SAND98-2466C

RECEIVED
NOV 17 1998
OSTI

Statistical Tests of System Linearity Based on the Method of Surrogate Data

Thomas Paez
Experimental Structural
Dynamics Dept., 9119
Mail Stop 0557
Sandia National Laboratories
Albuquerque, NM 87185-0557

Norman Hunter
Mechanical Testing
Engineering Sciences and
Applications Division
Mail Stop C-931
Los Alamos National Laboratory
Los Alamos, New Mexico 87545

John Red-Horse
Structural Dynamics
Department, 9234
Mail Stop 0439
Sandia National Laboratories
Albuquerque, NM 87185-0439

ABSTRACT

When dealing with measured data from dynamic systems we often make the tacit assumption that the data are generated by linear dynamics. While some systematic tests for linearity and determinism are available - for example the coherence function, the probability density function, and the bispectrum - further tests that quantify the existence and the degree of nonlinearity are clearly needed. In this paper we demonstrate a statistical test for the nonlinearity exhibited by a dynamic system excited by Gaussian random noise. We perform the usual division of the input and response time series data into blocks as required by the Welch method of spectrum estimation and search for significant relationships between a given input frequency and response at harmonics of the selected input frequency. We argue that systematic tests based on the recently developed statistical method of surrogate data readily detect significant nonlinear relationships. The paper elucidates the method of surrogate data. Typical results are illustrated for a linear single degree-of-freedom system and for a system with polynomial stiffness nonlinearity.

NOMENCLATURE

Δt	sampling interval
α, β	parameters of the nonlinear system
ω_n	resonant frequency (rad/sec)
$E[.]$	denotes expected value operation
F	force input
$G_{XX}(f)$	autospectral density of input signal
$H(f)$	equivalent linear transfer function
Q	intermediate variable – fn of input, output DFTs
$X(f)$	Fourier transform of input time series
$Y(f)$	Fourier transform of accel. response time series
Z	standard normal random variable
c	damping coefficient
f_n	resonant frequency (Hz)
k	stiffness
n	number pts per discrete Fourier transform block
$\ddot{x}, \dot{x}, x$	system acceleration, velocity, displacement

1. INTRODUCTION

The standard engineering measure for modeling experimental data from vibration systems is the transfer function, which is derived from a spectral ratio. Further processing of a set of transfer functions yields modal frequencies and mode shapes. Quite sophisticated modal models are developed for many systems. The accuracy of these models rests on the fundamental assumptions underlying transfer function computation. One of the most critical assumptions is that of linearity. For a nonlinear system the transfer function model yields the best "average" linear properties of the structure. These linear properties may or may not be sufficiently accurate to model system behavior, but in general, more complex and sophisticated models depend heavily on linearity. With increasing model complexity, detection of nonlinear behavior in vibration systems becomes increasingly important.

Numerous methods are available for the detection of nonlinearity including use of higher order spectra (Nikias, 1987); time series analysis, (Hunter, 1997); and use of the coherence function, Wirsching, Paez, and Ortiz (1995). Each method has advantages. One limitation is the heuristic manner in which these methods are often applied. For example, we review a coherence function and decide that the system is "fairly linear" or "clearly nonlinear." For

DISCLAIMER

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

basic modeling purposes this sort of judgement may be sufficient, but for sophisticated models, that depend on subtle data features, a systematic method is required.

In this paper we introduce a systematic method of detecting nonlinearity from the Fourier transform of blocks of the time series response. The basic question of nonlinearity is the focus. The method can readily be adapted to the systematic detection of other, more subtle features of the data. For example, a variation of this approach would allow validation of complex, higher frequency mode shapes in a sophisticated experimental model. Small nonlinearities in data may have surprisingly large effects on complex model structures. For such structures, systematic experimental validation is crucial. The development of this systematic approach follows.

In Section 2, the Method of Surrogate Data, basic arguments are developed for construction of surrogate data sets. The concept of statistical hypothesis testing is introduced. A step by step surrogate data procedure is delineated. Section 3 develops a linearity test using the definition of the transfer function. The test is based on considerations of expected interactions between a frequency and its harmonics. The distribution of a quantity that involves excitation and response Fourier transforms, a transfer function, and the response spectral density, is used to judge system linearity in the method of surrogate data framework. Section 4, Test Systems, describes the two systems tested, a linear single degree of freedom oscillator and an oscillator with a polynomial nonlinearity, and tests their linearities in an exercise of the current technique. Finally, conclusions are offered in assessment of the technique.

2. THE METHOD OF SURROGATE DATA

The method of surrogate data is a technique for testing hypotheses regarding measured random data and the physical systems from which they emanate. It has been developed in a sequence of papers by Tsay (1992) and Theiler, et al. (1991, 1992, 1992). The method of surrogate data can use the bootstrap, developed and described by Efron (1979) and Efron and Tibshirani (1993), or the Monte Carlo method in hypothesis testing. Some fundamental aspects of the technique are discussed here. The problem considered is the testing of the statistical hypothesis that a mechanical system excited by Gaussian excitation, and whose response is measured, is linear. For simplicity the discussion will be presented for the case of a single measured response.

Application of the method of surrogate data requires execution of the following steps.

1. Measure random data. In the present application it is assumed that a mechanical system is excited at a single point, and the system response at a single point is measured. The excitation is assumed to be a zero mean, normal random process. The characteristics of the response will be used to judge the system linearity.
2. Specify a hypothesis concerning the random source. The hypothesis must be specified in sufficient detail to generate data that satisfy the hypothesis. These are the surrogate data.
3. Specify statistics (measures) of the data (measured or generated surrogate data) that may distinguish measured from generated data if the hypothesis is not satisfied by the measured data.
4. Generate n realizations of data from the hypothetical source.
   - Use Monte Carlo approach, if possible.
   - Use the bootstrap when Monte Carlo is not feasible.
5. Compute data statistics (from Step 3 above) for each realization of the generated data.
6. Approximate the sampling probability distribution of the data statistics using, for example, the kernel density estimator (Silverman, 1986). Infer the $(1-\alpha)\times 100$ percent confidence region for the joint realization of the data statistics.
7. Compute data statistics from measured data. Observe where joint realization of statistics falls relative to $(1-\alpha)\times 100$ percent confidence region.
   - If measured-data-statistics fall outside confidence region, then reject the hypothesis at the α level of significance.
   - If measured-data-statistics fall inside confidence region, then do not reject hypothesis. (Strictly speaking, we do not "accept" the hypothesis because other statistics may lead to rejection. However, in a practical sense our confidence in a model may be augmented if we consider the statistical criteria to be well chosen and robust.)

Implementation of the method of surrogate data is completely arbitrary in the sense that the data measured, the hypothesis specified, and the statistics chosen to test the hypothesis are completely arbitrary. However, a key point is that the use of the Monte Carlo method or bootstrap to generate data realizations and the use of the kernel density estimator to characterize the sampling distribution of the data statistics make it possible to consider complicated measures of behavior in judging the measured data. For the present application this means that any reasonable measure of system excitation and response can be used to judge system linearity.

3. TEST OF SYSTEM LINEARITY

Though the method of surrogate data permits the testing of arbitrary hypotheses from the very specific to the very general, our objective here is to identify a test that can be used on a wide variety of systems to judge their linearity. The test of linearity specified assumes that a single wide-band input excites a mechanical system and that a single output is measured. Spectral statistics of the measured signals are used to judge the linearity of the system. In particular, spectral statistics that relate to the generation of harmonics in the responses of nonlinear systems are to be considered. Linear systems do not generate harmonics. Therefore, our scheme will be to (1) develop an hypothesis that the measured data come from a linear system, (2) establish the sampling distribution of spectral statistics that come from a linear system, then (3) test the estimated statistics of the measured data to determine whether or not they can be rejected as realizations of the sampling distribution related to the linear system data. The responses of a linear system excited by a single input are governed by

$$Y(f) = H(f)X(f) \qquad -\infty < f < \infty \tag{1}$$

where $Y(f)$ is the Fourier transform of the response, $H(f)$ is the system frequency response function, $X(f)$ is the Fourier transform of the excitation, and f denotes cyclic frequency. Another relation can be developed from (1) by evaluating the expression at a frequency $3f$, multiplying both sides by the complex conjugate of $X(f)$, and dividing through by $H(3f)$. The result is

$$\begin{aligned}\frac{Y(3f)X^*(f)}{H(3f)} &= X(3f)X^*(f)\\ &= X_R(3f)X_R(f) + X_I(3f)X_I(f) +\\ &\quad i[-X_R(3f)X_I(f) + X_I(3f)X_R(f)]\end{aligned} \tag{2}$$

where i is the imaginary unit and, on the second line, each complex quantity on the right-hand side has been replaced by its real and imaginary parts.

For discrete signals the continuous Fourier transform is replaced by the discrete Fourier transform (DFT). The n-point DFT corresponding to $X(f)$ is $X(f_k), k = 0,\ldots,n/2$. (The DFT is actually defined at n points, but half the information is redundant.) When the input excitation is a zero-mean, stationary random process with one-sided spectral density $G_{XX}(f), -\infty < f < \infty$, then an approximate relation between the spectral density and the DFT is

$$G_{XX}(f_k) = \frac{2\Delta t}{n} E\left[|X(f_k)|^2\right] \qquad k = 0,\ldots,n/2 \tag{3}$$

where Δt is the time increment, and $E[.]$ denotes the operation of mathematical expectation. When the excitation is a zero-mean, normally distributed random process, then the real and imaginary parts of the DFT, $X(f_k), k = 0,\ldots,n/2$, form zero-mean, normally distributed random processes. When the excitation is a wide-band random process, then the following correlations between random variable pairs from the DFT random processes hold (asymptotically).

$$\begin{aligned}&E[X_R(f_k)X_I(f_k)] \cong 0\\ &E[X_R(f_k)X_R(f_m)] \cong 0 \qquad k \neq m\\ &E[X_R(f_k)X_I(f_m)] \cong 0 \qquad k \neq m\\ &E[X_I(f_k)X_I(f_m)] \cong 0 \qquad k \neq m\end{aligned} \tag{4}$$

Further, the variances of the real and imaginary parts of the DFT random processes are

$$E\left[X_R^2(f_k)\right] \cong E\left[X_I^2(f_k)\right] \cong \frac{n}{4\Delta t} G_{XX}(f_k) \qquad k = 0,\ldots,n/2 \tag{5}$$

In view of these facts and accepting the assumptions offered above, we can divide both sides of Eq. (2) by the standard deviation of $X_R(f_k)$ (or $X_I(f_k)$) and the standard deviation of $X_R(f_{3k})$ (or $X_I(f_{3k})$) to obtain

$$\frac{Y(f_{3k})X^*(f_k)}{H(f_{3k})\frac{n}{4\Delta t}\sqrt{G_{XX}(f_k)G_{XX}(f_{3k})}} = Z_1Z_3 + Z_2Z_4 + i[-Z_1Z_4 + Z_2Z_3] \tag{6}$$

where $Z_j, j = 1,2,3,4$ are uncorrelated, standard normal random variables. Each of the Zs corresponds to a normalized real or imaginary part of $X(3f)$ or $X(f)$ in Eq. (2). The index k is limited to $n/6$ here because the index three times k cannot exceed $n/2$. Based on the same reasoning and assumptions as above, we can also write

$$\frac{|X(f_k)|^2}{\frac{n}{4\Delta t}G_{XX}(f_k)} = Z_3^2 + Z_4^2 \qquad k = 0,\ldots,n/2 \tag{7}$$

The relation in Eq. (6) holds for any pair of frequencies f_k and f_{3k} (or, in fact, for any pair of frequencies f_k and f_m, where $k \neq m$); however, it is particularly difficult for a system to satisfy Eq. (6) when f_k is a modal frequency. The quantities on the left-hand side of Eqs. (6) and (7) are relatively easy to estimate from measured excitation and response data, so we use the relations in Eqs. (6) and (7) to develop a surrogate data linearity test. The fundamental concepts underlying the developments summarized in Eqs.

(6) and (7) are presented in Wirsching, Paez, and Ortiz (1995).

The surrogate data generated come from a source whose magnitude is equal to the right-hand side of Eq. (6) divided by the right-hand side of Eq. (7). This quantity is denoted Q_Y, and we plot these surrogate data as a function of the right-hand side of Eq. (7). We denote this quantity Q_X. Two thousand realizations of these quantities were generated and are shown with points in Figure 1.

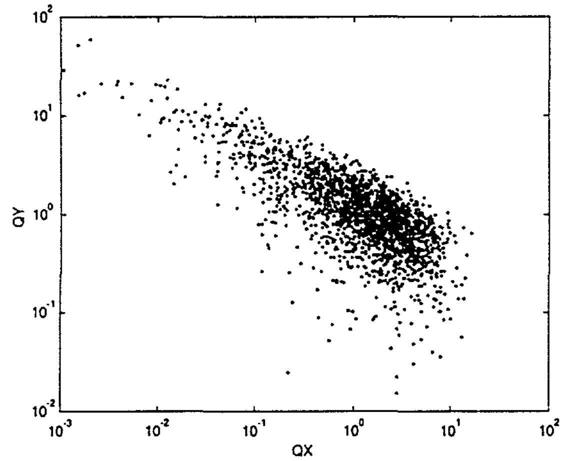


Figure 1. Two thousand surrogate data realizations of Q_Y versus Q_X.

Any mechanical system that is linear and excited by a zero mean, stationary, wide-band random excitation, and for which we compute the quantities on the left-hand sides of Eqs. (6) and (7) and plot

$$Q_Y = \left. \left| \frac{Y(f_{3k})X^*(f_k)}{\hat{H}(f_{3k})\frac{n}{4\Delta t}\sqrt{\hat{G}_{XX}(f_k)\hat{G}_{XX}(f_{3k})}} \right| \middle/ \frac{|X(f_k)|^2}{\frac{n}{4\Delta t}\hat{G}_{XX}(f_k)} \right.$$

$$\text{versus} \quad Q_X = \frac{|X(f_k)|^2}{\frac{n}{4\Delta t}\hat{G}_{XX}(f_k)} \tag{8}$$

where $\hat{G}_{XX}(f)$ is the spectral density estimator for the input random process, and $\hat{H}(f)$ is the frequency response function estimator, should yield a realization that comes from the same random source as the data in Figure 1. Our linearity test compares a function of the left sides of Eqs. (6) and (7) to a collection of realizations of the same function of the right sides of Eqs. (6) and (7).

To simplify the testing of a statistical hypothesis in this framework, we use the logarithms of the data in Figure 1 to form $(1-\alpha)\times 100$ percent upper confidence intervals on Q_Y given a particular value of Q_X. These intervals are plotted for three confidence levels in Figure 2. They were formed by writing the joint kernel density estimator for the logarithms of the data in Figure 1, then using this to estimate the conditional cumulative distribution function (CDF) of Q_Y given Q_X. (The kernel density estimator is described in Silverman, 1986.) Specific percentiles of the conditional CDF were identified via inversion of the approximate CDF, and these are plotted in Figure 2, along with the data upon which they are based.

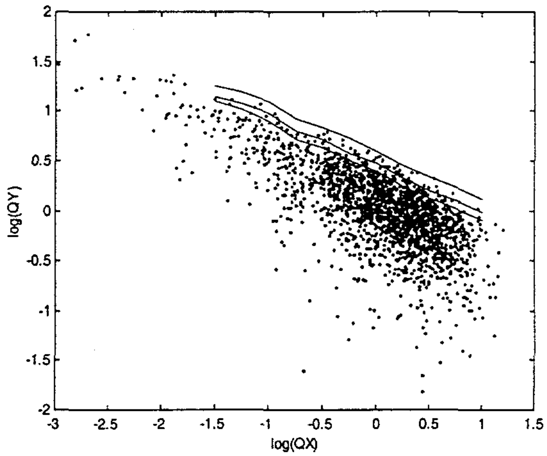


Figure 2. Upper confidence intervals on Q_Y, given Q_X at 90 percent level (lowest curve), 95 percent level (middle curve), and 99 percent level (highest curve).

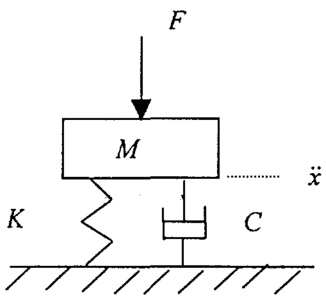
4. TEST SYSTEMS

The method of surrogate data technique for testing the linearity of systems is exercised in this section on two structures. The first is a simple linear system. The second is a nonlinear system whose response is simulated on an analog computer. The specific approach to the performance of the test of linearity hypothesis is demonstrated.

4.1 LINEAR OSCILLATOR

The first system is a linear single degree-of-freedom oscillator. This system is illustrated in Figure 3. The equation for this force-excited oscillator is

$$\ddot{x} + 2\zeta\omega_n\dot{x} + \omega_n^2 x = F/M \qquad \zeta = 0.02, \quad \omega_n = 2\pi(100) \tag{9}$$



Linear Single Degree-of-Freedom System
Figure 3

The parameters C and K are chosen to produce a resonant frequency of 100 Hz with an effective damping of 2%. The system is excited by band limited random process. The effective sample rate is 1000 Hz. A fourth order Runge-Kutta ODE solver simulates the system. The acceleration response of the mass is used for this analysis.

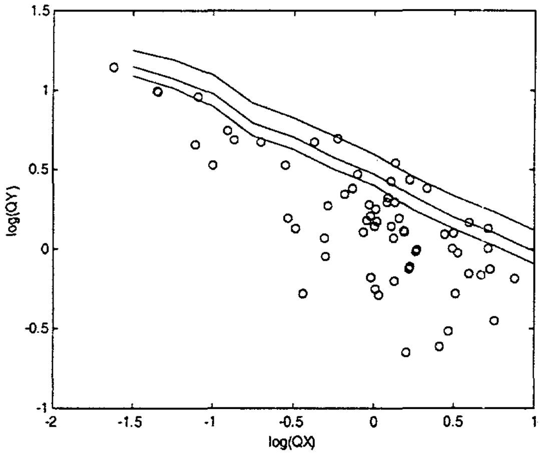
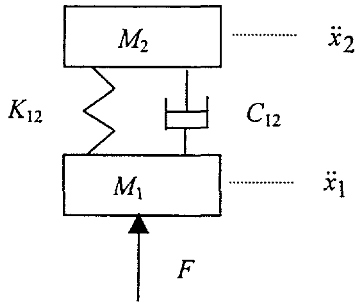


Figure 4. Realizations of Q_Y versus Q_X based on the linear system data and the confidence limits from Figure 2.

The number of points in the generated excitation and response time series used in the linearity assessment was 8192. Each collection of points was separated into 64 blocks of 128 points each. The DFTs of the excitation and response were computed and used to estimate the input autospectral density and the system linear frequency response function. The results were used to form 64 joint realizations of Q_X and Q_Y in Eq. (8) at frequency lines corresponding to f = 100 Hz. These realizations are plotted in Figure 4 along with the confidence limits from Figure 2. The interpretation of confidence limits indicates that approximately αX100 percent of realizations that satisfy a hypothesis will fall outside the (1-α)X100 percent confidence interval, and the remainder will fall inside the interval. Based on this criterion, the linearity hypothesis should not be rejected for these data.

4.2 NONLINEAR OSCILLATOR

The nonlinear oscillator is identical in general form to the linear oscillator of Figure 3, though the system is base excited by the acceleration $\ddot{x}_1$ rather than force excited, as the linear oscillator. The nonlinear hardening oscillator is illustrated in Figure 5.



A Nonlinear Hardening Oscillator
Figure 5

Equation 10 describes the system dynamics in detail.

$$\begin{aligned} &\ddot{x}_2 + 2\zeta\omega_n(\dot{x}_2 - \dot{x}_1) + \omega_n^2(x_2 - x_1) + \\ &\quad \alpha\omega_n^2(x_2 - x_1)^2 + \beta\omega_n^2(x_2 - x_1)|x_2 - x_1| = 0 \\ &\quad \alpha = 3000 \\ &\quad \beta = 3500 \\ &\quad \omega_n = 2\pi(11.5) \\ &\quad \zeta = 0.04 \end{aligned} \tag{10}$$

The $2\zeta\omega_n$ term is equal to the damping coefficient divided by the upper mass, C_{12}/M_2, and the terms that multiply the relative displacement $(x_2 - x_1)$ are equal to the nonlinear stiffness K_{12}. At low relative displacements the quadratic terms are small, and the resonant frequency is close to 100 Hz. At larger relative values of $(x_2 - x_1)$, the nonlinear terms dominate. The effective resonant frequency increases dramatically with increasing test level. The system is simulated using an analog computer. The mean square input acceleration is adjusted to produce approximately equal mean square levels for the linear and combined nonlinear responses. For positive excursions of $(x_2 - x_1)$ the quadratic and absolute value nonlinear terms add, whereas for negative values of $(x_2 - x_1)$, they subtract. Since $\beta > \alpha$ the net stiffness K_{12} increases for both positive and negative relative displacements $(x_2 - x_1)$, but less stiffness increase occurs for negative relative displacements. The system is driven well into the nonlinear range, but responses are not chaotic. Response is sampled at 1000 samples/second for 8192 acceleration response points.

Each collection of points was separated into 64 blocks of 128 points each. The DFTs of the excitation and response were computed and used to estimate the input autospectral density and the system linear equivalent frequency response function. The results were used to form 64 joint realizations of Q_X and Q_Y in Eq. (8) at frequency lines corresponding to $f = 100$ Hz. These realizations are plotted in Figure 6, along with the confidence limits from Figure 2. We interpret the confidence intervals as above to reject the hypothesis that the system is linear.

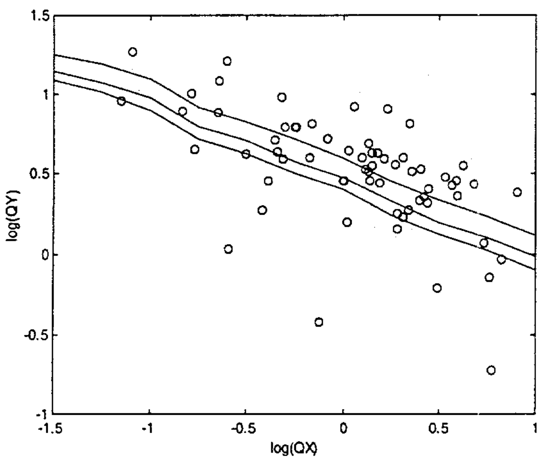


Figure 6. Realizations of Q_Y versus Q_X based on the linear system data and the confidence limits from Figure 2.

CONCLUSIONS

A technique for the assessment of system linearity has been developed. It is based on an analysis of data generated using the method of surrogate data. Its use depends upon the availability of spectral measures of experimentally collected time series of excitation and response. Because these quantities are usually obtained in the analysis of experimental data, it should be a very direct matter to perform the present test. Results from the examples performed here indicate that this approach can reliably judge the linearity of a system. In fact, it may be that the relative location of statistics Q_Y within or outside the various confidence intervals reflects, in a sense, the degree of nonlinearity of a system. Testing of other nonlinear systems may confirm this.

Other statistics like the ones in Eq. (8) can be used, along with the method of surrogate data, to assess not only system linearity but also other system characteristics.

ACKNOWLEDGMENT

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin company, for the United States Department of Energy under Contract DE-AC04-94AL85000. Los Alamos National Laboratory is operated by the University of California for the Department of Energy.

REFERENCES

Efron, B., (1979), "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, 1-26.

Efron, Bradley and Tibshirani, Robert J., (1993), *An Introduction to the Bootstrap*, Applied Monographs on Statistics and Applied Probability 57, Chapman and Hall, 1993.

Hunter, N.F. (1997) State Analysis of Nonlinear Systems Using Local Canonical Variate Analysis, Thirtieth Annual Hawaii Conference on System Sciences, 1997.

Moon, F., (1987), *Chaotic Vibrations*, Wiley, New York.

Nikias, C. L., 1987, and Raghuveer, Bispecturm Estimation: A Digital Signal Processing Framework, Proceedings of the IEEE, 75(7), 869, 1987.

Silverman, B. W., (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall.

Theiler, J., Galdirkian, B., Longtin, A., Eubank, S., Farmer, D., (1991), "Using Surrogate Data to Detect Nonlinearity in Time Series," Technical Report LA-UR-91-2615, Los Alamos National Laboratory, Los Alamos, New Mexico.

Theiler, J., Eubank, S., Longtin, A., Galdirkian, B., Farmer, D., (1992), "Testing for Nonlinearity in Time Series: The Method of Surrogate Data," *Physica D*, 58, North-Holland, pp. 77-94.

Theiler, J., Linsay, P., (1992), "Detecting Nonlinearity in Data with Long Coherence Times," *Proceedings of the NATO Workshop on Comparative Time Series Analysis*, Santa Fe Institute, Santa Fe, New Mexico.

Tsay, R. S., (1992), "Model Checking Via parametric Bootstraps in Time Series Analysis,' *Applied Statistics*, 41, No. 1, pp. 1-15.

Wirsching, P., Paez, T., Ortiz, (1995), *Random Vibrations: Theory and Practice*, Wiley, New York.