

Bayesian Additive Main Effects and Multiplicative Interaction Tensor Model

Antonia dos Santos¹, Rafael Moral¹, Danilo Sarti¹,
Alan Inglis¹ and Andrew Parnell¹

¹Maynooth University, Hamilton Institute

Overview

In the production of cultivars, several variables are important, such as genetic factors, location of cultivation, climates, and soils. Studying the individual effect and the interaction of these variables is therefore crucial for determining crop yield. We propose to extend the Additive Main Effects and Multiplicative Interaction (AMMI) model, to include, in addition to the effect of genotype and environment, the effect of other variables. The Bayesian Additive Main Effects and Multiplicative Interaction Tensor (BAMMIT) model, that can capture the effect of genotype, environment, and season and their interaction, and it is formulated as

$$y_{ijk} = \underbrace{\mu + g_i + e_j + t_k}_{\text{main effects}} + \underbrace{\sum_{q=1}^Q \lambda_q \gamma_{iq} \delta_{jq} \phi_{kq}}_{\text{interaction effect}} + \epsilon_{ijk},$$

where μ is the grand mean, g_i , $i \in \{1, \dots, I\}$, e_j , $j \in \{1, \dots, J\}$, and t_k , $k \in \{1, \dots, K\}$ represent the effect of the i -th genotype, j -th environment and the k -th time, respectively, λ_q , γ_{iq} , δ_{jq} and ϕ_{kq} are the parameters of the interaction term, and ϵ_{ijk} is noise distributed as $N(0, \sigma^2)$. We consider the following independent prior distributions:

$$\begin{aligned} p(\mu) &= \mathcal{N}(\mu_\mu; \sigma_\mu^2) \\ p(g) &= \prod_{i=1}^I \mathcal{N}(0; \sigma_g^2) \\ p(e) &= \prod_{j=1}^J \mathcal{N}(0; \sigma_e^2) \\ p(\lambda) &= \text{ordered sample of } \prod_{q=1}^Q \mathcal{N}^+(0; \sigma_\lambda^2) \\ p(\sigma^{-2}) &= \mathcal{G}(a; b), \end{aligned}$$

where \mathcal{N} , \mathcal{N}^+ and \mathcal{G} represent normal, truncated normal and

Gamma distributions, respectively. This model is overparametrised and it is necessary to add restrictions in order to obtain the estimates. In order to meet these model constraints, we perform a procedure on the interaction term. For example, to get γ_{iq} , let

$$\theta_{iq}^\gamma \sim N(0, \sigma_\theta^2), i = 1, \dots, I,$$

then we calculate

$$\gamma_{iq} = (\theta_{iq}^\gamma - m_q^\gamma) \left[\sum_i (\theta_{iq}^\gamma - m_q^\gamma)^2 \right]^{-1/2}.$$

A similar procedure is performed to obtain the other two parameters of the interaction term, that is, δ_{jq} and ϕ_{kq} .

Experiments

In this section, we illustrate the results for a simulation scenario, in which the data were generated considering $I = 12$, $J = 6$, $K = 6$, $\mu = 100$, $\sigma_g = 1$, $\sigma_e = 1$, $\sigma_t = 1$, $\sigma_y = 1$, $\sigma_\theta = 1$, $Q = 1$, $\lambda = 12$.

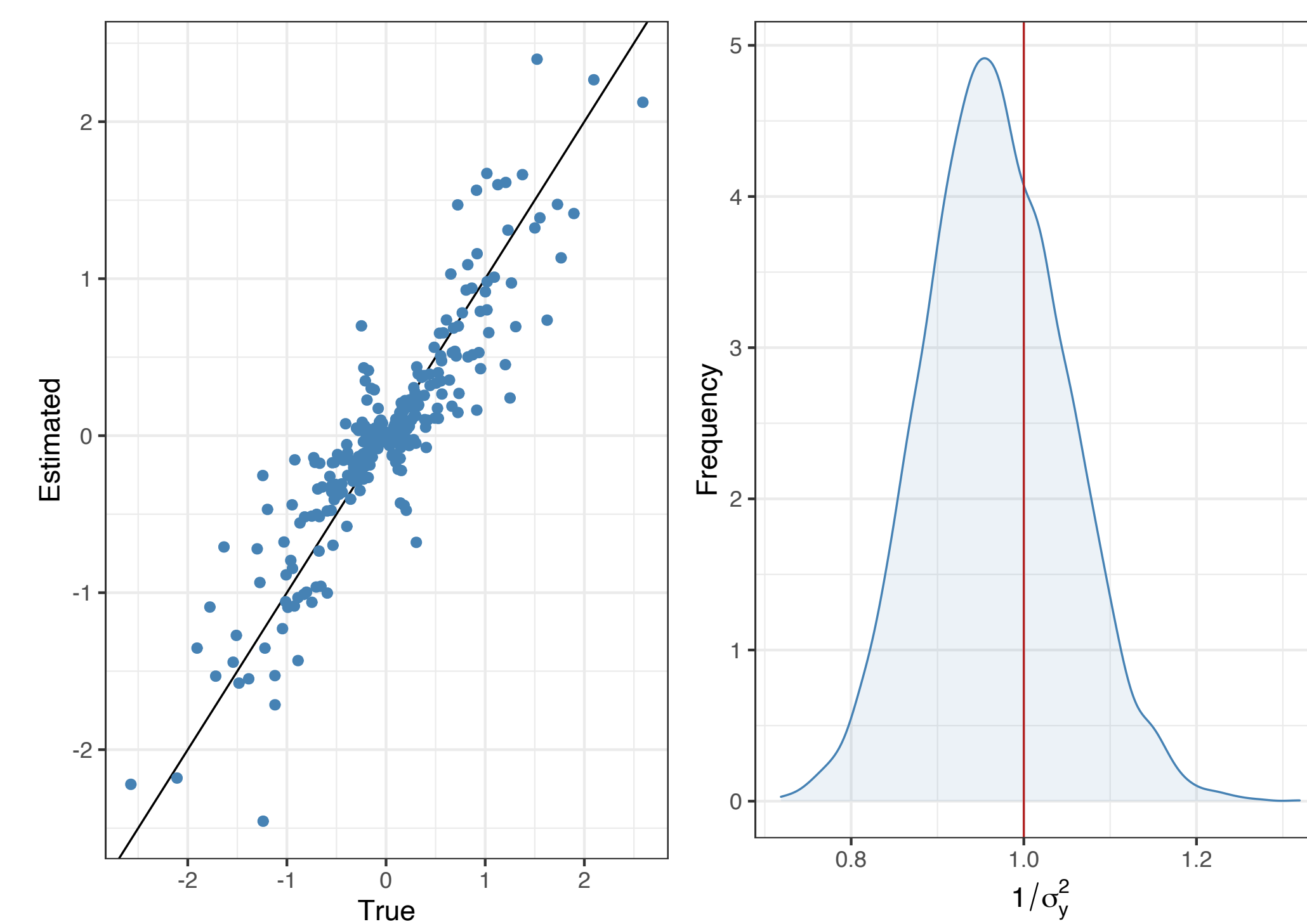


Figure 1: True interaction term vs estimated and posterior distribution of precision, $I = 12$, $J = 6$, $T = 6$, $Q = 1$, $\lambda = 12$. The red line represents the true value of the precision parameter.

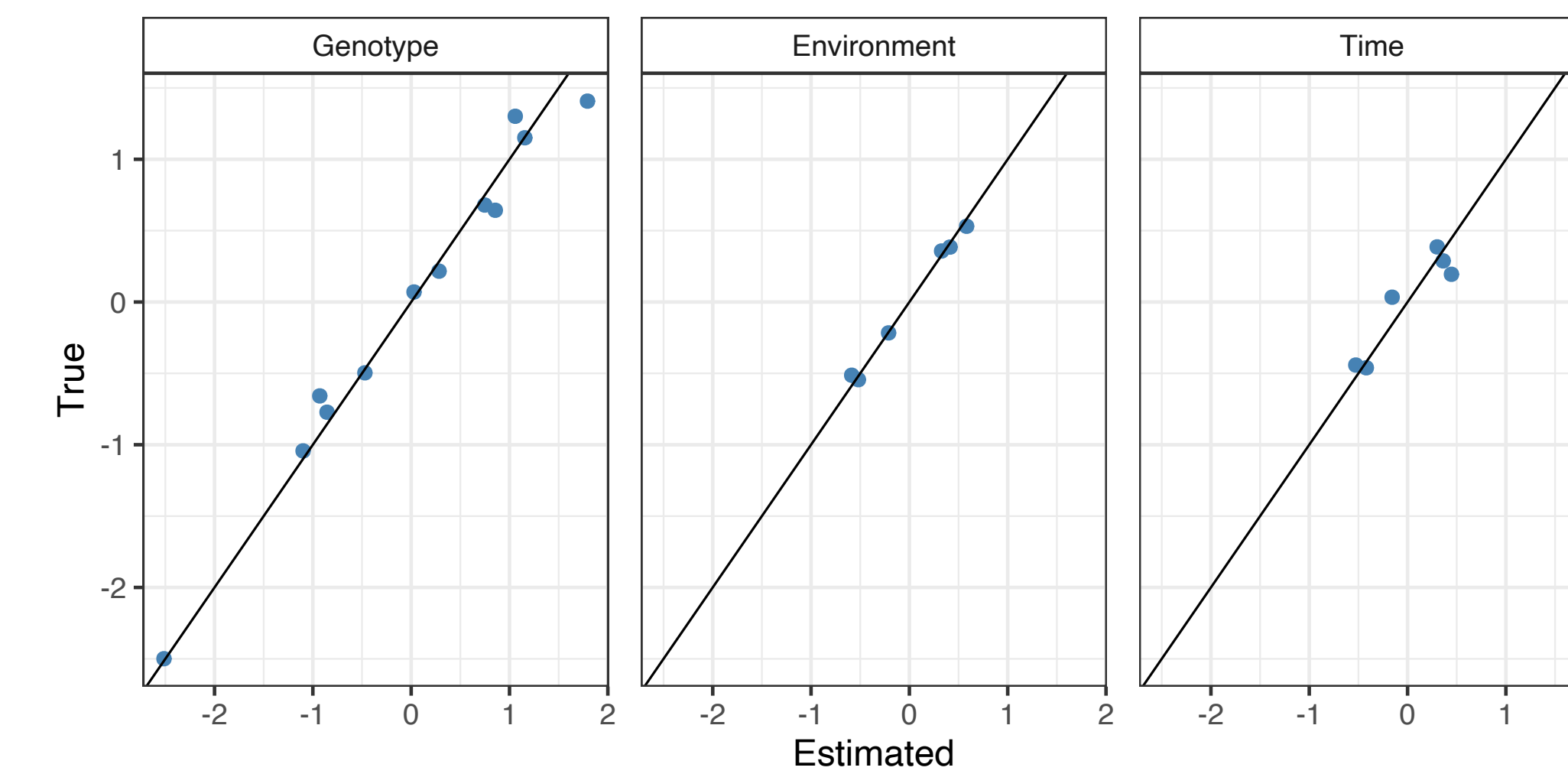


Figure 2: Graph of theoretical versus estimated values of genotypes, environments and time variables, $I = 12$, $J = 6$, $T = 6$, $Q = 1$, $\lambda = 12$.

Real Data

BAMMIT model was applied to a real data set from the Horizon2020 EU InnoVar project (www.h2020innovar.eu). We consider the production of a species of common wheat called *Triticum aestivum* L., in Ireland, where the response is the yield of wheat measured in tonnes per hectare. The data set contains 85 genotypes, 17 environments, over 10 years (2010 - 2019), and a total of 1445 observations. To run the model, we set $Q = 1$, obtain the posteriors, and from them, we take the last season to determine the best genotypes and environments.

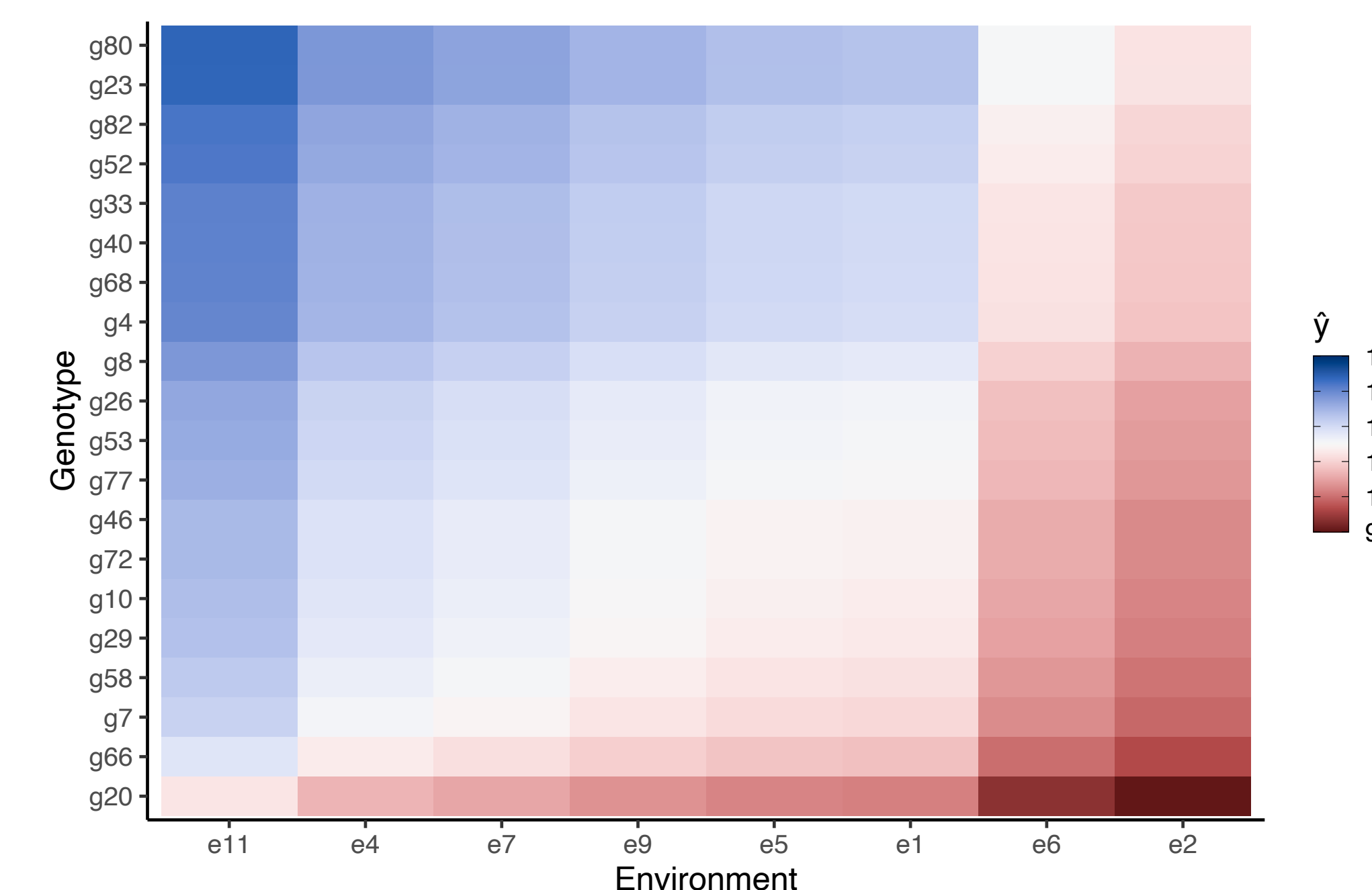


Figure 3: Predicted yields from the BAMMIT model for the data in 2019.

In Figure 3, we display the genotype x environment interaction for the data in the last season (2019), using the predictions from the BAMMIT model. The graph makes it easy to spot GxE interactions and see which environments or genotypes are best. For example, we can observe that genotype g_{23} and g_{80} in the environment e_{11} has a high predicted yield, while the genotype g_{20} in the environment e_2 has a low value. Wheat production is worst in environments e_6 and e_2 , and best in environment e_{11} .

Discussion

In this work, we propose the addition of two new parameters in the AMMI model, that measure the individual and interactive effect of a new variable, in this case, time. The results observed via simulation and the results obtained from using the real data set of the InnoVar project are shown to be good, with good accuracy and meeting the imposed restrictions.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 818144.

References

- [1] MANDEL, J. (1971). A new analysis of variance model for non-additive data. *Technometrics* 13 1–18.
- [2] Sarti, Danilo A., et al. "Bayesian Additive Regression Trees for Genotype by Environment Interaction Models." *bioRxiv* (2021).