Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 138 (2008) 93–104
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Robust estimators and designs for field experiments

Douglas P. Wiens[a,*], Julie Zhou[b]

[a]*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1*
[b]*Department of Mathematics and Statistics, University of Victoria, Canada*

Available online 16 May 2007

Abstract

We consider the construction of designs for test-control field experiments, with particular attention being paid to the effects of spatial correlation between adjoining plots. In contrast to previous approaches, in which very specific correlation structures were modelled, we explicitly allow a degree of uncertainty on the part of the experimenter. While fitting a particular correlation structure—and variance structure and regression response—the experimenter is thought to be seeking protection against other possible structures in full neighbourhoods of these particular choices. Robustness, in a minimax sense, is obtained through a modification of the kriging estimation procedure, and through the assignment of treatments to field plots.

MSC: Primary 62K05, 62F35; secondary 62P12

Keywords: Check variety; Crop; Kriging; Minimax; Regression; Simulated annealing; Spatial correlation

1. Introduction

We consider the construction of designs for test-control field experiments, with particular attention being paid to the effects of spatial correlation between adjoining plots. These effects are by now well documented—see for example Taplin (1999), Wu and Dutilleul (1999) and Legendre et al. (2004). Martin (1986) initiated a systematic study of designs which account for spatial correlation; see also Petraitis (2001) and Fagroud and Van Meirvenne (2002) for *ad hoc* recommendations. In contrast to these approaches (but see also Martin et al., 1993), in which very specific correlation structures were modelled, we explicitly allow a degree of uncertainty on the part of the experimenter. While fitting a particular correlation structure—and variance structure and regression response—the experimenter is thought to be seeking protection against other possible structures in full neighbourhoods of these choices. Robustness, in a minimax sense detailed below, is obtained through a modification of the kriging estimation procedure, and through the assignment of treatments to field plots.

To motivate the discussion, consider a rectangular layout of plots, with r rows and c columns. There are t crop varieties to be planted. One, denoted variety 1, is a control, or "check" variety. The $n = rc$ locations of plots are denoted by $\mathbf{t}_1, \ldots, \mathbf{t}_n$. A task of the experimenter is to allocate crop varieties to plots, subject perhaps to constraints on the frequencies $\{n_i\}_{i=1}^{t}$ with which each variety will appear; for instance we may require that they be equal, or that the frequency n_1 have special prominence.

* Corresponding author. Tel.: +1 780 492 4406; fax: +1 780 492 6823.
E-mail addresses: doug.wiens@ualberta.ca (D.P. Wiens), jzhou@math.uvic.ca (J. Zhou).

0378-3758/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2007.05.020

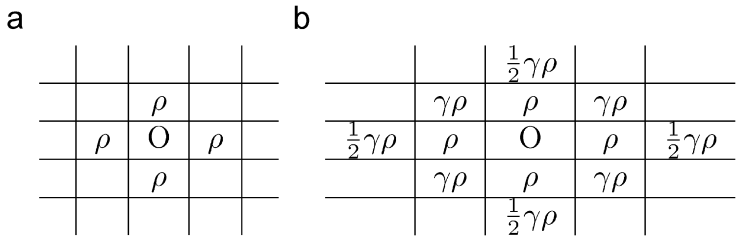

Fig. 1. Correlation structures for plot O: (a) NN(1); (b) MA(1).

In this paper treatments are not limited to crop varieties; they can be various nutrients, or combinations of crop varieties and nutrients, etc. One objective of the experiment is to model an observed random variable Y, e.g. crop yield, in terms of variety, and perhaps in terms of other design variables (added nutrients, water, etc.) as well. Denote by $\mathbf{x}_{q\times 1}$ the vector of such design variables. Each treatment determines a value of $\mathbf{x}$. We might also allow for a deterministic spatial trend. The experimenter intends to fit a linear regression model, with regressors $\mathbf{v}(\mathbf{x})$: $p_1 \times 1$, $\mathbf{w}(\mathbf{t})$: $p_2 \times 1$ and response

$$E[Y|\mathbf{x},\mathbf{t}] \approx \mathbf{v}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\theta}_{(1)} + \mathbf{w}^{\mathrm{T}}(\mathbf{t})\boldsymbol{\theta}_{(2)}. \tag{1}$$

As indicated by (1) the fitted response is acknowledged to be only an approximation, in which case ensuing estimates and predictions will be biased. In some cases, depending on the richness of the space over which $\mathbf{x}$ varies, some protection against this bias can be obtained through a judicious choice of the design variables.

Even in the simple case that $q = 1$ and x is merely the indicator of the variety, the allocation of varieties to plots should take into account the spatial correlations between the plots. Failure to do so may of course result in inefficient estimates. But the specification of an appropriate model for these correlations can be very difficult, and one typically posits a simple and intuitively pleasing model in the hopes that it will at least be adequate, if not particularly accurate. Robustness of design is again called for, as is robustness of the estimation procedure.

We anticipate that the experimenter will work with a particular correlation structure, but seek protection against a full neighbourhood of such structures. Two particular structures, appealing for their simplicity, are the first order nearest neighbour (NN(1)) model (Kiefer and Wynn, 1981) and the first order moving average (MA(1)) model (Haining, 1978). These are illustrated in Fig. 1 (a) and (b), respectively. Under the NN(1) structure the plot O has positive correlation ρ only with its first order neighbours, and $\rho \in (0, \frac{1}{4})$. Under the MA(1) structure O has positive correlation with 12 neighbours, described by a parameter $\gamma \in (0, \frac{1}{4})$ with $\rho = 2\gamma/(1+4\gamma^2)$.

A further call for robustness comes from the need to assume a variance structure on the measurement errors. One would generally fit a homoscedastic model, while wishing for some protection against heteroscedasticity. Again we will deal with this by seeking robustness against a class of such structures.

Specifically and more generally, we have a fixed set $\mathscr{T} = \{\mathbf{t}_1, \ldots, \mathbf{t}_n\}$ of locations of field plots. At location $\mathbf{t}_i$ one sets the levels of an independent variable $\mathbf{x}$, and observes

$$Y_i = E[Y|\mathbf{x}_i, \mathbf{t}_i] + \delta(\mathbf{t}_i) + \varepsilon(\mathbf{t}_i), \quad i = 1, \ldots, n.$$

Here:

(i) $\{\varepsilon(\mathbf{t})|\mathbf{t} \in \mathscr{T}\}$ is a set of uncorrelated zero-mean random errors, with variance matrix $\mathbf{F} = \mathrm{diag}(f(\mathbf{t}_1), \ldots, f(\mathbf{t}_n))$. We allow f to vary over a neighbourhood of a variance function f_0:

$$\mathscr{F}_\alpha = \left\{ f(\cdot) \left| \max_{\mathbf{t}\in\mathscr{T}} \right| f(\mathbf{t}) - f_0(\mathbf{t})| \leqslant \alpha,\ f(\mathbf{t}) \geqslant 0 \right\}.$$

(ii) $\{\delta(\mathbf{t})|\mathbf{t} \in \mathscr{T}\}$ is a zero-mean, spatially correlated stochastic process (uncorrelated with $\{\varepsilon(\mathbf{t})|\mathbf{t} \in \mathscr{T}\}$) with covariance function $g(\mathbf{t}_i, \mathbf{t}_j) = cov[\delta(\mathbf{t}_i), \delta(\mathbf{t}_j)]$ and covariance matrix $\mathbf{G}_{n\times n} = (g(\mathbf{t}_i, \mathbf{t}_j))_{i,j=1}^n$. We allow g to vary over

$$\mathscr{G}_\beta = \{g(\cdot,\cdot)|\mathbf{0} \leqslant \mathbf{G} \leqslant \mathbf{G}^{(0)} + \beta\mathbf{K}\},$$

where $\mathbf{G}^{(0)}$ is the covariance matrix under g_0 and $\mathbf{G} \geqslant \mathbf{0}$ refers to the ordering by non-negative definiteness (n.n.d.). The matrix $\mathbf{K}$ is a fixed n.n.d. matrix, typically $\mathbf{K} = \mathbf{I}_n$ or $\mathbf{K} = \mathbf{G}^{(0)}$. See Remark 3 of Section 2 for extensions.

(iii) $E[Y|\mathbf{x}, \mathbf{t}] = \psi(\mathbf{x}) + \mathbf{w}^{\mathrm{T}}(\mathbf{t})\boldsymbol{\theta}_{(2)}$ for some p_2-dimensional vectors $\mathbf{w}(\mathbf{t})$ of known regressors and $\boldsymbol{\theta}_{(2)}$ of regression parameters to be estimated. The function $\psi(\mathbf{x})$ is thought to be well approximated by a linear combination $\mathbf{v}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\theta}_{(1)}$ of p_1 regressors $v_j(\mathbf{x})$. Here the q-dimensional variable $\mathbf{x}$ lies in $\mathscr{S} = \{\mathbf{u}_1, \ldots, \mathbf{u}_N\}$, and $\boldsymbol{\theta}_{(1)}$ furnishes the 'best' approximation in that

$$\boldsymbol{\theta}_{(1)} = \arg\min_{\boldsymbol{\phi}} \sum_{\mathbf{x}\in\mathscr{S}} \{\psi(\mathbf{x}) - \mathbf{v}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\phi}\}^2.$$

Put

$$h(\mathbf{x}) = \psi(\mathbf{x}) - \mathbf{v}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\theta}_{(1)},$$

so that

$$E[Y|\mathbf{x}, \mathbf{t}] = \mathbf{v}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\theta}_{(1)} + \mathbf{w}^{\mathrm{T}}(\mathbf{t})\boldsymbol{\theta}_{(2)} + h(\mathbf{x}) \quad \text{and} \quad \sum_{\mathbf{x}\in\mathscr{S}} \mathbf{v}(\mathbf{x})h(\mathbf{x}) = 0.$$

We allow h to vary over

$$\mathscr{H}_\gamma = \left\{ h(\cdot) | N^{-1} \sum_{\mathbf{x}\in\mathscr{S}} h^2(\mathbf{x}) \leqslant \gamma, \sum_{\mathbf{x}\in\mathscr{S}} \mathbf{v}(\mathbf{x})h(\mathbf{x}) = \mathbf{0} \right\}.$$

This setup is superficially similar to that used in the study of spatial designs by Wiens (2005b). A crucial difference is that there, the locations $\mathbf{t}_1, \ldots, \mathbf{t}_n$ were *chosen* from a larger set of possible locations, at each of which there was a vector $\mathbf{x}$ of known covariates. Optimum choices $\mathbf{t}_1, \ldots, \mathbf{t}_n$ were to be made on the basis of $\mathbf{x}$, taking into account the spatial correlations. By contrast, in the current work the n locations of field plots are *fixed*, and each is to be assigned exactly one treatment—more generally, a value of $\mathbf{x}$. The design problem is that of *assigning* a choice $\mathbf{x}$, from $\mathscr{S}$, in some optimal manner.

Martin (2001) obtained designs for field experiments in which only the locations of the check varieties were chosen, thus allowing him to exploit a much closer analogy with the spatial design problem. Atkinson (1996) applied Bayesian design theory to the field of agricultural design, as an alternative to the criterion of local optimality when the regression response is nonlinear (entailing a dependence of the loss function on unknown parameters). A Bayesian approach could in principle complement the minimax approach taken by us; such an approach would involve placing reasonable priors on neighbourhoods such as $\mathscr{F}_\alpha$, $\mathscr{G}_\beta$ and $\mathscr{H}_\gamma$.

Note that we are assuming that any deterministic spatial effects, and the effects of any design variables, are additive. This assumption is commonly met in field trials.

Remark 1. If the set of possible regressors $\{\mathbf{v}(\mathbf{x})|\mathbf{x} \in \mathscr{S})\}$ contains a basis of $\mathbb{R}^N$, then the orthogonality condition in the definition of $\mathscr{H}_\gamma$ forces $\mathscr{H}_\gamma = \{0\}$. This is the case for instance if the only regressors are the indicators of the treatments, so that $p_1 = N$ and $(\mathbf{u}_1|\cdots|\mathbf{u}_N) = \mathbf{I}_N$.

For the post-design estimation the experimenter acts as though $\gamma = 0$, implying that $E[Y|\mathbf{x}, \mathbf{t}] = \mathbf{v}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\theta}_{(1)} + \mathbf{w}^{\mathrm{T}}(\mathbf{t})\boldsymbol{\theta}_{(2)}$. He seeks an unbiased linear estimator $\mathbf{A}_{M\times n}\mathbf{y}$ of $\mathbf{C}_{M\times n}\boldsymbol{\mu}$, where $\boldsymbol{\mu} = E[\mathbf{y}] = \mathbf{V}_{n\times p_1}\boldsymbol{\theta}_{(1)} + \mathbf{W}_{n\times p_2}\boldsymbol{\theta}_{(2)} = \mathbf{Z}_{n\times p}\boldsymbol{\theta}$, and where:

(i) $\mathbf{y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_{(1)}^{\mathrm{T}}, \boldsymbol{\theta}_{(2)}^{\mathrm{T}})^{\mathrm{T}}$;

(ii) $\mathbf{V}$ has rows $\mathbf{v}^{\mathrm{T}}(\mathbf{x}_i)$, $\mathbf{W}$ has rows $\mathbf{w}^{\mathrm{T}}(\mathbf{t}_i)$, $i = 1, \ldots, n$, and $\mathbf{Z}_{n\times p} = (\mathbf{V}\vdots\mathbf{W})$ with rows $\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i, \mathbf{t}_i) = (\mathbf{v}^{\mathrm{T}}(\mathbf{x}_i), \mathbf{w}^{\mathrm{T}}(\mathbf{t}_i))$ of length $p = p_1 + p_2$;

(iii) $\mathbf{C}$ is a fixed matrix, typically a contrast matrix;

(iv) 'Unbiased' means that $E[\mathbf{Ay}] = \mathbf{C}\boldsymbol{\mu}$ for all $\boldsymbol{\theta}$, requiring that

$$\mathbf{BZ} = \mathbf{0} \quad \text{where } \mathbf{B}_{M\times n} = \mathbf{C} - \mathbf{A}. \tag{2}$$

Loss is measured by the trace of the mean squared error matrix:

$$\text{MSE} = E[(\mathbf{Ay} - \mathbf{C}\boldsymbol{\mu})(\mathbf{Ay} - \mathbf{C}\boldsymbol{\mu})^{\mathrm{T}}].$$

At this point there are two ways to proceed. One can obtain a matrix $\mathbf{A}_{0,0}$ which minimizes the trace of MSE subject to (2), with MSE evaluated at f_0 and g_0:

$$\mathbf{A}_{0,0} = \arg\min_{\mathbf{A}} \operatorname{tr}(\text{MSE}|\alpha = \beta = \gamma = 0).$$

In this case we obtain the well known kriging estimate, which is the best linear unbiased estimate under f_0 and g_0. Alternatively one may take a *minimax robust* linear estimate defined by

$$\mathbf{A}_{\alpha,\beta} = \arg\min_{\mathbf{A}} \sup_{\mathscr{F}_\alpha, \mathscr{G}_\beta} \operatorname{tr}(\text{MSE}|\gamma = 0). \tag{3}$$

Note that this reduces to the first approach when $\alpha = \beta = 0$. Here we use that, although $\mathbf{G}_0$ is not the only member of $\mathscr{G}_0$, it is the least favourable member—this is brought out in the proof of Theorem 1 of the next section.

In Theorem 1 we will exhibit $\mathbf{A}_{\alpha,\beta}$, and the resulting loss

$$\mathscr{L}(\alpha, \beta) = \sup_{\mathscr{F}_\alpha, \mathscr{G}_\beta} \operatorname{tr}(\text{MSE}|\mathbf{A} = \mathbf{A}_{\alpha,\beta}),$$

in which the minimax estimate is used and the fitted response $\mathbf{z}^{\mathrm{T}}(\mathbf{x}, \mathbf{t})\boldsymbol{\theta}$ is assumed correct. We will then drop this assumption and obtain, in Theorem 2 of Section 2,

$$\mathscr{L}(\alpha, \beta, \gamma) = \sup_{\mathscr{H}_\gamma} \mathscr{L}(\alpha, \beta) = \sup_{\mathscr{H}_\gamma} \sup_{\mathscr{F}_\alpha, \mathscr{G}_\beta} \operatorname{tr}(\text{MSE}|\mathbf{A} = \mathbf{A}_{\alpha,\beta}).$$

Minimization of this quantity, through the assignment of treatments to field plots, gives the desired minimax robust design. This is a numerical problem for which algorithms are discussed in Section 3. Applications and examples are discussed as well in Section 3, and a case study is analysed in Section 4.

2. Minimax linear estimates and designs

Theorem 1. *The minimax linear estimate* $\mathbf{A}_{\alpha,\beta}\mathbf{y}$ *defined by* (2) *and* (3) *is given by*

$$\mathbf{A}_{\alpha,\beta} = \mathbf{C}\mathbf{R}_{\boldsymbol{\alpha},\boldsymbol{\beta}},$$

where $\mathbf{R}_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Lambda}_{\alpha,\beta}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Lambda}_{\alpha,\beta}^{-1}$ *and* $\boldsymbol{\Lambda}_{\boldsymbol{\alpha},\boldsymbol{\beta}} = (\mathbf{G}^{(0)} + \beta\mathbf{K}) + (\mathbf{F}^{(0)} + \alpha\mathbf{I}_n)$. *Minimax loss is*

$$\mathscr{L}(\alpha, \beta) = \operatorname{tr}[\mathbf{A}_{\alpha,\beta}\boldsymbol{\Lambda}_{\boldsymbol{\alpha},\boldsymbol{\beta}}\mathbf{A}_{\alpha,\beta}^{\mathrm{T}}] = \operatorname{tr}[\mathbf{C}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Lambda}_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}]. \tag{4}$$

Proof. Since we are assuming $\gamma = 0$,

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\delta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\delta}$, $\boldsymbol{\varepsilon}$ have elements $\delta(\mathbf{t}_i)$, $\varepsilon(\mathbf{t}_i)$, respectively. Then $\mathbf{Ay} - \mathbf{C}\boldsymbol{\mu} = \mathbf{A}(\boldsymbol{\delta} + \boldsymbol{\varepsilon}) - \mathbf{B}\mathbf{Z}\boldsymbol{\theta}$ and for any $\mathbf{A}$ we have

$$\begin{aligned}\text{MSE} &= \mathbf{A}E[(\boldsymbol{\delta} + \boldsymbol{\varepsilon})(\boldsymbol{\delta} + \boldsymbol{\varepsilon})^{\mathrm{T}}]\mathbf{A}^{\mathrm{T}} + (\mathbf{B}\mathbf{Z}\boldsymbol{\theta})(\mathbf{B}\mathbf{Z}\boldsymbol{\theta})^{\mathrm{T}} \\ &= \mathbf{A}\mathbf{G}\mathbf{A}^{\mathrm{T}} + \mathbf{A}\mathbf{F}\mathbf{A}^{\mathrm{T}} + (\mathbf{B}\mathbf{Z}\boldsymbol{\theta})(\mathbf{B}\mathbf{Z}\boldsymbol{\theta})^{\mathrm{T}},\end{aligned}$$

with

$$\operatorname{tr}(\text{MSE}) = \{\operatorname{tr}[\mathbf{A}\mathbf{G}\mathbf{A}^{\mathrm{T}}] + \operatorname{tr}[\mathbf{A}\mathbf{F}\mathbf{A}^{\mathrm{T}}]\} + \|\mathbf{B}\mathbf{Z}\boldsymbol{\theta}\|^2.$$

Maximum loss is easily seen to be attained at $\mathbf{G} = \mathbf{G}^{(0)} + \beta\mathbf{K}$ and $f(\mathbf{t}) = f_0(\mathbf{t}) + \alpha$, with

$$\begin{aligned}\mathscr{L}(\alpha, \beta) &= \{\operatorname{tr}[\mathbf{A}(\mathbf{G}^{(0)} + \beta\mathbf{K})\mathbf{A}^{\mathrm{T}}] + \operatorname{tr}[\mathbf{A}(\mathbf{F}^{(0)} + \alpha\mathbf{I}_n)\mathbf{A}^{\mathrm{T}}]\} + \|\mathbf{B}\mathbf{Z}\boldsymbol{\theta}\|^2 \\ &= \operatorname{tr}[\mathbf{A}\boldsymbol{\Lambda}_{\boldsymbol{\alpha},\boldsymbol{\beta}}\mathbf{A}^{\mathrm{T}}] + \|\mathbf{B}\mathbf{Z}\boldsymbol{\theta}\|^2.\end{aligned}$$

The minimax linear unbiased estimate $\mathbf{A}_{\alpha,\beta}\mathbf{y}$ is obtained by minimizing tr $\mathbf{A}\boldsymbol{\Lambda}_{\boldsymbol{\alpha},\boldsymbol{\beta}}\mathbf{A}^{\mathrm{T}}$, subject to the requirement that $\mathbf{BZ}\boldsymbol{\theta}$ vanish identically. It is thus the universal kriging estimate (see e.g. Cressie, 1993) computed assuming variance and covariance matrices $\mathbf{G}^{(0)}+\beta\mathbf{K}$ and $\mathbf{F}^{(0)}+\alpha\mathbf{I}_n$, i.e., is given by $\mathbf{A}_{\alpha,\beta}=\mathbf{CR}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. The second equality in (4) is now a simple calculation. □

Remark 2. Note that $\mathbf{A}_{\alpha,\beta}\mathbf{y}=\mathbf{CZ}\hat{\boldsymbol{\theta}}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\theta}}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is the generalized least squares estimate computed from $\boldsymbol{\Lambda}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. This suggests a way to further robustify the procedure—one might replace $\hat{\boldsymbol{\theta}}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ by a more robust estimate such as an M-estimate. This approach was explored in Wiens (2005a); for other robust approaches see Cressie and Hawkins (1980), Hawkins and Cressie (1984) and Genton (2001).

Remark 3. A referee has pointed out that Theorem 1 holds, with $\mathbf{K}=\mathbf{I}_n$, if the class $\mathscr{G}_\beta$ is replaced by $\mathscr{G}'_\beta=\{g(\cdot,\cdot)|\mathbf{G}\geqslant\mathbf{0},\ \|\mathbf{G}-\mathbf{G}^{(0)}\|\leqslant\beta\}$, where $\|\cdot\|$ is either $\|\cdot\|_1$ (the maximum absolute column sum norm) or $\|\cdot\|_\infty$(the maximum absolute row sum norm). In fact it holds for any induced matrix norm. This is because $\mathbf{G}^{(0)}+\beta\mathbf{I}_n\in\mathscr{G}'_\beta\subset\mathscr{G}_\beta$, so that $\mathbf{G}^{(0)}+\beta\mathbf{I}_n$ is the maximizer of the loss in the smaller class $\mathscr{G}'_\beta$ as well as in the larger class $\mathscr{G}_\beta$. The class $\mathscr{G}'_\beta$ has the property, not shared by $\mathscr{G}_\beta$, that it reduces to the singleton $\{\mathbf{G}^{(0)}\}$ when $\beta=0$. To see the inclusion $\mathscr{G}'_\beta\subset\mathscr{G}_\beta$, put $\mathbf{D}=\mathbf{G}-\mathbf{G}^{(0)}$. The inclusion will follow if $\mathbf{G}^{(0)}+\beta\mathbf{I}_n-\mathbf{G}=\beta\mathbf{I}_n-\mathbf{D}$ is n.n.d. whenever $\|\mathbf{D}\|\leqslant\beta$; this in turn follows from the well known result that the spectral radius $\rho(\mathbf{D})$ is at most $\|\mathbf{D}\|$ for any induced matrix norm $\|\cdot\|$.

Theorem 2. *The minimax design is obtained by choosing points* $\mathbf{x}_1,\ldots,\mathbf{x}_n\in\mathscr{S}$, *so as to minimize*

$$\mathscr{L}(\alpha,\beta,\gamma)=\mathrm{tr}[\mathbf{CZ}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}]+\gamma N\cdot\lambda_{\max}(\tilde{\mathbf{B}}^{\mathrm{T}}_{\alpha,\beta}\tilde{\mathbf{B}}_{\alpha,\beta}). \tag{5}$$

Here $\lambda_{\max}(\cdot)$ *denotes the maximum eigenvalue*,

$$\mathbf{B}_{\alpha,\beta}=\mathbf{C}(\mathbf{I}_n-\mathbf{R}_{\boldsymbol{\alpha},\boldsymbol{\beta}}):M\times n,$$

and the $M\times d$ *matrix* $\tilde{\mathbf{B}}_{\alpha,\beta}$ *is described as follows. Let* $\mathbf{b}_1,\ldots,\mathbf{b}_n$ *be the columns of* $\mathbf{B}_{\alpha,\beta}$. *Then the jth column of* $\tilde{\mathbf{B}}_{\alpha,\beta}$ *is* $\sum_{k=1}^n\mathbf{b}_k I(\mathbf{x}_k=\mathbf{x}_{i_j})$, $j=1,\ldots,d$, *where* $\{\mathbf{x}_{i_1},\ldots,\mathbf{x}_{i_d}\}$ *are the distinct design points* (*treatments*).

Proof. We are now taking $\gamma>0$, whence

$$\mathbf{y}=\mathbf{Z}\boldsymbol{\theta}+\mathbf{h}+\boldsymbol{\delta}+\boldsymbol{\varepsilon},$$

where $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are as in the proof of Theorem 1 and $\mathbf{h}$ has elements $h(\mathbf{x}_i)$. Then $\mathbf{Ay}-\mathbf{C}\boldsymbol{\mu}=\mathbf{A}(\boldsymbol{\delta}+\boldsymbol{\varepsilon})-\mathbf{BZ}\boldsymbol{\theta}-\mathbf{Bh}$. For any estimator satisfying (2) we have

$$\begin{aligned}\mathrm{tr(MSE)}&=\mathrm{tr}\{\mathbf{A}E[(\boldsymbol{\delta}+\boldsymbol{\varepsilon})(\boldsymbol{\delta}+\boldsymbol{\varepsilon})^{\mathrm{T}}]\mathbf{A}^{\mathrm{T}}+\mathbf{B}(\mathbf{Z}\boldsymbol{\theta}+\mathbf{h})(\mathbf{Z}\boldsymbol{\theta}+\mathbf{h})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\}\\&=\mathrm{tr}[\mathbf{AGA}^{\mathrm{T}}+\mathbf{AFA}^{\mathrm{T}}]+\|\mathbf{Bh}\|^2.\end{aligned}$$

Thus the maximum loss incurred by using the minimax estimator $\mathbf{A}_{\alpha,\beta}\mathbf{y}$ is, using Theorem 1,

$$\begin{aligned}&\sup_{\mathscr{H}_\gamma}\sup_{\mathscr{F}_\alpha,\mathscr{G}_\beta}\mathrm{tr}E[(\mathbf{A}_{\alpha,\beta}\mathbf{y}-\mathbf{C}\boldsymbol{\mu})(\mathbf{A}_{\alpha,\beta}\mathbf{y}-\mathbf{C}\boldsymbol{\mu})^{\mathrm{T}}]\\&\quad=\sup_{\mathscr{F}_\alpha,\mathscr{G}_\beta}\mathrm{tr}[\mathbf{A}_{\alpha,\beta}\mathbf{G}\mathbf{A}^{\mathrm{T}}_{\alpha,\beta}+\mathbf{A}_{\alpha,\beta}\mathbf{F}\mathbf{A}^{\mathrm{T}}_{\alpha,\beta}]+\sup_{\mathscr{H}_\gamma}\|\mathbf{Bh}\|^2\\&\quad=\mathrm{tr}[\mathbf{CZ}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}_{\alpha,\beta}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}]+\sup_{\mathscr{H}_\gamma}\|\mathbf{Bh}\|^2,\end{aligned} \tag{6}$$

where now $\mathbf{B}=\mathbf{B}_{\alpha,\beta}=\mathbf{C}-\mathbf{A}_{\alpha,\beta}=\mathbf{C}(\mathbf{I}_n-\mathbf{R}_{\boldsymbol{\alpha},\boldsymbol{\beta}})$.

To do the maximization over h, let $\{\mathbf{x}_{i_1},\ldots,\mathbf{x}_{i_d}\}$ be the distinct design points, with $i_1<\cdots<i_d$. Suppose that $\mathbf{x}_{i_j}$ is repeated n_j times, so that $\sum_{j=1}^d n_j=n$. Define

$$\mathbf{U}=\begin{pmatrix}\mathbf{1}_{n_1} & & 0\\ & \ddots & \\ 0 & & \mathbf{1}_{n_d}\end{pmatrix},\quad \mathbf{V}_d=\begin{pmatrix}\mathbf{v}^{\mathrm{T}}(\mathbf{x}_{i_1})\\ \vdots \\ \mathbf{v}^{\mathrm{T}}(\mathbf{x}_{i_d})\end{pmatrix}. \tag{7}$$

Then the rows of $\mathbf{U}\mathbf{V}_d$ are those of $\mathbf{V}$, so that there is an $n\times n$ permutation matrix $\mathbf{Q}$ such that

$$\mathbf{V}=\mathbf{Q}\mathbf{U}\mathbf{V}_d. \tag{8}$$

Define $\tilde{\mathbf{B}}=\mathbf{B}\mathbf{Q}\mathbf{U}$ and let $\boldsymbol{\alpha}_{d\times 1}$ be an eigenvector, of unit norm, of $\tilde{\mathbf{B}}^{\mathrm{T}}\tilde{\mathbf{B}}$, belonging to the maximum eigenvalue $\lambda_{\max}(\tilde{\mathbf{B}}^{\mathrm{T}}\tilde{\mathbf{B}})$. Define a function h_* by

$$h_*(\mathbf{x}_{i_j})=\sqrt{\gamma N}\alpha_j,\quad j=1,\ldots,d,$$

$$h_*(\mathbf{u})=0\quad \text{for } \mathbf{u}\in\mathscr{S}\backslash\{\mathbf{x}_{i_1},\ldots,\mathbf{x}_{i_d}\}.$$

Note first that $h_*\in\mathscr{H}_\gamma$: it satisfies $N^{-1}\sum_{\mathbf{u}\in\mathscr{S}}h_*^2(\mathbf{u})=N^{-1}\sum_{j=1}^d h_*^2(\mathbf{x}_{i_j})=\gamma$, and since $\tilde{\mathbf{B}}^{\mathrm{T}}\tilde{\mathbf{B}}\boldsymbol{\alpha}=\lambda_{\max}\boldsymbol{\alpha}$, h_* also satisfies

$$\sum_{\mathbf{u}\in\mathscr{S}}\mathbf{v}(\mathbf{u})h_*(\mathbf{u})=\sum_{j=1}^d\mathbf{v}(\mathbf{x}_{i_j})h_*(\mathbf{x}_{i_j})=\mathbf{V}_d^{\mathrm{T}}\sqrt{\gamma N}\boldsymbol{\alpha}=\sqrt{\gamma N}\mathbf{V}_d^{\mathrm{T}}\tilde{\mathbf{B}}^{\mathrm{T}}\tilde{\mathbf{B}}\boldsymbol{\alpha}/\lambda_{\max}=\sqrt{\gamma N}\mathbf{V}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\tilde{\mathbf{B}}\boldsymbol{\alpha}/\lambda_{\max}=\mathbf{0},$$

by (8) and (2), which entails $\mathbf{B}\mathbf{V}=\mathbf{0}$. Now for any $h\in\mathscr{H}_\gamma$ we have, with $\mathbf{h}_d=(h(\mathbf{x}_{i_1}),\ldots,h(\mathbf{x}_{i_d}))^{\mathrm{T}}$,

$$\|\mathbf{B}\mathbf{h}\|^2=\mathbf{h}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{h}=\mathbf{h}_d^{\mathrm{T}}\tilde{\mathbf{B}}^{\mathrm{T}}\tilde{\mathbf{B}}\mathbf{h}_d=\mathbf{h}_d^{\mathrm{T}}\mathbf{h}_d\cdot\frac{\mathbf{h}_d^{\mathrm{T}}\tilde{\mathbf{B}}^{\mathrm{T}}\tilde{\mathbf{B}}\mathbf{h}_d}{\mathbf{h}_d^{\mathrm{T}}\mathbf{h}_d}\leqslant\gamma N\cdot\lambda_{\max}=\|\mathbf{B}\mathbf{h}_*\|^2.$$

Thus $h_*\in\mathscr{H}_\gamma$ is the maximizer:

$$\sup_{\mathscr{H}_\gamma}\|\mathbf{B}\mathbf{h}\|^2=\|\mathbf{B}\mathbf{h}_*\|^2=\lambda_{\max}\cdot\gamma N;$$

this in (6) gives (5). □

3. Computation of minimax designs

The computation of a minimax design requires us to find d distinct design points (treatments) $\mathbf{x}_{i_1},\ldots,\mathbf{x}_{i_d}$ appearing with frequencies $n_1,\ldots,n_d$, and an optimal assignment of these $n=\sum n_i$ design points into field plots. Then as indicated in Theorem 2 and its proof—see in particular (7) and (8)—the design is specified by three matrices: $\mathbf{V}_d$, containing the regressors at the d distinct design points, $\mathbf{U}$, determining the frequency with which each such design point appears, and $\mathbf{Q}$, a permutation matrix determining the assignment of design points to plots.

In this section we discuss examples in which there is no deterministic spatial trend, i.e., $p_2=0$ and $p=p_1$. Computational techniques are discussed in detail for two scenarios:

(1) the set $\{\mathbf{v}(\mathbf{x})|\mathbf{x}\in\mathscr{S})\}$ contains a basis of $\mathbb{R}^N$ and $p=N$,
(2) the set $\{\mathbf{v}(\mathbf{x})|\mathbf{x}\in\mathscr{S})\}\subset\mathbb{R}^p$ and p is much smaller than N.

Most practical field experiments fall into one of these cases. Scenario (1) applies to experiments with qualitative regressors. For example, an experiment is conducted to compare t treatments and the linear model is

$$Y_i=\mathbf{v}^{\mathrm{T}}(x_i)\boldsymbol{\theta}+\delta(\mathbf{t}_i)+\varepsilon(\mathbf{t}_i),\quad i=1,\ldots,n,$$

where x_i is the treatment label and the regressors $\mathbf{v}(x)\in\mathbf{R}^t$ are the indicators of the treatments. Then

$$\{\mathbf{v}(x)|x\in\mathscr{S})\}=\left\{\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix},\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix},\ldots,\begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}\right\}\subset\mathbf{R}^t,$$

and $p = N = t$. Scenario (2) is for experiments with quantitative regressors, and is discussed in Section 3.2.

3.1. Scenario (1)

For scenario (1), as noted in Remark 1, $\mathscr{H}_\gamma = \{0\}$. Thus the minimax design minimizes the loss function $\mathscr{L}(\alpha, \beta)$ only. Furthermore the number of distinct design points equals the number of treatments, i.e., $d = t$, and $\mathbf{Z}_d = \mathbf{I}_t$. Hence there are only two components to be determined for the minimax designs: the design frequencies $n_1, \ldots, n_t$ and an optimal assignment of design points into n plots. Define

$$\mathscr{N}_t = \{(n_1, \ldots, n_t) | n_i \geqslant 1, \sum_{i=1}^{t} n_i = n, J(n_1, \ldots, n_t) = \mathbf{0}\},$$

a class of frequencies satisfying certain constraints J. For example, if we require equal numbers of observations for all treatments, then J embodies the constraints $n_1 - n_2 = 0, \ldots, n_{t-1} - n_t = 0$. Now the minimax design is a solution to the optimization problem

$$\min_{\mathscr{N}_t, \mathbf{Q}} \mathscr{L}(\alpha, \beta) = \min_{\mathscr{N}_t, \mathbf{Q}} \operatorname{tr}[\mathbf{CZ}(\mathbf{Z}^{\mathrm{T}} \boldsymbol{\Lambda}_{\alpha,\beta}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^{\mathrm{T}} \mathbf{C}^{\mathrm{T}}].$$

For small n—$n \lesssim 9$—the number of possible choices in $\mathscr{N}_t$ is small and the number of possible choices for $\mathbf{Q}$ is not very large. A direct search can be done to find the minimax design. In fact, for each choice in $\mathscr{N}_t$, we can evaluate $\min_{\mathbf{Q}} \mathscr{L}(\alpha, \beta)$. For moderate or large n the number of possible choices in $\mathscr{N}_t$ may be still small, however, the number $n!/\prod_{i=1}^{t} n_i!$ of possible choices for $\mathbf{Q}$ can be quite large. A direct search may not be feasible, but a simulated annealing algorithm can be applied to approximate $\min_{\mathbf{Q}} \mathscr{L}(\alpha, \beta)$ for each choice in $\mathscr{N}_t$. Simulated annealing algorithms are effective in searching for optimal or robust designs; see Haines (1987), Meyer and Nachtsheim (1988), Eccleston and Chan (1998), Elliott et al. (1999), Fang and Wiens (2000) and Zhou (2001).

In general, a simulated annealing algorithm to minimize a loss function $\mathscr{L}(\xi)$ over designs ξ consists of three steps.

1. Choose an initial design, say ξ^0.
2. Provide a scheme to modify ξ^0 to obtain a new design ξ^1.
3. Define a decision rule to accept or reject the new design. If $\mathscr{L}(\xi^1) < \mathscr{L}(\xi^0)$, then ξ^1 is accepted. Otherwise, with probability $\mathrm{e}^{-(\mathscr{L}(\xi^1) - \mathscr{L}(\xi^0))/T}$ parameterized by the 'temperature' T, ξ^1 is still accepted. If ξ^1 is accepted, then ξ^0 is replaced by ξ^1.

Through a large number of iterations between Steps 2 and 3 with decreasing T, the loss function is expected to converge to at least a near-minimum value. Indeed, Hàjek (1988) established that, in a situation such as ours with a finite number of possible states, convergence to a global minimum is guaranteed if the temperature decays sufficiently slowly. In our applications of the annealing algorithm we have set $T = 0.1$ and multiplied by .9 after every 200 iterations.

To find $\min_{\mathbf{Q}} \mathscr{L}(\alpha, \beta)$, a design is an assignment of treatments to n plots. New designs can be obtained by exchanging distinct treatments at two randomly selected plots.

Example 1. Consider a rectangular layout of plots with $r = 2$ rows and $c = 4$ columns, as in Section 1. There are $t = 3$ crop varieties to be planted; without loss of generality the first is the check variety. We are interested in comparing the other two varieties with the check variety, i.e., in estimating $\mathbf{C}_0\boldsymbol{\theta}$ for the contrast matrix

$$\mathbf{C}_0 = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

Putting $\mathbf{C} = \mathbf{C}_0(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}$, so that $\mathbf{CZ} = \mathbf{C}_0$, we can apply the results in Section 2 and get $\mathscr{L}(\alpha, \beta) = \operatorname{tr}[\mathbf{C}_0(\mathbf{Z}^{\mathrm{T}} \boldsymbol{\Lambda}_{\alpha,\beta}^{-1} \mathbf{Z})^{-1} \mathbf{C}_0^{\mathrm{T}}]$ with $\boldsymbol{\Lambda}_{\alpha,\beta} = \mathbf{G}^{(0)} + \beta\mathbf{K} + \mathbf{F}^{(0)} + \alpha\mathbf{I}_n$. Since the labels 2 and 3 are randomly assigned to the test varieties, it is reasonable to impose the constraint $n_2 - n_3 = 0$. Thus $\mathscr{N}_3 = \{(n_1, n_2, n_3) | n_i \geqslant 1, \sum_{i=1}^{3} n_i = 8, n_2 - n_3 = 0\}$. In fact there are only three possible choices in $\mathscr{N}_3$, i.e., $\mathscr{N}_3 = \{(6, 1, 1), (4, 2, 2), (2, 3, 3)\}$. In the covariance matrix $\boldsymbol{\Lambda}_{\alpha,\beta}$, we take $\mathbf{K} = \mathbf{I}_n$, $\mathbf{F}^{(0)} = \sigma_f^2 \mathbf{I}_n$, and $\mathbf{G}^{(0)} = \sigma_g^2 \mathbf{G}(\rho)$, where $\mathbf{G}(\rho)$ is the correlation matrix according to some spatial

Table 1
Values of $\min_{\mathbf{Q}} \mathscr{L}(\alpha, \beta)$ in Example 1 for NN(1) correlation with $\sigma_g^2 = 1$ and various values of ρ and $\alpha + \beta + \sigma_f^2$

(n_1, n_2, n_3)	$\alpha+\beta+\sigma_f^2=1$				$\alpha+\beta+\sigma_f^2=0$	
	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$	$\rho=0.1$	$\rho=0.2$
(6,1,1)	4.58	4.49	4.39	4.27	2.14	1.84
(4,2,2)	2.87	2.74	2.61	2.47	1.23	0.94
(2,3,3)	3.23	3.12	3.00	2.87	1.43	1.12

Table 2
Values of $\min_{\mathbf{Q}} \mathscr{L}(\alpha, \beta)$ in Example 1 for MA(1) correlation with $\sigma_g^2 = 1$ and various values of ρ and $\alpha + \beta + \sigma_f^2$.

(n_1, n_2, n_3)	$\alpha+\beta+\sigma_f^2=1$				$\alpha+\beta+\sigma_f^2=0$			
	$\rho=0.1$	$\rho=0.2$	$\rho=0.3$	$\rho=0.4$	$\rho=0.1$	$\rho=0.2$	$\rho=0.3$	$\rho=0.4$
(6,1,1)	4.49	4.28	4.03	3.76	2.14	1.86	1.50	1.08
(4,2,2)	2.75	2.51	2.29	2.09	1.25	0.98	0.72	0.47
(2,3,3)	3.12	2.89	2.65	2.42	1.44	1.16	0.83	0.50

Table 3
Values of $\min_{\mathbf{Q}} \mathscr{L}(\alpha, \beta)$ in Example 2 for $\sigma_g^2 = 1$, $\alpha + \beta + \sigma_f^2 = 1$, and various values of ρ

(n_1, n_2, n_3, n_4)	NN(1)		MA(1)	
	$\rho=0.1$	$\rho=0.2$	$\rho=0.2$	$\rho=0.4$
(9,1,1,1)	6.46	6.16	6.17	5.45
(6,2,2,2)	3.70	3.37	3.43	2.92
(3,3,3,3)	3.75	3.44	3.48	2.92

process (such as NN(1) or MA(1)) with ρ as the nearest neighbour correlation. Tables 1 and 2 present some numerical results for $\min_{\mathbf{Q}} \mathscr{L}(\alpha, \beta)$ from a direct search for NN(1) and MA(1) correlations, respectively. From these tables, it is clear that the minimax design has design frequencies $(n_1, n_2, n_3) = (4, 2, 2)$ for both NN(1) and MA(1) correlation structures, and for all the ρ values examined. The optimal assignment of the treatments to 8 plots is

2	1	3	1
1	2	1	3

.

Example 2. Consider the experiment having a rectangular layout of plots with $r = 3$ rows and $c = 4$ columns. There are $t = 4$ crop varieties to be planted, the first of which is a check variety. We are interested in estimating $\mathbf{C}_0\boldsymbol{\theta}$, where

$$\mathbf{C}_0 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}.$$

In $\mathscr{N}_4$, the constraints are $n_2 - n_3 = 0$ and $n_3 - n_4 = 0$. A simulated annealing algorithm is applied to obtain $\min_{\mathbf{Q}} \mathscr{L}(\alpha, \beta) = \min_{\mathbf{Q}} \operatorname{tr}[\mathbf{C}_0(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\Lambda}_{\alpha,\beta}^{-1}\mathbf{Z})^{-1}\mathbf{C}_0^{\mathrm{T}}]$. The covariance matrix $\boldsymbol{\Lambda}_{\alpha,\beta}$ is set up in a manner similar to that in Example 1. Table 3 gives some representative results for NN(1) and MA(1) correlation structures. The minimax design has frequencies $(n_1, n_2, n_3, n_4) = (6, 2, 2, 2)$ for all cases considered and the optimal treatment assignment is

3	1	2	1
1	3	1	2
4	1	4	1

.

If instead $\alpha+\beta+\sigma_f^2=0.2$ and $(\sigma_g^2,\rho)=(1,0.24)$, then for NN(1) correlation the minimax design is

2	4	1	3
4	1	3	4
3	2	1	2

,

with $(n_1,n_2,n_3,n_4)=(3,3,3,3)$.

Remark 4. The designs of Examples 1 and 2, in the case of NN(1) correlations, depend on the inputs α, β, σ_f^2 only through their sum. This is because we have taken $\mathbf{\Lambda}_{\alpha,\beta}=\sigma_g^2\mathbf{G}(\rho)+\beta\mathbf{I}_n+\sigma_f^2\mathbf{I}_n+\alpha\mathbf{I}_n$ where $\mathbf{G}(\rho)$ is the spatial correlation matrix. Thus $\mathbf{\Lambda}_{\alpha,\beta}=(\alpha+\beta+\sigma_f^2+\sigma_g^2)\mathbf{G}(\rho')$, where $\rho'=\sigma_g^2\rho/(\alpha+\beta+\sigma_f^2+\sigma_g^2)$. This implies that

$$\mathscr{L}(\alpha,\beta)=(\alpha+\beta+\sigma_f^2+\sigma_g^2)\,\mathrm{tr}[\mathbf{C}_0(\mathbf{Z}^{\mathrm{T}}\mathbf{G}^{-1}(\rho')\mathbf{Z})^{-1}\mathbf{C}_0^{\mathrm{T}}],$$

and the minimax design coincides with the optimal design minimizing the trace of the variance matrix of the generalized least squares estimate, with a correlation of ρ'. As a consequence, our designs in some cases coincide with those obtained by Martin (1986).

3.2. Scenario (2)

For scenario (2), suppose that the element $\mathbf{u}_i\in\mathscr{S}$ is selected as a design point n_i times. Let $\mathscr{N}_n=\{(n_1,\ldots,n_N)|n_i\geqslant 0,\ \sum_{i=1}^N n_i=n\}$. A minimax design is a solution to

$$\min_{\mathbf{x}_1,\ldots,\mathbf{x}_n\in\mathscr{S}}\mathscr{L}(\alpha,\beta,\gamma)=\min_{\mathscr{N}_n}\min_{\mathbf{Q}}\mathscr{L}(\alpha,\beta,\gamma).$$

This design is determined by an optimal design frequency vector $\mathbf{n}$ in $\mathscr{N}_n$ and an optimal assignment of design points into n field plots. The number of positive frequencies in $\mathbf{n}$ equals d. If the size of $\mathscr{N}_n$ and the order of $\mathbf{Q}$ are small, a direct search is effective in finding the minimax design. However, if $\mathscr{N}_n$ is large, a simulated annealing algorithm similar to those used in Fang and Wiens (2000) and Zhou (2001) can be applied to search for the minimax design. Here are some details. Any vector in $\mathscr{N}_n$ can be an initial design $\mathbf{n}^0=(n_1^0,\ldots,n_N^0)$. To generate a new design, randomly select two positive frequencies $n_{k_1}^0$ and $n_{k_2}^0$ and one zero frequency $n_{k_3}^0$ from $\mathbf{n}^0$ and a random number B from the Bernoulli distribution with probability of success equal to the fraction of zero frequencies in $\mathbf{n}^0$. Form a new design $\mathbf{n}^1=(n_1^1,\ldots,n_N^1)$, different from $\mathbf{n}^0$ only at the three selected frequencies

$$n_{k_1}^1=n_{k_1}^0-1,$$

$$n_{k_2}^1=n_{k_2}^0+1-B,$$

$$n_{k_3}^1=n_{k_3}^0+B.$$

The loss function for the annealing algorithm for fixed $\mathbf{n}$ is $\min_{\mathbf{Q}}\mathscr{L}(\alpha,\beta,\gamma)$, which may be calculated directly if the number of possible choices of $\mathbf{Q}$ is not very large. Otherwise a secondary annealing step over $\mathbf{Q}$ as discussed in scenario (1) can be used.

Example 3. To illustrate the annealing algorithm over $\mathscr{N}_n$, we consider a field experiment with a rectangular layout of plots with $r=2$ and $c=3$ to fit a quadratic model

$$Y_i=\theta_0+\theta_1x_i+\theta_2x_i^2+\delta(\mathbf{t}_i)+\varepsilon(\mathbf{t}_i),\quad i=1,\ldots,6,$$

where the quantitative regressor x can be selected from a set $\mathscr{S}$ of possible values. We take

$$\mathscr{S}=\{u_1,\ldots,u_{20}|u_i=1+(i-1)*9/19\},$$

consisting of 20 equally spaced points in the interval [1, 10]. In $\mathscr{L}(\alpha,\beta,\gamma)$, take $\mathbf{C}=(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}$, $(\alpha,\beta,\gamma)=(0,0,20)$, $\mathbf{G}^{(0)}=\mathbf{G}(0.2)$ the spatial NN(1) correlation matrix with parameter $\rho=0.2$ and $\mathbf{F}^{(0)}=\mathbf{I}_6$. The ordered minimax design points are $\langle 10,1,5.74,1,5.74,1\rangle$; this translates into the plots as

10	1	5.74
1	5.74	1

.

Example 4. Here we consider an agronomy trial on wheat as described in Petersen (1994, p. 79). There are two factors—row spacing (x_1) ranging from 15 to 30 cm and the amount of nitrogen (x_2) ranging from 0 to 50 kg/ha. The trial is run on a field with rectangular plots with $r=2$ rows and $c=5$ columns, and the results are used to fit a quadratic model with interaction

$$y_i = \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \theta_3 x_{1i} x_{2i} + \theta_4 x_{1i}^2 + \theta_5 x_{2i}^2 + \delta(\mathbf{t}_i) + \varepsilon(\mathbf{t}_i), \quad i = 1, \ldots, n.$$

The design space $\mathscr{S}$ contains all possible combinations of four row spacings: 15, 20, 25, 30 (cm), and six levels of nitrogen: 0, 10, 20, 30, 40, 50 (kg/ha). In $\mathscr{L}(\alpha, \beta, \gamma)$, take $\mathbf{C} = (\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}$, $(\alpha, \beta, \gamma) = (0, 0, 1)$, $\mathbf{G}^{(0)} = \mathbf{G}(\rho)$ with $\rho = 0.2$ from the NN(1) correlation model and $\mathbf{F}^{(0)} = \mathbf{I}_{10}$. Using the simulated annealing algorithm the minimum loss 54.25 is obtained at the following design for (x_1, x_2):

(15,50)	(25,0)	(30,50)	(25,50)	(15,0)
(25,50)	(30,10)	(25,0)	(15,0)	(20,30)

.

4. Case study

In this section we construct designs for an experiment which, while hypothetical, is based on the details of a set of yield trials described in Sarker et al. (2001). The purpose of the trials was to improve legumes such as lentil, and Sarker et al. (2001) investigated the fitting of various models to the data. The spatial correlations, even when fitted post-design, were not included in the design considerations. As in the 1997 'advanced' trials conducted at Tel Hadya, Syria, we consider the analysis of 25 genotypes, including three controls, with three replicates each. The layout of the plots is rectangular, with three rows and 25 columns. Each row is to contain a complete replicate. Fertilizer is to be used at the rate of 50 kg P_2O_5/ha before sowing.

There is an assumed linear trend along rows, so that the fitted linear response in row r, column c using genotype i is

$$E[Y|r, c, i] = \beta c + \theta_i.$$

The covariance between two observations in rows r_1 and r_2 and columns c_1 and c_2, respectively, is taken to be

$$\sigma_f^2(I(|r_1 - r_2| + |c_1 - c_2| = 0)) + \sigma_g^2 \mathrm{e}^{-(a|r_1-r_2|+b|c_1-c_2|)}.$$

On the basis of private communications with Sarker and Erskine we have taken $\sigma_f^2 = 30,000$, $\sigma_g^2 = 5000$, $a = 1.0236$ and $b = 1.3329$. The errors associated with the observations can be viewed as the sum of a white noise process and a directional exponential spatial process. A directional exponential spatial process is also called a doubly geometric process (Martin, 1979), and is the simplest planar autoregressive process. When $a = b$ and e^{-a} is small, the NN(1) model can approximate the process well. To estimate the effects θ_i, $i = 1, \ldots, 25$, of the 25 genotypes, the minimax design can be constructed as follows.

We have $p_1=25$, $p_2=1$, $n=75$, $N=25$, $d=25$, $\mathbf{V}_d=\mathbf{I}_d$ and $\gamma=0$. Because each row contains a complete replicate, the matrix $\mathbf{U}$ is fixed with frequencies $n_1 = \cdots = n_{25} = 3$ and there is a restriction on the permutation matrix $\mathbf{Q}$. Then $\mathbf{V}=\mathbf{QU}$ and $\mathbf{Z}=(\mathbf{V} : \mathbf{w})$ with $\mathbf{w}^{\mathrm{T}}=(1, \ldots, 25, 1, \ldots, 25, 1, \ldots, 25)$. The matrix $\mathbf{C}=\mathbf{C}_0(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}$ with $\mathbf{C}_0=(\mathbf{I}_d : \mathbf{0})$. The minimax design minimizes $\mathscr{L}(\alpha, \beta, 0)=\mathrm{tr}[\mathbf{C}_0(\mathbf{Z}^{\mathrm{T}}\mathbf{\Lambda}_{\alpha,\beta}^{-1}\mathbf{Z})^{-1}\mathbf{C}_0^{\mathrm{T}}]$ over $\mathbf{Q}$ only, and $\mathbf{\Lambda}_{\alpha,\beta}=(\alpha+\beta+\sigma_f^2)\mathbf{I}_{75}+\sigma_g^2\mathbf{G}(a, b)$ with $\mathbf{G}(a, b)$ being the directional exponential spatial correlation matrix. A simulated annealing algorithm can be applied to obtain $\min_{\mathbf{Q}} \mathscr{L}(\alpha, \beta, 0)$. In this algorithm a new design is obtained by exchanging two genotypes at two randomly selected columns in a randomly selected row. With $\alpha = \beta = 5$, the minimax design, with minimum loss 3.3943×10^5, is given in Table 4.

Suppose now that the agronomist can apply different amounts x of fertilizer to these plots, for example $x \in \{40, 50, 60\}$ (kg P_2O_5/ha), and that the response can be modelled by

$$E[Y|r, c, i, x] = \beta c + \theta_0 x + \theta_i.$$

We now have $p_1 = 26$, $p_2 = 1$, $n = 75$, $N = 75$ and $\gamma > 0$. Our interest is in the estimation of $(\theta_0, \theta_1, \ldots, \theta_{25})$, and so $\mathbf{C} = \mathbf{C}_0(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}$ with $\mathbf{C}_0 = (\mathbf{I}_{p_1} \vdots \mathbf{0})$. The minimax design minimizes $\mathscr{L}(\alpha, \beta, \gamma)$ at (5) of Theorem 2, and the

Table 4
Minimax design for case study with 'column' as the only covariate

	Column											
Row		1	2	3	4	5	6	7	8	9	10	11	12
1		12	7	11	2	13	8	20	1	3	6	22	21
2		24	6	19	16	9	1	10	14	25	17	13	20
3		23	15	5	14	19	4	17	18	21	8	25	22
Row	13	14	15	16	17	18	19	20	21	22	23	24	25
1	16	4	10	18	9	24	25	15	14	7	5	23	19
2	5	18	3	12	15	22	4	11	21	23	2	7	8
3	3	2	9	10	24	11	20	16	7	1	6	13	12

Table 5
Minimax design for case study with 'fertilizer' as an additional covariate

	Column											
Row		1	2	3	4	5	6	7	8	9	10	11	12
1		8	15	1	22	19	14	12	16	9	20	24	18
		40	40	40	40	40	40	40	40	40	60	40	40
2		17	14	7	5	11	20	25	23	19	2	10	21
		40	40	40	60	40	40	60	40	60	40	60	40
3		7	6	10	1	24	12	9	18	20	21	19	3
		40	40	40	40	40	40	40	60	40	60	40	60
Row	13	14	15	16	17	18	19	20	21	22	23	24	25
1	11	5	2	25	3	17	10	13	4	7	23	6	21
	40	40	60	40	60	60	40	60	40	60	40	60	60
2	8	12	18	22	4	6	16	3	24	15	9	13	1
	60	60	40	60	40	40	60	40	60	40	60	40	60
3	13	22	4	16	14	2	25	11	17	23	5	15	8
	40	40	60	40	60	60	40	60	40	60	40	60	40

covariance matrix $\mathbf{\Lambda}_{\alpha,\beta}$ is the same as above. In the simulated annealing algorithm, with probability 0.8 a new design is obtained by exchanging two genotypes and making changes for the amount of fertilizer at two randomly selected columns in a randomly selected row. With probability 0.2, a new design is obtained by making changes only for the amount of fertilizer at two randomly selected plots. At each selected plot, the amount of fertilizer is randomly selected from {40, 50, 60}. The algorithm has been found to work well. A minimax (or near) design with loss 4.6656×10^5 for $\gamma = 1$ and with other inputs as above is given in Table 5, where genotype is indicated by numbers $1, \ldots, 25$ and the amount of fertilizer is indicated below the genotypes. Note that $x = 50$ is not chosen in the minimax design.

5. Conclusions and summary

In this article we have exhibited methods of obtaining estimates and designs for field trials, robust against various model uncertainties. It is anticipated that such procedures will be useful to experimenters who are planning field trials, or other experiments in a spatial setting, for which the response and/or spatial correlation or variance structures may be only approximately known. Robustness has been obtained through a modified kriging estimation procedure—universal kriging evaluated at the least favourable variance/covariance structures—and through the use of minimax designs. These designs are obtained by minimizing the maximum loss through simulated annealing. As well as being quick and easy to programme, the annealing approach has been demonstrated to be quite flexible in that it can be carried out subject to any special constraints dictated by the particular experimental situation.

Acknowledgements

The research of both authors is supported by the Natural Sciences and Engineering Research Council of Canada. We are grateful for the insightful and helpful comments of the referees, and for the work of the Guest Editor in putting together this Special Issue of *JSPI*.

References

Atkinson, A.C., 1996. Bayesian optimum experimental designs in biology and agriculture. Tatra Mountains Math. Publications 7, 9–22.

Cressie, N., 1993. Statistics for Spatial Data. Wiley, New York.

Cressie, N.A.C., Hawkins, D.M., 1980. Robust estimation of the variogram I. Math. Geol. 12, 115–125.

Eccleston, J., Chan, B., 1998. Design algorithms for correlated data. In: COMPSTAT 1998: Proceedings in Computational Statistics, pp. 41–52.

Elliott, L.J., Eccleston, J.A., Martin, R.J., 1999. An algorithm for the design of factorial experiments when the data are correlated. Statist. Comput. 9, 195–201.

Fagroud, M., Van Meirvenne, M., 2002. Accounting for soil spatial autocorrelation in the design of experimental trials. Soil Sci. Soc. Amer. J. 66, 1134–1142.

Fang, Z., Wiens, D.P., 2000. Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximately linear models. J. Amer. Statist. Assoc. 95, 807–818.

Genton, M.G., 2001. Robustness problems in the analysis of spatial data. In: Moore, M. (Ed.), Spatial Statistics: Methodological Aspects and Applications. Springer, New York, pp. 21–37.

Haines, L.M., 1987. The application of the annealing algorithm to the construction of the exact optimal designs for linear regression models. Technometrics 29, 439–447.

Hàjek, B., 1988. Cooling schedules for optimal annealing. Math. Oper. Res. 13, 311–329.

Hawkins, D.M., Cressie, N., 1984. Robust kriging—a proposal. Math. Geol. 16, 3–18.

Haining, R.P., 1978. The moving average model for spatial interaction. Trans. Institute British Geographers 3, 202–225.

Kiefer, J., Wynn, H.P., 1981. Optimum balanced block and latin square designs for correlated observations. Ann. Statist. 9, 737–757.

Legendre, P., Dale, M.R.T., Fortin, M.-J., Casgrain, P., Gurevitch, J., 2004. Effects of spatial structures in the results of field experiments. Ecology 85, 3202–3214.

Martin, R.J., 1979. A subclass of lattice processes applied to a problem in planar sampling. Biometrika 66, 209–217.

Martin, R.J., 1986. On the design of experiments under spatial correlation. Biometrika 73, 247–277.

Martin, R.J., 2001. Comparing and contrasting some environmental and experimental design problems. Environmetrics 12, 273–287.

Martin, R.J., Eccleston, J.A., Gleeson, A.C., 1993. Robust linear block designs for a suspected LV model. J. Statist. Plann. Inference 34, 433–450.

Meyer, R.K., Nachtsheim, C.J., 1988. Constructing exact D-optimal experimental designs by simulated annealing. Amer. J. Math. Manage. Sci. 3/4, 329–359.

Petersen, R.G., 1994. Agricultural Field Experiments: Design and Analysis. Marcel Dekker, New York.

Petraitis, P., 2001. Designing experiments that control for spatial and temporal variation. In: Methodology Paper Series of the 4th International Conference on ILTER in East Asia and Pacific Region, Ulaan-Hatgal, Mongolia, vol. 12, pp. 273–287.

Sarker, A., Singh, M., Erskine, W., 2001. Efficiency of spatial methods in yield trials in lentil. J. Agricultural Sci. Cambridge 128, 427–438.

Taplin, R.H., 1999. An analysis of two-dimensional agricultural field trials in the presence of outliers and fertility jumps. Australia New Zealand J. Statist. 41, 285–297.

Wiens, D.P., 2005a. Robustness in spatial studies I: minimax prediction. Environmetrics 16, 191–203.

Wiens, D.P., 2005b. Robustness in spatial studies II: minimax design. Environmetrics 16, 205–217.

Wu, T., Dutilleul, P., 1999. Validity and efficiency of neighbor analyses in comparison with classical complete and incomplete block analyses of field experiments. Agronomy J. 91, 721–731.

Zhou, J., 2001. Integer-valued minimax robust designs for approximately linear models with correlated errors. Comm. Statist. Theory Methods 30, 21–39.