

Revision and Response to Reviewers' Comments

on the manuscript (FIELD-D-21-01010R1)

Bayesian Inference of Spatially Correlated Random Parameters for On-farm Experiment

Zhanglong Cao¹, Katia Stefanova¹, Mark Gibberd^{1,2}, and Suman Rakshit^{1,3}

¹SAGI West, School of Molecular and Life Sciences, Curtin University, Perth, Australia

²Centre for Crop and Disease Management, School of Molecular and Life Sciences, Curtin University, Perth, Australia

³School of Electrical Engineering, Computing, and Mathematical Sciences, Curtin University, Perth, Australia

February 3, 2022

We thank the Associate Editor and Reviewers for their helpful comments. We have tried to comply with all the recommendations, and the **red text** within our response refers to actions taken to modify the paper.

Comments

Reviewer 1 writes: *This revision is greatly improved. I few remaining issues are as follows:*

1. *The authors may have misunderstood my comment about resolution. This was not about the total number of data points as such but about the spatial level at which input rates are varied. The blocked design with three replicates varies treatment levels at a very high level because the same treatment is applied across a large plot. I think a smaller-scale variation of treatment levels would lead to more accurate estimates of treatment-level optima. That is why I had asked if the authors have access to another dataset where the spatial scale of treatment-level variation is smaller. I can understand if the authors do not have access to such a dataset and so want to stick with their analysis. This is fine, but perhaps the spatial scale issue could be briefly discussed.*

We first provide our quick and short response to the above comment by stating that, at the moment, we do not have access to any other data sets from large-scale on-farm experiments. Below we provide a detailed response addressing the comment about the spatial scale of the variation of treatment levels in large strip experiments.

We agree with the Reviewer-1 that if the treatments are varied at a smaller-scale, the treatment effects at a given location can be estimated more precisely. However, smaller-scale variation of treatment levels, similar to what is typically observed in a small-plot experiment, is often infeasible (if not impossible) for paddock-scale strip experiments. The spatial scale of the variation of treatment levels in a large strip experiment is given by the strip width used in the experiment, and the strip width in almost all cases is determined by the width of farmer's machinery, particularly the large spreaders and harvesters used in paddock-scale experiments. It is common these days to select the strip width based on a harvester's width, and this results in strips even wider than those used for the Las Rosas corn field experiment. The typical strip experiments that we are currently observing in Western Australia are using strip width between 24m and 36m.

Although the spatial scale of treatment variation in these large strip experiments is larger than that in small-plot experiments, this scale is practically tiny relative to the paddock size. Furthermore, this scale of variation is appropriate for implementing any changes in management practices at a local scale in large paddocks. For example, in a paddock of size 1000 hectares (10^7 square meter), it is more practical to implement any changes at the spatial scale of 100 or 250 square meter than at an extremely fine spatial scale of 15 or 20 square meter.

In response to the above comment, we have included the following sentences in lines 39-50 of our current draft of the paper:

"The spatial scale at which treatments are varied in these strip trials is larger than that is observed in typical small-plot trials. Because a finer spatial scale may lead to more precise estimates of the treatment effects, the aim always is to incorporate the narrowest possible treatment strips. However, in practice, the width of a strip in a paddock-scale strip experiment is determined by the size of the machinery (e.g., spreader's width or harvester's width) used in conducting the experiment. Other designs such as the *chequerboard* and *eggbox* designs may incorporate a relatively finer scale of treatment variation over space using the variable rate technology (Cook et al. 2018). However, the strip trials are often cheaper and easier to implement, and thus, more attractive

to farmers than these other designs. Furthermore, because the spatial scale of treatment variation in a large strip experiment is reasonably small relative to the size of the experiment, we can estimate the spatially-varying parameters quite successfully using such a trial. The resulting map of optimum treatment levels from such trials is often practically useful for farmers in terms of the spatial scale at which they are comfortable implementing any changes in management practices at the local scale within a paddock. In this paper, we focus on the analysis of large strip experiments.”

2. *I am not sure all the cited references advocate the usage of unrandomized designs for large strip trials in general. Such designs may be preferable in situations where they allow implementing a much finer-scale variation of treatment levels. When large plots are used that have the same treatment level applied everywhere on the same plot, then randomization has some clear advantages, especially when it comes to global inferences. Perhaps the authors can re-consider their statements in this regard, also in view of my comment (1).*

We agree with the Reviewer-1 that when the aim is global inference then randomisation has clear advantage. We do not discredit randomisation when the goal is the unbiased estimation of the global effects. In lines 64–68, we clearly mention the utility of randomisation for experiments aimed at global inference:

“The analysis of a small-plot trial typically assumes a spatially-invariant global treatment effect, as the main objective here is to obtain an unbiased estimate of the treatment effect. The unbiased estimation in small plot trials is ensured through appropriate *randomisation* in experimental designs, and the spatial variation is accounted for by fitting a spatially correlated covariance structure to the error terms (Gilmour et al. 1997; Stefanova et al. 2009). Randomisation does not play the same crucial role in the analysis of large strip experiments — a systematic design is more suitable for estimating spatially-varying treatment effects (Rakshit et al. 2020; Piepho et al. 2011; Evans et al. 2020).”

We have clearly stated that our aim is not global inference in this paper. The aim is to estimate the local effects using Bayesian inference for large strip experiments.

The references (Rakshit et al. 2020; Piepho et al. 2011; Evans et al. 2020) mentioned in the paper have advocated the use of systematic design when the goal is local inference.

Piepho et al. (2011) commented “...we see one exception, where a systematic design may be preferred over a randomized design, and this occurs when local predictions of response curves need to be derived ...”.

Rakshit et al. (2020) commented “We assume a systematic layout of the strips where within each replication the same randomized treatment arrangement is used. Although such layout is antithetical to traditional small-plot experimental design, systematic designs are extremely useful for estimating local treatment effects.”

Evans et al. (2020) commented “However, standard methods are frequently not suitable for the analysis of on-farm experiments that are conducted using large plots in simple designs that may be systematic rather than randomised, and where there are multiple geo-referenced measurements recorded for each plot.”

Furthermore, we are working on a paper that would report an extensive simulation study comparing the randomised versus systematic design for estimating spatially-varying treatment effects. Our initial simulation results have clearly identified the systematic design to be a better design for estimating local treatment effects. We plan to submit this paper for publication by the end of this year.

In conclusion, we do not feel that we need to change any of the statements in our current draft in regards to the design layout for large strip experiments.

3. *The Bayesian R^2 used by the authors does account for spatial covariance in that it is based on a model allowing for such covariance. However, the measure of total variance employed in this measure just evaluates the marginal variance and as such ignores spatial covariance. To illustrate, consider the case of two observations y_1 and y_2 . The authors measure would implicitly use the average marginal variance $1/2 * (\text{Var}(y_1) + \text{Var}(y_2))$, whereas Piepho (2019) would use the semi-variance $1/2 * (\text{Var}(y_1) + \text{Var}(y_2) - 2\text{Cov}(y_1, y_2))$, a very natural measure of total variance in a geostatistical setting. Note that the “res” in the authors’ R^2 are spatially correlated, and this spatial correlation is ignored, as far as I can make out from the equation given for R^2 .*

The Bayesian R^2 used in the manuscript is correspondence to the coefficient, which is used to determine the random effects in LMM, in Section 5 (Piepho 2019). Perhaps the term “ R^2 ” confuses readers, but to be consistent with Gelman et al. 2019; Selle et al. 2019, we use the term “Bayesian R^2 ” in our manuscript to distinguish conventional R^2 .

Regarding the semi-variance, we agree with the reviewer that it is a measure of variance commonly used for spatially correlated data, and the average semi-variance is more appropriate for LMM. With the formula that is presented by Piepho (2019), we calculated the average semi-variances (ASV) for four models. The results are consistent with Bayesian R^2 that Model 1 and Model 3 have large ASV, and Model 2 and Model 4, by taking into account the spatial covariance, have small ASV. Moreover, ASV of Model 4 is slightly smaller than that of Model 2. The conclusion from ASV and Bayesian R^2 is the same. We added the statement and discussion into lines 536–540:

A coefficient of determination for random effects of a linear mixed model and a generalised linear mixed model is proposed by Piepho (2019). The coefficient is corresponding to Bayesian R^2 . The author also proposed to use averaged semivariance (ASV), which is a measure of variance commonly used for spatially correlated data, and concluded that ASV is preferable for LMMs. We calculated ASV for four models and the results are consistent. The full results are listed in Table 1.

	Mean	SD	2.5%	Median	97.5%
Model 1	554.986	19.047	519.458	554.441	593.611
Model 2	85.340	5.844	74.484	85.037	97.151
Model 3	637.510	33.657	576.394	635.357	708.738
Model 4	76.441	5.617	66.449	76.042	88.233

Table 1: Summary statistics of the average semi-variances (ASV) calculated from posterior samples of four models. Mean, standard deviation (SD), 95% credibility interval (showing 2.5% and 97.5% sample quantiles) and median are reported.

Moreover, our point of using Bayesian R^2 in the manuscript is that the concept can mislead researchers and statistician in the case that a misspecified model has a larger Bayesian R^2 value. Hence, the Bayesian R^2 is not valid if the model is misspecified. Therefore, it should not be interpreted solely, and model checking/diagnostic should be conducted in advance.

In Lines 494, we state that “Hence, the potential model misspecification is not detected. Besides, some researchers use Bayesian R^2 as the index in model comparison. However, The Bayesian R^2 is misleading in some situations, and it should not be interpreted solely, such as the example in the paper. ”

References

- Cook, S., M. Lacoste, F. Evans, M. Ridout, M. Gibberd, and T. Oberthür (2018). “An on-farm experimental philosophy for farmer-centric digital innovation”. In: (cit. on p. 1).
- Evans, F. H., A. Recalde Salas, S. Rakshit, C. A. Scanlan, and S. E. Cook (Nov. 2020). “Assessment of the Use of Geographically Weighted Regression for Analysis of Large On-Farm Experiments and Implications for Practical Application”. en. In: *Agronomy* 10.11, p. 1720. DOI: [10.3390/agronomy10111720](https://doi.org/10.3390/agronomy10111720) (cit. on p. 2).
- Gelman, A., B. Goodrich, J. Gabry, and A. Vehtari (2019). “R-Squared for Bayesian Regression Models”. In: *American Statistician* 73.3, pp. 307–309. DOI: [10.1080/00031305.2018.1549100](https://doi.org/10.1080/00031305.2018.1549100) (cit. on p. 2).
- Gilmour, A. R., B. R. Cullis, and A. P. Verbyla (1997). “Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments”. In: *Journal of Agricultural, Biological, and Environmental Statistics* 2.3, pp. 269–293. DOI: [10.2307/1400446](https://doi.org/10.2307/1400446) (cit. on p. 2).
- Piepho, H. P. (2019). “A Coefficient of Determination (R^2) for Generalized Linear Mixed Models”. In: *Biometrical J.* 61.4, pp. 860–872. DOI: [10.1002/bimj.201800270](https://doi.org/10.1002/bimj.201800270) (cit. on pp. 2, 3).
- Piepho, H.-P., C. Richter, J. Spilke, K. Hartung, and A. Kunick (2011). “Statistical Aspects of On-Farm Experimentation”. In: *Crop & Pasture Science* 62, pp. 721–735. DOI: [10.1071/cp11175](https://doi.org/10.1071/cp11175) (cit. on p. 2).
- Rakshit, S., A. Baddeley, K. Stefanova, K. Reeves, K. Chen, Z. Cao, F. Evans, and M. Gibberd (2020). “Novel Approach to the Analysis of Spatially-Varying Treatment Effects in on-Farm Experiments”. In: *Field Crops Research* 255.October 2019, p. 107783. DOI: [10/gg2vv7](https://doi.org/10/gg2vv7) (cit. on p. 2).
- Selle, M. L., I. Steinsland, J. M. Hickey, and G. Gorjanc (2019). “Flexible Modelling of Spatial Variation in Agricultural Field Trials with the R Package INLA”. In: *Theoretical and Applied Genetics* 132.12, pp. 3277–3293. DOI: [10.1007/s00122-019-03424-y](https://doi.org/10.1007/s00122-019-03424-y) (cit. on p. 2).
- Stefanova, K. T., A. B. Smith, and B. R. Cullis (Dec. 2009). “Enhanced Diagnostics for the Spatial Analysis of Field Trials”. en. In: *JABES* 14.4, p. 392. DOI: [10.1198/jabes.2009.07098](https://doi.org/10.1198/jabes.2009.07098) (cit. on p. 2).