AMBARTI models for agricultural experiments

Prado, E. B., Santos, A. A. L

Hamilton Institute & Dept. of Mathematics and Statistics

Maynooth University
National University of Ireland Maynooth

January 26th, 2021 - Group meeting

Agenda

- Additive Main Effect interaction (AMMI) models
- AMMI + BART = AMBARTI
- Simulation results
- Next steps

Additive Main effects and Multiplicative Interactions (AMMI)

Linear–bilinear models are frequently used to analyse two-way data such as genotype-by-environment data.

An example of this class of models is the AMMI model:

$$y_{ij} = \mu + \alpha_i + \beta_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq} + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}\left(0, \tau^{-1}\right),$$

where

- y - response.
- α - effect of genotype.
- β - effect of environment.
- λ - singular value of the multiplicative component.
- γ - genotype singular value.
- δ - environment singular value.

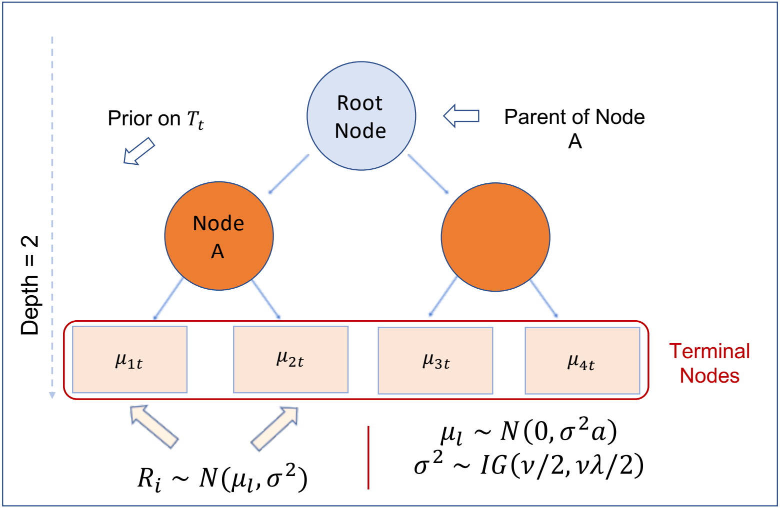
Additive Main effects and Multiplicative Interactions (AMMI)

The following priors are assumed in its Bayesian version (Josse et al, JABES, 2014):

$$
\begin{aligned}
\mu &\sim \mathsf{N}\left(m, s_\mu^2\right), \\
\alpha_i &\sim \mathsf{N}\left(0, s_g^2\right), \\
\beta_j &\sim \mathsf{N}\left(0, s_e^2\right), \\
(\lambda_q)_{q=1,\ldots,Q} &\sim \text{ ordered sample of } Q \text{ independent } \mathsf{N}^+\left(0, s_\lambda^2\right), \\
\gamma_{1q} &\sim \mathsf{N}^+(0,1) \quad \text{for } q = 1, \ldots, Q, \\
\gamma_{iq} &\sim \mathsf{N}(0,1) \quad \text{for } i > 1 \text{ and } q = 1, \ldots, Q, \\
\delta_{jq} &\sim \mathsf{N}(0,1) \quad \text{for } j \geq 1 \text{ and } q = 1, \ldots, Q, \\
\sigma_E &\sim \mathsf{U}\left(0, S_{\mathrm{ME}}\right).
\end{aligned}
$$

Additive Main Effect Bayesian Additive Regression Tree Interaction models (AMBARTI)

BART



Prior on T_t
Root Node
Parent of Node A
Depth = 2
Node A
μ_{1t}
μ_{2t}
μ_{3t}
μ_{4t}
Terminal Nodes
$R_i \sim N(\mu_l, \sigma^2)$
$\mu_l \sim N(0, \sigma^2 a)$
$\sigma^2 \sim IG(\nu/2, \nu\lambda/2)$

AMMI + BART

BART is a flexible tree-based method that can be used for predicting when there are interactions and non-linear relationships.

$$y_{ij}|\mathbf{x}_{ij}, \mathcal{T}, \mathcal{M}, \Theta, \sigma^2 \sim \mathrm{N}\left(\alpha_i + \beta_j + \sum_{t=1}^{T} h(\mathbf{x}_{ij}, \mathcal{M}_t, \mathcal{T}_t), \sigma^2\right),$$

where y_{ij} is the yield for genotype i and environment j, and α_i and β_j are the genotype and environment effects, respectively.

$$\begin{aligned}
\mu_{t\ell}|\mathcal{T}_t &\sim \mathrm{N}(\mu_\mu = 0, \sigma^2_\mu), \\
\alpha_i|\mathcal{T}_t &\sim \mathrm{N}(\mu_g, \sigma^2_g), \\
\beta_j|\mathcal{T}_t &\sim \mathrm{N}(\mu_e, \sigma^2_e), \\
\sigma^2_g &\sim \mathrm{IG}(a_g, b_g), \\
\sigma^2_e &\sim \mathrm{IG}(a_e, b_e), \\
\sigma^2 &\sim \mathrm{IG}(a, b).
\end{aligned}$$

Some preliminary results

Simulated example

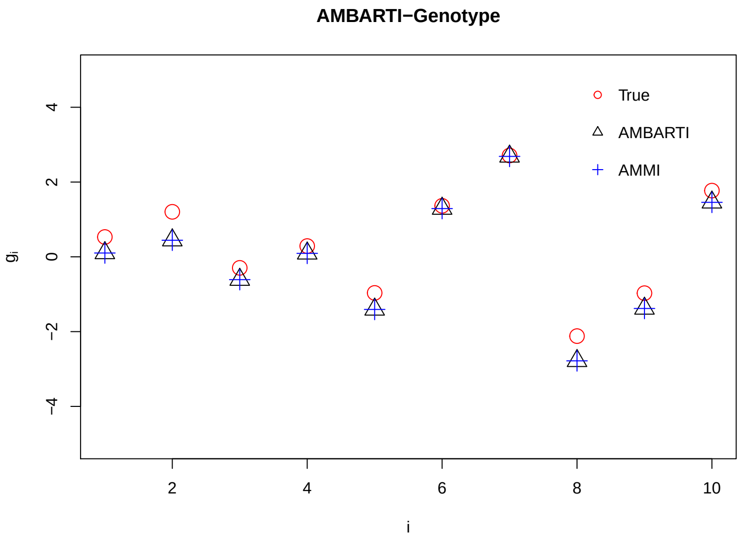
We consider the following setting to generate a simulated data set:

$$y_{ij} = \mu + \alpha_i + \beta_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq} + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}\left(0, \tau^{-1}\right),$$

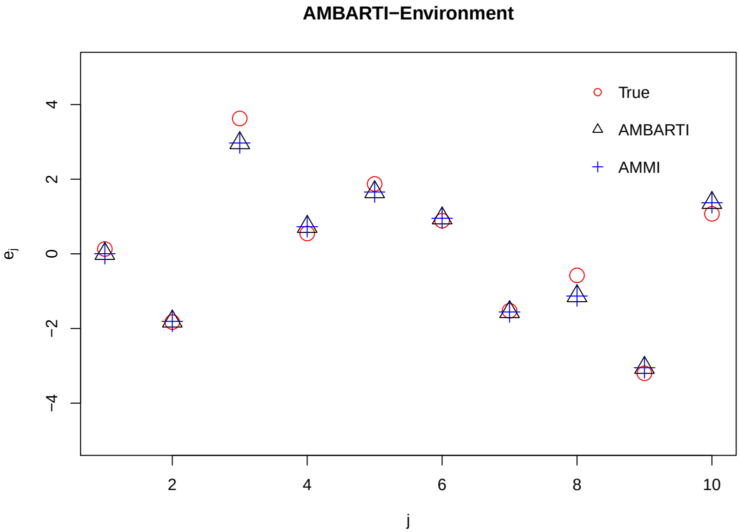
where

- $\mu = 10$, $\tau = 1$.
- $I = J = 10$.
- $Q = 1$ ($\lambda = 12$).
- $\alpha_i \sim \mathrm{N}(0, s_g)$, with $s_g = 1$.
- $\beta_j \sim \mathrm{N}(0, s_e)$, with $s_e = 2$.
- $\gamma = (-2, -1.5, ..., 1.5, 2) / \sqrt{(}10)$.
- $\delta = \left(\frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0, 0, -\frac{1}{2}, \frac{1}{2}\right)$.

Preliminary results



Preliminary results



Interesting fact. . .

Bayesian AMMI - Postprocessing (Josse et al, 2014)

Recall that

$$\mu_{ij} = \hat{y}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \sum_{q=1}^{Q} \hat{\lambda}_q \hat{\gamma}_{iq} \hat{\delta}_{jq}.$$

"This means that, concretely, S matrices of size $I \times J$ are available as draws from the posterior distributions of the μ_{ij}. Thus, it is possible to apply a postprocessing on each matrix $(s = 1, ..., S)$ performing the classical procedure (in accordance with the chosen constraints): each matrix is centered by row and by column, and an SVD is applied on the resulting matrix. Consequently, for each s, new parameters (μ, α, β, γ, δ, λ_q) meeting the constraints are available. Consequently, draws in the posterior distribution of the parameters (taking the S new values) are available. Such a postprocessing makes it easier to interpret the results."

Simulation scenarios (20 combinations)

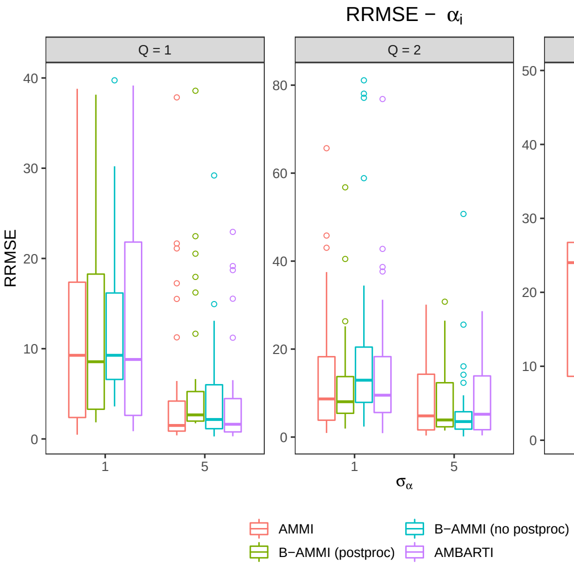
We consider the following setting to generate a set of simulated data sets:

$$y_{ij} = \mu + \alpha_i + \beta_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq} + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}\left(0, \tau^{-1}\right),$$

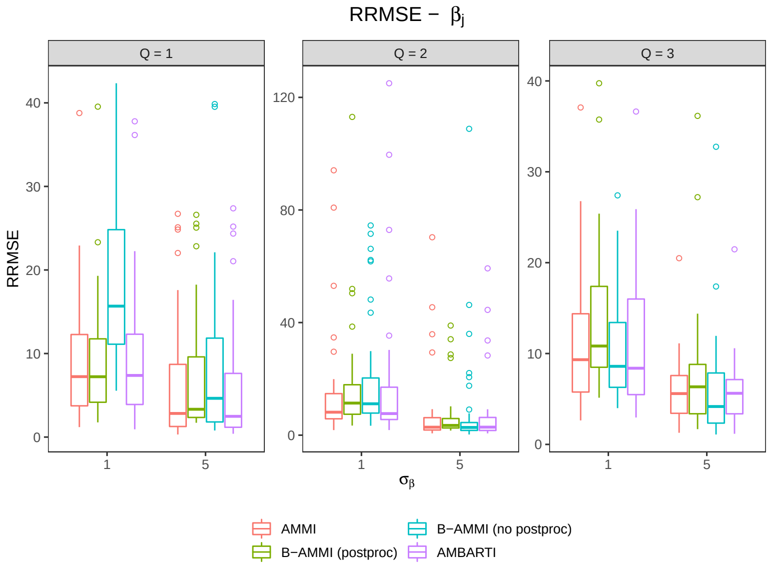
where

- $\mu = 100$, $\tau = 1$.
- $I = J = 10$ (without repetitions).
- $\beta_j \sim \mathrm{N}(0, s_e)$, with $s_e = c(1, 5)$.
- $\alpha_i \sim \mathrm{N}(0, s_g)$, with $s_g = c(1, 5)$.
- $Q = c(1, 2, 3)$, with $\lambda = c(8, 12, [8, 12], [10, 12], [8, 10, 12])$.
- We used RMSE (Root Mean Squared Error) and RRMSE (Relative Root Mean Squared Error).

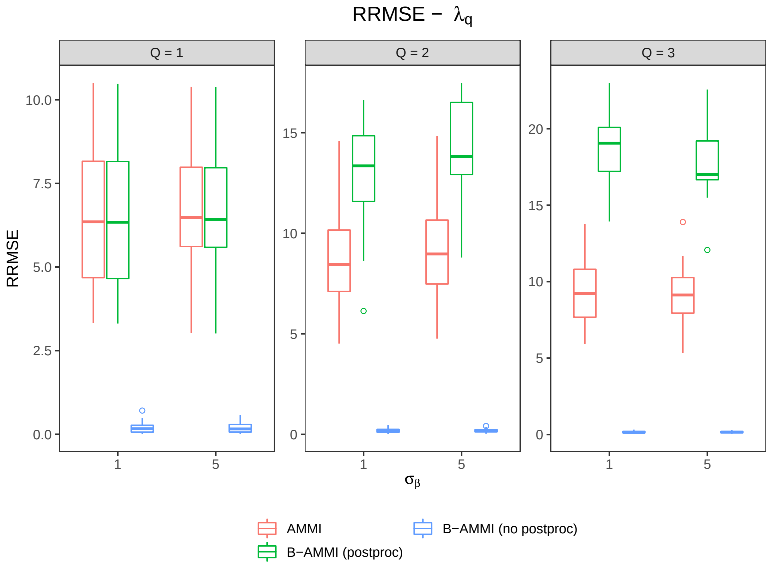
Simulation results ($y_{ij} = \mu + \alpha_i + \beta_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq}$)



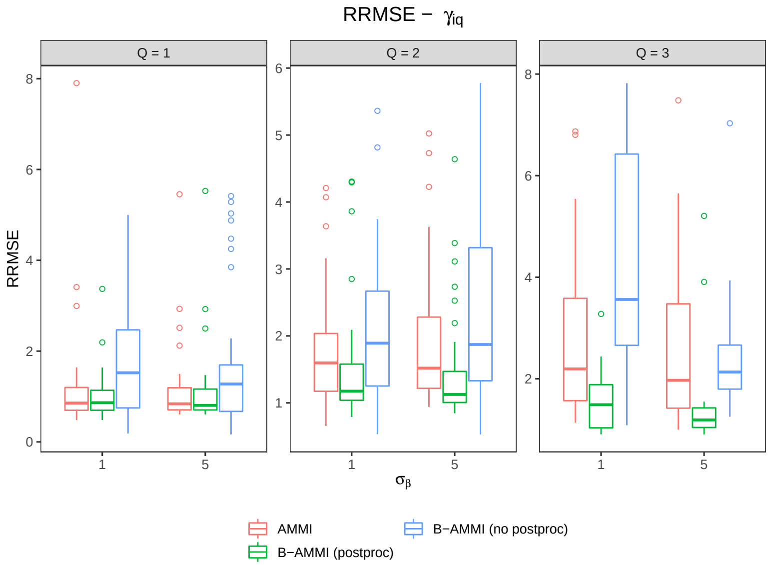
Simulation results ($y_{ij} = \mu + \alpha_i + \beta_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq}$)



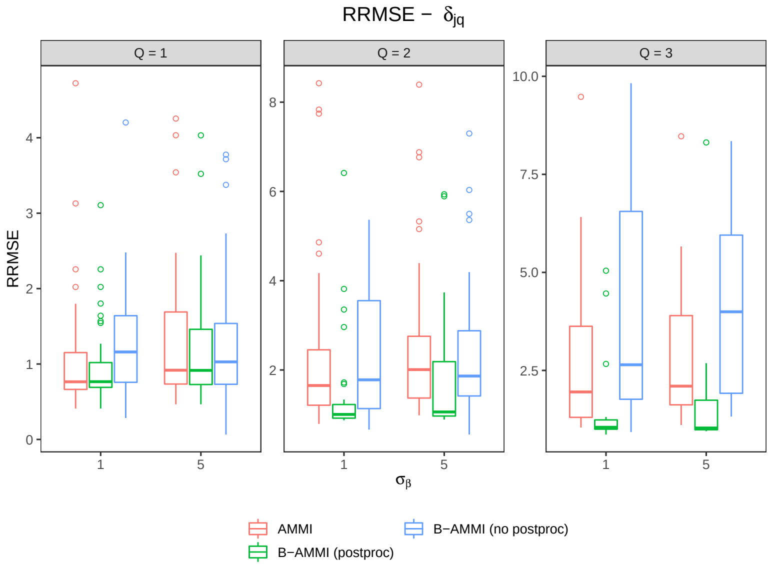
Simulation results ($y_{ij} = \mu + \alpha_i + \beta_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq}$)



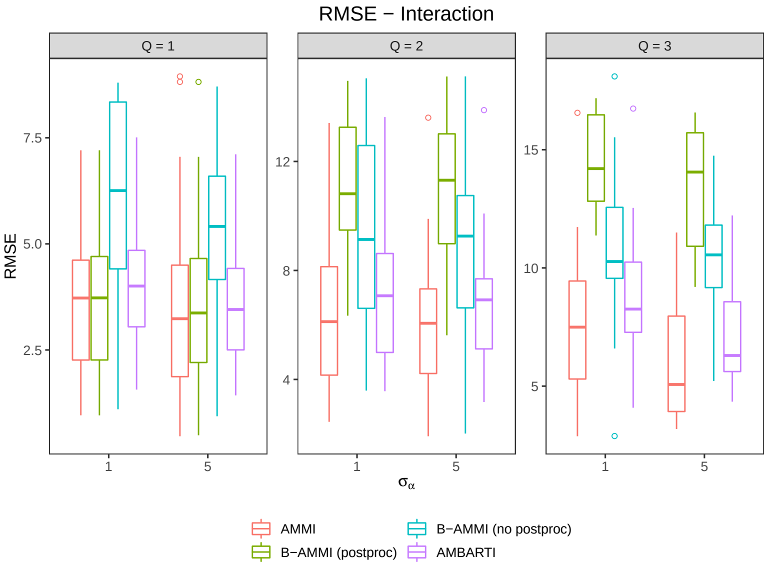
Simulation results ($y_{ij} = \mu + \alpha_i + \beta_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq}$)



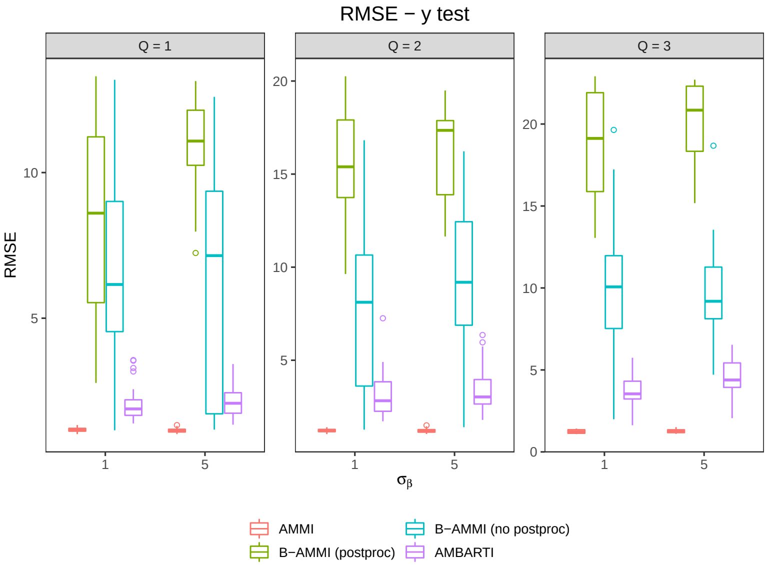
Simulation results ($y_{ij} = \mu + \alpha_i + \beta_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq}$)



Simulation results ($y_{ij} = \mu + \alpha_i + \beta_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq}$)



Simulation results ($y_{ij} = \mu + \alpha_i + \beta_j + \sum_{q=1}^{Q} \lambda_q \gamma_{iq} \delta_{jq}$)



Next steps

1. Perform more simulations, but now simulating from AMBARTI.
2. Adjust the R code to simulate large (> 15) I and J. We have talked to Rafael about it...
3. Perform more simulations considering (1) and (2).
4. Real data sets?

That's all, folks! Thank you!

This work was supported by a Science Foundation Ireland Career Development Award grant number: 17/CDA/4695

References