Proceedings of the 1992 Winter Simulation Conference
ed. J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson

VARIANCE REALLOCATION IN TAGUCHI'S ROBUST DESIGN FRAMEWORK

Lee W. Schruben

Operations Research and Industrial Engineering
Cornell University
Ithaca, New York 14853

Susan M. Sanchez

College of Business and Public Administration
The University of Arizona
Tucson, Arizona 85721

Paul J. Sanchez

Systems and Industrial Engineering
The University of Arizona
Tucson, Arizona 85721

Veronica A. Czitrom

Quality Progress Center
AT&T Bell Laboratories
Holmdell, New Jersey 07733

ABSTRACT

The appropriate use of antithetic random variates has been shown to improve the precision of response surface model estimation for simulation. We apply this approach to simulation experiments designed to determine operating conditions that reduce response variability by using Genichi Taguchi's parameter design framework. Antithetic random number streams can be viewed as another level of complexity in the experiment design: we call this class of simulation-specific factors *artificial factors*. A simple example illustrates how antithetic random variates may be beneficial for robust design in simulation settings.

1 INTRODUCTION

Genichi Taguchi pioneered a strategy for deliberately incorporating sources of variability into the evaluation of alternative manufactured product designs (Taguchi 1986, Taguchi and Wu 1980). He found that it was often more costly to control causes of manufacturing variation than to make a process insensitive to these variations.

Taguchi's three-stage approach for quality improvement activities consists of *system design*, *parameter design*, and *tolerance design*. In the manufacturing setting, system design is the application of scientific and engineering knowledge to produce a functional prototype model. This prototype model defines the product/process design characteristics (parameters) and their initial settings. In parameter design settings that reduce the variability in the response are identified; at the same time, the mean response may be maximized, minimized or adjusted to a target value. Tolerance design is a method for scientifically assigning tolerances in order to minimize total product manufacturing and lifetime costs. We concentrate on parameter design (also known as robust design) in this paper.

As part of parameter design, factors which are expected to affect the response of interest are classified into *control parameters* and *noise factors*. Control parameters are those factors which can be controlled, or set, during normal operating conditions in order to influence the response. Noise factors are sources of variability which are either not controllable or are too expensive to control under normal operating conditions, but can be controlled during experimentation. Parameter design is a methodology to reduce variability in the response by finding settings of the control parameters that are robust, or insensitive, to variations in the noise factors. Noise factors may either be internal sources of variation (such as drill bit wear, raw material composition, variations in operator skill and/or timing) or external sources (such as variation in temperature or humidity in a customer's environment over the lifetime of the product).

Taguchi advocates experimentation using noise factors to span the space of the noise—an active introduction of noise which is more effective than replication (Nair, 1992). Taguchi's use of loss functions to convert engineering deviations into costs has also made an important, although often underestimated, contribution by motivating management to carry out the robust design process and other quality improvement activities (Lucas 1985, Pignatiello and Ramberg 1991).

Although Taguchi's arena of application has been manufacturing, these concepts may also be applied to investigate the behavior of simulated systems (Ramberg et al. 1991). In the simulation context, system design might correspond to building and validating a functional model of an existing real-world system or a prospective new facility, process, or product. Parameter design is appropriate for attempting to "optimize" or "improve" performance of the simulation model by judiciously selecting settings for some of the decision factors in the model. Once these settings have been selected, a tolerance design experiment could be performed to provide further insight into the nature of the relationship between the noise factors, their interactions, and the response.

It might initially appear that the experiment design and analysis methods could be transferred directly to simulation environment. Product designers have performed parameter design experiments using computer models (CAD/CAM tools) in place of physical prototypes because of cost considerations, particularly in the semiconductor industry (Sacks et al. 1989). However, the concept of robust design presents both new challenges and new opportunities for many simulation analysts unfamiliar with the robust design philosophy. The use of a loss function, which incorporates the response variability as well as the response mean, challenges simulation analysts to view variability as an important characteristic of the system, rather than solely as a nuisance factor which complicates the comparison of mean responses. It is also an opportunity—it has the potential for improving the decision-making following experimentation on a simulated system.

Another opportunity arises because discrete event simulation output is typically a response stream rather than a single number. Thus the *system* variability for any specified combination of parameter settings can be estimated from a single run, using methods such as variance estimates computed from overlapping batches (Ramberg et al. 1991, Sanchez et al. 1992) or standardized time series (Schruben 1982, 1983). (Note that we need an estimate of the variance of an individual response, rather than an estimated variance of the mean of output stream.) These provide alternatives to the run-oriented variance estimation methods, such as replicating the entire experiment, or pooling effects associated with high-order interaction terms (which are deemed unimportant) to estimate error sums of squares. The former may be extremely costly for complex simulation experiments, while the latter may result in a misspecified model if lack-of-fit tests are not possible because of confounding. The usefulness of estimating system variability from a single run is even more apparent in the Taguchi context, where heterogeneity of variance across design points plays a crucial role in the final design selection.

Finally, additional opportunities arise from the increased control the analyst has of the experimental environment. In simulation, everything is controllable. This means that the classification of factors into parameters and noise factors must be based on controllability in the real-world, not the simulation environment. There is also another layer of controllability. For example, in addition to the parameter and noise factor settings, the analyst controls the initial state of the system (e.g., empty or capicitated queue), the warm-up period (truncation point), termination conditions (run duration), and random number stream(s) (seed, antithetic switch). We call all simulation-specific variables such as these the class of *artificial factors*.

2 RESPONSE SURFACE METAMODELS

We now consider response surface metamodels and their relation to the three factor classes: parameters, noise factors, and artificial factors. If a designed plan for the artificial factors is used, this complicates the overall design. This additional complexity is only justified if it improves the metamodeling capability.

Consider first a response surface metamodel in which the response (perhaps after suitable transformation) is a function only of the parameters with additive error. Mathematically, we have

$$\mathbf{Y} = f_1(\{X_i\}) + \epsilon \qquad (1)$$

where $\mathbf{Y}$ is the response vector, the $\{X_i\}$ are the independent variables, and ϵ is the error vector with mean zero. In response surface analysis, several assumptions are typically made. The errors are assumed to be iid normal random variables, a common error variance σ^2 is postulated, and $f_1(\{X_i\})$ is a linear function, i.e.,

$$f_1(\{X_i\}) = \mathbf{X}\beta$$

where the columns of $\mathbf{X}$ represent the model terms (including the intercept and any polynomial or interaction terms) and β is a column vector of the associated (constant) coefficients. Natural estimators of β and σ^2 are the least-squares estimators $\hat{\beta}$ and mean squared error (MSE), so the metamodel of the response mean at a point $\mathbf{x} = \{X_i\}$ is

$$\mathbf{Y_x} = \mathbf{X}\hat{\beta}$$

and the MSE provides the metamodel of the response variance for all parameter settings. Response surface methodology is often concerned with identifying the parameter values which lead to the 'optimal' response. In most applications, optimality means either the highest or lowest value of Y, thus we seek to find the x leading to high (low) values of $\hat{\mu}_{Y_x}$. Using Taguchi's terminology for the model of equation (1), the $\{X_i\}$ represent the parameters under investigation. No (controlled) noise factors exist.

Different metamodels result from a parameter design experiment, where some or all noise factor levels are controlled by design, rather than allowed to vary randomly. The underlying model is

$$\mathbf{Y} = f_2(\{X_i\}, \{W_j\}) + \xi \qquad (2)$$

where the $\{W_j\}$ are the controlled noise factors and the $\{\xi\}$ are independent random variables with mean zero. (Note that the addition of the $\{W_j\}$ has affected the error term.)

Two points merit further discussion. First, in the ideal situation where *all* sources of noise have been enumerated, $Var(\xi) = 0$ and the model becomes completely deterministic. While this is extremely unlikely in a real-world application, it may apply to certain situations in discrete event simulation (e.g., tolerance design example in Ramberg et al. 1991.). However, if *any* important noise factors have been included in the set $\{W_j\}$, then the variance of ξ in equation (2) is less than the variance of ϵ in equation (1). The second point is that we are *not* interested in assuming additive models when the noise factors are controlled. Interactions between the noise factors and the parameters lead to unequal error variances across parameter configurations. It is precisely this set of interactions which we hope to exploit in order to find robust parameter settings.

The analysis is also slightly different. Since the analyst is interested in a metamodel which contains only the parameters, the results can be averaged across the noise design before a metamodel is fit to the data. To illustrate, consider a simple example with a single parameter (X_1) and a single noise factor (W_1) controlled at one of w levels. For each design point x, we average across the noise factor space, e.g.,

$$\hat{\mu}_{Y_x} \approx \overline{Y}_x = \frac{1}{w} \sum_{k=1}^{w} \hat{f}_2(X_1, W_{1k}).$$

An estimate of the response variance at x can also be obtained:

$$s^2_{Y_x} = \frac{1}{w-1} \sum_{k=1}^{w} \left(\hat{f}_2(X_1, W_{1k}) - \overline{Y}_x \right)^2.$$

Regression can then be used to estimate linear metamodels for $\overline{Y}$ and $\log(s^2)$. (The logarithmic transformation is a variance stabilizing transformation.)

In parameter design problems involving physical experimentation, Taguchi uses saturated or nearly saturated orthogonal designs for the noise factors. The use of saturated designs may reduce the amount of experimentation required. Each noise factor that can be controlled during the experiment leaves less randomness to deal with. In this context, the ideal situation would be *no* remaining randomness associated with product performance. However, in the simulation environment, it may not be possible to remove all randomness from the system without changing it in a qualitative sense.

To clarify this distinction, we present a simple queueing example. The parameters are the number of servers and the service discipline (FIFO, LIFO). The noise factors might be the interarrival times of the customers and the service times, both random variables. Although it is typically impossible to control these noise factors in the real world setting, it would be possible to control these noise factors by removing all their variability while running a queueing system simulation. This would be extremely poor practice from a modeling perspective. It is a well-known fact that the behavior of queueing systems with deterministic interarrival and service times is radically different than when the inputs are stochastic. If we are interested in designing a robust queueing system, it may be appropriate to vary the nominal arrival rate and service rate (e.g., 'low' and 'high') and consider these as the noise factors.

In this scenario, it appears that (unlike Taguchi's strategy for product design using physical experimentation) the ideal situation is *not* one in which all randomness has been removed. However, the apparent remaining randomness is deceptive: it results from the use of pseudo-random number streams. Once a simulation run is fully specified (parameter settings, noise factor settings, initial conditions, run length, and random number streams) there is no more randomness in the system. The response can then be modeled as

$$\mathbf{Y} = f_3(\{X_i\}, \{W_j\}, \{A_k\}) + \zeta \qquad (3)$$

where $\{A_k\}$ denotes the class of artificial factors. As we shall show, the $\{A_k\}$ can be treated as a secondary category of noise factors. If we modify the data collection and analysis appropriately, we may be able to further improve the precision of the estimated coefficients.

The qualitative appeal for including artificial factors in the experiment design is readily apparent. Orthogonal experiment designs (a subset of which Taguchi popularized as "orthogonal arrays") have been around for many years. Taguchi realized dramatic gains in variability reduction by explicitly including noise factors that are varied systematically during physical experimentation. Since the noise factors are chosen to *span* the noise space, the analyst can gauge the sensitivity of the response to fluctuations in these factors. A designed noise plan, rather than simple random sampling, keeps data requirements from inflating too rapidly, particularly if saturated or near-saturated noise plans are used. We advocate an analogous step for simulation experiments: we will impose control over some artificial factors by specifying a designed data collection plan. This offers potential benefits in metamodel identification, particularly if the response is very sensitive to the random number stream(s) selected or if time and budget constraints prohibit extensive ad hoc sampling over the artificial factors.

3 DATA COLLECTION AND ANALYSIS

The impact on manufacturing of Taguchi's approach to variability reduction has been widely acknowledged. However, his implementation of some statistical techniques has sparked controversy in the statistical community (Kackar 1985, Box 1988, Pignatiello and Ramberg 1985, 1991, Nair 1992). While we follow Taguchi's philosophical approach in this paper, we recommend the well established statistical analysis methods presented by Sanchez et al. (1993) (see also Box, Hunter and Hunter 1978, Vining and Myers 1990, Ramberg et al. 1992). These methods utilize response surface methodology and modern graphical and data-analytic techniques for improved model identification.

Factorial and fractional factorial designs are often used for data collection. Two-level or three-level designs are selected for each factor based on the anticipated role for first-order and second-order terms in the metamodel. Parameter levels can be set to cover the region of interest. A noise factor with mean μ_N and variance σ_N^2 can be sampled equally at $\mu_n \pm \sigma_N$ for two-level plans, which results in a two-point distribution with mean μ_N and variance σ_N^2. For three-level factors, equal sampling can be performed at $\{\mu_N, \mu_N \pm \sigma_N\}$.

The parameter plan and noise plan (which Taguchi calls inner arrays and outer arrays) can be crossed to obtain an overall experiment plan. However, the number of runs required by such a crossed plan may be prohibitively large. One alternative is to fractionate the parameter and/or noise plans in order to economize on observations. Fractional factorials can be used to reduce the size of the parameter plan if high-order interactions are thought to be unimportant. In such cases, a center point can be added to the design to assess lack-of-fit. The noise factor plan will generally be more highly fractionated because noise×noise interactions are not of direct interest. For example, an experiment involving 2 parameters and 7 noise factors could be conducted as a 2^2 full factorial crossed with a 2^7 factorial, requiring a total of 128 runs for each of the four parameter configurations, or 512 runs in total. However, if a saturated eight-run design is used for the noise factors, the total number of runs required is 32. Alternatively, a combined plan can be used to accomodate both types of factors (parameters and noise). For the above example, a total of 9 factors are present so potential plans include a 2^9 full factorial, a 2^{9-1} half-fraction, etc., depending on the degree and nature of interactions to be modelled. Combined plans may be more economical than crossed plans unless the cost of a control run is much larger than the cost of a noise run (Shoemaker, Tsui and Wu 1991).

We remark that "replication" has two different meanings in the simulation setting. The output stream from any particular run can be reproduced exactly if identical inputs (parameter and noise factor settings, artifical factor settings) are used. In the robust design terminology, this means that replication error is zero. However, simulation analysts also use the term replication to refer to a single run of the simulation. Multiple replications for a particular set of parameter and controlled noise factors are made by changing some or all of the random number seeds.

A loss function should also be specified. For our example in the next section, we assume the goal of the parameter design process is to make the response as close as possible to a target value τ, and we use a quadratic loss function. The loss associated with a particular observation Y_x (taken at parameter configuration x) is

$$\ell_\tau(Y_x) = c\left[(Y_x - \tau)^2\right],$$

where c is a constant which converts the loss to monetary units. The expected loss, or risk, associated with parameter configuration x is then obtained by taking the expectation of the loss over the noise space. Mathematically,

$$\begin{aligned} R(x) &= E\left[\ell_\tau(Y_x)\right] \qquad (4) \\ &= c\left[\sigma_{Y_x}^2 + (\mu_{Y_x} - \tau)^2\right]. \end{aligned}$$

We refer to $R(x)/c$ as the *scaled loss* associated with parameter configuration x. This scaled loss can be used to compare alternatives when the response means cannot all be adjusted to the target value.

4 VARIANCE REALLOCATION

We now consider one type of artificial factor—the random number stream— and indicate how it can be incorporated into a robust design experiment. Consider a generic discrete-event simulation. One or more seeds are transformed algorithmically into sequences (streams) of (assumed) i. i. d. Uniform pseudo-random numbers. These streams are then transformed into sample paths for stochastic processes. Let $R_i = \{R_{ij}, j = 1, \ldots\}$ denote a set of such streams sufficient to execute a run of the simulation (i is a countable infinite index for the seed). Then let R_i^a denote the antithetic set for R_i, where $R_{ij}^a = 1 - R_{ij}$ The use of common and/or antithetic random number streams has received a great deal of attention in the simulation literature as a means of improving the estimation of a mean response or mean differences across alternative systems. (See Law and Kelton 1991 for an overview.)

We now illustrate the use of common and/or antithetic random number streams for robust design in simulation. Consider one independent variable X corresponding to a response Y and the first order model

$$Y_X = \beta_0 + \beta_1 X + \epsilon, \qquad (5)$$

where $\epsilon \sim N(0, \sigma^2)$. Suppose we are considering two alternative parameter settings, x and $-x$, and we collect $n/2$ observations at each of the two points.

If all independent random number streams are used, then the variances of the $\hat{\beta}$ in equation (5) are given by

$$\begin{aligned} Var(\hat{\beta}_0) &= \frac{2}{n} Var(Y_{-x} + Y_x) = \frac{\sigma^2}{n} \\ Var(\hat{\beta}_1) &= \frac{2}{n} Var(Y_{-x} - Y_x) = \frac{\sigma^2}{n} \end{aligned}$$

These variances change if common or antithetic random number streams are used. Following the notation of Schruben and Margolin (1979), let ρ^+ denote the correlation induced by common streams for different systems and $-\rho^-$ denote the correlation induced by antithetic streams. (The magnitudes of the induced correlation are unknown, and model dependent, although typically $|\rho^-| < |\rho^+|$.) Then if common random numbers (i.e., streams $\{R_i, i = 1 \ldots, n/2\}$) are used for both $-x$ and x, we find

$$\begin{aligned} Var(\hat{\beta}_0) &= \frac{\sigma^2}{n}(1 + \rho^+) \\ Var(\hat{\beta}_1) &= \frac{\sigma^2}{n}(1 - \rho^+). \end{aligned}$$

An alternative is to use antithetic sampling: streams $\{R_i, i = 1, \ldots, n/2\}$ are used for $-x$, and streams

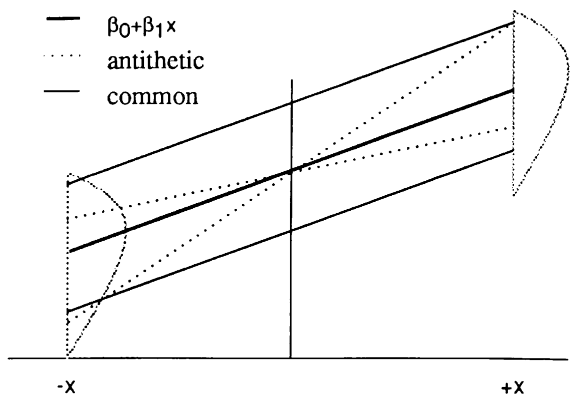

Figure 1: Comparison of Common and Antithetic Sampling for a First Order Linear Model

$\{R_i^a, i = 1, \ldots, n/2\}$ are used for x. This yields

$$\begin{aligned} Var(\hat{\beta}_0) &= \frac{\sigma^2}{n}(1 - \rho^-) \\ Var(\hat{\beta}_1) &= \frac{\sigma^2}{n}(1 + \rho^-). \end{aligned}$$

This means it is not possible to reduce the variability of both coefficient estimates simultaneously. If common random numbers are used, we have a better estimate of the slope but a poorer estimate of the mean. If antithetic random variates are used, then we have a better estimate of the mean response but a poorer estimate of the slope. This relationship is illustrated in Figure 1 for the model of equation (5). When common random numbers are used for an observation at each of $+x$ and $-x$, the responses are either both above or both below the true mean response. Since the error is additive and the error distribution is identical at the two points, the slope can be estimated exactly from such a pair of responses: the solid thin lines (joining pairs with common random numbers) are parallel to the underlying response (shown as a solid thick line). However, there is variability associated with the intercept estimate, particularly if the total sample size n is small. Conversely, antithetic random numbers pair positive errors at $-x$ with negative errors at $+x$ and vice versa. The dashed lines joining antithetic paired response have slopes that differ from the underlying model, but the intercept is much closer to the true value. (If the error distribution is symmetric, the intercept can be estimated exactly.)

Now consider two factors X_1 and X_2, and the corresponding first order model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \qquad (6)$$

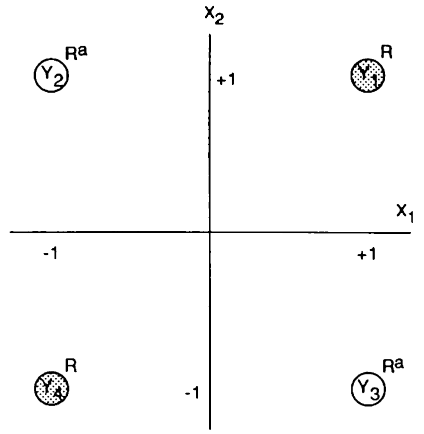

Figure 2: Common/Antithetic Sampling Strategy for Two Factor Model

where $\epsilon \sim N(0, \sigma^2)$. Suppose we wish to make a total of n runs covering 4 design points: $(-x_1, -x_2)$, $(-x_1, x_2)$, $(x_1, -x_2)$ and (x_1, x_2). One way of designing the experiment is to break the design into two half-fractions: one half-fraction uses the streams R_i $(i = 1, \ldots, n/4)$ and the other uses the the antithetic set R_i^a $(i = 1, \ldots, n/4)$. This correlation induction strategy, which is shown graphically for the points $Y_1, \ldots, Y_4$ in Figure 2, yields coefficient estimates with the following variances:

$$\begin{aligned} Var(\hat{\beta}_0) &= Var\left(\frac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4)\right) \\ &= \frac{\sigma^2}{4}(1 + \rho^+ - 2\rho^-) \\ Var(\hat{\beta}_1) &= Var\left(\frac{1}{4}(Y_1 - Y_2 + Y_3 - Y_4)\right) \\ &= \frac{\sigma^2}{4}(1 - \rho^+) \\ Var(\hat{\beta}_2) &= Var\left(\frac{1}{4}(Y_1 + Y_2 - Y_3 - Y_4)\right) \\ &= \frac{\sigma^2}{4}(1 - \rho^+) \end{aligned}$$

If $2\rho^- > \rho^+$ then the variance of *all* estimated coefficients has been reduced. However, the reduction is at the expense of *reallocation* (Schruben and Margolin 1978, Schruben 1979, see also Tew and Wilson 1991): the estimated variance of the coefficient of the interaction term which we chose not to include in the model has increased. In the full 2^2 factorial analysis,

$$\begin{aligned} Var(\hat{\beta}_{12}) &= Var\left(\frac{1}{4}(Y_1 - Y_2 - Y_3 + Y_4)\right) \\ &= \frac{\sigma^2}{4}(1 + \rho^+ + 2\rho^-). \end{aligned}$$

Schruben (1979) proved that the sum of coefficient variances is constant for *any* saturated, orthogonal design.

This reallocation has interesting implications for parameter design. Consider the crossed data collection plan (parameter plan × noise plan × artificial factor plan). The artificial factor plan should be chosen in order to induce correlations which reallocate variance from the interesting terms (parameters) to the uninteresting terms (noise factors). Since we will average the results over the noise space, obtaining precise estimates of the coefficients corresponding to the noise factors is unnecessary. Variance reallocation can also be conducted when a combined data collection plan is used.

The linear models in equations (5) and (6) are analytically tractable, and motivate the use of antithetic random number streams in general. However, they are not of particular interest in the robust design framework since the additive error has the same distribution at all parameter settings. As we illustrate in the next section, when the $\{W_j\}$ interact with the parameters this leads to nonhomogeneous variance across design points. Least-squares estimators of the βs in equations (5) and (6) may still be estimated more precisely if antithetic random number streams are used, but the response variances should be estimated separately at each design point from the *stream* of output available. Recall that the Y's analyzed in the response surface models above are themselves likely to be averages, (such as the average waiting time for a customer in a queueing system) for particular parameter configurations.

5 EXAMPLE

We illustrate the use of antithetic random variates for a simple, hypothetical system. We have a single parameter, X, which we will sample at one of three values: low ($X = 2$), middle ($X = 4$), and high ($X = 6$)). Our noise factor is the range of a Uniform[-c,c] random variable and represents the variability inherent in the system. Suppose we are uncertain of the range, but feel that it is equally likely to be any value between 0.5 and 3.5 (e.g., c equally likely between 0.25 and 1.75).

The response surface we model is given in equation (7) and shown in shown in Figure 3.

$$Y = \begin{cases} 100 - 50x & \text{if } x \leq 2 \\ (x-2)^2 & \text{if } x \geq 2 \end{cases} \tag{7}$$

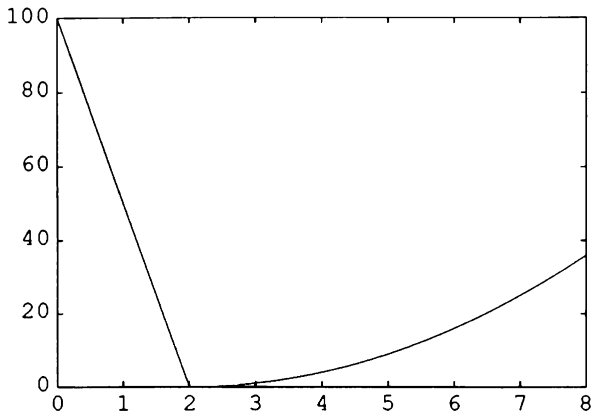

Figure 3: Response Surface for the Example

Note that the only source of variability is the uniform noise factor. Its variability is transmitted differentially to variability in the response, depending on the shape of the response surface near the value of X. For every one of the three X values, the response variability increases (though not necessarily by the same proportion) as c increases.

Although this is a hypothetical scenario, we stress that the classification of parameters, noise factors, and artificial factors cannot be done in a vacuum. In this example, we assume that all variability in W cannot be removed without qualitatively changing the system. For instance, we would not want to change interarrival times from stochastic to deterministic in a queueing application, but we might be unsure about the exact parameterization of the underlying distribution.

We sample from this system under three sets of conditions, using a common total sample size of 24 runs. First, we sample as though the parameter X is the sole factor of interest, and generate 8 samples (using independent random number streams) for each of the three design points. This is done in a two-stage process: a value of c is first generated from a Uniform[0.25,1.75] distribution, and then the noise W is generated from a Uniform$[-c,c]$ distribution. This value $X + W$ is then transformed according to equation (7) to yield a realization of Y. This corresponds to the model of equation (1), where we are only modeling the effect of the parameter.

Second, we consider both the parameter and noise factor, and select different random number seeds for all 24 runs: 4 runs for each of the six combinations of parameter and noise factor levels. This corresponds to the model of equation (2), which incorporates the noise factor into the design. The results obtained are averaged across the noise design to yield estimates of the mean response.

Finally, we combine the use of antithetic and common random number streams as discussed in the previous section (replicated twice): two pairs of antithetic streams are used for each of the six design points. Resulting values of Y are averaged over the noise factors and artifical factors for each design point to yield estimates of the mean response. This corresponds to a model of the form in equation (3), with parameters, noise factors, and artificial factors included.

The results for all three experiments are summarized in Table 1. Recall that our primary goal is the identification of robust design points. In our example, this corresponds to determining values of X for which the response is relatively insensitive to noise variation, while the mean response is close to a target value τ. The risk $R(x)$ of equation (4) (or the scaled loss $R(x)/c$) can be used to rank the alternatives. Which is better? That depends on the target value. Suppose we wish to make the response as small as possible ($\tau = 0$). From figure 3 it would appear that $X = 2$ yields the optimal design. However, for this design the response variability is extremely high. From Table 1, the loss associated with $x = 4$ (L_4), is an order of magnitude lower than L_2 and L_6. This indicates that $x = 4$ is clearly the best design in terms of overall performance.

The regression coefficients and associated standard errors provided in Table 1 indicate that the control of the random number stream has increased the predictive ability of the regression equation, and decreased the standard errors of all coefficients, despite the decrease in the degrees of freedom for error estimation. (24 independent observations are used to fit the first model; pairs of independent observations are averaged for low and high ranges to obtain 12 points for fitting the second model; the third model is fit using 6 points, obtained by averaging across low and high ranges and antithetic pairs.) This illustrates the benefit of using correlation induction strategies on the artificial factors.

6 CONCLUSIONS

The robust design approach mandates a new outlook on the response surface metamodeling problem. While ordinary least-squares regression and ANOVA techniques assume equal variance across design points, the success of Taguchi's robust design philosophy indicates the need for methods which incorporate variance heterogeneity into the analysis.

Table 1: Experiment Results: Metamodel is $\mu_{Y_x} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2$

	$Y = f(X)$			$Y = f(X, W)$			$Y = f(X, W, A)$		
Data	$x = 2$	$x = 4$	$x = 6$	$x = 2$	$x = 4$	$x = 6$	$x = 2$	$x = 4$	$x = 6$
$\overline{Y}_x$	16.78	4.12	14.20	27.63	3.92	14.91	11.12	4.25	16.29
s_{Y_x}	20.48	4.58	5.61	30.66	2.77	5.61	16.64	2.38	4.64
$\hat{L}_s$	701.0	38.0	233.1	1703.5	23.0	257.8	400.5	23.73	286.9
Metamodel	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
Fitted Value	4.199	-1.292	11.290	3.922	-6.361	17.352	4.294	2.586	9.414
Std. Error	(4.434)	(3.135)	(5.430)	(4.695)	(3.320)	(5.750)	(3.418)	(2.417)	(4.186)
R^2	0.1762			0.5867			0.6739		

One reason that robust design has only recently been discussed in the statistical literature is that primary attention has been given to transforming data until the standard assumptions are met—thus, heterogeneity of variance is considered a nuisance which complicates the ability to compare alternatives based on their means. This view has carried over in simulation as well although it is well-recognized that complex systems may have different variability at different model configurations (Law and Kelton 1991). We reiterate that variance estimation is *not* a nuisance for robust design identification. In fact, for many systems the differences in variability may overwhelm any differences in the mean responses.

Although the use of antithetic random variates holds promise for robust design experiments in the simulation setting, there are still some problems which need to be addressed. First, the extent to which this approach will improve the decision-making process for higher-order models is not readily apparent. The use of antithetic streams for quadratic models is a nontrivial problem even for traditional response-surface metamodeling (Tew 1989), although identifying good designs for higher-order metamodels is of special interest (Sargent 1991).

However, certain aspects of this variance reallocation problem are worth mentioning again. Every time a factor can be controlled in an experimental design, rather than sampled randomly, this decreases the underlying variance associated with all remaining sources of noise. As long as the noise factors chosen to be controlled do transmit variance to the response, the precision associated with estimating the parameter coefficients should improve.

ACKNOWLEDGMENTS

Susan M. Sanchez was supported in part by Grant DDM 9110573 from the National Science Foundation.

REFERENCES

Box, G. E. P. 1988. Signal-to-noise ratios, performance criteria, and transformations (with discussions), *Technometrics* 30:1–40.

Box, G. E. P., W. G. Hunter and J. S. Hunter. 1978. *Statistics for Experimenters*. New York: John Wiley and Sons, Inc.

Kackar, R. N. 1985. Off-line quality control, parameter design, and the Taguchi method, *Journal of Quality Technology* 17: 176–188.

Kelton, W. D. 1988. Designing computer simulation experiments, *1988 Winter Simulation Conference Proceedings*, ed. M. A. Abrams, P. L. Haigh and J. C. Comfort, 15–18. Institute of Electrical and Electronics Engineers, San Diego, California.

Law, A. M. and W. D. Kelton. 1991. *Simulation Modeling and Analysis*. 2nd ed. New York: McGraw-Hill.

Lucas, J. M. 1985. Comment on "Off-line quality control, parameter design, and the Taguchi method," by R. N. Kacker, *Journal of Quality Technology* 17: 195–197.

Nair, V. N. 1992. Taguchi's parameter design: a panel discussion, *Technometrics* 34: 127–161.

Pignatiello, J. J. Jr. and J. S. Ramberg. 1985. Discussion of Off-line quality control, parameter design, and the Taguchi method, *Journal of Quality Technology* 17: 198–206.

Pignatiello, J. J. Jr. and J. S. Ramberg. 1991. Top ten triumphs and tragedies of Genichi Taguchi, *Quality Engineering* 4: 211-225.

Ramberg, J. S., S. M. Sanchez, P. J. Sanchez and L. W. Hollick. 1991. Designing simulation experiments: Taguchi methods and response surface metamodels, *1991 Winter Simulation Conference Proceedings*, ed. B. Nelson, W. D. Kelton and G. M. Clark, 167–176. Institute of Electrical and Electronic Engineers, Phoenix, Arizona.

Ramberg, J. S., J. J. Pignatiello, Jr. and S. M. Sanchez. 1992. A critique and enhancement of the Taguchi method, *ASQC Quality Congress Transactions*, 491–498.

Sanchez, S. M., J. S. Ramberg, J. Fiero and J. J. Pignatiello, Jr. 1993. Quality By Design, forthcoming in *Concurrent Engineering*, ed. A. Kusiak, Chapter 10. New York: John Wiley and Sons, Inc.

Sanchez, S. M., P. J. Sanchez and J. S. Ramberg. 1992. Taguchi methods, response surface metamodels, and the design of simulation experiments. Technical report, College of Business and Public Administration, University of Arizona, Tucson, Arizona.

Sargent, R. G. 1991. Research issues in metamodeling, *1991 Winter Simulation Conference Proceedings*, ed. B. Nelson, W. D. Kelton and G. M. Clark, 887–893. Institute of Electrical and Electronic Engineers, Phoenix, Arizona.

Schruben, L. W. 1982. Detecting initialization bias in simulation output. *Operations Research* 30: 569–590.

Schruben, L. W. 1983.Confidence interval estimation using standardized Time Series. *Operations Research* 31: 1090–1108.

Schruben, L. W. 1979. Designing correlation induction strategies for simulation experiments, *Current Issues in Computer Simulation*, ed. N. R. Adam and A. Dogramaci, 235–256. New York: Academic Press.

Schruben, L. W. and B. H. Margolin. 1978. Pseudo-Random number assignment in statistically designed simulation and distribution sampling experiments, *Journal of the American Statistical Association*, 73, 504–525 (with discussions).

Shoemaker, A. C., K. L. Tsui and C. F. J. Wu. 1991. Economical experimentation methods for robust design, *Technometrics* 33: 415–427.

Taguchi, G. 1986. *Introduction to Quality Engineering*, White Plains, New York: UNIPUB/Krauss International.

Taguchi, G. and Y. Wu. 1980. *Introduction to off-line quality control,* Nagoya, Japan: Central Japan Quality Association.

Tew, J. D. 1989. Correlation-induction techniques for fitting second-order metamodels in simulation experiments, *1989 Winter Simulation Conference Proceedings*, ed. E. A. MacNair, K. J. Musselman and P. Heidelberger, 538–546. Institute of Electrical and Electronics Engineers, Washington, D. C.

Tew, J. D. and J. R. Wilson. 1991. Validation of simulation analysis methods for the Schruben-Margolin correlation-induction Strategy. *Operations Research* 40: 87–103.

Vining, G. G. and R. H. Myers. 1990. Combining Taguchi and response surface philosophies: a dual response approach. *Journal of Quality Technology* 22: 38–45.

AUTHOR BIOGRAPHIES

LEE W. SCHRUBEN is a Professor in the School of Operations Research and Industrial Engineering at Cornell University and currently the Visiting Distinguished Professor at SEMATECH. His research interests are in the design and analysis of large scale simulation experiments. He is a principal developer of the SIGMA simulation system. Three of his papers have received outstanding publications awards from the TIMS College on Simulation and the Chemical Division of the American Society of Quality Control.

SUSAN M. SANCHEZ is on the faculty of the Management and Policy Department at the University of Arizona. Beginning January 1993, she will be an Associate Professor of Management Science at the University of Missouri–St. Louis. Her research interests include applied statistics, selection, experimental design, and quality applications in manufacturing, health care, and simulation. She is a member of ORSA, TIMS, ASA, and ASQC.

PAUL J. SANCHEZ is on the faculty of the Systems and Industrial Engineering Department at The University of Arizona. Beginning January 1993, he will be starting a software design and consulting business. His research interests include various aspects of simulation. He is a member of TIMS, and was the local arrangements chair for the 1991 Winter Simulation Conference.

VERONICA A. CZITROM is currently on a two year assignment to SEMATECH from AT&T Bell Laboratories. She has published articles on robust design, mixture experiments, and estimation, is coauthor of four engineering textbooks, and is currently editing a book of statistical case studies in the semiconductor industry.