ESD TDR 64-338
ESTI FILE COPY

ESD-TDR-64-338

Group Report 1964-32

The Periodic Analysis of Sampled Data

E. Korngold

15 June 1964

Prepared under Electronic Systems Division Contract AF 19(628)-500 by

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Lexington, Massachusetts

AD0601942

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the support of the U.S. Air Force under Contract AF 19(628)-500.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

THE PERIODIC ANALYSIS OF SAMPLED DATA

E. KORNGOLD

Group 22

GROUP REPORT 1964-32

15 JUNE 1964

LEXINGTON MASSACHUSETTS

ABSTRACT

This report is a brief survey of some currently used techniques of periodic analysis of sampled data. Almost every periodic analysis of empirical data eventually relies on one of the following three techniques:

1. Regression analysis on a trigonometric polynomial
2. Periodogram analysis
3. The estimation of power spectra.

Each of these techniques is best applied under distinct modelling assumptions, and an attempt has been made to discuss their applicability and their range of validity under various circumstances.

After a brief introduction in Chapter I, Chapter II proceeds to describe the models in use and the statistical properties of the estimates associated with them The effects of sampling and some of their implications in practical situations are treated in Chapter III. Chapter IV relates several common question areas to the theory developed in the earlier portions of the report, and presents some illustrations of the power of the techniques discussed. Finally, Chapter V gives the briefest outline of the general requirements of a comprehensive program in time series data analysis. Numbers in square brackets refer to entries, in alphabetic order, in the appended bibliography.

Accepted for the Air Force
Franklin C. Hudson, Deputy Chief
Air Force Lincoln Laboratory Office

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd.)

LIST OF FIGURES

I. INTRODUCTION

In the processing of empirical data, a frequency analysis is often required as a partial characterization of the process generating the data. In particular, one hopes, on the basis of such an analysis, to draw some inference on the presence of periodicities in the underlying process. Two approaches to frequency analysis (harmonic analysis and spectral estimation) are commonly taken, sometimes without full awareness of their implication and range of validity. The sources of ambiguity are several:

1. There may be a basic ignorance about the nature of the underlying phenomenon, in which case the choice of a suitable model becomes the main stumbling block. Indeed, the following kinds of alternatives have to be considered, among others:
   - the phenomenon may be periodic, in which case all frequencies are multiples of a single base frequency;
   - the phenomenon may be aperiodic, in which case frequencies of interest may be either discrete or continuous;
   - the phenomenon may be thought of as a single unique occurrence, or it may be taken as one of many possible similar realizations.
2. Each of these alternatives represents a different point of view and requires different techniques in the analysis of data and the interpretation of results. In particular, there is an apparent formal similarity between conventional techniques of Fourier Analysis and the estimation of power spectra, which leads to the temptation of using one for the other although they apply to distinctly different models and have essentially different statistical properties. This gives rise to modelling uncertainty and the attendant problems in interpretation of results.
3. The current literature on the subject of frequency analysis of empirical time series draws its techniques and terminology from at least three fields:
   - Classical Statistical Theory, [3, 11, 12, 16]
   - Purely mathematical disciplines such as approximation theory, [1] representation theory, [6] Fourier Analysis and related fields, [19, 25, 29]

-The frequency analysis of linear systems in communications engineering, [14, 19, 29].

Here, unfamiliarity with one or another language is the stumbling block.

This report has been written in an attempt to elucidate the significant relations governing the applicability and validity of the theory. The subject to be covered is large and growing, so that our presentation is anything but complete. Moreover, the theory requires some fairly advanced mathematical tools; therefore the presentation is primarily discursive. Every argument can however be rigorously demonstrated without alteration of content.

We shall be considering primarily zero mean second order stationary stochastic processes, as there exists hardly any theory for evolutive* time series. Almost all references dealing with the latter generally convert them to something resembling stationarity by means of some elementary statistical device.

Chapter II describes the models which are most commonly employed in periodic analysis. These are a cyclic periodic model, a cyclic aperiodic model and an oscillatory model. The cyclic models are presented primarily because they offer a good background for the treatment of the more sophisticated oscillatory model. Emphasis will be placed on the statistical properties of the various estimation techniques presented.

Chapter III will treat several consequences of sampling and the implications which follow in practical applications. In particular the phenomena of folding and aliasing will be described.

Chapter IV has a two fold purpose: on the one hand several common question areas will be elucidated in the light of the foregoing theory; on the other, several illustrations will be given of the power of the techniques described in the earlier chapters.

Finally, Chapter V outlines briefly the necessary data processing and analysis areas which should be included in a comprehensive program of periodic analysis of empirical data.

*An evolutive time series is one with a non zero time varying mean.

II. MODELS

We shall examine three models which find most frequent application. The first two are cyclic models; they are the simpler of the three and have been studied extensively in the last forty years. They are based on the assumption that the time series to analyze is the sum of a deterministic component and white noise. The deterministic component is a finite sum of sine and cosine terms whose frequencies may or may not be multiples of a single base frequency. The third model, known as oscillatory, is based on the assumption that the time series is a realization of a stochastic process possessing a continuous power spectrum. A heuristic difference between cyclic and oscillatory models is that for cyclic processes, different realizations are expected to look alike except for small fluctuations, whereas two single realizations of an oscillatory process may look entirely different from each other. Another way of stating this is to say that different realizations of a cyclic process agree in mean whereas different realizations of an oscillatory process agree in quadratic mean. While the first two models rest mainly on the application of classical techniques, the third model has only been developed vigorously during the last fifteen years and is still the subject of current research.

A fourth model is of great practical interest and deals with processes which are a mixture of cyclic and oscillatory processes; it assumes a superposition of oscillatory and cyclic processes and a spectrum which need not be continuous. Analysis of this kind of model is almost non-existent in the literature but it appears* that work is currently in progress on spectral estimates which converge to the derivative of the integrated spectrum where it exists, and to the magnitude of the jump where continuity fails. This kind of model shall not be considered in this report.

In general, a realization of a continuous process $\{x_t\}$ will be represented as x(t). A discrete process $\{x_t\}$ will have a realization x_t. In the present chapter, we shall not distinguish between a realization of a discrete process and a sample of a realization of a continuous process; both will be written as x_t. For the present this is a minor distinction although we shall see later precisely under what conditions it becomes important.

A. First and Second Model: Cyclic Structure

We shall assume in this case that the time series x_t is generated by a process of the form:

*Doctoral Dissertion in progress under E. Parzen. (Private communication by L. Gardner).

$$x_t = \sum_{j=1}^{q} (a_j \cos \lambda_j t + b_j \sin \lambda_j t) + \xi_t$$

where* $0 \leq \lambda_j < \pi$, a_j and b_j are real and ξ_t is Gaussian, uncorrelated, with mean zero and variance σ^2. There are two possibilities open at this point:

(1) We can assume that the frequencies λ_j are all multiples of a base frequency known a priori.

(2) We know neither the frequencies nor the amplitudes of the process and are trying to estimate both, and in particular, the frequency associated with the peak amplitude.

The first possibility leads to a straightforward application of linear regression theory [16,Vol II, pp. 141-174] with trigonometric functions substituted for polynomials, and is equivalent to a least squares fit of an empirical function by a trigonometric polynomial [34, pp. 14-19] . The second possibility leads to the calculations of the periodogram and is commonly referred to as "Harmonic Analysis" of empirical data.

1. First Interpretation: Frequencies Known, Amplitudes Unknown [25, Ch I]

In this case, the process is a superposition of white noise and a periodic deterministic component. All frequencies are multiples of a base frequency and we can write

$$x_t = \alpha_0 + \sum_{j=1}^{q} (\alpha_{\nu_j} \cos \frac{2\pi\nu_j}{p} t + \beta_{\nu_j} \sin \frac{2\pi\nu_j}{p} t) + \xi_t .$$

---

* λ_j is restricted to be less than π because in the concrete situation of a sample of size T, we have $\lambda_j = \frac{2\pi j}{T}$ and j cannot be allowed to exceed $\frac{T}{2}$, (resulting in an upper bound of π for λ_j) lest there be more than T coefficients to evaluate on the basis of a sample of size T.

We assume $m_t = x_t - \xi_t$ to be periodic with period p, $m_{t+kp} = m_t$, and further, that the number of observations T is a multiple (large usually) of the period p. ξ_t is white noise as previously described. The ν_j, $j=1\rightarrow q$ are a subset of the sequence $1, 2 \ldots, p$ and are used to indicate that not every harmonic need be represented*. What is wanted is an estimate of the coefficients $\alpha_{\nu_j}, \beta_{\nu_j}$. Since ξ_t is Gaussian, the least squares estimate is the maximum likelihood estimate of α_{ν_j} and β_{ν_j} and we find that

$$A_o = \frac{1}{T} \sum_{t=1}^{T} x_t$$

$$A_{\nu_j} = \frac{2}{T} \sum_{t=1}^{T} x_t \cos \frac{2\pi\nu_j}{p} t$$

$$B_{\nu_j} = \frac{2}{T} \sum_{t=1}^{T} x_t \sin \frac{2\pi\nu_j}{p} t \quad , \quad j = 1\rightarrow q$$

are the estimates of the α's and β's.

Furthermore, the A_{ν_j} and B_{ν_j} are independent normal variables with means α_{ν_j} and β_{ν_j} respectively, and their variances are $\frac{2\sigma^2}{T}$, which go to 0 as sample size $T\rightarrow\infty$.

Setting the squared amplitude at a given frequency to

$$R^2_{\nu_j} = A^2_{\nu_j} + B^2_{\nu_j} \quad ,$$

it can be shown that $R^2_{\nu_j}$ is an estimate of the true squared amplitude $\rho^2_{\nu_j} = \alpha^2_{\nu_j} + \beta^2_{\nu_j}$ with the properties

---

* We have chosen the ν_j scheme of indexing the α's and β's rather than letting $j = 1\rightarrow p$ and allowing some a_j's and b_j's to be zero because the presence of the noise ξ_t will insure that none of the estimated α's and β's will be zero, whereas we might know a priori that some of them are.

$$\mathcal{E}(R^2_{\nu_j}) = \rho^2_{\nu_j} + \frac{4\sigma^2}{T}$$

$$\mathrm{Var}\, R^2_{\nu_j} = \frac{8\sigma^2}{T}\left[\rho^2_{\nu_j} + \frac{2\sigma^2}{T}\right]$$

In other words, $R^2_{\nu_j}$ is a consistent asymptotically unbiased estimate* of the squared amplitude of the ν_j th harmonic.

It may be pointed out that very rarely is there enough information to assume, in applications, that the fundamental frequency is known. The analysis of this model is thus primarily a theoretical exercise in regression analysis.

2. Second Interpretation: Amplitude and Frequencies Unknown

This model, as the previous one, is a finite parameter model of the form

$$x_t = \sum_{j=1}^{q} \left(\alpha_j \cos \omega_j t + \beta_j \sin \omega_j t\right) + \xi_t \quad , \quad t = 1 \rightarrow T \quad .$$

The constant term has been omitted as it can be estimated and removed. Again, ξ_t is white Gaussian noise; the only restriction** on the ω_j is

$$0 < \omega_j \leq \pi \;;$$

no longer are the ω_j harmonically related. The problem this time is to extract the periods $2\pi/\omega_j$ from the sequence x_t.

The technique of the periodogram proceeds as follows:

---

*Unbiased $\triangleq$ the mean value of the estimate equals the true value.
Consistent $\triangleq$ the estimate converges in probability to the true value. Sufficient for consistency is that Var(estimate) $\rightarrow 0$ as sample size $\rightarrow \infty$.
Asymptotically $\triangleq$ ... as the sample size goes to ∞.
The symbol $\triangleq$ is to be read as: "means by definition".

**cf. footnote page 4.

Calculate the periodogram

$$\Phi_T(\omega) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_t e^{i\omega t}\right|^2 = \frac{T}{8\pi}\,[A^2(\omega) + B^2(\omega)]$$

at all points

$$\omega_\nu = \frac{2\pi\nu}{T}, \quad \nu = 1 \rightarrow \frac{T}{2} - 1, \qquad \text{(even T)},$$

where

$$A(\omega) = \frac{2}{T}\sum_{t=1}^{T} x_t \cos \omega t$$

$$B(\omega) = \frac{2}{T}\sum_{t=1}^{T} x_t \sin \omega t$$

are the Fourier coefficients of the data relative to a trial frequency ω. In short, x_t is subjected to an ordinary Fourier Analysis as if the noise term ξ_t were absent.

Now, it is a characteristic property of the periodogram $\Phi_T(\omega)$ that for large samples it is large whenever ω is equal to one of the frequencies ω_j and that it is small whenever $\omega \neq \omega_j$ [11, p. 92]*. Thus, as one expects large peaks (for large samples) in $\Phi_T(\omega)$ whenever $\omega = \omega_j$, it is natural to say:

> "Let us calculate $\Phi_T(\omega)$ at all frequencies of interest and single out the peaks in $\Phi_T(\omega)$, inferring therefrom that a frequency component is predominantly present in the process x_t wherever a peak occurs in $\Phi_T(\omega)$."

The question of the statistical significance of the peaks in the periodogram arises immediately. Several tests of significance have been developed and are briefly described in [15]. Ideally, one should like to state that peaks occur if and only

---

*Heuristically, the periodogram is the square of the amplitude c_ν of the ν^{th} component; if we write the ν^{th} component as $c_\nu \cos(\alpha_\nu t + \varphi_\nu)$. It is known that when $\xi_t = 0$,

$$\frac{1}{T}\sum_{t=1}^{T} x_t^2 = \frac{1}{2}\sum_{\nu} c_\nu^2 \quad ,$$

so that c_ν^2 is proportional to the contribution of the ν^{th} component to the total energy dissipated by the process (cf. Parseval's relation and completeness).

if a corresponding frequency is present in the signal; this however is not so, as can be directly inferred from the degradation imposed by the noise term ξ_t. We shall mention only one test, due to Fisher, which is independent of sample size.

To assess the significance of the largest peak in the periodogram compute the statistic*

$$g = \Phi_T(\omega_{max}) \Big/ \sum_{j=1}^{\frac{1}{2}(T+1)} \Phi_T(\omega_j) ,$$

where ω_{max} is the frequency corresponding to the maximum amplitude of $\Phi_T(\omega)$, where T is odd and $\omega_j = \frac{2\pi j}{T}$. The distribution of g can be explicitly calculated. Details of usage are given in Ref. [11, pp. 91 ff]. Other tests are referred to in Ref. [15].

The distribution of the values of the periodogram of a set of T random, uncorrelated, normal (0, σ^2) numbers is $\frac{\sigma^2}{4\pi} \chi_2^2$. The expected value of the periodogram is $\frac{\sigma^2}{2\pi}$, its variance is $\frac{\sigma^4}{4\pi^2}$. The probability that the periodogram exceed k times its expected value is

$$P\left\{\Phi_T(\omega) > k\frac{\sigma^2}{2\pi}\right\}$$

$$= P\left\{\frac{\sigma^2}{4\pi}\chi_2^2 > k\,\frac{\sigma^2}{2\pi}\right\}$$

$$= P\left\{\frac{1}{2}\chi_2^2 > k\right\} = e^{-k} .$$

This is the basis of Shuster's original idea of using the periodogram to reject the hypothesis - whenever $\Phi_T(\omega)$ is large - that no periodic component is present in the data at frequency ω. An extensive exposition of periodogram analysis is given by Bartels in Ref. [15] and Stumpf in Ref [25].

B. <u>Third Model: Oscillatory Processes</u>

In very few cases is it reasonable on apriori grounds to decide that a process is of the finite parameter type. The previous section briefly described two such models:

- a periodic model where all frequencies were an integral multiple of a base frequency, sampled over an interval containing several periods;
- an aperiodic model where, while discrete, the spectral components were no longer harmonically related.

We will now consider the following:

- an aperiodic model with a continuous spectrum.

The cornerstone of the development is the well known theorem which asserts that for a wide class of stochastic processes the autocorrelation function and the power spectral density are completely equivalent descriptions of the second order properties of the process.

1. <u>Preliminaries</u>: In general, if $\hat{f}(\omega_o)$ is an estimate of $f(\omega_o)$, we shall use, as an evaluation criterion of $\hat{f}$, the mean square error

$$\text{MSE}\,\hat{f} = \text{Bias}^2\hat{f} + \text{Var}\,\hat{f}$$

where

$$\text{Bias}\,\hat{f} = \mathcal{E}(\hat{f} - f)$$
$$\text{Var}\,\hat{f} = \mathcal{E}(f - \mathcal{E}\hat{f})^2$$

There are several other criteria developed in the literature [15, 18, 31]. We shall not pursue these.

We recall that

$\hat{f}$ is unbiased if $\text{Bias}\,\hat{f} = 0$

$\hat{f}$ is asymptotically unbiased if $\text{Bias}\,\hat{f} \to 0$ as sample size $T \to \infty$

$\hat{f}$ is called consistent if it is unbiased and $\text{Var}(\hat{f}) \to 0$ as $T \to \infty$.

Sometimes, in the sequel, we shall write $f(\omega_o)$ instead of $f(\omega)$. This is done to stress the fact that we are concerned with pointwise estimates, i.e., we are estimating a single value, the value that f takes at ω_o. We are not evaluating a function. The estimation of the function is done pointwise, one by one, in practical estimations. In this chapter the emphasis is on statistical properties of estimates.

We shall be concerned with zero mean second order stationary processes having continuous spectra and possessing no periodic components corresponding to spectral lines*. Roughly speaking, a stochastic process is a collection of functions $\{x_t\}$. Each function is a possible realization of the process and it's occurrence is thought to be governed by a probability law. A zero mean process obeys the relation:

$$\mathcal{E} x_t = 0 \quad \text{for all } t.$$

Stationarity implies that the probabilities governing the occurrence of certain values

$$x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(n)}$$

for any n, are independent of the time origin, and that these values are thus just as likely to occur at one or the other of the two time sequences

$$t_1, t_2, \ldots, t_n$$

$$t_1+h, t_2+h, \ldots, t_n+h,$$

regardless of the shift h. Second order stationarity means that all moments up to moments of second order are stationary. In particular,

$$\mathcal{E}\left[x_t x_{t+|k|}\right] = \gamma(k)$$

depends only on the lag k. The function $\gamma(k)$ is the autocovariance function of the process. It is sometimes normalized by dividing by $\gamma(0) = \sigma^2$ and the resulting function is the autocorrelation function. A Gaussian process is completely specified by its first and second order properties. Most of the theory is developed for Gaussian processes. If the Gaussian assumption is dropped, results pertaining to expected values are not affected; but results pertaining to the variability of estimates become approximations.

*Processes with mixed spectra have been very scantily treated. One reference is [30]; also, see footnote on page 3.

There are two pairs of alternative ways of presenting results, corresponding to the real-complex presentation on the one hand, and the continuous-discrete cases on the other. The first dichotomy is purely a matter of convenience in notation. The second corresponds to the distinction between Fourier series and Integrals (the Fourier series' coefficients are the Fourier transform of a periodic function). We shall stay with the discrete case as this is the one most frequently encountered in practice.

The basic theorem states that under rather general conditions insuring that the covariance function $\gamma(k) = \mathcal{E}\, x_t x_{t+k} = \gamma(-k)$ goes to zero sufficiently rapidly as the lag goes to infinity, the second order zero mean stationary process $\{x_t\}$ possesses a spectrum

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\omega} \gamma(k)$$

which is the Fourier transform of the autocovariance $\gamma(k)$, and the converse relation,

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\omega} f(\omega)\, d\omega, \qquad k = 0, \pm 1, \pm 2, \ldots$$

also holds. (This is the discrete version of the Wiener-Khintchin theorem [34, pp. 66 ff]).

2. <u>An Algebraic Identity:</u> Given the discrete sample (at equidistant intervals)

$$x_1, x_2, x_3, \ldots, x_t, \ldots, x_T$$

the sample covariance at lag k is defined as

$$\varphi_T(k) = \begin{cases} \dfrac{1}{T} \displaystyle\sum_{t=1}^{T-|k|} x_t x_{t+|k|} & \text{for } k = 0, \pm 1, \ldots, \pm (T-1)^* \\ 0 & \text{for } |k| \geq T \end{cases} .$$

Recall that the periodogram is given by

$$\Phi_T(\omega) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} x_t e^{i\omega t} \right|^2 .$$

---

*Note that for large lags this formula for $\varphi(k)$ yields a poor estimate of $\gamma(k)$.

The identity in question reads

$$\Phi_T(\omega) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_t e^{i\omega t}\right|^2 = \frac{1}{2\pi}\sum_{|k|<T} \varphi_T(k) e^{-i\omega k} \tag{1}$$

regardless of assumptions on x_t [12, p. 52]. Thus $\Phi_T(\omega)$ and $\varphi_T(k)$ are a Fourier pair. Now, we have seen that for a zero mean, second order stationary process with covariance sequence

$$\gamma(k) = \mathcal{E}\left[x_t x_{t+|k|}\right] = \gamma(-k)$$

a spectral density exists as

$$f(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \gamma(k) e^{-i\omega k} \quad . \tag{2}$$

Two Fourier pairs — one for the sample, one for the process — are now in evidence, as shown in Table 1.

	Fourier Transform of Function	
Sample	$\Phi_T(\omega)$	$\varphi_T(k)$
Process	$f(\omega)$	$\gamma(k)$

Table 1 Two Fourier Pairs

Note that the $\varphi_T(k)$ provide an asymptotically unbiased estimate of the $\gamma(k)$ as follows:

$$\mathcal{E}\,[\varphi_T(k)] = \mathcal{E}\left[\frac{1}{T}\sum_{t=1}^{T-|k|} x_t x_{t+|k|}\right] = \frac{1}{T}\sum_{t=1}^{T-|k|} \mathcal{E}\left[x_t x_{t+|k|}\right]$$

$$= \frac{1}{T}\sum_{t=1}^{T-|k|} \gamma(k) = \left(1 - \frac{|k|}{T}\right)\gamma(k)$$

i.e., $\varphi_T(k) \xrightarrow{P} \gamma(k)$; the sample covariances converge in probability* to the process covariances. Taking expected values on both sides of (1), we have

$$\mathcal{E}\left[\Phi_T(\omega)\right] = \mathcal{E}\left[\frac{1}{2\pi}\sum_{|k|<T} \varphi_T(k)e^{-i\omega k}\right]$$

$$= \frac{1}{2\pi}\sum_{|k|<T} \left\{\mathcal{E}\ \varphi_T(k)\right\} e^{-i\omega k}$$

$$= \frac{1}{2\pi}\sum_{|k|<T} \left(1 - \frac{|k|}{T}\right)\gamma(k)e^{-i\omega k} \quad .$$

Comparing with (2), we obtain

$$\lim_{T\to\infty} \mathcal{E}\ \Phi_T(\omega) = f(\omega) \quad .$$

Hence, for large samples, the periodogram appears to be a natural estimator of the power spectrum. Surprisingly, this is not the case. It has been found by experience that a harmonic analysis of white noise (flat spectrum) produces a highly spiked spectrum (cf. IV, B, 4). The reason for this is that for fixed ω, regardless of the character of the noise,

$$\lim_{T\to\infty} \operatorname{Var} \Phi_T(\omega_o) = 2f^2(\omega_o),$$

whereas, for all well behaved estimators the variance should go to zero as sample size $T\to\infty$. It can in fact be shown that for Gaussian processes the ratio $2\Phi_T(\omega_o)/f(\omega_o)$ has a chi-square distribution (as $T\to\infty$) with 2 degrees of freedom. This result stresses the futility of appraising the significance of peaks in $\Phi_T(\omega)$ even as T increases, when

---

* $f_T \xrightarrow{P} f \triangleq \lim_{T\to\infty} P\{\,|f_T - f| > \epsilon\} = 0$; i.e., the probability that f_T differs from f by more than ϵ goes to 0 as T goes to ∞.

the spectrum of the process has a continuous component [2, p. 4].

We shall turn next to the development of consistent estimates of the spectrum.

3. Consistent Estimates (1): Variance Reduction: As an estimate of $f(\omega_o)$, the periodogram has a non-zero variance independent of sample size. Bartlett conceived the idea (1948) of using a variance reducing scheme in which he computes his estimate of the power spectrum as the average of $\frac{T}{m}$ periodograms, each computed for a sub-series of length m. If the variance of the periodogram estimate of $f(\omega_o)$ is σ_T^2, the variance of the new estimate may be expected to be of the order of $\frac{m}{T}\sigma_T^2$. We note here that as more segments are collected, the variance is reduced. In fact, we have a consistent estimate, for as $T \to \infty$ and as m remains constant, var $\to 0$. This fact has been the starting point of renewed interest in the theory of spectral estimation. We shall see later that a price is paid for the consistency in the guise of an asymptotic bias.

It can be shown that Bartlett's periodogram averaging scheme is equivalent to a modification of the classical periodogram computation consisting in the introduction of weighting factors for the sample autocovariances. Specifically, instead of calculating

$$\frac{1}{2\pi} \sum_{|k|<T} \varphi_T(k) e^{-i\omega_o k}$$

as an estimate of $f(\omega_o)$, one uses

$$f_T(\omega_o) = \sum_{|k|<T} h(k)\, \varphi_T(k) e^{-i\omega k}$$

where

$$h(k) = \begin{cases} 1 - \frac{|k|}{m} & \text{for } |k| \leq m \\ 0 & \text{for } |k| > m \end{cases} .$$

In this manner, an entire family of spectral estimates has been introduced depending on the properties of the weighting function h(k).

4. <u>Consistent Estimates (2): Smoothing Windows.</u> Although $\Phi_T(\omega)$ is not pointwise consistent; i.e., for fixed ω_o does not converge in probability* to $f(\omega)$ as $T \to \infty$, it is known that [12, p. 58]

$$\int_{\lambda_1}^{\lambda_2} \Phi_T(\omega)\, d\omega \xrightarrow{P} \int_{\lambda_1}^{\lambda_2} f(\omega)\, d\omega \quad ,$$

which shows that the averaged spectrum over an interval can be consistently estimated. The left side may be modified slightly to read

$$f(\omega_o) = \int_{-\infty}^{\infty} W(\omega)\, \Phi_T(\omega)\, d\omega \quad ,$$

with

$$W(\omega) = \begin{cases} \dfrac{1}{\lambda_2 - \lambda_1} & \text{for } \lambda_1 < \omega < \lambda_2 \\ 0 & \text{otherwise} \end{cases} \quad ,$$

and is a consistent estimate of $f(\omega_o)$ when ω_o is in the interval (λ_1, λ_2). It is, however, biased, since it estimates an averaged or smoothed value of $f(\omega)$ over a neighborhood of ω_o on the ω axis. We can now ask: Is there a way in which we can tamper with $W(\cdot)$ so as to maintain the consistency of the estimate and at the same time render $W(\cdot)$ dependent on sample size T in such a way as to reduce the bias as $T \to \infty$?

Heuristically, we started with $W(\cdot)$ as shown in Fig. 1a,

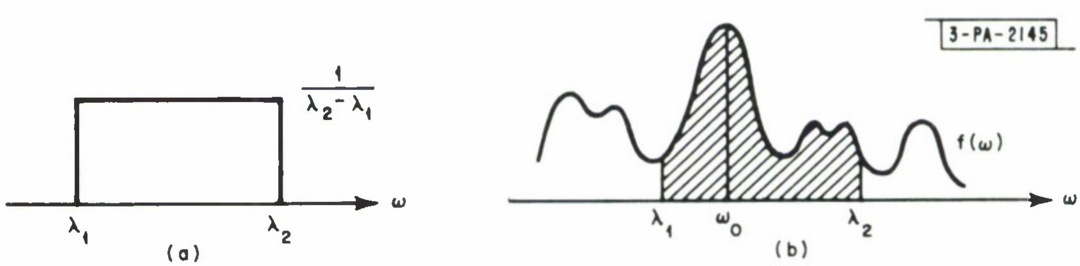

Fig. 1. Geometric interpretation of the periodogram as an estimate of $f(\omega_o)$.

*For definition, see footnote page 13.

and estimated the average of $f(\omega)$ over the shaded interval in Fig. 1(b). Can we suitably modify $W(\cdot)$ so that it depends on T in such a manner as to become increasingly peaked as T increases and assume the shape shown in Fig. 2, all the while satisfying the subsidiary condition $\int_{-\infty}^{\infty} W_T(\omega)\, d\omega = 1$? If this program can be carried out, the values of $f(\omega)$ at and near ω_o will be given more and more weight relative to values of $f(\omega)$ distant from

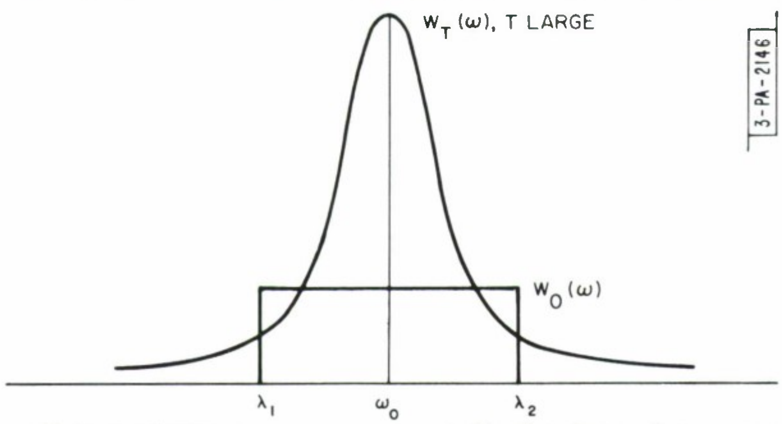

Fig. 2. Smoothing windows sequence and their dependence on sample size.

ω_o, and we may expect, as sample size increases, to reduce the bias introduced by averaging.

5. Consistent Estimates (3): Unified Treatment. Writing the periodogram $\Phi_T(\omega)$ as the Fourier transform of the sample autocovariances $\varphi_T(k)$ we have

$$\Phi_T(\omega) = \frac{1}{2\pi} \sum_{|k|<T} \varphi_T(k) e^{-ik\omega} \quad .$$

In Section B. 3, we had considered estimates $f_T(\omega)$ of the form

$$f_T(\omega) = \frac{1}{2\pi} \sum_{|k|<T} h_T(k)\, \varphi_T(k) e^{-ik\omega} \quad .$$

Let the Fourier transform of $h_T(k)$ be represented by

$$H_T(\omega) = \frac{1}{2\pi} \sum_{|k|<T} h_T(k) e^{-ik\omega} \quad .$$

Then, the convolution theorem applied to f_T gives

$$f_T(\omega) = \frac{1}{2\pi} \sum_{|k|<T} h_T(k)\, \varphi_T(k) e^{-ik\omega} = \int_{-\pi}^{\pi} H_T(\omega - \lambda)\, \Phi_T(\lambda)\, d\lambda \quad ,$$

which demonstrates the equivalence of the two approaches to consistent estimation of $f(\omega)$. $h_T(k)$ is called a covariance averaging kernel (lag window), $H_T(\omega)$ is called spectral window. Bartlett's scheme leads directly to a covariance averaging kernel. The use of smoothed spectra leads to spectral windows.

6. Statistical Properties of Consistent Estimates: The notion of bandwidth plays an important role in the formulation of the statistical properties of spectral estimates of the type we have just considered. The bandwidth B_T of a spectral window $W_T(\cdot)$ is defined as the base of a rectangle which has the same area and the same maximum height as the graph of $W_T(\cdot)$ in Fig. 2 (or $H_T(\cdot)$ in the previous section, where the notation H was adopted to highlight the fact that h and H are a Fourier pair). Formally,

$$B_T = \frac{\int_{-\infty}^{\infty} H_T(\omega)\, d\omega}{\operatorname{Max}_{\omega} |H_T(\omega)|} \quad ,$$

where the dependence on sample size T has been indicated. We would like $B_T \to 0$ as $T \to \infty$, to improve the focusing power of $H_T(\cdot)$ on the peak frequency ω_o. It can be shown that mean square error and bandwidth are related as

$$MSE = Bias^2 + Var = B_T^{2p} + \frac{1}{T B_T}$$

where p is some integer ≥ 1 (cf. [10]).*

The important point is to note the mutually antagonistic effects of bias and variance. Bandwidth should indeed go to zero as $T \to \infty$. But if $B_T \to 0$ faster than $\frac{1}{T}$, the variance term will become very large indeed.

---

* p is the order of the highest derivative of $f(\omega)$ at ω_o. Hence, the smoother $f(\omega)$ at ω_o, the larger is p.

The choice of a specific smoothing window or kernel averaging sequence in a practical situation will depend on how sharp an estimate of $f(\omega_o)$ one desires; with a broad bandwidth the variance of the estimate can be made small; on the other hand, if a sharp estimate is needed (narrow bandwidth, small bias) the variance of the estimate increases. We might further note that if the spectrum is relatively smooth (changes slowly) in the vicinity of ω_o, then its average value over an interval containing ω_o should differ but little from $f(\omega_o)$, thus insuring a small bias [20].

III. LIMITATIONS INTRODUCED BY SAMPLING

In the previous chapter we have concluded that a periodic regression model with known base frequency is of limited practical value, that classical periodogram analysis leads to spurious interpretations if a continuous component is present in the spectrum and that covariance averaging kernels (or smoothing windows) are of decided value in estimating spectra for processes which have a continuous spectrum. The presentation was made in terms of discrete processes because this is usually the case encountered in practice; but the statistical properties of estimates were stated in terms of their asymptotic behaviour. Discrete time-series often arise from sampling continuous processes and even in the case of discrete processes it may be that not every available piece of data is retained in the analysis. In the present chapter, we will briefly dwell on the effects of sampling, i.e., of retaining only a finite set of equi-distant sample values. Nor will the asymptotic character of the results of the previous chapter be strictly exact as we are always dealing with a finite sample; they serve principally as a guide in the choice of one estimation procedure over another.

We shall be dealing here with a specific finite sampled sequence of values rather than with an ensemble of such sequences. We shall see that the rejection of information inherent in sampling introduces ambiguities and limitations, but also, that in a profound sense practically all the information in the sample can be recovered.

A. What is Lost in Sampling (1), Uniqueness of Representation

The following illustration shows clearly the ambiguities which are introduced if a function on a domain is sampled at a subset of the domain of definition. In brief, we shall "reconstruct" a function from its sampled values in two different ways and see what happens to its expansion. Let x(t) be defined on [0,T] by

$$x(t) = \alpha_0 + \sum_{k=1}^{\infty} (\alpha_k \cos k\omega t + \beta_k \sin k\omega t), \quad \omega = \frac{2\pi}{T} \quad ,$$

and suppose we retain only values of x(t) at the discrete times $t_\nu = \nu, \; \nu = 1 \rightarrow T$.

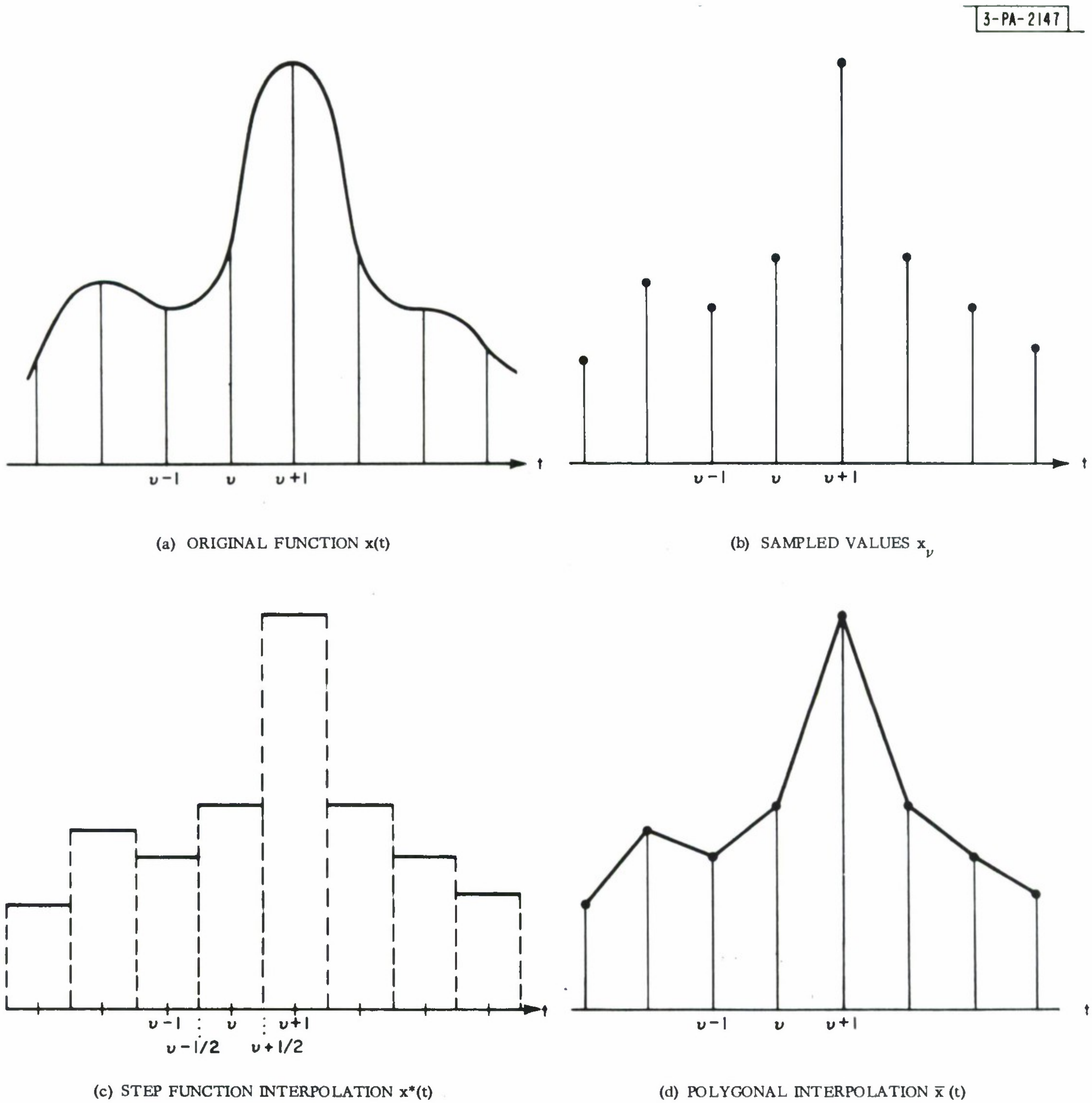

Fig. 3. Graphs of f(t) and its approximations.

Then

$$x_\nu = x(\nu) \; , \quad \nu = 1 \rightarrow T \; , \quad x_o = x_T \quad .$$

We shall reconstruct the Fourier coefficients of x(t) on the basis of the sample in two different manners:

1. By assuming that x is reconstructed as a step function (cf. Fig. 3c)

$$x^*(t) = x_\nu \quad \text{for} \quad \nu - \frac{1}{2} < t < \nu + \frac{1}{2} \quad .$$

2. By assuming that x is reconstructed as a polygonal line connecting successive ordinates

$$\overline{x}(t) = x_\nu + (x_{\nu+1} - x_\nu)\,(t - \nu) \text{ for } \nu < t < \nu + 1 \quad ,$$

as shown in Fig. 3d.

x(t)	x*(t)	$\overline{x}$ (t)
$\hat{\alpha}_o = A_o$	$\alpha^*_o = A_o$	$\overline{\alpha}_o = A_o$
$\hat{\alpha}_k = A_k$	$\alpha^*_k = A_k \cdot \left(\dfrac{\sin \frac{k\omega}{2}}{\frac{k\omega}{2}} \right)$	$\overline{\alpha}_k = A_k \cdot \left(\dfrac{\sin \frac{k\omega}{2}}{\frac{k\omega}{2}} \right)^2$
$\hat{\beta}_k = B_k$	$\beta^*_k = B_k \cdot \left(\dfrac{\sin \frac{k\omega}{2}}{\frac{k\omega}{2}} \right)$	$\overline{\beta}_k = B_k \cdot \left(\dfrac{\sin \frac{k\omega}{2}}{\frac{k\omega}{2}} \right)^2$

Table 2 Dependence of Fourier Coefficients on Interpolating Assumptions

The least squares estimates $(\hat{\alpha}_o, \hat{\alpha}_k, \hat{\beta}_k)$ of $(\alpha_o, \alpha_k, \beta_k)$ on the basis of the sample are given, as in (II, A, 1) by

$$A_o = \frac{1}{T} \sum_{\nu=1}^{T} x_\nu$$

$$A_k = \frac{2}{T} \sum_{\nu=1}^{T} x_\nu \cos k\omega\nu$$

$$B_k = \frac{2}{T} \sum_{\nu=1}^{T} x_\nu \sin k\omega\nu \quad , \quad k = 1 \rightarrow \frac{T-1}{2} \quad .$$

A straightforward computation yields the Fourier coefficients of $x^*(t)$ and $\bar{x}(t)$ in terms of the (A_o, A_k, B_k). These are given in Table 2.

Let us recapitulate. We have reconstructed a function from its sample in three different ways: (1) as a trigonometric series; (2) as a step function; (3) as a polygonal line, and we have calculated the Fourier coefficients of each reconstruction. It is now evident that these depend on the assumed behaviour of the function between the sampled values.

B. What is Lost in Sampling (2): Folding.

Suppose a set of discrete observations x_ν are obtained as a result of sampling x(t) at fixed intervals ΔT. The frequency components of x(t) with frequency $\omega > \frac{\pi}{\Delta T}$ rad/sec. cannot be distinguished from those with frequency ω in the range $(0, \frac{\pi}{\Delta T})$ on the sole basis of the sampled values.

To see this, we consider first the case of a single sinusoid and show that

$$\cos\left[\left(\frac{\pi}{\Delta T} + \epsilon\right) t + \varphi\right] = \cos\left[\left(\frac{\pi}{\Delta T} - \epsilon\right) t - \varphi\right]$$

for all $t = k(\Delta T)$, i.e., for all integral multiples of the sampling period ΔT. Indeed,

$$\left(\frac{\pi}{\Delta T} + \epsilon\right) k\Delta T + \varphi = k\pi + \left(\epsilon k\Delta T + \varphi\right)$$

$$\left(\frac{\pi}{\Delta T} - \epsilon\right) k\Delta T - \varphi = k\pi - \left(\epsilon k\Delta T + \varphi\right)$$

and the periodicity and symmetry of the cosine function with respect to $k\pi$ complete the proof.

Figure 4 illustrates the symmetry about multiples of $\frac{\pi}{\Delta T}$ and shows why this phenomenon is called folding.

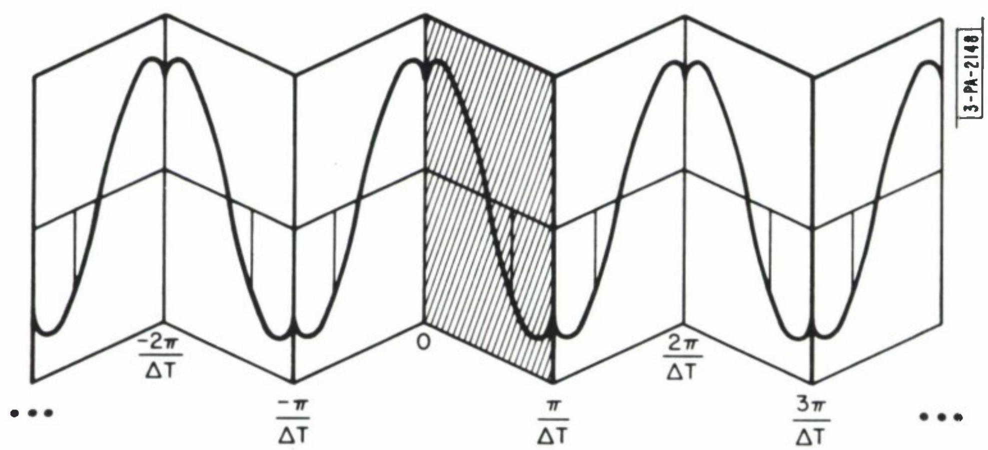

Fig. 4. Folding illustrated.

It is thus seen that two sinusoids having frequencies symmetrically placed with respect to $\frac{k\pi}{\Delta T}$ ($k = 0, \pm 1, \pm 2, \ldots$) cannot be distinguished on the basis of values sampled at $k\Delta T$. This has been called "aliasing" by J. W. Tukey in the sense that all frequencies differing from a given frequency ω by multiples of $\frac{2\pi}{\Delta T}$, $\{\omega \pm \frac{2\pi k}{\Delta T}\}$, are indistinguishable from ω_A the so called principal alias of ω, which is that frequency among all $\{\omega \pm \frac{2\pi k}{\Delta T}\}$ which falls within the interval $\left(0, \frac{\pi}{\Delta T}\right)$.

The argument just developed for single sinusoids applies equally well to any function x(t) which can be represented as a superposition of sinusoids.

In general, the spectrum $f^*(\omega)$ derived from a sample taken at rate $1/\Delta T$ is given in terms of the true spectrum $f(\omega)$ by

$$f^*(\omega) = \sum_{k=0}^{\infty} \left\{ f\left(\frac{2\pi k}{\Delta T} + \omega_A\right) + f\left(\frac{2\pi k}{\Delta T} - \omega_A\right) \right\} \quad .$$

The folding frequency $\frac{\pi}{\Delta T} = \omega_N$ is also known as Nyquist frequency. A graphical illustration of aliasing is shown in Fig. 5.

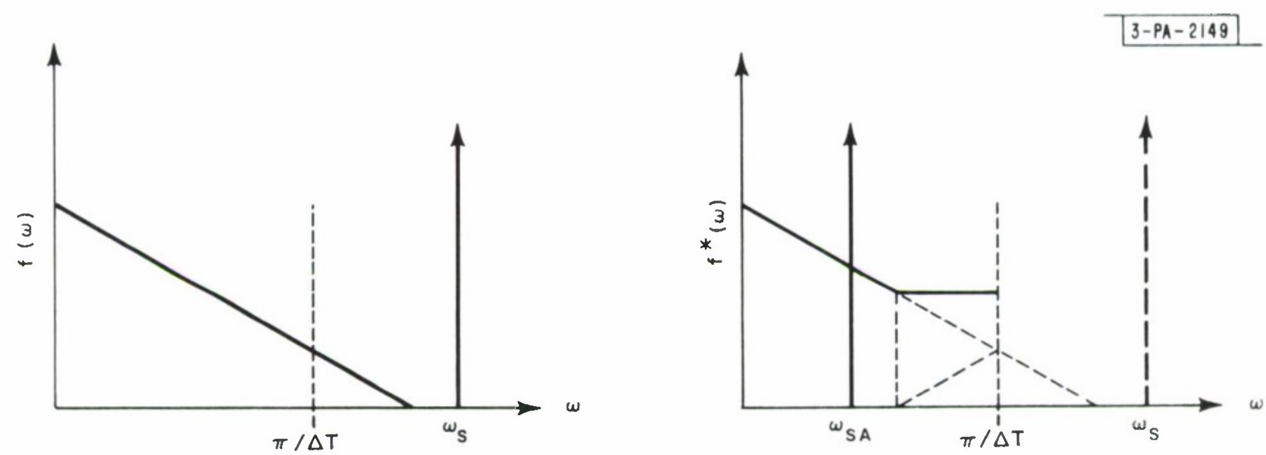

Fig. 5. Illustration of aliasing.

If $\frac{\pi}{\Delta T}$ is the folding frequency of the sampling process, the frequencies in $f(\omega)$ beyond $\frac{\pi}{\Delta T}$ are folded back and added to the original graph in Fig. 5b. A spike occurring at ω_s in $f(\omega)$ will appear in $f^*(\omega)$ at ω_{sA}, it's principal alias.

C. <u>What is Retained in Sampling: The Sampling Theorem</u>

Admittedly, sampling discards information. There is, however, one situation in which nothing is lost.

Suppose the function x(t) contains no frequency components above W cycles/second. Then x(t) is completely determined by its values at a sequence of points $\frac{1}{2W}$ apart.

In particular, if the sampled values are $\left(\frac{k}{2W}\right)$, $k = 0, \pm 1, \pm 2, \ldots$, then x(t) can be uniquely reconstructed as

$$x(t) = \sum_{k=-\infty}^{+\infty} x\left(\frac{k}{2W}\right) \frac{\sin \pi\,(2Wt - k)}{\pi\,(2Wt - k)} \quad , \quad [19, 33] \; .$$

Thus a band limited function can be restored completely if it is sampled at the proper frequency. Suppose now that x(t) is not band limited, i.e., there are frequency components in its spectrum beyond W. Then, what is lost by sampling at a rate of $\frac{1}{2W}$ is precisely what has been described as aliasing, and frequencies beyond W are folded back into the interval $(0, \omega_N)$. If, however, x(t) is sampled at only a finite number of points, say for $k=0, \pm 1, \pm 2, \ldots, \pm n$, then the reconstruction of x(t) as

$$x_1(t) = \sum_{k=-n}^{n} x\left(\frac{k}{2W}\right) \frac{\sin \pi\,(2Wt - k)}{\pi\,(2Wt - k)}$$

is but one of infinitely many possible reconstructions agreeing with the data in the sampled interval. Indeed, for any sequence y_k, $k = \pm (n+1); \pm (n+2), \ldots$ the function

$$x_2(t) = x_1(t) + \sum_{k=-\infty}^{-(n+1)} + \sum_{k=n+1}^{\infty} y_k \, \frac{\sin \pi\,(2Wt - k)}{\pi\,(2Wt - k)}$$

agrees with $x_1(t)$ for $k=0, \pm 1, \pm 2, \ldots, \pm n$, since $\frac{\sin \pi\,(2Wt - k)}{\pi\,(2Wt - k)}$ vanishes at all points of the form $t = \frac{k}{2W}$. Thus, values sampled at equidistant points within an interval do not determine values outside this interval. In practice, however, [5] empirical functions are always band limited, nor are measurements ever so precise that an accurate reconstruction could be obtained, even with an infinite sequence of values. In this sense, most of the information present in a continuous record can usually be extracted from a sample taken at sufficiently closely spaced ordinates.

D. Implications of Sampling

The practical implications of the ambiguities introduced by sampling are several. A good deal depends on how we conceive the underlying process to behave between the sampled points. If it is nonexistent there, we have an unlimited choice; but this is rarely the case. In general, then, aliasing has to be taken into account in the design of the data gathering procedure as well as in the evaluation of the results of the analysis.

Since the sampled spectrum $f^*(\omega)$ is obtained from the unsampled spectrum $f(\omega)$ by folding all frequencies about multiples of the folding frequency, and summing these contributions in the range $\left(0, \frac{\pi}{\Delta T}\right)$, we must insure, if we wish to estimate $f(\omega)$ in the range $\left(0, \frac{\pi}{\Delta T}\right)$, to have approximately $f(\omega) = 0$ for $\omega > \frac{\pi}{\Delta T}$. This can, in general, be done by filtering the data through a low pass filter before sampling. G. M. Jenkins [14] has suggested that in any practical problem there are 3 frequencies worth considering:

(1) the frequency ω_1 which is essentially the frequency beyond which the recording instruments' frequency response function is negligible (e.g., 1-2% of its maximum value, and goes to 0 asymptotically);

(2) the frequency ω_2 beyond which there is barely any power dissipated by the process $x(t)$ ($\sim$ 1 or 2% of the total);

(3) a good apriori opinion of the largest frequency ω_3 of interest in the study of the process at hand.

Frequency ω_1 has a transparent significance, as it limits the frequency content of the data. The interplay of ω_2 and ω_3 is somewhat more subtle. A suitable procedure might be to base the sampling rate on the larger of ω_2 and ω_3. If $\omega_3 > \omega_2$, well and good; $\omega_3 = \pi/\Delta T$, so that $\Delta T = \pi/\omega_3$ and no frequencies beyond those of interest are read; moreover, there is negligible aliasing since most of the power in the process has been dissipated before ω_2 ($< \omega_3$). If $\omega_2 > \omega_3$, components of f beyond ω_3 could be folded relative to ω_2 if ω_3 were determining the sampling rate. Thus we select ω_2 to determine $\Delta T = \pi/\omega_2$, providing we know enough about the process to choose ω_2 sensibly.

One of the costs of choosing the sampling interval on the basis of a frequency higher than one is really interested in, is that sampling then takes place at a much higher rate than really necessary, giving rise in turn to larger quantities of data to process.

Another cost factor is related to the nearness of the frequency of interest to the folding frequency; the closer to $\frac{\pi}{\Delta T}$, the harder it is to estimate, in the sense that a larger sample is required (for a given reliability of estimation) than for a frequency nearer the origin of the frequency axis. Intuitively, this can be seen in the following manner.

Suppose the folding frequency is ω_N (sampling interval $\Delta T = \frac{\pi}{\omega_N}$) and a component of frequency $\omega_o = \omega_N - \epsilon$ (ϵ small) is present in the signal. The sampling interval corresponding to ω_N is slightly smaller than the distance between successive zero crossings in $\sin \omega_o t$,

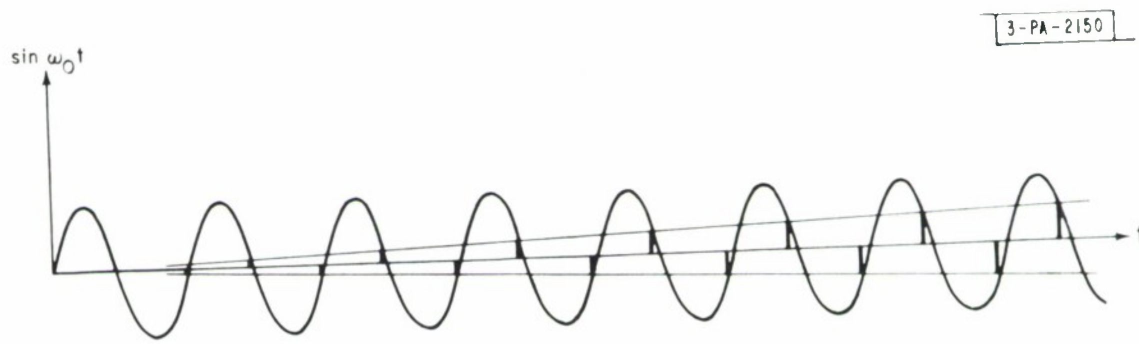

Fig. 6. Ill conditioning.

and ordinates at the sampled points creep up extremely slowly to the peak amplitude as illustrated in Fig. 6. Thus, a large sample may be needed to gain confidence in the estimate of amplitude*. As a rule of thumb, it is usually considered unwise to estimate spectra at frequencies exceeding 20 or 30% of the folding frequency, as the reliability of estimates beyond this point becomes questionable. There is a great deal of debate about the cutting off point, depending on specific situations [14, 28].

*The nature of the problem is one of ill conditioning. The smaller ϵ, the more difficult becomes the estimation. An analogy might clarify matters further: it is easy to locate the intercept of a straight line if its slope is large. It becomes much more difficult to do so if its slope is very small.

IV MISCELLANEOUS COMMENTS

This chapter deals with several points which could not properly be placed in the logical sequence of the previous development but which are nevertheless illuminating and of sufficient practical interest to be included. Some of these are direct consequences of the convolution theorem* as applied in (III, B, 5); these will be taken up first. The remainder of the chapter will present some useful procedures of interest in actual computations, and some examples illustrating the methods described earlier.

A. The Convolution Theorem and Some of its Consequences

Let us recall that we represented the periodogram

$$\Phi_T(\omega) = \frac{1}{2\pi} \sum_{|k|<T} \varphi_T(k) e^{-i\omega k} \tag{1}$$

as the Fourier transform of the sample covariances, and that, to improve its usefulness as an estimator, we introduced covariance averaging kernels (lag windows) $h_T(k)$ and obtained

$$\hat{f}_T(\omega) = \frac{1}{2\pi} \sum_{|k|<T} h_T(k)\, \varphi_T(k) e^{-i\omega k} \tag{2}$$

as an estimate of $f(\omega)$. Writing the Fourier transform of $h_T(k)$ as

$$H_T(\omega) = \frac{1}{2\pi} \sum_{|k|<T} h_T(k) e^{-i\omega k} \tag{3}$$

we noted that, by the convolution theorem applied to $h_T(k)$ and $\varphi_T(k)$

$$\hat{f}_T = \int_{-\pi}^{\pi} H_T(\omega-\lambda)\, \Phi_T(\lambda)\, d\lambda \quad . \tag{4}$$

*The convolution theorem states that if the following are Fourier pairs $u(t) \longleftrightarrow U(\omega)$
$v(t) \longleftrightarrow V(\omega)$

then the following is also a Fourier pair:

$$u(t) \cdot v(t) \longleftrightarrow \int_{-\infty}^{\infty} U(\omega-\lambda)\, V(\lambda) d\lambda.$$

We are now in a position to clarify:

-the occurrence of negative values in periodogram estimates of the spectrum although the latter is strictly non-negative;

-the necessity of removing D.C. components, deterministic components, or time dependent non-zero means from the data before estimating the spectrum;

-the difference in appearance between plots of periodogram and smoothed periodogram respectively.

1. Negative Values of the Periodogram

A comparison of equations (1) and (2) shows that the periodogram (1) can be expressed as (2) for a particular choice of covariance averaging kernel, namely

$$h_T(k) = \begin{cases} 1 & \text{for } |k| < T \\ 0 & \text{for } |k| > T \end{cases} \quad . \tag{5}$$

Figure 7 illustrates the continuous analog of (5) and its Fourier transform*, $\frac{T}{\pi} \frac{\sin \omega T}{\omega T}$.

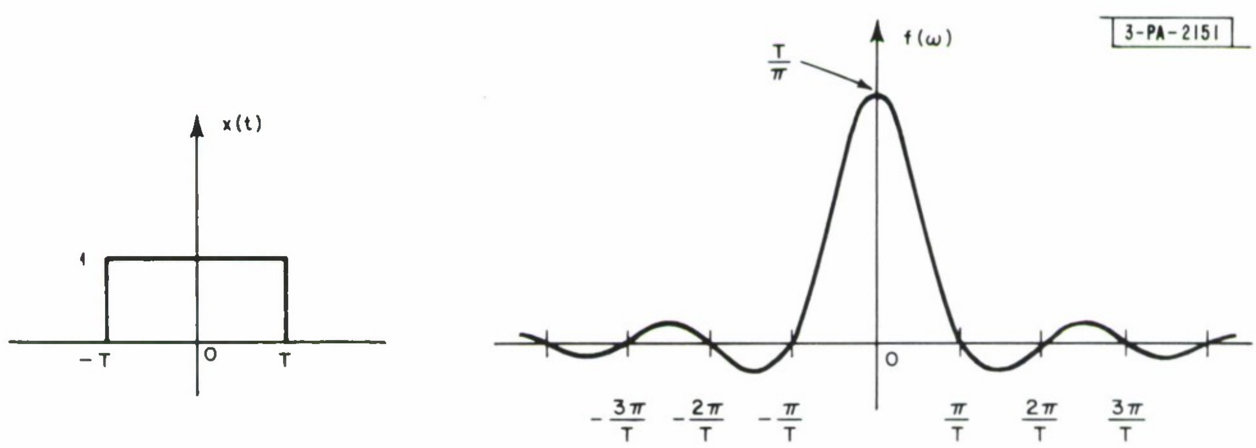

Fig. 7. A ubiquitous transform pair.

*There are some analytical complexities introduced by the "sampled nature" of $h_T(k)$. We do not go into full detail, but the argument is not invalidated. For complete details cf. Blackmann and Tukey [5].

The main point to notice is that sampling per se is equivalent to the introduction of a smoothing window $H_T(\omega)$ which assumes negative values, so that the spectrum estimate (4) allows negative weights for the periodogram. The conclusion is that $\Phi_T(\omega)$ assumes negative values with non-zero probability. Large negative values of the periodogram may either be ignored (e. g., if it is felt that the periodogram is well suited to the modelling assumptions) or they may be taken as an indication that other smoothing and estimation schemes should be attempted in order to ascertain the stability of the results.

2. The Removal of D.C. Components, Time Dependent Non-Zero Means, or Deterministic Components

By considering the effect of a smoothing window, it becomes understandable that D.C. components should be removed before processing. Indeed, every covariance averaging kernel corresponds to a window with some equivalent bandwidth over which the said D.C. component is smeared and thus made to contribute to the contamination of low frequency components. A similar argument may be made for deterministic components, periodic or otherwise, and non-zero means.

Admittedly, in the case of periodogram calculations the bandwidth of the smoothing window can be made arbitrarily small by taking a large enough observation interval (sample); but in general, since it is desirable to use covariance averaging kernels to obtain consistent estimates, it is advisable to remove all but the purely oscillatory components of the data.

3. Appearance of Periodogram and Smoothed Periodogram Plots

The shape of the smoothing window corresponding to the covariance averaging kernel used, is indicative of the effect of smoothing as compared to the "unsmoothed" estimate of the spectrum. If a spike appears in the unsmoothed spectrum estimate, the smoothed estimation procedure merely redistributes this spike over the entire axis with the smoothing window as envelope for the spectral components at discrete points. Since the smoothing window is generally sharp in practice, this redistribution takes place mainly over a small neighborhood of the frequency at which the spectrum is estimated.

4. The Effect of Non-Rational Frequencies

In general, when the sample size T is given, spectral analysis consists in calculating the periodogram (smoothed or unsmoothed) at a discrete set of trial frequencies. $\omega_j = \frac{2\pi j}{T}$, $j = 1 \rightarrow T/2$, under the assumption that x(t) repeats itself exactly after T observations, apart from an additive error term (cf. model in II, A, 2). We will now describe the effect of the presence of a frequency component whose frequency does not coincide with any of the trial frequencies.

Let a process be represented by a cyclical deterministic component, apart from additive noise ξ_t

$$x_t = \rho \cos\left(\frac{2\pi\lambda}{T} + \psi\right) + \xi_t$$

where k is the multiple of $\frac{2\pi}{T}$ nearest to $\frac{2\pi}{T}\lambda$ (cf. Fig. 8).

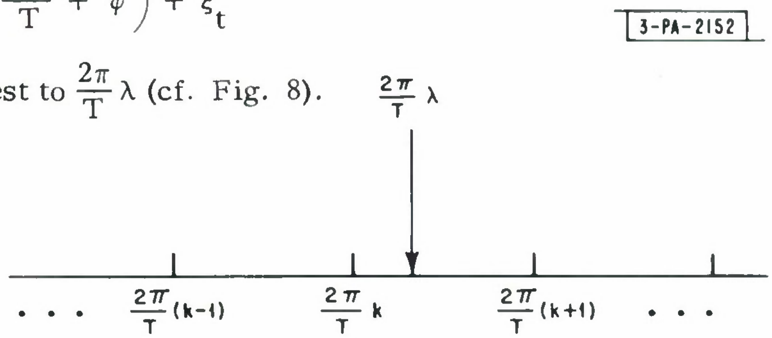

Fig. 8. A non-rational frequency.

It can be shown that R_{k+j}, the estimate of the process amplitude ρ_{k+j} (nonexistent!) depends asymptotically (as $T \rightarrow \infty$) both on j and on $\epsilon = \lambda - k$ in the following manner*

$$\mathcal{E}\left(R_{k+j}^2\right) \sim \left[\rho \frac{\sin \pi (j - \epsilon)}{\pi (j - \epsilon)}\right]^2 ,$$

so that the irrationality of λ has the effect of smearing the corresponding frequency component over the entire spectrum. The presence of "irrational" components helps further to explain why periodogram calculations should not be expected to yield zero

*This result is obtained by considering the Fourier coefficients of $\cos(\alpha t + \psi)$ on a base interval which is not an integral multiple of the period. In the worst case, when $\epsilon = \frac{1}{2}$

$$\mathcal{E}\left(R_{j+k}^2\right) \simeq \frac{4\rho^2}{\pi^2} \frac{1}{(2j+1)^2}$$

estimates for amplitudes, even in the absence of periodic components at the trial frequency (we recall that another source of non-zero values was the effect of the additive random noise treated in the model of Section II, A, 2).

B. Some Useful Procedures and Examples

Several procedural details are offered here as examples of the kind of improvements which can be brought to a general program of power spectrum estimation conceived within the framework of the theory described in the earlier parts of this report. Some concrete examples are provided to illustrate some of the developed concepts.

1. Direct Removal of Time Dependent Means

In the processing of time series which are suspected of having a non-zero time dependent mean, a very simple scheme exists to remove this component from the sample covariances. Instead of using the mean lagged product as an estimate of covariance, calculate instead

$$\varphi(k) = \mathcal{E}\, x_t x_{t+k} - \mathcal{E}\, x_t \mathcal{E}\, x_{t+k} \tag{1}$$

where the averages are taken over a moving strip of data.

Actually, this procedure assumes that the process x_t is the sum of a deterministic component m_t and a zero mean process y_t having a spectrum:

$$x_t = m_t + y_t \quad .$$

Taking expected values of lagged products, we obtain

$$\begin{aligned} \mathcal{E}\left[x_t x_{t+k}\right] &= \mathcal{E}\left[\left(m_t + y_t\right)\left(m_{t+k} + y_{t+k}\right)\right] \\ &= \mathcal{E}\left(m_t m_{t+k} + m_t y_{t+k} + y_t m_{t+k} + y_t y_{t+k}\right) \\ &= m_t m_{t+k} + \mathcal{E}\left[y_t y_{t+k}\right] \end{aligned}$$

Now formula (1) states that the sample autocovariance $\varphi(k)$ of the zero mean process y_t can be taken as the average lagged product of the sample values reduced by the product of the sample means. Thus the spectral analysis proceeds to be made on the portion of the signal for which the theory is applicable.

2. Calculation of Sample Covariances

There has been some argument that the inconsistency of the periodogram as an estimate of the power spectrum is due to the instability of the sample covariances as estimates of the process covariances as the lag becomes large. In practice, there are three candidates which have been used as sample covariances:

$$\text{(c1)} \qquad \varphi_T^{(1)}(k) = \frac{1}{T} \sum_{t=1}^{T-|k|} x_t x_{t+|k|}, \quad 0 \leq k \leq T-1$$

$$\text{(c2)} \qquad \varphi_T^{(2)}(k) = \frac{1}{T-|k|} \sum_{t=1}^{T-|k|} x_t x_{t+|k|}, \quad 0 \leq k \leq T-1$$

$$\text{(c3)} \qquad \varphi_T^{(3)}(k) = \frac{2}{T} \sum_{t=1}^{T/2} x_t x_{t+|k|}, \quad 0 \leq k \leq \frac{T}{2}$$

as estimates of the process covariance $\gamma(k)$. We can compare them briefly on the basis of their statistical properties:

Bias: c1 has a bias equal to $\frac{k}{T}$, which goes to zero, as $T \rightarrow \infty$, for every fixed k; for finite samples however, the bias persists. c2 is unbiased for any lag k and all sample sizes T.

Positive Definiteness: c1 is positive definite, which is a desirable property for an estimator of a positive quantity. c2 is not positive definite.

Mean Square Error: As an estimate of $\gamma(k)$, c1 has a smaller mean square error than c2 (cf. [20]).

As far as c3 is concerned, its statistical properties have not been worked out. However, as an estimate of $\gamma(k)$ it does not appear quite as good as c1 and c2 since it uses only half of the data available. But this in itself does not prove that it leads to poorer estimates of the power spectral density.

3. Examples

The importance of the specific choice of technique is illustrated in Figs. 9, 10, and 11. Figure 9 is a plot of the sample autocovariance of a set of data as calculated on the basis of formula c3 of Section IV, B, 1. The important point to note is that two periodic components are clearly present in the autocovariance (ACV) function, one of relatively high frequency, and one of low frequency.

Figures 10 and 11 are the results of estimating the power spectral density in two different ways. The method used to obtain Fig. 10 depends on formula c2 for the calculation of sample autocovariances; furthermore, the Fourier transform of the ACV function is smoothed with Hanning weights $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

The method used to obtain Fig. 11 proceeds to calculate and plot the Fourier transform of the ACV function shown in Fig. 9.

Note that to a hurried observer, the crispness of Fig. 11 would recommend itself in its clear identification of the frequency components one is led to expect on the basis of a glance at the ACV function shown in Fig. 9. A more seasoned analyst, however, will recall that the unsmoothed periodogram has a larger mean square error than the smoothed periodogram, as an estimate of the spectral density.

Admittedly, some crispness could be lent to the estimate of PSD shown in Fig. 10 by omitting the final smoothing with Hanning weights. Similarly, smoothing with Hanning weights would reduce the sharpness of the peaks in Fig. 11.

In conclusion, it must be said that there is at present no single optimal method of processing data when little is known apriori about either the process generating the data or the characteristics of the recording device. In any case, it is desirable to have several spectral estimation procedures on hand to gain familiarity with new data.

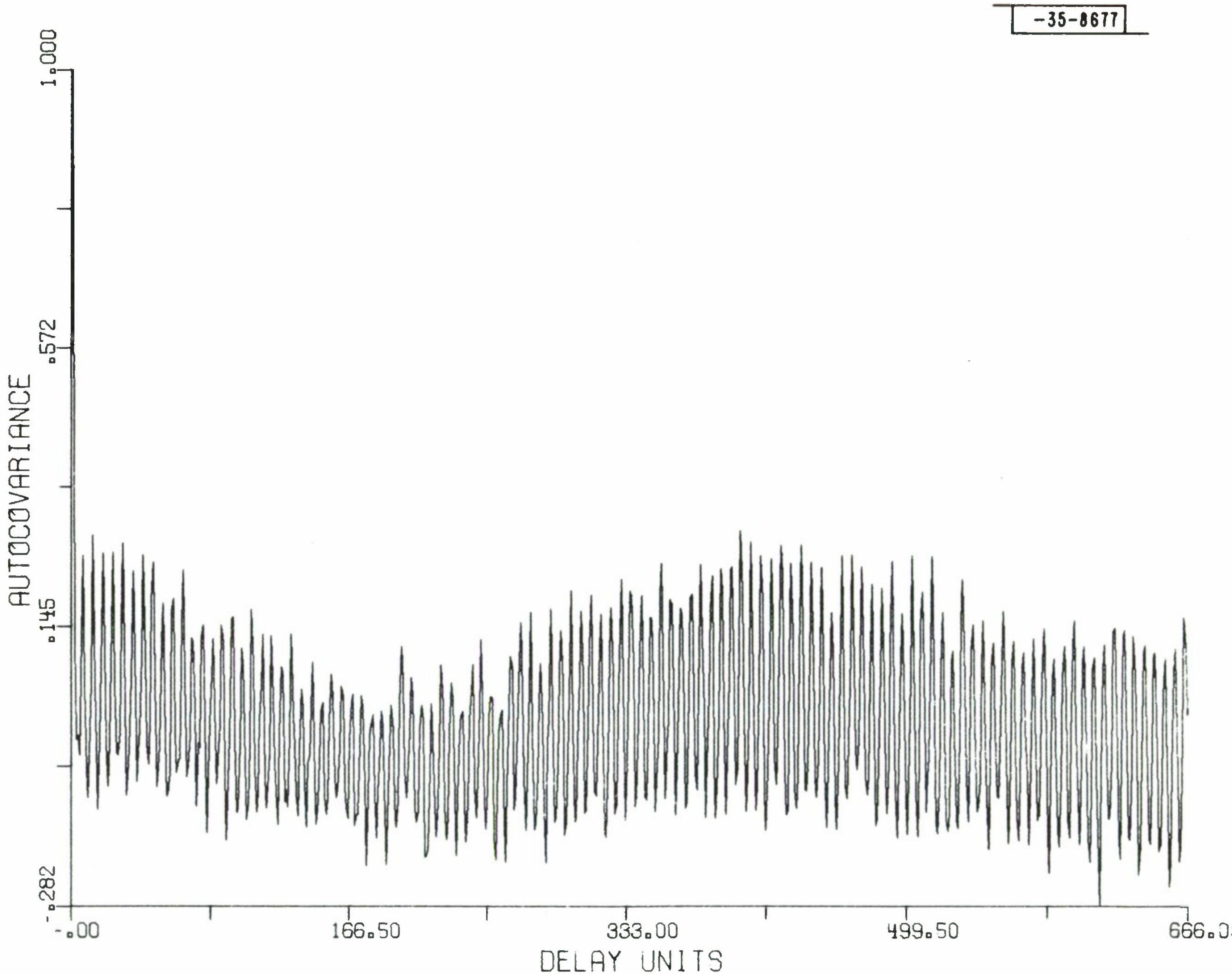

Fig. 9. Autocovariance function of a process.

-35-8678

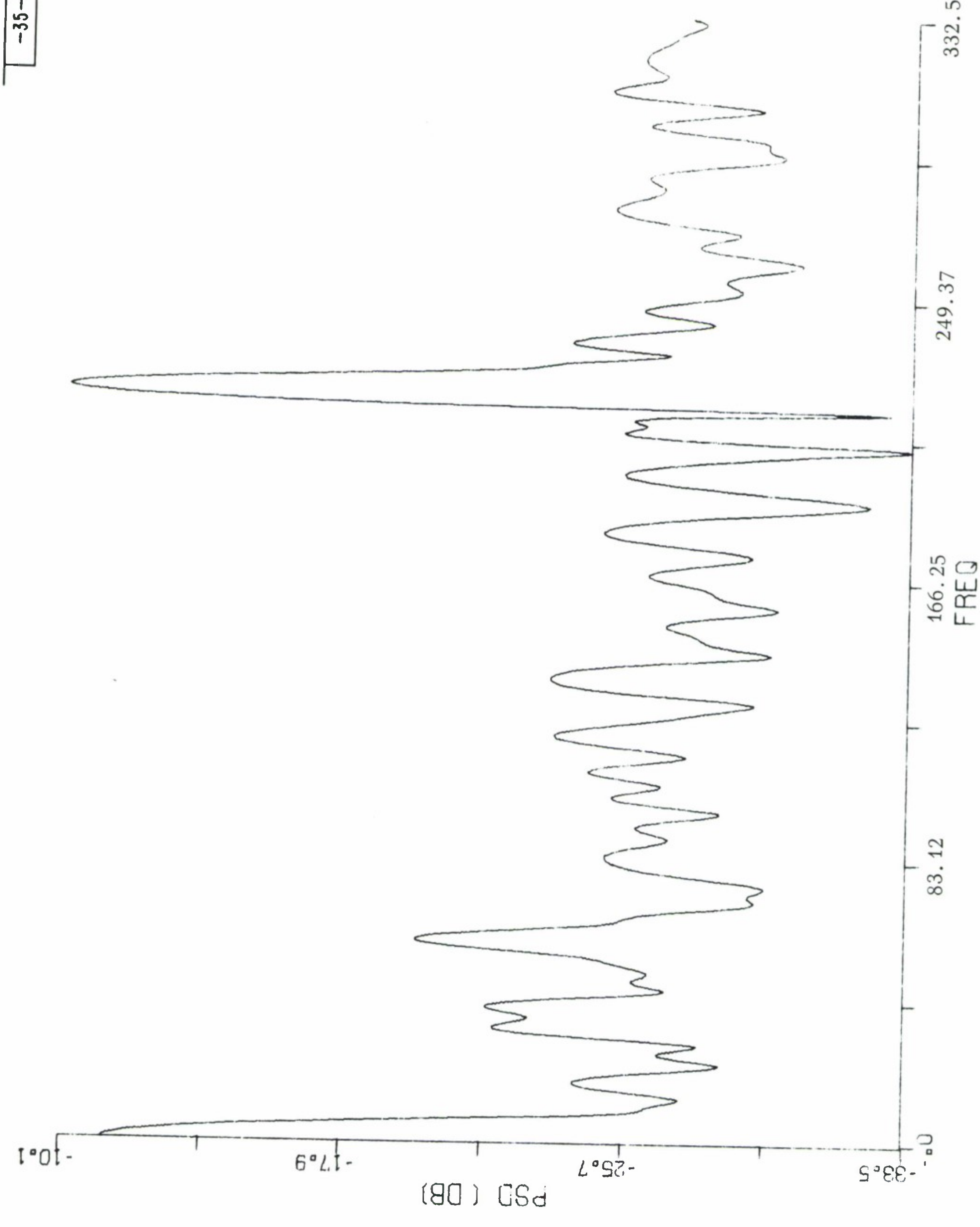

Fig. 10. Estimate of power spectrum by smoothed (Hanning weights) periodogram.

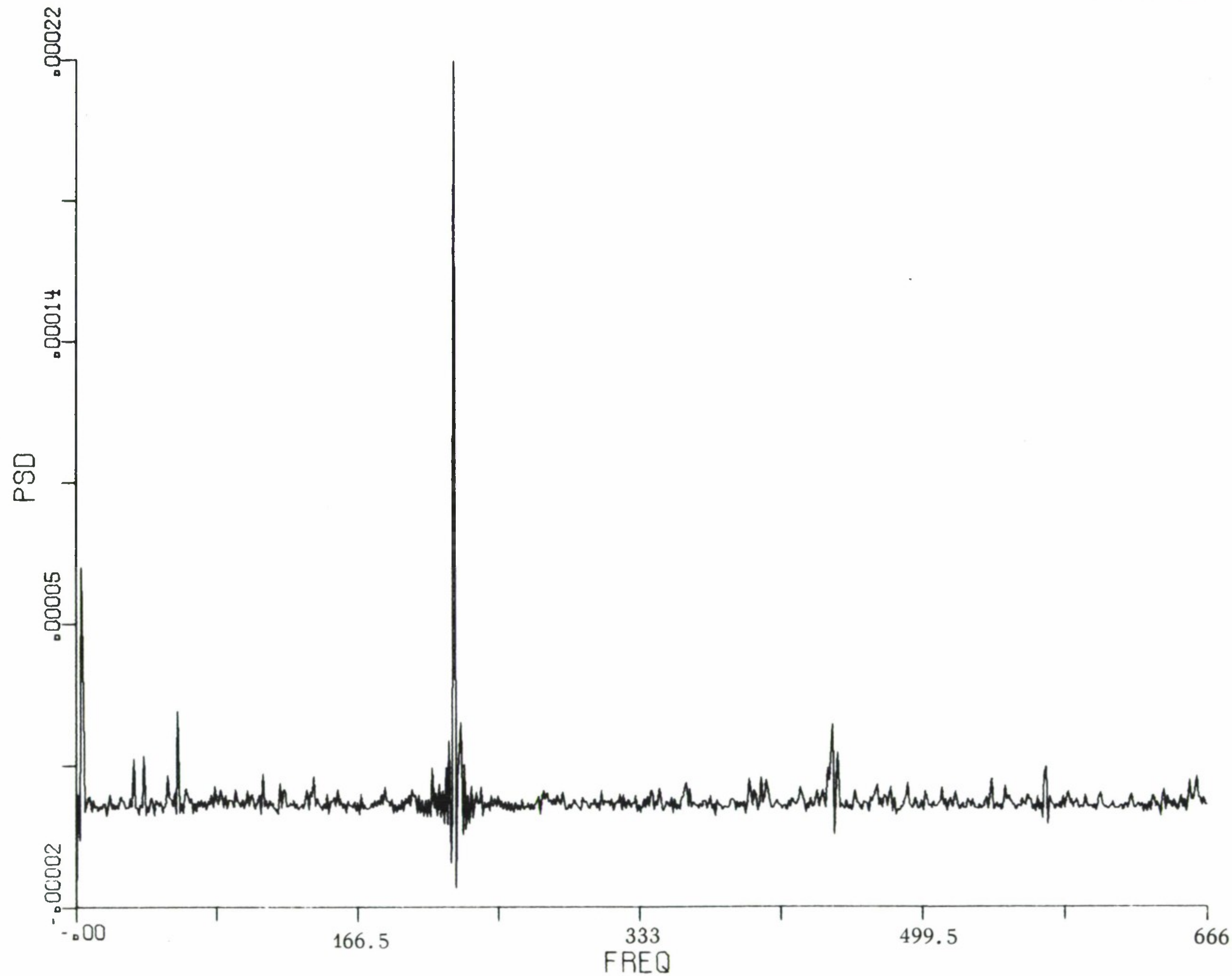

Fig. 11. Estimate of power spectrum after removal of time dependent mean (unsmoothed).

4. <u>Illustration of Periodogram Inconsistency for White Noise</u>

The probability distributions of spectral peaks for white noise are briefly described in (II, A, 2). In Section (II, B, 2) it was further mentioned that the periodogram is not a consistent estimate of the spectrum of a zero mean second order stationary process, and that as a consequence, the use of smoothed spectra came to be developed. The following figures (Figs. 12, 13, and 14) illustrate the situation rather clearly.

Figure 12 represents a sample of white noise; Fig. 13 is a plot of its autocorrelation; and, Fig. 14 is a plot of the estimated spectrum*. The main effect is to observe that the spectrum is highly peaked – for white noise – even in the case of a smoothed periodogram, as shown here. For a classical periodogram of white noise, the aspect of Fig. 14 would be more jagged still, whereas one should have expected a flat spectrum.

*The spectrum shown here is a smoothed periodogram with a Hanning window. A classical periodogram plot was not available at the time of writing.

-35-8680

Fig. 12. A sample of uncorrelated white noise.

-35-8681

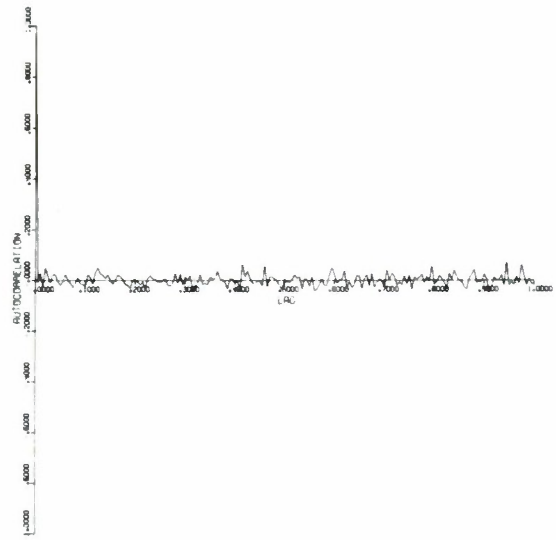

Fig. 13. Autocorrelation function of white noise.

-35-8682

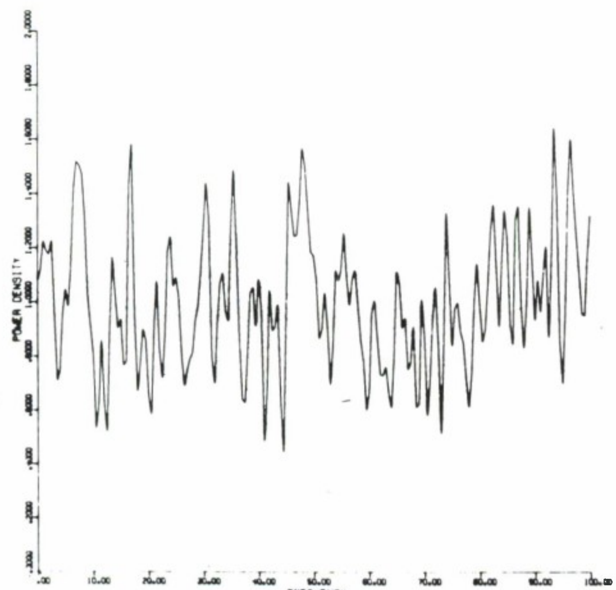

Fig. 14. Estimated power spectral density of white noise (smoothed periodogram, Hanning weights).

V. BRIEF OUTLINE OF A PROGRAM OF TIME SERIES ANALYSIS

The basic problem of time series analysis may be formulated as a question:

> "Given an empirical time series, by what process was it generated?" (P. Whittle).

This is a familiar question, commonly encountered in classical statistical estimation, and usually it is interpreted as

- selecting a hypothetical model for the process
- setting up criteria to measure the degree of agreement between the postulated model and the observed sequence.

Given two models to choose from, we can always decide, according to our established criteria, which is the more adequate. But to choose a reasonable model a priori requires some wisdom. In the present context (the analysis of radar data from re-entering vehicles), it would appear that a program of investigation might proceed as described below.

First, let us note that for processes with a discrete spectrum (e. g. , for which the cyclic models of Section (II, A, 1 and II, A, 2) are applicable) the calculation of smoothed spectra mars only slightly the occurrence of a peak by redistributing a portion of the peak amplitude over the width of the smoothing window.* Hence, in the case of cyclical processes, little is lost by using a smoothed periodogram instead of the periodogram, when the signal to noise ratio is not too small. On the other hand, we are hardly ever faced with purely cyclic data and since the statistical properties of the periodogram are so poor when the process to be analyzed is oscillatory, we feel that the basic spectral estimation tool should be the smoothed periodogram.**

---

*In general, only equidistant ordinates are considered in the estimation of spectra. If a harmonic term falls midway between two ordinates, its amplitude will be reduced by a factor of roughly $\frac{4}{\pi^2} \sim .41$ at these ordinates, in the case of classical periodogram calculations. This may or may not be enough to render a peak non-significant looking depending on the signal to noise ratio.

**Since the estimation will take place via covariance averaging kernels, several such kernels corresponding to several smoothing windows should be available as programmed options in any computer program for estimation of spectra.

The choice of a smoothing window and/or bandwidth remains, and should be made to depend on the data collection scheme and the resolution and precision requirements pertinent to the situation under investigation.

Second, since the theory is applicable only to zero mean second order stationary processes, we propose to use the techniques of spectral estimation only on the residuals of the data, after removal of deterministic components and non-zero means.

The following sequence of steps is envisaged.

A. The Search for Regression

As a first step, it is necessary to remove deterministic trends from the data. Some elegant methods exist to test for the presence of polynomial trends adaptively on the degree of the polynomial (up to fixed preassigned maximum degree) against the hypothesis that the data represents pure white noise.* A similar adaptive technique can be developed for other families of regression functions if it should appear that polynomials are not adequate for some reason or other. At any rate, the deterministic component is to be removed from the data, including a non-zero mean if it arises.

B. The Search for Periodic Components

It is clear that a periodogram analysis (smoothed or unsmoothed) will reveal the presence of deterministic cyclical components in the data. Once these are identified, they can be removed from the data. For identification of spurious components, see Section C.

C. Spectral Analysis

At this point, a spectral analysis can be made, using a variety of covariance averaging kernels, comparing the results critically among themselves and particularly against the results of step B to see whether the removal of a spurious cyclic component has not introduced an artificial periodicity in the data. The introduction of artificial periodicities in the data through smoothing is far from being an academic possibility, as evidenced by a theorem due to Slutsky (1927) [11, p. 95] which effectively demonstrates how to construct a linear smoothing scheme which introduces an arbitrarily selected spurious frequency.

D. Parameter Estimation

If it is known a priori, say on physical grounds, that a phenomenon exhibits a spectrum of a given functional form $f(\omega, \theta_1, \theta_2, \ldots, \theta_r)$ depending on several unknown parameters $\theta_1, \theta_2, \ldots, \theta_r$ these parameters may be estimated on the basis of the empirical time series.

---
*cf. Reference 35.

It should be possible to develop a testing scheme whereby one can compare an empirical spectrum to an a priori power spectrum. Such a technique has not been found in the literature to my knowledge.

E. Data Editing

There is a preliminary step which must be taken before any estimation and analysis can be carried out, and that is the removal of spurious data and of debris which usually finds its way past any recording device. Here, however, there are no general prescriptions as every case must be treated on its own demerits. For an example of semi-automatic data editing for trajectory data the reader is referred to Lincoln Laboratory Report PA-57 "A Trajectory Editing and Smoothing Program for FPS-16 Radars" (Unclassified) where some of the many possible vicissitudes of data editing are treated in detail.

ACKNOWLEDGEMENTS

I am grateful to many colleagues for patiently listening to my early gropings with an exposition of various portions of this report. I am indebted to Lee Gardner for the theoretical aspects, to Charles Arnold for the practical implications and to Fred Schweppe for many fruitful conversations.

Bino Nanni has supplied figures for Section IV, B, 4. Figures for Section IV, B, 3 are due to Charles Arnold. To all these my thanks, while I wish to retain the responsibility for whatever errors of omission or commission have crept past my hurried pen.

BIBLIOGRAPHY

1. N. I. Achieser, Vorlesungen uber Approximations Theorie, Akademie Verlag, Berlin 1953.

2. M. S. Bartlett, Periodogram Analysis of Continuous Spectra, Biometrika 37, pp. 1-16, 1950.

3. M. S. Bartlett, An Introduction to Stochastic Processes, Cambridge Univ. Press, Cambridge 1955.

4. M. S. Bartlett and J. Mehdi, On the Efficiency of Procedures for Smoothing Periodogram for Time Series with Continuous Spectra, Biometrika 42, pp. 143-150, 1955.

5. R. B. Blackmann and J. W. Tukey, The Measurement of Power Spectra, Dover Publications, New York, 1958.

6. S. Bochner, Vorlesgunen uber Fouriersche Integrale, Chelsea Publishing Co., New York 1948 (reprint of 1932 Edition).

7. Chapman and Bartels, Periodicities and Harmonic Analysis in Geophysics, Geomagnetism, Oxford University Press, Ch 16, Vol 2, pp. 545-605, 1951.

8. N. R. Goodman, Some Comments on Spectral Analysis of Time Series, Technometrics, Vol 3, No. 2, pp. 221-228, 1961.

9. U. Grenander and M. Rosenblatt, Spectral Analysis of Time Series Arising from Stationary Stochastic Processes, Ann. Math. Stat, 24, pp. 537-558, 1953.

10. U. Grenander and M. Rosenblatt, Some Problems in Estimating the Spectrum of a Time Series, Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, pp. 77-93, University of California Press, Berkeley, 1956.

11. U. Grenander and M. Rosenblatt, Statistical Analysis of Stationary Time Series, Wiley, New York, 1957.

12. E. J. Hannan, Time Series Analysis, John Wiley, New York, 1960.

13. H. O. Hartley, Tests of Significance in Harmonic Analysis, Biometrika 36, pp. 194-201, 1949.

14. G. M. Jenkins, General Considerations in the Analysis of Spectra, Technometrics, Vol 3, No. 2, pp. 133-166, 1961.

15. G. M. Jenkins and M. B. Priestley, The Spectral Analysis of Time Series, J. Roy. Stat. Soc. B, pp. 13-37, 1957.

16. M. G. Kendall and A. Stuart, The Advanced Theory of Statistics, Griffin-London, Vol I (1958), Vol II (1951).

17. C. Lanczos, Trigonometric Interpolation of Empirical and Analytical Functions, J. Math, Phys. 17, pp. 123-199, 1938.

18. Z. A. Lomnicki and S. K. Zaremba, On Estimating the Spectral Density Function of a Stochastic Process, J. Roy. Stat. Soc. B, 19, pp. 13-37, 1957.

19. K. S. Miller, Engineering Mathematics, Rhinehart and Co., New York, 1956, Section 7, p. 159.

20. E. Parzen, Mathematical Considerations in the Estimation of Spectra, Technometrics, Vol 3 No. 2, pp. 167-190, 1961.

21. E. Parzen, On Consistent Estimates of the Spectrum of a Stationary Time Series, Ann. Math. Stat., 28, pp. 329-348. 1957.

22. E. Parzen, On Choosing an Estimate of the Spectral Density Function of a Stationary Time Series, Ann. Math. Stat., 28, pp. 921-932, 1957.

23. H. Press and J. W. Tukey, Power Spectral Methods of Analysis and Application in Airplane Dynamics, AGARD Flight Test Manual, Vol. 4, Part C, NATO June 1956.

24. D. Ross and A. Suskind, Sampling and Quantizing, Ch. II Notes on Analog-Digital Conversion Techniques, A. Suskind, Ed., John Wiley, New York, 1957.

25. Karl Stümpff, Grundlagen and Methoden der Periodenforschung, J. Springer, Berlin, 1937.

26. J. W. Tukey, The Sampling Theory of Power Spectral Estimates, Symposium on Applications of Autocorrelation Analysis to Physical Problems, Woods Hole, 1949.

27. U. Grenander, J. W. Tukey, An Introduction to the Measurement of Power Spectra Probability and Statistics, Edited by U. Grenander, McGraw-Hill, New York 1959, pp. 300-330.

28. J. W. Tukey, Discussion, Emphasizing the Connection Between Analysis of Variance and Spectrum Analysis, Technometrics, Vol. 3, No. 2, pp. 191-219, 1961.

29. J. W. Tukey, The Estimation of (Power) Spectra and Related Quantities, On Numerical Approximation, R. E. Langer, Ed., 389-411, University of Wisc. Press, Madison, 1959.

30. P. Whittle, The Simultaneous Estimation of a Time Series Harmonic Component and Covariance Structure, Trab. Estad., 3, pp. 43-57, 1952.

31. P. Whittle, Curve and Periodogram Smoothing, J. Roy. Stat. Soc. B, 19, pp. 38-47, 1957.

32. N. Wiener, Generalized Harmonic Analysis, Acta Mathematics 55, pp. 117-258, 1930.

33. J. M. Whittaker, Interpolatory Function Theory, Cambridge Univ. Press, Cambridge, 1935.

34. H. Wold, A Study in the Analysis of Stationary Time Series, Almquist and Wiksell, Stockholm, 1953.

35. T. W. Anderson, The Choice of the Degree of a Polynomial Regression as a Multiple Decision Problem, Columbia University, Math. Stat, March 1961 ASTIA # 259759.