

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

3 Zhanglong Cao<sup>1</sup>, Suman Rakshit<sup>1</sup>, and XXX<sup>2</sup>

4 <sup>1</sup>SAGI

5 <sup>2</sup>LATEX University

6 **Abstract**

7 ~~Accommodating spatial variation is common in analysing field trials, and has become a challenge.~~ Accounting for  
8 ~~spatial variability is crucial in analysing agricultural trial data, and is more difficult~~ in large on-farm experiments.  
9 ~~Simple linear mixed models and Gaussian distribution are incapable of dealing with complex data sets and results~~  
10 ~~in misspecification. This paper describes trials than small greenhouse trials. Different approaches were proposed to~~  
11 ~~assess spatial variation in recent years. However, the potential model misspecification is inadequately discussed in~~  
12 ~~literature. In this paper, we use a Bayesian framework to analyse the model with spatially correlated random~~  
13 ~~parameters for large on-farm experiments~~~~strip trials and implement the approach in the R package rstan.~~ With  
14 ~~advanced model post sampling~~ diagnostic tools, we found that the model with Gaussian assumption is misspecified.  
15 ~~Therefore, we use A more robust Student-t distribution with which the model is augmented. The open-source~~  
16 ~~R code that implements our proposed method for analysing on-farm data is provided for further use.~~ augments the  
17 ~~model and avoids misspecification.~~ We also discuss the difference of the Bayesian approach and GWR geographically  
18 ~~weighted regression (GWR)~~, and compare the results from these two approaches. The open-source R code is provided  
19 ~~for further use.~~

20 **Keywords:** Spatial variation, geostatistics, No-U-turn sampler.

21 **1 Introduction**

22 Traditional agricultural experiments are conducted ~~as replicated small plots in replicated small plots~~ at one or  
23 multiple ~~designated~~ locations for a ~~period of time~~~~number of years~~. However, farmers prefer to conduct experiments ~~to~~  
24 ~~test and verify systems on their own farms~~~~even though the data are correlated as well as heterogeneous across space~~  
25 ~~and the inference would be considered invalid under classical statistics [Griffin et al., 2008]~~~~in their farms.~~ Therefore,  
26 ~~the use of~~ on-farm (OFE) trials or on-farm experiments (OFE) has been ~~developed~~ ~~conducted~~ in recent years  
27 ~~[Troyer and Wellin, 2009, Yan et al., 2002]~~ and statistical models for analysing the data that are collected from OFE  
28 are required [Yan et al., 2002].

29 ~~A problem of~~

30 ~~A problem with~~ OFE data is the inherent spatial variation, which ~~occurs when the values of variables sampled at~~  
31 ~~nearby locations are not independent from each other and~~ may seriously bias treatment estimates and inflate standard  
32 errors [Tobler, 1970]. Spatial variation arises from small changes, such as fertility, soil moisture, light ~~et al~~ and so on, or  
33 is aligned with the rows and columns of a field trial. It could also arise from experimental procedures or management  
34 practices that have a recurrent pattern, known as extraneous variation [Gilmour et al., 1997, Hinkelmann, 2012].  
35 ~~Appropriate statistical approaches for analysing OFE data are in high demand.~~

36 ~~A few researches have been conducted on this topic and discussed the accommodation of spatial variation, such~~  
37 ~~as [Montesinos-López et al., 2018, Selle et al., 2019]. These approaches~~ In previous work, Selle et al. [2019] compared  
38 established spatial models that account for spatial variability in the research of plant breeding. The proposed Bayesian  
39 framework and model are good for variety selection, where a variety replicates two or more times and the results are  
40 averaged across the field. However, ~~their approaches are it is~~ incapable of addressing the objective of precision farming:

41 how to maximise the output and/or profit by developing management practices that lead to optimal utilisation of  
42 resources across variable environments [Cook et al., 2013].

43 Rakshit et al. [2020] proposed a geographically weighted regression (GWR) approach which is adapted to obtain  
44 spatially-varying estimates of treatment effects for OFE. They addressed the problem with the proposed model and  
45 found optimal Nitrogen rate for each single plot on the field. A problem with Local *t*-scores and *p*-values are introduced  
46 for inference. A crucial step in their approach is that the choice of bandwidth of kernel functions affects the results  
47 and may bandwidth selection for kernel functions. False selection affects the estimation and cause biases.

48 In this paper, we use a Bayesian approach to fit the geo-referenced model for OFE data and to find the optimal  
49 treatment for each plot. Instead of imposing spatially correlated residuals, we found that the model with model that  
50 explores spatial nonstationarity by incorporating spatially correlated random parameters is good enough for capturing  
51 the spatial variation. Besides, the . The adoption of the Bayesian approaches simplifies the interpretation of the results  
52 and augments the inference [Che and Xu, 2010]. Compared with REML (restricted maximum likelihood ) (REML)  
53 approach, Guan et al. [2017], Omer and Singh [2017] demonstrate the advantages in terms of variation control and  
54 powerful inference.

55 Markov chain Monte Carlo (MCMC) methods have become important in Bayesian computation, and they allow  
56 inferences to be drawn from complex posterior distribution where analytical or numerical integration techniques cannot  
57 be applied [Sorensen and Gianola, 2007]. By using MCMC, complex formulations can be analysed with comparative  
58 ease [Besag and Higdon, 1999]. Apart from the classical Metropolis-Hastings (MH) algorithm [Hastings, 1970, Metropolis et al., 1  
59 , the Gibbs sampling [Gelfand and Smith, 1990, Geman and Geman, 1984] is a special case of MH and is a method for  
60 sampling from distributions over at least two dimensions. See [Che and Xu, 2010, Chen et al., 2019, Montesinos López et al., 201  
61 for examples. However, the main problem of Gibbs sampling is that it update a single parameter at one time based on  
62 current status and end up with random walk behaviour, higher auto-correlation and slow mixing. So it has insufficient  
63 capability in dealing with correlated parameters for high-dimensional models [Brooks et al., 2011], and may require  
64 an unacceptably long time to converge to the target distribution [Hoffman and Gelman, 2014].

65 Alternatively, the For Bayesian inference, we use the No-U-Turn Sampler (NUTS), which is an extension to the  
66 Hamiltonian Monte Carlo (HMC), by Hoffman and Gelman [2014] in the R package rstan. The HMC, originally  
67 called hybrid Monte Carlo [Duane et al., 1987], is a MCMC algorithm that makes use of gradient information of the  
68 objective function to reduce random walk behaviour. HMC It uses “momentum”, a definition from physics, variables  
69 that accelerate each iteration within a parameter space to allow faster mixing and convergence [Brooks et al., 2011,  
70 Ng’ombe et al., 2020]. A problem with HMC is that it is highly sensitive to two user-specified parameters: a step  
71 size  $\epsilon$  and a desired number of steps  $L$ . A poor choice decreases the efficiency of HMC dramatically. Therefore, the  
72 No-U-Turn Sampler (NUTS) was developed as an extension to HMC by Hoffman and Gelman [2014]. The advantage  
73 of NUTS is that it Alternatively, NUTS determines the step size by adapting it during the warm-up (burn-in) phase  
74 to aiming at a target acceptance rate, and then used uses it for all sampling iterations [Monnahan et al., 2017]. It  
75 is a promising an useful sampling method because of its good sampling qualities: large effective sample sizes, low  
76 auto-correlations, and low skewness of marginal posterior distributions [Nishio and Arakawa, 2019]. It is embedded  
77 with the R package rstan. The advantage of using rstan rather than other “black box” packages, is that it has  
78 flexibility for researchers to implement complex and ad hoc models for particular problems. Besides, the Besides,  
79 NUTS does not require conjugate property on priors. As a summary, the advantages of NUTS are faster convergence  
80 for multi-parameters and has higher flexibility for self-defined models by researchers.

81 We estimated approximate the Bayesian inference and estimate the posterior distribution of the model that in-  
82 corporates spatially correlated random parameters of OFE data in rstan. Usually, that is the end of the story in  
83 most circumstances. However, with for OFE data. With advanced model diagnostic tools, such as probability integral  
84 transformation (PIT) checks [Gabry et al., 2019], and evaluation methods, such as Bayesian leave-one-out (LOO) cross  
85 validation (CV) for model comparison [Vehtari et al., 2017], we found that the model with Gaussian distribution is  
86 misspecified, and the influential points/outliers are inevitable. Therefore, we suggest to use Student-*t* distribution and  
87 the results are improved.

88 The objective of this paper is to demonstrate the capability of the Bayesian approach in modelling spatial  
89 nonstationarity and analysing spatially correlated data for OFE and the power of the approach for precision agri-  
90 culture. We suggest researchers run comprehensive model diagnosis and “think out of the box” that sometimes

91 Gaussian distribution does not work well. Our `rstan` script is open to public and can be applied to similar data  
 92 by other users. The paper is organised as follow: in Section 2 we propose a generic spatial model for OFE data;  
 93 in Section 3 we discuss the prior and posterior distribution for the model, and explained the mechanism of NUTS  
 94 sampler; in Section 4 we discuss the post-sampling model checking and diagnostic process; finally, in Section 5 we  
 95 apply the proposed model and Bayesian framework to Las Rosas corn yield data set, and compare it with GWR.

## 96 2 Statistical models

97 Piepho et al. [2011] summarise three general options regarding the range of spatial correlation for modelling geo-  
 98 referenced measurements: spatial correlation within but not between plots [Ritter et al., 2008], spatial correlation  
 99 across a whole block but not between blocks [Piepho et al., 2008] and spatial correlation across the whole field [Hong  
 100 et al., 2005, Hurley et al., 2004]. For OFE data, we focus on the third option.

### 101 2.1 Generic statistical model

102 Let  $y$  denote the response variable, it is assumed that  $y$  varies on  $\mathbf{Y} = \{y_1, \dots, y_n\}$  denotes the vector of  $n$   
 103 observations from an experimental field and spatially correlated. The distribution is assumed a multivariate distribution  
 104 denoted as. We assume that it follows the multivariate distribution  $\mathcal{MD}$ , with the predictor  $\eta$  and covariance with  
 105 predictors  $\eta = \{\eta_1, \dots, \eta_n\}$  and  $n \times n$  variance-covariance matrix  $\Sigma$ . The distribution  $\mathcal{MD}$  is usually a Gaussian or a  
 106 student-t distribution. We write It can be written as

$$y \sim \mathcal{M}\mathcal{MD}(\eta(\phi), \Sigma(\sigma)), \quad (1)$$

107 where  $\theta = \{\phi, \sigma\}$  is a set of parameters. The predictor  $\eta$  can be either linear or non-linear. The structure of the  
 108 variance-covariance matrix  $\Sigma(\sigma)$  covariance matrix  $\Sigma$  may be written as

$$\Sigma(\sigma) = V_s + \sigma^2 I, \quad (2)$$

109 where  $V_s$  is some spatial covariance structure and  $\sigma^2$  is a nugget variance. Alternatively, Gilmour et al. [1997] suggest  
 110 to use a separable first-order autoregressive process  $-AR1 \times AR1$ , to commence the spatial modelling process in such  
 111 way that

$$\Sigma(\sigma) = \sigma_s^2 V_s = \sigma_s^2 V_r(\rho_r) \otimes V_c(\rho_c), \quad (3)$$

112 which is a Kronecker product of two correlation matrices  $V_r(\rho_r)$  and  $V_c(\rho_c)$  for rows and columns by sorting the  
 113 residuals columns within rows.

114 Let  $x_1, \dots, x_l$  denote the explanatory variables at the global level (fixed effects) and  $z_1, \dots, z_k$  denote  
 115 the explanatory variables at the local level (random effects). Suppose we have  $n$  measurements from  
 116  $n$  locations  $s_1, \dots, s_n$  within a study region  $\mathcal{S}$ . For a given location  $s_i \in \mathcal{S}$ , the underlying template predictor term  $\eta$   
 117 of equation (1) can generally be written as

$$\eta(s_i) = \sum_{m=1}^l b_m x_m(s_i) + \sum_{j=1}^k u_j(s_i) z_j(s_i), \quad (4)$$

118 where the unknown  $b_m$  and  $u_j(s)$  are the coefficients that the latter one is associated with spatial variation, for the  
 119 fixed and random terms respectively, where  $i = 1, \dots, n$ .

120 In the matrix notation, the model given in above above in equations (1) and (4) including all the information is  
 121 expressed as

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (5)$$

122 where  $\mathbf{Y}$  is the  $n \times 1$  vector of observation,  $\mathbf{b}$  and  $\mathbf{u}$  are vectors of fixed and random effects, respectively, and  $\mathbf{X}$  and  
 123  $\mathbf{Z}$  are the associated design matrices. Hence, the distribution of  $\mathbf{Y}$  is a multivariate distribution with mean  $\mathbf{X}\mathbf{b}$  and  
 124 variance  $\mathbf{Z}\Sigma_u\mathbf{Z}^\top + \Sigma(\sigma)$ .

## 125 2.2 Model with spatially correlated random parameters

126 In the context of OFE or precision farming, [Rakshit et al. \[2020\]](#) used the GWR approach and analysed yield  
 127 monitor data for a corn field in Argentina [\[Edmondson, 2014\]](#). They ~~found assumed~~ that the nitrogen rate is in  
 128 quadratic form against yield, and the model is estimated through local regression.

129 As a complementary approach to GWR, we use model (5) by splitting all  $\beta$ s in GWR into  $bs$  and  $us$  and rewrite  
 130 the model as

$$131 \quad y_i = \mathbf{x}_i^\top \mathbf{b} + \mathbf{z}_i^\top \mathbf{u}_i + e_i, \quad (6)$$

131 where, at any location  $s_i \in \mathcal{S}$ ,  $y_i$  is the observation,  $\mathbf{x}_i$  and  $\mathbf{z}_i$  are vectors of treatment effects with three levels, and  
 132  $e_i$  is a residual error term that  $e_i \sim N(0, \sigma_e^2)$ .

133 Stacking random vectors  $\mathbf{u}_i$  for all  $y_i$  at location  $s_i$  into a vector  $\mathbf{u}$ , we assume that the random vector  $\mathbf{u}$  are  
 134 from the multivariate normal distribution, and different levels  $j$  of the same plot  $i$  are correlated. In this case, at any  
 135 location  $s_i$ , the covariance matrix of  $\mathbf{u}_i$  becomes  $V_u$ , which could be a known or unknown parameter in the model.

136 Without spatial correlation, the variance matrix of random parameters is

$$\Sigma_u = I_n \otimes V_u. \quad (7)$$

137 On the contrary, if the random parameters are spatially correlated with the spatial covariance matrix  $V_s$ , then the  
 138 above covariance matrix is re-written as

$$\Sigma_u = V_s \otimes V_u, \quad (8)$$

139 where  $V_s$  is the  $AR1(\rho_r) \otimes AR1(\rho_c)$  spatial variance matrix or a weighted distance matrix, and  $\text{Var}(\mathbf{e}) = \sigma^2 I$ . Hence,  
 140  $\mathbf{u} \sim N(0, \Sigma_u)$ .

141 Despite the fact that only a single treatment is directly observed at each plot, with this model, the estimation of  
 142 localised treatment effects  $\mathbf{u}_i$  is possible because the spatial model allows the exploiting of information from neigh-  
 143 bouring positions with other treatments [\[Piepho et al., 2011\]](#). Further, it is better to point out that the above variance  
 144 nugget  $\sigma^2$  could vary on the field, and thus for the term  $\sigma^2 I$  of a hierarchical model, one may assign a distribution to  
 145 it, such as  $\sigma_i \sim N(\mu_\sigma, sd_\sigma)$ , or in an alternative way that it is replaced by a diagonal matrix with  $\sigma^2(s_i)$  along the  
 146 diagonal. Then one may assume that  $\sigma \sim N(0, \sigma_s I)$ . In most scenarios, the latter option only increases the complexity  
 147 and is not necessary in our study.

## 148 3 Bayesian process

### 149 3.1 Prior specification

150 Being different from REML approaches, Bayesian approaches assume parameters are random variables and start  
 151 the estimation from a prior distribution, which summarises the previous knowledge or the experience before the  
 152 experiment about the parameters to be estimated [\[Onofri et al., 2019\]](#). The prior should be clearly defined, even  
 153 before the experiment has been conducted.

154 The selection of priors in Bayesian inference has been discussed for a long time. In usual cases, it is common to  
 155 use a flat non-informative prior, also known as “improper prior” that  $p(\theta) \propto \text{constant}$ , if we know nothing from earlier  
 156 studies [\[Gelman et al., 2006\]](#). In other circumstances, a Cauchy or Gamma prior remains a reasonable candidate.  
 157 Some researchers prefer inverse Wishart (IW) or inverse Gamma distributions for the standard deviation of a hierar-  
 158 chical model. However, [\[Gelman et al., 2017\]](#) suggested to use weakly informative priors instead, and half- $t$  family is  
 159 recommended for standard deviation when the number of groups is small. The general idea is that such a prior affects  
 160 the information in the likelihood as weakly as possible.

161 For the covariance matrices  $V_u$ , it is possible to use an inverse Wishart distribution [\[Kass and Natarajan, 2006\]](#).  
 162 However, there are good arguments against this prior and, in some settings, a weakly informative prior is desired.

McElreath [2015], Sorensen et al. [2016] use the following weakly informative prior for the variance matrix, that

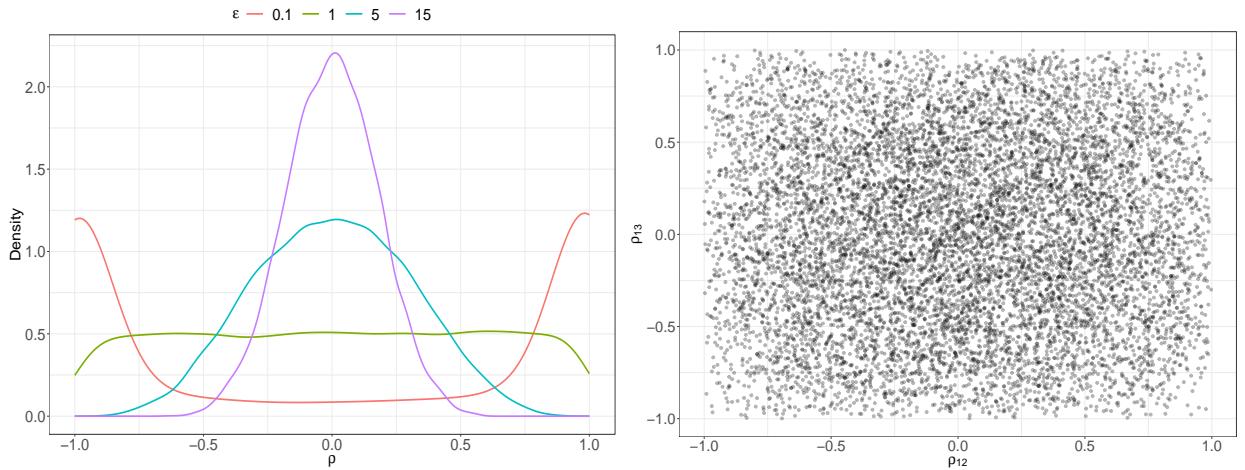
$$V_u = B(\sigma_u)R_uB(\sigma_u) \quad (9)$$

$$R_u \sim \text{LKJcorr}(\epsilon) \quad (10)$$

where  $B(\sigma_u)$  denotes the diagonal matrix with diagonal elements  $\sigma_u$ , which could be half- $t$ , half-Cauchy or inverse-Gamma distribution and  $R_u$  is a weak LKJ-prior identity matrix with respect to a positive value  $\epsilon$ . This statement of prior will adaptively regularise the individual coefficients of random effects and the correlation among them. See [Gabry et al., 2019, Gelman et al., 2017] for some discussions. A  $3 \times 3$  symmetric correlation matrix  $R$ , for instance, is as

$$R = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix}, \quad (11)$$

where  $\rho$ s are the pairwise correlation parameters. Figure 1 demonstrates how the distribution of  $\rho$  is influenced by  $\epsilon$ . Small  $\epsilon$  leads a wider tail and big  $\epsilon$  narrows down the tail for  $\rho$ .



(a) Distribution of correlation coefficients  $\rho$  extracted from random  $2 \times 2$  correlation matrices with different  $\epsilon$ s. (b) Visualisation of  $\rho_{12}$  against  $\rho_{13}$  from a  $3 \times 3$  correlation matrix with  $\epsilon = 1$ .

Figure 1: LKJcorr( $\epsilon$ ) probability density. When  $\epsilon = 1$ , all correlations are equally plausible. As  $\epsilon$  increases, extreme correlations become less plausible.

### 3.2 Posterior distribution

Suppose that we are interested in estimating  $\theta$  from data  $\mathbf{Y}$  using the probability density  $p(\mathbf{Y} | \theta)$ , the Bayes theorem tells us that the joint posterior density of the parameters becomes

$$p(\theta | \mathbf{Y}) = \frac{p(\mathbf{Y} | \theta)\pi(\theta)}{p(\mathbf{Y})}, \quad (12)$$

where  $\pi(\theta)$  is known as the prior distribution that represents our previous knowledge or “best guess”,  $p(\mathbf{Y}) = \int p(\mathbf{Y} | \theta)\pi(\theta)d\theta$  is the normalising constant and does not affect the inference. Therefore, people prefer to re-write the equation to the form

$$p(\theta | \mathbf{Y}) \propto p(\mathbf{Y} | \theta)\pi(\theta), \quad (13)$$

which is also called the marginal posterior distribution. The distribution of  $p(\theta | \mathbf{Y})$  is the “Bayesian inference” of the parameter because all information about  $\theta$  is contained in the distribution [Che and Xu, 2010]. Then, by taking natural logarithm on the posterior of (13), we will have

$$\ln L(\theta) \propto -\frac{1}{2}(\mathbf{Y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{u})^\top \Sigma(\sigma)^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{u}) + \ln \pi(\theta). \quad (14)$$

<sup>177</sup> Because of  $\mathbf{u} \sim N(0, \Sigma_u)$ , for faster gradient evaluation and sampling, we impose Cholesky decomposition, such that

$$\Sigma_u = \Sigma(\rho_r) \otimes \Sigma(\rho_c) \otimes V_u = (L_r L_r^\top) \otimes (L_c L_c^\top) \otimes (L_u L_u^\top) = (L_r \otimes L_c \otimes L_u)(L_r \otimes L_c \otimes L_u)^\top. \quad (15)$$

<sup>178</sup> Therefore,

$$\tilde{\mathbf{u}} = (L_r \otimes L_c \otimes L_u)(z_r \otimes z_c \otimes z_u) = (L_r z_r) \otimes (L_c z_c) \otimes (L_u z_u), \quad (16)$$

<sup>179</sup> where  $z_r, z_c, z_u$  are vectors of i.i.d. random variables sampled from  $N(0, 1)$ .

### <sup>180</sup> 3.3 No U-turn sampler

<sup>181</sup> Suppose  $\theta$  is a set of parameters of interests and  $f(\theta)$  is the density function. In Hamiltonian system, a set of  
<sup>182</sup> auxiliary momentum variables  $r$  that are drawn independently from the standard normal distribution is introduced in  
<sup>183</sup> order to keep the system invariant. Then the joint distribution of  $f(\theta, r)$  is

$$f(\theta, r) = \exp\{-U(\theta) - K(r)\} = \exp\{-H(\theta, r)\}, \quad (17)$$

<sup>184</sup> where  $H(\theta, r)$  is the Hamiltonian system dynamic equation with potential energy  $U(\theta)$  and kinetic energy  $K(r)$ ,  
<sup>185</sup> which represent the density functions of  $\theta$  and  $r$  respectively. The property of the dynamics is that it keeps the joint  
<sup>186</sup> distribution invariant [Nishio and Arakawa, 2019].

<sup>187</sup> When new samples  $(\theta^*, r^*)$  are drawn, while Hamiltonian dynamics system is numerically approximated in discrete  
<sup>188</sup> time space with the leapfrog method to maintain the total energy. The leapfrog methods requires two parameters: the  
<sup>189</sup> step size  $\epsilon$ , how far the next draw is from current sample, and  $L$ , the number of steps to move in the process. When  
<sup>190</sup> the process is finished, new samples are accepted with probability

$$\alpha = \min\left\{1, \frac{f(\theta^*, r^*)}{f(\theta, r)}\right\}. \quad (18)$$

<sup>191</sup> Because HMC is highly sensitive to  $\epsilon$  and  $L$ , [Hoffman and Gelman, 2014] proposed the No-U-Turn Sampler  
<sup>192</sup> (NUTS), which determines the step size by adapting it during the warm-up (burn-in) phase to a target acceptance  
<sup>193</sup> rate, and then used it for all sampling iterations [Monnahan et al., 2017]. It also gets rid of  $L$  by using the criterion

$$\frac{d}{dt} \frac{(\theta^* - \theta) \cdot (\theta^* - \theta)}{2} = (\theta^* - \theta) \cdot \frac{d}{dt}(\theta^* - \theta) = (\theta^* - \theta) \cdot r^* < 0, \quad (19)$$

<sup>194</sup> where  $r^*$  is the current momentum and  $(\theta^* - \theta)$  is the distance from initial position to current position. The idea is  
<sup>195</sup> that the trajectory will keep exploring the space until  $\theta^*$  starts to move back towards to  $\theta$ .

<sup>196</sup> To guarantee time reversibility and to converge to the correct distribution, NUTS overcomes this issue by means of  
<sup>197</sup> a recursive algorithm that preserves reversibility by running the Hamiltonian simulation both forward and backward  
<sup>198</sup> in time [Hoffman and Gelman, 2014]. This process starts with a slice variable  $u$  which is uniformly distributed as  
<sup>199</sup>  $p(u | \theta, r) = U(0, f(\theta, r))$ , and generates a finite set of all  $(\theta, r)$  in the doubling size process by randomly taking forward  
<sup>200</sup> and backward leapfrog steps until

$$(\theta^+ - \theta^-) \cdot r^- < 0 \quad \text{or} \quad (\theta^+ - \theta^-) \cdot r^+ < 0, \quad (20)$$

<sup>201</sup> where  $\theta^-, r^-$  and  $\theta^+, r^+$  are the leftmost and rightmost of  $\theta$  and  $r$ . The best candidate  $(\theta^*, r^*)$  is uniformly sampled  
<sup>202</sup> from the subset of all candidate  $(\theta, r)$ .

<sup>203</sup> Precise definition and description of NUTS algorithm can be found at [Hoffman and Gelman, 2014].

## <sup>204</sup> 4 Post-sampling checking

<sup>205</sup> Gelman [2003] suggests a few strategies for Bayesian model checks: (1) checking that the posterior inferences are  
<sup>206</sup> reasonable, given the substantive context of the model; (2) examining the sensitivity of the inferences to reasonable  
<sup>207</sup> changes in the prior distribution and the likelihood; and (3) checking that the model is capable of generating data

like the observed data. One may refer to [Congdon, 2019, Gelman, 2004, Gelman et al., 2013, Weiss, 2016] for explanations and applications. We focus the third strategy by graphically checking the replicate and observed data for the reproducible capability of our model.

In ideal situations, researcher are able to use independent data set which is not used in modelling process to test the performances of different models. Alternatively, one may split one data set into training and testing data sets, and use the former one for model training and latter one for testing. However, it is not feasible for some experimental agricultural data sets. Nevertheless, we can still check the model by using the data that we already have.

## 4.1 Posterior predictive checking

The posterior predictive checking uses the posterior of the parameter in the model to regenerate the observations. The idea behind the concept is that if a model is a good fit we should be able to use it to generate data that resemble the data that we observed [Gabry et al., 2019]. In other words, let  $\mathbf{Y}^{rep}$  denotes the replicate data if the process that generated the data  $\mathbf{Y}$  is replicated with the same value of  $\theta$  that generated the observed data. Then  $\mathbf{Y}^{rep}$  is governed by the posterior predictive distribution

$$p(\mathbf{Y}^{rep} | \mathbf{Y}) = \int p(\mathbf{Y}^{rep} | \theta) p(\theta | \mathbf{Y}) d\theta. \quad (21)$$

The samples  $\mathbf{Y}^{rep}$  is checked against the data  $\mathbf{Y}$  [Congdon, 2019, Dipak Dey and C.R. Rao, 2005].

The application of posterior predictive distributions is more robust than prior specification, because the details of the prior are washed out by the likelihood [Gelman et al., 2017] if the model is good enough.

## 4.2 Model diagnosis and evaluation

Leave-one-out (LOO) cross validation (CV), where one data is omitted at one time and the fitted model based on the remaining data is the best predictor in terms of mean square error, is widely used for model evaluation and selection. In Bayesian statistics, to measure the predictive accuracy, researchers use the expected log LOO predictive density (ELPD), which is as:

$$\text{elpd}_{\text{loo}} = \sum_{i=1}^n \log p(y_i | y_{-i}), \quad (22)$$

where  $p(y_i | y_{-i}) = \int p(y_i | \theta)p(\theta | y_{-i})d\theta$  is the LOO predictive density with the  $i$ -th data omitted from the data set [Vehtari et al., 2017]. However, it is too expensive in computing because it requires refitting the model  $n$  times. Bürkner et al. [2020] proposed an approximated LOO CV using only a single model fit by instead calculating the pointwise log predictive density as a fast approximation to the exact LOO CV. It uses the Pareto-smoothed importance-sampling algorithm [Vehtari et al., 2017] on the pointwise log-likelihood matrix for each draw  $m$  from the full posterior distribution and obtains a PSIS-LOO-CV estimate, which is

$$\widehat{\text{elpd}}_{\text{psis-loo}} = \sum_{i=1}^n \log \left( \frac{\sum_{m=1}^M p(y_i | \theta^{(m)}) w_i^{(m)}}{\sum_{m=1}^M w_i^{(m)}} \right), \quad (23)$$

where  $w_i^{(m)}$  are stabilised weights by PSIS,  $m = 1, \dots, M$ .

The resulting PSIS LOO CV approximations 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 to the LOO predictive distribution for each left out point [Gabry et al., 2019], and the estimated tail shape parameter  $\hat{k}$  of the generalised Pareto distribution can be used for assessing the reliability and approximate convergence rate of the estimates. If  $\hat{k} < 0.5$  then the distribution of raw importance ratios has finite variance and the central limit theorem holds. But in practice the model is still 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. It is the variance of the

244 predicted values divided by the variance of predicted values plus the expected variance of the errors in such way

$$\text{Bayesian } R^2 = \frac{\text{Var}(\mathbf{Y}^{pred})}{\text{Var}(\mathbf{Y}^{pred}) + \text{Var}(e)}. \quad (24)$$

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

## 247 5 Analysis of Las Rosas data

### 248 5.1 Data visualisation

249 A part of Las Rosas data set, which is publicly available by the name of `lasrosas.corn` in the R-package `agridat`  
250 [Edmondson, 2014], was used in our study. The yield monitor data for a corn field was conducted by incorporating six  
251 nitrogen rate treatments in three replicated blocks comprising 18 strips across the field in Argentina. It also consists  
252 a topographic factor with four levels: W (West slope), HT (Hilltop), E (East slope) and LO (Low East). An obvious  
253 spatial variation in the corn yield can be seen in Figure 2.

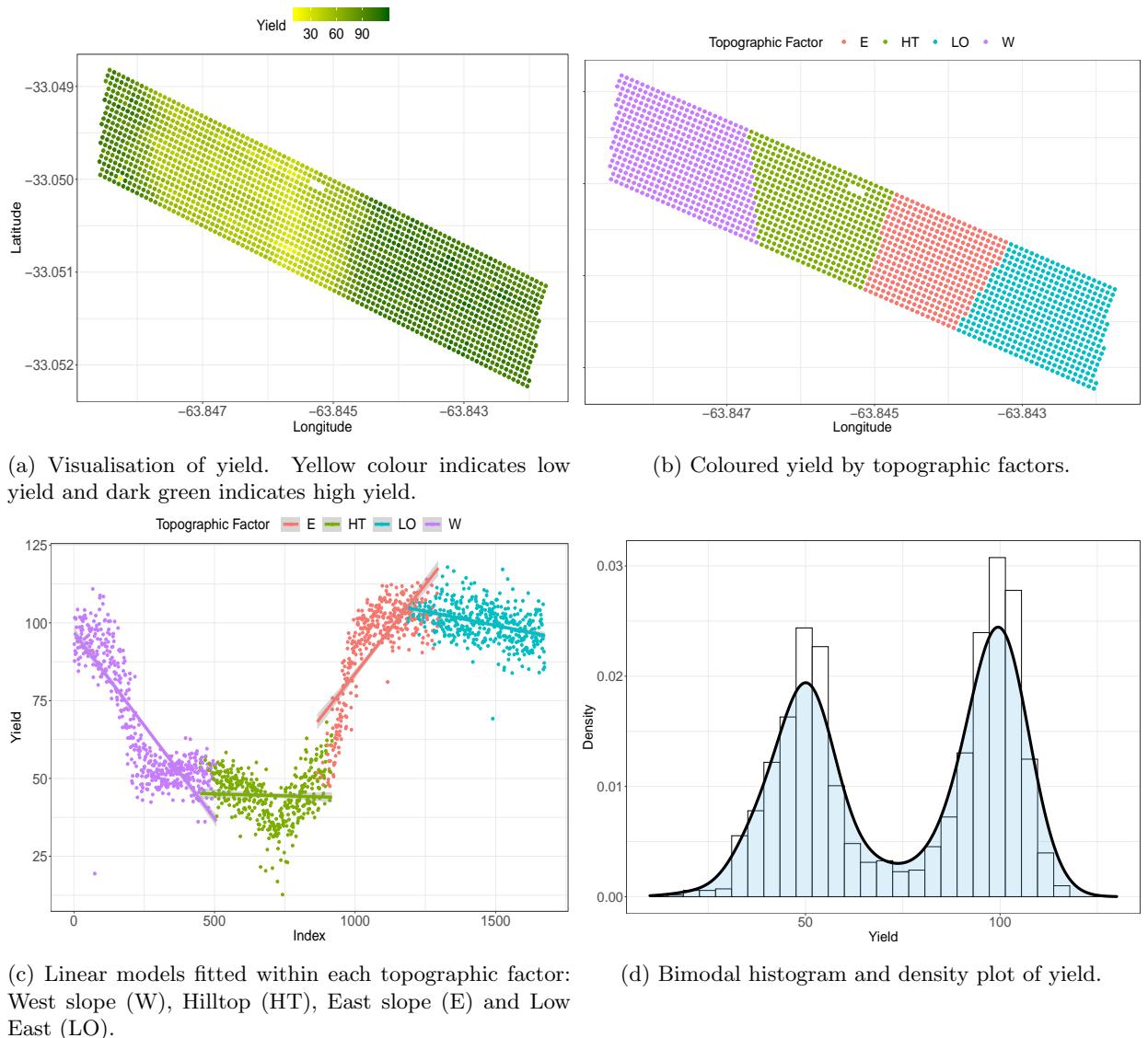


Figure 2: Visualisation of Las Rosas yield monitor data for harvests in 2001.

254 In order to analyse the data set, we apply a geographic projection which transforms the geo-spatial coordinates  
255 to planar coordinates expressed in meters. The Las Rosas experiment is approximately 810m long and 150m wide  
256 [Rakshit et al., 2020].

## 257 5.2 Statistical models and prior predictive simulations

258 To demonstrate the power of the proposed model (6) which has spatially correlated random parameters  $\mathbf{u}$ , we  
 259 put the model without spatial correlation as the benchmark model. We will also prove that, with the assumption  
 260 of Gaussian distribution, the model is mis-specified. Instead, we use student- $t$  distribution in the final model. To  
 261 summarise, we have four models illustrating in Table 1.

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

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

262 We start from choosing weakly informative priors for model 1. A weakly informative prior has the capability that  
 263 samples from the prior distribution through data-generating process could represent any data set that is plausibly  
 264 observed.

265 Vague priors, for instance, are  $b_0 \sim N(\mu, 100)$ ,  $b_1, b_2 \sim N(0, 100)$  and  $\sigma_e \sim IG(1, 100)$ , where  $IG$  is the inverse  
 266 Gamma distribution. Similarly, we assume  $u_{ki} \sim N(0, \sigma_k^2)$  with  $\sigma_k^2 \sim IG(1, 100)$  and  $k = 0, 1, 2$  at location  $s_i$ .  $\mu$  is  
 267 either mean or median of the observed data. Alternatively, weakly informative priors that we choose are  $b_0 \sim N(80, 10)$ ,  
 268  $b_1 \sim N(0, 0.01)$ ,  $b_2 \sim N(0, 0.001)$ ,  $\sigma_0 \sim N_+(0, 1)$ ,  $\sigma_1 \sim N_+(0, 0.01)$ ,  $\sigma_2 \sim N_+(0, 0.001)$ ,  $R_u \sim LKJcorr(1)$  and  
 269  $\sigma_e \sim N_+(0, 1)$ , where  $N_+(\cdot)$  is the positive half Gaussian distribution.

270 Figure 3 compares the regenerated data with vague and weakly informative priors. When the vague priors are  
 271 applied, model 1 regenerates extremely small and large values which are implausible for yield data. This is mostly  
 272 because the vague priors disregard practical knowledge. Applying weakly informative priors avoids negative values  
 273 and keeps the simulations within reasonable interval. Even though some simulations are not perfect, the priors are  
 274 overall good enough according to our common knowledge. If the priors, on the other hand, are too informative, they  
 275 may badly influence the posterior distribution and result in partial exploration in the posterior space.

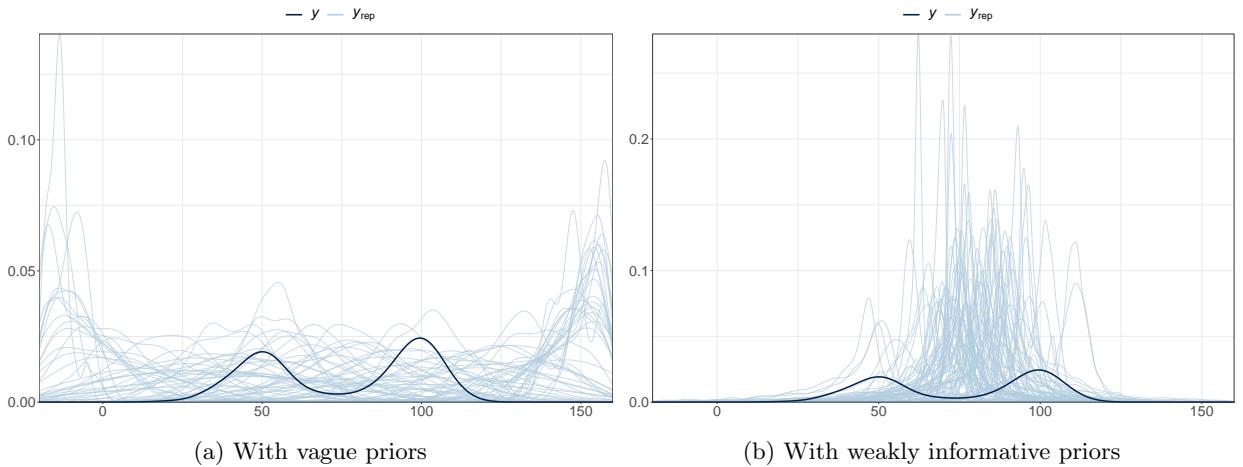


Figure 3: Capability of re-generating observed data with different priors by running 100 simulations. Vague priors failed in regenerating and leads to extreme values. Weakly informative priors give plausible regenerated data.

276 For model 2, besides the priors that are used in model 1, we add the following priors:  $\rho_c, \rho_r \sim U(0, 1)$  a uniform  
 277 distribution between 0 and 1. For model 3 and 4, the prior was assigned to the extra parameter of degrees of freedom  
 278 as  $\nu \sim \text{Gamma}(2, 0.1)$ , suggested by Juárez and Steel [2010] that we keep the prior as easy as possible.

279 A summary of the priors are listed in Table 2. Please note that it is usually not a good idea to use the same priors  
 280 for different models, and prior predictive checking is recommended for each model if a new prior is proposed for a new  
 281 parameter. In our study, with the same weakly informative priors, our models regenerate plausible outcomes, and  
 282 hence we have confidence that we only need to add extra priors to new parameters and the rest priors stay unchanged.

283     Additionally, NUTS does not require conjugate priors, which means the results are still valid if our priors are  
 284     Gaussian but the data is Student distribution.

	Model 1	Model 2	Model 3	Model 4
$b_0$			$N(80, 10)$	
$b_1$			$N(0, 0.01)$	
$b_2$			$N(0, 0.001)$	
$\sigma_0$			$N_+(0, 1)$	
$\sigma_1$			$N_+(0, 0.01)$	
$\sigma_2$			$N_+(0, 0.001)$	
$\sigma_e$			$N_+(0, 1)$	
$R_u$	—	LKJcorr(1)	—	LKJcorr(1)
$\rho_c$	—	$U(0, 1)$	—	$U(0, 1)$
$\rho_r$	—	$U(0, 1)$	—	$U(0, 1)$
$\nu$	—	—	$Gamma(2, 0.1)$	$Gamma(2, 0.1)$

Table 2: Priors of four models.

285     With the above priors, the models were run on four parallel Markov chains in `rstan` with each chain contains 2000  
 286     iterations where the first 1000 iterations are in warm-up phase. The final posterior contains 4000 samples for each  
 287     parameter.

### 288     5.3 Posterior checking

289     The prior predictive checking is a powerful tool for understanding the model. However, it is not able to extend this  
 290     technique to model selection and performance evaluation. We use MCMC diagnostic tools and posterior predictive  
 291     checking for further analyses and comparison.

292     We firstly use posterior predictive checking to visualise the performance of the four models. Figure 4 displays the  
 293     results of the PP checking. If the random parameters are not spatially correlated, model 1 and 3 are incapability in  
 294     regenerating the data and capture the feature of the observations. But for model 2 and 4, the results are promising.

295     Figure 5 is the investigation the skewness of the posterior predictive distribution. Model 2 and 4 capture observed  
 296     skewness, and model 1 and 3 provides, indicated by the plots, are terribly misspecified.

297     To visualise the performance of the models, we use the LOO CV predictive cumulative density plots, which are  
 298     asymptotically uniform (for continuous data) if the model is well calibrated [Gabry et al., 2019, Gelman et al., 2013].  
 299     Figure 6 compare the density of the computed LOO PIT (the thick dark curve) versus 100 simulated data sets from a  
 300     standard uniform distribution (the thin light curves). Model 1 and 3 are obviously mis-calibrated. The frown shape  
 301     by model 2 indicates that the model is overall okay but still a little mis-calibrated comparing with Model 4. It is either  
 302     mis-specified or too flexible. Model 4 performs the best among all four models.

303     Pareto  $\hat{k}$  diagnosis is vital in Bayesian analysis as well. Model 1 has too many large  $\hat{k}$  values, which indicates that  
 304     the model is either misspecified or too flexible. Similarly for model 3 that there are 24 “bad” values. These results  
 305     should not be interpreted solely but together with LOO PIT and  $p_{loo}$  values in Table 4. If we look at Figure 4, it  
 306     is clearly that model 1 and model 3 are misspecified. The LOO PIT plots 6 confirm it even if there are no “bad”  $\hat{k}$   
 307     values. If there are high Pareto  $\hat{k}$  values but the model fitting is okay, it means the model is both misspecified and  
 308     flexible, which indicates that the model has a good capability in predicting unknown data. In this scenario,  $K$ -fold  
 309     CV is recommended.

310     For model 2, there is one high  $\hat{k}$  value, which could be a high influential point or an outlier. However, it still  
 311     indicates that there might be an issue with this model. Therefore, instead of using Gaussian distribution, model 4 uses  
 312     Student- $t$  distribution and all  $\hat{k}$  values are small (less than 0.7). Then LOO CV does the balancing of goodness-of-fit  
 313     and parsimony automatically, and  $elpd_{loo}$  and Bayesian  $R^2$  are valid.

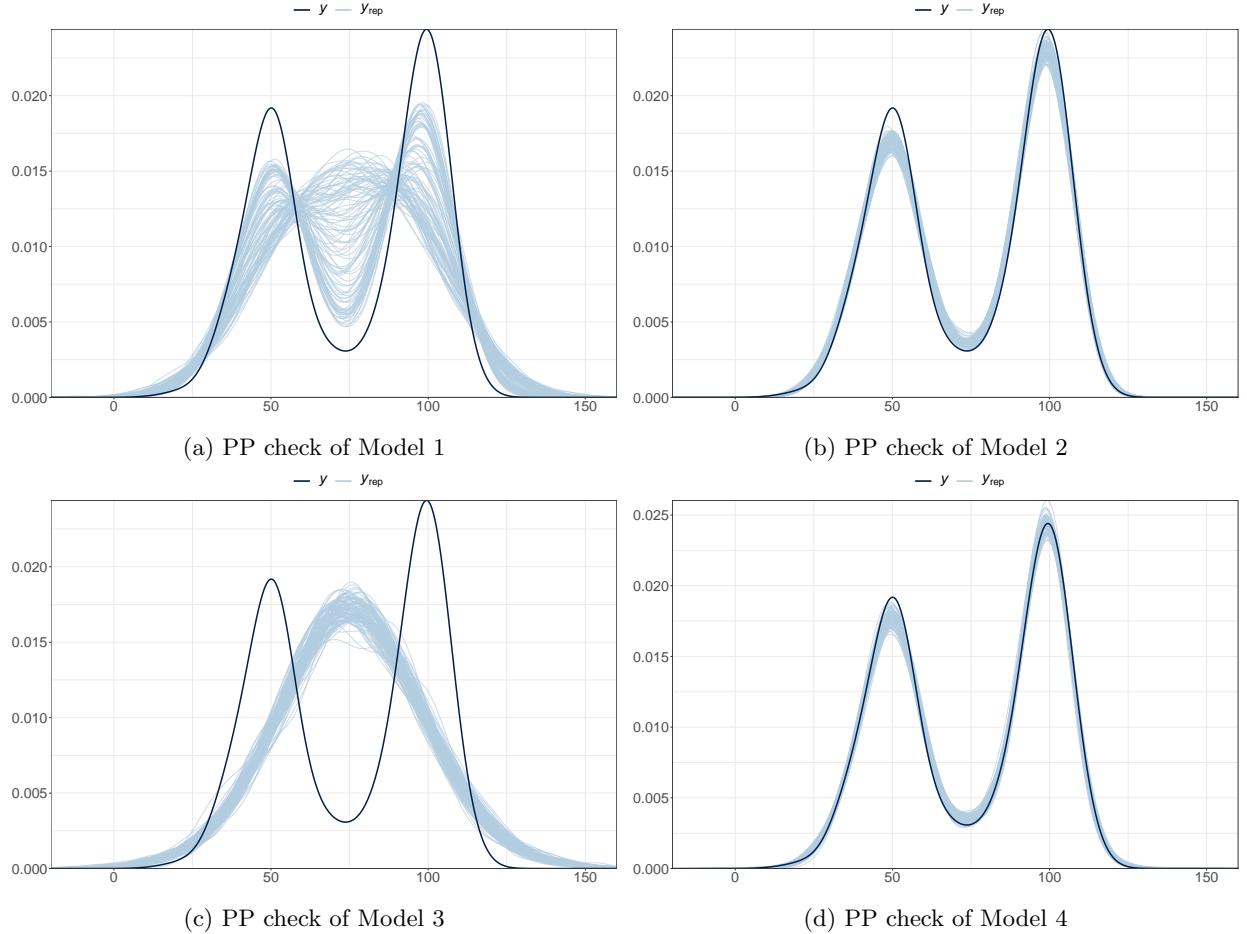


Figure 4: Posterior predictive checking for simple linear and the proposed spatial models with 100 simulations.

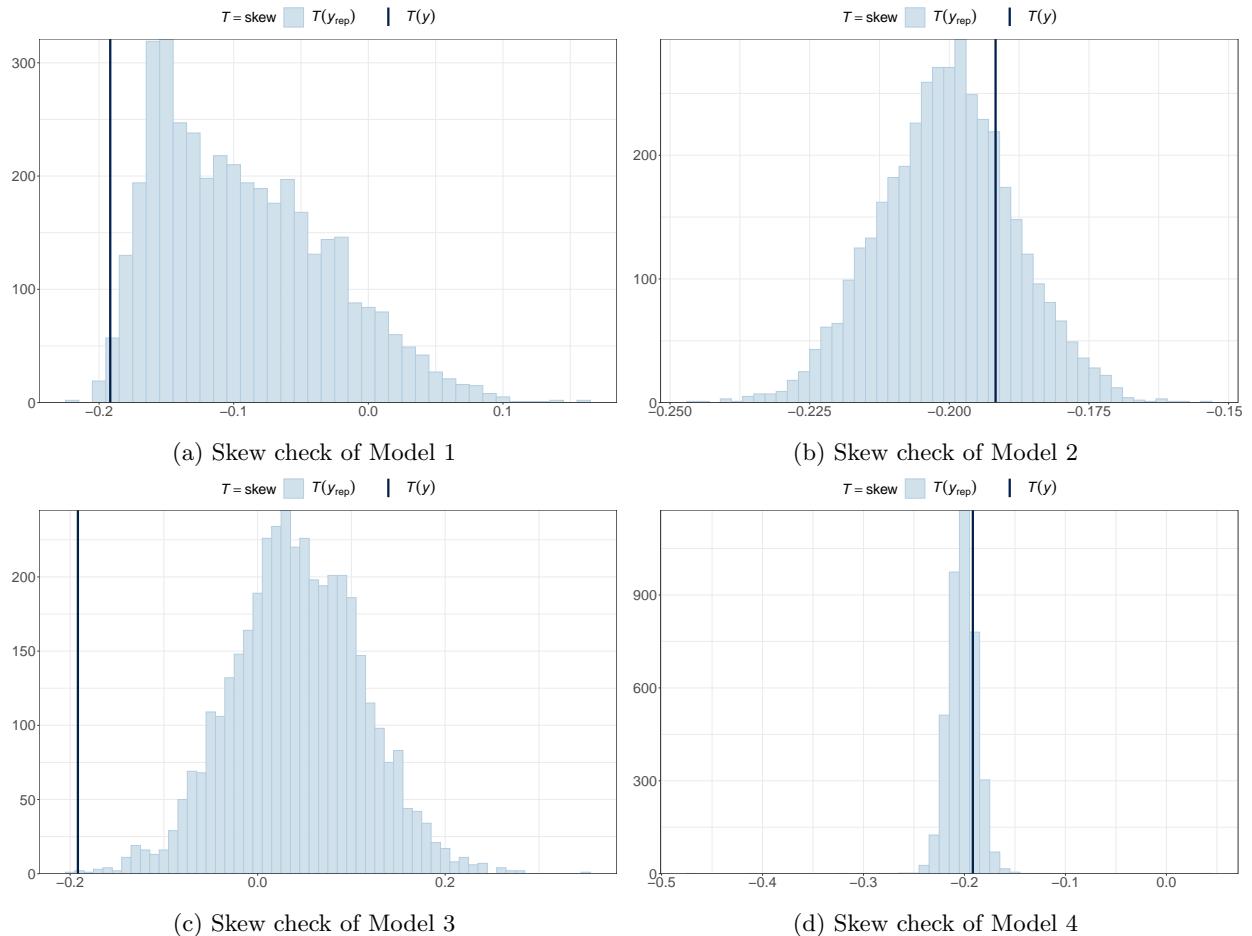


Figure 5: Histograms of skewness for 4000 draws from the posterior predictive distribution.

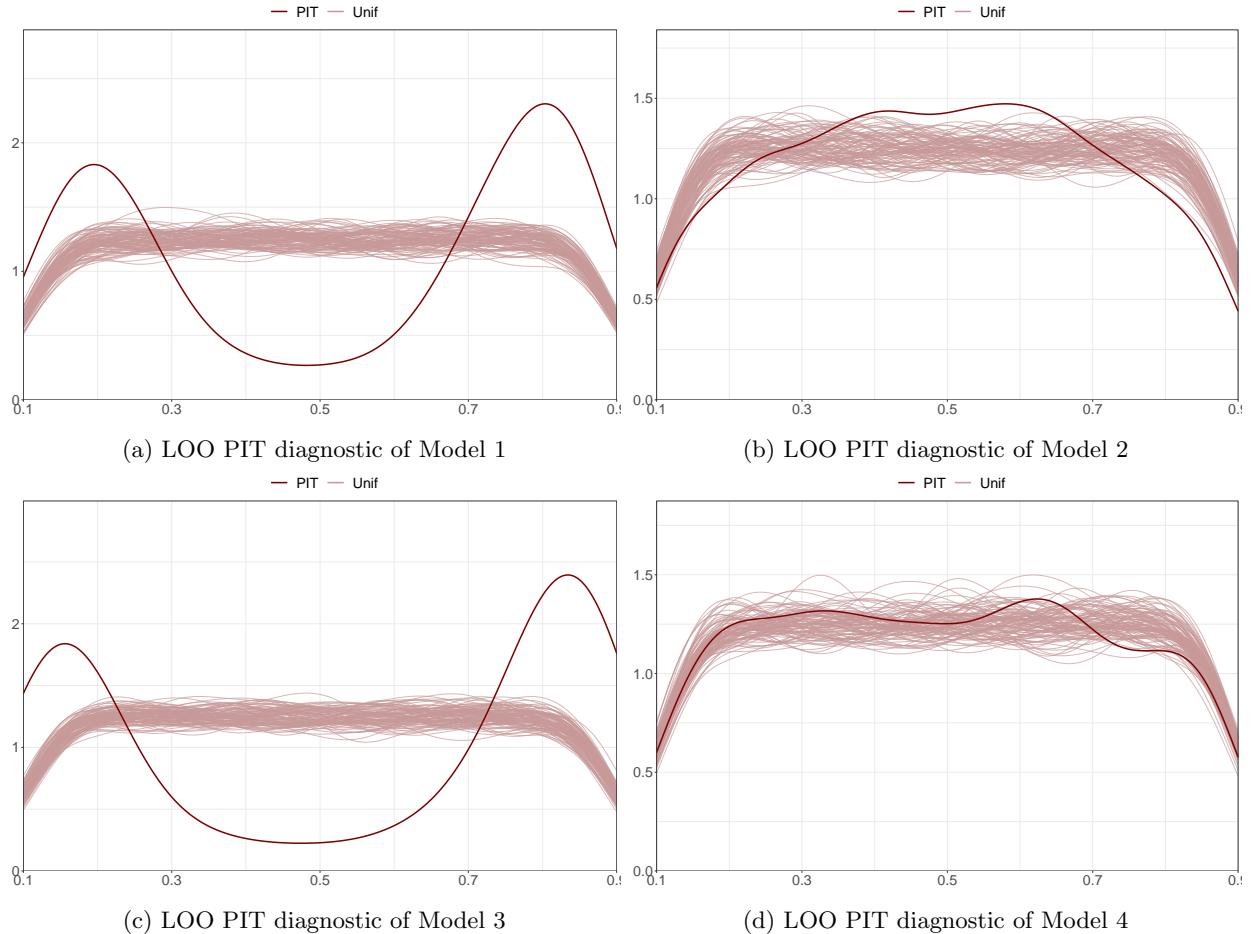


Figure 6: LOO PIT plots of the four models.

	Model 1			Model 2			Model 3			Model 4		
	Count	Per	M.Eff									
(-Inf, 0.5] (good)	28	1.7%	457	1673	99.9%	831	1474	88.1%	494	1673	99.9%	2990
(0.5, 0.7] (ok)	372	22.2%	112	0	0%	—	176	10.5%	254	1	0.1%	589
(0.7, 1] (bad)	1138	68.0%	18	0	0%	—	24	1.4%	170	0	0.0%	—
(1, Inf) (very bad)	136	8.1%	8	1	0.1%	11	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.

### 314 5.4 Model evaluation

315 We use ELPD, LOO CV scores and Bayesian  $R^2$  to evaluate and compare the performance of different models.  $p_{\text{loo}}$   
 316 is the effective number of parameters and is calculated as the subtraction of  $\text{elpd}_{\text{loo}}$  and the non-cross-validated log  
 317 posterior predictive density. In Bayesian analysis, even if there are no high Pareto  $\hat{k}$  values,  $R^2$  can't be trusted if  $p_{\text{loo}}$   
 318 is relatively high compared to the total number of parameters or the number of observations, which indicates weakly  
 319 predictive capability and probably a model misspecification.

320 The results from our study are shown in Table 4. Other than mean and standard deviation of the posterior  
 321 distribution, the credibility interval (CI) is reported in the table as well. The level  $\alpha$  equal tail credibility interval is  
 322 the interval bracketed by the  $\alpha/2$  and the  $1 - \alpha/2$  quantiles of the posterior samples, where  $\alpha \in (0, 1)$ . Like frequentist  
 323 way, we chose a typical  $\alpha = 0.05$ .

	Model 1		Model 2		Model 3		Model 4	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
elpd <sub>loo</sub>	-7236.2	13.4	-5253.4	100.6	-7848.4	17.1	-5133.0	37.8
p <sub>loo</sub>	1487.1	11.7	91.9	16.6	241.2	6.8	97.9	2.4
looic	14472.5	26.7	10506.7	201.1	15696.8	34.3	10266.0	75.7
	Median	CI	Median	CI	Median	CI	Median	CI
Bayesian $R^2$	0.842	0.563~0.965	0.956	0.953~0.959	0.190	0.135~0.251	0.974	0.970~0.977

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

324 Even if model 1 has higher  $R^2$  values compared to model 3, because of a few high Pareto  $\hat{k}$  and large  $p_{\text{loo}}$  values,  
 325 the  $R^2$  should not be interpreted. Then we should only focus on model 2 and 4. Apparently, Student- $t$  is better than  
 326 Gaussian distribution in terms of smaller LOO CV score and higher  $R^2$  value.

## 327 6 Results

328 Table 5 illustrates the statistics of the posterior distribution of all parameters and Figure 7 display the overall  
 329 coefficients, fixed terms plus random terms, for the intercept  $\hat{\beta}_0 = \hat{b}_0 + \tilde{u}_0$ , linear term  $\hat{\beta}_1 = \hat{b}_1 + \tilde{u}_1$  and quadratic  
 330 term  $\hat{\beta}_2 = \hat{b}_2 + \tilde{u}_2$ . In the Hilltop (middle area) and eastern of West slope (left side) areas, the quadratic term  
 331  $\beta_2$  is negative, which means the optimal treatment level exists. Due to the factors of brightness, sunshine or wind  
 332 magnitude, the production of yield will decrease if it is over treated.

### 333 6.1 Yield prediction

334 With the results from the previous section, we are now able to predict yield with given nitrogen rates. Suppose we  
 335 apply the medium nitrogen rate 75.4 kg/ha on the entire farm, the yield is calculated by substituting it into the model,  
 336 where parameters are drawn from the posterior distribution. After a few iterations, we may calculate the median of  
 337 yield for each grid. The result is shown in Figure 8.

### 338 6.2 Comparing to GWR

339 GWR approach, proposed by [Rakshit et al. \[2020\]](#), uses geographic weighted local regression to estimate the optimal  
 340 nitrogen rate for each plot and to predict the yield for OFE. Compared to GWR, the Bayesian approach uses Bayes  
 341 theorem and NUTS sampler to explore the posterior distribution of the objective and, with Bayesian inference, to  
 342 estimate the parameters for each plot.

343 They are two different paths but aiming at the same goal: providing easy interpretation and visual contour plots  
 344 for plant growers. Each of these approaches has its own advantages.

345 For GWR, the reliability of the results relies on the bandwidth of the moving window. The optimal bandwidth is  
 346 selected by cross validation. All data points within the window are used for inferring the information of the quarry  
 347 plot. Other data points, which are out of the window, contributes nothing to the target plot. It has "higher resolution"

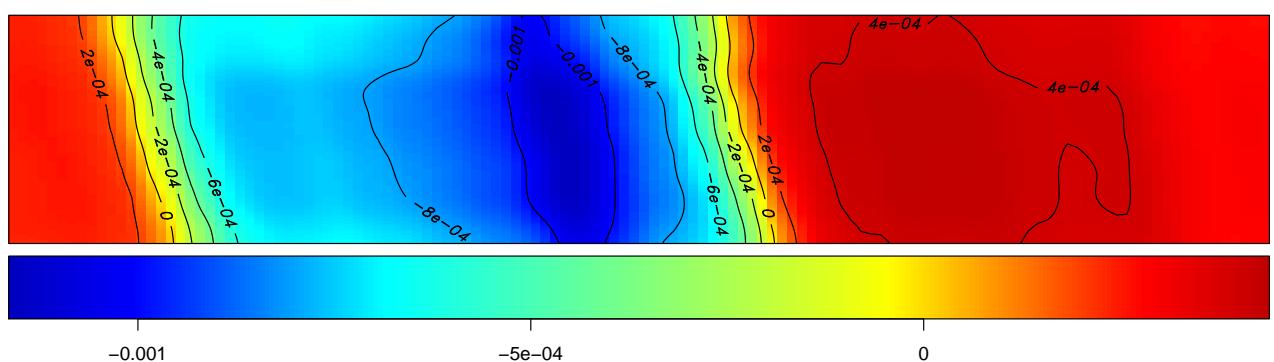
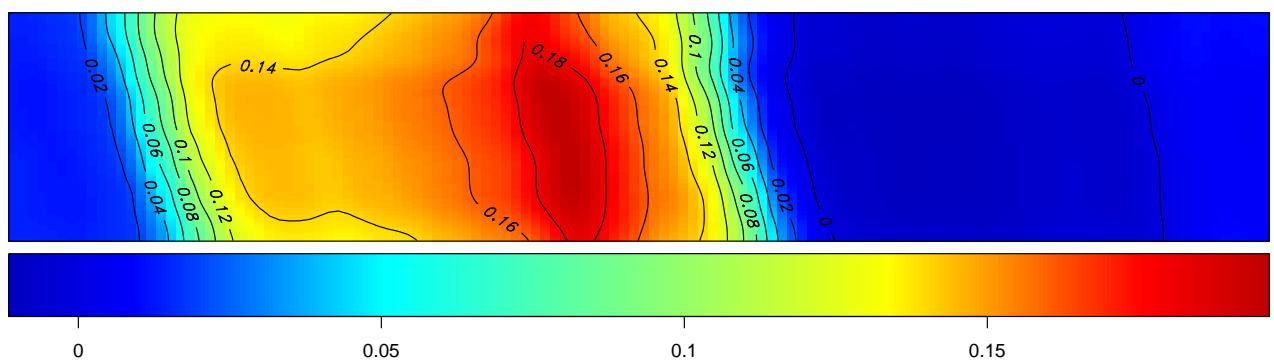
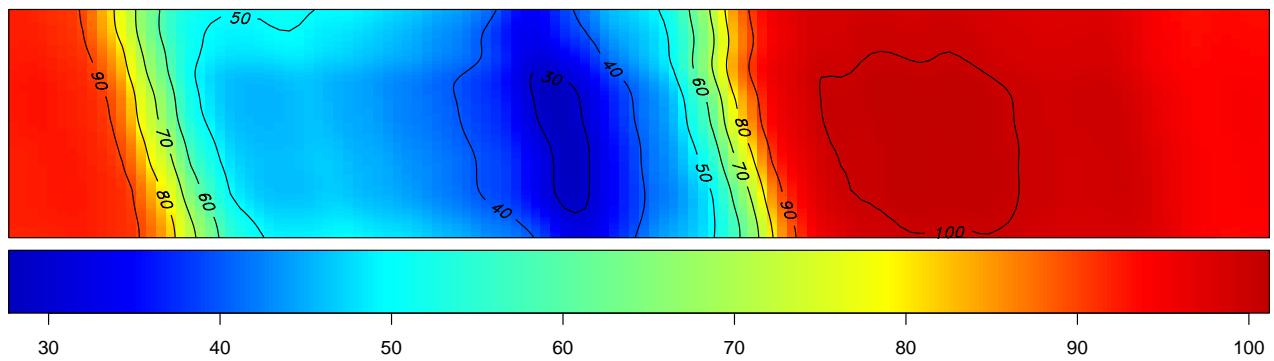


Figure 7: Contour plots of overall  $\hat{\beta}_0$  (top),  $\hat{\beta}_1$  (middle) and  $\hat{\beta}_2$  (bottom).

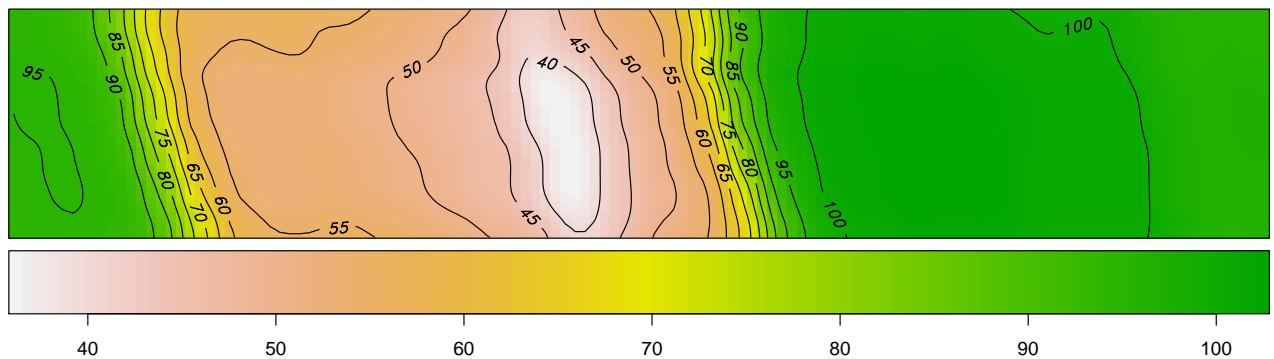


Figure 8: Predicted yield with medium nitrogen rate 75.4 kg/ha across the field.

Parameter	Mean	SD	Credibility interval		
			2.5%	Median	97.5%
$\hat{b}_0$	88.700	2.834	83.473	88.549	94.627
$\hat{b}_1$	0.031	0.009	0.013	0.031	0.047
$\hat{b}_2(\times 10^{-4})$	1.140	0.690	-0.180	1.131	2.545
$\hat{\sigma}_0$	2.453	0.564	1.487	2.412	3.670
$\hat{\sigma}_1$	0.011	0.005	0.004	0.010	0.022
$\hat{\sigma}_2(\times 10^{-4})$	3.876	3.785	0.415	2.510	14.370
$\hat{\sigma}_e$	4.113	0.136	3.853	4.113	4.386
$\hat{\rho}_{12}$	-0.464	0.313	-0.931	-0.504	0.217
$\hat{\rho}_{13}$	0.210	0.359	-0.500	0.221	0.863
$\hat{\rho}_{23}$	-0.135	0.458	-0.897	-0.154	0.721
$\hat{\rho}_r$	0.977	0.008	0.960	0.977	0.990
$\hat{\rho}_c$	0.990	0.007	0.973	0.992	0.998
$\hat{\nu}$	5.40	0.783	4.122	5.318	7.140

Table 5: Summary statistics of the posterior samples from model 4.

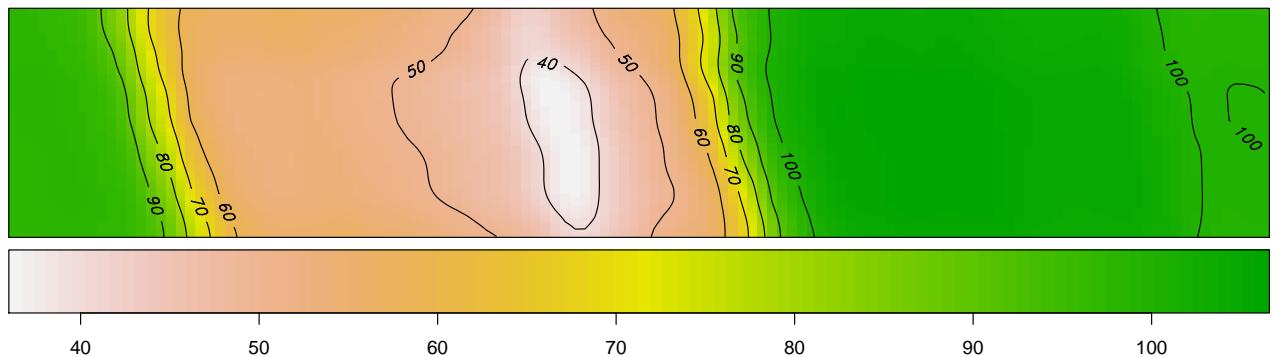
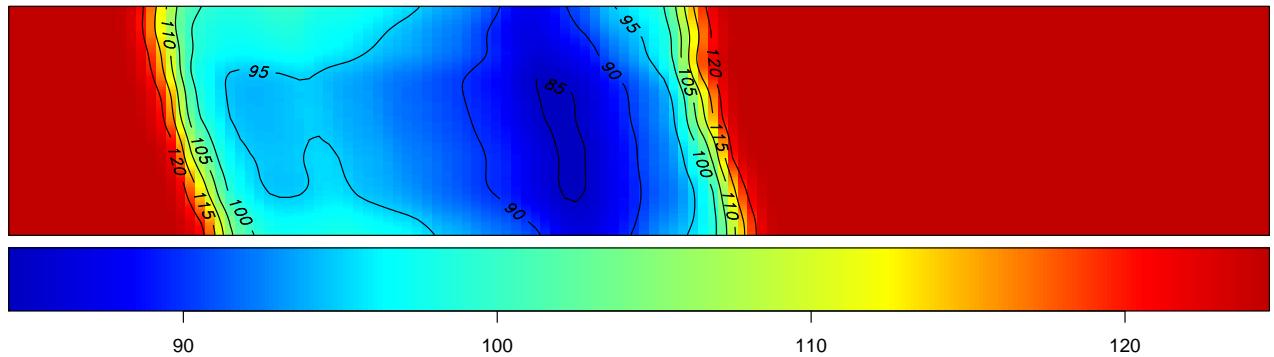


Figure 9: Optimal nitrogen rates (top) and estimated yield with the rates (bottom).

348 that the optimal treatments at pairwise plots are distinguishing, even though the adjusted- $p$  values are large. On the  
 349 contrary, the proposed model with Bayesian approach uses all data with the spatial variance matrix on the entire field  
 350 to estimate the quarry plot. The nearer plots contributes more and further plots contributes less in the inference. So  
 351 it has “lower resolution”, which can be seen from Figure 9, that the the difference of optimal nitrogen rates in West  
 352 and East ares are not significant. Additionally, the reliability of the Bayesian approach is affected by the priors and  
 353 the model itself, whereas the influence of the priors is “washed out” if the model is good enough. Additionally, GWR  
 354 is an ad hoc approach for addressing a particular question. The Bayesian approach has more flexibility to extend and  
 355 to broadly applied to other questions.

356 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
Distinguishability	High	Low
Evaluation	$t$ scores and $p$ -values	credible intervals PP check and LOO PIT Pareto $k$ diagnostic Bayesian $R^2$

Table 6: Comparison of GWR and Bayesian approach. Each of these two approach has its own advantages and disadvantages.

## 357 7 Conclusion

358 This paper demonstrate the Bayesian framework to analyse spatial varying on farm trial experiment. Starting  
 359 from the prior selection, we explain the mechanism of Bayesian approach and NUTS sampler. We also present the  
 360 model checking and diagnostic process for post-sampling. Compared with GWR, Bayesian approach does not require  
 361 bandwidth selection, but requires pre-specified priors. The results from Bayesian approach are similar with GWR and  
 362 is able to capture more detailed information on the field. We open our `rstan` function to public, with which one my  
 363 apply to their own data sets.

## 364 Acknowledgement

365 SAGI West gratefully acknowledges the support from the Grains Research and Development Corporation of Aus-  
 366 tralia. The computation in this paper has been performed using the R-packages `rstan`.

## 367 Appendix

### 368 Prior predictive checking

#### 369 Faster Cholesky factor for $AR1(\rho)$

370 The  $AR1(\rho)$  correlation matrix with correlation factor  $\rho$  is defined by  $\rho_{ij} = \rho^{|i-j|}$ . Not well-known but real helpful  
 371 that a simple form of Cholesky factor for the  $AR1(\rho)$  structure is given by [Madar, 2015] that the simple form is

$$l_{ij} = \begin{cases} \rho^{j-1} & j \geq i = 1 \\ \rho^{j-i}\sqrt{1-\rho^2} & j \geq i \geq 2 \end{cases} \quad (25)$$

372 which significantly improve the computation efficiency in `rstan`.

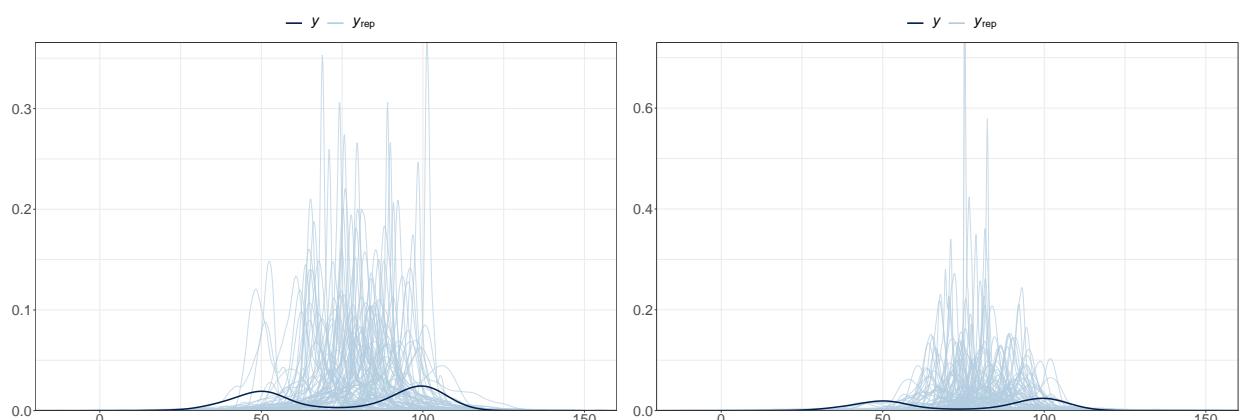
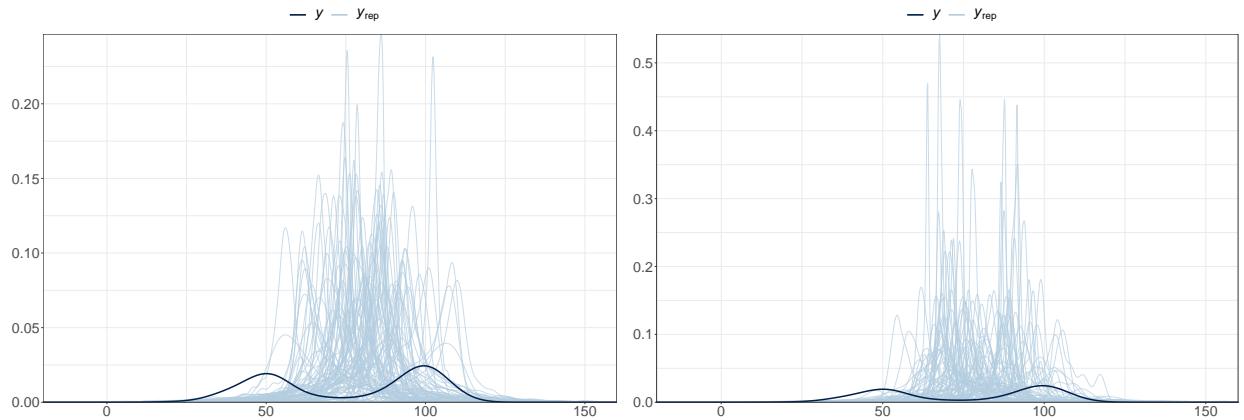


Figure 10: Weakly informative priors checking on four models.

### 373 Faster Kronecker product

374 Suppose  $A = L_A L_A^\top$  and  $B = L_B L_B^\top$  are respectively  $N \times N$  and  $M \times M$  matrices with Cholesky decomposition  
 375  $L_A$  and  $L_B$ , Kronecker product has a good property that

$$C = A \otimes B = (L_A L_A^\top) \otimes (L_B L_B^\top) = (L_A \otimes L_B)(L_A \otimes L_B)^\top, \quad (26)$$

376 where the new matrix  $C$  is  $NM \times NM$ , and the elements of the new matrix  $c_{p,q} = a_{i,j} b_{k,l}$ , where  $p = M(i - 1) + k$   
 377 and  $q = M(j - 1) + l$ . Similarly, the Kronecker product of three matrices, with a  $K \times K$  matrix  $C$ , will be

$$D = A \otimes B \otimes C = (L_A \otimes L_B \otimes L_C)(L_A \otimes L_B \otimes L_C)^\top, \quad (27)$$

378 and the new elements are  $d_{p,q} = a_{i,j} b_{k,l} c_{m,n}$ , where  $p = K(M(i - 1) + k - 1) + m$  and  $q = K(M(j - 1) + l - 1) + n$ .

379 Additionally, with the above properties, we use the following formula to increase the computation efficiency, that

$$(L_A \otimes L_B \otimes L_C)(Z_1 \otimes Z_2) = ((L_A \otimes L_B)Z_1) \otimes (L_C Z_2), \quad (28)$$

380 where  $Z_1$  is a vector with length  $N + M$  and  $Z_2$  is a vector with length  $K$ . This formula is used in `rstan` and  
 381 considerably saves computation time. For interests of other properties of Kronecker product, one can refer to [Zhang  
 382 and Ding, 2013].

## 383 References

- 384 Besag, J. and Higdon, D. (1999). Bayesian analysis of agricultural field experiments. *J Royal Statistical Soc B*,  
 385 61(4):691–746.
- 386 Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. CRC press.
- 387 Bürkner, P.-C., Gabry, J., and Vehtari, A. (2020). Efficient leave-one-out cross-validation for Bayesian non-factorized  
 388 normal and Student-t models. *arXiv:1810.10559 [stat]*.
- 389 Che, X. and Xu, S. (2010). Bayesian data analysis for agricultural experiments. *Can. J. Plant Sci.*, 90(5):575–603.
- 390 Chen, K., O’Leary, R. A., and Evans, F. H. (2019). A simple and parsimonious generalised additive model for  
 391 predicting wheat yield in a decision support tool. *Agricultural Systems*, 173(August 2018):140–150.
- 392 Congdon, P. D. (2019). *Bayesian Hierarchical Models: With Applications Using R, Second Edition*. CRC Press, second  
 393 edition.
- 394 Cook, S., Cock, J., Oberthür, T., and Fisher, M. (2013). On-farm experimentation. *Better Crop. Plant Food*, 97:17–20.
- 395 Dipak Dey and C.R. Rao, editors (2005). *Bayesian Thinking, Modeling and Computation*, volume 25. Elsevier, 1st  
 396 edition.
- 397 Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics letters B*,  
 398 195(2):216–222.
- 399 Edmondson, R. N. (2014). Agridat. *The Journal of Agricultural Science*, 152(1):2–2.
- 400 Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019). Visualization in Bayesian workflow.  
 401 *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2):389–402.
- 402 Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the  
 403 American statistical association*, 85(410):398–409.
- 404 Gelman, A. (2003). A Bayesian Formulation of Exploratory Data Analysis and Goodness-of-fit Testing. *International  
 405 Statistical Review*, 71(2):369–382.

- 406 Gelman, A. (2004). Exploratory Data Analysis for Complex Models. *Journal of Computational and Graphical Statistics*,  
407 13(4):755–779.
- 408 Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*.  
409 Chapman & Hall CRC Texts in Statistical Science. CRC Press, third edition edition.
- 410 Gelman, A. et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by  
411 Browne and Draper). *Bayesian analysis*, 1(3):515–534.
- 412 Gelman, A., Goodrich, B., Gabry, J., and Vehtari, A. (2019). R-squared for bayesian regression models. *American  
413 Statistician*, 73(3):307–309.
- 414 Gelman, A., Simpson, D., and Betancourt, M. (2017). The Prior Can Often Only Be Understood in the Context of  
415 the Likelihood. *Entropy*, 19(10):555.
- 416 Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images.  
417 *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.
- 418 Gilmour, A. R., Cullis, B. R., and Verbyla, A. P. (1997). Accounting for natural and extraneous variation in the  
419 analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics*, 2(3):269–293.
- 420 Griffin, T. W., Dobbins, C. L., Vyn, T. J., Florax, R. J. G. M., and Lowenberg-Deboer, J. M. (2008). Spatial analysis of  
421 yield monitor data: Case studies of on-farm trials and farm management decision making. *Precis. Agric.*, 9:269–283.
- 422 Guan, J., Wang, W., Hu, Y., Wang, M., Tian, T., and Kong, J. (2017). Estimation of genetic parameters for growth  
423 trait of turbot using Bayesian and REML approaches. *Acta Oceanol. Sin.*, 36(6):47–51.
- 424 Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*,  
425 57(1):97–109.
- 426 Hinkelmann, K. (2012). *Design and Analysis of Experiments*, volume 3 of *Wiley Series in Probability and Statistics*.  
427 John Wiley & Sons, Inc.
- 428 Hoffman, M. D. and Gelman, A. (2014). The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian  
429 Monte Carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623.
- 430 Hong, N., White, J. G., Gumpertz, M. L., and Weisz, R. (2005). Spatial analysis of precision agriculture treatments  
431 in randomized complete blocks: Guidelines for covariance model selection. *Agronomy Journal*, 97(4):1082–1096.
- 432 Hurley, T. M., Malzer, G. L., and Kilian, B. (2004). Estimating site-specific nitrogen crop response functions: A  
433 conceptual framework and geostatistical model. *Agronomy Journal*, 96(5):1331–1343.
- 434 Juárez, M. A. and Steel, M. F. J. (2010). Model-Based Clustering of Non-Gaussian Panel Data Based on Skew-  $t$   
435 Distributions. *Journal of Business & Economic Statistics*, 28(1):52–66.
- 436 Kass, R. E. and Natarajan, R. (2006). A default conjugate prior for variance components in generalized linear mixed  
437 models (Comment on Article by Browne and Draper). *Bayesian Analysis*, 1(3):535–542.
- 438 Madar, V. (2015). Direct formulation to Cholesky decomposition of a general nonsingular correlation matrix. *Statistics  
439 and Probability Letters*, 103(1):142–147.
- 440 McElreath, R. (2015). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, volume 122 of  
441 *Chapman and Hall/CRC Texts in Statistical Science Ser.* CRC Press LLC, first edition.
- 442 Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state  
443 calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.
- 444 Monnahan, C. C., Thorson, J. T., and Branch, T. A. (2017). Faster estimation of Bayesian models in ecology using  
445 Hamiltonian Monte Carlo. *Methods in Ecology and Evolution*, 8(3):339–348.

- 446 Montesinos-López, O. A., Montesinos-López, A., Hernández, M. V., Ortiz-Monasterio, I., Pérez-Rodríguez, P., Bur-  
447 gueño, J., and Crossa, J. (2018). Multivariate Bayesian Analysis of On-Farm Trials with Multiple-Trait and Multiple-  
448 Environment Data. *Agron.j.*, 111(6):2658–2669.
- 449 Ng’ombe, J. N., Tembo, M. C., and Masasi, B. (2020). “Are They Aware, and Why?” Bayesian Analysis of Predictors  
450 of Smallholder Farmers’ Awareness of Climate Change and Its Risks to Agriculture. *Agronomy*, 10(3):376.
- 451 Nishio, M. and Arakawa, A. (2019). Performance of Hamiltonian Monte Carlo and No-U-Turn Sampler for estimating  
452 genetic parameters and breeding values. *Genet Sel Evol*, 51(1):73.
- 453 Omer, S. O. and Singh, M. (2017). Comparing Bayesian and Frequentist Approaches for GGE Bi-plot Analysis in  
454 Multi-Environment Trials in Sorghum. *Eur Exp Biol*, 7(6):40.
- 455 Onofri, A., Piepho, H.-P., and Kozak, M. (2019). Analysing censored data in agricultural research: A review with  
456 examples and software tips. *Annals of Applied Biology*, 174(1):3–13.
- 457 Piepho, H.-P., Richter, C., Spilke, J., Hartung, K., and Kunick, A. (2011). Statistical aspects of on-farm experimen-  
458 tation. *Crop & Pasture Science*, 62:721–735.
- 459 Piepho, H.-P., Richter, C., and Williams, E. (2008). Nearest neighbour adjustment and linear variance models in plant  
460 breeding trials. *Biometrical Journal*, 50(2):164–189.
- 461 Rakshit, S., Baddeley, A., Stefanova, K., Reeves, K., Chen, K., Cao, Z., Evans, F., and Gibberd, M. (2020). Novel  
462 approach to the analysis of spatially-varying treatment effects in on-farm experiments. *Field Crops Research*,  
463 255(October 2019):107783.
- 464 Ritter, C., Dicke, D., Weis, M., Oebel, H., Piepho, H. P., Büchse, A., and Gerhards, R. (2008). An on-farm approach  
465 to quantify yield variation and to derive decision rules for site-specific weed management. *Precision Agriculture*.
- 466 Selle, M. L., Steinsland, I., Hickey, J. M., and Gorjanc, G. (2019). Flexible modelling of spatial variation in agricultural  
467 field trials with the R package INLA. *Theoretical and Applied Genetics*, 132(12):3277–3293.
- 468 Sorensen, D. and Gianola, D. (2007). *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer  
469 Science & Business Media.
- 470 Sorensen, T., Hohenstein, S., and Vasishth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for  
471 psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology*, 12(3):175–200.
- 472 Theobald, C. M., Talbot, M., and Nabugoomu, F. (2002). A bayesian approach to regional and local-area prediction  
473 from crop variety trials. *JABES*, 7(3):403–419.
- 474 Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Econ. Geogr.*, 46:234.
- 475 Troyer, A. F. and Wellin, E. J. (2009). Heterosis decreasing in hybrids: Yield test inbreds. *Crop Science*, 49(6):1969–  
476 1976.
- 477 Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation  
478 and WAIC. *Stat Comput*, 27(5):1413–1432.
- 479 Weiss, R. (2016). Pediatric Pain, Predictive Inference, and Sensitivity Analysis:. *Evaluation Review*.
- 480 Yan, W., Hunt, L. A., Johnson, P., Stewart, G., and Lu, X. (2002). On-farm strip trials vs. replicated performance  
481 trials for cultivar evaluation. *Crop Science*, 42(2):385–392.
- 482 Zhang, H. and Ding, F. (2013). On the Kronecker Products and Their Applications. *Journal of Applied Mathematics*,  
483 2013:1–8.