NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Report No. 32-744

A Summary of Methods for Analyzing Nonstationary Data

J. S. Bendat

G. P. Thrall

JET PROPULSION LABORATORY

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA

September 1, 1965

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Report No. 32-744

A Summary of Methods for Analyzing Nonstationary Data

J. S. Bendat

G. P. Thrall

W. S. Shipley, Manager
Environmental Requirements Section

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

September 1, 1965

Prepared Under Contract No. NAS 7-100
National Aeronautics & Space Administration

CONTENTS

FIGURES

PREFACE

The work presented in this Report was performed for Jet Propulsion Laboratory by J. S. Bendat and G. P. Thrall of the Measurement Analysis Corporation, Los Angeles, California, under Contract NAS 7-100, sponsored by the National Aeronautics and Space Administration.

ABSTRACT

N65-33142

This Report summarizes theoretical ideas and techniques used for the analysis of nonstationary data. Details are presented for estimation of nonstationary mean values, correlation functions, and spectral density functions. Three methods are discussed for estimating nonstationary mean values: ensemble averaging, short-time averaging, and orthogonal-function averaging. Three methods are discussed for estimating nonstationary spectral density functions: double-frequency spectra, time-varying power spectra, and time-averaged power spectra.

Author

I. INTRODUCTION

A random process $\{x_i(t)\}$, $i = 1, 2, 3, \cdots$, is an ensemble of functions of a single variable t which can be characterized through its statistical properties. A typical random process is pictured in Fig. 1. The variable t is time or any other parameter of interest.

The set of amplitude values at a particular time t_1, denoted by $x_i(t_1)$, can be combined together in appropriate ways to determine their mean value, mean square value, and higher moments, and thus a complete probability distribution at t_1, where the probability distribution yields the probability that the amplitude values at t_1 will lie in any specified amplitude range. For example, the mean value at t_1 is defined by the ensemble average, denoted by the expected value $E[\ ]$, namely,

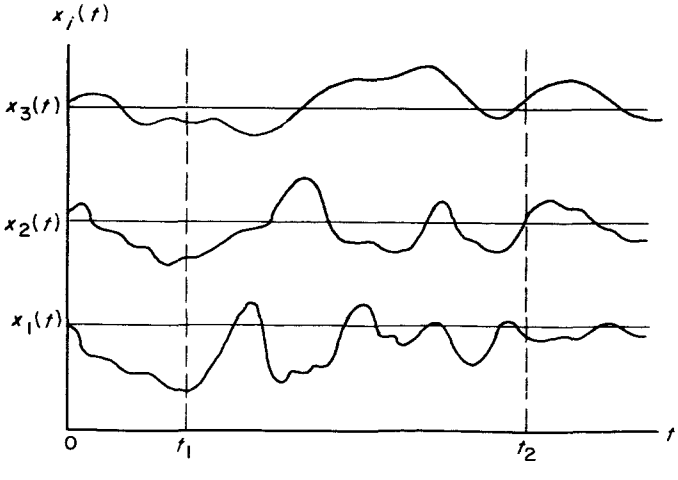

Fig. 1. Random process

$$\mu(t_1) = E[x(t_1)] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i(t_1) \qquad (1)$$

A different $t = t_2$ can be selected, and similar statistical calculations may be carried out for the set of amplitude values $\{x_i(t_2)\}$, $i = 1, 2, \cdots$. In general, significantly different results would be obtained for the two sets $\{x_i(t_1)\}$ and $\{x_i(t_2)\}$. That is to say, the statistical results would not be invariant with respect to translations in time. Random processes of this general category are known as *nonstationary random processes.* A process is said to be *stationary* when statistical properties do not change with time. Much past analytical work assumed a stationary hypothesis because it simplified the further derivations. This Report is concerned with methods for analyzing nonstationary data.

II. TESTING FOR STATIONARITY

The first step in the analysis of data suspected of having nonstationary characteristics is to examine or test the data for stationarity. This is very important for two reasons: Nonstationary data require much more processing to estimate statistical parameters of interest, such as mean and mean square values. Secondly, treating nonstationary data as stationary can lead to highly erroneous results and interpretations.

Stationarity is technically a property of a collection (ensemble) of time-history records which together represent a random process. A random process is said to be weakly stationary if the mean value and autocovariance (autocorrelation) function computed over the ensemble are invariant with shifts in the time origin (note that the mean square value will also be time invariant if these conditions exist). The process is said to be strongly stationary if the joint density functions of all orders are time invariant. When only a single record from a random process is available, which is often true in practice, a property called self-stationarity is of interest. A single time-history record is said to be weakly self-stationary if the mean value and autocovariance function computed over short time intervals are statistically equivalent for intervals with different starting times. The single time-history record is said to be strongly self-stationary if this is true for all higher order joint moments.

Most practical tests for stationarity are really tests for self-stationarity of single records. If the record in question is a sample from an ergodic random process, then self-stationarity of the single record implies stationarity for the entire random process. An experimental verification of the ergodic hypothesis is usually not possible since a collection of records would be required for such a verification. However, the ergodic hypothesis can often be justified for practical situations from a knowledge of the physics of the specific problem. Hence, tests for self-stationarity often suffice as tests for stationarity and will be referred to as such in the discussions to follow.

There are many general considerations in attempting to develop a test for stationarity. In the past a test for time invariance of mean square values has often been accepted as a test for time invariance of an autocovariance function. The justification for this assumption is that it would be extremely unlikely in practical problems to have a variation of the autocovariance function without a variation of the mean square value, which is the peak value of the autocovariance function; i.e., the value of the autocovariance function for zero time displacement. Another way of looking at this statement is in terms of the power spectrum as follows. A change in the covariance function without a change in the mean square value is equivalent to a change in the general shape of the power spectrum without a change in the total area under the power spectrum. This clearly would be an unlikely occurrence (although not impossible) for data representing the vibration response at some point on a structure. Hence, the current stationarity tests which consider only mean and mean square values are reasonable procedures to establish weak stationarity.

Generally speaking, a verification of weak stationarity is sufficient for many practical applications. For example, only weak stationarity is required to properly measure and interpret power spectra information in conventional stationary terms. However, there are certain specialized applications where an assumption of strong stationarity is of importance. For example, the measurement of probability density functions and the application of resulting probability statements requires a time invariance of higher moments for the data in question.

For the special case of data with a Gaussian probability density function, the verification of weak stationarity also establishes strong stationarity since all higher moments of a Gaussian density function are defined by the first two (mean and mean square values). For non-Gaussian data, however, a proper verification of strong stationarity would technically require a test for time invariance of the entire probability density function for the data. Nonparametric tests, as well as special parametric tests, for stationarity are discussed and verified in Ref. 1, 2, 3.

III. ESTIMATION OF MEAN VALUES

The estimation of any parameter of a random process $x(t)$ may be related to the estimation of the mean value of a second random process $y(t)$, which is derived from $x(t)$. This is a rather strong statement so that before proceeding further a few examples will be provided.

1. Mean value of $x(t)$

   Let $\quad y(t) = x(t)$

   Then $\quad E[y(t)] = E[x(t)]$

2. Mean square value of $x(t)$

   Let $\quad y(t) = x^2(t)$

   Then $\quad E[y(t)] = E[x^2(t)]$

3. Autocorrelation function of $x(t)$

   Let $\quad y(t) = x(t)\,x(t-\tau)$

   Then

   $$E[y(t)] = E[x(t)\,x(t-\tau] = R_x(t, t-\tau)$$

   where $R_x(t_1, t_2)$ is the autocorrelation function of $x(t)$ and τ is a given time delay.

4. Time-varying power spectrum of $x(t)$

   Let $\quad y(t) = \dfrac{x^2(t, f, B)}{B}$

   Then

   $$\lim_{B\to 0} E[y(t)] = \lim_{B\to 0} \frac{E[x^2(t, f, B)]}{B} = G_x(t, f)$$

   where $G_x(t, f)$ is the physically realizable one-sided ($f \geq 0$) time-varying power spectrum of $x(t)$, and $x(t, f, B)$ is the result of passing $x(t)$ through a narrow bandpass filter of bandwidth B centered at the frequency f.

Thus, detailed consideration should be given to proper estimation of mean values. However, when determining the errors caused by sample size, record length, bandwidth resolution, etc., changes in the probability distributions caused by transforming $x(t)$ into $y(t)$ must be taken into account. These effects will be discussed in subsequent Sections.

A. Ensemble Averaging

Mean values can be estimated by using an average-response computer that performs the following operation to calculate a sample mean value from a sample of size N (Fig. 2). For N records $\{x_i(t);\ 0 \leq t \leq T;\ i = 1, 2, \cdots, N\}$ from a nonstationary process $x(t)$, compute the estimate

$$m(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t) \tag{2}$$

The quantity $m(t)$ will differ over different choices of the N samples $\{x_i(t)\}$. Consequently, one must investigate how closely an arbitrary measurement m(t), approximates the true mean value $\mu(t)$, which is given by the expected value

$$\mu(t) = E[m(t)] = \frac{1}{N}\sum_{i=1}^{N} E[x_i(t)] \tag{3}$$

Note that $m(t)$ is an unbiased estimate of the true mean value for all t, independent of N.

A measure of the error involved in estimating $\mu(t)$ by $m(t)$ is the variance of $m(t)$ given by

$$\sigma_m^2(t) = E\left[\{m(t) - \mu(t)\}^2\right] \tag{4}$$

The square root of the variance $\sigma_m(t)$, called the standard deviation, provides the indicator that determines how closely a set of measurements of $m(t)$ clusters about its mean value $\mu(t)$.

In many practical applications, the N samples used to compute $m(t)$ are statistically independent and this will be assumed here. Upon expanding Eq. (4), it is seen that

$$\sigma_m^2(t) = E\left[\frac{1}{N}\sum_{i=1}^{N}\{x_i(t) - \mu(t)\}\right]^2$$

$$= \frac{1}{N^2}\sum_{i,j=1}^{N} E\left[\{x_i(t) - \mu(t)\}\{x_j(t) - \mu(t)\}\right]$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\sigma_x^2(t) = \frac{\sigma_x^2(t)}{N} \tag{5}$$

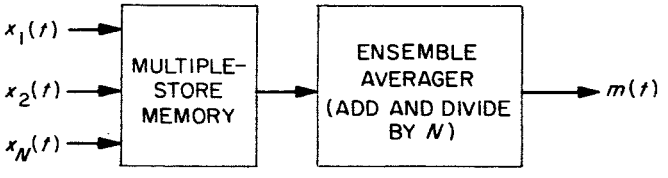

Fig. 2. Nonstationary mean value measurement

where $\sigma_x^2(t)$ is the variance associated with the nonstationary process $x(t)$. It should be noted that assuming that the samples are independent causes all the cross-product terms ($i \neq j$) in the double sum of Eq. (5) to be zero.

As shown in Ref. 2, a knowledge of the mean value and variance for the random variable $m(t)$ at any time t enables one to answer questions concerning the range of the results at any time t without knowing the exact probability distribution function for $m(t)$. From the Chebyshev inequality, which applies to arbitrary general situations, one may state with 89% confidence, for example, that $\mu(t)$ lies inside the range $m(t) \pm 3\sigma_m(t)$. In equation form, for any constant $c > 0$, the Chebyshev inequality is

$$\text{Prob}\left[\,|\, m(t) - \mu(t) \,|\, \geq c\sigma_m(t)\right] \leq \frac{1}{c^2} \tag{6}$$

Thus, for $c = 3$, this probability is at most (1:9), giving the 89% confidence limits as illustrated in Fig. 3.

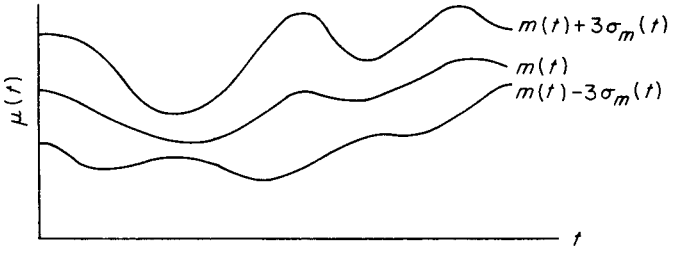

Fig. 3. 89% confidence limits for arbitrary distribution based upon the Chebyshev inequality

A stronger statement can be made if one can justify an assumption that $m(t)$ follows a normal (Gaussian) distribution at any value of t. For this special case, a 95% confidence band is given by the range $m(t) \pm 2\sigma_m(t)$. Thus, an observed measurement for $m(t)$ in the Gaussian case yields a greater confidence of being close to the theoretical mean value than in the case where the underlying probability distribution is unknown.

Two main steps are involved in a measurement of $m(t)$. The first step is to obtain and store each record $x_i(t)$ as a function of t. This may be done continuously for all t in the range $0 \leq t \leq T$, or discretely by some digitizing procedure. After this has been done for N records, the next step is to perform an ensemble averaging by adding the records together and dividing by N. If each $x_i(t)$ is digitized in, say M steps, then the total number of stored values would be MN.

B. Short-Time Averaging

One special technique currently being employed to estimate nonstationary mean values when only one or a few samples are available is based upon obtaining a continuous short-time average. The effect of a time average is to smooth the random fluctuations over the averaging interval and thus reduce the uncertainty in the estimate. This technique is discussed in Ref. 4 and 7 for mean square value measurements.

Suppose that $x(t)$ is averaged over a time $2s$. The resulting random process $z(t)$ is thus defined by

$$z(t) = \frac{1}{2s}\int_{t-s}^{t+s} x(v)\, dv \tag{7}$$

The expected value of $z(t)$ is easily seen to be

$$E\left[z(t)\right] = \frac{1}{2s}\int_{t-s}^{t+s} \mu(v)\, dv \tag{8}$$

where $\mu(v)$ is the true mean value of $x(t)$ at $t = v$. In general, the expected value of $z(t)$ will be a biased estimate of $\mu(t)$ because of the integration. To determine the bias error, assume that $\mu(t)$ can be represented in the interval $(t-s, t+s)$ by a power series of order N, namely

$$\mu(v) = \sum_{n=0}^{N} a_n(t)\,(v-t)^n \tag{9}$$

where $t - s < v < t + s$.

Upon substitution of Eq. (9) into Eq. (8), it follows that

$$E\left[z(t)\right] = \sum_{n=0}^{N}{}^{e} \frac{a_n(t)}{n+1}\, s^n \tag{10}$$

where the "e" within the summation sign means that only even values are used.

The bias error in estimating $\mu(t)$ is defined by

$$\begin{aligned} b(t,s) &= E\left[z(t)\right] - \mu(t) \\ &= \sum_{n=2}^{N}{}^{e} \frac{a_n(t)}{n+1}\, s^n \end{aligned} \tag{11}$$

It is clear that if $\mu(t)$ is a polynomial of moderate order, a large bias error can be introduced into the estimation procedure. In most practical applications, however, the variations in $\mu(t)$ are not too complicated within an interval of length $2s$, since s is usually small. This means that, within each interval of length $2s$, $\mu(t)$ can be considered to be a polynomial of degree two or less and the resulting bias error will be negligible or zero.

The variance in estimating $\mu(t)$ by a short-time average is given by

$$\sigma(t) = E_z^2 \left[\{z(t) - \mu(t)\}^2\right] \tag{12}$$

Upon carrying out the operations indicated in Eq. (12) it is seen that for short-time averaging

$$\sigma_z^2(t) = \frac{1}{4s^2}\iint_{t-s}^{t+s} C_x(\nu, \eta)\, d\nu\, d\eta + b^2(t, s) \tag{13}$$

where $C_x(\nu, \eta)$ is the covariance function of $x(t)$ at $t_1 = \nu$ and $t_2 = \eta$, defined by Eq. (28). The double integral occurring in Eq. (13) is quite difficult to evaluate even for simple cases. The optimum averaging time s for minimum variance $\sigma_z^2(t)$ can be defined from Eq. (13) by setting the derivative with respect to s equal to zero.

There is one reasonably general case, however, where reduction to a single integral may be achieved. Suppose that the bias term can be neglected and the random process $x(t)$ is stationary except for the time-dependent mean value. In this situation, the double integral may be reduced to a single integral involving the physically realizable one-sided ($f \geq 0$) power spectral density function $G_x(f)$ of the random process, namely

$$\sigma_z^2(t) = \int_0^\infty G_x(f)\, \frac{\sin^2(2\pi sf)}{(2\pi sf)^2}\, df \tag{14}$$

Equation (14) shows that in this situation the variance is independent of time since the variable t does not appear on the right-hand side.

C. Orthogonal Function Averaging

Another special technique is based upon fitting a Kth-order orthogonal function to $x(t)$ and using the resulting expansion to estimate the mean value. This technique is discussed in Ref. 4 and 7.

Let $\{P_k(t), k = 0, \cdots, K\}$ be a sequence of ortho-normal functions defined on the interval $(0, T)$. Thus

$$\int_0^T P_i(t)\, P_j(t)\, dt = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \tag{15}$$

Let $L_K(t)$ be a linear combination of the $P_k(t)$:

$$L_K(t) = \sum_{k=0}^{K} a_k P_k(t) \tag{16}$$

The coefficients a_k are to be determined by multiplying $x(t)$ by $P_k(t)$ and integrating over $(0, T)$. Thus

$$a_k = \int_0^T x(t)\, P_k(t)\, dt \tag{17}$$

The expected value of a_k is given by

$$E[a_k] = \int_0^T \mu(t)\, P_k(t)\, dt = b_k \tag{18}$$

where b_k is defined as the kth coefficient in the orthogonal expansion of $\mu(t)$. The expected value of $L_K(t)$ is thus

$$E[L_k(t)] = \sum_{k=0}^{K} b_k P_k(t) \tag{19}$$

Equation (19) indicates that unless $\mu(t)$ can be represented exactly by a Kth order expansion, $L_K(t)$ is a biased estimate of $\mu(t)$.

The integrated mean square error in approximating $\mu(t)$ by $L_K(t)$ is

$$\epsilon^2 = E\int_0^T \left[L_K(t) - \mu(t)\right]^2 dt \tag{20}$$

Upon carrying out the operations indicated in Eq. (20), it is seen that the expression for ϵ^2 becomes

$$\epsilon^2 = \int_0^T \mu^2(t)\, dt - \sum_{k=0}^{K} b_k^2 + \sum_{k=0}^{K}\int_0^T\int_0^T C_x(u, v)\, P_k(u)\, P_k(v)\, du\, dv \tag{21}$$

where $C_x(u, v)$ is the covariance function defined by Eq. (28). Equation (21) has several interesting properties which result from the use of the orthogonal expansion. Let ϵ_K^2 represent the first two terms of ϵ^2, called the truncation error. It is clear that these two terms involve only $\mu(t)$ and the coefficients in the orthogonal expansion of $\mu(t)$, and thus are independent of the higher moments of $x(t)$. The truncation error ϵ_K^2 in using a finite orthogonal expansion is thus completely isolated from the error caused by higher order effects. This property is of great practical importance since it permits independent investigations of the "signal" and "noise" to be carried out.

It should be noted that ϵ_K^2 is always positive or zero. This follows from the fact that, for any ortho-normal system, Bessel's inequality (Ref. 5) applies so that

$$\sum_{k=0}^{K} b_k^2 \leq \int_0^T \mu^2(t)\, dt \tag{22}$$

where the equality sign holds if $\mu(t)$ can be represented exactly by a Kth degree expansion.

The final term in the expression for ϵ^2 is the contribution of the noise to the mean square error and will be denoted by ϵ_0^2. Unless $C_x(u, v)$ is known, it is not possible to

evaluate the double integrals; however, an upper bound may be obtained by using the fact that $C_x(u,v) \leq C_x(u,u)$. Thus,

$$\epsilon_0^2 = \sum_{k=0}^{K} \int_0^T \int_0^T C_x(u,v)\, P_k(u)\, P_k(v)\, du\, dv$$

$$\leq \sum_{k=0}^{K} \int_0^T C_x(u,u)\, P_k(u) \int_0^T P_k(v)\, dv\, du \tag{23}$$

Since $P_0(v) = 1/\sqrt{T}$, the orthogonality property of the set $\{P_k(v)\}$ shows that for $k \neq 0$,

$$\sqrt{T} \int_0^T P_k(v)\, P_0(v)\, dv = \int_0^T P_k(v)\, dv = 0 \tag{24}$$

Thus all the terms in the second line of Eq. (23) are zero except the $k = 0$ term, and

$$\frac{\epsilon_0^2}{N} \leq \frac{1}{N} \int_0^T C_x(u,u)\, du = \frac{1}{N} \int_0^T \sigma_x^2(u)\, du \tag{25}$$

If the integrated mean-square-error criterion is applied to the ensemble-averaging technique, it is easily shown that the resulting mean square error is exactly equal to the right side of Eq. (25). This means that by choosing a suitable value of K to minimize ϵ_K, an improved estimate of $\mu(t)$ will be obtained in almost all cases through the use of the orthogonal expansion.

As a final note, it should be mentioned that the preceding analysis is unchanged if the measured data are discrete rather than continuous. For the discrete case, the integral signs should be replaced by summations over the data points and discrete orthogonal functions used.

IV. ESTIMATION OF CORRELATION FUNCTIONS

For nonstationary random processes $x(t)$ and $y(t)$, the correlation functions at arbitrary fixed values of t_1 and t_2 are defined by the expected values (ensemble averages)

$$\begin{aligned} R_x(t_1,t_2) &= E[x(t_1)\, x(t_2)] \\ R_y(t_1,t_2) &= E[y(t_1)\, y(t_2)] \end{aligned} \tag{26}$$

$$R_{xy}(t_1,t_2) = E[x(t_1)\, y(t_2)] \tag{27}$$

The quantities $R_x(t_1, t_2)$ and $R_y(t_1, t_2)$ are called *nonstationary autocorrelation functions*, while $R_{xy}(t_1, t_2)$ is called a *nonstationary cross-correlation function*. Note that $R_{xy}(t_1, t_2)$ includes $R_x(t_1, t_2)$ as special cases when $x = y$. In practical problems, it is convenient to allow these correlation functions to include delta functions.

The *covariance functions* are defined by

$$\begin{aligned} C_x(t_1,t_2) &= R_x(t_1,t_2) - \mu_x(t_1)\, \mu_x(t_2) \\ C_y(t_1,t_2) &= R_y(t_1,t_2) - \mu_y(t_1)\, \mu_y(t_2) \\ C_{xy}(t_1,t_2) &= R_{xy}(t_1,t_2) - \mu_x(t_1)\, \mu_y(t_2) \end{aligned} \tag{28}$$

Note that covariance functions and correlation functions are identical when the mean values are zero.

An upper bound for the nonstationary cross-correlation (or cross-covariance) function is given by the inequality

$$|R_{xy}(t_1,t_2)|^2 \leq R_x(t_1,t_1)\, R_y(t_2,t_2) \tag{29}$$

From the original definitions, one sees that the following symmetry properties are satisfied:

$$\begin{aligned} R_x(t_2,t_1) &= R_x(t_1,t_2) \\ R_y(t_2,t_1) &= R_y(t_1,t_2) \\ R_{xy}(t_2,t_1) &= R_{yx}(t_1,t_2) \end{aligned} \tag{30}$$

Thus, the correlation structure of nonstationary random processes $x(t)$ and $y(t)$ may be described by the four functions $R_x(t_1, t_2)$, $R_y(t_1, t_2)$, $R_{xy}(t_1, t_2)$, and $R_{yx}(t_1, t_2)$. These need be calculated only for values of $t_1 \leq t_2$ since the symmetry properties of Eq. (30) yield results for $t_1 > t_2$.

Consider the problem of measuring $R_x(t_1, t_2)$ using a set of N sample functions $x_i(t)$; $i = 1, 2, \cdots, N$, from the nonstationary random process. In place of Eq. (26), one should compute the ensemble-average estimate

$$\hat{R}_x(t_1,t_2) = \frac{1}{N} \sum_{i=1}^{N} x_i(t_1)\, x_i(t_2) \tag{31}$$

A recommended procedure is to hold t_1 fixed and vary t_2. Let $t_1 = t$ and let $t_2 = t - \tau$ where τ is a fixed time-delay value. This yields

$$\hat{R}_x(t, t-\tau) = \frac{1}{N} \sum_{i=1}^{N} x_i(t)\, x_i(t-\tau) \qquad (32)$$

which for stationary processes would be a function of τ only, but for nonstationary processes would be a function of both t and τ. For each fixed delay value τ and each record $x_i(t)$, calculate and store the product $x_i(t)\, x_i(t-\tau)$. Repeat for all N records and then perform an ensemble average to yield the estimate of Eq. (32). This whole operation must be repeated for every different τ of concern. Figure 4 illustrates this procedure for measuring nonstationary autocorrelation functions. A similar procedure may be followed for nonstationary cross-correlation function measurements.

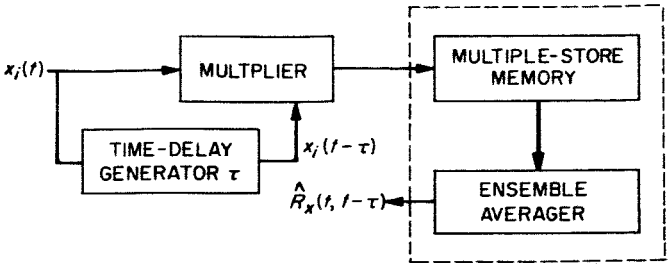

Fig. 4. Nonstationary autocorrelation measurement

Instead of storing the records and carrying out an ensemble averaging on the products $x_i(t)\, x_i(t-\tau)$, the methods of short-time averaging and orthogonal-function averaging may be used to estimate $\hat{R}_x(t, t-\tau)$. This is indicated in Fig. 4 by the dotted lines. For the special delay value $\tau = 0$, any of these methods will provide an estimate of the nonstationary mean square value of $x(t)$.

V. ESTIMATION OF SPECTRAL DENSITY FUNCTIONS

Three different methods to describe nonstationary spectra will now be discussed. Each method has very special features and properties. These three methods are:

1. Double-Frequency Spectra
2. Time-Varying Power Spectra
3. Time-Averaged Power Spectra

Methods 1 and 2 are considered to be the significant theoretical ways to analyze nonstationary spectra. The time-varying and time-averaged power spectra of Methods 2 and 3 are measurable experimentally. This is not possible for the theoretical spectra of Method 1. Method 3 is included because it may be helpful for certain applications where the detailed time structure is not of importance. Further discussion on these methods appears in Ref. 6.

A. Double-Frequency Spectra

Assume that the nonstationary autocorrelation and cross-correlation functions exist and that they have double Fourier transforms as defined below:

$$S_x(f_1, f_2) = \int_{-\infty}^{\infty}\int R_x(t_1, t_2) \exp\left[j2\pi(f_1 t_1 - f_2 t_2)\right] dt_1\, dt_2 \qquad (33)$$

$$S_y(f_1, f_2) = \int_{-\infty}^{\infty}\int R_y(t_1, t_2) \exp\left[j2\pi(f_1 t_1 - f_2 t_2)\right] dt_1\, dt_2$$

$$S_{xy}(f_1, f_2) = \int_{-\infty}^{\infty}\int R_{xy}(t_1, t_2) \exp\left[j2\pi(f_1 t_1 - f_2 t_2)\right] dt_1\, dt_2 \qquad (34)$$

The quantities $S_x(f_1, f_2)$ and $S_y(f_1, f_2)$ are called *generalized (nonstationary) spectral density functions*, while $S_{xy}(f_1, f_2)$

is called a *generalized (nonstationary) cross-spectral density function.* They are defined for both positive and negative f_1 and f_2. Observe that the definition for $S_{xy}(f_1, f_2)$ includes $S_x(f_1, f_2)$ and $S_y(f_1, f_2)$ as special cases. The inverse double Fourier transform relations for $S_{xy}(f_1, f_2)$ is

$$R_{xy}(t_1, t_2) = \int_{-\infty}^{\infty}\int S_{xy}(f_1, f_2) \exp\left[-j2\pi(f_1 t_1 - f_2 t_2)\right] df_1\, df_2 \tag{35}$$

A major limitation in the usefulness of the double-frequency spectra in engineering applications is that these quantities are not subject to direct measurement. Instead, the associated nonstationary correlation functions must be computed for all possible time pairs, and then a double Fourier transform must be computed for the various frequency pairs of interest. In spite of this limitation, however, significant theoretical studies can be carried out which depend on this representation. This is indicated in Section VI, which deals with input-output relations. In practical problems, it is convenient to allow these spectral density functions to include delta functions.

B. Time-Varying Power Spectra

Let $x_i(t)$ be a sample function of a zero mean, nonstationary process $\{x(t)\}$ which is operated upon as shown in Fig. 5.

Assume that the filter is an ideal rectangular filter with a frequency response function given by

$$\begin{aligned} H(\nu) &= 1 \quad , \quad f - \frac{B}{2} < |\nu| < f + \frac{B}{2} \\ &= 0 \quad , \quad \text{elsewhere} \end{aligned} \tag{36}$$

where B is the filter bandwidth and f is the center frequency. Although the ideal filter is not physically realizable, it is a good approximation of many narrow-band filters.

The corresponding weighting function is simply the inverse Fourier transform of $H(\nu)$ so that

$$\begin{aligned} h(\tau) &= \int_{-\infty}^{\infty} H(\nu)\, e^{i2\pi\nu\tau}\, d\nu \\ &= \frac{\sin(\pi B\tau)}{\pi\tau} \cos(2\pi f\tau) \end{aligned} \tag{37}$$

The filter output is given by

$$x_i(t, f, B) = \int_{-\infty}^{\infty} h(\xi)\, x_i(t-\xi)\, d\xi \tag{38}$$

and is the instantaneous value of that part of $x_i(t)$ which lies within the bandwidth B centered at the frequency f. The output of the squaring device is

$$\begin{aligned} x_i^2(t, f, B) &= \left(\int_{-\infty}^{\infty} h(\xi)\, x_i(t-\xi)\, d\xi\right)^2 \\ &= \int_{-\infty}^{\infty} h(\xi)\, h(\eta)\, x_i(t-\xi)\, x_i(t-\eta)\, d\xi\, d\eta \end{aligned} \tag{39}$$

It should be noted that the right side of Eq. (39) is always non-negative since it represents the square of a real function of time. Also, f and B are contained in $h(\xi)$ and $h(\eta)$, and $f \geq 0$ for real narrow-band filters. This requirement on f will be assumed henceforth so as to derive a one-sided spectrum.

By recording $x_i^2(t, f, B)$ as a function of time, repeating the filter-square-record operation N times, and ensemble averaging, an estimate of the average value of $x^2(t, f, B)$ is obtained. Letting N become arbitrarily large will cause the estimate to converge to the true average value. Thus,

$$E\left[x^2(t, f, B)\right] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i^2(t, f, B) \tag{40}$$

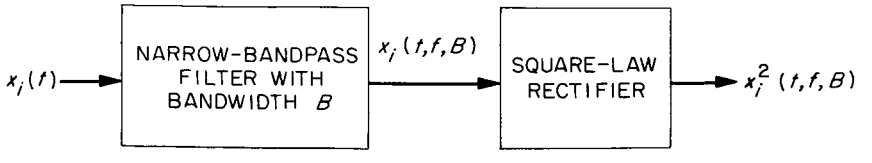

Fig. 5. Filter-squaring operation

Division by the bandwidth B and letting B approach zero now yields a physically meaningful nonstationary time-varying power spectrum:

$$G_x(t,f) = \lim_{B\to 0} \frac{E\left[x^2(t,f,B)\right]}{B} = \lim_{\substack{N\to\infty \\ B\to 0}} \frac{1}{BN} \sum_{i=1}^{N} x_i^2(t,f,B) \tag{41}$$

where the one-sided power spectral notation $G_x(t,f)$ states the fact that $f \geq 0$. If it is desired for mathematical reasons to include negative f, then $G_x(t,f)$ would be replaced in the usual way by the two-sided power spectrum $S_x(t,f)$, where half the power is distributed in the negative frequency range. Observe that $G_x(t,f)$ or $S_x(t,f)$ will be non-negative for all t and f, in agreement with physical requirements for a meaningful spectrum. In terms of $G_x(t,f)$ or $S_x(t,f)$, the mean square value $E\left[x^2(t)\right]$ can be calculated by carrying out the integration

$$E\left[x^2(t)\right] = \int_0^\infty G_x(t,f)\,df = \int_{-\infty}^{\infty} S_x(t,f)\,df \tag{42}$$

The mean-square-value portion in any positive frequency range (f_1, f_2) can be found by integrating $G_x(t,f)$ over that frequency range.

For finite B, an estimate of $G_x(t,f)$ from a set of N records is given clearly by

$$\hat{G}_x(t,f) = \frac{1}{BN} \sum_{i=1}^{N} x_i^2(t,f,B) \tag{43}$$

The estimate will always be non-negative for all t and f, and may be measured as illustrated in Fig. 6. Equation (41) shows that this estimate approximates the true value only if both $N\to\infty$ and $B\to 0$. Thus, a bias error will occur as well as a statistical error for finite B and N.

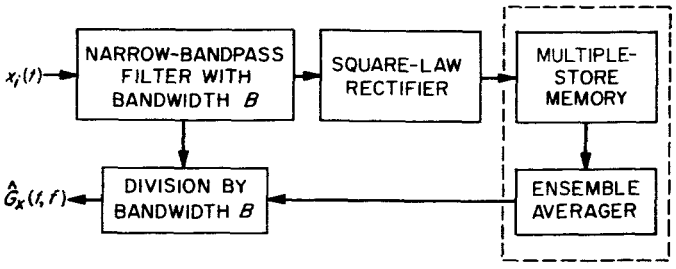

Fig. 6. Time-varying power spectrum measurement

To measure $\hat{G}_x(t,f)$ accurately, one would need a large collection of records if an ensemble averaging is to be performed. This may be a major problem for some applications. Another limitation on the usefulness of $\hat{G}_x(t,f)$ is that it must be determined for each frequency of interest. Thus, the entire sequence of filtering, squaring, storing, and averaging must be performed at some initial frequency f_1, and then repeated at all other frequencies of interest. The result of this processing is to generate a family of one-dimensional functions, $G_x(t,f_j)$, $j = 1, \cdots, M$, which approximates the two-dimensional function $G_x(t,f)$. Short-time averaging or orthogonal-function averaging can replace ensemble averaging, as indicated by the dotted lines in Fig. 6.

C. *Time-Averaged Power Spectra*

A function related to the time-varying power spectrum $G_x(t,f)$, and which may have important applications in the characterization of nonstationary processes, is the time-averaged power spectrum defined for $f \geq 0$ by

$$\bar{G}_x(f,T) = \frac{1}{T}\int_T G_x(t,f)\,dt \tag{44}$$

where T is the interval of interest. The value of T could be the entire duration of the nonstationary process, in which case a limiting operation on T may be involved, or some smaller interval, such as a tape-loop length, in which the spectral structure of the process is of particular interest. Clearly, $\bar{G}_x(f,T)$ is non-negative for all f and T.

When T approaches infinity, $\bar{G}_x(f,T)$ will be replaced by $\bar{G}_x(f)$ where

$$\bar{G}_x(f) = \lim_{T\to\infty} \bar{G}_x(f,T) = \lim_{T\to\infty} \frac{1}{T}\int_T G_x(t,f)dt \tag{45}$$

If the two-sided spectrum $S_x(t,f)$ is used in place of $G_x(f,t)$, then $\bar{G}_x(f,T)$ should be replaced by $\bar{S}_x(f,T)$. This in turn, becomes $\bar{S}_x(f)$ as T approaches infinity.

For a sample function $x(t)$ of length T from a nonstationary process, an estimate of $\bar{G}_x(f,T)$ is given by

$$\hat{\bar{G}}_x(f,T) = \frac{1}{B_e T}\int_T x^2(t,f,B)\,dt \tag{46}$$

Equations (41) and (46) show

$$\bar{G}_x(f,T) = \lim_{B\to 0} \frac{E[\hat{\bar{G}}_x(f,T)]}{B} \tag{47}$$

Thus, $\hat{\bar{G}}_x(f,T)$ will be a biased estimate of $\bar{G}_x(f,T)$ for finite B. A procedure for measuring $\hat{\bar{G}}_x(f,T)$ is illustrated in Fig. 7. This is precisely the same procedure as one would follow for stationary data.

Problems which are of concern here deal with questions of the length of T required to give an accurate estimate of $\bar{G}_x(f,T)$, and the sampling distribution of these estimates for different nonstationary processes. The bias effects of finite B need to be investigated. These matters are not easy to solve analytically and may require considerable experimental testing for various cases of interest. The time-averaged power spectrum of Fig. 7, however, is simple to measure and can result in substantial savings in processing costs in some cases. The finding of appropriate short averaging times and bandwidths to use for different nonstationary processes can lead to a meaningful interpretation of how nonstationary power spectra change with time.

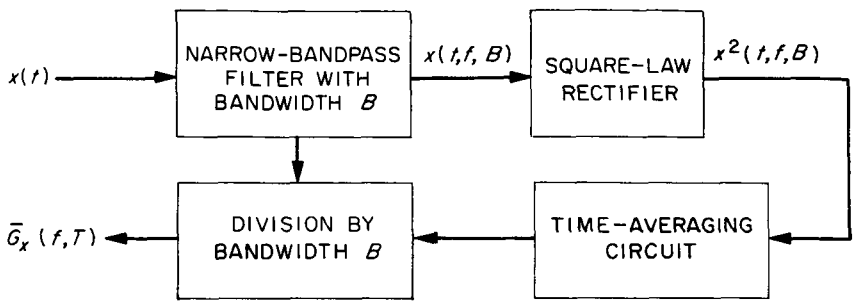

Fig. 7. Time-averaged power spectrum measurement

VI. INPUT-OUTPUT RELATIONS FOR NONSTATIONARY DATA

Consider sample functions $x(t)$ from a nonstationary random process acting as input to a constant-parameter linear system with weighting function $h(\tau)$ and frequency-response function $H(f)$. By definition,

$$H(f) = \int_0^\infty h(\tau)\, e^{-j2\pi f\tau}\, d\tau \tag{48}$$

with $h(\tau) = 0$ for $\tau < 0$ when the system is physically realizable, which will be assumed here. The output $y(t)$ will also belong to a nonstationary random process and is given in the time domain by

$$y(t) = \int_0^\infty h(\tau)\, x(t-\tau)\, d\tau \tag{49}$$

and in the frequency domain by

$$Y(f) = H(f)\, X(f) \tag{50}$$

where $X(f)$ and $Y(f)$ are the Fourier transforms of $x(t)$ and $y(t)$, respectively, assuming, of course, that they exist.

A more general result occurs for nonstationary double-frequency spectra which does not require existence of $X(f)$ and $Y(f)$, but merely existence of the nonstationary spectra. In terms of the double-frequency spectra defined by Eq. (33) and (34), the following generalized input-output nonstationary spectral relation is derived in Ref. 2:

$$S_y(f_1,f_2) = H^*(f_1)\, H(f_2)\, S_x(f_1,f_2) \tag{51}$$

where $H^*(f)$ is the complex conjugate of $H(f)$. For the special case of stationary data, it is shown that $S_x(f_1, f_2)$ and $S_y(f_1, f_2)$ are nonzero only for $f = f_1 = f_2$. Here Eq. (51) reduces to the familiar result

$$S_y(f) = |H(f)|^2 S_x(f) \tag{52}$$

The generalized nonstationary cross-spectral relation between input and output is given by

$$S_{xy}(f_1, f_2) = H(f_2) S_x(f_1, f_2) \tag{53}$$

For the special case of stationary data, this reduces to the familiar result

$$S_{xy}(f) = H(f) S_x(f) \tag{54}$$

Equations (51) and (53) indicate how double-frequency spectra should be used to describe input-output relations for nonstationary data. Their calculation for actual problems can become quite complicated in spite of the simple appearance of these formulas.

REFERENCES

1. Bendat, J. S., L. D. Enochson, G. H. Klein, and A. G. Piersol, *The Application of Statistics to the Flight Vehicle Vibration Problem,* ASD TR 61-123, Air Force Systems Command, Aeronautical Systems Division, USAF, Wright-Patterson AFB, Ohio, December 1961 (AD 271 913).

2. Bendat, J. S., L. D. Enochson, G. H. Klein, and A. G. Piersol, *Advanced Concepts of Stochastic Processes and Statistics for Flight Vehicle Vibration Estimation Measurement,* ASD TDR 62-973, Air Force Systems Command, Aeronautical Systems Division, USAF, Wright-Patterson AFB, Ohio, December 1962 (AD 297 031).

3. Piersol, A. G., *Nonparametric Tests for Equivalence of Vibration Data,* Paper No. 748C, SAE National Meeting, Los Angeles, Calif. September 1963.

4. Thrall, G. P., *Mean Square Measurements of Nonstationary Random Processes,* Paper No. 925D, SAE National Meeting, Los Angeles, Calif., October 1964.

5. Courant, R., and D. Hilbert, *Methods of Mathematical Physics,* Vol. I, Interscience Publishers, New York, 1953, p. 51.

6. Bendat, J. S., and G. P. Thrall, *Spectra of Nonstationary Random Processes,* Report No. AFFDL-TR-64-198, Air Force Systems Command, Research and Technology Division, Wright-Patterson AFB, Ohio, February 1965.

7. Thrall, G. P., and J. S. Bendat, *Mean and Mean Square Measurements of Nonstationary Random Processes,* NASA CR-226, National Aeronautics and Space Administration, Washington, D. C., May 1965.