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

Zhanglong Cao[1], Katia Stefanova[1], Mark Gibberd[1,2], and Suman Rakshit[1,3]

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

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

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

Abstract

Accounting for spatial variability is crucial while estimating treatment effects in large on-farm trials. It allows to determine the optimal treatment for every part of a paddock, resulting in a management strategy that improves sustainability and profitability of the farm. We specify a model with spatially correlated random parameters to account for the spatial variability in large on-farm trials. A Bayesian framework has been adopted to estimate the posterior distribution of these parameters. By accounting for spatial variability, this framework allows the estimation of spatially-varying treatment effects in large on-farm trials. Several approaches have been proposed in the past for assessing spatial variability. However, these approaches lack an adequate discussion of the potential problem of model misspecification. Often the Gaussian distribution is assumed for the response variable, and this assumption is rarely investigated. Using Bayesian post sampling tools, we show how to diagnose the problem of model misspecification. To illustrate the applicability of our proposed method, we analysed a real on-farm strip trial from Las Rosas, Argentina, with the main aim of obtaining a spatial map of locally-varying optimal nitrogen rates for the entire paddock. The analysis of these data revealed that the assumption of Gaussian distribution for the response variable is unsatisfactory; the Student-t distribution provides a more robust inference. We finish the paper by discussing the difference between the proposed Bayesian approach and geographically weighted regression, and comparing the results of these two approaches.

Keywords: Geographically weighted regression, Geostatistics, Large strip trials, No-U-turn sampler, Precision agriculture, Site-specific management.

1 Introduction

Traditional agricultural experiments are often conducted on small plots, and more often than not, these experiments do not address the main concerns of an individual farmer. For a farmer, one of the main motivations of conducting an experiment is to identify a management strategy that could improve the profitability of a farm. This closely aligns

with the objective of *site-specific farming*, typically enabled by *precision agriculture* technologies (Cook and Bramley 1998). The main aim is to identify the optimal strategy of input utilisation for every part of a large paddock. Because of inherent spatial variation in a large paddock, a uniform management strategy for the entire paddock is sub-optimal. An optimal strategy may require the identification of location-specific optimal treatments that could vary across the paddock. Small plot experiments are inadequate for obtaining a spatially-varying map of optimal treatments for a paddock (Rakshit et al. 2020; Evans et al. 2020). Consequently, there is an increasing trend to conduct on-farm experiments (OFE) using large strips across farmers' paddocks, utilising their own tools and machinery (Yan et al. 2002; Rakshit et al. 2020; Evans et al. 2020).

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.

Spatial variation in OFE may introduce bias while estimating treatment effects and inflate associated standard errors if not accounted for in fitted models. Spatial variation may be caused by environmental factors such as soil fertility, moisture trends, and light exposure (Selle et al. 2019), or it could also arise due to management practices with reoccurring patterns (Gilmour et al. 1997; Hinkelmann 2012). Two common approaches of tackling spatial variation are through the modelling of a *nonstationary* mean structure or modelling of a spatially *autocorrelated* error structure (Fotheringham 2009; Harris 2019). However, these two forms of spatial variation are quite difficult to disentangle from each other. The following statement from (Cressie 1993) articulates this point: *"What is one person's (spatial) covariance structure may be another person's mean structure."*

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. This further allows an investigation of the central hypothesis of precision agriculture that the optimum treatment varies spatially within a paddock (Páez et al. 2002; Brunsdon et al. 1999; Lark and Wheeler 2003; Pringle et al. 2010). To obtain spatially-varying treatment effects, we incorporate spatial heterogeneity in our modelling framework, which is quite different than the traditional models used for analysing small plot trials (Rakshit et al. 2020; Piepho et al. 2011). 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 propose a Bayesian framework for modelling the nonstationary first-order effect, characterised by the conditional mean of the response variable, for any location within a paddock. We first specify a regression function with spatially-varying coefficients, representing local departures of treatment effects from their global estimates (Banerjee et al. 2004). Appropriate *prior* distributions are considered next for the model parameters, and finally, the spatially-varying estimates are computed by sampling from the *posterior* distributions. The proposed modelling framework is used to determine the locally-varying optimum nitrogen rates for a real-life large strip experiment from Las Rosas, Argentina (details of the analysis are provided below in Section 5 and the results in Section 6).

There have been efforts in the recent past to estimate spatially-varying treatment effects for large strip experiments (Lawes and Bramley 2012; Marchant et al. 2019; Rakshit et al. 2020; Evans et al. 2020). However, some of these approaches can be considered as merely ad hoc solutions to the problem, particularly restricted to comparing adjacent strips in a large strip trial (Lawes and Bramley 2012). A more statistically principled approach, called *geographically weighted regression* (GWR), is proposed by Rakshit et al. (2020) for estimating spatially-varying treatment effects in large strip experiments, based on the general theory of *local likelihood estimation* (Hastie and Loader 1993). GWR is fairly easy to implement using open-source software and provides a pragmatic solution to support on-farm decision making (Evans et al. 2020). However, a crucial step in GWR is the bandwidth selection for kernel functions. Inaccurate bandwidth may introduce unknown bias in estimated coefficients. Because the optimal bandwidth size would always be unknown for a given dataset, one needs to use some data-based methods to select an appropriate bandwidth. See Rakshit et al. (2020) for a discussion on the topic of bandwidth selection for on-farm strip experiments.

The Bayesian framework proposed in this paper simplifies statistical inference by providing straightforward interpretation of the results (Che and Xu 2010). Statistical inference using GWR is not straightforward, as it involves adjusting for the problem of multiple testing. In particular, localised p-values are required to be adjusted to avoid a large number of false positives in the spatial map of treatment effects; see Rakshit et al. (2020) for the details of computing adjusted p-values in GWR. Due to the availability of adequate computing resources and due to the fact that both model fitting and statistical inference under Bayesian framework are extremely intuitive, Bayesian modelling has become popular for analysing agricultural field trials in the last few years (Besag and Higdon 1999; Theobald et al. 2002; Che and Xu 2010; Donald et al. 2011; Montesinos-López et al. 2018; Selle et al. 2019; Shirley et al. 2020). Montesinos-López et al. (2018) proposed a multivariate Bayesian analysis to estimate multiple-trait and multiple-environment on-farm data. Selle et al. (2019) compared popular spatial models and proposed a Bayesian modelling framework for variety selection in plant breeding experiments. Jiang et al. (2009) used Bayesian conditional auto-regressive models to account for spatial autocorrelation in OFE data. However, none of these approaches is useful to fit a regression function with spatially-varying coefficients. These methods are also inadequate for developing a management practice that may lead to the optimal use of input resources.

For modelling spatial nonstationarity, we adopted a Bayesian hierarchical model with spatially correlated random parameters. We use the No-U-Turn Sampler (NUTS) (Hoffman and Gelman 2014) for performing Bayesian inference. NUTS is an efficient sampler that allows quick exploration of the posterior distribution in high dimensional space. NUTS was developed by extending the popular Hamiltonian Monte Carlo (HMC) algorithm to address a crucial

drawback of HMC — it is highly sensitive to two user-specified parameters: a step size ϵ and the desired number of steps L. NUTS determines the step size during the warm-up (burn-in) phase while aiming at a target acceptance rate, and then uses the chosen step size for all subsequent sampling iterations (Monnahan et al. 2017). It also eliminates the need to set the number of steps L; see Hoffman and Gelman 2014 for a detailed discussion on this topic.

We investigate the potential problem of model misspecification during the stages of post-sampling posterior diagnoses and model evaluation. To this end, we utilised advanced model diagnostic tools, such as probability integral transformation (PIT) checks (Gabry et al. 2019), and model evaluation methods, such as Bayesian leave-one-out (LOO) cross validation (CV) (Vehtari et al. 2017) and Bayesian R^2 (Gelman et al. 2019).

The paper is organised as follows. In Section 2, we specify the regression model for analysing the data from large strip experiments; in Section 3 we describe the prior and posterior distribution for the model, and explain the mechanism of NUTS sampler; in Section 4 we discuss the post-sampling model checking and diagnostic process; finally, in Section 5 and Section 6, we apply the proposed model to Las Rosas corn yield data set, and compare with the results obtained from GWR.

2 Statistical models

We describe here a Bayesian hierarchical regression model for analysing data from a large strip experiment. We start by first introducing in Section 2.1 the linear mixed effects model used to analyse typical small-plot field experiments. We extend this model to analyse large strip experiments under a Bayesian hierarchical modelling framework in Section 2.2, and finally in Section 2.3 we show how to incorporate a spatially-correlated structure for model parameters into the Bayesian modelling framework.

2.1 Statistical model for field experiments

A field experiment can be considered as a rectangular array, consisting of r rows and c columns, where the total number of observed data points is $n = r \times c$. We adopt the notation used by Zimmerman and Harville (1991), in which $s_i \in \mathbb{R}^2$, $i = 1, \ldots, n$, is a two-cell vector of the Cartesian coordinates of the plot centroid corresponding to the ith plot. Let $y(s_i)$ be the real-valued response variable corresponding to the ith plot, and let $\boldsymbol{Y}$ denote the vector consisting of response data from all n plots, ordered as rows within columns. Then a linear mixed effects model for $\boldsymbol{Y}$, using the matrix notation, is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{b}$ and $\boldsymbol{u}$ are vectors of fixed and random effects, respectively, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the associated design matrices, and $\boldsymbol{e}$ is the residual error vector. It is typically assumed that the vectors $\boldsymbol{u}$ and $\boldsymbol{e}$ are distributed independently of each other, and that their joint distribution is a multivariate Gaussian distribution such that

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_u & 0 \\ 0 & \Sigma_e \end{bmatrix} \right), \tag{2}$$

and

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{b}, \boldsymbol{Z}\Sigma_u\boldsymbol{Z}^\top + \Sigma_e), \tag{3}$$

where Σ_u and Σ_e are variance-covariance matrices corresponding to the random vectors $\boldsymbol{u}$ and $\boldsymbol{e}$, respectively. The fixed term $\boldsymbol{b}$ in (1) typically represents the treatment effects under consideration, and the random term $\boldsymbol{u}$ represents the effects of blocking units imposed by an experimental design (Piepho et al. 2003). The residuals $\boldsymbol{e}$ are often assumed spatially correlated.

The covariance matrix Σ_e can accommodate a separable first or second order autoregressive process to model the spatial correlation of plot residuals. Gilmour et al. (1997) suggested a separable first-order autoregressive AR1 $\times$ AR1 process with column and row correlation matrices Σ_c and Σ_r, respectively, to model the residual covariance structure, and this is given by

$$\Sigma_e = \sigma^2\Sigma_c(\rho_c) \otimes \Sigma_r(\rho_r), \tag{4}$$

where $\otimes$ denotes the Kronecker product, σ^2 is the residual variance component, and the parameters ρ_c and ρ_r determine the strengths of spatial correlations in the column and row directions, respectively. Note that sorting the data rows within columns produces the neat representation (4) of Σ_e in terms of two correlation matrices. In the case where there is no spatial autocorrelation, the residual variance-covariance matrix becomes $\Sigma_e = \sigma^2 I_n$.

In the context of our large strip-trial example in Section 5, the fixed effects $\boldsymbol{b}$ would correspond to the (global) regression parameters for the entire trial, and the random effects $\boldsymbol{u}$ would correspond to the local departures from these global parameters. When analysing a small plot experiment using a linear mixed effects model with h random terms in which each term is of dimension $m \times 1$, it is typically assumed that these terms are independently distributed of each other, imposing a direct sum structure on the variance matrix $\Sigma_u = \oplus_{j=1}^{h}\sigma^2_{u_j} I_m$ for $\boldsymbol{u}$ with variance components $\sigma^2_{u_j}$, $j = 1, \ldots, h$ (Butler et al. 2009). In contrast, any analysis of a large strip experiment would require to incorporate correlated random effects in the regression model to account for the spatial correlations amongst treatment effects.

2.2 Bayesian hierarchical model

In the context of large strip experiments, we have n grid points instead of n plots, as defined for general field experiments in the section above. At each of these n grid points, the response variable is measured, and the values of treatment factor and other spatial covariates are recorded. This is similar to the setup considered by Rakshit et al. (2020). These authors proposed a GWR model for analysing data arising from large strip experiments. GWR allows spatial nonstationarity in modelled relationships and estimates spatially-varying parameters governing these relationships by maximising local loglikelihoods. The regression function defined in GWR can also be written in the form of a linear mixed effects model, given in (1). The main difference between the linear mixed effects model (1) and the Bayesian approach is that, in the Bayesian model, we treat both model components $\boldsymbol{b}$ and $\boldsymbol{u}$ as random vectors, i.e., some (prior) distributions are specified for both $\boldsymbol{b}$ and $\boldsymbol{u}$, along with a distribution for the error term (Bürkner 2017). Consequently, the uncertainty associated with the estimates of the model parameters can be derived using posterior distributions.

Using the grid point specific notation of the response variable (i.e., $y(s_i)$) introduced in the previous section, the

underlying model for analysing a large strip experiment is given by

$$\begin{aligned} y(s_i) &= \sum_{m=1}^{l} b_m x_m(s_i) + \sum_{j=1}^{h} u_j(s_i) z_j(s_i) + e(s_i), \\ \boldsymbol{u}_i \mid \theta_u &\sim \mathcal{N}(0, V_u(\theta_u)), \\ e(s_i) \mid \sigma_e &\sim \mathcal{N}(0, \sigma_e^2), \end{aligned} \tag{5}$$

where $b_1, \ldots, b_l$ are global effects corresponding to the l explanatory variables $x_1, \ldots, x_l$; $z_1, \ldots, z_h$ denote h variables whose effects are fitted as local effects; $u_j(s_i)$ denotes the local effect corresponding to z_j at grid $s_i \in \mathcal{S}$; $\boldsymbol{u}_i = (u_1(s_i), \ldots, u_h(s_i))^\top$ is the vector of all local effects at s_i, $i = 1, \ldots, n$; θ_u is a set of parameters of the covariance matrix V_u, and σ_e is the error standard deviation, assumed to be distributed as either Gamma, half-Cauchy, or half-normal.

Because both components $\boldsymbol{b}$ and $\boldsymbol{u}$ under the Bayesian framework are considered random, the use of the term "fixed effects" when referring to the vector $\boldsymbol{b}$ may seem inappropriate. However, it is common to use this term when linear mixed effects models are fitted under a Bayesian framework; see Zhao et al. (2006) and Fong et al. (2010) for details. In this paper we shall use both terms "fixed effects" and "global effects" to refer to the model parameters b_m, $m = 1, \ldots, l$.

A regression model of particular interest is the quadratic response model, used to model the example data set in Section 5. The term associated with the global effects in (5) would take the form:

$$b_1 + b_2 x(s_i) + b_3 x^2(s_i),\ i = 1, \ldots, n, \tag{6}$$

where $x(s_i)$ is the particular level of some controllable treatment applied at location s_i. Local departures from the global treatment effects b_2 and b_3 take the form:

$$u_1(s_i) + u_2(s_i) x(s_i) + u_3(s_i) x^2(s_i),\ i = 1, \ldots, n, \tag{7}$$

where $u_1(s_i)$, $u_2(s_i)$, and $u_3(s_i)$ are spatially correlated local effects corresponding to the location s_i. See Piepho et al. (2011) for detailed description of the quadratic response model.

2.3 Model with spatially correlated random parameters

To incorporate spatial correlation amongst the model parameters in our Bayesian hierarchical modelling framework, we investigate here how the variance-covariance matrix of $\boldsymbol{u}$ can be specified to represent the spatial correlation across all the grid points s_i, $i = 1, \ldots, n$. Note that, at location s_i, the covariance matrix of $\boldsymbol{u}_i$ is V_u (5).

Without any spatial correlation between grid points, the variance-covariance matrix of the random parameters is

$$\Sigma_u = I_n \otimes V_u. \tag{8}$$

If the correlation between grid points is characterised by a spatial variance-covariance matrix V_s, the variance-

covariance matrix of $\boldsymbol{u}$ is given by

$$\Sigma_u = V_s \otimes V_u, \tag{9}$$

where V_s may be considered either a AR1 $\times$ AR1 spatial variance-covariance matrix or a weighted distance matrix. The model (8) implies correlation within grid points, but not between grid points. This is a simple model to fit, but may be unrealistic for modelling treatment effects of a large strip experiment. The model (9) imposes spatial correlation both within and between grid points, and thus, allows us to estimate the spatially-varying treatment effects across the whole field. Despite that only a single treatment is directly observed at each grid point, the estimation of localised treatment effects $\boldsymbol{u}_i$ is possible due to the fact that the spatial model (9) allows the use of information from neighbouring plots with other treatments (Piepho et al. 2011). In what follows, we incorporate the spatial structure specified in (9) into our Bayesian modelling framework (5).

3 Bayesian process

Suppose $\theta \in \Theta$ is the set of all parameters under consideration in (5). For a given $f : \Theta \to \mathbb{R}$, the main focus in the Bayesian approach is to estimate $f(\theta)$, typically by its conditional expectation, which is given by

$$\mathrm{E}[f(\theta) \mid \boldsymbol{Y}] = \int_{\Theta} f(\theta) p(\theta \mid \boldsymbol{Y}) d\theta. \tag{10}$$

Assuming a prior distribution for θ and applying the Bayes theorem we obtain the posterior density function $p(\theta \mid \boldsymbol{Y})$, which, subsequently, leads to the solution in (10).

In the rest of this section, we discuss the analytical tools that are essential for our Bayesian modelling of the real-life on-farm data from Las Rosas, Argentina, described below in Section 5.

3.1 Prior specification

The main difference between the REML and Bayesian estimation is that, in Bayesian modelling, we assume that the model parameters are random variables and estimate them using their posterior distributions. The estimation starts with the specification of a prior distribution, which may summarise the previous knowledge about the parameters (Onofri et al. 2019). Therefore, the prior distributions can be specified even before conducting the experiment.

The selection of priors in Bayesian inference has been discussed for a long time. Usually, if nothing is known from earlier studies, we can use a flat non-informative prior $p(\theta)(\propto$ constant), also called an "improper prior" (Gelman et al. 2006). In many circumstances, a Cauchy or Gamma prior is a reasonable candidate for regression coefficients. Some researchers prefer inverse Wishart (IW) or inverse Gamma as the prior distribution for the standard deviation parameter of a hierarchical model, while Gelman et al. (2006) and Gelman et al. (2017) suggested using weakly informative priors for variance parameters for Bayesian analyses of hierarchical linear model. In the cases when the number of groups is small, a half-t family is also recommended.

To specify a prior distribution for the parameters associated with the variance-covariance matrix V_u, note that the matrix can be decomposed as follows:

$$V_u = B(\sigma_u) R_u B(\sigma_u), \tag{11}$$

where $B(\sigma_u)$ denotes the diagonal matrix with diagonal elements $\sigma_{u_1}, \ldots, \sigma_{u_h}$, the standard deviation of $u_1, \ldots, u_h$, and R_u is the matrix whose diagonal elements are equal to unity and off-diagonal elements are the correlation coefficients (details are given in (29)) between the random effects. The prior distribution of V_u can now be specified by specifying priors separately for $B(\sigma_u)$ and R_u (McElreath 2015). A possible choice of a prior for the standard deviation parameters σ_{u_j} in $B(\sigma_u)$ is an inverse Wishart distribution (Kass and Natarajan 2006); another choice is an inverse Gamma distribution. However, in our setting, a weakly informative prior is preferred. We adopted the half-normal distribution in our work for all σ_{u_j}, $j = 1, \ldots, h$. For the matrix R_u with correlation coefficients, we specify the Lewandowski-Kurowicka-Joe (LKJ) distribution (Lewandowski et al. 2009) as the prior distribution, and this specification is given by

$$R_u \sim \text{LKJcorr}(\epsilon), \tag{12}$$

where LKJcorr(ϵ) is a positive definite correlation matrix sampled from the LKJ distribution that depends on the value of a positive parameter ϵ. The parameter ϵ controls the correlations in a way that, as the value of ϵ increases, the correlations amongst parameters decrease. An useful feature of our prior selection process is that the selected priors would adaptively regularise the individual coefficients of random effects and the associated correlation coefficients; see Gelman et al. (2017) and Gabry et al. (2019) for more details.

3.2 Likelihood and posterior distribution

In precision agriculture, the focus is on, firstly, determining the optimal treatment (e.g., the most productive nitrogen rate) for every part of the field, and then applying the spatially-varying optimal treatments to the entire field as part of a site-specific management strategy. To this end, an important quantity is

$$p(\boldsymbol{X} \mid \boldsymbol{Y}) = \int p(\boldsymbol{X} \mid \boldsymbol{Y}, \theta) p(\boldsymbol{Y}, \theta) d\theta, \tag{13}$$

the conditional probability of $\boldsymbol{X}$ given the response, computed by integrating out the set of unknown parameters θ.

In order to estimate θ conditional on $\boldsymbol{Y}$, we use the Bayes theorem to obtain the joint posterior density of the parameters in terms of the likelihood $p(\boldsymbol{Y} \mid \theta)$ and the prior $\pi(\theta)$ as follows:

$$p(\theta \mid \boldsymbol{Y}) = \frac{p(\boldsymbol{Y} \mid \theta)\pi(\theta)}{p(\boldsymbol{Y})}, \tag{14}$$

where $p(\boldsymbol{Y}) = \int p(\boldsymbol{Y} \mid \theta)\pi(\theta)d\theta$ is the normalising constant, which is often difficult to compute. Because this constant does not affect the inference, we can ignore it while computing the posterior distribution. Consequently, the equation (14) is often written as

$$p(\theta \mid \boldsymbol{Y}) \propto p(\boldsymbol{Y} \mid \theta)\pi(\theta). \tag{15}$$

The distribution $p(\theta \mid \boldsymbol{Y})$ is the key ingredient for "Bayesian inference" of the parameter θ. The posterior distribution $p(\theta \mid \boldsymbol{Y})$ provides all information about θ conditional on the observed data (Che and Xu 2010).

Below we specify the Gaussian and Student-t log likelihoods for our problem. We obtain for multivariate Gaussian

distribution

$$\log p(\boldsymbol{Y}|\,\theta) \propto -\ \frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{b}-\boldsymbol{Z}\boldsymbol{u})^{\top}\Sigma_e^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{b}-\boldsymbol{Z}\boldsymbol{u}) - \ \frac{1}{2}\ln\det\Sigma_e, \tag{16}$$

and for multivariate Student-t distribution

$$\begin{aligned}\log p(\boldsymbol{Y}|\,\theta) \propto -\ \frac{\nu+n}{2}\ln\left(1+\frac{1}{\nu}\quad(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{b}-\boldsymbol{Z}\boldsymbol{u})^{\top}\Sigma_e^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{b}-\boldsymbol{Z}\boldsymbol{u})\right) - \frac{n}{2}\ln\nu \\ + \ln\Gamma(\frac{\nu+n}{2}) - \ln\Gamma(\frac{\nu}{2}) - \frac{1}{2}\ln\det\Sigma_e,\end{aligned} \tag{17}$$

where $\nu \geq 1$ is the degrees of freedom.

Then the posterior distribution can be calculated by combining the likelihood and prior distribution using equation (15) (Besag and Higdon 1999; Tsionas 2002).

Assuming $\boldsymbol{u} \sim \mathcal{N}(0, \Sigma_u)$, for faster gradient evaluation and sampling we impose Cholesky decomposition, such that

$$\Sigma_u = \Sigma_c \otimes \Sigma_r \otimes V_u = (L_c L_c)^{\top} \otimes (L_r L_r)^{\top} \otimes (L_u L_u)^{\top} = (L_c \otimes L_r \otimes L_u)(L_c \otimes L_r \otimes L_u)^{\top}, \tag{18}$$

where L_c, L_r, and L_u are the lower triangular Cholesky decomposition factors of the matrices Σ_c, Σ_r, and V_u, respectively. Moreover, to improve the efficiency of sampling, we also impose the following formula based on the Kronecker product property, shown in Appendix C, that

$$\tilde{\boldsymbol{u}} = (L_r \otimes L_c \otimes L_u) z_u = (L_c \otimes L_u)\tilde{z}_u L_r,^{\top} \tag{19}$$

where z_u is the length $r \times c \times k$ vector of i.i.d. samples from $\mathcal{N}(0, 1)$ and $\tilde{z}_u$ is the transformation of z_u with size $(k \times c) \times r$. The order the matrices L_c and L_r has been swapped as columns are nested within rows. It is because $c \ll r$, and the Kronecker product of small matrices is faster to compute than that of large matrices.

The predictive distribution for a new query location s^*, based on the aforementioned posterior distribution, is obtained by marginalizing over θ and is written as

$$p(y(s^*) \mid x(s^*), z(s^*), \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}) = \int p(y(s^*) \mid x(s^*), z(s^*), \theta)) p(\theta \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}) d\theta. \tag{20}$$

3.3 No U-turn sampler

Hamiltonian Monte Carlo (HMC) (Brooks et al. 2011; Duane et al. 1987) is an efficient Markov chain Monte Carlo (MCMC) method that overcomes the inefficiency associated with the random walk and with the sensitivity to correlated parameters. An important step in HMC is the drawing of a set of auxiliary momentum variables $r = \{r_1, \ldots, r_d\}$, independently from the standard normal distribution for each parameter in the set $\theta = \{\theta_1, \ldots, \theta_d\}$. The joint density function $f(\theta, r)$ of θ and r is given by

$$f(\theta, r) \propto \exp\{L(\theta) - K(r)\} = \exp\{-H(\theta, r)\}, \tag{21}$$

where $H(\theta, r)$ is the Hamiltonian system dynamics (HSD) equation with potential energy $L(\theta)$ and kinetic energy $K(r)$. An useful property of the dynamics is that it keeps the joint distribution invariant (Nishio and Arakawa 2019).

The HSD is numerically approximated in discrete time space with the leapfrog method to maintain the total energy when a new sample (θ^*, r^*) is drawn. The leapfrog method requires two parameters: (i) a step size ϵ, representing the distance between two consecutive draws, and (ii) a desired number of steps L, required to complete the process. A new sample is accepted with the probability

$$\alpha = \min\left\{1, \frac{f(\theta^*, r^*)}{f(\theta, r)}\right\}. \tag{22}$$

Because HMC can be highly sensitive to the choice of ϵ and L, and in turn, may affect the results crucially, Hoffman and Gelman (2014) proposed the No-U-Turn Sampler (NUTS), which determines the step size adaptively during the warm-up (burn-in) phase to a target acceptance rate and uses it then for all sampling iterations (Monnahan et al. 2017). The NUTS also eliminates the need to specify a value of L by using the criterion

$$\frac{d}{dt}\frac{(\theta^* - \theta)\cdot(\theta^* - \theta)}{2} = (\theta^* - \theta)\cdot\frac{d}{dt}(\theta^* - \theta) = (\theta^* - \theta)\cdot r^* < 0, \tag{23}$$

where r^* is the current momentum and $(\theta^* - \theta)$ is the distance from the initial position to the current position. The idea is that the trajectory will keep exploring the space until θ^* starts to move back towards θ.

To guarantee time reversibility and convergence to the correct distribution, NUTS uses a recursive algorithm that preserves reversibility by running the Hamiltonian simulation in both forward and backward time directions (Hoffman and Gelman 2014). This process starts by introducing a slice variable w with conditional distribution $p(w \mid \theta, r) = U(0, f(\theta, r))$, where $U(0, f(\theta, r))$ is the uniform distribution between the bounds zero and $f(\theta, r)$. The slice sampling generates a finite set of samples of the form (θ, r) during the doubling procedure and the binary tree building process by randomly taking forward and backward leapfrog steps until

$$(\theta^+ - \theta^-)\cdot r^- < 0 \quad \text{or} \quad (\theta^+ - \theta^-)\cdot r^+ < 0, \tag{24}$$

where (θ^-, r^-) and (θ^+, r^+) are the leftmost and rightmost leaves, respectively, in the subtree. The best candidate (θ^*, r^*) is uniformly sampled from the subset of all candidate values of (θ, r).

4 Post-sampling checking

A few common strategies for Bayesian model checking, as suggested by Gelman (2003), are: (1) ensuring that the posterior inference is reasonable, given the substantive context of the model; (2) assessing the sensitivity of the inference to reasonable changes in the prior distribution and the likelihood; and (3) examining whether the model is capable of generating data similar in characteristics to the observed data. See Gelman (2004), Weiss (1994), Gelman et al. (2013), and Congdon (2019) for an overview of the topic. To examine the suitability of our Bayesian model for analysing an on-farm strip experiment, we particularly focused on the third strategy of graphically checking the similarities between the observed and simulated data from the fitted model.

In an ideal situation, researchers would be able to use an independent data set, which is not used in the modelling process, to test the predictive performance of the fitted model. Alternatively, in the absence of such independent data,

one may split the observed data into training and testing data sets, and use the training data for model fitting and the test data for evaluating the predictive performance of the fitted model. However, it may not even be feasible for many experimental data in agricultural applications to reasonably split into training and testing data sets. We illustrate our Bayesian model checking procedure below in Section 5.3 using a real-life on-farm data from Las Rosas, Argentina.

4.1 Posterior predictive checking

The posterior predictive (PP) checking uses the posterior distribution of the model parameters to regenerate the observations. The idea behind this concept is that, if a model is a good fit, we should be able to use it to generate data that resemble the observed data (Gabry et al. 2019). Let $\boldsymbol{Y}^{rep}$ denote a simulated or replicated data set, generated using the posterior predictive distribution

$$p(\boldsymbol{Y}^{rep} \mid \boldsymbol{Y}) = \int p(\boldsymbol{Y}^{rep} \mid \theta) p(\theta \mid \boldsymbol{Y}) d\theta. \tag{25}$$

To assess the fitted model, several data sets are simulated from $p(\boldsymbol{Y}^{rep} \mid \boldsymbol{Y})$, and each of them is compared with the observed data $\boldsymbol{Y}$ (Dipak Dey and C.R. Rao 2005; Congdon 2019). The application of posterior predictive distributions is robust to prior specification because the details of the prior are washed out by the likelihood (Gelman et al. 2017).

4.2 Model diagnosis and evaluation

The leave-one-out (LOO) cross validation (CV) is widely used for model evaluation. It is performed by first omitting an observation and fitting the model based on the remaining data, and then by computing the predictive error associated with the omitted observation. The process is repeated for all observations, omitting one observation at a time. The predictive errors from the LOO CV are finally used to compute an estimate of the average out-of-sample predictive error for a given model. In Bayesian statistics, the expected log LOO predictive density (ELPD) is used to measure the predictive accuracy :

$$\text{elpd}_{\text{loo}} = \sum_{i=1}^{n} \log p(y_i \mid y_{-i}), \tag{26}$$

where $p(y_i \mid y_{-i}) = \int p(y_i \mid \theta) p(\theta \mid y_{-i}) d\theta$ is the LOO predictive density with the i-th observation omitted from the data set (Vehtari et al. 2017). One disadvantage of this measure is the high computational cost due to the model being refitted n times. Recently, an approximated LOO CV has been proposed by Bürkner et al. (2021), using only a single model fit and calculating the pointwise log predictive density as a fast approximation to the exact LOO CV. It uses the Pareto-smoothed importance-sampling (PSIS) algorithm (Vehtari et al. 2017), which draws n samples, each of size M, from the posterior distribution. For each observation, then the pointwise log-likelihood is computed based on the M sampled values, and the PSIS-LOO-CV estimate is computed taking a weighted sum over all n pointwise log-likelihood as follows:

$$\widehat{\text{elpd}}_{\text{psis-loo}} = \sum_{i=1}^{n} \log \left(\frac{\sum_{m=1}^{M} p(y_i \mid \theta^{(m)}) w_i^{(m)}}{\sum_{m=1}^{M} w_i^{(m)}} \right), \tag{27}$$

where $w_i^{(m)}$ are stabilised weights computed during PSIS, $m = 1, \ldots, M$. See Vehtari et al. (2017) for the details of computing the stabilised weights in PSIS.

The resulting PSIS-LOO-CV (27) can be used for model diagnosis and comparison. The advantage of PSIS is that

it automatically computes an empirical similarity between the full data predictive distribution and the LOO predictive distribution for each omitted observation in LOO CV (Gabry et al. 2019). Another useful quantity, obtained during PSIS, is the estimated tail shape parameter $\hat{k}$ of the generalised Pareto distribution. This estimate can also be used for assessing the reliability of the model. If $\hat{k} < 0.5$ the distribution of raw importance ratios has finite variance and the central limit theorem holds; see Vehtari et al. (2017) for a detailed discussion on the raw importance ratio. In practice, however, the model may still be robust for $\hat{k}$ values up to 0.7. Otherwise the variance and the mean of the raw ratios distribution do not exist (Vehtari et al. 2017).

The Bayesian R^2, proposed by Gelman et al. (2019), is used for model evaluation as well. R^2 is presented as the variance of the predicted values divided by the variance of predicted values plus the expected residual variance

$$\text{Bayesian } R^2 = \frac{\text{Var}(\boldsymbol{Y}^{pred})}{\text{Var}(\boldsymbol{Y}^{pred}) + \text{Var}(\boldsymbol{res})}. \tag{28}$$

However, it should not be interpreted solely if the model has a large number of bad Pareto $\hat{k}$ values, i.e., values greater than 0.7 or, even worse, greater than 1.

5 Analysis of a real-life large strip experiment

A part of Las Rosas data set, which is publicly available by the name of `lasrosas.corn` in the `R`-package `agridat` (Edmondson 2014), was used in our study. In this section, we adopt the proposed Bayesian approach to analyse the data set.

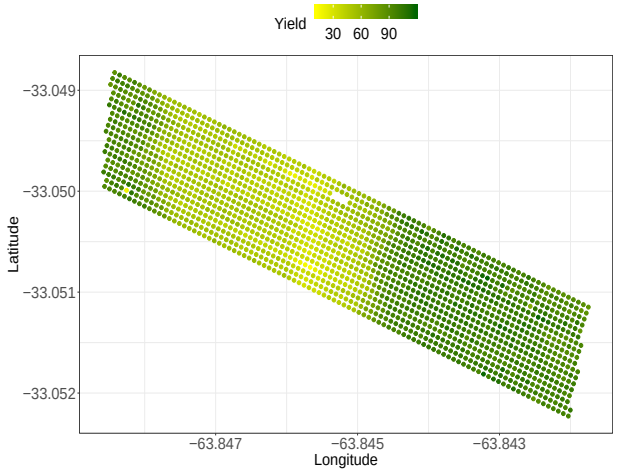
5.1 Las Rosas data

The data were produced by a yield monitor in an Argentinian corn field trial conducted by incorporating six nitrogen rates 0, 39.0, 50.6, 75.4, 99.8, and 124.6 kg/ha, which are systematically allocated in three replicated blocks comprising 18 strips (columns) and 93 rows. In order to account for some of the spatial variation (Figure 1), a four-level topographic factor was defined: W (West slope), HT (Hilltop), E (East slope) and LO (Low East).

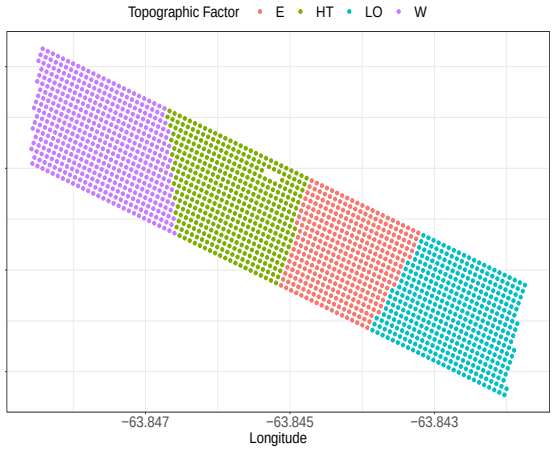
Additionally, a geographic projection was applied to the data. It transforms the geo-spatial coordinates to planar coordinates expressed in meters and assists with the model fitting (Rakshit et al. 2020). The field area of the Las Rosas experiment is approximately 810 metres long and 150 metres wide.

5.2 Statistical models and prior predictive simulations

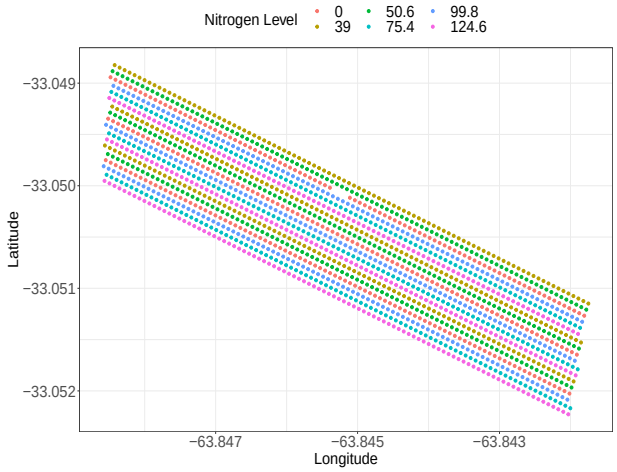
To obtain the map of locally varying optimal input rates, we specified a quadratic regression model, in which the corn yield is modelled as a quadratic function of the nitrogen rate. The optimal treatment can be determined by estimating the coefficients of the quadratic regression model at each grid point. To demonstrate the flexibility of the proposed model (5), in which the random parameters $\boldsymbol{u}$ are spatially correlated, we compare it with the one without spatial correlation, used as a benchmark model for the rest of the analysis. We also compare two distributional assumptions in the context of specifying the likelihood – the popular Gaussian likelihood has been compared with the Student-t distribution in order to assess whether the Gaussian model, often chosen as the default model, is misspecified



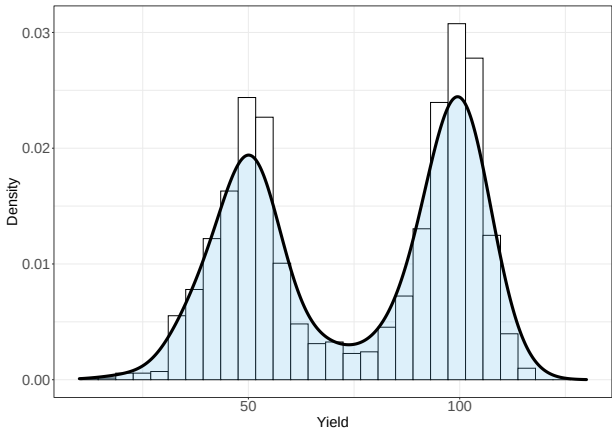
(a) Visualisation of yield. Yellow colour indicates low yield and dark green indicates high yield.



(b) Coloured by different topographic factors: West slope (W), Hilltop (HT), East slope (E) and Low East (LO).



(c) Six nitrogen treatment levels 0, 39.0, 50.6, 75.4, 99.8, and 124.6 kg/ha are systematically allocated into three replicates; the ordering, from left-to-right, used within each replication is 124.6–75.4–99.8–0–50.6–39.0.



(d) Bimodal histogram and density plot of yield.

Figure 1: Visualisation of Las Rosas yield monitor data for harvests in 2001.

for our example data set. We define our four models below in Table 1.

	Model 1	Model 2	Model 3	Model 4
Spatial correlation	No	Yes	No	Yes
Var($\boldsymbol{u}$)	$I_{n\times n} \otimes V_u$	$V_s \otimes V_u$	$I_{n\times n} \otimes V_u$	$V_s \otimes V_u$
Distribution	Gaussian	Gaussian	Student-t	Student-t

Table 1: Four models that are fitted in our study.

The modelling process starts by selecting appropriate priors for the model parameters by comparing the simulated responses and the observed responses graphically, as shown in Figure 3. In the right panel of Figure 3, we have chosen weakly informative priors to simulate responses based on the quadratic regression function. In the left panel, we show the simulated responses obtained using vague priors for the regression coefficients.

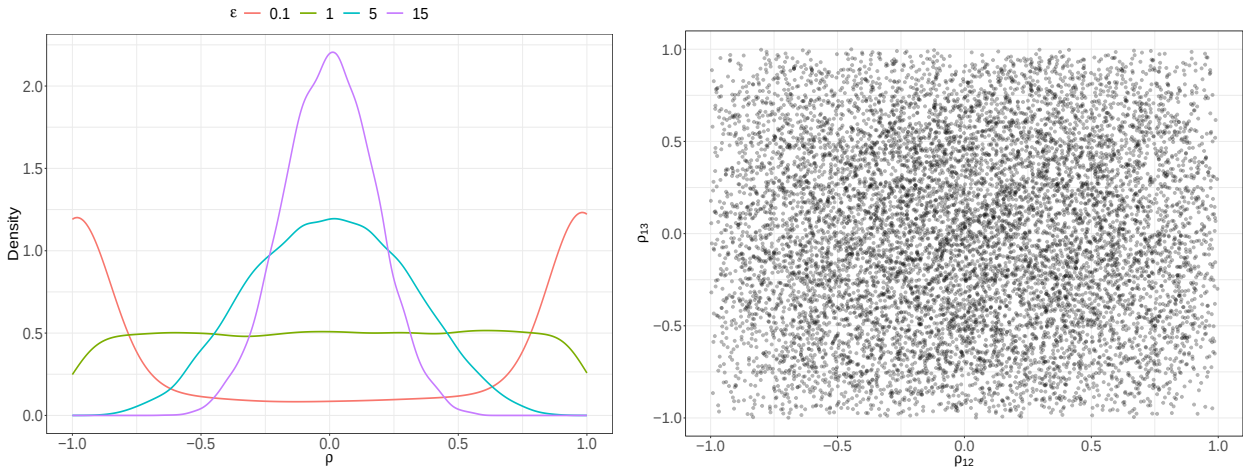
The vague priors used in our analysis are $b_0 \sim \mathcal{N}(\mu, 100)$, $b_1, b_2 \sim \mathcal{N}(0, 100)$ and $\sigma_e \sim IG(1, 100)$, where μ is the median of the observed responses and IG refers to the inverse Gamma distribution. We assume $u_{i_h} \sim \mathcal{N}(0, \sigma_h^2)$ with $\sigma_h^2 \sim IG(1, 100)$ and $h = 0, 1, 2$ at grid s_i. Alternatively, we can choose weakly informative priors $b_0 \sim \mathcal{N}(80, 10)$, $b_1 \sim \mathcal{N}(0, 0.01)$, $b_2 \sim \mathcal{N}(0, 0.001)$, $\sigma_0 \sim \mathcal{N}_+(0, 1)$, $\sigma_1 \sim \mathcal{N}_+(0, 0.01)$, $\sigma_2 \sim \mathcal{N}_+(0, 0.001)$, $R_u \sim$ LKJcorr(1) and $\sigma_e \sim \mathcal{N}_+(0, 1)$, where $\mathcal{N}_+(\cdot)$ is the positive half Gaussian distribution.

The correlation matrix R_u, defined in (11), is given by

$$R_u = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix}, \tag{29}$$

where ρs are the pairwise correlation parameters. For the correlation matrix R_u, we select LKJcorr(ϵ) with $\epsilon = 1$, which represents weak correlation amongst $\boldsymbol{u}_i$ values at grid i, $i = 1, \ldots, n$.

Figure 2 demonstrates how the distribution of ρ is influenced by ϵ. A small ϵ leads to a wider tail and a big ϵ typically narrows down the tail. In the case of $\epsilon = 1$, all correlations are equally plausible. As ϵ increases, the variables are more likely to be independent.

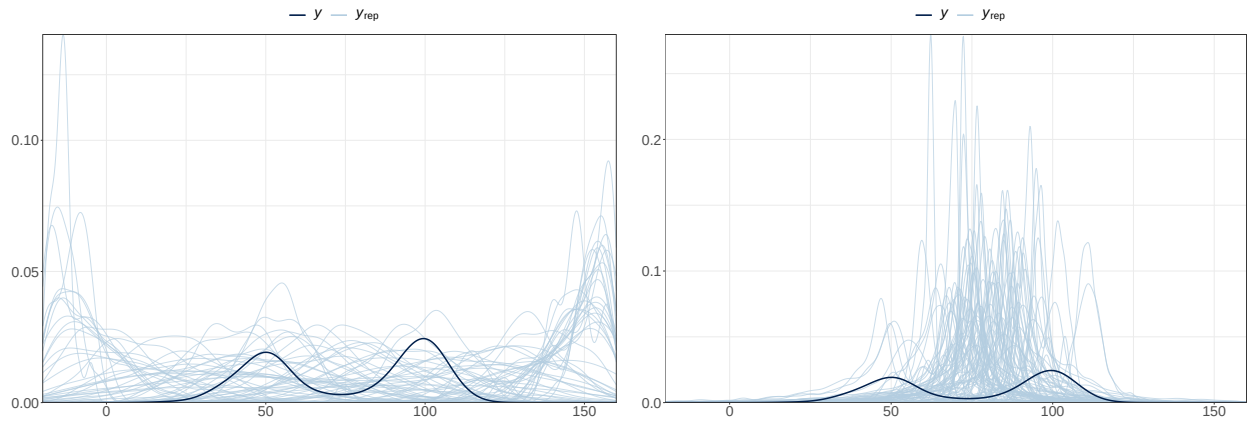


(a) Distribution of correlation coefficients ρ extracted from random 2×2 correlation matrices with different values of ϵ.

(b) Visualisation of ρ_{12} against ρ_{13} from a 3×3 correlation matrix with $\epsilon = 1$.

Figure 2: LKJcorr(ϵ) probability density.

Figure 3 compares the simulated data with vague and weakly informative priors. When the vague priors are applied, Model 1 generates extremely small and large values, which are highly unlikely for our corn yield data set. This is mostly because the vague priors disregard practical knowledge. The use of weakly informative priors avoids negative values and keeps the simulations within a reasonable interval. Even though some simulations are not perfect, the weakly informative priors overall exhibit good results that reflect commonsense knowledge about the yield response. On the other hand, if the priors are too informative, the posterior distribution maybe badly influenced and result in partial exploration of the posterior space.



(a) With vague priors

(b) With weakly informative priors

Figure 3: Capability of regenerating observed data with different priors by running 100 simulations. Vague priors failed in regenerating and lead to extreme values. Weakly informative priors give plausible regenerated data.

In Model 2, in addition to the priors used in Model 1, we need priors for the parameters ρ_c and ρ_r, and suppose $\rho_c, \rho_r \sim U(0,1)$, where $U(0,1)$ is the uniform distribution between 0 and 1. In Model 3 and 4, we have an extra parameter $\nu(\geq 1)$ for the degrees of freedom, and we specify a Gamma prior $\nu \sim \Gamma(2, 0.1)$, as suggested by Juárez and Steel (2010).

In Table 2, we present the complete list of priors selected for our study. In general, it is not recommended to use the same priors across all the different models listed in Table 2. Furthermore, if a new prior is proposed for a new parameter, examining the suitability of that prior is recommended for each model. In this study, we have checked the suitability of all the priors for all our four models, and it turns out that the same priors, listed in the top-half of Table 2, work well for all the four models (see Figure 10 for further details). Consequently, we built the models by using the same weakly informative priors for a number of common parameters, and only adding new priors (listed in the bottom-half of Table 2) for the additional parameters.

Using the above priors, our proposed hierarchical Bayesian models were run on four parallel Markov chains using the `R`-package `rstan` with each chain having a warmup period of 1000 iterations and post-warmup period of another 1000 iterations. Consequently, for each parameter, we generated 4000 samples (1000 samples from each of the four chains) from its posterior distribution.

5.3 Posterior checking

The prior predictive checking is a powerful tool for understanding the structure of the model. However, it is not possible to extend this technique to choose between competing models for the data and evaluate their predictive

	Model 1	Model 2	Model 3	Model 4
b_0		$\mathcal{N}(80, 10)$		
b_1		$\mathcal{N}(0, 0.01)$		
b_2		$\mathcal{N}(0, 0.001)$		
σ_0		$\mathcal{N}_+(0, 1)$		
σ_1		$\mathcal{N}_+(0, 0.01)$		
σ_2		$\mathcal{N}_+(0, 0.001)$		
σ_e		$\mathcal{N}_+(0, 1)$		
R_u	—	LKJcorr(1)	—	LKJcorr(1)
ρ_c	—	$U(0, 1)$	—	$U(0, 1)$
ρ_r	—	$U(0, 1)$	—	$U(0, 1)$
ν	—	—	$\Gamma(2, 0.1)$	$\Gamma(2, 0.1)$

Table 2: Priors of four models. Top-half: priors for common parameters of four models; Bottom-half: new priors for additional parameters.

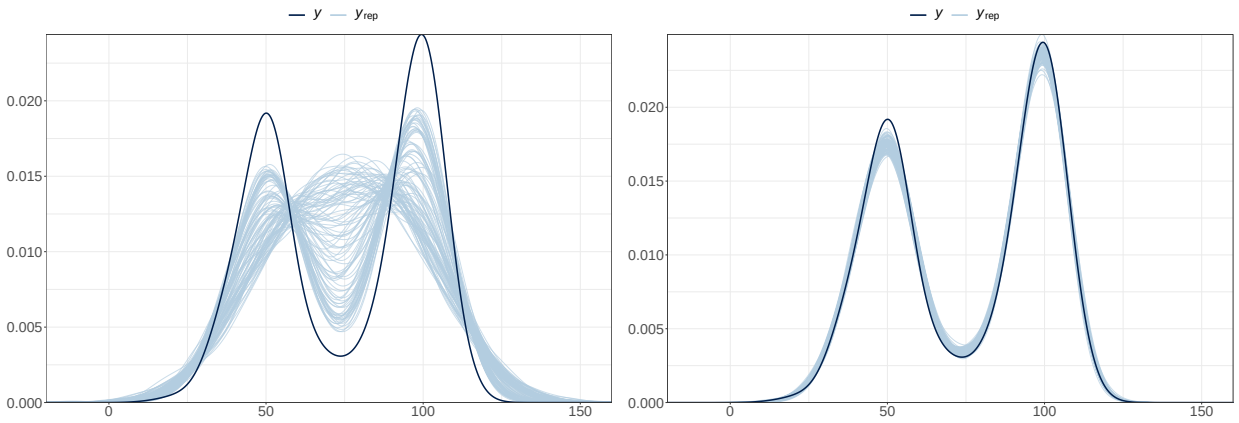
performances. To assess the performance of a fitted model and diagnose potential model misspecifications, it is crucial to include posterior checking in the Bayesian modelling workflow. We use MCMC and PP diagnostic tools as part of our posterior checking.

We start with the PP checking to visualise the performance of the four models described in Table 1. Figure 4 displays the results of the PP checking. It appears that, if we do not take into account the spatial correlation of parameters $\boldsymbol{u}$, we are incapable of simulating the data that adequately capture the distribution of the observations (see plots of models 1 and 3 in Figure 4). On the contrary, because models 2 and 4 incorporate spatial correlation, simulations from these models closely mimic the distribution of observed yield from the Las Rosas experiment.

Figure 5 illustrates the observed skewness of the posterior predictive distribution for the four models. While models 2 and 4 capture the skewness of the observed corn yield, the plots of models 1 and 3 indicate that these models may be misspecified. See Gabry et al. (2019, p. 397) for more details on the use of such skewness plots for model selection in a Bayesian workflow.

LOO CV predictive cumulative density plots can also be used to assess the performance of fitted models. A model is well calibrated for continuous responses when the corresponding plot shows asymptotically uniform behaviour (Gabry et al. 2019; Gelman et al. 2013). Figure 6 compares the density of the computed *leave one out probability integral transformation* (LOO PIT) (the thick dark curve) with the 100 simulated data sets from a standard uniform distribution (the thin light curves). It is evident from Figure 6 that Model 1 and 3 are miscalibrated. Although the Model 2 fit seems good, the frown shape of the curve indicates inferior calibration than Model 4. This implies that Model 2 is either misspecified or too flexible. A flexible model often has the capability of predicting successfully out-of-sample data. However, amongst the four fitted models, Model 4 demonstrates the best fit for the Las Rosas data set.

Pareto $\hat{k}$ diagnostic value is also important, as shown in Table 3. Model 1 has too many large $\hat{k}$ values, which indicates that the model is either misspecified or too flexible. A similar interpretation can be made for Model 3 where a few "bad" values can be observed. These results should not be interpreted solely based on the computed $\hat{k}$ values in Table 3, but we need to take into account the values of LOO PIT and the effective number of parameters $\mathrm{p}_{\mathrm{loo}}$ in Table 4. The $\mathrm{p}_{\mathrm{loo}}$ is calculated by subtracting the $\mathrm{elpd}_{\mathrm{loo}}$ from the full log posterior predictive density. Figure 4 shows that Model 1 and 3 are misspecified. The LOO PIT plots (Figure 6) also confirm that these two models are



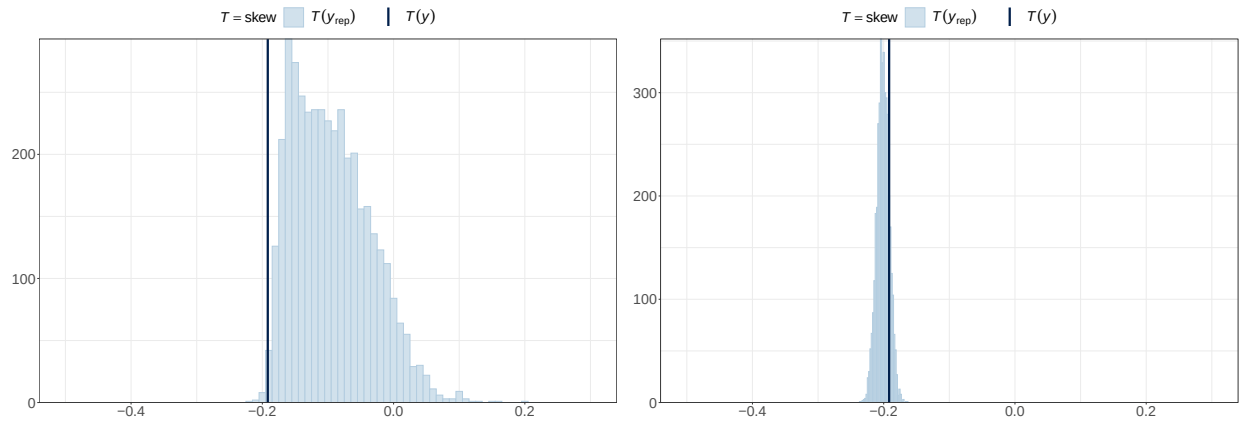
(a) PP check of Model 1

(b) PP check of Model 2

(c) PP check of Model 3

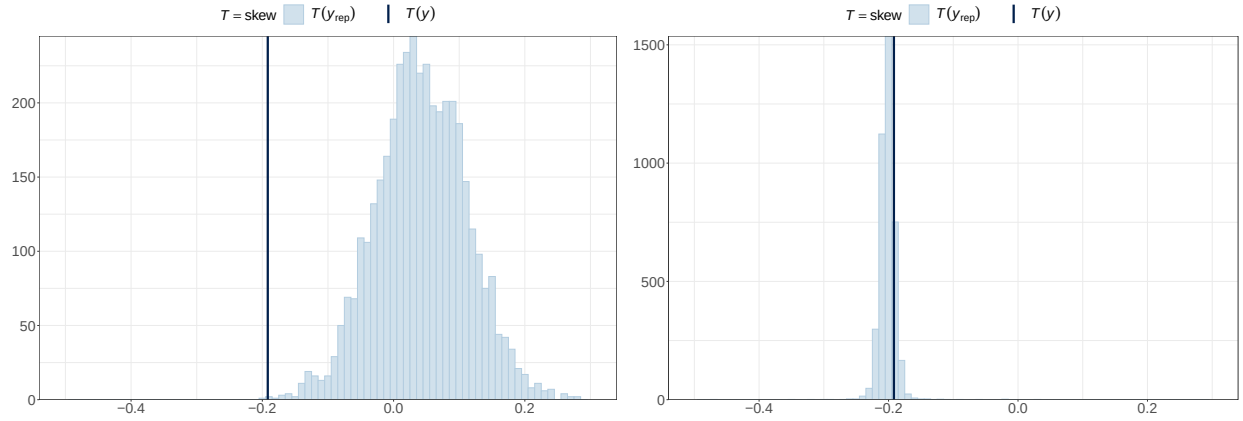
(d) PP check of Model 4

Figure 4: Posterior predictive checking for simple linear and the proposed spatial models with 100 simulations (blue lines) comparing to the observed data (black line).



(a) Skew check of Model 1

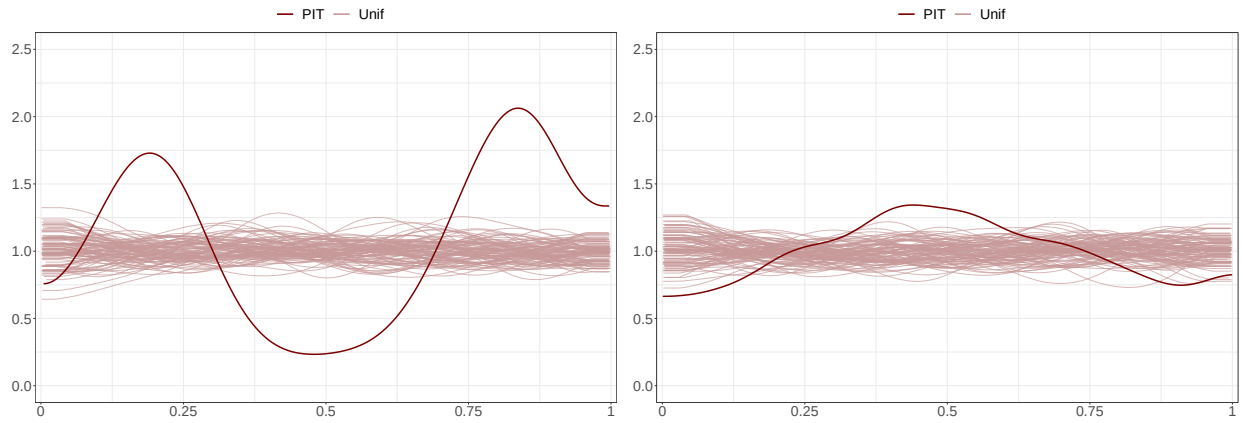
(b) Skew check of Model 2



(c) Skew check of Model 3

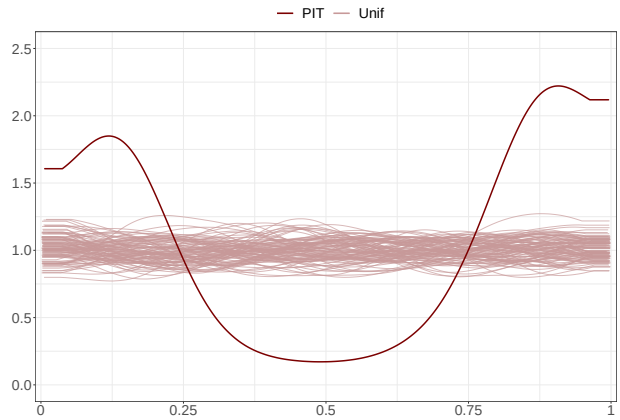
(d) Skew check of Model 4

Figure 5: Histograms of skewness for 4000 draws (blue) from the posterior predictive distribution comparing to the observed data (black).

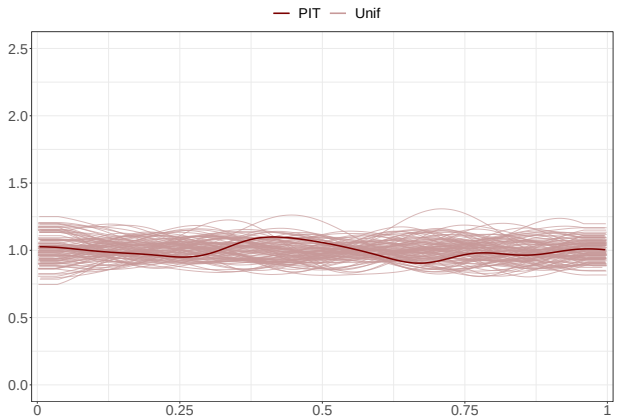


(a) LOO PIT diagnosis of Model 1

(b) LOO PIT diagnosis of Model 2



(c) LOO PIT diagnosis of Model 3



(d) LOO PIT diagnosis of Model 4

Figure 6: LOO PIT plots of the four models. The thick dark line is the density of the LOO PIT for each candidate model, and the thin lines are simulated data from a standard uniform distribution.

misspecified, even though there are no "very bad" $\hat{k}$ values. In the case where high Pareto $\hat{k}$ values are observed but the model fit is good, one can conclude that the model is both misspecified and flexible. In this scenario, a K-fold CV is recommended instead of LOO CV for some $K \geq 5$.

For Model 2, there are six "bad" and "very bad" $\hat{k}$ values, which might be due to highly influential points or outliers. These large $\hat{k}$ values also indicate the potential misspecification of the Gaussian likelihood. Therefore, instead of using Gaussian distribution, Model 4 uses the Student-t distribution. The selection of the Student-t distribution resulted in improvement in all $\hat{k}$ values, as these are estimated to be less than the threshold value of 0.70. Then the elpd_{loo} and Bayesian R^2 are valid.

	Model 1			Model 2			Model 3			Model 4		
	Count	Per	M.Eff	Count	Per	M.Eff	Count	Per	M.Eff	Count	Per	M.Eff
(-Inf, 0.5] (good)	28	1.7%	457	1585	94.7%	432	1474	88.1%	494	1672	99.9%	868
(0.5, 0.7] (ok)	372	22.2%	112	83	5.0%	103	176	10.5%	254	2	0.1%	1733
(0.7, 1] (bad)	1138	68.0%	18	4	0.2%	70	24	1.4%	170	0	0.0%	—
(1, Inf) (very bad)	136	8.1%	8	2	0.1%	4	0	0%	—	0	0%	—

Table 3: Pareto $\hat{k}$ diagnostic values including count, percentage (Per) and minimal effective sample sizes (M.Eff) for all models.

5.4 Model evaluation

We use elpd_{loo}, p_{loo}, LOO information criterion (looic), which is $-2\times\text{elpd}_{\text{loo}}$ in deviance scale, and Bayesian R^2 to evaluate and compare the performance of different models. In Bayesian analysis, even if there are no high Pareto $\hat{k}$ values, R^2 is not indicative if p_{loo} is relatively high compared to the total number of parameters or the number of observations. High p_{loo} and looic values imply weak predictive capability and potential model misspecification.

The results for each of our four fitted models are presented in Table 4. The mean and standard deviation of the posterior distribution, along with the 95% credibility interval (CI) are reported. The lower and upper limits of the CI are given by the 2.5% and 97.5% quantiles of the posterior samples, respectively.

	Model 1		Model 2		Model 3		Model 4	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
elpd_{loo}	-7236.2	13.4	-4945.2	134.8	-7848.4	17.1	-4734.3	38.3
p_{loo}	1487.1	11.7	341.8	41.3	241.2	6.8	516.1	10.5
looic	14472.5	26.7	9890.4	269.6	15696.8	34.3	9468.7	76.7
	Median	CI	Median	CI	Median	CI	Median	CI
Bayesian R^2	0.842	0.563∼0.965	0.974	0.972∼0.977	0.190	0.135∼0.251	0.989	0.987∼0.991

Table 4: LOO CV estimates with standard errors, medians of Bayesian R^2, and 95% credibility intervals.

The R^2 is valid only when the model is not misspecified. Table 4 shows that Model 1 is a better fit than Model 3 in terms of the R^2 value. But these two models are misspecified, as evidenced by their high Pareto $\hat{k}$ values and large p_{loo} values. Therefore, we shall only focus on Model 2 and 4. Model 4 with Student-t distribution is better than Model 2 with Gaussian distribution in terms of smaller looic and higher R^2 value. The bad Pareto $\hat{k}$ values in Model 2 are eliminated by fitting Model 4. Therefore, we use Model 4 to fit the Las Rosas data, and the results are presented in the section below.

6 Results

In the previous section, through model selection and evaluation process, we concluded that Model 4 is the best fit for our example data set. It shows the capability of spatially correlated random parameters in capturing the spatial variation. Using the posterior distribution of the model parameters, we are able to produce the spatially-varying maps of the regression coefficients and subsequently, obtain a smooth map of optimal treatment levels across the whole field. We have also produced the estimated yield map for spatially-varying optimal nitrogen rates.

6.1 Model assessment

Table 5 presents the summary statistics of the posterior distribution of all parameters from Model 4. It should be noted that the means and the medians for all parameters are very close or identical which indicates robust results. Another feature is that the magnitude of the values of $\hat{b}_2$ and $\hat{\sigma}_2$ are very small. It indicates a week influence of the quadratic term of the regression. The pattern of coefficients magnitude is well illustrated in Figure 7.

Parameter	Mean	SD	Credibility interval		
			2.5%	Median	97.5%
$\hat{b}_0$	78.7361	3.0680	72.8282	78.6723	84.8108
$\hat{b}_1$	0.0126	0.0091	-0.0049	0.0127	0.0303
$\hat{b}_2(\times 10^4)$	1.9850	1.0945	-1.3057	1.9776	4.1425
$\hat{\sigma}_0$	9.1322	0.3902	8.4027	9.1271	9.9447
$\hat{\sigma}_1$	0.0173	0.0071	0.0034	0.0174	0.0314
$\hat{\sigma}_2(\times 10^4)$	1.7157	0.7151	0.3935	1.6742	3.2388
$\hat{\sigma}_e$	2.6399	0.1244	2.3953	2.6398	2.8905
$\hat{\rho}_{12}$	-0.6493	0.2467	-0.9623	-0.7005	-0.0115
$\hat{\rho}_{13}$	0.5367	0.2481	0.0193	0.5514	0.9480
$\hat{\rho}_{23}$	-0.4282	0.3754	-0.9361	-0.5033	0.4732
$\hat{\rho}_c$	0.9076	0.0115	0.8835	0.9080	0.9287
$\hat{\rho}_r$	0.9274	0.0074	0.9120	0.9275	0.9410
$\hat{\nu}$	4.1321	0.5503	3.2098	4.0861	5.3573

Table 5: Summary statistics of the posterior samples from Model 4. Mean, standard deviation (SD), 95% credibility interval (showing 2.5% and 97.5% sample quantiles) and median of posterior samples are reported.

Figure 7 displays the maps of the spatially-varying regression coefficients, estimated using Model 4. The top, middle, and bottom panels of Figure 7 show the intercept $\hat{\boldsymbol{\beta}}_0 = \hat{\boldsymbol{b}}_0 + \tilde{\boldsymbol{u}}_0$, the linear term $\hat{\boldsymbol{\beta}}_1 = \hat{\boldsymbol{b}}_1 + \tilde{\boldsymbol{u}}_1$ and the quadratic term $\hat{\boldsymbol{\beta}}_2 = \hat{\boldsymbol{b}}_2 + \tilde{\boldsymbol{u}}_2$, respectively. The plots cover the whole trial area, as presented in Figure 1a and 1c. The contour maps are aligned with the topology of the area. It can be observed that the Hilltop area and small part of the neighbouring areas on the left and right (see Figure 1c) are exhibiting different pattern in comparison to the other three topological regions, for all of the $\hat{\boldsymbol{\beta}}$ coefficients. The linear component coefficient for the Hilltop area is the highest, in the range of 0.02 – 0.08, while for the other three areas is around -0.01. The quadratic component coefficient for the Hilltop area is negative, which indicates that an optimal treatment in the area is available. However, in other areas, the coefficients are positive and a linear pattern is sufficient in model fitting.

The result is consistent with the discovery by Rakshit et al. (2020) that the quadratic pattern is strong in Hilltop region but weak in other regions. Even though a quadratic pattern is identified in the East slope and Low East, the adjusted-p values indicate non-significant for these areas.

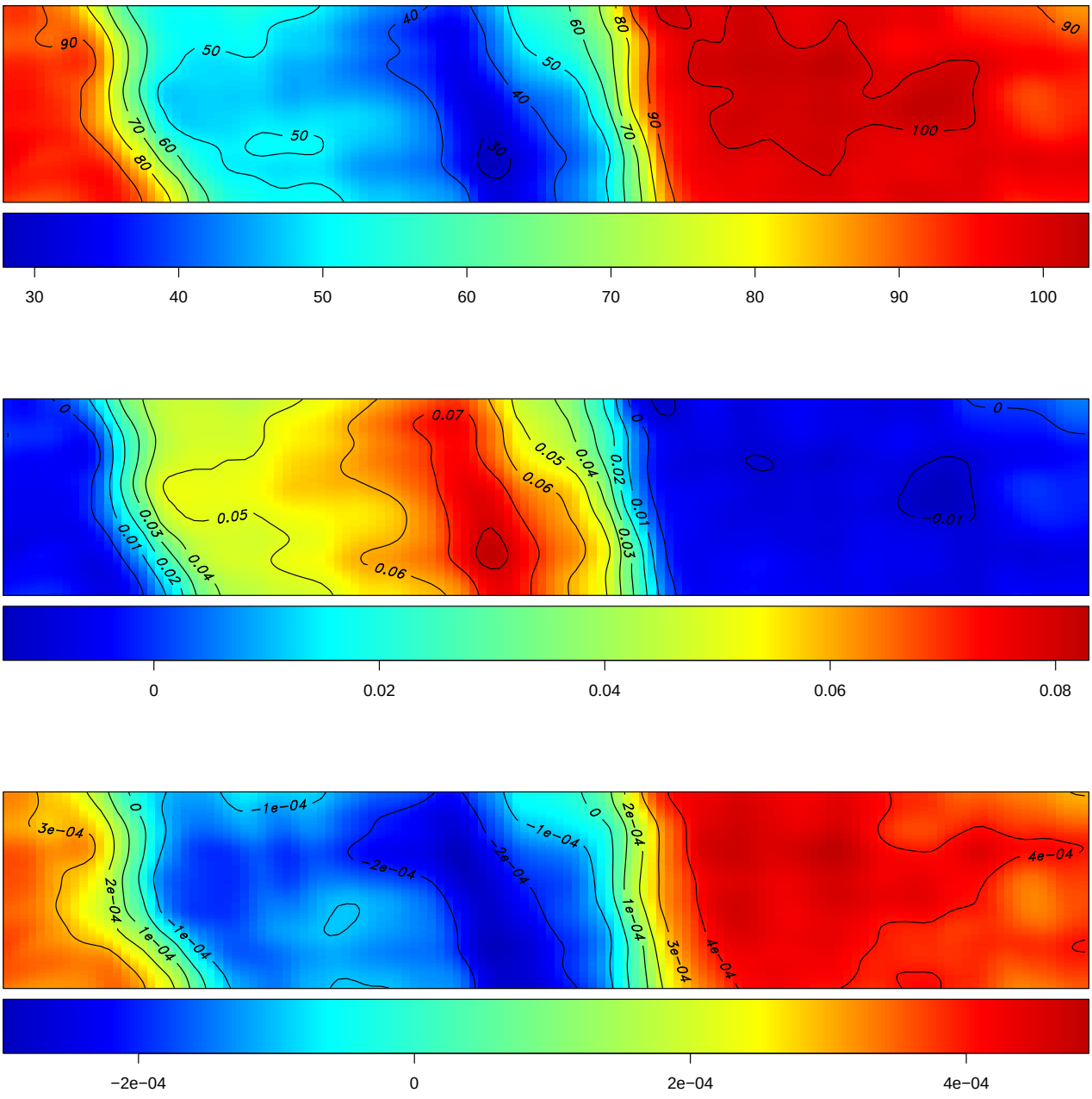


Figure 7: Contour plots of spatial-varying coefficients $\hat{\boldsymbol{\beta}}_0$ (top), $\hat{\boldsymbol{\beta}}_1$ (middle) and $\hat{\boldsymbol{\beta}}_2$ (bottom) for Las Rosas data. Negative $\hat{\boldsymbol{\beta}}_2$ is available in the Hilltop region, where optimal treatments exist. For other regions, linear response is sufficient.

6.2 Yield prediction

Because we have fitted a quadratic response of yield to nitrogen rates, we can compute the optimal nitrogen rate $\tilde{N}_i$ for the ith grid point using $\tilde{N}_i = -\hat{\beta}_1/(2\hat{\beta}_2)$, $i = 1, \ldots, n$, given $\hat{\beta}_2 < 0$. However, if the optimum rate exceeds the maximum rate $N_{\max} = 124.6$ kg/ha used in the trial, the maximum rate has been chosen as the optimal rate. Therefore, we can compute the adjusted optimal rate $\hat{N}_i = \min\{\tilde{N}_i, N_{\max}\}$ for $i = 1, \ldots, n$. Figure 8 depicts the map of the adjusted optimal treatment and estimated yield corresponding to the spatially-varying adjusted optimal treatment rates across the field.

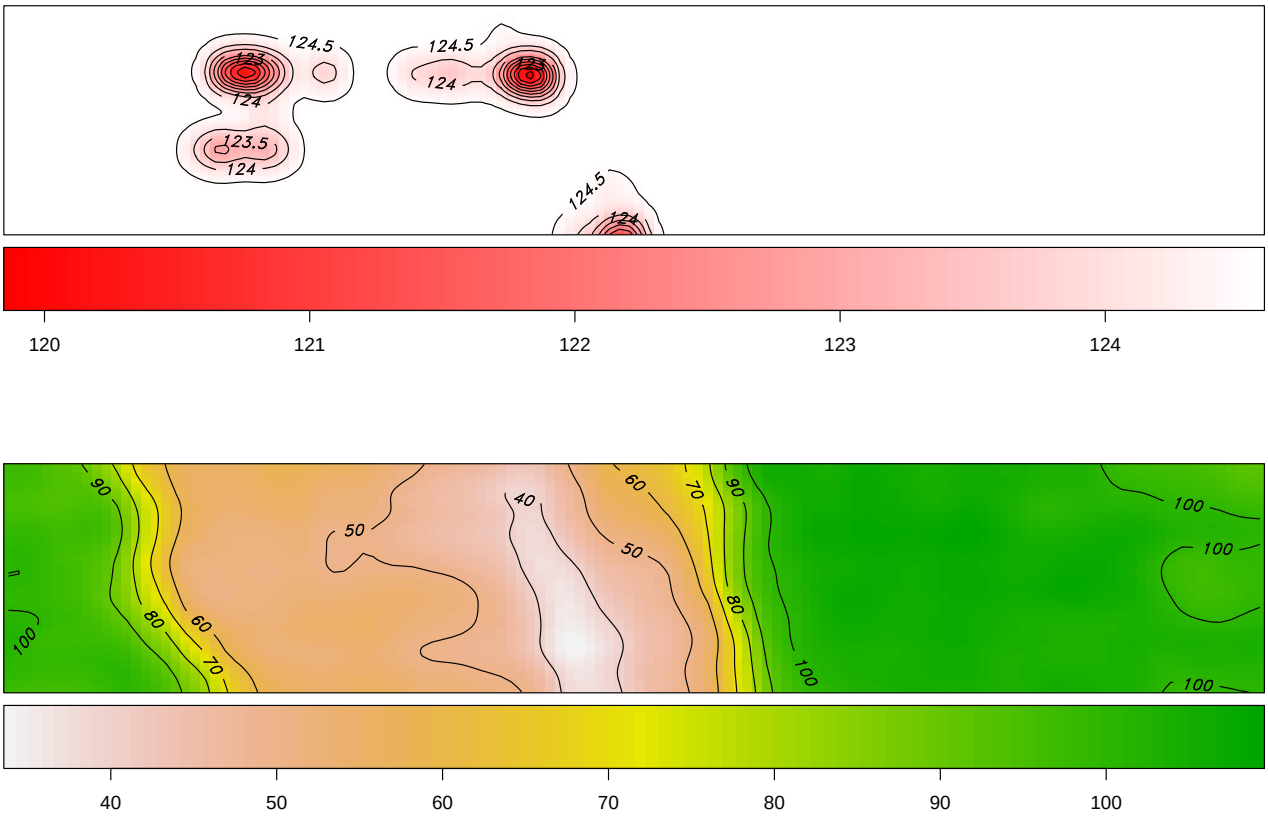


Figure 8: *Top*: adjusted optimal nitrogen rates ($\hat{N}$); *Bottom*: estimated yield corresponding to the adjusted optimal rates.

Figure 9 shows the difference between the predicted yield for the adjusted optimal treatments and the observed yield. As expected, the difference is positive, indicating a higher yield prediction under the optimal nitrogen treatment.

6.3 Comparing to the GWR approach

Rakshit et al. (2020) suggested a GWR based analysis for the same problem considered in this paper, and estimated the spatially-varying coefficients by maximising the local loglikelihoods. GWR is also used to estimate the optimal nitrogen rate for each grid and to predict the yield for the Las Rosas data set.

In GWR, the results crucially depends on the bandwidth of the selected kernel function. Although an appropriate bandwidth can be selected using spatial cross validation, it is computationally challenging for large data sets. To estimate the regression parameters for a query location, the neighbouring observations are given more weight than the distant ones in GWR. On the contrary, the proposed Bayesian approach uses all data in one go to produce estimates

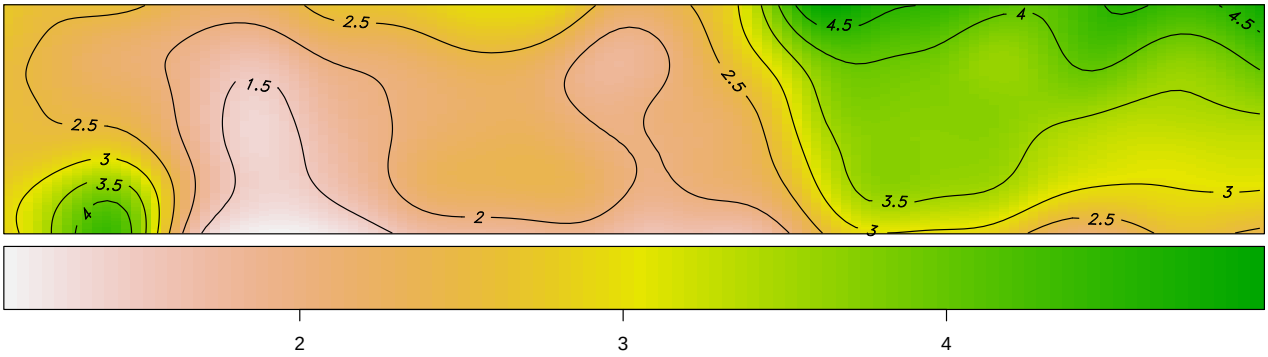


Figure 9: Yield difference computed between predicted yield with optimal nitrogen and observed yield.

for all grid point, based on a spatial variance matrix defined for the entire field. The Bayesian inference is affected by the choice of priors and the likelihood. However, the influence of the prior reduces if the amount of data increases. The Bayesian approach in general is more flexible than GWR, as it can be easily extended and applied broadly to other applications.

A comparison of these two approaches is summarised in Table 6.

	GWR	Bayesian
Inference	with neighbouring data	with all data
Initialisation	bandwidth selection	prior specification
Objective	local log-likelihood	global log-likelihood
Evaluation	t scores and p-values	credible intervals PP check and LOO PIT Pareto k diagnosis Bayesian R^2

Table 6: Comparison of GWR and Bayesian approach.

7 Discussion

In this paper, we developed a Bayesian hierarchical model for estimating spatially-varying treatment effects and mapping locally-varying optimum treatments for large strip experiments. We maximised yield in order to determine the spatially-varying optimum nitrogen rates for the Las Rosas data example. However, another choice, particularly desirable from a farmer's perspective, could be to maximise profit in order to determine the spatial map of optimum nitrogen rates. Such an analysis would require authentic data on economic variables, including treatment cost and revenue from yield, and our proposed Bayesian framework could be easily adapted to incorporate these information into the analysis of on-farm data sets. It is crucial in Bayesian inference to be able to sample from posterior distributions. In order to analyse the Las Rosas data set, we have used the NUTS sampler to sample from highly correlated high-dimensional posterior distributions. NUTS exhibits excellent sampling qualities in terms of generating large effective sample sizes, producing low autocorrelation, and obtaining low skewness of marginal posterior distributions (Nishio and Arakawa 2019). Moreover, NUTS does not require conjugate priors, exhibits faster convergence for multi-parameters and has considerable flexibility for fitting user-specified models by researchers using the `R`-package `rstan`. However,

if the data set is large, computing the inverse of the covariance matrix, which is three times the size of the data, is extremely time consuming by conventional algorithm. Therefore, we implement a faster algorithm for calculating the autocorrelation matrix and develop a faster algorithm for computing the Kronecker product of three matrices. The details of these algorithms are presented in the Appendix.

Other covariance structures, including the Matérn class of covariance functions (Cressie and Huang 1999), also can be used for capturing spatial variation in OFE (Selle et al. 2019). The Matérn covariance structure can be incorporated in our Bayesian modelling framework. However, the main drawback of implementing the Matérn covariance is that it takes a large amount of time to calculate the inverse of the covariance matrix when the data size is large. For implementing the Matérn covariance function, we have to either wait for a long time to obtain converged MCMC chains or reduce the effective sample size and terminate the sampling process earlier, which increases the risk of obtaining non-converged chains and leaving parts of the posterior space unexplored. In practice, however, the difference between the results due to AR1 $\times$ AR1 and due to Matérn covariance is not significant, as shown in (Selle et al. 2019). For most gridded OFE data sets, the AR1 $\times$ AR1 covariance structure is a reasonable choice in terms of both efficiency and accuracy.

The model checking and diagnostic process for post-sampling were presented as well. In order to check the appropriateness of spatially-correlated regression parameters, we considered models without any spatial correlation as benchmark models (see Table 1). Using posterior model checking in Section 5.3, we showed that the models with spatially-correlated parameters performed much better than the models without spatially-correlated parameters for the Las Rosas data. Without any prior knowledge of the data, one may wish to first investigate the spatial variability by comparing a model with local effects with a model with only fixed regression coefficients. Conventionally, one may only check the divergence of MCMC chains and insufficiently diagnose the model and its assumption. 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. The Gaussian assumption of the model for the Las Rosas data is misspecified even though the Bayesian R^2 value is relatively high. Therefore, other than checking the behaviour of MCMC chains, candidate models should be diagnosed with advance diagnostic tools, such as PP check, LOO CV, Pareto k, etc, in the first place. With the help of these diagnostic tools, we discover that Student-t distribution provides a more robust inference.

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 7.

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

Finally, in Section 6.3, we explained the difference between our proposed Bayesian approach and the GWR method. However, the results from the Bayesian approach are very similar to those from GWR, reported in Rakshit et al. (2020). Another potential method of analysis is based on the residual maximum likelihood (REML). The estimation of regression coefficients under the REML framework would require the development of a computing algorithm that would take into account the spatial correlation of the random effects while computing the best linear unbiased predictors of the treatment effects.

8 Conclusion

The novelty of our work can be summarised as follows:

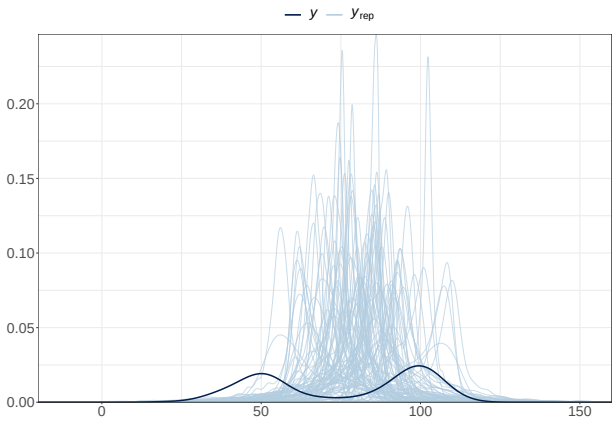
- A Bayesian hierarchical model is adopted to analyse large on-farm strip trials.
- Spatial variation is accounted for by incorporating spatially correlated random terms in the model.
- The posterior samples of all parameters were obtained by utilising faster Kronecker product computing algorithms in `rstan`.
- Advanced diagnostic tools were used to guard against the crucial problem of model misspecification.
- The real-life OFE data set from Las Rosas, Argentina, was analysed to obtain the spatially-varying optimum nitrogen rates for maximising corn yield across the entire field.

Authors' contribution

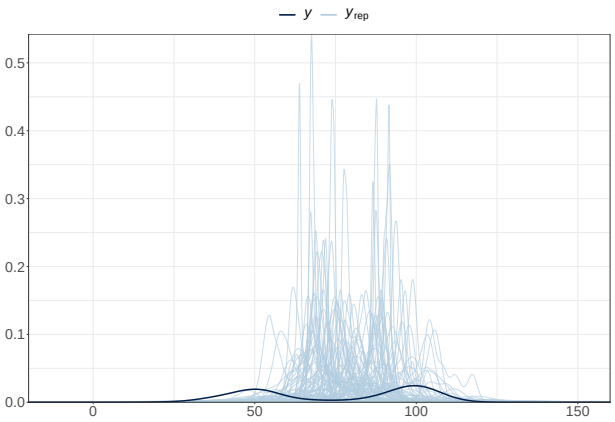
Zhanglong Cao: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing, Visualization; Katia Stefanova: Writing - Original Draft, Writing - Review & Editing; Mark Gibberd: Writing - Review & Editing, Project administration; Suman Rakshit: Conceptualization, Methodology, Writing - Review & Editing, Supervision.

Acknowledgement

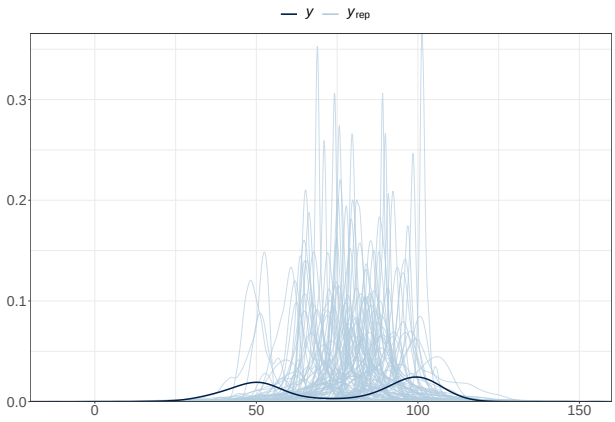
The authors gratefully acknowledge the support from the Grains Research and Development Corporation of Australia (GRDC).



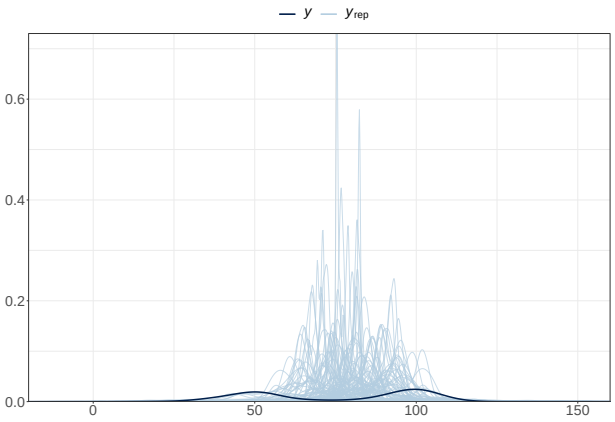
(a) Model 1: Gaussian distribution without spatial correlation.



(b) Model 2: Gaussian distribution with spatial correlation.



(c) Model 3: Student distribution without spatial correlation.



(d) Model 4: Student distribution with spatial correlation.

Figure 10: Weakly informative priors checking for four models.

Appendix

A Prior predictive checking

B Faster Cholesky factor for AR1(ρ)

The AR1(ρ) correlation matrix with correlation coefficient ρ is defined as $\rho_{ij} = \rho^{|i-j|}$. A simple form of Cholesky factor for the AR1(ρ) structure, given by Madar (2015), was used

$$l_{ij} = \begin{cases} \rho^{j-1} & j \geq i = 1 \\ \rho^{j-i}\sqrt{1-\rho^2} & j \geq i \geq 2 \end{cases}, \tag{30}$$

which significantly improved the computational efficiency in `rstan`.

C Fast Kronecker product

Let $A = [a_1, a_2, \ldots, a_n] \in \mathbb{R}^{m \times n}$, where $a_j \in \mathbb{R}^m, j = 1, 2, \ldots, n$. Then the vector vec($A$) is defined as

$$\text{vec}(A) = [a_1, a_2, \ldots, a_n]^\top \in \mathbb{R}^{mn}, \tag{31}$$

which vec-permutes the given matrix. With the vector-valued operator, we have the "Vec Trick" theorem:

Lemma 1. *(Roth's Column Lemma: "Vec Trick" (Roth 1934; Airola and Pahikkala 2018)): Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, and $C \in \mathbb{R}^{p \times q}$ be matrices. Then*

$$vec(ABC) = (C^\top \otimes A)vec(B). \tag{32}$$

The above property and theorem are implemented in `rstan` and considerably saved computation time. For other properties of the Kronecker product see Zhang and Ding (2013).

References

Airola, A. and T. Pahikkala (Aug. 2018). "Fast Kronecker Product Kernel Methods via Generalized Vec Trick". In: *IEEE Transactions on Neural Networks and Learning Systems* 29.8, pp. 3374–3387. DOI: 10.1109/TNNLS.2017.2727545.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2004). *Hierarchical Modeling and Analysis for Spatial Data*. en. Monographs on Statistics and Applied Probability 101. Boca Raton, Fla: Chapman & Hall/CRC.

Besag, J. and D. Higdon (Nov. 1999). "Bayesian Analysis of Agricultural Field Experiments". en. In: *J Royal Statistical Soc B* 61.4, pp. 691–746. DOI: 10.1111/1467-9868.00201.

Brooks, S., A. Gelman, G. Jones, and X.-L. Meng (2011). *Handbook of Markov Chain Monte Carlo*. CRC press.

Brunsdon, C., A. S. Fotheringham, and M. Charlton (Aug. 1999). "Some Notes on Parametric Significance Tests for Geographically Weighted Regression". en. In: *Journal of Regional Science* 39.3, pp. 497–524. DOI: 10.1111/0022-4146.00146.

Bürkner, P. C. (Aug. 2017). "Brms: An R Package for Bayesian Multilevel Models Using Stan". In: *Journal of Statistical Software* 80.1, pp. 1–28. DOI: 10.18637/jss.v080.i01.

Bürkner, P.-C., J. Gabry, and A. Vehtari (June 2021). "Efficient Leave-One-out Cross-Validation for Bayesian Non-Factorized Normal and Student-t Models". In: *Comput Stat* 36.2, pp. 1243–1261. DOI: 10.1007/s00180-020-01045-4.

Butler, D., B. Cullis, A. Gilmour, B. Gogel, and R Thompson (2009). "ASReml-R reference manual version 4". In: *The State of Queensland, Department of Primary Industries and Fisheries: Brisbane, Qld.*

Che, X. and S. Xu (Sept. 2010). "Bayesian Data Analysis for Agricultural Experiments". In: *Can. J. Plant Sci.* 90.5, pp. 575–603. DOI: 10.4141/cjps10004.

Congdon, P. D. (Sept. 2019). *Bayesian Hierarchical Models: With Applications Using R, Second Edition.* English. Second. CRC Press.

Cook, S. E. and R. G. V. Bramley (1998). "Precision Agriculture — Opportunities, Benefits and Pitfalls of Site-Specific Crop Management in Australia". en. In: *Aust. J. Exp. Agric.* 38.7, pp. 753–763. DOI: 10.1071/ea97156.

Cook, S., M. Lacoste, F. Evans, M. Ridout, M. Gibberd, and T. Oberthür (June 2018). "An on-farm experimental philosophy for farmer-centric digital innovation". In: *14th International Conference on Precision Agriculture.* Montreal, Quebec, Canada.

Cressie, N. and H.-C. Huang (Dec. 1999). "Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions". In: *Journal of the American Statistical Association* 94.448, pp. 1330–1339. DOI: 10.1080/01621459.1999.10473885.

Cressie, N. A. C. (1993). "Statistics for Spatial Data". en. In: *Statistics for Spatial Data.* John Wiley & Sons, Ltd. Chap. 1, pp. 1–26. DOI: 10.1002/9781119115151.ch1.

Dipak Dey and C.R. Rao, eds. (Nov. 2005). *Bayesian Thinking, Modeling and Computation.* en. 1st. Vol. 25. Elsevier.

Donald, M., C. L. Alston, R. R. Young, and K. L. Mengersen (Dec. 2011). "A Bayesian Analysis of an Agricultural Field Trial with Three Spatial Dimensions". en. In: *Computational Statistics & Data Analysis* 55.12, pp. 3320–3332. DOI: 10.1016/j.csda.2011.06.022.

Duane, S., A. D. Kennedy, B. J. Pendleton, and D. Roweth (1987). "Hybrid Monte Carlo". In: *Physics letters B* 195.2, pp. 216–222. DOI: 10.1016/0370-2693(87)91197-x.

Edmondson, R. N. (Feb. 2014). "Agridat". en. In: *The Journal of Agricultural Science* 152.1, pp. 2–2. DOI: 10.1017/S0021859613000920.

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.

Fong, Y., H. Rue, and J. Wakefield (2010). "Bayesian inference for generalized linear mixed models". In: *Biostatistics* 11.3, pp. 397–412.

Fotheringham, A. S. (2009). ""The Problem of Spatial Autocorrelation" and Local Spatial Statistics". en. In: *Geographical Analysis* 41.4, pp. 398–403. DOI: 10.1111/j.1538-4632.2009.00767.x.

Gabry, J., D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman (2019). "Visualization in Bayesian Workflow". en. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182.2, pp. 389–402. DOI: 10.1111/rssa.12378.

Gelman, A. (2003). "A Bayesian Formulation of Exploratory Data Analysis and Goodness-of-Fit Testing". en. In: *International Statistical Review* 71.2, pp. 369–382. DOI: 10.1111/j.1751-5823.2003.tb00203.x.

— (Dec. 2004). "Exploratory Data Analysis for Complex Models". en. In: *Journal of Computational and Graphical Statistics* 13.4, pp. 755–779. DOI: 10.1198/106186004x11435.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (Nov. 2013). *Bayesian Data Analysis*. English. Third Edition. Chapman & Hall CRC Texts in Statistical Science. CRC Press.

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.

Gelman, A., D. Simpson, and M. Betancourt (Oct. 2017). "The Prior Can Often Only Be Understood in the Context of the Likelihood". en. In: *Entropy* 19.10, p. 555. DOI: 10.3390/e19100555.

Gelman, A. et al. (2006). "Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper)". In: *Bayesian analysis* 1.3, pp. 515–534. DOI: 10.1214/06-BA117A.

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.

Harris, P. (2019). "A Simulation Study on Specifying a Regression Model for Spatial Data: Choosing between Autocorrelation and Heterogeneity Effects". en. In: *Geographical Analysis* 51.2, pp. 151–181. DOI: 10.1111/gean.12163.

Hastie, T. and C. Loader (1993). "Local Regression: Automatic Kernel Carpentry". In: *Statistical Science* 8.2, pp. 120–129.

Hinkelmann, K. (2012). *Design and Analysis of Experiments*. Vol. 3. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc. DOI: 10.1002/9781118147634.

Hoffman, M. D. and A. Gelman (2014). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." In: *J. Mach. Learn. Res.* 15.1, pp. 1593–1623.

Jiang, P., Z. He, N. R. Kitchen, and K. A. Sudduth (Apr. 2009). "Bayesian Analysis of Within-Field Variability of Corn Yield Using a Spatial Hierarchical Model". en. In: *Precision Agric* 10.2, pp. 111–127. DOI: 10.1007/s11119-008-9070-4.

Juárez, M. A. and M. F. J. Steel (Jan. 2010). "Model-Based Clustering of Non-Gaussian Panel Data Based on Skew-t Distributions". en. In: *Journal of Business & Economic Statistics* 28.1, pp. 52–66. DOI: 10.1198/jbes.2009.07145.

Kass, R. E. and R. Natarajan (2006). "A Default Conjugate Prior for Variance Components in Generalized Linear Mixed Models (Comment on Article by Browne and Draper)". In: *Bayesian Analysis* 1.3, pp. 535–542. DOI: 10.1214/06-ba117b.

Lark, R. and H. Wheeler (2003). "A Method to Investigate Within-Field Variation of the Response of Combinable Crops to an Input". English. In: *Agronomy Journal* 95, pp. 1093–1104. DOI: 10.2134/AGRONJ2003.1093.

Lawes, R. A. and R. G. V. Bramley (2012). "A Simple Method for the Analysis of On-Farm Strip Trials". English. In: *Agronomy Journal* 104.2, pp. 371–377. DOI: 10.2134/agronj2011.0155.

Lewandowski, D., D. Kurowicka, and H. Joe (2009). "Generating random correlation matrices based on vines and extended onion method". In: *Journal of Multivariate Analysis* 100.9, pp. 1989–2001. DOI: https://doi.org/10.1016/j.jmva.2009.04.008.

Madar, V. (2015). "Direct Formulation to Cholesky Decomposition of a General Nonsingular Correlation Matrix". In: *Statistics and Probability Letters* 103.1, pp. 142–147. DOI: 10.1016/j.spl.2015.03.014.

Marchant, B., S. Rudolph, S. Roques, D. Kindred, V. Gillingham, S. Welham, C. Coleman, and R. Sylvester-Bradley (Jan. 2019). "Establishing the Precision and Robustness of Farmers' Crop Experiments". en. In: *Field Crops Research* 230, pp. 31–45. DOI: 10.1016/j.fcr.2018.10.006.

McElreath, R. (Dec. 2015). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. English. First. Vol. 122. Chapman and Hall/CRC Texts in Statistical Science Ser. CRC Press LLC.

Monnahan, C. C., J. T. Thorson, and T. A. Branch (2017). "Faster Estimation of Bayesian Models in Ecology Using Hamiltonian Monte Carlo". en. In: *Methods in Ecology and Evolution* 8.3, pp. 339–348. DOI: 10.1111/2041-210X.12681.

Montesinos-López, O. A., A. Montesinos-López, M. V. Hernández, I. Ortiz-Monasterio, P. Pérez-Rodríguez, J. Burgueño, and J. Crossa (Nov. 2018). "Multivariate Bayesian Analysis of On-Farm Trials with Multiple-Trait and Multiple-Environment Data". en. In: *Agron.j.* 111.6, pp. 2658–2669. DOI: 10.2134/agronj2018.06.0362.

Nishio, M. and A. Arakawa (Dec. 2019). "Performance of Hamiltonian Monte Carlo and No-U-Turn Sampler for Estimating Genetic Parameters and Breeding Values". en. In: *Genet Sel Evol* 51.1, p. 73. DOI: 10.1186/s12711-019-0515-1.

Onofri, A., H.-P. Piepho, and M. Kozak (2019). "Analysing Censored Data in Agricultural Research: A Review with Examples and Software Tips". en. In: *Annals of Applied Biology* 174.1, pp. 3–13. DOI: 10.1111/aab.12477.

Páez, A., T. Uchida, and K. Miyamoto (Apr. 2002). "A General Framework for Estimation and Inference of Geographically Weighted Regression Models: 1. Location-Specific Kernel Bandwidths and a Test for Locational Heterogeneity". In: *Environ Plan A* 34.4, pp. 733–754. DOI: 10.1068/a34110.

Piepho, H. P. (2019). "A Coefficient of Determination (R2) for Generalized Linear Mixed Models". In: *Biometrical J.* 61.4, pp. 860–872. DOI: 10.1002/bimj.201800270.

Piepho, H.-P., A Büchse, and K. Emrich (2003). "A hitchhiker's guide to mixed models for randomized experiments". In: *Journal of Agronomy and Crop Science* 189.5, pp. 310–322.

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.

Pringle, M. J., T. F. A. Bishop, R. M. Lark, B. M. Whelan, and A. B. McBratney (2010). "The Analysis of Spatial Experiments". en. In: *Geostatistical Applications for Precision Agriculture*. Ed. by M. Oliver. Dordrecht: Springer Netherlands, pp. 243–267. DOI: 10.1007/978-90-481-9133-8_10.

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.

Roth, W. E. (1934). "On Direct Product Matrices". In: *Bulletin of the American Mathematical Society* 40.6, pp. 461–468. DOI: 10.1090/S0002-9904-1934-05899-3.

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.

Shirley, R., E. Pope, M. Bartlett, S. Oliver, N. Quadrianto, P. Hurley, S. Duivenvoorden, P. Rooney, A. B. Barrett, C. Kent, and J. Bacon (Jan. 2020). "An Empirical, Bayesian Approach to Modelling Crop Yield: Maize in USA". en. In: *Environ. Res. Commun.* 2.2, p. 025002. DOI: 10.1088/2515-7620/ab67f0.

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.

Theobald, C. M., M. Talbot, and F. Nabugoomu (Sept. 2002). "A Bayesian Approach to Regional and Local-Area Prediction from Crop Variety Trials". en. In: *JABES* 7.3, pp. 403–419. DOI: 10.1198/108571102230.

Tsionas, E. G. (2002). "Bayesian Inference in the Noncentral Student-t Model". In: *Journal of Computational and Graphical Statistics* 11.1, pp. 208–221. DOI: 10.1198/106186002317375695.

Vehtari, A., A. Gelman, and J. Gabry (Sept. 2017). "Practical Bayesian Model Evaluation Using Leave-One-out Cross-Validation and WAIC". en. In: *Stat Comput* 27.5, pp. 1413–1432. DOI: 10.1007/s11222-016-9696-4.

Weiss, R. (Dec. 1994). "Pediatric Pain, Predictive Inference, and Sensitivity Analysis". en. In: *Evaluation Review* 18.6, pp. 651–677. DOI: 10.1177/0193841x9401800601.

Yan, W., L. A. Hunt, P. Johnson, G. Stewart, and X. Lu (2002). "On-Farm Strip Trials vs. Replicated Performance Trials for Cultivar Evaluation". In: *Crop Science* 42.2, pp. 385–392. DOI: 10.2135/cropsci2002.0385.

Zhang, H. and F. Ding (2013). "On the Kronecker Products and Their Applications". en. In: *Journal of Applied Mathematics* 2013, pp. 1–8. DOI: 10.1155/2013/296185.

Zhao, Y., J. Staudenmayer, B. A. Coull, and M. P. Wand (2006). "General design Bayesian generalized linear mixed models". In: *Statistical science*, pp. 35–51.

Zimmerman, D. L. and D. A. Harville (1991). "A Random Field Approach to the Analysis of Field-Plot Experiments and Other Spatial Experiments". In: *Biometrics* 47.1, pp. 223–239. DOI: 10.2307/2532508.