Indian Journal of Animal Sciences **92** (5): 645–648, May 2022/Short communication

Estimation of heritability of Karan Fries cattle using Bayesian procedure

HIMADRI SHEKHAR ROY[1]✉, AMRIT KUMAR PAUL[1], RANJIT KUMAR PAUL[1], RAMESH KUMAR SINGH[2], PRAKASH KUMAR[1] and MD YEASIN[1]

ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012 India

Received: 25 October 2021 ; Accepted: 17 February 2022

Keywords: Bayesian, Credible interval, Heidelberg stationarity test, Heritability, Karan fries, Linear mixed model

Recent progress in the field of quantitative genetics for both animals and plants has done an immense contribution to increasing the production and productivity of animals and plants. Quantitative genetics is a model-based study that deals with the genetics of complex traits or characters in which genes, as well as non-genetic factors, influence the trait of interest. The origin of quantitative genetics started with the development of the basic statistical tools which measure the targeted traits which are influenced by both genotypes and by the environment. It finds a way to improve animal populations by inferring the breeding value from the phenotypic value to maximize the probability of selecting better parents.

As stated by Dairy and Milk Processing Market in India (2018-2023), as of 2018, India leads in the milk production in the world, accounting for ~19% of the global market share. According to Basic Animal Husbandry Statistics, DAHD&F, GoI, the Milk production and per capita availability of milk in India for the year 2018-19 has increased to 187.70 Million tonnes and 394 gms/day. The report published in 'Food Outlook, 2018' by the United Nations Food and Agriculture Organization, milk production for the world scenario has increased from 800.2 MMT in 2016 to 811.9 MMT in 2017 with a growth rate of 1.46%.

Statistical methods using a linear mixed model (LMM) are diverse and applied in various fields (Brown and Prescott 1999, Demidenko 2004). Estimation of heritability and breeding values (BV) therefore received much attention in the quantitative genetic literature (Sorensen and Gianola 2007). Among the class of linear mixed model (LMM), the animal model becomes one of the popular methods which has been used for many decades in the field of animal breeding (Henderson 1975, Wang *et al.* 1993). It combines phenotypic records of an individual with pedigree and/or genetic marker information to draw inferences about the parameters of interest. Animal model uses pedigree / genetic marker information as a form of an additive genetic matrix (Ahlinder *et al.* 2013). Recently, estimation of genetic parameters using the Bayesian model gained much more reputation (Sorensen and Gianola 2007, Hadfield 2010) and Gianola *et al.* (1990) pointed out the framework of Bayesian methodology for the estimation of breeding values when variances are not known.

Let σ be the unknown variance components of a mixed model, the marginal posterior density i.e. f(σ/y) is given by- $f(\sigma/y) = \int_b f(,\sigma/y)$; where the joint posterior density and $f(,\sigma/y) \propto f(y/,\sigma)*f(,\sigma)$; where f(y/,σ) and f(,σ) corresponds to the likelihood and the joint prior density of and σ respectively. In Bayesian Structure, the specification of prior distributions is not straightforward. Gianola and Fernando (1986) also stated that the prior distribution depends on many factors such as information contained in past data, theoretical considerations, and personal beliefs which affect Posterior inference. Bayesian inference also required a complex computation procedure via Markov chain Monte Carlo (MCMC) (Gilks *et al*. 1995). The two commonly used MCMC methods are the Gibbs sampler and the Metropolis-Hastings (M–H) algorithm. Wang *et al*. (1993) applied MCMC methods in the standard additive polygenic model. Breslow and Clayton (1993) helped to popularize GLMMs and emphasized likelihood-based inference via penalized quasi-likelihood (PQL). Ahlinder *et al.* (2013) proposed an analytic Bayesian implementation of the mixed linear model for estimation of heritability in animal models without convergence problems. The breeding values and residual variance component are analytically integrated out from the model and utilized Gibbs sampling distribution. Meyer Karin (2007) developed Wombat, a software package for quantitative genetic analyses of continuous traits using linear mixed model. The developed package accommodate a variety of models for numerous traits, multiple fixed and random effects, selected genetic covariance structures, random regression models etc. Holand *et al.* (2013) studied mixed model approach using integrated nested Laplace approximations (INLA) in Bayesian paradigm using pedigree structure. Singh *et al.* (2016) applied single trait linear mixed random regression

Present address: [1]ICAR-Indian Agricultural Statistics Research Institute, Pusa, New Delhi. [2]Division of Animal Genetics and Breeding, Bihar Veterinary College, Patna, Bihar. ✉Corresponding author email: himadriiasri@gmail.com

model for analyzing the first lactation monthly test-day milk yield records in Karan Fries cattle. Singh *et al.* (2020) estimated genetic parameters of first lactation 305-day milk yield and energy traits in karan fries cattle.

In this paper, we have used Bayesian Linear mixed model for estimation of heritability using pedigree data. Linear mixed models provide a flexible framework for modeling a wide range of pedigree data but in some practical situations, response variables do not follow normality assumptions. Hence for non-Gaussian response variables, a generalized linear mixed model solves the complex architecture. Generalized Linear Mixed Models using MCMC (Hadfield 2009) is applied and MCMC algorithm is used to approximate the posterior distribution of the parameter of interest. Diagnostic test of the MCMC was done graphically as well as by the Heidelberg stationarity test. Estimation of Variance estimates of the random effects (V_A) and residual variance estimation (V_R) and Variance estimates location effects i.e. fixed effects were done by using the Bayesian procedure. Finally, the posterior estimate of heritability (h^2) for first lactation 305 days or less milk yield (FL305DMY) was estimated along with its Highest Posterior Density (HPD) intervals.

A general form of the linear mixed model is represented as follows:

$$y = X\beta + Zu + e \quad (1)$$

Where, y, vector of response variables of size nx1; β, vector of fixed effects with size p × 1; *u,* vector of random genetic effects with size n × 1. Here both *X* and *Z* are known incidence matrices relating phenotypic records to respective parameters included in (1), and e is a vector of errors that follow a multivariate normal distribution with zero mean vector, and covariance structure $I\sigma^2_E$, where I is the identity matrix of order n.

Let the responses vector be $Y=[Y_1, \ldots, Y_n]^T$ with corresponding means $\mu=[\mu 1, \ldots, \mu n]^T$. For generalized linear models, the marginal mean μ of the response Y is related to a linear predictor through a link function $g(\mu)$. The observation vector Y is not necessarily Normal (e.g., Gamma, Inverse Gaussian, etc.) E(Y|U)= μ and g(μ)= Xβ+ZU, here g is assumed as link function.

Suppose represents the probability distribution of i^{th} observation with latent variable *l*. The linear model for the latent variable with known design matrices *X* and *Z* with parameter vectors and is given by:

$$l = X\beta + Zu + E$$

Here it is assumed that the location effects (β and *u*) and the residuals (*e*) follow a multivariate normal distribution-

$$\begin{bmatrix}\beta \\ \mu \\ e\end{bmatrix} \sim N\left(\begin{bmatrix}\beta_0 \\ 0 \\ 0\end{bmatrix} \begin{bmatrix}B & 0 & 0 \\ 0 & G & 0 \\ 0 & 0 & R\end{bmatrix}\right)$$

Here β_0 denotes the prior means of fixed effect and B, G and R denote the expected (co)variances of the fixed effect, random effects, and residuals respectively. Here it is also seen that fixed effects, random effects, and residuals are independent. Generally, they are unknown and must be estimated from the data, usually by assuming they are structured in a way that they can be parameterized by a few parameters.

Here, we have applied the Bayesian Animal model using the following information of Bayesian structure-

Prior distribution: Prior distributions for the fixed effects are assumed to follow Normal distribution. The random effects are assumed to be Inverse-Gamma distribution with parameters =1, =0.002, and finally, the prior distribution for the residual variance is also assumed to follow Inverse-Gamma distribution with the same parameters. Inverse Gamma is parameterized differently using α and β, where $\alpha = \frac{\nu}{2}$ and $\beta = \frac{\nu}{2}$. Hence, the actual value becomes an inverse-Gamma (0.001; 0.001).

The Bayesian paradigm requires other important parameters also, i.e. a total number of iterations=100000, burn-in period i.e. the number of iterations to drop at the beginning (convergence)=10000, and thin, i.e. the number of iterations stored in memory=10 are applied. Here, Lag 10 states the values of autocorrelation for every 10 iteration values. Since our thinning parameter was 10, this refers actually to the correlation of every sampled value with the following one. Theoretically, it should be good to re-run a longer MCMC to increase the effective sample size. The effective sample size of the mean (Intercept) is larger than the effective sample size for variance components, for which the autocorrelation is greater.

The records of first lactation data of production (305-day or less milk yield and daily milk yield) and reproduction traits (AFC, FSP) on 1481 Karan-Fries cows were collected over 26 years from 1984 to 2009 at Dairy Cattle Breeding Division (DCB), National Dairy Research Institute (NDRI), Karnal. FL305MY, i.e. First lactation 305 days or less milk yield was collected and analyzed for the above-discussed procedure.

Results and diagnostic of the MCMC output: Here MCMC algorithm was used to solve Bayesian linear mixed model. MCMC algorithm (Markov Chain Monte Carlo) used in Bayesian model helps to approximate the posterior distribution of the parameters of interest. Here, the diagnostic of MCMC by graphically and Heidelberg stationarity test is shown below.

Diagnostic of the MCMC: The pattern of behaviour of the MCMC algorithm i.e. trace and the convergence and autocorrelation of our 'chain' of samples are graphically represented in Fig. 1 and 2. The posterior density function for each component (Intercept, animal, and units or residual) is graphically represented from the right side of Fig.1 and 2.

Results suggest the effective sample size of the mean (Intercept), variance components for animal and units were 8115.481, 920.306, and 1169.66 respectively. The diagnostic tests of convergence were done by the Heidelberg stationarity test. Table 1 shows that for both the cases i.e. Animal and Residual, the condition of stationarity is satisfied.

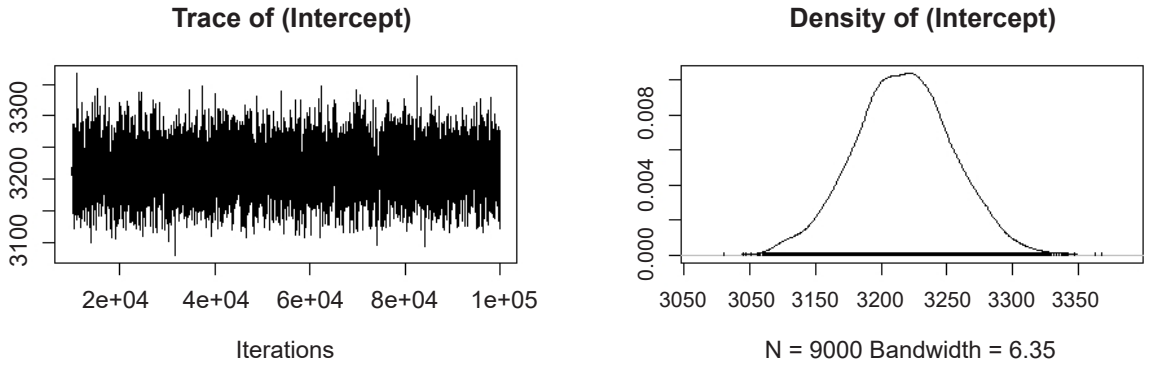

Fig. 1. Trace of the mean (or intercept).

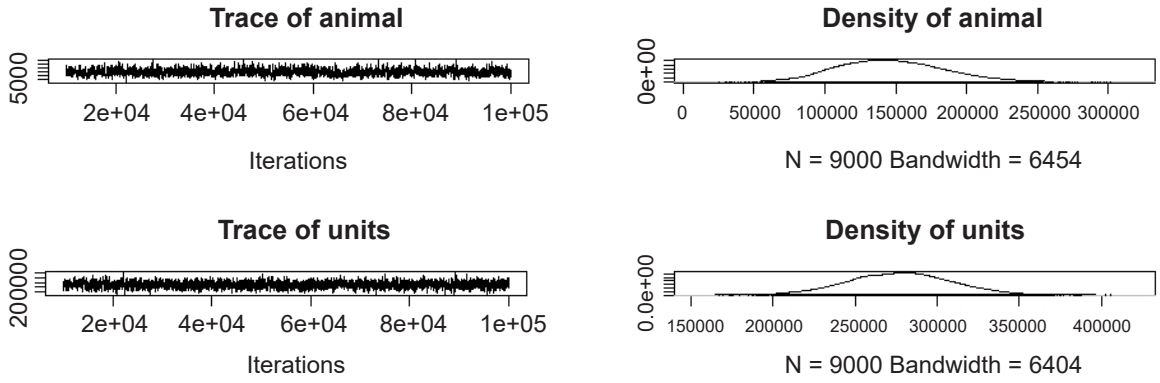

Fig. 2. Trace of the variances. (Animal, V_A ; units,V_R).

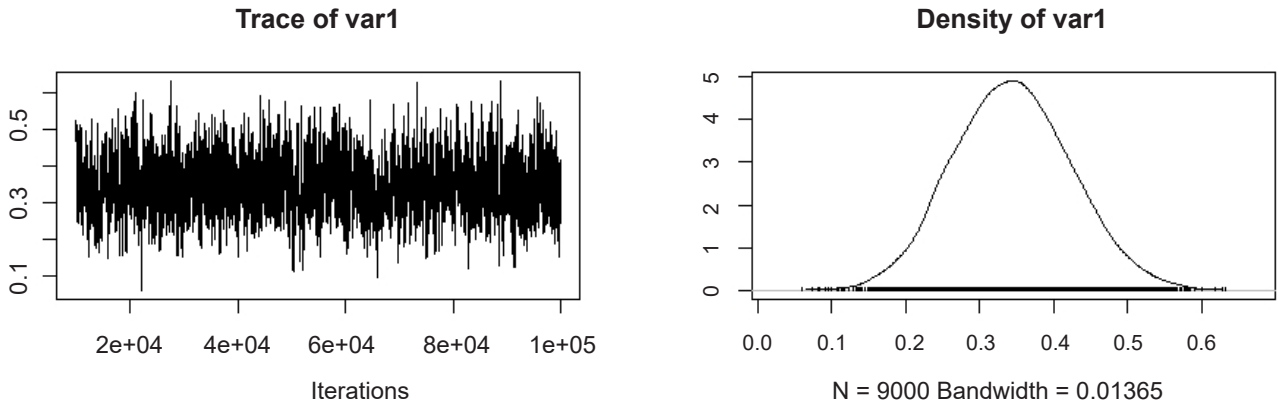

Fig. 3. Trace (left) and posterior density (right) of the heritability.

Table 1. Diagnostic tests of convergence (Heidelberg stationarity test)

	Stationarity	Start iteration	p-value
Animal	Passed	1	0.773
Residual	Passed	1	0.902

For model selection, DIC (Deviance Information Criterion) associated with the model is calculated, which is 13552.65. Variance component estimates of the random effects (V_A), residual variance estimation (V_R), and location effects i.e. fixed effects (the populations mean is called as Intercept) are given in Table 2.

Table 2. Variance estimates

Random effects (V_A) and Residual variance estimation (V_R)				
	Posterior mean	Lower 95% CI	Upper 95% CI	Effective sample size
Animal	146819	78027	224891	920.3
Residual	277574	217266	339986	1170
Location effects i.e. fixed effects				
	Posterior mean	L-95% CI	U-95% CI	Effective sample size	pMCMC
Intercept	3217	3144	3292	8115	<1e-04 ***

The value of pMCMC is the posterior probability associated with the event which is not a p-value but provides the same kind of information. Here, the pMCMC is very weak indicating that the population mean is very different from zero. Here the posterior estimate of the heritability was calculated along with its lower HPD interval and Upper HPD interval, which is presented in Table 3. The heritability of the trait is about 0.34 with 95% of probability to lie between 0.19 and 0.50. The plot of the trace and the density function is given in Fig. 3.

Table 3. Heritability estimates along with HPD interval

	Mean	Lower HPD interval	Upper HPD interval
Heritability	0.34	0.19	0.50

SUMMARY

In this study, Bayesian model was applied for analyzing the first lactation in Karan Fries cattle. First lactation data of production (305-day or less milk yield and daily milk yield) were collected from the history-cum pedigree sheet and daily milk yield registers of the division of Dairy Cattle Breeding (DCB), National Dairy Research Institute, Karnal. In the Bayesian paradigm, MCMC method was applied to solve complex mathematical problems to estimate a large number of unknown parameters. Assuming linear mixed model and using the different prior set up, diagnostic of MCMC (Markov Chain Monte Carlo) was carried out graphically as well as by Heidelberg stationarity test. Variance estimates of the random effects (VA) and residual variance estimation (VR) and Variance estimate location effects, i.e. fixed effects were calculated along with effective sample size. Finally, heritability (h^2) estimate for First lactation 305 days or less milk yield (FL305DMY) was estimated along with its credible interval.

REFERENCES

Ahlinder J and Sillanpää M J. 2013. Rapid Bayesian inference of heritability in animal models without convergence problems. *Methods in Ecology and Evolution* **4**(11): 1037–46.

Basic Animal Husbandry Statistics. DAHD&F, GoI.

Blasco A. 2001. The Bayesian controversy in animal breeding. *Journal of Animal Science* **79**(8): 2023–46.

Breslow N E and Clayton D G. 1993. Approximate inference in generalized linear mixed models. *Journal of the American statistical Association* **88**(421): 9–25.

Brown H and Prescott R. 1999. *Applied Mixed Models in Medicine.* John Wiley & Sons, New York.

Chib S and Greenberg E. 1995. Understanding the metropolis-hastings algorithm. *The American Statistician* **49**(4): 327–35.

Demidenko E. 2004. *Mixed Models: Theory and Application*. John Wiley & Sons, New Jersey.

Fong Y, Rue H, and Wakefield J. 2010. Bayesian inference for generalized linear mixed models. *Biostatistics* **11**(3): 397–412.

Gianola D and Fernando R L. 1986. Bayesian methods in animal breeding theory. *Journal of Animal Science* **63**(1): 217–44.

Gianola D and Foulley J L. 1990. Variance estimation from integrated likelihoods (VEIL). *Genetics Selection Evolution* **22**(4): 1–15.

Gianola D, Im S and Macedo F W. 1990. A framework for prediction of breeding value, pp. 210-238. *Advances in Statistical Methods for Genetic Improvement of Livestock*. Springer, Berlin, Heidelberg.

Geman S and Geman D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–41.

Gilks W, Richardson S and Spiegelhalter D. 1995. *Markov Chain Monte Carlo in Practice.* Interdisciplinary Statistics. Chapman & amp.

Hadfield J D. 2010. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software* **33**(2): 1–22.

Henderson C R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 423–47.

Holand A M, Steinsland I, Martino S and Jensen H. 2013. Animal models and integrated nested Laplace approximations. *G3: Genes, Genomes, Genetics* **3**(8): 1241–51.

Kaam V. 1997. GIBANAL-Analyzing program for Markov Chain Monte Carlo sequences.

Mathew B, Bauer A M, Koistinen P, Reetz T C, Léon J and Sillanpää M J. 2012. Bayesian adaptive Markov chain Monte Carlo estimation of genetic parameters. *Heredity* **109**(4): 235–45.

Meyer K. 2007. WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *Journal of Zhejiang University Science B* **8**(11): 815–21.

Nustad H E, Page C M, Reiner A H, Zucknick M and LeBlanc M. 2018. A Bayesian mixed modeling approach for estimating heritability. *BMC Proceedings* **12**: 117–22.

Singh A P, Chakravarty A K, Mir M A and Arya A. 2020. Genetic parameters and animal model evaluation of first lactation 305-d milk yield and energy traits in Karan Fries cattle. *Indian Journal of Animal Research* **54**: 405–08.

Singh A, Singh A, Singh M, Prakash V, Ambhore G S, Sahoo S K and Dash S. 2016. Estimation of genetic parameters for first lactation monthly test-day milk yields using random regression test day model in Karan Fries cattle. *Asian-Australasian Journal of Animal Sciences* **29**(6): 775–81.

Singh R K. 2013. 'Genetic evaluation of Karan-Fries sires using multiple trait models.' Ph.D. Thesis, National Dairy Research Institute (Deemed University), Karnal, India.

Sorensen D and Gianola D. 2007. *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer Science and Business Media.

Wang C S, Rutledge J J and Gianola D. 1993. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genetics Selection Evolution* **25**(1): 1–22.