IMPLEMENTATION ISSUES IN BAYESIAN ANALYSIS IN ANIMAL BREEDING

C.S. Wang

Central Research, Pfizer Inc., Groton, CT 06385, USA

SUMMARY
Contributions of Bayesian methods to animal breeding are reviewed briefly. Emphasis is placed on using Gibbs sampling to arrive at marginal posterior distributions needed for making Bayesian inferences. Improper posteriors can occur due to improper priors employed or insufficient data. Caution should be exercised when improper priors are used.
Keywords: genetic parameters, Bayesian methods, Gibbs sampling

INTRODUCTION
Best linear unbiased prediction (BLUP) (Goldberger 1962; Henderson 1963, 1973) has been used extensively in animal breeding to predict merit of animals that are candidate for selection. Bulmer (1980), Goffinet (1983) and Fernando and Gianola (1986) showed that under multivariate normality, BLUP is the conditional mean of merit given a set of linearly independent error contrasts. The conditional mean maximizes expected response to selection if k best individuals are selected from n available candidates. The conditional inference interpretation of BLUP reveals its correction with Bayesian procedures.

BLUP assumes known covariance matrices. If these matrices are estimated and the estimated values are used in the mixed model equations (MME), the solutions to the MME are no longer BLUP, and ranking properties of the predictors are unknown. Resulting estimators of genetic and environmental trends have unknown distributional properties. Also, the assumption of normality is violated when data are categorical in nature (e.g., calving ease scores), or counts (e.g., litter size) or censored (e.g., survival data). Hence, a more general formulation is needed.

A seminal paper of Gianola and Fernando (1986) laid foundations to solve animal breeding problems beyond BLUE and BLUP using Bayesian methods in an unified manner. Inferences about parameters or functions thereof are based on their respective marginal posterior distributions such that uncertainty about other parameters in the model is taken into account.

Developments of Bayesian methods in animal breeding can be roughly divided into two phases: pre-Gibbs and Gibbs. The pre-Gibbs phase involved Bayesian model building and used analytical approximations to arrive at desired posterior distributions. Bayesian linear models were elucidated by Dempfle (1977), Gianola and Fernando (1986) and Gianola et al. (1990), following Lindley and Smith (1972) as points of departure. Notable contributions in threshold models were Gianola and Foulley (1983) and Harville and Mee (1984). Foulley et al. (1987) developed a Poisson model for count data, which was extended to model overdispersed count data by Tempelman and Gianola (1996). Mixed effects survival models were adapted in animal breeding by Smith and Quaas (1984), Ducrocq et al. (1988) and Ducrocq and Casella (1996).

Models for major genes and genetic marker data were developed by Hoeschele (1988), and Hoeschele and VanRaden (1993a, 1993b). Difficulties in applying the theoretically appealing Bayesian methods were primarily numerical because integration in high dimensions is needed to obtain marginal posterior distributions, aside from problems in specifying prior distributions for unknown parameters.

While continuing model development as in the pre-Gibbs phase, characteristics of the Gibbs phase are mainly numerical, the objective being to obtain desired marginal posterior densities without analytical approximations. Markov Chain Monte Carlo (MCMC) procedures (Metropolis *et al*, 1953; Hastings 1970) and in particular Gibbs sampling (Geman and Geman 1984; Gelfand and Smith 1990), for sampling from joint distributions provided the solution. Gibbs sampling coupled with Bayesian analyses has been applied to a wide range of animal breeding problems.

This paper gives a brief review of contributions of Bayesian methods to animal breeding and discusses issues when applying MCMC methods to solve animal breeding problems under a Bayesian framework.

BAYESIAN INFERENCES IN ANIMAL BREEDING

A linear model is used as a basis for discussion. Let $\theta' = [\beta', \mathbf{u}', \Sigma]$, where β and $\mathbf{u}$ are location parameters and Σ are dispersion parameters. Under normal priors for β and $\mathbf{u}$, and a proper prior (e.g., scaled inverted Wishart) for Σ, and assuming a normal likelihood, the joint posterior density of θ is formed as:

$$p(\beta, \mathbf{u}, \Sigma | y) \propto p(y | \beta, \mathbf{u}, \Sigma) p(\beta, \mathbf{u}, \Sigma) \tag{1}$$

A Bayesian must express uncertainty about unknown parameters before data are observed in terms of a prior distribution, $p(\beta, \mathbf{u}, \Sigma)$. All unknowns are treated as random variables. This uncertainty is incorporated formally into the posterior density. The prior density of θ is subjective, and so is the posterior density. Prior information can include personal beliefs, be data derived or be justified theoretically. Inferences about location parameters are based on:

$$p(\beta, \mathbf{u} | y) = \int_{\Sigma} p(\beta, \mathbf{u}, \Sigma | y) d\Sigma \tag{2}$$

and about dispersion parameters on

$$p(\Sigma | y) = \int_{\beta, \mathbf{u}} p(\beta, \mathbf{u}, \Sigma | y) d\beta d\mathbf{u} \tag{3}$$

The expected value of distribution [2] minimizes expected posterior loss and expected risk over repeated sampling under a quadratic loss function (e.g., Gianola *et al*, 1990), provided the prior is proper. Selecting individuals with ordered conditional means maximizes expected genetic merit in a sample having a fixed number of selected individuals (Goffinet 1983; Fernando and Gianola 1986). However, if improper priors are employed , the expected value of [2] only minimizes expected posterior loss; there might exist other estimators with lower expected risk over repeated sampling (Gianola *et al*, 1990).

Under selection, when all information used to make selection decision is available, joint and marginal posterior densities constructed considering selection and non-random mating are the same as when selection and non-random mating are ignored. The results hold for any distribution, any selection type or non-random mating, and any number of generations of selection (Gianola and Fernando 1986; Fernando and Gianola 1990). Again, under selection, the expected value of [2] only minimizes expected posterior loss even with proper priors (Fernando and Gianola 1990). Distributional properties under repeated sampling and selection are unknown.

The connection between posterior means of location parameters in linear models and BLUP is significant. In [1], if an uniform prior is assigned to β, and assuming Σ is known, the expected value of the conditional posterior distribution $[\beta, \mathbf{u} | \mathbf{y}, \Sigma]$ is identical to BLUE(β) and BLUP($\mathbf{u}$), under normality, (Dempfle 1977; Gianola and Fernando 1986). This not only gives a Bayesian interpretation of BLUP, but also points out short comings of BLUP. When Σ is unknown, a two stage procedure for predicting breeding values is adopted by first estimating Σ, and then using the estimated values in the MME. Hence, uncertainty in estimating Σ is ignored in predicting breeding values. This practice can be justified by an appropriate Bayesian argument (Gianola and Fernando 1986), because

$$p(\beta, \mathbf{u} | \mathbf{y}) = \int_{\Sigma} p(\beta, \mathbf{u}, \Sigma | \mathbf{y}) d\Sigma \approx p(\beta, \mathbf{u} | \mathbf{y}, \hat{\Sigma}) \qquad [4]$$

where $\hat{\Sigma}$ is estimated based on $p(\hat{\Sigma} | \mathbf{y})$. If an uniform prior is employed for Σ, the mode of $p(\hat{\Sigma} | \mathbf{y})$ corresponds to REML of Σ (Harville, 1974). Approximation [4] is good, provided Σ is well estimated. Densities [2] and [3] are always preferred for making inferences whenever feasible. Computing [2] and [3] requires successive integration, and these usually cannot be done analytically. MCMC methods draw samples from [1], and such samples can be used to make inferences

METROPOLIS-HASTINGS AND GIBBS SAMPLING ALGORITHMS
MCMC methods allow to draw random samples from desired posterior distributions; samples are then used to make inferences about parameters of interest. A useful class of MCMC algorithms is Metropolis-Hastings (MH) algorithm with the Gibbs sampler being a special case. MH consists of two steps: generation and selection. MH draws a sample from a candidate distribution $q(\theta)$ (dropping $\mathbf{y}$ for conciseness), and the sample drawn is accepted or rejected based on a random mechanism. The process is repeated indefinitely, thus creating a Markov chain having $p(\theta)$ as its equilibrium distribution. MH works as follows. Let θ^n be the value at the current step, θ^{n+1} be the value at the next step, and $\theta*$ be a possible (proposed) value, of θ. Let $q(\theta^n, \theta*)$ be a transition probability function, such that: $q(\theta^n, \theta*) = \Pr[\text{generating } \theta* | \theta^n]$. After drawing a sample $\theta*$ from $q(\theta^n, \theta*)$, compute the ratio

$$r = r(\theta^n, \theta*) = \begin{cases} \min\left[\dfrac{p(\theta*)}{p(\theta^n)} \dfrac{q(\theta*, \theta^n)}{q(\theta^n, \theta*)}, 1\right], & \text{if } p(\theta^n) q(\theta^n, \theta*) > 0 \\ 1, & \text{otherwise} \end{cases} \qquad [5]$$

Then, set $\theta^{n+1} = \theta*$ with probability r and set $\theta^{n+1} = \theta^n$ with probability 1-r. In summary, the algorithm is,
- Draw a sample $\theta*$ from $q(\theta^n, \theta*)$
- Compute the ratio r as in [5]
- Sample an u from Uniform(0,1)
- If $r \geq u$, set $\theta^{n+1} = \theta*$
- Otherwise set $\theta^{n+1} = \theta^n$
- Repeat.

It should be noted that $p(\theta)$ needs to be specified only up to a multiplicative constant, since the constant cancels in the ratio calculation, a great advantage of the algorithm.

Gibbs sampling repeatedly draws samples from a set of fully conditional densities, $p(\theta_i | \theta_{-i}), i = 1, 2, \ldots s$, where θ_i is an element (or a subvector) of θ, and $\theta' = (\theta_i' \; \theta_{-i}')$. At the limit, samples drawn are from the equilibrium distribution $p(\theta)$. As mentioned above, Gibbs sampler is a special case of MH with $p(\theta_i | \theta_{-i})$ as the driving densities. The ratio r under such a set of driving densities is always 1, i.e., there is no rejection.

This connection between the Gibbs sampler and MH is useful in at least two ways. First, in using Gibbs sampling, some of the fully conditional densities may not be in recognizable forms, MH can be used to sample from these. Second, even if all conditional densities in a Gibbs sampler are recognizable, the sampler has shown slow mixing properties in some problems. A MH step for pre-chosen parameters at pre-defined sampling intervals may break the lag correlations among sequential Gibbs samples.

PRACTICAL ISSUES OF RUNNING MCMC

Convergence and convergence checking: There are two issues. For an irreducible and aperiodic chain, Markov chain theory guarantees that the chain converges to its equilibrium distribution. A weak law of large numbers guarantees that the average of a sufficient number of samples drawn converges to some expectation of interest.

Practically, the first convergence issue means that sampling usually starts somewhere outside of the joint distribution, so a number of iterations is needed such that the sampler forgets the starting values and samples from the joint distribution. This period is called "burn-in". Once the sampler gets inside the joint distribution, it stays there. After burn-in, a large sample can be drawn for computing summary statistics such that the weak law of large numbers is applicable.

How should the Gibbs chains be run? There is some debate, but a general consensus seems to be emerging (Cowles and Carlin 1996). The multiple short chain method of Gelfand and Smith (1990) is too wasteful and less favored; the single long chain method advocated by Geyer (1992) is unassured. A middle ground is to run a few "long" chains, say 2 to 4 chains. These chains should use deliberately overdispersed starting values (guesses) with respect to the unknown joint distribution. Parameters with slow mixing (high lag correlations) should be

monitored. Convergence of a chain is constrained by the slow mixing parameters, i.e., if such parameters appear not to be converged, the whole chain is not converged. In linear models, Σ tends to have slow mixing properties (Wang *et al*, 1993, 1994a), while thresholds and Σ mixed slowly (Sorensen *et al*, 1995; Wang *et al*, 1997). In generalized linear models, Janss *et al*. (1995) and Thaller and Hoeschele (1996a, 1996b) observed extremely slow mixing for univariate sampling of genotypes.

Chains should be plotted and inspected visually to detect slow mixing parameters. Visual inspection along with Raftery and Lewis's (1992) NBURN statistics can be used to determine the burn-in period. Gelman and Rubin (1992) statistics can reveal slow mixing and multimodality, and lag correlations should be calculated. Monte Carlo variances and effective sample sizes (Geyer 1992; Sorensen *et al*, 1995) for slow mixing parameters should be calculated to assess if chain is "long" enough.

Post Gibbs analysis: After burn-in, samples from several long chains should be combined to estimate characteristics of desired distributions. These samples are identically, though dependently, distributed. Subsampling of Gibbs chains is not necessary (Geyer 1992). Subsampling is a concern only if cost for storage and processing of Gibbs samples is excessive. Let the Gibbs samples, after burn-in, for x be: $x_i, i = 1,2,\ldots,n$. The expected value and variance of x are estimated by $\hat{E}(x) = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\hat{V}ar(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$. The density of x can be estimated by the normal kernel estimator: $\hat{p}(z) = \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{z-x_i}{h}\right)^2\right]$, where h is a constant (window width). The mode of p(x) can be located using $\hat{p}(z)$. Probability statements including the highest posterior density (HPD) region can be computed as: $\hat{p}(a < z < b) = \int_{a}^{b} \hat{p}(z)dz$.

THE GIBBS SAMPLER IN ANIMAL BREEDING
To apply Gibbs sampling, fully conditional posterior densities need to be derived analytically. In animal breeding, these have been derived for univariate mixed linear models (Wang *et al*, 1993, 1994a), for inferences about response to selection (Sorensen *et al*, 1994; Wang *et al*, 1994b), for maternal effects models (Jensen *et al*, 1994), for multivariate linear models (Varona et al, 1994; Van Tassell and Van Vleck 1995), for univariate threshold models (Sorensen *et al*, 1995), for a bivariate threshold and Gaussian model (Jensen 1994; Wang *et al*, 1997), for segregation analysis (Janss *et al*, 1995) and linkage analysis (Thaller and Hoeschele 1996a, 1996b), and in survival models (Korsgaard 1996), among others. A Bayesian implementation of univariate and multivariate mixed effects models with t-distributions is in Gianola et al. (1995), Gianola and Stranden (1996) and Stranden (1996). Gianola and Sorensen (1996) developed a threshold model with t-distributed liabilities and random effects. Tempelman and Gianola (1996) pointed out difficulties when applying Gibbs sampling in Poisson and negative binomial models due to unrecognizable conditional distributions for the location parameters.

Data augmentation (Tanner 1993) has been useful for constructing joint posterior densities and for deriving fully conditional posterior densities to facilitate Gibbs sampling. Appropriate use of data augmentation may lead to fully conditional distributions that are recognizable and easy to sample from. In univariate threshold models, Sorensen *et al.* (1995) used underlying variables as augmented data. Van Tassell and Van Vleck (1995) augmented residuals associated with missing data in multi-trait Gaussian models. In segregation and linkage analyses, marker-QTL genotypes were augmented data (Janss *et al*, 1995; Thaller and Hoeschele 1996a, 1996b). For a particular problem, the data augmentation procedure may not be unique. Wang *et al.* (1997) discussed various data augmentation strategies in modeling a mixture of continuous and threshold variables. Gianola et al. (1995), Gianola and Sorensen (1996) and Stranden (1996) used augmentation in connection with robust estimation.

IMPROPER PRIOR AND POSTERIOR DISTRIBUTIONS

Improper posterior distributions can result from using an improper prior or from a deficient data structure. In Gaussian linear models, Hobert and Casella (1996) showed that a frequently used improper prior $1/\sigma_i^2$ for variances always leads to improper posteriors. Another convenient improper prior for variances is the flat prior . This prior may or may not cause an improper posterior depending on data structure, so it is unsafe to use. What happens if a MME do not have a unique solution? Sahu and Gelfand (1994) pointed out that though the sampler may not converge to a stationary distribution in a usual sense, inferences about estimable functions are still valid, and Gianola (personal communication, 1997) has a simple proof of this. It is clear that monitoring such a chain is difficult, so it is advantageous to reparameterize the model such that the incidence matrix for the fixed effects has full column rank.

In a simple probit or logit model for binomial data with a single variance component, Natarajan and McCulloch (1995) demonstrated that the $1/\sigma_i^2$ prior or flat prior for the variance component leads to improper posteriors under certain conditions. We conjecture that this conclusion may apply to more general generalized linear models. For a simple one-way analysis of variance model, with probit or logit link, they showed that a success and a failure for at least one level of the random effect is needed to ensure a proper posterior distribution. This effectively rules out animal threshold models! This fact was evident in Hoeschele and Tier (1995); their Gibbs sampler "blew up" when analyzing simulated data sets with an animal model. A well known extreme category problem (ECP) in threshold models, i.e., a level of a fixed factor has only one extreme response category in threshold models, also induces improper posteriors if the fixed effects are assigned improper priors (Hoeschele and Tier 1995). A bounded uniform priors for the fixed effects, however, was not sufficient to overcome the problem caused by ECP, while a normal prior did (Hoeschele and Tier 1995). Running Gibbs chains without checking that the posterior is proper is dangerous.

One way to solve the problem of an improper posterior is to use proper priors. The Normal distribution used for breeding values is well justified by genetic considerations. Hoeschele and

VanRaden (1993a) derived priors for gene effects and recombination rate based on genetic principles. Wang *et al.* (1994b) used a method of moments fit to derive priors for variance components from experimental data.

To represent prior ignorance, proper priors with large variances have been suggested. However, prior variances should not be too diffuse with respect to the corresponding posterior variances to assure numerical stability. A theoretically appealing noninformative prior is reference prior (Bernardo 1979). Unfortunately, reference priors are unavailable for general models.

In summary, the Bayesian paradigm has helped to extend the range of models with animal breeding problems, but care should be exercised in: a) the construction of the Bayesian models (priors and sampling assumptions), and b) the implementation via MCMC methods.

ACKNOWLEDGMENTS

REFERENCES
Bernardo JM (1979) *J R Stat Soc Series B* **41**:113-147
Bulmer MG (1980) The Mathematical Theory of Quantitative Genetics. Oxford
Cowles MK, Carlin BP (1996) *J Am Stat Assoc* **91**: 883-904
Dempfle L (1977) *Ann Genet Sel Anim* **9**: 27-32
Ducrocq V, Quaas RL, Pollak EJ, Casella G (1988) *J Dairy Sci* **71**: 3071-3079
Ducrocq V, Casella G (1996) *Genet Sel Evol* **28**: 505-529
Fernando RL, Gianola D (1986)) *Theor Appl Genet* **72**: 395-404
Fernando RL, Gianola D (1990) In: *Adv in Stat Methods for Genet Improve of Livestock* (D Gianola, K Hammond eds), Springer-Verlag, Berlin, 437-453
Foulley JL, Gianola D, Im S (1987) *Theor Appl Genet* **73**: 870-877
Gelfand AE, Smith AFM (1990) *J Am Stat Assoc* **85**: 398-409
Gelman A, Rubin DB (1992) *Stat Sci* **7**: 457-511
Geman S, Geman D (1984) *IEEE Trans on Pattern Analy and Machine Intellig* **6**: 721-741
Geyer CJ (1992) *Stat Sci* **7**: 467-511
Gianola D, Foulley JL (1983) *Genet Sel Evol* **15**: 201-224
Gianola G, Fernando RL (1986) *J Animal Sci* **63**: 217-244
Gianola D, Im S, Macedo FW (1990) In: *Adv in Stat Methods for Genet Improve of Livestock* (D Gianola, K Hammond eds), Springer-Verlag, Berlin, 210-238
Gianola D, Stranden I, Foulley JL (1995) 5th Spanish Biometrics Conf, p3-4, Valencia, Spain
Gianola D, Stranden I (1996) 4th World Meeting of the International Soc for Bayesian Analysis, p12 , Cape Town, South Africa.
Gianola D, Sorensen DA (1996) Abstracts, European Association for Animal Production, **2**:48
Goffinet B (1983) *Genet Sel Evol* **15**: 91-98
Goldberger AS (1962) *J Am Stat Assoc* **57**: 369-375
Harville DA (1974) *Biometrika* **61**: 383-385
Harville DA, Mee RW (1984) *Biometrics* **40**: 393-408

Hastings W (1970) *Biometrika* **57**: 97-109

Henderson CR (1963) In: *Statistical Genetics and Plant Breeding (Hanson WD, Robinson HF eds)*, Washington, D.C. NAS-NRC **982**: 141-153

Henderson CR (1973) In: *Proc Animal Breeding and Genetics Symposium in Honor of Dr Jay L Lush.* ASAS and ADSA, Champaign, IL, pp 10-41

Hobert JP, Casella G (1996) *J Am Stat Assoc* **91**: 1461-1473

Hoeschele I (1988) *Theor Appl Genet* **76**: 311-319

Hoeschele I, VanRaden PM (1993a) *Theor Appl Genet* **85**: 953-960

Hoeschele I, VanRaden PM (1993b) *Theor Appl Genet* **85**: 946-952

Hoeschele I, Tier B (1995) *Genet Sel Evol* **27**: 519-540

Janss LLG, Thompson R, Van Arendonk JAM (1995) *Theor Appl Genet* **91**: 1137-1147

Jensen J (1994) In: *Proc 5th World Congr on Genet Appl to Livest Prod*, Guelph, Canada, Univ of Guelph, **18**: 333-336

Jensen J, Wang CS, Sorensen DA, Gianola D (1994) *Acta Agric Scandinavia, Sect A, Anim Sci* **44**: 193-201

Korsgaard IR (1996) International Workshop on genetic Improvement of Functional Traits in Cattle, Faculte universitaire des sciences agronomiques, Gembloux, Belgium, Jan 21-23

Lindley DV, Smith AFM (1972) *J R Stat Soc Series B* **34**: 1-18

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller A, Teller E (1953) *J Chem Physics* **21**: 1087-1092

Natarajan R, McCulloch CE (1995) *Biometrika* **82**: 639-643

Raftery AE, Lewis SM (1992) In Bayesian Statistics 4 (Bernardo JM, Berger JO, Dawid AP, Smith AFM eds), Oxford, Oxford Univ Press, 765-776

Sahu SK, Gelfand AE (1994) On propriety of posteriors and Bayesian identifiability in Generalized Linear Models. Univ of Cambridge and Univ of Conn

Smith SP, Quaas RL (1984) *J Dairy Sci* **67**: 2999-3007

Sorensen DA, Wang CS, Jensen J, Gianola D (1994) *Genet Sel Evol* **26**: 333-360

Sorensen DA, Anderson S, Gianola D, Korsgaard I (1995) *Genet Sel Evol* **27**: 229-249

Stranden I (1996) Ph.D. thesis. University of Wisconsin-Madison

Tanner MA (1993) *Tools for Statistical Inference.* Springer-Verlag, NY

Tempelman RJ, Gianola D (1996) *Biometrics* **52**: 265-279

Thaller G, Hoeschele I (1996a) *Theor Appl Genet* **93**: 1161-1166

Thaller G, Hoeschele I (1996b) *Theor Appl Genet* **93**: 1167-1174

Van Tassell CP, Van Vleck LD (1996) *J Animal Sci* **74**: 2586-2597

Varona L, Moreno C, Garcia-Cortes LA, Altarriba J (1994) *Revista Portuguesa de Zootecnia* **1**: 185-195

Wang CS, Rutledge JJ, Gianola D (1993) *Genet Sel Evol* **25**: 41-62

Wang CS, Rutledge JJ, Gianola D (1994a) *Genet Sel Evol* **26**: 91-115

Wang CS, Gianola D, Sorensen D, Jensen J, Christensen A, Rutledge JJ (1994b) *Theor Appl Genet* **88**: 220-230

Wang CS, Quaas RL, Pollak EJ (1997) *Genet Sel Evol* **29**: 117-143