Optimal design for on-farm strip trials — systematic or randomised?

Zhanglong Cao[1], Jordan Brown[1], Mark Gibberd[1], Julia Easton[1], and Suman Rakshit[1,2]

[1]Curtin Biometry and Agricultural Data Analytics, Centre for Crop and Disease Management, Curtin University, Perth, Australia

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

Abstract

CONTEXT OR PROBLEM: Randomised designs are often preferred over systematic designs by agronomists and biometricians. However, for on-farm trials, the choice may depend on the objective of the experiments. If the purpose is to create a prescription map of a continuous input for each plot of a grid covering a large strip trial, a systematic design may be a better choice, but it attracts less discussion and attention.

OBJECTIVE OR RESEARCH QUESTION: This study aims to evaluate the performance of systematic designs with geographically weighted regression (GWR) models in addressing spatial variation and estimating continuous treatment effects in large strip trials through numeric simulations.

METHODS: A Bayesian hierarchical model with spatially correlated random parameters is utilised to generate simulated data for various scenarios of large strip on-farm trials. The study employs GWR models to analyse the simulated data for two assumptions: a linear response and a quadratic response to the treatment effects.

RESULTS: With the assumption of linear response, the difference between a systematic design and a randomised design is not significant regardless of the presence of spatial variation. However, with the assumption of quadratic response, a systematic design is superior to a randomised design concerning the lower mean squared errors (MSE) by GWR.

CONCLUSIONS: The findings highlight the superiority of systematic designs for producing smooth maps of optimal input levels and the impact of spatial variation on the performance of quadratic models in strip trials. Additionally, the study recommends fixed bandwidths based on experimental designs for GWR models. For a large strip trial to create a varying treatment map, a systemic design should be used as it has more flexibility in post-experiment statistical modelling.

IMPLICATIONS OR SIGNIFICANCE: The findings offer practical recommendations for designing large strip trials. By drawing attention to the considerations, especially regarding the experiment's purpose, this research contributes valuable insights for improving the efficacy and planning of future large strip trials.

Keywords: yield map, optimal treatment, spatially varying coefficients.

1 Introduction

The principles of randomisation were first expounded in 1925 by Fisher (1934), who analysed a few systematically arranged experiments and pointed out that randomisation can provide valid tests of significance subject to appropriate restrictions, such as experimental units arranged in blocks or in rows and columns of a Latin square (Verdooren 2020). Traditionally, small-plot trials for agriculture are designed to accentuate treatment effects with the completely randomised design (CRD) as the most straight forward, least restrictive experimental design. More complex designs, such as a randomised complete block design (RCBD), a split-plot design, a strip-plot design and a Latin square design, are also widely used in agricultural experiments (Petersen 1994). Following these principles, randomised designs are routinely used for on-farm strip trials, while systematic designs are rarely used.

OFE enables farmers the flexibility to implement large-scale experiments in order to test management practices on their farms (Evans et al. 2020). The goal of OFE is to help farmers better understand uncertainties and leverage their existing strengths in managing translational and structural uncertainty (Cook et al. 2013). In the situation that the goal is to compare yield responses between management classes or to select individuals with the best performance as new market varieties, a randomised design is superior to a systematic design (Pringle et al. 2004; Selle et al. 2019).

While randomisation is often considered a crucial prerequisite for obtaining valid statistical inferences (Piepho et al. 2013), this is not always the case when the goal of on-farm experiments (OFE) shifts from the conventional analysis. In the application of precision agriculture (PA), the variable rate applicators (VRA) require a prescription map of the experiment before the start of the operation (Pringle et al. 2004). Therefore, in this scenario, the goal of OFE becomes obtaining a smooth map showing the optimal level of a controllable input, such as nitrogen rates, across a grid made of rows and columns covering the whole field. For this objective, Piepho et al. (2011) stresses that only a single level of treatments can be directly observed at any one point on the grid and the response for other levels at the same grid must be interpolated. If a randomised design is conducted, the interpolation distances to locations with treatment levels of interest will vary throughout the field. Such heterogeneous distances increase the uncertainty in the analysis and reduce the efficiency of local prediction. As a result, a systematic design is preferable to a randomised design in this scenario. Unfortunately, this perspective has often been overlooked by researchers, leading to the widespread use of randomised designs.

Analysing a systematic design for the creation of an optimal treatment map for on-farm experiments (OFE) is a statistically challenging task. The truly localised estimation at each point on the grid is unknown, and the optimum treatment response varies continuously across the field. Cao et al. (2022) implemented a Bayesian approach using spatially correlated random parameters for large systematic OFE strip trials, assuming a quadratic response model with both global and local spatially varying components. However, the Bayesian approach is time-consuming and requires preliminary knowledge of Bayesian statistics for farmers and agronomists. Alternatively, Rakshit et al. (2020) adopt a local regression approach, called geographically weighted regression (GWR), to obtain spatially-varying estimates of treatment effects for OFE. Additionally, Evans et al. (2020) conclude through simulation studies that GWR is a simple method for OFE data analysis and is capable of accurately separating yield variation that is not due to applied treatment from yield response due to treatment. However, their study was limited by the use of a randomised design and the assumption of a linear response to fertiliser treatment. Alesso et al. (2021) simulated corn yield with four nitrogen levels assigned systematically and randomly in the chessboard designs and fitted the true coefficients using GWR. They concluded that systematic designs achieved the best results in most cases. However, in these simulation studies, the use of chessboard design presents a problem: harvesters can smooth over abrupt treatment changes between plots, potentially leading to misleading results unless there are constraints on the field (Pringle et al. 2004). Additionally, the quadratic or plateau feature was not considered a factor in the simulations.

Piepho and Edmondson (2018) demonstrate an example that a linear model is lacking in fitting to the sugar beet data (Petersen 1994). Glynn (2007) show that many curves exist beyond a linear trend for nutrient-response relationships. The curve depends on the current availability of other macro and micronutrients in the soil (Marschner 2011), meaning that a linear relationship is unlikely to be consistent across a large trial. For this reason, it is important to consider models with degrees higher than 1, which means a quadratic model was found to be suitable for relationships (Piepho and Edmondson 2018; Liben et al. 2019).

Our study uses simulation examples to demonstrate that randomisation is not essential for large strip trials. For the purpose of obtaining a treatment map, a systematic design is superior to a randomised design, subject to appropriate restrictions. We also test the power of GWR to determine if it can successfully estimate spatially varying treatment effects for both linear and quadratic responses. Our results show that the optimal bandwidth found by AICc is not the best for GWR. Instead, a fixed bandwidth based on the experimental design is recommended.

The structure of the paper is organised as follows: In Section 2, we describe the statistical model for generating simulated data, which has spatially varying coefficients of treatments, and the GWR model to fit OFE data. In Section 3, we generate simulated data for the combination of the following scenarios: randomised and systematic designs; linear and quadratic response; low and high coefficient correlations; spatial variation among grids is identity (no spatial trend), AR1 $\otimes$ AR1 and Matérn form. Finally, in Sections 4 and 5, we illustrate the results and discuss their importance with respect to OFE, and how the findings should influence future trial designs.

2 Methods

This section describes the statistical model used in the simulation study. It outlines the basic model (Subsection 2.1) followed by the methodology for the spatially correlated treatment parameters (Subsection 2.2), and finally GWR (Subsection 2.3).

2.1 Basic statistical model

In a conventional agricultural study, a field experiment can be considered as a rectangular matrix consisting of r rows and c columns, where the total number of plots in the experiment is $n = r \times c$. The notation $s_i \in \mathcal{R}^2, i = 1, \ldots, n$ is a two-cell vector of the Cartesian coordination of the plot centroids, located on a regular grid (Zimmerman and Harville 1991). Hence, $y(s_i)$ denotes the dependent variable at a query location/grid i.

With the assumption that $\boldsymbol{Y}$ is the vector of the plot data ordered as rows nested within ranges (columns), then the matrix notation of the model 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 error vector. We further assume that $\boldsymbol{u}$ and $\boldsymbol{e}$ are pairwise independent and that their joint distribution is

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

2.2 Spatially correlated treatment parameters

Cao et al. (2022) implemented the Bayesian hierarchical model with spatially correlated random parameters on large OFE strip trials. Here, for the simulation study, we use the same model to generate the synthetic data.

With the same notation given in the reference and previous section, the underlying model is represented as

$$\begin{aligned} y(s_i) \mid \boldsymbol{u}_i, \theta_u, \sigma_e &\sim \mathcal{N}\left(\sum_{m=1}^{l} b_m x_m(s_i) + \sum_{j=1}^{k} u_j(s_i) z_j(s_i),\ e(s_i) \right) \\ \boldsymbol{u}_i \mid \theta_u &\sim \mathcal{N}(0, V_u(\theta_u)) \\ e(s_i) \mid \sigma_e &\sim \mathcal{N}(0, \sigma_e^2) \end{aligned} \tag{3}$$

where: $x_1, \ldots, x_l$ denote l fixed effects and $z_1, \ldots, z_k$ denote k random effects; b_m and $u_j(s)$ are the coefficients for the fixed and random terms, respectively; $\boldsymbol{u}_i$ is a vector of all random effects at grid $s_i \in \mathcal{S}$, $i = 1, \ldots, n$; θ_u is a set of parameters of the covariance matrix V_u; and σ_e is a positive latent variable.

In model (3), the structure of the covariance matrix $V_u(\theta_u)$ of $\boldsymbol{u}_i$ can be either diagonal, which implies the treatments at grid i are independent, or in general form, which means a correlation exists. McElreath (2015) suggest that the covariance of $\boldsymbol{u}_i$ can be $V_u = B(\sigma_u) R_u B(\sigma_u)$, where $B(\sigma_u)$ denotes the diagonal matrix of elements σ_{u_j}, $j = 1, \ldots, k$. For the matrix R_u with correlation coefficients, we specify the Lewandowski-Kurowicka-Joe (LKJ) distribution (Lewandowski et al. 2009), which is given by

$$R_u \sim \mathrm{LKJcorr}(\epsilon), \tag{4}$$

where $\mathrm{LKJcorr}(\epsilon)$ is a positive definite correlation matrix sampled from the LKJ distribution controlled by a positive parameter ϵ. As ϵ increases, a high correlation among parameters becomes less likely.

Furthermore, by incorporating a spatial correlation structure V_s, the complete form of the covariance matrix of $\boldsymbol{u}$ is presented as

$$\Sigma_u = V_s \otimes V_u. \tag{5}$$

In fact, V_s is the covariance matrix of all grids on the field. For example, if $V_s = I_{n\times n}$ is an identity matrix, each grid is independent even though the treatments within each grid are correlated. However, the correlation among grids is ubiquitous. Hence, we introduce a simple spatial covariance matrix such that

$$V_s = \mathrm{AR1}(\rho_c) \otimes \mathrm{AR1}(\rho_r), \tag{6}$$

where AR1 is the separable first-order auto-regressive model in the column and row direction which is controlled by the correlation parameters ρ_c and ρ_r, respectively (Butler et al. 2017).

Besides the above AR1 $\otimes$ AR1 covariance, the Matérn class covariance

$$V_s(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{d}{r}\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}\frac{d}{r}\right) \tag{7}$$

is also used in spatial analysis (Cressie and Huang 1999) and in capturing spatial variation in OFE (Selle et al. 2019). Here, d is the space lag or distance; r is a non-negative scaling parameter; $\nu > 0$ is a smoothness parameter determining the mean-square differentiability of the field; σ_d^2 is the variance of the process; Γ is the Gamma function; and K_ν is the modified Bessel function of the second kind. If $\nu = r + \frac{1}{2}$, then the Matérn covariance can be expressed as a product of an exponential and a polynomial of order r (Pandit and Infield 2019; Abramowitz 1974), which simplifies the model and the computation process. The Matérn class $\nu = \frac{3}{2}$ and $\nu = \frac{5}{2}$ are common in application.

Model (3) has the advantage of reproducibility in the simulation study and robustness in estimation. It is possible even though only a single treatment is directly observed at each plot, and the responses of the other levels need to be estimated by interpolation because the spatial model allows using information from neighbouring plots with other treatments (Panten et al. 2010; Piepho et al. 2011).

2.3 Geographically weighted regression (GWR)

Geographically weighted regression (GWR) is a local regression approach and is adapted to obtain spatially-varying estimates of treatment effects for OFE (Rakshit et al. 2020). It is seen as a locally weighted regression method that operates by assigning a weight to each observation i depending on its distance from the query grid on the field (Páez et al. 2002).

The underlying template model for the GWR, according to Leung et al. (2000), is given by

$$y(s_i) = \beta_0 + \sum_{j=1}^{k} \beta_j z_j(s_i) + \varepsilon_i, \tag{8}$$

where $\boldsymbol{\beta}$ and $\varepsilon \sim \mathcal{N}(0, \tau^2)$ are the model parameters for the k levels treatments and error terms, respectively, at grid $i,\ i = 1, \ldots, n$. The estimator of this model is given by the geographically weighted expression in

$$\hat{\boldsymbol{\beta}}(s) = \left(Z^\top W(s) Z\right)^{-1} Z^\top W(s) Y, \tag{9}$$

where $W(s)$ is an $n \times n$ diagonal matrix of weights, $\boldsymbol{w}$. Then it can be found by maximising the local log-likelihood

$$\log L(s;\beta) = -\frac{1}{2\tau^2} \sum_{i=1}^{n} K(s, s_i) \left(y(s_i) - \beta_0 - \sum_{j=1}^{k} \beta_j z_j(s_i) \right)^2 \tag{10}$$

with a given kernel function $K(\cdot,\cdot)$, such as Gaussian, exponential, bi-square or tri-cube (Gollini et al. 2015). In the simulation study, we use a Gaussian kernel. In fact, the kernel is not the crucial factor in the GWR model fitting on OFE data. By contrast, the factor bandwidth has a higher influence on the estimation.

The optimal bandwidth of a GWR model is usually given by the lowest AICc where

$$\text{AICc} = 2n\log(\tau^2) + n\log(2\pi) + n\frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)}, \tag{11}$$

and S is the matrix with the i-th row given by $Z_i\left(Z^\top W(s_i)Z\right)^{-1} Z^\top W(s_i)$ (Evans et al. 2020). Alternatively, as suggested by Rakshit et al. (2020), it can be based on the experimental design such that the local regressions capture data covering the full range of treatments.

The GWR model in this paper is implemented with the `R`-package `GWmodel` (Lu et al. 2014; Gollini et al. 2015).

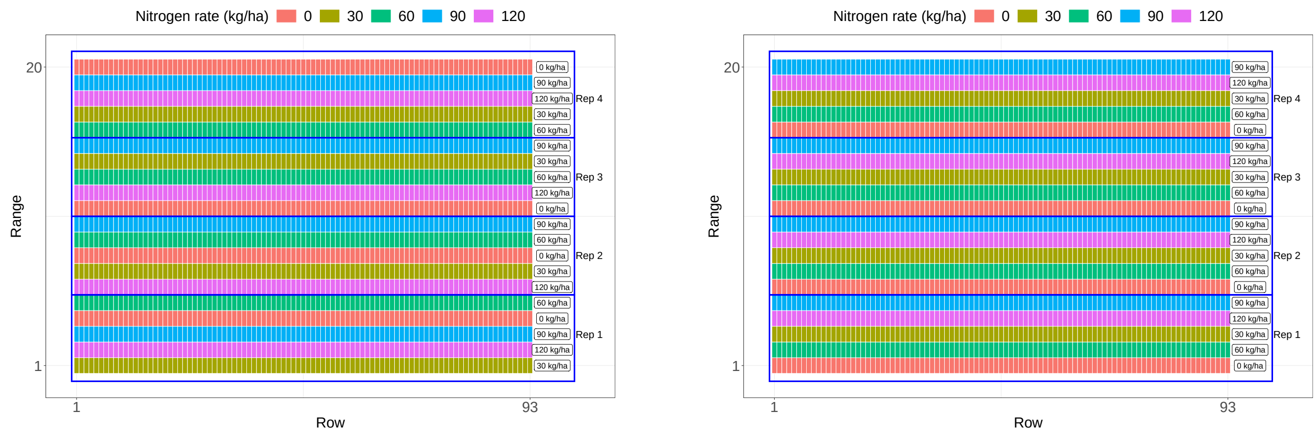
3 Simulation study

To study the effect of randomised designs and systematic designs and to evaluate the power of GWR, we simulate spatially correlated large strip trials. The advantage of the simulation study is that the actual coefficients of the models are known, so adverse effects of model misspecification can be ruled out (Piepho et al. 2013).

We investigate the combination of the following factors: types of design with two levels: randomised and systematic; response relationship with two levels: linear and quadratic; correlation of coefficients with two levels: low and high; spatial variation with three levels: identity (no spatial trend), AR1 $\otimes$ AR1 and Matérn form; GWR bandwidth with three levels: 5, 9 and optimum given by AICc. The fixed bandwidth 5 in the simulation study covers all treatment levels (five nitrogen levels) in a systematic design, where the information is adequate in the inference of a quadratic curve. Similarly, a fixed bandwidth of 9 is essential to cover all possible treatment levels in a randomised design. This necessity stems from situations where treatments are randomly distributed across strips. In particular, if identical treatments are positioned at the far left edge of the first replicate block and the far right edge of the second replicate block, a bandwidth smaller than 9 would lead to the GWR model capturing only the treatments between these boundaries, thereby missing the treatment level at the extremes.

With model (3), we generate the yield as a response to nitrogen treatment in two scenarios that cover whether the yield has a linear or quadratic relationship with the nitrogen rate. The nitrogen rates are treated as continuous observations with five levels: 0, 35, 75, 105 and 140 kg/ha. A strip plot structure was used to allocate the five nitrogen levels, where each level is assigned to one strip. Then we assume that the experimental design of the trial consists of four replicates, each containing 5 ranges (columns) by 93 rows. The nitrogen rates are allocated randomly or systematically in the whole field. The overall layout of the trial is 20 ranges by 93 rows. Examples of a randomly and systematically allocated treatment map are presented in Figure 1.

With model (3), the nitrogen rates are treated as continuous observations with five levels. For a linear relationship, we assume that the true global intercept is $b_0 = 65$ and the slope is $b_1 = 0.05$. The parameters are chosen according to the estimates by Rakshit et al. (2020) and Cao et al. (2022) on the Las Rosas corn field data, which is initially provided by Anselin et al. (2004) and embedded in the `R`-package `agridat` (White and Evert 2008). The unit of yield is quintals per hectare. The variances of $\boldsymbol{u}_i$ are $\sigma_{u_0} = 5$ and $\sigma_{u_1} = 0.01$. For the AR1 $\otimes$ AR1 covariance matrix in (6), the two correlation parameters for column and row are $\rho_c = 0.15$ and $\rho_r = 0.5$. We have assumed a higher correlation in the rows due to the fact that the crop is traditionally sown and harvested along ranges where the correlation is higher in the perpendicular direction than the travel direction (Marchant et al. 2019). For the Matérn covariance



(a) Treatments are randomly allocated into large strips in each replicate block.

(b) Treatments are systematically allocated into large strips in each replicate block.

Figure 1: The nitrogen treatments with five levels (0, 35, 70, 105 and 140 kg/ha) randomly (1a) and systematically (1b) allocated into strips in each replicate block.

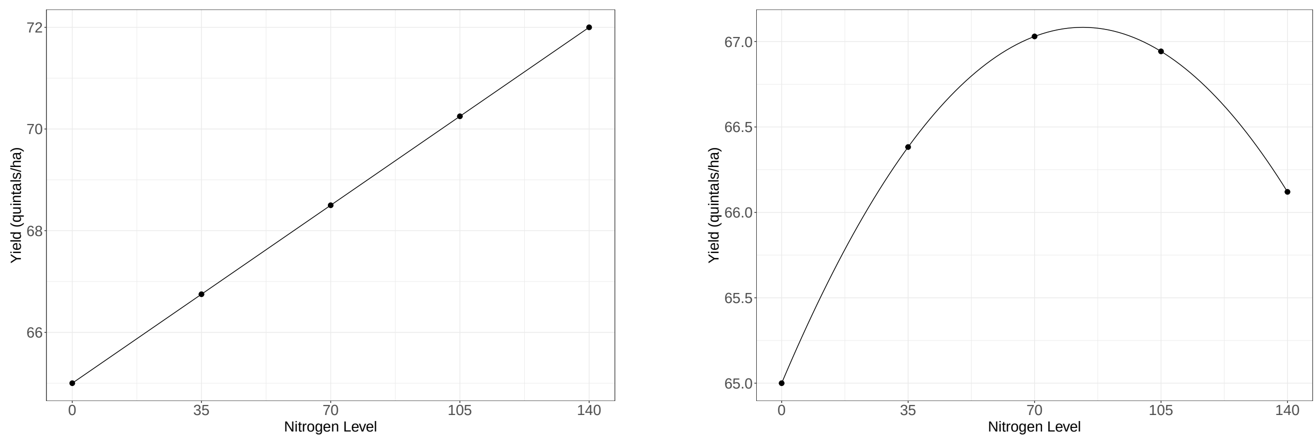
matrix (7), we set $\sigma_d^2 = 1$, $r = 1$ and $\nu = \frac{3}{2}$. After drawing samples of $\boldsymbol{u}$ from $\mathcal{N}(0, \Sigma_u)$, the true spatially varying coefficients are $\boldsymbol{\beta}_0 = b_0 + \boldsymbol{u}_0$ and $\boldsymbol{\beta}_1 = b_1 + \boldsymbol{u}_1$.

Similarly, for the quadratic relationship, we have the true global intercept $b_0 = 65$, coefficients $b_1 = 0.05$ and $b_2 = -0.0003$, making the curve concave down. We keep $\sigma_{u_0} = 5$, $\sigma_{u_1} = 0.01$ and add $\sigma_{u_2} = 0.0001$ because the true values are small. The rest of the parameters remain unchanged. Therefore, the true spatially varying coefficients are $\boldsymbol{\beta}_0 = b_0 + \boldsymbol{u}_0$, $\boldsymbol{\beta}_1 = b_1 + \boldsymbol{u}_1$ and $\boldsymbol{\beta}_2 = b_2 + \boldsymbol{u}_2$.

To summarise, with the true spatially varying coefficients of the treatments, the simulated yield is obtained by

$$\begin{cases} \text{Linear} & y_i = b_0 + u_{0i} + (b_1 + u_{1i})N_i + e_i \\ \text{Quadratic} & y_i = b_0 + u_{0i} + (b_1 + u_{1i})N_i + (b_2 + u_{2i})N_i^2 + e_i \end{cases}, \tag{12}$$

where N_i is the nitrogen rate, $e_i \sim \mathcal{N}(0, 1)$ is the error term at grid i, $i = 1, \ldots, n$. Figure 2 illustrates how these curves behave for the linear and quadratic relationships.



(a) Linear relationship between crop yield and nitrogen levels, where $y = 65 + 0.05x$.

(b) Quadratic relationship between crop yield and nitrogen levels, where $y = 65 + 0.05x - 0.003x^2$.

Figure 2: Noise-free linear and quadratic relationships between crop yield and nitrogen levels.

The purpose of the simulation study is to test the effect of different types of designs in coefficients estimation by

GWR. Identical coefficients were used in one comparison process for two types of designs, and the yield reflects the effect of nitrogen rates.

4 Results

Running the simulation 100 times, we assessed the performance of the randomised and systematic designs for linear and quadratic responses. In subsection 4.1 the mean squared errors are compared between the two designs for different bandwidths, parameter correlations and spatial covariance matrices. Subsection 4.2 uses an analysis of variance (ANOVA) test to explore the significance of the factors in the simulation, while subsection 4.3 states the performance of bandwidth selection using AICc.

4.1 Mean squared error

We evaluated the true mean squared error (MSE) of the estimated coefficient differences. This was calculated by the difference of true coefficients, $\boldsymbol{\beta} = b + \boldsymbol{u}$, and estimated spatially varying coefficients, $\hat{\boldsymbol{\beta}} = \hat{b} + \hat{\boldsymbol{u}}$, for each grid, and then squared the discrepancy and averaged across the field for comparison. The results are shown in Figures 3 and 4 where "NS" stands for no spatial variation ($V_s = I_{n\times n}$), "AR1" is for AR1(0.15) $\otimes$ AR1(0.5) and "Matern" is the Matérn covariance with $\nu = \frac{3}{2}$. Since the MSEs of β_0, β_1, and β_2 are small, we take the natural logarithm for better visualisation.

With the assumption of a linear response, both randomised and systematic designs perform similarly, meaning GWR is able to partition the local varying intercept and treatment coefficient. Figure 3 shows that if the response is linear, the MSEs of $\hat{\beta}_0$ and $\hat{\beta}_1$ estimated by GWR, for all bandwidths, is not distinguishing between randomised and systematic designs. This is true regardless of the type of spatial covariance matrix when the correlation within grids is small ($\epsilon = 1$), or high ($\epsilon = 0.1$) as is shown in Figure 6. The GWR with bandwidth selected by AICc has a smaller MSE for its coefficients than the model with a fixed bandwidth (Figure 3).

However, if the assumption is a quadratic response, Figures 4 and 7 show that the GWR with a fixed bandwidth of a systematic design outperforms a randomised design if the spatial correlation is taken into account. With the optimal bandwidth, GWR successfully estimates the global intercepts β_0 but failed in estimating local varying coefficients β_1 and β_2, where the MSE is relatively larger than with fixed bandwidth. However, if we only compare across two types of designs, it still proves that GWR is robust to fit a systematic design rather than a randomised design if the assumption is a quadratic response, regardless of the intensity of the correlation within grids.

MSE analyses revealed that the choice of bandwidth significantly influenced the relative performance based on both the spatial covariance matrix and the coefficient characteristics. In scenarios without spatial variation for β_1 and β_2, the AICc-selected bandwidth demonstrated the lowest MSE, followed by fixed bandwidths of 9 and 5, respectively. However, when the spatial variation was present (utilising either AR1$\otimes$AR1 or Matérn covariance structures), the bandwidth of 9 consistently produced the most accurate estimations for β_1 and β_2, outperforming the bandwidths of 5 and the one chosen by AICc. Notably, for the intercept (β_0), significant changes in MSE amongst the bandwidths were observed only in the context of the Matérn spatial covariance. Here, the bandwidth selected by AICc resulted in the smallest MSE, highlighting its efficacy when dealing with spatially correlated residuals following the Matérn

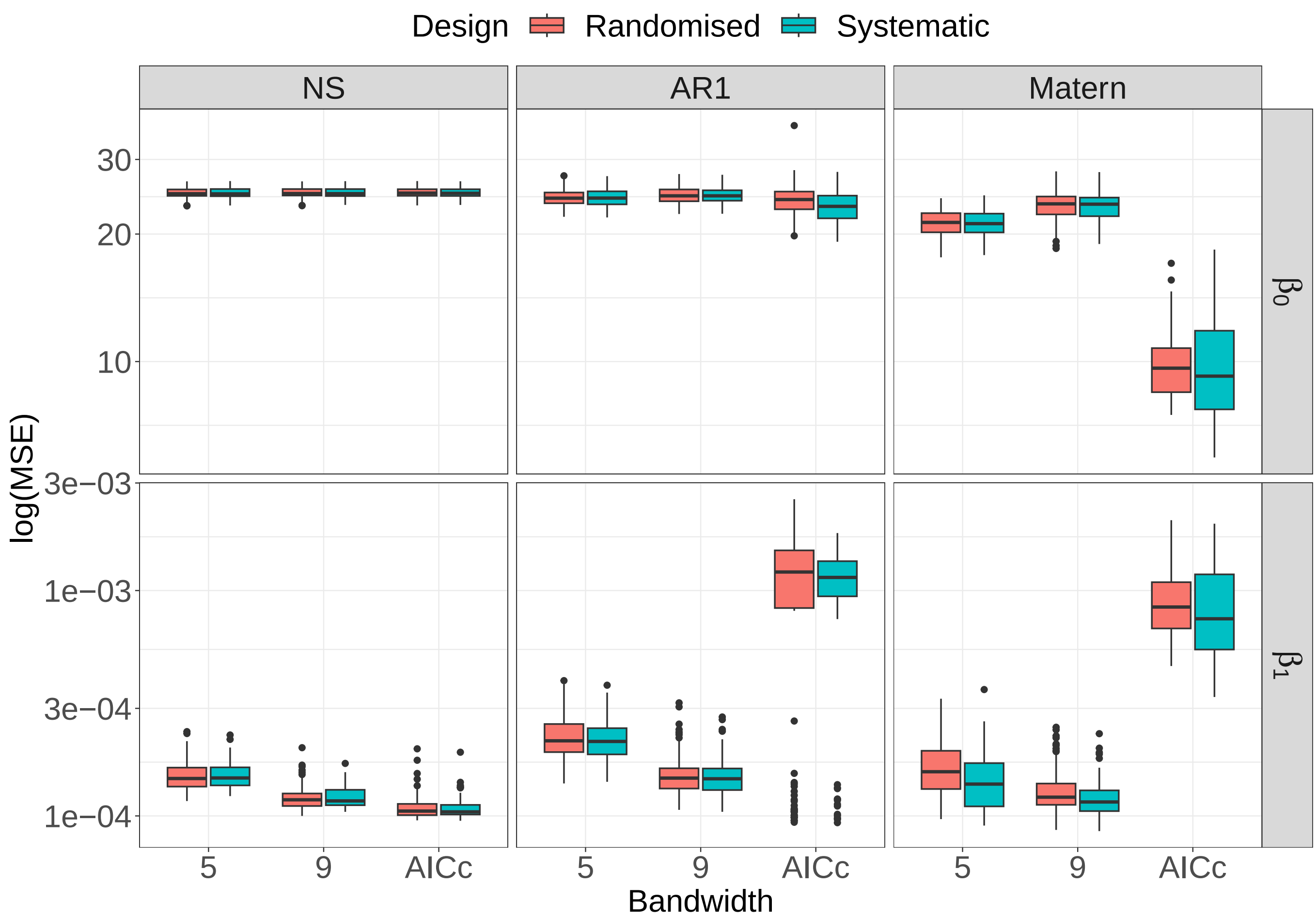


Figure 3: Boxplots of the logarithm of MSE for $\hat{\beta}_0$ and $\hat{\beta}_1$ in GWR models using different bandwidths for the simulated data with a linear response. The simulated data had different spatial covariance matrices (NS, AR1⊗AR1 and Matérn) and a low correlation between the parameters ($\epsilon = 1$).

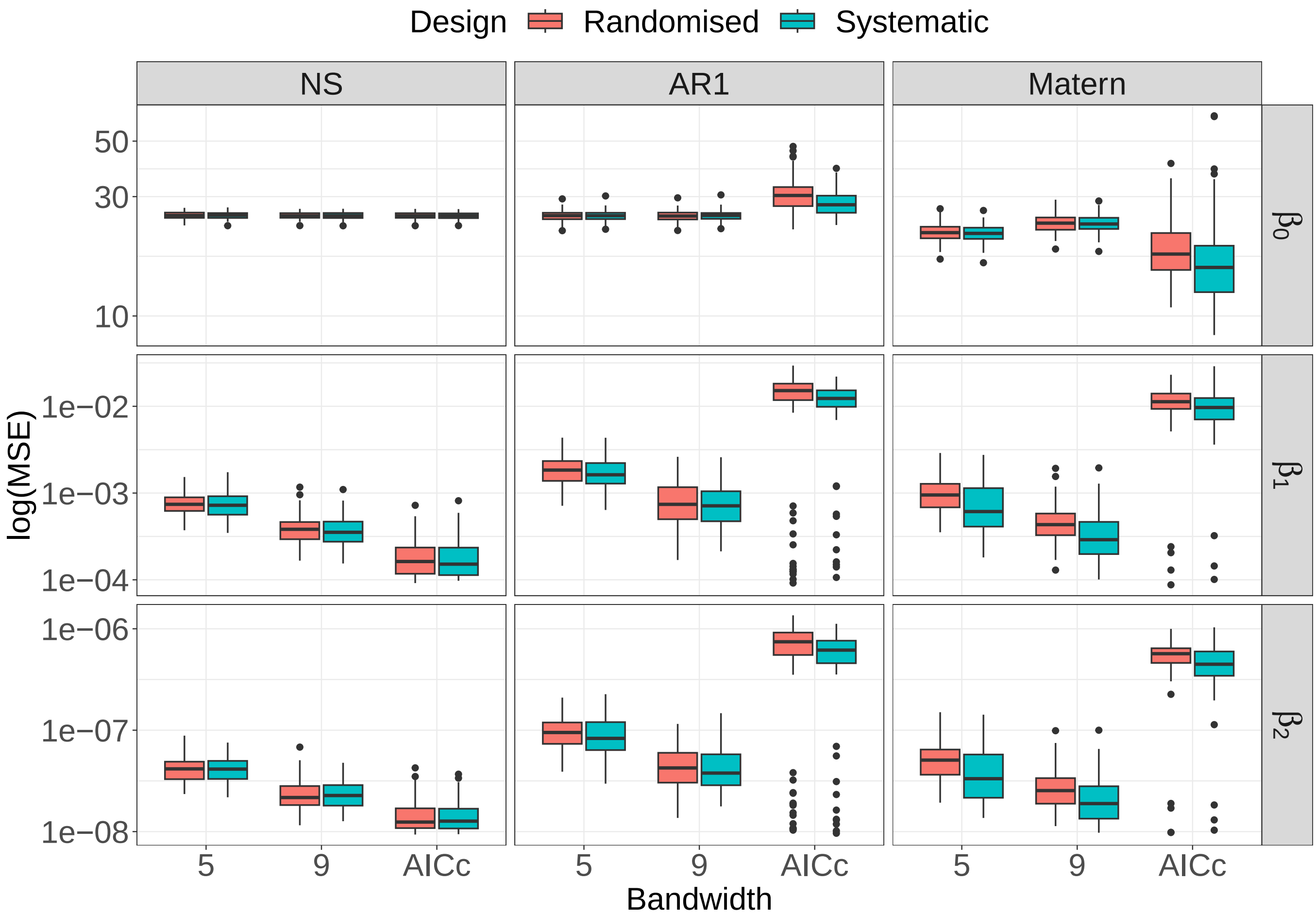


Figure 4: Boxplots of the logarithm of MSE for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ in GWR models using different bandwidths for the simulated data with a quadratic response. The simulated data had different spatial covariance matrices (NS, AR1$\otimes$AR1 and Matérn) and a low correlation amongst the parameters ($\epsilon = 1$).

covariance structure.

Tables 1 and 2 are the median MSEs of the linear response for two scenarios: the correlation is low and the correlation is high.

Table 1: Median MSE of GWR coefficient estimates for a linear response when the correlation between the parameters is low ($\epsilon = 1$).

		Randomised			Systematic		
	Coefficient	5	9	AICc	5	9	AICc
NS	$\hat{\beta}_0$	24.903	24.924	24.983	24.886†	24.911	24.965
	$\hat{\beta}_1(\times 10^3)$	0.147	0.118	0.105	0.147	0.117	0.104†
AR1	$\hat{\beta}_0$	24.308	24.617	24.126	24.319	24.617	23.246†
	$\hat{\beta}_1(\times 10^3)$	0.215	0.147	1.208	0.214	0.146†	1.143
Matérn	$\hat{\beta}_0$	21.303	23.566	9.647	21.164	23.526	9.239†
	$\hat{\beta}_1(\times 10^3)$	0.157	0.121	0.845	0.138	0.115†	0.749

† Indicates the smallest MSE for the row.

Table 2: Median MSE of GWR coefficient estimates of linear response when the correlation between the parameters is high ($\epsilon = 0.1$).

		Randomised			Systematic		
	Coefficient	5	9	AICc	5	9	AICc
NS	$\hat{\beta}_0$	24.961	24.991	24.958	24.934†	24.992	25.020
	$\hat{\beta}_1(\times 10^3)$	0.144	0.115	0.103†	0.145	0.116	0.104
AR1⊗AR1	$\hat{\beta}_0$	24.234	24.518	23.696	24.171	24.497	23.421†
	$\hat{\beta}_1(\times 10^3)$	0.216	0.144†	1.024	0.211	0.146	1.027
Matérn	$\hat{\beta}_0$	20.882	22.811	9.823	20.770	22.789	9.257†
	$\hat{\beta}_1(\times 10^3)$	0.152	0.120	0.939	0.140	0.112†	0.815

† Indicates the smallest MSE for the row.

Tables 3 and 4 are the median MSE of the quadratic response for two scenarios: correlation is low and correlation is high. Despite the intensity of the correlation, the data fitted by GWR from systematic designs are superior to randomised designs, in terms of having a lower MSE.

Table 3: Median MSE of GWR coefficient estimates of quadratic response when the correlation amongst the parameters is low ($\epsilon = 1$).

		Randomised			Systematic		
	Coefficient	5	9	AICc	5	9	AICc
NS	$\hat{\beta}_0$	25.218	25.150	25.152	25.263	25.179	25.138†
	$\hat{\beta}_1(\times 10^4)$	7.417	3.823	1.625	7.233	3.529	1.516†
	$\hat{\beta}_2(\times 10^8)$	4.157	2.168	1.242†	4.135	2.269	1.269
AR1⊗AR1	$\hat{\beta}_0$	25.185	25.092†	30.315	25.166	25.230	27.831
	$\hat{\beta}_1(\times 10^4)$	18.395	7.414	151.595	16.243	7.124†	123.181
	$\hat{\beta}_2(\times 10^8)$	9.491	4.244	74.420	8.305	3.777†	61.619
Matérn	$\hat{\beta}_0$	21.532	23.502	17.680	21.384	23.319	15.631†
	$\hat{\beta}_1(\times 10^4)$	9.502	4.326	112.914	6.121	2.901†	96.829
	$\hat{\beta}_2(\times 10^8)$	5.071	2.537	56.789	3.324	1.889†	44.707

† Indicates the smallest MSE for the row.

Table 4: Median MSE of GWR coefficient estimates of quadratic response when the correlation amongst the parameters is high ($\epsilon = 0.1$).

		Randomised			Systematic		
$\epsilon = 0.1$	Coefficients	5	9	AICc	5	9	AICc
NS	$\hat{\beta}_0$	25.075	25.067	25.015	25.082	25.060	25.012†
	$\hat{\beta}_1(\times 10^4)$	6.683	3.466	1.478†	7.353	3.472	1.506
	$\hat{\beta}_2(\times 10^8)$	3.779	2.101	1.222†	3.806	2.124	1.284
AR1⊗AR1	$\hat{\beta}_0$	25.103	25.223	29.266	25.033	25.032†	27.378
	$\hat{\beta}_1(\times 10^4)$	16.260	6.845	130.335	16.228	6.314†	112.599
	$\hat{\beta}_2(\times 10^8)$	8.488	3.931	61.765	7.915	3.533†	54.866
Matérn	$\hat{\beta}_0$	21.780	23.622	18.832	21.409	23.296	15.728†
	$\hat{\beta}_1(\times 10^4)$	11.367	5.085	122.638	6.205	2.892†	88.256
	$\hat{\beta}_2(\times 10^8)$	5.979	2.981	60.298	3.025	1.803†	43.156

† Indicates the smallest MSE for the row.

4.2 ANOVA

Furthermore, ANOVA techniques were used for the analyses of the above results. The analyses were performed for two scenarios of different responses separately. For each response, the coefficients were also taken into account in the model. The objective of the analysis was to investigate the five main factors: two types of design, three bandwidths, three covariance matrices, coefficients $\boldsymbol{\beta}$ and correlation ϵ. The significance patterns of the second-order interactions were also of interest. The results are listed in Table 5.

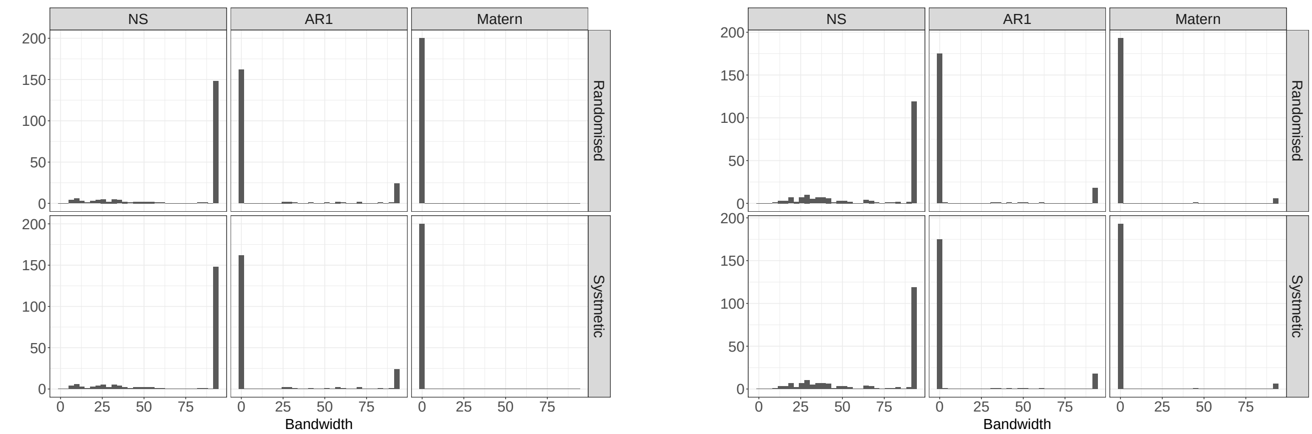
The results are consistent with what was observed in the previous subsection that, for the linear response, the difference between the randomised and systematic designs was not significant. However, for the quadratic response, the design and its interactions with the bandwidth and the coefficients were significant. For both scenarios, the intensity of the correlation and all its interactions were not significant (Table 5). Therefore, the GWR performs similarly with a low or high correlation between coefficients. Also of note was that bandwidth and its second-order interactions with variables other than correlation were found to be significant for both response types.

Table 5: ANOVA analyses were conducted on the main factors and their second order interactions. The table lists degrees of freedom (Df), sum of squared errors (Sum Sq) and p-values of F tests (Pr(>F)).

	Linear			Quadratic		
	Df	Sum Sq	Pr(>F)	Df	Sum Sq	Pr(>F)
Design	1	8.46	0.0935	1	74.03	<0.001
Bandwidth	2	7453.09	<0.001	2	90.29	<0.001
Covariance (V_s)	2	16958.42	<0.001	2	6122.84	<0.001
Coefficients (β)	1	903705.52	<0.001	2	1419372.21	<0.001
Correlation (ϵ)	1	5.63	0.171	1	0.07	0.8976
Design:Bandwidth	2	9.21	0.216	2	113.48	<0.001
Design:Covariance	2	3.91	0.5217	2	37.55	0.0134
Design:Coefficients	1	8.45	0.0935	2	147.55	<0.001
Design:Correlation	1	0	0.9832	1	0.01	0.9541
Bandwidth:Covariance	4	12283.13	<0.001	4	3135.47	<0.001
Bandwidth:Coefficients	2	7456.81	<0.001	4	179.38	<0.001
Bandwidth:Correlation	2	5.95	0.3716	2	0.85	0.9073
Covariance:Coefficients	2	16959.68	<0.001	4	12248.74	<0.001
Covariance:Correlation	2	0.26	0.9572	2	11.57	0.2649
Coefficients:Correlation	1	5.63	0.171	2	0.15	0.9834

4.3 AICc bandwidth selection

From the simulation study, we found that the bandwidth given by AICc skewed to 1 if spatial covariance was included in the model, for all types of the design and the nature of the response. If spatial covariance was not introduced, the GWR tended to use all data in one row (Figure 5).



(a) Histogram of optimal bandwidth for linear response.

(b) Histogram of optimal bandwidth for quadratic response.

Figure 5: Histogram of optimal bandwidth found by AICc for linear and quadratic response.

5 Discussion

Agronomists and biometricians generally prefer randomised designs for OFE trials. According to the performance metrics used, our simulation study shows a systematic design performs either preferably or similarly to a randomised design for the purposes of creating a varying treatment map. The differentiating factors included primarily the response type and the spatial covariance model, while the correlation amongst the treatment coefficients was not found to be important. These are factors that can be assessed by the farmer beforehand, and this should dictate which design should be used. However, given that a systematic design is easier to implement in the field, and shows little downside for the purposes of creating a varying treatment map, we will advocate for the use of systematic designs.

The response type was the main differentiating factor between randomised and systematic designs. When the response was quadratic, the systematic design performed favourably, which contrasted the result for the linear design. Given this, if a farmer expects an approximately linear response in the field, then the selection of the design may not be important. However, given the variable nature of the relationship between response and treatment over a large field (see Rakshit et al. (2020)), it may be wise to implement a systematic design for the potential outcome of a quadratic relationship.

Another consideration for farmers as to which design to use is the expected spatial covariance structure in the field. When no spatial structure was simulated, the differences between the prediction from the systematic and random design were minimal. This result should be expected given that if there are no spatial autocorrelations, then the individual query grids are independent observations, and therefore the design is not important. However, when a first-order auto-regressive structure was simulated, the differences were noticeable when a quadratic response was used, showing systematic designs to be preferential. The largest difference between the two designs occurred when considering the Matérn spatial covariance structure, which showed a clear preference for systematic designs when a quadratic response was considered, and also a small preference for systematic designs for a linear response. Therefore, only if spatial variability was predicted to be negligible in the field would using a randomised design be reasonable given a quadratic response. This assumption of negligible spatial variability would be difficult to reason with given the large fields used in on-farm experimentation (add reference), meaning that in the application a systematic design should be used.

There were found to be significant deficiencies in using AICc for bandwidth selection. The AICc-minimising bandwidths skewed to 1 and, in a few cases, ended in 93 (number of rows). Even though the bandwidth was optimal according to AICc, the MSE was higher than when using a fixed bandwidth. Therefore, we recommended using the fixed bandwidth based on the experimental design (5 or 9 in this case), rather than the recommended bandwidth from AICc which is prone to either over-fitting or being too generalised. Selecting the bandwidth based on the experimental design is also theoretically better since only a single measurement is observed in each grid, all levels of the treatment factor should be included in a GWR window at the same time to interpolate the relationship. Otherwise, the interpolation is incomplete if more than one level is missing.

Given the scope of the paper, some designs and factors were not considered. Designs such as chequerboard or wave designs have been suggested for on-farm experiments (Bramley et al. 1999), however, were not considered here. Topographical factors (spatial zones) were also not entertained in our study. Since GWR estimates a global template model and then adjusts it at a local scale across the study region, the variation between zones is "flushed out" by the spatial covariance.

6 Conclusion

Agronomists and biometricians generally prefer randomised designs for OFE trials. With the purpose of creating a varying treatment map, our simulation study proves that a systematic design produces better performance metrics, under particular circumstances, than a randomised design for large on-farm trials in terms of robustness and smaller MSE on coefficients. On the other hand, if spatial variation is not considered or if researchers believe in linear response, a systematic or a randomised design could be implemented because the difference is not significant. We recommend that, for a large OFE strip trial with the goal to create a varying treatment map, a systemic design should be used as it has more flexibility in post-experiment statistical modelling.

Acknowledgements

Curtin Biometry and Data Analytics (CBADA) gratefully acknowledges the support from the Grains Research and Development Corporation of Australia (GRDC).

CRediT authorship contribution statement

Zhanglong Cao: Conceptualization; Formal analysis; Methodology; Software; Validation; Visualization; Roles/ Writing - original draft; and Writing - review & editing. **Jordan Brown**: Investigation; Methodology; Visualization;

Roles/Writing - original draft; and Writing - review & editing. **Mark Gibberd**: Project administration; Roles/Writing - original draft; and Writing - review & editing. **Julia Easton**: Roles/Writing - original draft; and Writing - review & editing. **Suman Rakshit**: Conceptualization; Methodology; Supervision; Roles/Writing - original draft; and Writing - review & editing.

A Figures

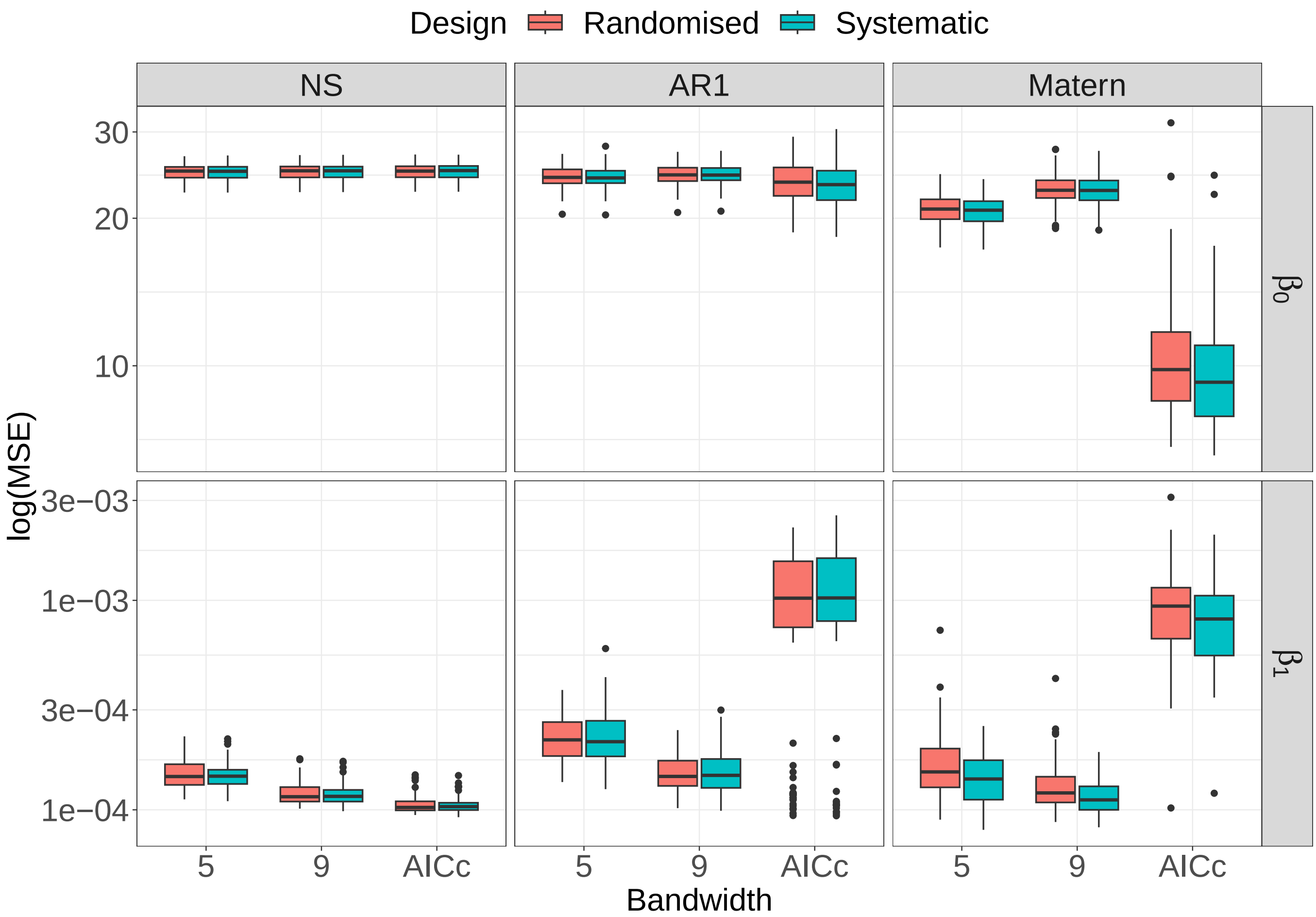


Figure 6: Boxplots of the logarithm of MSE for $\hat{\beta}_0$ and $\hat{\beta}_1$ in GWR models using different bandwidths for the simulated data with a linear response. The simulated data had different spatial covariance matrices (NS, AR1⊗AR1 and Matérn) and a high correlation between the parameters ($\epsilon = 0.1$).

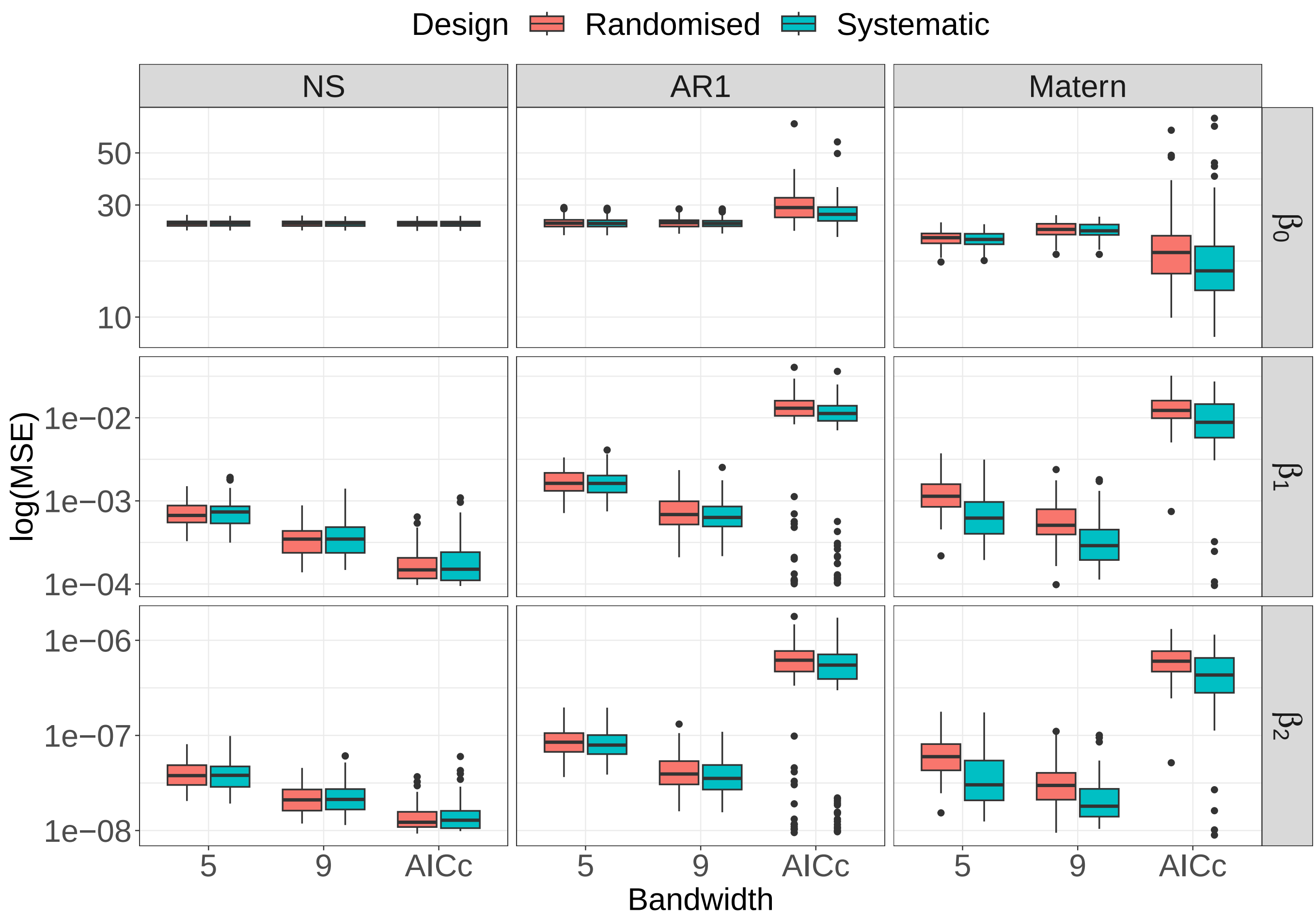


Figure 7: Boxplots of the logarithm of MSE for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ in GWR models using different bandwidths for the simulated data with a quadratic response. The simulated data had different spatial covariance matrices (NS, AR1⊗AR1 and Matérn) and a high correlation amongst the parameters ($\epsilon = 0.1$).

References

Abramowitz, M. (1974). *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables.* USA: Dover Publications, Inc. (cit. on p. 5).

Alesso, C. A., Cipriotti, P. A., Bollero, G. A., and Martin, N. F. (2021). "Design of on-farm precision experiments to estimate site-specific crop responses". In: *Agronomy Journal* 113.2, pp. 1366–1380. DOI: https://doi.org/10.1002/agj2.20572 (cit. on p. 2).

Anselin, L., Bongiovanni, R., and Lowenberg-DeBoer, J. (2004). "A Spatial Econometric Approach to the Economics of Site-Specific Nitrogen Management in Corn Production". In: *American Journal of Agricultural Economics* 86.3, pp. 675–687. DOI: https://doi.org/10.1111/j.0002-9092.2004.00610.x (cit. on p. 6).

Bramley, R., Cook, S., Adams, M., and Corner, R. (1999). "Designing your own on-farm experiments: How precision agriculture can help". In: (cit. on p. 14).

Butler, D., Cullis, B., Gilmour, A., Gogel, B., and Thompson, R. (2017). *ASReml-R Reference Manual Version 4* (cit. on p. 4).

Cao, Z., Stefanova, K., Gibberd, M., and Rakshit, S. (2022). "Bayesian inference of spatially correlated random parameters for on-farm experiment". In: *Field Crops Research* 281, p. 108477. DOI: https://doi.org/10.1016/j.fcr.2022.108477 (cit. on pp. 2, 4, 6).

Cook, S., Cock, J., Oberthür, T., and Fisher, M. (2013). "On-Farm Experimentation". In: *Better Crop. Plant Food.* 4th ser. 97, pp. 17–20 (cit. on p. 2).

Cressie, N. and Huang, H.-C. (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 (cit. on p. 5).

Evans, F. H., Recalde Salas, A., Rakshit, S., Scanlan, C. A., and Cook, S. E. (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 (cit. on pp. 2, 6).

Fisher, R. A. (1934). *Statistical Methods for Research Workers.* English. fifth. Edinburgh: Oliver and Boyd (cit. on p. 2).

Glynn, C. (2007). "Testing the Growth-Differentiation Balance Hypothesis: Dynamic Responses of Willows to Nutrient Availability". In: *New Phytologist* 176.3, pp. 623–634. DOI: 10.1111/j.1469-8137.2007.02203.x (cit. on p. 3).

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., and Harris, P. (Feb. 2015). "GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models". en. In: *Journal of Statistical Software* 63.1, pp. 1–50. DOI: 10.18637/jss.v063.i17 (cit. on pp. 5, 6).

Leung, Y., Mei, C.-L., and Zhang, W.-X. (Jan. 2000). "Statistical Tests for Spatial Nonstationarity Based on the Geographically Weighted Regression Model". In: *Environment and Planning A: Economy and Space* 32.1, pp. 9–32. DOI: 10.1068/a3162 (cit. on p. 5).

Lewandowski, D., Kurowicka, D., and Joe, H. (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 (cit. on p. 4).

Liben, F. M., Midega, T., Tufa, T., and Wortmann, C. S. (2019). "Soil Fertility & Crop Nutrition Barley and wheat nutrient responses for Shewa, Ethiopia". In: *Agronomy Journal.* DOI: 10.1002/agj2.20020 (cit. on p. 3).

Lu, B., Harris, P., Charlton, M., and Brunsdon, C. (2014). "The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models". In: *Geo-spatial Information Science* 17.2, pp. 85–101. DOI: 10.1080/10095020.2014.917453 (cit. on p. 6).

Marchant, B., Rudolph, S., Roques, S., Kindred, D., Gillingham, V., Welham, S., Coleman, C., and Sylvester-Bradley, R. (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 (cit. on p. 6).

Marschner, H. (2011). *Marschner's Mineral Nutrition of Higher Plants.* San Diego: San Diego: Elsevier Science & Technology. DOI: 10.1016/C2009-0-63043-9 (cit. on p. 3).

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 (cit. on p. 4).

Páez, A., Uchida, T., and Miyamoto, K. (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: *Environment and Planning A: Economy and Space* 34.4, pp. 733–754. DOI: 10.1068/a34110 (cit. on p. 5).

Pandit, R. K. and Infield, D. (Sept. 2019). "Comparative Analysis of Gaussian Process Power Curve Models Based on Different Stationary Covariance Functions for the Purpose of Improving Model Accuracy". en. In: *Renewable Energy* 140, pp. 190–202. DOI: 10.1016/j.renene.2019.03.047 (cit. on p. 5).

Panten, K., Bramley, R. G. V., Lark, R. M., and Bishop, T. F. A. (Apr. 2010). "Enhancing the Value of Field Experimentation through Whole-of-Block Designs". en. In: *Precision Agriculture* 11.2, pp. 198–213. DOI: 10.1007/s11119-009-9128-y (cit. on p. 5).

Petersen, R. G. (1994). *Agricultural Field Experiments: Design and Analysis*. en. 1st. Boca Raton: CRC Press (cit. on pp. 2, 3).

Piepho, H. P. and Edmondson, R. N. (2018). "A tutorial on the statistical analysis of factorial experiments with qualitative and quantitative treatment factor levels". In: *Journal of Agronomy and Crop Science* 204.5, pp. 429–455. DOI: https://doi.org/10.1111/jac.12267 (cit. on p. 3).

Piepho, H. P., Möhring, J., and Williams, E. R. (2013). "Why Randomize Agricultural Experiments?" In: *Journal of Agronomy and Crop Science* 199.5, pp. 374–383. DOI: 10.1111/jac.12026 (cit. on pp. 2, 6).

Piepho, H.-P., Richter, C., Spilke, J., Hartung, K., and Kunick, A. (2011). "Statistical Aspects of On-Farm Experimentation". In: *Crop & Pasture Science* 62, pp. 721–735. DOI: 10.1071/cp11175 (cit. on pp. 2, 5).

Pringle, M. J., Cook, S. E., and McBratney, A. B. (Dec. 2004). "Field-Scale Experiments for Site-Specific Crop Management. Part I: Design Considerations". en. In: *Precision Agriculture* 5.6, pp. 617–624. DOI: 10.1007/s11119-004-6346-1 (cit. on pp. 2, 3).

Rakshit, S., Baddeley, A., Stefanova, K., Reeves, K., Chen, K., Cao, Z., Evans, F., and Gibberd, M. (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 (cit. on pp. 2, 5, 6, 13).

Selle, M. L., Steinsland, I., Hickey, J. M., and Gorjanc, G. (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 (cit. on pp. 2, 5).

Verdooren, L. R. (Dec. 2020). "History of the Statistical Design of Agricultural Experiments". en. In: *Journal of Agricultural, Biological and Environmental Statistics* 25.4, pp. 457–486. DOI: 10.1007/s13253-020-00394-3 (cit. on p. 2).

White, J. W. and Evert, F. K. van (2008). "Publishing Agronomic Data". In: *Agronomy Journal* 100.5, pp. 1396–1400. DOI: https://doi.org/10.2134/agronj2008.0080F (cit. on p. 6).

Zimmerman, D. L. and Harville, D. A. (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 (cit. on p. 3).