DE GRUYTER DOI 10.1515/sagmb-2012-0042 — Statistical Applications in Genetics and Molecular Biology 2013; 12(3): 375–391

Christina Lehermeier, Valentin Wimmer, Theresa Albrecht[a], Hans-Jürgen Auinger, Daniel Gianola, Volker J. Schmid and Chris-Carolin Schön*

Sensitivity to prior specification in Bayesian genome-based prediction models

Abstract: Different statistical models have been proposed for maximizing prediction accuracy in genome-based prediction of breeding values in plant and animal breeding. However, little is known about the sensitivity of these models with respect to prior and hyperparameter specification, because comparisons of prediction performance are mainly based on a single set of hyperparameters. In this study, we focused on Bayesian prediction methods using a standard linear regression model with marker covariates coding additive effects at a large number of marker loci. By comparing different hyperparameter settings, we investigated the sensitivity of four methods frequently used in genome-based prediction (Bayesian Ridge, Bayesian Lasso, BayesA and BayesB) to specification of the prior distribution of marker effects. We used datasets simulated according to a typical maize breeding program differing in the number of markers and the number of simulated quantitative trait loci affecting the trait. Furthermore, we used an experimental maize dataset, comprising 698 doubled haploid lines, each genotyped with 56110 single nucleotide polymorphism markers and phenotyped as testcrosses for the two quantitative traits grain dry matter yield and grain dry matter content. The predictive ability of the different models was assessed by five-fold cross-validation. The extent of Bayesian learning was quantified by calculation of the Hellinger distance between the prior and posterior densities of marker effects. Our results indicate that similar predictive abilities can be achieved with all methods, but with BayesA and BayesB hyperparameter settings had a stronger effect on prediction performance than with the other two methods. Prediction performance of BayesA and BayesB suffered substantially from a non-optimal choice of hyperparameters.

Keywords: genome-based prediction; genomic selection; Bayesian learning; shrinkage prior; plant breeding.

[a]Current address: Institute for Crop Production and Plant Breeding, Bavarian State Research Center for Agriculture, Am Gereuth 6, 85354 Freising, Germany

***Corresponding author: Chris-Carolin Schön,** Plant Breeding, Technische Universität München, Emil-Ramann-Straße 4, 85354 Freising, Germany, e-mail: chris.schoen@tum.de

Christina Lehermeier, Valentin Wimmer, Theresa Albrecht and Hans-Jürgen Auinger: Plant Breeding, Technische Universität München, Emil-Ramann-Straße 4, 85354 Freising, Germany

Daniel Gianola: Department of Animal Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA; and Institute for Advanced Study, Technische Universität München, Lichtenbergstraße 2a, 85748 Garching, Germany

Volker J. Schmid: Department of Statistics, Ludwig-Maximilians-Universität München, Ludwigstraße 33, 80539 München, Germany

Introduction

The term genomic selection was introduced by Meuwissen et al. (2001) to denote the prediction of breeding values using molecular markers covering the whole genome. Genomic selection is principally based on two steps. In the first step, effects are estimated for all markers simultaneously based on a large phenotyped and genotyped training population, using some appropriate genome-based prediction model. In the second step, genomic estimated breeding values (GEBVs) of selection candidates are calculated based on their genomic information by using marker effects estimated in the first step. Candidates are selected based on their GEBVs and not on their phenotypes (Jannink et al., 2010). First used in animal breeding [e.g., Schaeffer (2006)], genome-based prediction of breeding values has been investigated in plant populations recently (Heffner et al., 2009; Albrecht et al., 2011; Heslot et al., 2012; Riedelsheimer et al., 2012), yielding encouraging results for its application in breeding.

Many statistical models that can handle the problem of having more marker effects (p) than phenotypes (n) and that differ with respect to the assumption of marker effect distribution, have been proposed for genome-based prediction. In a plant breeding context Crossa et al. (2010) and Heslot et al. (2012) compared the performance of various models proposed for genome-based prediction using several datasets and crops. However, the use of Bayesian models requires specifying prior distributions with hyperparameters, and comparisons have been made based on a single set of hyperparameters. Limited efforts have been made towards evaluating impact of the choice of model hyperparameters, although this is likely to influence the performance of the model. In our study, we investigate the sensitivity of the performance of Bayesian genome-based prediction models with respect to hyperparameter choice. We focused on quantitative traits affected by a large number of loci and an advanced cycle breeding population of maize with strong linkage disequilibrium (LD) between markers. We employed the models known as Bayesian Ridge (Kneib et al., 2011), Bayesian Lasso (Park and Casella, 2008), BayesA and BayesB (Meuwissen et al., 2001), as these are frequently used for genome-based prediction. The models differ in the prior specification for the marker effects with hyperparameters controlling the amount of shrinkage of the effects. With an increasing number of observations n, the influence of the prior on the posterior distribution should vanish, which is known as the concept of Bayesian learning (Sorensen and Gianola, 2002). However, when the number of parameters p exceeds n ($p \gg n$), which is typical in genome-based prediction, the prior setting will always matter. Thus, to obtain good predictive abilities an adequate choice of hyperparameters is necessary to prevent both over- and underfitting. This study was motivated by theoretical results of Gianola et al. (2009), who pointed out that BayesA and BayesB are strongly influenced by hyperparameter choice due to a limited extent of Bayesian learning. Our objectives were to evaluate the sensitivity of these four models to the choice of hyperparameters. The goal was to identify a model with both good predictive ability and robustness over a wide range of hyperparameters in practice. In order to quantify the Bayesian learning ability of the models, we used the Hellinger distance as distance measure between prior and posterior densities of marker effects.

The superiority of specific Bayesian models has been shown in many studies using simulated data with prior specifications according to simulation parameters. In this study we evaluated the four models based on two datasets simulated according to a typical maize breeding program, differing in the number of markers and in the number of simulated quantitative trait loci (QTL). Results were compared to those from an experimental maize dataset, comprising 11646 polymorphic high-quality single nucleotide polymorphism markers (SNPs) and 698 doubled haploid (DH) lines, phenotyped for two quantitative traits (grain dry matter yield and content). The predictive ability of the models was assessed by five-fold cross-validation.

Datasets

Simulated datasets

We simulated two maize datasets (maizeA & B), starting with a base population similar to that described in Meuwissen et al. (2001). One thousand individuals were generated with diploid genomes having a length of 16 Morgan (M) assuming 10 equally sized chromosomes with two haplotypes each. The coding in the haplotype sequence was 0 or 1 with equal probability to simulate biallelic markers. A mutation-drift equilibrium was expected to be reached after 1000 generations of random mating. We simulated different marker densities by assigning 800 and 64,000 equidistant markers per chromosome to datasets maizeA and B, respectively. A random subset of 1000 (500) markers was initially selected as QTL in dataset maizeA (B). Each QTL was assigned an additive effect of equal magnitude on the phenotype.

From generation 1000 we randomly selected 10 individuals to be the founders of a breeding program. This mimics the typically small effective population size and large LD of maize breeding populations. Homozygous recombinant inbred lines (n=1250) were derived from crossings of the founders followed by six generations

of selfing. For line i (i=1, 2, …, 1250) three environmental errors (ϵ_{il}, l=1, …, 3) were sampled from the same normal distribution $N(0, \sigma^2_\epsilon)$ to represent three phenotypic observations per line i. The variance σ^2_ε was chosen to be three times the variance of the true breeding values (TBV), with TBV being the sum of the QTL effects times the number of alleles per individual. The variance of the environmental errors σ^2_ϵ produced a repeatability parameter of 0.25. The sum of the individual TBV and the mean error $(1/3\sum_{l=1}^{3}\epsilon_{il})$ formed the phenotypic value y_i for each line i: y_i=TBV$_i$+$1/3\sum_{l=1}^{3}\epsilon_{il}$. This procedure mimics a typical approach in plant breeding to observe mean phenotypic values from several observations per line. In each of six breeding cycles, the 25 lines with the largest phenotypic values were selected as parents and recombined to produce 1250 inbred lines in the next cycle. The final datasets maizeA and B consisted of the genotypes and phenotypes from the seventh cycle. Due to selection and drift a subset of markers and QTL was monomorphic in the final datasets maizeA and B. A summary of the number of polymorphic markers, the number of polymorphic QTL and the number of inbred lines n in the datasets is given in Table 1. Additionally the trait heritabilities h^2, calculated from the squared Pearson correlation between phenotypic values and TBV, are listed for the datasets maizeA and B. The dataset maizeA is publicly available within the R package `synbreedData` (Wimmer et al., 2012).

Experimental dataset

The experimental dataset comprised 752 fully homozygous DH lines of maize (*Zea mays* L.), derived from 122 different crosses with 27 inbred lines and nine single crosses as parents. The number of DH lines derived from each cross ranged from 1 to 63 with an average of six DH lines per cross. The lines were phenotyped in testcrosses with a single-cross tester in four different European locations in 2010 for the traits grain dry matter yield (GDY; dt/ha) and grain dry matter content (GDC; %). For the sake of simplicity, hereinafter testcross values are denoted as breeding values. Each location comprised eight sets each replicated twice. Sets were arranged in a 10×10 lattice design containing 94 entries and six checks. Outliers were identified and removed based on maximum deviant residuals of a full stage model according to Grubbs (1950). In each location, entries were adjusted for set, replication and block effects and, in a second stage, adjusted means were calculated over locations. The generalized heritability was $\hat{h}^2_{GDY}$=0.74 for GDY and $\hat{h}^2_{GDC}$=0.94 for GDC, estimated according to Cullis et al. (2006). For 698 of the 752 DH lines tested in the field marker data were generated. Lines were genotyped with the MaizeSNP50 BeadChip from Illumina® (Ganal et al., 2011) containing 56 110 SNPs. SNPs with a GTScore <0.7, a call rate <0.9, a minor allele frequency (MAF) <0.01 and identical SNPs were excluded from the analysis. Missing values were imputed using the function `codeGeno()` from R package `synbreed` (Wimmer et al., 2012) with option "beagleAfterFamily." Here, missing values of markers that were monomorphic within one family were imputed with the fixed allele, and missing values in families segregating for the marker were imputed based on information of flanking markers using the software BEAGLE (Browning and Browning, 2009). Finally 11,646 high quality SNPs were used for further analysis. We measured pair-wise LD between SNPs as squared correlation (r^2) between allelic states according to Hill and

Table 1 Summary of the simulated and experimental datasets for grain dry matter yield (GDY) and grain dry matter content (GDC). Represented are the number of polymorphic single nucleotide polymorphism markers (no. SNP), number of polymorphic quantitative trait loci in the simulated datasets (no. QTL), number of lines (n) and the trait heritability (h^2). U represents an unknown number of QTL.

	No. SNP	No. QTL	n	h^2
Simulated datasets				
maizeA	1117	500	1250	0.46
maizeB	7425	369	1250	0.64
Experimental dataset				
GDY	11,646	U	698	0.74
GDC	11,646	U	698	0.94

Robertson (1968). The decay of pair-wise LD between SNPs on the same chromosome with increasing physical distance is depicted in Figure 1. We observed a high long-range LD between SNPs, which is typical for advanced cycle breeding populations in maize (Ching et al., 2002).

Genome-based prediction model and Bayesian regularization

We used the standard linear model for genome-based prediction of breeding values:

$$\boldsymbol{y}=\mathbf{1}_n\beta_0+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{y}$ is the n-dimensional vector of phenotypes, $\mathbf{1}_n$ an n-dimensional vector of ones, β_0 is the intercept, $\mathbf{X}$ is an $n\times p$ incidence matrix with elements x_{ij} (i=1, …, n; j=1, …, p) representing the genotype score of line i at marker j, defined by the number of copies of the minor allele at marker j. For fully homozygous lines such as DH lines, the entries can only take values 0 or 2. The vector $\boldsymbol{\beta}$ is a p-dimensional vector of marker effects and $\boldsymbol{\epsilon}$ is the n-dimensional vector of residuals. The residuals were assumed to be independent and normally distributed with mean 0 and equal variance σ^2_ϵ, so $\boldsymbol{\epsilon}\sim \mathrm{N}\left(0, \mathbf{I}\sigma^2_\epsilon\right)$ with $\mathbf{I}$ being the $n\times n$ identity matrix. This leads to the following data distribution:

$$f\left(\boldsymbol{y}\,|\,\beta_0,\boldsymbol{\beta},\sigma^2_\epsilon\right)=\prod_{i=1}^{n}\mathrm{N}\left(y_i\,|\,\beta_0+\sum_{j=1}^{p}x_{ij}\beta_j,\sigma^2_\epsilon\right), \tag{2}$$

where the notation $\mathrm{N}(y_i|\mu,\sigma^2)$ denotes the conditional density of the normal distribution of y_i given the mean μ and variance σ^2 (de los Campos et al., 2013). In the Bayesian framework, prior distributions have to be specified for all unknown parameters. By using Bayes' Theorem the posterior distribution of the parameters is assessed by combining the prior and the data distribution (Bernardo and Smith, 2002). A frequently used approach is to use conjugate prior distributions, in order to obtain conditional posterior distributions from the same family of distributions as the prior with updated parameters. For genome-based prediction, with $p\gg n$, shrinkage priors are required, which shrink effects towards zero. Meuwissen et al. (2001) proposed the Bayesian models BayesA and BayesB for genome-based prediction, which are frequently used. We compared

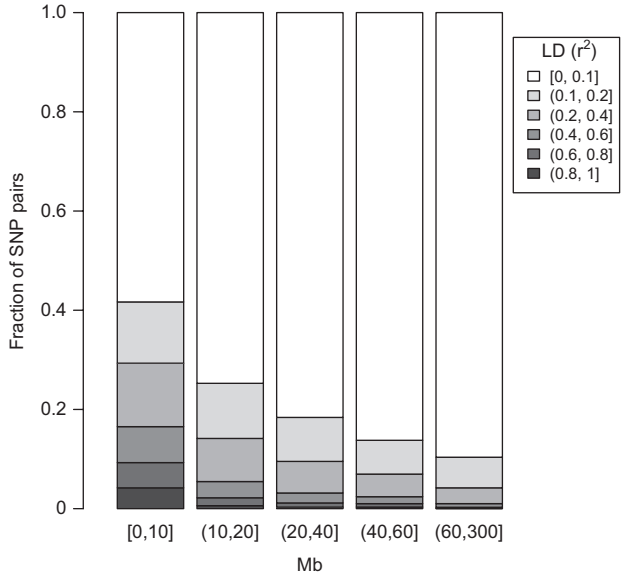

Figure 1 Linkage disequilibrium in the experimental dataset, measured as squared correlation (r^2) of pair-wise marker combinations on the same chromosome against physical distance between them.

the predictive abilities and the sensitivity of these models to the choice of hyperparameters with well known shrinkage models, such as the Bayesian Ridge (Kneib et al., 2011) and the Bayesian Lasso (Park and Casella, 2008). The four models differ with respect to their shrinkage properties. Whereas Bayesian Lasso, BayesA and BayesB induce marker-specific shrinkage, Bayesian Ridge uses the same shrinkage parameter for all markers.

Bayesian Ridge

In the Bayesian Ridge model (Kneib et al., 2011) a Gaussian prior with mean 0 and common variance σ^2_β is assigned to all marker effects β_j (j=1, …, p):

$$f\left(\boldsymbol{\beta}|\sigma^2_\beta\right)=\prod_{j=1}^{p}\mathrm{N}\left(\beta_j|0,\sigma^2_\beta\right). \tag{3}$$

With a smaller variance σ^2_β the prior density of marker effects is more concentrated around 0 and so effects are shrunken to a greater extent than when the variance is larger. The same prior variance σ^2_β is assigned to all marker effects, so the shrinkage is marker-homogeneous. Each of the variance parameters σ^2_β and σ^2_ϵ is assigned a scaled inverse-χ^2 prior, $f\left(\sigma^2_\epsilon\right)=\chi^{-2}\left(df_\epsilon, S_\epsilon\right)$ and $f\left(\sigma^2_\beta\right)=\chi^{-2}\left(df_\beta, S_\beta\right)$, respectively, with df_ϵ and df_β being the degrees of freedom, and S_ϵ and S_β the scale parameters of the corresponding scaled inverse-χ^2 prior distribution. A flat prior distribution is assigned to the intercept β_0, with $f(\beta_0) \propto$ const. A Gibbs sampler is employed for posterior inference.

A question remains: how hyperparameters of prior distributions $f\left(\sigma^2_\epsilon\right)$ and $f\left(\sigma^2_\beta\right)$ should be chosen? According to Pérez et al. (2010), parameters can be chosen on prior beliefs that suggest a certain partition of the phenotypic variance into residual and genotypic variance, respectively. Here, we focus on the hyperparameter setting of $f\left(\sigma^2_\beta\right)$. In principle, the approach for σ^2_β can also be adapted for σ^2_ϵ. The mean and mode of the scaled inverse-χ^2 distribution are mean $\left(\sigma^2_\beta\right)=df_\beta S_\beta/\left(df_\beta-2\right)$ and mode $\left(\sigma^2_\beta\right)=df_\beta S_\beta/\left(df_\beta+2\right)$, respectively. The larger the scale S_β, for fixed df_β, the larger the mean and mode of the distribution. To get an a priori guess for σ^2_β, a prior expectation about the genotypic variance is used. According to classical quantitative genetics, the phenotypic value of an individual y_i can be expressed as the sum of its genotypic value and an environmental deviation (Falconer and Mackay, 1996). In the simulated datasets the genotypic value of line i is the true breeding value TBV_i. In a population the total phenotypic variance is given as the sum of the genotypic and the environmental variance components. Assuming model (1) holds it follows that $\mathrm{TBV}_i=\beta_0+\sum_{j=1}^{p}x_{ij}\beta_j$. Thus, with marker scores x_{ij} assumed to be fixed the genotypic variance of an individual i in the Bayesian Ridge model is derived as

$$\sigma^2_{G_i}=\mathrm{Var}\left(\sum_{j=1}^{p}x_{ij}\beta_j\right)=\sum_{j=1}^{p}x^2_{ij}\sigma^2_\beta,\ i=1,\ \ldots,\ n. \tag{4}$$

Following Pérez et al. (2010), substituting the individual marker score x_{ij} by the average marker score over all individuals, $\bar{x}_j=1/n\sum_{i=1}^{n}x_{ij}\ (j=1,\ \ldots,\ p)$, approximates $\sigma^2_G=\sum_{j=1}^{p}\bar{x}^2_j\sigma^2_\beta$. Thus, the variance of effects σ^2_β is assessed as:

$$\sigma^2_\beta=\frac{\sigma^2_G}{\sum_{j=1}^{p}\bar{x}^2_j}. \tag{5}$$

Setting df_β to a small value, e.g., df_β=4, to get a relatively flat prior distribution and applying the a priori expected genotypic variance for σ^2_G from (5), the scale S_β can be calculated by using the prior mode as:

$$S_\beta=\frac{\sigma^2_\beta\left(df_\beta+2\right)}{df_\beta}=\frac{\sigma^2_G\left(df_\beta+2\right)}{\sum_{j=1}^{p}\bar{x}^2_j\ df_\beta}. \tag{6}$$

Bayesian Lasso

The second model we explored for genome-based prediction is the Bayesian Lasso (Park and Casella, 2008; de los Campos et al., 2009). Here, conditional Gaussian priors with mean 0 are assigned to the marker effects. In the Bayesian Lasso the variance $\sigma_\epsilon^2\tau_j^2$ of a marker effect is peculiar to marker locus j (j=1, …, p) so that the joint prior distribution, given σ_ϵ^2 and τ_j^2, is

$$f\left(\boldsymbol{\beta}|\boldsymbol{\tau}^2, \sigma_\epsilon^2\right)=\prod_{j=1}^{p}\mathrm{N}\left(\beta_j|0, \sigma_\epsilon^2\tau_j^2\right). \tag{7}$$

Thus, contrary to marker-homogeneous shrinkage like in Bayesian Ridge, Bayesian Lasso has the potential of providing marker-heterogeneous shrinkage. The extent of shrinkage depends on τ_j^2, with smaller values of τ_j^2 producing more shrinkage of effect β_j. For each of the variance parameters $\boldsymbol{\tau}^2=\left\{\tau_j^2\right\}$ the same exponential prior distribution is used with $f\left(\boldsymbol{\tau}^2|\lambda\right)=\prod_{j=1}^{p}\mathrm{Exp}\left(\tau_j^2|\lambda^2\right)$, under independence assumptions. Here, we denote with $\mathrm{Exp}\left(\tau_j^2|\lambda^2\right)$ the conditional density of the exponential distribution of τ_j^2 given the rate parameter λ^2. The shrinkage parameter λ can then be either set to a fixed value, or a prior distribution can be assigned to λ. Park and Casella (2008) suggested using a Gamma prior for λ^2, that is, $f(\lambda^2)$=Gamma(r, δ). For the residual variance σ_ϵ^2 and for the intercept the same prior distributions as in Bayesian Ridge are used. As in Bayesian Ridge, a Gibbs sampler is employed in Bayesian Lasso. The fully conditional posterior distributions of the unknown parameters are given in de los Campos et al. (2009). To find an appropriate λ, the prior variance of the effects $\sigma_{\beta_j}^2=2\sigma_\epsilon^2\lambda^{-2}$ is used, as derived by Pérez et al. (2010). By using the genotypic variance σ_G^2, an "optimal" a priori value of λ can be arrived at through the relationship

$$\hat{\lambda}=\sqrt{2\frac{\sigma_\epsilon^2}{\sigma_G^2}\sum_{j=1}^{p}\overline{x}_j^2}=\sqrt{2\frac{1-h^2}{h^2}\sum_{j=1}^{p}\overline{x}_j^2}, \tag{8}$$

with $h^2=\sigma_G^2/\left(\sigma_G^2+\sigma_\epsilon^2\right)$ being the trait heritability (Falconer and Mackay, 1996).

BayesA

Meuwissen et al. (2001) suggested the model BayesA for inferring marker effects in genome-based prediction. Here each of the effects β_j (j=1, …, p) is assigned a Gaussian prior with mean 0 and variance $\sigma_{\beta_j}^2$, leading to the joint prior distribution

$$f\left(\boldsymbol{\beta}|\sigma_{\beta_1}^2, ..., \sigma_{\beta_p}^2\right)=\prod_{j=1}^{p}\mathrm{N}\left(\beta_j|0, \sigma_{\beta_j}^2\right). \tag{9}$$

As opposed to Bayesian Ridge each marker effect β_j is assigned a marker-specific variance $\sigma_{\beta_j}^2$. This leads to marker-heterogeneous shrinkage as in Bayesian Lasso. However, differently from the Bayesian Lasso where the prior distribution of the variances of marker effects is exponential, in BayesA the same scaled inverse-χ^2 prior $f\left(\sigma_{\beta_j}^2\right)=\chi^{-2}\left(df_\beta, S_\beta\right)$ is used. For the intercept β_0 a flat prior distribution like in Bayesian Ridge is used, and for the residual variance σ_ϵ^2 the same scaled inverse-χ^2 prior as in Meuwissen et al. (2001). A Gibbs sampler is employed for sampling from the joint posterior distribution and the required fully conditional posterior distributions are given in Meuwissen et al. (2001). For finding appropriate values for the hyperparameters df_β and S_β the same guidelines as in Bayesian Ridge can be used, as the prior distributions are also scaled inverse-χ^2.

BayesB

The model BayesB was suggested by Meuwissen et al. (2001) as an extension to BayesA. Due to the assumption that many SNPs are expected to have no effect on the trait, in difference to BayesA the variances of marker effects $\sigma_{\beta_j}^2$ are assigned the following mixture distribution as prior:

$$\sigma^2_{\beta_j} = 0 \quad \text{with probability } \pi,$$
$$\sigma^2_{\beta_j} \sim \chi^{-2}\left(df_\beta, S_\beta\right) \quad \text{with probability } (1-\pi).$$

Generally $\sigma^2_{\beta_j}=0$ implies, that the effect β_j is set to a constant value. Here the conditional prior distributions of the effects β_j are normal as in BayesA, with mean 0 and variance $\sigma^2_{\beta_j}$. Thus, if the variance $\sigma^2_{\beta_j}$ is set to zero, the effect β_j is set to zero in turn. A Gibbs sampler with a Metropolis-Hastings (MH) step is used, as described in Meuwissen et al. (2001), for posterior inference. The MH step is needed here, because no closed form of the fully conditional posterior distribution of β_j can be derived due to the mixture distribution. In principle, the same guidelines as in Bayesian Ridge can be used to choose the hyperparameters df_β and S_β, but the probability π of setting $\sigma^2_{\beta_j}$ to zero must be taken into account. Considering this in equation (5), $\sigma^2_{\beta_j}$ can be denoted as

$$\sigma^2_{\beta_j} = \frac{\sigma^2_G}{(1-\pi)\sum_{j=1}^{p}\overline{x}^2_j}. \tag{10}$$

By using the mode of the scaled inverse-χ^2 distribution a scale parameter S_β can be arrived at by writing

$$S_\beta = \frac{\sigma^2_G\left(df_\beta+2\right)}{df_\beta(1-\pi)\sum_{j=1}^{p}\overline{x}^2_j}, \tag{11}$$

according to equation (6).

Model comparison

Hellinger distance

The Hellinger distance $H(f, g)$ (Le Cam, 1986), which is also used in Roos and Held (2011) to evaluate the sensitivity of models with respect to the choice of prior distributions, measures the distance between two densities f and g:

$$H\left(f, g\right) = \sqrt{\frac{1}{2}\int_{-\infty}^{\infty}\left(\sqrt{f(u)}-\sqrt{g(u)}\right)^2 \mathrm{d}u}. \tag{12}$$

$H(f, g)$ is a symmetric measure, which takes its maximum value of 1, if the density f assigns probability 0 to every data point to which g assigns a positive value, and vice versa. The minimum value of H is 0, if $f=g$. In Gianola et al. (2009) it has been pointed out that the results of BayesA and BayesB are highly influenced by the choice of hyperparameters that are assigned to the prior distributions. This lack of Bayesian learning prevents the posterior distribution to move far away from the prior distribution, at least for some parameters such as $\sigma^2_{\beta_j}$ in BayesA and BayesB. To evaluate this for marker effects, we calculated the Hellinger distances between marginal prior and posterior densities of the marker effects numerically. The marginal posterior density was estimated from the posterior MCMC samples using kernel density estimation with a Gaussian kernel and bandwidth chosen based on Silverman's rule of thumb (Silverman, 1986). To approximate the integral in H the trapezoidal rule was used (Atkinson, 1989).

Effective number of parameters

To obtain an estimate of the effective number of parameters in the different models, we calculated the p_D values, as suggested by Spiegelhalter et al. (2002). The p_D statistic is calculated as $p_D = 1/L\sum_{l=1}^{L} D\left(\boldsymbol{y}, \boldsymbol{\theta}^l\right) - D\left(\boldsymbol{y}, \hat{\boldsymbol{\theta}}(\boldsymbol{y})\right)$, with $\boldsymbol{\theta}^l$ being the MCMC sample from the l-th iteration ($l=1, \ldots, L$) of $\boldsymbol{\theta}$ $\left(\boldsymbol{\theta}=\left(\beta_0, \boldsymbol{\beta}, \sigma^2_\epsilon\right)\right)$, $\boldsymbol{y}$ being the data vector,

$\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ being the posterior mean of $\boldsymbol{\theta}$, and $D(\boldsymbol{y}, \boldsymbol{\theta})=-2 \log f(\boldsymbol{y}|\boldsymbol{\theta})$ is the residual deviance. With an informative prior, the effective number of parameters p_D is generally smaller than the total number of parameters in $\boldsymbol{\theta}$ (Gelman et al., 2004).

Cross-validation

We used five-fold cross-validation (CV) to assess the predictive ability of the different models. For five-fold CV, each dataset was randomly divided into five subsets, with four subsets forming the training set and the fifth subset forming the test set. The training set was used to infer the marker effects with the different models, and the test set was used to predict the breeding values based on the genotypic data. Each of the five subsets formed the test set once. The correlation between estimated breeding values $\hat{\boldsymbol{y}}^{(k)}$ (for $k=1, \ldots, 5$) and observed phenotypic values $\boldsymbol{y}^{(k)}$ $\text{cor}(\hat{\boldsymbol{y}}^{(k)}, \boldsymbol{y}^{(k)})$ yields an estimate of the predictive ability of a model in fold k, where $\hat{\boldsymbol{y}}^{(k)}$ is obtained from $\hat{\boldsymbol{y}}^{(k)}=\mathbf{1}_n\hat{\beta}_0+\mathbf{X}^{(k)}\hat{\boldsymbol{\beta}}$, where the matrix $\mathbf{X}^{(k)}$ includes the marker genotypes of the test set. The effects $\hat{\boldsymbol{\beta}}$ were estimated posterior means based on the training set. For the simulated datasets true breeding values (TBVs) are available and the accuracy can be calculated directly. No TBVs are available for experimental data, so here the accuracy is approximated according to Dekkers (2007) by dividing the predictive correlation by the square-root of the heritability h^2: $\text{cor}\left(\hat{\boldsymbol{y}}^{(k)}, \text{TBV}^{(k)}\right) \approx \text{cor}\left(\hat{\boldsymbol{y}}^{(k)}, \boldsymbol{y}^{(k)}\right) \Big/ \sqrt{h^2}$.

Model overview and computation

For every combination of model and dataset, a scenario with hyperparameter setting as denoted earlier as "optimal" was computed. Therefore, for Bayesian Ridge, BayesA and BayesB, we set the degrees of freedom of the scaled inverse-χ^2 distribution of σ^2_β to $df_\beta=4$, for BayesB π was set to 0.8. The "optimal" scale parameter S_β was then calculated using equations (6) and (11), respectively. The required genotypic variance σ^2_G was calculated from the product of the trait heritability h^2 and the phenotypic variance σ^2_P. For Bayesian Lasso, an "optimal" λ was calculated by using equation (8). To judge the influence of the hyperparameters on predictive ability, scenarios with altered hyperparameters were computed for every model. Table 2 gives an overview of the hyperparameter settings for the different datasets. The models Bayesian Ridge, Bayesian Lasso, BayesA and BayesB are abbreviated as BR, BL, BA and BB, respectively, and different numbers indicate different model scenarios. In the case of BR1, BA1 and BB1 the scale parameter S_β was set to the value from the formulas given previously. For the scenarios BR2, BA2 and BB2, the chosen scale parameter was ten times the "optimal" value, and for BR3, BA3 and BB3 the "optimal" value was divided by ten, respectively. In the BL1 scenarios the parameter λ was set to the "optimal" value, in BL2 the "optimal" value was divided by $\sqrt{10}$ and in BL3 multiplied by $\sqrt{10}$. This range of hyperparameters was chosen to make prior distributions comparable across scenarios for the four different models. In BL4-6 a Gamma prior was assigned to λ^2, with shape parameter $\delta=0.52$ and rate parameter r chosen according to a density of λ with a mode corresponding to the λ values in BL1-3.

Computation was done using the software package R (R Development Core Team, 2012). For Bayesian Ridge and Bayesian Lasso the function `BLR()` of the R package `BLR` (de los Campos and Pérez, 2012) was used. For computation of BayesA and BayesB we implemented our own algorithm within R. We used 13,000 iterations for the MCMC algorithms of the different models, with the first 3000 iterations discarded as burn in. Only every 10th sample was retained for storage reasons. Hence, 1000 MCMC samples were used for the calculation of posterior means and densities. The convergence of the Markov chains was checked by visualizing the sampling paths and by Geweke's diagnostic of the R package `coda` (Plummer et al., 2006).

Table 2 Parameter setting for all models in the different datasets. For Bayesian Ridge, BayesA and BayesB the value of S_β varied, with df_β=4 in all scenarios. In BayesB parameter π was set to 0.8. In scenarios BL1-3 different values for λ were chosen, and in BL4-6 the parameter r of the Gamma prior for λ^2 varied, with δ=0.52 in all scenarios.

Model	maizeA	maizeB	GDY	GDC
Bayesian Ridge (S_β)				
BR1	0.15	0.013	0.012	0.00049
BR2	1.45	0.129	0.123	0.00486
BR3	0.01	0.001	0.001	0.00005
Bayesian Lasso with fixed λ (λ)				
BL1	30	52	44	17
BL2	9.5	16.4	13.9	5.4
BL3	95	164	139	54
Bayesian Lasso with random λ (r)				
BL4	2×10^{-5}	7×10^{-6}	1×10^{-5}	7×10^{-5}
BL5	2×10^{-4}	7×10^{-5}	1×10^{-4}	7×10^{-4}
BL6	2×10^{-6}	7×10^{-7}	1×10^{-6}	7×10^{-6}
BayesA (S_β)				
BA1	0.15	0.013	0.012	0.00049
BA2	1.45	0.129	0.123	0.00486
BA3	0.01	0.001	0.001	0.00005
BayesB (S_β)				
BB1	0.73	0.064	0.062	0.0024
BB2	7.26	0.645	0.615	0.0243
BB3	0.07	0.006	0.006	0.0002

Simulation results

Full datasets and Hellinger distance

The results from the full (not cross-validated) simulated datasets are shown in Table 3. For all models the correlation between phenotypic values $\boldsymbol{y}$ and estimated breeding values $\hat{\boldsymbol{y}}$ and the correlation between TBVs and $\hat{\boldsymbol{y}}$ is given. Taking the correlation between the true and estimated breeding value cor(TBV, $\hat{\boldsymbol{y}}$) as a reference, we interpreted an increase of cor($\boldsymbol{y}$, $\hat{\boldsymbol{y}}$) accompanied by a decrease of cor(TBV, $\hat{\boldsymbol{y}}$) as a sign of model overfitting. If both parameters decreased we interpreted this to be the result of model underfitting. Furthermore, we report the effective number of parameters in the models (p_D). For Bayesian Lasso models, we also report the posterior mean of λ.

The fixed λ value in BL1-3 had a great impact on the estimation. Bayesian Ridge and Bayesian Lasso scenarios with random λ (BL4-6) showed similar correlations and p_D values in both simulated maize datasets. The correlation between TBV and $\hat{\boldsymbol{y}}$ was highest in these models. In BL3 with a large λ value, effects were shrunken strongly, which is reflected in the small p_D value and the low correlation between $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$, which indicates underfitting. Here the correlation with TBVs is also low. In BL2 with small λ, the p_D statistic and cor($\boldsymbol{y}$, $\hat{\boldsymbol{y}}$) were high but cor(TBV, $\hat{\boldsymbol{y}}$) was decreasing, indicating overfitting on the training data, especially in the dataset maizeB, where the number of SNPs strongly exceeded the number of observations ($p \gg n$). It should be noted that the posterior mean of λ in the Bayesian Lasso models with random λ exceeded the "optimal" λ calculated a priori in all datasets. Hence, fixing λ to an "optimal" a priori value might distort inference relative to the situations where λ is inferred from the data. We conclude that assigning hyperpriors to λ should be preferred to assigning a fixed value, as it seemingly allows for Bayesian learning, providing that the posterior distribution of λ is sharp. Similar to the Bayesian Lasso models with different fixed λ, the scale parameter S_β in BayesA and BayesB had a strong impact on the estimation procedure. In scenarios BA2 and BB2 a higher scale S_β was assigned to the inverse-χ^2 prior of σ^2_β than in the other BayesA and BayesB

Table 3 Results from the full (not cross-validated) simulated datasets. The correlations between phenotypic values $\boldsymbol{y}$ and estimated breeding values $\hat{\boldsymbol{y}}$ as well as the correlations between true breeding values (TBV) and $\hat{\boldsymbol{y}}$, the effective number of parameters (p_D) and the posterior means of λ from Bayesian Lasso models are given.

			maizeA					maizeB	
Model	**cor(y, $\hat{y}$)**	**cor(TBV, $\hat{y}$)**	p_D	$\hat{\lambda}$	**Model**	**cor(y, $\hat{y}$)**	**cor(TBV, $\hat{y}$)**	p_D	$\hat{\lambda}$
Bayesian Ridge					Bayesian Ridge				
BR1	0.72	0.85	196		BR1	0.86	0.89	357	
BR2	0.73	0.86	212		BR2	0.86	0.89	366	
BR3	0.72	0.86	193		BR3	0.86	0.89	354	
Bayesian Lasso					Bayesian Lasso				
BL1	0.74	0.86	232		BL1	0.87	0.89	389	
BL2	0.82	0.84	450		BL2	0.96	0.87	803	
BL3	0.62	0.80	80		BL3	0.73	0.83	132	
BL4	0.73	0.86	198	37.21	BL4	0.86	0.89	350	59.65
BL5	0.73	0.86	199	36.83	BL5	0.86	0.89	353	58.99
BL6	0.73	0.86	201	36.73	BL6	0.86	0.89	351	59.56
BayesA					BayesA				
BA1	0.78	0.86	322		BA1	0.92	0.89	577	
BA2	0.86	0.82	568		BA2	1.00	0.82	1115	
BA3	0.66	0.82	111		BA3	0.77	0.85	178	
BayesB					BayesB				
BB1	0.74	0.85	232		BB1	0.90	0.89	508	
BB2	0.74	0.84	249		BB2	0.97	0.86	870	
BB3	0.67	0.83	136		BB3	0.79	0.87	210	

scenarios, leading to less shrinkage of marker effects. This is reflected in the high effective number of parameters in these models (p_D), and in an overfitting of the data, indicated by the high value of cor($\boldsymbol{y}$, $\hat{\boldsymbol{y}}$) but lower value of cor(TBV, $\hat{\boldsymbol{y}}$) compared to the other scenarios. However, the overfitting is less pronounced in BayesB than in BayesA. Vice versa in BA3 and BB3, the small scale value S_β led to high shrinkage of marker effects and did not result in overfitting.

For maizeA, we calculated the Hellinger distance $H(f, g)$ between marginal prior (f) and approximated marginal posterior density (g) of the marker effects. In Figure 2, the distribution of the Hellinger distance for the different model scenarios is visualized by boxplots for all 1117 SNPs. For Bayesian Ridge and Bayesian

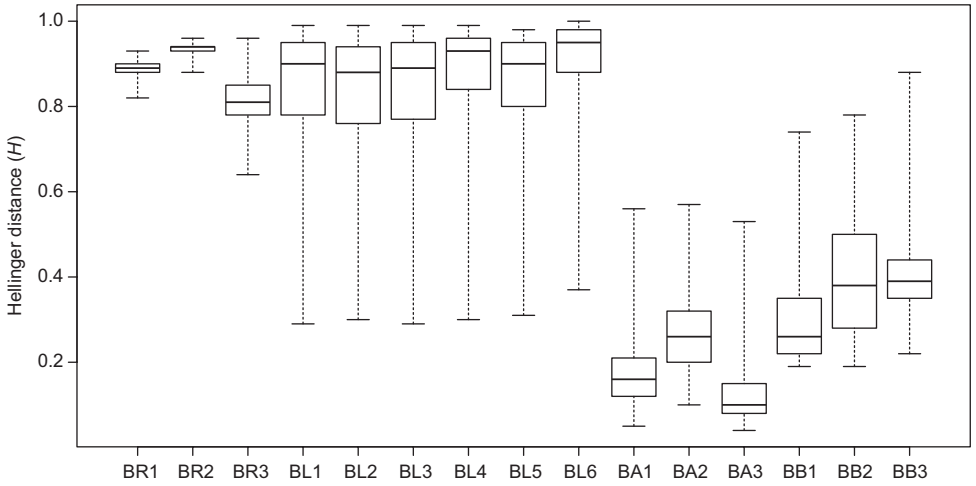

Figure 2 Distribution of Hellinger distance (H) between the marginal prior and posterior densities of marker effects $\boldsymbol{\beta}$ from different model scenarios, calculated with simulated dataset maizeA. Each boxplot displays the distribution of Hellinger distances of the 1117 marker effects out of each model.

Lasso models H was considerably higher than for BayesA and BayesB, indicating that there was less Bayesian learning in BayesA and BayesB, as there is more similarity between the marginal prior and posterior densities of the effects β_j. The Bayesian Ridge scenarios showed less variability with respect to Hellinger distances of the marker effects than the other models. We conjecture that the marker-heterogeneous shrinkage in Bayesian Lasso, BayesA and BayesB caused the higher variability in these models.

Cross-validation

We used cross-validation to assess the predictive ability on simulated data. Figure 3 shows the predictive ability and the accuracy from five-fold CV, for all model scenarios in the simulated datasets maizeA and maizeB. The mean p_D value from each CV-fold is depicted as a measure of model complexity. In the dataset maizeA, BL1 yielded the highest predictive ability and accuracy with mean 0.53 and 0.79, respectively. The mean predictive ability of all Bayesian Ridge models and Bayesian Lasso models with random λ was 0.52, and the mean accuracy was approximately 0.79. The Bayesian Lasso models with fixed non-"optimal" λ (BL2-3) showed lower predictive abilities and accuracies. The mean predictive ability of BayesA and BayesB models did not exceed 0.52, the highest accuracy with these models was 0.78. Thus, no BayesA or BayesB scenario exceeded the performance of Bayesian Ridge and Bayesian Lasso models with random λ, as well as BL1, in terms of predictive ability. However, the predictive ability and accuracy of BayesA models with a

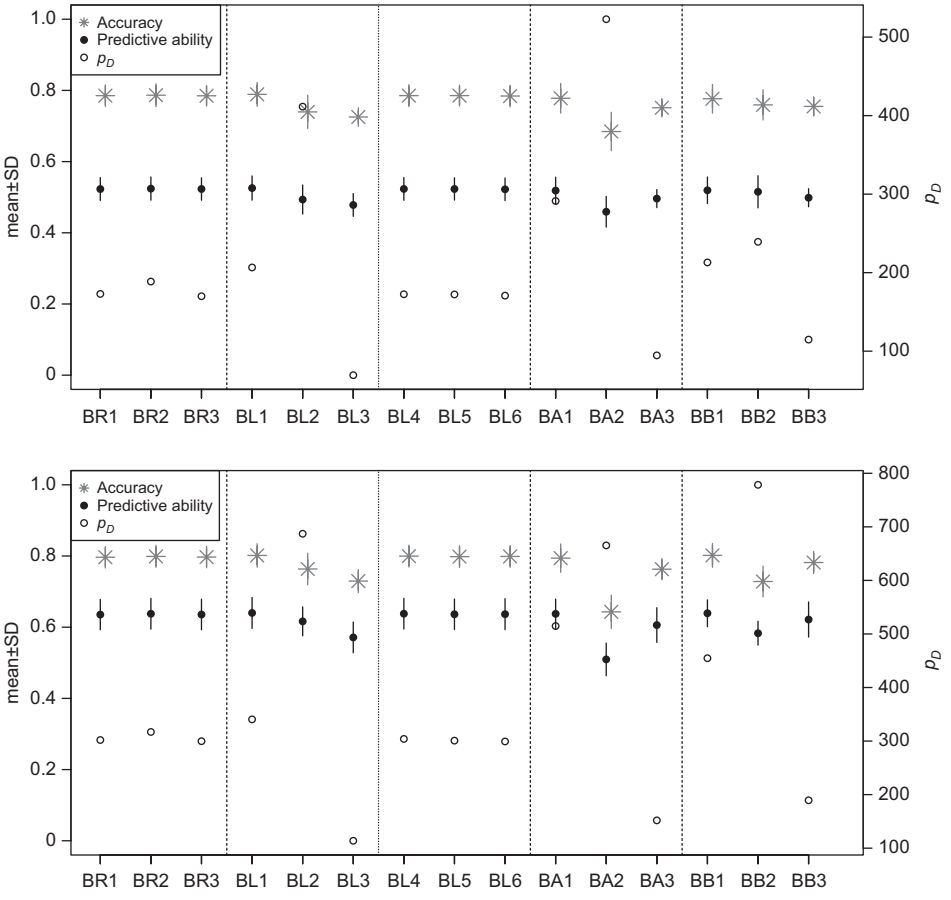

Figure 3 Predictive abilities and accuracies from five-fold cross-validation, calculated with simulated datasets maizeA (upper panel) and maizeB (lower panel). The black dots show the mean predictive ability from five-fold cross-validation for the different scenarios, the gray stars show the mean accuracy. The lines define the interval [mean−SD; mean+SD]. The open circles denote the mean effective number of parameters p_D in the models.

non-"optimal" scale (BA2-3) was significantly reduced when compared to BA1 (tested with Student's paired t-test, α=0.05). In BayesB models the scenarios with non-"optimal" scale also showed a reduced predictive ability and accuracy, but not as strong as in BayesA models. The highest mean predictive ability in maizeB was 0.64 and the highest mean accuracy was 0.80. All Bayesian Ridge models, Bayesian Lasso models with random λ and BL1 yielded the same predictive ability and accuracy. BayesA and BayesB with "optimal" scale had the same predictive ability, but the predictive ability of BayesA and BayesB models with non-"optimal" S_β was significantly reduced. It is worth mentioning that the marker-heterogeneous shrinkage in Bayesian Lasso, BayesA and BayesB did not outperform the marker-homogeneous shrinkage of Bayesian Ridge. This did not happen in dataset maizeA nor in dataset maizeB, where the number of markers strongly exceeded the number of observations. The higher values of predictive ability and accuracy in the dataset maizeB than in maizeA are presumably due to higher heritability, fewer simulated QTL and more markers. As seen in the p_D values, both over- and underfitting led to reduced predictive abilities and accuracies.

Results: experimental dataset

Full dataset

The results from the full, not cross-validated, experimental dataset, for both phenotypic traits GDY and GDC, are shown in Table 4. Here the same parameters as for the simulated data are given, except for the correlation between estimated and true breeding values. In BL2, BA2 and BB2 cor($\boldsymbol{y}$, $\hat{\boldsymbol{y}}$)=1.00 indicates a strong overfitting, because of $p \gg n$ and little shrinkage of the effects. For both traits the posterior mean of λ of BL models with random λ (BL4-6) was higher than the calculated "optimal" value, which was already seen with the simulated datasets.

Table 4 Results from the experimental data. The correlations between phenotypic values $\boldsymbol{y}$ and estimated breeding values $\hat{\boldsymbol{y}}$, the effective number of parameters (p_D) and the posterior means of λ of Bayesian Lasso models are given.

	Grain dry matter yield (GDY)				Grain dry matter content (GDC)		
Model	**cor(y,$\hat{y}$)**	**p_D**	**$\hat{\lambda}$**	**Model**	**cor(y,$\hat{y}$)**	**p_D**	**$\hat{\lambda}$**
Bayesian Ridge				Bayesian Ridge			
BR1	0.86	208		BR1	0.97	377	
BR2	0.89	249		BR2	0.97	391	
BR3	0.86	203		BR3	0.97	364	
Bayesian Lasso				Bayesian Lasso			
BL1	0.94	358		BL1	0.99	561	
BL2	0.99	593		BL2	1.00	676	
BL3	0.81	136		BL3	0.96	315	
BL4	0.86	207	90.25	BL4	0.97	372	41.28
BL5	0.87	215	86.76	BL5	0.97	378	40.27
BL6	0.87	213	87.24	BL6	0.97	361	43.21
BayesA				BayesA			
BA1	0.98	510		BA1	1.00	595	
BA2	1.00	487		BA2	1.00	340	
BA3	0.83	162		BA3	0.96	310	
BayesB				BayesB			
BB1	0.97	469		BB1	1.00	576	
BB2	1.00	483		BB2	1.00	583	
BB3	0.84	180		BB3	0.95	281	

Cross-validation

CV results for the two experimental traits GDY and GDC are shown in Figure 4. For GDY the highest predictive ability was achieved with BL5 (mean≈0.65). All Bayesian Ridge models, all Bayesian Lasso models with random λ and BL1 had an average predictive ability of approximately 0.65. Predictive abilities of BayesA and BayesB models were lower but in BA3 and BB3 only slightly. The scenarios with S_β set to its calculated "optimal" value (BA1 and BB1) gave lower predictive abilities, than models where a 10 times smaller scale was used (BA3 and BB3). The effective number of parameters p_D in BA3 and BB3 were in a similar range as in Bayesian Ridge and Bayesian Lasso models with random λ. For the Bayesian Ridge and Bayesian Lasso models the estimated p_D values met our expectations. However, for BA2 and BB2 they were surprisingly low. Due to the reduced shrinkage, the p_D values should be higher than in BA1 and BB1. The p_D statistic is based on a number of assumptions (Spiegelhalter et al., 2002; Celeux et al., 2006), which are partially violated in BayesA and BayesB and thus the p_D statistic might not be a meaningful estimate of the effective number of parameters here. The predictive ability for GDC was higher than for GDY, most likely due to its higher heritability. Here, BL5 also yielded the highest predictive ability, with a mean of 0.83. The other Bayesian Lasso models with random λ and the Bayesian Ridge models yielded a similar predictive ability. As in the case of GDY, in GDC BayesA and BayesB scenarios also showed more variability in their predictive abilities. The scenarios BA3 and BB3 gave the highest predictive abilities among BayesA and BayesB models, respectively, which was not expected because "optimal" values for S_β had been chosen for BA1 and BB1. It seems that the

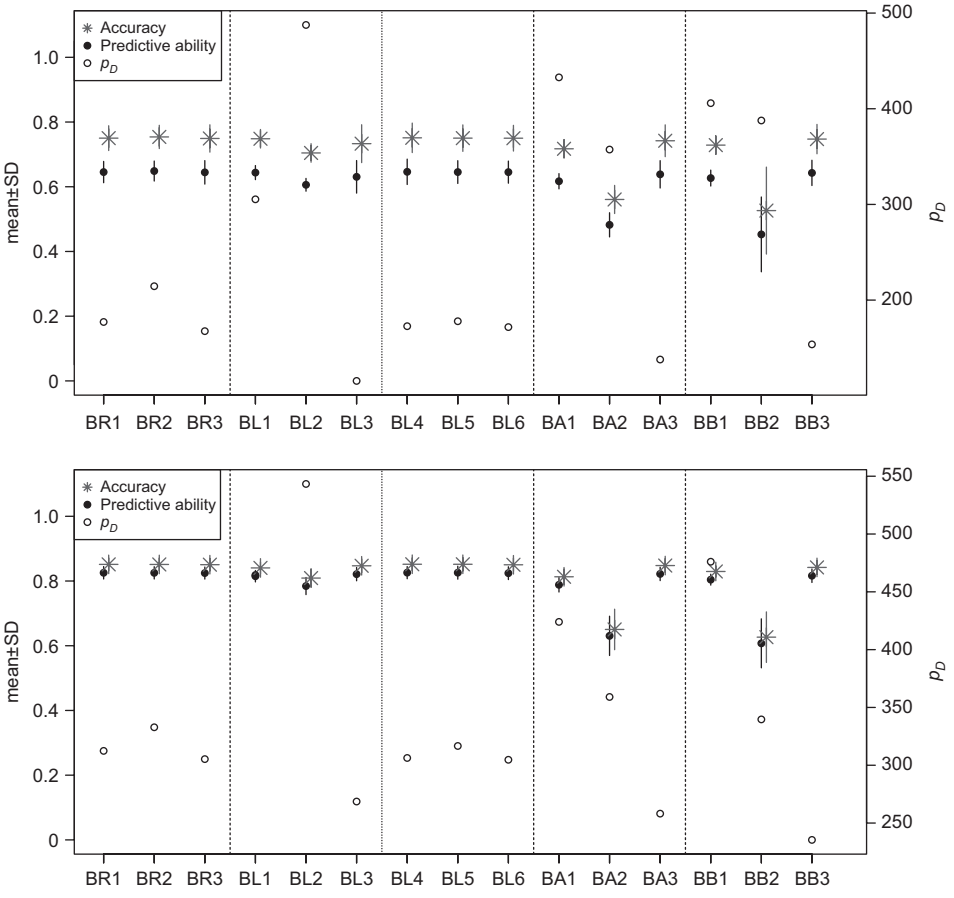

Figure 4 Predictive abilities and accuracies from five-fold cross-validation, calculated with experimental data, for traits grain dry matter yield (upper panel) and grain dry matter content (lower panel). The black dots show the mean predictive ability from five-fold cross-validation for the different scenarios, the gray stars show the mean accuracy. The lines define the interval [mean−SD; mean+SD]. The open circles denote the mean effective number of parameters p_D in the models.

proposed equation for finding an "optimal" scale S_β overestimated the "true" value. Neither for GDY, nor for GDC marker-heterogeneous shrinkage outperformed marker-homogeneous shrinkage, which corroborated the results from the simulated datasets.

Discussion

We examined the performance of four Bayesian models for genome-based prediction of breeding values in maize, as well as their sensitivity with respect to the choice of hyperparameters. The models were evaluated with simulated data resembling a commercial maize breeding program and with experimental maize data. A strong influence of prior parameters on the predictive ability was seen in BayesA and BayesB models, as well as in the Bayesian Lasso models with fixed λ. The variation of the scale parameter S_β in BayesA and BayesB had a strong impact on prediction. Choosing a too large scale S_β for the prior distribution of variance $\sigma^2_{\beta_j}$ led to an overfitting of the data, whereas a too small scale parameter led to underfitting, due to too much shrinkage of the effects. In both cases the predictive ability is considerably reduced.

The strong influence of the choice of hyperparameters indicates a lack of Bayesian learning ability in BayesA and BayesB, which was already pointed out by Gianola et al. (2009). We quantified the ability of Bayesian learning of the respective models by calculating the Hellinger distance between marginal prior and posterior densities of marker effects for the simulated dataset maizeA. A larger distance indicates that the posterior density moved away from the prior density, and that Bayesian learning has taken place. For BayesA and BayesB models we observed quite small distances, whereas in Bayesian Ridge and Bayesian Lasso the distance between prior and posterior density was much larger. We are aware that a small distance between prior and posterior density can also emerge if a perfect prior density is assigned, however, this happens with probability close to zero if prior knowledge is scant. In combination with a lower predictive ability and the fact, that all BayesA and BayesB scenarios yielded a small Hellinger distance between prior and posterior density, this is very unlikely to be the reason for the small distances found here. From the Hellinger distances it can be seen that Bayesian learning is smaller in BayesA and BayesB than in Bayesian Ridge and Bayesian Lasso and, hence, the influence of the choice of hyperparameters on prediction is large. In BayesA and BayesB the degrees of freedom of the fully conditional posterior distribution of $\sigma^2_{\beta_j}$ are df_β+1, and thus only one degree of freedom higher than the prior degrees of freedom, independently of the number of observations (n) or markers (p) in the model (Gianola et al., 2009). In contrast, in Bayesian Ridge the degrees of freedom increase with the number of markers in the model (p). In genomic datasets, Bayesian learning is limited due to the $p \gg n$ situation. In real life, as with next generation sequencing data, p will get even larger, and is expected to increase much more than n. Thus, models with a strong Bayesian learning ability such as the Bayesian Ridge and Bayesian Lasso seem useful.

If the prior parameter setting was appropriate, predictive abilities and accuracies of BayesA and BayesB were high and equal to those of Bayesian Ridge and Bayesian Lasso with random λ. However, finding "optimal" parameters is not straightforward. The proposed guidelines for finding an "optimal" scale S_β and λ, according to Pérez et al. (2010), did not always yield the best parameter setting in terms of predictive ability. Alternative formulas have been proposed for finding an "optimal" scale parameter S_β, e.g., in Habier et al. (2010, 2011). Both formulas are based on strong assumptions. The formula used in our study suggested by Pérez et al. (2010) as well as the formula according to Habier et al. (2010, 2011), assume independence of marker effects, which may be inadequate if strong LD among markers translates into a joint dependence of their effects.

In experimental data there is additional uncertainty in the proposed formulas for finding "optimal" parameters, because the variance components σ^2_G and σ^2_ϵ need to be estimated. In practical application of genome-based prediction, variance components can only be estimated based on the training dataset and not on phenotypic values of the test dataset, as these are unknown. Thus, there may be additional uncertainty

as the data distribution may change from training to test set. We have chosen hyperparameters based on the mode of the prior distributions. An option would be to choose hyperparameters based on the mean of the prior densities which would change the numerator of formulas (6) and (11) to $\sigma_G^2\left(df_\beta - 2\right)$. In the case of df_β=4 this would lead to an "optimal" scale S_β which would be 1/3 of the "optimal" S_β derived from the mode of the inverse-χ^2 distribution. Thus, our variation of hyperparameter settings is in the order of magnitude of uncertainty due to the choice of formula for hyperparameter calculation and estimation of variance components. An alternative to using an ad hoc formula would be to find hyperparameters iteratively via cross-validation, but this would have high computational costs. Hence, Bayesian models that are less sensitive with respect to the choice of hyperparameters are highly desirable.

The Bayesian Ridge model with marker-homogeneous shrinkage was in all datasets among the models with the highest predictive ability. Irrespective of the number of markers and observations, marker-specific shrinkage did not outperform marker-homogeneous shrinkage. The performance in the experimental dataset was similar as in the simulated datasets for both traits. One reason for the good performance of Bayesian Ridge may be the large number of QTL affecting the target traits. The simulated datasets comprised more than 300 segregating QTL and, also, the two quantitative traits in the experimental datasets have been found to be affected by many QTL (Schön et al., 2004). Furthermore, substantial long-range LD exists in maize breeding populations, which was also shown in our data. If there is strong LD, many SNPs are expected to be in LD with at least one QTL, and therefore to have non-zero effects. We conjecture that the large number of QTL and the strong correlation between markers are reasons for the superiority of the Bayesian Ridge model in terms of predictive abilities. The Bayesian Ridge model is similar to genome enabled best linear unbiased prediction (Piepho, 2009). These mixed models have also been shown in data with similar genetic architecture to perform as well as Bayesian models with a more complex prior setting, e.g., with marker-specific shrinkage (Heslot et al., 2012). The superiority of Bayesian models with marker-heterogeneous shrinkage has been shown mainly in simulation studies with a few simulated QTL (Meuwissen et al., 2001; Habier et al., 2007). However, such a genetic architecture seems to be unrealistic for truly quantitative traits. If interest lies mainly in the prediction of phenotypic traits, there may be little difference if a small effect is assigned to a group of highly correlated markers, or a larger effect is assigned to only one of them. On the other hand, marker-specific shrinkage models may be advantageous, if one is interested in specific marker effects. It is conjectured, that the performance of Ridge regression type models may change compared to that of marker-specific shrinkage models when marker coverage is more dense and LD is less pronounced (de los Campos et al., 2013). However, with denser marker coverage the ratio p/n will further increase and, thus, the influence of the prior density. Hence, the Bayesian Lasso may be advantageous over BayesA and BayesB due to its stronger Bayesian learning ability.

In our study, Bayesian Ridge and the Bayesian Lasso with assigning a hyperprior on λ were quite robust, whereas BayesA and BayesB showed a strong sensitivity with respect to the choice of hyperparameters. The ability of Bayesian learning is reduced in these models, as indicated by the Hellinger distance between prior and posterior densities of marker effects. No superiority of models with marker-specific shrinkage (Bayesian Lasso, BayesA, BayesB) was seen in our maize datasets, with a large number of QTL affecting the quantitative traits and a high long-range LD. Considering also the higher computing efforts of models with marker-specific shrinkage, we recommend Bayesian Ridge as a robust model for genome-based prediction, if one is mainly interested in the prediction of breeding values in datasets with a similar genetic architecture as those analyzed in this study.

Author notes: This research was funded by the German Federal Ministry of Education and Research (BMBF) within the AgroClustEr "Synbreed – Synergistic plant and animal breeding" (FKZ: 0315528A). This research was carried out with the support of the Technische Universität München – Institute for Advanced Study, funded by the German Excellence Initiative. We gratefully acknowledge the KWS SAAT AG for providing the experimental data.

References

Albrecht, T., V. Wimmer, H.-J. Auinger, M. Erbe, C. Knaak, M. Ouzunova, H. Simianer and C.-C. Schön (2011): "Genome-based prediction of testcross values in maize," Theor. Appl. Genet., 123, 339–350.

Atkinson, K. E. (1989): An introduction to numerical analysis, New York: Wiley.

Bernardo, J. M. and A. F. M. Smith (2002): Bayesian theory, Chichester: Wiley.

Browning, B. L. and S. R. Browning (2009): "A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals," Am. J. Hum. Genet., 84, 210–223.

Celeux, G., F. Forbes, C. P. Robert and D. M. Titterington (2006): "Deviance information criteria for missing data models," Bayesian Anal., 1, 651–673.

Ching, A., K. S. Caldwell, M. Jung, M. Dolan, O. S. Smith, S. Tingey, M. Morgante and A. J. Rafalski (2002): "SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines," BMC Genet., 3. DOI:10.1186/1471-2156-3-19.

Crossa, J., G. de los Campos, P. Pérez, D. Gianola, J. Burgueño, J. L. Araus, D. Makumbi, R. P. Singh, S. Dreisigacker, J. Yan, V. Arief, M. Banziger and H.-J. Braun (2010): "Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers," Genetics, 186, 713–724.

Cullis, B. R., A. B. Smith and N. E. Coombes (2006): "On the design of early generation variety trials with correlated data," J. Agric. Biol. Environ. Stat., 11, 381–393.

Dekkers, J. C. M. (2007): "Marker-assisted selection for commercial crossbred performance," J. Anim. Sci., 85, 2104–2114.

de los Campos, G. and P. Pérez (2012): BLR: Bayesian Linear Regression, URL http://CRAN.R-project.org/package=BLR, R package version 1.3. Accessed on November 30, 2012.

de los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, K. Weigel and J. M. Cotes (2009): "Predicting quantitative traits with regression models for dense molecular markers and pedigree," Genetics, 182, 375–385.

de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler and M. P. L. Calus (2013): "Whole genome regression and prediction methods applied to plant and animal breeding," Genetics, 193, 327–345.

Falconer, D. S. and T. F. C. Mackay (1996): Introduction to Quantitative Genetics, Essex: Longman.

Ganal, M. W., G. Durstewitz, A. Polley, A. Bérard, E. S. Buckler, A. Charcosset, J. D. Clarke, E.-M. Graner, M. Hansen, J. Joets, M.-C. L. Paslier, M. D. McMullen, P. Montalent, M. Rose, C.-C. Schön, Q. Sun, H. Walter, O. C. Martin and M. Falque (2011): "A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome," PLoS ONE, 6, 1–15.

Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin (2004): Bayesian data analysis, London: Chapman and Hall.

Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi and R. Fernando (2009): "Additive genetic variability and the Bayesian alphabet," Genetics, 138, 347–363.

Grubbs, F. E. (1950): "Sample criteria for testing outlying observations," Ann. Math. Stat., 21, 27–58.

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

Habier, D., J. Tetens, F.-R. Seefried, P. Lichtner and G. Thaller (2010): "The impact of genetic relationship information on genomic breeding values in German Holstein cattle," Genet. Sel. Evol., 42, 1–12.

Habier, D., R. L. Fernando, K. Kizilkaya and D. J. Garrick (2011): "Extension of the Bayesian alphabet for genomic selection," BMC Bioinformatics, 12, 1–24.

Heffner, E. L., M. E. Sorrells and J.-L. Jannink (2009): "Genomic selection for crop improvement," Crop Sci., 49, 1–12.

Heslot, N., H.-P. Yang, M. E. Sorrells and J.-L. Jannink (2012): "Genomic selection in plant breeding: A comparison of models," Crop. Sci., 52, 146–160.

Hill, W. G. and A. Robertson (1968): "Linkage disequilibrium in finite populations," Theor. Appl. Genet., 38, 226–231.

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

Kneib, T., S. Konrath and L. Fahrmeir (2011): "High dimensional structured additive regression models: Bayesian regularization, smoothing and predictive performance," J. Roy. Stat. Soc. C-App., 60, 51–70.

Le Cam, L. (1986): Asymptotic methods in statistical decision theory, New York: Springer-Verlag.

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

Park, T. and G. Casella (2008): "The Bayesian Lasso," J. Am. Stat. Assoc., 103, 681–686.

Pérez, P., G. de los Campos, J. Crossa and D. Gianola (2010): "Genomic-enabled prediction based on molecular markers and pedigree using the BLR package in R," Plant Genome, 3, 106–116.

Piepho, H.-P. (2009): "Ridge regression and extensions for genomewide selection in maize," Crop. Sci., 49, 1165–1176.

Plummer, M., N. Best, K. Cowles and K. Vines (2006): "CODA: convergence diagnosis and output analysis for MCMC," R News, 6, 7–11, URL http://CRAN.R-project.org/doc/Rnews/. Accessed on December 10, 2012.

Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, R. Sulpice, T. Altmann, M. Stitt, L. Willmitzer and A. E. Melchinger (2012): "Genomic and metabolic prediction of complex heterotic traits in hybrid maize," Nat. Genet., 44, 217–220.

Roos, M. and L. Held (2011): "Sensitivity analysis in Bayesian generalized linear mixed models for binary data," Bayesian Anal., 6, 259–278.

R Development Core Team (2012): R: A language and environment for statistical computing, R foundation for statistical computing, Vienna, Austria, URL http://www.R-project.org/, ISBN 3-900051-07-0. Accessed on November 30, 2012.

Schaeffer, L. R. (2006): "Strategy for applying genome-wide selection in dairy cattle," J. Anim. Breed. Genet., 123, 218–223.

Schön, C.-C., H. F. Utz, S. Groh, B. Truberg, S. Openshaw and A. E. Melchinger (2004): "Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits," Genetics, 167, 485–498.

Silverman, B. W. (1986): Density estimation, London: Chapman and Hall.

Sorensen, D. and D. Gianola (2002): Likelihood, Bayesian, and MCMC methods in quantitative genetics, New York: Springer.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin and A. van der Linde (2002): "Bayesian measures of model complexity and fit," J. R. Stat. Soc. B, 64, 583–639.

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