

SCHOOL OF OPERATIONS RESEARCH  
AND INDUSTRIAL ENGINEERING  
COLLEGE OF ENGINEERING  
CORNELL UNIVERSITY  
ITHACA, NEW YORK 14853

TECHNICAL REPORT NO. 784

April 1988

FREQUENCY DOMAIN SIMULATION EXPERIMENTS:  
THE FALSE NEGATIVES PROBLEM

by

Douglas Morrice  
Lee W. Schruben

## **Abstract**

Frequency domain simulation experiments were introduced by Schruben and Cogliano (1987) as an efficient method for detecting relationships between the input processes and an output process of a simulation model. One objective for conducting such experiments is to identify those input factors that significantly influence the output and screen out the unimportant input factors. Further experimentation with the model can then be focused on the important factors.

In factor screening experiments two types of errors can occur: false negative errors and false positive errors. Under certain circumstances, the experiment may indicate a relationship exists when it does not (a false positive result) or it may fail to identify an important relationship that does exist (a false negative result). In this article we are concerned with the potentially more serious problem of false negatives. We show that for a general class of simulation experiments the hypothesis test proposed by Schruben and Cogliano will consistently identify any significant input factors; the asymptotic power of the test is equal to one. This implies that the probability of a false negative error in a frequency domain factor screening simulation experiment goes to zero as the simulation run duration is increased.

### **1. Background and Notation:**

For analysis it is convenient to describe a simulation model in terms of relationships between its input processes  $\{x(t)\}$  and an output process  $\{y(t)\}$ . Here we will assume that the simulation model acts as a general linear filter with additive randomness  $\{\varepsilon(t)\}$  (Brillinger (1981), Chapter 6). Such a filter

can be represented as

$$y(t) = \sum_{u=-\infty}^{\infty} g(u)x(t-u) + \varepsilon(t) \quad (1)$$

where

$\{y(t)\}$  is a real-valued stochastic series,

$\{x(t)\}$  is a real-valued deterministic series,

$\{g(u)\}$  is an impulse response (memory weighting) function,

and

$\{\varepsilon(t)\}$  is a stationary stochastic series with mean zero.

If  $\{x(t)\}$  satisfies the quasi-stationary assumptions given in Ljung (1987) on page 27, then (1) can be expressed in the frequency domain as

$$f_y(\omega) = |\Gamma(\omega)|^2 f_x(\omega) + f_\varepsilon(\omega) \quad (2)$$

where

$f_\varepsilon(\omega)$  is the power spectrum of  $\{\varepsilon(t)\}$

$f_y(\omega)$  is the power spectrum of  $\{y(t)\}$

$f_x(\omega)$  is the power spectrum of  $\{x(t)\}$

and

$$\Gamma(\omega) \equiv \sum_{u=-\infty}^{\infty} g(u)e^{-iu\omega}$$

Note that even though the model given in (1) has a very simple form, it is sufficient for our purposes. Schruben and Cogliano (1987) consider a multivariate linear model for which the results in this article are still valid since all of the spectral estimators considered here are asymptotically independent at different values of  $\omega$ . They also point out that polynomial functions of the input variables are equivalent in the frequency domain as sets of linear inputs; they refer to these terms as pseudo-linear inputs.

Serious simulation models typically involve a great many input variables. Efficient identification of those variables that have a truly significant

influence on the output is the motivation for running frequency domain simulation experiments. In a straightforward frequency domain simulation experiment, input factors and parameters of the simulation model are oscillated at distinct frequencies during a run. The selection of these input *driving* frequencies must be done carefully (see Jacobson, Buss, and Schruben (1986)). The sample power spectrum of the output process,  $\{y(t)\}$ , is examined for the presence of sinusoidal components at input driving frequencies. The strength of such signals are used to infer the relative importance of each of the input factors in the simulation model. Only two runs of the simulation (discussed below) are typically required in a frequency domain experiment to screen a large number of input factors. This is in contrast with conventional simulation experiments where factors are constant for each run and additional runs are required to test the influence of each factor. See Sanchez and Schruben (1987) for a detailed example of a frequency domain simulation factor screening experiment. See, in addition, Schruben (1986) for extensions of the method where estimation and optimization are the experimental objectives.

If we assume the model given in (1) and oscillate  $\{x(t)\}$  according to

$$x(t) = \alpha \cos(\omega_0 t + \varphi)$$

where

$\alpha$  is the amplitude

$\omega_0$  is the angular frequency ( radians per unit time )

$\varphi$  is a phase shift

Then

$$f_x(\omega) = (\alpha / 4)^2 (\delta(\omega - \omega_0) + \delta(\omega + \omega_0))$$

where

$\delta(\ )$  is the Dirac  $\delta$  - function (see Chatfield (1984), Appendix 2 )

From (2), the power spectrum of  $\{y(t)\}$  is

$$f_y(\omega) = |\Gamma(\omega)|^2 (\alpha/4)^2 (\delta(\omega - \omega_0) + \delta(\omega + \omega_0)) + f_\varepsilon(\omega) \quad (3)$$

Hence, if the input series is related to the output series by (1), then the output power spectrum should illustrate the presence of periodicities at  $\pm\omega_0$  in the output time series.

## 2. The Asymptotic Power of the Test:

Certain complications arise in frequency domain experiments. Expression (3) illustrates that the presence of periodicities at frequency  $\pm\omega_0$  may be masked by memory,  $|\Gamma(\omega)|^2$ , and random noise,  $f_\varepsilon(\omega)$ . If  $|\Gamma(\omega_0)|^2 = 0$  then the periodicity at  $\omega_0$  in the output spectrum is eliminated. Since  $\{\varepsilon(t)\}$  is a random noise process, hence a realization of  $f_\varepsilon(\omega)$  may mask the presence of a periodicity at  $\omega_0$ . The result would be a false negative error; the experimenter would wrongly conclude that an important input factor was not significant. Clearly in a factor screening experiment we would like to insure that this serious error will not occur. We demonstrate in this section if the experiment is run long enough the probability of this type of error goes to zero.

The function  $\Gamma$  is typically unknown. We assume that  $\Gamma$  belongs to a family of functions such that the roots of  $\Gamma$  are isolated and hence countable. A class of such functions includes analytic functions; this is discussed further in Section 3. If it is further assumed that the roots of  $\Gamma(\omega)$  are distributed according to a continuous probability measure, then the probability of generating a root which coincides with at most one of the term indicator frequencies is zero since the set of driving frequencies is finite. If we

assume that  $\Gamma(\omega)$  belongs to such a family of functions, then for any fixed  $\omega$ ,  $\omega_0$  say, then  $\Gamma(\omega_0) \neq 0$  w.p. 1 (see Section 3 for more details).

In order to overcome the potential problem of masking due to possible dependencies in the random noise (i.e., non-constant  $f_\varepsilon(\omega)$ ), the experiment is run twice. The first run (called a control run) is a conventional simulation run where the values of all the input factors are held fixed throughout the run; no sinusoidal components are added. The output power spectrum for the control run (denoted as  $\hat{f}_y^c(\omega)$ ) from (3) is given by

$$\hat{f}_y^c(\omega) = f_\varepsilon(\omega) \quad (4)$$

In the second run (called the spectral run), the input process is oscillated sinusodally and the resulting output power spectrum (denoted as  $\hat{f}_y^s(\omega)$ ) is given by (3). Corresponding superscripts will denote the cross-spectra and input spectra for the signal and the control run. Spectral estimators will be denoted with a  $\hat{\cdot}$ ; the class of estimators and their properties is presented in Brillinger, section 6.5. We can now set up the following hypothesis:

$$H_0: \hat{f}_y^s(\omega) = f_\varepsilon(\omega)$$

$$H_a: \hat{f}_y^s(\omega) = |\Gamma(\omega)|^2 (\alpha/4)^2 (\delta(\omega - \omega_0) + \delta(\omega + \omega_0)) + f_\varepsilon(\omega).$$

$H_0$  says that oscillation of the the input factor,  $\{x(t)\}$ , has no effect on  $\{y(t)\}$ .  $H_a$  says that  $\{x(t)\}$  has an effect on  $\{y(t)\}$  and that this is apparent by the presence of a periodicities at frequencies  $\pm\omega_0$ . Assuming that  $H_0$  is true, the test statistic proposed by Schruben and Cogliano (1987) is

$$\frac{\hat{f}_y^s(\omega)}{\hat{f}_y^c(\omega)} \stackrel{D}{\approx} F_{\nu\nu} \quad (5)$$

('  $\stackrel{D}{\approx}$  ' means 'approximately equal in distribution').

Here  $F_{\nu\nu}$  has a F-distribution with  $\nu$  and  $\nu$  degrees of freedom. The degrees of

freedom depend on the spectral window used in the estimation of  $\hat{f}_y^c(\omega)$  and  $\hat{f}_y^s(\omega)$ .

If  $H_0$  is false then the numerator of (5) is not estimating  $f_\varepsilon(\omega)$  but something larger, hence the objective is to reject  $H_0$  if the observed value of (5) is "too large". In order to determine what is "too large", an F-table may be used since the statistic in (5) has an approximate F-distribution. It is desirable to examine the asymptotic properties of (5) and determine the power of the test as the sample size goes to infinity.

Under  $H_0$ , both  $\hat{f}_y^c(\omega)$  and  $\hat{f}_y^s(\omega)$  are consistent and asymptotically unbiased estimators of  $f_\varepsilon(\omega)$  (this follows from the window used in the construction of the estimates - see Brillinger (1981), page 149). Define  $M$  as the truncation point of the spectral window and  $N$  as the sample size. If  $M \rightarrow \infty$  as  $N \rightarrow \infty$  in such a way that  $M/N \rightarrow 0$ , then (Brillinger (1981), page 149),

$$\lim_{n \rightarrow \infty} E |\hat{f}_\varepsilon(\omega) - f_\varepsilon(\omega)|^2 = 0 \quad (6)$$

(where  $\hat{f}_\varepsilon(\omega) = \hat{f}_y^c(\omega) = \hat{f}_y^s(\omega)$ ).

Since convergence in mean square implies convergence in probability then from (6), we get

$$\hat{f}_\varepsilon(\omega) \xrightarrow{P} f_\varepsilon(\omega)$$

(where ' $\xrightarrow{P}$ ' means 'goes in probability to').

Hence,

$$\frac{\hat{f}_y^s(\omega)}{\hat{f}_y^c(\omega)} \xrightarrow{P} 1 \quad \text{under } H_0.$$

Under  $H_a$ , (6) implies that

$$\hat{f}_y^c(\omega) \xrightarrow{P} f_\varepsilon(\omega).$$

Also,

$$\lim_{n \rightarrow \infty} E |\hat{f}_y^s(\omega) - [|\Gamma(\omega)|^2 (\alpha/4)^2 (\delta(\omega - \omega_0) + \delta(\omega + \omega_0)) + f_\varepsilon(\omega)]|^2 = 0. \quad (7)$$

Sections 6.5 and 6.6 in Brillinger (1981) show that an asymptotically unbiased and consistent estimator for  $\Gamma(\omega)$  is

$$\hat{\Gamma}(\omega) = \hat{f}_{yx}^s(\omega) (\hat{f}_x^s(\omega))^{-1} \quad (8)$$

where  $\hat{f}_{yx}^s(\omega)$  is a cross-spectrum estimator for  $f_{yx}^s(\omega)$ .

Also, an asymptotically unbiased and consistent estimator of  $f_\varepsilon(\omega)$  is,

$$\hat{f}(\omega) = \hat{f}_y^s(\omega) - \hat{f}_{yx}^s(\omega) (\hat{f}_x^s(\omega))^{-1} \hat{f}_{xy}^s(\omega) \quad (9)$$

where

$\hat{f}_{xy}^s(\omega)$  is the complex conjugate of  $\hat{f}_{yx}^s(\omega)$ .

It then follows from (8) and (9) that

$$\hat{f}_y^s(\omega) = |\hat{\Gamma}(\omega)|^2 \hat{f}_x^s(\omega) + \hat{f}_\varepsilon(\omega)$$

is an asymptotically unbiased and consistent estimator for  $f_y^s(\omega)$  under  $H_a$ , which justifies the result given in (7). Again using the result that convergence in mean square implies convergence in probability, we have

$$\hat{f}_y^s(\omega) \xrightarrow{p} |\Gamma(\omega)|^2 (\alpha/4)^2 (\delta(\omega-\omega_0) + \delta(\omega+\omega_0)) + f_\varepsilon(\omega).$$

Also, since  $\hat{f}_y^c(\omega) \xrightarrow{p} f_\varepsilon(\omega)$ , then

$$\frac{\hat{f}_y^s(\omega)}{\hat{f}_y^c(\omega)} \xrightarrow{p} \frac{|\Gamma(\omega)|^2 (\alpha/4)^2 (\delta(\omega-\omega_0) + \delta(\omega+\omega_0))}{f_\varepsilon(\omega)} + 1.$$

Define the hypothesis test outcome indicator function as,

$$\hat{I}_{[\text{sig}]} = 1 \text{ if } \frac{\hat{f}_y^s(\omega)}{\hat{f}_y^c(\omega)} > F_{\nu\nu;\alpha} \text{ and 0 otherwise.} \quad (10)$$

As  $N \rightarrow \infty$ , then  $\nu \rightarrow \infty$ , which in turn implies  $F_{\nu\nu;\alpha} \rightarrow 1$ . If  $H_0$  is false, then



$$\frac{\hat{f}_y^s(\omega)}{\hat{f}_y^c(\omega)} \xrightarrow{P} \frac{|\Gamma(\omega)|^2 (\alpha/4)^2 (\delta(\omega - \omega_0) + \delta(\omega + \omega_0))}{f_\varepsilon(\omega)} + 1 > 1.$$

Hence from (10)

$$\hat{I}_{[\text{sig}]} \xrightarrow{P} 1. \quad (11)$$

By definition, the power of the test =  $E[\hat{I}_{[\text{sig}]}]$ . From (10) and (11), we get

$$E[\hat{I}_{[\text{sig}]}] \rightarrow 1 \text{ as } N \rightarrow \infty,$$

i.e., the power of the test goes to 1 as  $N \rightarrow \infty$ .

### 3. The Effect of Aliasing:

The results of Section 2 are based on the assumption that the transfer function is non-zero almost everywhere. When the output is discrete the effects of aliasing (Priestley (1981), page 224) must be considered. We will now show that the class of functions which satisfy our requirements include analytic functions. The set of analytic functions include, for example, polynomials, rational functions and exponential functions. Furthermore, we show that aliasing does not change our conclusions concerning the asymptotic power of the test studied in this article.

Theorem 3.1 (Kaplun (1966), page 53)

The zeros of an analytic function are isolated, unless the function is identically zero, ie, if  $f(z)$  is analytic on its domain, then for each zero  $z_0$  of  $f(z)$  there exists a deleted neighborhood of  $z_0$  in which  $f(z) \neq 0$ .

Proposition 3.2

The zeros of an analytic function are countable in the domain of the function,

which is a subset of  $\mathcal{R}$  (where  $\mathcal{R} \equiv$  Real numbers).

Proof:

Since the zeros are isolated in the domain, there exists a deleted neighborhood of each zero that does not contain any zeros of  $f$ . Furthermore the deleted neighborhood of each zero can be chosen small enough so that they are all disjoint (by shrinking the deleted neighborhood to eliminate overlap, no roots are lost since, by definition, the deleted neighborhoods do not contain any roots). Since the deleted neighborhoods are disjoint, we may pick a unique rational number from each one (which is possible by the density of the rationals). Since the rationals are countable, then so are the deleted neighborhoods and hence the roots of the function are countable.

The motivation for the next theorem comes from the fact that in discrete sampling, aliasing occurs. In a real-valued process, all frequencies  $\omega$  (here  $\omega$  is measured in units of cycles per observation) are mapped back to the interval  $[0, 0.5]$ . Therefore, the zeros of the transfer function  $\Gamma$  are mapped from the interval  $(-\infty, \infty)$  to the interval  $[0, 0.5]$ . Theorem 3.3 shows that the resulting transfer function on the interval  $[0, 0.5]$  is non-zero almost everywhere if  $\Gamma$  is non-zero almost everywhere on  $\mathcal{R}$ .

Theorem 3.3

If the transfer function  $\Gamma$  for a continuous real valued process is analytic on  $(-\infty, \infty)$ , then the alias of  $\Gamma$ , resulting from discrete sampling and defined on  $[0, 0.5]$ , is non-zero almost everywhere.

Proof:

Consider subdividing  $\mathcal{R}$  into the following intervals,

$$\dots[-2,-1.5][-1.5,-1][-1,-0.5][-0.5,0][0,0.5][0.5,1][1,1.5]\dots$$

The alias mapping is defined as  $A:\mathcal{R} \rightarrow [0,0.5]$  to be

$$A(\omega) = |\omega| \quad \text{if } 0 \leq |\omega| \leq 0.5$$

$$= 1 - |\omega| \quad \text{if } 0.5 < |\omega| \leq 1$$

$$= |\omega - [\omega]| \quad \text{if } |\omega| > 1$$

Hence  $A$  is a many to one mapping of  $\mathcal{R}$  to  $[0,0.5]$ , but  $A$  is a one-to-one mapping of any one of the defined intervals onto  $[0,0.5]$ .

Consider an arbitrary interval  $I_\alpha$ , then  $A:I_\alpha \rightarrow [0,0.5]$  is one-to-one and onto. Let  $K_\alpha$  be the set of zeros of  $\Gamma$  contained in  $I_\alpha$  and define  $L_\alpha$  such that  $A(K_\alpha) = L_\alpha$ .  $K_\alpha$  is countable. We next show that  $L_\alpha$  is countable.

Since  $K_\alpha$  is countable there exists a one-to-one, onto mapping  $B:\mathcal{Q} \rightarrow K_\alpha$  (where  $\mathcal{Q} \equiv$  Rational numbers), hence the composition  $A \circ B:\mathcal{Q} \rightarrow L_\alpha$  has the following properties:

1) *one-to-one* since

$$A \circ B(r_1) = A \circ B(r_2),$$

$$B(r_1) = B(r_2) \quad (\text{since } A \text{ is one-to-one), and}$$

$$r_1 = r_2 \quad (\text{since } B \text{ is one-to-one}).$$

2) onto since

$$l = A(k) \text{ for } l \in L_\alpha \text{ and} \\ \text{for some } k \in K_\alpha \text{ (since } A \text{ is onto),}$$

$$k = B(r) \text{ for some } r \in \mathcal{Q} \quad (\text{since } B \text{ is onto), and}$$

$$l = A \circ B(r).$$

Since  $A \circ B$  is one-to-one and onto, therefore  $L_\alpha$  is countable.

By construction  $\mathcal{R} = \bigcup_{n=-\infty}^{\infty} I_n$ . For each  $I_n$ , the set  $L_n$  is the range of the alias mapping of zeros contained in  $I_n$ . Hence  $\bigcup_{n=-\infty}^{\infty} L_n$  is the range of the alias mapping of the zeros in  $\mathcal{R}$ . This is a countable union of countable of countable sets, which is countable. Hence the zeros of the alias of  $\Gamma$ , defined on the set  $[0, 0.5]$ , are countable and therefore belong to a null set in  $[0, 0.5]$ .

#### 4. Summary:

In this article we have shown that the probability of a false negative error occurring in a frequency domain simulation factor screening experiment goes to zero as the run length is increased. In factor screening experiments a false negative error (missing something significant) is of primary importance. False positives (a much less serious error) can and do occur although some effort must be expended to construct such situations.

### Bibliography

1. Brillinger, D.R. Time Series: Data Analysis and Theory. (Expanded Edition). Holden-Day Inc, San Francisco, 1981.
2. Chatfield, C. The Analysis of Time Series: An Introduction. (Third Edition). Chapman and Hall, New York, 1984.
3. Jacobson, S., Schruben, L., and Buss, A., The Frequency Selection Problem in Frequency Domain Simulation Experiments, Tech. Rpt. 714, SORIE, Cornell University, Ithaca, NY, August, 1986.
4. Kaplan, W. Introduction to Analytic Functions. Addison-Wesley, Reading Mass., 1966.
5. Ljung, L. System Identification: Theory for the User. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1987.
6. Priestly, M.B. Spectral Analysis and Time Series. Volume I: Univariate Series. Academic Press Inc., London, 1981.
7. Sanchez, P.J., and Schruben, L.W., Simulation Factor Screening using Frequency Domain Methodology: An Illustrative Example, (submitted for publication) 1987.
8. Schruben, L.W., and Cogliano, V.J. An Experimental Procedure for Simulation Response Surface Model Identification. CACM, Vol.30, No.8 (August 1987).

9. Schruben, L., Simulation Optimization Using Frequency Domain Methods, Proc. 1986 Winter Simulation Conf.. (J.R.Wilson, J.O.Henriksen, and S.D.Roberts, Eds.), pp. 366-369. 1986.