Sequential Design of Computer Experiments to Minimize Integrated Response Functions

Brian J. Williams, Thomas J. Santner, and William I. Notz

The Ohio State University

Abstract: In the last ten to fifteen years many phenomena that could only be studied using physical experiments can now be studied by computer experiments. Advances in the mathematical modeling of many physical processes, in algorithms for solving mathematical systems, and in computer speeds, have combined to make it possible to replace some physical experiments with computer experiments. In a computer experiment, a deterministic output, $y(\boldsymbol{x})$, is computed for each set of input variables, $\boldsymbol{x}$. This paper is concerned with the commonly occuring situation in which there are two types of input variables: suppose $\boldsymbol{x} = (\boldsymbol{x}_c, \boldsymbol{x}_e)$ where $\boldsymbol{x}_c$ is a set of "manufacturing" (control) variables and $\boldsymbol{x}_e$ is a set of "environmental" (noise) variables. Manufacturing variables can be controlled while environmental variables are not controllable but have values governed by some distribution. We introduce a sequential experimental design for finding the optimum of $\ell(\boldsymbol{x}_c) = E\{y(\boldsymbol{x}_c, \boldsymbol{X}_e)\}$, where the expectation is taken over the distribution of the environmental variables. The approach is Bayesian; the prior information is that $y(\boldsymbol{x})$ is a draw from a stationary Gaussian stochastic process with correlation function from the Matérn class having unknown parameters. The idea of the method is to compute the posterior expected "improvement" over the current optimum for each untested site; the design selects the next site to maximize the expected improvement. The procedure is illustrated with examples from the literature.

Key words and phrases: Computer experiments; control variables; expected improvement; noise variables; optimization; sequential design.

1 Introduction

Many physical systems can be modeled mathematically so that responses are computable for specified inputs, using numerical methods that are implemented by (complex) computer codes. For example, Bernardo, Buck, Liu, Nazaret, Sacks and Welch (1992) used computer-aided design simulators to model current reference and voltage shifter circuits. Haylock and O'Hagan (1996) modelled the radiation dose received by body organs after ingesting radioactive iodine. Chang, Williams, Notz, Santner and Bartel (1999) used finite-element methods to model proximal bone stress shielding resulting from an in vivo hip prosthesis. We refer to such settings as *computer experiments*, to contrast them with physical experiments.

Unlike physical experiments, where random error is a fundamental part of the model, the output of computer experiments is generally deterministic. In addition, observations from computer experiments can be expensive and time consuming to collect. This is a consequence of the complexity of the code and/or the large number of input variables. Thus, in many applications, it is impractical to compute the response on a large grid of input variable values; this difficulty has led investigators to develop statistical methodology to allow the response to be accurately predicted throughout the input variable space based on a small training sample of computed responses.

Two basic statistical frameworks have been explored in developing predictors for application to computer experiments. Sacks, Schiller and Welch (1989) and Welch, Buck, Sacks, Wynn, Mitchell and Morris (1992) adopt a "modeling approach" in that they regard the deterministic response as a realization of a random function. From an alternate philosophical

viewpoint, Currin, Mitchell, Morris and Ylvisaker (1991) and O'Hagan (1992) make clear that the previous approach is essentially a Bayesian formulation of the problem where a random function represents prior information about the deterministic function.

One important goal of computer experiments is to find input values that optimize the response. The problem of minimizing deterministic responses has been studied extensively in the mathematical programming literature but essentially all such techniques require far too many function evaluations to be used in most computer experiments. Instead, various statistical predictors have been used in conjunction with traditional numerical algorithms to solve such problems. One example is Bernardo et al. (1992) who carry out a sequential strategy for response minimization. The response is predicted throughout the full input space based on an initial design. If the predictor is sufficiently accurate, then it is minimized. Otherwise, a promising subregion of the input space is determined and the response is predicted throughout the subregion based on a second-stage design. If the predictor is sufficiently accurate in the subregion, it is minimized. Otherwise, this process moves to a third stage and continues in this fashion until adequate prediction accuracy is obtained. Jones, Schonlau and Welch (1998) and Schonlau, Welch and Jones (1998) introduced a criterion-based sequential strategy for response minimization. This also begins with an initial design, but proceeds subsequently by choosing points one at a time, or in groups, to maximize a criterion that favors inputs in regions where either the predicted response is small or where there is relatively large prediction uncertainty. The true response is calculated at each selected point, the predictor is updated, and the algorithm continues until relative

changes in the criterion value become negligible.

This paper considers the frequently occurring situation in which the input variables can be divided into two classes, *control* variables and *environmental* variables. For computational convenience we assume that the environmental variables have finite support. Our objective function depends only on the control variables: for each fixed setting of the control variables the objective function is the mean of the deterministic response over the distribution of the environmental variables. Thus we seek to minimize a weighted average of the response over the values of the environmental variables.

For example, in the hip prosthesis problem of Chang et al. (1999), the control variables specify the geometry of the implant and the environmental variables account for variability in patient bone properties and activity. The deterministic response was proximal bone stress shielding, and the goal of the problem was to determine the combination of control variables that minimized stress shielding averaged over a discrete probability distribution for the environmental variables. More generally, the classification of inputs as control variables or environmental variables applies in many manufacturing settings where some inputs affecting a process can be controlled while others cannot. For example, Welch, Yu, Kang and Sacks (1990) were interested in minimizing the clock skews of a very large scale integrated circuit. The control variables were the widths of six transistors, and the environmental variables were qualitative indicators of their current–driving capabilities.

We consider an extension of the expected improvement algorithm of Jones et al. (1998) to carry out this type of objective function minimization. We cannot use their expected

improvement algorithm because it requires direct observation of the objective function at each selected point. This will not be possible in many applications due to the enormous computational burden of calculating the average response. As an example, consider again the hip replacement problem described in the previous paragraph. The support of the environmental variables was restricted to twelve points. Thus, *twelve runs* of the code are required to calculate the objective function at each control variable setting of interest. Each run of the finite element code requires five to ten hours so that it could take five days to calculate a single value of the objective function.

Our algorithm attempts to optimize the objective function using a predictor of the mean function. In brief, the algorithm proceeds as follows:

1. Calculate the responses on an initial space-filling design.

2. Use the information from these runs to select the next point according to a modified expected improvement criterion.

3. Continue selecting points using the necessary information from all of the previous runs until a stopping criterion is met.

In Section 2 we discuss the statistical approach taken in this paper. A modified expected improvement algorithm is presented in detail in Section 3, and three examples are given in Section 4. These examples use closed–form test functions to allow the optimum found by the algorithm to be checked with the true optimum. Section 5 contains a discussion of several important issues regarding implementation and extensions of the algorithm. The appendix

presents details that we have found useful in implementing the algorithm.

2 Modeling

We follow the Bayesian framework proposed by many authors and assume that prior uncertainty about the deterministic response $y(\boldsymbol{x})$ for compact $\mathcal{X} \subset \mathbb{R}^p$ is represented by the stationary random function $Y(\boldsymbol{x}) = \beta_0 + Z(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathcal{X}$, where β_0 is an unknown constant and $Z(\cdot)$ is taken to be a zero-mean stationary Gaussian stochastic process with unknown variance σ^2 and correlation function $R(\cdot)$. The parameter β_0 represents the global mean of the Y process. The function $R(\cdot)$ determines the correlation between the random responses at any two input sites in $\mathcal{X}$. Stationarity implies that $R(\cdot)$ is translation invariant, so that the correlation calculated for any two input sites $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{X}$ depends only on $\boldsymbol{u} - \boldsymbol{v}$.

The correlation functions commonly used in practice are members of some parametric family. The development of Section 3 assumes that the correlation function depends on an unknown parameter vector $\boldsymbol{\zeta}$. In the examples of Section 4, we follow Handcock and Stein (1993) and use the **Matérn** class of correlation functions:

$$R(\boldsymbol{u} - \boldsymbol{v}) = \prod_{i=1}^{p} \frac{1}{\Gamma(\nu) 2^{\nu-1}} \left(\frac{2\sqrt{\nu}\,|u_i - v_i|}{\theta_i} \right)^{\nu} K_\nu \left(\frac{2\sqrt{\nu}\,|u_i - v_i|}{\theta_i} \right) ,$$

where u_i and v_i are the i-th coordinates of $\boldsymbol{u}$ and $\boldsymbol{v}$. Here $\nu > 0$, $\theta_i > 0$ and $K_\nu(\cdot)$ is the modified Bessel function of order ν. The parameter θ_i controls the strength of correlations in the i-th input variable dimension. Larger values of these range parameters are associated with increased dependence between the random responses at any two fixed input sites. The

parameter ν controls the smoothness of the random field. The Y process is $\lceil \nu \rceil - 1$ times mean square differentiable, where $\lceil \cdot \rceil$ denotes the integer ceiling function. In fact, the sample paths of the Y process are almost surely $\lceil \nu \rceil - 1$ times continuously differentiable (see Cramér and Leadbetter (1967), Secs. 4.2, 7.3, and 9.2–9.5).

We adopt a Bayesian viewpoint and assume the noninformative prior distribution,

$$[\, \beta_0, \sigma^2, \boldsymbol{\zeta} \,] \;\propto\; \frac{1}{\sigma^2}\,, \tag{1}$$

for the parameter vector $(\beta_0, \sigma^2, \boldsymbol{\zeta}^\top)^\top$.

Let $\boldsymbol{x}_c$ and $\boldsymbol{x}_e$ represent the control and environmental variable vectors, and denote their corresponding domains by $\mathcal{X}_c$ and $\mathcal{X}_e$. We assume that the environmental variables have (approximately) a joint probability distribution with finite support $\{\boldsymbol{x}_{e,i}\}_{i=1}^{n_e}$ and associated weights $\{w_i\}_{i=1}^{n_e}$. The *objective function* $\ell(\cdot)$ is given by

$$\ell(\boldsymbol{x}_c) = \sum_{i=1}^{n_e} w_i \, y(\boldsymbol{x}_c, \boldsymbol{x}_{e,i})\,. \tag{2}$$

Our goal is to identify the control variable settings $\boldsymbol{x}_c^*$ that minimize $\ell(\cdot)$, $\boldsymbol{x}_c^* = \mathrm{argmin}_{\boldsymbol{x}_c \in \mathcal{X}_c}$ $\ell(\boldsymbol{x}_c)$. Prior uncertainty in $\ell(\cdot)$ is induced directly from the Y process; the prior of $\ell(\cdot)$ can be described by the distribution of $L(\boldsymbol{x}_c) = \sum_{i=1}^{n_e} w_i \, Y(\boldsymbol{x}_c, \boldsymbol{x}_{e,i})\,.$

3 The Minimization Algorithm

3.1 Overview

As discussed in the introduction, the first stage of the modified expected improvement algorithm involves observing the response at each site in an initial space-filling design

$S_n = \{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n\}$. Denote the control variable portion of S_n by $S_n^c = \{\boldsymbol{t}_{c,1}, \ldots, \boldsymbol{t}_{c,n}\}$. With this notation, let $\boldsymbol{Y}_{S_n}$ denote the random vector of responses associated with S_n and $\boldsymbol{L}_{S_n^c}$ the random vector of objective function values associated with S_n^c. Setting $L_{1:n} = \min\{L(\boldsymbol{t}_{c,1}), \ldots, L(\boldsymbol{t}_{c,n})\}$, the *improvement* at control variable site $\boldsymbol{x}_c$ is defined to be $I_n(\boldsymbol{x}_c) = \max\{0, L_{1:n} - L(\boldsymbol{x}_c)\}$.

Before proceeding, we note two key ways in which the specification of improvement given in Jones et al. (1998) differs from that used in this paper. First, they replace the *random variable* $L_{1:n}$ with the *known* quantity $y_{\min} \equiv \min\{y(\boldsymbol{t}_1), \ldots, y(\boldsymbol{t}_n)\}$, the minimum of the observed responses on S_n; as noted above, $L_{1:n}$ is unknown because there are no direct observations on the objective function. Second, they replace $L(\boldsymbol{x}_c)$ by $Y(\boldsymbol{x})$. Our changes reflect the fact that we are concerned with minimization of the mean of $y(\cdot)$ over the environmental variables.

With the above notation, we can summarize the proposed algorithm.

S0: Choose the initial set of design points $S_n = \{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n\}$ according to a space-filling criterion. We use the ACED software of Welch (1985) to choose a maximin distance design in the set of Latin Hypercube Sampling (LHS) designs.

S1: Estimate the correlation parameter vector $\boldsymbol{\zeta}$ by the maximizer of the posterior density of $\boldsymbol{\zeta}$ given $\boldsymbol{Y}_{S_n}$ from (16).

S2: Choose the $(n+1)$-st *control* variable site, $\boldsymbol{t}_{c,n+1}$, to maximize the *posterior expected*

improvement given the current data, i.e.,

$$\boldsymbol{t}_{c,n+1} = \underset{\boldsymbol{x}_c \in \mathcal{X}_c}{\operatorname{argmax}} \, \mathrm{E}\{ I_n(\boldsymbol{x}_c) \mid \boldsymbol{Y}_{S_n}, \boldsymbol{\zeta} \}, \tag{3}$$

where $\mathrm{E}\{ \cdot \mid \boldsymbol{Y}_{S_n}, \boldsymbol{\zeta} \}$ denotes the posterior conditional mean given the observed data $\boldsymbol{Y}_{S_n}$ and the correlation parameter vector $\boldsymbol{\zeta}$.

S3: Choose the *environmental* variable site corresponding to the control site $\boldsymbol{t}_{c,n+1}$ to minimize the *posterior mean square prediction error* given the current data, i.e.

$$\boldsymbol{t}_{e,n+1} = \underset{\boldsymbol{x}_e \in \mathcal{X}_e}{\operatorname{argmin}} \, \mathrm{E}\{ [\widehat{L}_{n+1}(\boldsymbol{t}_{c,n+1}) - L(\boldsymbol{t}_{c,n+1})]^2 \mid \boldsymbol{Y}_{S_n}, \boldsymbol{\zeta} \}, \tag{4}$$

where $\widehat{L}_{n+1}(\cdot)$ is the posterior mean $m_e(\boldsymbol{\zeta})$ given in (13), based on the $(n+1)$-point design $S_n \cup (\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_e)$.

S4: Determine if the algorithm should be stopped. If the stopping criterion is not met, set $S_{n+1} = S_n \cup (\boldsymbol{t}_{c,n+1}, \boldsymbol{t}_{e,n+1})$ and calculate the underlying response $y(\cdot)$ at $(\boldsymbol{t}_{c,n+1}, \boldsymbol{t}_{e,n+1})$. Then set n to $(n+1)$ and continue with *S1*. If the criterion is met, the global minimizer is set to be the minimizer of the empirical best linear unbiased predictor (EBLUP) based on the current design. Specific stopping criteria are discussed in the examples of Section 4.

The optimizations required in (3) and (4) are carried out using the simplex algorithm of Nelder and Mead (1965). The starting simplex is determined randomly; repeated attempts are made to find an optimal solution to avoid getting trapped in local optima.

3.2 Some Details

The following result, discussed in O'Hagan (1992), is used throughout.

Result 1. Let $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ denote $q_1 \times 1$ and $q_2 \times 1$ random vectors having the Gaussian distribution

$$\begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{pmatrix} \mid \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}_{q_1+q_2}\left[\begin{pmatrix} \boldsymbol{F}_1 \\ \boldsymbol{F}_2 \end{pmatrix} \boldsymbol{\beta}\,,\ \sigma^2 \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{12}^\top & \boldsymbol{R}_{22} \end{pmatrix} \right],$$

where $\boldsymbol{\beta} \in \mathbb{R}^k$ and $\sigma^2 > 0$. It is assumed that the elements of each $\boldsymbol{F}_i$ and $\boldsymbol{R}_{ij}$ are known, each $\boldsymbol{F}_i$ has full column rank, and the correlation matrix is positive definite. Let the parameter vector $(\boldsymbol{\beta}, \sigma^2)$ have the noninformative prior distribution $[\boldsymbol{\beta}, \sigma^2] \propto 1/\sigma^2$ for $\boldsymbol{\beta} \in \mathbb{R}^k$ and $\sigma^2 > 0$. The posterior distribution of $\boldsymbol{U}_1$ given $\boldsymbol{U}_2$ is q_1-variate t: $[\boldsymbol{U}_1|\boldsymbol{U}_2] \sim \mathcal{T}_{q_1}(\boldsymbol{m}_{1|2}, \widehat{\sigma^2}\boldsymbol{R}_{1|2}, q_2 - k)$, where $\mathcal{T}_{q_1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes a q_1-variate shifted t-distribution with location shift (mean) $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and ν degrees of freedom. We have $\boldsymbol{m}_{1|2} = \boldsymbol{F}_1\widehat{\boldsymbol{\beta}} + \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}(\boldsymbol{U}_2 - \boldsymbol{F}_2\widehat{\boldsymbol{\beta}})$ for $\widehat{\boldsymbol{\beta}} = (\boldsymbol{F}_2^\top \boldsymbol{R}_{22}^{-1}\boldsymbol{F}_2)^{-1}\boldsymbol{F}_2^\top \boldsymbol{R}_{22}^{-1}\boldsymbol{U}_2$, $\widehat{\sigma^2} = [\boldsymbol{U}_2^\top \boldsymbol{R}_{22}^{-1}\boldsymbol{U}_2 - \widehat{\boldsymbol{\beta}}^\top(\boldsymbol{F}_2^\top \boldsymbol{R}_{22}^{-1}\boldsymbol{F}_2)\widehat{\boldsymbol{\beta}}]/(q_2 - k)$ and $\boldsymbol{R}_{1|2} = \boldsymbol{R}_{11} - \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{12}^\top + (\boldsymbol{F}_1 - \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{F}_2)(\boldsymbol{F}_2^\top \boldsymbol{R}_{22}^{-1}\boldsymbol{F}_2)^{-1}(\boldsymbol{F}_1 - \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{F}_2)^\top$.

3.2.1 Selection of control variables

We obtain a simplified expression for the posterior expected improvement. Note first that

$$\mathrm{E}\{\, I_n(\boldsymbol{x}_c) \mid \boldsymbol{Y}_{S_n}, \boldsymbol{\zeta} \,\} = \mathrm{E}_{\boldsymbol{L}_{S_n^c} \mid \boldsymbol{Y}_{S_n}, \boldsymbol{\zeta}} \left\{ \mathrm{E}\{ I_n(\boldsymbol{x}_c) \mid \boldsymbol{Y}_{S_n}, \boldsymbol{L}_{S_n^c}, \boldsymbol{\zeta} \} \right\}. \tag{5}$$

To evaluate the inner expectation, we require the posterior distribution of $L(\boldsymbol{x}_c)$ given $\boldsymbol{Y}_{S_n}$, $\boldsymbol{L}_{S_n^c}$ and $\boldsymbol{\zeta}$. Define the $n_e \times 1$ vector $\boldsymbol{Y}_1$ and the $(n \cdot n_e) \times 1$ vector $\boldsymbol{Y}_2$ as follows: $\boldsymbol{Y}_1 = [Y(\boldsymbol{x}_c, \boldsymbol{x}_{e,1}), \ldots, Y(\boldsymbol{x}_c, \boldsymbol{x}_{e,n_e})]^\top$ and $\boldsymbol{Y}_2 = [Y(\boldsymbol{t}_{c,1}, \boldsymbol{x}_{e,1}), \ldots, Y(\boldsymbol{t}_{c,1}, \boldsymbol{x}_{e,n_e}), \ldots, Y(\boldsymbol{t}_{c,n}, \boldsymbol{x}_{e,1}),$

$\ldots, Y(\boldsymbol{t}_{c,n}, \boldsymbol{x}_{e,n_e})]^\top$. Here $\boldsymbol{Y}_1$ is the random vector of responses at control site $\boldsymbol{x}_c$ paired with each environmental support point, and $\boldsymbol{Y}_2$ is the random vector of responses at the n control sites combined with each of the environmental support points. Given $(\beta_0, \sigma^2, \boldsymbol{\zeta}^\top)^\top$, the random vector $(\boldsymbol{Y}_1^\top, \boldsymbol{Y}_{S_n}^\top, \boldsymbol{Y}_2^\top)^\top$ has a joint Gaussian distribution with mean $(\mathbf{1}_{n_e}^\top, \mathbf{1}_n^\top, \mathbf{1}_{n\cdot n_e}^\top)^\top \beta_0$ and covariance matrix $\sigma^2\,((\,\boldsymbol{R}_{\boldsymbol{\zeta},ij}\,))$ for $i, j \in \{1, S_n, 2\}$, where $\mathbf{1}_r$ is the $r \times 1$ vector of ones and $\boldsymbol{R}_{\boldsymbol{\zeta},ij}$ is the matrix of correlations between the responses in the corresponding random vectors. Because Gaussian random vectors remain Gaussian under linear transformations, we see easily that $(L(\boldsymbol{x}_c), \boldsymbol{Y}_{S_n}^\top, \boldsymbol{L}_{S_n^c}^\top)^\top$ given $(\beta_0, \sigma^2, \boldsymbol{\zeta}^\top)^\top$ has a Gaussian distribution with mean $\beta_0 \mathbf{1}_{2n+1}$ and covariance matrix

$$\sigma^2 \begin{pmatrix} \boldsymbol{w}^\top \boldsymbol{R}_{\boldsymbol{\zeta},11} \boldsymbol{w} & \boldsymbol{w}^\top \boldsymbol{R}_{\boldsymbol{\zeta},(1,S_n)} & \boldsymbol{w}^\top \boldsymbol{R}_{\boldsymbol{\zeta},12}(\boldsymbol{I}_n \otimes \boldsymbol{w}) \\ \cdot & \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)} & \boldsymbol{Q}_{23} \\ \cdot & \cdot & \boldsymbol{Q}_{33} \end{pmatrix}, \tag{6}$$

where $\boldsymbol{Q}_{23} = \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,2)}(\boldsymbol{I}_n \otimes \boldsymbol{w})$ and $\boldsymbol{Q}_{33} = (\boldsymbol{I}_n \otimes \boldsymbol{w}^\top)\boldsymbol{R}_{\boldsymbol{\zeta},22}(\boldsymbol{I}_n \otimes \boldsymbol{w})$. Here, $\boldsymbol{I}_r$ denotes the $r \times r$ identity matrix, $\boldsymbol{w} = (w_1, \ldots, w_{n_e})^\top$ is the vector of probabilities (weights) defining the uncertainty distribution of the environmental variables, and $\otimes$ denotes the Kronecker (outer) product operator. The missing entries in the covariance matrix are defined by symmetry.

The posterior distribution of $L(\boldsymbol{x}_c)$ given $\boldsymbol{Y}_{S_n}$, $\boldsymbol{L}_{S_n^c}$ and $\boldsymbol{\zeta}$, is a shifted, univariate t. Specifically, let $t(\mu,\, v,\, \nu)$ denote a univariate t-distribution with mean μ, scale parameter v and ν degrees of freedom and $\boldsymbol{Z}_c^\top = (\boldsymbol{Y}_{S_n}^\top, \boldsymbol{L}_{S_n^c}^\top)$, $\boldsymbol{c}_{\boldsymbol{\zeta},12}^\top = \boldsymbol{w}^\top[\boldsymbol{R}_{\boldsymbol{\zeta},(1,S_n)} \quad \boldsymbol{R}_{\boldsymbol{\zeta},12}(\boldsymbol{I}_n \otimes \boldsymbol{w})]$ and

$$\boldsymbol{C}_{\boldsymbol{\zeta},22} = \begin{pmatrix} \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)} & \boldsymbol{Q}_{23} \\ \cdot & \boldsymbol{Q}_{33} \end{pmatrix}.$$

It follows from Result 1 that

$$L(\boldsymbol{x}_c) \,|\, \boldsymbol{Y}_{S_n}, \boldsymbol{L}_{S_n^c}, \boldsymbol{\zeta} \;\sim\; t(m_c(\boldsymbol{\zeta}),\, \widehat{\sigma_c^2(\boldsymbol{\zeta})}\, R_c(\boldsymbol{\zeta}),\, 2n-1)\,, \tag{7}$$

where $m_c(\boldsymbol{\zeta}) = \widehat{\beta}_{c,0}(\boldsymbol{\zeta}) + \boldsymbol{c}_{\boldsymbol{\zeta},12}^{\top}\boldsymbol{C}_{\boldsymbol{\zeta},22}^{-1}\left(\boldsymbol{Z}_c - \widehat{\beta}_{c,0}(\boldsymbol{\zeta})\mathbf{1}_{2n}\right)$, $\widehat{\beta}_{c,0}(\boldsymbol{\zeta}) = (\mathbf{1}_{2n}^{\top}\boldsymbol{C}_{\boldsymbol{\zeta},22}^{-1}\boldsymbol{Z}_c)/(\mathbf{1}_{2n}^{\top}\boldsymbol{C}_{\boldsymbol{\zeta},22}^{-1}\mathbf{1}_{2n})$, $\widehat{\sigma_c^2(\boldsymbol{\zeta})} = [\boldsymbol{Z}_c^{\top}\boldsymbol{C}_{\boldsymbol{\zeta},22}^{-1}\boldsymbol{Z}_c - \widehat{\beta}_{c,0}^2(\boldsymbol{\zeta})\,(\mathbf{1}_{2n}^{\top}\boldsymbol{C}_{\boldsymbol{\zeta},22}^{-1}\mathbf{1}_{2n})]/(2n-1)$, and $R_c(\boldsymbol{\zeta}) = \boldsymbol{w}^{\top}\boldsymbol{R}_{\boldsymbol{\zeta},11}\boldsymbol{w} - \boldsymbol{c}_{\boldsymbol{\zeta},12}^{\top}\boldsymbol{C}_{\boldsymbol{\zeta},22}^{-1}\boldsymbol{c}_{\boldsymbol{\zeta},12} + (1 - \boldsymbol{c}_{\boldsymbol{\zeta},12}^{\top}\boldsymbol{C}_{\boldsymbol{\zeta},22}^{-1}\mathbf{1}_{2n})^2/(\mathbf{1}_{2n}^{\top}\boldsymbol{C}_{\boldsymbol{\zeta},22}^{-1}\mathbf{1}_{2n})$.

The distribution (7) results in the posterior mean of $I_n(\boldsymbol{x}_c)$ given $\boldsymbol{Y}_{S_n}$, $\boldsymbol{L}_{S_n^c}$ and $\boldsymbol{\zeta}$:

$$\begin{aligned}\mathrm{E}\left\{\, I_n(\boldsymbol{x}_c) \mid \boldsymbol{Y}_{S_n}, \boldsymbol{L}_{S_n^c}, \boldsymbol{\zeta} \,\right\} &= (L_{1:n} - m_c)\, T_{2n-1}\left(\frac{L_{1:n} - m_c}{\sqrt{\widehat{\sigma_c^2}\, R_c}}\right) + \\ &\frac{1}{2(n-1)}\left[(2n-1)\sqrt{\widehat{\sigma_c^2}\, R_c} + \frac{(L_{1:n} - m_c)^2}{\sqrt{\widehat{\sigma_c^2}\, R_c}}\right] t_{2n-1}\left(\frac{L_{1:n} - m_c}{\sqrt{\widehat{\sigma_c^2}\, R_c}}\right),\end{aligned} \tag{8}$$

where $T_\nu(\cdot)$ and $t_\nu(\cdot)$ denote the standard t cumulative distribution function and density function with ν degrees of freedom. Appendix A.1 contains a discussion of some computational simplifications regarding this calculation. The two terms in (8) have simple intuitive interpretations. The first term is "large" when the prediction m_c of $L(\boldsymbol{x}_c)$ is "small." The second term is "large" when the prediction uncertainty $\widehat{\sigma_c^2}\, R_c$ of $L(\cdot)$ at $\boldsymbol{x}_c$ is "large." Thus, the posterior expected improvement criterion in (5) will choose $\boldsymbol{t}_{c,n+1}$ roughly in an area of the control variable space where $L(\cdot)$ is predicted to be small *or* where there is high uncertainty in the prediction of $L(\cdot)$.

The posterior expected improvement of (5) is estimated by Monte Carlo simulation. A random sample of size N_c is obtained from the posterior distribution of $\boldsymbol{L}_{S_n^c}$ given $\boldsymbol{Y}_{S_n}$ and $\boldsymbol{\zeta}$. For each sample, the minimum loss $L_{1:n}$ is obtained and the expectation in (8) is computed. The estimate of the posterior expected improvement is taken to be the average of these quantities over all N_c observations. Note that the posterior expected improvement can be

estimated at any control site $\boldsymbol{x}_c$ using the *same* Monte Carlo sample. This follows from the fact that the posterior distribution of $\boldsymbol{L}_{S_n^c}$ given $\boldsymbol{Y}_{S_n}$ and $\boldsymbol{\zeta}$ does not depend on $\boldsymbol{x}_c$. We now obtain this posterior distribution and describe how to sample from it.

Proceeding along the same lines as before we find that, given $(\beta_0, \sigma^2, \boldsymbol{\zeta}^\top)^\top$, the random vector $(\boldsymbol{L}_{S_n^c}^\top, \boldsymbol{Y}_{S_n}^\top)^\top$ has a joint Gaussian distribution with mean $\beta_0 \mathbf{1}_{2n}$ and covariance matrix

$$\sigma^2 \begin{pmatrix} \boldsymbol{Q}_{33} & \boldsymbol{Q}_{23}^\top \\ \cdot & \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)} \end{pmatrix}.$$

From Result 1, the posterior distribution of $\boldsymbol{L}_{S_n^c}$ given $\boldsymbol{Y}_{S_n}$ and $\boldsymbol{\zeta}$ is n-variate t:

$$\boldsymbol{L}_{S_n^c} \,|\, \boldsymbol{Y}_{S_n}, \boldsymbol{\zeta} \;\sim\; \mathcal{T}_n\left(\boldsymbol{m}_{\boldsymbol{\zeta},n},\, \widehat{\sigma_n^2(\boldsymbol{\zeta})}\boldsymbol{R}_{\boldsymbol{\zeta},n},\, n-1\right), \tag{9}$$

where $\boldsymbol{m}_{\boldsymbol{\zeta},n} = \widehat{\beta}_{n,0}(\boldsymbol{\zeta})\,\mathbf{1}_n + \boldsymbol{Q}_{23}^\top \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)}^{-1}(\boldsymbol{Y}_{S_n} - \widehat{\beta}_{n,0}(\boldsymbol{\zeta})\,\mathbf{1}_n)$, $\widehat{\sigma_n^2(\boldsymbol{\zeta})} = [\boldsymbol{Y}_{S_n}^\top \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)}^{-1}\boldsymbol{Y}_{S_n} - \widehat{\beta}_{n,0}^2(\boldsymbol{\zeta})\,(\mathbf{1}_n^\top \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)}^{-1}\mathbf{1}_n)]/(n-1)$, and $\widehat{\beta}_{n,0}(\boldsymbol{\zeta}) = (\mathbf{1}_n^\top \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)}^{-1}\boldsymbol{Y}_{S_n})/(\mathbf{1}_n^\top \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)}^{-1}\mathbf{1}_n)$. Also, $\boldsymbol{R}_{\boldsymbol{\zeta},n} = \boldsymbol{Q}_{33} - \boldsymbol{Q}_{23}^\top \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)}^{-1}\boldsymbol{Q}_{23} + [\mathbf{1}_n - \boldsymbol{Q}_{23}^\top \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)}^{-1}\mathbf{1}_n][\mathbf{1}_n - \boldsymbol{Q}_{23}^\top \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)}^{-1}\mathbf{1}_n]^\top/(\mathbf{1}_n^\top \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)}^{-1}\mathbf{1}_n)$. We sample from the posterior distribution of (9) in two steps.

1. Sample from a chi-square distribution with $n-1$ degrees of freedom and denote the result by χ_{n-1}^2. Set $\widetilde{\sigma^2(\boldsymbol{\zeta})} = (n-1)\,\widehat{\sigma_n^2(\boldsymbol{\zeta})}/\chi_{n-1}^2$.

2. Sample $\boldsymbol{L}_{S_n^c}$, given $\boldsymbol{Y}_{S_n}$ and $\boldsymbol{\zeta}$, from a n-variate normal distribution with mean $\boldsymbol{m}_{\boldsymbol{\zeta},n}$ and covariance matrix $\widetilde{\sigma^2(\boldsymbol{\zeta})}\,\boldsymbol{R}_{\boldsymbol{\zeta},n}$.

3.2.2 Selection of environmental variables

The selection of the environmental variable site requires the evaluation of the expectation in (4). Letting $J_n(\boldsymbol{x}_e) = [\widehat{L}_{n+1}(\boldsymbol{t}_{c,n+1}) - L(\boldsymbol{t}_{c,n+1})]^2$ be the squared prediction error at $\boldsymbol{t}_{c,n+1}$

and $\widehat{L}_{n+1}(\cdot)$, this is performed by first noting that

$$\mathrm{E}\{\, J_n(\boldsymbol{x}_e) \,|\, \boldsymbol{Y}_{S_n}, \boldsymbol{\zeta} \,\} = \mathrm{E}_{Y(\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_e) \,|\, \boldsymbol{Y}_{S_n}, \boldsymbol{\zeta}} \left\{ \mathrm{E}\{\, J_n(\boldsymbol{x}_e) \,|\, \boldsymbol{Y}_{S_n}, Y(\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_e), \boldsymbol{\zeta} \,\} \right\} . \tag{10}$$

Recall that $\widehat{L}_{n+1}(\boldsymbol{t}_{c,n+1})$ is taken to be the posterior mean of $L(\boldsymbol{t}_{c,n+1})$ given $\boldsymbol{Y}_{S_n}$, $Y(\boldsymbol{t}_{c,n+1}, \cdot)$ and $\boldsymbol{\zeta}$. Hence, an analytic expression for the inner expectation can be obtained upon specification of this posterior distribution. Define the $n_e \times 1$ vector $\boldsymbol{Y}_3$ as $\boldsymbol{Y}_3 = [Y(\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_{e,1}), \ldots, Y(\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_{e,n_e})]^\top$. Here $\boldsymbol{Y}_3$ is the random vector of responses at the $(n+1)$-st control site $\boldsymbol{t}_{c,n+1}$ and each environmental support point. We require the following correlation matrices and vectors:

$$\begin{array}{cc} \mathrm{Corr}\,[\boldsymbol{Y}_3, \boldsymbol{Y}_3^\top] = \boldsymbol{R}_{\boldsymbol{\zeta},33}\,, & \mathrm{Corr}\,[\boldsymbol{Y}_3, \boldsymbol{Y}_{S_n}^\top] = \boldsymbol{R}_{\boldsymbol{\zeta},(3,S_n)}\,, \\ \mathrm{Corr}\,[\boldsymbol{Y}_3, Y(\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_e)] = \boldsymbol{r}_{\boldsymbol{\zeta},3}\,, & \mathrm{Corr}\,[\boldsymbol{Y}_{S_n}, Y(\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_e)] = \boldsymbol{r}_{\boldsymbol{\zeta},S_n}\,. \end{array} \tag{11}$$

In addition, let $\boldsymbol{Z}_e^\top = (\boldsymbol{Y}_{S_n}^\top, Y(\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_e))$, $\boldsymbol{e}_{\boldsymbol{\zeta},12}^\top = \boldsymbol{w}^\top [\boldsymbol{R}_{\boldsymbol{\zeta},(3,S_n)} \quad \boldsymbol{r}_{\boldsymbol{\zeta},3}]$ and

$$\boldsymbol{E}_{\boldsymbol{\zeta},22} = \begin{pmatrix} \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)} & \boldsymbol{r}_{\boldsymbol{\zeta},S_n} \\ \cdot & 1 \end{pmatrix} .$$

It follows from Result 1 that the posterior distribution of $L(\boldsymbol{t}_{c,n+1})$, given $\boldsymbol{Y}_{S_n}$, $Y(\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_e)$ and $\boldsymbol{\zeta}$, is univariate t:

$$L(\boldsymbol{t}_{c,n+1}) \,|\, \boldsymbol{Y}_{S_n}, Y(\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_e), \boldsymbol{\zeta} \;\sim\; t(m_e(\boldsymbol{\zeta}),\, \widehat{\sigma_e^2(\boldsymbol{\zeta})}\, R_e(\boldsymbol{\zeta}),\, n)\,, \tag{12}$$

where

$$m_e(\boldsymbol{\zeta}) = \widehat{\beta}_{e,0}(\boldsymbol{\zeta}) + \boldsymbol{e}_{\boldsymbol{\zeta},12}^\top \boldsymbol{E}_{\boldsymbol{\zeta},22}^{-1} \left(\boldsymbol{Z}_e - \widehat{\beta}_{e,0}(\boldsymbol{\zeta}) \mathbf{1}_{n+1}\right), \tag{13}$$

$\widehat{\sigma_e^2(\boldsymbol{\zeta})} = [\boldsymbol{Z}_e^\top \boldsymbol{E}_{\boldsymbol{\zeta},22}^{-1} \boldsymbol{Z}_e - \widehat{\beta}_{e,0}^2(\boldsymbol{\zeta})\, (\mathbf{1}_{n+1}^\top \boldsymbol{E}_{\boldsymbol{\zeta},22}^{-1} \mathbf{1}_{n+1})]/n$, $R_e(\boldsymbol{\zeta}) = \boldsymbol{w}^\top \boldsymbol{R}_{\boldsymbol{\zeta},33} \boldsymbol{w} - \boldsymbol{e}_{\boldsymbol{\zeta},12}^\top \boldsymbol{E}_{\boldsymbol{\zeta},22}^{-1} \boldsymbol{e}_{\boldsymbol{\zeta},12} + (1 - \boldsymbol{e}_{\boldsymbol{\zeta},12}^\top \boldsymbol{E}_{\boldsymbol{\zeta},22}^{-1} \mathbf{1}_{n+1})^2 / (\mathbf{1}_{n+1}^\top \boldsymbol{E}_{\boldsymbol{\zeta},22}^{-1} \mathbf{1}_{n+1})$, and $\widehat{\beta}_{e,0}(\boldsymbol{\zeta}) = (\mathbf{1}_{n+1}^\top \boldsymbol{E}_{\boldsymbol{\zeta},22}^{-1} \boldsymbol{Z}_e)/(\mathbf{1}_{n+1}^\top \boldsymbol{E}_{\boldsymbol{\zeta},22}^{-1} \mathbf{1}_{n+1})$. The

posterior mean $m_e(\boldsymbol{\zeta})$ is the best linear unbiased predictor of $L(\boldsymbol{t}_{c,n+1})$ based on the design $S_n \cup (\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_e)$. Then,

$$\mathrm{E}\,\{\, J_n(\boldsymbol{x}_e) \,|\, \boldsymbol{Y}_{S_n}, Y(\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_e), \boldsymbol{\zeta}\,\} = \frac{n\,\widehat{\sigma_e^2(\boldsymbol{\zeta})}}{n-2}\, R_e(\boldsymbol{\zeta})\,, \tag{14}$$

which is the variance of the posterior distribution in (12).

A closed form expression for the posterior mean square prediction error of (10) at $\boldsymbol{x}_e$ can be obtained. Let $m_1(\boldsymbol{\zeta})$ denote the posterior mean of $Y(\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_e)$, given $\boldsymbol{Y}_{S_n}$ and $\boldsymbol{\zeta}$, $m_1(\boldsymbol{\zeta}) = \widehat{\beta}_{n,0}(\boldsymbol{\zeta}) + \boldsymbol{r}_{\boldsymbol{\zeta},S_n}^{\top} \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)}^{-1} (\boldsymbol{Y}_{S_n} - \widehat{\beta}_{n,0}(\boldsymbol{\zeta})\mathbf{1}_n)$. Define $\boldsymbol{M}_e^{\top} = (\boldsymbol{Y}_{S_n}^{\top}, m_1(\boldsymbol{\zeta}))$, which is just $\boldsymbol{Z}_e$ with $m_1(\boldsymbol{\zeta})$ in place of $Y(\boldsymbol{t}_{c,n+1}, \boldsymbol{x}_e)$. Taking the outer expectation in (10) of the quantity in (14) gives

$$\begin{aligned}
&\mathrm{E}\,\{\, J_n(\boldsymbol{x}_e) \,|\, \boldsymbol{Y}_{S_n}, \boldsymbol{\zeta}\,\} = \\
&\frac{1}{n-2}\left[\boldsymbol{M}_e^{\top}\left(\boldsymbol{E}_{\boldsymbol{\zeta},22}^{-1} - \frac{\boldsymbol{E}_{\boldsymbol{\zeta},22}^{-1}\mathbf{1}_{n+1}\mathbf{1}_{n+1}^{\top}\boldsymbol{E}_{\boldsymbol{\zeta},22}^{-1}}{\mathbf{1}_{n+1}^{\top}\boldsymbol{E}_{\boldsymbol{\zeta},22}^{-1}\mathbf{1}_{n+1}}\right)\boldsymbol{M}_e + \frac{n-1}{n-3}\,\widehat{\sigma_n^2(\boldsymbol{\zeta})}\right] R_e(\boldsymbol{\zeta})\,.
\end{aligned} \tag{15}$$

The formulas for $\widehat{\beta}_{n,0}(\boldsymbol{\zeta})$ and $\widehat{\sigma_n^2(\boldsymbol{\zeta})}$ were given in conjunction with the posterior distribution of (9). Some computational simplifications involving calculation of the posterior mean square prediction error are discussed in Appendix A.2.

In this presentation, all posterior distributions are given up to the unknown correlation parameter vector $\boldsymbol{\zeta}$. The probability density function of the posterior distribution of $\boldsymbol{\zeta}$, given $\boldsymbol{Y}_{S_n}$, is

$$p(\boldsymbol{\zeta} \,|\, \boldsymbol{Y}_{S_n}) \propto p(\boldsymbol{\zeta})\, \frac{[\widehat{\sigma_n^2(\boldsymbol{\zeta})}]^{-(n-1)/2}}{\sqrt{\mathbf{1}_n^{\top}\boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)}^{-1}\mathbf{1}_n}}\, |\boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)}|^{-1/2}\,, \tag{16}$$

where $p(\boldsymbol{\zeta})$ is a prior distribution on the permissible range of values for the correlation parameters (see Handcock and Stein (1993)). Recall from (1) that $p(\boldsymbol{\zeta}) \propto 1$. It is possible to carry out a fully Bayesian analysis by using (16) to integrate $\boldsymbol{\zeta}$ out of the posterior distributions given above. However, we adopt the simpler approach of setting $\boldsymbol{\zeta}$ equal to its posterior mode and then proceed by substituting this mode for $\boldsymbol{\zeta}$ wherever necessary. The posterior mode is the restricted maximum likelihood (REML) estimator of $\boldsymbol{\zeta}$ (see Cressie (1991), Sec. 2.6.1).

4 Examples

The following examples illustrate the operation of the modified expected improvement algorithm. All calculations are made with the Matérn family of correlation functions described in Section 2. The correlation parameter vector is given by $\boldsymbol{\zeta} = (\theta_1, \ldots, \theta_p, \nu)^\top$. The test functions are taken from Dixon and Szego (1978).

4.1 Branin function

In this example, we assume the response $y(\cdot)$ is the product $y(\boldsymbol{x}) = y_b(15x_1 - 5, 15x_2) \times$ $y_b(15x_3 - 5, 15x_4)$, where

$$y_b(u, v) = (v - \frac{5.1}{4\pi^2}\, u^2 + \frac{5}{\pi}\, u - 6)^2 + 10(1 - \frac{1}{8\pi})\, \cos(u) + 10$$

is the Branin function and $x_1, \ldots, x_4$ lie in $[0, 1]^4$. The Branin function is defined on the (u, v) domain $[-5, 10] \times [0, 15]$ in $\mathbb{R}^2$. We take x_1 and x_4 to be the control variables, $\boldsymbol{x}_c = \{x_1, x_4\}$,

and x_2 and x_3 to be the environmental variables, $\boldsymbol{x}_e = \{x_2, x_3\}$.

The joint distribution of the environmental variables is given in Table 1. The true objective function is obtained from (2) using these weights.

		x_3			
		0.2	0.4	0.6	0.8
x_2	0.25	0.0375	0.0875	0.0875	0.0375
	0.5	0.0750	0.1750	0.1750	0.0750
	0.75	0.0375	0.0875	0.0875	0.0375

Table 1: Probability distribution for x_2 and x_3.

The modified expected improvement algorithm was run twice, once to predict the global maximizer and once to predict the global minimizer of this objective function. Each run of the algorithm was started with the same 40-point maximin distance LHS design generated by the software package ACED. Figure 1 gives perspective plots of the true objective function and the EBLUP of this function based on the 40-point initial design. The global maximizer $\boldsymbol{x}^*$ of the true objective function is located at $(0, 1)$ with $\ell(\boldsymbol{x}^*) = 16,261.37$. The global minimizer $\boldsymbol{x}_*$ is located at $(0.20263, 0.25445)$ with $\ell(\boldsymbol{x}_*) = 323.01174$. Note that the response surface is poorly predicted from the initial design.

Figures 2 and 3 show contour plots of the true objective function along with the projection of (some of) the points added by the modified expected improvement algorithm when finding the maximum and minimum, respectively. The algorithm was run with $N_c = 100$ Monte Carlo samples to estimate the posterior expected improvement. In both cases, the projections of the (same) initial 40-point design onto $x_1 \times x_4$ space, the control variable space, are denoted by open circles. The modified expected improvement algorithm added 19 points to predict

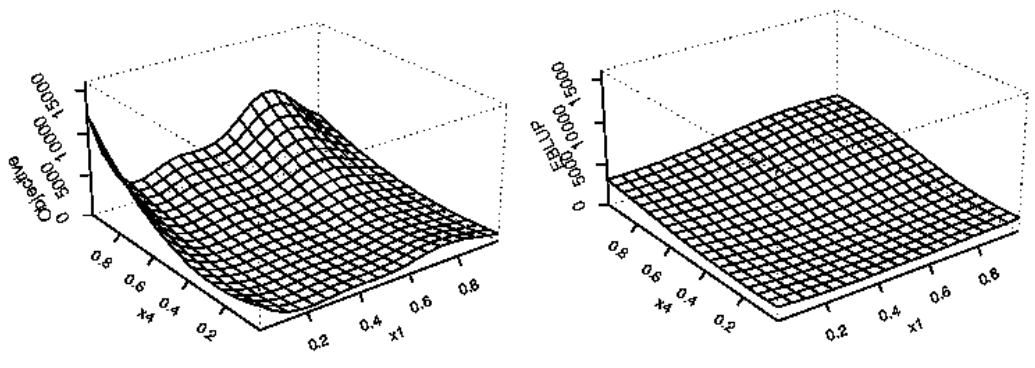

Figure 1: True objective function (left panel) and EBLUP based on initial 40-point design (right panel).

the global maximizer, while it added 116 points to predict the global minimizer. To enhance visibility, a selection of these points are indicated on the plots according to the order in which they were added to the design.

In its search for the global maximum, the modified expected improvement algorithm focused on the regions of the control variable space near the global maximum and the prominent local maximum at $(0.75011, 1)$, with a few searches along the lower boundary due to the large standard errors of prediction there. The algorithm was stopped at the 59-point design because the expected improvements of the points 57–59 are small relative to the expected improvements, on the order of 10^{-1}–10^{2}, observed in previous steps. The algorithm should not be terminated after the observation of a single small expected improvement, because it can get trapped in a local optimum. However, a longer series of small expected improvements suggests that the algorithm can be terminated. Note that "smallness" of the expected improvement is established relative to previously observed expected improvements.

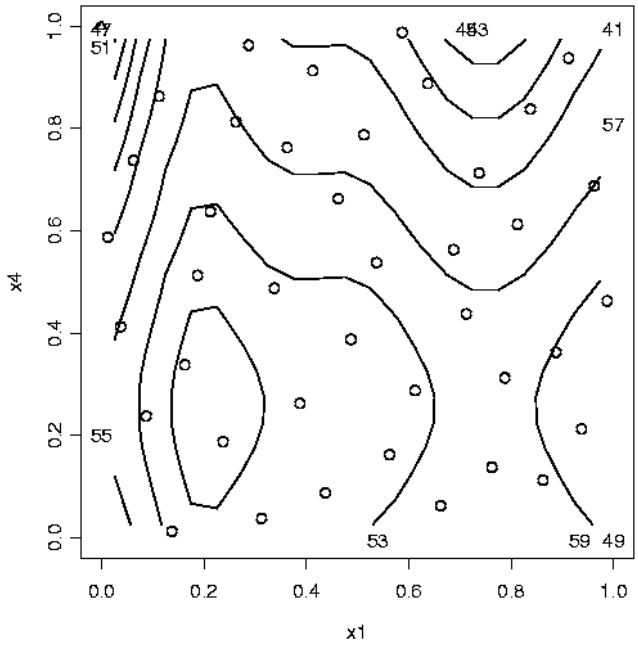

Figure 2: Projection of a selection of points from the sequential design for predicting the global maximizer of the Branin function. The integers sequentially identify points chosen by the algorithm.

The predictor of the global maximizer is taken to be a point $\widehat{\boldsymbol{x}^*}$ that maximizes the EBLUP based on the final 59-point design. This point is $\widehat{\boldsymbol{x}^*} = (0, 1)$.

In searching for the global minimum, the modified expected improvement algorithm heavily visited the region of the true global minimum, and frequently visited the regions of local minima at $(1, 0.25445)$ and $(0.46287, 0.25445)$. Table 2 gives the expected improvement for the last ten points added by the algorithm. The algorithm was terminated at point 156 because the expected improvements appear to have stabilized, and are small relative to pre-

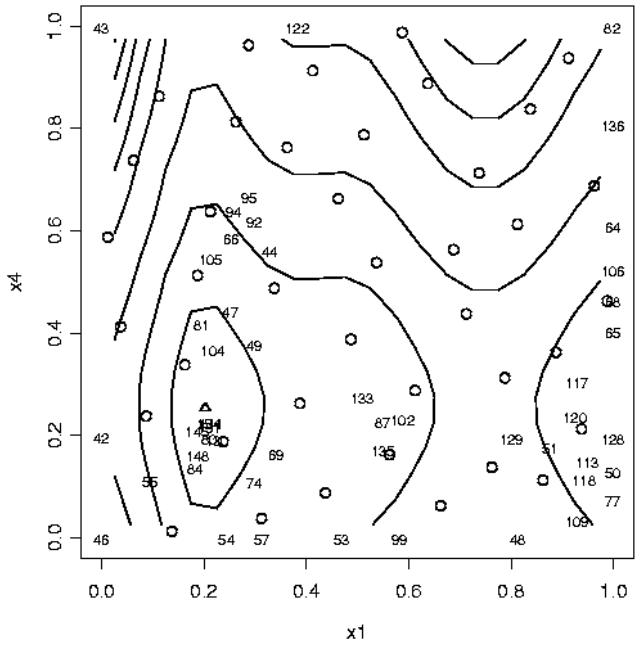

Figure 3: Projection of a selection of points from the sequential design for predicting the global minimizer of the Branin function. The integers sequentially identify points chosen by the algorithm.

vious values. The predictor of the global minimizer based on the EBLUP, calculated from the final 156-point design, is $\widehat{\boldsymbol{x}_*} = (0.21096, 0.23324)$ with $\ell(\widehat{\boldsymbol{x}_*}) = 326.67005$. Thus, the predicted global minimum is within 1.15% of the true global minimum.

All computations were performed on a Sun Ultra 5. For minimization of the Branin function, once correlation parameter estimates had been obtained, the search component of the modified expected improvement algorithm required 35 s to find the first site added and 520 s to find the final site added. This time increases with design size due to the larger

Point	Expected Improvement	Point	Expected Improvement
147	0.49574	152	5.20332×10^{-2}
148	3.651×10^{-2}	153	0.21567
149	7.00628×10^{-2}	154	3.61129×10^{-2}
150	0.16862	155	5.91403×10^{-2}
151	8.05449×10^{-2}	156	2.48648×10^{-2}

Table 2: Expected improvement for last ten points of final design.

systems of linear equations that must be solved. Our correlation parameter estimation algorithm required 270 s to obtain initial REML estimates of the correlation parameters, and 6930 s for final estimates. If the power exponential class of correlation functions given by Welch et al. (1992) is assumed for these calculations, our times are 145 s and 6690 s. If the GaSP (Gaussian Stochastic Process) software developed by W. J. Welch is used to obtain maximum likelihood estimates of the power exponential correlation parameters, these times can be reduced further to 45 s and 2150 s.

4.2 Hartman 6 function

The function

$$z(x_1, \ldots, x_6) = -\sum_{i=1}^{4} c_i \exp\left[-\sum_{j=1}^{6} \alpha_{ij} \left(x_j - p_{ij}\right)^2\right],$$

defined on the six-dimensional unit hypercube $[0,1]^6$ is known as the Hartman 6 function, where the quantities $\{c_i\}$, $\{\alpha_{ij}\}$ and $\{p_{ij}\}$ are given in Table 3. The underlying response is taken to be a logarithmic transformation of the Hartman 6 function, $y(\boldsymbol{x}) = -\log(-z(\boldsymbol{x}))$. The control variables are taken to be $\boldsymbol{x}_c = \{x_1, x_2, x_4, x_6\}$ and the environmental variables are $\boldsymbol{x}_e = \{x_3, x_5\}$. We assume environmental variables are distributed independently, with the marginal distribution of each variable given in Table 4. It is a discretization of a triangular

i	$\alpha_{ij}, j = 1, \ldots, 6$						c_i
1	10	3	17	3.5	1.7	8	1
2	.05	10	17	.1	8	14	1.2
3	3	3.5	1.7	10	17	8	3
4	17	8	.05	10	.1	14	3.2

i	$p_{ij}, j = 1, \ldots, 6$					
1	.1312	.1696	.5569	.0124	.8283	.5886
2	.2329	.4135	.8307	.3736	.1004	.9991
3	.2348	.1451	.3522	.2883	.3047	.6650
4	.4047	.8828	.8732	.5743	.1091	.0381

Table 3: Coefficients for Hartman 6 function.

distribution with support on seven points. The resulting joint distribution provides the weights used in (2) to obtain the true objective function.

x	0.125	0.25	0.375	0.5	0.625	0.75	0.875
Probability	9/128	1/8	3/16	15/64	3/16	1/8	9/128

Table 4: Marginal probability distribution for x_3 (and x_5).

The modified expected improvement algorithm was run to predict the global minimizer of this objective function. The posterior expected improvement was estimated with $N_c = 100$ Monte Carlo samples. The algorithm started with a 50-point maximin distance LHS design generated by the software package ACED, and it added 32 points until it stopped. The algorithm was stopped at the 82-point design because the expected improvements of the last three points added are small relative to the expected improvements, on the order of 10^{-3}–10^{-1}, observed previously. Table 5 gives the expected improvement for the last ten points added by the algorithm.

Point	Expected Improvement	Point	Expected Improvement
73	5.01422×10^{-5}	78	2.57626×10^{-3}
74	2.92048×10^{-3}	79	9.67359×10^{-4}
75	8.58409×10^{-3}	80	1.60879×10^{-6}
76	8.59993×10^{-7}	81	2.03093×10^{-6}
77	6.19091×10^{-3}	82	2.22035×10^{-7}

Table 5: Expected improvement for last ten points of final design.

The algorithm was not stopped at points 73 or 76 because the small expected improvements at these points were followed by substantially larger expected improvements. The true global minimizer of the objective function is $\boldsymbol{x}_* = (0.40459,\ 0.88231,\ 0.57389,\ 0.03865)$, with $\ell(\boldsymbol{x}_*) = -1.13630$ (-3.11522 on the original scale). The minimizer of the EBLUP based on the 82-point final design is $\widehat{\boldsymbol{x}_*} = (0.38928,\ 0.87683,\ 0.58822,\ 0.03835)$, with $\ell(\widehat{\boldsymbol{x}_*}) = -1.13018$ (-3.09621 on the original scale). Thus, the predicted global minimum is within 1% of the true global minimum.

The search component of the modified expected improvement algorithm required 50 s to find the first site added, and 220 s to find the final site added. Our REML estimation of the Matérn correlation parameters required 2140 s and 4230 s, and our times were 1105 s and 3100 s for the power exponential correlation parameters. Maximum likelihood estimation of the power exponential parameters using GaSP required 115 s and 415 s.

5 Discussion

To obtain estimates of the correlation parameters, the modified expected improvement algorithm requires the specification of an initial experimental design at which the responses are calculated. We have used maximin distance LHS designs; however, other initial designs have equal intuitive appeal. For example, the cascading LHS designs of Handcock (1991) contain both space-filling and local components. The latter designs may yield estimates of the process variance and correlation smoothness parameters that are superior to those that are obtained from designs that only contain a space-filling component (such as maximin

distance LHS designs). Multi-stage initial designs that contain a space-filling stage to obtain estimates of the correlation parameters, followed by a prediction-based stage to improve the quality of the predictor, could also be investigated. The choice of initial designs is an area of active research.

An important issue not considered in this paper is the choice of sample size for the initial design. The initial design should not be too small, because this can result in a poor estimate of the correlation parameters and can substantially increase the number of points that will need to be added by the sequential optimization strategy. An initial design that is too large risks wasting observations that are not needed. The sizes of the initial designs we used were chosen based on informal guidelines from the literature for predicting the response (Jones et al. (1998)). To our knowledge there are no formal results regarding the choice of sample size for computer experiments, and this is also an area for further research.

It is important to note that no formal rule for stopping the modified expected improvement algorithm has been given. There are two major reasons for this. First, the expected improvements at each stage of the algorithm are not monotone decreasing. The estimates of the correlation parameters are updated after each new point is added and this, combined with information in the new observed response, affects the prediction quality and uncertainty throughout the control variable space. Thus, although the expected improvement generally decreases, the circumstances present at any particular stage of the algorithm do not prohibit finding expected improvements that are larger than previously observed. Second, the size of the expected improvements depends on the scale of the response, as can be seen from

the examples in Section 4. These two factors preclude easy identification of a relative or an absolute stopping criterion.

The number of points that need to be added by the algorithm before termination can depend heavily on the properties of the objective near the global optimum. If the objective is relatively flat with little curvature near the global optimum, or has several local optima with objective function values near that of the global optimum, the algorithm will run much longer than if the global optimum is clearly identifiable. The Branin function example of Section 4 illustrates this phenomenon, as the global maximum stands out clearly while the surface is flat around the global minimum, with two local minima having similar function values to the global minimum.

This paper focuses on optimizing the mean of $y(\boldsymbol{x}_c, \boldsymbol{X}_e)$. In practice, other functionals of this distribution may be of greater importance. For example, the median or other quantiles of $y(\boldsymbol{x}_c, \boldsymbol{X}_e)$, or $\mathrm{Var}(\, y(\boldsymbol{x}_c, \boldsymbol{X}_e)\,)$, may be of interest.

Work is underway to extend the modified expected improvement algorithm to the problem of optimizing one objective function subject to a constraint on another objective. This setting occurs often in practice, see for example Chang et al. (1999) and Schonlau et al. (1998). This problem is complicated by the need to estimate the correlation structure between the two objectives.

ACKNOWLEDGMENT

This work was sponsored, in part, by the National Institutes of Health, Grant AR42737-01. We thank the reviewers and editor, whose suggestions resulted in improvements to the paper.

A Computational considerations

A.1 Improvement criterion

Each component of the matrix in (6) is needed in the computation of the improvement criterion (8). It is possible to derive simpler expressions for some of these components by taking advantage of the product structure of the Matérn correlation function. More generally, suppose the correlation function has the following form: $R(\boldsymbol{u}, \boldsymbol{v}) = \prod_{j=1}^{p} R_j(u^{(j)}, v^{(j)})$, where $u^{(j)}$ $(v^{(j)})$ is the j-th coordinate of the p-dimensional vector $\boldsymbol{u}$ $(\boldsymbol{v})$, and the R_j are one-dimensional correlation functions. Without loss of generality, we suppose that the p_c control and p_e environmental components of the input vector $\boldsymbol{u}$ are grouped as follows: $\boldsymbol{u}^\top = (\boldsymbol{u}_c^\top, \boldsymbol{u}_e^\top)$, where $p = p_c + p_e$. We can write $R(\cdot)$ as

$$R(\boldsymbol{u}, \boldsymbol{v}) = \prod_{j=1}^{p_c} R_j(u_c^{(j)}, v_c^{(j)}) \prod_{j=1}^{p_e} R_{j+p_c}(u_e^{(j)}, v_e^{(j)}), \tag{17}$$

where $u_c^{(j)}$ $(v_c^{(j)})$ and $u_e^{(j)}$ $(v_e^{(j)})$ are the j-th control and environmental components of the partitioned vector $\boldsymbol{u}$ $(\boldsymbol{v})$.

Let

$$\boldsymbol{R}_{\boldsymbol{\zeta},(S_n^c,S_n^c)} = \left(\prod_{k=1}^{p_c} R_k(\boldsymbol{t}_{c,i}^{(k)}, \boldsymbol{t}_{c,j}^{(k)}) \right) \quad \text{and} \quad \boldsymbol{R}_{\boldsymbol{\zeta},ee} = \left(\prod_{k=1}^{p_e} R_{k+p_c}(\boldsymbol{x}_{e,i}^{(k)}, \boldsymbol{x}_{e,j}^{(k)}) \right). \tag{18}$$

Here $\boldsymbol{R}_{\boldsymbol{\zeta},(S_n^c,S_n^c)}$ is the $n \times n$ correlation matrix formed from the control variable components of the n-point design S_n, i.e. from the points in S_n^c. The $n_e \times n_e$ correlation matrix $\boldsymbol{R}_{\boldsymbol{\zeta},ee}$ is formed from the support points of the environmental variable distribution. From (17), $\boldsymbol{R}_{\boldsymbol{\zeta},11} = \boldsymbol{R}_{\boldsymbol{\zeta},ee}$ and $\boldsymbol{R}_{\boldsymbol{\zeta},22} = \boldsymbol{R}_{\boldsymbol{\zeta},(S_n^c,S_n^c)} \otimes \boldsymbol{R}_{\boldsymbol{\zeta},ee}$. The $(3,3)$ entry in the matrix

of (6) simplifies nicely: $(\boldsymbol{I}_n \otimes \boldsymbol{w}^\top)\boldsymbol{R}_{\boldsymbol{\zeta},22}(\boldsymbol{I}_n \otimes \boldsymbol{w}) = (\boldsymbol{I}_n \otimes \boldsymbol{w}^\top)(\boldsymbol{R}_{\boldsymbol{\zeta},(S_n^c,S_n^c)} \otimes \boldsymbol{R}_{\boldsymbol{\zeta},ee})(\boldsymbol{I}_n \otimes \boldsymbol{w}) = (\boldsymbol{w}^\top \boldsymbol{R}_{\boldsymbol{\zeta},ee}\boldsymbol{w})\, \boldsymbol{R}_{\boldsymbol{\zeta},(S_n^c,S_n^c)}$.

Let $S_n^e = \{\boldsymbol{t}_{e,1}, \ldots, \boldsymbol{t}_{e,n}\}$ denote the environmental variable portion of the n-point design S_n. Define the $n \times n_e$ matrix $\boldsymbol{R}_{\boldsymbol{\zeta},(S_n^e,e)}$ as follows:

$$\boldsymbol{R}_{\boldsymbol{\zeta},(S_n^e,e)} = \left(\prod_{k=1}^{p_e} R_{k+p_c}(\boldsymbol{t}_{e,i}^{(k)}, \boldsymbol{x}_{e,j}^{(k)}) \right) . \tag{19}$$

This is the cross-correlation matrix formed between the environmental variable components of the design and the support points of the environmental variable distribution. With $\boldsymbol{R}_{\boldsymbol{\zeta},(S_n^e,e)}[\, i\, ;\, \cdot\,]$ denoting the i-th row of this matrix, (17) can be used to show that

$$\boldsymbol{R}_{\boldsymbol{\zeta},(S_n,2)}(\boldsymbol{I}_n \otimes \boldsymbol{w}) = \text{diag}\, \left(\boldsymbol{R}_{\boldsymbol{\zeta},(S_n^e,e)}[\, i\, ;\, \cdot\,]\, \boldsymbol{w} \right)\, \boldsymbol{R}_{\boldsymbol{\zeta},(S_n^c,S_n^c)}\,, \tag{20}$$

where $\text{diag}\,(\, a_i\,)$ denotes a diagonal matrix having the $\{a_i\}$ as elements. The diagonal matrix in (20) is denoted by $\boldsymbol{D}_{\boldsymbol{\zeta},(S_n^e,e)}$.

Let $\boldsymbol{r}_{\boldsymbol{\zeta},(S_n^c,c)} = \left(\prod_{k=1}^{p_c} R_k(\boldsymbol{t}_{c,1}^{(k)}, \boldsymbol{x}_c^{(k)}), \ldots, \prod_{k=1}^{p_c} R_k(\boldsymbol{t}_{c,n}^{(k)}, \boldsymbol{x}_c^{(k)})\right)^\top$ be the vector of correlations involving the elements of S_n^c and an arbitrary control site $\boldsymbol{x}_c$. Application of (17) establishes the following: $\boldsymbol{R}_{\boldsymbol{\zeta},(1,S_n)}^\top = \text{diag}\, \left(\prod_{k=1}^{p_c} R_k(\boldsymbol{t}_{c,i}^{(k)}, \boldsymbol{x}_c^{(k)}) \right)\, \boldsymbol{R}_{\boldsymbol{\zeta},(S_n^e,e)}$ and $(\boldsymbol{I}_n \otimes \boldsymbol{w}^\top)\boldsymbol{R}_{\boldsymbol{\zeta},12}^\top \boldsymbol{w} = (\boldsymbol{w}^\top \boldsymbol{R}_{\boldsymbol{\zeta},ee}\boldsymbol{w})\, \boldsymbol{r}_{\boldsymbol{\zeta},(S_n^c,c)}$.

Several inversions of the matrix $\boldsymbol{C}_{\boldsymbol{\zeta},22}$ are needed to compute the expected improvement. These inversions are carried out by solving appropriate systems of linear equations using iterative refinement. Incorporating some of the simplifications presented above, the generic system that must be solved is:

$$\begin{pmatrix} \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)} & \boldsymbol{D}_{\boldsymbol{\zeta},(S_n^e,e)}\boldsymbol{R}_{\boldsymbol{\zeta},(S_n^c,S_n^c)} \\ \boldsymbol{R}_{\boldsymbol{\zeta},(S_n^c,S_n^c)}\boldsymbol{D}_{\boldsymbol{\zeta},(S_n^e,e)} & (\boldsymbol{w}^\top \boldsymbol{R}_{\boldsymbol{\zeta},ee}\boldsymbol{w})\, \boldsymbol{R}_{\boldsymbol{\zeta},(S_n^c,S_n^c)} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{pmatrix} .$$

This $2n \times 2n$ system is solved by finding the solutions to two $n \times n$ systems of linear equations. First, $\boldsymbol{x}_1$ is obtained by solving

$$\left(\boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)} - \frac{\boldsymbol{D}_{\boldsymbol{\zeta},(S_n^e,e)} \boldsymbol{R}_{\boldsymbol{\zeta},(S_n^c,S_n^c)} \boldsymbol{D}_{\boldsymbol{\zeta},(S_n^e,e)}}{\boldsymbol{w}^\top \boldsymbol{R}_{\boldsymbol{\zeta},ee} \boldsymbol{w}} \right) \boldsymbol{x}_1 = \boldsymbol{b}_1 - \frac{\boldsymbol{D}_{\boldsymbol{\zeta},(S_n^e,e)} \boldsymbol{b}_2}{\boldsymbol{w}^\top \boldsymbol{R}_{\boldsymbol{\zeta},ee} \boldsymbol{w}} .$$

Second, the vector $\boldsymbol{x}_3$ is obtained as the solution to the system $\boldsymbol{R}_{\boldsymbol{\zeta},(S_n^c,S_n^c)} \boldsymbol{x}_3 = \boldsymbol{b}_2$. Then $\boldsymbol{x}_2$ is computed as $\boldsymbol{x}_2 = (\boldsymbol{x}_3 - \boldsymbol{D}_{\boldsymbol{\zeta},(S_n^e,e)} \boldsymbol{x}_1)/(\boldsymbol{w}^\top \boldsymbol{R}_{\boldsymbol{\zeta},ee} \boldsymbol{w})$.

A.2 Posterior mean square prediction error criterion

The correlation matrices and vectors in (11) are needed for the calculation of the posterior mean square prediction error of (15). We present simplified expressions for these quantities, assuming the product correlation structure introduced in the previous section.

It is clear that $\boldsymbol{R}_{\boldsymbol{\zeta},33} = \boldsymbol{R}_{\boldsymbol{\zeta},ee}$, where $\boldsymbol{R}_{\boldsymbol{\zeta},ee}$ is defined in (18). Let $\boldsymbol{r}_{\boldsymbol{\zeta},(S_n^e,e)} = \left(\prod_{k=1}^{p_e} R_{k+p_c}(\boldsymbol{t}_{e,1}^{(k)}, \boldsymbol{x}_e^{(k)}), \ldots, \prod_{k=1}^{p_e} R_{k+p_c}(\boldsymbol{t}_{e,n}^{(k)}, \boldsymbol{x}_e^{(k)}) \right)^\top$ denote the vector of correlations involving the elements of S_n^e and an arbitrary environmental site $\boldsymbol{x}_e$, and $\boldsymbol{D}_{\boldsymbol{\zeta},(S_n^c,n+1)} = \text{diag}\left(\prod_{k=1}^{p_c} R_k(\boldsymbol{t}_{c,i}^{(k)}, \boldsymbol{t}_{c,n+1}^{(k)}) \right)$ denote the diagonal matrix of correlations involving the elements of S_n^c and the $(n+1)$-st control site $\boldsymbol{t}_{c,n+1}$ found according to (3). Then $\boldsymbol{R}_{\boldsymbol{\zeta},(3,S_n)}^\top = \boldsymbol{D}_{\boldsymbol{\zeta},(S_n^c,n+1)} \boldsymbol{R}_{\boldsymbol{\zeta},(S_n^e,e)}$ and $\boldsymbol{r}_{\boldsymbol{\zeta},S_n} = \boldsymbol{D}_{\boldsymbol{\zeta},(S_n^c,n+1)} \boldsymbol{r}_{\boldsymbol{\zeta},(S_n^e,e)}$, where $\boldsymbol{R}_{\boldsymbol{\zeta},(S_n^e,e)}$ is defined in (19). Finally, we note that $\boldsymbol{r}_{\boldsymbol{\zeta},3} = \left(\prod_{k=1}^{p_e} R_{k+p_c}(\boldsymbol{x}_{e,1}^{(k)}, \boldsymbol{x}_e^{(k)}), \ldots, \prod_{k=1}^{p_e} R_{k+p_c}(\boldsymbol{x}_{e,n_e}^{(k)}, \boldsymbol{x}_e^{(k)}) \right)^\top$ is the vector of correlations involving the support points of the environmental variable distribution and an arbitrary environmental site $\boldsymbol{x}_e$.

Several inversions of the matrix $\boldsymbol{E}_{\boldsymbol{\zeta},22}$ are needed to compute the posterior mean square prediction error. These inversions are carried out by solving appropriate systems of linear

equations using iterative refinement. Incorporating some of the simplifications presented above, the generic system that must be solved is:

$$\begin{pmatrix} \boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)} & \boldsymbol{D}_{\boldsymbol{\zeta},(S_n^c,n+1)}\boldsymbol{r}_{\boldsymbol{\zeta},(S_n^e,e)} \\ \boldsymbol{r}^{\top}_{\boldsymbol{\zeta},(S_n^e,e)}\boldsymbol{D}_{\boldsymbol{\zeta},(S_n^c,n+1)} & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{b}_1 \\ b_2 \end{pmatrix} .$$

This $(n+1)\times(n+1)$ system is solved by finding the solution $\boldsymbol{x}_1$ to an $n \times n$ system of linear equations and then calculating x_2. First, $\boldsymbol{x}_1$ is obtained by solving $\Big(\boldsymbol{R}_{\boldsymbol{\zeta},(S_n,S_n)} - \boldsymbol{D}_{\boldsymbol{\zeta},(S_n^c,n+1)} \boldsymbol{r}_{\boldsymbol{\zeta},(S_n^e,e)}\boldsymbol{r}^{\top}_{\boldsymbol{\zeta},(S_n^e,e)}\boldsymbol{D}_{\boldsymbol{\zeta},(S_n^c,n+1)}\Big)\boldsymbol{x}_1 = \boldsymbol{b}_1 - b_2\boldsymbol{D}_{\boldsymbol{\zeta},(S_n^c,n+1)}\boldsymbol{r}_{\boldsymbol{\zeta},(S_n^e,e)}$. Second, x_2 is computed as $x_2 = b_2 - \boldsymbol{r}^{\top}_{\boldsymbol{\zeta},(S_n^e,e)}\boldsymbol{D}_{\boldsymbol{\zeta},(S_n^c,n+1)}\boldsymbol{x}_1$.

References

Bernardo, M. C., Buck, R., Liu, L., Nazaret, W. A., Sacks, J. and Welch, W. J. (1992). Integrated circuit design optimization using a sequential strategy. *IEEE Transactions on Computer-Aided Design* **11**, 361–372.

Chang, P. B., Williams, B. J., Notz, W. I., Santner, T. J. and Bartel, D. L. (1999). Robust optimization of total joint replacements incorporating environmental variables. *Journal of Biomechanical Engineering* **121**, 304–310.

Cramér, H. and Leadbetter, M. R. (1967). *Stationary and Related Stochastic Processes*. J. Wiley, New York.

Cressie, N. A. C. (1991). *Statistics for Spatial Data*. J. Wiley, New York.

Currin, C., Mitchell, T. J., Morris, M. D. and Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association* **86**, 953–963.

Dixon, L. C. W. and Szego, G. P. (1978). The global optimisation problem: an introduction. In *Towards Global Optimisation*, Vol. 2 (L. C. W. Dixon and G. P. Szego (eds)), pp. 1–15, North Holland, Amsterdam.

Handcock, M. S. (1991). On cascading latin hypercube designs and additive models for experiments. *Commun. Statist.—Theory Meth.* **20**, 417–439.

Handcock, M. S. and Stein, M. L. (1993). A Bayesian analysis of kriging. *Technometrics* **35**, 403–410.

Haylock, R. G. and O'Hagan, A. (1996). On inference for outputs of computationally expensive algorithms with uncertainty on the inputs. In *Bayesian Statistics*, Vol. 5 (J. Bernardo, J. Berger, A. Dawid and A. Smith (eds)), pp. 629–637, Oxford University Press.

Jones, D. R., Schonlau, M. and Welch, W. J. (1998). Efficient global optimization of expensive black–box functions. *Journal of Global Optimization* **13**, 455–492.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal* **7**, 308–313.

O'Hagan, A. (1992). Some Bayesian numerical analysis. In *Bayesian Statistics*, Vol. 4 (J. Bernardo, J. Berger, A. Dawid and A. Smith (eds)), pp. 345–363, Oxford University Press.

Sacks, J., Schiller, S. B. and Welch, W. J. (1989). Designs for computer experiments. *Technometrics* **31**, 41–47.

Schonlau, M., Welch, W. J. and Jones, D. R. (1998). Global versus local search in constrained optimization of computer models. In *New Developments and Applications in Experimental Design*, Vol. 34 (N. Flournoy, W. Rosenberger and W. Wong (eds)), pp. 11–25, Institute of Mathematical Statistics.

Welch, W. J. (1985). ACED: Algorithms for the construction of experimental designs. *The American Statistician* **39**, 146.

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J. and Morris, M. D. (1992). Screening, predicting, and computer experiments. *Technometrics* **34**, 15–25.

Welch, W. J., Yu, T.-K., Kang, S. M. and Sacks, J. (1990). Computer experiments for quality control by parameter design. *Journal of Quality Technology* **22**, 15–22.

Department of Statistics, Cockins Hall, 1958 Neil Ave., Columbus, OH 43210, U.S.A.

E-mail: williams@stat.ohio-state.edu (BJW), tjs@stat.ohio-state.edu (TJS), win@stat.ohio-state.edu (WIN)