

A Bayesian Updating Framework for Long-term Multi-Environment Trial Data in Plant Breeding – LinStat 2026

Stephan Bark, Maryna Prus and Hans-Peter Piepho

University of Hohenheim, Germany

Abstract

Multi-environment trial (METs) are essential in plant breeding to assess genotype performance across diverse climatic zones ([2]). Consequently, the key question of the design regarding the optimal allocation of a fixed total number of trials across climatic zones arises. An optimal allocation allows for more precise estimates of genotype performance within each zone, taking into account local environmental conditions ([3]). Their A-optimal design calculation starts with a linear mixed model (LMM) analysis estimated via residual (restricted) maximum likelihood (REML). Some variance components may be estimated as exactly zero when signals are weak, unintentionally removing random effects and reducing accuracy.

A fully Bayesian reformulation calculated via Markov chain Monte Carlo (MCMC) methods uses priors on variance components to keep them positive, yielding more realistic estimates ([6]). This has been addressed in this study ([1]) applied to the MET data of [4]. The corresponding Bayesian Conditional Mixed Model (BCMM) reformulation incorporates historical prior information and quantifies uncertainty in the design. To the best of the available knowledge, this study comprises one of the first serious attempts to objectively inform priors in the context of historical MET data subdivided into L successive multi-year data windows. Following the terminology of [5], this approach is referred to as the Bayesian updating approach.

It is crucial that the structural assumptions of the variance parameter priors are maintained throughout the windows to remain within the Bayesian mixed model framework. Posterior Gibbs samples of variance components are used to approximate prior distributions for the next successive data window via maximum likelihood estimation.

Our results indicate that Bayesian methods can enhance the quality of variance component estimates and consequently of optimal design strategies. The findings provide insights into improving the efficiency and accuracy in MET data analysis for more reliable genotype selection. Key challenges include prior assumptions and residual variance heterogeneity, which can pose convergence issues in Bayesian posterior sampling. Future research should reformulate design criteria to address variance heterogeneity, explore alternative prior specifications,

and enhance sampling strategies. Moreover, extending the modeling to include environmental covariates could lead to a remarkable gain in accuracy.

Keywords

Bayesian linear mixed model; multi-environment trial; residual (restricted) maximum likelihood; Markov chain Monte Carlo; optimal design.

Acknowledgments

We are grateful to the Bangladesh Rice Research Institute (BRRI) for generating, maintaining, and generously providing access to the long-term multi-environment trial data used in this study. The exploration of our statistical method would not have been possible without all the efforts of BRRI scientists, field staff, and technical personnel!

We thank Reinhard Meister for suggesting a fully Bayesian analysis of historical MET data at a seminar held by Hans-Peter Piepho at the Berliner Kolloquium “Statistische Methoden in der empirischen Forschung” in 1998. It has taken quite some time to get there, but better late than never!

References

- [1] Bark, S., Malik, W. A., Prus, M., Piepho, H.-P. and Schmid, V. (2026). A Bayesian Updating Framework for Long-term Multi-Environment Trial Data in Plant Breeding. Submitted manuscript, arXiv:2604.16203.
- [2] Kleinknecht, K., Möhring, J., Singh, K.-P., Zaidi, P.-H., Atlin, G.-N. and Piepho, H.-P. (2013). Comparison of the Performance of Best Linear Unbiased Estimation and Best Linear Unbiased Prediction of Genotype Effects from Zoned Indian Maize Data. *Crop Science* 53, 1384–1391.
- [3] Prus, M. and Piepho, H.-P. (2024). Optimizing the Allocation of Trials to Sub-regions in Crop Variety Testing with Multiple Years and Locations. *Journal of Agricultural, Biological and Environmental Statistics*.
- [4] Rahman, N. M. F., Malik, W. A., Kabir, M. S., Baten, M. A., Hossain, M. I., Rudra Paul, D. N. et al. (2023). 50 years of rice breeding in Bangladesh: genetic yield trends. *Theoretical and Applied Genetics* 136, 18.
- [5] Sorensen, D. and Gianola, D. (2002). *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer, New York.
- [6] Studnicki, M. and Piepho, H.-P. (2024). Hierarchical modelling of variance components makes analysis of resolvable incomplete block designs more efficient. *Theoretical and Applied Genetics* 137, 134.