples from the fully conditional distributions of each parameter or groups of model parameters depending on whether univariate or joint updating is performed, respectively. These distributions can be obtained up to proportionality by retaining from the joint posterior distribution (2) the terms depending on the parameter of interest. The notation, where $\boldsymbol{\Delta}_{\backslash\varphi}$ denotes the parameter vector $\boldsymbol{\Delta} = (\boldsymbol{\theta}, \mathbf{e}_2)$ except the parameter φ, will be used. Here it is assumed that matrices being inverted all have full rank. The elements of $\mathbf{R}_e^{-1}$ and $\mathbf{G}^{-1}$ are denoted R_e^{ij} and G^{ij} for $i, j = 1, 2$.

The fully conditional posterior distribution of the vector of regression parameters $\boldsymbol{\beta}_1$ of the linear Gaussian trait is multivariate normal $N\left(\boldsymbol{\mu}_{\boldsymbol{\beta}_1}, \mathbf{V}_{\boldsymbol{\beta}_1}\right)$, where

$$\boldsymbol{\mu}_{\boldsymbol{\beta}_1} = \mathbf{V}_{\beta_1}\left(R_e^{11}\mathbf{X}_1'\left(\mathbf{y}_1 - \sum_{j\in Q(1)\cup Q(1,2)} \mathbf{Z}_1^j\mathbf{u}_1^j - \mathbf{Z}_1^g\mathbf{a}_1\right) + R_e^{12}\mathbf{X}_1'\mathbf{e}_2\right) \tag{4}$$

$$\mathbf{V}_{\boldsymbol{\beta}_1} = \left(R_e^{11}\mathbf{X}_1'\mathbf{X}_1\right)^{-1}.$$

The fully conditional posterior distribution of the Weibull baseline parameter ρ is up to proportionality given by

$$\rho^{\sum_{i=1}^{n}\delta_{2i}}\left(\prod_{i=1}^{n} y_{2i}^{\delta_{2i}}\right)^{(\rho-1)}$$
$$\times\exp\left\{-\sum_{i=1}^{n}\sum_{m=1}^{M_i}\exp\left(\mathbf{x}_{2i}'(h_{i(m)})\boldsymbol{\beta}_2 + \mathbf{z}_{2i}'(h_{i(m)})\mathbf{u}_2 + \mathbf{z}_{2i}'^{g}\mathbf{a}_2 + e_{2i}\right)\left(h_{i(m)}^{\rho} - h_{i(m-1)}^{\rho}\right)\right\}. \tag{5}$$

The fully conditional posterior distribution of a regression parameter β_{2b} for $b = 1, ..., p_2$ of the survival trait associated with possibly time-dependent covariates is up to proportionality given by

$$\exp\left\{\sum_{i\in\Psi(\beta_{2b})}\delta_{2i}\mathbf{x}_{2i}'(y_{2i})\boldsymbol{\beta}_2\right\}\exp\left\{-\sum_{i\in\Psi(\beta_{2b})}\sum_{m=1}^{M_i}\Omega_{im}\right\}$$

where $\Psi(\beta_{2b})$ is the set of animals affected by the bth element of $\boldsymbol{\beta}_2$ during a period of their observed lifetime, and

$$\Omega_{im} = \exp\Bigg(x_{2ib}(h_{i(m)})\boldsymbol{\beta}_{2b}$$
$$+ \sum_{k:k\neq b} x_{2ik}(h_{i(m)})\boldsymbol{\beta}_{2k} + \mathbf{z}_{2i}'(h_{i(m)})\mathbf{u}_2 + \mathbf{z}_{2i}'^{g}\mathbf{a}_2 + e_{2i}\Bigg)\left(h_{i(m)}^{\rho} - h_{i(m-1)}^{\rho}\right).$$

For $j \in Q(1)$ the fully conditional posterior distribution of the vector $\mathbf{u}_1^j$ of the linear trait is multivariate normal $N\left(\boldsymbol{\mu}_{\mathbf{u}_1^j}, \mathbf{V}_{\mathbf{u}_1^j}\right)$, where

$$\boldsymbol{\mu}_{\mathbf{u}_1^j} = \mathbf{V}_{\mathbf{u}_1^j}\left(R_e^{11}\mathbf{Z}_1'^{j}\left(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1 - \sum_{k\in Q(1)\cup Q(1,2)\setminus\{j\}} \mathbf{Z}_1^k\mathbf{u}_1^k - \mathbf{Z}_1^g\mathbf{a}_1\right) + R_e^{12}\mathbf{Z}_1'^{j}\mathbf{e}_2\right)$$

$$\mathbf{V}_{\mathbf{u}_1^j} = \left(R_e^{11}\mathbf{Z}_1'^{j}\mathbf{Z}_1^j + \mathbf{I}_{q_{1j}}\sigma_{1j}^{-2}\right)^{-1}. \tag{6}$$

For $j \in Q(1,2)$ the fully conditional posterior distribution of the vector $\mathbf{u}_1^j$ of the linear trait is multivariate normal $N\left(\boldsymbol{\mu}_{\mathbf{u}_1^j}, \mathbf{V}_{\mathbf{u}_1^j}\right)$, where

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{u}_1^j} &= \mathbf{V}_{\mathbf{u}_1^j}\left(R_e^{11}\mathbf{Z}_1'^j\left(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1 - \sum_{k\in Q(1)\cup Q(1,2)\setminus\{j\}} \mathbf{Z}_1^k\mathbf{u}_1^k - \mathbf{Z}_1^g\mathbf{a}_1\right)\right.\\
&\qquad \left. + R_e^{12}\mathbf{Z}_1'^j\mathbf{e}_2 - R_e^{12}\mathbf{u}_2^j\right)\\
\mathbf{V}_{\mathbf{u}_1^j} &= \left(R_e^{11}\mathbf{Z}_1'^j\mathbf{Z}_1^j + \mathbf{R}_j^{11}\mathbf{I}_{q_{cj}}\right)^{-1}.
\end{aligned} \tag{7}$$

For $j \in Q(2)$ the fully conditional posterior distribution of a random effect u_{2w}^j for $w = 1, ..., q_{2j}$ of the survival trait associated with possibly time-dependent incidence arrays is up to proportionality given by

$$\exp\left\{u_{2w}^j d\left(u_{2w}^j\right)\right\}\exp\left\{-\sum_{i\in\Psi(u_{2w}^j)}\sum_{m=1}^{M_i}\Omega_{im}^j\right\}\exp\left\{-\frac{\left(u_{2w}^j\right)^2}{2\sigma_{2j}^2}\right\} \tag{8}$$

where

$$\begin{aligned}
\Omega_{im}^j = \exp\Bigg(&\mathbf{x}_{2i}'(h_{i(m)})\boldsymbol{\beta}_2 + z_{2iw}^j(h_{i(m)})u_{2w}^j + \sum_{k:k\neq w} z_{2ik}^j(h_{i(m)})u_{2k}^j\\
&+ \sum_{l\in Q(2)\cup Q(1,2)\setminus\{j\}} \mathbf{z}_{2i}'^l(h_{i(m)})\mathbf{u}_2^l + \mathbf{z}_{2gi}'\mathbf{a}_2 + e_{2i}\Bigg)\left(h_{i(m)}^{\rho} - h_{i(m-1)}^{\rho}\right)
\end{aligned}$$

and $d\left(u_{2w}^j\right)$ is the number of animals that have an observed lifetime, y_{2i}, with $z_{2iw}'(y_{2i}) = 1$, and $\Psi(u_{2w}^j)$ is the set of animals affected by the wth element of $\mathbf{u}_2^j$ during a period of their observed lifetime.

For $j \in Q(1,2)$ the fully conditional posterior distribution of a time-independent random effect u_{2w}^j for $w = 1, ..., q_{cj}$ of the survival trait is up to proportionality given by

$$\exp\left\{u_{2w}^j d\left(u_{2w}^j\right)\right\}\exp\left\{-\sum_{i\in\Psi(u_{2w}^j)}\sum_{m=1}^{M_i}\Omega_{im}^j\right\}\exp\left\{-\frac{1}{2}\left(R_j^{22}\left(u_{2w}^j\right)^2 + 2R_j^{12}u_{1w}^j u_{2w}^j\right)\right\}. \tag{9}$$

The fully conditional posterior distribution of the vector of additive genetic effects $\mathbf{a}_1$ of the linear trait is multivariate normal $N(\boldsymbol{\mu}_{\mathbf{a}_1}, \mathbf{V}_{\mathbf{a}_1})$, where

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{a}_1} &= \mathbf{V}_{\mathbf{a}_1}\left(R_e^{11}\mathbf{Z}_{1g}'(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1 - \mathbf{Z}_1\mathbf{u}_1) + R_e^{12}\mathbf{Z}_{1g}'\mathbf{e}_2 - G^{12}\mathbf{A}^{-1}\mathbf{a}_2\right)\\ \mathbf{V}_{\mathbf{a}_1} &= \left(R_e^{11}\mathbf{Z}_1'^{g}\mathbf{Z}_1^{g} + \mathbf{G}^{11}\mathbf{A}^{-1}\right)^{-1}.\end{aligned} \tag{10}$$

For w, an animal with records, the fully conditional posterior distribution of an additive genetic effect a_{2w} of the survival trait is up to proportionality given by

$$\begin{aligned}&\exp\left\{a_{2w}\delta_{2g(w)}\right\}\\ &\times\exp\left\{-\sum_{m=1}^{M_{g(w)}}\exp\left(\mathbf{x}_{2g(w)}'(h_{g(w)(m)})\boldsymbol{\beta}_2 + \mathbf{z}_{2g(w)}'(h_{g(w)(m)})\mathbf{u}_2 + a_{2w} + e_{2g(w)}\right)\right.\\ &\left.\times\left(h^{\rho}_{g(w)(m)} - h^{\rho}_{g(w)(m-1)}\right)\right\}\\ &\times\exp\left\{-\frac{1}{2}G^{22}\left(a_{2w}^2A^{ww} + 2a_{2w}\sum_{k:k\neq w}a_{2k}A^{wk}\right)\right\}\\ &\times\exp\left\{-G^{21}\left(a_{2w}\sum_{k}a_{1k}A^{wk}\right)\right\}\end{aligned} \tag{11}$$

where $g(w)$ is a function relating an animal effect to its observation.

For w, an animal without records, the fully conditional posterior distribution of an additive genetic effect a_{2w} of the survival trait is normally distributed

$$N\left(-\frac{1}{G^{22}A^{ww}}\left(G^{22}\sum_{k:k\neq w}a_{2k}A^{wk} + G^{21}\sum_{k}a_{1k}A^{wk}\right), \left(G^{22}A^{ww}\right)^{-1}\right). \tag{12}$$

The fully conditional posterior distribution of a residual effect e_{2i} for $i = 1, ..., n$ of the survival trait is up to proportionality given by

$$\begin{aligned}&\exp\{e_{2i}\delta_{2i}\}\\ &\times\exp\left\{-\sum_{m=1}^{M_i}\exp\left(\mathbf{x}_{2i}'(h_{i(m)})\boldsymbol{\beta}_2 + \mathbf{z}_{2i}'(h_{i(m)})\mathbf{u}_2 + \mathbf{z}_{2i}'^{g}\mathbf{a}_2 + e_{2i}\right)\left(h^{\rho}_{i(m)} - h^{\rho}_{i(m-1)}\right)\right\}\\ &\times\exp\left\{-\frac{1}{2}\left(R_e^{22}e_{2i}^2 + 2R_e^{12}e_{2i}\left(y_{1i} - \mathbf{x}_{1i}'\boldsymbol{\beta}_1 - \mathbf{z}_{1i}'\mathbf{u}_1 - \mathbf{z}_{1i}'^{g}\mathbf{a}_1\right)\right)\right\}.\end{aligned} \tag{13}$$

The fully conditional posterior distribution of variance components σ^2_{1j} for $j \in Q(1)$ and of σ^2_{2j} for $j \in Q(2)$ are inverse Gamma distributed, according to.

$$\begin{aligned}\sigma^2_{1j}|(\boldsymbol{\theta}, \mathbf{e}_2)_{\backslash\sigma^2_{2j}}, data &\sim IG\left[q_{1j}/2 + v_{1j}, \left(f^{-1}_{1j} + \mathbf{u}'^{j}_1\mathbf{u}^j_1/2\right)^{-1}\right] \\ \sigma^2_{2j}|(\boldsymbol{\theta}, \mathbf{e}_2)_{\backslash\sigma^2_{2j}}, data &\sim IG\left[q_{2j}/2 + v_{2j}, \left(f^{-1}_{2j} + \mathbf{u}'^{j}_2\mathbf{u}^j_2/2\right)^{-1}\right].\end{aligned} \tag{14}$$

The fully conditional posterior distribution of covariance matrices $(\mathbf{R}_j)_{j\in Q(1,2)}$, $\mathbf{G}$ and $\mathbf{R}_e$ are inverse Whishart distributed, according to

$$\begin{aligned}\mathbf{R}_j|(\boldsymbol{\theta}, \mathbf{e}_2)_{\backslash\mathbf{R}_j}, data &\sim IW\left[\left(\mathbf{F}^{-1}_j + \mathbf{S}_j\right)^{-1}, f_j + q_{cj}\right] \\ \mathbf{G}|(\boldsymbol{\theta}, \mathbf{e}_2)_{\backslash\mathbf{G}}, data &\sim IW\left[\left(\mathbf{F}^{-1}_g + \mathbf{S}_g\right)^{-1}, f_g + q_g\right] \\ \mathbf{R}_e|(\boldsymbol{\theta}, \mathbf{e}_2)_{\backslash\mathbf{R}_e}, data &\sim IW\left[\left(\mathbf{F}^{-1}_e + \mathbf{S}_e\right)^{-1}, f_e + n\right]\end{aligned} \tag{15}$$

where $\mathbf{S}_j = \begin{bmatrix}\mathbf{u}'^{j}_1\mathbf{u}^j_1 & \mathbf{u}'^{j}_1\mathbf{u}^j_2 \\ \mathbf{u}'^{j}_2\mathbf{u}^j_1 & \mathbf{u}'^{j}_2\mathbf{u}^j_2\end{bmatrix}$ for $j \in Q(1,2)$,

$\mathbf{S}_g = \begin{bmatrix}\mathbf{a}'_1\mathbf{A}^{-1}\mathbf{a}_1 & \mathbf{a}'_1\mathbf{A}^{-1}\mathbf{a}_2 \\ \mathbf{a}'_2\mathbf{A}^{-1}\mathbf{a}_1 & \mathbf{a}'_2\mathbf{A}^{-1}\mathbf{a}_2\end{bmatrix}$ and $\mathbf{S}_e = \begin{bmatrix}\mathbf{e}'_1\mathbf{e}_1 & \mathbf{e}'_1\mathbf{e}_2 \\ \mathbf{e}'_2\mathbf{e}_1 & \mathbf{e}'_2\mathbf{e}_2\end{bmatrix}$.

2.2. Model for missing observations

Missing observations for one of the two traits are often a fact that must be dealt with in field data. In what follows, we assume that such missingness is completely at random [20]. In Bayesian analysis, this type of misingness often is dealt with by augmenting the joint posterior distribution with the residuals associated with missing observations (*e.g.* [31]). The augmented residuals are treated as unknown parameters, so at each iteration of the Gibbs sampler the augmented residual effects are generated from their fully conditional posterior distributions, which are specified below. Note that this is the only additional step necessary in order to account for missing data. Now, let $\mathbf{e}_{m1}$ and $\mathbf{e}_{m2}$ be vectors of size n_{m1} and n_{m2} including residual effects associated with missing observations of the linear Gaussian trait and survival trait, respectively. The augmented joint posterior distribution is given by

$$p(\boldsymbol{\theta}, \mathbf{e}_2, \mathbf{e}_{m1}, \mathbf{e}_{m2}|\mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\delta}_2) \propto p(\mathbf{y}_2, \boldsymbol{\delta}_2|\mathbf{e}_2, \boldsymbol{\theta}, \mathbf{e}_{m2})p(\mathbf{y}_1, \mathbf{e}_2, \mathbf{e}_{m1}, \mathbf{e}_{m2}|\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

Combining this joint posterior with the previous model specification, it follows that the fully conditional posterior distribution of an augmented residual effect for a missing linear Gaussian record is normally distributed $N(\mu_{e_{m1j}}, V_{e_{m1j}})$ for $j = 1, ..., n_{m1}$, where $\mu_{e_{m1j}} = (R_{e_{12}}/R_{e_{22}})e_{2j}$ and $V_{e_{m1j}} = R_{e_{11}} - R^2_{e_{12}}/R_{e_{22}}$. Similarly, it follows that the fully conditional posterior distribution of an augmented residual effect for a missing survival record is normally distributed $N(\mu_{e_{m2j}} V_{e_{m2j}})$ for $j = 1, ..., n_{m2}$, where $\mu_{e_{m2j}} = (R_{e_{12}}/R_{e_{11}})(y_{1j} - \mathbf{x}'_{1i}\boldsymbol{\beta}_1 - \mathbf{z}'_{1i}\mathbf{u}_1 - \mathbf{z}'^{g}_{1i}\mathbf{a}_1)$ and $V_{e_{m2j}} = R_{e_{22}} - R^2_{e_{12}}/R_{e_{11}}$.

2.3. Implementation of the Gibbs sampler

A Gibbs sampler of the bivariate animal model (1) with no random environmental effects (*i.e.* no $\mathbf{u}'s$) was implemented in Fortran 90 for data without missing observations for any of the two traits. Inferences of model parameters were based on a single chain in which systematic effects of the linear Gaussian trait were the only parameters that were jointly updated. Sampling from closed form fully conditional posterior distributions were performed using standard methods. The fully conditional posterior distributions of $(\rho, (\beta_{2i})_{i=1,...,p_2}, (a_{2i})_{i=1,...,n}, (e_{2i})_{i=1,...,n})$ did not reduce to well-known distributions and other sampling procedures were required. In a first implementation, adaptive rejection sampling (ARS) [13] was used to update all these parameters univariately. In order to use ARS the distributions from which one wants to sample must be log-concave. This condition was satisfied for all these parameters with the exception of a special case for ρ. For ρ, the log-concavity condition is only satisfied if the time points at which possible time-dependent covariates change status are greater than one unit of measurement. ARS was easily implemented applying the subroutine provided by Wild and Gilks [33]. Unfortunately, slow mixing properties were observed especially for the two Weibull baseline parameters ρ and β_{21} and the residual variance component R_{e22}.

In the final implementation and in order to improve mixing, we changed from univariate to a joint updating of ρ and β_{21} by introducing a Metropolis-Hasting step within the Gibbs sampler [15, 24]. As a proposal distribution, we used a large sample bivariate normal approximation (see *e.g.* [31]) given by

$$N\left(\widehat{\boldsymbol{\gamma}}, -\left[\left|\frac{\partial^2 l(\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right|_{\boldsymbol{\gamma}=\widehat{\boldsymbol{\gamma}}}\right]^{-1}\right) \tag{16}$$

where $\boldsymbol{\gamma} = (\rho, \beta_{21})'$ and $\widehat{\boldsymbol{\gamma}}$ is the value of the vector at the joint mode of $l(\boldsymbol{\gamma}) \propto p(\rho, \beta_{21}|\boldsymbol{\Delta}_{\setminus(\rho,\beta_{21})}, data)$. Preliminary simulation studies showed that the

performance of the proposal distribution could be improved by lowering the correlation between ρ and β_{21}. For all the analyses performed in this study we simply multiplied the covariance elements of 16 by a factor of 0.7. The same studies also showed that the implemented Metropolis-Hasting step generally worked satisfactorily, except for a few occasions where the chain remained in the same state for up to about 10 000 iterations. This problem of an occasionally high rejection rate was alleviated by changing to one round of univariate updating of ρ and β_{21} using ARS if the chain for ρ and β_{21} remained in the same state for 100 iterations. The final implementation led to substantially better mixing of the three parameters ρ, β_{21} and R_{22} compared to the initial implementation where ρ and β_{21} were updated univariately based on ARS (up to a 15 fold increase in the number of effective samples).

3. RESULTS

The bivariate model was illustrated in two simulation studies in which the models used to simulate and analyze data were the same. Thus, the focus was on validating the estimation of parameters under conditions where all model assumptions were satisfied, and not on issues related to the robustness of the model.

3.1. Simulation study 1

The records of 4000 animals after 100 unrelated sires each having 40 offspring were simulated from the bivariate model (1). For both traits, the model included a systematic effect with two levels (β_{11},β_{12}) and (β_{22},β_{23}), an animal effect, and a residual effect. In addition, the model of the survival trait included the two Weibull baseline parameters ρ and $\beta_{21} = \rho \log(\lambda)$. Lifetimes higher than or equal to 1510 were right censored, leading to a data set with approximately 11% right censored records. This simple censoring mechanism satisfies the independent and non-informative censoring assumption and it is easy to implement. Data were analyzed with the model that was used to generate data parameterized both as an animal and as a sire model. Note that given a half-sib design, the animal model was equivalent to the sire model. The sire model has not previously been specified in this paper. However, when applying the notation provided earlier, the sire covariance matrix ($\mathbf{G}_s$) and the residual covariance matrix ($\mathbf{R}_{\bar{e}}$) of the sire model are defined by ($\mathbf{G}_s + \mathbf{R}_{\bar{e}}$) = ($\mathbf{R}_e + \mathbf{G}$) and $\mathbf{G}_s = \mathbf{G}/4$. The remaining model parameters are identical to parameters of the animal model [1]. Improper uniform priors were used for $\mathbf{G}_s$ and $\mathbf{R}_{\bar{e}}$.

The model parameters used in the simulation of data and results from the Bayesian analysis are given in Table I.

Starting values of the Gibbs sampler of the additive genetic covariance matrix, the residual covariance matrix and the two Weibull parameters were set equal to the true values used in the simulation of data. All of the remaining model parameters were started at zero. For the regression parameters of the survival trait, β_{23} was used as a reference group and was fixed to zero. Although the Gibbs sampler seemingly converged within the first few iterations, the first 10 000 iterations of the Gibbs sampler were considered as burnin for all model parameters. Altogether, 4900 samples of model parameters were saved with a sampling interval of 100; *i.e.* a total of 500 000 rounds were run. The effective number of samples (N_e) of each parameter was calculated by the method of batching based on 30 batches (*e.g.* [31]). To facilitate comparison between the animal and sire parameterization, the results obtained from the sire model were transformed to the equivalent animal model at each saved iteration of the Gibbs sampler. To perform the 500 000 iterations the program ran for 7 to 8 days on a IBM POWRE3 4-way 375 nHz. Silver node computer.

Results

The results of simulation study 1 showed that the estimated marginal posterior distributions covered the parameter values used for simulating data well (Tab. I). All the true values were within the 95% central posterior density (CPD) regions [12] defined by the 2.5% and 97.5% percentiles. The posterior summary statistics given for the sire and the animal model were very similar. The only substantial difference between the two models was the number of effective samples, which in particular for the elements of the covariance matrices was higher for the sire model. This suggests that the sire parameterization provides better mixing properties. Even though the posterior uncertainty for the additive genetic and residual correlations was relatively high, the marginal posterior densities still assigned relative high probability to values in the neighborhood of the true values used in the simulation of data.

3.2. Simulation study 2

In simulation study 2, the data generated from the bivariate model (1) consisted of records of 5000 animals after 100 unrelated sires each having 50 offspring (balanced half-sib design). Again, data were analyzed with the model that was used to generate data, but in this study we only applied a sire model.

Table I. Simulation study 1, true value, marginal posterior mode, mean, 2.5% and 97.5% percentiles, and effective sample size (N_e) of fixed effects of the linear Gaussian trait (β_{11}, β_{12}), of Weibull parameters (ρ, β_{21}), of time-independent fixed effects (β_{22}, $\beta_{23} = 0$), of additive genetic variance components (G_{11}, G_{22}), and correlation (ρ_G), of residual variance components (R_{e11}, R_{e22}), and correlation (ρ_{R_e}), (*) indicates that results were obtained from the sire model and thereafter transformed to parameters of the equivalent animal model.

Parameter	True	Mode	Mean	2.5%	97.5%	N_e
β_{11}	860.00	860.80	860.80	858.16	863.34	3618
β^*_{11}	–	860.92	860.78	858.18	863.44	4068
β_{12}	920.00	919.47	919.67	917.11	922.28	4530
β^*_{12}	–	919.83	919.67	917.13	922.22	4042
ρ	1.80	1.79	1.75	1.64	1.87	329
ρ^*	–	1.75	1.74	1.63	1.88	347
β_{21}	−12.00	−11.74	−11.60	−12.44	−10.90	336
β^*_{21}	–	−11.42	−11.56	−12.44	−10.84	340
β_{22}	0.20	0.19	0.19	0.090	0.29	2001
β^*_{22}	–	0.19	0.19	0.088	0.28	1825
G_{11}	500	481.50	491.34	344.31	688.82	1584
G^*_{11}	–	477.62	492.87	345.79	691.57	7815
G_{22}	0.80	0.72	0.75	0.49	1.07	1201
G^*_{22}	–	0.77	0.75	0.49	1.11	1369
ρ_G	−0.15	−0.18	−0.15	−0.38	0.090	875
ρ^*_G	–	−0.14	−0.15	−0.39	0.10	13758
R_{e11}	625.00	649.72	633.89	479.40	754.12	1742
R^*_{e11}	–	677.68	633.47	477.68	757.46	8714
R_{e22}	0.4	0.29	0.35	0.098	0.66	201
R^*_{e22}	–	0.34	0.32	0.018	0.67	738
ρ_{R_e}	−0.49	−0.51	−0.60	−0.93	−0.33	223
$\rho^*_{R_e}$	–	−0.52	−0.60	−0.93	−0.33	2757

In addition to the model studied in simulation study 1, we included a time-dependent systematic effect with three levels ($\beta_{24}, \beta_{25}, \beta_{26}$), where β_{25} was used as the reference group and fixed to zero. Also note that in this second simulation study, data was simulated with a positive additive genetic correlation (0.24) and a negative residual correlation (−0.24) (Tab. II).

The time-dependent covariates changed status according to the following fixed time points (90, 200, 305, 395, 505, 610, 700, 810, 915, 1005, 1115, 1220, 1310, 1420, 1520). β_{24} (β_{25}, β_{26}) was the systematic effect associated

Table II. Simulation study 2, true value, marginal posterior mode, mean, 2.5% and 97.5% percentiles, and effective sample size (N_e) of fixed effects of linear Gaussian trait (β_{11}, β_{12}), of Weibull parameters (ρ, β_{21}), of time-independent fixed effects (β_{22}, $\beta_{23} = 0$), of time-dependent fixed effects ($\beta_{24}, \beta_{25} = 0, \beta_{26}$), of additive genetic sire variance components (G_{s11}, G_{s22}), and correlation (ρ_{G_s}), of residual variance components ($R_{\bar{e}11}$, $R_{\bar{e}22}$), and correlation ($\rho_{R_{\bar{e}}}$).

Parameter	True	Mode	Mean	2.5%	97.5%	N_e
β_{11}	860.00	861.20	860.84	858.32	863.37	1358
β_{12}	920.00	919.48	919.23	916.67	921.72	1979
ρ	1.80	1.73	1.80	1.67	1.94	167
β_{21}	−12.00	−11.51	−12.06	−13.00	−11.22	174
β_{22}	0.30	0.26	0.26	0.17	0.36	938
β_{24}	−0.14	−0.14	−0.22	−0.22	−0.058	1140
β_{26}	0.3	0.32	0.32	0.25	0.38	1423
G_{s11}	100.00	124.98	127.17	89.84	172.08	1827
G_{s22}	0.10	0.086	0.097	0.058	0.15	1114
ρ_{G_s}	0.24	0.12	0.16	−0.10	0.43	1232
$R_{\bar{e}11}$	1000.00	974.66	971.00	933.77	1010.04	3009
$R_{\bar{e}22}$	1.00	0.89	1.01	0.69	1.40	183
$\rho_{R_{\bar{e}}}$	−0.24	−0.21	−0.22	−0.27	−0.17	1250

with the 1st (2th, 3th), 4th (5th, 6th), 7th (8th, 9th) and 10th (11th, 12th) time period. The time-dependent systematic effect may be thought of as a "stage of lactation effect" applied in many studies addressing longevity of dairy cows [9]. The true conditional distribution of lifetimes is given in Figure 1, which shows the hazard function of the first 12 time periods conditional on a zero value for the time-independent systematic effect, the sire effect and the residual effect. The empirical distribution of simulated lifetimes is given in Figure 2. Lifetimes higher than or equal to 1520 were right censored, leading to a data set with approximately 11% right censored records. The model parameters used to simulate data and the results from the Bayesian analysis are given in Table II. The starting values of the Gibbs sampler were as described in simulation study 1. The interval between saved sampled values was 100, the length of burnin was 10 000, and the Gibbs sampler was run for a total of 200 000 iterations giving 1900 saved samples of each parameter and functions thereof. To perform the 200 000 iterations, the program ran for 5 to 6 days on the same computer as simulation study 1.

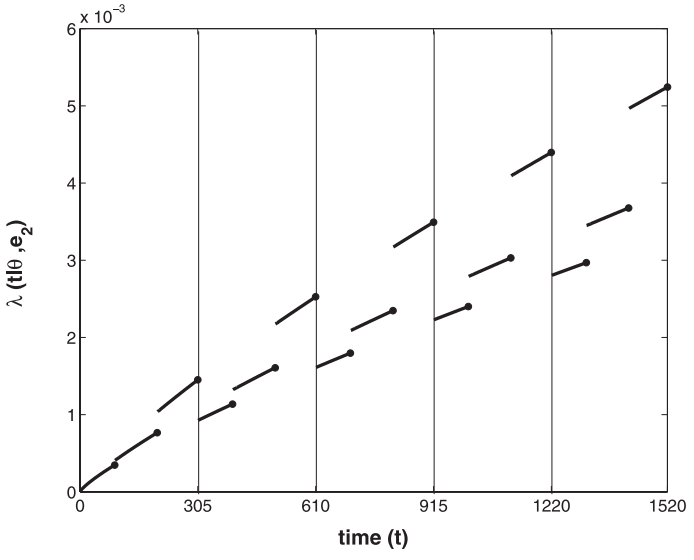

Figure 1. Hazard function based on true values conditional on a zero value for the systematic time-independent effect, the sire effect and the residual effect.

Results

In agreement with simulation study 1, the results of simulation study 2 also showed that the estimated marginal posterior distributions covered the parameter values used for simulating data well (Tab. II). There was close agreement between the marginal posterior mode (mean) of the time-dependent systematic effects and the true values. Note also that the positive additive genetic correlation and the negative residual correlation could be correctly inferred from the marginal posterior summary statistics.

4. DISCUSSION

In this study we describe a Gibbs sampler for joint Bayesian analysis of a linear Gaussian trait and a survival trait. The method was tested in two simulation studies, which both established that the estimated marginal posterior distributions covered the true values used in the simulation of data well. In conclusion, model parameters and functions thereof including the genetic correlation can be correctly inferred applying the method proposed.

The bivariate model was derived under the assumption of an additive genetic infinitesimal model [3, 11]. The additive genetic correlation is a measure of the extent to which the two traits are controlled by the same genes. Therefore, it

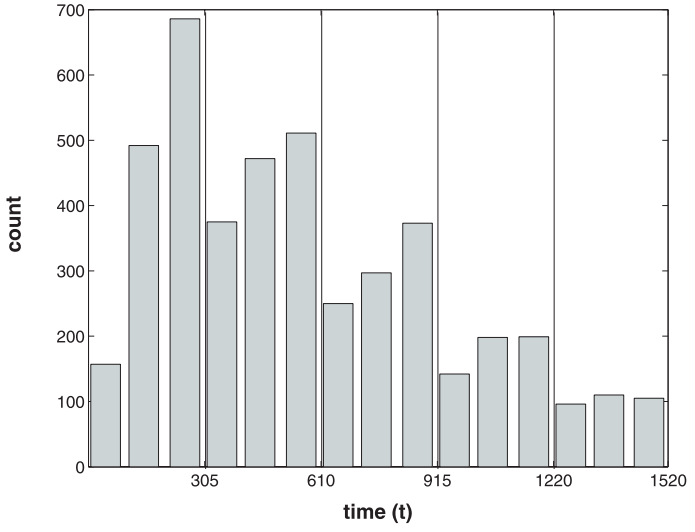

Figure 2. Number of observed lifetimes in the periods defined by the time-dependent systematic effect.

provides information about how selection for a survival trait simultaneously may affect a linear Gaussian trait and vice versa [10, 23]. The bivariate model allows for a more accurate genetic evaluation of animals for a correlated linear Gaussian trait and a survival trait owing to the shared information between traits [10, 23].

So far additive genetic correlations have mainly been approximated by product-moment correlations between estimated sire effects obtained from univariate analyses of the individual traits (*e.g.* [16, 26, 28]). In order to investigate the performance of this simple approach, we applied it to the two simulated data sets. The univariate analyses were performed using the model proposed in this study by constraining the additive genetic and residual covariances to zero. In each saved round of the Gibbs sampler the product-moment correlations between sire effects and residual effects were calculated to obtain the marginal posterior distributions of the product-moment correlations.

The results from these analyses indicated clearly that the approximative method can be very imprecise. For the first simulation study, the 95% CPD region of the additive genetic product-moment correlation ranged between −0.24 and 0.01 (true value; −0.15), and the residual product-moment correlation ranged between −0.16 and −0.10 (true value; −0.49). The credibility interval for the additive genetic correlations included the true value, whereas this was not the case for the residual correlation. For the second simulation study, the

95% CPD region of the additive genetic product-moment correlation ranged between −0.070 and 0.20, and of the residual product-moment correlation ranged between −0.11 and −0.063. None of these two credibility intervals include the corresponding true values of 0.24 for the genetic correlation and −0.24 for the residual correlation.

Gibbs sampling combined with ARS is conceptually straightforward and easily implemented applying the subroutine provided by Wild and Gilks [33]. A sufficient condition for applying ARS is that the distribution from which one wants to sample is log-concave. Gilks and Best [14] extended this ARS algorithm to deal with non-log-concave distributions. In this extended version of ARS, the piecewise exponential envelope function of the ARS algorithm takes the role of the proposal distribution in the Metropolis-Hastings algorithm from which generated candidate values are accepted according to the acceptance probability [15, 24].

The updating strategy chosen here for implementing the Gibbs sampler is just one out of many possible. In the first implementation in which all parameters without closed form fully conditional distributions were updated univariately based on ARS, we observed slow mixing properties of the two Weibull baseline parameters ρ, β_{21} and the residual variance component on the log-frailty scale R_{e22}. A plausible explanation may be that these parameters in the posterior distribution are highly correlated [22, 29]. On the contrary to univariate updating, joint updating takes advantage of the posterior correlation structure induced by the model and the data. For a broad class of models, mixing has been improved by changing to joint updating of highly correlated parameters [21, 22, 29]. However, no general rule exists and Liu *et al.* [22] and Roberts and Sahu [29] provide examples for non-hierarchical models where joint updating actually slowed down mixing. In the final implementation of the model, we changed to joint updating of ρ, β_{21} by introducing a Metropolis-Hastings step within the Gibbs sampler [15, 24]. The joint updating of the two Weibull parameters was observed to considerably improve mixing properties of the Gibbs sampler.

The computation time given for the simulation studies analyzed in this study is most likely not representative of a very efficient implementation of the model. Therefore, in order to evaluate the potential use of the model in large scale evaluation, more studies are required. However, unpublished simulation studies showed that under the assumption of known Weibull parameter rho and variance components, the mixing of the Gibbs sampler was significantly improved. This suggests that the number of required iterations necessary to obtain representative samples of genetic effects may be much smaller than 200 000,

leaving us positive with respect to the application of the model in large scale genetic evaluation.

In this study, the additive genetic effect of the survival trait was assumed to be time-independent. It is thereby assumed that an animal is born at a certain level of the relative genetic effect and stays at this level for all its lifetime. A more reasonable way of modeling the genetic effect would allow for age-related changes as suggested by Ducrocq [6]. As an example, consider longevity of dairy cows. Here it seems likely that the risk of culling involves different genes at different lactations and at different stages within lactation. An age-related genetic effect can be modeled by assuming that the genetic effect is a time-dependent random effect. For example, suppose that lactations are divided into two periods with associated additive genetic effects of the survival trait given by the two vectors $\mathbf{a}_{21}$, $\mathbf{a}_{22}$. From genetic theory, the three additive genetic effects associated with the linear Gaussian trait and the survival trait are multivariate normal distributed $\left(\mathbf{a}_1', \mathbf{a}_{21}', \mathbf{a}_{22}'\right)' | \mathbf{G} \sim N\left(\mathbf{0}, \mathbf{G} \otimes \mathbf{A}_{q_g}\right)$, where $\mathbf{G}$ is a 3×3 genetic covariance matrix. If an inverse Whishart distribution is used as prior for $\mathbf{G}$ then the fully conditional distribution of $\mathbf{G}$ will also be inverse Whishart distributed.

It follows that the framework given in this study still applies for drawing inferences in the case of time-dependent additive genetic effects, although the joint posterior distribution would differ slightly as well as the implementation.

In a very similar way the bivariate model can be extended with one or more dependent random effects including, for example, a maternal additive genetic effect, which is relevant for studying *e.g.* postnatal survival. Along the same line it should be straightforward to fit a QTL effect, which allows searching for areas on chromosomes with large effect on a survival trait and a linear Gaussian trait jointly.

A Weibull distributed baseline hazard function restricts the form of the function to be either monotone increasing, decreasing or constant. An alternative is to use the more general semi-parametric Cox model [5] in which the baseline hazard function is left completely arbitrary analogue to the univariate frailty model proposed by Korsgaard *et al.* [17]. Finally, the bivariate model can be extended to a multivariate Bayesian analysis of several traits jointly. One example is to extend the bivariate model with an ordered categorical trait by assuming that the additive genetic and residual effects on the linear liability scale, on the log-frailty scale, and of the linear Gaussian trait are multivariate normally distributed. In principle it should be straightforward to generalize the bivariate model to an arbitrary number of ordered categorical characters, survival traits and linear Gaussian traits along the framework given in this study.

REFERENCES

[1] Andersen A.H., Korsgaard I.R., Jensen J., Idenfiability of parameters in -and equivalence of animal and sire models for Gaussian and threshold characters, traits following a Poisson mixed model and survival traits, Department of Theoretical Statistics, University of Aahus, Research reports 417 (2000) 1–36.

[2] Andersen P.K., Borgan Ø., Gill R.D., Keiding N., Statistical models based on counting processes, Springer, New York, 1992.

[3] Bulmer M.G., The mathematical theory of quantitative genetics, Clarendon Press, Oxford, 1980.

[4] Carlin B.P., Louis T.A., Bayes empirical bayes methods for data analysis, Chapman and Hall, Boca Raton, 2000.

[5] Cox D.R., Regression models and life tables, J. Royal Stat. Soc. Series B 34 (1972) 187–220.

[6] Ducrocq V., Topics that may deserve further attention in survival analysis applied to dairy cattle breeding -some suggestions, in: Proceedings of the 21th Interbull Meeting, May, 1999, Vol. 21, Jouy-en-Josas, pp. 181–189.

[7] Ducrocq V., Casella G., A Bayesian analysis of mixed survival models, Genet. Sel. Evol. 28 (1996) 505–529.

[8] Ducrocq V., Sölkner J., "The Survival Kit V3.0", a package for large analyses of survival data, in: Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, 11–16 January 1998, Vol. 27, University of New England, Armidale, pp. 447–450.

[9] Ducrocq V., Quaas R.L., Pollak E.J., Casella G., Length of productive life of dairy cows. 2. Variance component estimation and sire evaluation, J. Dairy Sci. 71 (1988) 3071–3079.

[10] Falconer D.S., Mackay T.F.C., Introduction to quantitative genetics, Longman Group Ltd., Harlow, 1996.

[11] Fisher B.A., The correlation between relatives on the supposition of Mendelian inheritance, Transactions of the Royal Society of Edinburgh 52 (1918) 399–433.

[12] Gelman A., Carlin J.B., Stern H.S., Rubin D.B., Bayesian data analysis, Chapman & Hall, Boca Raton, 1995.

[13] Gilks W.R., Wild P., Adaptive rejective sampling for Gibbs sampling, Appl. Statist. 41 (1992) 337–348.

[14] Gilks W.R., Best N.G., Adaptive rejection Metropolis sampling within Gibbs sampling, Appl. Statist. 44 (1995) 455–472.

[15] Hastings W.K., Monte Carlo sampling methods using Markov chains and their application, Biometrika 57 (1970) 97–109.

[16] Henryon M., Berg P., Jensen J., Andersen S., Genetic variation for resistance to clinical and subclinical diseases exist in growing pigs, Anim. Sci. 73 (2001) 375–387.

[17] Korsgaard I.R., Madsen P., Jensen J., Bayesian inference in the semi-parametric log normal frailty model using Gibbs sampling, Genet. Sel. Evol. 30 (1998) 241–256.

[18] Korsgaard I.R., Lund M.S., Sorensen D., Gianola D., Madsen P., Jensen J., Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical, and binary traits using Gibbs sampling, Genet. Sel. Evol. 35 (2003) 159–183.
[19] Larroque H., Ducrocq V., An indirect approach for estimation of genetic correlations between longevity and other traits, in: Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, 11–16 January 1998, Vol. 27, University of New England, Armidale, pp. 128–136.
[20] Little R.J.A., Rubin D.B., Statistical analysis with missing data, Wiley, New York, 1987.
[21] Liu J.S., The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem, J. Am. Stat. Assoc. 89 (1994) 958–966.
[22] Liu J.S., Wong H.W., Kong A., Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes, Biometrika 81 (1994) 27–40.
[23] Lynch M., Walsh B., Genetics and analysis of quantitative traits, Sinauer Associates, Sunderland, 1998.
[24] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E., Equations of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1092.
[25] Mrode R.A., Linear models for the prediction of animal breeding values, CAB International, New York, 2000.
[26] Neerhoof H.J., Madsen P., Ducrocq V.P., Vollema A.R., Jensen J., Korsgaard I.R., Relationships between mastitis and functional longevity in Danish Black and White dairy cattle estimated using survival analysis, J. Dairy. Sci. 83 (2000) 1064–1071.
[27] Patterson H.D., Thompson R., Recovery of inter-block information when block sizes are unequal, Biometrika 58 (1971) 545–554.
[28] Pedersen O.M., Nielsen U.S., Avlsværdital for holdbarhed, Avlsværdivurdering. Forbedringer indført i 2000–2001, Landbrugets Rådgivningscenter, Skejby, Rapport 99 (2002) 42–69.
[29] Roberts G.O., Sahu S.K., Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler, J. Royal Stat. Soc. Series B 59 (1997) 291–317.
[30] Schaeffer L.R., Multiple-country comparison of dairy sires, J. Dairy Sci. 77 (1994) 2671–2678.
[31] Sorensen D., Gianola D., Likelihood, Bayesian, and MCMC methods in quantitative genetics, Springer, New York, 2002.
[32] Van Tassell C.P., Van Vleck L.D., Gregory K.E., Bayesian analysis of twinning and ovulation rates using multiple-trait threshold model and Gibbs sampling, J. Anim. Sci. 76 (1998) 2048–2061.
[33] Wild P., Gilks W.R., Adaptive rejection sampling for Gibbs sampling, Appl. Statist. 42 (1993) 701–709.