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

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

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

$$\operatorname{cov}(Y(x), Y(w)) = \sum C_j(\alpha_j | w_j - x_j |)^{\nu} K_{\nu}(\alpha_j | w_j - x_j |) + D \prod (\beta_j | w_j - x_j |)^{\nu} K_{\nu}(\beta_j | w_j - x_j |).$$

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

the optimization problem is still tractable. Suitably scaled half fractions of maximum resolution are apparently optimal or very close to optimal for the IMSE criterion when the model has only a constant term for the regression, and the correlation family is (9) with $p = 2$. The scaling of the design depends on the value of θ. (We regret that we did not include more of these anecdotes.) Extrapolating these results to the six-dimensional problem at hand, we tried a half fraction (I = ABCDEF) on $\{-0.37, 0.37\}^6$, a scaling associated with a small value of θ. As seen in Table R1, this design performs much the same as the complementary quarter fractions. It is quite likely that a Faure sequence, favored by Owen, Koehler and Sharifzadeh, or a geometrical design, as in Johnson, Moore and Ylvisaker (1988), would also perform similarly.

These very different designs produce rather similar results, then. In this respect, our choice of example was less revealing than we had hoped. Whether this is a feature of the particular surface or something more general we can only guess.

In other examples optimal design provides greater benefits. The 4^2 design suggested by Easterling for the problem in Currin, Mitchell, Morris and Ylvisaker (1988) performs relatively poorly. For the cubic correlation function (11) with parameters determined by maximum likelihood, the empirical average squared error of prediction on a 21×21 grid is $(0.62)^2$ for the optimal design shown in Easterling's figure, compared to $(0.94)^2$ for the 4^2 design. Although this is just one example, it does indicate that the well balanced, symmetric design does not necessarily perform best, and the difference is not necessarily trivial. Incidentally to us, the Currin, Mitchell, Morris and Ylvisaker design seems rather elegant—beauty of a design is in the eye of the beholder.

One very important place for the use of optimal design is for less-regular regions where intuition is lacking. Ongoing work by D. Cox, J. Park and C. Singer on a computer model for a nuclear-fusion device (Tokamak) has a six-dimensional region which is determined by linear constraints. The cost of generating data is also fairly high, 3–5 minutes on a Cray 2 per run. Here, no regular, symmetric designs easily come to mind.

TABLE R1
Empirical average squared error of prediction at the 100 random points

Design	Average squared error
Authors'	$(.122)^2$
Complementary quarter fractions	$(.146)^2$
Latin hypercube	$(.136)^2$
Half fraction on $\{-0.37, 0.37\}^6$	$(.143)^2$
Authors' (Stein's correlation function)	$(.115)^2$

We are intrigued by O'Hagan's experience in applying similar models to the estimation of integrals. We have some unpublished results relating to quadrature in two dimensions. Low-discrepancy sequences (e.g., Halton sequences) perform well on the average for functions generated by the model (9) when $p = 1$. However, for functions generated by (9) with $p = 2$, which are much smoother, the average performance of the Halton sequences is poor relative to the optimal designs obtainable for small problems or relative to various *ad hoc* schemes for larger problems.

Morris' connections between interactions and the correlation parameters θ in (9) suggest small rather than large values of θ, a view shared by O'Hagan. Our experience with estimating these parameters in a number of examples, using models with no regression terms other than a constant, is also consistent with smaller values. This stands in contrast to Johnson and Ylvisaker's results that *designs* based on large θ have certain robustness properties. How efficient their designs are when small values of θ are appropriate, or when linear models are incorporated, is not clear.

To sum up on the choice of design, we suspect with Easterling that, providing that the design does a reasonable job of infiltrating the space, the predictor is probably more important than the design. Unfortunately, we do not have enough evidence to strengthen these suspicions, nor to make notions like infiltration more precise. Sometimes standard designs like fractional factorials do fairly well; for other problems they perform rather poorly. To know which case is true, the optimal design is necessary as a benchmark. Johnson and Ylvisaker correctly point out that we do not yet have catalogs of useful designs that can be automatically applied. Clearly, more work is needed, and advanced computations seem indispensable.

Stein suggests a more flexible class of correlation functions. We took the design and data of Table 1 and maximized the likelihood over $\alpha_1, \ldots, \alpha_6$ and $\nu = 1, 2, 3$. Predictions based on $\hat{\nu} = 1$ and $\hat{\alpha} = (.260, .255, .446, .566, .466, .934)$ at the 100 random points give the results reported in Table R1. There is some improvement and this family may be worth pursuing further. We do note, however, that optimization of the likelihood is more costly, and there may be numerical instabilities associated with computing K_ν as ν grows.

Stein's additive covariance models seem promising. Though they introduce a number of additional parameters (the C's and D), we agree they may be useful when the output is nearly additive. There may also be important design considerations.

Any help in estimating the correlation parameters is welcome, and we look forward to seeing further

details of the method outlined by Owen, Koehler and Sharifzadeh. These discussants also note that smaller θ's may indicate inactive factors, with implications for dimension reduction. This is usually valid, but we have found, for example, cases in which $\hat{\theta}$ is close to zero for a variable with a strong linear effect. This is backed up by theoretical work on asymptotic behavior as $\theta \to 0$ when $p = 2$ in (9), as mentioned in Section 7.4. To avoid overinterpretation of the θ's we endorse the plotting of the main effects, interactions and so on defined in Section 6.

Black box or gray box? We could not agree more with Easterling about the need to employ subject-matter expertise. Progress in applications and new methodologies requires two-way exchange between statisticians and the scientists conducting these experiments. In our experience, as in the example of Section 6, the expert has usually reduced the number of factors to a small set, most of which are active for one response or another. To ensure that all important factors are included, however, a screening stage might be used to determine the active set empirically. In this case, as Owen, Koehler and Sharifzadeh point out, designs that project well for one or relatively few active factors will also be more useful for prediction. Overly symmetric designs like fractional factorials may have replicates when projected in this way. It might be helpful to incorporate prior information on effects sparsity into the assumed models, with implications for design.

O'Hagan sheds some more light on the Bayesian viewpoint here, to which he has made important contributions, and Morris points out some difficulties with the frequentist interpretation. In earlier drafts we did attempt to discuss these philosophical matters more fully, but we gave up due to differences amongst the authors! A full Bayesian viewpoint might offer some advantages. Unfortunately, as O'Hagan points out, unknown correlation parameters are not easy to deal with in a full Bayesian framework.

We are grateful that the discussants share and reinforce our excitement in developing this area. It is clear to us that there is much work to be done; we hope that there will be many to do it.