Reproducing Perez-Elizalde et al. (2012)

Muhammad Yaseen

2026-11-04

Table of contents

1 Abstract **1**

2 The fit **1**

3 Table B.1: the linear effects **2**
3.1 Why the posterior means are averages of the two years 4

4 Table 2: the residual standard deviation and the singular values **4**

5 Table 2: the significant scores **5**

6 Figure 4: the biplot **6**

7 References **7**

1 Abstract

Perez-Elizalde et al. (2012, Section 3.2) analysed the second year of the maize trial with the first year as prior and all 11 bilinear terms. This vignette repeats their analysis and compares it with Table B.1 (the 38 linear effects), Table 2 (σ, the singular values and the significant scores) and Figure 4 (the biplot).

2 The fit

The paper used two chains of 20,000 iterations. Here two chains of 5000 after a burn-in of 1000 suffice for the comparison.

```
fm <-
  bayes_ammi(
      .data       = cultivo2009
    , .y          = y
    , .gen        = entry
```

```
    , .env        = site
    , .rep        = rep
    , .nIter      = 5000
    , .prior_data = cultivo2008
    , .n0         = NULL
    , .model      = "AMMI"
    , .k          = NULL
    , .burnin     = 1000
    , .thin       = 1
    , .chains     = 2
    , .priors     = bys_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       = 2012
    , .align      = TRUE
    , .waic       = TRUE
    )
```

```
s <- fm$summary
c(max_rhat = max(s$rhat), min_ess_linear = min(s[block %in% c("mu",
 ↪  "tau", "delta"), ess]))
```

```
      max_rhat min_ess_linear
      1.003199    9323.595868
```

3 Table B.1: the linear effects

```
pub <- utils::read.csv(file = system.file("extdata", "perez2012.csv",
 ↪  package = "baystability"))
pos <- c("mu", sprintf("tau[%d]", 1:12), sprintf("delta[%d]", 1:25))
lin <- s[block %in% c("mu", "tau", "delta")]
b1 <- data.table::data.table(
    parameter = c("mu", sprintf("alpha%d", 1:12), sprintf("beta%d",
     ↪  1:25))
  , published = pub$mean[match(x = pos, table = pub$parameter)]
  , here      = round(x = lin$mean, digits = 1)
  , sd_pub    = pub$sd[match(x = pos, table = pub$parameter)]
```

```
  , sd_here   = round(x = lin$sd, digits = 2)
  )
b1[, diff_in_sd := round(x = (here - published) / sd_pub, digits = 3)]
b1
```

```
    parameter published    here sd_pub sd_here diff_in_sd
 1:        mu   3848.00  3848.6  30.68   30.59      0.020
 2:    alpha1    135.90   136.2 102.92  102.17      0.003
 3:    alpha2    317.50   316.5 102.29  102.25     -0.010
 4:    alpha3    -64.68   -63.1 103.26  101.36      0.015
 5:    alpha4      2.75     5.9 102.62  102.19      0.031
 6:    alpha5     93.39    90.9 101.90  101.13     -0.024
 7:    alpha6    181.00   181.5 101.81  102.52      0.005
 8:    alpha7    499.30   500.7 103.21  101.03      0.014
 9:    alpha8     64.30    64.3 102.72  101.15      0.000
10:    alpha9    129.30   127.7 102.34  101.78     -0.016
11:   alpha10   -479.00  -481.6 103.10  101.06     -0.025
12:   alpha11   -270.10  -270.1 103.02  101.04      0.000
13:   alpha12   -609.70  -608.8 102.61  101.67      0.009
14:     beta1    134.50   131.8 151.21  150.78     -0.018
15:     beta2   -533.60  -533.9 151.39  151.41     -0.002
16:     beta3    864.20   864.8 151.39  150.44      0.004
17:     beta4   -377.40  -376.9 151.37  152.35      0.003
18:     beta5   1537.00  1537.2 150.94  151.72      0.001
19:     beta6    -96.78   -95.3 150.28  149.07      0.010
20:     beta7  -1761.00 -1759.2 151.22  150.60      0.012
21:     beta8    348.20   349.2 151.61  152.07      0.007
22:     beta9  -1468.00 -1465.4 150.96  151.53      0.017
23:    beta10   -745.10  -747.4 150.44  149.38     -0.015
24:    beta11  -1202.00 -1201.7 151.64  150.92      0.002
25:    beta12  -1900.00 -1902.1 152.11  150.28     -0.014
26:    beta13   -971.60  -970.3 150.75  149.92      0.009
27:    beta14   -548.60  -549.3 151.84  152.49     -0.005
28:    beta15  -2215.00 -2217.2 151.62  151.60     -0.015
29:    beta16     87.58    87.8 150.95  149.10      0.001
30:    beta17    160.40   160.8 152.11  149.39      0.003
31:    beta18   1848.00  1847.9 152.18  148.98     -0.001
32:    beta19   -865.20  -867.4 150.62  152.42     -0.015
33:    beta20  -1112.00 -1110.5 151.82  151.23      0.010
34:    beta21  -1627.00 -1626.7 151.76  148.66      0.002
35:    beta22   4917.00  4918.7 151.56  149.98      0.011
36:    beta23   2865.00  2865.7 151.21  151.83      0.005
37:    beta24    911.80   910.0 149.76  151.62     -0.012
38:    beta25   1749.00  1749.4 150.27  151.06      0.003
    parameter published    here sd_pub sd_here diff_in_sd
```

All 38 posterior means agree with the published ones to within 0.031 posterior standard deviations, and the standard deviations agree too.

3.1 Why the posterior means are averages of the two years

With $n_0 = B = 3$ the earlier trial weighs as much as the new one, so the posterior mean of every linear effect is the average of its two least-squares estimates:

```
both <-
  bys_ls_means(
      .data = rbind(
          cultivo2008[c("site", "rep", "entry", "y")]
        , cultivo2009[c("site", "rep", "entry", "y")]
        )
    , .y    = y
    , .gen  = entry
    , .env  = site
    )
c(average_mu = both$grand_mean, published = 3848)
```

```
average_mu  published
  3848.275   3848.000
```

4 Table 2: the residual standard deviation and the singular values

```
sig <- sqrt(unlist(fm$tau1))
lam <- s[block == "lambda"]
t2 <- data.table::data.table(
    parameter = c("sigma", sprintf("lambda%d", 1:11))
  , published = pub$mean[match(x = c("sigma", sprintf("lambda[%d]",
↪  1:11)), table = pub$parameter)]
  , here      = round(c(mean(sig), lam$mean))
  , sd_pub    = pub$sd[match(x = c("sigma", sprintf("lambda[%d]", 1:11)),
↪  table = pub$parameter)]
  , sd_here   = round(c(sd(sig), lam$sd))
  )
t2[, ratio := round(x = here / published, digits = 2)]
t2
```

```
    parameter published here sd_pub sd_here ratio
 1:     sigma    1311.0 1305  39.13      39  1.00
 2:   lambda1    2572.0 3658 451.76     397  1.42
 3:   lambda2    2354.0 3280 388.40     353  1.39
 4:   lambda3    2123.0 2822 340.34     313  1.33
 5:   lambda4    1815.0 2501 328.85     295  1.38
 6:   lambda5    1576.0 2225 317.49     297  1.41
```

```
 7:   lambda6    1262.0 1856 311.66     297  1.47
 8:   lambda7     992.0 1536 297.34     288  1.55
 9:   lambda8     762.6 1248 283.71     286  1.64
10:   lambda9     506.1  945 260.09     285  1.87
11:  lambda10     315.4  648 217.51     276  2.05
12:  lambda11     167.4  328 153.72     234  1.96
```

The residual standard deviation agrees: 1305.1 against 1311.00, 0.15 posterior standard deviations apart.

The singular values do not agree: here they are 1.33 to 2.05 times the published ones. By eq. 2.15 the posterior mean of λ_k is close to $(n_0\lambda_{0k} + B\,\boldsymbol{\alpha}_k^\top \mathbf{Y}\boldsymbol{\gamma}_k)/(n_0 + B)$. Here $\lambda_{01} = 5306$ is the first singular value of the first year. With equal weights, a published $\lambda_1 = 2572$ below $\lambda_{01}/2 = 2653$ would need $\boldsymbol{\alpha}_1^\top \mathbf{Y}\boldsymbol{\gamma}_1$ to be negative on average, that is, the second-year data pulling the first term against itself. The leading published values are close to $\lambda_{0k}/2$, as if the second year added nothing to them.

Two explanations were checked with the sampler of this package (`dev/lambda-checks.R` in the source) and neither gives the published values:

- Score columns that do not move (no pair moves, as in the column-by-column sampler of the paper), started at the first year's fit or at random orthonormal scores: λ_1 is then 3613 or 3313, not 2572, because the environment scores still follow the data.
- A prior mean of zero for the singular values with the same weight: λ_1 is then 1203.

The linear effects and σ, which come from the same draws, are reproduced; the sampler gives the same draws as an independent R version of eq. 2.15 and recovers the true values of simulated trials (validation vignette). The published singular values therefore disagree with the method as the paper states it, and the difference lies in the article, through a step it does not report or an error in its Table 2, rather than in the package.

5 Table 2: the significant scores

The paper lists the scores whose bivariate 0.95 HPD region excludes the origin: genotype 12 on the second term and environments 5, 9, 23 and 25 on the first.

```
sc <- s[block %in% c("alpha", "gamma") & axis %in% 1:2]
sc[(block == "alpha" & label == levels(factor(cultivo2009$entry))[12]) |
     (block == "gamma" & label %in% c("5", "9", "23", "25")), list(
         parameter
       , axis
       , mean
       , sd
       , lower
```

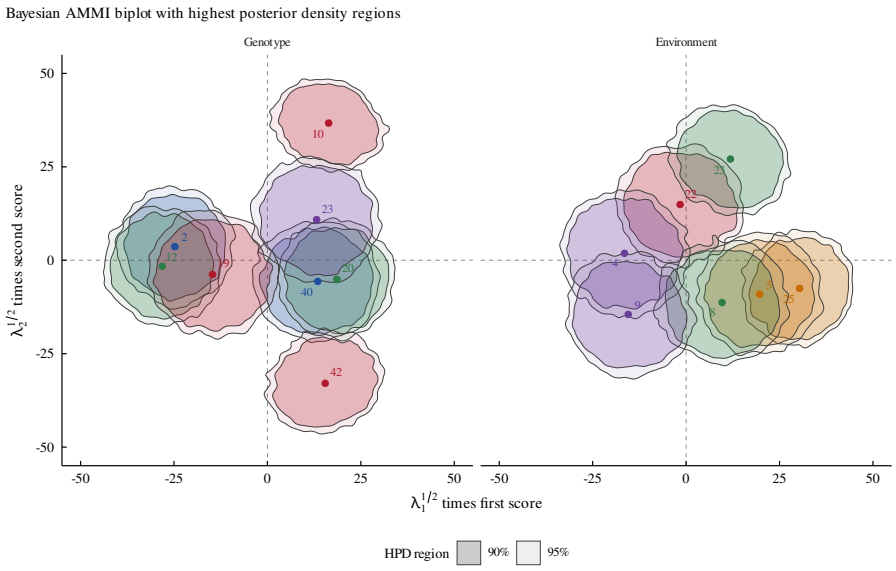
```
      , upper
      )]
```

```
      parameter axis       mean         sd       lower       upper
 1: alpha[42,1]    1  0.2561431 0.10545654  0.04358435  0.45406065
 2: alpha[42,2]    2 -0.5756696 0.09304491 -0.75843874 -0.39343712
 3:  gamma[5,1]    1  0.3253565 0.11155397  0.09532592  0.53119878
 4:  gamma[9,1]    1 -0.2567318 0.11568836 -0.46952401 -0.01959994
 5: gamma[23,1]    1  0.1969576 0.10973267 -0.01683380  0.40928783
 6: gamma[25,1]    1  0.5029371 0.09801811  0.30471768  0.68974470
 7:  gamma[5,2]    2 -0.1590733 0.11546738 -0.38297736  0.06846025
 8:  gamma[9,2]    2 -0.2529788 0.11685961 -0.47634328 -0.01716198
 9: gamma[23,2]    2  0.4735830 0.10219009  0.27225302  0.67003633
10: gamma[25,2]    2 -0.1319041 0.11086797 -0.35726222  0.07672915
```

On the first term the environment scores agree in sign and size with the published values (0.288, -0.277, 0.278 and 0.459 for environments 5, 9, 23 and 25; here 0.325, -0.257, 0.197, 0.503). Terms beyond the first are weakly identified when 11 terms of similar size are fitted: their order and orientation can differ between analyses.

6 Figure 4: the biplot

```
bp <-
  bayes_ammi_biplot(
      fit           = fm
    , burnin        = 0
    , thin          = 1
    , pb            = c(0.10, 0.05)
    , plot_stable   = FALSE
    , plot_unstable = TRUE
    , ncolors       = 5
    , .which        = "both"
    , .grid         = 100
    )
bp$biplot
```



7 References

Perez-Elizalde, S., Jarquin, D. and Crossa, J. (2012). A general Bayesian estimation method of linear-bilinear models applied to plant breeding trials with genotype x environment interaction. *Journal of Agricultural, Biological, and Environmental Statistics*, 17, 15–37. doi:10.1007/s13253-011-0063-9

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