Nuclear Mathematical and Computational Sciences: A Century in Review, A Century Anew
Gatlinburg, Tennessee, April 6-11, 2003, on CD-ROM, American Nuclear Society, LaGrange Park, IL (2003)

THE STOCHASTIC PARAMETER SIMULATION SYSTEM: A WAVELET-BASED METHODOLOGY FOR "PERFECT" SIGNAL RECONSTRUCTION

Adrian Miron
Argonne National Laboratory
Reactor Analysis and Engineering Division
9700 S. Cass Ave.
Argonne, Illinois 60439
amiron@anl.gov

John M. Christenson
University of Cincinnati
College of Engineering, Mechanical, Industrial and Nuclear Engineering Department
Cincinnati, Ohio 45221
John.Christenson@uc.edu

ABSTRACT

This paper describes some of the capabilities of a newly developed signal processing tool, the Stochastic Parameter Simulation System (SPSS), designed to decompose any nuclear power plant signal into its deterministic and stochastic component and to reconstruct a simulated signal with the same statistical characteristics as the original signal. SPSS uses wavelet, filter banks and threshold theory to accomplish its design objectives. SPSS results were compared to a base-case methodology, which uses the optimal-order Fourier series approximation. The analyzed signals were real sensor signal outputs, independently recorded at three nuclear power plants. For all of the recorded signals, SPSS produced a better approximation of the original signal than the Fourier series-based procedure. For each signal, SPSS found many wavelet-based decompositions, all of which produced white and normally distributed residuals. In most cases, the Fourier-based analysis failed to completely eliminate the original signal serial-correlation in the residuals. The SPSS can be used to generate large data sets of simulated sensor signals for which the physical data collection would be very expensive. Also, SPSS can be used to increase the reliability of estimation and validation systems that monitor sensor signals and perform fault detection.

Key Words: Wavelets, Signal Processing, Nuclear Power Plant Applications, Stochastic Process, Sensor Estimation and Validation Applications

1. INTRODUCTION

In this paper, the results produced by two systems (Reactor Parameter Signal Simulator (RPSS), developed by Argonne National Laboratory (ANL) [1] and a newly developed system, Stochastic Parameter Simulation System (SPSS) [2]) are compared. Both systems were designed to meet two key functional requirements:

1. To analyze any steady state plant signal, decompose it into its deterministic and stochastic components, and then reconstruct a new, simulated signal that possesses the same statistical noise characteristics as the actual signal; and

2. To be able to filter out the principal serially-correlated, deterministic components from the analyzed signal so that the remaining stochastic signal can be analyzed with signal validation tools that are designed for signals drawn from independent random distributions.

The difference between the two systems is that the RPSS uses Fourier techniques and the SPSS uses wavelet methodology to achieve the above objectives.

An SPSS-type system has the potential for a number of significant applications. An important set of applications involves generating large data sets of simulated sensor signals for which the data collection/gathering of the physical signals would be very expensive. Using the SPSS, only several thousand data points, obtained over a reasonable time interval, can establish the dynamic and statistical structure of the signals. Then, millions of data points, representing a much longer time span and having the same statistical idiosyncrasies as the physical signals, can be generated using SPSS. When the difference between the real and the estimated signal contains serial correlation, the SPSS can be also integrated with estimation and validation systems that monitor sensor signals and perform fault detection [2]. Recently, the SPSS methodology was successfully used for determining the ANL's Multivariate State Estimation Technique (MSET) estimation uncertainty for nuclear power plant applications. Assuming that the sensor noise characteristics are known, the SPSS was used to determine the "true" process value required in uncertainty computations. Here, the SPSS optimization procedure [2] was used for finding the "true" signal deterministic component that produced the assumed noise characteristics.

2. THEORY

The idea behind the RPSS/SPSS is that any steady-state signal, $\boldsymbol{S_t}$, can be considered as a sum of two components: a deterministic component that includes the signal serial-correlation, $\boldsymbol{A_t}$, and some random contribution, $\boldsymbol{e_t}$, so that [3]:

$$S_t = A_t + e_t. \tag{1}$$

If the difference between the original signal, $\boldsymbol{S_t}$, and the deterministic component (approximation), $\boldsymbol{A_t}$, is proven to be white and to have a normal probability distribution function (pdf), the original signal can be reconstructed by adding to the approximation a white and Gaussian randomly generated function. Since the approximation is a deterministic function and the white, Gaussian function is randomly generated, the two functions are independent. Therefore, the randomly generated function is chosen to have the mean 0 and the variance equal to the difference between the variance of the original signal and the one given by the approximation. Consequently, the reconstructed signal has at least the first two moments (mean and variance) identical to the original signal since, as it will be shown here, the deterministic component has the mean very close or equal to the original signal mean.

The second functional requirement described above is easily accomplished by subtracting from the original signal the deterministic component.

In RPSS, the approximation A_t in eq. (1) is seen as a composite function containing k Fourier modes given by the most dominant frequencies w, and e_t, as a discrete function of random **residuals**. Therefore, the signal, S_t in eq. (1), is approximated by:

$$A_t = \frac{a_0}{2} + \sum_{k=1} (a_k \cos(w_k t) + b_k \sin(w_k t)) \quad (2)$$

where $a_0/2$ is the mean value of the signal and a_k and b_k are the Fourier coefficients corresponding to the Fourier frequency w_k. In this research, up to 10 most dominant frequencies, determined by the highest periodogram values are considered [2]. Although no exhaustive study was performed to identify the optimum number of modes for the signal reconstruction, it was noticed that if the number of harmonics is stepwise increased from four to eight, the serial correlation of the residuals is reduced. Also, for the variables studied in reference [1], increasing the number of harmonics from eight to ten introduces nonphysical high frequency oscillations in A_t due to the errors in computing the periodogram of a noisy signal.

RPSS also uses an iterative procedure that optimizes and balances the degrees in reduction in both serial correlation and non-normality by incrementing k from 1 until the optimal composite function is found. This loop continues until a user-defined number for k is reached (10 in this research). The k modes with the largest values of the periodogram function are combined to create the composite function A_t in eq. (2) and during each step of the loop, the residual functions, e_t, are calculated. Several statistical tests are used to determine the whiteness and normality of the residuals. If any of the tests fail, the loop variable is incremented and the loop is reiterated. Once all of the tests are passed, the composite function identified by the algorithm can be finally used to filter out the serially-correlated contamination from the original signal. If the user-defined number for k is reached and none of the residuals thus calculated are white and normally distributed, the Fourier composite that contains most of the original signal serial-correlation is used to approximate the signal in eq. (1) [4].

SPSS uses wavelet-based decomposition techniques to find the deterministic part of the signal. The wavelet and filter banks theories decompose and perfectly reconstruct a signal S_t, in the following form [5]:

$$S_t = A_J + \sum_{i=1}^{J} D_i \quad (3)$$

where A_J is the **wavelet approximation** at the user-defined level of decomposition J and D_i are the signal **details**. The approximation and the details are completely determined by the wavelet approximation and detail coefficients respectively, obtained in the decomposition filter bank. One way of solving the problem is to let the wavelet approximation, A_J, in eq. (3) to be the deterministic part of the signal, A_t, in eq. (1). Then, all the details grouped together represent the functional e_t. One advantage of this approach is that we are dealing only with two parameters:

the level of decomposition, J, and the wavelet family, which is completely determined by the decomposition and reconstruction filter banks. The major disadvantage is that there is no guarantee that all the details grouped together would consist of white and normally distributed residuals.

We can also define a noise model in which a signal S_t, can be seen as [6]:

$$S_t = f_t + \sigma e_t \tag{4}$$

where f_t is the de-noised signal, σ is the noise level and e_t is the standard white and normally distributed noise with mean 0 and variance 1.

By applying the thresholding theory on the detail coefficients obtained in the wavelet decomposition filter bank, the **wavelet-based de-noised signal**, f_t, can be expressed as [6]:

$$f_t = A_J + \sum_{i=1}^{J} D_i' \tag{5}$$

where D_i' are now the new details completely determined by the thresholded coefficients.

We are now letting the deterministic part of the signal, A_t, in eq. (1) to be equal to the wavelet-based de-noised signal, f_t, in eq. (5) through the noise model defined in eq. (4). In this case, our guarantee that the residuals are white and normally distributed is increased if we manage to find the appropriate constant noise level, σ. Although the SPSS calculates both the wavelet approximation and the wavelet-based de-noised signal, it was found that the former produces superior results for finding white and normally distributed residuals and therefore, the wavelet-based de-noise signal was used in the comparison between the RPSS and SPSS results.

Two statistical tests were used to determine the normality (Pearson's χ^2 test [2], [7]) and the whiteness (Bartlett-Durbin procedure [2], [8], [9], [10]) of the analyzed data. In this research, the χ^2 test reports a normality indicator which, if it is negative, the data is assumed to be normally distributed and if it is positive, the data is assumed to have a non-normal pdf [2]. The whiteness test reports a normalized Kolmogorov-Smirnov (K-S) indicator which if it smaller than 1, the data is assumed to be white and if it is greater than 1, the data is assumed to contain serial-correlation [2].

3. RESEARCH PROCEDURE

The primary research objective was to show that the wavelet technique used to approximate a plant process signal can produce better results than the commonly used Fourier approximation. Both SPSS and RPSS are designed to meet the same design requirements as discussed in the first section. Therefore, an appropriate way to decide which system accomplishes its design objectives better is to compare the results when the **same** signals and the **same** statistical indicators, calculated at the **same** significance level are analyzed by the two systems.

This research used real data from three different power plants. The data was recorded at full power, steady state operation at each plant and represents sensor output readings of the plant parameters such as pressure, temperature, flow, level, flux etc., which are continuously monitored by the plant computer for analyzing and controlling the plant state.

Each signal in the three processed data sets was run through the SPSS and RPSS programs. The Pearson's χ^2 indicator for normality and the normalized K-S indicator for whiteness given by the Bartlett–Durbin procedure (see previous section) were used for determining the normality and whiteness of the residuals. The same statistical indicators were used in both optimization procedures of RPSS (as described in previous section) and SPSS programs. The SPSS optimization procedure [2] finds the decomposition parameters for all the wavelet-based approximations that, when subtracted from the original signal, produces residuals that are white and normally distributed, as determined by the statistical tests.

Since the results were so similar to all three sets of signals we discuss only one set of signals in the next section. First, the whiteness and normality of RPSS vs. SPSS residuals are compared and discussed. Then, the mean and variance of the original signal are compared with the mean and variance of the reconstructed signals. Finally, the serial-correlation statistical indicators for the original and reconstructed signals are compared.

4. RESULTS

Using the SPSS optimization function [2] for each analyzed signal, tens if not hundreds of wavelet decompositions were found that produced similar results. Also, the same conclusions are valid for all three sets of analyzed data. Therefore, we pick only one wavelet-based decomposition for each signal and only one set of data (DB data) for illustration.

The first statistical parameters compared with respect to the RPSS and SPSS residuals (F.Res. and W.Res., respectively) were the mean and variance. The analysis shows that the W.Res. variance is usually smaller than F.Res. variance and, as expected, the means are comparable and close to zero. Therefore, both systems provided good estimates for the analyzed signals. A statistical parameter that is a measure of the estimates' quality is the root mean squares (rms) of the residuals. The smaller the residual rms, the more accurate is the estimation. Therefore, the residual rms values for all of the DB signals were plotted in bar graphs to provide a comprehensive illustration of the rms comparison in Figure 1.

All the residual rms values are extremely small when compared with the signal mean (0.09 % in average for SPSS and 0.2 % in average for RPSS), and therefore both systems performed an excellent job in estimating the original signals. However, it can be seen from Figure 1 that, for **all** the analyzed signals, the SPSS residual rms values are smaller than the RPSS residual rms values and therefore the wavelet-based estimation is more accurate than the Fourier-based approximation.

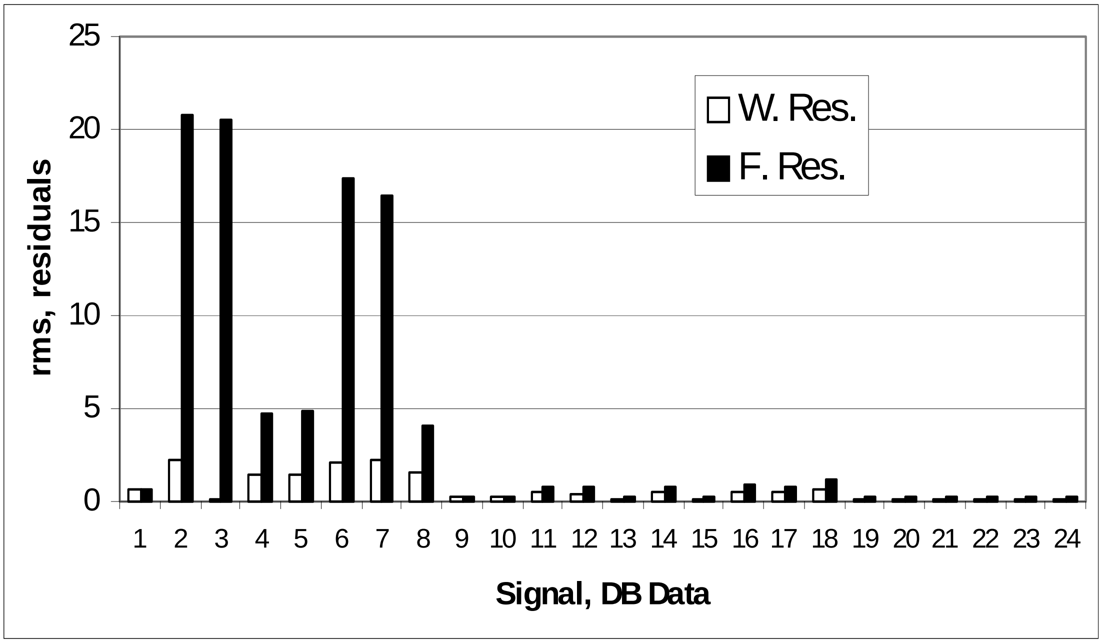

Figure 1 - DB data residual root mean squared (rms) comparison

The RPSS and SPSS residual whiteness and normality parameters are summarized in the bar graph presented in Figure 2. The ordinate bars of the figures display the normalized K-S whiteness indicator of the original signal (in the middle), the K-S indicator of the SPSS residual (at the left) and the K-S indicator of the RPSS residual (at the right). The black dot indicates non-normally distributed residuals according to the Pearson's χ^2 test.

Figure 2 show that usually the RPSS analysis reduces, but does not eliminate the serial-correlation from the analyzed signals. With RPSS analysis, the serial-correlation was completely eliminated according to the K-S indicator (very close to or under one) for 5 out of 44 analyzed signals (see DB signals 1, 6, 8, and 9 in Figure 2). For signals that are white but not normally distributed (DB signal 10), the RPSS analysis managed to normalize the data only for signal # 10 in DB data. Moreover, in 15 out of 44 cases, the RPSS analysis did not normalize the residuals (see the black dot on top of RPSS residuals for signals 2, 4, and 16 in DB data).

As illustrated in Figure 2, according to the normalized K-S indicator for whiteness, the SPSS analysis managed to eliminate the serial-correlation from all analyzed sensors. Also, all the SPSS residuals are found to be normally distributed according to the Pearson's χ^2 tests for normality. Although the figure present the results for only one wavelet-based decomposition, many models that produced similar results were found for each signal considered in this research.

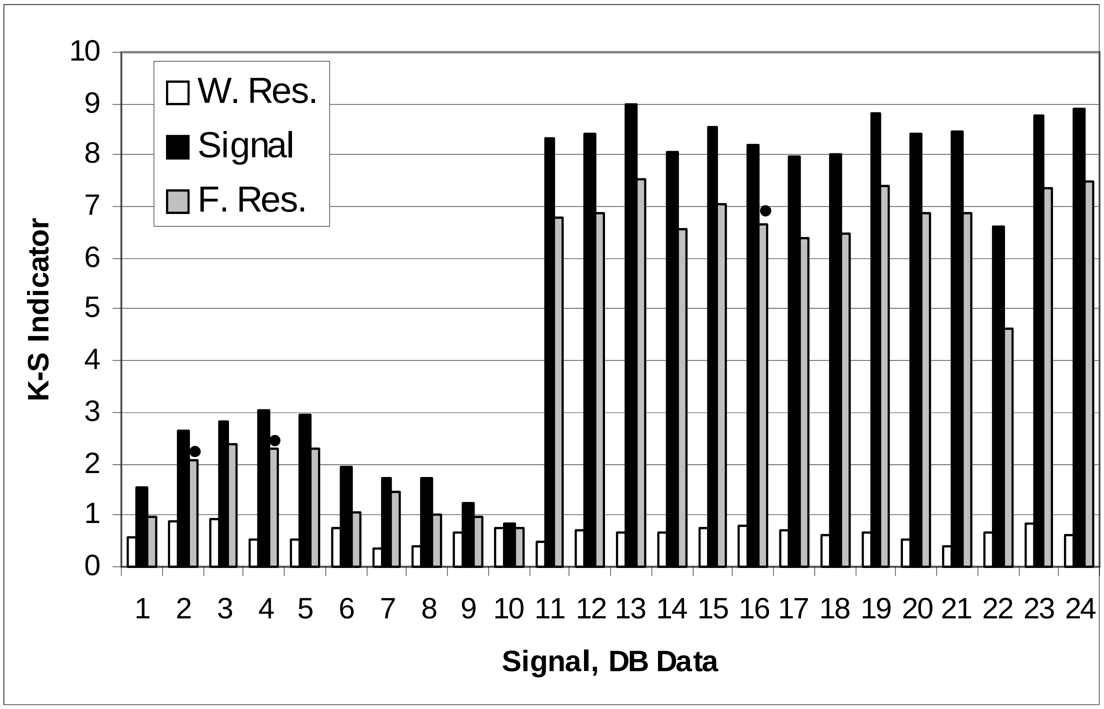

Figure 2 – Whiteness and normality comparison between RPSS and SPSS residual results for DB data.

Next, the statistical parameters of the original signal were analyzed and compared with similar parameters of the RPSS and SPSS reconstructed signals. The analysis showed that both systems reconstructed signals with very close if not identical mean and variance as the original signal. For the SPSS reconstructed signal mean, it was calculated that the average of the relative deviations from the original signal mean is $2\ 10^{-3}$ %. For the RPSS reconstructed signal mean, the average of the relative deviations from the original signal mean was found to be $1\ 10^{-2}$ %. Although there is almost an order of magnitude in favor of the SPSS reconstruction, the numbers are so small that, in terms of signal mean reconstruction comparison, both systems performed very well. In terms of variance, it was found that, for the SPSS reconstructed signal variance, the average of relative deviation from the original signal variance is 2 % and for the RPSS reconstructed signal variance, the same quantity is 6 %. Therefore, in terms of mean and variance, both systems provided excellent results, although overall, the SPSS produced reconstructed signals with closer original signal means and variances than RPSS reconstruction.

Since the mean and variance of the original signals are so well preserved by the reconstructed signals, other statistical indicators were used to compare the original signal with the reconstructed ones. The normalized K-S indicator for whiteness is also an indicator of the degree of serial-correlation in the analyzed data [2]. The larger this indicator, the higher is the degree of serial-correlation in the data. The K-S indicators for the original signal and the reconstructed ones are plotted for each analyzed signal of DB data in Figure 3. The figure shows that, except for DB signal 1, all the K-S indicators of the SPSS reconstructed signals are closer to the original signal K-S indicators than the RPSS reconstructed signal K-S indicators. For DB signal 1, the RPSS reconstructed signal is white, whereas the original signal is not, and therefore, the SPSS did again a better reconstruction than RPSS by preserving the initial serial-correlation of the signal.

Thus, it can be concluded that the SPSS preserved better than RPSS the degree of serial-correlation of the original signal in the reconstructed signals. This result is due to the fact that the SPSS residuals are white and normally distributed, whereas most of the RPSS residuals are not. Although the residual non-normality does not influence the RPSS reconstruction [4], the fact that the residuals are not white affects the signal reconstruction.

Since the SPSS residuals are white, the wavelet-based approximation of the original signal contains most of the original signal serial-correlation. The Fourier approximation contains only a part of the original serial correlation, the rest being contained in the RPSS residuals. In the reconstruction process a white noise is added to the approximations. Therefore, in terms of preserving serial-correlation in the original signal, better results are to be expected for the SPSS reconstruction since the wavelet-based approximations contain almost all of the original serial-correlation. Signals with a smaller degree of serial-correlation than the original ones are expected in the RPSS reconstruction since a part of serial correlation is lost in the RPSS residuals. This conclusion can be clearly seen in Figure 3 in which the normalized K-S indicator for RPSS reconstruction is always smaller than the K-S indicator of the original signal.

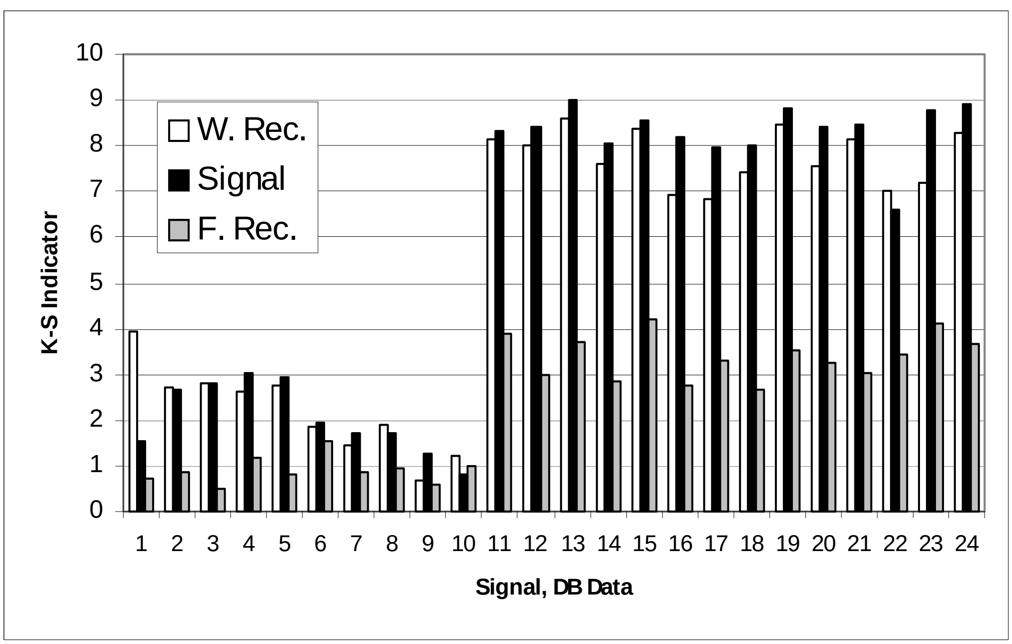

Figure 3 – Comparison of the K-S indicators for the original DB signals, SPSS and RPSS reconstructed signals

5. CONCLUSIONS

In this paper, the results of a newly developed system were reported and compared to ANL's base-case methodology (optimal-order Fourier series approximation). The analyzed signals represent sensor readings independently recorded at three nuclear power plants. For all of the recorded signals, the wavelet technique provided a better approximation of the original signal than the Fourier procedure. For each signal, many wavelet-based decompositions were found by the SPSS methodology, all of which produced white and normally distributed residuals. In most cases, the Fourier-based analysis failed to completely eliminate the original signal serial-correlation in the residuals. The reconstructed signals produced by SPSS are also statistically closer to the original signal than the RPSS reconstructed signal.

REFERENCES

1. K. C. Gross, K. K. Hoyer, "Reactor Parameter Signal Simulator", *Fourth International Conference on Simulation Methods in Nuclear Engineering*, Montreal, Canada, June 1993, Vol. 1, Session 3C (1993).
2. A. Miron, "A Wavelet Approach for Development and Application of a Stochastic Parameter Simulation System", Ph.D. Dissertation, University of Cincinnati, Cincinnati, www.ohiolink.edu/etd (2001).
3. G. E. P. Box, G. M Jenkins, G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd Edition, Prentice Hall Inc. (1994).
4. K. C. Gross, K. K. Hoyer, "Spectrum-Transformed Sequential Testing Method for Signal Validation Application", *8th Power Plant Dynamics, Control and Testing Symposium Proceedings*, Knoxville, Tennessee (1992).
5. G. Strang, T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press (1996).
6. M. Misiti, Y. Misiti, G. Oppenheim, J. M. Poggi, *Wavelet Toolbox For Use with MATLAB*, User's Guide, Version 1, The MathWorks, Inc. (1996).
7. B. W. Lindgren, *Statistical Theory*, 2nd Edition, The McMillan Company (1968).
8. M. S. Bartlett, *An Introduction to Stochastic Processes with Special Reference to Methods and Applications*, 2nd Edition, Cambridge University Press (1966).
9. J. Durbin, "Test of Serial Independence Based on the Cumulated Periodogram", *Bull. Int. Statist. Inst.*, **Vol. 42** (1967).
10. J. Durbin, "Tests for Serial Correlation in Regression Analysis Based on the Periodogram of Least-Squares Residuals", *Biometrika*, **Vol. 56** (1969).