be possible to take account of this in the search algorithm, to both speed the search and evade numerical problems on the way. Nevertheless, large n must always present problems.

The only comment in the paper which jars with my own experience is the reference to designing for very large θ, in the Currin, Mitchell, Morris and Ylvisaker (1988) paper. When θ is large, you cannot estimate $Z(\cdot)$ except very locally to each design point. The second part of (7), which smooths the residuals, consists of zero almost everywhere except for blips at each design point to make $y(x)$ pass through the observation. Designs for this case will be exclusively concerned with estimating the regression function and, like classical optimal design for regression, will place clusters of points at the boundaries of the design region. Such designs must be very poor when θ is in reality not large.

I was very intrigued to see the decomposition of $Y(\cdot)$ into main effects, interactions, etc. In my context where $Y(\cdot)$ is a multivariate density function, the main effects are just marginal densities. The interactions as defined, however, have no particular value. Instead I would define

$$\mu_{ij}(x_i, x_j) = \int y(x)\textstyle\prod_{h\neq i,j} dx_h - \mu_i(x_i)\mu_j(x_j),$$

representing non-independence between x_i and x_j.

It should be clear from my remarks how much I have enjoyed reading this paper. The wealth of detail and the authors' breadth of knowledge make it one that I am sure to turn to repeatedly.

ADDITIONAL REFERENCES

GOLDSTEIN, M. (1979). The variance modified linear Bayes estimator. *J. Roy. Statist. Soc. Ser. B* **41** 96–100.

O'HAGAN, A. (1987). Bayes linear estimators for randomized response models. *J. Amer. Statist. Assoc.* **82** 580–585.

O'HAGAN, A. (1988). Bayesian quadrature. Warwick Statistics Research Report 159, Univ. Warwick.

Comment

Michael L. Stein

I wholeheartedly agree with the authors that statisticians can and should contribute to the design and analysis of computer experiments. Too often statisticians shy away from problems that do not fit into the standard statistical frameworks; the authors are to be congratulated for their trailblazing efforts. Furthermore, I agree that a sensible way to approach these problems is to view the output from the computer model as a realization of a stochastic process. Where I think further work is needed is in the development of appropriate stochastic models.

The model given by (9) in this article by Sacks, Welch, Mitchell and Wynn has some undesirable properties. For $0 < p < 2$, a stochastic process with this covariance function will not be mean square differentiable. As noted by the authors, for $p = 2$, the process is infinitely mean square differentiable. Not allowing processes that are differentiable but not infinitely differentiable strikes me as unnecessarily restrictive. A more flexible class of correlation functions is (Yaglom, 1987, page 139)

$$\prod \frac{1}{\Gamma(\nu)2^{\nu-1}} (\alpha_j \,|\, w_j - x_j \,|)^\nu K_\nu(\alpha_j \,|\, w_j - x_j \,|),$$

where K_ν is a modified Bessel function of order ν (Abramowitz and Stegun, 1965, page 374). A stochastic process with this covariance function will be m times mean square differentiable if and only if $\nu > m$. The α_js measure the range of the correlation: a large α_j indicates that correlations die out quickly in the x_j direction.

A problem with all of the correlation functions used by Sacks, Welch, Mitchell and Wynn is that they do not allow for the inclusion of prior knowledge such as that most of the variation in the output $y(\cdot)$ can probably be explained by main effects plus perhaps some low order interactions, which in fact occurred in the circuit simulator example they discuss. If we expected most of the variation in $y(\cdot)$ could be explained by main effects, we might want to model $Y(x)$ as

$$(1) \qquad Y(x) = \sum Y_j(x_j) + Z(x),$$

Michael L. Stein is Assistant Professor, Department of Statistics, University of Chicago, 5734 University Avenue, Chicago, Illinois 60637.

where the Y_j's and Z are independent Gaussian processes with covariance functions $\sigma_j(x_j - w_j)$ and $\sigma_z(x - w)$ respectively, so that

$$\mathrm{cov}(Y(x), Y(w)) = \sum \sigma_j(x_j - w_j) + \sigma_z(x - w).$$

One specific parametric form of this model that might be worth exploring is

$$\begin{aligned}\mathrm{cov}(Y(x), Y(w)) \\ &= \sum C_j(\alpha_j \,|\, w_j - x_j \,|)^\nu K_\nu(\alpha_j \,|\, w_j - x_j \,|) \\ &\quad + D \prod (\beta_j \,|\, w_j - x_j \,|)^\nu K_\nu(\beta_j \,|\, w_j - x_j \,|).\end{aligned}$$

A large C_j would correspond to an important main effect. The model for $Z(\cdot)$ is somewhat problematic as it allows $Z(\cdot)$ to have an additive component. Following the decomposition into main effects and interactions from Section 6 of the article by Sacks, Welch, Mitchell and Wynn, it might be more satisfying to define $Z(\cdot)$ to be a stochastic process with no additive component:

$$Z(x) = Z^*(x) - \sum_j \int Z^*(x) \prod_{h \neq j} dx_h + (d - 1) \int Z^*(x)\,dx,$$

where d is the number of dimensions of x and $Z^*(x)$ is a Gaussian process with some simple covariance function. I think it would be very interesting to find optimal designs under some models of the general form given by (1). If the optimal designs are very different from those obtained by Sacks, Welch, Mitchell and Wynn for their models, that would call into question the effectiveness of their designs for processes where most of the variation can be explained by main effects.

ADDITIONAL REFERENCES

ABRAMOWITZ, M. and STEGUN, I. (1965). *Handbook of Mathematical Functions*, 9th ed. Dover, New York.

YAGLOM, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions* **1**. Springer, New York.

Rejoinder

Jerome Sacks, William J. Welch, Toby J. Mitchell and Henry P. Wynn

We thank the discussants for their incisive comments, suggestions and questions. Nearly all the discussants have been key participants at the workshops mentioned by Johnson and Ylvisaker; all have been instrumental in the development of new methodologies for the design and analysis of computer experiments. Most of the comments and our responses are concerned with the choice of the experimental design and the choice of the correlation function.

We had hoped that the example of Section 6 would attract some suggestions from the discussants, and in this we are not disappointed. Morris' results on the first-stage, 16-point design are interesting—in particular, they indicate that the concentration of the design in the center of the region also occurs for the much rougher process corresponding to $p = 1$ in (9). As this is only a preliminary stage, and there is not much to be lost by using a cheaper design anyway, his scaled quarter fraction makes a lot of sense. In a seven-dimensional problem, Sacks, Schiller and Welch (1989) similarly reduced the optimization problem by restricting attention to scaled central-composite designs. Without doing the optimization or amassing experience from many problems, though, we cannot know when the relative performance of cheap designs will be satisfactory.

For all 32 runs, Easterling recommends two complementary quarter fractions. He rightly points out the advantage of not having to optimize anything, and we tried these fractions on $\{-\frac{1}{2}, \frac{1}{2}\}^6$ and $\{-\frac{1}{4}, \frac{1}{4}\}^6$. In some recent applications where data are cheap to generate, we have been using Latin hypercube designs, and for comparison we also report results for a 32-run Latin hypercube. The six factors have the same 32 equally spaced values, $-\frac{31}{64}, -\frac{29}{64}, \ldots, \frac{31}{64}$, but in different random orders. For both designs, the predictor is based on model (14) after re-estimating the parameters $\theta_1, \ldots, \theta_6$ and p in the correlation function (9). Table R1 shows the average squared error of prediction at the same 100 random points we used previously. For ease of comparison, the results for our original design are repeated. The complementary quarter fractions and the Latin hypercube perform similarly, with our design showing a modest advantage.

It is of interest to note that, for certain values of n and d, scaled standard designs can be optimal. For 8 points in 4 dimensions and 16 points in 5 dimensions