Designs for Computer Experiments
Author(s): Jerome Sacks, Susannah B. Schiller, William J. Welch
Source: *Technometrics*, Vol. 31, No. 1 (Feb., 1989), pp. 41-47
Published by: American Statistical Association and American Society for Quality
Stable URL: http://www.jstor.org/stable/1270363
Accessed: 22/05/2011 12:48

Designs for Computer Experiments

Jerome Sacks and Susannah B. Schiller

Department of Statistics
University of Illinois
Champaign, IL 61820

William J. Welch

Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada

A computer experiment generates observations by running a computer model at inputs x and recording the output (response) Y. Prediction of the response Y to an untried input is treated by modeling the systematic departure of Y from a linear model as a realization of a stochastic process. For given data (selected inputs and the computed responses), best linear prediction is used. The design problem is to select the inputs to predict efficiently. The issues of choice of stochastic-process model and computation of efficient designs are addressed, and applications are made to some chemical kinetics problems.

KEY WORDS: Computer-aided design; Experimental design; Prediction; Response surface; Spatial statistics; Supercomputing.

1. INTRODUCTION

In a computer experiment, observations are made on a response function Y by running a (typically complex) computer model at various choices of input factors, x. For example, in the chemical kinetics of methane combustion, x can be a set of rate constants in a system of differential equations and Y can be a concentration of a chemical species some time after combustion. Solving the differential equations numerically for specified x yields a value for Y. Because running the equations solver is expensive, the aim is to estimate the relationship between x and Y from a moderate number of runs so that Y can be predicted at untried inputs.

Another important application area is the computer simulation of integrated circuits. Here x defines various circuit parameters, such as transistor characteristics, and Y is a measurement of the circuit performance, such as an output voltage [e.g., the differential operational amplifier of Phadke (1986)]. In fact, the widespread use of computer models and experiments for simulating real phenomena generates examples in virtually all areas of science and engineering.

In this article, we consider the prediction problem. In some applications, however, the actual goal may be to evaluate or optimize a particular functional of the output, in which case prediction may be viewed as an intermediate step toward achieving that goal. For example, this is an approach that was taken in ongoing research on a very large scale integrated circuit design problem by Welch, Yu, Kang, and Sacks (1988).

If the response is modeled by Response = Linear model + Departures, that is,

$$Y(x) = \sum_{j=1}^{k} \beta_j f_j(x) + Z(x), \qquad (1.1)$$

then traditional response-surface methodology would treat Z as white noise corresponding to measurement error. In a computer experiment, however, often there is no measurement error, and $Z(x)$ is then systematic departure from the assumed linear model. The form of $Z(x)$ is, of course, usually unknown. Despite the lack of measurement error, $Z(x)$ has sometimes been taken as white noise.

Our approach is to model the systematic departure Z as a realization of a stochastic process in which the covariance structure of Z relates to the smoothness of the response. In some instances it is useful to eliminate the linear model altogether and regard the entire response as a realization of the stochastic process. This approach was also used in related work by Currin, Mitchell, Morris, and Ylvisaker (1988).

Concern about systematic departures from a linear model and its effect on designs originated with Box and Draper (1959, 1963). They modeled departures from simple polynomials by higher-order terms. Welch (1983) and Sacks and Ylvisaker (1984) modeled the departures nonparametrically. In these, as in other treatments (see the last two citations for references), measurement error may be important and induces the model

$$Y(x) = \sum \beta_j f_j(x) + Z(x) + \varepsilon(x), \qquad (1.2)$$

where Z is the systematic departure and ε is the measurement error.

Sacks and Ylvisaker (1985) modeled Z as a random function—that is, a realization of a stochastic process. There is earlier work in the context of numerical integration or predicting integrals in which Z is modeled as a random function and there is no linear model; see Sacks and Ylvisaker (1970), Suldin (1959, 1960), and a survey by Diaconis (1988). In these references, the random function was defined on the real line; the more difficult, multidimensional case was broached by Ylvisaker (1975). Related work also was done by Speckman (1976) and more currently by Ylvisaker (1987).

Linear prediction from (1.2), where Z is a stochastic process, is known in the geostatistics and other spatial statistics literature as kriging (Cressie 1986; Ripley 1981). Our discussion then relates to designs for spatial problems as well as computer experiments. Sacks and Schiller (1988) and Shewry and Wynn (1987) discussed designs for spatial problems.

Earlier research on designs for computer experiments includes Latin hypercube sampling (Iman and Conover 1980; McKay, Conover, and Beckman 1979; Stein 1987). These sampling designs are discussed further in Section 4.

Section 2 considers designing for Model (1.1) with Z a covariance process and addresses the issue of how to choose the unknown covariance structure. Section 3 describes two examples involving the chemical kinetics of methane combustion. Compared with factorial designs and least squares estimation, our design and prediction strategy based on (1.1) with Z a stochastic process reduces the actual squared error of prediction by factors of 8–10 in the first example. A similar comparison can be made in the second example, but, of greater interest, there is evidence that inclusion of a stochastic-process term may reduce the need for complex linear models in (1.1).

2. PREDICTION AND DESIGN

Central to our methodology is the covariance structure of the stochastic process Z of (1.1). In the examples of Section 3, we use

$$\text{cov}(Z(t), Z(u)) = V(t, u)$$

$$= \sigma_Z^2 \exp\left(-\theta \sum_{j=1}^{d} (t_j - u_j)^2\right)$$

$$= \sigma_Z^2 R(t, u) \qquad (2.1)$$

between d-dimensional inputs $t = (t_1, \ldots, t_d)$ and $u = (u_1, \ldots, u_d)$. The parameter $\theta > 0$ defines the correlation structure of Z; σ_Z^2 is a scale factor. We shall always assume that $EZ(x) = 0$, and it is convenient also to assume that the joint distributions of

the $Z(x)$'s are Gaussian—that is, Z is a Gaussian process. The correlation function is called R.

The appropriateness of this choice of Z for the examples of Section 3 stems from its smooth structure; the realizations of Z are infinitely differentiable with probability 1 (see Parzen 1967). The responses Y in the examples of Section 3 are solutions to systems of differential equations and depend smoothly on the rate constants x that form the inputs. The structure (2.1) is a very strong one and helps explain why our methods work effectively in these examples.

For applications with more erratic responses, we would employ different correlation structures. For example, the process with correlation function $R(t, u) = \exp(-\theta\Sigma|t_j - u_j|)$ can be thought of as a model for functions only required to have one-sided first-order derivatives (see Sacks and Ylvisaker 1966). Integrating this process constructs one that is smoother but less smooth than (2.1) and that may be useful for applications in which some differentiability is present but analyticity may be too strong an assumption. Mitchell, Morris, and Ylvisaker (1988) addressed stationarity issues in such constructions.

It is convenient to use one-dimensional correlation functions to construct d-dimensional ones by defining

$$R(t, u) = \text{corr}(Z(t), Z(u)) = \prod_{j=1}^{d} R_j(t_j, u_j).$$

Moreover, these product correlations will lead to the simpler computations that follow.

The parameter θ in the preceding correlation structures is critical. Prediction is harder when θ is large (small correlations between observations) than when θ is small (large correlations between observations). There is no added difficulty in treating the model of (1.2) and incorporating measurement errors independent of each other and independent of Z. The relevant covariance structure must reflect

$$\text{var}(Y(t)) = \sigma_Z^2 + \sigma_\varepsilon^2. \qquad (2.2)$$

Once θ is specified, (2.1) can be used in (1.1) to provide predictions of $Y(x)$ from data $Y(s_1), \ldots, Y(s_n)$ drawn from a design $S = \{s_1, \ldots, s_n\}$, a set of eligible inputs. The best linear predictor of $Y(x)$ and its mean squared error (MSE) can be obtained as follows. Introduce the notation

$$f_x' = [f_1(x), \ldots, f_k(x)]$$

$$V = [\text{cov}(Y(s_i), Y(s_j))]_{1 \leq i \leq n, 1 \leq j \leq n}$$

$$v_x' = [V(s_1, x), \ldots, V(s_n, x)]$$

$$y' = [Y(s_1), \ldots, Y(s_n)]$$

$$F = [f_l(s_i)]_{1 \leq i \leq n, 1 \leq l \leq k}. \qquad (2.3)$$

If $c'y$ is a linear predictor of $Y(x)$, then its MSE is

$$E[c'y - Y(x)]^2 = (c'F\beta - f'_x\beta)^2$$
$$+ [c', -1] \begin{bmatrix} V & v_x \\ v'_x & \sigma^2_Z \end{bmatrix} \begin{bmatrix} c \\ -1 \end{bmatrix}. \quad (2.4)$$

Typically, an "unbiasedness" requirement leads to the constraints $F'c = f_x$. An alternative minimax argument is that the maximum over β of the right side of (2.4) is infinite unless the constraints are satisfied. In any case, we adopt them.

Thus the best linear predictor (BLP) of $Y(x)$ is obtained by minimizing $E(c'y - Y(x))^2$ subject to $F'c = f_x$. Use of Lagrange multipliers produces $Vc - v_x - F\lambda = 0$ and $F'c = f_x$, or

$$\begin{bmatrix} 0 & F' \\ F & V \end{bmatrix} \begin{bmatrix} -\lambda \\ c \end{bmatrix} = \begin{bmatrix} f_x \\ v_x \end{bmatrix}. \quad (2.5)$$

The BLP is then

$$c'y = [-\lambda', c'] \begin{bmatrix} 0 \\ y \end{bmatrix}$$
$$= [f'_x, v'_x] \begin{bmatrix} 0 & F' \\ F & V \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix}, \quad (2.6)$$

which can also be written as $\hat{Y}(x) = f'_x\hat{\beta} + v'_xV^{-1}(y - F\hat{\beta})$, where $\hat{\beta} = (F'V^{-1}F)^{-1}F'V^{-1}y$ is the generalized least squares estimate of β. Equations (2.4) and (2.5) imply that

$$MSE(\hat{Y}(x)) = \sigma^2_Z - [f'_x, v'_x]$$
$$\times \begin{bmatrix} 0 & F' \\ F & V \end{bmatrix}^{-1} \begin{bmatrix} f_x \\ v_x \end{bmatrix} \quad (2.7)$$

is the MSE of prediction.

In the absence of measurement error [$\sigma^2_\varepsilon = 0$ in (2.2)], the estimated response surface interpolates the observations because the predictor $\hat{Y}(s_i) = Y(s_i)$ at a design point s_i has MSE 0.

Of the many possible design criteria, we minimize the integrated mean squared error (IMSE) of prediction. Given the true correlation parameter θ and a design-prediction strategy $(S, \hat{Y})$, the normalized (for σ^2_Z) IMSE is

$$J_0(S, \hat{Y}) = (1/\sigma^2_Z) \int E_\theta(\hat{Y}(x) - Y(x))^2 dx. \quad (2.8)$$

The integration is over the region of interest and could be adapted to include a weight function. The predictor $\hat{Y}$ will usually be the BLP (2.6), which depends on θ. Therefore, the design problem is to choose a design S minimizing (2.8). In the following examples, we always restrict the design points to lie in the prediction region. For a given design, the integral in the criterion (2.8) can be written using (2.7)

as

$$\sigma^2_Z - tr\left\{ \begin{bmatrix} 0 & F' \\ F & V \end{bmatrix}^{-1} \int \begin{bmatrix} f_xf'_x & f_xv'_x \\ v_xf'_x & v_xv'_x \end{bmatrix} dx \right\}. \quad (2.9)$$

The integrals involving f_x and v_x simplify to products of one-dimensional integrals if the elements of f_x and v_x are products of functions of a single input factor. Thus polynomial linear models, with each term of the form $x_1^{q_1} \cdots x_d^{q_d}$, and a product covariance such as (2.1) are numerically convenient.

To minimize the criterion (2.8) as a function of the $n \times d$ design-point coordinates, we have mainly used a quasi-Newton optimizer. Computations were performed on the Cray X-MP/48 computer at the National Center for Supercomputing Applications at the University of Illinois. The Cray vectorizing architecture is well suited to the linear algebra necessary for evaluations of (2.9). As often happens with optimization, only a local minimum may be found, so several random starting designs were tried.

Note that σ^2_Z plays no role in this minimization; knowledge of θ is, however, critical. Because θ is generally not available for the design stage, a robustness study, described in Section 3, is used to choose θ. In Example 1 (Sec. 3) there are inadequate data for estimating θ and the robust θ is also used for prediction. In Example 2 (Sec. 3), however, there are enough data to estimate θ for prediction.

3. EXAMPLES

Example 1

The first example, homogeneous pyrolysis of propane, was discussed by Miller and Frenklach (1983). This chemical kinetics problem is modeled by a linear system of 11 differential equations:

$$\dot{\eta}_j(x, t) = g_j(\eta, x, t), \quad j = 1, \ldots, 11, \quad (3.1)$$

where x is a set of rate constants, the inputs to the system. A solution to (3.1) can be obtained numerically for any input x by use of a differential-equations solver, yielding concentrations of five chemical species at a reaction time of 7×10^{-4} seconds. All but two rate constants are hypothesized to have been established by previous work, so x is two-dimensional.

Miller and Frenklach (1983) took a nine-point design and approximated each of the five log concentrations with a two-dimensional quadratic. Similarly, our model is

$$Y(x) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{11}x_1^2$$
$$+ \beta_{22}x_2^2 + \beta_{12}x_1x_2 + Z(x). \quad (3.2)$$

where $Y(x)$ is the log concentration, but we model the departure Z as a stochastic process with corre-

lation function given by (2.1). No measurement error is involved here. The inputs x are transformed so that the prediction region is the square of side $[-\frac{1}{2}, \frac{1}{2}]$. We present the results of modeling one of the Y's; similar results were obtained for the others.

To select a design by minimizing (2.8) we need to specify θ. We will also use this θ in (2.6) to give the BLP $\hat{Y}_\theta$. Unfortunately, before taking observations from an experiment, there is no way to know θ or even make an intelligent guess. Therefore, we need to choose an assumed θ_A that will give a design-prediction strategy that performs well for a wide range of true (but unknown) θ_T.

To do this, we select several assumed values θ_A; for each, we find the optimal design $S^*_{\theta_A}$ and BLP $\hat{Y}_{\theta_A}$. Figure 1 gives the designs for $\theta_A = 1$ and $\theta_A = 100$. The $\theta_A = 1$ design is typical of those for $\theta_A \leq 1$. As θ_A increases, the designs become more regular as depicted in the $\theta_A = 100$ design. We want to select a strategy that performs well over a wide range of true θ_T. Table 1 gives values of $J_{\theta_T}(S^*_{\theta_A}, \hat{Y}_{\theta_A})$, the measurement of performance of the strategy $(S^*_{\theta_A}, \hat{Y}_{\theta_A})$ when θ_T is true. Since $J_{\theta_T}(S^*_{\theta_T}, \hat{Y}_{\theta_T})$ is the best one can do when θ_T is true, the ratio

$$J_{\theta_T}(S^*_{\theta_T}, \hat{Y}_{\theta_T})/J_{\theta_T}(S^*_{\theta_A}, \hat{Y}_{\theta_A})$$

is the relative efficiency of $(S^*_{\theta_A}, \hat{Y}_{\theta_A})$ at θ_T. These efficiencies appear in parentheses in Table 1.

On the basis of the relative efficiencies in Table 1, the strategy $(S^*_1, \hat{Y}_1)$ for $\theta_A = 1$ appears most robust in the sense of maximizing minimum efficiency. It is interesting that large absolute differences are guarded against by the strategy for $\theta_A = 100$. The IMSE using $\theta_A = 100$, however, is uniformly poor

and improvements from $1.55\sigma_Z^2$ to $1.20\sigma_Z^2$ are of little value.

We compared the strategy $(S^*_1, \hat{Y}_1)$ with the Miller and Frenklach (1983) strategy of a 3×3 factorial design $\{-.5, 0, .5\}^2$ and least squares (LS) prediction, denoted by $(3 \times 3, \hat{Y}_{LS})$. We took a 101×101 grid of points, G, on the unit square and ran the computational model to produce true Y's at each $x \in G$ and then computed the empirical integrated squared error

$$\text{EISE} = (1/(101)^2) \sum (\hat{Y}(x) - Y(x))^2 \quad (3.3)$$

for the two strategies.

The design S^*_1 is not completely symmetric, and there are three equivalent designs obtained by reflections in the unit square. Therefore, four EISE values for strategy $(S^*_1, \hat{Y}_1)$ are 1.6, 1.7, 1.9, and 2.2. The EISE for $(3 \times 3, \hat{Y}_{LS})$, 16.3, shows that our method reduces the EISE in this example by a factor of 8–10, depending on which one of the S^*_1 designs is (arbitrarily) chosen.

Example 2

The second example is more complicated than Example 1. In this instance, there is a large linear system of differential equations describing a methane combustion process. Seven rate constants are deemed active in the solution for a given set of initial conditions. The output of interest is the induction–delay time.

Modeling proceeds as in Example 1 using quadratic approximation (following discussion with M. Frenklach). In seven dimensions, the quadratic model has 36 terms. Again, the stochastic process is
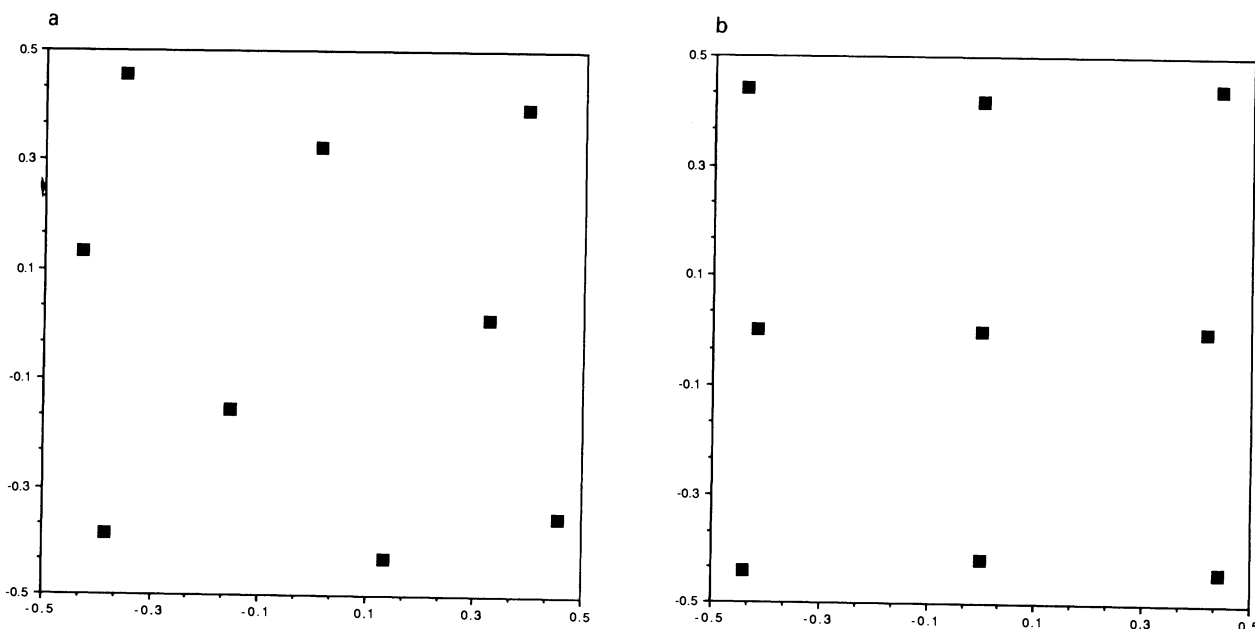

Figure 1. Optimal Nine-Point Designs for a Quadratic Model: (a) $\theta_A = 1$; (b) $\theta_A = 100$.

Table 1. Integrated Mean Squared Error of Strategy $(S_{\theta_A}^*, \hat{Y}_{\theta_A})$ When the True Correlation Parameter Is θ_T

Assumed θ_A	True θ_T						
	.25	.5	1	2	5	10	100
.25	.17 × 10⁻⁴ (100%)	.15 × 10⁻³ (99%)	.00128 (96%)	.0100 (89%)	.099 (78%)	.33 (74%)	1.55 (78%)
.5	.17 × 10⁻⁴ (99%)	.15 × 10⁻³ (100%)	.00125 (98%)	.0096 (92%)	.095 (81%)	.32 (77%)	1.54 (78%)
1	.18 × 10⁻⁴ (94%)	.15 × 10⁻³ (98%)	.00122 (100%)	.0092 (97%)	.089 (87%)	.31 (81%)	1.51 (80%)
2	.21 × 10⁻⁴ (79%)	.17 × 10⁻³ (87%)	.00128 (96%)	.0089 (100%)	.082 (94%)	.29 (87%)	1.48 (82%)
5	.33 × 10⁻⁴ (51%)	.24 × 10⁻³ (60%)	.00164 (74%)	.0099 (90%)	.077 (100%)	.26 (96%)	1.41 (85%)
10	.48 × 10⁻⁴ (35%)	.34 × 10⁻³ (43%)	.00221 (55%)	.0122 (73%)	.081 (95%)	.25 (100%)	1.36 (89%)
100	1.46 × 10⁻⁴ (11%)	1.02 × 10⁻³ (14%)	.00631 (19%)	.0317 (28%)	.166 (47%)	.39 (64%)	1.20 (100%)

defined by (2.1) but with $d = 7$. Frenklach adopted a "box and star" design, usually called a central-composite design (Box and Hunter 1957). The box is a half-fraction of a 2^7 design on the vertices of the 7-cube, and the star points are the 14 points $(\pm\frac{1}{2}, 0, \ldots, 0), (0, \pm\frac{1}{2}, 0, \ldots, 0), \ldots$, and the origin. The design consists of these 79 points. Such designs are commonly used for fitting quadratic responses with independent errors. Because of the computational complexity of minimizing over designs with 79 × 7 coordinates, we also restricted attention to box and star designs and sought best designs of the form $(\alpha_1 \cdot \text{box}, \alpha_2 \cdot \text{star})$, where α_1 and α_2 are between 0 and 1. By $\alpha_1 \cdot$ box, we mean a scale shrinkage of the box design by α_1 with a similar meaning for $\alpha_2 \cdot$ star.

Let $S_{\theta_A}^B$ be the best 79-point box–star design for a

quadratic model assuming θ_A; that is, $S_{\theta_A}^B$ minimizes $J_{\theta_A}(S, \hat{Y}_{\theta_A})$ over $S \in (\alpha_1 \text{box}, \alpha_2 \text{star})$. Enough data are available to estimate the true θ for the prediction stage, so the robustness study is slightly different from that in Example 1. Here we evaluate $J_{\theta_T}(S_{\theta_A}^B, \hat{Y}_{\theta_T})$ for the efficient predictor $\hat{Y}_{\theta_T}$. This leads to Table 2 and the conclusion that S_5^B with .64 · box, 1.00 · star is reasonably robust as are the designs for θ_A between 1 and 5.

For brevity, Table 2 is based on the assumption of a second-order linear model, but it is possible to perform an expanded robustness study that also allows for various assumed and true linear models as well as assumed and true values of θ. Thus the IMSE of a design optimal for assumed θ_A and assumed linear model is evaluated under various θ_T and var-

Table 2. Integrated Mean Squared Error of Strategy $(S_{\theta_A}^*, \hat{Y}_{\theta_T})$ When the True Correlation Parameter Is θ_T

Design assumed θ_A	Shrinkage factors		True θ_T				
	α_1	α_2	.1	.5	1.0	5.0	25.0
.1	.67	.80	.80 × 10⁻⁴ (100%)	.0099 (98%)	.065 (96%)	1.16 (86%)	1.94 (71%)
.5	.65	.93	.81 × 10⁻⁴ (98%)	.0098 (100%)	.063 (99%)	1.04 (96%)	1.68 (83%)
1.0	.64	1.0	.83 × 10⁻⁴ (96%)	.0098 (100%)	.062 (100%)	1.00 (100%)	1.60 (87%)
5.0	.64	1.0	.82 × 10⁻⁴ (97%)	.0098 (100%)	.063 (100%)	1.00 (100%)	1.59 (87%)
25.0	1.0	1.0	1.52 × 10⁻⁴ (52%)	.0254 (38%)	.147 (42%)	1.14 (88%)	1.39 (100%)

ious true linear models. Table 2 is actually just part of such a study considering an eight-parameter first-order model and a one-parameter constant model, in addition to the second-order model. Simpler linear models might seem reasonable because of the flexibility allowed by the stochastic process, but the more extensive comparisons showed that the design for the second-order model and $\theta_A = 5$ continues to be robust.

From design S_5^β, we generated data with the computer model and estimated σ_Z^2 and θ by maximum likelihood; $\hat{\sigma}_Z^2 = 4.6 \times 10^{-5}$ and $\hat{\theta} = 12.4$. For this purpose Z is assumed to be Gaussian process. (Only $\hat{\theta}$ is used for prediction, but $\hat{\sigma}_Z^2$ is required for inference.) The IMSE evaluations in Table 2 indicate that the strategy $(S_5^\beta, \hat{Y}_{12.4})$ has an IMSE between $1.00\sigma_Z^2$ and $1.59\sigma_Z^2$ and σ_Z^2 is about 5×10^{-5} from the maximum likelihood estimation. When the true Y was evaluated at 200 random inputs, the EISE computed analogously to (3.3) was 15.2×10^{-5} with standard error 3×10^{-5}. The IMSE, being an expectation over hypothetical realizations of the stochastic process, is not necessarily a proper measure of uncertainty in the prediction. In this case, however, its order of magnitude is correct; the IMSE is about $\frac{1}{3}$ of the EISE.

Concern that the use of the 36-parameter quadratic model leads to overfitting prompted us to evaluate the simpler first-order and constant models. In each case we retained the design S_5^β based on the second-order model but reestimated the parameters θ and σ_Z^2. The estimates and EISE (including those for the quadratic model for comparison) are shown in Table 3. Interestingly, the three linear models have similar EISE's. This suggests a trade-off between the complexity of the model and the magnitude of θ and σ_Z^2. Further work is needed to determine the generality of this phenomenon.

The designs in Table 2 differ little in their shrinkage parameters for the box points (which predominate) except for the $\theta_A = 25$ design. Using a quadratic model and reestimating θ by $\hat{\theta} = 2.57$, the $\theta_A = 25$ design has an EISE of 40.8×10^{-5} or 37% efficiency relative to the (robust) $\theta_A = 5$ design. A similar comparison shows that a 79-point Latin hypercube design (McKay et al. 1979) supplied to us

Table 3. Parameter Estimates and Empirical Integrated Squared Error of Prediction (EISE) for the $\theta_A = 5$ Box–Star Design

Linear model	$\hat{\theta}$	$\hat{\sigma}^2$	EISE
Constant	.080	1.75	10.7×10^{-5}
First-order	.28	.053	13.0×10^{-5}
Quadratic	12.4	4.6×10^{-5}	15.2×10^{-5}

by R. Iman has an efficiency of 58%. (The Latin hypercube method gives very variable results in the first example when the design is reflected, but the efficiency averages about 30%.)

4. DISCUSSION

Although we have restricted attention to issues in which the design space and the region of interest coincide, it is clear that similar tactics work with contexts in which, for example, the region of interest is larger than the design region. The IMSE criterion can be replaced by other criteria. The maximum mean squared error (MMSE) is a natural criterion, but it involves a multidimensional optimization at every design step. Comparisons of IMSE and MMSE for discrete regions were made by Sacks and Schiller (1988).

Design optimization can be computationally formidable for the models considered in this article. With n design points and d-dimensional inputs, full optimization requires the solution of a large system of equations to compute the IMSE. The $n \times n$ covariance matrix V involved in these equations is poorly conditioned for small correlation parameter θ, which is often the case of interest (e.g., Table 3).

For larger problems, therefore, heuristics like the box–star (central-composite) designs employed in Section 3 may be useful. Sequential design is also natural; information about the unknown θ can be acquired, thus bypassing the design-robustness issue. Linear-model adaptation might also be possible. Our experience, however, is that simple-minded sequential rules, such as looking one point ahead, are ineffective. Another approach that circumvents the problem of not knowing θ at the design stage is to design for asymptotic θ's (θ approaching 0). Recent results obtained with Y. B. Lim and W. J. Studden on asymptotic ($\theta \to 0$) designs makes this method appear promising as a tool for design construction.

ACKNOWLEDGMENTS

REFERENCES

Box, G. E. P., and Draper, N. R. (1959), "A Basis for the Selection of a Response Surface Design," *Journal of the American Statistical Association*, 54, 622–654.

—— (1963), "The Choice of a Second Order Rotatable Design," *Biometrika*, 50, 335–352.

Box, G. E. P., and Hunter, J. S. (1957), "Multi-Factor Experimental Designs for Exploring Response Surfaces," *The Annals of Mathematical Statistics*, 28, 195–241.

Cressie, N. (1986), "Kriging Nonstationary Data," *Journal of the American Statistical Association*, 81, 625–634.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1988), "A Bayesian Approach to the Design and Analysis of Computer Experiments," Technical Report 6498, Oak Ridge National Laboratory, Oak Ridge, TN (available from National Technical Information Service, Springfield, VA 22161).

Diaconis, P. (1988), "Bayesian Numerical Analysis," in *Statistical Decision Theory and Related Topics IV* (Vol. 1), eds. S. S. Gupta and J. O. Berger, New York: Springer-Verlag, pp. 163–175.

Iman, R. L., and Conover, W. J. (1980), "Small Sample Sensitivity Analysis Techniques for Computer Models, With an Application to Risk Assessment" (with discussion), *Communications in Statistics, Part A—Theory and Methods*, 9, 1749–1874.

McKay, M. D., Conover, W. J., and Beckman, R. 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.

Miller, D., and Frenklach, M. (1983), "Sensitivity Analysis and Parameter Estimation in Dynamic Modeling of Chemical Kinetics," *International Journal of Chemical Kinetics*, 15, 677–696.

Mitchell, T., Morris, M., and Ylvisaker, D. (1988), "Existence of Smoothed Stationary Processes on an Interval," unpublished manuscript.

Parzen, E. (1967), "Statistical Inference on Time Series by Hilbert Space Methods I," in *Time Series Analysis Papers by Emanuel Parzen*, San Francisco: Holden-Day, pp. 251–382.

Phadke, M. S. (1986), "Design Optimization Case Studies," *AT&T Technical Journal*, 65, 51–68.

Ripley, B. D. (1981), *Spatial Statistics*, New York: John Wiley.

Sacks, J., and Schiller, S. (1988), "Spatial Designs," in *Statistical Decision Theory and Related Topics IV* (Vol. 2), eds. S. S. Gupta and J. O. Berger, New York: Springer-Verlag, pp. 385–399.

Sacks, J., and Ylvisaker, D. (1966), "Designs for Regression Problems With Correlated Errors," *The Annals of Mathematical Statistics*, 37, 66–89.

—— (1970), "Statistical Designs and Integral Approximation," in *Proceedings of the 12th Biennial Seminar of the Canadian Mathematical Congress*, ed. R. Pyke, Montreal: Canadian Mathematical Congress, pp. 115–136.

—— (1984), "Some Model Robust Designs in Regression," *The Annals of Statistics*, 12, 1324–1348.

—— (1985), "Model Robust Design in Regression: Bayes Theory," in *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (Vol. 2), eds. L. M. Le Cam and R. A. Olshen, Monterey, CA: Wadsworth, pp. 667–679.

Shewry, M. C., and Wynn, H. P. (1987), "Maximum Entropy Sampling," *Journal of Applied Statistics*, 14, 165–170.

Speckman, P. L. (1976), "Approximation of Autoregressive Gaussian Processes," unpublished thesis, University of California, Los Angeles, Dept. of Mathematics.

Stein, M. (1987), "Large Sample Properties of Simulations Using Latin Hypercube Sampling," *Technometrics*, 29, 143–151.

Suldin, A. V. (1959), "Wiener Measure and Its Applications to Approximation Methods, I," *Izvestiia Vysshikh Uchebnykh Zavedenii Matematika*, 6 (13), 145–158.

—— (1960), "Wiener Measure and Its Applications to Approximation Methods, II," *Izvestiia Vysshikh Uchebnykh Zavedenii Matematika*, 5 (18), 165–179.

Welch, W. J. (1983), "A Mean Squared Error Criterion for the Design of Experiments," *Biometrika*, 70, 205–213.

Welch, W. J., Yu, T. K., Kang, S. M., and Sacks, J. (1988), "Computer Experiments for Quality Control by Parameter Design," Technical Report 4, University of Illinois, Dept. of Statistics.

Ylvisaker, D. (1975), "Designs on Random Fields," in *A Survey of Statistical Design and Linear Models*, ed. J. N. Srivastava, Amsterdam: North-Holland, pp. 593–607.

—— (1987), "Prediction and Design," *The Annals of Statistics*, 15, 1–19.