JUN 6 1991

ORNL
MASTER COPY

ORNL/TM-11699

ornl

OAK RIDGE
NATIONAL
LABORATORY

MARTIN MARIETTA

Bayesian Design and Analysis of Computer Experiments: Use of Derivatives in Surface Prediction

M. D. Morris
T. J. Mitchell
D. Ylvisaker

MANAGED BY
MARTIN MARIETTA ENERGY SYSTEMS, INC.
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831; prices available from (615) 576-8401, FTS 626-8401.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

ORNL/TM-11699

ENGINEERING PHYSICS AND MATHEMATICS DIVISION

BAYESIAN DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS: USE OF DERIVATIVES IN SURFACE PREDICTION

Max D. Morris
Toby J. Mitchell
Mathematical Sciences Section, EPM Division
Oak Ridge National Laboratory
P.O. Box 2008, Bldg. 6012, MS 6367
Oak Ridge, Tennessee 37831

Don Ylvisaker
Department of Mathematics
University of California
Los Angeles, California 90024

DATE PUBLISHED: June 1991

Research sponsored by the Applied Mathematical Sciences Research Program, Office of Energy Research, U. S. Department of Energy under contract DE-AC05-84OR21400 with the Martin Marietta Energy Systems, Inc. and in part by NSF Grant DMS 89-02494 between NSF and UCLA.

Prepared by the
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831
Managed by
MARTIN MARIETTA ENERGY SYSTEMS, INC.
for the
U.S. DEPARTMENT OF ENERGY
under Contract No. DE-AC05-84OR21400

TABLE OF CONTENTS

LIST OF TABLES iv

LIST OF FIGURES iv

ABSTRACT v

I. INTRODUCTION 1

II. METHODOLOGY 2

III. EXAMPLE: BOREHOLE MODEL 6

IV. OPTIMAL DESIGN 11

V. EXAMPLE REVISITED 14

VI. CONCLUSION 20

REFERENCES 21

LIST OF TABLES

Table 1. Design and data for a simple example. 9

Table 2. A maximin distance design in $[0,1]^8$ for $n = 10$. ($d^* = 2, J = 36$) 16

Table 3. Comparison of predictive errors for two test sets: Latin hypercube design and maximin design. 17

Table 4. Comparison of design to prediction-site distances for two test sets: Latin hypercube design and maximin design. 17

Table 5. Comparison of predictive errors for two test sets: Two compromise Latin hypercube designs. 18

Table 6. Comparison of design to prediction-site distances for two test sets: Two compromise Latin hypercube designs. 18

LIST OF FIGURES

Figure 1. Contours of output Q as a function of t_1 and t_8, in the borehole example of Section III. 8

Figure 2. Contours of predicted output $\hat{Q}$ as a function of t_1 and t_8, in the borehole example of Section III. 12

BAYESIAN DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS: USE OF DERIVATIVES IN SURFACE PREDICTION

Max D. Morris, Toby J. Mitchell, and Don Ylvisaker

ABSTRACT

The work of Currin et al. (1988) and others in developing "fast predictive approximations" of computer models is extended for the case in which derivatives of the output variable of interest with respect to input variables are available. In addition to describing the calculations required for the Bayesian analysis, the issue of experimental design is also discussed, and an algorithm is described for constructing "maximin distance" designs. An example is given based on a demonstration model of eight inputs and one output, in which predictions based on a maximin design, a Latin hypercube design, and two "compromise" designs are evaluated and compared.

Key Words: Bayesian Prediction, Computer Experiment, Computer Model, Interpolation, Latin Hypercube Design, Maximin Design, Random Functions.

I. INTRODUCTION

In the past few decades, computer models have become important tools in virtually all fields of scientific research. As surrogates for physical or behavioral systems, computer models can be subjected to experimentation, the goal being to predict how the corresponding real system would behave under certain conditions. Here we regard a computer model as a computer program that maps a vector of input variables t into a vector of output variables y, where t and y are physically meaningful. We view y as a function y(t) over some domain T in the space of the input variables. This function is deterministic; if the program is run twice on the same computer using the same value of t, the same value of y will result. We are specifically interested here in computer models which, in addition to calculating the response at a given set of input values, also provide derivatives of y with respect to elements of t.

We consider a *computational experiment* to be a collection of runs of the computer model, made for the purpose of investigating y(t) for $t \in T$. For convenience, we shall consider T to be defined only by the *design variables*, i.e., those variables that are changed during the course of the experiment. In a typical experiment of n runs, the ith computer run is made using inputs $t^{(i)} \in T$, $i = 1, 2, \ldots, n$; this collection of input configurations is called the *experimental design.*

One of the most fundamental problems that can be approached through computational experiments is the prediction of y(t) at sites t that have not been directly observed. This is motivated by applications requiring a large number of evaluations of y, such as numerical optimization, in which repeated execution of the model may be prohibitive due to computing expense. Hence, we seek to develop a *fast predictive approximation* to a computer model, that is sufficiently accurate for many purposes, based on relatively few actual runs. In this paper, we shall restrict our attention to a single output variable, i.e. scalar y.

Recently, there has emerged an interest in using random functions (stochastic processes, random fields) as a structure on which to base the design and analysis of computational experiments, particularly for the purpose of constructing fast predictive approximations. Sacks, Schiller, and Welch (1989) and Sacks, Welch, Mitchell, and Wynn (1989) use an approach similar in some ways to the spatial modeling techniques of kriging for prediction of a computer model. Currin, Mitchell, Morris, and Ylvisaker (1988) also use random functions as a basis for model prediction, but they formulate the problem from a Bayesian point of view, under which uncertainty about the

function y is expressed by means of a probability distribution over all possible response functions. In these approaches, the values of y generated by the computational model are regarded as "data" that, unlike most physical measurements, are exactly reproducible.

Many modern computer models have the potential for providing not only the output values themselves, but also partial derivatives of outputs with respect to inputs, $\partial y(t)/\partial t_j$. Recent research has produced computer-automated methods for "enhancing" computer codes; i.e. expanding codes that compute only outputs so that they also compute derivatives (e.g., Griewank (1988), Oblow, Pin, and Wright (1986), and Worley, Wright, Pin, and Harper (1986)). These derivatives are of interest to users of computer models who are often concerned about the "sensitivity" of model outputs to inputs -- the amount of change induced in y by a small change in one or more elements of t. When derivatives are calculated in the execution of a computer model, they are a source of additional information that may be useful in computer experiments. For example, Worley (1987) discusses a method of uncertainty analysis (determination of the probability distribution of y that results from a known probability distribution of t) that uses values both of the output of interest and its derivatives.

In this report, we will discuss a natural extension of our work in fast predictive approximation based on the stochastic process model, that allows use of derivative information about y. Our approach to Bayesian prediction is outlined in Section II, and an example of a fast predictive approximation based on output values and their derivatives is presented in Section III. In Section IV, we discuss the experimental design problem - how the values of t can be chosen for the needed runs of the computer model. Our emphasis here is on D-optimal design based on weak prior information. In Section V, the example problem is continued with an examination of fast predictive approximations based on data from several different designs.

II. METHODOLOGY

Details of our approach to the Bayesian prediction problem where derivative values are not used may be found in Currin, et al. (1988). Here, we offer a brief description and indicate the modifications necessary for incorporation of derivative information.

We represent prior "knowledge" about the unknown function $y(t)$, $t \in T$, by the Gaussian process $Y = \{Y(t), t \in T\}$, with mean function $M(t) = E[Y(t)]$ and positive definite covariance function

$K(t, s) = Cov[Y(t), Y(s)]$. For every finite set $S \subset T$, the response vector $Y_S = (Y(s_1), Y(s_2), \cdots, Y(s_m))^T$, where m is the number of sites in S, is multinormal with mean $E[Y_S] = \mu_S$ and covariance matrix $Cov[Y_S, Y_S] = \sigma_{SS}$. Normality is chosen for convenience; the posterior process, given the vector of observed responses y_D on the set of n design sites $D \subset T$, is well known and is also Gaussian, with mean and covariance:

$$\mu_{S|D} = E[Y_S | y_D] = \mu_S + \sigma_{SD}\sigma_{DD}^{-1}(y_D - \mu_D) \tag{2.1}$$

$$\sigma_{SS|D} = Cov[Y_S, Y_S | y_D] = \sigma_{SS} - \sigma_{SD}\sigma_{DD}^{-1}\sigma_{DS} \tag{2.2}$$

where σ_{SD} is, for example, the covariance matrix $Cov[Y_S, Y_D]$, μ_S and μ_D are the expectations of Y_S and Y_D respectively, and y_D is the observed value of Y_D. Following execution of the computer model at each site in design D, we use the posterior mean of Y(s) as the fast predictive approximation for the true response y(s) at any site s, and the posterior standard deviation of Y(s) as a measure of the uncertainty of prediction there. Of course, the specification of the prior is the central issue in practice. We simplify matters by adopting various stationarity restrictions: For any pair of sites t, s in T, $M(t) = M(s) = \mu$ and $K(t, s) = \sigma^2 R(s - t)$, where R is a "correlation function" that depends only on the difference vector $(s - t)$, with $R(0) = 1$ and $R(s-t) = R(t-s)$. Further simplification comes from adoption of the "product correlation rule":

$$R(s-t) = \prod_{j=1}^{k} R_j(s_j - t_j) \tag{2.3}$$

The R_j's are chosen from a parametric family of correlation functions on the real line; the correlation parameters, as well as μ and σ, are chosen using maximum likelihood or cross-validation.

The specification of a prior process, with appropriate mean and covariance functions, determines also various derivative processes. (See Parzen 1962, p. 83, for formal definitions and conditions for existence.) For example, the first partial of Y with respect to t_1 is:

$$Y^{(10..0)}(t) = \lim_{h \to 0} \frac{Y(t_1+h, t_2, \ldots, t_k) - Y(t_1, t_2, \ldots, t_k)}{h} \quad .$$

A given stochastic process may then have a number of derivative processes, associated with different orders of derivatives involving different subsets of the elements of t. We shall denote by $Y^{(a_1,a_2,\cdots,a_k)}$ the derivative of order a_i in input t_i, i.e.

$$Y^{(a_1,a_2,\cdots,a_k)} = \frac{\partial^{a_1+a_2+\cdots+a_k}}{\partial t_1^{a_1}\partial t_2^{a_2}\cdots\partial t_k^{a_k}} Y .$$

It can be shown that for Gaussian processes Y, the existing derivative processes are also Gaussian, with

$$E[Y^{(a_1,a_2,\cdots,a_k)}(t)] = M^{(a_1,a_2,\cdots,a_k)}(t) \tag{2.4}$$

and

$$\mathrm{Cov}[Y^{(a_1,a_2,\cdots,a_k)}(t), Y^{(b_1,b_2,\cdots,b_k)}(s)] = K^{(a_1,\cdots,a_k,b_1,\cdots,b_k)}(t,s) \quad . \tag{2.5}$$

Since Y and its derivative processes are jointly Gaussian, prediction using derivative values as well as response values can be done using appropriately modified versions of (2.1) and (2.2). Here we treat the case where the available derivatives are the k first-order partial derivatives with respect to each of the elements of t. Let the random variables that correspond to the response and derivative values at the design sites be held in the n(k+1) vector:

$$Y_D^* = [Y^{(00..0)}(t^{(1)}), Y^{(10..0)}(t^{(1)}), \cdots, Y^{(00..1)}(t^{(1)}), Y^{(00..0)}(t^{(2)}), \cdots, Y^{(00..1)}(t^{(n)})]^T \quad . \tag{2.6}$$

The prior mean μ_D^* and covariance matrix σ_{DD}^* of Y_D^* can be calculated via (2.4) and (2.5), as can the matrix σ_{SD}^* of covariances between Y_S and Y_D^*, where S is any finite set of sites in T. Prediction at S can then be made be substituting y_D^*, μ_D^*, σ_{DD}^*, and σ_{SD}^* for y_D, μ_D, σ_{DD}, and σ_{SD} in (2.1) and (2.2).

Simpler versions of (2.4) and (2.5) result from our stationarity restrictions and the product correlation rule (2.3). Specifically

$$E[Y^{(00..0)}(t)] = E[Y(t)] = \mu \tag{2.7a}$$

and

$$E[Y^{(a_1, a_2, \cdots, a_k)}(t)] = 0 \tag{2.7b}$$

if at least one $a_j > 0$. Also,

$$\mathrm{Cov}[Y^{(a_1, a_2, \cdots, a_k)}(t), Y^{(b_1, b_2, \cdots, b_k)}(s)] = \sigma^2(-1)^{\Sigma a_j} \prod_{j=1}^{k} R_j^{(a_j+b_j)}(s_j - t_j). \tag{2.8}$$

Of course, the chosen R_j's must correspond to differentiable processes. For example, Gaussian processes with the correlation function used by Sacks et al. (1989)

$$R_j(s_j - t_j) = e^{-\theta_j |s_j - t_j|^{p_j}} \tag{2.9}$$

with $\theta_j > 0$ and $0 < p_j \leq 2$, are infinitely differentiable for $p_j = 2$ but not differentiable at all for $p_j < 2$. In his discussion of that paper, Michael Stein refers to an alternative class of processes that is exactly m times differentiable, $m > 1$. In the present context, we require that Y be at least once differentiable. A useful way to derive such processes is by integrating known processes -- see Mitchell, Morris, and Ylvisaker (1990) for some examples that are stationary on an interval. In the examples of this paper, we shall use (2.9) with $p_j = 2$.

The expressions above for posterior means and covariances require specification of the scalars μ and σ and the functions R_j. In practice, we choose a parametric family for each R_j a priori, but allow its parameters, and also μ and σ, to be determined by the data, usually by maximum likelihood. We can isolate the parameters μ and σ in the likelihood by defining

$$C^*_{DD} = \sigma^{-2}\sigma^*_{DD}$$

and

$$v = \mu^{-1}\mu^*_{DD}\,;$$

it is evident that C^*_{DD} depends on the correlation parameters (but not on μ or σ), and v is a binary vector with 1 in position $(i-1)(k+1)+1$, $i=1,\ldots,n$, i.e., in each position corresponding to the mean of some $Y(t^{(i)})$, and 0 everywhere else. The log likelihood is, apart from additive and multiplicative constants,

$$L(\mu, \sigma, \theta) = -n(k+1)\ln\sigma^2 - \ln|C^*_{DD}(\theta)| - \frac{1}{\sigma^2}(y^*_D-\mu v)^T[C^*_{DD}(\theta)]^{-1}(y^*_D-\mu v) \tag{2.10}$$

where dependence on the correlation parameters, collectively denoted as θ here, is now explicitly indicated. For fixed θ, maximization of L over μ and σ^2 is obtained by:

$$\hat{\mu}(\theta) = \frac{v^T[C^*_{DD}(\theta)]^{-1}y^*_D}{v^T[C^*_{DD}(\theta)]^{-1}v} \tag{2.11}$$

$$\hat{\sigma}^2(\theta) = \frac{1}{n(k+1)}(y^*_D - \hat{\mu}(\theta)v)^T[C^*_{DD}(\theta)]^{-1}(y^*_D - \hat{\mu}(\theta)v) \quad . \tag{2.12}$$

Determination of $\hat{\theta}$, which requires maximization of $L(\hat{\mu}(\theta), \hat{\sigma}(\theta), \theta)$ is usually done by constrained iterative search. Although this can be done using routines from standard mathematical software libraries, it may require a considerable amount of computation, depending on the dimension of θ.

Generalization to the case where μ and σ have the usual "noninformative" prior distributions, i.e. μ and $\log\sigma$ have independent uniform prior distributions over arbitrarily large domains, is relatively straightforward, but a fully Bayesian approach, in which vague priors are also attached to the correlation parameters, appears difficult to implement.

III. EXAMPLE: BOREHOLE MODEL

In his discussion of a method of uncertainty analysis, Worley (1987) uses a simple demonstration model of the flow of water through a borehole that is drilled from the ground surface through two aquifers. (His use of this particular model follows that of Harper and Gupta (1983), who use it in demonstrating other methods of uncertainty analysis.) The response variable he examines from this model is Q, the flow rate through the borehole in m^3/yr, which is determined by the equation:

$$Q = \frac{2\pi T_u(H_u - H_l)}{\ln(r/r_w)\left[1 + \frac{2LT_u}{\ln(r/r_w)r_w^2K_w} + \frac{T_u}{T_l}\right]} , \tag{3.1}$$

where the eight inputs and their respective ranges of interest and units are:

r_w = radius of borehole, 0.05 to 0.15 m
r = radius of influence, 100 to 50,000 m
T_u = transmissivity of upper aquifer, 63,070 to 115,600 m^2/yr
H_u = potentiometric head of upper aquifer, 990 to 1110 m
T_l = transmissivity of lower aquifer, 63.1 to 116 m^2/yr
H_l = potentiometric head of lower aquifer, 700 to 820 m
L = length of borehole, 1120 to 1680 m
K_w = hydraulic conductivity of borehole, 9855 to 12,045 m/yr .

Since Q can be expressed as a simple, explicit equation in the "inputs", it certainly is not typical of the computationally intensive computer models that motivate this work. However, it is useful for demonstration purposes, since its simplicity will allow us to quickly assess the accuracy of predictions at a large number of test sites via direct evaluation. In Section V, we shall demonstrate the use of our methodology for predicting Q as a function of all 8 inputs. Here, to illustrate the pattern of required calculations, we shall consider only two, r_w and K_w, and fix the remaining outputs at their respective lowest values. The range of K_w has been extended (for this calculation only) to [1500 , 15,000] to produce a somewhat more nonlinear, nonadditive function. Also, the two input variables considered here have been scaled so that each takes its values from the unit interval; the scaled versions of r_w and K_w are denoted by t_1 and t_8, respectively. Figure 1 is a contour graph of Q as a function of t_1 and t_8 over the region of interest.

For demonstration purposes, consider the experimental design at the 3 sites marked as heavy dots on Figure 1. The data, Q and its first derivatives with respect to t_1 and t_8, are displayed in Table 1. We place these values in the data vector y_D^*, as indicated at (2.6):

$$y_D^* = (\,3.0489\,,\,12.1970\,,\,27.4428\,,\,71.6374\,,\,\cdots\,,\,244.4854\,)^T \;. \tag{3.2}$$

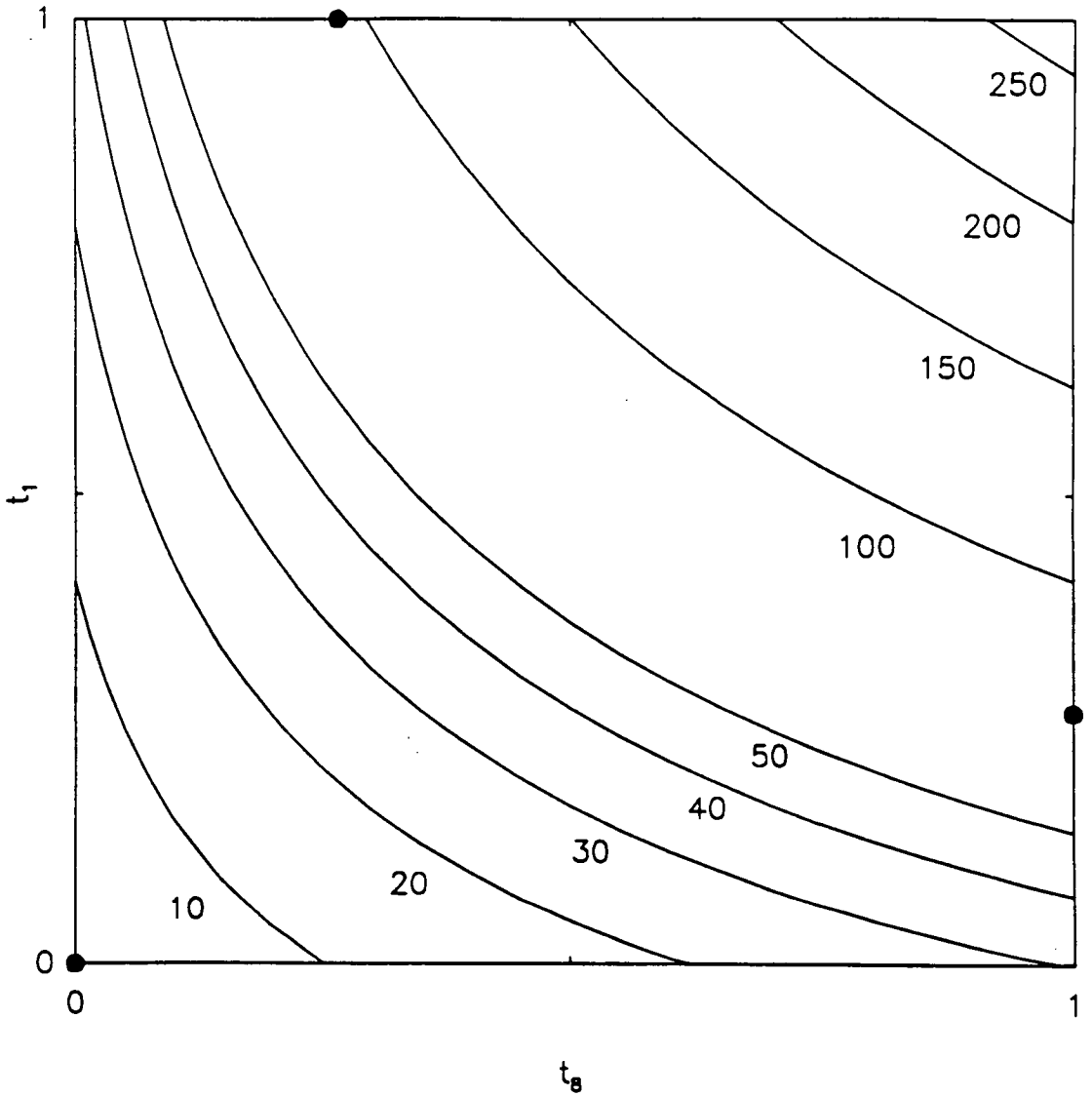

Figure 1.

Table 1. Design and data for a simple example.

site	t_1	t_8	Q	$\partial Q/\partial t_1$	$\partial Q/\partial t_8$
$t^{(1)}$	0.0000	0.0000	3.0489	12.1970	27.4428
$t^{(2)}$	0.2680	1.0000	71.6374	185.7917	64.185
$t^{(3)}$	1.0000	0.2680	93.1663	123.6169	244.4854

Note that the data are organized into segments of $k+1 = 3$ elements each, where the ith segment corresponds to $t^{(i)}$, the ith design site. Conforming to this pattern, the prior covariance matrix σ^*_{DD} is organized as a $n\times n = 3\times 3$ array of $(k+1)\times(k+1) = 3\times 3$ blocks, where the ith diagonal block holds the within-site covariances at $t^{(i)}$ and the $(i,j)^{th}$ off-diagonal block holds the between-site covariances corresponding to the pair $(t^{(i)}, t^{(j)})$. That is,

$$\sigma^*_{DD} = \sigma^2 C^*_{DD} = \sigma^2 \begin{bmatrix} C^{*(1,1)}_{DD} & C^{*(1,2)}_{DD} & C^{*(1,3)}_{DD} \\ C^{*(2,1)}_{DD} & C^{*(2,2)}_{DD} & C^{*(2,3)}_{DD} \\ C^{*(3,1)}_{DD} & C^{*(3,2)}_{DD} & C^{*(3,3)}_{DD} \end{bmatrix} ,$$

where, for example,

$$C^{*(1,2)}_{DD} = \sigma^{-2} \begin{bmatrix} \mathrm{Cov}[Y^{(00)}(t^{(1)}), Y^{(00)}(t^{(2)})] & \mathrm{Cov}[Y^{(00)}(t^{(1)}), Y^{(10)}(t^{(2)})] & \mathrm{Cov}[Y^{(00)}(t^{(1)}), Y^{(01)}(t^{(2)})] \\ \mathrm{Cov}[Y^{(10)}(t^{(1)}), Y^{(00)}(t^{(2)})] & \mathrm{Cov}[Y^{(10)}(t^{(1)}), Y^{(10)}(t^{(2)})] & \mathrm{Cov}[Y^{(10)}(t^{(1)}), Y^{(01)}(t^{(2)})] \\ \mathrm{Cov}[Y^{(01)}(t^{(1)}), Y^{(00)}(t^{(2)})] & \mathrm{Cov}[Y^{(01)}(t^{(1)}), Y^{(10)}(t^{(2)})] & \mathrm{Cov}[Y^{(01)}(t^{(1)}), Y^{(01)}(t^{(2)})] \end{bmatrix} .$$

To compute the elements of C^*_{DD}, we use (2.8), where the correlation functions R_j are given at (2.9) with $p_j = 2$. It is evident that C^*_{DD} is independent of σ^2. For example, if $\theta_1 = 0.4$ and $\theta_8 = 0.5$, the (1,3) and (2,2) elements of $C^{*(1,2)}_{DD}$ are, respectively,

$$(C^{*(1,2)}_{DD})_{1,3} = R_1(t_1^{(2)} - t_1^{(1)})\, R_8'(t_8^{(2)} - t_8^{(1)}) = R_1(.2680)\, R_8'(1) = -0.58935$$

$$(C^{*(1,2)}_{DD})_{2,2} = -R_1''(t_1^{(2)} - t_1^{(1)})\, R_8(t_8^{(2)} - t_8^{(1)}) = -R_1''(.2680)\, R_8(1) = 0.44439$$

where we have used the fact that

$$R_j'(x) = -2\theta_j x e^{-\theta_j x^2} , \quad R_j''(x) = (-2\theta_j + 4\theta_j^2 x^2) e^{-\theta_j x^2} .$$

Noting that $v = (1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0)^T$ here, we maximize the likelihood for fixed $(\theta_1, \theta_8) = (.4, .5)$, using (2.11) and (2.12); this yields $\hat{\mu}(.4, .5) = 70.77$ and $\hat{\sigma}(.4, .5) = 135.70$. Substitution into (2.10) gives the maximum log likelihood at $(\theta_1, \theta_8) = (.4, .5)$. A search of the (θ_1, θ_8) - space using a standard numerical optimization routine finds that the greatest log likelihood occurs at $\hat{\theta}_1 = .429$ and $\hat{\theta}_8 = .467$; the corresponding maximum likelihood values for μ and σ are $\hat{\mu} = 69.15$, $\hat{\sigma} = 135.47$.

Now suppose we wish to calculate the posterior mean and variance of Y(s) at an arbitrary site s. Equations (2.1) and (2.2) imply

$$\mu_{s|D} = \hat{\mu} + C^*_{sD} C^{*-1}_{DD} (y^*_D - \hat{\mu} v) \tag{3.3}$$

$$\sigma_{ss|D} = \hat{\sigma}^2 (1 - C^*_{sD} C^{*-1}_{DD} C^*_{Ds}), \tag{3.4}$$

where we have chosen $\mu = \hat{\mu}$ and $\sigma = \hat{\sigma}$. The data vector y^*_D is given at (3.1) and the matrix C^*_{DD} is available from the final iteration of the likelihood maximization. The one remaining ingredient is the $1 \times n(k+1) = 1 \times 9$ vector $C^*_{sD} = \sigma^{-2} \sigma^*_{sD}$, which can be partitioned into n=3 segments:

$$C^*_{sD} = [C^{*(1)}_{sD}\ C^{*(2)}_{sD}\ C^{*(3)}_{sD}]$$

where, for instance,

$$C^{*(3)}_{sD} = \sigma^{-2} \Big[\mathrm{Cov}[Y^{(00)}(s), Y^{(00)}(t^{(3)})]\ \ \mathrm{Cov}[Y^{(00)}(s), Y^{(10)}(t^{(3)})]\ \ \mathrm{Cov}[Y^{(00)}(s), Y^{(01)}(t^{(3)})] \Big]$$

(Because of the constant multiplier, C^*_{sD} is independent of σ^2.) These values are computed in the same way as the entries of C^*_{DD} were computed above, with the correlation parameters θ_1 and θ_8 fixed at their maximum likelihood values.

Now predictions can be made at any set of sites S, using (3.3) and (3.4). Here we find, for example, the posterior mean at $s = (.5,.5)$ is 69.4 with a posterior standard deviation of 2.7. At $s = (1,1)$, the posterior mean is 230.0 and the posterior standard deviation is 19.2. The posterior means can be computed quickly, once the $n(k+1) \times n(k+1) = 9 \times 9$ linear system

$$C^*_{DD} g = y^*_D - \hat{\mu} v$$

is solved for the 9×1 vector g. Predictions may then be made at any site s by

$$\hat{y}(s) = \mu_{s|D} = \hat{\mu} + C^*_{sD}g \quad .$$

Predictions on a 21×21 grid were generated in this way and used to produce the contour graph of $\hat{Q}(t_1, t_8)$ over the region of interest, as shown in Figure 2.

IV. OPTIMAL DESIGN

An advantage to the use of random functions for prediction is that the posterior variability of Y can be used to provide measures of uncertainty, and designs can be sought to minimize the expected uncertainty in some sense. See Ylvisaker (1987) and Sacks et al. (1989) for references to some previous work along these lines. Criteria that have been considered are G-optimality (minimization of the maximum variance of $Y_{s|D}$ for $s \in T$), A-optimality (minimization of the average variance of $Y_{s|D}$ for $s \in T$), and D-optimality (minimization of the generalized variance of $Y_{S|D}$ for a specific $S \subset T$).

Johnson, Moore, and Ylvisaker (1990) establish an interesting link between these criteria and the geometric properties of certain designs, for the case in which only the response is observed. They show that, when the prior correlation between sites is extremely weak and is a decreasing function of an appropriately defined intersite distance, necessary conditions for a design to be D-optimal are (1.) the minimum distance between pairs of points in the design is maximized, and (2.) the number of point pairs separated by this distance is minimized.

Mitchell, Morris, and Ylvisaker (1991) discuss similar results for the case in which both the response and its derivatives are observed. Arguments are given there to justify the maximization of $|C^*_{DD}|$ as a convenient way to minimize the generalized variance of $Y_{S|D}$ or $Y^*_{S|D}$; we therefore refer to designs that maximize this determinant as D-optimal. Here, we briefly summarize without proof one of their results. Define a product correlation function such that R(s–t) is a decreasing function of the Euclidean norm of its argument, $\|s-t\|$; for

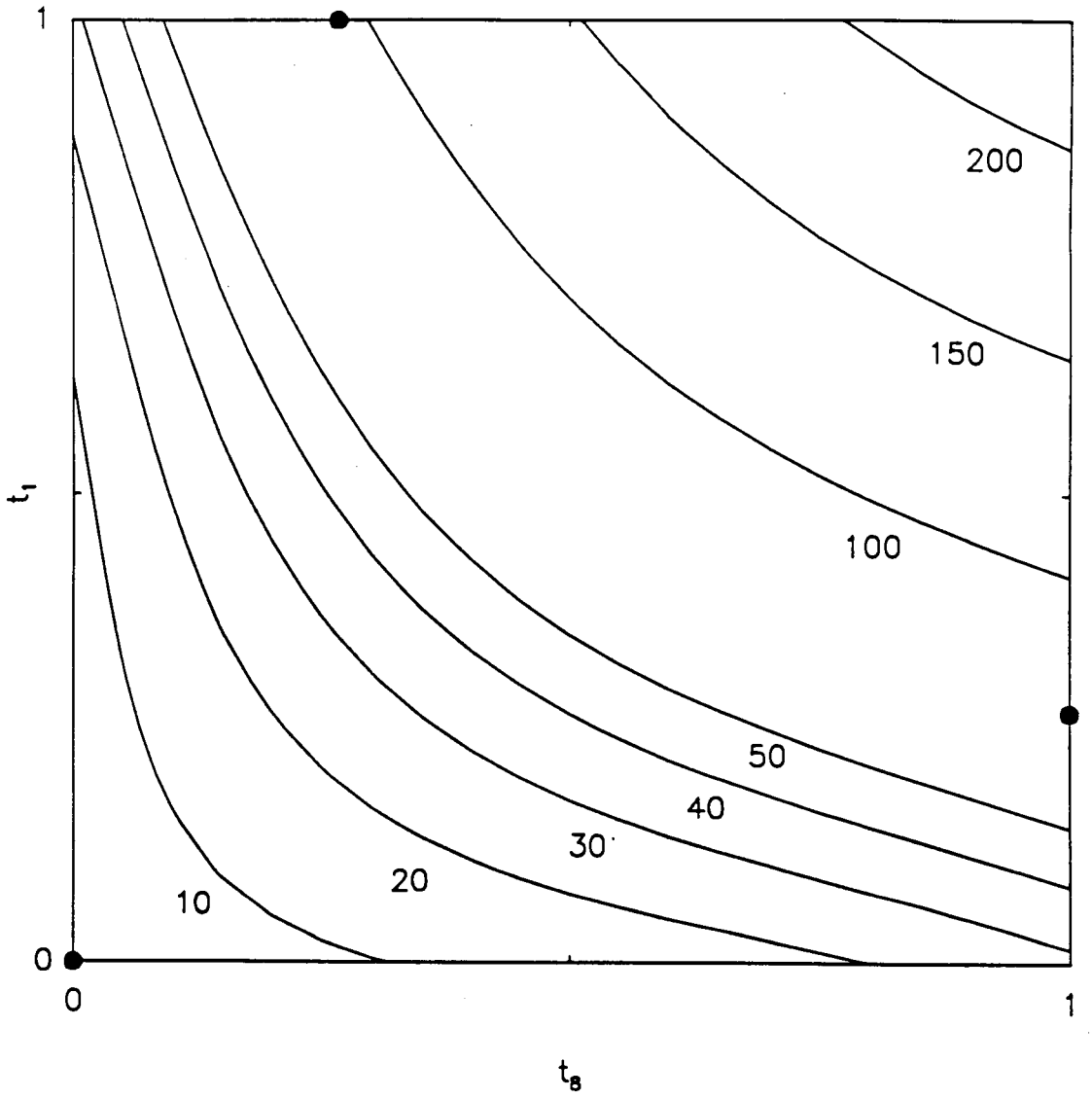

Figure 2.

example, the correlation defined by equations (2.3) and (2.9) has this property if $\theta_j = 1$ and $p_j = 2$ for all j. Next, define a family of correlation functions:

$$R_\kappa(s-t) = [R(s-t)]^\kappa$$

indexed by the positive integer κ. For any design D, let $d^*(D)$ be the smallest Euclidean distance between any two sites. Finally, let J(D), called the index of the design, be the number of site pairs separated by this distance $d^*(D)$. Then the following result holds:

Theorem: As κ tends to infinity, the designs that maximize $|C^*_{DD}|$ are such that (1.) $d^*(D)$ is maximized, and among the designs for which this is true, (2.) J(D) is minimized.

Designs for which conditions (1.) and (2.) of the theorem are satisfied will be called *maximin distance designs*, or simply *maximin designs*.

A practical weakness of design optimality for a given R is that one seldom knows, at design time, what correlation function will be selected for analysis. This difficulty has a parallel in optimal design for regression experiments, where the optimal design is highly dependent on the choice of regression model, which is not usually made until the data are analyzed. A pragmatic approach there is to base the design on weaker prior information than one expects to invoke in the analysis, e.g. use a cubic rather than a linear or quadratic polynomial model for design. Our use of maximin designs in this context is similarly motivated. Although the limiting (weak) correlation needed to link the maximin property to D-optimality is not useful for analysis, a somewhat "pessimistic" design strategy based on weak correlation seems prudent.

In order to construct specific maximin distance designs, we wrote a computer program to find designs that minimize a surrogate criterion function:

$$\phi_p(D) = [\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d_{ij}^{-p}(D)]^{1/p} \quad , \tag{5.1}$$

where $d_{ij}(D)$ is the Euclidean distance between the ith and jth design points in design D. To see the motivation for this, first rewrite (5.1) as

$$\phi_p(D) = \frac{1}{d^*(D)} [\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\frac{d^*(D)}{d_{ij}(D)})^p]^{1/p} \quad .$$

For pairs of sites separated by distance $d^*(D)$, the corresponding term in the sum is one. If a large value is chosen for p, pairs of site separated by greater distances will have associated terms in the sum that are approximately zero. Hence, for large p,

$$\phi_p(D) \approx \frac{J^{1/p}(D)}{d^*(D)} \quad .$$

For large enough p, minimizing ϕ_p is primarily accomplished by maximizing d^*, and to a much smaller degree by minimizing J.

Our computer program for minimizing ϕ_p implements a simple point-exchange algorithm based on the optimization technique of simulated annealing. (See Kirkpatrick, Gelatt, and Vechhi (1983) for a discussion of simulated annealing, or Bohachevsky, Johnson, and Stein (1986) for a generalization of this technique applied to a statistical problem.) Briefly, a search begins with a randomly constructed design, which is sequentially modified as follows. First, one site from the design is randomly selected, and each coordinate of that site is subjected to a (trial) random perturbation. (Specific distributions of perturbations used, and other particulars of the search, are specified for the example application of Section V.) If the modification is such as to decrease the value of ϕ_p, the change is made. If the value of ϕ_p would not be decreased by the change, a random choice is made either to make the change anyway or to ignore the change. The probability that such a change (to a design with higher ϕ_p) is made decreases with the amount of increase which would result in ϕ_p. The probability that a change resulting in any given increase in ϕ_p is made also decreases as the optimization proceeds, according to an "annealing schedule".

V. EXAMPLE REVISITED

We now return to our example model, described by equation (3.1), to demonstrate an application of this methodology for an 8-dimensional input vector. For this purpose, all eight inputs were scaled as in Section III so that the range of each t_i was the unit interval, $T=[0,1]^8$. Initially, we applied the prediction method to the design used by Worley (1987) in his demonstration of a methodology he calls *deterministic uncertainty analysis*, which uses both observed values of Q and its first derivatives. Although his primary interest is in exploring how a specified probability distribution on t is propagated to Q, his analysis includes an interim step that involves prediction

of Q at sites not included in the design, using essentially a first order Taylor series expansion of Q. In his demonstration, Worley's experimental design was a 10-run Latin hypercube sample, generated using a non-uniform distribution across T, and he compared predictions of Q with its actual value at sites in a 50-run Latin hypercube test set, generated using the same distribution. Worley reports root mean square errors, over these 50 sites, of 1.89, 2.45, and 2.37 for three versions of his method; for comparison, the range of true values of Q over the test set is 24.97 to 144.57. Using our procedure, the corresponding root mean square error is 0.610. Encouraged by this result, we undertook a more extensive investigation using the same model.

We first tried two different experimental designs, each of 10 runs. The first design is a Latin hypercube sample, as introduced by McKay, Beckman, and Conover (1979). (Actually, our version of the Latin hypercube is a bit different than what is described in this reference, since we used 10 equally-spaced values within the range of each input, instead of randomly chosen values from 10 non-overlapping intervals of equal length; we suspect that this modification has very little effect in this application.) One hundred random Latin hypercube samples were generated, and the one used here was selected so as to minimize the largest correlation between any two columns of the design matrix.

We also generated a maximin distance design in 10 runs using the algorithm described in Section IV. After some initial experimenting to find annealing parameters that appeared to be effective for this problem, our first attempt at finding an optimal design consisted of ten searches, in which each element of the starting design matrix was chosen from the unit interval, perturbations were normally distributed with a standard deviation of 0.3 (except when this would result in a value outside the unit interval, in which case the change was modified to yield either 0 or 1), $p = 1000$, and a schedule of 100 "temperatures" was used. Five of these searches resulted in designs with $d^* = 2$, and the other five produced smaller values. Of the five with $d^* = 2$, one had an index of 42, three had indexes of 38, and one had an index of of 37. These five designs (unlike the others) placed all sites in the corners of T.

Following this last observation, ten additional searches were made using a modified search in which only designs on the 2^8 corners of T were considered, i.e. each coordinate in the initial design was 0 or 1 with equal probability. Here, once a site was selected for modification, the level of an individual coordinate was reversed with probability 0.3; again $p = 1000$. Of these searches, nine yielded designs with $d^* = 2$, one of these with index 40 and the remaining eight

with index 36. Our (tentative) conclusion based on this search was that these last eight designs are optimal, although they are not all equivalent. The design we chose from this set is given in Table 2.

The correlation function associated with each of the eight inputs was chosen to be of the form shown in equation (2.9) with $p_j=2$. Using each of the two designs individually, the parameters of the process (μ, σ, and θ_j, $j=1,2,\cdots,8$) were estimated by the method of maximum likelihood, as described in Section III. In order to evaluate how well our predictions match the true model, we selected two "test sets" of sites at which to compare $\hat{Q}$ and Q. The first of these is a random sample of 400 sites, selected from the uniform distribution over T. The second set of test sites is the 256 corners of T, i.e. those points at which each of the inputs takes either the high or low extreme value in its range. The first set is intended to provide an indication of how well each predictor does throughout the interior of T, while the second allows us to compare their performance at the extreme sites. Values of Q range from 12.4035 to 230.6478 in the first test set, and from 7.8197 to 309.5756 in the second. Predictions were made for each design, at the sites in each test set, and errors of prediction ($\hat{Q}$ - Q) were calculated. These errors are summarized in Table 3; errors for the maximin design are based on the 246 corner points not included in the design.

Table 3 indicates mixed results for the predictors based on these two designs. The Latin hypercube appears to produce better predictions on the random test set, while the maximin

Table 2. A maximin distance design in $[0,1]^8$ for n = 10.
($d^* = 2$, J = 36)

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
1	1	0	0	1	0	1	1
1	1	1	1	0	0	1	0
1	0	0	1	1	0	0	0
0	1	0	0	1	1	0	0
1	1	0	1	0	1	0	1
0	1	1	0	0	0	0	1
0	0	1	1	1	0	1	1
0	0	0	0	0	1	1	1
0	0	1	1	0	1	0	0
1	0	1	0	1	1	1	0

Table 3. Comparison of predictive errors for two test sets: Latin hypercube design and maximin design.

Design	Test set: 400 random points in T	Test set: 256 corners of T
Latin hypercube	maximum error = 5.68 minimum error = -11.82 root m.s.e. = 3.162 Corr[Q,$\hat{Q}$] = 0.998438	maximum error = 15.73 minimum error = -40.69 root m.s.e. = 9.217 Corr[Q,$\hat{Q}$] = 0.995634
maximin design	maximum error = 5.12 minimum error = -28.34 root m.s.e. = 9.547 Corr[Q,$\hat{Q}$] = 0.990784	maximum error = 15.67 minimum error = -12.78 root m.s.e. = 4.681 Corr[Q,$\hat{Q}$] = 0.998860

distance design does a somewhat better job on the corners of T. These results may be due in part to the proximity of each design to the two test sets. To investigate this, we calculated Euclidean distances from each site in the test sets to each of the two experimental designs. The minimum, maximum, and average distances from points in each test set to each design are given in Table 4; again, distances from the maximin design are calculated only on the 246 corner points not included in the design. As we suspected, the Latin hypercube design is considerably closer (on average) to the random test set than is the maximin design, and this may partially explain the difference in performance we see here. Surprisingly, the Latin hypercube design is also somewhat closer (on average) to the 256-site test set than is the maximin design. On this test set, however, the maximin design yielded a smaller root mean square error by a factor of about 2, and a smaller absolute error by a factor of about 3.

Table 4. Comparison of design to prediction-site distances for two test sets: Latin hypercube design and maximin design.

Design	Test set: 400 random points in T	Test set: 256 corners of T
Latin hypercube	minimum distance = 0.187 maximum distance = 1.148 average distance = 0.788	minimum distance = 0.903 maximum distance = 1.478 average distance = 1.107
maximin design	minimum distance = 0.591 maximum distance = 1.445 average distance = 1.142	minimum distance = 1.000 maximum distance = 1.732 average distance = 1.342

Another relevant consideration in the comparison of the two designs is the nature of the true function Q, which appears to be mainly a function of the first two inputs. When one considers the projection of the designs into the low dimensional subspaces defined by the coordinates, it is clear that the Latin hypercube design does better than the maximin design, even if one adheres to the maximin criterion. (For example, the projection of the Latin hypercube design onto any coordinate is a maximin design in that one-dimensional space.) On the other hand the maximin design, by definition, spreads the design sites as much as possible in the full k-dimensional space, whereas the intersite distances in the Latin hypercube design are to some extent left to chance. We therefore continued our investigation by generating a couple of "compromise" designs, both of which are Latin hypercube designs in that their one-dimensional projections onto every coordinate produces equispaced sites on [0,1].

The design we shall call compromise #1 is a modification of our maximin design to give it the desired one-dimensional projections. Starting with each column of the design matrix of Table 2, the five 0's were replaced with the values 0, 1/9, 2/9, 3/9, and 4/9, assigned in random order, and the five 1's were similarly replaced with 5/9, 6/9, 7/9, 8/9, and 1. This procedure is essentially the same as randomly selecting a Latin hypercube design from among those which, if each entry were rounded to 0 or 1, would be the maximin design given in Table 2.

Compromise #2 is intended to apply the maximin criterion *within the class of Latin hypercube designs.* It was constructed using our simulated annealing algorithm with minor modifications. A randomly constructed Latin hypercube was used as the starting design in each search, and trial perturbations were created by exchanging two entries in a randomly chosen column of the design matrix; the result of any such exchange is another Latin hypercube. Although 20 searches were attempted, the apparent maximin design in this class was generated only once, so it is quite possible that the result is only a "near maximin" design.

Summaries of the prediction errors generated at the test sets by these two designs, and of distances from the test sets to the designs, are given in Tables 5 and 6. Compromise #1 seems to be particularly successful in comparison to our first two designs, in that root mean square errors are smaller than were attained previously for either design, for both test sets. Although the error summaries are somewhat less impressive for compromise #2, they still represent general improvement over the first two designs examined, yielding a somewhat larger root mean square error than the maximin design on the corners of T, and more accurate predictions on the interior of T than either the maximin or (original) Latin hypercube designs. Distances from these

Table 5. Comparison of predictive errors for two test sets: two compromise Latin hypercube designs.

Design	Test set: 400 random points in T	Test set: 256 corners of T
compromise #1	maximum error = 5.76 minimum error = -2.43 root m.s.e. = 0.765 Corr[$Q,\hat{Q}$] = 0.999862	maximum error = 10.04 minimum error = -10.28 root m.s.e. = 3.222 Corr[$Q,\hat{Q}$] = 0.999414
compromise #2	maximum error = 5.66 minimum error =-6.63 root m.s.e. = 1.584 Corr[$Q,\hat{Q}$] = 0.999485	maximum error = 14.19 minimum error = -27.94 root m.s.e. = 6.242 Corr[$Q,\hat{Q}$] = 0.997846

Table 6. Comparison of design to prediction-site distances for two test sets: two compromise Latin hypercube designs.

Design	Test set: 400 random points in T	Test set: 256 corners of T
compromise #1	minimum distance = 0.325 maximum distance = 1.168 average distance = 0.774	minimum distance = 0.910 maximum distance = 1.552 average distance = 1.124
compromise #2	minimum distance = 0.378 maximum distance = 1.112 average distance = 0.772	minimum distance = 0.903 maximum distance = 1.445 average distance = 1.100

Latin hypercubes to the points in the test sets appear to be similar for each test set. They are also similar to the distances reported for the first Latin hypercube, with the exception of minimum distance to the 400 point test set, which is somewhat larger for the compromise designs.

Finally, to see how much was gained by using the derivative information, we repeated the example for a couple of the designs, using only the observed Q at each of the ten design sites, with the same type of correlation function. For the Latin hypercube design, the maximum absolute errors and root mean squared errors were roughly four times larger than those shown in Table 3, while for compromise #1, these errors were roughly ten times larger than those shown for the same design in Table 5.

VI. CONCLUSION

We have described a generalization of the methods of Currin et al. (1988), for developing a fast predictive approximation of a computer model based on evaluations of the output and its partial derivatives at a set of design sites. A generalization of the work of Johnson, Moore, and Ylvisaker (1990), as elaborated in Mitchell, Morris, and Ylvisaker (1991), provides an asymptotic argument for the use of maximin distance designs in this context. In the example calculation described here, a comparison of predictions based on a Latin hypercube design and a maximin design lead to mixed results, with the Latin hypercube performing better on the interior of the input domain and the maximin design performing better at the extremes of the region. Two compromise designs, which are constructed in an effort to preserve the strengths of both the Latin hypercube structure and the maximin criterion, are more generally successful.

While the design and analysis procedures described here are straightforward in principle, some questions will require further attention. In particular, the type of correlation function used in an analysis may have considerable influence on the predictions. Our somewhat arbitrary selection of a correlation function has produced reasonably good results in the example of Section V. However, more complicated computer models with inputs of higher dimension will undoubtedly pose more difficult challenges, and both theoretical consideration and empirical investigation of the effect of different correlation functions will be important. Further investigation of what may be expected from maximin (and other "optimal") designs, Latin hypercube designs, and designs constructed from other approaches, is also in order. When these and other issues are better resolved, it will eventually be important to consider how both design and analysis may be implemented sequentially as data are collected.

REFERENCES

Bohachevsky, I.O., Johnson, M.E., and Stein, M.L. (1986). "Generalized Simulated Annealing for Function Optimization," *Technometrics*, **28**, 209-217.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1988). "A Bayesian Approach to the Design and Analysis of Computer Experiments," ORNL-6498, available from National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.

Griewank, A. (1988). "On Automatic Differentiation," *Proceedings of the 13th Mathematical Programming Symposium*, Tokyo, Japan.

Harper, W.V., and S.K. Gupta (1983). "Sensitivity/Uncertainty Analysis of a Borehole Scenario Comparing Latin Hypercube Sampling and Deterministic Sensitivity Approaches," BMI/ONWI-516, Office of Nuclear Waste Isolation, Battelle Memorial Institute, Columbus, Ohio.

Johnson, M., Moore, L., and Ylvisaker, D. (1990). "Minimax and Maximin Distance Designs," to appear in *Journal of Statistical Planning and Inference.*

Kirkpatrick, S., Gelatt, C.D., Jr., and Vechhi, M.P. (1983). "Optimization by Simulated Annealing," *Science*, **220**, 671-680.

McKay, M.D., Beckman, R.J., and Conover, W.J. (1979). "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, **21**, 239-245.

Mitchell, T.J., Morris, M.D., and Ylvisaker, D. (1990). "Existence of Smoothed Stationary Processes on an Interval," *Stochastic Processes and their Application*, **35**, 109-119.

Mitchell, T.J., Morris, M.D., and Ylvisaker, D. (1991). "Minimax and Maximin Distance Designs for Computer Experiments with Derivative Information," UCLA technical report, to appear.

Oblow, E.M., Pin, F.G., and Wright, R.Q. (1986). "Sensitivity Analysis Using Computer Calculus: A Nuclear Waste Application," *Nucl. Sci. Eng.*, **94**, 46.

Parzen, E. (1962). *Stochastic Processes*, Holden-Day, Inc, San Francisco.

Sacks, J., Schiller, S.B., and Welch, W.J. (1989). "Designs for Computer Experiments," *Technometrics*, **31**, 41-47.

Sacks, J., Welch, W.J., Mitchell, T.J., and Wynn, H.P. (1989). "Design and Analysis of Computer Experiments," *Statistical Science*, **4**, 409-423.

Stein, M.L. (1989). Discussion of "Design and Analysis of Computer Experiments" by Sacks et al., *Statistical Science*, **4**, 432-433.

Worley, B.A. (1987). "Deterministic Uncertainty Analysis," ORNL-6428, available from National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161.

Worley, B.A., Wright, R.Q., Pin, F.G., and Harper, W.V. (1986). "Application of an Automated Procedure for Adding a Comprehensive Sensitivity Calculation Capability to the ORIGEN2 Point Depletion and Radioactivity Decoy Code," *Nucl. Sci. Eng.*, **94**, 180.

Ylvisaker, D. (1987). "Prediction and Design," *Ann. Statist.*, **15**, 1-19.

INTERNAL DISTRIBUTION

1. B. R. Appleton
2-3. T. Darland/ Mathematical Sciences Library
4. D. J. Downing
5. L. J. Gray
6-10. T. J. Mitchell
11-15. M. D. Morris
16-20. S. Raby
21-25. R. F. Sincovec
26-30. R. C. Ward
31. B. Worley
32. Central Research Library
33. K-25 Plant Library
34. ORNL Patent Office
35. Y-12 Technical Library Document Reference Section
36-37 Laboratory Records Dept.
38. Laboratory Records - RC

EXTERNAL DISTRIBUTION

39. Professor Roger W. Brockett (EPMD Advisory Committee), Wang Professor of Electrical Engineering and Computer Science, Division of Applied Sciences, Harvard University, Cambridge, MA 02138.

40. Dr. John Cavallini, Scientific Computing Staff, Applied Mathematical Sciences, Office of Energy Research, U.S. Department of Energy, Washington, D.C. 20585.

41. Dr. John J. Doming (EPMD Advisory Committee), Department of Nuclear Engineering Physics, Thornton Hall, McCormick Road, University of Virginia, Charlottesville, Virginia 22901.

42. Dr. Robert Easterling, Sandia National Laboratory, Division 7223, Albuquerque, New Mexico 87185.

43. Dr. Jerome Friedman, Department of Statistics, Sequoia Hall, Stanford University, Stanford, California 94305.

44. Ms. Karole Herzog, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 20461-0439.

45. Dr. Ronald Iman, Sandia National Laboratory, Division 6415, Albuquerque, New Mexico 87185.

46. Professor Mark Johnson, Department of Statistics, University of Central Florida, Orlando, Florida 32816.

47. Dr. James E. Leiss (EPMD Advisory Committee), Route 2, Box 142C, Broadway, Virginia 22815.

48. Dr. L. M. Moore, Los Alamos National Laboratory, P. O. Box 1663, Los Alamos, New Mexico 87545.

49. Professor Neville Moray (EPMD Advisory Committee), Department of Mechanical and Industrial Engineering, University of Illinois, 1206 West Green Street, Urbana, Illinois 61801.

50. Dr. Art Owen, Stanford University, Department of Statistics, Sequoia Hall, Stanford, California 94305-4065.

51. Professor Friedrich Pukelsheim, Institut für Mathematik, Universität Augsburg, Universitätsstrasse, D-8900, Augsburg, Germany.

52. Dr. L. R. Shenton, Office of Computing and Information Service, Boyd Graduate Studies Building, University of Georgia, Athens, Georgia 30602.

53. Dr. Daniel L. Solomon, Department of Statistics, North Carolina State University, P. O. Box 5457, Raleigh, North Carolina 27650.

54. Professor Michael Stein, Department of Statistics, University of Chicago, 5734 University Avenue, Chicago, Illinois 60637.

55. Dr. Werner Stuetzle, Department of Statistice, GN-22, University of Washington, Seattle, Washington 98195.

56. Dr. Ray A. Waller, S-1, Statistics, Los Alamos National Laboratory, P. O. Box 1663, Los Alamos, New Mexico 87545.

57. Professor William Welch, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.

58. Professor Mary F. Wheeler (EPMD Advisory Committee), Rice University, Department of Mathematical Sciences, P. O. Box 1892, Houston, Texas 77251.

59. Dr. D. George Wilson, Numerically Intensive Computing, Applications Support Center, Department 41UD, MS 276, Neighborhood Road, Kingston, New York 12401.

60. Professor Henry Wynn, Department of Mathematics, The City University, Northhampton Square, London ECIV OHB, England.

61-65. Professor Don Ylvisaker, Department of Mathematics, University of California, Los Angeles, California 90024.

66. Office of Assistant Manager for Energy Research and Development, U.S. Department of Energy, Oak Ridge Operations Office, P. O. Box 2001, Oak Ridge, Tennessee 37831-8600.

67-76. Office of Scientific and Technical Information, P. O. Box 62, Oak Ridge, Tennessee 37831-0062.