Bayesian stability models: Eberhart-Russell, Finlay-Wilkinson and Shukla

Muhammad Yaseen

2026-11-04

Table of contents

1 Abstract **1**

2 Model **2**
2.1 Priors 2
2.2 The sampler 3

3 The Eberhart-Russell model on Maiz **3**
3.1 Least squares and Bayes 7

4 The safety-first index **7**

5 Which model? **8**

6 The stability fit and the decision tools **10**

7 Interpretation and sources **10**

8 References **12**

1 Abstract

The joint regression of Finlay and Wilkinson (1963) and Eberhart and Russell (1966) and the stability variance of Shukla (1972) are the most used measures of adaptability and stability. Their classical estimates come with tests, not with the uncertainty a breeder needs to choose. `bayes_stability()` fits the three models by Gibbs sampling, so every slope, deviation variance and stability variance has a posterior distribution. This vignette gives the model, the sampler, an analysis of `Maiz`, a comparison of the three models and the safety-first index of Eskridge (1990).

2 Model

For genotype i, environment j and plot l,

$$y_{ijl} = \mu + g_i + (1 + b_i)h_j + d_{ij} + e_{ijl}, \qquad d_{ij} \sim N(0, \sigma^2_{d,i}), \quad e_{ijl} \sim N(0, \sigma^2),$$

with $\sum_i g_i = \sum_j h_j = \sum_i b_i = 0$. The environmental index h_j is a parameter, estimated jointly with the others, as in the Bayesian Finlay-Wilkinson regression of Lian and de los Campos (2016). The classical analysis plugs in the environment means instead and so ignores their error. The slope of genotype i is $1 + b_i$, with average 1.

`.model`	Slopes	Deviation variance	Classical counterpart
`"ER"`	yes	$\sigma^2_{d,i}$ for every genotype	Eberhart and Russell (1966); Bayesian version of Oliveira et al. (2018)
`"FW"`	yes	one σ^2_d	Finlay and Wilkinson (1963)
`"Shukla"`	no ($b_i = 0$)	$\sigma^2_{d,i}$ for every genotype: the stability variance	Shukla (1972); Edwards and Jannink (2006)

The replicates separate the deviations d_{ij} from the plot error e_{ijl}, so the data must have replicates.

2.1 Priors

`bam_stab_priors()` sets normal priors on μ and on the contrasts $\mathbf{K}_G\mathbf{g}$, $\mathbf{K}_E\mathbf{h}$ and $\mathbf{K}_G\mathbf{b}$ (so the sums are zero), and scaled inverse chi-squared priors on σ^2 and $\sigma^2_{d,i}$. The defaults are vague and in the unit of the trait: 10^4 times the variance of the response for the effects, and the least-squares variances as scales with one degree of freedom.

```
bam_stab_priors(
    mu_var = NULL
  , g_var  = NULL
  , h_var  = NULL
  , b_var  = 100
  , nu     = 1
  , s2     = NULL
  , nu_d   = 1
  , s2_d   = NULL
  )
```

```
Priors of the Bayesian stability models
  mu       ~ N(0, 1e4 x var(y))
  K_G g    ~ N(0, 1e4 x var(y) I)
  K_E h    ~ N(0, 1e4 x var(y) I)
  K_G b    ~ N(0, 100 I)   (slopes 1 + b_i)
  sigma^2  ~ Inv-chi^2(1, within-cell mean square)
  sigma_d^2 ~ Inv-chi^2(1, least-squares deviation variance)
```

2.2 The sampler

Every conditional distribution is normal or scaled inverse chi-squared. Drawing the deviations d_{ij} one at a time next to the genotype effects would mix slowly, because the mean of the deviations of a genotype and its effect g_i trade off. The sampler therefore draws μ, $\mathbf{g}$, $\mathbf{h}$ and $\mathbf{b}$ with the deviations integrated out: each observed cell mean then has variance $\sigma^2_{d,i} + \sigma^2/n_{ij}$. The deviations come next, then the variances. This is a partially collapsed Gibbs sampler: each sweep is still an exact draw from the posterior. Unbalanced data need nothing more, as the weights use the number of plots of every cell, and an empty cell has weight zero. The sampler is compiled; vignette 13 checks it against a plain R version.

3 The Eberhart-Russell model on Maiz

```
fs <-
  bayes_stability(
      .data   = Maiz
    , .y      = y
    , .gen    = entry
    , .env    = site
    , .rep    = rep
    , .model  = "ER"
    , .nIter  = 3000
    , .burnin = 500
    , .thin   = 1
    , .chains = 2
    , .priors = bam_stab_priors(
          mu_var = NULL
        , g_var  = NULL
        , h_var  = NULL
        , b_var  = 100
        , nu     = 1
        , s2     = NULL
        , nu_d   = 1
        , s2_d   = NULL
        )
    , .level  = 0.95
    , .alpha  = 0.05
    , .seed   = 1
```

```
    , .waic   = TRUE
    )
fs
```

```
Bayesian Eberhart-Russell stability model: 9 genotypes, 20 environments, 4 replicates
2 chain(s) x 3000 kept draws (iterations 3000 after a burn-in of 500, thinning 1); 1.8 s

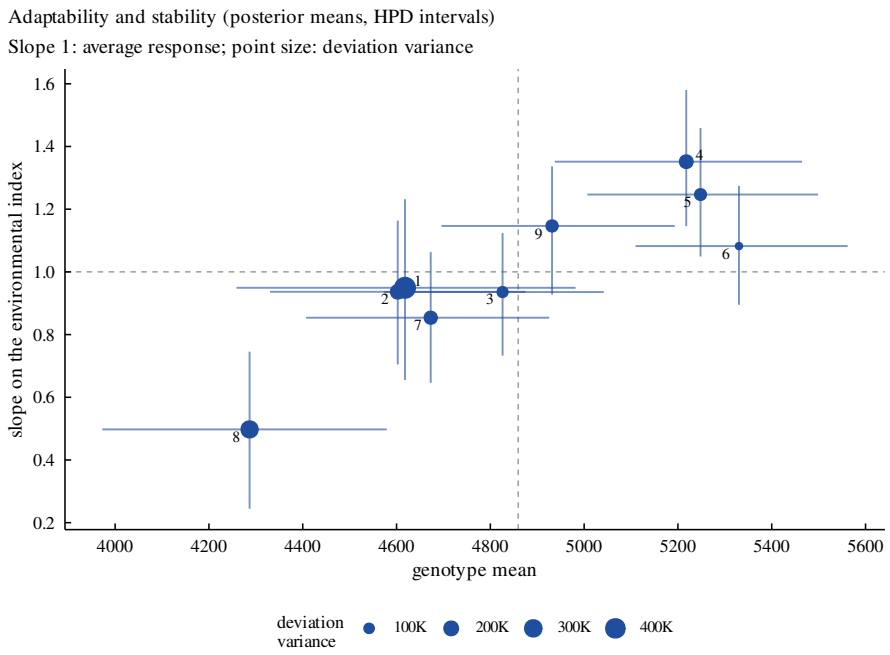
Genotypes: posterior means and 95% HPD intervals; safety-first index at alpha = 0.05
   Gen mean mean_lower mean_upper  slope slope_lower slope_upper
1:   1 4618       4259       4982 0.9493      0.6548      1.2320
2:   2 4603       4330       4875 0.9366      0.7046      1.1640
3:   3 4826       4574       5042 0.9357      0.7329      1.1240
4:   4 5218       4938       5465 1.3520      1.1460      1.5800
5:   5 5248       5007       5499 1.2470      1.0490      1.4590
6:   6 5330       5110       5562 1.0820      0.8950      1.2740
7:   7 4673       4407       4926 0.8538      0.6461      1.0630
8:   8 4287       3973       4579 0.4975      0.2446      0.7457
9:   9 4932       4696       5193 1.1470      0.9280      1.3370
   prob_slope_above_1 dev_var dev_var_lower dev_var_upper     R2 safety_first
1:          0.3590000  483200         85600       1036000 0.7349         3477
2:          0.2830000  193900         14610        464200 0.8689         3875
3:          0.2487000  108600         11620        273700 0.9216         4268
4:          0.9973000  167800         13440        450100 0.9408         4267
5:          0.9900000  124100         10230        302700 0.9478         4499
6:          0.8060000   77570          9611        192600 0.9559         4835
7:          0.0811700  152400         11200        364100 0.8760         3975
8:          0.0003333  288100         31390        618100 0.5723         2978
9:          0.9222000  130800         10310        345400 0.9364         4278
   safety_first_lower safety_first_upper prob_best_safety
1:               2757               4116        0.0000000
2:               3339               4383        0.0001667
3:               3820               4675        0.0103300
4:               3737               4759        0.0216700
5:               4038               4925        0.1113000
6:               4464               5202        0.8407000
7:               3482               4439        0.0005000
8:               2393               3514        0.0000000
9:               3775               4722        0.0153300

Overall mean 4859, error variance 756700; largest potential scale reduction factor 1.017
All parameters: x$summary; environmental index: x$environments

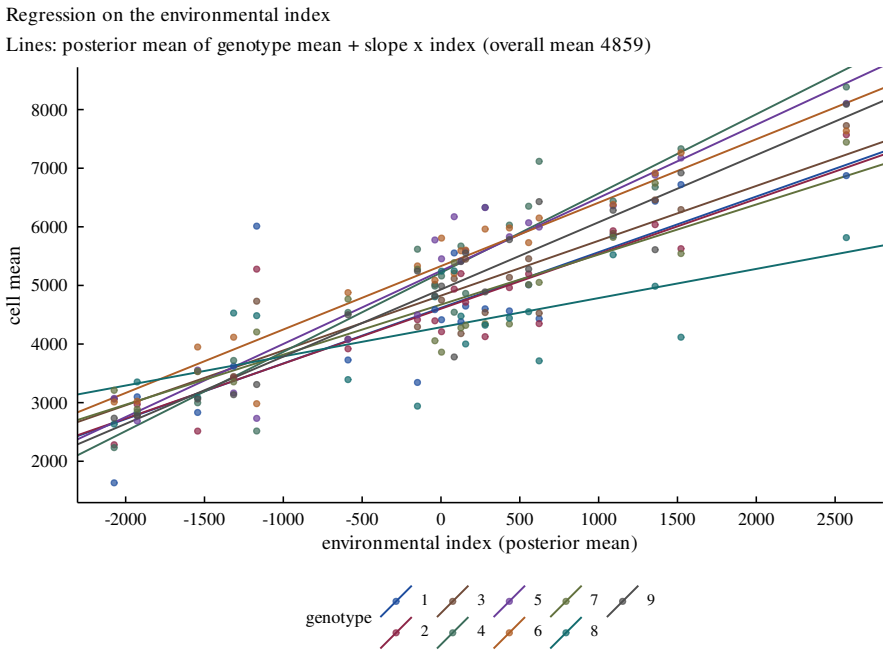
Sources: Eberhart and Russell (1966); Oliveira et al. (2018); Lian and de los
Campos (2016); Eskridge (1990); Spiegelhalter et al. (2002); Watanabe (2010).
Full references: bam_cite(x).
```

The slopes of genotypes 4 and 5 lie above 1 with 95% posterior probability: they respond more than the average genotype to good environments. The slope of 8 lies below 1. The posterior probability that a slope exceeds 1 is in the column `prob_slope_above_1`. A breeder can use it directly, which a test of $b_i = 1$ does not allow.

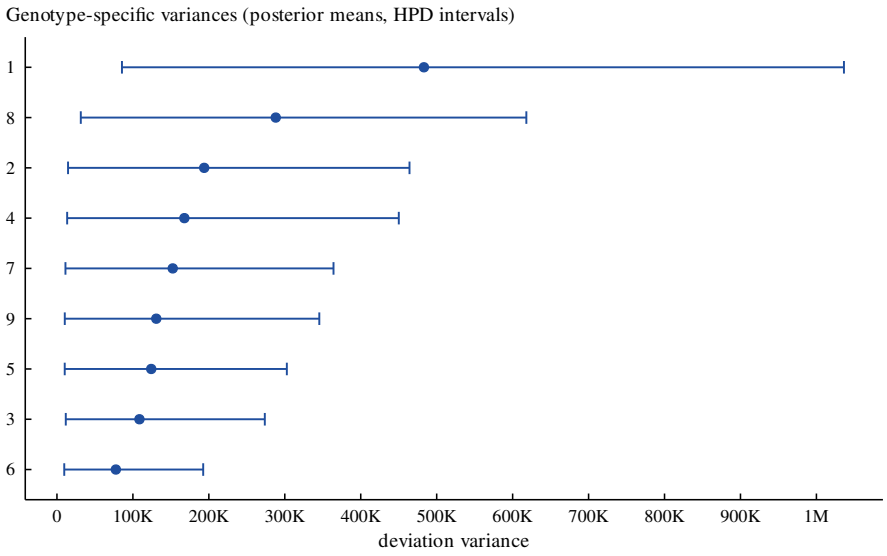
```
plot(x = fs, which = "slopes")
```




```
plot(x = fs, which = "lines")
```




```
plot(x = fs, which = "deviations")
```



The deviation variances are poorly determined from 20 environments, so their HPD intervals are wide and overlap. The Bayesian analysis shows this directly; a classical F test of $\sigma^2_{d,i} = 0$ would hide it.

3.1 Least squares and Bayes

```
data.table::data.table(
    Gen        = s$Gen
  , slope_ls   = round(x = fs$ls$slope, digits = 3)
  , slope      = round(x = s$slope, digits = 3)
  , dev_var_ls = round(fs$ls$dev_var)
  , dev_var    = round(s$dev_var)
  )
```

```
   Gen slope_ls slope dev_var_ls dev_var
1:   1    0.981 0.949     298673  483210
2:   2    0.965 0.937      49012  193865
3:   3    0.954 0.936      18897  108601
4:   4    1.314 1.352     250809  167772
5:   5    1.228 1.247      96052  124129
6:   6    1.057 1.082      23383   77575
7:   7    0.858 0.854      93153  152443
8:   8    0.524 0.498     212947  288101
9:   9    1.118 1.147     145199  130778
```

The posterior means of the slopes are close to the least-squares slopes (correlation 0.997). The deviation variances differ more. The posterior of a variance estimated from 20 environments is skewed to the right, so its mean lies above its mode. The prior, with its scale at the pooled least-squares value, also pulls the small variances up. The order of the genotypes changes less (rank correlation 0.73).

4 The safety-first index

Eskridge (1990) ranks genotypes by a lower quantile of their yield across environments, $\mu + g_i - z_{1-\alpha}\sqrt{V_i}$, where V_i is the variance of genotype i across environments under the model:

Model	V_i
Eberhart-Russell	$b_i^2 s_h^2 (1 - 1/E) + \sigma_{d,i}^2$
Finlay-Wilkinson	$b_i^2 s_h^2 (1 - 1/E)$
Shukla	$s_h^2 + \sigma_{d,i}^2$

Here s_h^2 is the variance of the environmental index. The index is computed on every draw, so it has an HPD interval and every genotype has a probability of being the best on it. Here 6 is most probably the best (0.84).

```
s[order(-safety_first), list(
    Gen
```

```
  , mean
  , safety_first
  , safety_first_lower
  , safety_first_upper
  , prob_best_safety
  )]
```

```
   Gen     mean safety_first safety_first_lower safety_first_upper
1:   6 5330.120     4834.788           4464.444           5202.405
2:   5 5248.139     4499.144           4037.793           4924.656
3:   9 4931.551     4278.128           3775.390           4722.382
4:   3 4826.080     4268.433           3820.211           4675.155
5:   4 5217.723     4266.802           3737.334           4759.201
6:   7 4672.868     3974.792           3482.001           4439.187
7:   2 4602.566     3875.111           3338.771           4382.980
8:   1 4618.040     3477.059           2757.315           4115.888
9:   8 4286.651     2978.271           2393.288           3514.087
   prob_best_safety
1:     0.8406666667
2:     0.1113333333
3:     0.0153333333
4:     0.0103333333
5:     0.0216666667
6:     0.0005000000
7:     0.0001666667
8:     0.0000000000
9:     0.0000000000
```

5 Which model?

```
ff <-
  bayes_stability(
      .data   = Maiz
    , .y      = y
    , .gen    = entry
    , .env    = site
    , .rep    = rep
    , .model  = "FW"
    , .nIter  = 3000
    , .burnin = 500
    , .thin   = 1
    , .chains = 2
    , .priors = bam_stab_priors(
          mu_var = NULL
        , g_var  = NULL
        , h_var  = NULL
```

```
        , b_var  = 100
        , nu     = 1
        , s2     = NULL
        , nu_d   = 1
        , s2_d   = NULL
        )
    , .level  = 0.95
    , .alpha  = 0.05
    , .seed   = 1
    , .waic   = TRUE
    )
fh <-
  bayes_stability(
      .data   = Maiz
    , .y      = y
    , .gen    = entry
    , .env    = site
    , .rep    = rep
    , .model  = "Shukla"
    , .nIter  = 3000
    , .burnin = 500
    , .thin   = 1
    , .chains = 2
    , .priors = bam_stab_priors(
          mu_var = NULL
        , g_var  = NULL
        , h_var  = NULL
        , b_var  = 100
        , nu     = 1
        , s2     = NULL
        , nu_d   = 1
        , s2_d   = NULL
        )
    , .level  = 0.95
    , .alpha  = 0.05
    , .seed   = 1
    , .waic   = TRUE
    )
cp <-
  bam_compare(
      ER     = fs
    , FW     = ff
    , Shukla = fh
    )
cp
```

```
Comparison of Bayesian models (smaller is better)

    model terms                variances      DIC       pD     WAIC   p_waic
```

```
1:     ER    NA   deviation per genotype 11893.02 104.1660 11897.91  97.2984
2:     FW    NA   one deviation variance 11896.06 101.2211 11901.92  95.4625
3: Shukla    NA interaction per genotype 11897.47 108.9504 11903.37 101.7497
        se lambda_last_lower lambda_last_mean  dWAIC dWAIC_se
1: 46.1831                NA               NA 0.0000   0.0000
2: 45.9799                NA               NA 4.0128   5.0903
3: 45.4214                NA               NA 5.4558   7.7438

Sources: Spiegelhalter et al. (2002); Watanabe (2010); Vehtari et al. (2017).
Full references: bam_cite(x).
```

WAIC prefers the ER model; the next, FW, is 4 above it (0.8 standard errors of the difference). The standard errors do not separate the three models here; the choice can follow the question, the slopes for adaptability and the variances for stability. The models are fitted to the same plots, so they can also be compared with the AMMI and GGE fits of `bayes_ammi()` (vignette 9).

6 The stability fit and the decision tools

The draws of the cell means $\mu + g_i + (1 + b_i)h_j + d_{ij}$ feed the decision tools of vignette 9:

```
bam_superiority(
    fit           = fs
  , .s            = 0.2
  , higher_better = TRUE
  )$marginal
```

```
   Gen     mean prob_performance prob_stability   prob_joint
1:   6 5329.591     0.8831666667    0.584666667 0.5163581111
2:   5 5248.083     0.6336666667    0.048500000 0.0307328333
3:   4 5218.175     0.4790000000    0.002500000 0.0011975000
4:   9 4929.700     0.0038333333    0.179333333 0.0006874444
5:   3 4824.993     0.0003333333    0.676333333 0.0002254444
6:   1 4617.564     0.0000000000    0.007166667 0.0000000000
7:   2 4602.071     0.0000000000    0.311000000 0.0000000000
8:   7 4672.272     0.0000000000    0.190500000 0.0000000000
9:   8 4281.845     0.0000000000    0.000000000 0.0000000000
```

7 Interpretation and sources

```
bam_interpret(
    x             = fs
  , higher_better = TRUE
  , top           = 5
  )
```

```
Bayesian Eberhart-Russell stability model

Best by the safety-first index (alpha = 0.05): 6, 5, 9, 3 and 4; most
probably the best: 6 (0.84).

Findings:
- The chains have converged (largest potential scale reduction factor 1.017;
  smallest effective sample size of the main parameters 574).
- Slopes whose HPD interval lies above 1 (respond more than average to good
  environments): 4 and 5; below 1 (respond less, relatively better in poor
  environments): 8.
- Smallest deviation variance: 6 (77,600); largest: 1 (483,000).
- Share of each genotype's variation explained by the regression (R2): 1
  0.73, 2 0.87, 3 0.92, 4 0.94, 5 0.95, 6 0.96, 7 0.88, 8 0.57, 9 0.94.

Recommendations:
- For a wide recommendation choose genotypes with a high mean, a slope near 1
  and a small deviation variance; for favourable environments a slope above 1
  with a high mean.
- The safety-first index combines the two: 6, 5 and 9 lead.

Cautions:
- The environmental index is estimated from the same trial; with few
  environments the slopes are uncertain (see the HPD intervals).

Sources:
- Eberhart, S. A. and Russell, W. A. (1966). Stability parameters for
  comparing varieties. Crop Science, 6(1), 36-40.
  doi:10.2135/cropsci1966.0011183X000600010011x
- Oliveira, T. R. A., Carvalho, H. W. L., Nascimento, M., Costa, E. F. N.,
  Amaral Junior, A. T., Gravina, G. A. and Carvalho Filho, J. L. S. (2018).
  The Eberhart and Russell's Bayesian method used as an instrument to select
  maize hybrids. Euphytica, 214, 64. doi:10.1007/s10681-018-2146-y
- Lian, L. and de los Campos, G. (2016). FW: an R package for
  Finlay-Wilkinson regression that incorporates genomic/pedigree information
  and covariance structures between environments. G3: Genes, Genomes,
  Genetics, 6(3), 589-597. doi:10.1534/g3.115.026328
- Eskridge, K. M. (1990). Selection of stable cultivars using a safety-first
  rule. Crop Science, 30, 369-374.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A.
  (2002). Bayesian measures of model complexity and fit. Journal of the Royal
  Statistical Society B, 64(4), 583-639. doi:10.1111/1467-9868.00353
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and
  widely applicable information criterion in singular learning theory.
```

Journal of Machine Learning Research, 11, 3571-3594.

```
bam_cite(
    x     = fs
  , style = c("text", "bibtex")[1]
  )
```

Eberhart, S. A. and Russell, W. A. (1966). Stability parameters for comparing varieties. Cro

Oliveira, T. R. A., Carvalho, H. W. L., Nascimento, M., Costa, E. F. N., Amaral Junior, A. T

Lian, L. and de los Campos, G. (2016). FW: an R package for Finlay-Wilkinson regression that

Eskridge, K. M. (1990). Selection of stable cultivars using a safety-first rule. Crop Scienc

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian meas

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable

8 References

Finlay, K. W. and Wilkinson, G. N. (1963). The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research*, 14(6), 742–754. doi:10.1071/AR9630742

Eberhart, S. A. and Russell, W. A. (1966). Stability parameters for comparing varieties. *Crop Science*, 6(1), 36–40. doi:10.2135/cropsci1966.0011183X000600010011x

Shukla, G. K. (1972). Some statistical aspects of partitioning genotype-environmental components of variability. *Heredity*, 29(2), 237–245.

Eskridge, K. M. (1990). Selection of stable cultivars using a safety-first rule. *Crop Science*, 30, 369–374.

Lian, L. and de los Campos, G. (2016). FW: an R package for Finlay-Wilkinson regression that incorporates genomic/pedigree information and covariance structures between environments. *G3: Genes, Genomes, Genetics*, 6(3), 589–597. doi:10.1534/g3.115.026328

Oliveira, T. R. A., Carvalho, H. W. L., Nascimento, M., Costa, E. F. N., Amaral Junior, A. T., Gravina, G. A. and Carvalho Filho, J. L. S. (2018). The Eberhart and Russell's Bayesian method used as an instrument to select maize hybrids. *Euphytica*, 214, 64. doi:10.1007/s10681-018-2146-y

Edwards, J. W. and Jannink, J.-L. (2006). Bayesian modeling of heterogeneous error and genotype x environment interaction variances. *Crop Science*, 46(2), 820. doi:10.2135/cropsci2005.0164