

Chapter 6

Bayesian Genomic Linear Regression



6.1 Bayes Theorem and Bayesian Linear Regression

Unlike classic statistical inference, which assumes that the parameter θ that defines the model is a fixed unknown quantity, in Bayesian inference it is considered as a random variable whose variation tries to represent the knowledge or ignorance about this, before the data points are collected (Box and Tiao 1992). The probability density function which describes such variation is known as the prior distribution and is an additional component in the specification of the complete model in a Bayesian framework.

Given a data set $\mathbf{y}_n = (y_1, \dots, y_n)$ whose distribution is assumed to be $f(\mathbf{y}|\theta)$, and a prior distribution for the parameter θ , $f(\theta)$, the Bayesian analysis uses the Bayes theorem to combine these two pieces of information to obtain the posterior distribution of the parameters, on which the inference is fully based (Christensen et al. 2011):

$$f(\theta|\mathbf{y}) = \frac{f(\mathbf{y}, \theta)}{f(\mathbf{y})} = \frac{f(\theta)f(\mathbf{y}|\theta)}{f(\mathbf{y})} \propto f(\theta)L(\theta; \mathbf{y}),$$

where $f(\mathbf{y}) = \int f(\mathbf{y}|\theta)f(\theta)d\theta = E_{\theta}[f(\mathbf{y}|\theta)]$ is the marginal distribution of θ . This conditional distribution describes what is known about θ after data is collected and can be thought of as the updated prior knowledge about θ with the information contained in the data, which is done through the likelihood function $L(\theta; \mathbf{y})$ (Box and Tiao 1992).

In general, because the posterior distribution doesn't always have a recognizable form and it is often not easy to simulate from this, numerical approximation methods are employed. Once a sample of the posterior distribution is obtained, estimation of a parameter is often found by averaging the sample values or averaging a function of the sample values when another quantity is of interest. For example, in genomic prediction with dense molecular markers, the main interest is to predict the trait of

interest of the non-phenotyped individuals that have only genotypic information, environment variables, or other information (covariates). In this situation, a convenient practice is to include the individuals to be predicted (\mathbf{y}_p) in the posterior distribution to be sampled.

Specifically, a standard Bayesian framework for a normal linear regression model (see Chap. 3)

$$Y = \beta_0 + \sum_{j=1}^p X_j \beta_j + \epsilon \quad (6.1)$$

with ϵ a random error with normal distribution with mean 0 and variance σ^2 , is fully specified by assuming the next non-informative prior distribution: $\boldsymbol{\beta}$ and $\log(\sigma)$ approximately independent and locally uniform.

$$f(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-2} \quad (6.2)$$

which is not a proper distribution because it does not integrate to 1 (Box and Tiao 1992; Gelman et al. 2013). However, when \mathbf{X} is of full column rank, the posterior distribution is a proper distribution and is given by

$$\begin{aligned} f(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) &\propto (\sigma^2)^{-\frac{p}{2}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] (\sigma^2)^{-1} \\ &\propto (\sigma^2)^{-\frac{p}{2}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} - 2\mathbf{y}^T \mathbf{X} \boldsymbol{\beta} + \mathbf{y}^T \mathbf{y})\right] (\sigma^2)^{-1} \\ &\propto (\sigma^2)^{-\frac{p+1}{2}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) - \frac{1}{2\sigma^2}(\mathbf{y}^T \mathbf{y} - \tilde{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \tilde{\boldsymbol{\beta}})\right] (\sigma^2)^{-1-(n-p-1)/2} \\ &\propto (\sigma^2)^{-\frac{p+1}{2}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})\right] (\sigma^2)^{-1-(n-p-1)/2} \exp\left[-\frac{1}{2\sigma^2} \mathbf{y}^T (\mathbf{I} - \mathbf{H}) \mathbf{y}\right] \\ &\propto (\sigma^2)^{-\frac{p+1}{2}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})\right] (\sigma^2)^{-1-\frac{n-p-1}{2}} \exp\left(-\frac{(n-p-1)\tilde{\sigma}^2}{2\sigma^2}\right), \end{aligned}$$

where $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$, $\tilde{\sigma}^2 = \frac{1}{n-p-1} \mathbf{y}^T (\mathbf{I} - \mathbf{H}) \mathbf{y}$, and $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$. From here the marginal posterior distribution of σ^2 is $\sigma^2 | \mathbf{y}, \mathbf{X} \sim IG\left((n-p-1)/2, \frac{(n-p-1)\tilde{\sigma}^2}{2}\right)$ with mean $\frac{(n-p-1)\tilde{\sigma}^2}{2} / [(n-p-1)/2] = \tilde{\sigma}^2$, and given σ^2 , the posterior conditional distribution of $\boldsymbol{\beta}$ is given by $\boldsymbol{\beta} | \sigma^2, \mathbf{y}, \mathbf{X} \sim N\left(\tilde{\boldsymbol{\beta}}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}\right)$.

6.2 Bayesian Genome-Based Ridge Regression

When $p > n$, \mathbf{X} is not of full column rank and the posterior of model (6.1) may not be proper (Gelman et al. 2013), so a solution is instead to consider independently proper prior distributions, $\boldsymbol{\beta} \sim N(0, \mathbf{I}\sigma^2)$ and $\sigma^2 \sim IG(\alpha_0, \alpha_0)$, which for large values of σ^2

(10^6) and small values of α_0 (10^{-3}) is an approximation to the standard non-informative prior given in (6.1) (Christensen et al. 2011). A similar prior specification is taken in genomic prediction where different models are obtained by adopting different prior distributions of the parameters. For example, the Bayesian Linear Ridge Regression (Pérez and de los Campos 2014) with standardized covariates (X'_j/s) is given by

$$Y = \mu + \sum_{j=1}^p X_j \beta_j + \epsilon \quad (6.3)$$

with a flat prior for mean parameter (μ), $f(\mu) \propto 1$, which can be approximately specified by $\mu \sim N(0, \sigma_0^2)$, with a large value of σ_0^2 (10^{10}), a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\sigma_\beta^2 \mathbf{I}_p$ on the beta coefficients, $\boldsymbol{\beta}_0 = (\beta_1, \dots, \beta_p)^\top \mid \sigma_\beta^2 \sim N_p(\mathbf{0}, \mathbf{I}_p \sigma_\beta^2)$, and scaled inverse Chi-square distributions as priors for the variance component: $\sigma_\beta^2 \sim \chi_{v_\beta, S_\beta}^{-2}$ (prior for the variance of the regression coefficients β_j) and $\sigma^2 \sim \chi_{v, S}^{-2}$ (prior for the variance of random errors, ϵ), where $\chi_{v, S}^{-2}$ denotes the scaled inverse Chi-squared distribution with shape parameter v and scale parameter S . The joint posterior distribution of the parameters in this model, $\boldsymbol{\theta} = (\mu, \boldsymbol{\beta}_0^\top, \sigma^2, \sigma_\beta^2)^\top$, is given by

$$\begin{aligned} f(\mu, \boldsymbol{\beta}_0^\top, \sigma_\beta^2, \sigma^2 \mid \mathbf{y}, \mathbf{X}) &\propto L(\mu, \boldsymbol{\beta}_0^\top, \sigma^2; \mathbf{y}) f(\boldsymbol{\theta}) \\ &\propto L(\mu, \boldsymbol{\beta}_0^\top, \sigma^2; \mathbf{y}) f(\mu) f(\boldsymbol{\beta}_0 \mid \sigma_\beta^2) f(\sigma_\beta^2) f(\sigma^2) \\ &\propto \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{1}_n \mu - \mathbf{X}_1 \boldsymbol{\beta}_0\|^2\right] \times \exp\left(-\frac{1}{2\sigma_0^2} \mu^2\right) \\ &\times \frac{1}{(\sigma_\beta^2)^{\frac{v_\beta}{2}}} \exp\left(\left[-\frac{1}{2\sigma_\beta^2} \boldsymbol{\beta}_0^\top \boldsymbol{\beta}_0\right]\right) \times \frac{\left(\frac{S_\beta}{2}\right)^{\frac{v_\beta}{2}}}{\Gamma\left(\frac{v_\beta}{2}\right) (\sigma_\beta^2)^{1+\frac{v_\beta}{2}}} \exp\left(-\frac{S_\beta}{2\sigma_\beta^2}\right) \\ &\times \frac{\left(\frac{S}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right) (\sigma^2)^{1+\frac{v}{2}}} \exp\left(-\frac{S}{2\sigma^2}\right). \end{aligned}$$

This has no known form and it is not easy to simulate values of it, so numerical methods are required to explore it. One way to simulate values of this distribution is by means of the Gibbs sampler method, which consists of alternatingly generating samples of the full conditional distributions of each variable (or block of variables) given the rest of the parameters (Casella and George 1992).

The full conditional posteriors to implement the Gibbs sampler are obtained in the lines below.

The conditional posterior distribution of β_0 is given by

$$\begin{aligned}
 f(\beta_0 | -) &\propto L(\mu, \beta_0^T, \sigma^2; \mathbf{y}) f(\beta_0 | \sigma^2) \\
 &\propto \exp \left[-\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{1}_n \mu - \mathbf{X}_1 \beta_0\|^2 - \frac{1}{2\sigma_\beta^2} \beta_0^T \beta_0 \right] \\
 &\propto \exp \left\{ -\frac{1}{2} \left[\beta_0^T (\sigma_\beta^{-2} \mathbf{I}_p + \sigma^{-2} \mathbf{X}_1^T \mathbf{X}_1) \beta_0 - 2\sigma^{-2} (\mathbf{y} - \mathbf{1}_n \mu)^T \mathbf{X}_1 \beta_0 \right] \right\} \\
 &\propto \exp \left\{ -\frac{1}{2} \left[(\beta_0 - \tilde{\beta}_0)^T \tilde{\Sigma}_0^{-1} (\beta_0 - \tilde{\beta}_0) \right] \right\}
 \end{aligned}$$

$\tilde{\Sigma}_0 = (\sigma_\beta^{-2} \mathbf{I}_p + \sigma^{-2} \mathbf{X}_1^T \mathbf{X}_1)^{-1}$ and $\tilde{\beta}_0 = \sigma^{-2} \tilde{\Sigma}_0 \mathbf{X}_1^T (\mathbf{y} - \mathbf{1}_n \mu)$. That is, $\beta_0 | - \sim N_p(\tilde{\beta}_0, \tilde{\Sigma}_0)$. Similarly, the conditional distribution of μ is $\mu | - \sim N(\tilde{\mu}, \tilde{\sigma}_0^2)$, where $\tilde{\sigma}_0^2 = \frac{\sigma^2}{n}$ and $\tilde{\mu} = \frac{1}{n} \mathbf{1}_n^T (\mathbf{y} - \mathbf{X}_1 \beta_0)$.

The conditional distribution of σ^2 is

$$\begin{aligned}
 f(\sigma^2 | -) &\propto L(\mu, \beta_0^T, \sigma^2; \mathbf{y}) f(\sigma^2) \\
 &\propto \frac{1}{(\sigma^2)^{\frac{v}{2}}} \exp \left[-\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{1}_n \mu - \mathbf{X}_1 \beta_0\|^2 \right] \frac{\left(\frac{S}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right) (\sigma^2)^{1+\frac{v}{2}}} \exp \left(-\frac{S}{2\sigma^2} \right) \\
 &\propto \frac{\left(\frac{\tilde{S}}{2}\right)^{\frac{\tilde{v}}{2}}}{(\sigma^2)^{1+\frac{\tilde{v}}{2}}} \exp \left(-\frac{\tilde{S}}{2\sigma^2} \right),
 \end{aligned}$$

where $\tilde{v} = v + n$ and $\tilde{S} = S + \|\mathbf{y} - \mathbf{1}_n \mu - \mathbf{X}_1 \beta_0\|^2$. So $\sigma^2 | - \sim \chi_{\tilde{v}, \tilde{S}}^{-2}$, where $\chi_{v, S}^{-2}$ denotes a scaled inverse Chi-squared distribution with parameters v and S . Similarly, $\sigma_\beta^2 | - \sim \chi_{\tilde{v}_\beta, \tilde{S}_\beta}^{-2}$, where $\tilde{v}_\beta = v_\beta + p$ and $\tilde{S}_\beta = S_\beta + \beta_0^T \beta_0$.

In summary, for the Ridge regression model, a Gibbs sampler consists of the following steps:

1. Choose initial values for μ , β_0 , and σ^2 .
2. Simulate a value of the full conditional distribution of σ_β^2 :

$$\sigma_\beta^2 | \mu, \beta_0, \sigma^2 \sim \chi_{\tilde{v}_\beta, \tilde{S}_\beta}^{-2},$$

where $\chi_{\tilde{v}_\beta, \tilde{S}_\beta}^{-2}$ denotes a scaled inverse Chi-square distribution with shape parameter $\tilde{v}_\beta = v_\beta + p$ and scale parameter $\tilde{S}_\beta = S_\beta + \beta_0^T \beta_0$.

3. Simulate the full conditional posterior distribution of β_0 :

$$\beta_0 \mid \mu, \sigma_\beta^2, \sigma^2 \sim N_p(\tilde{\beta}_0, \tilde{\Sigma}_0),$$

$$\text{where } \tilde{\Sigma}_0 = (\sigma_\beta^{-2} \mathbf{I}_p + \sigma^{-2} \mathbf{X}_1^T \mathbf{X}_1)^{-1} \text{ and } \tilde{\beta}_0 = \sigma^{-2} \tilde{\Sigma}_0 \mathbf{X}_1^T (\mathbf{y} - \mathbf{1}_n \mu)$$

4. Simulate the full conditional distribution of μ :

$$\mu \mid \beta_0, \sigma_\beta^2, \sigma^2 \sim N(\tilde{\mu}, \tilde{\sigma}_\mu^2),$$

$$\text{where } \tilde{\sigma}_\mu^2 = \frac{\sigma^2}{n} \text{ and } \tilde{\mu} = \mathbf{1}_n^T (\mathbf{y} - \mathbf{X}_1 \beta_0).$$

5. Simulate the full conditional distribution of σ^2 :

$$\sigma^2 \mid \mu, \beta_0, \sigma_\beta^2 \sim \chi_{\tilde{v}, \tilde{S}}^{-2},$$

$$\text{where } \tilde{v} = v + n \text{ and } \tilde{S} = S + \|\mathbf{y} - \mathbf{1}_n \mu - \mathbf{X}_1 \beta_0\|^2.$$

6. Repeat steps 2–5 depending on how many values of the parameter vector $(\beta^T, \sigma_\beta^2, \sigma^2)$ you wish to simulate. Usually a large number of iterations are needed and an early part of them are discarded, to finally average the rest of each parameter to obtain estimates of them.

The Gibbs sampler described above can be implemented easily with the BGLR R package: if the hyperparameters S - v and S_{β} - v_{β} are not specified, by default the BGLR function assigns $v = v_{\beta} = 5$, and to S and S_{β} assigns values such that the mode of the priors of σ^2 and σ_{β}^2 (inverse scaled Chi-square) matches a certain proportion of the total variance ($1 - R^2$ and R^2): $S = \text{Var}(Y) \times (1 - R^2) \times (v + 2)$ and $S_{\beta} = \text{Var}(Y) \times R^2 \times (v_{\beta} + 2)$ (see Appendix 2 for more details). Explicitly, in BGLR this model can be implemented by running the following R code:

```
ETA = list(list(model = 'BRR', X = X1, df0 = v_beta, S0 = S_beta, R2 = 1 - R^2) )
A = BGLR(y=y, ETA=ETA, nIter = 1e4, burnIn = 1e3, S0 = S, df0 = v, R2 = R^2)
```

where `nIter = 1e4` and `burnIn = 1e3` are the desired number of iterations and the number of them to be discarded when computing the estimates of the parameters. Remember that when the hyperparameter values are not given, they are set up in the default values, as previously described.

A sub-model of the BRR that does not induce shrinkage of the beta coefficients is obtained by assuming that $(\beta_1, \dots, \beta_p)^T \mid \sigma_\beta^2 \sim N_p(\mathbf{0}, \mathbf{I}_p \sigma_\beta^2)$, ignoring the prior distribution of σ_β^2 and setting this at a very high value (10^{10}). Note that this model is very similar to the Bayesian model obtained by adopting the prior (6.2), under which the beta coefficients are estimated solely with the information contained in the likelihood function (Pérez and de los Campos 2014). This prior model can also be implemented in the BGLR package and is called **FIXED**. Certainly, the Gibbs

sampler steps for its implementation are the same as the steps described before for the BRR, except that step 2 is removed (no simulations are obtained from σ_β^2) and σ_β^{-2} is set equal to zero in the full conditional of β_0 (step 3).

6.3 Bayesian GBLUP Genomic Model

In genomic-enabled prediction, the number of markers used to predict the performance of a trait of interest is often very large compared to the number of individuals phenotyped in the sample ($p \gg n$); for this reason, some computational difficulties may arise when exploring the posterior distribution of the beta coefficients. When the main objective is to use this model for predictive purposes, a solution consists of reducing the dimension of the problem by directly simulating values of $\mathbf{g} = \mathbf{X}_1\beta_0$ (breeding values or genomic effects, Lehermeier et al. 2013) instead of only from β_0 . To do this, first note that because $\beta_0 | \sigma_\beta^2 \sim N_p(\mathbf{0}, \mathbf{I}_p\sigma_\beta^2)$, to induce a prior for \mathbf{g} , this is defined as $\mathbf{g} = \mathbf{X}_1\beta_0 | \sigma_\beta^2 \sim N_n(\mathbf{0}, \sigma_\beta^2\mathbf{X}_1\mathbf{X}_1^T) = N_n(\mathbf{0}, \sigma_g^2\mathbf{G})$, where $\sigma_g^2 = p\sigma_\beta^2$ and $\mathbf{G} = \frac{1}{p}\mathbf{X}_1\mathbf{X}_1^T$, which is known as the genomic relationship matrix (VanRaden 2007). Then, under this parameterization ($\mathbf{g} = \mathbf{X}_1\beta_0$ and $\sigma_g^2 = p\sigma_\beta^2$), the model specified in (6.3), in matrix notation takes the following form:

$$\mathbf{Y} = \mathbf{1}_n\mu + \mathbf{g} + \epsilon \quad (6.4)$$

with a flat prior to mean parameter (μ), $\sigma^2 \sim \chi_{v,S}^{-2}$, and the induced priors: $\mathbf{g} = \mathbf{X}_1\beta_0 | \sigma_g^2 \sim N_n(\mathbf{0}, \sigma_g^2\mathbf{G})$ and $\sigma_g^2 \sim \chi_{v_g}^{-2} S_g$ ($v_g = v_\beta$, $S_g = pS_\beta$).

Similarly to what was done for model (6.3), the full conditional posterior distribution of \mathbf{g} in model (6.4) is given by

$$\begin{aligned} f(\mathbf{g}|-) &\propto L(\mu, \mathbf{g}, \sigma^2; \mathbf{y})f(\sigma_g^2) \\ &\propto \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{1}_n\mu - \mathbf{g}\|^2\right] \frac{1}{(\sigma_g^2)^{\frac{n}{2}}} \exp\left(\left[-\frac{1}{2\sigma_g^2}\mathbf{g}^T\mathbf{G}^{-1}\mathbf{g}\right]\right) \\ &\propto \exp\left\{-\frac{1}{2}\left[(\mathbf{g} - \tilde{\mathbf{g}})^T\tilde{\mathbf{G}}^{-1}(\mathbf{g} - \tilde{\mathbf{g}})\right]\right\}, \end{aligned}$$

where $\tilde{\mathbf{G}} = (\sigma_g^{-2}\mathbf{G}^{-1} + \sigma^{-2}\mathbf{I}_n)^{-1}$ and $\tilde{\mathbf{g}} = \sigma^{-2}\tilde{\mathbf{G}}(\mathbf{y} - \mathbf{1}_n\mu)$, and from here $\mathbf{g} | - \sim N_n(\tilde{\mathbf{g}}, \tilde{\mathbf{G}})$. Then, the mean/mode of $\mathbf{g} | -$ is $\tilde{\mathbf{g}} = \sigma^{-2}\tilde{\mathbf{G}}(\mathbf{y} - \mathbf{1}_n\mu)$, which is also the best linear unbiased predictor (BLUP) of \mathbf{g} under the mixed model equation of Henderson (1975) using the machinery of a classic linear mixed model described in the previous chapter for model (6.4), after recognizing the

prior distribution of \mathbf{g} as the distribution of the random effects in a model, ignoring the priors specification of the rest of the parameters and assuming that they are known (Henderson 1975). For this reason, model (6.4) is often referred to as GBLUP. If \mathbf{G} is replaced by the pedigree matrix \mathbf{A} , the resulting model is known as PBLUP or ABLUP.

The full conditional posterior of the rest of parameters is similar to the BRR model: $\mu \mid - \sim N(\tilde{\mu}, \tilde{\sigma}_0^2)$, where $\tilde{\sigma}_0^2 = \frac{\sigma^2}{n}$ and $\tilde{\mu} = \frac{1}{n} \mathbf{1}_n^T (\mathbf{y} - \mathbf{g})$; $\sigma^2 \mid - \sim \chi_{\tilde{\nu}, \tilde{S}}^{-2}$, where $\tilde{\nu} = \nu + n$ and $\tilde{S} = S + \|\mathbf{y} - \mathbf{1}_n \mu - \mathbf{g}\|^2$; and $\sigma_g^2 \mid - \sim \chi_{\tilde{\nu}_g, \tilde{S}_g}^{-2}$, where $\tilde{\nu}_g = \nu_g + n$ and $\tilde{S}_g = S_g + \mathbf{g}^T \mathbf{G}^{-1} \mathbf{g}$.

Note that when $p \gg n$, then the dimension of the parameter space of the posterior of GBLUP model is lower than the BRR.

The GBLUP model (6.4) also can be implemented easily with the BGLR R package, and when the hyperparameters S - ν and S_g - ν_g are not specified, $\nu = \nu_g = 5$ is used by default and the scale parameters are settled similarly as in the BRR.

The BGLR code to fit this model:

```
ETA = list( list( model = 'RHKS', K = G, df0 = v_g, S0 = S_g, R2 = 1-R^2) )
A = BGLR(y=y, ETA = ETA, nIter = 1e4, burnIn = 1e3, S0 = S, df0 = v, R2 = R^2)
```

The GBLUP can be equivalently expressed and consequently fitted with the BRR model by making the design matrix equal to the lower triangular factor of the Cholesky decomposition of the genomic relationship matrix, i.e., $\mathbf{X} = \mathbf{L}$, where $\mathbf{G} = \mathbf{L}\mathbf{L}'$. So, with the BGLR package, the BRR implementation of a GBLUP model is.

```
L = t(chol(G))
ETA = list( list( model = 'BRR', X = L, df0 = v_beta, S0 = S_beta, R2 = 1-R^2) )
A = BGLR(y=y, ETA = ETA, nIter = 1e4, burnIn = 1e3, S0 = S, df0 = v, R2 = R^2)
```

When there is more than one repetition of an individual in the data at hand, or a more sophisticated design is used in the data collection, model (6.4) can be specified in a more general way to take into account this structure, as follows:

$$\mathbf{Y} = \mathbf{1}_n \mu + \mathbf{Zg} + \boldsymbol{\epsilon} \quad (6.5)$$

with \mathbf{Z} the incident matrix of the genotypes. This model cannot be fitted directly in the BGLR and some precalculus is needed first to compute the ‘‘covariance’’ matrix of the predictor \mathbf{Zg} in model (6.5): $\mathbf{K}_L = \text{Var}(\mathbf{Zg}) = \mathbf{ZGZ}^T$. The BGLR code for implementing this model is the following:

```
Z = model.matrix(~0+GID, data=dat_F, xlev = list(GID=unique
(dat_F$GID)))
K_L = Z%*%G%*%t(Z)
ETA = list( list( model = 'RHKS', K = K_L, df0 = v_g, S0 = S_g, R2 = 1-R^2) )
A = BGLR(y=y, ETA = ETA, nIter = 1e4, burnIn = 1e3, S0 = S, df0 = v, R2 = R^2)
```

where **dat_F** is the data set that contains the necessary phenotypic information (GID: Lines or individuals; y: response variable of the trait of interest).

6.4 Genomic-Enabled Prediction BayesA Model

Another variant to the standard Bayesian model (6.1) is the BayesA model proposed by Meuwissen et al. (2001), which is a slight modification of the BRR model obtained with the same prior distributions, except that now a specific variance $\sigma_{\beta_j}^2$ is assumed for each of the covariate (marker) effects, that is, $\beta_j | \sigma_{\beta_j}^2 \sim N(0, \sigma_{\beta_j}^2)$, and these variance parameters are supposed to be independent random variables with a scaled inverse Chi-square distribution with parameters ν_β and S_β , $\sigma_{\beta_j}^2 \sim \chi^{-2}(\nu_\beta, S_\beta)$. These specific variances for each marker effect provide covariate heterogeneous shrinkage estimation. Furthermore, a gamma distribution is assigned to S_β , $S_\beta \sim G(r, s)$, where $G(r, s)$ denotes a gamma distribution where r and s are the rate and shape parameters, respectively. By providing a different prior variance for each β_j , this model has the potential of inducing covariate-specific shrinkage of estimated effects (Pérez and de los Campos 2013).

Note that choosing $r = r^*/\nu_\beta$ and taking very large values of ν_β , the prior of $\sigma_{\beta_j}^2$ collapses to a degenerate distribution at S_β , and the BRR model is obtained, but with a gamma distribution with parameters r^* and s as priors to the common variance of the effects $\sigma_\beta^2 = \text{Var}(\beta_j) = S_\beta$, instead of $\chi^{-2}(\nu_\beta, S_\beta)$. Furthermore, the marginal distribution of each beta coefficient β_j , that is, the unconditional distribution of $\beta_j | S_\beta$, is a scaled- t -student distribution (scaled by $\sqrt{S_\beta/\nu_\beta}$). These distributions, compared to the normal, have heavier tails and put higher mass around 0, which compared to the BRR, induce fewer shrinkage estimates of covariates with sizable effects, and induce strong shrinkage toward zero estimates of covariates with smaller effects, respectively (de los Campos et al. 2013).

A Gibbs sampler implementation for estimating the parameters of this model can be done following steps 1–6 of the BRR model, where the second step is replaced by the next 2.1 and 2.2 steps, the third and the last step are replaced and modified by the next steps 3 and 6:

2.1 Simulate from the full conditional of each $\sigma_{\beta_j}^2$

$$\sigma_{\beta_j}^2 | \mu, \beta_0, \sigma_{-j}^2, S_\beta, \sigma^2 \sim \chi_{\nu_j, S_{\beta_j}}^{-2},$$

where $\tilde{\nu}_{\beta_j} = \nu_\beta + 1$ and scale parameter $\tilde{S}_{\beta_j} = S_\beta + \beta_j^2$, where σ_{-j}^2 is the vector $\sigma_\beta^2 = (\sigma_{\beta_1}^2, \dots, \sigma_{\beta_p}^2)$ but without the j th entry.

2.2 Simulate from the full conditional of S_β

$$\begin{aligned}
f(S_\beta | -) &\propto \left[\prod_{j=1}^p f(\sigma_{\beta_j}^2 | S_\beta) \right] f(S_\beta) \\
&\propto \prod_{j=1}^p \left[\frac{\left(\frac{S_\beta}{2}\right)^{\frac{v_\beta}{2}}}{\Gamma\left(\frac{v_\beta}{2}\right) (\sigma_{\beta_j}^2)^{1+\frac{v_\beta}{2}}} \exp\left(-\frac{S_\beta}{2\sigma_{\beta_j}^2}\right) \right] S_\beta^{s-1} \exp(-rS_\beta) \\
&\propto S_\beta^{s+\frac{pv_\beta}{2}-1} \exp\left[-\left(r + \frac{1}{2} \sum_{j=1}^p \frac{1}{\sigma_{\beta_j}^2}\right) S_\beta\right]
\end{aligned}$$

which corresponds to the kernel of the gamma distribution with rate parameter $\tilde{r} = r + \frac{1}{2} \sum_{j=1}^p \frac{1}{\sigma_{\beta_j}^2}$ and shape parameter $\tilde{s} = s + \frac{pv_\beta}{2}$, and so $S_\beta | - \sim \text{Gamma}(\tilde{r}, \tilde{s})$.

3. Simulate the full conditional posterior distribution of β_0 :

$$\beta_0 | \mu, \sigma_\beta^2, \sigma^2 \sim N_p(\tilde{\beta}_0, \tilde{\Sigma}_0),$$

where $\tilde{\Sigma}_0 = \left(D_p^{-1} + \sigma^{-2} X_1^T X_1\right)^{-1}$ and $\tilde{\beta}_0 = \sigma^{-2} \tilde{\Sigma}_0 X_1^T (y - \mathbf{1}_n \mu)$, $D_p = \text{Diag}(\sigma_{\beta_1}^2, \dots, \sigma_{\beta_p}^2)$.

6. Repeat steps 2–5 (given in the BRR method) depending on how many values of the parameter vector $(\beta^T, \sigma_\beta^2, \sigma^2, S_\beta)$ we wish to simulate.

When implementing this model in the BGLR package, by default $v = v_\beta = 5$ are used and $S = \text{Var}(Y) \times (1 - R^2) \times (v + 2)$, which makes the mode of the priors of σ^2 ($\chi^{-2}(v, S)$) match a certain proportion of the total variance $(1 - R^2)$. If the hyperparameters of the a priori for S_β , s , and r , are not specified, by default BGLR takes $s = 1.1$ to have an a priori ($S_\beta \sim G(s, r)$) that is relatively non-informative with a coefficient of variation $(1/\sqrt{s})$ of approximately 95%. Then, to the rate parameter it assigns the value $r = (s - 1)/S_\beta$, with $S_\beta = \text{Var}(y) \times R^2 \times (v_\beta + 2)/S_x^2$, where S_x^2 is the sum of the variances of the columns of X and R^2 is the percentage of the total variation that a priori is required to attribute to the covariates in X . The BGLR code for implementing this model is

```

ETA = list(list(model = 'BayesA', X=X1, df0 = vβ, rate0 = r, shape0 = s,
R2 = 1-R2))
A = BGLR(y=y, ETA = ETA, nIter = 1e4, burnIn = 1e3, S0 = S, df0 = v, R2 =
R2)

```

6.5 Genomic-Enabled Prediction BayesB and BayesC Models

Other variants of the model (6.1) are the BayesC and the BayesB models, which in turn can be considered as direct extensions of the BRR and BayesA models, respectively, by adding a parameter π that represents the prior proportion of covariates with nonzero effect (Pérez and de los Campos 2014).

The **BayesC** is the same as the BRR, but instead of assuming a priori that the beta coefficients are independent normal random variables with mean 0 and variance σ_β^2 , it assumes that with probability π each β_j comes from a $N(0, \sigma_\beta^2)$, and with probability $1 - \pi$ comes from a degenerate distribution (DG) at zero, that is, $\beta_1, \dots, \beta_p \mid \sigma_\beta^2, \pi \stackrel{iid}{\sim} \pi_p N(0, \sigma_\beta^2) + (1 - \pi_p) DG(0)$ (mix of a normal distribution with mean 0 and variance σ_β^2 , and degenerate distribution at zero). In addition, for the parameter π_p , a beta distribution is assigned as prior, that is, $\pi_p \sim \text{Beta}(\pi_{p0}, \phi_0)$, where $\pi_{p0} = E(\pi_p)$ represents the mean and ϕ_0^{-1} is the “dispersion” parameter ($\text{var}(\pi_p) = \frac{\pi_{p0}(1-\pi_{p0})}{\phi_0+1}$). If $\phi_0 = 2$ and $\pi_{p0} = 0.5$, the prior for π_p is a uniform distribution in $(0,1)$. For large values of ϕ_0 , the distribution for π_p is highly concentrated around π_{p0} , and so the BayesC is reduced to BRR when $\pi_{p0} = 1$ for large values of ϕ_0 .

For this model, the full conditional distributions of μ and σ_β^2 are the same as the model described before, that is, $\mu \mid - \sim N(\tilde{\mu}, \tilde{\sigma}_\mu^2)$ and $\sigma^2 \mid - \sim \chi_{v,S}^{-2}$. However, for the rest of the parameters, this does not have a known form and is not easy to simulate from them. A solution is to introduce a latent variable to represent the prior distribution of each β_j , and compute all the conditional distributions in this augmented scheme, including the distribution corresponding to the latent variable. To do this, note that this prior can be specified by assuming that conditional to a binary latent variable Z_j ,

$$\beta_j \mid \sigma_\beta^2, Z_j = z \sim \begin{cases} N(0, \sigma_\beta^2), \\ DG(0), \end{cases}$$

where Z_j is a Bernoulli random variable with parameter π_p ($Z_j \sim \text{Ber}(\pi_p)$). With this introduced latent variable, all the full conditionals can be derived, as is described next.

If the current value of z_j is 1, the full conditional posterior of β_j is

$$\begin{aligned}
f(\beta_j | -) &\propto L(\mu, \beta_0, \sigma^2; \mathbf{y}) f(\beta_j | \sigma_\beta^2, z_j) \\
&\propto \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{1}_n \mu - \mathbf{X}_1 \beta_0\|^2\right) \frac{1}{\sqrt{2\pi\sigma_\beta^2}} \exp\left(-\frac{\beta_j^2}{2\sigma_\beta^2}\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_{ij} - x_{ij}\beta_j)^2 - \frac{1}{2\sigma_\beta^2} \beta_j^2\right) \\
&\propto \exp\left\{-\frac{1}{2} \left[\left(\sigma_\beta^{-2} + \sigma^{-2} \sum_{i=1}^n x_{ij}^2\right) \beta_j^2 - 2\sigma^{-2} \sum_{i=1}^n x_{ij} y_{ij} \beta_j + \frac{1}{\sigma^2} \sum_{i=1}^n y_{ij}^2 \right]\right\} \\
&\propto \exp\left[-\frac{(\beta_j - \tilde{\beta}_j)^2}{2\tilde{\sigma}_j^2}\right],
\end{aligned}$$

where $y_{ij} = y_i - \sum_{k=1}^p x_{ik}\beta_k$, $\tilde{\sigma}_j^2 = \left(\sigma_\beta^{-2} + \sigma^{-2} \sum_{i=1}^n x_{ij}^2\right)^{-1}$, and $\tilde{\beta}_j =$
 $k \neq j$

$\sigma^{-2} \tilde{\sigma}_j^2 \sum_{i=1}^n x_{ij} y_{ij}$. That is, when the current value of z_j is 1, $\beta_j | - \sim N(\tilde{\beta}_j, \tilde{\sigma}_j^2)$. However, if $z_j = 0$, the full conditional posterior of β_j is a degenerate random variable at 0, that is, $\beta_j | - \sim \text{DG}(0)$.

The full conditional distribution of Z_j is

$$\begin{aligned}
f(z_j | -) &\propto f(\beta_j | \sigma_\beta^2, z_j) f(z_j) \\
&\propto f(\beta_j | \sigma_\beta^2, z_j) \pi_p^{z_j} (1 - \pi_p)^{1-z_j}
\end{aligned}$$

from which, conditional on the rest of the parameters, Z_j is a Bernoulli random

variable with parameter $\tilde{\pi}_{pj} = \frac{\frac{\pi_p}{\sqrt{2\pi\sigma_\beta^2}} \exp\left(-\frac{\beta_j^2}{2\sigma_\beta^2}\right)}{\frac{\pi_p}{\sqrt{2\pi\sigma_\beta^2}} \exp\left(-\frac{\beta_j^2}{2\sigma_\beta^2}\right) + (1-\pi_p)\delta_0(\beta_j)}$. Note however that, if

$\beta_j \neq 0$, $\tilde{\pi}_{pj} = 1$, and then $Z_j = 1$ with probability 1, when simulating from the full conditional posterior of β_j , we will always obtain values different from zero, and this cyclic behavior will remain permanent. On the other hand, note that if $\beta_j = 0$,

$\tilde{\pi}_{pj} = \frac{\frac{\pi_p}{\sqrt{2\pi\sigma_\beta^2}}}{\frac{\pi_p}{\sqrt{2\pi\sigma_\beta^2}} + (1-\pi_p)}$ is not 0, then the next simulated value of β_j will be different

from 0, and in this scenario, in the next steps of the ‘‘Gibbs,’’ Z_j will at all times be 1, and so the chain has absorbing states and will not explore the entire sampling space. A solution to this problem consists of trying to simulate from the joint conditional distribution of β_j and Z_j , that is, from $\beta_j, Z_j | -$. This full joint conditional distribution can be computed as

$$f(\beta_j, z_j | -) \propto f_*(z_j) f(\beta_j | z_j, -) \propto f_*(z_j) L(\mu, \boldsymbol{\beta}_0, \sigma^2; \mathbf{y}) f(\beta_j | \sigma_\beta^2, z_j),$$

where $f_*(z_j)$ is the marginal conditional distribution of Z_j conditioned to all parameters except β_j ($Z_j | -j \sim f_*(\cdot)$). Specifically, this is given by “integrating” ($f(\beta_j, z_j | -)$) with respect to β_j)

$$\begin{aligned} f_*(z_j) &\propto \int_{-\infty}^{\infty} f(\beta_j, z_j | -) d\beta_j \propto \int_{-\infty}^{\infty} L(\mu, \boldsymbol{\beta}_0, \sigma^2; \mathbf{y}) f(\beta_j | \sigma_\beta^2, z_j) f(z_j | \pi_p) d\beta_j \\ &\propto \begin{cases} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_{ij} - x_{ij}\beta_j)^2\right] \frac{1}{\sqrt{2\pi\sigma_\beta^2}} \exp\left(-\frac{\beta_j^2}{2\sigma_\beta^2}\right) \pi_p d\beta_j \\ \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_{ij} - x_{ij}\beta_j)^2\right] \delta_0(\beta_j) (1 - \pi_p) d\beta_j \\ \pi_p \sqrt{\frac{\tilde{\sigma}_j^2}{\sigma_\beta^2}} \\ 1 - \pi_p. \end{cases} \end{aligned}$$

From here, $Z_j | -$ is a Bernoulli random distribution with parameter $\tilde{\pi}_p =$

$\left(\pi_p \sqrt{\frac{\tilde{\sigma}_j^2}{\sigma_\beta^2}}\right) / \left(\pi_p \sqrt{\frac{\tilde{\sigma}_j^2}{\sigma_\beta^2}} + 1 - \pi_p\right)$. With this and the full conditional distribution derived above for β_j , an easy way to simulate values from $\beta_j, Z_j | -$ consists of first simulating a z_j value from $Z_j | -j \sim \text{Ber}(\tilde{\pi}_p)$, and then, if $z_j = 1$, simulating a value of β_j from $\beta_j | - \sim N(\tilde{\beta}_j, \tilde{\sigma}_j^2)$, otherwise take $\beta_j = 0$.

Now, note that the full conditional distribution of σ_β^2 is

$$\begin{aligned} f(\sigma_\beta^2 | -) &\propto \left\{ \prod_j^p [f(\beta_j | \sigma_\beta^2, z_j)]^{z_j} \right\} f(\sigma_\beta^2) \\ &\propto \frac{1}{(\sigma_\beta^2)^{p\tilde{z}_p}} \exp\left(-\frac{1}{2\sigma_\beta^2} \sum_{j=1}^p z_j \beta_j^2\right) \frac{\left(\frac{S_\beta}{2}\right)^{\frac{v_\beta}{2}}}{\Gamma\left(\frac{v_\beta}{2}\right) (\sigma_\beta^2)^{1+\frac{v_\beta}{2}}} \exp\left(-\frac{S_\beta}{2\sigma_\beta^2}\right) \\ &\propto \frac{1}{(\sigma_\beta^2)^{p\tilde{z}_p}} \exp\left(-\frac{1}{2\sigma_\beta^2} \sum_{j=1}^p z_j \beta_j^2\right) \frac{\left(\frac{S_\beta}{2}\right)^{\frac{v_\beta}{2}}}{\Gamma\left(\frac{v_\beta}{2}\right) (\sigma_\beta^2)^{1+\frac{v_\beta}{2}}} \exp\left(-\frac{S_\beta}{2\sigma_\beta^2}\right) \\ &\propto \frac{1}{(\sigma_\beta^2)^{1+\frac{v_\beta}{2}}} \exp\left(-\frac{\tilde{S}_\beta}{2\sigma_\beta^2}\right), \end{aligned}$$

where $\bar{z}_p = 1/p \sum_{j=1}^p z_j$, $\tilde{S}_\beta = S_\beta + \sum_{j=1}^p z_j \beta_j^2$, and $\tilde{v}_\beta = v_\beta + p\bar{z}_p$. That is, $\sigma_\beta^2 \mid \mu, \beta_0, \sigma^2, \mathbf{z} \sim \chi_{v, S_\beta}^{-2}$. The full conditional distributions of μ and σ^2 are the same as BRR, that is, $\mu \mid - \sim N(\tilde{\mu}, \tilde{\sigma}_\mu^2)$ and $\sigma^2 \mid - \sim \chi_{v, \tilde{S}}^{-2}$, with $\tilde{\sigma}_\mu^2 = \frac{\sigma^2}{n}$, $\tilde{\mu} = \mathbf{1}_n^\top (\mathbf{y} - \mathbf{X}_1 \beta_0)$, $\tilde{v} = v + n$, and $\tilde{S} = S + \|\mathbf{y} - \mathbf{1}_n \mu - \mathbf{X}_1 \beta_0\|^2$.

The full conditional distribution of π_p is

$$\begin{aligned} f(\pi_p \mid -) &\propto \left[\prod_{j=1}^p f(z_j \mid \pi_p) \right] f(\pi_p) \\ &\propto \pi_p^{p\bar{z}_p} (1 - \pi_p)^{p(1-\bar{z}_p)} \pi_p^{\phi_0 \pi_{p0} + 1} (1 - \pi_p)^{\phi_0 (1 - \pi_{p0}) - 1} \\ &\propto \pi_p^{\phi_0 \pi_{p0} + p\bar{z}_p - 1} (1 - \pi_p)^{\phi_0 (1 - \pi_{p0}) + p(1 - \bar{z}_p) - 1} \end{aligned}$$

which means that $\pi_p \mid - \sim \text{Beta}(\tilde{\pi}_{p0}, \tilde{\phi}_0)$, with $\tilde{\phi}_0 = \phi_0 + p$ and $\tilde{\pi}_{p0} = \frac{\phi_0 \pi_{p0} + p\bar{z}_p}{\phi_0 + p}$.

The **BayesB** model is a variant of BayesA that assumes almost the same prior models to the parameters, except that instead of assuming independent normal random variables with common mean 0 and common variance σ_β^2 for the beta coefficients, this model adopts a mixture distribution, that is, $\beta_j \mid \sigma_\beta^2, \pi \stackrel{iid}{\sim} \pi N(0, \sigma_\beta^2) + (1 - \pi) \text{DG}(0)$, with $\pi \sim \text{Beta}(\pi_0, p_0)$. This model has the potential to perform variable selection and produce covariate-specific shrinkage estimates (Pérez et al. 2010).

This model can also be considered an extension of the BayesC model with a gamma distribution as prior to the scale parameter of the a priori distribution of the variance of the beta coefficients, that is, $S_\beta \sim G(s, r)$. It is interesting to point out that if $\pi = 1$, this model is reduced to BayesA, which is obtained by taking $\pi_0 = 1$ and letting ϕ_0 go to ∞ . Also, this is reduced to the BayesC by setting $s/r = S_\beta^0$ and choosing a very large value for r .

To explore the posterior distribution of this model, the same Gibbs sampler given for BayesC can be used, but adding to the process the full conditional posterior distribution of S_β : $S_\beta \mid - \sim \text{Gamma}(\tilde{r}, \tilde{s})$, where $\tilde{r} = r + \frac{1}{2\sigma_\beta^2}$ and shape parameter $\tilde{s} = s + \frac{v_\beta}{2}$.

When implementing both these models in the BGLR R package, by default this assigns $\pi_{p0} = 0.5$ and $\phi_0 = 10$, for the hyperparameters of the prior model of π_p , Beta(π_{p0}, ϕ_0), which results in a weakly informative prior. For the remaining hyperparameters of the **BayesC** model, by default BGLR assigns values like those assigned to the BRR model, but with some modifications to consider because a priori now only a proportion π_0 of the covariates (columns of \mathbf{X}) has nonzero effects:

$$\begin{aligned}
v &= v_\beta = 5, \\
S &= \text{Var}(Y) \times (1 - R^2) \times (v + 2), \\
S_\beta &= \text{Var}(y) \times R^2 \times \frac{(v_\beta + 2)}{S_x^2 \pi_0}.
\end{aligned}$$

While for the remaining hyperparameters of **BayesB**, by default BGLR also assigns values similar to **BayesA**: $v = v_\beta = 5$, $S = \text{Var}(Y) \times (1 - R^2) \times (v + 2)$, $r = (s - 1)/S_\beta$, with $S_\beta = \text{Var}(y) \times R^2 \times \frac{(v_\beta + 2)}{S_x^2 \pi_0}$, where S_x^2 is the sum of the variances of the columns of X .

The BGLR codes to implement these models are, respectively:

```
ETA = list( list( model = 'BayesC', X=X1 ) )
A = BGLR(y=y, ETA=ETA, nIter=1e4, burnIn=1e3, df0=v, S0=S, probIn
= pi_p0, counts = phi_0, R2 = R^2)
```

and

```
ETA = list( list( model = 'BayesB', X=X1 ) )
A = BGLR(y=y, ETA=ETA, nIter=1e4, burnIn=1e3, df0=v, rate0=r,
shape0=s, probIn=pi_p0, counts = phi_0, R2 = R^2)
```

6.6 Genomic-Enabled Prediction Bayesian Lasso Model

Another variant of the model (6.1) is the Bayesian Lasso linear regression model (**BL**). This model assumes independent Laplace or double-exponential distributions with location and scale parameters 0 and $\frac{\sqrt{\sigma^2}}{\lambda}$, respectively, for the beta coefficients, that is, $\beta_1, \dots, \beta_p \mid \sigma^2, \lambda \stackrel{iid}{\sim} L\left(0, \frac{\sqrt{\sigma^2}}{\lambda}\right)$. Furthermore, the priors for parameters μ and σ^2 are the same as in the models described before, while for λ^2 , a gamma distribution with parameters s_λ and r_λ is often adopted.

Because compared to the normal distribution, the Laplace distribution has fatter tails and puts higher density around 0, this prior induces stronger shrinkage estimates for covariates with relatively small effects and reduced shrinkage estimates for covariates with larger effects (Pérez et al. 2010).

A more convenient specification of the prior for the beta coefficients in this model is obtained with the representation proposed by Park and Casella (2008), which is a continuous scale mixture of a normal distribution: $\beta_j \mid \tau_j \sim N(0, \tau_j \sigma^2)$ and $\tau_j \sim \text{Exp}(2/\lambda^2)$, $j = 1, \dots, p$, where $\text{Exp}(\theta)$ denotes an exponential distribution with scale parameter θ . So, unlike the prior used by the BRR model, this prior distribution also puts a higher mass at zero and has heavier tails, which induce stronger shrinkage estimates for covariates with relatively small effect and less shrinkage estimates for markers with sizable effect (Pérez et al. 2010).

Note that the prior distribution for the beta coefficients and the prior variance of this distribution in BayesB and BayesC can be equivalently expressed as a mixture of a scaled inverse Chi-squared distribution with parameters ν_β and S_β , and a degenerate distribution at zero, that is, $\beta_j \sim N(0, \sigma_\beta^2)$ and $\sigma_\beta^2 \sim \pi_p \chi^{-2}(\nu_\beta, S_\beta) + (1 - \pi_p) \text{DG}(0)$. So, based on this result and the connections between the models described before, the main difference between all these models is the manner in which the prior variance of the predictor variable is modelled.

Example 1 To illustrate how to use the models described before, here we consider the prediction of grain yield (tons/ha) based on marker information. The data set used consists of 30 lines in four environments with one and two repetitions and the genotyped information contains 500 markers for each line. The numbers of lines with one (two) repetition are 6 (24), 2 (28), 0 (30), and 3 (27) in Environments 1, 2, 3, and 4, respectively, resulting in 229 observations. The performance prediction of all these models was evaluated with 10 random partitions in a cross-validation strategy, where 80% of the complete data set was used to fit the model and the rest to evaluate the model in terms of the mean squared error of prediction (MSE).

The results for all models (shown in Table 6.1) were obtained by iterating 10,000 times the corresponding Gibbs sampler and discarding the first 1000 of them, using the default hyperparameter values implemented in BGLR. This indicates that the behavior of all the models is similar, except the BayesC, where the MSE is slightly greater than the rest.

The R code to obtain the results in Table 6.1 is given in Appendix 3.

What happens when using other hyperparameter values? Although the ones used here (proposed by Pérez et al. 2010) did not always produce the best prediction performance (Lehermeier et al. 2013) and there are other ways to propose the hyperparameter values in these models (Habier et al. 2010, 2011), it is important to point out that the values used by default in BGLR work reasonably well and that it

Table 6.1 Mean squared error (MSE) of prediction across 10 random partitions, with 80% for training and the rest for testing, in five Bayesian linear models

PT	BRR	GBLUP	BayesA	BayesB	BayesC
1	0.5395	0.5391	0.5399	0.5415	0.5470
2	0.3351	0.3345	0.3345	0.3411	0.3511
3	0.4044	0.4065	0.4056	0.4050	0.4106
4	0.3540	0.3583	0.3571	0.3561	0.3559
5	0.3564	0.3555	0.3556	0.3549	0.3604
6	0.7781	0.7772	0.7702	0.7777	0.7776
7	0.7430	0.7357	0.7380	0.7384	0.7431
8	0.3662	0.3717	0.3695	0.3669	0.3661
9	0.3065	0.3026	0.3056	0.3021	0.3030
10	0.5842	0.5822	0.5846	0.5860	0.6079
Mean (SD)	0.4767 (0.1742)	0.4763 (0.1724)	0.476 (0.1716)	0.4769 (0.1734)	0.4822 (0.1744)

is not easy to find other combinations that work better in all applications, and when you want to use other combinations of hyperparameters you need to be very careful because you can dramatically affect the predictive performance of the model that uses the default hyperparameters.

Indeed, by means of simulated and experimental data, Lehermeier et al. (2013) observed a strong influence on the predictive performance of the hyperparameters given to the prior distributions in BayesA, BayesB, and the Bayes Lasso with fixed λ . Specifically, in the first two models, they observed that the scale parameter S_β of the prior distribution of variance of β_j had a strong effect on the predictive ability because overfitting in the data occurred when a too large value of this value was chosen, whereas underfitting was observed when too small values of this parameter were used. Note that this is expected approximately by seeing that in both models (BayesA and BayesB), $\text{Var}(\beta_j) = E(\sigma_{\beta_j}^2) = S_\beta / (v_\beta - 1)$, which is almost the inverse of the regularization parameter in any type of Ridge regression model.

6.7 Extended Predictor in Bayesian Genomic Regression Models

All the Bayesian formulations of the model (6.1) described before can be extended, in terms of the predictor, to easily take into account the effects of other factors. For example, effects of environments and environment–marker interaction can be added as

$$\mathbf{y} = \mathbf{1}_n\mu + X_E\boldsymbol{\beta}_E + X\boldsymbol{\beta} + X_{EM}\boldsymbol{\beta}_{EM} + \boldsymbol{\epsilon}, \quad (6.6)$$

where X_E and X_{EM} are the design matrices of the environments and environment–marker interactions, respectively, while $\boldsymbol{\beta}_E$ and $\boldsymbol{\beta}_{EM}$ are the vectors of the environment effects and the interaction effects, respectively, with a prior distribution that can be specified as was done for $\boldsymbol{\beta}$. Indeed, with the BGLR function all these things are possible, and all the options described before can also be adopted for the rest of effects added in the model: FIXED, BRR, BayesA, BayesB, BayesC, and BL.

Under the RKHS model with genotypic and environment–genotypic interaction effects, in the predictor, the modified model (6.6) is expressed as

$$\mathbf{Y} = \mathbf{1}_n\mu + X_E\boldsymbol{\beta}_E + \mathbf{Z}_L\mathbf{g} + \mathbf{Z}_{LE}\mathbf{g}\mathbf{E} + \boldsymbol{\epsilon}, \quad (6.7)$$

where \mathbf{Z}_L and \mathbf{Z}_{LE} are the incident matrices of the genomic and environment–genotypic interaction effects, respectively. Similarly to model (6.5), this model cannot be fitted directly in the BGLR and some precalculations are needed first to compute the “covariance” matrix of the predictors $\mathbf{Z}_L\mathbf{g}$ and $\mathbf{Z}_{LE}\mathbf{g}\mathbf{E}$, which are $\mathbf{K}_L = \sigma_g^{-2} \text{Var}(\mathbf{Z}_L\mathbf{g}) = \mathbf{Z}_L\mathbf{G}\mathbf{Z}_L^T$ and $\mathbf{K}_{LE} = \sigma_{gE}^{-2} \text{Var}(\mathbf{Z}_{LE}\mathbf{g}\mathbf{E}) = \mathbf{Z}_{LE}(\mathbf{I}_I \otimes \mathbf{G})\mathbf{Z}_{LE}^T$,

respectively, where I is the number of environments. The BGLR code for implementing this model is the following:

```

I = length(unique(dat_F$Env))
XE = model.matrix(~0+Env, data=dat_F)[, -1]
Z_L = model.matrix(~0+GID, data=dat_F, xlev = list(GID=unique
(dat_F$GID)))
K_L = Z_L %*%G%*t(Z_L)
Z_LE = model.matrix(~0+GID:Env, data=dat_F,
xlev = list(GID=unique(dat_F$GID), Env = unique(dat_F$Env)))
K_LE = Z_LE%*%kronecker(diag(I), G)%*%t(Z_LE)
ETA = list(list(model='FIXED', X=XE),
list(model='RKHS', K=K_L, df0 = v_g, S0 = S_g, R2 = 1-R^2)),
list(model='RKHS', K=K_LE))
A = BGLR(y, ETA=ETA, nIter = 1e4, burnIn = 1e3, S0 = S, df0 = v, R2 = R^2)

```

where **dat_F** is the data set that contains the necessary phenotypic information (GID: Lines or individuals; Env: Environment; y: response variable of trait under study).

Example 2 (Includes Models with Only Env Effects and Models with Env and LinexEnv Effects) To illustrate how to fit the extended genomic regression models described before, here we consider the prediction of grain yield (tons/ha) based on marker information and the genomic relationship derived from it. The data set used consists of 30 lines in four environments, and the genotyped information of 500 markers for each line. The performance prediction of 18 models was evaluated with a five-fold cross-validation, where 80% of the complete data set was used to fit the model and the rest to evaluate the model in terms of the mean squared error of prediction (MSE). These models were obtained by considering different predictors (marker, environment, or/and environment–marker interaction) and different prior models to the parameters of each predictor included.

The model M1 only considered in the predictor the marker effects, from which six sub-models were obtained by adopting one of the six options (BRR, RKHS, BayesA, BayesB, BayesC, or BL) to the prior model of β (or g). Model M2 is model M1 plus the environment effects with a FIXED prior model, for all prior sub-models in the marker predictor. Model M3 is model M2 plus the environment–marker interaction, with a prior model of the same family as those chosen for the marker predictor (see in Table 6.2 all the models we compared).

The performance prediction of the models presented in Table 6.2 is shown in Table 6.3. The first column represents the kind of prior model used in both marker effects and env:marker interaction terms, when the latter is included in the model. In each of the first five prior models, model M2 resulted in better MSE performance, while when the BL prior model was used, model M3, the model with the interaction term, was better. The greater difference is between M1 and M2, where the average MSE across all priors of the first model is approximately 21% greater than the corresponding average of the M2 model. Similar behavior was observed with Pearson's correlation, with the average of this criterion across all priors about 32%

Table 6.2 Fitted models: M_{md} , $m = 1, 2, 3$, $d = 1, \dots, 6$

Model	Sub-model	Predictor		
		Marker or G: $X\beta$	Environment: $X_E\beta_E$	Interaction Env-Markers: $X_{EM}\beta_{EM}$
M1	M11	BRR		
	M12	RKHS		
	M12	BayesA		
	M14	BayesB		
	M15	BayesC		
	M16	BL		
M2	M21	BRR	FIXED	
	M22	RKHS	FIXED	
	M23	BayesA	FIXED	
	M24	BayesB	FIXED	
	M25	BayesC	FIXED	
	M26	BL	FIXED	
M3	M31	BRR	FIXED	BRR
	M32	RKHS	FIXED	RKHS
	M33	BayesA	FIXED	BayesA
	M34	BayesB	FIXED	BayesB
	M35	BayesC	FIXED	BayesC
	M36	BL	FIXED	BL

The third to fifth columns are the considered predictors corresponding to marker (genetic effect), environment, and environment–marker (genetic effect) effects, respectively. The cells in row 2 and column 3 indicate the prior distribution model specified for the beta parameters in each predictor when that parameter is present. M1 : $Y = \mathbf{1}\mu + X\beta + \epsilon$, M2 : $Y = \mathbf{1}\mu + X_E\beta_E + X\beta + \epsilon$, and M3 : $Y = \mathbf{1}\mu + X_E\beta_E + X\beta + X_{EM}\beta_{EM} + \epsilon$. For the RKHS prior model, predictors $X\beta$ and $X_{EM}\beta_{EM}$ are replaced by g and Eg , respectively

greater in model M2 than in M1. So the inclusion of the environment effect was important, but not the environment:marker interaction.

The best prediction performance in terms of MSE was obtained with sub-model M25 (M2 with a BayesC prior) followed by M21 (M2 with a BRR prior). However, the difference between those and sub-models M22, M23, and M24, also derived from M2, is only slight and a little more than with M26, which as commented before, among the models that assume a BL prior, showed a worse performance than M36 (M3 plus a BL prior for marker effects and environment–marker interaction).

6.8 Bayesian Genomic Multi-trait Linear Regression Model

The univariate models described for continuous outcomes do not exploit the possible correlation between traits, when the selection of better individuals is based on several traits and these univariate models are fitted separately to each trait. Relative

Table 6.3 Performance prediction of models in Table 6.2: Mean squared error of prediction (MSE) and average Pearson's correlation (PC), each with its standard deviation across the five partitions

	Markers			Env + Markers			Env + Markers + Env:Markers		
	M1			M2			M3		
	MSE (SD)	PC (SD)		MSE (SD)	PC (SD)		MSE (SD)	PC (SD)	
Prior									
BRR	0.5384 (0.08)	0.4678 (0.11)		0.4446 (0.11)	0.6112 (0.1)		0.4782 (0.12)	0.579 (0.11)	
RKHS	0.5509 (0.09)	0.4505 (0.13)		0.4542 (0.12)	0.6012 (0.1)		0.51 (0.13)	0.5271 (0.12)	
BayesA	0.5441 (0.08)	0.4622 (0.12)		0.4452 (0.11)	0.616 (0.1)		0.4806 (0.15)	0.5964 (0.11)	
BayesB	0.5462 (0.08)	0.4595 (0.12)		0.4455 (0.11)	0.6141 (0.1)		0.4669 (0.13)	0.5981 (0.11)	
BayesC	0.5449 (0.08)	0.4615 (0.12)		0.4416 (0.11)	0.614 (0.1)		0.474 (0.12)	0.5864 (0.11)	
BL	0.5621 (0.1)	0.4529 (0.12)		0.4719 (0.14)	0.5871 (0.09)		0.4537 (0.12)	0.606 (0.1)	

The first column represents the prior used in the model in both marker main effects and env:marker interaction term (when included in the model)

to this, some advantages of jointly modelling the multi-traits is that in this way the correlation among the traits is appropriately accounted for and can help to improve statistical power and the precision of parameter estimation, which are very important in genomic selection, because they can help to improve prediction accuracy and reduce trait selection bias (Schaeffer 1984; Pollak et al. 1984; Montesinos-López et al. 2018).

An example of this is when crop breeders collect phenotypic data for multiple traits such as grain yield and its components (grain type, grain weight, biomass, etc.), tolerance to biotic and abiotic stresses, and grain quality (taste, shape, color, nutrient, and/or content) (Montesinos-López et al. 2016). In this and many other cases, sometimes the interest is to predict traits that are difficult or expensive to measure with those that are easy to measure or the aim can be to improve all these correlated traits simultaneously (Henderson and Quaas 1976; Calus and Veerkamp 2011; Jiang et al. 2015). In these lines, there is evidence of the usefulness of multi-trait modelling. For example, Jia and Jannink (2012) showed that, compared to single-trait modelling, the prediction accuracy of low-heritability traits could be increased by using a multi-trait model when the degree of correlation between traits is at least moderate. They also found that multi-trait models had better prediction accuracy when phenotypes were not available on all individuals and traits. Joint modelling also has been found useful for increasing prediction accuracy when the traits of interest are not measured in the individuals of the testing set, but this and other traits were observed in individuals in the training set (Pszczola et al. 2013).

6.8.1 Genomic Multi-trait Linear Model

The genomic multi-trait linear model adopts a univariate genomic linear model structure for each trait but with correlated residuals and genotypic effects for traits in the same individual. Assuming that for individual j , n_T traits (Y_{jt} , $t = 1, \dots, n_T$) are measured, this model assumes that

$$\begin{bmatrix} Y_{j1} \\ Y_{j2} \\ \vdots \\ Y_{jn_T} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{n_T} \end{bmatrix} + \begin{bmatrix} \mathbf{x}_j^T \boldsymbol{\beta}_1 \\ \mathbf{x}_j^T \boldsymbol{\beta}_2 \\ \vdots \\ \mathbf{x}_j^T \boldsymbol{\beta}_{n_T} \end{bmatrix} + \begin{bmatrix} g_{j1} \\ g_{j2} \\ \vdots \\ g_{jn_T} \end{bmatrix} + \begin{bmatrix} \epsilon_{j1} \\ \epsilon_{j2} \\ \vdots \\ \epsilon_{jn_T} \end{bmatrix},$$

where μ_t , $t = 1, \dots, n_T$, are the specific trait intercepts, \mathbf{x}_j is a vector of covariates equal for all traits, g_{jt} , $t = 1, \dots, n_T$, are the specific trait genotype effects, and ϵ_{jt} , $t = 1, \dots, n_T$ are the random error terms corresponding to each trait. In matrix notation, it can be expressed as

$$\mathbf{Y}_j = \boldsymbol{\mu} + \mathbf{B}^T \mathbf{x}_j + \mathbf{g}_j + \boldsymbol{\epsilon}_j, \quad (6.8)$$

where $\mathbf{Y}_j = [Y_{j1}, \dots, Y_{jn_T}]^T$, $\boldsymbol{\mu} = [\mu_1, \dots, \mu_{n_T}]^T$, $\mathbf{B} = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{n_T}]$, $\mathbf{g}_j = [g_{j1}, \dots, g_{jn_T}]^T$, and $\boldsymbol{\epsilon}_j = [\epsilon_j, \dots, \epsilon_{n_T}]^T$. The residual vectors are assumed to be independent with multivariate normal distribution, that is $\boldsymbol{\epsilon}_j \sim N_{n_T}(\mathbf{0}, \mathbf{R})$, and all the random genotype effects are assumed to be $\mathbf{g} = [g_1^T, \dots, g_{n_T}^T]^T \sim N(\mathbf{0}, \mathbf{G} \otimes \boldsymbol{\Sigma}_T)$, \otimes being the Kronecker product. For a full Bayesian specification of this model, we suppose that $\boldsymbol{\beta} = \text{vec}(\mathbf{B}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_\beta)$, that is, marginally, for the fixed effect of each trait, a prior multivariate normal distribution is adopted, $\boldsymbol{\beta}_t \sim N_p(\boldsymbol{\beta}_{t0}, \boldsymbol{\Sigma}_{\beta_t})$, $t = 1, \dots, n_T$; a flat prior for the intercepts, $f(\boldsymbol{\mu}) \propto 1$; and independent inverse Wishart distributions for the covariance matrix of residuals \mathbf{R} and for $\boldsymbol{\Sigma}_T$, that is, $\boldsymbol{\Sigma}_T \sim \text{IW}(v_t, \mathbf{S}_t)$ and $\mathbf{R} \sim \text{IW}(v_R, \mathbf{S}_R)$.

Putting all the information together where the measured traits of each individual (\mathbf{Y}_j) are accommodated in the rows of a matrix response (\mathbf{Y}), model (6.8) can be expressed as

$$\mathbf{Y} = \mathbf{1}_J \boldsymbol{\mu}^T + \mathbf{X}\mathbf{B} + \mathbf{Z}_1 \mathbf{b}_1 + \mathbf{E}, \quad (6.9)$$

where $\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_J]^T$, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_J]^T$, $\mathbf{b}_1 = [g_1, \dots, g_J]^T$, and $\mathbf{E} = [\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_J]^T$. Note that under this notation, $\mathbf{E}^T \sim MN_{n_T \times J}(\mathbf{0}, \mathbf{R}, \mathbf{I}_J)$ or equivalently $\mathbf{E} \sim MN_{J \times n_T}(\mathbf{0}, \mathbf{I}_J, \mathbf{R})$, and $\mathbf{b}_1^T \sim MN_{n_T \times J}(\mathbf{0}, \boldsymbol{\Sigma}_T, \mathbf{G})$ or $\mathbf{b}_1 \sim MN_{J \times n_T}(\mathbf{0}, \mathbf{G}, \boldsymbol{\Sigma}_T)$. Here $\mathbf{Z} \sim MN_{J \times n_T}(\mathbf{M}, \mathbf{U}, \mathbf{V})$ means that the random matrix \mathbf{Z} follows the matrix variate normal distribution with parameters \mathbf{M} , \mathbf{U} , and \mathbf{V} , or equivalently, that the Jn_T random vector $\text{vec}(\mathbf{Z})$ is distributed as $N_{Jn_T}(\text{vec}(\mathbf{M}), \mathbf{V} \otimes \mathbf{U})$, with $\text{vec}(\cdot)$ denoting the vectorization of a matrix that stacks the columns of this in a single column. Note that when $\boldsymbol{\Sigma}_T$ and \mathbf{R} are diagonal matrices, model (6.9) is equivalent to separately fitting a univariate GBLUP model to each trait.

The conditional distribution of all traits is given by

$$\begin{aligned} f(\mathbf{Y} | \boldsymbol{\mu}, \boldsymbol{\beta}, \mathbf{b}_1, \boldsymbol{\Sigma}_T, \mathbf{R}) &= \frac{|\mathbf{R}|^{-\frac{J}{2}}}{(2\pi)^{Jn_T}} \exp \left\{ -\frac{1}{2} \text{tr} \left[\mathbf{R}^{-1} (\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^T - \mathbf{X}\mathbf{B} - \mathbf{Z}_1 \mathbf{b}_1)^T \mathbf{I}_J (\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^T - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_1 \mathbf{b}_1) \right] \right\} \\ &= \frac{|\mathbf{R}|^{-\frac{J}{2}}}{(2\pi)^{Jn_T}} \exp \left\{ -\frac{1}{2} [\text{vec}(\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^T - \mathbf{X}\mathbf{B} - \mathbf{Z}_1 \mathbf{b}_1)]^T (\mathbf{R}^{-1} \otimes \mathbf{I}_J) [\text{vec}(\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^T - \mathbf{X}\mathbf{B} - \mathbf{Z}_1 \mathbf{b}_1)] \right\} \end{aligned}$$

and the joint posterior of parameters $\boldsymbol{\mu}$, \mathbf{B} , \mathbf{b}_1 , $\boldsymbol{\Sigma}_T$, and \mathbf{R} is given by

$$f(\boldsymbol{\mu}, \mathbf{B}, \mathbf{b}_1, \boldsymbol{\Sigma}_T, \mathbf{R} | \mathbf{Y}) \propto f(\mathbf{Y} | \boldsymbol{\mu}, \mathbf{B}, \mathbf{b}_1, \boldsymbol{\Sigma}_T, \mathbf{R}) f(\mathbf{b}_1 | \boldsymbol{\Sigma}_T) f(\boldsymbol{\Sigma}_T) f(\boldsymbol{\beta}) f(\mathbf{R}),$$

where $f(\mathbf{b}_1 | \boldsymbol{\Sigma}_T)$ denotes the conditional distribution of the genotype effects, and $f(\boldsymbol{\Sigma}_T)$, $f(\boldsymbol{\beta})$, and $f(\mathbf{R})$ denote the prior density distribution of $\boldsymbol{\Sigma}_T$, \mathbf{B} , and \mathbf{R} , respectively. This joint posterior distribution of the parameters doesn't have closed form; for this reason, next are derived the full conditional distributions for Gibbs sampler implementation.

Let $\boldsymbol{\beta}_0$ and $\boldsymbol{\Sigma}_\beta$ be the prior mean and variance of $\boldsymbol{\beta} = \text{vec}(\mathbf{B})$. Because $\text{tr}(\mathbf{A}\mathbf{B}) = \text{vec}(\mathbf{A}^\top)^\top \text{vec}(\mathbf{B}) = \text{vec}(\mathbf{B})^\top \text{vec}(\mathbf{A}^\top)$ and $\text{vec}(\mathbf{A}\mathbf{X}\mathbf{B}) = (\mathbf{B}^\top \otimes \mathbf{A})\text{vec}(\mathbf{X})$, we have that

$$\begin{aligned} P(\boldsymbol{\beta}|-) &\propto f(\mathbf{Y}|\boldsymbol{\mu}, \mathbf{B}, \mathbf{b}_1, \boldsymbol{\Sigma}_T, \mathbf{R})f(\boldsymbol{\beta}) \\ &\propto \exp \left\{ -\frac{1}{2} [\text{vec}(\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^\top - \mathbf{Z}_1 \mathbf{b}_1) - (\mathbf{I}_{n_T} \otimes \mathbf{X})\text{vec}(\mathbf{B})]^\top (\mathbf{R}^{-1} \otimes \mathbf{I}_J) \right. \\ &\quad \left. \times [\text{vec}(\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^\top - \mathbf{Z}_1 \mathbf{b}_1) - (\mathbf{I}_{n_T} \otimes \mathbf{X})\text{vec}(\mathbf{B})] - \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \boldsymbol{\Sigma}_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} [\text{vec}(\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^\top - \mathbf{Z}_1 \mathbf{b}_1) - (\mathbf{I}_{n_T} \otimes \mathbf{X})\boldsymbol{\beta}]^\top (\mathbf{R}^{-1} \otimes \mathbf{I}_J) \right. \\ &\quad \left. \times [\text{vec}(\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^\top - \mathbf{Z}_1 \mathbf{b}_1) - (\mathbf{I}_{n_T} \otimes \mathbf{X})\boldsymbol{\beta}] - \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \boldsymbol{\Sigma}_\beta^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} [\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_0]^\top \tilde{\boldsymbol{\Sigma}}_\beta^{-1} [\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_0] \right\}, \end{aligned}$$

where $\tilde{\boldsymbol{\Sigma}}_\beta = [\boldsymbol{\Sigma}_\beta^{-1} + (\mathbf{R}^{-1} \otimes \mathbf{X}^\top \mathbf{X})]^{-1}$ and $\tilde{\boldsymbol{\beta}}_0 = \tilde{\boldsymbol{\Sigma}}_\beta [\boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\beta}_0 + (\mathbf{R}^{-1} \otimes \mathbf{X}^\top) \text{vec}(\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^\top - \mathbf{Z}_1 \mathbf{b}_1)]$. So, the full conditional distribution of $\boldsymbol{\beta}$ is $N_p(\tilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{\Sigma}}_\beta)$. Similarly, the full conditional distribution of $\mathbf{g} = \text{vec}(\mathbf{b}_1)$ is $N_J(\tilde{\mathbf{g}}, \tilde{\mathbf{G}})$, with $\tilde{\mathbf{G}} = [(\boldsymbol{\Sigma}_T^{-1} \otimes \mathbf{G}^{-1}) + (\mathbf{R}^{-1} \otimes \mathbf{Z}_1^\top \mathbf{Z}_1)]^{-1}$ and $\tilde{\mathbf{g}} = \tilde{\mathbf{G}}(\mathbf{R}^{-1} \otimes \mathbf{Z}_1^\top) \text{vec}(\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^\top - \mathbf{X}\mathbf{B})$. Now, because $\text{vec}(\mathbf{1}_J \boldsymbol{\mu}^\top) = (\mathbf{I}_{n_T} \otimes \mathbf{1}_J) \boldsymbol{\mu}$, similarly as before, the full conditional of $\boldsymbol{\mu}$ is $N_{n_T}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}_\mu)$, where $\tilde{\boldsymbol{\Sigma}}_\mu = \mathbf{J}^{-1} \mathbf{R}$ and $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\Sigma}}_\mu (\mathbf{R}^{-1} \otimes \mathbf{1}_J) \text{vec}(\mathbf{Y} - \mathbf{X}\mathbf{B} - \mathbf{Z}_1 \mathbf{B})$.

The full conditional distribution of $\boldsymbol{\Sigma}_T$

$$\begin{aligned} P(\boldsymbol{\Sigma}_T|-) &\propto P(\mathbf{b}_1|\boldsymbol{\Sigma}_T)P(\boldsymbol{\Sigma}_T) \\ &\propto |\boldsymbol{\Sigma}_T|^{-\frac{J}{2}} |\mathbf{G}|^{-\frac{n_T}{2}} \exp \left\{ -\frac{1}{2} \text{tr}[\mathbf{b}_1^\top \mathbf{G}^{-1} \mathbf{b}_1 \boldsymbol{\Sigma}_T^{-1}] \right\} P(\boldsymbol{\Sigma}_T) \\ &\propto \boldsymbol{\Sigma}_T^{-\frac{v_T + J + n_T + 1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{b}_1^\top \mathbf{G}^{-1} \mathbf{b}_1 + \mathbf{S}_T) \boldsymbol{\Sigma}_T^{-1} \right\}. \end{aligned}$$

From here we have that $\boldsymbol{\Sigma}_T | - \sim \text{IW}(v_T + J, \mathbf{b}_1^\top \mathbf{G}^{-1} \mathbf{b}_1 + \mathbf{S}_T)$. Now, because

$$\begin{aligned} P(\mathbf{R}|-) &\propto f(\mathbf{Y}|\boldsymbol{\mu}, \mathbf{B}, \mathbf{b}_1, \boldsymbol{\Sigma}_T, \mathbf{R})f(\mathbf{R}) \\ &\propto |\mathbf{R}|^{-\frac{J}{2}} \exp \left\{ -\frac{1}{2} \text{tr}[\mathbf{R}^{-1} (\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^\top - \mathbf{X}\mathbf{B} - \mathbf{Z}_1 \mathbf{b}_1)^\top \mathbf{I}_J (\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^\top - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_1 \mathbf{b}_1)] \right\} \\ &\quad |\mathbf{R}|^{-\frac{v_R + n_T + 1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{S}_T \mathbf{R}^{-1}) \right\} \\ &\propto |\mathbf{R}|^{-\frac{v_R + J + n_T + 1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}[\mathbf{S}_T + (\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^\top - \mathbf{X}\mathbf{B} - \mathbf{Z}_1 \mathbf{b}_1)^\top (\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^\top - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_1 \mathbf{b}_1)] \mathbf{R}^{-1} \right\} \end{aligned}$$

the full conditional distribution of \mathbf{R} is $\text{IW}(\tilde{v}_R, \tilde{\mathbf{S}}_R)$, where $\tilde{v}_R = v_R + J$ and $\tilde{\mathbf{S}}_R = \mathbf{S}_T + (\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^\top - \mathbf{X}\mathbf{B} - \mathbf{Z}_1 \mathbf{b}_1)^\top (\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^\top - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_1 \mathbf{b}_1)$.

In summary, a Gibbs sampler exploration of the joint posterior distribution of $\boldsymbol{\mu}$, $\boldsymbol{\beta}$, \mathbf{g} , $\boldsymbol{\Sigma}_T$, and \mathbf{R} can be done with the following steps:

1. Simulate $\boldsymbol{\beta}$ from a multivariate normal distribution $N_p(\tilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{\Sigma}}_\beta)$, where $\tilde{\boldsymbol{\Sigma}}_\beta = \left[\boldsymbol{\Sigma}_\beta^{-1} + (\mathbf{R}^{-1} \otimes \mathbf{X}^T \mathbf{X}) \right]^{-1}$ and $\tilde{\boldsymbol{\beta}}_0 = \tilde{\boldsymbol{\Sigma}}_\beta \left[\boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\beta}_0 + (\mathbf{R}^{-1} \otimes \mathbf{X}^T) \text{vec}(\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^T - \mathbf{Z}_1 \mathbf{b}_1) \right]$.
2. Simulate $\boldsymbol{\mu}$ from $N_{n_T}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}_\mu)$, where $\tilde{\boldsymbol{\Sigma}}_\mu = \mathbf{J}^{-1} \mathbf{R}$ and $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\Sigma}}_\mu (\mathbf{R}^{-1} \otimes \mathbf{1}_J) \text{vec}(\mathbf{Y} - \mathbf{X} \boldsymbol{\beta} - \mathbf{Z}_1 \mathbf{b}_1)$.
3. Simulate $\mathbf{g} = \text{vec}(\mathbf{b}_1)$ from $N_J(\tilde{\mathbf{g}}, \tilde{\mathbf{G}})$, where $\tilde{\mathbf{G}} = [(\boldsymbol{\Sigma}_T^{-1} \otimes \mathbf{G}^{-1}) + (\mathbf{R}^{-1} \otimes \mathbf{Z}_1^T \mathbf{Z}_1)]^{-1}$ and $\tilde{\mathbf{g}} = \tilde{\mathbf{G}} (\mathbf{R}^{-1} \otimes \mathbf{Z}_1^T) \text{vec}(\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^T - \mathbf{X} \boldsymbol{\beta})$.
4. Simulate $\boldsymbol{\Sigma}_T$ from $\text{IW}(v_T + J, \mathbf{b}_1^T \mathbf{G}^{-1} \mathbf{b}_1 + \mathbf{S}_T)$.
5. Simulate \mathbf{R} from $\text{IW}(\tilde{v}_R, \tilde{\mathbf{S}}_R)$, where $\tilde{v}_R = v_R + J$ and $\tilde{\mathbf{S}}_R = \mathbf{S}_T + (\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^T - \mathbf{X} \boldsymbol{\beta} - \mathbf{Z}_1 \mathbf{b}_1)^T (\mathbf{Y} - \mathbf{1}_J \boldsymbol{\mu}^T - \mathbf{X} \boldsymbol{\beta} - \mathbf{Z}_1 \mathbf{b}_1)$.
6. Return to step 1 or terminate when chain length is adequate to meet convergence diagnostics and the required sample size is reached.

An implementation of this model can be done using the github version of the BGLR R library, which can be accessed from <https://github.com/gdlc/BGLR-R> and can be installed directly in the R console by running the following commands: `install.packages('devtools'); library(devtools); install_git('https://github.com/gdlc/BGLR-R')`. This implementation also uses a flat prior for the fixed effect regression coefficients $\boldsymbol{\beta}$, and in such a case, the corresponding full conditional of this parameter is the same as step 1 of the Gibbs sampler given before, but removing $\boldsymbol{\Sigma}_\beta^{-1}$ and $\boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\beta}_0$ from $\tilde{\boldsymbol{\Sigma}}_\beta$ and $\tilde{\boldsymbol{\beta}}_0$, respectively. Specifically, model (6.8) can be implemented with this version of the BGLR package as follows:

```
ETA = list( list( X=X, model='FIXED' ), list( K=Z1GZ1T, model='RKHS' ) )
A = Multitrait(y=Y, ETA=ETA, resCov=list( type='UN', S0=S_R, df0=v_R ), nIter=nI, burnIn=nb)
```

The first argument in the Multitrait function is the response variable which many times is a phenotype matrix where each row corresponds to the measurement of n_T traits in each individual. The second argument is a list predictor in which the first sub-list specifies the design matrix and prior model to the fixed effects part of the predictor in model (6.9), and in the second sub-list are specified the parameters of the distribution of random genetic effects of \mathbf{b}_1 , where it is specified the $K = \mathbf{G}$ genomic relationship matrix, that accounts for the similarity between individuals based on marker information, $\text{df0} = v_T$ and $S0 = \mathbf{S}_T$ are the degrees of freedom parameter (v_T) and the scale matrix parameter (\mathbf{S}_T) of the inverse Wishart prior distribution for $\boldsymbol{\Sigma}_T$, respectively. In the third argument (resCOV), $S0$ and df0 are the scale matrix parameter (\mathbf{S}_R) and the degree of freedom parameter (v_R) of the inverse Wishart

prior distribution for \mathbf{R} . The last two arguments are the required number of iterations (\mathbf{nI}) and the burn-in period (\mathbf{nb}) for running the Gibbs sampler.

Similarly to the univariate case, model (6.9) can be equivalently described and implemented as a multivariate Ridge regression model, as follows:

$$\mathbf{Y} = \mathbf{1}_J \boldsymbol{\mu}^T + \mathbf{X}\mathbf{B} + \mathbf{X}_1 \mathbf{B}_1 + \mathbf{E}, \quad (6.10)$$

where $\mathbf{X}_1 = \mathbf{Z}_1 \mathbf{L}_G$, $\mathbf{G} = \mathbf{L}_G \mathbf{L}_G^T$ is the Cholesky factorization of \mathbf{G} , $\mathbf{B}_1 = \mathbf{L}_G^{-1} \mathbf{b}_1 \sim MN_{J \times n_T}(\mathbf{0}, \mathbf{I}_J, \boldsymbol{\Sigma}_T)$, and the specifications for the rest of parameters and prior distribution are the same as given in model (6.8). A Gibbs sampler implementation of this model is very similar to the one described before, with little modification. Indeed, a Gibbs implementation with the same multi-trait function is as follows:

```

L_G = t(chol(G))
X_1 = Z_1 L_G
ETA = list(list(X=X, model='FIXED'), list(X=X_1, model='BRR'))
A = Multitrait(y=Y, ETA=ETA, resCov=list(type='UN', S0=S_R, df0=
v_R), nIter=nI, burnIn=nb)

```

with the only change in the second sub-list predictor, where now the design matrix \mathbf{X}_1 and the Ridge regression model (BRR) are specified.

Example 3 To illustrate the performance in terms of the prediction power of these models and how to implement this in R software, we considered a reduced data set that consisted of 50 wheat lines grown in two environments. In each individual, two traits were measured: FLRSDS and MIXTIM. The evaluation was done with a five-fold cross-validation, where lines were evaluated in some environments with all traits but are missing for all traits in other environments. Model (6.9) was fitted and the environment effect was assumed a fixed effect.

The results are shown in Table 6.4, where the average (standard deviation) of two performance criteria is shown for each trait in each environment: average Pearson's correlation (PC) and the mean squared error of prediction (MSE). Table 6.4 shows good correlation performance in both traits and in both environments, and better predictions were obtained in environment 2 with both criteria. The magnitude of the

Table 6.4 Average Pearson's correlation (PC) and mean squared error of prediction (MSE) between predicted and observed values across five random partitions where lines were evaluated in some environments with all traits but are missing for all traits in other environments, SD represent standard deviation across partitions

Env	Trait	PC (SD)	MSE (SD)
1	FLRSDS	0.6252 (0.276)	4.3009 (2.343)
	MIXTIM	0.6709 (0.367)	0.457 (0.343)
2	FLRSDS	0.6895 (0.219)	3.9553 (2.171)
	MIXTIM	0.7523 (0.157)	0.3764 (0.249)

MSE in the first trait is mainly because the measurement scale is greater than in the second trait.

The R codes to reproduce these results (Table 6.4) are shown in Appendix 5.

6.9 Bayesian Genomic Multi-trait and Multi-environment Model (BMTME)

Model (6.9) does not take into account the possible trait–genotype–environment interaction ($T \times G \times E$), when environment information is available. An extension of this model is the one proposed by Montesinos-López et al. (2016), who added this interaction term to vary the specific trait genetic effects (\mathbf{g}_j) across environments. If the information of n_T traits of J lines is collected in I environments, this model is given by

$$\mathbf{Y} = \mathbf{1}_{IJ}\boldsymbol{\mu}^T + \mathbf{X}\mathbf{B} + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{Z}_2\mathbf{b}_2 + \mathbf{E}, \quad (6.11)$$

where $\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_{IJ}]^T$, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_{IJ}]^T$, \mathbf{Z}_1 and \mathbf{Z}_2 are the incident lines and the incident environment–line interaction matrices, $\mathbf{b}_1 = [\mathbf{g}_1, \dots, \mathbf{g}_J]^T$, $\mathbf{b}_2 = [\mathbf{g}_{21}, \dots, \mathbf{g}_{2IJ}]^T$, and $\mathbf{E} = [\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_{IJ}]^T$. Here, $\mathbf{b}_2 \mid \boldsymbol{\Sigma}_T, \boldsymbol{\Sigma}_E \sim MN_{IJ \times n_T}(\mathbf{0}, \boldsymbol{\Sigma}_E \otimes \mathbf{G}, \boldsymbol{\Sigma}_T)$, and similar to model (6.2), $\mathbf{b}_1 \mid \boldsymbol{\Sigma}_T \sim MN_{J \times n_T}(\mathbf{0}, \mathbf{G}, \boldsymbol{\Sigma}_T)$ and $\mathbf{E} \sim MN_{IJ \times n_T}(\mathbf{0}, \mathbf{I}_{IJ}, \mathbf{R})$. The complete Bayesian specification of this model assumes independent multivariate normal distributions for the columns of \mathbf{B} , that is, for the fixed effect of each trait a prior multivariate normal distribution is adopted, $\boldsymbol{\beta}_t \sim N_p(\boldsymbol{\beta}_{t0}, \boldsymbol{\Sigma}_{\beta_t})$, $t = 1, \dots, n_T$; a flat prior for the intercepts, $f(\boldsymbol{\mu}) \propto 1$; and independent inverse Wishart distributions for the covariance matrices of residuals \mathbf{R} and for $\boldsymbol{\Sigma}_T$, $\boldsymbol{\Sigma}_T \sim IW(v_T, \mathbf{S}_T)$ and $\mathbf{R} \sim IW(v_R, \mathbf{S}_R)$, and also an inverse Wishart distribution for $\boldsymbol{\Sigma}_E$, $\boldsymbol{\Sigma}_E \sim IW(v_E, \mathbf{S}_E)$.

The full conditional distributions of $\boldsymbol{\mu}$, \mathbf{B} , \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{R} can be derived as in model (6.9). The full conditional distribution of $\boldsymbol{\Sigma}_T$ is

$$\begin{aligned} f(\boldsymbol{\Sigma}_T | -) &\propto f(\mathbf{b}_1 | \boldsymbol{\Sigma}_T) P(\mathbf{b}_2 | \boldsymbol{\Sigma}_T, \boldsymbol{\Sigma}_E) P(\boldsymbol{\Sigma}_T) \\ &\propto |\boldsymbol{\Sigma}_T|^{-\frac{J}{2}} |\mathbf{G}|^{-\frac{IJ}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\mathbf{b}_1^T \mathbf{G}^{-1} \mathbf{b}_1 \boldsymbol{\Sigma}_T^{-1}] \right\} \\ &\quad \times |\boldsymbol{\Sigma}_T|^{-\frac{IJ}{2}} |\boldsymbol{\Sigma}_E \otimes \mathbf{G}|^{-\frac{n_T}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\mathbf{b}_2^T (\boldsymbol{\Sigma}_E^{-1} \otimes \mathbf{G}^{-1}) \mathbf{b}_2 \boldsymbol{\Sigma}_T^{-1}] \right\} P(\boldsymbol{\Sigma}_T) \\ &\propto \boldsymbol{\Sigma}_T^{\frac{v_T + J + IJ + n_T + 1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} (\mathbf{b}_1^T \mathbf{G}^{-1} \mathbf{b}_1 + \mathbf{b}_2^T (\boldsymbol{\Sigma}_E^{-1} \otimes \mathbf{G}^{-1}) \mathbf{b}_2 + \mathbf{S}_T) \boldsymbol{\Sigma}_T^{-1} \right\}, \end{aligned}$$

that is, $\boldsymbol{\Sigma}_T \mid - \sim IW(v_T + J + IJ, \mathbf{b}_1^T \mathbf{G}^{-1} \mathbf{b}_1 + \mathbf{b}_2^T (\boldsymbol{\Sigma}_E^{-1} \otimes \mathbf{G}^{-1}) \mathbf{b}_2 + \mathbf{S}_T)$.

Now, let be \mathbf{b}_2^* a $Jn_T \times I$ matrix such that $\text{vec}(\mathbf{b}_2^{\text{T}}) = \text{vec}(\mathbf{b}_2^*)$. Then because $\mathbf{b}_2^{\text{T}} \mid \Sigma_T, \Sigma_E \sim MN_{n_T \times IJ}(\mathbf{0}, \Sigma_T, \Sigma_E \otimes \mathbf{G})$, $\text{vec}(\mathbf{b}_2^{\text{T}}) \mid \Sigma_T, \Sigma_E \sim N(\mathbf{0}, \Sigma_E \otimes (\mathbf{G} \otimes \Sigma_T))$, and $\mathbf{b}_2^* \mid \Sigma_T, \Sigma_E \sim MN_{n_T \times IJ}(\mathbf{0}, \mathbf{G} \otimes \Sigma_T, \Sigma_E)$, the full conditional posterior distribution of Σ_E is

$$\begin{aligned} P(\Sigma_E | \text{ELSE}) &\propto P(\mathbf{b}_2 | \Sigma_E) P(\Sigma_E) \\ &\propto |\Sigma_E|^{-\frac{IJ}{2}} |\mathbf{G} \otimes \Sigma_T|^{-\frac{I}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma_E^{-1} \mathbf{b}_2^{*\text{T}} (\mathbf{G}^{-1} \otimes \Sigma_T^{-1}) \mathbf{b}_2^*] \right\} \\ &\quad |\mathbf{S}_E|^{\frac{v_E + I - 1}{2}} \times |\Sigma_E|^{-\frac{v_E}{2}} \exp \left\{ -\frac{1}{2} \text{tr} (\mathbf{S}_E \Sigma_E^{-1}) \right\} \\ &\propto |\Sigma_E|^{\frac{v_E + JL + I + 1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [(\mathbf{b}_2^{*\text{T}} (\mathbf{G}^{-1} \otimes \Sigma_T^{-1}) \mathbf{b}_2^* + \mathbf{S}_E)] \Sigma_E^{-1} \right\} \end{aligned}$$

which means that $\Sigma_E \mid - \sim \text{IW}(v_E + JL, \mathbf{b}_2^{*\text{T}} (\mathbf{G}^{-1} \otimes \Sigma_T^{-1}) \mathbf{b}_2^* + \mathbf{S}_E)$.

A Gibbs sampler to explore the joint posterior distribution of parameters of model (6.11), μ , β , $\mathbf{b}_1, \mathbf{b}_2, \Sigma_T, \Sigma_E$, and \mathbf{R} , can be implemented with the following steps:

1. Simulate β from a multivariate normal distribution $N_p(\tilde{\beta}_0, \tilde{\Sigma}_\beta)$, where $\tilde{\Sigma}_\beta = [\Sigma_\beta^{-1} + (\mathbf{R}^{-1} \otimes \mathbf{X}^{\text{T}} \mathbf{X})]^{-1}$ and $\tilde{\beta}_0 = \tilde{\Sigma}_\beta [\Sigma_\beta^{-1} \beta_0 + (\mathbf{R}^{-1} \otimes \mathbf{X}^{\text{T}}) \text{vec}(\mathbf{Y} - \mathbf{1}_{IJ} \mu^{\text{T}} - \mathbf{Z}_1 \mathbf{b}_1 - \mathbf{Z}_2 \mathbf{b}_2)]$.
2. Simulate μ from $N_{n_T}(\tilde{\mu}, \tilde{\Sigma}_\mu)$, where $\tilde{\Sigma}_\mu = (IJ)^{-1} \mathbf{R}$ and $\tilde{\mu} = \tilde{\Sigma}_\mu (\mathbf{R}^{-1} \otimes \mathbf{1}_{IJ}) \text{vec}(\mathbf{Y} - \mathbf{X} \mathbf{B} - \mathbf{Z}_1 \mathbf{b}_1 - \mathbf{Z}_2 \mathbf{b}_2)$.
3. Simulate $\mathbf{g}_1 = \text{vec}(\mathbf{b}_1)$ from $N_J(\tilde{\mathbf{g}}_1, \tilde{\mathbf{G}})$, where $\tilde{\mathbf{G}} = [(\Sigma_T^{-1} \otimes \mathbf{G}^{-1}) + (\mathbf{R}^{-1} \otimes \mathbf{Z}_1^{\text{T}} \mathbf{Z}_1)]^{-1}$ and $\tilde{\mathbf{g}} = \tilde{\mathbf{G}} (\mathbf{R}^{-1} \otimes \mathbf{Z}_1^{\text{T}}) \text{vec}(\mathbf{Y} - \mathbf{1}_{IJ} \mu^{\text{T}} - \mathbf{X} \mathbf{B} - \mathbf{Z}_2 \mathbf{b}_2)$.
4. Simulate $\mathbf{g}_2 = \text{vec}(\mathbf{b}_2)$ from $N_J(\tilde{\mathbf{g}}_2, \tilde{\mathbf{G}}_2)$, where $\tilde{\mathbf{G}}_2 = [(\Sigma_T^{-1} \otimes \Sigma_E^{-1} \otimes \mathbf{G}^{-1}) + (\mathbf{R}^{-1} \otimes \mathbf{Z}_2^{\text{T}} \mathbf{Z}_2)]^{-1}$ and $\tilde{\mathbf{g}}_2 = \tilde{\mathbf{G}}_2 (\mathbf{R}^{-1} \otimes \mathbf{Z}_2^{\text{T}}) \text{vec}(\mathbf{Y} - \mathbf{1}_{IJ} \mu^{\text{T}} - \mathbf{X} \mathbf{B} - \mathbf{Z}_1 \mathbf{b}_1)$.
5. Simulate Σ_T from $\text{IW}(v_T + J + IJ, \mathbf{b}_1^{\text{T}} \mathbf{G}^{-1} \mathbf{b}_1 + \mathbf{b}_2^{\text{T}} (\Sigma_E^{-1} \otimes \mathbf{G}^{-1}) \mathbf{b}_2 + \mathbf{S}_T)$.
6. Simulate Σ_E from $\text{IW}(v_E + JL, \mathbf{b}_2^{*\text{T}} (\mathbf{G}^{-1} \otimes \Sigma_T^{-1}) \mathbf{b}_2^* + \mathbf{S}_E)$.
7. Simulate \mathbf{R} from $\text{IW}(\tilde{v}_R, \tilde{\mathbf{S}}_R)$, where $\tilde{v}_R = v_R + IJ$ and $\tilde{\mathbf{S}}_R = \mathbf{S}_T + (\mathbf{Y} - \mathbf{1}_{IJ} \mu^{\text{T}} - \mathbf{X} \mathbf{B} - \mathbf{Z}_1 \mathbf{b}_1 - \mathbf{Z}_2 \mathbf{b}_2)^{\text{T}} (\mathbf{Y} - \mathbf{1}_{IJ} \mu^{\text{T}} - \mathbf{X} \mathbf{B} - \mathbf{Z}_1 \mathbf{b}_1 - \mathbf{Z}_2 \mathbf{b}_2)$.
8. Return to step 1 or terminate when chain length is adequate to meet convergence diagnostics and the required sample size is reached.

A similar Gibbs sampler is implemented in the BMTME R package, with the main difference, that this package does not allow specifying a general fixed effect design matrix \mathbf{X} , only the corresponding to the design matrix for the environment effects, and also the intercept vector μ is ignored because it is included in the fixed

environment effects. Specifically, to fit model (6.11) where the only fixed effects to be taken into account are the environment's effects, the R code to implement this with the BMTME package is as follows:

```

XE  = model.matrix(~Env, data=dat_F)
Z1  = model.matrix(~0+GID, data=dat_F)
Lg  = t(chol(G))
Z1_a = Z1%*%Lg
Z2  = model.matrix(~0+GID:Env, data=dat_F)
G2  = kronecker(diag(dim(XE)[2]), Lg)
Z2_a = Z2%*%G2
A = BMTME(Y = Y, X = XE, Z1 = Z1_a, Z2 = Z2_a, nIter = nI, burnIn = nb, thin =
2, bs = 50)

```

where **Y** is the matrix of response variables where each row corresponds to the measurement of n_T traits in each individual, **XE** is a design matrix for the environment effects, **Z1** is the incidence matrix of the genetics effects, **Z2** is the design matrix of the genetic–environment interaction effects, **nI** and **nb** are the required number of interactions and the burn-in period, and *bs* is the number of blocks to use internally to sample from $\text{vec}(\mathbf{b}_2)$.

Example 4 To illustrate how to implement this model with the BMTME R package, we considered the data in Example 2, but now the explored model includes the trait–genotype–environment interaction.

The average results of the prediction performance in terms of PC and MSE for implementing the same five-fold cross-validation used in Example 3 are shown in Table 6.5. These results show an improvement in terms of prediction performance with this model in all trait environments combinations and in both criteria (PC and MSE) to measure the prediction performance, except in trait MIXTIM and Env 2, where the MSE is slightly greater than the one obtained with model (6.9), which does not take into account the triple interaction (trait–genotype–environment).

The R code used to obtain these results is given in Appendix 5.

Table 6.5 Average Pearson's correlation (PC) and mean squared error of prediction (MSE) between predicted and observed values across five random partitions where lines were evaluated in some environments with all traits but are missing for all traits in other environments

Env	Trait	PC (SD)	MSE (SD)
1	FLRSDS	0.668 (0.243)	3.4483 (1.9)
	MIXTIM	0.6913 (0.35)	0.4255 (0.325)
2	FLRSDS	0.7218 (0.219)	3.2304 (2.034)
	MIXTIM	0.7667 (0.188)	0.3806 (0.287)

Appendix 1

- The probability density function (pdf) of the scaled inverse Chi-square distribution with ν degrees of freedom and scale parameter S , $\chi^{-2}(\nu, S)$, is given by

$$f(\sigma^2; \nu, S) = \frac{\left(\frac{S}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)(\sigma^2)^{1+\frac{\nu}{2}}} \exp\left(-\frac{S}{2\sigma^2}\right).$$

and the mean, mode, and variance of this distribution are given by $\frac{S}{\nu-2}$, $\frac{S}{\nu+2}$, and $\frac{2S^2}{(\nu-2)^2(\nu-4)}$, respectively.

- The pdf of the gamma distribution with shape parameter s and rate parameter r :

$$f_{S_p}(x; s, r) = \frac{r^s x^{s-1}}{\Gamma(s)} \exp(-rx).$$

The mean, mode, and variance of this distribution are s/r , $(s-1)/r$, and s/r^2 , respectively.

- The pdf of a beta distribution with mean μ and precision parameter ϕ (“dispersion” parameter ϕ^{-1}) is given by

$$f(x; \mu, \phi) = \frac{1}{B[\mu\phi, (1-\mu)\phi]} x^{\mu\phi-1} (1-x)^{(1-\mu)\phi-1},$$

where the relation with the standard parameterization of this distribution, Beta (α, β) , is

$$\mu = \frac{\alpha}{\alpha + \beta}, \phi = \alpha + \beta.$$

- The pdf of a Laplace distribution is

$$f(x; b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right), b > 0, -\infty \leq x \leq \infty.$$

The mean and variance of this distribution are 0 and $2b^2$.

- A random matrix $\mathbf{\Sigma}$ of dimension $p \times p$ is distributed as inverse Wishart distribution with parameter ν and \mathbf{S} , $\mathbf{\Sigma} \sim \text{IW}(\nu, \mathbf{S})$, if it has a density function

$$f(\mathbf{\Sigma}) = \frac{1}{2^{\frac{p\nu}{2}} \Gamma_p(\nu/2)} |\mathbf{S}|^{\frac{\nu}{2}} |\mathbf{\Sigma}|^{-\frac{\nu+p+1}{2}} \exp\left[-\frac{1}{2} \text{tr}(\mathbf{S}\mathbf{\Sigma}^{-1})\right],$$

where $\Gamma_p(\nu/2)$ is the multivariate gamma function, $\nu > 0$, and $\mathbf{\Sigma}$ and \mathbf{S} are positive defined matrices. The mean matrix of this distribution is $\frac{\mathbf{S}}{\nu-p-1}$.

- A $(p \times q)$ random matrix \mathbf{Z} follows the matrix normal distribution with matrix parameters \mathbf{M} ($p \times q$), \mathbf{U} ($p \times p$), and \mathbf{V} ($q \times q$), $\mathbf{Z} \sim MN_{p, q}(\mathbf{M}, \mathbf{U}, \mathbf{V})$, if it has density

$$f(\mathbf{Z}|\mathbf{M}, \mathbf{U}, \mathbf{V}) = \frac{\exp \left\{ -\frac{1}{2} \text{tr} \left[\mathbf{V}^{-1} (\mathbf{Z} - \mathbf{M})^T \mathbf{U}^{-1} (\mathbf{Z} - \mathbf{M}) \right] \right\}}{(2\pi)^{pq/2} |\mathbf{V}|^{p/2} |\mathbf{U}|^{q/2}}.$$

Appendix 2: Setting Hyperparameters for the Prior Distributions of the BRR Model

The following rules are those used in Pérez and de los Campos (2014), and provide proper but weakly informative prior distributions. In general, this consists of assigning a certain proportion of the total variance of the phenotypes, to the different components of the model.

Specifically, for model (6.3), first the total variance of y is partitioned into two components: (1) the error and (2) the linear predictor:

$$\text{Var}(y_j) = \text{Var}(\mathbf{x}_j^T \boldsymbol{\beta}_0) + \sigma^2$$

Therefore, the average of the variance of the individuals, called total variance, is equal to

$$\frac{1}{n} \sum_{j=1}^n \text{Var}(y_j) = \frac{1}{n} \sum_{j=1}^n \text{Var}(\mathbf{x}_j^T \boldsymbol{\beta}_0) + \sigma^2 = \frac{1}{n} \text{tr}(\mathbf{X}\mathbf{X}^T) \sigma_\beta^2 + \sigma^2 = V_M + V_\epsilon.$$

Then, by setting R_1^2 as the proportion of the total variance (\mathbf{V}_y), that is explained by markers a priori, $V_M = R_1^2 \mathbf{V}_y$, and replacing σ_β^2 in V_M by its prior mode, $\frac{S_\beta}{v_\beta + 2}$, we have that

$$\frac{1}{n} \text{tr}(\mathbf{X}\mathbf{X}^T) \left(\frac{S_\beta}{v_\beta + 2} \right) = R_1^2 \mathbf{V}_y.$$

From here, once we have set a value for v_β , the scale parameter is given by

$$S_\beta = \frac{R_1^2 \mathbf{V}_y}{\frac{1}{n} \text{tr}(\mathbf{X}\mathbf{X}^T)} (v_\beta + 2).$$

A commonly used value of the shape parameter is $\nu_\beta = 5$ and the value for the proportion of explained variances is $R_1^2 = 0.5$.

Because the model only has two predictors and R_1^2 was set as the proportion of the total variance that is explained by markers a priori, the corresponding proportion that is explained by error a priori is $R_2^2 = 1 - R_1^2$. Then, similar to what was done before, once there is a value for the shape parameter of the prior distribution of σ^2 , ν , the value of the scale parameter is given by

$$S = (1 - R_1^2) \mathbf{V}_y (\nu + 2).$$

By default, $\nu = 5$ is often used.

Appendix 3: R Code Example 1

```
rm(list=ls())
library(BGLR)
load('dat_ls_E1.RData', verbose=TRUE)
#Phenotypic data
dat_F = dat_ls$dat_F
head(dat_F)
#Marker data
dat_M = dat_ls$dat_M
dim(dat_M)

dat_F = transform(dat_F, GID = as.character(GID))
head(dat_F, 5)

#Matrix design of markers
Pos = match(dat_F$GID, row.names(dat_M))
XM = dat_M[Pos,]
XM = scale(XM)
dim(XM)

n = dim(dat_F)[1]
y = dat_F$y

#10 random partitions
K = 10
set.seed(1)
PT = replicate(K, sample(n, 0.20*n))

#BRR
ETA_BRR = list(list(model='BRR', X=XM))
Tab = data.frame(PT = 1:K, MSEP = NA)
set.seed(1)
for(k in 1:K)
{
```

```

  Pos_tst = PT[,k]
  y_NA = y
  y_NA[Pos_tst] = NA
  A = BGLR(y=y_NA,ETA=ETA_BRR,nIter = 1e4,burnIn = 1e3,verbose = FALSE)
  yp_ts = A$yHat[Pos_tst]
  Tab$MSEP[k] = mean((y[Pos_tst]-yp_ts)^2)
}

#GBLUP
dat_M = scale(dat_M)
G = tcrossprod(XM)/dim(XM)[2]
dim(G)
#Matrix design of GIDs
Z = model.matrix(~0+GID,data=dat_F,xlev = list(GID=unique
(dat_F$GID)))
K_L = Z%*%G%*%t(Z)
ETA_GB = list(list(model='RKHS',K = K_L))
#Tab = data.frame(PT = 1:K,MSEP = NA)
set.seed(1)
for(k in 1:K)
{
  Pos_tst = PT[,k]
  y_NA = y
  y_NA[Pos_tst] = NA
  A = BGLR(y=y_NA,ETA=ETA_GB,nIter = 1e4,burnIn = 1e3,verbose = FALSE)
  yp_ts = A$yHat[Pos_tst]
  Tab$MSEP_GB[k] = mean((y[Pos_tst]-yp_ts)^2)
}

#BA
ETA_BA = list(list(model='BayesA',X=XM))
#Tab = data.frame(PT = 1:K,MSEP = NA)
set.seed(1)
for(k in 1:K)
{
  Pos_tst = PT[,k]
  y_NA = y
  y_NA[Pos_tst] = NA
  A = BGLR(y=y_NA,ETA=ETA_BA,nIter = 1e4,burnIn = 1e3,verbose = FALSE)
  yp_ts = A$yHat[Pos_tst]
  Tab$MSEP_BA[k] = mean((y[Pos_tst]-yp_ts)^2)
}

#BB
ETA_BB = list(list(model='BayesB',X=XM))
#Tab = data.frame(PT = 1:K,MSEP = NA)
set.seed(1)
for(k in 1:K)
{
  Pos_tst = PT[,k]
  y_NA = y
  y_NA[Pos_tst] = NA
  A = BGLR(y=y_NA,ETA=ETA_BB,nIter = 1e4,burnIn = 1e3,verbose = FALSE)

```

```

yp_ts = A$yHat[Pos_tst]
Tab$MSEP_BB[k] = mean((y[Pos_tst]-yp_ts)^2)
}

#BC
ETA_BC = list(list(model='BayesC',X=XM))
#Tab = data.frame(PT = 1:K,MSEP = NA)
set.seed(1)
for(k in 1:K)
{
  Pos_tst = PT[,k]
  y_NA = y
  y_NA[Pos_tst] = NA
  A = BGLR(y=y_NA,ETA=ETA_BC,nIter = 1e4, burnIn = 1e3, verbose = FALSE)
  yp_ts = A$yHat[Pos_tst]
  Tab$MSEP_BC[k] = mean((y[Pos_tst]-yp_ts)^2)
}

#BL
ETA_BL = list(list(model='BL',X=XM))
#Tab = data.frame(PT = 1:K,MSEP = NA)
set.seed(1)
for(k in 1:K)
{
  Pos_tst = PT[,k]
  y_NA = y
  y_NA[Pos_tst] = NA
  A = BGLR(y=y_NA,ETA=ETA_BL,nIter = 1e4, burnIn = 1e3, verbose = FALSE)
  yp_ts = A$yHat[Pos_tst]
  Tab$MSEP_BL[k] = mean((y[Pos_tst]-yp_ts)^2)
}
Tab

#Mean and SD across the five partitions
apply(Tab[, -1], 2, function(x) c(mean(x), sd(x)))

```

Appendix 4: R Code Example 2

```

rm(list=ls())
library(BGLR)
library(BMTME)
load('dat_ls_E2.RData', verbose=TRUE)
#Phenotypic data
dat_F = dat_ls$dat_F
head(dat_F)
dim(dat_F)
#Marker data
dat_M = dat_ls$dat_M
dim(dat_M)

```



```

dat_F = transform(dat_F, GID = as.character(GID))
head(dat_F, 5)

#Matrix design for markers
Pos = match(dat_F$GID, row.names(dat_M))
XM = dat_M[Pos, ]
dim(XM)
XM = scale(XM)
#Environment design matrix
XE = model.matrix(~0+Env, data=dat_F) [, -1]
head(XE)
#Environment-marker design matrix
XEM = model.matrix(~0+XM:XE)
#GID design matrix and Environment-GID design matrix
#for RKHS models
Z_L = model.matrix(~0+GID, data=dat_F, xlev = list(GID=unique
(dat_F$GID)))
Z_LE = model.matrix(~0+GID:Env, data=dat_F,
xlev = list(GID=unique(dat_F$GID), Env = unique(dat_F$Env)))
#Genomic relationship matrix derived from markers
dat_M = scale(dat_M)
G = tcrossprod(dat_M)/dim(dat_M) [2]
dim(G)
#Covariance matrix for Zg
K_L = Z_L%*%G%*%t(Z_L)
#Covariance matrix for random effects ZEG
K_LE = Z_LE%*%kronecker(diag(4), G)%*%t(Z_LE)
n = dim(dat_F) [1]
y = dat_F$y

#Number of random partitions
K = 5
PT = CV.KFold(dat_F, DataSetID = 'GID', K=5, set_seed = 1)
Models = c('BRR', 'RKHS', 'BayesA', 'BayesB', 'BayesC', 'BL')
Tab = data.frame()
for(m in 1:6)
{
ETA1 = list(list(model=Models[m], X=XM))
ETA2 = list(list(model='FIXED', X=XE), list(model=Models[m], X=XM))
ETA3 = list(list(model='FIXED', X=XE), list(model=Models[m], X=XM),
list(model=Models[m], X=XEM))
if(Models[m] == 'RKHS')
{
ETA1 = list(list(model='RKHS', K=K_L))
ETA2 = list(list(model='FIXED', X=XE), list(model='RKHS', K=K_L))
ETA3 = list(list(model='FIXED', X=XE), list(model='RKHS', K=K_L),
list(model='RKHS', K=K_LE))
}
}
Tab1_m = data.frame(PT = 1:K, MSEP = NA)
set.seed(1)
Tab2_m = Tab1_m
Tab3_m = Tab2_m
for(k in 1:K)

```

```

{
  Pos_tst = PT$CrossValidation_list[[k]]
  y_NA = y
  y_NA[Pos_tst] = NA
  A = BGLR(y=y_NA, ETA=ETA1, nIter = 1e4, burnIn = 1e3, verbose = FALSE)
  yp_ts = A$yHat[Pos_tst]
  Tab1_m$MSEP[k] = mean((y[Pos_tst] - yp_ts)^2)
  Tab1_m$Cor[k] = cor(y[Pos_tst], yp_ts)

  A = BGLR(y=y_NA, ETA=ETA2, nIter = 1e4, burnIn = 1e3, verbose = FALSE)
  yp_ts = A$yHat[Pos_tst]
  Tab2_m$MSEP[k] = mean((y[Pos_tst] - yp_ts)^2)
  Tab2_m$Cor[k] = cor(y[Pos_tst], yp_ts)

  A = BGLR(y=y_NA, ETA=ETA3, nIter = 1e4, burnIn = 1e3, verbose = FALSE)
  yp_ts = A$yHat[Pos_tst]
  Tab3_m$MSEP[k] = mean((y[Pos_tst] - yp_ts)^2)
  Tab3_m$Cor[k] = cor(y[Pos_tst], yp_ts)
}
Tab = rbind(Tab, data.frame(Model=Models[m], Tab1_m, Tab2_m, Tab3_m))
}
Tab

```

Appendix 5

R Code Example 3

```

rm(list=ls(all=TRUE))
library(BGLR)
library(BMTME)
library(dplyr)

load('dat_ls.RData', verbose=TRUE)
dat_F = dat_ls$dat_F
head(dat_F)
Y = as.matrix(dat_F[, -(1:2)])
dat_F$Env = as.character(dat_F$DS)
G = dat_ls$G
J = dim(G)[1]
XE = matrix(model.matrix(~0+Env, data=dat_F)[, -1], nc=1)
Z = model.matrix(~0+GID, data=dat_F)
K = Z%*%G%*%t(Z)
#Partitions for a 5-FCV
PT_ls = CV.KFold(dat_FF, DataSetID='GID', K=5, set_seed = 123)
PT_ls = PT_ls$CrossValidation_list
#Predictor BGLR
ETA = list(list(X=XE, model='FIXED'), list(K=K, model='RKHS'))
#Function to summarize the performance prediction: PC_MM_f
source('PC_MM.R') #See below
Tab = data.frame()

```


R Code for Example 4

```

rm(list=ls(all=TRUE))
library(BMTME)
library(dplyr)
load('dat_ls.RData', verbose=TRUE)
dat_F = dat_ls$dat_F
head(dat_F)
Y = as.matrix(dat_F[, -(1:2)])
dat_F$Env = as.character(dat_F$DS)
G = dat_ls$G
Lg = t(chol(G))
XE = model.matrix(~Env, data=dat_F)
Z1 = model.matrix(~0+GID, data=dat_F)
Z1_a = Z1*%Lg
Z2 = model.matrix(~0+GID:Env, data=dat_F)
L2 = kronecker(diag(dim(XE)[2]), Lg)
Z2_a = Z2*%L2
#Partitions for a 5-FCV
PT_ls = CV.KFold(dat_FF, DataSetID='GID', K=5, set_seed = 123)
PT_ls = PT_ls$CrossValidation_list
source('PC_MM.R') #See file R "PC_MM.R" defined in example 6.1
Tab = data.frame()
set.seed(1)
for(p in 1:5)
{
  Y_NA = Y
  Pos_NA = PT_ls[[p]]
  Y_NA[Pos_NA,] = NA
  A = BMTME(Y = Y_NA, X = XE, Z1 = Z1_a, Z2 = Z2_a,
            nIter = 3e3, burnIn = 5e2, thin = 2, bs = 50)
  PC = PC_MM_f(Y[Pos_NA,], A$yHat[Pos_NA,], Env=dat_F$Env[Pos_NA])
  Tab = rbind(Tab, data.frame(PT=p, PC))
  cat('PT=', p, '\n')
}
Tab_R = Tab%>%group_by(Env, Trait)%>%select(Cor, MSEP)%>%summarise
(Cor_mean = mean(Cor),
                                     Cor_sd = sd(Cor),
                                     MSEP_mean = mean(MSEP),
                                     MSEP_sd = sd(MSEP))
Tab_R = as.data.frame(Tab_R)
Tab_R

```

References

- Box GEP, Tiao GC (1992) Bayesian inference in statistical analysis. Wiley, New York
- Calus MP, Veerkamp RF (2011) Accuracy of multi-trait genomic selection using different methods. *Genet Sel Evol* 43(1):26
- Casella G, George EI (1992) Explaining the Gibbs sampler. *Am Stat* 46(3):167–174
- Christensen R, Johnson W, Branscum A, Hanson TE (2011) Bayesian ideas and data analysis: an introduction for scientists and statisticians. Chapman & Hall/CRC, Stanford, CA
- de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193(2):327–345
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian data analysis. Chapman and Hall/CRC, Stanford, CA
- Habier D, Tetens J, Seefried FR, Lichtner P, Thaller G (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet Sel Evol* 42(1):5
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ (2011) Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12(1):186
- Henderson C (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31(2):423–447. <https://doi.org/10.2307/2529430>
- Henderson CR, Quaas RL (1976) Multiple trait evaluation using relative's records. *J Anim Sci* 43: 11–88
- Jia Y, Jannink JL (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192(4):1513–1522
- Jiang J, Zhang Q, Ma L, Li J, Wang Z, Liu JF (2015) Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. *Heredity* 115(1):29–36
- Lehermeier C, Wimmer V, Albrecht T, Auinger HJ, Gianola D, Schmid VJ, Schön CC (2013) Sensitivity to prior specification in Bayesian genome-based prediction models. *Stat Appl Genet Mol Biol* 12(3):375–391
- Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–1829
- Montesinos-López OA, Montesinos-López A, Crossa J, Toledo FH, Pérez-Hernández O, Eskridge KM, Rutkoski J (2016) A genomic Bayesian multi-trait and multi-environment model. *G3* 6(9):2725–2744
- Montesinos-López OA, Montesinos-López A, Montesinos-López JC, Crossa J, Luna-Vázquez FJ, Salinas-Ruiz J (2018) A Bayesian multiple-trait and multiple-environment model using the matrix normal distribution. In: Physical methods for stimulation of plant and mushroom development. IntechOpen, Croatia, p 19
- Park T, Casella G (2008) The Bayesian lasso. *J Am Stat Assoc* 103(482):681–686
- Pérez P, de los Campos G (2013) BGLR: a statistical package for whole genome regression and prediction. R package version 1(0.2)
- Pérez P, de los Campos G (2014) Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198(2):483–495
- Pérez P, de los Campos G, Crossa J, Gianola D (2010) Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. *Plant Genome* 3(2):106–116
- Pollak EJ, Van der Werf J, Quaas RL (1984) Selection bias and multiple trait evaluation. *J Dairy Sci* 67(7):1590–1595
- Pszczola M, Veerkamp RF, De Haas Y, Wall E, Strabel T, Calus MPL (2013) Effect of predictor traits on accuracy of genomic breeding values for feed intake based on a limited cow reference population. *Animal* 7(11):1759–1768
- Schaeffer LR (1984) Sire and cow evaluation under multiple trait models. *J Dairy Sci* 67(7):1567–1580
- VanRaden PM (2007) Genomic measures of relationship and inbreeding. *Interbull Bull* 7:33–36

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

