Reviewer #1: The authors implement a spatially explicit random coefficient model for data from on-farm trials that allows mapping the optimal levels of an agronomic input. They use a Bayesian framework and illustrate the approach using a randomized complete block trial with six nitrogen treatment levels.

I applaud the authors for implementing this model, which is a major step forward in the analysis of spatially referenced on-farm trial data. I have a couple of comments for the authors.

(1) Why did they not simply use a REML package to fit the model? If there are any major obstacles, these should be discussed, as many users would use REML as a default approach.

The use of the REML package is discussed in our *Random Coefficient Regression* paper, which is in preparation. To implement the model in the REML package, we split the process into two phases: in phase 1, we estimated the correlation parameters of AR1xAR1 covariance through a two-dimensional grid search; in phase 2, the final model is estimated. The entire process is prolonged compared to the Bayesian model in our paper. Additionally, all parameters are estimated at the same time by the NUTS sampler. We compare and discuss the two approaches in the RCR paper. This paper compares the Bayesian approach with GWR and overcomes the disadvantages, such as bandwidth selection.

(2) The example is a randomized complete block design with three reps and six nitrogen levels. I would think that this design is superbly inadequate for assessing locally varying optimal input rates. Nitrogen levels would need to be varied on a finer grid, as the authors hint themselves in the introduction. Can they not find a more suitable dataset to illustrate their modelling approach? If not, the paper is best characterized as a proof-of-concept.

A few references have stated that to obtain an optimal treatment map, a systmatic design overperforms to a randomised design. We found that the results are consistent through simulation studies. However, the model of obtaining such a treatment map was not properly developed until the GWR paper (Suman, 2020). With the adoption of Bayesian approach, we are comparing the results with GWR on the same data set.

Besides, we proposed the Bayesian workflow and discussed potential model misspecification, which are inadequately discussed in references.

(3) The paper has a lot of mathematical detail. This is unavoidable with this kind of modelling and an absolute necessity. That being said, it seems to me that the example could play a much more central role in guiding the uninitiated reader. For example, the authors never even state that the purpose of their analysis is to determine the locally varying rate of nitrogen input! This key objective, along with a brief sketch of the example, would sit well in the introduction, thus triggering the readers' interest. Without such a teaser, my prediction is that 99.9% of the readers will put the paper down after the first few equations. Also, when stating the model for the first time, explain what is in Zu!! This is the key component of the model. Moreover, explain that you will be fitting a quadratic polynomial, as this is essential in order to be able to determine an optimum! Also tell readers how this is done, even if it seems too obvious to a statistician: 99.9% of your readership will not be statisticians.

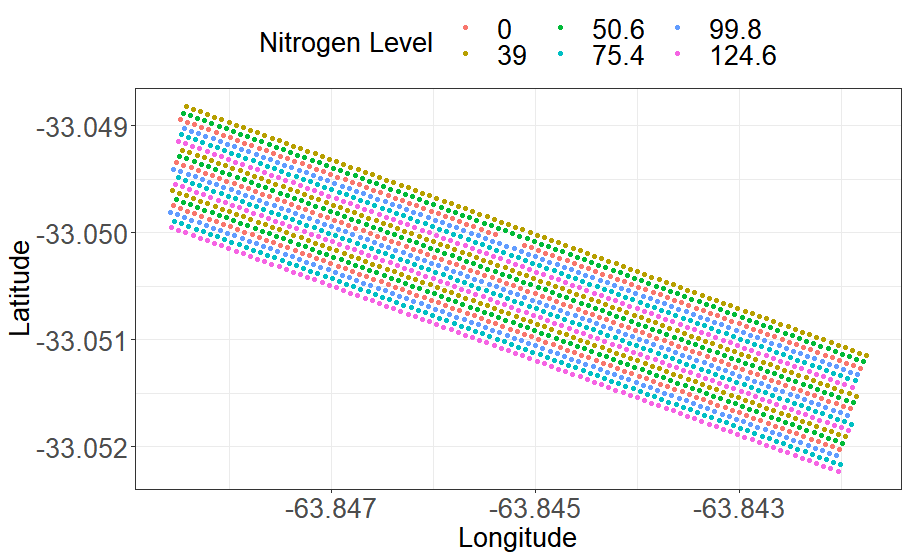
We introduced our research object in Line 47: Our aim in this paper is to obtain spatially-varying estimates of treatment effects, which in turn enables the creation of spatial maps of optimum treatment levels for large paddocks.

(4) The Bayesian R2 in eq. (27) does not account for spatial covariance. It uses just the marginal variance. This seems inappropriate for the data at hand, which are spatially correlated, a feature that is, in fact, central to the whole modelling approach. In linear mixed models in general, and in a spatial context in particular, there are pairwise correlations among observations. A natural way to account for this is the semivariance and averaging this across all pairs of observations, or integrating across the random field, naturally leads to an R2 that does account for correlation. See

Piepho, H.P. (2019): A coefficient of determination (R2) for generalized linear mixed models. Biometrical Journal 61, 860-872.

(5) The field layout of the treatment design should be shown. Were treatments randomized within complete reps? How many observations were there per plot? What were the six treatment levels? This is absolutely key information, without which it is impossible to judge the results of the analysis. For example, it is unclear that the predicted optima of nitrogen input are within the observed range of levels.

Need to add a figure of the treatment.



Added line 274: which are systematically allocated.

Added figure 1 (b)

(6) It would be useful to show a model with fixed regression terms, i.e. V\_u = 0, as a benchmark, as this is what would be routinely used for this kind of experiment. In fact, showing that V\_u is not zero would be central in order to demonstrate that it is worthwhile to fit spatially varying coefficients and that the central hypothesis of precision farming is valid (see Piepho et al. 2011, whom the authors are citing).

Our model didn’t include the conventional fixed and random terms, such as the replicate structure. It is because the replicate factor is not significant. Alternatively, we use nitrogen treatment levels as both fixed and random terms. Our assumption is that there is a global trend on the treatment against yield, and the local treatment in each grid is adjusted by the model.

(7) How did the authors determine the optimal inputs? Some explanation of the algebra would be useful. I suspect these are inputs maximizing yield, but as a farmer I'd be more concerned with economic optima. What about credible intervals for the predicted optima? How wide are they? Can this width be mapped as well?

Added: Line 281 To obtain the map of locally varying optimal input rates, we assume a quadratic form of the response in model \eqref{eq:underlying}. In other words, the levels of fixed and random effects are $l=3$ and $k=3$. Hence, the optimal treatment is possible by calculating a simple quadratic equation for each grid when the coefficients become available.

(8) How was the topographic factor incorporated into the model, and how does this factor align with blocks? It would help to see the estimates of all the fixed effects of the models fitted.

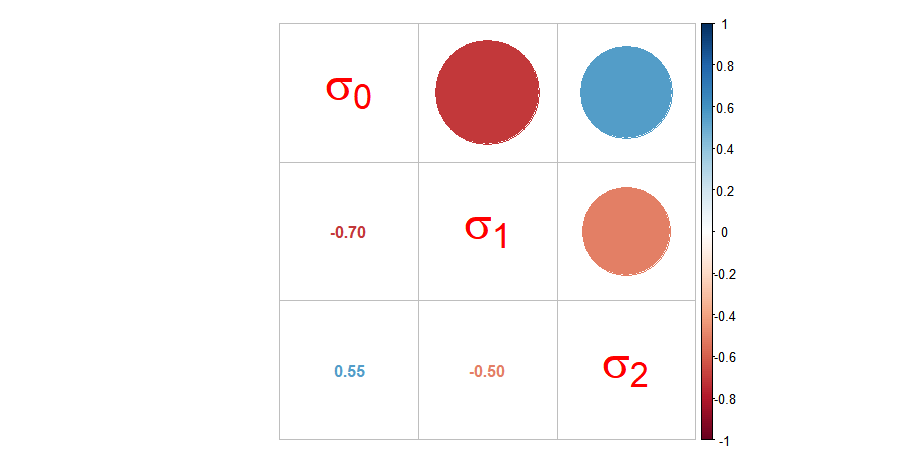
The block/replicate factor is not significant compared to the topographic factor. We used topographic factors as both fixed and random terms in our study, and the posterior checking shows severe model misspecification even though the R2 value is high. The model might be overfitted.

Being more ambitious, sigma\_u ~ topo. But the model will be more complex and slows down the computing speed.

(9) Eq. (11): What does LKJ stand for? What are the parameters of this model? I did not find and output for the actual correlations or covariances among the random coefficients. Can these be reported? Random coefficient models are notoriously difficult to fit, and proper scaling of the covariates is usually essential. Did the authors encounter such issues and how did they deal with them?

Added to line 174: a positive definite correlation matrix sampled from the Lewandowski-Kurowicka-Joe (LKJ) distribution.

As shown in Figure 2. The correlation is reported in Table 5. But the plot of correlation was not included.



Further comments:

Eq. (17): The ARxAR model vor V\_s has different parameters than the one for Sigma\_e. Somehow, this should be reflected in the notation of both model components.

In conventional statistical model in equation (3-5), the covariance matrix is imposed to \Sigma\_e. But for the proposed Bayesian model, the covariance matrix is incorporated with the random parameter u. So for Sigma\_e, it is just sigma\_e\*I.

Figure 8: This map only shows optimum levels for a very limited part of the field. What about the big white area??!!

This proves that the quadratic relationship is not significant. It is more like a linear relationship. The results are consistent with GWR, where the adjusted P-values are more than 0.05 for the quadratic term. It is explained in line 381.

L405: I do not think it is correct to say that the influence of the prior is washed out if the model is good enough. It is the amount of data that determines the influence of the prior.

It can be “washed out” from the reference Gabry2019Visualization

Reviewer2

The manuscript deals with an important topic of modelling spatial variability in large on-farm trials. A Bayesian framework is adopted to estimate the posterior distribution of parameters. Also, the proposed method is applied on a real on-farm strip trial from Las Rosas, Argentina, with the aim of obtaining a spatial map of optimal nitrogen rates for the entire paddock. The manuscript requires revision before it can be accepted for publication. My specific comments are listed below:

1) Why was weakly informative prior preferred? How do you define a weakly or strongly informative prior?

Ref Gabry2019Visualization. In particular, we say that a prior leads to a weakly informative joint prior data-generating process if draws from the prior data-generating distribution could represent any data set that could plausibly be p(y) observed. A strongly informative prior might have a smaller deviation and/or high influence on the posterior distribution, which is constrained in exploring the posterior space.

2) Four models are used in the analysis with conditions of with or without spatial correlation. Why was the model uncertainty of each of them not characterised?

Reference Gelman et al. (2019), The uncertainty and misspecification are identified by LOO PIT and Pareto k values.

3) Figure 3 presents the realisations of 100 simulations. It is known that more informative prior will give better results than vague prior. The results do not suggest anything new. The authors could have analysed the uncertainty of model prediction.

It is not new, but it provides a visualisation perspective on prior checking, rather than picking priors from references and evaluating them by posterior checking.

4) What is expected error variance in equation (27)? How to determine it? Will the error in spatial variability problem be linear?

Reference Gelman et al. (2019) Line 267: residual variance.

5) In equation (28) how are correlation parameters determined?

Example in figure 2. It is a positive definite correlation matrix sampled from the Lewandowski-Kurowicka-Joe (LKJ) distribution.

6) In Figure 4, PP(posterior predictive) checking is done against the observed data Y. But you have used the same dataset to update the parameters and obtained posterior distribution of parameters. Hence, the posterior predictive results will be close to the observations. The authors could have checked the reliability of the model by performing PP checking on some other dataset obtained from the same site.

The PP check is like a simulation checking method that uses posterior distribution to generates “new” data and compare it with the observed data set. It is used to check the performance of the parameters by comparing two data. Equation 24.