Statistical and Computational Challenges in Whole Genome Prediction and Genome-Wide Association Analyses for Plant and Animal Breeding

Robert J. TEMPELMAN

Whole genome prediction (WGP) modeling and genome-wide association (GWA) analyses are big data issues in agricultural quantitative genetics. Both areas require meaningful input from the statistical scholarly community in order to further improve the accuracy of prediction of genetic merit and inference on putative causal variants as well as improving the computational efficiency of existing methods and algorithms. These concerns have become increasingly critical as new sequencing technologies will only exacerbate current model dimensionality problems. We focus primarily on mixed model and hierarchical Bayesian analyses which have been most commonly pursued by animal and plant breeders for WGP thus far. We draw attention to our observation that many such previous analyses have not carefully inferred upon hyperparameters defined at the top levels of the Bayesian model hierarchy, but simply arbitrarily specify their values. We also reassess previous discussions on WGP model dimensionality, believing that useful data augmentation schemes utilized in various Markov Chain Monte Carlo (MCMC) schemes have led to a general misunderstanding that heavy-tailed or variable selection-based WGP models may be highly parameterized relative to more standard mixed model representations. Computational efficiency is addressed with respect to MCMC and competitive, albeit approximate, alternatives. Furthermore, GWA analyses are reassessed, encouraging a greater reliance on shrinkage-based inferences based on critically chosen priors, instead of potentially nonreproducible fixed effects *P* value-based inference.

Key Words: Bayesian modeling; Genomics; Hyperparameters; Shrinkage.

1. BACKGROUND

Genetic evaluation of livestock has inspired many statistical developments with a very good and extensive historical account written by Gianola and Rosa (2015) and a much

Robert J. Tempelman (✉) Department of Animal Science and Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824-1225, USA (E-mail: *tempelma@msu.edu*).

Journal of Agricultural, Biological, and Environmental Statistics
DOI: 10.1007/s13253-015-0225-2

Published online: 28 September 2015

shorter yet effective summary provided in Gianola (2000). Indeed, it can be readily argued that the need to provide such genetic evaluations provided the early motivation in formalizing prediction of random effects (Hazel 1943), followed by the development of mixed effects models to additionally adjust for nongenetic systematic or "fixed" effects (Henderson et al. 1959). Animal geneticists were also early adopters of Bayesian inference (Gianola and Fernando 1986), even before these methods became heavily popularized by introduction of the Markov Chain Monte Carlo (MCMC) algorithm (Gelfand et al. 1990). Of course, hierarchical modeling inference, whether based on classical mixed models or Bayesian analyses, is now pervasively used in many areas of agricultural, biological, and environmental statistics research and applications, including those published in this journal.

During the latter part of the twentieth century, animal and plant breeders valiantly strived for algorithmic and computing strategy improvements, whether it pertained to the estimation of variance components (Gilmour et al. 1995; Johnson and Thompson 1995; Meyer 1989; Misztal 1990; Misztal and Perez-Enciso 1993) or solving large systems of mixed model equations (Schaeffer and Kennedy 1986; Strandén and Lidauer 1999). These developments were driven by increasing sizes of databases as the number (n) of records and the number (q) of animals (i.e., number of levels of random effects) processed in these mixed model equations exponentially increased over time with both n and q having exceeded 10 million (M) for American Holstein dairy cattle genetic evaluations several decades ago (Wiggans et al. 1988)! These challenges were further compounded by the fact that typically $q > n$ as historical pedigree information can be extensive and because of the desire to genetically evaluate both male and female breeding stock for sex-limited traits (e.g., dairy bulls do not produce milk). In other words, predictions of genetic merit are typically desired for many animals that have no records of their own. Hence, animal breeders have been historically accustomed to solving "big data" problems, in no small part because of their passion for reliable quantitative genetic inference, not much unlike the physicists, engineers, or neuroscientists earnestly trying to confront their own statistical inferential challenges as described in Brown and Kass (2009).

The twenty-first century generation of quantitative geneticists face far greater challenges, primarily because of the advent of genomic selection based on whole genome prediction (WGP), a concept commonly attributed to Meuwissen et al. (2001) although some earlier credit should also be given to Nejati-Javaremi et al. (1997). WGP is defined as the prediction of genetic merit of individuals based on dense genotypic marker information. Some very good recent reviews on WGP have been provided in the context of applications to aquaculture (Taylor 2014), livestock production (Blasco and Toro 2014; Garrick et al. 2014; Goddard et al. 2010), crop production (Desta and Ortiz 2014; Jannink et al. 2010; Jonas and de Koning 2013), and forestry (Grattapaglia and Resende 2010). Meuwissen et al. (2001) foresaw that genome sequencing technologies would make genotypes based on thousands (K) and now even millions (Druet et al. 2014), of single nucleotide polymorphism (SNP) markers widely available for livestock breeding; nevertheless, current genetic evaluation systems are typically based on genotypes derived from tens of thousands of SNP markers (Wiggans et al. 2011).

SNP markers are biallelic markers such that the provided SNP-specific covariates are typically either of value 0, 1, or 2, i.e., the number of copies of an arbitrarily chosen reference allele. Assuming that these polymorphic SNP markers are in close proximity or association, more specifically linkage disequilibrium (LD), with polymorphic genes or quantitative trait loci (QTL) that influence the trait of interest, Meuwissen et al. (2001) conjectured and demonstrated by simulation that the jointly estimated effects of all SNP markers in a penalized regression analysis, whether based on classical mixed model or more extensive hierarchical Bayesian specifications, should provide reliable estimates (i.e., WGP) of the total genetic merit of animals for the trait of interest.

There have been a large number of comparisons made between various WGP methods and models against each other based on simulation and/or analysis of real data, some of which are summarized in recent reviews (de los Campos et al. 2013; Gianola 2013). The use of hierarchical Bayesian analyses for WGP has particularly proliferated in the animal breeding and genetics community (Gianola 2013; Gianola et al. 2009). Increasingly, others are focusing their attention on discovering segregating genes with major effects as in genome-wide association (GWA) analyses (Hayes 2013) based either on just using phenotypic data as the response variables or further integrating these phenotypes with other 'omics' response variables such as gene expression or proteomics data (Kadarmideen 2014). It is not the intent of this review to necessarily endorse one statistical paradigm, method, model or computing strategy over another except to suggest that WGP and GWA research is in critical need of intellectual input from the broader statistical community, particularly given that there is currently a dire shortage of quantitative geneticists able to work on these issues (Eisen 2008; Misztal 2007).

In this review, nevertheless, we will highlight that the rigor of many previous WGP model assessments may have been compromised due to improper specification or tuning of key hyperparameters at the higher levels of the model hierarchy. Hyperparameters typically refer to those parameters that define the distribution of effects specified in the first stage (likelihood) or even later stages of a Bayesian hierarchical model. For the simplest hierarchical model, being the linear mixed model, hyperparameters are merely synonymous with variance components; hence, hyperparameter inference in mixed models then equates to variance component estimation, an already established research area (Searle et al. 1992). However, hyperparameter inference is further extended to include other parameters that characterize the nature of heavy-tailedness or variable selection in more flexible hierarchical WGP model specifications.

We focus primarily on parametric WGP in the context of prediction but also discuss the implications of these efforts in GWA analyses which are often based on the same statistical models as WGP except that the focus is on inferring putative causal variants instead of the prediction of total genetic merit. We furthermore reconsider the issue of model dimensionality in the context of augmented models, computational aspects, including alternatives to MCMC inference, and point to future yet unresolved areas of work, again hoping to harness the intellectual energy of the broader statistics community that might develop a passion for quantitative genetics research.

2. AN INTRODUCTION TO WHOLE GENOME PREDICTION MODELING

The first-stage specification of virtually all parametric WGP models can be written as follows:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{z}_i'\mathbf{g} + e_i, i = 1, 2, \ldots, n, \tag{1}$$

or jointly for all n records as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \mathbf{e} \tag{2}$$

where $\mathbf{y} = \{y_i\}_{i=1}^n, \mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2 \cdots \mathbf{x}_n]'$, and $\mathbf{Z} = [\mathbf{z}_1\ \mathbf{z}_2 \cdots \mathbf{z}_n]'$ are known, and $\mathbf{e} = \{e_i\}_{i=1}^n$. Here y_i represents the record on subject i whereas $\mathbf{g} = \{g_j\}_{j=1}^m$ such that g_j represents the random effect of SNP j and considered to be a random draw from distribution $p\left(g_j|\vartheta\right)$, where ϑ represents a vector of hyperparameters, and e_i is typically $NIID\left(0, \sigma_e^2\right)$. Furthermore, $\mathbf{z}_i' = [z_{i1}\ z_{i2} \ldots z_{im}]$ represents the m-dimensional row vector of SNP genotypes on subject i where $z_{ij} \subset \{0, 1, 2\}$ denotes the number of copies of a reference allele at SNP j on subject i. Finally, $\boldsymbol{\beta}$ represents a vector of unknown fixed effects with $\mathbf{x}_i'$ being the corresponding known p-dimensional row vector of covariates.

Noting that typically $m \gg n$, Meuwissen et al. (2001) recognized the need to specify a penalty or prior, i.e., $p\left(g_j|\vartheta\right)$, in order to be able to even estimate $\mathbf{g}$; i.e., otherwise, treating $\mathbf{g}$ as fixed would make $\mathbf{g}$ nonestimable (Gianola 2013). Yet even with $m < n$ (e.g., Wiggans et al. 2015), it is reasonable to assume exchangeability amongst the elements of $\mathbf{g}$ and hence specify a prior. Typically, elements of $\mathbf{g}$ are specified to be independent a priori; i.e., $p(\mathbf{g}|\vartheta) = \prod_{j=1}^m p\left(g_j|\vartheta\right)$. If $p\left(g_j|\vartheta\right) \equiv N(0, \sigma_g^2)$, then the sole hyperparameter is $\vartheta = \sigma_g^2$. A commonly used inferential strategy in this case might to use restricted maximum likelihood (REML) to estimate variance components like σ_g^2 and σ_e^2 (Searle et al. 1992) followed by the use of best linear unbiased predictions (BLUP) of $\mathbf{g}$, often referred to as ridge regression BLUP (rrBLUP) of $\mathbf{g}$, and, subsequently, genomic BLUP (GBLUP) of the genetic merit $\mathbf{u} = \mathbf{Z}\mathbf{g}$ of all animals conditional on these REML estimates. We should quickly add, nevertheless, that the term GBLUP is generally reserved for the mixed model reparameterization used to directly solve for BLUP of $\mathbf{u}$ instead of directly solving for BLUP of $\mathbf{g}$. That is, in GBLUP, the genotypes $\mathbf{Z}$ are typically rescaled such that $\mathbf{u} = \mathbf{Z}^*\mathbf{g}$ with $\text{var}(\mathbf{u}) = \mathbf{Z}^*\mathbf{Z}^{*\prime}\sigma_g^2$ where $\mathbf{Z}^*$ has elements $z_{ij}^* = \frac{z_{ij} - 2\hat{p}_j}{k}$ with $k = \sqrt{2\sum_{j=1}^m \hat{p}_j(1 - \hat{p}_j)}$ based on $\hat{p}_j = \frac{1}{2n}\sum_{i=1}^n z_{ij}$ denoting the estimate of the frequency of the reference allele for SNP j. That is, $\mathbf{Z}^*\mathbf{Z}^{*\prime}$ can be regarded as a genomic marker derived matrix of realized additive genetic relationships between animals (Habier et al. 2007) as opposed to a matrix of expected additive genetic relationships derived simply from a pedigree (Henderson 1976). Computational advantages of the GBLUP as opposed to the rrBLUP solving approach are quite apparent when $m \gg q$; furthermore, it is computationally feasible using GBLUP to backsolve for BLUP ($\mathbf{g}$) from BLUP ($\mathbf{u}$) (Gauldron-Duarte et al. 2014; Strandén and Garrick 2009). An important point worth noting is that $\mathbf{u}$ as defined above only pertains to the portion of the genetic merit that is marked by genotypes (i.e., $\mathbf{Z}$ or $\mathbf{Z}^*$) such that it

has been often recommended to additionally model polygenic effects based on the use of pedigree information (Calus and Veerkamp 2007).

In addition to the rrBLUP specification provided above, Meuwissen et al. (2001) further conjectured that $\mathbf{g}$ might be better characterized by heavier tailed specifications like a scaled Student t density, i.e., $p(g_j|\vartheta) = t_\nu(0, \sigma_g^2)$, thereby requiring an additional degrees of freedom hyperparameter ν such that $\vartheta = [\,\sigma_g^2 \;\nu\,]'$. They labeled this model as BayesA. Meuwissen et al. (2001) further extended BayesA with a variable selection specification called BayesB such that $p(g_j|\vartheta) \equiv t_\nu(0, \sigma_g^2)$ with probability π for non-zero g_j and a point mass on zero with probability $1 - \pi$; in other words, $\vartheta = [\,\sigma_g^2 \;\nu\;\pi\,]'$ extends by yet one more hyperparameter π. Note that Meuwissen et al. (2001) did not work explicitly with a scaled Student t formulation but its well known hierarchical specification based on a scale mixture of normals (Andrews and Mallows 1974). That is, $g_j|\vartheta \sim t_\nu(0, \sigma_g^2)$ is marginally equivalent to $g_j|\sigma_{g_j}^2 \sim N(0, \sigma_{g_j}^2)$ followed by $\sigma_{g_j}^2|\sigma_g^2, \nu \sim \chi^{-2}(\nu, \nu\sigma_g^2)$ with $E(\sigma_{g_j}^2|\sigma_g^2, \nu) = \frac{\nu}{\nu-2}\sigma_g^2$. Meuwissen et al. (2001) and most other subsequent researchers refer to $\sigma_{g_j}^2$ as SNP-specific variances; however, since the dimension of each g_j is simply 1, the $\sigma_{g_j}^2$ should perhaps be merely characterized as augmented variables as discussed later. Also, Meuwissen et al. (2001) defined π as $\text{Prob}(\sigma_{g_j}^2 = 0) = \text{Prob}(g_j = 0)$ whereas we prefer the opposite definition of $\pi = \text{Prob}(g_j = 1)$ in this review, similar to what is more commonly defined in the conventional variable selection literature (O'Hara and Sillanpää 2009).

Since Meuwissen et al. (2001), there have been various other extensions including the use of alternative heavy-tailed specifications for $\mathbf{g}$ such as, for example, the Laplace distribution in Bayesian LASSO (de los Campos et al. 2009b), specifications based on modeling elements of $\mathbf{g}$ as correlated (Yang and Tempelman 2012), a mixture based on a normal distribution for non-zero $\mathbf{g}$ and a point mass on 0 in BayesC (Habier et al. 2011), a closely related mixture on two normals (high variance σ_g^2 for non-zero g_j, low variance $\frac{\sigma_g^2}{c}$ with $c \gg 1$ for "zero" g_j, both distributions centered on zero) as in stochastic search and variable selection or SSVS (George and McCulloch 1993; Verbyla et al. 2009), or even more extensive variable selection based on the mixture of three or more distributional components for $\mathbf{g}$ such as in BayesR (Erbe et al. 2012). This ever expanding library of Bayesian parametric approaches to WGP, with labels that typically involve various letter/symbol suffixes to "Bayes," has been satirically referred to as "Bayesian alphabet" analyses (Gianola 2013; Gianola et al. 2009). Criticisms of these approaches have been followed by a stronger advocacy for alternative approaches based on machine learning and/or nonparametric techniques (de Los Campos et al. 2009a; De los Campos et al. 2010b; González-Recio et al. 2014; Morota and Gianola 2014) with the basic premise being that these latter methods are better designed for prediction rather than for inference.

3. INFERENCE ON HYPERPARAMETERS

3.1. General Issues Concerning Specification of Hyperparameters

An important precursor for linear mixed model inference is how to specify variance components such as σ_g^2 or σ_e^2. As indicated previously, these hyperparameters are typically

estimated using REML such that one should really then refer to the corresponding estimates of $\mathbf{g}$ as empirical (e-) rrBLUP (Robinson 1991) leading to e-GBLUP of $u_i = \mathbf{z}_i' \mathbf{g}$ as being the estimated genetic merit of subject i; however, we will not make the distinction between rrBLUP and e-rrBLUP, for example, in this review. Typically, e-rrBLUP($\mathbf{g}$) is generally not meaningfully different from posterior means of $\mathbf{g}$ based on fully Bayes inference using MCMC on a rrBLUP model, provided that the posterior marginal density of the variance components σ_g^2 and σ_e^2 (i.e., $p(\sigma_g^2, \sigma_e^2|\mathbf{y})$) is reasonably symmetric (Gianola et al. 1986).

As with any hierarchical Bayesian model analysis, it is important to determine which hyperparameters are estimable and which are not. If hyperparameters are used to *structurally* specify distributional forms for exchangeable elements, such as $\mathbf{g}$ (and augmented variables $\{\sigma_{g_j}^2\}_{j=1}^m$ as another example) then they are generally estimable, although then arbitrarily vague or, conversely, justifiably informative priors with known parameters should be specified, in turn, on these hyperparameters. Examples not only include σ_g^2 in linear mixed models (e.g., rrBLUP or GBLUP), but additionally ν in BayesA and π in BayesB, for example. Arbitrarily specifying these hyperparameters can lead to a substantial degradation in performance as previously noted in WGP modeling (Lehermeier et al. 2013; Yang et al. 2015). Conversely, independent *subjective* priors with known hyperparameters are typically specified separately for parameters that are not considered to be exchangeable such as, for example, elements of $\boldsymbol{\beta}$. Further illustrations of structural versus subjective priors is provided in Bello et al. (2010) and Kizilkaya and Tempelman (2005).

For WGP models with hierarchical specifications beyond rrBLUP, such as BayesA and BayesB, it has not generally been too obvious how to estimate or specify other hyperparameters, even though these specifications have been widely recognized to be vitally important (Gianola 2013; Gianola et al. 2009; Hill 2014). That such specification or tuning is important should be intuitive. For example, as it pertains to the use of two different marker densities (i.e., different m or lengths of $\mathbf{z}_i'$, $i = 1, 2, \ldots, n$) for the same set of phenotypes (i.e., same n), it should be obvious that the genomic variance component σ_g^2 should differ between the two analyses accordingly, assuming that $\mathbf{z}_i'$ has not been rescaled any differently (e.g., all elements of $\mathbf{z}_i'$ are either 0, 1, or 2 in both cases). That is, using Eq. (1) and following de los Campos et al. (2013), $\text{var}(u_i) = \text{var}(\mathbf{z}_i'\mathbf{g}) = \mathbf{z}_i'\mathbf{z}_i\sigma_g^2$ using rrBLUP or $\text{var}(\mathbf{z}_i'\mathbf{g}) = \mathbf{z}_i'\mathbf{z}_i \text{E}(\sigma_{g_j}^2|\sigma_g^2, \nu) = \mathbf{z}_i'\mathbf{z}_i \frac{\nu}{\nu-2}\sigma_g^2$ using BayesA, should not inherently differ a great deal between two analyses differing only in the row dimension, m, of $\mathbf{z}_i'$. Hence, σ_g^2, being the typical or average SNP variance component, should decrease with larger m as the cumulative genetic variability would need to be distributed over a greater number of markers. Similarly, the specification of π should also be expected to decrease as m increases based on a variable selection procedure such as BayesB, as has been demonstrated in previous studies (Yang et al. 2015; Yang and Tempelman 2012). That is, the ratio of the "true" number (m_o) of QTL relative to m should decrease with increasing m. Furthermore, with $\pi < 1, \sigma_g^2$ should be substantially greater in BayesB than what is specified for BayesA since σ_g^2 is spread over a smaller number of non-zero markers in BayesB. Finally, as also determined by Yang and Tempelman (2012) and Yang et al. (2015), heavier tailed specifications (i.e., lower ν) are generally inferred in models without variable selection such as BayesA compared to BayesB. So as $\pi \cdot m$ increases (noting again that $\pi = 1$ in BayesA), the distribution of $\mathbf{g}$ should be inferred to be more extreme (i.e., heavier tailed) than those

obtained at lower marker densities and/or with BayesB variable selection; independently, de los Campos et al. (2013) also conjectured a more extreme distribution of effects with higher marker densities as well. To recapitulate, it should be obvious that these hyperparameters should not be arbitrarily specified but that attempts be made to estimate at least some of them.

3.2. METHOD OF MOMENTS APPROACHES

A seemingly reasonable strategy for estimating or tuning hyperparameters is based on a methods of moments strategy, often referred to as the heritability-based rules (de los Campos et al. 2013; Lehermeier et al. 2013; Pérez and de los Campos 2014). Recall again that $\text{var}(\mathbf{z}_i'\mathbf{g}) = \mathbf{z}_i'\mathbf{z}_i\sigma_g^2$ under a rrBLUP specification. Averaging across all n subjects then, the average or overall genetic variance is defined to be $MS_X\sigma_g^2$ where $MS_X = (1/n)\sum_{i=1}^{n}\mathbf{z}_i'\mathbf{z}_i$; for other hierarchical Bayesian models, the corresponding genetic variance depends upon other hyperparameters in $\boldsymbol{\vartheta}$ in addition to σ_g^2. Hence, if the heritability (h^2) of a trait is "known" which may be a dubious assumption for most field data analyses, one could multiply h^2 by the phenotypic variance σ_y^2 and equate it to $MS_X\sigma_g^2$ to solve or tune for σ_g^2 in a rrBLUP specification; de los Campos et al. (2013) further demonstrated how this same approach could be used to tune σ_g^2 in BayesA or BayesB models, holding constant ν and π.

However, as Lehermeier et al. (2013) duly noted, this strategy is based on rather strong assumptions, the most prominent being the independence of marker effects which may be tenuous if strong LD (i.e., strong multicollinearity) exists among the SNP markers. For BayesA and BayesB, Lehermeier et al. (2013) demonstrated that using methods of moments to provide these "optimal" values, as they described them, for σ_g^2 may lead to poorer cross-validation performance compared to specifying σ_g^2 to be 1/10 of these optimal values. Also, it may not be completely obvious whether to base hyperparameter specifications on means or modes using this approach (although the label "methods of moments" would imply the use of means), as they can be quite different from each other. Using the BayesA example described by Lehermeier et al. (2013), the mean of a $\sigma_g^2\chi^{-2}(\nu, \nu)$ is $\frac{\nu}{\nu-2}\sigma_g^2$ whereas the mode is $\frac{\nu}{\nu+2}\sigma_g^2$. So if, $\nu = 4$ as often arbitrarily specified, then the mean of the SNP specific variances is 3 times greater than that of the mode; however, it is not totally obvious which might be the better specification for σ_g^2 if one was going to equate either the mean or the mode to a "typical" value for SNP-specific variances.

3.3. CROSS-VALIDATION APPROACHES

Some researchers (Heslot et al. 2012; Usai et al. 2010) have suggested to use cross-validation based on a preselected grid of values to estimate or tune hyperparameters $\boldsymbol{\vartheta}$; tuning is then based on some criteria such as predictive performance in a validation subset based on estimates derived from a training subset. The utility of this approach is somewhat vulnerable to the choice of range and coarseness of grid of values of $\boldsymbol{\vartheta}$ chosen, which in turn is partly determined by the dimension of $\boldsymbol{\vartheta}$ and the available computing resources. Hence, as also noted by de los Campos et al. (2013) and Lehermeier et al. (2013), cross-validation can be prohibitively expensive as a strategy for hyperparameter inference.

As an interesting sidenote, several areas of study that have historically depended upon cross-validation-based inference for tuning parameters, including smoothing splines (Ruppert et al. 2009), have increasingly relied more upon REML-like approaches to estimate tuning parameters, particularly given the remarkable similarity to mixed model inference for some of these semiparametric models. In fact, estimates based on REML tend to display greater stability than estimates derived from cross-validation tuning (Wang 1998). This advantage of REML has been increasingly recognized for nonparametric WGP approaches based on, for example, kernel regression (Endelman 2011).

3.4. Formal Inference on Hyperparameters

Suggesting then that REML is a suitable strategy for estimating variance components in rrBLUP (Endelman 2011), how should hyperparameter inference be conducted in WGP models facilitating heavy-tails (BayesA, Bayesian LASSO) and/or variable selection (BayesB, BayesC)? In principle, hyperparameter inference should not really differ from variance component estimation, such that likelihood and/or Bayesian inference strategies should be considered. For example, Pinheiro et al. (2001) describe likelihood-based strategies for estimating hyperparameters in the situation where both $\mathbf{g}$ and $\mathbf{e}$ are scaled Student t distributed; in other words, their model is BayesA-like except that robust heavy-tailed behavior is also conferred on the residuals. Subsequently, MCMC strategies for inferring hyperparameters in BayesA and Bayesian LASSO models (Yi and Xu 2008), as well as BayesB models (Yang et al. 2015; Yang and Tempelman 2012) have been provided. Conceptually, if fully Bayes inference (i.e., using MCMC) is applied to a BayesB model, model averaging is essentially conducted over several models since BayesA (i.e., $\pi = 1$), rrBLUP ($\pi = 1, \nu \to \infty$), and BayesC ($\nu \to \infty$) are all special cases of BayesB.

However, there have been some challenges in attempting to estimate all hyperparameters, most notably ν (Habier et al. 2011) because of the naturally high posterior correlation between some of these hyperparameters. As another example, Gianola (2013) cite Duchemin et al. (2012) who suggested identifiability problems with BayesC, noting that as π increased, estimates of σ_g^2 went down; in other words, the conclusion was drawn that π was not separately estimable from σ_g^2. However, similar phenomena have been characteristic of many other quantitative genetic models such that innovative MCMC strategies have been proposed to help alleviate similar problems (Shariati and Sorensen 2008). Yang et al. (2015) also noted the high posterior correlation between ν and σ_g^2 in both BayesA and BayesB models, yet proposed effective strategies based on the collapsed Gibbs sampler (Liu 1994) to improve MCMC mixing for jointly inferring both hyperparameters. Conceptually, MCMC inference on ν and σ_g^2 in BayesA or BayesB should not be any different, for example, from inferring upon similarly defined parameters in scaled Student t residual models as has been successfully demonstrated previously (Rosa et al. 2003). Admittedly, however, inferential issues would be more challenging for hyperparameters characterizing the distribution of $\mathbf{g}$ than that of $\mathbf{e}$, not just necessarily because of the difference in dimensionality ($m \gg n$), but also because second stage prior specifications (i.e., on $\mathbf{g}$) are further removed away from $\mathbf{y}$ relative to the first stage likelihood specifications (i.e., on $\mathbf{e}$) in the model hierarchy. In other words, there is typically less information to estimate the (hyper)parameters that characterize

the distribution of $\mathbf{g}$ relative to those that characterize the distribution of $\mathbf{e}$. Many researchers arbitrarily set ν in BayesA or BayesB models to values within $4 \leq \nu \leq 6$, as those values represent specifications with rather heavy tails but where the first three central moments are defined for a scaled Student t. However, Nadaf et al. (2012) has demonstrated that posterior inferences of ν concentrated near 1 can lead to better WGP performance of BayesA models compared to arbitrarily setting $\nu = 4$.

One criticism that might be directed at much of the current MCMC-based research on WGP modeling is the general lack of accountability regarding whether or not one has demonstrably drawn a sufficient number of MCMC samples, including the number of "burn-in" samples to ensure that the sampler has converged in distribution to the true joint posterior density (Gelman et al. 2014), before samples are even saved for inference. It is well known that MCMC samples are highly autocorrelated over adjacently drawn cycles, particularly in models that are highly parameterized; hence Monte Carlo error can be substantial if not enough samples are drawn. A popular MCMC diagnostic metric is the effective sample size (ESS) which estimates the number of effectively independent random samples from the joint posterior distribution (Plummer et al. 2006). It is not unusual to find WGP studies where less than a total of 100,000 MCMC samples have been drawn even though Yang and Tempelman (2012) has noted the need for far more MCMC cycles in some WGP analyses just to ensure ESS $>$ 100 for some hyperparameters. Nevertheless, it should be quickly added that 100 might be considered to be only a barely sufficient ESS to provide reliable point estimates (i.e., posterior means), but certainly not reliable posterior densities. This problem is particularly exacerbated for higher density SNP marker panels. Hence, further research for improving computational efficiency of MCMC schemes in WGP modeling is badly needed with emerging marker technologies (Yang et al. 2015).

4. DATA AUGMENTATION VERSUS THE CURSE OF DIMENSIONALITY

4.1. HIERARCHICAL CONSTRUCTIONS

As indicated earlier, many of the proposed Bayesian alphabet models are merely extensions of rrBLUP (or, equivalently, GBLUP) whereby $p(g_j|\vartheta)$ is often specified to have distributional assumptions other than Gaussian but for which the Gaussian is a special case subset model. Indeed, most of these models are typically constructed by specifying Gaussian densities for $\mathbf{g}$ conditional on SNP-specific *augmented variables* $\boldsymbol{\delta} = \{\delta_j\}$, i.e., $p(g_j|\sigma_g^2, \delta_j) \equiv N(0, \sigma_g^2\delta_j)$. Using $\boldsymbol{\delta}$ then, we reparameterize the hierarchical construction for BayesA very slightly from Meuwissen et al. (2001) using $\delta_j|\nu \sim \chi^{-2}(\nu, \nu)$, i.e., a scaled inverted chi-square density defined such that $\mathrm{E}(\delta_j) = \frac{\nu}{\nu-2}$. Nevertheless, it can be readily demonstrated that marginally $p(g_j|\vartheta) \equiv t_\nu(0, \sigma_g^2)$ using this specification as with BayesA. BayesB additionally builds on top of BayesA by adding yet an another layer of augmented, or rather, *indicator variables* which define membership conditionally into either one of two mixture components: one defining a scaled Student t specification for non-zero g_j and the other being a point mass for g_j on 0. Data augmentation is considered to be extremely useful in MCMC inference because of the convenience of creating full conditional densities

that are readily recognizable, albeit at the expense of increasing the number of variables or working parts (Tanner and Wong 1987).

Now Gianola et al. (2009) and Habier et al. (2011) warned not only about the sensitivity of $\mathbf{g}$ but also the sensitivity of augmented variables $\boldsymbol{\delta} = \{\delta_j\}_{j=1}^{m}$ to hyperparameter specifications. It is commonly perceived that the use of these augmented variables unduly leads to highly parameterized models; for example, BayesA is typically characterized as having at least $2m$ "parameters" if one counts both the dimension of $\mathbf{g}$ and $\boldsymbol{\delta}$. Hence BayesA, for example, is typically perceived to not allow sufficient "Bayesian learning" from $\mathbf{y}$ for meaningful inference on SNP-specific variances or, equivalently, $\boldsymbol{\delta}$ (Gianola et al. 2009). However, it is important to remember again that the marginal distribution of $\mathbf{g}$ is scaled Student t in BayesA, such that, in principle, there really is only one additional parameter (i.e., ν) relative to rrBLUP. That is, a scaled Student t distribution on the random effects really only constitutes one additional parameter (i.e., ν) relative to a normal distribution once $\boldsymbol{\delta}$ is integrated out, recalling that as $\nu \to \infty$, BayesA converges to rrBLUP.

Data augmentation has been extensively used for MCMC inference in animal breeding outside of WGP applications, including probit mixed models (Sorensen et al. 1995), censored data models (Sorensen et al. 1998) and robust error models (Rosa et al. 2003) without similar concerns. Perhaps one useful approach to address this dimensionality concern is to consider methods that infer upon the effective number of parameters such as the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002) as has been used by Yi et al. (2014) for WGP models. It is intriguing that currently there are proposals to use estimates of augmented variables from previous analyses as inputs to subsequent or updated WGP analyses (Calus et al. 2014); however, this practice deserves further study over and beyond the realization that LD associations between SNP markers and causal variants will break down over subsequent generations (Fragomeni et al. 2014).

4.2. Effects of Determinedness and Model Complexity

Recently, Wimmer et al. (2013) has suggested that determinedness, defined as the ratio n/m, and model complexity, defined as the ratio m_o/n may influence the relative utility of hierarchical Bayesian specifications for WGP analyses. Specifically, Wimmer et al. (2013) concluded that as m_o/n increased and n/m decreased, the effectiveness of hierarchical Bayesian WGP models, whether based on heavy-tailed and/or variable selection specifications, for improving prediction accuracy beyond rrBLUP diminished. Furthermore, their performance relative to rrBLUP particularly degraded with higher average levels of pairwise LD between the markers, an observation also independently noted by de los Campos et al. (2013). However, in Wimmer et al. (2013), specifications on various elements of the hyperparameters were either arbitrarily chosen, with different values of π chosen for different species/traits based on somewhat ad hoc criteria, or based on the aforementioned methods of moments approach for σ_g^2. It might be curious to reassess these comparisons based on the more formal inferential approaches as described earlier. At any rate, there appears to be, on average, a slight advantage in predictive accuracy for heavy-tailed and/or variable selection specifications for WGP based on an extensive review of the literature thus far (de los Campos et al. 2013). As an interesting sidenote, Vattikuti et al. (2014) also focused their

attention on the joint effects of determinedness and model complexity but for GWA which is addressed in a subsequent section.

4.3. INCREASING MARKER DENSITIES

O'Hara and Sillanpää (2009) suggested that m should be no more than 10–15 times greater than n when using any penalized regression analyses such as those suggested in this review whereas González-Recio et al. (2014) has suggested a corresponding m:n ratio cap of 50–100. With a 777K SNP chip as becoming increasingly common for Holstein cattle (Wiggans et al. 2011), this conservatively would imply that the number of animals should be no less than, say, 50,000. But with emerging gene chips based on 28M+ different SNP markers based on sequencing technologies (Daetwyler et al. 2014), it seems apparent that many more phenotypic records will be needed than are currently available for many livestock species, particularly in light of the fact that multicollinearity among adjacent markers will be typically extreme. Within a completely different context, Allenby et al. (2014) wrote that focusing on statistical sufficiency or data compression will become increasingly important, drawing the analogy that inferences will be akin to finding a needle in a haystack that keeps getting increasingly bigger. Yang et al. (2015) recognized that reliable inferences on hyperparameters could be rather formidable with extremely high marker densities, particularly as MCMC diagnostic performance (i.e., ESS) degrades substantially for the same number of MCMC cycles with a higher m/n while the computational cost of each MCMC cycle increases proportionally to m. They proposed that inferences on $\mathbf{g}$, regardless of which WGP model is used, could be based on values of $\boldsymbol{\vartheta}$ extrapolated from inferences based on lower marker density panels; nevertheless, further work is badly needed in this area.

It is also important to recognize, as elucidated by de los Campos et al. (2013) that the implications of increasing marker densities are much different in animal and plant breeding, where LD is extensive and large, compared to human genetics where LD is considerably lower. That is, de los Campos et al. (2013) suggested that pursuing $m > 10,000$ may not lead to any sizeable increases in prediction accuracy for some dairy cattle traits whereas human research continues to benefits from m exceeding hundreds of thousands. However, these results did not factor into account the potential impact of haplotyping which is addressed subsequently in this review.

4.4. MODELING CORRELATION

Most WGP researchers do not model correlation between elements of $\mathbf{g}$, particularly those pertaining to SNP markers in close proximity to each other. Yang and Tempelman (2012) proposed modeling nonstationary correlation among the elements of $\mathbf{g}$ using a first-order antedependence specification based on the order of SNP markers within chromosomes, believing that any such dependence between elements of $\mathbf{g}$ in the vicinity of a QTL was not simply due to multicollinearity. One proposed model, ante-BayesA, was based on superimposing first-order antedependence relationships on $\mathbf{g}$ within a scaled Student t distributional framework with a second model, ante-BayesB, further superimposing variable selection on top of ante-BayesA. They determined that prediction accuracy was enhanced in these mod-

els relative to their conventional BayesA and BayesB counterparts, using both simulation and application to weight data on mice.

de los Campos et al. (2013), citing work from Hastie (2009), noted that when predictors are correlated, something that occurs when LD span over long regions, methods performing variable selection were often outperformed by rrBLUP. Since QTL in large LD blocks are usually characterized by a high degree of correlation between elements of $\mathbf{g}$ in those regions, it might seem worthy to further investigate how specifying antedependence might facilitate variable selection in these situations using, for example, ante-BayesB. Furthermore, it also seems reasonable to consider using nonstationary correlation structures like first order antedependence to estimate the effective number of markers, in much the same manner that the numerator relationship matrix between animals can be used to infer the effective population size or effective number of unrelated individuals (Pérez-Enciso 1995).

5. COMPUTATIONAL EFFICIENCY

5.1. IMPROVING MCMC SCHEMES

As noted previously, animal breeders are naturally concerned as much about computational efficiency as they are about statistical modeling given ever increasing m and n. This may again explain the partial reluctance, for example, to conduct formal inference on $\boldsymbol{\vartheta}$ or modeling correlation between elements of $\mathbf{g}$ in Bayesian alphabet models. Furthermore, many scientists working most closely with the industry on very large datasets are rather content with the relative performance of GBLUP based on standard REML/BLUP methods, particularly since GBLUP has been demonstrated to be amenable to including information on non-genotyped animals as described later. Other than GBLUP, most other hierarchical Bayesian WGP models have been analyzed using MCMC which is often deemed to be computationally expensive. Furthermore, MCMC does not allow a form of 'memory' when re-running or updating analyses based on the collection of more data. As Allenby et al. (2014) attests, "real-time" posterior densities are very difficult to get with big data using MCMC in an era where just-in-time analysis updates are increasingly demanded (Liu et al. 2014). This is particularly true in animal breeding where inferences on breeding values are continually updated (Wiggans et al. 2015).

To address this concern, it has been suggested to run multiple shorter chain analyses and using meta-analytic methods to combine the results based on previous MCMC analyses (Allenby et al. 2014). In the meantime, researchers continue to strive to improve computational efficiency of MCMC sampling schemes in animal breeding (Shariati and Sorensen 2008; Shariati et al. 2009; Yang et al. 2015). Animal breeding scientists have been particularly astute in developing computational enhancements for WGP models including parallel computing (Fernando et al. 2014; Wu et al. 2012), Gauss Seidel residual updating (Legarra and Misztal 2008), or other right-hand-side updating strategies (Calus 2014). Although most code created for public domain use is written using efficient low-level programming languages such as FORTRAN, C or C++ (Aguilar et al. 2014; Fernando 2009), often with convenient R wrapper packages (Endelman 2011; Pérez and de los Campos 2014; Wimmer et al. 2012), the animal breeding community nevertheless has also been conscientious to

provide lucid R code for pedagogical purposes as well (Gondro et al. 2013) in an attempt to train and engage future researchers to further pursue these challenges.

5.2. APPROXIMATE ANALYTICAL APPROACHES

Provided that it is carefully conducted, MCMC provides a current gold standard for fully Bayesian inference. However, because of its computational burden, there is a desire to consider other approximate analytical approaches for Bayesian alphabet models beyond GBLUP. Strategies based on expectation maximization (EM) have drawn increasing interest for WGP models in animal breeding (Kärkkäinen and Sillanpää 2012). In fact, EM has been characterized as a form of "big data Bayes" (Allenby et al. 2014) whereby the primary focus is adapting computationally efficient strategies to compute joint posterior modes. Furthermore, EM is amenable to analysis updates whereby one can readily use solutions from a previous analysis as starting values for the current analysis.

Nevertheless, EM implementations in WGP modeling have faced several criticisms. Firstly, it has been noted by de los Campos et al. (2013) that the vast majority of current EM WGP proposals do not provide estimates of uncertainty about the estimated marker effects (**g**) or breeding values (**u = Zg**). This is in spite of the fact that there are well established strategies to provide the observed information matrix or asymptotic standard errors for such estimates when using EM (Louis 1982; Meng and Rubin 1991). However, perhaps an even more serious concern with these EM schemes is that they are known to converge or get trapped at local modes that can be very different from global maxima (de los Campos et al. 2013; Gianola 2013). This is a particular concern as much previous WGP EM-based research has been based on using **g = 0** as starting values as noted by Chen and Tempelman (2015).

As illustrated by Kärkkäinen and Sillanpää (2012), the Bayesian LASSO prior on g_j can be constructed in at least two different ways. Firstly, it can be constructed directly or marginally as $g_j \sim \text{Laplace}(g_j|\lambda)$, i.e., a Laplace distribution characterized by hyperparameter λ. Kärkkäinen and Sillanpää (2012) labeled the corresponding model based on the direct use of this prior as the nonhierarchical Laplace approach. However, as similarly demonstrated previously for BayesA, the Bayesian LASSO prior can be also constructed hierarchically, i.e., $g_j \sim N(0, \sigma_g^2\delta_i)$ followed by $\delta_i \sim \text{Exp}(\lambda))$ where $\boldsymbol{\delta} = \{\delta_j\}$ again represents a vector of augmented variables that one must contend with in this so-called hierarchical Laplace approach. Based on a simulation study, Kärkkäinen and Sillanpää (2012) concluded that use of the hierarchical versus nonhierarchical Laplace models lead to substantially different EM-based inferences on **g**. That this could be even possible is rather difficult to understand given that marginally both constructions essentially define the same model; however, perhaps convergence of **g** was to low probability posterior modes for one or both (i.e., nonhierarchical versus hierarchical) of the alternative constructions.

Recently, a potential solution to identifying sparse high posterior probability modes that appears to be resilient to starting values has been proposed for high dimensional variable selection models based on the use of SSVS modeling within an EM framework (Rockova and George 2014a) and has been adapted for WGP in this special issue (Chen and Tempelman 2015). Subsequent developments used to help mitigate the effects of high levels of

multicollinearity for high dimensional SSVS regression are further proposed in Rockova and George (2014b) and warrant further investigation for WGP modeling.

One final critical issue is how to estimate hyperparameters ($\boldsymbol{\vartheta}$) in EM-like implementations. Kärkkäinen and Sillanpää (2012) deemed such a process to be nearly impossible. Chen and Tempelman (2015) propose a REML like strategy to estimate all elements of $\boldsymbol{\vartheta}$ in a SSVS or BayesA WGP within an EM inference framework. An obvious limitation using EM is that uncertainty in the hyperparameters may not be adequately accounted for as it would be in a fully Bayesian analysis. For example, Lehermeier et al. (2013) noted that formally allowing for uncertainty on the key hyperparameter in Bayesian LASSO was preferred to assigning fixed "optimal" values to that same hyperparameter, as the former strategy better allowed for Bayesian learning from the data. At any rate, alternative "empirical Bayes" like strategies for inferring hyperparameters in WGP models seem warranted, such as the strategy proposed by Perez-Elizalde et al. (2015) for kernel regression WGP models in this special issue, the integrated nested Laplace approximation (Rue et al. 2009) or Laplace's method (Tierney and Kadane 1986). Variational Bayes (Logsdon et al. 2010) and approximate Bayesian computation (Technow et al. 2015) are emerging analytical approaches worthy of further pursuit in WGP modeling as they may overcome the noted limitations of EM-based estimation while allowing inference that more closely matches MCMC but at a fraction of the computational cost.

6. GENOME-WIDE ASSOCIATION ANALYSES

GWA analysis refers to quantitative genetic analysis that is focused on inferences on SNP effects (i.e., $\mathbf{g}$) rather than on predicting cumulative genetic merit (i.e., $\mathbf{u} = \mathbf{Zg}$) of all individuals. This area is widely recognized to be plagued by lack of reproducibility across studies (Li and Meyre 2012). Before the advent of WGP models, many GWA analyses were simply based on a series of single marker regression analyses. However, this was eventually recognized to be suboptimal for various reasons, including the inability to account for population structure (Hayes 2013). Perhaps one of the most popular tests that account for population structure is called EMMA (Kang et al. 2008). In EMMA, the basic linear model [2], based on the rrBLUP assumptions, is adapted except that the effect g_j for the SNP marker of interest is treated as fixed in a series of mixed model analyses for $j = 1, 2, \ldots, m$; i.e.,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z}_j g_j + \mathbf{Z}_{-j}\mathbf{g}_{-j} + \mathbf{e}. \tag{3}$$

Here $\mathbf{Z}_{-j}$ refers to $\mathbf{Z}$ without column j where $\mathbf{z}_j$ refers to column j of $\mathbf{Z}$ and not to be confused with $\mathbf{z}_i'$ from Eq. (1) which refers to row i of $\mathbf{Z}$. Typically in EMMA and other related work (Gauldron-Duarte et al. 2014), Eq. (3) is rewritten using $\mathbf{u} = \mathbf{Z}_{-j}\mathbf{g}_{-j}$ such that $\text{var}(\mathbf{u}) = \mathbf{Z}_{-j}\mathbf{Z}_{-j}'\sigma_g^2$ although the typical simplifying assumption is made that $\mathbf{Z}_{-j}\mathbf{Z}_{-j}' \approx \mathbf{ZZ}'$ for all m analyses involving Eq. (3) in order to save computing time. Like $\boldsymbol{\beta}$, g_j is treated as a fixed effect such that mixed model inferences for testing H_o: $g_j = 0$, $j = 1, 2, \ldots, m$ is then typically based on a t- or z-based *fixed effects test*. This test statistic is derived by dividing the generalized least squares (GLS) estimate of g_j by its estimated

standard error, typically basing the P value on a Bonferroni correction or other multiple testing adjustments (Storey and Tibshirani 2003) as determined by the total number (m) of SNP markers tested.

It might seem vexing that one would treat all elements of **g** as random (or having some specified distributional form) in a WGP analysis whereas the element of **g** that is of specific interest is momentarily treated as fixed in a GWA analysis. Alternatively, a *random effects test* might be based on reverting from Eq. (3) back to Eq. (1) or (2), thereby treating all elements of **g** as random, such that the point estimate of g_j would be BLUP and the denominator of a corresponding GWA test statistic would be the standard error of prediction for BLUP(g_j) defined as the square root of the prediction error variance (Gauldron-Duarte et al. 2014). Note that the latter is also synonymous with the posterior standard deviation whereas BLUP(g_j) equates to the posterior mean from a Bayesian perspective; hence, the corresponding GWA test statistic could be referred to as a posterior z score (Gelman et al. 2012). One obvious reason why the random effects test is often avoided by GWA researchers that the shrinkage of BLUP(g_j) to zero tends to be very "hard" (Hayes 2013) such that the random effects test is rather conservative. In an example applied to backfat thickness in a pig resource population based on a 44K (44,055) SNP marker panel, Gauldron-Duarte et al. (2014) observed that the unadjusted (i.e., with no Bonferroni correction) shrinkage or random effects test P values never went beneath 0.20! Conversely, statistically significant (Bonferroni adjusted $P < 0.05$) SNP effects were determined using the fixed effects or EMMA test for two different genomic regions previously determined as being important for that trait. The EMMA test also appeared to preserve Type I error rates based on genome reshuffling or permutation (Gauldron-Duarte et al. 2014).

Gianola (2013) wrote that BLUP is biased with respect to specific marker effects. Does this reminder and the evidence provided by Gauldron-Duarte et al. (2014) then refute that BLUP or "shrinkage is a good thing" (Allison et al. 2006; Robinson 1991), particularly as even greater shrinkage of point estimates of **g** toward zero is to be far more likely as one increases m/n? Gelman et al. (2012) seem to convincingly argue that shrinkage inference in the context of high dimensional covariate inference is actually highly preferable to fixed effects hypothesis testing, even when inference is focused on the individual regression coefficients. With shrinkage comes borrowing of information across SNP such that point estimates of g_j are shifted toward each other and toward zero; hence, it is far less likely for false positive results to occur with random effects testing as opposed to fixed effects testing, even when a Bonferroni adjustment is invoked for the latter, albeit not for the shrinkage-based random effects test. GWA actually characterizes a situation where it is rather unlikely for the global null hypotheses to be true; that is, most g_j will truly be non-zero, not only because most traits are influenced by many genes (often thousands), but also because most SNP markers are in non-zero LD with at least one of these genes. Hence, the generally desired preservation of global Type I error may be somewhat misguided for GWA studies. Gelman et al. (2012) further argue that the shrinkage penalty itself further builds multiple testing concerns directly into the model such that there is no need to invoke multiple testing adjustments on a random effects or posterior z score test.

However, since shrinkage estimation inherently involves borrowing of information across exchangeable entities, a greater burden is placed on the model, i.e., careful specification of

the prior on **g**. Recalling the harsh shrinkage provided by a Gaussian prior with rrBLUP as described by Hayes (2013), one should alternatively consider pursuing sparser prior specifications (i.e., heavy-tailed, variable selection or both) for GWA in a manner similar to that described in Yi and Banerjee (2009) and Yi et al. (2014). This effectively facilitates far less shrinkage for truly large values of g_j whereas smaller values of g_j are shrunk even harder to 0 compared to the use of a Gaussian prior. The use of variable selection priors in GWA are particularly attractive in that one can estimate the posterior probability that g_j is non-zero; this estimate might be considered an even more robust measure of association as opposed to the use of P values (Hayes 2013). Furthermore, these posterior probabilities have a strong theoretical connection with false discovery rate reporting (Käll et al. 2008).

Because of potentially high correlation between inferences on elements of **g** deriving from adjacent SNP markers in the vicinity of a QTL, some individuals have advocated the use of GWA inferences based on average estimates of g_j contained within moving genomic "windows" (Fernando and Garrick 2013). A window would be defined as a subset of adjacent markers such that the estimate of g_j for the SNP in the middle of that window is based on an average of the estimates of all g_j within that window. Since it can be rather challenging on how to properly specify the widths or boundaries of such windows, it might also be useful to consider formal correlation modeling strategies, such as first order antedependence, for example, to adaptively infer upon the appropriate width of such windows. Alternatively a procedure based on smoothing splines has been recently proposed by Beissinger et al. (2015).

7. MISCELLANEOUS ISSUES

7.1. Combining Information on Genotyped and Non-genotyped Animals

As indicated earlier, animal breeders have been particularly resourceful in combining developments in statistical, algorithmic, and computational sciences in addressing the needs of the livestock industries. One particular WGP need that required a solution was how to combine genetic evaluations on animals that have genomic marker information with animals that do not, particularly since high genotyping costs have generally precluded the use of genotyping on all animals. Christensen and Lund (2010) and Legarra et al. (2009) independently derived a model now known as single-step GBLUP or ss-GBLUP (Legarra et al. 2014) by partitioning $\mathbf{u}$ into components from non-genotyped ($\mathbf{u}_1$) and genotyped animals ($\mathbf{u}_2$), i.e., $\mathbf{u} = [\,\mathbf{u}_1' \;\mathbf{u}_2'\,]'$. The multivariate Gaussian prior density of $\mathbf{u}$ specified in ss-GBLUP is loosely derived by conditioning arguments, i.e., writing $p(\mathbf{u}) = p(\mathbf{u}_1|\mathbf{u}_2)\, p(\mathbf{u}_2)$ whereby $p(\mathbf{u}_2)$ is based on using the realized genomic relationship matrix between genotyped animals, as indicated previously with GBLUP, whereas $p(\mathbf{u}_1|\mathbf{u}_2)$ is based on imputing realized relationships involving non-genotyped animals based on their pedigree-based relationships with each other and with genotyped animals. The resulting hybrid genetic/genomic relationships, or correlation matrix, between all of $\mathbf{u} = [\,\mathbf{u}_1' \;\mathbf{u}_2'\,]'$ is typically referred to as the "**H**" matrix, thereby making its usage particularly amenable for large scale BLUP-based genetic/genomic evaluations (Fragomeni et al. 2015) within a GBLUP modeling framework.

The ss-GBLUP has been recently extended to a fully hierarchical Bayesian framework by Fernando et al. (2014) with extensions for GWA analysis and an approximate "weighted" ss-GBLUP approach that closely emulates BayesA both provided by Wang et al. (2012).

7.2. NONPARAMETRIC VERSUS PARAMETRIC APPROACHES?

Various nonparametric approaches to WGP have been more extensively advocated such as kernel regression (Morota and Gianola 2014) and various other machine learning approaches (González-Recio et al. 2014). Although they are not nearly used as widely as the Bayesian alphabet models described thus far, they have been periodically demonstrated to be superior with respect to predictive accuracy, particularly when epistasis is extensive (Howard et al. 2014). Nevertheless, it may be possible that some comparisons may have been affected by improper tuning/inference of hyperparameters in both nonparametric and Bayesian alphabet models as discussed previously. At any rate, nonparametric methods, which typically involves specifying predictions as nonlinear functions of the SNP genotypes, appear to have very attractive WGP properties and indeed deserve greater attention.

A current conceptual limitation is that it has not been clearly elucidated how these nonparametric approaches could be modified for inference, specifically for GWA types of analyses. Superior prediction properties for WGP (i.e., on $\mathbf{u}$) may not necessarily equate with efficient inference on SNP-specific regression coefficients (i.e., on $\mathbf{g}$) which does not appear to be currently well defined, if at all, with nonparametric methods for GWA. This may be another fruitful area for further study.

7.3. HAPLOTYPING

Animal breeders are gradually moving away from biallelic SNP-based modeling, as predominantly described in this review, to multi-allelic haplotype effects in $\mathbf{Z}$ realizing that SNP haplotypes (i.e,. groupings of SNP in high LD with each other) are likely to be in higher LD with QTL than the individual biallelic SNP themselves (Cuyabano et al. 2014; Hayes 2013). Now, this is not necessarily a trivial task as this implies that elements of $\mathbf{Z}$ needs to be inferred based on a process called phasing, often with considerable uncertainty particularly since the size of the blocks is often based on an arbitrary LD threshold. From a modeling perspective, this involves modifying Model [2] such that $\mathbf{g}$ is now partitioned into $b < m$ genome-location defined haplotype blocks $\mathbf{g} = [\,\mathbf{g}_1'\ \mathbf{g}_2'\ \cdots\ \mathbf{g}_b'\,]'$ with the dimension of each such component $\mathbf{g}_j$ of $\mathbf{g}$ being defined by the number of unique haplotypes in that block. The corresponding columns of $\mathbf{Z} = [\,\mathbf{Z}_1\ \mathbf{Z}_2\ \cdots\ \mathbf{Z}_b\,]$ again represent 0–1–2 covariate assignments (number of copies of each corresponding haplotype) within each of the b blocks.

It is curious that whereas fixed effects hypothesis testing has been advocated for GWA biallelic marker modeling as described earlier, there has been no philosophical difficulty treating elements of $\mathbf{g}$ as random effects for GWA analyses based on the use of multi-allelic haplotypes (Hayes 2013). This may relate to the fact that the number of haplotypes within each block is typically large enough to conveniently treat $\mathbf{g}_j$, $j = 1, 2, \ldots, b$ as random. Doing so allows GWA evidence to be based on a likelihood ratio test for that block-specific variance component providing, in essence, a "blessing of dimensionality."

Indeed, haplotyping-based inference will blur the distinction between what might be considered augmented variables versus haplotype block-specific variance components. At any rate, it seems increasingly likely that haplotype-based modeling, whether for WGP or for GWA, will be much more informative than biallelic SNP modeling, as has been described previously, and thereby likely improve the relative performance of, say, variable selection-based strategies to determine genomic regions of interest. In particular, haplotyping-based WGP or GWA analyses will make far more efficient use of information deriving from high marker density panels or sequencing technologies, i.e., much larger m. Even so, it will still be imperative to properly estimate or tune hyperparameter specifications in these analyses as well, although haplotype-based inference should lead to greater success with stable hyperparameter inference than what has been attained up until now.

7.4. Criteria for Assessing Goodness of Fit or Superiority of Performance

Even with simulation studies based on generating historical LD between SNP and additive SNP effects, we truly do not have a reasonable way of indicating what the "true" values of the hyperparameters should be with parametric models. For example, π cannot be really interpreted as the proportion of all SNPs that are QTL, for example, because π innately depends upon the strength of LD associations between markers. Hence simulation studies have primarily focused on comparisons between methods for prediction of genetic merit (i.e., $\mathbf{u}$) where analyses of real data have focused on cross-validation measures of accuracy such as correlation which can be faulty measures of reproducibility (Irizarry 2015).

Given these concerns, it may be useful to explore other metrics for performance, including those discussed near the end of González-Recio et al. (2014) or the Hellinger distance as proposed by Lehermeier et al. (2013). Plasmode-based permutations of the data are also useful to address false positive performance rates between competing methods for GWA (Gauldron-Duarte et al. 2014). At any rate, this is an area that is dire need of further study and investigation.

7.5. Structured Populations and Genotype by Environment Interaction

In heterogeneous populations where different strains or groups exist, marker effects ($\mathbf{g}$) should be allowed to vary between groups. Various promising statistical models are presented and tested by de los Campos et al. (2015) to allow for such heterogeneity. Similarly, when the same strains or breeds are distributed across various environments, one should allow for potential genetic by environment (G×E) interactions, particularly in an era of potentially disruptive climate changes (Pregitzer et al. 2013). Nevertheless, G×E modeling only further intensifies the curse of dimensionality. Recently, Jarquin et al. (2014) presented a WGP model for G×E based on high dimensional environmental covariates and genotypes using kernel-based covariance functions. Nevertheless, they suggested that their Gaussian prior specifications might be far too limiting, such that future research in variable selection and/or heavy-tailed specifications are warranted.

7.6. USING BIOLOGICAL INFORMATION TO AID WGP AND GWA

As the number of markers increase, it may be useful to consider how prior information on known coding regions or QTL could be used to selectively choose or upweight SNP markers in those regions relative to other markers. Although Schoen et al. (2014) have noted such advantages in doing so, Morota et al. (2014) and Janss (2015) separately concluded that markers in functionally annotated coding regions did not necessarily provide superior predictive ability for WGP relative to SNP markers in noncoding regions. At any rate, it is becoming increasingly important to link inferences on **g** to their cellular functions or to their corresponding transcript or protein expression, being a basic objective of 'omics research (Kadarmideen 2014). Given the increasing emphasis on cross-disciplinarity as it pertains to statistical research (Brown and Kass 2009), a statistician or quantitative geneticist working in WGP or GWA will have far greater impact once he or she studies the biological context of the work so that he/she is better able to generate ideas for future scientific investigation.

8. CONCLUDING REMARKS

Formidable challenges exist for WGP and GWA modeling, particularly, as real-time analyses are becoming increasingly desired for both focus areas while, at the same time, the number of marker covariates and phenotypes continue to increase. Although research on alternative inferential approaches are needed, greater care is needed in correctly inferring parameters in current WGP models, particularly those defined at the highest levels of the model hierarchy (i.e., hyperparameters). Data augmentation schemes will require further refinements (Van Dyk and Meng 2001) to order to facilitate computation for hierarchical Bayesian WGP modeling. Furthermore, hypothesis testing strategies for GWA, perhaps away from P value-based inferences (Halsey et al. 2015), should be reconsidered in favor of shrinkage-based inferences in order to help improve the reproducibility of these studies. Of course, any proposed method also needs to respect the overriding computational efficiency concerns so that it can be pragmatically adopted! It should be also quickly added that any future enhancements to both WGP and GWA modeling extend readily to human health applications, particularly given the increasing interest in harnessing WGP for personalized medicine, for example (de los Campos et al. 2010a).

Although intellectual property and/or economic considerations have limited the accessibility of genomic marker data (Jorasch 2005), more SNP data is becoming publicly available in an effort to further harness the intellectual energy of the statistical and quantitative genetics community (de Koning and McIntyre 2012). Other useful resources include the 1000 genomes project (http://www.1000genomes.org/) and the Framingham Heart Study (https://www.framinghamheartstudy.org/researchers/description-data/genetic-data.php).

ACKNOWLEDGEMENTS

The author gratefully acknowledges support from Agriculture and Food Research Initiative Competitive Grant No. 2011-67015-30338 from the USDA National Institute of Food and Agriculture.

[Received September 2015. Accepted September 2015.]

REFERENCES

Aguilar, I., Misztal, I., Tsuruta, S., Legarra, A., and Wang, H. (2014). PREGSF90–POSTGSF90: computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production, ASAS, August 17-22, 2014, Vancouver, BC, Canada*.

Allenby, G. M., Bradlow, E. T., George, E. I., Liechty, J., and McCulloch, R. E. (2014). Perspectives on Bayesian Methods and Big Data. *Customer Needs and Solutions* **1**, 169-175.

Allison, D. B., Cui, X., Page, G. P., and Sabripour, M. (2006). Microarray data analysis: from disarray to consolidation and consensus. *Nature Review Genetics* **7**, 55-65.

Andrews, D. F., and Mallows, C. L. (1974). Scale Mixtures of Normal Distributions. *Journal of the Royal Statistical Society. Series B (Methodological)* **36**, 99-102.

Beissinger, T. M., Rosa, G. J., Kaeppler, S. M., Gianola, D., and de Leon, N. (2015). Defining window-boundaries for genomic analyses using smoothing spline techniques. *Genetics Selection Evolution* **47**, 30.

Bello, N. M., Steibel, J. P., and Tempelman, R. J. (2010). Hierarchical Bayesian modeling of random and residual variance-covariance matrices in bivariate mixed effects models. *Biometrical Journal* **52**, 297-313.

Blasco, A., and Toro, M. A. (2014). A short critical history of the application of genomics to animal breeding. *Livestock Science* **166**, 4-9.

Brown, E. N., and Kass, R. E. (2009). What Is Statistics? *American Statistician* **63**, 105-123.

Calus, M. (2014). Right-hand-side updating for fast computing of genomic breeding values. *Genetics Selection Evolution* **46**, 1-11.

Calus, M. P. L., Schrooten, C., and Veerkamp, R. F. (2014). Genomic prediction of breeding values using previously estimated SNP variances. *Genetics Selection Evolution* **46**, 13.

Calus, M. P. L., and Veerkamp, R. F. (2007). Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. *Journal of Animal Breeding and Genetics* **124**, 362-368.

Chen, C., and Tempelman, R. J. (2015). An integrated approach to empirical Bayesian whole genome prediction modeling *Journal of Agricultural Biological and Environmental Statistics* **(submitted for this special issue)**.

Christensen, O. F., and Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* **42**, 2.

Cuyabano, B. C., Su, G., and Lund, M. S. (2014). Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population. *BMC Genomics* **15**, 1171.

Daetwyler, H. D., Capitan, A., Pausch, H., *et al.* (2014). Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics* **46**, 858-865.

de Koning, D.-J., and McIntyre, L. (2012). Setting the Standard: A Special Focus on Genomic Selection in GENETICS and G3. *Genetics* **190**, 1151–1152.

de los Campos, G., Gianola, D., and Allison, D. B. (2010a). Predicting genetic predisposition in humans: the promise of whole-genome markers. *Nature Reviews Genetics* **11**, 880–886.

de Los Campos, G., Gianola, D., and Rosa, G. J. (2009a). Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. *Journal of Animal Science* **87**, 1883–1887.

De los Campos, G., Gianola, D., Rosa, G. J., Weigel, K. A., and Crossa, J. (2010b). Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genetics Research* **92**, 295–308.

de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., and Calus, M. P. L. (2013). Whole Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* **193**, 327–345.

de los Campos, G., Naya, H., Gianola, D., *et al.* (2009b). Predicting Quantitative Traits With Regression Models for Dense Molecular Markers and Pedigree. *Genetics* **182**, 375–385.

de los Campos, G., Veturi, Y., Vazquez, A. I., Lehermeier, C., and Perez-Rodriguez, P. (2015). Incorpating genetic heterogeneity in whole-genome regressions using interactions. *Journal of Agricultural Biological and Environmental Statistics* **(submitted for this special issue)**.

Desta, Z. A., and Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. *Trends in Plant Science* **19**, 592-601.

Druet, T., Macleod, I. M., and Hayes, B. J. (2014). Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity* **112**, 39-47.

Duchemin, S. I., Colombani, C., Legarra, A.*, et al.* (2012). Genomic selection in the French Lacaune dairy sheep breed. *Journal of Dairy Science* **95**, 2723-2733.

Eisen, E. J. (2008). Can we rescue an endangered species? *Journal of Animal Breeding and Genetics* **125**, 1-2.

Endelman, J. B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome Journal* **4**, 250.

Erbe, M., Hayes, B. J., Matukumalli, L. K.*, et al.* (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science* **95**, 4114-4129.

Fernando, R., and Garrick, D. (2013). Bayesian Methods Applied to GWAS. In *Genome-Wide Association Studies and Genomic Prediction*, C. Gondro, J. van der Werf, and B. Hayes (eds), 237-274: Humana Press.

Fernando, R. L. (2009). GenSel: User Manual for a Portfolio of Genomic Selection Related Analyses. Iowa State University, Ames, IA, U.S.A.

Fernando, R. L., Dekkers, J. C. M., and Garrick, D. J. (2014). A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genetics Selection Evolution* **46**, 13.

Fragomeni, B. O., Lourenco, D. A. L., Tsuruta, S.*, et al.* (2015). Hot topic: Use of genomic recursions in single-step genomic best linear unbiased predictor (BLUP) with a large number of genotypes. *Journal of Dairy Science* **98**, 4090-4094.

Fragomeni, B. O., Misztal, I., Lourenco, D. L., Aguilar, I., Okimoto, R., and Muir, W. M. (2014). Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. *Frontiers in Genetics* **5**, 332.

Garrick, D., Dekkers, J., and Fernando, R. (2014). The evolution of methodologies for genomic prediction. *Livestock Science* **166**, 10-18.

Gauldron-Duarte, J. L., Cantet, R. J. C., Bates, R. O., Ernst, C. W., Raney, N. E., and Steibel, J. P. (2014). Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. *BMC Bioinformatics* **15**, 11.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association* **85**, 972-985.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*: Taylor & Francis.

Gelman, A., Hill, J., and Yajima, M. (2012). Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness* **5**, 189-211.

George, E. I., and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**, 881 - 889.

Gianola, D. (2000). Statistics in animal breeding. *Journal of the American Statistical Association* **95**, 296-299.

——— (2013). Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics* **194**, 573-596.

Gianola, D., de los Campos, G., Hill, W. G., Manfredi, E., and Fernando, R. (2009). Additive Genetic Variability and the Bayesian Alphabet. *Genetics* **183**, 347–363.

Gianola, D., and Fernando, R. L. (1986). Bayesian methods in animal breeding theory. *Journal of Animal Science* **63**, 217-244.

Gianola, D., Foulley, J. L., and Fernando, R. (1986). Prediction of breeding values when variances are not known. *Genetics, Selection, Evolution* **18**, 485-498.

Gianola, D., and Rosa, G. J. (2015). One Hundred Years of Statistical Developments in Animal Breeding. *Annual Review of Animal Biosciences* **3**, 19-56.

Gilmour, A. R., Thompson, R., and Cullis, B. R. (1995). Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models. *Biometrics* **51**, 1440-1450.

Goddard, M. E., Hayes, B. J., and Meuwissen, T. H. (2010). Genomic selection in livestock populations. *Genetics Research* **92**, 413-421.

Gondro, C., Van der Werf, J., and Hayes, B. (2013). *Genome-wide association studies and genomic prediction*: Springer: Humana Press, New York.

González-Recio, O., Rosa, G. J. M., and Gianola, D. (2014). Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livestock Science* **166**, 217-231.

Grattapaglia, D., and Resende, M. D. V. (2010). Genomic selection in forest tree breeding. *Tree Genetics & Genomes* **7**, 241-255.

Habier, D., Fernando, R. L., and Dekkers, J. C. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* **177**, 2389 - 2397.

Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* **12**, 186.

Halsey, L. G., Curran-Everett, D., Vowler, S. L., and Drummond, G. B. (2015). The fickle P value generates irreproducible results. *Nature Methods* **12**, 179-185.

Hayes, B. (2013). Overview of Statistical Methods for Genome-Wide Association Studies (GWAS). In *Genome-Wide Association Studies and Genomic Prediction*, C. Gondro, J. van der Werf, and B. Hayes (eds), 149-169: Humana Press.

Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics* **28**, 476-490.

Henderson, C. R. (1976). A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. *Biometrics* **32**, 69-83.

Henderson, C. R., Kempthorne, O., Searle, S. R., and Krosigk, C. M. v. (1959). The Estimation of Environmental and Genetic Trends from Records Subject to Culling. *Biometrics* **15**, 192–218.

Heslot, N., Yang, H.-P., Sorrells, M. E., and Jannink, J.-L. (2012). Genomic Selection in Plant Breeding: A Comparison of Models. *Crop Science* **52**, 146.

Hill, W. G. (2014). Applications of population genetics to animal breeding, from wright, fisher and lush to genomic prediction. *Genetics* **196**, 1-16.

Howard, R., Carriquiry, A. L., and Beavis, W. D. (2014). Parametric and Nonparametric Statistical Methods for Genomic Selection of Traits with Additive and Epistatic Genetic Architectures. *G3: Genes/Genomes/Genetics* **4**, 1027-1046.

Irizarry, R. (2015). Correlation is not a measure of reproducibility. In http://simplystatistics.org/2015/08/12/correlation-is-not-a-measure-of-reproducibility/.

Jannink, J. L., Lorenz, A. J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics* **9**, 166-177.

Janss, L. (2015). Experiences with bioinformatics. *Journal of Animal Science* **93**, 198.

Jarquin, D., Crossa, J., Lacaze, X.*, et al.* (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics* **127**, 595-607.

Johnson, D. L., and Thompson, R. (1995). Restricted Maximum Likelihood Estimation of Variance Components for Univariate Animal Models Using Sparse Matrix Techniques and Average Information. *Journal of Dairy Science* **78**, 449-456.

Jonas, E., and de Koning, D. J. (2013). Does genomic selection have a future in plant breeding? *Trends in Biotechnology* **31**, 497-504.

Jorasch, P. (2005). Intellectual Property Rights in the Field of Molecular Marker Analysis. In *Molecular Marker Systems in Plant Breeding and Crop Improvement*, H. Lörz, and G. Wenzel (eds), 433-471: Springer Berlin Heidelberg.

Kadarmideen, H. N. (2014). Genomics to systems biology in animal and veterinary sciences: Progress, lessons and opportunities. *Livestock Science* **166**, 232-248.

Käll, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008). Posterior Error Probabilities and False Discovery Rates: Two Sides of the Same Coin. *Journal of Proteome Research* **7**, 40-44.

Kang, H. M., Zaitlen, N. A., Wade, C. M., *et al.* (2008). Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics* **178**, 1709-1723.

Kärkkäinen, H. P., and Sillanpää, M. J. (2012). Back to Basics for Bayesian Model Building in Genomic Selection. *Genetics* **191**, 969-987.

Kizilkaya, K., and Tempelman, R. J. (2005). A general approach to mixed effects modeling of residual variances in generalized linear mixed models. *Genetics Selection Evolution* **37**, 31-56.

Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* **92**, 4656-4663.

Legarra, A., Christensen, O. F., Aguilar, I., and Misztal, I. (2014). Single Step, a general approach for genomic selection. *Livestock Science* **166**, 54-65.

Legarra, A., and Misztal, I. (2008). Technical Note: Computing Strategies in Genome-Wide Selection. *Journal of Dairy Science* **91**, 360-366.

Lehermeier, C., Wimmer, V., Albrecht, T., *et al.* (2013). Sensitivity to prior specification in Bayesian genome-based prediction models. *Statistical Applications in Genetics and Molecular Biology* **12**, 375-391.

Li, A., and Meyre, D. (2012). Challenges in reproducibility of genetic association studies: lessons learned from the obesity field. *International Journal of Obesity* **2012**, 1-9.

Liu, J. S. (1994). The Collapsed Gibbs Sampler in Bayesian Computations with Applications to a Gene-Regulation Problem. *Journal of the American Statistical Association* **89**, 958-966.

Liu, Z., Goddard, M. E., Reinhardt, F., and Reents, R. (2014). A single-step genomic model with direct estimation of marker effects. *Journal of Dairy Science* **97**, 5833-5850.

Logsdon, B. A., Hoffman, G. E., and Mezey, J. G. (2010). A variational Bayes algorithm for fast and accurate multiple locus genome-wide association analysis. *BMC Bioinformatics* **11**, 58.

Louis, T. A. (1982). Finding the Observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **44**, 226-233.

Meng, X.-L., and Rubin, D. B. (1991). Using EM to Obtain Asymptotic Variance-Covariance Matrices: The SEM Algorithm. *Journal of the American Statistical Association* **86**, 899-909.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819-1829.

Meyer, K. (1989). Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genetics Selection Evolution* **21**, 317-340.

Misztal, I. (1990). Restricted Maximum Likelihood Estimation of Variance Components in Animal Model Using Sparse Matrix Inversion and a Supercomputer. *Journal of Dairy Science* **73**, 163-172.

——— (2007). Shortage of quantitative geneticists in animal breeding. *Journal of Animal Breeding and Genetics* **124**, 255-256.

Misztal, I., and Perez-Enciso, M. (1993). Sparse Matrix Inversion for Restricted Maximum Likelihood Estimation of Variance Components by Expectation-Maximization. *Journal of Dairy Science* **76**, 1479-1483.

Morota, G., Abdollahi-Arpanahi, R., Kranis, A., and Gianola, D. (2014). Genome-enabled prediction of quantitative traits in chickens using genomic annotation. *BMC Genomics* **15**, 109.

Morota, G., and Gianola, D. (2014). Kernel-based whole-genome prediction of complex traits: a review. *Frontiers in Genetics* **5**.

Nadaf, J., Riggio, V., Yu, T.-P., and Pong-Wong, R. (2012). Effect of the prior distribution of SNP effects on the estimation of total breeding value. *BMC Proceedings* **6**, S6.

Nejati-Javaremi, A., Smith, C., and Gibson, J. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of Animal Science* **75**, 1738-1745.

O'Hara, R. B., and Sillanpää, M. J. (2009). A Review of Bayesian Variable Selection Methods: What, How and Which. *Bayesian Analysis* **4**, 85-118.

Perez-Elizalde, S., Cuevas, J., Perez-Rodriguez, P., and Crossa, J. (2015). Selection of the bandwidth parameter in a Bayesian kernel regression model for genomic-enabled prediction. *Journal of Agricultural Biological and Environmental Statistics* **(submitted for this special issue)**.

Pérez-Enciso, M. (1995). Use of the uncertain relationship matrix to compute effective population size. *Journal of Animal Breeding and Genetics* **112**, 327-332.

Pérez, P., and de los Campos, G. (2014). Genome-Wide Regression & Prediction with the BGLR Statistical Package. *Genetics*.

Pinheiro, J. C., Liu, C., and Wu, Y. N. (2001). Efficient Algorithms for Robust Estimation in Linear Mixed-Effects Models Using the MultivariatetDistribution. *Journal of Computational and Graphical Statistics* **10**, 249-276.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnostics and output analysis for MCMC. *R News* **6**, 7-11.

Pregitzer, C. C., Bailey, J. K., and Schweitzer, J. A. (2013). Genetic by environment interactions affect plant–soil linkages. *Ecology and Evolution* **3**, 2322-2333.

Robinson, G. K. (1991). That BLUP is a Good Thing: The Estimation of Random Effects. 15-32.

Rockova, V., and George, E. I. (2014a). EMVS: The EM Approach to Bayesian Variable Selection. *Journal of the American Statistical Association* **109**, 828-846.

——— (2014b). Negotiating Multicollinearity with Spike-and-Slab Priors. *Metron* **72**, 217-229.

Rosa, G. J. M., Padovani, C. R., and Gianola, D. (2003). Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. *Biometrical Journal* **45**, 573-590.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B-Statistical Methodology* **71**, 319-392.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2009). Semiparametric regression during 2003–2007. *Electronic Journal of Statistics* **3**, 1193-1256.

Schaeffer, L. R., and Kennedy, B. W. (1986). Computing Strategies for Solving Mixed Model Equations. *Journal of Dairy Science* **69**, 575-579.

Schoen, C.-C., Wimmer, V., and Lehermeier, C. (2014). Efficiency of Variable Selection in Genome-Wide Prediction for Traits of Different Genetic Architecture. In *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production, ASAS, August 17-22, 2014, Vancouver, BC, Canada*: Asas.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. New York: John Wiley and Sons.

Shariati, M., and Sorensen, D. (2008). Efficiency of alternative MCMC strategies illustrated using the reaction norm model. *Journal of Animal Breeding and Genetics* **125**, 176-186.

Shariati, M. M., Korsgaard, I. R., and Sorensen, D. (2009). Identifiability of parameters and behaviour of MCMC chains: a case study using the reaction norm model. *Journal of Animal Breeding and Genetics* **126**, 92-102.

Sorensen, D., Andersen, S., Gianola, D., and Korsgaard, I. (1995). Bayesian inference in threshold models using Gibbs sampling. *Genetics Selection Evolution* **27**, 229-249.

Sorensen, D. A., Gianola, D., and Korsgaard, I. R. (1998). Bayesian mixed-effects model analysis of a censored normal distribution with animal breeding applications. *Acta Agriculturae Scandinavica A—Animal Sciences* **48**, 222-229.

Spiegelhalter, D. J., Best, N. G., Carlin, B. R., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **64**, 583-616.

Storey, J. D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* **100**, 9440-9445.

Strandén, I., and Garrick, D. J. (2009). Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *Journal of Dairy Science* **92**, 2971-2975.

Strandén, I., and Lidauer, M. (1999). Solving Large Mixed Linear Models Using Preconditioned Conjugate Gradient Iteration. *Journal of Dairy Science* **82**, 2779-2787.

Tanner, M. A., and Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association* **82**, 528-540.

Taylor, J. F. (2014). Implementation and accuracy of genomic selection. *Aquaculture* **420-421**, S8-S14.

Technow, F., Messina, C. D., Totir, L. R., and Cooper, M. (2015). Integrating Crop Growth Models with Whole Genome Prediction through Approximate Bayesian Computation. *PLoS ONE* **10**, e0130855.

Tierney, L., and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**, 82-86.

Usai, M. G., Goddard, M. E., and Hayes, B. J. (2010). LASSO with cross-validation for genomic selection. *Genetics Research* **91**, 427.

Van Dyk, D. A., and Meng, X.-L. (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics* **10**, 1-50.

Vattikuti, S., Lee, J. J., Chang, C. C., Hsu, S. D., and Chow, C. C. (2014). Applying compressed sensing to genome-wide association studies. *Gigascience* **3**, 10.

Verbyla, K. L., Hayes, B. J., Bowman, P. J., and Goddard, M. E. (2009). Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. *Genetics Research* **91**, 307 - 311.

Wang, H., Misztal, I., Aguilar, I., Legarra, A., and Muir, W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetics Research* **94**, 73-83.

Wang, Y. (1998). Smoothing Spline Models with Correlated Random Errors. *Journal of the American Statistical Association* **93**, 341-348.

Wiggans, G. R., Misztal, I., and Van Vleck, L. D. (1988). Implementation of an Animal Model for Genetic Evaluation of Dairy Cattle in the United States. *Journal of Dairy Science* **71**, 54-69.

Wiggans, G. R., VanRaden, P. M., and Cooper, T. A. (2011). The genomic evaluation system in the United States: Past, present, future. *Journal of Dairy Science* **94**, 3202-3211.

——— (2015). Technical note: Rapid calculation of genomic evaluations for new animals. *Journal of Dairy Science* **98**, 2039-2042.

Wimmer, V., Albrecht, T., Auinger, H.-J., and Schön, C.-C. (2012). synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* **28**, 2086-2087.

Wimmer, V., Lehermeier, C., Albrecht, T., Auinger, H.-J., Wang, Y., and Schön, C.-C. (2013). Genome-Wide Prediction of Traits with Different Genetic Architecture Through Efficient Variable Selection. *Genetics* **195**, 573-587.

Wu, X.-L., Sun, C., Beissinger, T. M., *et al.* (2012). Parallel Markov chain Monte Carlo-bridging the gap to high-performance Bayesian computation in animal breeding and genetics. *Genetics Selection Evolution* **44**, 29.

Yang, W., Chen, C., and Tempelman, R. J. (2015). Improving the computational efficiency of fully Bayes inference and assessing the effect of misspecification of hyperparameters in whole-genome prediction models. *Genetics Selection Evolution* **47**, 13.

Yang, W., and Tempelman, R. J. (2012). A Bayesian Antedependence Model for Whole Genome Prediction. *Genetics* **190**, 1491-1501.

Yi, N. J., and Xu, S. H. (2008). Bayesian LASSO for quantitative trait loci mapping. *Genetics* **179**, 1045-1055.

Yi, N. J., and Banerjee, S. (2009). Hierarchical Generalized Linear Models for Multiple Quantitative Trait Locus Mapping. *Genetics* **181**, 1101-1113.

Yi, N., Xu, S., Lou, X. Y., and Mallick, H. (2014). Multiple comparisons in genetic association studies: a hierarchical modeling approach. *Statistical Applications in Genetics and Molecular Biology* **13**, 35-48.