NOTES: MINIMAX PREDICTION DESIGNS, ROBUST AGAINST MISSPECIFIED RESPONSE AND ERROR STRUCTURES

Douglas P. Wiens[1]

July 2, 2019

Abstract

Key words and phrases

1 Introduction

Some motivation:

Example 1: Model robust design of experiments. There is a considerable literature on this problem under independence. Numerous authors (...) have studied design problems with correlated errors and a variety of choices of correlation functions, but with the fitted regression model generally assumed to be correct. In this case the interpretation is that the estimation is carried out assuming that $E[Y(\boldsymbol{x})] = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta}$ exactly, and that any departures are due to randomness. The estimate $\hat{\boldsymbol{\theta}}$ is then computed by least squares, and $\hat{Y}(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\hat{\boldsymbol{\theta}}$. The experimenter should however aim for some protection against errors in these assumptions.

Example 2: Spatial design. Here, the covariates $\boldsymbol{x}$ will contain information about the physical locations, and observations at different locations are expected to be correlated. The predictions $\hat{Y}(\boldsymbol{x})$ are obtained by (universal) kriging; this presupposes prior knowledge of, or some model for the covariance function $\text{COV}[Y(\boldsymbol{x}), Y(\boldsymbol{x}')]$ and for the error variance $\sigma_\varepsilon^2 = \text{VAR}[\varepsilon(\boldsymbol{x})]$. The 'classical' spatial design problem, in which the choices of these functions, and of the fitted regression model, are assumed to be exactly correct, is well studied – ... . An experimenter wishing robustness against misspecifications in these choices might seek predictions that minimize the prediction mean squared error (PMSE) in neighbourhoods of the assumed values; see Wiens (2005).

Example 3: Computer experimentation. In this context $\boldsymbol{x}$ is often viewed as a vector of control variables. Correlation models commonly used are as for those in spatial design, but with $\sigma_\varepsilon^2 = 0$, reflecting the deterministic nature of the response. See Santner, Williams and Notz (2003), Chen, Loeppky, Sacks and Welch (2016), ...

[1]Department of Mathematical and Statistical Sciences; University of Alberta, Edmonton, Alberta; Canada T6G 2G1. e-mail: doug.wiens@ualberta.ca

2 Problem formulation

We consider design problems in 'approximately linear' models, for which the response variable Y is observed subject to random error at covariates $\boldsymbol{x} \in \chi \subset \mathbb{R}^q$, for a finite 'design space' $\chi = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$. The response has 'approximate' mean $\boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta}$ for p-dimensional regressors $\boldsymbol{f}(\boldsymbol{x})$ and some unknown parameter $\boldsymbol{\theta}_{p\times 1}$.

More precisely, we consider a collection $\{Y_i = Y(\boldsymbol{x}_i)\}_{i=1}^N$ of r.v.s, satisfying

$$E[Y(\boldsymbol{x}_i)] = \boldsymbol{f}'(\boldsymbol{x}_i)\boldsymbol{\theta} + \psi(\boldsymbol{x}_i), i = 1, ..., N, \tag{1}$$

for a function $\psi(\cdot)$ quantifying the approximate nature of the experimenter's, possibly incorrect, model assumption that $E[Y(\boldsymbol{x})] = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta}$. For identifiability we define

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\eta}} \sum_{\boldsymbol{x}\in\chi} \left(E[Y(\boldsymbol{x})] - \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\eta}\right)^2.$$

Defining $\psi(\boldsymbol{x}) = E[Y(\boldsymbol{x})] - \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta}$ then leads to (1) and to the orthogonality requirement

$$\mathbf{F}'\boldsymbol{\psi}_N = \mathbf{0}_{p\times 1}; \tag{2}$$

here $\boldsymbol{\psi}_N$ is the $N \times 1$ vector with elements $\{\psi(\boldsymbol{x}_i)\}_{i=1}^N$ and $\mathbf{F}$ is the $N \times p$ matrix with rows $\{\boldsymbol{f}'(\boldsymbol{x}_i)\}_{i=1}^N$. We assume that $\mathbf{F}$ has full column rank.

The Y_i are possibly dependent; with $\boldsymbol{Y} = (Y_1, ..., Y_N)'$ we represent the covariance matrix as

$$\boldsymbol{C}_N = \textsc{cov}[\boldsymbol{Y}] \stackrel{def}{=} (\sigma_{ij})_{i,j=1,...,N}.$$

The intention is to predict unobserved values of $\{Y_i\}$. To this end, for each i we may observe $n_i \geq 0$ copies $\{Y_{i_k}(\boldsymbol{x}_i)\}_{i_k=1}^{n_i}$ distributed in the same manner as Y_i, but observed with i.i.d. measurement error with common variance σ_ε^2.

We suppose that $\textsc{cov}[Y_{i_k}, Y_j] = \sigma_{ij}$ and that

$$\textsc{cov}[Y_{i_k}, Y_{j_l}] = \sigma_{ij} + \begin{cases} \sigma_\varepsilon^2, & (i,k) = (j,l), \\ 0, & \text{otherwise.} \end{cases}$$

(So, in particular, the variance of one observation at $\boldsymbol{x}_i$ is $\textsc{var}[Y(\boldsymbol{x}_i)] = \sigma_{ii} + \sigma_\varepsilon^2$.) We impose the bound

$$\|\boldsymbol{\psi}_N\|^2 \leq \alpha^2/n, \tag{3}$$

where n is the study size and $\|\cdot\|$ is the Euclidean norm. For an induced matrix norm $\|\cdot\|_M$ we impose a bound

$$\|\boldsymbol{C}_N\|_M \leq \beta^2/n. \tag{4}$$

In both (3) and (4) the dependence on n is to ensure that, in cases where replication is possible, the various components of the loss have the same asymptotic order. A common choice of matrix norm is the spectral radius, which for symmetric matrices is the maximum eigenvalue $\|\boldsymbol{C}_N\|_M = ch_{\max}(\boldsymbol{C}_N)$. In this case the interpretation is that the variance of a linear combination $\sum_{i=1}^N t_i Y_i$, with $\sum_{i=1}^N t_i^2 = 1$, is bounded by β^2/n.

We note that these dependencies on n are only for the asymptotics; for finite n they can be absorbed into $\boldsymbol{\psi}_N$ and $\boldsymbol{C}_N$.

We define Ψ to be the class of functions $\psi(\cdot)$ satisfying (2) and (3), and $\mathcal{C}$ to be the class of positive semi-definite matrices satisfying (4).

Let $\boldsymbol{y}$ be the $n \times 1$ vector of observations with subvectors $\{\boldsymbol{y}_i\}_{n_i>0}$, where, when $n_i > 0$, $\boldsymbol{y}_i = \left(Y_{i_1}, ..., Y_{i_{n_i}}\right)'$. Then $n = \sum n_i$ and $E[\boldsymbol{y}] = \boldsymbol{X}_{n\times p}\boldsymbol{\theta}$ for

$$\boldsymbol{X}_{n\times p} = \begin{pmatrix} \mathbf{1}_{n_1}\boldsymbol{f}'(\boldsymbol{x}_1) \\ \vdots \\ \mathbf{1}_{n_N}\boldsymbol{f}'(\boldsymbol{x}_N) \end{pmatrix}_{n_i>0}.$$

The covariances between $\boldsymbol{Y}$ and the data form the matrix

$$\boldsymbol{C}_{N,n} = \text{COV}[\boldsymbol{Y}, \boldsymbol{y}'] : N \times n;$$

this has i^{th} row

$$\text{COV}[Y_i, \boldsymbol{y}'] = \left(\sigma_{i1}\mathbf{1}'_{n_1}, ..., \sigma_{ii}\mathbf{1}'_{n_i}, ..., \sigma_{iN}\mathbf{1}'_{n_N}\right),$$

with the understanding that if $n_j = 0$ then the j^{th} block is absent.

[The following is crucial. As I say below, the matrix E determines the design, so that the problem is to determine it – equivalently the weights $\{\xi_i = n_i/n\}$ – so as the minimize the maximum loss at (11) or (12).]

It is useful to introduce the incidence matrix $\boldsymbol{E}_{N\times n} = (e_{ij})$, with

$$e_{ij} = I\left(\text{the } j^{th} \text{ element of } \boldsymbol{y} \text{ is observed at } \boldsymbol{x}_i\right).$$

An alternate expression is

$$\boldsymbol{E} = \begin{pmatrix} \boldsymbol{e}'_1 \\ \vdots \\ \boldsymbol{e}'_N \end{pmatrix}, \text{ where } \boldsymbol{e}'_i = \begin{cases} \mathbf{0}'_{1\times n}, & n_i = 0, \\ \left(\mathbf{0}'_{\sum_{j<i} n_j} \vdots \mathbf{1}'_{n_i} \vdots \mathbf{0}'_{\sum_{j>i} n_j}\right), & n_i > 0. \end{cases}$$

In this notation

$$\boldsymbol{X} = \boldsymbol{E}'\boldsymbol{F} \text{ and } \boldsymbol{C}_{N,n} = \boldsymbol{C}_N\boldsymbol{E}. \tag{5}$$

The matrix $\boldsymbol{E}$ determines the design: if $\{\xi_i = n_i/n\}$ are the design weights – the proportion of observations made at $\boldsymbol{x}_i$ – then

$$\boldsymbol{D}_\xi \stackrel{def}{=} \oplus_{i=1}^N \xi_i = n^{-1}\boldsymbol{E}\boldsymbol{E}'. \tag{6}$$

Note also that $\boldsymbol{E}'\,\boldsymbol{E} = \oplus_{n_i>0}\mathbf{1}_{n_i}\mathbf{1}_{n_i}' : n \times n$.

We define the covariance matrix of the data by

$$\boldsymbol{C}_n = \text{COV}\left[\boldsymbol{y}\right] : n \times n.$$

This is a block matrix in which most blocks – those corresponding to $n_i = 0$ or $n_j = 0$ – are typically absent; if $n_i, n_j > 0$ the $(i,j)^{th}$ block is

$$\boldsymbol{C}_{n,ij} = \text{COV}\left[\boldsymbol{y}_i, \boldsymbol{y}_j'\right] = \sigma_{ij}\mathbf{1}_{n_i}\mathbf{1}_{n_j}' + \begin{cases} \sigma_\varepsilon^2 \boldsymbol{I}_{n_i}, & i = j, \\ 0, & i \neq j. \end{cases}$$

Thus

$$\boldsymbol{C}_n = \boldsymbol{E}'\boldsymbol{C}_{\boldsymbol{N}}\boldsymbol{E} + \sigma_\varepsilon^2 \boldsymbol{I}_n. \tag{7}$$

We suppose that the investigator computes estimates and inferences under the assumption that $\psi \equiv 0$, so that $E\left[Y\left(\boldsymbol{x}\right)\right] = \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta}$, and the assumption that the true covariance matrix is a particular member $\boldsymbol{C}_{0;N}$ of $\mathcal{C}$.

The investigator seeks a set of linear predictors $\hat{\boldsymbol{Y}} = \boldsymbol{L}_0\boldsymbol{y}$ of $\boldsymbol{Y} = (Y_1, ..., Y_N)'$ that are unbiased – $E\left[\hat{\boldsymbol{Y}}\right] = E\left[\boldsymbol{Y}\right]$ – and that minimize the prediction mean squared error (PMSE), defined as

$$\text{PMSE} = \sum E\left[\left(Y_i - \hat{Y}_i\right)^2\right] = E\left(\left\|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right\|^2\right).$$

The solution to this problem is the universal kriging estimate – see, e.g. Cressie (1993) – given in part (i) in the following theorem, with further details in part (ii). Part (iii) gives the PMSE under the general mean/covariance structures discussed above. In §3 this PMSE is expressed explicitly in terms of the design, for various cases. In Theorem 2 we give the maximum of this PMSE, over Ψ and $\mathcal{C}$.

Theorem 1 *(i) The linear predictors* $\hat{\boldsymbol{Y}} = \boldsymbol{L}_0\boldsymbol{y}$ *minimizing the* PMSE *(under the experimenter's model assumptions) are given by* $\boldsymbol{L}_0 = \boldsymbol{L}_1 + \boldsymbol{L}_2$*, with*

$$\begin{aligned} \boldsymbol{L}_1 &= \boldsymbol{F}\left(\boldsymbol{X}'\boldsymbol{C}_{0;n}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{C}_{0;n}^{-1}, \\ \boldsymbol{L}_2 &= \boldsymbol{C}_{0;N,n}\boldsymbol{C}_{0;n}^{-1}\left[\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{C}_{0;n}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{C}_{0;n}^{-1}\right]; \end{aligned}$$

thus

$$\hat{\boldsymbol{Y}} = \boldsymbol{F}\hat{\boldsymbol{\theta}}_0 + \boldsymbol{C}_{0;N,n}\boldsymbol{C}_{0;n}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\theta}}_0\right),$$

where $\hat{\boldsymbol{\theta}}_0 = \left(\boldsymbol{X}'\boldsymbol{C}_{0;n}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{C}_{0;n}^{-1}\boldsymbol{y}$ *is the generalized least squares estimator.*

(ii) Define $\boldsymbol{V}_0 = \boldsymbol{E}\boldsymbol{C}_{0;n}^{-1}\boldsymbol{E}' = \boldsymbol{E}\left(\boldsymbol{E}'\boldsymbol{C}_{0;N}\boldsymbol{E} + \sigma_\varepsilon^2\boldsymbol{I}_n\right)^{-1}\boldsymbol{E}'$ *and* $\boldsymbol{A}_0 = \boldsymbol{I}_N - \boldsymbol{L}_0\boldsymbol{E}'$*. Then*

$$\begin{aligned} \boldsymbol{L}_0 &= \left[\left(\boldsymbol{I}_N - \boldsymbol{C}_{0;N}\boldsymbol{V}_0\right)\boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{V}_0\boldsymbol{F}\right)^{-1}\boldsymbol{F}' + \boldsymbol{C}_{0;N}\right]\boldsymbol{E}\boldsymbol{C}_{0;n}^{-1}, \\ \boldsymbol{A}_0 &= \left(\boldsymbol{I}_N - \boldsymbol{C}_{0;N}\boldsymbol{V}_0\right)\left(\boldsymbol{I}_N - \boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{V}_0\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\boldsymbol{V}_0\right). \end{aligned}$$

Note that

$$\boldsymbol{A}_0\boldsymbol{F} = \boldsymbol{0}_{N\times p}. \tag{8}$$

(iii) If ψ *and* $\text{COV}[\mathbf{Y}] = \boldsymbol{C}_N$ *are arbitrary members of* Ψ *and* $\mathcal{C}$ *respectively, then*

$$\text{PMSE} = \left\|\boldsymbol{A}_0\boldsymbol{\psi}_N\right\|^2 + tr\left\{\boldsymbol{A}_0\boldsymbol{C}_N\boldsymbol{A}_0'\right\} + \sigma_\varepsilon^2 tr\left\{\boldsymbol{L}_0\boldsymbol{L}_0'\right\}. \tag{9}$$

Thus, under the model assumptions, the minimized PMSE *is*

$$\text{PMSE}_0 = tr\left\{\boldsymbol{A}_0\boldsymbol{C}_{0;N}\boldsymbol{A}_0'\right\} + \sigma_\varepsilon^2 tr\left\{\boldsymbol{L}_0\boldsymbol{L}_0'\right\}.$$

The maximum value of PMSE at (9), over Ψ and $\mathcal{C}$, is developed in the following two lemmas and summarized in Theorem 2. The proof of Lemma 1 is given in the Appendix. Lemma 2 is proved in Welsh and Wiens (2013), but was previously noted in Wiens and Zhou (2008).

Lemma 1 *For an* $N \times N$ *matrix* $\mathbf{A}$ *satisfying* $\boldsymbol{AF} = \boldsymbol{0}_{N\times p}$*, the maximum of* $\boldsymbol{\psi}_N'\mathbf{A}'\boldsymbol{A}\boldsymbol{\psi}_N$ *over* $\boldsymbol{\psi}_N$*, subject to (2) and (3), is*

$$\left(\alpha^2/n\right)\cdot ch_{\max}\left\{\boldsymbol{A}\boldsymbol{A}'\right\}. \tag{10}$$

Lemma 2 *Suppose that* $\mathcal{L}(\boldsymbol{C})$ *is a function of positive semi-definite matrices* $\boldsymbol{C}_{N\times N}$ *which is monotonic with respect to the ordering by positive semi-definiteness, in that* $\boldsymbol{C}_1 \geq \boldsymbol{C}_2 \Rightarrow \mathcal{L}(\boldsymbol{C}_1) \geq \mathcal{L}(\boldsymbol{C}_2)$*. For any induced matrix norm* $\|\cdot\|$*, define a class of matrices*

$$\mathcal{C} = \left\{\boldsymbol{C} \mid \boldsymbol{C} \text{ positive semi-definite and } \|\boldsymbol{C}\| \leq \tau^2\right\},$$

and define the class

$$\mathcal{C}' = \left\{\boldsymbol{C} \mid \boldsymbol{C} \text{ positive semi-definite and } \boldsymbol{0} \leq \boldsymbol{C} \leq \tau^2\mathbf{I}_N\right\}.$$

Then in all such classes

$$\max_{\mathcal{C}}\mathcal{L}\left(\boldsymbol{C}\right) = \max_{\mathcal{C}'}\mathcal{L}\left(\boldsymbol{C}\right) = \mathcal{L}\left(\tau^2\mathbf{I}_N\right).$$

An implication of Lemma 2 is that a least favourable model of dependence in this problem is in fact independence.

The following is now immediate.

Theorem 2 *Recall (8). Then the maximum, over* Ψ *and* $\mathcal{C}$*, of the* PMSE *at (9) is* $\left(\alpha^2+\beta^2+\sigma_\varepsilon^2\right)/n$ *times*

$$\mathcal{L}\left(\xi\right) = \left(1-\nu-\omega\right)ch_{\max}\boldsymbol{A}_0\boldsymbol{A}_0' + \nu tr\left\{\boldsymbol{A}_0\boldsymbol{A}_0'\right\} + \omega tr\left\{n\boldsymbol{L}_0\boldsymbol{L}_0'\right\}, \tag{11}$$

where $\nu = \beta^2/\left(\alpha^2+\beta^2+\sigma_\varepsilon^2\right)$ *and* $\omega = \sigma_\varepsilon^2/\left(\alpha^2+\beta^2+\sigma_\varepsilon^2\right)$.

Remark 1: An alternate class of covariance structures is

$$\mathcal{C}_\gamma = \left\{ (1-\gamma)\, \boldsymbol{C}_{0;N} + \gamma \boldsymbol{C}_N \mid \|\boldsymbol{C}_N\|_M \leq \beta^2/n \right\},$$

for $\gamma \in [0,1]$. For this class the theory above continues to hold, the only change being that the least favourable $\boldsymbol{C}_N$ in (9) is $(1-\gamma)\, \boldsymbol{C}_{0;N} + \left(\gamma\beta^2/n\right) \boldsymbol{I}_N$. This yields that $\max_{\Psi,\mathcal{C}_\gamma} \text{PMSE}$ is $\left(\alpha^2 + \beta^2\gamma + \sigma_\varepsilon^2 + 1 - \gamma\right)/n$ times

$$\mathcal{L}_1(\xi) = (1-a-b-c)\, ch_{\max} \boldsymbol{A}_0 \boldsymbol{A}_0' + a \cdot tr\left\{\boldsymbol{A}_0 \boldsymbol{A}_0'\right\} + b \cdot tr\left\{n \boldsymbol{A}_0 \boldsymbol{C}_{0;N} \boldsymbol{A}_0'\right\} + c \cdot tr\left\{n \boldsymbol{L}_0 \boldsymbol{L}_0'\right\}, \tag{12}$$

where $a = \beta^2\gamma / \left(\alpha^2 + \beta^2\gamma + \sigma_\varepsilon^2 + 1 - \gamma\right)$, $b = (1-\gamma) / \left(\alpha^2 + \beta^2\gamma + \sigma_\varepsilon^2 + 1 - \gamma\right)$, $c = \sigma_\varepsilon^2 / \left(\alpha^2 + \beta^2\gamma + \sigma_\varepsilon^2 + 1 - \gamma\right)$. Note that $\mathcal{L}_1(\xi) = \mathcal{L}(\xi)$ when $b = 0$. We will discuss solutions for all $a, b, c \in [0,1]$, $a + b + c \leq 1$.

Remark 2: In studying applications in which replication is possible (implying that the elements of $\boldsymbol{D}_\xi$ are $O(1)$ as $n \to \infty$), we shall take $\boldsymbol{C}_{0;N} = \left(\sigma_0^2/n\right) \boldsymbol{P}_{0;N}$, for a specified σ_0^2 and specified correlation matrix $\boldsymbol{P}_{0;N}$. One implication of this is that, if $\boldsymbol{y} = (y_1, .., y_n)'$ is a vector of observations then, in the case that $\|\cdot\|_M$ is the spectral radius,

$$\sup_{\|\boldsymbol{a}\| \leq 1} \text{VAR}\left[\boldsymbol{a}'\boldsymbol{y}\right] = ch_{\max} \boldsymbol{C}_{0;n} = \sigma_\varepsilon^2 + \sigma_0^2 ch_{\max} \boldsymbol{D}_\xi^{1/2} \boldsymbol{P}_{0;N} \boldsymbol{D}_\xi^{1/2} = O(1).$$

Further implications are that $\boldsymbol{V}_0$ is $O(n)$ so that $\boldsymbol{A}_0$ and $tr\left(n\boldsymbol{L}_0\boldsymbol{L}_0'\right)$ are $O(1)$. Thus all three components of the loss in (11) are $O(1)$. The first of these is the contribution to PMSE arising from the misspecified mean structure, the second is that arising from the dependencies, and the third is that arising from measurement errors in the observations. The designer chooses $\nu, \omega \in [0,1]$ $(\nu + \omega \leq 1)$ according to the emphasis that he/she wishes to place on these contributions. Similarly, all four components in (12) are $O(1)$.
If, as is typically the case in spatial studies, no replication is possible, then since the design space χ is being held fixed an asymptotic treatment is not possible. (Note: But perhaps we should take $\boldsymbol{C}_{0;N} = \sigma_0^2 \boldsymbol{P}_{0;N}$, i.e. $\sigma_0^2 = O(n)$, and $\omega = O(n^{-1})$ for balance, since $tr\left(n\boldsymbol{L}_0\boldsymbol{L}_0'\right) = O(n)$.)

3 Special cases

Interesting and relevant examples should be found, and the designs computed, in each of the following cases. [**Some algorithms that have worked well for me are described in Wiens (2017) and Kong and Wiens (2015).**]

3.1 Model robust design of experiments

[**Possible example: Experimenter assumes MA(1) errors and a bivariate regression without interactions; we could compare with the**

designs in Figure 1 of Wiens and Zhou (1997). We should use a range of values of a, b, c as in Remark 1 above.]

If $\sigma_\varepsilon^2 > 0$ then useful (?) formulas are given in the following proposition.

Proposition 1 *Assume that $\sigma_\varepsilon^2 > 0$; recall (6). Then*

$$\begin{aligned}
\boldsymbol{V}_0 &= \boldsymbol{D}_\xi^{1/2}\left(\frac{\sigma_\varepsilon^2}{n}\boldsymbol{I}_N + \boldsymbol{D}_\xi^{1/2}\boldsymbol{C}_{0;N}\boldsymbol{D}_\xi^{1/2}\right)^{-1}\boldsymbol{D}_\xi^{1/2}, \\
tr\left\{n\boldsymbol{L}_0\boldsymbol{L}_0'\right\} &= tr\left\{\begin{array}{c}
\left[\left(\boldsymbol{I}_N - \boldsymbol{C}_{0;N}\boldsymbol{V}_0\right)\boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{V}_0\boldsymbol{F}\right)^{-1}\boldsymbol{F}' + \boldsymbol{C}_{0;N}\right] \\
\cdot\left(\frac{\sigma_\varepsilon^2}{n}\boldsymbol{I}_N + \boldsymbol{D}_\xi\boldsymbol{C}_{0;N}\right)^{-1}\boldsymbol{D}_\xi\left(\frac{\sigma_\varepsilon^2}{n}\boldsymbol{I}_N + \boldsymbol{C}_{0;N}\boldsymbol{D}_\xi\right)^{-1}\cdot \\
\left[\left(\boldsymbol{I}_N - \boldsymbol{C}_{0;N}\boldsymbol{V}_0\right)\boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{V}_0\boldsymbol{F}\right)^{-1}\boldsymbol{F}' + \boldsymbol{C}_{0;N}\right]'
\end{array}\right\}.
\end{aligned}$$

Remark 3: The computations involve

$$\begin{aligned}
\boldsymbol{G}_1 &= \left(\frac{\sigma_\varepsilon^2}{n}\boldsymbol{I}_N + \boldsymbol{D}_\xi^{1/2}\boldsymbol{C}_{0;N}\boldsymbol{D}_\xi^{1/2}\right)^{-1} = \frac{n}{\sigma_\varepsilon^2}\left(\boldsymbol{I}_N + \boldsymbol{S}_1\right)^{-1}, \\
\boldsymbol{G}_2 &= \left(\frac{\sigma_\varepsilon^2}{n}\boldsymbol{I}_N + \boldsymbol{D}_\xi\boldsymbol{C}_{0;N}\right)^{-1} = \frac{n}{\sigma_\varepsilon^2}\left(\boldsymbol{I}_N + \boldsymbol{S}_2\right)^{-1},
\end{aligned}$$

for

$$\boldsymbol{S}_1 = n\boldsymbol{D}_\xi^{1/2}\frac{\boldsymbol{C}_{0;N}}{\sigma_\varepsilon^2}\boldsymbol{D}_\xi^{1/2},\ \boldsymbol{S}_2 = \frac{n\boldsymbol{C}_{0;N}}{\sigma_\varepsilon^2}\boldsymbol{D}_\xi.$$

Typically $\boldsymbol{D}_\xi$ is quite sparse; denote by 'pos' the locations of the N_+ non-zero diagonal elements ($N_+ \le \min(N, n)$). Let $\boldsymbol{J} : N \times N_+$ consist of those columns of $\boldsymbol{I}_N$ in 'pos'. Let

$$\boldsymbol{S}_{1+} = \boldsymbol{S}_1\left[pos, pos\right] : N_+ \times N_+$$

consist only of these rows and columns, and let

$$\boldsymbol{S}_{2+} = \boldsymbol{S}_2\left[:, pos\right] : N \times N_+$$

consist only of these columns. Then $\boldsymbol{J}'\boldsymbol{J} = \boldsymbol{I}_{N_+}$ and

$$\boldsymbol{S}_1 = \boldsymbol{J}\boldsymbol{S}_{1+}\boldsymbol{J}',\ \boldsymbol{S}_2 = \boldsymbol{S}_{2+}\boldsymbol{J}';$$

thus

$$\begin{aligned}
\boldsymbol{G}_1 &= \frac{1}{\sigma_\varepsilon^2}\left(\boldsymbol{I}_N + \boldsymbol{J}\boldsymbol{S}_{1+}\boldsymbol{J}'\right)^{-1} = \frac{1}{\sigma_\varepsilon^2}\left[\boldsymbol{I}_N - \boldsymbol{J}\left(\boldsymbol{I}_{N_+} + \boldsymbol{S}_{1+}\right)^{-1}\boldsymbol{S}_{1+}\boldsymbol{J}'\right], \\
\boldsymbol{G}_2 &= \frac{n}{\sigma_\varepsilon^2}\left(\boldsymbol{I}_N + \boldsymbol{S}_{2+}\boldsymbol{J}'\right)^{-1} = \frac{n}{\sigma_\varepsilon^2}\left[\boldsymbol{I}_N - \boldsymbol{S}_{2+}\left(\boldsymbol{I}_{N_+} + \boldsymbol{J}'\boldsymbol{S}_{2+}\right)^{-1}\boldsymbol{J}'\right].
\end{aligned}$$

These involve only the inversion of matrices of order N_+; this reduces (?) the computing time substantially.

3.2 Spatial design

[**For interesting examples, I think I'd start by looking through Cressie (1993). That's where I found the 'coal ash' example I worked on in Wiens (2005).**] When there are no replicates, as for spatial design, the expressions in Proposition 1 simplify. One can replace $n\boldsymbol{D}_\xi = \sqrt{n}\boldsymbol{D}_\xi^{1/2}$ by $\mathbf{I}_{(N)} = diag(\cdots, I(\boldsymbol{x}_i \text{ is in the design}), \cdots) : N \times N$.

Remark 4: We might restrict to covariance structures which are isotropic, i.e. structures for which $\text{COV}[Y(\boldsymbol{x}_1), Y(\boldsymbol{x}_2)]$ depends only on the distance $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$. In this case the theory above continues to hold, since the maximizing structure $\tau^2\mathbf{I}_N$ in Lemma 2 is an isotropic covariance matrix.

3.3 Computer experimentation

[**For examples you might start by looking at the two 'Example 3' references in §1.**]

Here, in addition to there being no replicates, we take $\sigma_\varepsilon^2 = 0$. Then (12) becomes

$$\mathcal{L}_1(\xi) = (1-a-b)\, ch_{\max}\boldsymbol{A}_0\boldsymbol{A}_0' + a \cdot tr\left\{\boldsymbol{A}_0\boldsymbol{A}_0'\right\} + b \cdot tr\left\{n\boldsymbol{A}_0\boldsymbol{C}_{0;N}\boldsymbol{A}_0'\right\},$$

with the calculation of $\boldsymbol{A}_0$ employing $\boldsymbol{V}_0 = \boldsymbol{E}\left(\boldsymbol{E}'\boldsymbol{C}_{0;N}\boldsymbol{E}\right)^{-1}\boldsymbol{E}'$.

Remark 5: In computer experimentation it is common to fit only a constant mean, so that $\boldsymbol{F} = \mathbf{1}_N$ and $\boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{V}_0\boldsymbol{F}\right)^{-1}\boldsymbol{F}' = \frac{\mathbf{1}_N\mathbf{1}_N'}{\mathbf{1}_N'\boldsymbol{V}_0\mathbf{1}_N}$.

Appendix: Derivations

Proof of Theorem 1. (i) For notational convenience we temporarily drop the subscript '0'. The requirement of unbiasedness is $E\left[\hat{\mathbf{Y}}\right] = \boldsymbol{L}\boldsymbol{X}\boldsymbol{\theta} = \boldsymbol{F}\boldsymbol{\theta} = E[\boldsymbol{Y}]$ for all $\boldsymbol{\theta}$; this entails $\boldsymbol{L}\boldsymbol{X} = \boldsymbol{F}$ and so $\boldsymbol{L} = \boldsymbol{L}_1 + \boldsymbol{L}_2$, where $\boldsymbol{L}_1 = \boldsymbol{F}\left(\boldsymbol{X}'\boldsymbol{C}_n^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{C}_n^{-1}$ and

$$\boldsymbol{L}_2\mathbf{X} = \mathbf{0}. \tag{A.1}$$

Thus

$$\hat{\mathbf{Y}} = \boldsymbol{F}\hat{\boldsymbol{\theta}}_{GLS} + \boldsymbol{L}_2\boldsymbol{y},$$

where $\hat{\boldsymbol{\theta}}_{GLS} = \left(\boldsymbol{X}'\boldsymbol{C}_n^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{C}_n^{-1}\boldsymbol{y}$ is unbiased for $\boldsymbol{\theta}$ and $E[\boldsymbol{L}_2\boldsymbol{y}] = \mathbf{0}$.

The PMSE to be minimized is

$$\begin{aligned}
\text{PMSE} &= tr\left\{\text{COV}\left[\boldsymbol{Y} - \hat{\mathbf{Y}}\right]\right\} \\
&= tr\text{COV}[\mathbf{Y}] + tr\text{COV}\left[\hat{\mathbf{Y}}\right] - 2tr\text{COV}\left[\mathbf{Y}, \hat{\mathbf{Y}}'\right] \\
&= tr\boldsymbol{C}_N + tr\boldsymbol{L}\boldsymbol{C}_n\boldsymbol{L}' - 2tr\text{COV}\left[\boldsymbol{Y}, \boldsymbol{y}'\boldsymbol{L}'\right] \\
&= tr\boldsymbol{C}_N + tr\boldsymbol{F}\left(\boldsymbol{X}'\boldsymbol{C}_n^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{F}' - 2tr\boldsymbol{C}_{N,n}\boldsymbol{L}_1' + \left[tr\boldsymbol{L}_2\boldsymbol{C}_n\boldsymbol{L}_2' - 2tr\boldsymbol{C}_{N,n}\boldsymbol{L}_2'\right];
\end{aligned}$$

here we use that $\boldsymbol{L}_1\boldsymbol{C}_n\boldsymbol{L}_2' = \boldsymbol{L}_2\boldsymbol{C}_n\boldsymbol{L}_1' = \boldsymbol{0}_{N\times N}$ and that $\boldsymbol{L}_1\boldsymbol{C}_n\boldsymbol{L}_1' = \boldsymbol{F}\left(\boldsymbol{X}'\boldsymbol{C}_n^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{F}'$.

We are now to minimize $tr\boldsymbol{L}_2\boldsymbol{C}_n\boldsymbol{L}_2' - 2tr\boldsymbol{C}_{N,n}\boldsymbol{L}_2'$ subject to (A.1). This orthogonality condition, which we now write as $\boldsymbol{L}_2\boldsymbol{C}_n^{1/2}\boldsymbol{C}_n^{-1/2}\boldsymbol{X} = \boldsymbol{0}$, states that the rows of $\boldsymbol{L}_2\boldsymbol{C}_n^{1/2}$ lie in the row space of the orthogonal projector $\boldsymbol{I}_n - \boldsymbol{H}$, where $\boldsymbol{H} = \boldsymbol{C}_n^{-1/2}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{C}_n^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{C}_n^{-1/2}$, so that if the rows of $\boldsymbol{\Pi} : (n-p)\times n$ form an orthogonal basis for this space (so that $\boldsymbol{\Pi}\boldsymbol{\Pi}' = \boldsymbol{I}_{n-p}$ and $\boldsymbol{\Pi}'\boldsymbol{\Pi} = \boldsymbol{I}_n - \boldsymbol{H}$), we have that, for some $\boldsymbol{M} : N\times(n-p)$, $\boldsymbol{L}_2\boldsymbol{C}_n^{1/2} = \boldsymbol{M}\boldsymbol{\Pi}$. Thus we minimize

$$\begin{aligned}
& tr\boldsymbol{L}_2\boldsymbol{C}_n\boldsymbol{L}_2' - 2tr\boldsymbol{C}_{N,n}\boldsymbol{L}_2' \\
= \;& tr\left[\boldsymbol{M}\boldsymbol{M}' - \boldsymbol{C}_{N,n}\boldsymbol{C}_n^{-1/2}\boldsymbol{\Pi}'\boldsymbol{M}' - \boldsymbol{M}\boldsymbol{\Pi}\boldsymbol{C}_n^{-1/2}\boldsymbol{C}_{N,n}'\right] \\
= \;& tr\left[\left(\boldsymbol{M} - \boldsymbol{C}_{N,n}\boldsymbol{C}_n^{-1/2}\boldsymbol{\Pi}'\right)\left(\boldsymbol{M}' - \boldsymbol{\Pi}'\boldsymbol{C}_n^{-1/2}\boldsymbol{C}_{N,n}'\right) - \boldsymbol{C}_{N,n}\boldsymbol{C}_n^{-1/2}\boldsymbol{\Pi}'\boldsymbol{\Pi}\boldsymbol{C}_n^{-1/2}\boldsymbol{C}_{N,n}'\right]
\end{aligned}$$

over $\boldsymbol{M}$, unconditionally. The solution is clearly $\boldsymbol{M} = \boldsymbol{C}_{N,n}\boldsymbol{C}_n^{-1/2}\boldsymbol{\Pi}'$, whence

$$\boldsymbol{L}_2 = \boldsymbol{C}_{N,n}\boldsymbol{C}_n^{-1}\left[\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{C}_n^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{C}_n^{-1}\right].$$

The statements in (ii) follow by a direct calculation.
(iii) Put $\boldsymbol{\psi}_n = \left(\boldsymbol{1}_{n_1}'\psi(\boldsymbol{x}_1), \cdots, \boldsymbol{1}_{n_N}'\psi(\boldsymbol{x}_N)\right)_{n_i>0}'$, so that $E[\boldsymbol{Y}] = \boldsymbol{F}\boldsymbol{\theta} + \boldsymbol{\psi}_N$, $E[\boldsymbol{y}] = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\psi}_n$. Define $\boldsymbol{m} = E\left[\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right]$, then $\text{PMSE} = \|\boldsymbol{m}\|^2 + tr\text{COV}\left[\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right]$. Since $\boldsymbol{\psi}_n = \boldsymbol{E}'\boldsymbol{\psi}_N$ and $\boldsymbol{L}_0\boldsymbol{X} = \boldsymbol{F}$, we have that

$$\boldsymbol{m} = E[\boldsymbol{Y} - \boldsymbol{L}_0\boldsymbol{y}] = (\boldsymbol{I}_N - \boldsymbol{L}_0\boldsymbol{E}')\,\boldsymbol{\psi}_N.$$

Furthermore, and using (5) and (7),

$$\begin{aligned}
tr\text{COV}\left[\boldsymbol{Y} - \hat{\boldsymbol{Y}}\right] &= tr\text{COV}\left[\boldsymbol{Y} - \boldsymbol{L}_0\boldsymbol{y}\right] \\
&= tr\left\{\boldsymbol{C}_N - \boldsymbol{C}_{N,n}\boldsymbol{L}_0' - \boldsymbol{L}_0\boldsymbol{C}_{N,n}' + \boldsymbol{L}_0\boldsymbol{C}_n\boldsymbol{L}_0'\right\} \\
&= tr\left\{\left(\boldsymbol{I}_N \vdots -\boldsymbol{L}_0\right)\left[\begin{array}{c}\begin{pmatrix}\boldsymbol{I}_N \\ \boldsymbol{E}'\end{pmatrix}\boldsymbol{C}_N\left(\boldsymbol{I}_N \vdots \boldsymbol{E}\right) \\ +\sigma_\varepsilon^2\left(0_{N\times N}\oplus\boldsymbol{I}_n\right)\end{array}\right]\begin{pmatrix}\boldsymbol{I}_N \\ -\boldsymbol{L}_0'\end{pmatrix}\right\} \\
&= tr\left\{(\boldsymbol{I}_N - \boldsymbol{L}_0\boldsymbol{E}')\,\boldsymbol{C}_N\,(\boldsymbol{I}_N - \boldsymbol{E}\boldsymbol{L}_0')\right\} + \sigma_\varepsilon^2 tr\boldsymbol{L}_0\boldsymbol{L}_0'.
\end{aligned}$$

□

Proof of Lemma 1: We first classify the solutions to (2). Let $\boldsymbol{F} = \boldsymbol{Q}_1\boldsymbol{R}$ be the qr-decomposition of $\boldsymbol{F}$, so that $\boldsymbol{Q}_1 : N\times p$ satisfies $\boldsymbol{Q}_1'\boldsymbol{Q}_1 = \boldsymbol{I}_p$ and $\boldsymbol{R} : p\times p$ is upper triangular and non-singular. Augment $\boldsymbol{Q}_1$ by $\boldsymbol{Q}_2 : N\times(N-p)$ in such a way that $\boldsymbol{Q} = \left(\boldsymbol{Q}_1 \vdots \boldsymbol{Q}_2\right)$ is an orthogonal matrix. Then the columns of $\boldsymbol{Q}_2$ form an orthogonal basis for the orthogonal complement of the column space of $\boldsymbol{F}$, to which $\boldsymbol{\psi}_N$ belongs by virtue of (2). Thus $\boldsymbol{\psi}_N = \boldsymbol{Q}_2\boldsymbol{c}$ for some $\boldsymbol{c} \in \mathbb{R}^{N-p}$, and we maximize $\boldsymbol{\psi}_N'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{\psi}_N = \boldsymbol{c}'\boldsymbol{Q}_2'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}_2\boldsymbol{c}$, subject to $\|\boldsymbol{\psi}_N\| = \|\boldsymbol{c}\| \leq \alpha/\sqrt{n}$. The maximizing $\boldsymbol{c}$ is $\alpha/\sqrt{n}$ times the unit eigenvector of $\boldsymbol{Q}_2'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}_2$ corresponding to the maximum eigenvalue $ch_{\max}\boldsymbol{Q}_2'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}_2$, and then

$$\max \boldsymbol{\psi}_N'\boldsymbol{A}\boldsymbol{\psi}_N = \left(\alpha^2/n\right) \cdot ch_{\max}\boldsymbol{Q}_2'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}_2.$$

Now (10) follows from

$$\begin{aligned} ch_{\max}\boldsymbol{Q}_2'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{Q}_2 &= ch_{\max}\boldsymbol{A}\boldsymbol{Q}_2\boldsymbol{Q}_2'\boldsymbol{A}' = ch_{\max}\boldsymbol{A}\left(\boldsymbol{I}_N - \boldsymbol{Q}_1\boldsymbol{Q}_1'\right)\boldsymbol{A}' \\ &= ch_{\max}\boldsymbol{A}\left(\boldsymbol{I}_N - \boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1}\boldsymbol{F}'\right)\boldsymbol{A}' = ch_{\max}\boldsymbol{A}\boldsymbol{A}'. \end{aligned}$$

□

Proof of Proposition 1: The identities follow from

$$\begin{aligned} \boldsymbol{C}_{0;n}^{-1} &= \frac{1}{\sigma_\varepsilon^2}\left[\boldsymbol{I}_n - \boldsymbol{E}'\frac{\boldsymbol{C}_{0;N}}{\sigma_\varepsilon^2}\left(\boldsymbol{I}_N + n\boldsymbol{D}_\xi\frac{\boldsymbol{C}_{0;N}}{\sigma_\varepsilon^2}\right)^{-1}\boldsymbol{E}\right], \\ \boldsymbol{E}\boldsymbol{C}_{0;n}^{-1} &= \left(\sigma_\varepsilon^2\boldsymbol{I}_N + n\boldsymbol{D}_\xi\boldsymbol{C}_{0;N}\right)^{-1}\boldsymbol{E}. \end{aligned}$$

□

Possibly useful (?) notes to myself:

1. Put $\boldsymbol{A}_0 = \boldsymbol{C}_{0;N,n}\boldsymbol{C}_{0;n}^{-1}$ and $\boldsymbol{B}_0 = \left(\boldsymbol{F} - \boldsymbol{C}_{0;N,n}\boldsymbol{C}_{0;n}^{-1}\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{C}_{0;n}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{C}_{0;n}^{-1}$; note that $\boldsymbol{B}_0\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{C}_{0;n}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{C}_{0;n}^{-1} = \boldsymbol{B}_0$. In these terms we have that $\boldsymbol{L}_0 = \boldsymbol{A}_0 + \boldsymbol{B}_0$ and the error vector decomposes into uncorrelated components
$$\boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{Y} - \boldsymbol{A}_0\boldsymbol{y}) - \boldsymbol{B}_0\boldsymbol{y}.$$

2. Suppose the Y_i are i.i.d., so that $\boldsymbol{C}_{0;N} = \sigma_0^2\boldsymbol{I}_N$. Then
$$\begin{aligned}
\boldsymbol{C}_{0;n} &= \sigma_\varepsilon^2\left(\boldsymbol{I}_n + \frac{\sigma_0^2}{\sigma_\varepsilon^2}\boldsymbol{E}'\boldsymbol{E}\right),\\
\boldsymbol{C}_{0;n}^{-1} &= \frac{1}{\sigma_\varepsilon^2}\left(\boldsymbol{I}_n - \frac{\sigma_0^2}{\sigma_\varepsilon^2}\boldsymbol{E}'\left(\boldsymbol{I}_N + \frac{\sigma_0^2}{\sigma_\varepsilon^2}n\boldsymbol{D}_\xi\right)^{-1}\boldsymbol{E}\right),\\
\boldsymbol{C}_{0;N,n} &= \sigma_0^2\boldsymbol{E},\\
\boldsymbol{C}_{0;N,n}\boldsymbol{C}_{0;n}^{-1} &= \frac{\sigma_0^2}{\sigma_\varepsilon^2}\left(\boldsymbol{I}_N + \frac{\sigma_0^2}{\sigma_\varepsilon^2}n\boldsymbol{D}_\xi\right)^{-1}\boldsymbol{E}\\
&= diag\left(\cdots,\frac{\sigma_0^2}{\sigma_\varepsilon^2 + n_i\sigma_0^2},\cdots\right)\boldsymbol{E},\\
\boldsymbol{V}_0 &= \boldsymbol{D}_\xi^{1/2}\left(\frac{\sigma_\varepsilon^2}{n}\boldsymbol{I}_N + \sigma_0^2\boldsymbol{D}_\xi\right)^{-1}\boldsymbol{D}_\xi^{1/2}\\
&= diag\left(\cdots,\frac{n_i}{\sigma_\varepsilon^2 + n_i\sigma_0^2},\cdots\right)\\
&= diag\left(\cdots,\frac{1}{\frac{\sigma_\varepsilon^2}{n_i} + \sigma_0^2}I\left(n_i > 0\right),\cdots\right).
\end{aligned}$$

   Then if $n_i > 0$, the prediction is
$$\hat{Y}_i = \frac{\sigma_\varepsilon^2\boldsymbol{f}'\left(\boldsymbol{x}_i\right)\hat{\boldsymbol{\theta}}_0 + n_i\sigma_0^2\bar{y}_i}{\sigma_\varepsilon^2 + n_i\sigma_0^2},$$
   a weighted average of the regression estimate $\boldsymbol{f}'\left(\boldsymbol{x}_i\right)\hat{\boldsymbol{\theta}}_0$ and $\bar{y}_i$, with the weight on $\bar{y}_i$ increasing with n_i but decreasing with σ_ε^2. If $n_i = 0$, the prediction is $\boldsymbol{f}'\left(\boldsymbol{x}_i\right)\hat{\boldsymbol{\theta}}_0$.

Acknowledgements

This work was carried out with the support of the Natural Sciences and Engineering Research Council of Canada.

References

Chen, H., Loeppky, J. L., Sacks J., and Welch, W. J. (2016), "Analysis Methods for Computer Experiments: How to Assess and What Counts?," *Statistical Science*, 31, 40-60.

Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley.

Kong, L., and Wiens, D. P. (2015), "Model-Robust Designs for Quantile Regression," *Journal of the American Statistical Association*, 110, 233-245.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, Springer Verlag: New York.

Welsh, A. H., and Wiens, D. P. (2013), "Robust Model-based Sampling Designs," *Statistics and Computing*, 23, 689-701.

Wiens, D. P. (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference,* 31, 353-371.

Wiens, D. P. (2005); "Robustness in Spatial Studies II: Minimax Design," *Environmetrics*, 16, 205-217.

Wiens, D. P. (2017), "I-Robust and D-Robust Designs on a Finite Design Space," *Statistics and Computing*, in press.

Wiens, D. P., and Zhou, J. (1997), "Robust Designs Based on the Infinitesimal Approach," *Journal of the American Statistical Association*, 92, 1503-1511.

Wiens, D. P., and Zhou, J. (2008), "Robust Estimators and Designs for Field Experiments," *Journal of Statistical Planning and Inference*, 138, 93-104.