Robust estimators and designs for field experiments

Douglas P. Wiens\textsuperscript{a,}*, Julie Zhou\textsuperscript{b}

\textsuperscript{a}Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1
\textsuperscript{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, 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 \( t_1, \ldots, t_n \). 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).
As indicated by (1) the fitted response is acknowledged to be only an approximation, in which case ensuing estimates
of the variance function may be biased. In some cases, depending on the richness of the space over which \( x \) 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
with its first order neighbours, and
\[
\begin{align*}
\rho & , \quad \rho \\
\rho & , \quad \rho \\
\rho & , \quad \rho 
\end{align*}
\]

and predictions will be biased. In some cases, depending on the richness of the space over which \( x \) varies, some
protection against this bias can be obtained through a judicious choice of the design variables.

We allow \( \rho \) only with its first order neighbours, and \( \rho \in (0, \frac{1}{2}) \). Under the MA(1) structure \( O \) has positive correlation with 12
neighbours, described by a parameter \( \gamma \in (0, \frac{1}{2}) \) 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 \( \mathcal{I} = \{t_1, \ldots, t_n\} \) of locations of field plots. At location \( t_i \) one
sets the levels of an independent variable \( x \), and observes
\[
Y_i = E[Y|x_i, t_i] + \varepsilon(t_i) + \delta(t_i), \quad i = 1, \ldots, n.
\]
Here:

(i) \( \{\varepsilon(t)|t \in \mathcal{I}\} \) is a set of uncorrelated zero-mean random errors, with variance matrix \( F = \text{diag}(f(t_1), \ldots, f(t_n)) \). We allow \( f \) to vary over a neighbourhood of a variance function \( f_0 \):
\[
\mathcal{F}_x = \left\{ f(\cdot) \mid \max_{t \in \mathcal{I}} |f(t) - f_0(t)| \leq \varepsilon, \quad f(t) \geq 0 \right\}.
\]

(ii) \( \{\delta(t)|t \in \mathcal{I}\} \) is a zero-mean, spatially correlated stochastic process (uncorrelated with \( \{\varepsilon(t)|t \in \mathcal{I}\} \)) with
covariance function \( g(t_i, t_j) = \text{cov}[(\delta(t_i), \delta(t_j))] \) and covariance matrix \( G_{n \times n} = (g(t_i, t_j))_i,j=1 \). We allow \( g \) to vary over
\[
G_\beta = \{g(\cdot, \cdot)|0 \leq G \leq G^{(0)} + \beta K\},
\]
where $G^{(0)}$ is the covariance matrix under $g_0$ and $G \succeq 0$ refers to the ordering by non-negative definiteness (n.n.d.).

The matrix $K$ is a fixed n.n.d. matrix, typically $K = I_n$ or $K = G^{(0)}$. See Remark 3 of Section 2 for extensions.

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

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

Put

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

so that

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

We allow $h$ to vary over

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

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 $t_1, \ldots, t_n$ were chosen from a larger set of possible locations, at each of which there was a vector $\mathbf{x}$ of known covariates. Optimum choices $t_1, \ldots, t_n$ 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 $\mathcal{S}$, 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 neighourhoods such as $\mathcal{F}_\beta$, $\mathcal{H}_\beta$ and $\mathcal{H}_\gamma$.

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

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

(i) $\mathbf{y} = (Y_1, \ldots, Y_n)^T$, $\theta = (\theta^T(1), \theta^T(2))^T$;
(ii) $V$ has rows $v^T(x_i)$, $W$ has rows $w^T(t_i)$, $i = 1, \ldots, n$, and $Z_{n \times p} = (V; W)$ with rows $z^T(x_i, t_i) = (v^T(x_i), w^T(t_i))$ of length $p = p_1 + p_2$;
(iii) $C$ is a fixed matrix, typically a contrast matrix;
(iv) ‘Unbiased’ means that $E[A\mathbf{y}] = C\mu$ for all $\theta$, requiring that

$$
BZ = 0 \quad \text{where} \quad B_{M \times n} = C - A.
$$

(2)
Loss is measured by the trace of the mean squared error matrix:

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

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

\[ A_{0,0} = \arg\min_A \text{tr}(\text{MSE}|x = \beta = \gamma = 0). \]

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 \textit{minimax robust} linear estimate defined by

\[ A_{x,\beta} = \arg\min_A \mathcal{L}(x, \beta), \]

where \( \mathcal{L}(x, \beta) = \mathcal{L}(x, \beta|\gamma = 0) \).

Note that this reduces to the first approach when \( x = \beta = 0 \). Here we use that, although \( G_0 \) is not the only member of \( G_0 \), 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 \( A_{x,\beta} \), and the resulting loss

\[ \mathcal{L}(x, \beta) = \text{tr}(\text{MSE}|A = A_{x,\beta}), \]

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

\[ \mathcal{L}(x, \beta, \gamma) = \sup_{\theta} \mathcal{L}(x, \beta) = \sup_{\theta} \sup_{\gamma} \text{tr}(\text{MSE}|A = A_{x,\beta}). \]

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. \textbf{Minimax linear estimates and designs}

**Theorem 1.** The minimax linear estimate \( A_{x,\beta} \) defined by (2) and (3) is given by

\[ A_{x,\beta} = CR_{x,\beta}, \]

where \( R_{x,\beta} = Z(Z^T A_{x,\beta}^{-1} Z)^{-1} Z^T A_{x,\beta}^{-1} \) and \( A_{x,\beta} = (G^{(0)} + \beta K) + (F^{(0)} + z_I n) \). Minimax loss is

\[ \mathcal{L}(x, \beta) = \text{tr}[A_{x,\beta} A_{x,\beta}^T] = \text{tr}[CZ(Z^T A_{x,\beta}^{-1} Z)^{-1} Z^T C^T]. \]

**Proof.** Since we are assuming \( \gamma = 0 \),

\[ y = Z\theta + \delta + \varepsilon, \]

where \( \delta, \varepsilon \) have elements \( \delta(t_i), \varepsilon(t_i) \), respectively. Then \( Ay - C\mu = A(\delta + \varepsilon) - BZ\theta \) and for any \( A \) we have

\[ \text{MSE} = AE[(\delta + \varepsilon)(\delta + \varepsilon)^T]A^T + (BZ\theta)(BZ\theta)^T = AGA^T + AFA^T + (BZ\theta)(BZ\theta)^T, \]

with

\[ \text{tr}(\text{MSE}) = [\text{tr}[AGA^T] + \text{tr}[AFA^T]] + \|BZ\theta\|^2. \]

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

\[ \mathcal{L}(x, \beta) = [\text{tr}[A(G^{(0)} + \beta K)A^T] + \text{tr}[A(F^{(0)} + z_I n)A^T]] + \|BZ\theta\|^2 \]

\[ = \text{tr}[AA_{x,\beta}A^T] + \|BZ\theta\|^2. \]
Theorem 2. The minimax design is obtained by choosing points
\[ \hat{x}_{\min} \]
where \( \theta \) is the generalized least squares estimate computed from \( A_{x, \beta} \).

This suggests a way to further robustify the procedure—one might replace \( \hat{x}_{\min} \) 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 2. Note that \( A_{x, \beta} y \) is, using Theorem 1, and the \( \beta \)-treatments matrix \( \hat{B}_{x, \beta} \) is described as follows. Let \( b_1, \ldots, b_n \) be the columns of \( B_{x, \beta} \). Then the \( j \)th column of \( \hat{B}_{x, \beta} \) is \( \sum_{k=1}^{n} b_k I(\xi_k = x_j) \), \( j = 1, \ldots, d \), where \( \{x_1, \ldots, x_n\} \) are the distinct design points (treatments).

Proof. We are now taking \( \gamma > 0 \), whence
\[ y = Z\theta + h + \delta + \varepsilon, \]
where \( \delta \) and \( \varepsilon \) are as in the proof of Theorem 1 and \( h \) has elements \( h(x_i) \). Then \( A_{x, \beta} y = A_{x, \beta} (Z\theta + h + \delta + \varepsilon) - BZ\theta - Bh \). For any estimator satisfying (2) we have
\[
\begin{align*}
\text{tr}(\text{MSE}) &= \text{tr} \{AE[(\delta + \varepsilon)(\delta + \varepsilon)^T]\} + B(Z\theta + h)(Z\theta + h)^T B^T \\
&= \text{tr} \{AGA^T + AFA^T\} + \|Bh\|^2.
\end{align*}
\]
Thus the maximum loss incurred by using the minimax estimator \( A_{x, \beta} y \) is, using Theorem 1,
\[
\begin{align*}
\sup_{\gamma} \sup \left\{ \text{tr}(A_{x, \beta} y - C\mu)(A_{x, \beta} y - C\mu)^T \right\} &= \sup \left\{ \text{tr}(A_{x, \beta} y - C\mu)(A_{x, \beta} y - C\mu)^T \right\} + \sup \|Bh\|^2 \\
&= \text{tr} \{C(Z\theta + h)^T \} + \sup \|Bh\|^2,
\end{align*}
\]
where now \( B = B_{x, \beta} = C - A_{x, \beta} = C(I_n - R_{x, \beta}) \).

To do the maximization over \( h \), let \( \{x_{i_1}, \ldots, x_{i_d}\} \) be the distinct design points, with \( i_1 < \cdots < i_d \). Suppose that \( x_{i_j} \) is repeated \( n_j \) times, so that \( \sum_{j=1}^{d} n_j = n \). Define
\[
\begin{align*}
\mathcal{A}(z, \beta, \gamma) &= \min \left\{ \text{tr}[CZ(Z\theta + h)^T C^T] + \sup \|Bh\|^2 \right\} \quad \text{for} \sup \|Bh\|^2
\end{align*}
\]
Thus he then the rows of \( n \times d \) regressors. For example, an experiment is conducted to compare \( x_i \) where

\[
\begin{align*}
U = \begin{pmatrix}
1_{n_1} & 0 \\
\vdots & \ddots \\
0 & 1_{n_d}
\end{pmatrix}, & \quad V_d = \begin{pmatrix}
\mathbf{v}^T(x_{i_1}) \\
\vdots \\
\mathbf{v}^T(x_{i_d})
\end{pmatrix}.
\end{align*}
\]

(7)

Then the rows of \( UV_d \) are those of \( V \), so that there is an \( n \times n \) permutation matrix \( Q \) such that

\[
V = QUV_d.
\]

(8)

Define \( B = BQU \) and let \( z_{d \times 1} \) be an eigenvector, of unit norm, of \( B^T \tilde{B} \), belonging to the maximum eigenvalue \( \lambda_{\text{max}}(B^T \tilde{B}) \). Define a function \( h_* \) by

\[
h_*(x_j) = \sqrt{\gamma N} z_j, \quad j = 1, \ldots, d,
\]

\[
h_*(u) = 0 \quad \text{for} \quad u \in \mathcal{S} \setminus \{x_{i_1}, \ldots, x_{i_d}\}.
\]

Note first that \( h_* \in \mathcal{K}_\gamma \); it satisfies \( N^{-1} \sum_{u \in \mathcal{S}} h_*^2(u) = N^{-1} \sum_{j=1}^d h_*^2(x_j) = \gamma \), and since \( B^T \tilde{B} z = \lambda_{\text{max}} z, h_* \) also satisfies

\[
\sum_{u \in \mathcal{S}} v(u) h_*(u) = \sum_{j=1}^d v(x_j) h_*(x_j) = V_d^T \sqrt{\gamma N} z = \sqrt{\gamma N} V_d^T B^T \tilde{B} z / \lambda_{\text{max}} = \sqrt{\gamma N} V_d^T B^T \tilde{B} z / \lambda_{\text{max}} = 0,
\]

by (8) and (2), which entails \( B V = 0 \). Now for any \( h \in \mathcal{K}_\gamma \) we have, with \( h_d = (h(x_{i_1}), \ldots, h(x_{i_d}))^T \),

\[
\|Bh\|^2 = h^T B^T B h = h_d^T \tilde{B} \tilde{B} h_d = h_d^T h_d \cdot \frac{h_d^T \tilde{B} \tilde{B} h_d}{h_d^T h_d} \leq \gamma N \cdot \lambda_{\text{max}} = \|Bh_*\|^2.
\]

Thus \( h_* \in \mathcal{K}_\gamma \) is the maximizer:

\[
\sup_{\mathcal{K}_\gamma} \|Bh\|^2 = \|Bh_*\|^2 = \lambda_{\text{max}} \cdot \gamma N;
\]

this in (6) gives (5). \( \square \)

3. Computation of minimax designs

The computation of a minimax design requires us to find \( d \) distinct design points (treatments) \( x_{i_1}, \ldots, x_{i_d} \) 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: \( V_d \), containing the regressors at the \( d \) distinct design points, \( U \), determining the frequency with which each such design point appears, and \( Q \), 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 \( \{v(x) | x \in \mathcal{S}\} \) contains a basis of \( \mathbb{R}^N \) and \( p = N \),
2. the set \( \{v(x) | x \in \mathcal{S}\} \subset \mathbb{R}^p \) 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

\[
Y_i = v^T(x_i) \theta + \delta(t_i) + \epsilon(t_i), \quad i = 1, \ldots, n,
\]

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

\[
\{v(x) | x \in \mathcal{S}\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \right\} \subset \mathbb{R}^t.
\]
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, \( \mathcal{H}_T = \{0\} \). Thus the minimax design minimizes the loss function \( \mathcal{L}(\alpha, \beta) \) only. Furthermore the number of distinct design points equals the number of treatments, i.e., \( d = t \), and \( Z_d = I_t \). 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

\[
\mathcal{N}_t = \{(n_1, \ldots, n_t) | n_t \geq 1, \sum_{i=1}^{t} n_i = n, J(n_1, \ldots, n_t) = 0\},
\]

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

\[
\min_{\mathcal{N}_t, Q} \mathcal{L}(\alpha, \beta) = \min_{\mathcal{N}_t, Q} \text{tr}[CZ(Z^T\alpha^{-1}Z)^{-1}Z^TC^T].
\]

For small \( n - n \leq 9 \)—the number of possible choices in \( \mathcal{N}_t \) is small and the number of possible choices for \( Q \) is not very large. A direct search can be done to find the minimax design. In fact, for each choice in \( \mathcal{N}_t \), we can evaluate \( \min_{Q} \mathcal{L}(\alpha, \beta) \). For moderate or large \( n \) the number of possible choices in \( \mathcal{N}_t \) may be still small, however, the number \( n! / \prod_{i=1}^{t} n_i! \) of possible choices for \( Q \) can be quite large. A direct search may not be feasible, but a simulated annealing algorithm can be applied to approximate \( \min_{Q} \mathcal{L}(\alpha, \beta) \) for each choice in \( \mathcal{N}_t \). 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 \( \mathcal{L}(\xi) \) over designs \( \xi \) consists of three steps.

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

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, Hájek (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 \( \min_{Q} \mathcal{L}(\alpha, \beta) \), 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 \( C_0\theta \) for the contrast matrix

\[
C_0 = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.
\]

Putting \( C = C_0(Z^TZ)^{-1}Z^T \), so that \( CZ = C_0 \), we can apply the results in Section 2 and get \( \mathcal{L}(\alpha, \beta) = \text{tr}[C_0(Z^T\alpha^{-1}Z)^{-1}C_0^T] \) with \( \alpha_{x, \beta} = G^{(0)} + \beta K + F^{(0)} + zI_n. \) 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 \( \mathcal{A}_3 = \{(n_1, n_2, n_3) | n_1 \geq 1, \sum_{i=1}^{3} n_i = 8, n_2 - n_3 = 0 \} \). In fact there are only three possible choices in \( \mathcal{A}_3 \), i.e., \( \mathcal{A}_3 = \{(6, 1, 1), (4, 2, 2), (2, 3, 3)\} \). In the covariance matrix \( \alpha_{x, \beta} \), we take \( K = I_n, F^{(0)} = \sigma^2_f I_n, \) and \( G^{(0)} = \sigma^2_g G(\rho), \) where \( G(\rho) \) is the correlation matrix according to some spatial
Table 1
Values of $\min_{\mathbf{Q}} \mathcal{L}(x, \beta)$ in Example 1 for NN(1) correlation with $\sigma^2 = 1$ and various values of $\rho$ and $x + \beta + \sigma^2$

<table>
<thead>
<tr>
<th>$(n_1, n_2, n_3)$</th>
<th>$x + \beta + \sigma^2 = 1$</th>
<th>$x + \beta + \sigma^2 = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho = 0.05$</td>
<td>$\rho = 0.10$</td>
</tr>
<tr>
<td>(6,1,1)</td>
<td>4.58</td>
<td>4.49</td>
</tr>
<tr>
<td>(4,2,2)</td>
<td>2.87</td>
<td>2.74</td>
</tr>
<tr>
<td>(2,3,3)</td>
<td>3.23</td>
<td>3.12</td>
</tr>
</tbody>
</table>

Table 2
Values of $\min_{\mathbf{Q}} \mathcal{L}(x, \beta)$ in Example 1 for MA(1) correlation with $\sigma^2 = 1$ and various values of $\rho$ and $x + \beta + \sigma^2$.

<table>
<thead>
<tr>
<th>$(n_1, n_2, n_3)$</th>
<th>$x + \beta + \sigma^2 = 1$</th>
<th>$x + \beta + \sigma^2 = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho = 0.1$</td>
<td>$\rho = 0.2$</td>
</tr>
<tr>
<td>(6,1,1)</td>
<td>4.49</td>
<td>4.28</td>
</tr>
<tr>
<td>(4,2,2)</td>
<td>2.75</td>
<td>2.51</td>
</tr>
<tr>
<td>(2,3,3)</td>
<td>3.12</td>
<td>2.89</td>
</tr>
</tbody>
</table>

Table 3
Values of $\min_{\mathbf{Q}} \mathcal{L}(x, \beta)$ in Example 2 for $\sigma^2 = 1$, $x + \beta + \sigma^2 = 1$, and various values of $\rho$.

<table>
<thead>
<tr>
<th>$(n_1, n_2, n_3, n_4)$</th>
<th>NN(1)</th>
<th>MA(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho = 0.1$</td>
<td>$\rho = 0.2$</td>
</tr>
<tr>
<td>(9,1,1,1)</td>
<td>6.46</td>
<td>6.16</td>
</tr>
<tr>
<td>(6,2,2,2)</td>
<td>3.70</td>
<td>3.37</td>
</tr>
<tr>
<td>(3,3,3,3)</td>
<td>3.75</td>
<td>3.44</td>
</tr>
</tbody>
</table>

process (such as NN(1) or MA(1)) with $\rho$ as the nearest neighbour correlation. Tables 1 and 2 present some numerical results for $\min_{\mathbf{Q}} \mathcal{L}(x, \beta)$ 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 $\rho$ values examined. The optimal assignment of the treatments to 8 plots is

$$
\begin{bmatrix}
2 & 1 & 3 & 1 \\
1 & 2 & 1 & 3
\end{bmatrix}
$$

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 $\mathbf{C}_0 \theta$, where

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

In $\mathcal{N}_4$, the constraints are $n_2 - n_3 = 0$ and $n_3 - n_4 = 0$. A simulated annealing algorithm is applied to obtain $\min_{\mathbf{Q}} \mathcal{L}(x, \beta) = \min_{\mathbf{Q}} \text{tr}[\mathbf{C}_0 (\mathbf{Z}^T \mathbf{L}_{x, \beta}^{-1} \mathbf{Z})^{-1} \mathbf{C}_0^T]$. The covariance matrix $\mathbf{L}_{x, \beta}$ 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

$$
\begin{bmatrix}
3 & 1 & 2 & 1 \\
1 & 3 & 1 & 2 \\
4 & 1 & 4 & 1
\end{bmatrix}.
$$
If instead \( x + \beta + \sigma_{ij}^2 = 0.2 \) and \( (\sigma_{ij}^2, \rho) = (1, 0.24) \), then for NN(1) correlation the minimax design is

\[
\begin{bmatrix}
2 & 4 & 1 & 3 \\
4 & 1 & 3 & 4 \\
3 & 2 & 1 & 2
\end{bmatrix},
\]

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 \( x, \beta, \sigma_{ij}^2 \) only through their sum. This is because we have taken \( \Lambda_{x,\beta} = \sigma_{ij}^2 G(\rho) + \beta I_n + \sigma_{ij}^2 I_n + z I_n \) where \( G(\rho) \) is the spatial correlation matrix. Thus \( \Lambda_{x,\beta} = (x + \beta + \sigma_{ij}^2 + \sigma_{ij}^2)G(\rho') \), where \( \rho' = \sigma_{ij}^2 \rho / (x + \beta + \sigma_{ij}^2 + \sigma_{ij}^2) \). This implies that

\[
\mathcal{L}(x, \beta) = (x + \beta + \sigma_{ij}^2 + \sigma_{ij}^2) \text{tr}[C_0(Z^T G^{-1}(\rho')Z)^{-1}C_0^T],
\]

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 \( \rho' \). 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 \( u_i \in \mathcal{S} \) is selected as a design point \( n_i \) times. Let \( \mathcal{N}_n = \{(n_1, \ldots, n_N) | n_i \geq 0, \sum_{i=1}^N n_i = n\} \). A minimax design is a solution to

\[
\min_{x_1, \ldots, x_0 \in \mathcal{S}} \mathcal{L}(x, \beta, \gamma) = \min_{\mathcal{N}_n} \min_{Q} \mathcal{L}(x, \beta, \gamma).
\]

This design is determined by an optimal design frequency vector \( n \) in \( \mathcal{N}_n \) and an optimal assignment of design points into \( n \) field plots. The number of positive frequencies in \( n \) equals \( d \). If the size of \( \mathcal{N}_n \) and the order of \( Q \) are small, a direct search is effective in finding the minimax design. However, if \( \mathcal{N}_n \) 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 \( \mathcal{N}_n \) can be an initial design \( n^0 = (n_1^0, \ldots, n_N^0) \). To generate a new design, randomly select two positive frequencies \( n_{k1}^0 \) and \( n_{k2}^0 \) and one zero frequency \( n_{k3}^0 \) from \( n^0 \) and a random number \( B \) from the Bernoulli distribution with probability of success equal to the fraction of zero frequencies in \( n^0 \). Form a new design \( n^1 = (n_1^1, \ldots, n_N^1) \), different from \( n^0 \) only at the three selected frequencies

\[
\begin{align*}
n_{k1}^1 &= n_{k1}^0 - 1, \\
n_{k2}^1 &= n_{k2}^0 + 1 - B, \\
n_{k3}^1 &= n_{k3}^0 + B.
\end{align*}
\]

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

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

\[
Y_i = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \delta(t_i) + \varepsilon(t_i), \quad i = 1, \ldots, 6,
\]

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

\[
\mathcal{S} = \{u_1, \ldots, u_{20} | u_j = 1 + (i - 1) \times 9/19\},
\]

consisting of 20 equally spaced points in the interval [1, 10]. In \( \mathcal{S}(x, \beta, \gamma) \), take \( C = (Z^T Z)^{-1} Z^T \), \( (x, \beta, \gamma) = (0, 0, 20), G^{(0)} = G(0.2) \) the spatial NN(1) correlation matrix with parameter \( \rho = 0.2 \) and \( F^{(0)} = I_n \). The ordered minimax design points are \((10, 1, 5.74, 1, 5.74, 1)\); this translates into the plots as...
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

\[
y_i = \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \theta_3 x_{1i} x_{2i} + \theta_4 x_{1i}^2 + \theta_5 x_{2i}^2 + \delta(t_i) + \epsilon(t_i), \quad i = 1, \ldots, n.
\]

The design space \(\mathcal{S}\) 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 \(\mathcal{L}(x, \beta, \gamma)\), take \(C = (Z^T Z)^{-1} Z^T\), \((x, \beta, \gamma) = (0, 0, 1)\), \(G^{(0)} = G(\rho)\) with \(\rho = 0.2\) from the NN(1) correlation model and \(F^{(0)} = I_{10}\). Using the simulated annealing algorithm the minimum loss 54.25 is obtained at the following design for \((x_1, x_2)\):

\[
\begin{array}{cccc}
(15,50) & (25,0) & (30,50) & (25,50) \\
(25,50) & (30,10) & (25,0) & (15,0) \\
& & (20,30) & \\
\end{array}
\]

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 P2O5/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

\[
\sigma^2_f (I(|r_1 - r_2| + |c_1 - c_2| = 0)) + \sigma^2_e e^{-(a|r_1 - r_2| + b|c_1 - c_2|)}.
\]

On the basis of private communications with Sarker and Erskine we have taken \(\sigma^2_f = 30,000\), \(\sigma^2_e = 5000\), \(a = 1.0236\) and \(b = 1.3329\). The errors associated with these 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 \(\theta_i\), \(i = 1, \ldots, 25\), of the 25 genotypes, the minimax design can be constructed as follows.

We have \(p_1 = 25\), \(p_2 = 1\), \(n = 75\), \(N = 25\), \(d = 25\), \(V_d = I_d\) and \(\gamma = 0\). Because each column contains a complete replicate, the matrix \(U\) is fixed with frequencies \(n_1 = \cdots = n_{25} = 3\) and there is a restriction on the permutation matrix \(Q\). Then \(V = QU\) and \(Z = (V : w)\) with \(w^T = (I, \ldots, 25, 1, \ldots, 25, 1, \ldots, 25)\). The matrix \(C = C_0(Z^T Z)^{-1} Z^T\) with \(C_0 = (I_d : 0)\). The minimax design minimizes \(\mathcal{L}(x, \beta, 0) = \text{tr}[C_0 (Z^T Z)^{-1} C_0] \text{tr}\) over \(Q\) only, and \(A_{x, \beta} = (x + \beta + \sigma^2 F)I_{15} + \sigma^2 G(a, b)\) with \(G(a, b)\) being the directional exponential spatial correlation matrix. A simulated annealing algorithm can be applied to obtain \(\min_Q \mathcal{L}(x, \beta, 0)\). In this algorithm a new design is obtained by exchanging two genotypes at two randomly selected columns in a randomly selected row. With \(x = \beta = 5\), the minimax design, with minimum loss 3.3943 \times 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 P2O5/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 \(C = C_0(Z^T Z)^{-1} Z^T\) with \(C_0 = (I_{p_1}, 0)\). The minimax design minimizes \(\mathcal{L}(x, \beta, 0)\) at (9) of Theorem 2, and the
covariance matrix $\Lambda_{x,\beta}$ 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 $4.6656 \times 10^5$ for $\gamma = 1$ and with other inputs as above is given in Table 5, where genotype is indicated by numbers 1,..., 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


