Interpreting the results

Muhammad Yaseen

2026-11-04

Table of contents

1 Abstract **1**

2 The Bayesian AMMI fit **2**

3 The facts behind the sentences **4**

4 The biplot **4**

5 Stability models **5**

6 Decisions and model choice **7**

7 Markdown for reports **9**
7.1 Bayesian AMMI analysis 9

8 How each method is read **10**

9 References **10**

1 Abstract

`bam_interpret()` writes a plain-language interpretation of a result of `bayes_ammi()`, `biplots()` or `ge_ammi()`. Every sentence is computed from the numbers of the result. This vignette shows the interpretations and the rules behind them.

```
fit <-
  bayes_ammi(
      .data   = Maiz
    , .y      = y
    , .gen    = entry
    , .env    = site
    , .rep    = rep
```

```
    , .nIter  = 4000
    , .k      = 2
    , .burnin = 1000
    , .thin   = 1
    , .chains = 2
    , .priors = bam_priors(
          mu_mean     = 0
        , mu_var      = 1e+8
        , tau_mean    = 0
        , tau_var     = 1e+8
        , delta_mean  = 0
        , delta_var   = 1e+8
        , lambda_mean = 0
        , lambda_var  = 1e+8
        , nu          = 1
        , s2          = NULL
        )
    , .level  = 0.95
    , .seed   = 8
    , .align  = TRUE
    , .model  = "AMMI"
    , .het    = FALSE
    , .waic   = TRUE
    )
```

2 The Bayesian AMMI fit

```
txt <-
  bam_interpret(
      x             = fit
    , higher_better = TRUE
    , top           = 5
    )
txt
```

```
Bayesian AMMI analysis

2 bilinear term(s): lambda_1 = 5,466, lambda_2 = 1,646; the chains have
converged.

Findings:
- The chains have converged by the criteria of Crossa et al. (2011) (largest
  potential scale reduction factor 1.003); the smallest effective sample size
  of the main parameters is 862.
- 5 of 131 Heidelberger-Welch stationarity and half-width tests failed (about
  5% fail by chance at level 0.05).
```

```
- Overall mean 4,858 (95% HPD 4,793 to 4,920); error variance 763,800, a
  coefficient of variation of 18.0%.
- Genotypes significantly above average (HPD of the effect excludes zero): 4,
  5 and 6; below average: 1, 2, 7 and 8.
- Environments with significantly high means: 4, 6, 11, 12, 13 and 2 more;
  low: 1, 2, 7, 8, 10 and 2 more.
- Singular values with 95% HPD intervals: lambda_1 = 5,466 (4,538 to 6,343);
  lambda_2 = 1,646 (15.93 to 2,835).
- On the first bilinear term, genotypes with significantly positive scores:
  4, 5, 6 and 9; negative: 1, 2, 3 and 8.
- Environments with significantly positive first scores: 11, 12, 13, 15 and
  18; negative: 1, 3, 8 and 10.
- Genotypes and environments with significant first scores of the same sign
  interact positively: the genotype yields more there than its main effect
  predicts; of opposite sign, less.

Recommendations:
- Recommend the genotypes with significant first scores for the environments
  whose scores have the same sign (biplots() shows the regions).
- Draw biplots() to see which genotypes and environments form homogeneous
  groups (overlapping HPD regions).

Cautions:
- The HPD interval of lambda_2 reaches down to near zero: the term is weakly
  supported and its scores are uncertain.

Sources:
- Crossa, J., Perez-Elizalde, S., Jarquin, D., Cotes, J. M., Viele, K., Liu,
  G. and Cornelius, P. L. (2011). Bayesian estimation of the additive main
  effects and multiplicative interaction model. Crop Science, 51(4),
  1458-1469. doi:10.2135/cropsci2010.06.0343
- 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.
```

The rules:

- **Convergence**: potential scale reduction factor below 1.1, effective sample size of the main parameters at least 400, Raftery-Lewis dependence factor below 5, and no more than 10% of the Heidelberger-Welch tests failed.
- **Main effects**: a genotype or environment differs from the average when the HPD interval of its effect excludes zero.
- **Bilinear terms**: a term whose HPD interval of λ_k reaches below a tenth of its posterior mean is weakly supported.

- **Interaction**: significant first scores of the same sign mean a positive interaction.

3 The facts behind the sentences

```
str(object = txt$facts[c("converged", "max_rhat", "genotypes_above",
 ↪  "axis1_genotypes_positive",
                          "axis1_environments_positive", "weak_terms")])
```

```
List of 6
 $ converged                   : logi TRUE
 $ max_rhat                    : num 1
 $ genotypes_above             : chr [1:3] "4" "5" "6"
 $ axis1_genotypes_positive    : chr [1:4] "4" "5" "6" "9"
 $ axis1_environments_positive: chr [1:5] "11" "12" "13" "15" ...
 $ weak_terms                  : int 2
```

4 The biplot

```
bp <-
  biplots(
      model         = fit
    , burnin        = 0
    , thin          = 0.5
    , pb            = 0.05
    , plot_stable   = TRUE
    , plot_unstable = TRUE
    , ncolors       = 5
    , .which        = "both"
    , .grid         = 100
    )
bam_interpret(
    x             = bp
  , higher_better = TRUE
  , top           = 5
  )
```

```
Biplot with highest posterior density regions

7 of 9 genotypes and 8 of 20 environments have a 95% HPD region that excludes
the origin (significant interaction).

Findings:
- Genotypes whose region contains the origin (interaction not different from
  zero, stable): 3 and 7.
```

- Groups on the first term (marginal HPD interval excludes zero): positive 4, 5, 6 and 9; negative 1, 2, 3 and 8.
- Environments on the first term: positive 4, 11, 12, 13, 15 and 1 more; negative 1, 3, 8 and 10; the others do not differ from zero.
- Overlapping regions mark genotypes (environments) whose interaction parameters do not differ: a homogeneous group.

Recommendations:
- Genotypes 4, 5, 6 and 9 are adapted to environments 4, 11, 12, 13, 15 and 1 more (both positive on the first term).
- Genotypes 1, 2, 3 and 8 are adapted to environments 1, 3, 8 and 10 (both negative on the first term).

Cautions:
- The regions are estimated from the draws by a kernel density; with few draws their outline is rough.
- The second term is usually estimated with much more uncertainty than the first: long regions along the second axis.

Sources:
- Crossa, J., Perez-Elizalde, S., Jarquin, D., Cotes, J. M., Viele, K., Liu, G. and Cornelius, P. L. (2011). Bayesian estimation of the additive main effects and multiplicative interaction model. Crop Science, 51(4), 1458-1469. doi:10.2135/cropsci2010.06.0343

5 Stability models

```
fs <-
  bayes_stability(
      .data   = Maiz
    , .y      = y
    , .gen    = entry
    , .env    = site
    , .rep    = rep
    , .model  = "ER"
    , .nIter  = 2000
    , .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   = 8
    , .waic   = TRUE
    )
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, 3, 9 and 4; most
probably the best: 6 (0.84).

Findings:
- The chains may not have converged (largest potential scale reduction factor
  1.011; smallest effective sample size of the main parameters 383).
- 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 (74,500); largest: 1 (502,000).
- Share of each genotype's variation explained by the regression (R2): 1
  0.73, 2 0.88, 3 0.93, 4 0.94, 5 0.95, 6 0.96, 7 0.86, 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 3 lead.

Cautions:
- The chains have not converged by the criteria used; run longer chains
  before using the estimates.
- 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.

Every interpretation ends with the sources of the method; `bam_cite()` gives them as BibTeX.

6 Decisions and model choice

The results of the decision tools and of the model choice functions have interpretations too:

```
bam_interpret(
    x             = bam_superiority(
        fit           = fit
      , .s            = 0.2
      , higher_better = TRUE
      )
  , higher_better = TRUE
  , top           = 5
  )
```

```
Probabilities of superior performance and stability

Selecting 2 genotype(s): the highest joint probability of superior
performance and stability is that of 6 (0.15).

Findings:
- Probability of superior performance, largest first: 6 0.89, 5 0.64, 4 0.47;
  under 0.005: 9, 1, 2, 3, 7, 8.
- Probability of superior stability, largest first: 7 0.71, 3 0.70, 2 0.24, 6
  0.16, 5 0.11, 9 0.06, 1 0.02; under 0.005: 8, 4.
- Most likely superior in performance: 6; in stability: 7.

Recommendations:
```

- For a wide recommendation consider 6 and 5 (joint probability of at least 0.05); the probabilities by environment (component conditional) show where each genotype is superior.

Cautions:
- The joint probability multiplies the two probabilities as in Dias et al. (2022), which treats them as independent.

Sources:
- Dias, K. O. G., dos Santos, J. P. R., Krause, M. D., Piepho, H.-P., Guimaraes, L. J. M., Pastina, M. M. and Garcia, A. A. F. (2022). Leveraging probability concepts for cultivar recommendation in multi-environment trials. Theoretical and Applied Genetics. doi:10.1007/s00122-022-04041-y

```
bam_interpret(
    x             = bam_ppc(
        fit    = fit
      , .draws = 300
      , .seed  = 1
      )
  , higher_better = TRUE
  , top           = 5
  )
```

Posterior predictive checks

The model does not reproduce: largest / smallest environment variance.

Findings:
- variance of cell means: observed 1,842,000, replicated mean 1,917,000, p = 0.713.
- largest cell mean: observed 8,385, replicated mean 8,636, p = 0.730.
- smallest cell mean: observed 1,632, replicated mean 1,731, p = 0.650.
- variance of genotype means: observed 124,700, replicated mean 146,900, p = 0.723.
- first singular value: observed 5,923, replicated mean 5,957, p = 0.540.
- within-cell variance: observed 755,900, replicated mean 771,700, p = 0.567.
- largest / smallest environment variance: observed 25.82, replicated mean 2.944, p = 0.000.

Recommendations:
- Fit environment-specific variances (.het = TRUE).

Cautions:
- A p-value near 0 or 1 marks a misfit; values in between do not prove the model right.

```
Sources:
- Gelman, A., Meng, X.-L. and Stern, H. (1996). Posterior predictive
  assessment of model fitness via realized discrepancies. Statistica Sinica,
  6(4), 733-807.
```

7 Markdown for reports

```
cat(format(x = txt, style = "markdown"), sep = "\n")
```

7.1 Bayesian AMMI analysis

2 bilinear term(s): lambda_1 = 5,466, lambda_2 = 1,646; the chains have converged.

7.1.1 Findings

- The chains have converged by the criteria of Crossa et al. (2011) (largest potential scale reduction factor 1.003); the smallest effective sample size of the main parameters is 862.
- 5 of 131 Heidelberger-Welch stationarity and half-width tests failed (about 5% fail by chance at level 0.05).
- Overall mean 4,858 (95% HPD 4,793 to 4,920); error variance 763,800, a coefficient of variation of 18.0%.
- Genotypes significantly above average (HPD of the effect excludes zero): 4, 5 and 6; below average: 1, 2, 7 and 8.
- Environments with significantly high means: 4, 6, 11, 12, 13 and 2 more; low: 1, 2, 7, 8, 10 and 2 more.
- Singular values with 95% HPD intervals: lambda_1 = 5,466 (4,538 to 6,343); lambda_2 = 1,646 (15.93 to 2,835).
- On the first bilinear term, genotypes with significantly positive scores: 4, 5, 6 and 9; negative: 1, 2, 3 and 8.
- Environments with significantly positive first scores: 11, 12, 13, 15 and 18; negative: 1, 3, 8 and 10.
- Genotypes and environments with significant first scores of the same sign interact positively: the genotype yields more there than its main effect predicts; of opposite sign, less.

7.1.2 Recommendations

- Recommend the genotypes with significant first scores for the environments whose scores have the same sign (biplots() shows the regions).
- Draw biplots() to see which genotypes and environments form homogeneous groups (overlapping HPD regions).

7.1.3 Cautions

- The HPD interval of lambda_2 reaches down to near zero: the term is weakly supported and its scores are uncertain.

7.1.4 Sources

- Crossa, J., Perez-Elizalde, S., Jarquin, D., Cotes, J. M., Viele, K., Liu, G. and Cornelius, P. L. (2011). Bayesian estimation of the additive main effects and multiplicative interaction model. Crop Science, 51(4), 1458-1469. doi:10.2135/cropsci2010.06.0343
- 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.

8 How each method is read

```
cat(strwrap(x = txt$guide, width = 80), sep = "\n")
```

```
Bayesian AMMI: every parameter has a posterior distribution, summarised by its
mean, standard deviation and highest posterior density (HPD) interval. Main
effects whose HPD interval excludes zero differ from the average. Singular
values measure the strength of each bilinear term. Genotype and environment
scores with HPD intervals excluding zero interact significantly; same signs
mean positive interaction. Check convergence first: potential scale reduction
factor near 1, large effective sample sizes, Raftery-Lewis dependence factor
below 5, Heidelberger-Welch tests passed.
```

9 References

Crossa, J., Perez-Elizalde, S., Jarquin, D., Cotes, J. M., Viele, K., Liu, G. and Cornelius, P. L. (2011). Bayesian estimation of the additive main effects and multiplicative interaction model. *Crop Science*, 51(4), 1458–1469. doi:10.2135/cropsci2010.06.0343