Simulation Modeling and Statistical Computing

Robert G. Sargent Editor

An Experimental Procedure for Simulation Response Surface Model Identification

LEE W. SCHRUBEN and V. JAMES COGLIANO

ABSTRACT: An experimental method for identifying an appropriate model for a simulation response surface is presented. This technique can be used for globally identifying those factors in a simulation that have a significant influence on the output. The experiments are run in the frequency domain. A simulation model is run with input factors that oscillate at different frequencies during a run. *The functional form of a response surface model for the simulation is indicated by the frequency spectrum of the output process. The statistical significance of each term in a prospective response surface model can be measured. Conditions are given for which the frequency domain approach is equivalent to ranking terms in a response surface model by their correlation with the output. Frequency domain simulation experiments typically will require many fewer computer runs than conventional run-oriented simulation experiments.*

1. INTRODUCTION TO FREQUENCY DOMAIN EXPERIMENTS

A feature of simulated systems is that one has complete control over the input factors. Frequency domain experiments exploit this feature. In a frequency domain experiment, values for continuous factors are varied *during a run* according to sinusoidal oscillations. Different frequencies are assigned to each factor. If the simulation response is sensitive to changes in a particular factor, then oscillating that factor induces predictable oscillations in the response. Varying the values of unimportant factors does not alter the response. The spectrum of the simulation output process measures the relative strength of these oscillations. Spectral analysis of the series of simulation output observations permits one to identify an appropriate polynomial model for the simulation output as a function of the input levels over some experimental region. Such a model is called a simulation response surface regression model or metamodel. Response surface model identification using frequency domain methods will typically require far fewer runs of the simulation model than would be necessary in a conventional run-oriented experimental approach.

In this paper no distinction is made between simulation model parameters and factors; both can be treated in much the same manner in simulation experiments. All of the inputs to a simulation are generically referred to as input factors.

Simulation response surface models are high-level mathematical relationships that aid in a global understanding of the complex relationships between the inputs and the output of a simulation program. By far the most common simulation response surface model is a polynomial that expresses the expected output as a weighted sum of products of powers of the input levels. The value of such response surface models is well established (e.g., see [8] and [9]).

A practical use of response surface models is to guide in the design of conventional run-oriented simulation experiments by screening out insignificant input factors or identifying important interactions between factors. Input factors that are not in any significant terms might be excluded from further experiments. The designs chosen for these conventional experiments should not confound the estimators of coefficients of significant terms. For example, if a particular product of factors is shown to be significant in the response model, it would not be advisable to employ a fractional factorial design with this product as the defining interaction. A response surface model can also help focus data collection efforts on the estimation of input parameters and guide in the application of variance reduction techniques such as correlation induction [16]. A direct application of the experimental technique presented in this paper is *global* simulation sensitivity analysis. This

This research was supported in part by a grant from AT&T Bell Information Systems to Cornell University and by the Air Force Systems Command, Rome Air Development Center, Griffiss Air Force Base, Rome, NY through a subcontract to Cornell University from Syracuse University.

technique complements available methods for *local* simulation sensitivity analysis via response gradient estimation (e.g., see the heuristic described in [17]).

There are two basic tasks in simulation response surface modeling: First, the functional form of an adequate response surface model must be identified. Second, the coefficients of the response surface model must be estimated. This paper focuses on the identification of the appropriate functional form of a response surface model. Only after a response surface model is identified can an experiment be properly designed to estimate the values of its coefficients. As often as not, the experimenter must rely on both extensive experimentation and intuition in selecting an adequate response surface model.

This paper considers the situation where the simulation output is a scalar time series and the controllable input factors are continuous. Discrete valued input factors can be included in frequency domain simulation experiments by randomizing over their values as described in [14]. The remainder of this paper is organized as follows: The response surface model identification problem is formulated and theoretical support for frequency domain simulation experiments is presented in Section 2. In Sections 3 and 4, the design of fequency domain experiments in a simulation program is motivated and discussed. A method for statistically analyzing such experiments is given in Section 5. In Section 6 an explicit, easily implemented procedure for frequency domain response surface model identification is presented. Section 7 contains several examples that illustrate the application of the method. The final section contains some conclusions and suggestions for running frequency domain simulation experiments.

2. PROBLEM STATEMENT

Consider a simulation program with p continuous input factors, $x_1, x_2, \ldots, x_p$. We wish to identify significant terms in a polynomial model of the expected response, $E[y]$, as a function of the xs. Specifically, we consider as *prospective response surface models* the class of all k-order polynomials given by

$$E[y] = \beta_0 + \sum_{j=1}^{q} \beta_j \tau_j.$$

Here,

- $E[y]$ is the expected simulation response;
- τ_j is a term in the k-order polynomial, that is, a particular product of nonnegative integer powers of the input factors, where the sum of the exponents is not greater than k (e.g., if $k = 7$, $x_1^2 x_2^3$ is such a term);
- β_j is the coefficient for term τ_j; and
- q is the number of potential terms in the prospective model.

Each term in this polynomial is a candidate for inclusion in a simulation response surface model. Our objective is to select the set of terms (τ_js) that have coefficients (β_js) that are significantly different from zero. We assume that the experimenter knows nothing about the response surface except an *upper bound* on the order, k, of the approximating polynomial. A check for lack of fit of any model in the class of k-order polynomials is available and is discussed in Section 5.

2.1 Qualitative Analysis of Frequency Domain Experiments

Before discussing frequency domain simulation experiments in detail, the concept is illustrated by qualitatively examining a "black box" system. The simulation output sequence at time t, $y(t)$, is a function of four input processes, $x_1(t)$, $x_2(t)$, $x_3(t)$, and $x_4(t)$, and a stochastic input process, $\varepsilon(t)$. Suppose, unknown to the experimenter, the response is given by

$$\begin{aligned} y(t) = {} & x_1(t-5) + x_1^2(t-5) \\ & + \tfrac{1}{3}(x_2(t-1) + x_2(t-2) + x_2(t-3)) \\ & + \tfrac{1}{3}(x_2(t-1)x_3(t-1) \\ & + x_2(t-2)x_3(t-2) + x_2(t-3)x_3(t-3)) \\ & + \varepsilon(t). \end{aligned}$$

Here $\varepsilon(t)$ is the first-order autoregressive noise process given by $\varepsilon(t) = .6\varepsilon(t-1) + .8e(t)$ with $e(t)$ independently sampled from a normal distribution with zero mean and unit variance. The response contains five-period lags of x_1 and x_1^2, three-period moving averages of x_2 and x_2x_3, and autoregressive noise. The response is independent of x_4. Time lags in the response are reflected as phase shifts in the frequency domain. These lags should have no effect since the frequency spectrum is insensitive to phase shifts. A response surface model identification procedure should indicate that x_1, x_1^2, x_2, and x_2x_3 are the only important terms in the response surface model.

A frequency domain response surface model identification experiment requires at least two runs of the simulation, a *control run* and a *signal run*. For this example, the run lengths (denoted by n) were 300 observations of $y(t)$.

A control run is just a conventional simulation run with the input factors held constant at their nominal values. The control run identifies natural cycles in the response. The 300 control-run observations were used to estimate the output frequency spectrum. The spectrum estimator for the control run is the *control spectrum* and is denoted $\hat{f}^c(\omega)$.

In a signal run, the input factors vary according to sinusoidal oscillations during the run. The frequency assigned to a particular input factor in a signal run is called its *driving frequency*. In this example the four input factors were oscillated at frequencies of .06, .20, .29, and .39 cycles per time unit, respectively. This set of driving frequencies distinguishes all potential terms in the prospective second-order response surface model (driving frequency selection is discussed in Section 4).

The values of the input processes at time t are given by $x_1(t) = \cos(2\pi(.06)t)$, $x_2(t) = \cos(2\pi(.20)t)$, $x_3(t) = \cos(2\pi(.29)t)$, and $x_4(t) = \cos(2\pi(.39)t)$. The spectrum estimator for the signal run is the *signal spectrum* denoted as $\hat{f}^s(\omega)$.

The *spectral signal-to-noise ratio*, $R(\omega) = \hat{f}^s(\omega)/\hat{f}^c(\omega)$, is the basis for the analysis of the experiments described in this paper. Figure 1 is a plot of the log of $R(\omega)$. There is a peak at frequency .06, which is the oscillation frequency of the factor x_1. This peak indicates that the response is, in part, a linear function of x_1. Similarly, the peak at frequency .20 indicates the presence of a

Spectral ratio	Frequency	
1.282131	0.010	**
0.769167	0.020	*
1.540611	0.030	**
1.932238	0.040	**
3.549101	0.050	****
22.632690	0.060	**********
1.840299	0.070	**
2.194231	0.080	***
28.159400	0.090	**********
6.448475	0.100	******
4.235514	0.110	****
54.441450	0.120	**********
2.674181	0.130	***
2.254318	0.140	***
0.954398	0.150	**
1.116599	0.160	**
1.543496	0.170	**
1.445948	0.180	**
1.922813	0.190	**
48.873300	0.200	**********
2.754482	0.210	***
0.931734	0.220	*
1.330020	0.230	**
1.266759	0.240	**
1.357243	0.250	**
1.127451	0.260	**
1.000037	0.270	*
1.104708	0.280	**
1.049220	0.290	*
1.175703	0.300	**
1.161654	0.310	**
1.140434	0.320	**
1.012885	0.330	*
1.034468	0.340	*
1.071303	0.350	*
0.997419	0.360	*
0.913147	0.370	*
1.026011	0.380	*
0.779825	0.390	*
0.843078	0.400	*
1.055776	0.410	*
1.062831	0.420	*
1.129120	0.430	**
1.092276	0.440	**
1.296505	0.450	**
1.236115	0.460	**
1.129132	0.470	**
5.080697	0.480	*****
32.294870	0.490	**********
1.081355	0.500	*

FIGURE 1. Spectral Ratio from the Model Identification Example

linear term for x_2. The peak at frequency .12, twice the oscillation frequency of x_1, indicates that a quadratic term in x_1 is present. (A rule for identifying the "indicator frequencies" for different terms in the response surface model is given in Section 4.) The remaining peaks are at frequencies .09 and .49 (the sum and difference of the oscillation frequencies of x_2 and x_3). These peaks indicate that the interaction term x_2x_3 is present. The true response surface model, containing the terms x_1, x_2, x_1^2, and x_2x_3, is easily identified.

An important practical point concerns the sampling of the input and output time series and their relationship. The index, t, was called "time" in the above example. This index could represent different quantities depending on the particular application. Frequencies are expressed in this paper in cycles per time unit; a more descriptive measure of frequency would be in cycles per output observation. For example, in a simulated queueing system where the output is some function of customer delays, an appropriate index would count the customers as they enter the system. The index, t, would then be the *customer arrival sequence number*. The input processes for the signal run would oscillate with this index, not, say, with simulated clock time. Frequency domain simulation experiments of a computer system have been successfully run indexing output series (after sorting) by the input job number (personal communication from R. G. Sargent, 1985). Specifically, for a computer network where job delays are measured, an input factor x_j might be a parameter of some CPU or memory unit processing time distribution. The nominal value for x_j is given as μ. For the signal run, the processing time for the tth job would be generated using a parameter value given by $\mu + \alpha_j \cos(2\pi\omega_j t)$, where ω_j is the driving frequency assigned to input factor x_j and α_j is its amplitude. Similarly, the time between job arrivals would be generated using a parameter that oscillates with the job arrival sequence number. Before computing the output sample spectrum, the series of observed job delay times must be ordered according to the job *arrival* sequence number. This is because in a complex network jobs may exit the system in a different order than that in which they entered the system. Keeping a common index for driving oscillations and output analysis is important.

Although it is not central to the ideas presented in this paper, it is worth noting that common pseudorandom number streams were used in the previous example for both the signal run and the control run. This results in a reduction in the variance of the estimated signal-to-noise ratio, $R(\omega)$. This variance reduction technique appears to be very helpful in frequency domain experiments.

2.2 Theoretical Requirements for Frequency Domain Experiments

Frequency spectra are defined only for stationary stochastic processes. Depending on how the simulation run is initialized and terminated, an assumption of stationarity may or may not be justifiable. For instance, there may be an initial transient early in a run if the initial conditions are far from typical. Fortunately, an initial transient in the output process does not appear to cause any practical problems with the methodology presented in this paper. In fact, the examples presented in Section 7 involve short runs of transient simulations (sometimes called "terminating" simulations in the literature) where no initial warm-up period is used. The frequency domain approach was quite effective in identifying an appropriate response surface model for these simulations. In the limited empirical studies done to date, frequency domain response model identification appears to be equally effective for both transient and stationary simulations. This phenomenon can be explained by noting that initial transients in a simulation output series appear in the frequency domain as partially completed cycles at very low frequencies. These cycles can be viewed as naturally occurring low-frequency components of a stationary stochastic process. In the time domain, this is equivalent to modeling a transient as a sum of small pieces of orthogonal sinusoids with very long periods. The analysis is done using spectral ratios where "true" spectra values (including those at low-frequency components due to transients) cancel out of statistics used to detect the significance of terms in the response surface model.

In order for the frequency domain approach to be used to identify a model for a system, the system must have

(1) parameter settings that can be changed during an experimental trial;
(2) a response that can be observed at periodic intervals; and
(3) a response that can be adequately modeled as a time-invariant linear combination of products of powers of the factors—that is, given p factors observed at equally spaced intervals

$$x_i(t) \quad \text{for} \quad i = 1, \ldots, p, \quad t = 1, 2, \ldots, n,$$

we can form q polynomial terms by taking products of powers of the factors

$$\tau_j(t) = \prod_{i=1}^{p} [x_i(t)]^{a_{ij}} \quad \text{for} \quad j = 1, \ldots, q$$

and a_{ij}, some nonnegative integer;

the response is modeled as a time-invariant linear combination (filter) of these terms and noise that is independent of the factors—that is,

$$y(t) = \sum_{j=1}^{q} \sum_{k=-\infty}^{\infty} h_{jk}\tau_j(t-k) + \varepsilon(t); \qquad (1)$$

the weights, $\{h_{jk}\}$, are often referred to as an impulse response function corresponding to the input $\tau_j(t)$ [10, p. 14].

We refer to Eq. (1) as the *dynamical* response surface model for the system. Such a system can be studied as a "black box" with two types of input factors: $x_i(t)$, which may change during an experimental trial, and random

noise, $\varepsilon(t)$. Whatever series of values are assigned to the factors, the rule for forming the response remains unchanged because the *system is deterministic*. A simulation computer program is an important special class of systems where the frequency domain approach is applicable. Simulation computer programs are almost always time invariant and deterministic; the computational rules do not change during execution of the program. These programs can be written to vary factor settings during a run and to periodically sample the response. If the response can be modeled as a polynomial function of the factors (at least in the region of interest), then all three requirements are met, and the frequency domain approach should identify a response model.

3. MOTIVATION FOR THE FREQUENCY DOMAIN APPROACH

The frequency domain approach to simulation sensitivity analysis has two major advantages: First, several input factors can be studied in the same run. Second, nonlinear effects, such as products of integer powers of the input factors, can be detected with no additional experimentation.

In frequency domain experiments, frequency bands rather than simulation runs are the experimental units. Each run of the simulation provides a large number of essentially independent frequency bands. The experiment is designed so that each term in a prospective response surface model is assigned to a distinct set of frequency bands; many factors can be studied at once. The advantage of this property can be demonstrated with a simple system for which the response is a linear function of one factor with added noise (a special case of Eq. (1)).

Denote the linear system by

$$y(t) = \sum_{k=-\infty}^{\infty} h_k x(t-k) + \varepsilon(t),$$

where $y(t)$, $x(t)$, h_k, and $\varepsilon(t)$ are as defined earlier. The functions $f_y(\omega)$ and $f_x(\omega)$ will denote the spectral functions of the output process $y(t)$ and the input process $x(t)$, respectively (see Section 3.2). The response spectrum $f_y(\omega)$ and the factor spectrum $f_x(\omega)$ are related by

$$f_y(\omega) = G^2(\omega) f_x(\omega) + f_\varepsilon(\omega), \tag{2}$$

where $G(\omega)$ is the *gain*. The gain function describes how the linear system amplifies or attenuates oscillations at different frequencies. The gain is zero and there is no oscillation-induced peak if the response does not depend on the factor. Suppose the factor oscillates with amplitude α and frequency ω; $x(t) = \alpha \cos(2\pi\omega t + \delta)$, where δ is the *phase shift*. The factor spectrum has a peak at ω independent of δ. The response spectrum, defined by Eq. (2), also has a peak at ω, scaled by the gain and masked by the noise. A response spectrum peak at ω is evidence that the factor affects the response; no peak is evidence that it does not. If $\varepsilon(t)$ is a white noise process, then $f_\varepsilon(\omega)$ is constant for all ω, and in theory no masking should occur. In practice, however, the sample spectrum computed from a particular realization of $\varepsilon(t)$ is not necessarily constant. In addition, $\varepsilon(t)$ may not be a white noise process and so may not have a constant spectrum. The "black box" example presented in Section 2 is an example of a linear system with an added autoregressive noise process; the noise spectrum is not constant for that example.

Consider next a *multiple-factor linear system*

$$y(t) = \sum_{i=1}^{p} \sum_{k=-\infty}^{\infty} h_{ik} x_i(t-k) + \varepsilon(t).$$

Here the response spectrum and the factor spectra, $f_{x_i}(\omega)$, are related by

$$f_y(\omega) = \sum_{i=1}^{p} G_i^2(\omega) f_{x_i}(\omega) + f_\varepsilon(\omega), \tag{3}$$

where each $G_i(\omega)$ is the gain for oscillations of x_i. Suppose each factor x_i oscillates at the frequency ω_i:

$$x_i(t) = \alpha_i \cos(2\pi\omega_i t + \delta_i). \tag{4}$$

Then the factor spectrum $f_{x_i}(\omega)$ has a peak at ω_i. If $G_i(\omega_i)$ is not zero, the response spectrum has a peak at ω_i. A response spectrum peak at ω_i is evidence that x_i affects the response; no peak is evidence that it does not. One of course cannot guarantee that the gain might not be zero; however, it is extremely unlikely in practice that the finite set of discrete driving frequencies used will contain a frequency with zero gain (this can be argued to be an event with zero probability). In Section 4.3 we offer several practical approaches for dealing with the problem of nonuniform system gain.

3.1 Higher Order Terms in the Response Surface Model

A second advantage of frequency domain simulation experiments is that high-order terms in a response surface model can be identified without additional runs of the simulation program. This is due to the fact that *products* of factors in a response model become *sums* of frequencies in the frequency domain. The presence of a high-order term or factor interaction in a general polynomial response surface is equivalent to a set of additional *linear pseudofactors* for the model. Therefore, the input/output spectrum relationships applicable to linear response surfaces are also applicable to general polynomial response surfaces. Every potential term in the prospective polynomial response surface model is some product of integral powers of the input factors. Each product of powers of oscillating input factors will itself oscillate. The frequencies of these oscillations are called the *indicator frequency set* for a potential term in the response surface model. The linear pseudofactors equivalent to a high-order response surface term each oscillate at one of the frequencies in the indicator frequency set for that term; a term's indicator frequencies are the "driving frequencies" for the equivalent linear pseudofactors. Linear inputs and outputs oscillate at the same frequencies (see [2, chap. 9]. Thus changes in the response spectrum at frequencies in a high-order term's

indicator frequency set indicate that the term should be included in the response surface model.

Indicator frequencies are easily computed from the oscillation frequencies of the individual input factors. The computations are based on elementary trigonometric identities. For example, suppose x_i and x_j are input factors with driving frequencies ω_i and ω_j; that is, $x_i(t) = \alpha_i \cos(2\pi\omega_i t)$ and $x_j(t) = \alpha_j \cos(2\pi\omega_j t)$. The quadratic term $x_i^2(t)$ can be expressed

$$\begin{aligned} x_i^2(t) &= \alpha_i^2 \cos^2(2\pi\omega_i t) \\ &= \tfrac{1}{2}\alpha_i^2(1 + \cos(2\pi(2\omega_i)t)) \\ &= \tfrac{1}{2}\alpha_i^2 \cos(2\pi(0)t) + \tfrac{1}{2}\alpha_i^2 \cos(2\pi(2\omega_i)t). \end{aligned}$$

The presence of $x_i^2(t)$ in the response model is indicated by peaks in the response spectrum at the indicator frequencies 0 and $2\omega_i$. Thus the set of indicator frequencies for the term x_i^2 is $\{0, 2\omega_i\}$.

The product $x_i x_j$ can be expressed

$$x_i x_j = \tfrac{1}{2}\alpha_i\alpha_j(\cos 2\pi(\omega_i + \omega_j)t + \cos 2\pi(\omega_i - \omega_j)t).$$

The presence of $x_i x_j$ in the response model is indicated by peaks in the response spectrum at both $\omega_i + \omega_j$ and $\omega_i - \omega_j$. Thus the set of indicator frequencies for the term $x_i x_j$ is $\{\omega_i + \omega_j, \omega_i - \omega_j\}$.

Sets of indicator frequencies for all high-order polynomial terms in the prospective response model can be similarly established using elementary trigonometric identities. These sets are given by the following indicator frequency rule:

Rule. The set of indicator frequencies for the term

$$\tau_j = \prod_{i=1}^{p} [x_i]^{a_{ij}}$$

is the direct sum of the sets

$$S_i = \{a_{ij}\omega_i, (a_{ij} - 2)\omega_i, \ldots, -a_{ij}\omega_i\}. \qquad (5)$$

The direct sum of sets is formed by taking all combinations of one element from each set and adding. This rule extends the results in [4], which apply frequency domain experiments to fit planar response surface models to chemical reactions, but do not consider powers or interactions of the input factors.

3.2 Spectrum Amplification, Correlations, and Model Coefficients

The heights of spectrum peaks are related to the coefficients of the linear response model

$$E[y] = \beta_0 + \sum_{i=1}^{p} \beta_i x_i.$$

Consider the relationship between the response spectrum and the factor spectra in a multiple-factor linear system with added noise, a system represented by Eq. (3). The gain $g_i(\omega)$ is scaled by the coefficient β_i to define the *standardized gain function* $g_i(\omega) = G_i(\omega)/\beta_i$. The input factor $x_i(t)$ oscillates as a sinusoid with amplitude α_i and frequency ω_i. Following [2, Appendix II], we formally define the input factor spectrum as $f_{x_i}(\omega) = \alpha_i^2 s(\omega_i)$, where $s(\omega_i)$ is the spectrum of a sinusoid with frequency ω_i and amplitude 1; $s(\omega_i)$ has a sharp peak at ω_i and is zero elsewhere. (The reader may note that the proper mathematical interpretation of $s(\omega_i)$ is as a generalized delta function.) Formally substituting these expressions into Eq. (3) gives

$$f_y(\omega) = \sum_{i=1}^{p} \beta_i^2 g_i^2(\omega)\alpha_i^2 s(\omega_i) + f_\epsilon(\omega).$$

For purely mathematical reasons, the phase angles (in Eq. 4) of all driving frequencies are assumed to be uniformly distributed between 0 and 2π. The phase angle has no effect on the frequency spectrum. Again, to simplify this theoretical discussion, we will assume that all driving frequencies are Fourier frequencies (frequencies that are multiples of $1/n$). It is not critical that Fourier driving frequencies be used in applications. The sample variance of each linear term in the response surface metamodel follows from the usual computational definition,

$$\sigma_{x_i}^2 = \frac{1}{n}\sum_{t=1}^{n} (\alpha_i \cos(2\pi\omega_i t) - \mu_{x_i})^2 = \frac{\alpha_i^2}{2}, \qquad i = 1, \ldots, k_j,$$

where the average input signal (for oscillations at Fourier frequencies) is

$$\mu_{x_i} = \frac{1}{n}\sum_{t=1}^{n} \alpha_i \cos(2\pi\omega_i t) = 0.$$

The sample correlation between input factor values is zero when Fourier frequencies are used. The sample correlation between the output sequence, $y(t)$, and the input sequence, $x_i(t)$, for $(t = 1, \ldots, n)$, is

$$\rho_i = \beta_i \sigma_{x_i}/\sigma_y.$$

Substituting the above into the general response relationship of Eq. (3) yields

$$f_y(\omega) = 2\sigma_y^2 \sum_{i=1}^{p} \rho_i^2 g_i^2(\omega) s(\omega_i) + f_\epsilon(\omega).$$

This equation states that the height of the output spectrum peak at ω_i is scaled by $2\sigma_y^2\rho_i^2 g_i^2(\omega)$. The relative heights are proportional to $\rho_i^2 g_i^2(\omega)$. If $g_i(\omega)$ is constant, then the heights of the output spectrum peaks are proportional to ρ_i^2. This result is stated as follows:

Lemma. If the standardized gain $g_i(\omega)$ is constant for every x_i and ω, then ranking the factors according to the heights of their output spectrum peaks is equivalent to ranking the factors according to the magnitudes of their sample correlations with the response.

Dividing the heights of the peaks by the squares of the input amplitudes compensates for possible differences in these amplitudes. These are proportional to

$$\frac{2\sigma_y^2\rho_i^2 g_i^2(\omega)}{\alpha_i^2} = \frac{2\sigma_y^2\rho_i^2 g_i^2(\omega)}{2\sigma_{x_i}^2} = \beta_i^2 g_i^2(\omega).$$

If the gain is constant, then these are proportional to β_i^2. Thus we have the following corollary:

Corollary. If the standardized gain $g_i(\omega)$ is constant for every x_i and ω, then ranking the linear factors according to the heights of their peaks divided by the squares of their driving oscillation amplitudes is equivalent to ranking the factors according to the magnitudes of their coefficients in the linear response surface model.

As shown in Section 3.1, a general polynomial term in the response surface model is equivalent in the frequency domain to a set of linear pseudofactors each oscillating at the term's indicator frequencies. Therefore, with suitable modifications (scaling by the driving amplitudes of the equivalent linear pseudofactors) the above results can be applied to rankings of general polynomial response surface terms.

4. DESIGNING FREQUENCY DOMAIN EXPERIMENTS

There are three steps to designing a frequency domain simulation experiment:

(1) the selection of a set of driving frequencies for the input factors,
(2) the determination of the amplitudes of the driving frequencies, and
(3) the assignment of driving frequencies to each input factor.

4.1 Selection of Factor Driving Frequencies

Frequencies that are used to drive oscillations of the input factors in a frequency domain simulation experiment have upper and lower bounds. When the simulation output consists of a finite-length, discrete-indexed time series (as is the case for most discrete event simulation programs), the spectrum cannot be estimated at all frequencies. The lowest detectable frequency will complete one full cycle during the simulation run. The n observations in the simulation output series cannot be used to accurately estimate spectrum components at frequencies below $1/n$ cycles per observation. The highest detectable frequency completes one-half cycle per observation. Higher frequencies cannot be observed directly, but instead appear as frequencies between zero and one-half. This phenomenon, known as *aliasing*, is described in [1]. Let ω' denote the fractional part of $|\omega|$. If $0 \leq \omega' \leq .5$, then the observed alias frequency for ω is ω'. If $.5 \leq \omega' \leq 1$, then the observed alias frequency for ω is $1 - \omega'$. The effects of aliasing must be considered in the simulation experiments proposed in this paper.

Partial confounding of indicator frequencies may occur if driving frequencies are selected carelessly. For example, suppose the frequencies $\omega_1 = .3$ and $\omega_2 = .4$ cycles per time unit are assigned to input factors x_1 and x_2. If the interaction x_1x_2 was present in the response surface model, it would be indicated by response oscillations at the indicator frequencies of $\{.1, -.1, .7, -.7\}$, which alias to the observable frequency set $\{.1, .3\}$. A spike in the output spectrum at a frequency of .3 cycles per observation indicates the presence of both the linear term, x_1, and in part the presence of the interaction term, x_1x_2, in the response model. The selection of driving frequencies for the input factors can easily be made to avoid such confounding.

Independent spectrum estimates for different indicator frequencies is desirable, so indicator frequencies should be as widely separated as possible in the interval $(0, \frac{1}{2}]$. The frequency selection problem depends on the number of input factors and the list of terms in the prospective response model. The problem of selecting driving frequencies can be formulated as a mixed integer linear program [3]. Further discussion of the frequency selection problem is given in [6]. However, frequency selection is not critical. Term indicator frequencies and their aliases should be at least one bandwidth apart. The selection of bandwidth is under the control of the experimenter and will be discussed in Section 5. If bandwidth is decreased, the same estimator precision can be obtained by increasing the run length.

4.2 Amplitudes for the Driving Frequencies

As with classical experimental designs, the experimenter must specify a range of values for each input factor. The experimental region takes the form of a p-dimensional rectangle,

$$\{(x_1, \ldots, x_p) \mid L_i \leq x_i \leq U_i\}. \tag{6}$$

The oscillation amplitudes affect the size of changes in the response spectrum. The height of the response spectrum is directly proportional to the height of the input spectrum at the same frequency, and the height of the input spectrum is scaled by the square of the oscillation amplitude. If the amplitudes are too small, the oscillation effects can be difficult to detect. If the amplitudes are too large, the input factor values can exceed the range for which the simulation model makes sense or for which the response surface can be fit by a polynomial of the assumed order. Detection of this lack of fit is discussed in Section 5. Amplitudes are chosen so that each input factor varies over its whole range of values. For example, the value for input factor x_i at simulated time t is

$$x_i(t) = \tfrac{1}{2}(U_i + L_i) + \tfrac{1}{2}(U_i - L_i)\cos(2\pi\omega_i t). \tag{7}$$

4.3 Assignment of Driving Frequencies to Factors

A potential problem with the frequency domain approach is gain. Gain describes how a system amplifies or attenuates input oscillations at each frequency. Systems sometimes tend to filter out particular frequencies and amplify others. A low-pass filter, such as an exponentially smoothed average, suppresses high frequencies. A high-pass filter, such as a period-to-period difference, suppresses low frequencies. Parameters and terms might appear to be more or less important de-

pending on their assigned frequencies of oscillation. If the gain lowers the spectrum at one of the indicator frequencies, an important term in the response model might be overlooked.

One way to deal with gain is to estimate it. Gain can be estimated by driving each input factor with *white noise* (i.e., with a sequence of independent, identically distributed values) and estimating the response spectrum. This approach requires additional simulation runs and is complicated by the possible interaction between input factor and frequency; for example, the system might act as a low-pass filter for one input factor and as a high-pass filter for another. This approach is discussed in [15].

A more systematic approach is to treat gain as an unknown nuisance factor and *block* on it. A few independently seeded simulation runs can be made, changing between runs the input factor assigned to each driving frequency. A Latin square design can be used so that each input factor is assigned once to each driving frequency. Intuitively, each factor will then have equal exposure to gain across the different frequencies. The motivation for this approach is similar to that for the classical analysis of variance (ANOVA) framework. No claim is made that most simulation models satisfy the classical ANOVA assumptions. It is more realistic to appeal to intuition for support than to try to justify ANOVA-like assumptions (see [13] for a further discussion of the connection with ANOVA).

Because response spectrum peaks can occur regardless of whether the input factors oscillate, one must determine whether the peaks result from induced input factor oscillations or from natural conditions. For this purpose the concept of a *control frequency* is introduced. A control frequency determines the level of the response spectrum under natural conditions. A control frequency is not assigned during a run; no input factor oscillations occur and no response spectrum changes are expected at a control frequency. The output spectrum at a control frequency estimates the natural or background level of the response spectrum. This allows one to make statistical inferences about whether or not response spectrum peaks are induced by input factor oscillations.

		Driving frequency		
		.11	.18	.43
Run number	1	λ	μ	*
	2	μ	*	λ
	3	*	λ	μ

The asterisk (*) denotes a frequency not used in a run—that is, a control frequency.

FIGURE 2. Assignment of Parameter Driving Frequencies for the Queue Simulation

The design of Figure 2 uses three driving frequencies for a model with two input factors. Each driving frequency serves as the control frequency for one of the runs. This provides a standard for comparing response spectrum peaks at that frequency. With one control frequency per run, a Latin square design requires $p + 1$ runs to study p input factors. This design has considerably fewer runs than are required by most conventional run-oriented designs.

In the terminology of experimental design, the factors represent treatments, the frequency bands represent random block effects, and gain is roughly analogous to an interaction between the factors and the frequency bands. The possible presence of system gain deserves further consideration because Latin square experimental designs confound interactions with main effects. In these designs there are three main effects: a factor effect, a frequency band effect, and a run effect. These effects are confounded as follows: The factor effect is confounded with the band-run interaction and the three-way factor-band-run interaction, the band effect is confounded with the factor-run interaction and the three-way interaction, and the run effect is confounded with the factor-band interaction (gain) and the three-way interaction. By independently seeding each of the runs, we are effectively randomizing this factor in an attempt to avoid any systematic main effect or interactions involving runs. In the absence of band-run and three-way interactions, the factor effect can be observed without confounding. The gain effect can also be observed without confounding.

Not all Latin square designs are equally suited for frequency domain experiments. A Latin square design assigns driving frequencies to the various input factors in the simulation. Implicit in this assignment is the determination of indicator frequencies for each of the high-order terms in the response model. Some Latin square designs assign the same frequencies repeatedly to a particular interaction, whereas other Latin square designs assign a greater variety of frequencies to interactions. A variety of frequencies should be assigned because it makes the blocking scheme more effective (this is discussed by Cogliano [3]).

Before running a full Latin square experiment like that just described, the experimenter should make a pair of runs assigning driving frequencies from the interval (0, .25] for the first run and from the interval (.25, .5] for the second run. The sets of term indicator frequencies from the two runs should be disjoint. Spectrum estimates for unused frequency bands from one run can provide the control spectrum for the other run. Spectral ratios, $R(\omega)$, can be computed as described in Section 2. If the runs are independently seeded, then one can test the statistical significance of terms in the metamodel as described in Section 5. Using common random number streams for the pair of runs sharpens the contrasts by reducing the variance of the spectral ratios. However, a conventional statistical analysis of spectral ratios from runs using common random number streams is not valid. Nevertheless, a qualitative screening of factors can be extremely valuable.

5. ANALYSIS OF FREQUENCY DOMAIN EXPERIMENTS

The response from a simulation run is often a time series, denoted by $y(1), \ldots, y(n)$. The frequency spectrum measures the relative strength of sinusoidal oscillations in a time series. In the examples to be presented in Section 7, we estimate the frequency spectrum with the *sample spectrum* [1] given by

$$\hat{f}(\omega) = \sum_{k=-m}^{m} \lambda_k c_k \cos(2\pi\omega k), \qquad 0 \le |\omega| \le \tfrac{1}{2} \tag{8}$$

with the usual autocovariance estimators given by

$$c_k = \frac{1}{n} \sum_{t=1}^{n-k} (y(t) - \bar{y})(y(t+k) - \bar{y})$$

$$\text{if} \quad 0 \le k \le (n-1)$$

and $\qquad c_k = c_{-k} \qquad \text{if} \quad -(n-1) \le k \le 0.$

We use the weights (called the Tukey lag window of size m) given by

$$\lambda_K = \tfrac{1}{2}(1 + \cos\pi k/m), \qquad \text{for} \quad |k| \le m.$$

The Tukey window is used here because it is well known and has desirable statistical properties such as consistency. Other windows such as the Parzen window might be used as well. More important than the window type is the *truncation point, m* [7, p. 252]. There is a trade-off in truncation point selection. A small value of m gives an estimator with high precision (low variance), and a large value of m gives an estimator with high resolution (small bandwidth). The bandwidth of the Tukey window is $4/3m$ (see [7, p. 255]). Following [7, p. 286] we will treat spectral estimators that are at least one bandwidth apart as approximately independent.

Lack of fit by any polynomial function of order less than or equal to that of the prospective response surface model (1) is indicated by significant output spectrum peaks at frequencies that are not indicator frequencies for terms in the prospective response model. Instead of using the full-frequency spectrum in the analysis, a harmonic analysis at only the indicator frequencies for terms in the prospective response surface model can be performed. However, estimating the full spectrum provides the lack-of-fit check for any model in the class of prospective response surface models.

Statistical analysis of response spectrum changes uses the fact that $(8n/3m)(\hat{f}(\omega)/f(\omega))$ has an approximate chi-square distribution with $v = 8n/3m$ degrees of freedom [7, p. 253]. Here $f(\omega)$ is the true spectrum at ω and $\hat{f}(\omega)$ is its estimator. This approximate result has been shown to be quite good even for short time series [11].

Consider the situation in which there are several independently seeded runs of the simulation experiment. The minimum number of runs is two. The statistical analysis proceeds in nested steps:

(1) *For each potential response surface model term* τ_j, compile the list of its indicator frequencies $\omega_1, \omega_2, \ldots, \omega_{k_j}$.

(2) *For each indicator frequency* ω_i, compute the output spectrum estimates, $\hat{f}_1(\omega_i), \ldots, \hat{f}_r(\omega_i)$, *for each of the r runs* of the simulation experiment where ω_i is an indicator frequency for term τ_j. (In practice there would probably be only one such run.) Compute also the output spectrum estimates $[\hat{f}^c_1(\omega_i), \ldots, \hat{f}^c_s(\omega_i)]$ for each of the s runs in the experiment, where ω_i is not an indicator frequency for any term in the prospective response surface model. The superscript, c, on these estimators indicates that these are considered "control" estimates. Since runs are independently seeded and spectrum estimators that are more than one frequency bandwidth apart are approximately independent, the spectrum estimators are considered independent. Under the hypothesis that the term τ_j should not be included in the response surface model, the true value for the spectrum at ω_i is constant in all the above runs. This true spectrum value will cancel out of the spectral ratio,

$$f_j(\omega_i) = \frac{1/vr \sum_{k=1}^{r} \hat{f}_k(\omega_i)}{1/vs \sum_{k=1}^{s} \hat{f}^c_k(\omega_i)}, \tag{9}$$

which is the ratio of two approximately independent chi-square random variables each divided by their degrees of freedom (the quantity v was defined earlier and depends on the lag window used in spectrum estimation). Therefore, $F_j(\omega_i)$ can be considered as having an F distribution with vr and vs degrees of freedom. Let $p_j(\omega_i)$ denote the observed significance level of $F_j(\omega_i)$. This value is the probability that an $F_{vr,vs}$ distributed random variable exceeds $F_j(\omega_i)$.

(3) Combine the significance levels for all indicator frequencies for the term τ_j. There are several methods for combining significance levels from independent tests; these are discussed and compared in [12]. In the examples in Section 7, we arbitrarily used the most popular method, Fisher's method, to compute the combined significance level given by the quantity

$$p_j^* = \text{Prob.}\{X \ge -2 \sum_{i=1}^{k_j} \log p_j(\omega_i)\}, \tag{10}$$

where X has a chi-square distribution with $2k_j$ degrees of freedom and $\{\omega_i;\ i = 1, 2, \ldots, k_j\}$ is the set of indicator frequencies for the term τ_j. A low combined significance level (say, $p_j^* \le .1$) is evidence that the term τ_j should be *included* in the response surface model. That is, the hypothesis that there is no amplification at indicator frequencies for the response model term, τ_j, is rejected at the p_j^* level of significance.

The power of the above test for a particular term in the prospective response surface metamodel increases as the driving frequency amplitude increases. Comparing p_j^* values for different terms is similar to comparing the significance of regression coefficients when the ranges of the independent variables are different. Comparisons of p_j^* values are appropriate only over the experimental region given by Eq. (6).

If one is primarily concerned with failing to detect a term in the response surface model, then all runs in the

experiment should be made with common random number streams. When common streams are used, the p_j*s should be regarded as only qualitative measures of a term's significance since independently seeded runs are necessary to justify the assumption that the spectral ratios will have an F distribution. Finally, one should also note that (as in classical ANOVA) the observed significance levels for different terms are not independent when the same control frequencies are used in the denominator of Eq. (9).

6. EXPERIMENTAL PROCEDURE FOR METAMODEL IDENTIFICATION

The following steps summarize the frequency domain approach to simulation sensitivity analysis:

Step 1. Determine an appropriate common index for the output and input series. The output may need to be sorted so that its index is in the same order as the input. (See the discussion in Section 2.1.)

Step 2. Select a range of interest (Eq. (6)) for each continuous input factor. The larger the region, the more power there is to detect input factor effects.

Step 3. Select $p + 1$ driving frequencies between 0 and ½. Determine the principal aliases of the indicator frequencies (Eq. (5)) for each term in the prospective response model. Compute the minimum spacing, b, between these frequencies—this will be the bandwidth necessary for the spectral estimators. Some effort should be made to select driving frequencies that maximize this spacing.

Step 4. Choose a window truncation point m and a run length n such that $m \geq 4/3b$. Chatfield [2, p. 141] recommends that n be roughly $(m/2)^2$. Empirical work with this method suggests that n be large enough to include at least 10 full cycles of the lowest term indicator frequency. The larger the value of n, the smaller the variance of the spectral estimators. For simulation models the run lengths typically can be made quite large for very little marginal increase in experimental cost. For terminating simulations where the run length cannot be controlled, several independently seeded replications can be run, and their spectra averaged across the replicates to reduce estimator variance.

Step 5. Run $p + 1$ independently seeded replications of the simulation program using a Latin square design to assign input factors to driving frequencies as in Section 4.3. For each run the input factors oscillate according to Eq. (7).

Step 6. Compute the sample spectrum (Eq. (8)) for each response series.

Step 7. Compute the spectrum ratios (Eq. (9)) and the combined significance level (Eq. (10)) to evaluate each term for inclusion in the prospective simulation response surface model.

7. SOME EXAMPLES

Two detailed simulation response surface model identification studies are presented in order to illustrate the procedure presented in the previous section. The first example is an M/M/1 queue, and the second is a more complex model of an inventory system with repairable items.

7.1 Example: A Single-Server Queue

Jobs arrive at a single service center according to a Poisson process with an arrival rate of $\lambda = .8$. Service times are exponentially distributed with a mean of $1/\mu = 1$. Service is given on a first-come-first-served basis. The system starts empty, and the performance measure is the average waiting times of the first 300 jobs. The goal of the simulation study is to identify a response model of at most second order consisting of perhaps linear terms (λ and μ), quadratic terms (λ^2 and μ^2), and the interaction term ($\lambda\mu$). The following steps illustrate how the steps of the frequency domain procedure were followed.

Step 1. The index used here for the input and output series is the customer arrival sequence number; the "time" index for the parameter oscillations is customer arrival number. Since the queue is FIFO, there is no need to sort the output series by arrival number before analysis.

Step 2. The ranges of interest for the input factors were taken to be λ between 0.4 and 1.2 and μ between 0.6 and 1.4. Note that, since we are sampling only the (transient) means of the first $n = 300$ jobs, the stationarity restriction $\lambda < \mu$ is not required.

Step 3. The three driving frequencies selected were $\omega_1 = .11$, $\omega_2 = .18$, and $\omega_3 = .43$ cycles per job. These driving frequencies achieve the maximal spacing between all term indicator frequencies for a prospective second-order response surface model. The resulting term indicator frequencies (after aliasing) were {.11, .18, .43} for the linear terms, {.22, .36, .14} for the quadratic terms, and {.07, .29, .32, .46, .25, .39, .00, .50} for the two-way interaction terms. The minimum spacing between these frequencies is $b = .03$.

Step 4. The window truncation point was chosen to be $m = 100$, equal to one-third the number of observations in the output series. The resulting bandwidth is $1.33/m = .0133$ which is considerably less than the minimum spacing between indicator frequencies. This permits treating the spectrum estimators for different indicator frequencies as being essentially independent.

Step 5. Three independently seeded runs of the simulation were made. The input factors were assigned to the driving frequencies according to the Latin square design in Figure 2. For each run the waiting times of the first 300 jobs were collected.

Step 6. The output series sample spectrum for each run was estimated using Eq. (8).

TABLE I. Results of the Sensitivity Experiment for the Queue Simulation

Model term τ_j	Indicator frequencies ω_i	$F_j(\omega_i)$	Degrees of freedom for $F_j(\omega_i)$	$p_j(\omega_i)$	Combined χ^2 values	χ^2 degrees of freedom	p_j^*
λ	0.110	2.523	8, 8	0.106	11.127	6	.08
	0.180	3.606	8, 8	0.045			
	0.430	0.524	8, 8	0.811			
μ	0.110	3.211	8, 8	0.060	15.111	6	.02
	0.180	4.440	8, 8	0.026			
	0.430	1.345	8, 8	0.342			
λ^2	0.220	0.818	8, 8	0.609	6.128	6	>.5
	0.360	2.563	8, 8	0.102			
	0.140	0.611	8, 8	0.750			
μ^2	0.220	0.710	8, 8	0.681	10.628	6	.10
	0.360	4.458	8, 8	0.025			
	0.140	1.510	8, 8	0.286			
$\lambda\mu$	0.070	0.760	8, 16	0.642	10.680	12	>.5
	0.290	1.029	8, 16	0.454			
	0.320	0.850	8, 16	0.576			
	0.460	1.033	8, 16	0.452			
	0.250	0.874	8, 16	0.558			
	0.390	2.000	8, 16	0.113			

Step 7. The combined significance level of each term, p_j^*, was computed as described in Section 5. A summary of these calculations is given in Table I. The terms that were found to be significant at, say, the .1 level are λ, μ, and μ^2.

For comparison with conventional run-oriented designs, a 10-run response surface experiment (four corner points plus a four-pointed star plus two center points) was conducted. The same experimental region was used as for the frequency domain experiment. The experimental design is given in the left three columns and the experimental results are given in the rightmost column of Table II.

All possible polynomial regression models of order less than or equal to two were fit to the output. Terms were selected (based on the R^2 criterion, [5]) to be entered into the regression model in the *same order* as their significance levels in the frequency domain experiment. This is in agreement with the lemma in Section 3.2. Table III shows the fraction of the response variability, R^2, explained by the sequence of regression models. There are other measures of the goodness of a regression model; however, R^2 is one of the more commonly accepted measures. The first three regression models were the best (highest R^2) one-variable, two-variable, and three-variable regression models. These best models had R^2 values equal to .513, .694, and .865, respectively.

The frequency domain approach worked very well here. It identified with three runs the best second-order polynomial regression models that could be found with the output from a 10-run conventional simulation experiment.

TABLE II. Results of Response Surface Experiment for the Queue Simulation

Run	λ	μ	Average wait
1	1.0	1.2	3.646
2	0.6	1.2	0.822
3	0.6	0.8	3.525
4	1.0	0.8	51.592
5	1.2	1.0	36.893
6	0.8	1.4	0.970
7	0.4	1.0	0.614
8	0.8	0.6	79.741
9	0.8	1.0	2.735
10	0.8	1.0	3.653

TABLE III. Sequence of Regression Models for the Queue Experiment

Terms in regression model	R^2
μ	.513
μ, λ	.694
μ, λ, μ^2	.865
μ, λ, μ^2, λ^2	.903
μ, λ, μ^2, λ^2, $\lambda\mu$	.975

7.2 Example: An Inventory System with Repairable Items

Our second example is an inventory system with repairable items. The system consists of a supply facility and a repair facility. Items are demanded according to a Poisson process with a mean of 15 items per day. Items

are returned for repair according to a Poisson process with a mean of 6 items per day. Repair times are geometrically distributed with a mean of 3 days. After being repaired, items are placed in inventory. Whenever the inventory position (defined as the number of items in supply, in repair, or on order) reaches or falls below 36 items at the end of a day, an additional lot of 24 items is ordered. The lot arrives in 3 days. Excess demand that cannot be filled immediately is back ordered. At the end of each day, a holding cost of \$1 is assessed for each item in supply, and a back-order cost of \$5 is assessed for each item back ordered. Placing an order costs \$10. The inventory system is pictured in Figure 3.

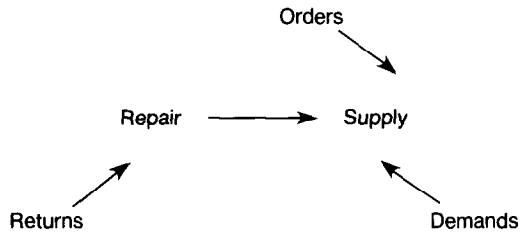

FIGURE 3. An Inventory System with Returns

The goal of the study is to assess the sensitivity of the operating costs to three input factors: the demand rate λ, the return rate γ, and the mean repair time r. We are interested in identifying the significant terms in a second-order polynomial regression model of the operating costs. The potential terms in the regression model are the linear terms (λ, γ, and r), the quadratic terms (λ^2, λ^2, and r^2), and the interaction terms ($\lambda\gamma$, λr, and γr). The following steps give the computational details of the procedure of Section 6.

Step 1. The index used here is simulated time, measured in days.

Step 2. The ranges of interest for each input factor were arbitrarily chosen to be between two-thirds and four-thirds of its mean value. This is, λ ranged between 10 and 20, γ between 4 and 8, and r between 2 and 4.

Step 3. The four driving frequencies selected were $\omega_1 = .06$, $\omega_2 = .20$, $\omega_3 = .29$, and $\omega_4 = .39$. The resulting indicator frequencies were {.06, .20, .29, .39} for the linear terms, {.12, .40, .42, .22} for the quadratic terms, and {.14, .26, .23, .35, .33, .45, .09, .49, .19, .41, .10, .32} for the interaction terms. The minimum spacing between the indicator frequencies was .01.

Step 4. A window truncation point of $m = 100$ and a run length of $n = 300$ days were chosen. The resulting bandwidth allows us to treat the spectral estimates at each of the term indicator frequencies as independent. Simulated clock time (measured in days) was the index for the driving frequencies and output in this example. A series of 300 daily operating costs was recorded from each run.

Step 5. Four independently seeded runs of the simulation model were made. The input factors were assigned to the driving frequencies according to the Latin square design of Table IV. The resulting indicator frequency assignments are given in Table V. To cover the experimental region, each input factor oscillated at its assigned driving frequency with an amplitude of one-third of its mean value. The initial inventory was set at 60, and no warm-up period was used.

TABLE IV. Latin Square Design for the Inventory Experiment

	Driving frequencies			
Run	.06	.20	.29	.39
1	*	λ	γ	r
2	r	*	λ	γ
3	γ	r	*	λ
4	λ	γ	r	*

The asterisk (*) denotes a frequency not used in a run—that is, a control frequency.

Step 6. The response spectrum was computed for the operating cost series from each of the four runs. The spectrum estimates at the indicator frequencies are arranged into two-way layouts by term and frequency in Table VI.

Step 7. Table VII presents the analysis of the spectral estimates. For each term several F-ratios were computed by dividing the spectral estimates from Table VI in a term's row by the corresponding spectral estimates in the control row. For example, the four F-ratios for λ were computed as follows: 20.389 = 4873/239, 5.445 = 893/164, 1.331 = 173/130, and 2.944 = 683/232. For cells with more than one entry, such as those in the interaction table, the entries were averaged over the degrees of freedom. For example, the first F-ratio for $\lambda\gamma$ was computed as

$$1.715 = \frac{385}{\frac{1}{2}(291 + 158)}.$$

This F-ratio has 8 and 16 degrees of freedom. The re-

TABLE V. Indicator Frequency Assignments for the Inventory Experiment

Run	.06	.20	.29	.39	.12	.40	.42	.22	.14	.26	.23	.35	.33	.45	.09	.49	.19	.41	.10	.32
	*	λ	γ	r	*	λ^2	γ^2	r^2	*	*	*	*	*	*	$\lambda\gamma$	$\lambda\gamma$	λr	λr	γr	γr
1																				
2	r	*	λ	γ	r^2	*	λ^2	γ^2	*	*	λr	λr	γr	γr	*	*	*	*	$\lambda\gamma$	$\lambda\gamma$
3	γ	r	*	λ	γ^2	r^2	*	λ^2	γr	γr	*	*	$\lambda\gamma$	$\lambda\gamma$	*	*	λr	λr	*	*
4	λ	γ	r	*	λ^2	γ^2	r^2	*	$\lambda\gamma$	$\lambda\gamma$	λr	λr	*	*	γr	γr	*	*	*	*

TABLE VI. Spectral Estimates for the Inventory Simulation

	Linear effects frequency					Quadratic effects frequency			
	.06	.20	.29	.39		.06	.20	.29	.39
Control	239	164	130	232	Control	80	408	411	170
λ	4873	893	173	683	λ^2	1725	635	96	68
γ	203	233	106	298	γ^2	262	501	185	202
r	560	129	362	293	r^2	93	514	184	99

	Interaction effects frequency											
	.14	.26	.23	.35	.33	.45	.09	.49	.19	.41	.10	.32
Control	291 158	73 139	45 93	92 59	112 92	383 296	232 149	186 314	222 136	230 609	130 307	59 173
$\lambda\gamma$	385	381			36	166	231	259			220	96
λr			241 509	102 140					146 133	418 248		
γr	128	70			192	66	366	106			400	76

TABLE VII. Results of the Sensitivity Experiment for the Inventory Simulation

Model term τ_j	Indicator frequencies ω_i	$F_j(\omega_i)$	Degrees of freedom for $F_j(\omega_i)$	$(p_j\,\omega_i)$	Combined χ^2 values	χ^2 degrees of freedom	p_j^*
λ	.06	20.389	8, 8	.001	29.677	8	.001
	.20	5.445		.014			
	.29	1.331		.347			
	.39	2.944		.074			
γ	.06	0.849	8, 8	.589	6.365	8	>.5
	.20	1.421		.315			
	.29	0.815		.611			
	.39	1.284		.366			
r	.06	2.343	8, 8	.125	11.983	8	.20
	.20	0.787		.629			
	.29	2.785		.085			
	.39	1.263		.374			
λ^2	.12	21.563	8, 8	.001	16.697	8	.05
	.40	1.556		.273			
	.42	0.234		.972			
	.22	0.400		.892			
γ^2	.12	2.900	8, 8	.077	9.121	8	>.5
	.40	1.228		.389			
	.42	0.448		.862			
	.22	1.188		.406			
r^2	.12	1.163	8, 8	.418	4.526	8	>.5
	.40	1.260		.375			
	.42	0.448		.862			
	.22	0.582		.770			
$\lambda\gamma$	.14	1.715	8, 16	.170	18.814	16	.30
	.26	3.594		.014			
	.33	0.353		.930			
	.45	0.489		.847			
	.09	1.213		.352			
	.49	1.036		.450			
	.10	1.007		.468			
	.32	0.828		.591			

(continued)

TABLE VII. Results of the Sensitivity Experiment for the Inventory Simulation (continued)

Model term τ_j	Indicator frequencies ω_i	$F_j(\omega_i)$	Degrees of freedom for $F_j(\omega_i)$	$(p_j\omega_i)$	Combined χ^2 values	χ^2 degrees of freedom	p_j^*
λr	.23	5.435	16, 16	.001	18.810	8	.025
	.35	1.603		.177			
	.19	0.779		.689			
	.41	0.794		.675			
γr	.14	0.570	8, 16	.788	14.067	16	>.5
	.26	0.660		.719			
	.33	1.882		.134			
	.45	0.194		.987			
	.09	1.921		.127			
	.49	0.424		.890			
	.10	1.831		.144			
	.32	0.655		.723			

maining columns in Table VII show the significance level from the independent experiments, the degrees of freedom for this χ^2 statistic, and the resulting overall significance levels (p_j^*s).

For comparison, a conventional 3^3 complete factorial run-oriented simulation experiment was conducted, and regression models fit to the results. The corner points of the factorial design corresponded to the extreme points of the sinusoidal oscillations of the input factors. That is, we used the same experimental region as for the frequency domain experiment. The experimental design is given in the left four columns and the results of the experiment are given the rightmost column of Table VIII.

TABLE VIII. Complete Factorial Design and Results

Run	Parameter settings λ	γ	r	Average response (daily cost)
1	10	4	2	27.603
2	10	4	3	23.813
3	10	4	4	20.327
4	10	6	2	28.463
5	10	6	3	22.633
6	10	6	4	18.277
7	10	8	2	31.630
8	10	8	3	23.753
9	10	8	4	17.020
10	15	4	2	23.760
11	15	4	3	26.930
12	15	4	4	33.683
13	15	6	2	21.967
14	15	6	3	25.237
15	15	6	4	37.240
16	15	8	2	20.767
17	15	8	3	23.270
18	15	8	4	40.297
19	20	4	2	59.347
20	20	4	3	73.937
21	20	4	4	92.410
22	20	6	2	50.793
23	20	6	3	72.643
24	20	6	4	99.227
25	20	8	2	43.993
26	20	8	3	71.597
27	20	8	4	109.043

All possible polynomial regression models of order less than or equal to two were fit to the results in Table VIII. As in the previous example, terms were selected to be entered into the regression model in the *same order* as their significance levels in the frequency domain experiment. The order of significance for the terms is λ, λr, λ^2, r, and $\lambda\gamma$. Table IX shows the proportion of response variability explained by the sequence of regression models. The four terms with the highest overall significance in the frequency domain experiment provide a regression model that explains virtually all of the response variability ($R^2 = .98$). For comparison, a regression model including all nine potential terms explains only slightly more of the response variability.

TABLE IX. Results of the Regression Study

Terms in the regression model	R^2
λ	.622
λ, λr	.753
λ, λr, λ^2	.895
λ, λr, λ^2, r	.980
λ, λr, λ^2, r, $\lambda\gamma$	.980
λ, γ, r, λ^2, γ^2, r^2, $\lambda\gamma$, λr, γr	.987 (maximum)

The frequency domain experiment worked very well here for model identification. It identified with 4 runs the same models that a complete factorial experiment identified with 27 runs.

8. CONCLUDING REMARKS

A frequency domain experimental method for selecting a simulation response surface model has been presented. A limited comparison of this technique with conventional run-oriented simulation experiments was made. In each case the polynomial model selected was

identical to the one selected using a conventional design for the simulation runs. Conventional simulation experiments, however, typically require many more runs of the simulation. The examples in the Introduction and in Section 7 are typical of our experiences with the frequency approach. The experiments reported here involved only a few short runs. Only when very small driving amplitudes were used were the results unsatisfactory. The relative efficiency of the frequency domain approach as compared to conventional experimental methods increases with the number of input factors. The full potential of this approach is better appreciated in large-scale simulation experiments. Simulation practitioners should consider frequency domain response surface model identification as a means of quickly focusing their attention on important input factors. Initially the two (commonly seeded) runs required for a qualitative analysis like that for the example in Section 2.1 should be made.

Acknowledgments. The authors would like to thank Robert Sargent of Syracuse University and David Heath and Arnold Buss of Cornell University along with three anonymous referees for their comments on earlier drafts of this paper.

REFERENCES

1. Bloomfield, P. *Fourier Analysis of Time Series: An Introduction.* Wiley, New York, 1976.
2. Chatfield, C. *The Analysis of Time Series: An Introduction.* 3rd. ed. Chapman and Hall, London, 1984.
3. Cogliano, V.J. Sensitivity analysis and model identification in simulation studies. Ph.D. dissertation, School of Operations Research and Industrial Engineering, Cornell Univ., Ithaca, N.Y., 1982.
4. Cukier, R.I., Fortuin, C.M., Shuler, K.E., Petschek, A.G., and Schaibly, J.H. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I: Theory. *J. Chem. Phys. 59,* 8 (Aug. 1973), 3873–3878.
5. Draper, N.R., and Smith, H. *Applied Regression Analysis.* Wiley, New York, 1966.
6. Jacobson, S., Buss, A., and Schruben, L. The frequency selection problem in frequency domain experiments. Tech. Rep. 714, School of Operations Research and Industrial Engineering, Cornell Univ., Ithaca, N.Y., Aug. 1986.
7. Jenkins, G.M., and Watts, D.G. *Spectral Analysis and Its Applications.* Holden-Day, San Francisco, Calif., 1968.
8. Kleijnen, J.P.C. Experimentation with models: Statistical design and analysis techniques. In *Progress in Modelling and Simulation,* F.E. Cellier, Ed. Academic Press, New York, 1982, pp. 173–185.
9. Kleijnen, J.P.C., and van den Burg, A.J., and van der Ham, R.T. Generalization of simulation results: Practicality of statistical methods. *Eur. J. Oper. Res. 3,* 1 (1979), 50–64.
10. Ljung, L. *System Identification: Theory for the User.* Prentice-Hall, Englewood Cliffs, N.J., 1987.
11. Neave, H.R. Observations on "Spectral Analysis of Short Time Series—A Simulation Study" by Granger and Hughes. *J. R. Stat. Soc. A,* 135 (1972), 393–405.
12. Rosenthal, R. (1978). Combining results of independent studies. *Psychol. Bull. 85,* 1 (Jan. 1978), 185–193.
13. Sanchez, P.J. Design and analysis of frequency domain experiments. Ph.D. thesis, School of Operations Research and Industrial Engineering, Cornell Univ., Ithaca, N.Y., 1987.
14. Schruben, L. Simulation optimization using frequency domain methods. In *Proceedings of the 1986 Winter Simulation Conference,* J.R. Wilson, J.O. Henriksen, and S.D. Roberts, Eds. IEEE, 1986, pp. 366–369.
15. Schruben, L.W., and Cogliano, V.J. Simulation sensitivity analysis: A frequency domain approach. In *Proceedings of the 1981 Winter Simulation Conference,* T.I. Oren, C.M. Delfosse, and C.M. Shub, Eds. IEEE, 1981, pp. 455–460.
16. Schruben, L.W., and Margolin, B.H. Pseudo-random number assignment in statistically designed simulation and distribution sampling experiments. *J. Am. Stat. Assoc. 73,* 363 (Sep. 1978), 504–520.
17. Suri, R. Implementation of sensitivity calculations on a Monte Carlo experiment. *J. Optim. Theory Appl. 40,* 4 (Aug. 1983), 625–630.

CR Categories and Subject Descriptors: I.6.4 [**Simulation and Modeling**]: Model Validation and analysis; G.3 [**Probabilistic Algorithms**]
General Terms: Experimentation
Additional Key Words and Phrases: Response surface methodology, sensitivity analysis, simulation, spectral analysis

Received 8/85; revised 5/86; accepted 2/87

Authors' Present Addresses: Lee W. Schruben, College of Engineering, School of Operations Research and Industrial Engineering, Upson Hall, Cornell University, Ithaca, NY 14853-7501; V. James Cogliano, U.S. Environmental Protection Agency, 401 M Street SW, Washington, D.C. 20460.