## Missing values

The FW function can take missing values, but OLS method and Bayesian method treated missing values differently. For OLS, only the observed values were used to fit the model, the $\hat{\boldsymbol{y}}$ for both the observed and the missing values were obtained as $\hat{y}_{ij} = \hat{\mu} + \hat{g}_i + \hat{h}_j + \hat{b}_i\hat{h}_j$. However, if a variety i is completely missing across all environments, $\hat{g}_i$ and $\hat{b}_i$ will be NA for the missing variety and the corresponding $\hat{y}_{ij}$ is NA; Similarly, if an environment j is completely missing in all the observed data, $\hat{h}_j$ will be NA for the missing environment and the corresponding $\hat{y}_{ij}$ will be NA.

For Gibbs method, missing values are not removed from the model fitting, but sampled from the corresponding full conditional density. If a variety i is completely missing, Gibbs sampler can still estimate its value of $\hat{g}_i$ and $\hat{b}_i$ based on information borrowed from other varieties with the covariance structure **A**. However, if the information of **A** is not available, Gibbs sampler will be dominated by the prior distribution, which will lead to $\hat{g}_i$ and $\hat{b}_i$ close to 0. Similar treatment was done for $\hat{h}_j$. The estimates of posterior means of $\hat{\boldsymbol{y}}$ were obtained by averaging over the samples.

## Hyper-parameter setup

In the main article, we mentioned that the prior distributions for $\sigma_\varepsilon^2$, $\sigma_g^2$, $\sigma_b^2$ and $\sigma_h^2$ are all set to be scaled-inverse $\chi^2$ distributions: $\sigma_\varepsilon^2 \sim (\nu_\varepsilon, S_\varepsilon^2)$, $\sigma_g^2 \sim \chi^{-2}(\nu_g, S_g^2), \sigma_b^2 \sim \chi^{-2}(\nu_b, S_b^2), \sigma_h^2 \sim \chi^{-2}(\nu_h, S_h^2)$. The set of hyper-parameter that control the shape of these prior distributions are the degrees of freedom ($\nu_\varepsilon$, $\nu_g$, $\nu_b$ and $\nu_h$) and scale parameters ($S_\varepsilon^2$, $S_g^2$, $S_b^2$ and $S_h^2$).

The default setup was similar as discussed in ( Pérez et al., 2010). Small values of degree of freedom ($\nu_\varepsilon$, $\nu_g$, $\nu_b$, $\nu_h$) for the scaled inverse chi-square distributions are generally selected. The default is $\nu_\varepsilon = \nu_g = \nu_b = \nu_h = 5$, but the user can choose any small positive numbers.

The scale parameter $S^2$ is set to meet the mode of density: mode $(\sigma^2) = \frac{\nu S^2}{\nu+2}$. If we have some prior estimate of $\sigma^2$ as $\hat{\sigma}^2$, we will have $S^2 = \hat{\sigma}^2 \frac{\nu+2}{\nu}$. By default, the FW package sets the prior estimates of $\sigma_\varepsilon^2$, $\sigma_g^2$, $\sigma_b^2$ and $\sigma_h^2$ as $\hat{\sigma}_\varepsilon^2 = 0.5\ \hat{\sigma}_p^2$, $\hat{\sigma}_g^2 = 0.25\hat{\sigma}_p^2$, $\hat{\sigma}_b^2 = \hat{\sigma}_h^2 = 0.5\ \hat{\sigma}_p$, where $\hat{\sigma}_p^2$ is the estimate of the total phenotypic variance. When assuming independent random effects, the variance of the interaction term will be $\sigma_b^2\sigma_h^2$. Therefore, setting $\hat{\sigma}_b^2 = \hat{\sigma}_h^2 = 0.5\ \hat{\sigma}_p$ will lead to a $0.25\hat{\sigma}_p^2$ interaction variance. The scale parameters can then be set up as $S_\varepsilon^2 = \frac{\hat{\sigma}_\varepsilon^2(\nu_\varepsilon+2)}{\nu_\varepsilon}$, $S_g^2 = \frac{\hat{\sigma}_g^2(\nu_g+2)}{\nu_g}$, $S_b^2 = \frac{\hat{\sigma}_b^2(\nu_b+2)}{\nu_b}$ and $S_h^2 = \frac{\hat{\sigma}_h^2(\nu_h+2)}{\nu_h}$.

The FW package provides arguments to setup hyper parameters: `dfe`, `dfg`, `dfb`, `dfh` will set up the degrees of freedom $\nu_\varepsilon$, $\nu_g$, $\nu_b$ and $\nu_h$ respectively; `priorVARe`, `priorVARg`, `priorVARb`, `priorVARh` will setup the prior estimates for $\sigma_\varepsilon^2$, $\sigma_g^2$, $\sigma_b^2$ and $\sigma_h^2$; the scale parameter $S^2$ is then automatically reset based on the mode of scaled inverse $\chi^2$ distribution. For example, if we set `dfe=5`, and `priorVARe=0.2`, then $S_\varepsilon^2$ will be automatically set to be `priorVAR*(dfe+2)/dfe=0.2*(5+2)/5=0.28`. The code in Box S1 shows an example of setting user-defined hyper-parameters.

**Box S1: Changing hyper parameters in FW.**

```
lm1=FW(y=yNA,VAR=VAR,ENV=ENV,dfe=1,dfg=1,dfb=1,dfh=0.1,
priorVARe=0.36,priorVARg=0.18,priorVARh=0.74,priorVARb=0.18)
```

## Initial values

The FW package sets default initial values for all the chains. The default initial values for chain 1 are set as $\mu = 0$, **g**=**0**, **b**=**0**, **h**=**0**, $\sigma_\varepsilon^2$ =0.5 $\hat{\sigma}_p^2$ , $\sigma_g^2 = 0.25\hat{\sigma}_p^2$, $\sigma_b^2 = \sigma_h^2$ $0.5\hat{\sigma}_p$, where $\hat{\sigma}_p^2$ is the estimate of the total phenotypic variance. The default initial values for all other chains are generated from random sampling: $\mu$ from N(0, $0.5\hat{\sigma}_p^2$), **g** from N(**0**, $0.25\hat{\sigma}_p^2\mathbf{I}$), **b** and **h** from N(**0**, $0.5\hat{\sigma}_p\mathbf{I}$), the variance components are generated from uniform distribution whose boundaries are set as half and twice the values used in chain 1: $\sigma_\varepsilon^2$ from $0.5\hat{\sigma}_p^2$U(0.5, 2) , $\sigma_g^2$ from $0.25\hat{\sigma}_p^2$U(0.5, 2), $\sigma_b^2$ and $\sigma_h^2$ from $0.5\hat{\sigma}_p\mathrm{U}(0.5, 2)$.

The FW package provided the `inits` argument to set user defined initial values for Gibbs sampler. The parameters taken by R (the corresponding math representation in in parentheses) are: `mu` ($\mu$), `g` (**g**), `b` (**b**), `h`(**h**), `var_e` ($\sigma_\varepsilon^2$), `var_g` ($\sigma_g^2$), `var_b` ($\sigma_b^2$), `var_h` ($\sigma_h^2$). These values must be supplied in a list as `inits=list(inits1=list(mu=,g=,b=,h=,var_e=,var_g=,var_b=,var_h=),inits2=list(...),inits3=NULL)`. If any of the chains is set to be `NULL`, the initial values for that chain will be set up with default initial values from random sampling. The example code in Box S2 fitted a FW model with two chains for Gibbs sampler, the first chain is provided user-defined initial values, and by setting `inits2=NULL`, the second chain would have default initial values from random sampling.

**Box S2: Fitting more than one chain with user defined initial values.**

```
ng=length(unique(VAR));nh=length(unique(ENV));
fit2=FW(y=yNA,VAR=VAR,ENV=ENV, nchain=2, seed=c(1,2),
inits=list(inits1=list(mu=0,g=rep(0,ng),b=rep(0,ng),h=rep(0,nh),v
ar_e=0.36, var_g=0.18,var_b=0.18,var_h=0.74),inits2=NULL))
```

## References

Pérez, P., G. de Los Campos, J. Crossa, and D. Gianola. 2010. Genomic-enabled prediction based on molecular markers and pedigree using the bayesian linear regression package in R. The Plant Genome 3: 106-116.