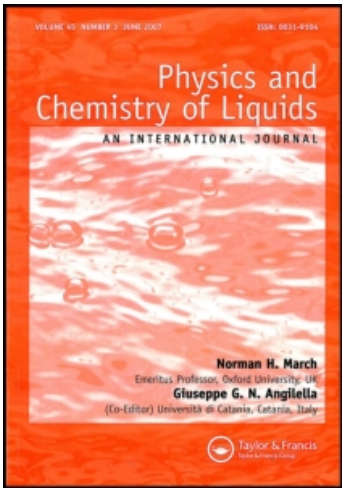

Physics and Chemistry of Liquids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713646857

Statistical Spectral Analysis of liquid X-ray Diffraction Data

A. J. Starshak[ab]; R. D. Larsen[c]
[a] Department of Chemistry, Illinois Institute of Technology, Chicago, Illinois, U.S.A. [b] Exploratory Research, American Cam Company, Neenah, Wis [c] Department of Chemistry, Texas A&M University, College Station, Tezxas, U.S.A.

To cite this Article Starshak, A. J. and Larsen, R. D.(1975) 'Statistical Spectral Analysis of liquid X-ray Diffraction Data', Physics and Chemistry of Liquids, 4: 4, 255 — 265
To link to this Article: DOI: 10.1080/00319107508083829
URL: http://dx.doi.org/10.1080/00319107508083829

Phys. Chem. Liq., 1975, Vol. 4, pp. 255-265

Printed in Dordrecht, Holland

Statistical Spectral Analysis of Liquid X-ray Diffraction Data

A. J. STARSHAK†

Department of Chemistry, Illinois Institute of Technology, Chicago, Illinois 60616 U.S.A.

and

R. D. LARSEN

Department of Chemistry, Texas A&M University, College Station, Texas 77843 U.S.A.

(Received February 21, 1974)

The experimental procedure of determining the structure of a liquid by diffraction techniques is reformulated herein as a stochastic experiment subject to the data analysis formalism of statistical spectral analysis. Observed in such an experiment are averaged local microscopic fluctuations from the bulk density. The intensity function then represents a stochastic spectrum and it becomes necessary to statistically estimate a minimum bias, minimum variance covariance function which is the net radial distribution function. A low-pass tapered data window produces such an optimum estimate.

1. INTRODUCTION

Several decades of X-ray and neutron diffraction experimentation on simple liquids and solutions has given rise to structural data which has been embedded within and analysed in terms of a "deterministic" Fourier analysis formalism originally suggested by Zernicke and Prins. The current state of the art and relevant theory in this field have been elucidated in a recent exposition by Pings.[1]

We suggest herein an alternative data analysis scheme for liquid structure

† Present Address: Exploratory Research, American Can Company, Neenah, Wis. 54956

determination which is a direct consequence of considering that the measurement process inherent in a diffraction experiment, the data arising therefrom, and the molecular physics of the liquid state constitute stochastic processes. It is then natural to employ the formalism of spectral analysis of time series in determining a "radial distribution function" characteristic of a specific liquid.

Spectral analysis is a fundamental reformulation of the techniques of deterministic Fourier analysis for stochastic processes. This reformulation allows a recorded data set to be characterized by its individual frequency components. In the present application, the frequency components themselves are of direct interest; in other applications they are used to produce a mathematical model of the stochastic process.

We consider that the random density fluctuations of a liquid are a stochastic process which is able to be subjected to the formalism of statistical spectral analysis.

Figure 1 exhibits a plot of the net radial distribution function of an ammonium fluoride solution.[2] In this figure, the three distinct maxima are interpreted as three distinct solvation spheres in the hydrogen-bonded

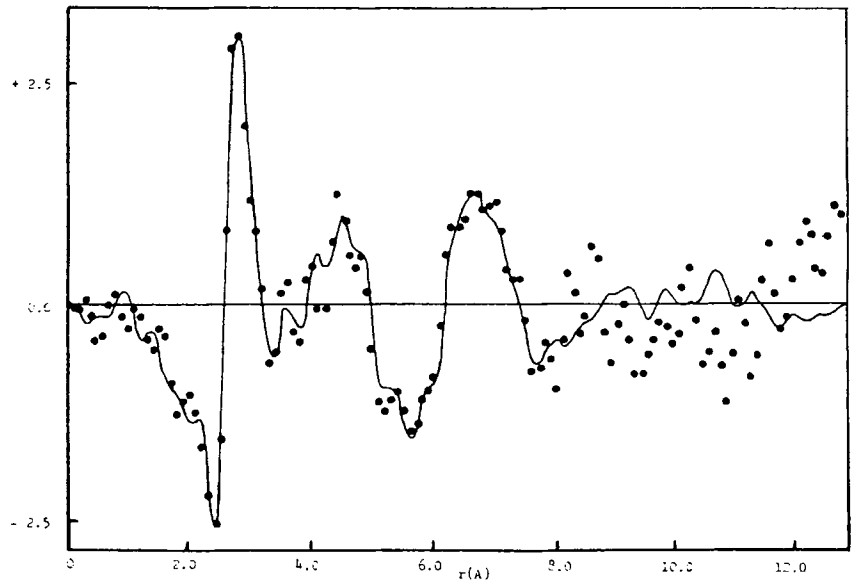

FIGURE 1 Net radial distribution function for $NH_4F{\cdot}3.57H_20$ from data taken at Oak Ridge National Laboratory. The points represent the calculated values for a single experiment and the line represents the average of several experiments.

structure. These distinct solvation spheres are the averaged microscopic density fluctuations in the liquid as observed experimentally.

To illustrate the stochastic nature of this experimental data, Figure 2 presents some of the detected x-rays scattered from an aqueous solution of ammonium fluoride.[2] The mean value is used as the estimator of the center of tendency, and the standard error is shown. It is evident from Figure 2 that, even with extreme care, the data set will not be reproduced exactly. Each point will cluster about some mean value, but deviations will occur. From such a series of determinations, each data point can be represented by a probability density function which indicates the most probable mean value

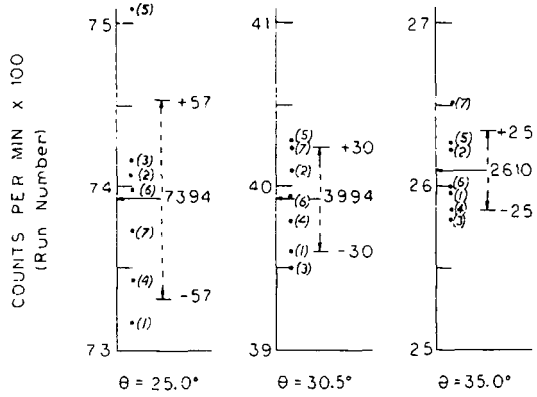

FIGURE 2 Mean value and standard error of detected x-rays scattered from ammonium fluoride solution for several experimental determinations.

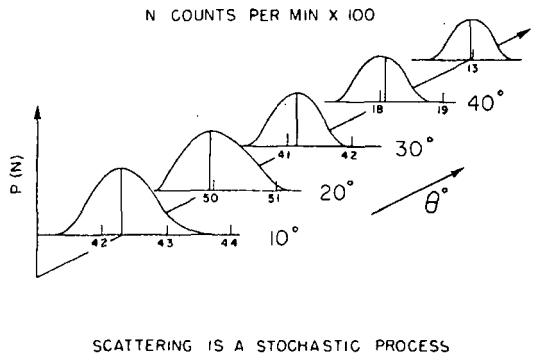

FIGURE 3 Scattering is a stochastic process. Measured value of intensity at any angle is governed by probability density function at that angle.

and some measure of the scatter. In Figure 3 the x-ray diffraction data set is shown as a series of probability density functions, one for each value of the scattering angle θ and each centering on the estimated mean value at that angle.

We shall find it convenient to refer to any specific data set as a "time series". A time series (being a measured realization of a stochastic process) is a random function of an independent variable; this variable is usually time and hence the name, but it may be any other physical parameter such as distance, angle, or frequency.†

While each time series has a Fourier transform each gives rise to frequency components quite different from other time series. Thus, the desired spectrum of a stochastic process is a consistent estimate (as defined statistically) of the frequency components of the stochastic process itself.[3]

A central task of spectral analysis is the estimation from a finite time series of the asymptotic covariance function $\gamma(r)$ of the underlying stochastic process. Then the Fourier transform of $\gamma(r)$ is the spectrum, $\Gamma(f)$. We have achieved consistent covariance estimates having a known bias and small mean-square error by employing a specific class of data windows which are discussed in Section 3.

2. STOCHASTIC DENSITY FLUCTUATIONS

The central problem of ascertaining the local structure of a liquid from the point of view of this manuscript is the determination of the radial distribution function from a *limited set*‡ of diffraction data. The Fourier transform of the diffraction data is usually expressed in terms of the net radial distribution function (1) (Figure 1) given by:

$$r^2 D(r) = 4\pi r^2 [\rho(r) - \bar{\rho}] = \frac{2r}{\pi} \int_0^\infty si(s) \sin rs \, ds \qquad (1)$$

Statistical spectral analysis addresses itself to precisely this same type of problem: the *estimation* of the Fourier transform of a limited empirical data set.

---

† It is important to note that we are not herein considering a time dependent process.

‡ By "limited set of data" in this case we do not refer to the experimental absence of low and high angle data which theory, in part, provides. Rather, each intensity function is a realization of an ensemble of such intensity functions and as such, is an estimate of an asymptotic intensity function which is not experimentally accessible.

That which is measured is the scattered x-ray intensity, representing in reciprocal space the location of individual scatterers, averaged over both time and the sample. The intensity function is, from this viewpoint, a stochastic spectrum. What distinguishes this application from other time series applications is that the intensity data (the spectrum) is directly measured and does not have to be statistically estimated.† Thus its accuracy is characteristic of the experimental procedure alone and is not based on the validity of any data-handling techniques. Note, however, that the intensity data do represent an estimate of the ensemble of all possible intensity realizations or the ergodic equivalent of an infinite frequency record. That which *is* statistically estimated, however, is the net radial distribution function which is a covariance function.

3. DATA WINDOWS

The use of convergence factors in the Fourier inversion of diffraction data assures the convergence of the Fourier integral over a finite range of the independent variable. Waser and Schomaker document thirteen such "modification functions" for which the diffraction integral converges and then suggest how others may be constructed in a similar way.[4] Two of these modification functions are referred to herein as the rectangular and Bartlett windows, and a third is similar to the Parzen window to the 3/4 power. Another is called the "temperature factor" but is essentially a Gaussian function and is called so here. The convergence factors used by Waser and Schomaker are an application of standard techniques of *deterministic* Fourier transform theory to the diffraction integral.

The effect of calculating a biased estimate of the covariance function may be shown to be the same as multiplying the unbiased estimate by the Bartlett data window (1-r/M).[5] In the transform domain each spectral estimate (that is, each calculated spectral point) is the result of a convolution of the spectral window and the neighboring spectral points.

The bandwidth of the spectral window determines the number of neighboring points averaged in the convolution and the side lobes of the window specify their positive and negative weights.

† It may be argued correctly that a measurement is itself an estimate but we attempt to distinguish here between an experimental realization and a post-experimental estimation procedure.

The design of an optimum data window has been discussed throughout the literature of spectral analysis.[6,7] Six data windows are shown in Figure 4; their transforms are the spectral windows shown in Figure 5. Four of these windows, the rectangular, Bartlett, Tukey and Parzen, have a rich history in spectral analysis; they have been extensively characterized.[8] The Gaussian window has been a very important modification function in deterministic

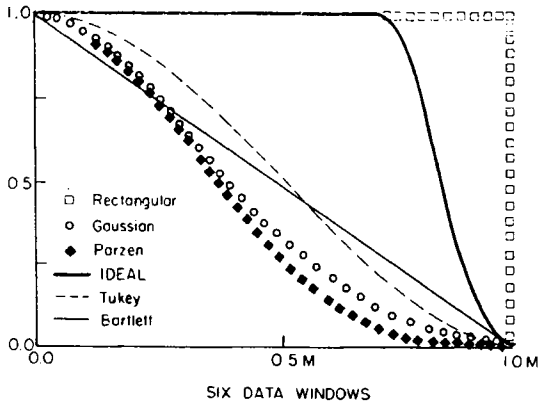

FIGURE 4 Data windows.

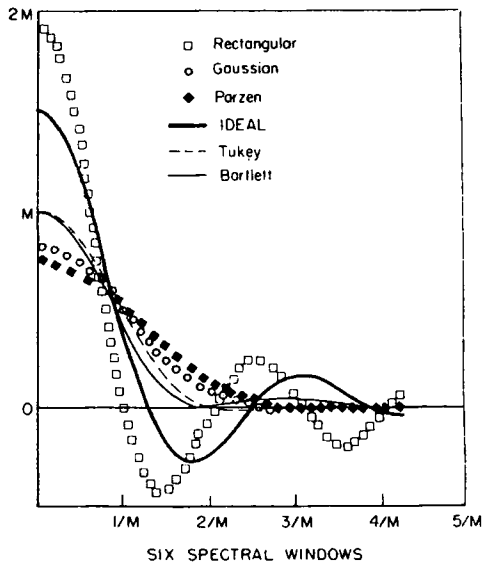

FIGURE 5 Spectral windows.

diffraction data inversion.[9] Finally, the "ideal window", upon which we wish to focus the most attention, corresponds to the shape of the ideal, low-pass filter used by electrical engineers.[10] This window has a flat pass band, a smooth transition band, and an effective stop band.

The above data windows were used in the following way. The modified net radial distribution functions $r^2\hat{D}(r)$, as calculated† from the variously modified reduced intensity functions si(s), were compared with the unmodified net radial distribution function $r^2D(r)$. The results group the Bartlett and Tukey windows, the rectangular and ideal windows, and the Parzen and Gaussian windows in three recognizable categories. That the Parzen and the Gaussian windows are quite similar is satisfying, since the mathematical form for the Parzen spectral window is $(\sin x/x)^4$ and the Gaussian window is $(\sin x/x)^n$, $n>5$. That the Tukey window is so similar to the Bartleet window is sur-

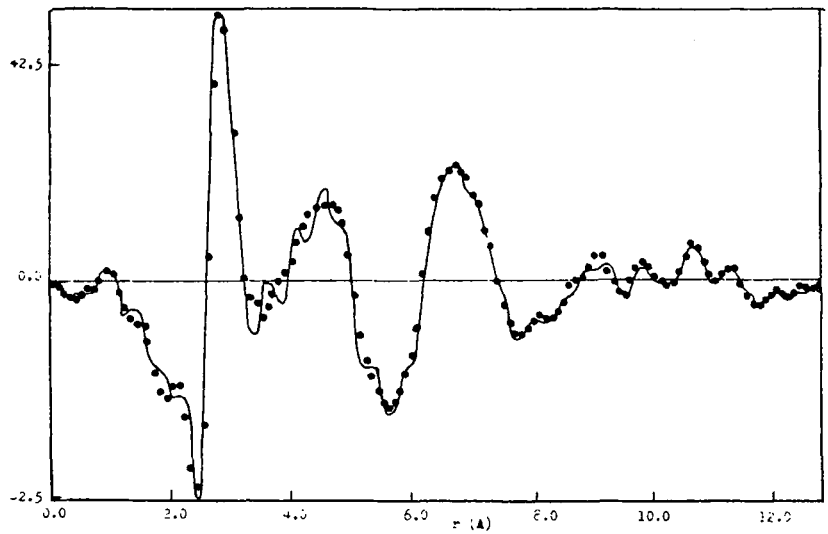

FIGURE 6 The net radial distribution function for $NH_4F{\cdot}3.57H_2O$ showing effect of smoothing by the ideal window. The points represent the best smoothed $r^2D^*(r)$ and the line represents the unmodified $r^2D(r)$.

† In all cases a fast Fourier transform (FFT) algorithm was applied to compute the radial distribution function.[11] The advantage of the algorithm for this purpose is its ability to transform from one domain to another with great accuracy because a discretely sampled continous function is expanded over a complete set of orthonormal exponential functions (see Ref. 11 for details).

prising since the Tukey window has a smooth second derivative. Finally, that the ideal window is so similar to the rectangular window is initially disappointing until the $r^2 \hat{D}(r)$ curves from each are more closely examined.

Figure 6 shows these two curves. It is apparent that there is little bias stemming from the application of the ideal window. Figure 7 shows an enlarged view of the region from 2.7 to 6.0 Å; the decrease in the variance of the net radial distribution function is most apparent in this region. From a comparison of these curves, it would appear that the ideal window produces a minimum-bias, minimum-variance estimate (relative to other such windows) of the theoretically perfect net radial distribution function.†

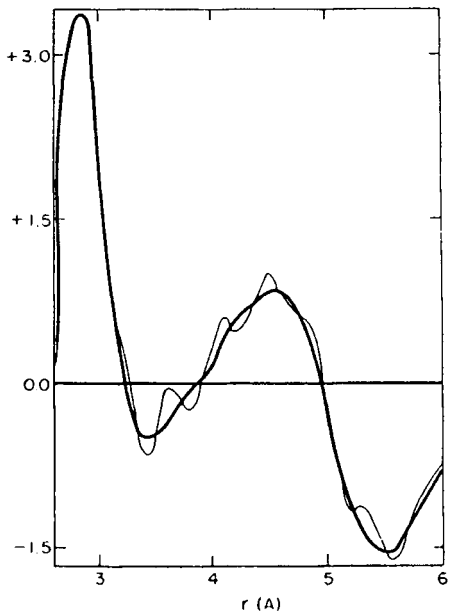

FIGURE 7 Enlarged view of smoothed $r^2 D^*(r)$ from 2.7 to 6.0 A showing the decrease in variance of $r^2 D(r)$ due to application of the ideal window.

† A measure of the bias of the smoothing process can be obtained from the absolute value of the mean deviation of $r^2 D(r)$ from $r^2 D(r)$. Plots of these values along with plots of correlation coefficients comparing $r^2 \hat{D}(r)$ smoothed by six windows and standard deviations of residual deviations between various smoothed values of $r^2 \hat{D}(r)$ are contained in the dissertation of A.J.S. Copies of these plots and other related tabulated quantities may be obtained from the authors.

4. THE OPTIMUM DATA WINDOW

Most of the data windows suggested for use with liquid diffraction data tend to oversmooth. Although they eliminate most of the variance, they produce a bias which seems unacceptable. The ideal window, on the other hand, is basically a class of rounded or tapered rectangular functions for which the degree of smoothing can be varied. In attempting to ascertain the degree of smoothing which produces the best $r^2 D^*(r)$, the ideal window was constructed in the following manner: a rectangular window was terminated in a Parzen window at successively shorter distances, designated: L = 120, 112, 104, 96, 80, 64, and 32. (Thus, a monotone varying series of windows was produced between the extremes of the rectangular (L = 128) and the Parzen windows (L = 0); had the ideal window been constructed from the rectangular and Tukey windows, such a monotone series would have been formed between those two as the extremes.) The unmodified $r^2 D(r)$ in the region $2.7 \leq r \leq 6.0$ was compared with $r^2 \hat{D}(r)$ curves calculated from $\hat{si}(s)$ modified with an ideal window for various L values. An inspection of the results showed that apparently for L = 112 and 104 no bias disrupts the major maximum at 2.8 A. When L = 64, in addition to the bias, the broader bandwidth shifted the maximum to a slightly larger r-value. Despite the first appearance of the bias in the peak at 2.8 A for L = 96 the modified curve flows through the minimum of the unmodified $r^2 D(r)$ demonstrably more smoothly than for L = 80. Thus, it appears that if the tapering is initiated anywhere between L = 80 and L = 104 an excellent (in the sense described above) $r^2 D^*(r)$ is obtained. For convenience, L = 96 or L/M = 3/4 was chosen as the optimum point to begin tapering of the ideal window.

5. SMOOTHING AND THE SINGLE EXPERIMENT

The traditional way of the experimentalist to reduce the variance in empirical data is to average N independent determinations. In this way, the variance is reduced by the factor $N^{-1/2}$. Another way to reduce the variance in this context is the Bartlett technique: smooth the covariance function of the data before taking the transform.[12] This was done with a single experimental data set in the same manner as outlined in the last section.

The results of using different smoothing windows parallel the results of the last section. As expected, the effect of the rectangular window was

exactly the same as the effect of using no window. The results of using the Bartlett, Tukey, Parzen, and Gaussian windows were to oversmooth the variance and generate much bias.

A thorough examination of the influence of the ideal window on the transform of the single data set showed the best results came from an ideal window with L = 64. Figure 8 gives the net radial distribution function with the line representing the averaged $r^2\bar{D}(r)$ unsmoothed and with the points representing the best $r^2 D_1^*(r)$ from the single experimental data set smoothed with L = 64. This figure shows that the single experimental data set can be smoothed to give an $r^2 D_1^*(r)$ as smooth as that from the averaged data set (but not as smooth as can be calculated from the averaged data set smoothed with an ideal window.)

6. SUMMARY

Upon identification of (a) the net radial distribution function $r^2 D(r)$ as the covariance function and (b) the reduced intensity function si(s) as the spectrum known techniques for treating a stochastic process can be applied to the analysis of liquid diffraction data.

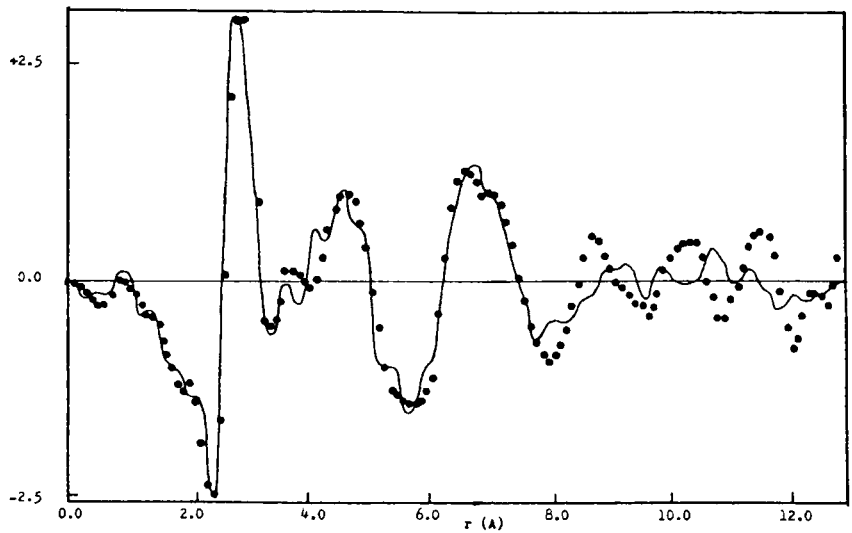

FIGURE 8 The net radial distribution function for $NH_4F{\cdot}3.57H_2O$ showing the influence of the ideal window on a single data set. The points represent the calculated values $r^2 D_1(r)$ from a single experiment and the line represents the smoothed $r^2 D_1^*(r)$ for L = 64.

This analysis correctly identifies the underlying molecular process as a stochastic process, not a deterministic process as has been tacitly assumed in the standard analysis. Thus, a distinguishing feature of a liquid – translational molecular motion – becomes an essential factor in the treatment and interpretation of the data.

A low-pass tapered data window has been seen to produce a relatively minimum-bias, minimum-variance estimate of the transform. As such it proves to be an aid in enhancing the information content of the empirical data.

Finally, it was seen that a smoothing of the data from a single experiment gave a result virtually as reliable as that obtained from the average of several experiments.

Our recommendation based on these results is that experiments be designed to collect data that is relatively closely spaced and of relatively long record length and then to employ covariance smoothing of the resulting data using an appropriate tapered data window of the type described herein.

Acknowledgements

This research was submitted by A.J.S. in partial fulfillment of the requirements for the Ph.D. degree at Illinois Institute of Technology. The authors wish to thank Dr. A.H. Narten of the Oak Ridge National Laboratory for his constructive comments on this work. Mr. Nate Thomas of the IIT Computation Center assisted in scheduling many of the computer calculations. This work was supported, in part, by the National Science Foundation through Grant No. GP-16760.

References

1. C. J. Pings, "Structure of simple liquids by x-Ray diffraction", in *Physics of Simple Liquids*, H. N. V. Temperley, J. S. Rowlinson, and G. S. Rushbrooke, Eds., Wiley Interscience and North Holland Publishing Co., p. 387, (1969).
2. M. D. Danford, *ORNL* – 4244, 14 (1968).
3. G. M. Jenkins and D. G. Watts, *Spec. Anal.*, Holden-Day, San Francisco (1968).
4. J. Waser and V. Schomaker, *Rev. Mod. Phys.*, 25, 671 (1953).
5. G. M. Jenkins and D. G. Watts, *Spec. Anal.*, Holden-Day, San Francisco, p. 240, (1968).
6. H. Akaike, Low Pass Filter Design, *Tech Rept. #20*, Dept. Stat., Stanford, California, (1967).
7. E. Parzen, *Time Series Analysis Papers*, Holden-Day, San Francisco, p. 190 (1967).
8. G. M. Jenkins and D. G. Watts, *Spec. Anal.*, Holden-Day, San Francisco, pp. 243-257 (1968).
9. B. E. Warren, *X-Ray Diffraction*, Addison-Wesley, Reading, Mass., pp. 127, 137 (1969).
10. B. Gold and C. M. Rader, *Dig. Proc. Sign.*, McGraw-Hill, New York, p. 228, (1969).
11. A. J. Starshak and R. D. Larsen, *Phys. Chem. Liq.*, 2, 45, (1970).
12. B. Gold and C. M. Rader, *Dig. Proc. Sign.* McGraw-Hill, New York, p. 280, (1969).