January 1970

# Topics in Time Series Analysis of Water Quality Data

Manhattan College

Environmental Engineering and Science Program

TABLE OF CONTENTS

# VARIANCE SPECTRA ANALYSIS

# AND APPLICATIONS TO

# NATURAL WATER SYSTEMS

D. M. Di Toro and D. J. O'Connor

Environmental Engineering and Science Program

Manhattan College, Bronx, New York

*Presented at American Association of Professors in Sanitary Engineering Fourth Annual Workshop "Applications of Systems Analysis in Sanitary Engineering" Myrtle Beach, S.C., June 25-27, 1969.

D. M. DiToro and D. J. O'Conner
*Civil Engineering Dept.*
*Manhattan College, Bronx, New York*

VARIANCE SPECTRA ANALYSIS AND APPLICATIONS TO NATURAL WATER SYSTEMS

## INTRODUCTION

As an introduction to Fourier analysis, i.e., the analysis of data in terms of sinusoidal functions, and to variance spectra analysis, i.e., the frequency analysis of variance, consider the following least squares fitting problem: A set of data points $g_j$ is obtained which are samples of a function g (t) at g(0), g($\Delta$t), g( 2$\Delta$ t)..... g((N-1) $\Delta$ t), and for convenience assume N is an even integer. As an example, consider daily average water temperature over a year. Since there are physical reasons for expecting a yearly periodicity in the data, it is required to fit the function:

$$\hat{g}_j = \frac{A_0}{2} + A_1 \cos(2\pi j/N) + B_1 \sin(2\pi j/N) \qquad j = 0, 1, \ldots, N-1$$

to the data $g_j$ in such a way that the least mean square error is obtained. That is the criteria to be satisfied is to minimize

$$\sum_{j=0}^{N-1} \left[ g_j - \hat{g}_j \right]^2 \tag{1}$$

with respect to $A_0$, $A_1$, $B_1$. Taking partial derivatives of eq. 1 with respect to $A_0$, $A_1$, $B_1$, and setting the result equal to zero yields the following simultaneous linear equations:

$$\frac{A_0}{2} N + A_1 \sum_{j=0}^{N-1} \cos(j\alpha) + B_1 \sum_{j=0}^{N-1} \sin(j\alpha) = \sum_{j=0}^{N-1} g_j$$

*Manhattan College, Bronx, N.Y.

$$\frac{A_0}{2}\sum_{j=0}^{N-1}\cos(j\alpha) + A_1\sum_{j=0}^{N-1}\cos^2(j\alpha) + B_1\sum_{j=0}^{N-1}\cos(j\alpha)\sin(j\alpha) = \sum_{j=0}^{N-1} g_j\cos(j\alpha)$$

$$\frac{A_0}{2}\sum_{j=0}^{N-1}\sin(j\alpha) + A_1\sum_{j=0}^{N-1}\sin(j\alpha)\cos(j\alpha) + B_1\sum_{j=0}^{N-1}\sin^2(j\alpha) = \sum_{j=0}^{N-1} g_j\sin(j\alpha)$$

where $\alpha = 2\pi/N$.

The following general relations; called orthogonality relations, are true: (1)

$$\sum_{j=0}^{N-1}\sin(jk\alpha)\sin(j\ell\alpha) = \begin{cases} 0 & k\neq\ell \\ N/2 & k=\ell\neq 0\end{cases}$$

$$\sum_{j=0}^{N-1}\sin(jk\alpha)\cos(j\ell\alpha) = 0$$

$$\sum_{j=0}^{N-1}\cos(jk\alpha)\cos(j\ell\alpha) \begin{cases} = 0 & k\neq\ell \\ = N/2 & k=\ell\neq 0 \\ = N & k=\ell=0\end{cases}$$

Therefore the equations become

$$\frac{A_0}{2}N + A_1(0) + B_1(0) = \sum_{j=0}^{N-1} g_j$$

$$\frac{A_0}{2}(0) + A_1 \cdot \frac{N}{2} + B_1(0) = \sum_{j=0}^{N-1} g_j \cos(j\alpha)$$

$$\frac{A_0}{2}(0) + A_1(0) + B_1 \cdot \frac{N}{2} = \sum_{j=0}^{N-1} g_j \sin(j\alpha)$$

Hence:

$$A_0 = \frac{2}{N} \sum_{j=0}^{N-1} g_j$$

$$A_1 = \frac{2}{N} \sum_{j=0}^{N-1} g_j \cos(j\alpha)$$

$$B_1 = \frac{2}{N} \sum_{j=0}^{N-1} g_j \sin(j\alpha)$$

are the least mean square choice for $A_0$, $A_1$, $B_1$. Notice in this case the solution of the simultaneous equations is trivial.

UNITS OF FREQUENCY

In the previous example, the data is taken daily over a period of one year. To include the explicit units of time in the analysis, some conventional normalizations are used. The period of $\sin(x)$ is $2\pi$ radians i.e., $\sin(x + 2\pi) = \sin(x)$. Since the argument of sin must be unitless, a radians has the dimensions L/L (length of circumference of a circle subtended by the angle in question) / (length of the radius of the circle) and $2\pi$ corresponds to one complete revolution or period. If the independent variable being considered is time = t, then the argument of the sin must be multiplied by a quantity such that the resulting product has units radians: for example sin(wt) where w has units radians / time. Another normalization often used is sin $(2\pi ft)$ where f has units cycle/time and $2\pi$ is the conversion from cycles to radians. Further, if the period of interest is T then the normalization is sin $(2\pi t/T)$, which has period T, i.e., sin $(2\pi(t + T)/T)$ = sin $(2\pi t/T + 2\pi)$ = sin $(2\pi t/T)$. So that for period T, the corresponding frequency is f = 1/T cycles/time and the radian frequency is w = $2\pi/T$ radians/time.

Continuing the example, g(t) is temperature and t is in days. Let $g(j\Delta t) = g_j$ where j = 0, 1, ... N-1 correspond to the data at time t = 0, $\Delta t$, $2\Delta t$, ... , $(N-1)\Delta t$ days. Using this notation, the period is $T = N\Delta t$ days so that the approximation $\hat{g}(j\Delta t)$ can be written:

$$\hat{g}(j\Delta t) = \frac{A_0}{2} + A_1 \cos\left(\frac{2\pi}{T} j\Delta t\right) + B_1 \sin\left(\frac{2\pi}{T} j\Delta t\right) \qquad (2)$$

The sinusoids have period T = 366 days and frequency f = 1/T = 1/366 cycles/day. Eq. (2) can be put into a more convenient form using a simple trigonometric identity:

$$\hat{g}(j\Delta t) = \frac{A_0}{2} + C_1 \cos\left(\frac{2\pi}{T} j\Delta t - \theta_1\right)$$

where
$$C_1 = \sqrt{A_1^2 + B_1^2}$$
$$\theta_1 = \tan^{-1}\left(B_1/A_1\right)$$

$C_1$ is called the amplitude of the sinusoid and $\theta_1$ is the phase in radians. $C_1$ is the amplitude of the yearly cycle in the temperature data and a plot of $C_1$ versus frequency or period is usually drawn to illustrate this dependence.

## FINITE FOURIER SERIES

At this point it is natural to ask whether there is any significant 1/2 year period, 1/3 year period, 1/4 year period, etc.; or equivalently, any significant frequency components at 2 cycles/year, 3 cycles/yr, 4 cycles/yr, etc. These frequencies are called the harmonics of the fundamental frequency 1 cycle/year and such an analysis is called a harmonic analysis of the data. The answer can be obtained by fitting sinusoids at these higher frequencies to the data. Further, if enough sin and cosine terms are added so that the number of unknown Fourier coefficients is equal to the number of data points, then exact equality rather than a least mean square approximation can be achieved.

The specific form of the Fourier series which achieves equality between the N data points $g(j\Delta t)$ and the Fourier series is [2]:

$$g(j\Delta t) = \frac{A_0}{2} + \sum_{k=1}^{N/2-1} A_k \cos\left(\frac{2\pi k}{T} j\Delta t\right) + B_k \sin\left(\frac{2\pi k}{T} j\Delta t\right)$$
$$+ \frac{A_{N/2}}{2} \cos\left(\frac{\pi N}{T} j\Delta t\right) \qquad (3)$$

where:

$$\begin{matrix} A_k \\ B_k \end{matrix} = \frac{2}{N} \sum_{j=0}^{N-1} g(j\Delta t) \begin{matrix} \cos \\ \sin \end{matrix} \left( \frac{2\pi k}{T} j\Delta t \right) \qquad (4)$$

The N data points g (j $\Delta$ t) are transformed into N Fourier coefficients $A_0$, $A_1$, $B_1$, ... , $A_{N/2}$, and all the information contained in the record g (0), g ( $\Delta$ t), g (2 $\Delta$ t), ... g [ (N-1) $\Delta$ t] is also contained in the Fourier coefficients. Eq. 3 , gives the data in terms of the Fourier coefficients and eq. 4 gives the Fourier coefficients in terms of the data.

The amplitudes and phases are given as follows:

$$g(j\Delta t) = \frac{C_0}{2} + \sum_{k=1}^{N/2} C_k \cos\left( \frac{2\pi k}{T} j\Delta t - \theta_k \right)$$

where

$$C_0 = A_0$$

$$C_k = \sqrt{A_k^2 + B_k^2} \qquad k = 1, 2, \ldots, \frac{N}{2} - 1$$

$$C_{N/2} = \frac{A_{N/2}}{2}$$

and

$$\theta_k = \tan^{-1} \left( B_k / A_k \right) \qquad k = 1, 2, \ldots, \frac{N}{2} - 1$$

$$\theta_{N/2} = 0.$$

EXAMPLES:

The amplitude of the Fourier coefficients for the yearly temperature records and daily average dissolved oxygen data from the Delaware Estuary are shown in Fig. 1 [3].

As a second example, consider 3 days of solar radiation data, sampled every 1/2 hour. The fundamental period of the data is $T = 3$ days and the sampling interval $\Delta t = 1/48$ day. There are $N = T/\Delta t = 144$ data points. Fig. 2 presents an idealization of the data that has been used for photosynthetic oxygen analysis in streams, and the actual data, as well as the calculated amplitudes of the Fourier coefficients in each case.

Consider the actual record of solar radiation intensity and the calculated Fourier coefficients. The components corresponding to 1 cycle/day, and 2 cycles/day are quite large. But the other components are not zero. In fact they are due to the random fluctuations associated with real solar radiation, e.g. the effect of clouds passing overhead. In order to analyze their significance, it is necessary to characterize these random influences in some way.

DATA WITH RANDOM COMPONENTS - THE PERIODOGRAM

The first statistics that are usually calculated from a set of data with random components is the mean and variance. From eq. (4) it is clear that:

$$E(g_j) = \frac{1}{N} \sum_{j=0}^{N-1} g_j = \frac{A_0}{2}$$

A useful result is obtained from the calculation of the variance of $g_j$ :

$$\sigma^2 = \frac{1}{N}\sum_{j=0}^{N-1}\left[g_j - E(g_j)\right]^2$$

$$= \frac{1}{N}\sum_{j=0}^{N-1}\left[\sum_{k=1}^{N/2-1} A_k \cos\left(\frac{2\pi kj}{N}\right) + B_k \sin\left(\frac{2\pi kj}{N}\right) + \frac{A_{N/2}}{2}\cos(\pi j)\right]^2$$

$$= \frac{1}{N}\left\{\sum_{k=1}^{N/2-1} A_k^2 + B_k^2 + \frac{A_{N/2}^2}{2}\right\}\frac{N}{2}$$

$$= \frac{1}{2}\left\{\sum_{k=1}^{N/2-1} C_k^2 + 2\,C_{N/2}^2\right\} \qquad (5)$$

This result, known as Parseval's theorem [(4)], forms the basis of Variance spectra analysis. Eq. 5 is a formula which gives the variance in a record as a sum of components, each of which is associated with a particular frequency. To write this sum in the form of an integral, define the periodogram $I_k$ as:

$$I_k = \frac{N}{2}\Delta t\left(A_k^2 + B_k^2\right) \qquad k = 1, 2, \ldots, N/2 - 1$$

$$I_{N/2} = \frac{N}{2}\Delta t\left(\frac{A_{N/2}^2}{2}\right)$$

The units of $I_k$ are variance/frequency, and the integral of $I_k$ from k = 0 to N/2 gives the total variance in the record, i.e.

$$\sigma^2 = \frac{1}{N\Delta t}\sum_{k=1}^{N/2} I_k$$

The periodogram plays the same role in Variance analysis as the amplitudes play in Harmonic analysis, that is, $I_k$ is the variance density at the frequency $k/T$. A plot of $I_k$ versus the frequency $k/T$ is called the periodogram of the data, and it was first introduced into data analysis by Schuster [5] in 1898.

There is one major difficulty with periodogram analysis and that is the statistical behavior of the $I_k$ if the data being analyzed has significant random components.

Consider the following data points: $n_j$, $j = 0, \ldots, N-1$, where $n_j$ are independent Gaussian random variables with zero mean and variance $= \sigma^2$. To estimate $\sigma^2$ from the data, the sample variance $\hat{\sigma}^2$ is used:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{j=0}^{N-1} n_j^2$$

It is easy to show that $\hat{\sigma}^2$ has a chi square distribution with N degrees of freedom. The coefficient of variation of $\hat{\sigma}^2$ is $\sqrt{2/N}$ so that as N increases $\hat{\sigma}^2$ approaches the actual variance $\sigma^2$. If, however, the periodogram of $n_j$ is obtained, $I_k$, $k = 1, \ldots, N/2$, it is shown in the appendix that $I_k$ has a chi square distribution with 2 degrees of freedom, and the coefficient of variation is $\sqrt{2/2} = 1$. so that N increases, the $I_k$ do not fluctuate less violently and do not eventually approach the theoretical value, which in this case is $2\sigma^2 \Delta t$.

Fig. 3 presents two periodograms for N = 25 and N = 50. Notice the large scatter of the estimates; the mean value is shown as a dotted line in both cases. Notice also that in the case of N = 50, there are twice as many estimates as the N = 25 case.

Thus increasing the record length increases the number of estimates but not their statistical stability.

A solution to this dilemma has been provided by Daniell (1946)(6) who suggested that the $I_k$ be averaged in groups of 2L + 1. Thus, the estimate of the actual $I_k$ at frequency k/T, which is called the modified periodogram $\hat{I}_k$, is given as the average of the L adjacent raw periodogram estimates $I_k$ i.e.:

$$\hat{I}_k = \frac{1}{2L+1} \sum_{j=k-L}^{k+L} I_j$$

Now $\hat{I}_k$ is a chi square random variable with 2( 2L + 1 ) degrees of freedom and the coefficient of variation of $I_k$ is:

$$C.V.\left(\hat{I}_k\right) = \sqrt{\frac{2}{2(2L+1)}} = 1/\sqrt{2L+1}$$

A reasonable value for L would be L = 5 and the resulting estimate has a coefficient of variation of $1/\sqrt{11} = .30$. For the modified periodogram, an increase in the length of the record available can be used either to obtain twice as many estimates of the variance spectra with the same coefficient of variation, or to double the value of L and increase the statistical stability of the estimates of the variance spectra. Therefore, using the modified periodogram, a consistant estimate of the variance spectra can be obtained and the sampling fluctuations of this estimate can be specified.

## APPLICATIONS OF VARIANCE SPECTRA ANALYSIS

The applications that have been proposed for spectral analysis can be put into three categories: 1) To verify theoretical models of random processes by comparing the theoretical spectra to the spectra estimated from the data. An example of this type, taken from Taylor, is shown in Fig. The spectrum of longitudinal velocity in turbulent flow is compared to a theoretical prediction. 2) To unravel the periodicities inherent in a data record in order to understand further at what frequencies the major contributions to the variance occur and what their causes might be. 3) To predict the variance spectra of the output of a linear system given the variance spectra of the input. Examples of these applications will be presented during the lecture.

ACKNOWLEDGEMENT: The assistance provided by Prof. Robert V. Thomann of Manhattan College is gratefully acknowledged.

## APPENDIX

### DISTRIBUTION OF THE PERIODOGRAM

To calculate the probability distribution of the periodogram $I_k$, consider a sequence $n_j$, $j = 0, 1, \ldots, N-1$, of independent Gaussian random variables with zero mean and variance $= \sigma^2$.

The Fourier coefficients for this sequence have the form:

$$\begin{matrix} A_k \\ B_k \end{matrix} = \frac{2}{N} \sum_{j=0}^{N-1} n_j \begin{matrix} \cos \\ \sin \end{matrix} \left( \frac{2\pi k}{T} j \Delta t \right) \tag{6}$$

The probability distribution of $A_k$ and $B_k$ is found as follows: $n_j$ , $j = 0, \ldots, N-1$, are independent Gaussian random variables and $A_k$ , $B_k$ being linear combinations of independent Gaussian random variables are therefore themselves Gaussian random variables. It can also be shown that $A_k$ and $B_k$ are uncorrelated and therefore, being Gaussian, are independent (7)
The periodogram $I_k$ is defined as the sum of the square of two independent Gaussian random variables. It, therefore, has a chi square distribution with 2 degrees of freedom i.e. an exponential distribution. Finding the expected value of $I_k$ :

$$E(I_k) = \frac{N}{2} \Delta t \left[ E(A_k^2) + E(B_k^2) \right]$$

from Eq 6 :

$$\begin{aligned} E(A_k^2) &= \frac{4}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} E(n_i n_j) \cos\left(\frac{2\pi k j}{N}\right) \cos\left(\frac{2\pi k i}{N}\right) \\ &= \frac{4}{N^2} \sum_{j=0}^{N-1} E(n_j^2) \cos^2\left(\frac{2\pi k j}{N}\right) \\ &= \frac{4\sigma^2}{N^2} \sum_{j=0}^{N-1} \cos^2\left(\frac{2\pi k j}{N}\right) \\ &= \frac{2\sigma^2}{N} \end{aligned}$$

and similarly:

$$E\left(B_k^2\right) = \frac{2\sigma^2}{N}$$

so that:

$$E\left(I_k\right) = \frac{N}{2}\Delta t\left(\frac{4\sigma^2}{N}\right) = 2\sigma^2 \Delta t$$

Hence the distribution function of $I_k$ is:

$$\mathrm{Prob}\left\{I_k \geqslant x\right\} = \exp\left\{-\frac{x}{2\sigma^2 \Delta t}\right\}$$

the standard deviation is:

$$\text{s.d.}\ (I_k) = 2\sigma^2 \Delta t$$

and the coefficient of variation of $I_k$ :

$$\text{c.v.}\ (I_k) = \frac{\text{s.d.}\ (I_k)}{E\ (I_k)} = 1$$

## REFERENCES

1. Hamming, R.W. Numerical Methods for Scientists and Engineers. P. 68, McGraw-Hill Book Co., N.Y. 1962

2. ibid. p. 69

3. Thomann, R. V. "Time-Series Analyses of Water-Quality Data" Proc. A.S.C.E., SA 1, Feb. 1967

4. Hamming, op. cit. p. 70

5. Schuster, A. On the Investigation of Hidden Periodicities with Application to a Supposed 26 Day Period of Meteorological Phenomena. Terr. Magn. 3. 13-41 1898

6. Daniell, P. J. Discussion on Symposium on Autocorrelation in Time-Series, Supplement to the J. Roy Statist. Soc. 8, p. 88-90 1946

7. Davenport, W. B., and W. L. Root
An Introduction to the Theory of Random Signals and Noise, McGraw-Hill Book Co., N.Y. 1958 p. 149

# BIBLIOGRAPHY

## FAST FOURIER TRANSFORM


1. J. W. Cooley and J. W. Tukey "An Algorithm for the Machine Calculation of Complex Fourier Series" Math of Comput. Vol 19 pp. 297-301, April 1965.

    The paper describing the "Cooley-Tukey" algorithm for calculating Fourier coefficients, which made possible the computation of all the Fourier coefficients of very long sequences of numbers.

2. IEEE Transactions on Audio and Electro Acoustics. June 1967 Vol AU-15 No. 2. "Special Issue on Fast Fourier Transforms"

    Contains papers dealing with the application of the Fast Fourier Transform to Digital Filtering and Spectral Analysis.

3. D. R. Cox and P. A. W. Lewis. The Statistical Analysis of Series of Events. Methuen & Co. Ltd. 1966 pp. 96-107.

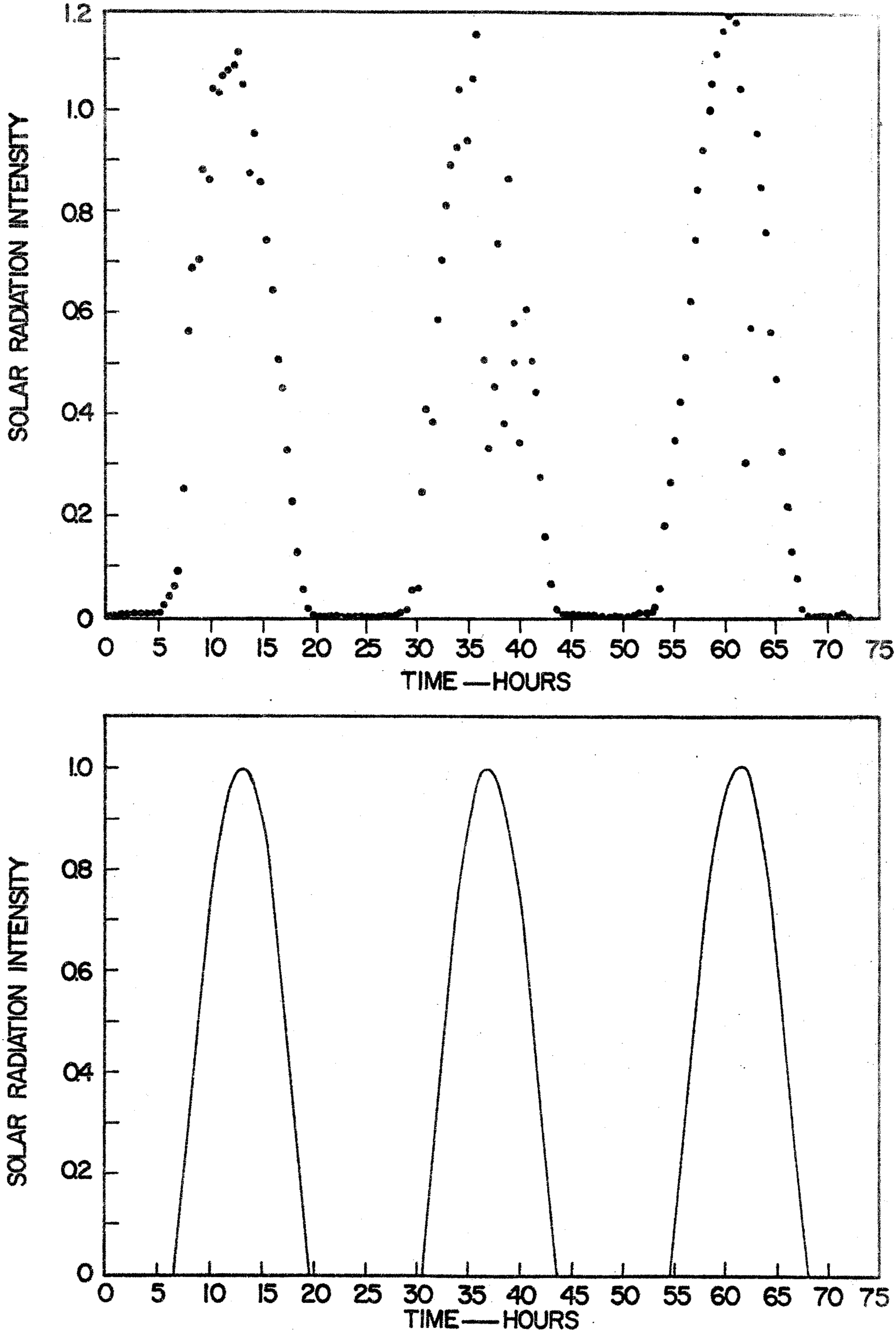
    A good high level discussion of the periodogram, it's statistics and various smoothing techniques.

4. HARM, RHARM. IBM System 360 Scientific Subroutine Package.

    Fortran IV Subroutines implimenting the Cooley-Tukey algorithm.

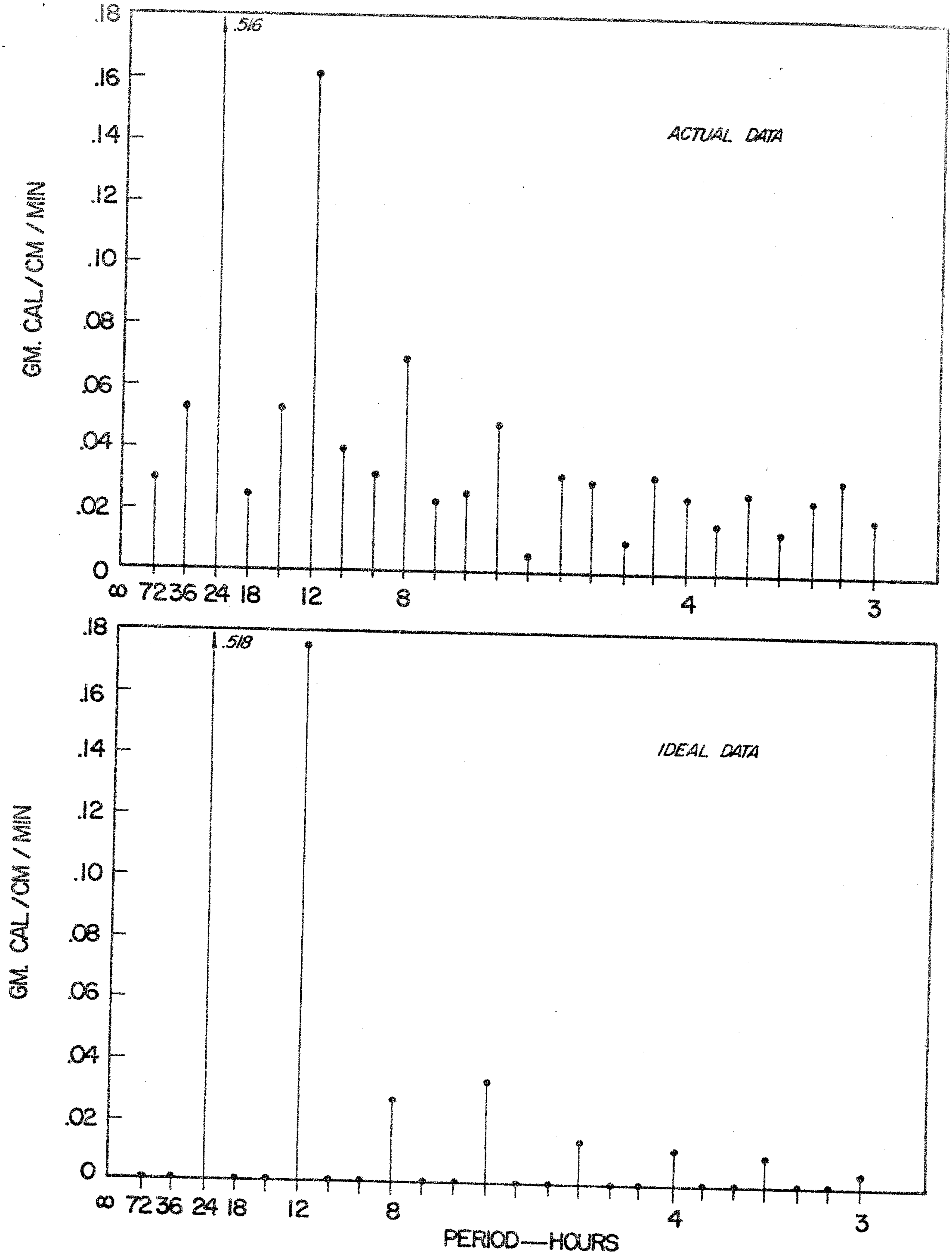
5. John W. Tukey "An Introduction to the Calculation of Numerical Spectrum Analysis", in Spectral Analysis of Time Series ed. B. Harris. J. Wiley & Sons.

    A very interesting overview from periodogram analysis, to the most modern techniques.

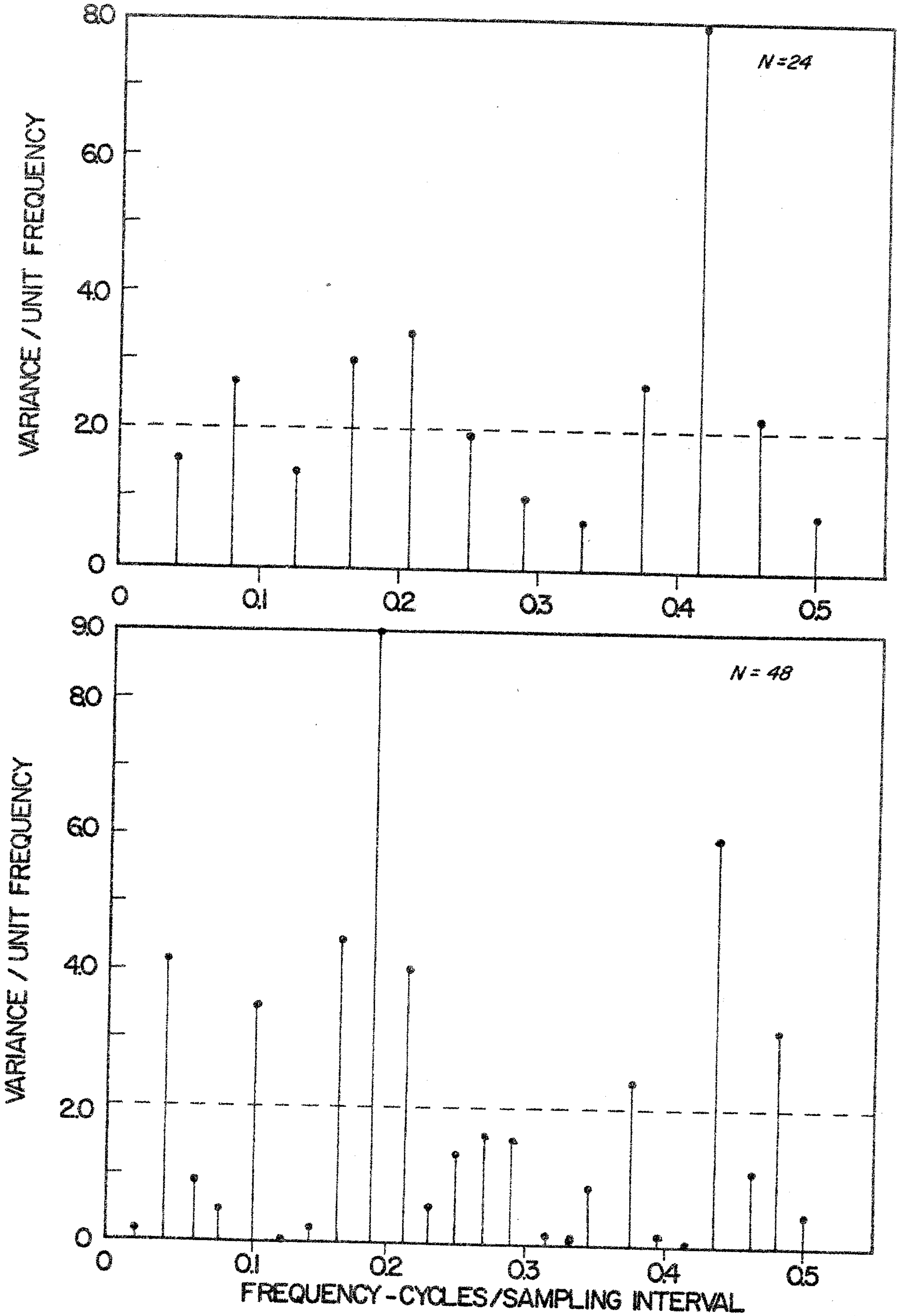


SOLAR RADIATION INTENSITY
1.2
1.0
0.8
0.6
0.4
0.2
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75
TIME — HOURS
SOLAR RADIATION INTENSITY
1.0
0.8
0.6
0.4
0.2
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75
TIME — HOURS

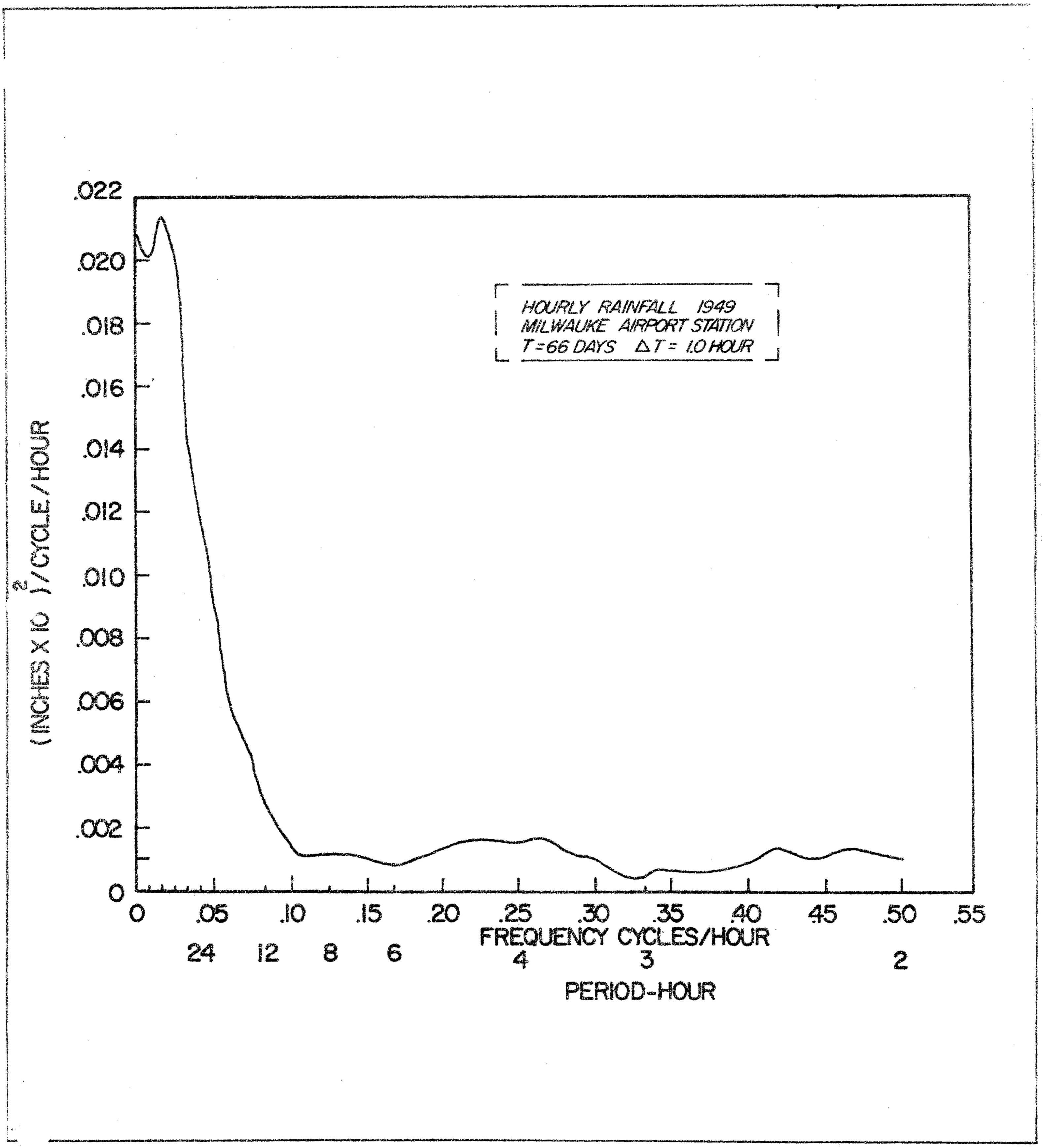

RARITAN BAY




AMPLITUDES OF FOURIER COEFFICIENTS
SOLAR RADIATION




PERIODOGRAM OF INDEPENDENT GAUSSIAN RANDOM VARIABLES




VARIANCE SPECTRA

# ANALYSIS OF VARIABILITY IN WASTE TREATMENT PLANT PERFORMANCE USING TIME SERIES TECHNIQUES [a]

by

Robert V. Thomann [b]

## Introduction

As nationwide interest has focused on water quality and pollution control, regulatory agencies have become increasingly aware of the need to continually examine the performance characteristics of waste treatment plants. Regulations regarding the efficiency of plants may specify ranges in effluent waste load, maximum discharge loads or specifications of allowable seasonal variations. For example, the Delaware River Basin Commission Regulations [1] require that the average number of pounds of ultimate carbonaceous oxygen demand in waste effluent to the Delaware Estuary for any five consecutive days during any month shall not exceed by more than fifty percent a waste allocation established by the Commission. An average of an unspecified number of samples is to be obtained monthly from May 1 through December 31. Design engineers will be required more and more to examine and estimate the expected variation in waste load from a given treatment process. Estimates of average performance while necessary will not be sufficient.

The reasons for increased attention to waste load variability are found in the effect that such variability has on stream water quality. It has always been known that treatment plant performance is variable. Detailed analysis of the variation and the development of means for predicting the effect of seasonal or random oscillations on river quality has not been carried out to any great extent.

---

(a) Presented at ASCE Second Nat. Symposium on Sanitary Engineering Research, Development and Design, July 15, 1969
(b) Assoc. Prof. Civil Engineering, Manhattan College, Bronx, N.Y.

Several statistical techniques can be applied to the time variations of waste treatment processes. The analyses in this paper are restricted to municipal plants. The purposes of time series analysis of existing municipal waste treatment plant data are to a) gain further insight into the nature of the variability of flow and biochemical oxygen demand (BOD) b) develop additional measures of performances of treatment plants which may prove useful in estimating future variability and c) relate the behavior of the treatment plant to resulting stream quality variations.

Because the variables under consideration are functions of time, care must be taken in the data analysis. Periodicities, trends and persistence in the records must be recognized and incorporated in any analysis scheme. For example, if there is a strong periodicity in the data, plotting on normal probability paper is not appropriate and may give misleading results. The techniques of classical statistics must therefore be used with caution when dealing with time variable quantities.

## Observed Treatment Plant Variability

In order to provide a setting for the analyses to follow, a visual examination of the time series of several plants is helpful. Data were obtained from eight plants ranging in flow from less than 10 MGD to 140 MGD and included municipalities with a relatively large percentage of industrial wastes. The plants were not limited geographically and included sources in California, Pennsylvania and Ohio among other locations. The two smallest plants had many data gaps especially in BOD values which precluded complete application of the techniques discussed below. However, some statistical analyses was performed on the data from these two plants. Sufficient data records were available from six of the eight plants to permit a complete examination of the time variable behavior. In all cases, one year of average daily data was used as the basic analysis unit. The initial step in the analysis approach was to therefore review machine plotted data to qualitatively determine significant periodicities or other trends.

Figure 1 is a time series plot of the final effluent 5-day BOD (lbs/day) for 1965 from a California activated sludge plant. This plant is subjected to a dominant seasonal influence resulting from canning operations. The substantial increase in load toward the latter part of 1965 reflects this additional input. There are other less well defined oscillations of higher frequency (shorter period) superimposed on the general trend of the series. This plot illustrates the hazards of failing to recognize time dependence in the data. A calculation of the overall mean and variance for the year would reflect the effects of the two month increase. The variance would be distorted on the high side and unless the seasonal oscillation was recognized, the variance could be misinterpreted.

Figure 2 shows a detailed time series plot of the first one-half year of data for a mid-West primary plant. The annual mean and range of $\pm 1$ standard deviation are shown. It is clear that significant short term oscillations exist in the record. Inspection indicates that the data are not distributed normally so that the estimate of the standard deviation, while important as a measure of dispersion should not be utilized with a Gaussian density function. The purpose of time series analyses is made clear by this example. If one attempts to follow the oscillation by eye, confusion results. The periodicity persists for a time, but then the maximum or minimun occasionally "disappears" or the periodicity does not persist or one notes that other oscillations appear to be superimposed on the approximately 7-10 day periodicity. Time series analysis represents a methodology which draws on a number of statistical techniques to aid in understanding this and other types of time variable phenomena.

## Theoretical Background

A complete review of the technique of time series analysis is given by Bendat and Piersol[2] and will not be repeated here. Applications have also been made to the analysis of stream and estuarine water quality data[3,4,5] Only the general outline of the theory is presented. Equally spaced data is assumed throughout.

For records with reasonably well-defined periodicities (see Fig. 1) it is advantageous to first perform an harmonic or Fourier analysis. The purpose of this analysis is to determine quantitatively the amplitudes and phase angles of the important periodic components and the contribution to the total variance of each of the harmonics.

If then the x (t) is represented as

$$x(t) = \bar{x} + \sum_{i=1}^{M} A_i \sin(i\omega t) + B_i \cos(i\omega t) + x_r(t) \quad (1)$$

where $\omega$ is the fundamental frequency $= 2\pi/T$ where T is the fundamental period (365 days in this work), t is time, i is the harmonic number, $\bar{x}$ is the mean value, $x_r$ is the residual variation not accounted for by the M harmonics and $A_i$ and $B_i$ are the Fourier or harmonic coefficients and are given by

$$A_i = 2/N \cdot \sum_{t=1}^{N} x_t \sin(i\omega t) \quad (2a)$$

$$B_i = 2/N \cdot \sum_{t=1}^{N} x_t \cos(i\omega t) \quad (2b)$$

where N is the total number of data points.

With $A_i$ and $B_i$ determined from the observed data, the amplitude and phase angles for the ith harmonic are given by

$$C_i = \sqrt{A_i^2 + B_i^2} \quad (3a)$$

$$\theta_i = \arctan A_i/B_i \quad (3b)$$

The variance accounted for by the ith harmonic is

$$a_i^2 = C_i^2/2 \quad , \quad i < N/2$$

$$a_i^2 = C_i^2 \quad , \quad i = N/2 \tag{4}$$

This analysis is relatively easy to program and provided that M is not too large, the analysis uses only a small amount of time on a medium size computer. For larger M, specialalgorithms can be used that compute the Fourier coefficients. In any event, the analysis provides a rapid means for estimating the amplitude, phase angle and variance of any dominant periodicities. Harmonic analysis can then be followed by construction of the residual record $X_h(t)$ with subsequent random process analysis of that record.

For records that appear to oscillate randomly, i.e. where no periodicities are evident and for residual records after harmonic removal, a series of analyses can be performed which essentially detect any statistical persistence and the distribution of variance over frequency. The determination of persistence is given by the auto-correlation function which is computed by

$$R(\tau) = \frac{1}{N - m} \sum_{\tau=0}^{m} X_t \cdot X_{t+\tau} \tag{5}$$

where x (t) is here understood to have zero mean, $\tau$ is the "lag number" and the maximum number of lags is m. The Fourier transformation of $R_x$ ($\tau$) yields the power spectrum (more properly the variance spectrum) given by the computational formula

$$\tilde{G}_x\left(\frac{k f_c}{m}\right) = 2h\left[R_0 + 2\sum_{r=1}^{m-1} R_r \cos\left(\frac{\pi r k}{m}\right) + (-1)^k R_m\right] \tag{6}$$

where h is the time interval between samples, k is the discrete harmonic number and $f_c$ is the so-called Nyquist frequency given by

$$f_c = 1/2h$$

and represents the highest frequency about which information will be obtained. In this work, h = 1 day and $f_c$ is one cycle per (2) days. A final smoothing of the estimate given by Eq.(6) is necessary.(2)

The importance and utility of spectrum analysis is that it is an estimate of the distribution of variance over frequency. The labelling of the abscissa is either in terms of frequency or period. The ordinate is variance/cycle per time. Thus, in the records analyzed here, h = 1 day, m (the maximum number of lags) is 30 so that the frequency scale is cycles/60 days and runs from 0 to 30. The period scale therefore extends from infinitely long periods to a period of 2 days. The choice of the maximum number of lags depends on several factors including recognition of confidence limits on the estimates. The value of m = 30 provides for a reasonable statistical estimate. The total area under the spectralcurve is equal to the total variance of the record. Any fractional area between two frequencies therefore represents the fraction of the total variance occuring in the continuous (as opposed to discrete) frequency band.

The above analysis dealt with the spectrum of single records. Often it is of interest to examine the cross-correlation and cross-spectral relationships between two variables both of which are varying as functions of time. It is important to recognize that two times series may be correlated only at particular frequencies through a specific phase shift.

The cross-correlation function $R_{xy}(\tau)$ is given formally by

$$R_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)\, y(t+\tau)\, dt \qquad (8)$$

where T is the record length. The Fourier transform of $R_{xy}(\tau)$ is the cross-spectral density function

$$G_{xy}(f) = 2\int_{-\infty}^{\infty} R_{xy}(\tau) \exp(-i2\pi f\tau)\, d\tau \qquad (9)$$

which is a complex function. The real part $C_{xy}(f)$ is called the cospectrum and the imaginary part $Q_{xy}(f)$ is the quadrature spectrum. Computional formulae are given in (2).

A useful measure of the degree of correlation between two time series is the coherency function given by

$$\gamma^2_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_x(f)\, G_y(f)} \qquad (10)$$

where $\gamma^2$ is the coherency, $|G_{xy}(f)|$ is the modulus of $G_{xy}(f)$ and is given as

$$|G_{xy}(f)| = \sqrt{C_{xy}^2(f) + Q_{xy}^2(f)}$$

Note that in an analogous fashion to simple correlation, the coherency function is bounded between zero and one.

In addition to gaining insight into the time variable behavior of treatment plants, time series analyses can form an important input into theoretical models of waste plants and models of stream and estuarine water quality. This is illustrated in Figure 3. If data were available on the influent and effluent of a primary plant, one could consider the plant to be a "black box" and compute the overall behavior of the plant from its measured input and output. A similar analysis can be made of a secondary plant or any stage between raw influent and a treatment process. Thus, one knows the input and output of a complicated system and is attempting, without writing down a deterministic model, to obtain information on how the "black box" transforms an input into an output.

Figure 4 is an example of where one knows the input, say the waste effluent spectrum, and in addition has a deterministic model of the "black box" and wishes to predict the output dissolved oxygen (DO) spectrum. Here one operates on the input to produce an output in contrast to the previous example where input and output are known and the "model" connecting the two is desired.

The relationship between a single random input and output of a fixed parameter linear system is given by [2]

$$G_y(f) = |H(f)|^2 G_x(f) \tag{11}$$

$$G_{xy}(f) = H(f)\, G_x(f) \tag{12}$$

where H ( $f$ ) is a complex valued function called the frequency transfer function. One can compute this transfer function in two ways: (a) by observing input and output spectra and cross-spectra and using Eq. (12), or (b) by constructing a theoretical model of the system from which H ( $f$ ) can be determined analytically, independent of measured data. Eq (12) is generally to be preferred for estimating H ( $f$ ) since it provides an unbiased estimate. The frequency transfer function provides information on how the system attenuates and phase shifts the input. Procedure for calculating H ( $f$ ) from measured data are given in (2).

Where a theoretical model of a stream or estuary is available, the transfer function H ( $f$ ) can be obtained analytically [6] and can be used to predict the distribution of variance of BOD or dissolved oxygen (DO). For a non-dispersive stream, the modulus of H ( $f$ ), designated $|H(f)|$ represents the amplitude attenuation and for BOD is given by

$$|H(f)|_L = \frac{\exp(-K_r x/u)}{Q} \tag{13}$$

where x is distance downstream from the outfall, $K_r$ is the BOD decay coefficient, u is the stream velocity, Q is the river flow and the subscript, L, represents BOD. For DO

$$|H(f)|_D = \frac{K_d}{K_a - K_r}\left(\exp(-K_r x/u) - \exp(-K_a x/u)\right) \tag{14}$$

Where $K_a$ is the reaeration coefficient, $K_d$ is the deoxygenation coefficient and the subscript D represents DO deficit. Thus for DO, the amplitude attenuation is zero at the outfall, builds to a maximum at some distance downstream and then decreases again to zero. Eqs. (11) and (12) show that the output spectrum is given by the input spectrum times the square of the transfer function modulus. It is interesting to note that Eq. (14) is the same form as the classical Streeter-Phelps equation but has a quite different interpretation.

## Results of Analyses

In order to provide initial insight into plant behavior, several preliminary statistical analyses were performed. These included computation of means, $\bar{X}$, standard deviations, s, coefficient of variation, $s/\bar{x}$ and estimates of probability density functions. With respect to the latter analyses, no consistent pattern emerged from the eight plants. Probability histograms ranged from almost uniform to normal and log-normal density functions. This was to some extent a consequence of the periodicitis inherent in several of the records.

Table 1 shows some of the statistical properties of the records. The coefficients of variation for the BOD from the secondary plants are interesting and reflect wide variability in the records. A value of greater than 1 ($s=\bar{x}$) was computed for one plant and three other secondary plants had values greater than 0.5. This is important since it is an indication of the variability to be expected in downstream water quality. It is also obvious that a meaningful probability density function for the secondary effluent must be everywhere positive.

Figure 5 graphically displays some of the statistical analyses. The upper plot indicates a general tendency for the coefficient of variation (C.V.) to increase as the size of the plant (as measured by the % reduction of the mean) increases up to the secondary treatment range. The C.V. holds fairly constant from the raw waste (0% reduction) to the primary treatment range. Another and different measure of the performance of the plants is given in the lower plot of Figure 5. This graph shows the % reduction of the variance versus the % reduction of the mean. In both cases, the reduction is referenced to the incoming waste. There is a general tendency for the plants at the primary treatment level to be somewhat more efficient in the reduction of variance than in the reduction of the mean. In one case however, (Plant #4) there was an increase of the primary effluent variance over the variance in the raw waste. At the secondary treatment level, variance reduction is about equal to reduction of the mean. These simple results indicate the wide range in flow and BOD variability and highlight the need to further investigate plant behavior.

Harmonic and Spectral Analysis

Only Plants #3-#8 had sufficient BOD data to permit further detailed time series analysis. As indicated previously, where dominant periodicities are evident in the data, an harmonic analysis should first be performed to obtain information on amplitudes, phase angles and variance. Figure 1 illustrated such a record for plant #4 which is subject to seasonal waste load fluctuations. Eqs. (2a) and (2b) can be used to obtain the necessary coefficients. The results of the harmonic analysis of this record are shown in Table 2. Note that in terms of pounds/day in the effluent, the coefficient of variation is 1.03. The dominance of the seasonal variation is evident in the total variance of 82.3% accounted for by the first five harmonics. It can also be noted that because the phenomena is not sinusoidal, but instead is a time function with a positive increase over only one-quarter year (from day 180-270), the harmonics 2-5 are important. The higher harmonics modify the 365 day harmonic by "flatening" out the record at the beginning and end of the year. The amplitude of the second harmonic is particularly interesting.

Spectral plots of the effluent BOD at Plant #4 before and after harmonic removal are shown in Figure 6. The "before harmonic removal" spectrum represents an analysis of the record as shown in Figure 1. The "after harmonic removal" spectrum is a result of the analysis of the residual record after the dominant seasonal influence was removed. It is seen that the removal of the first five harmonics accounts for the major portion of the low frequency variance (from zero to about 4 cycles/60 days). A dominant peak in the spectrum occurs at a period of 7 days with secondary peaks at 3.5 and 2.3 days, the principal harmonics of the 7 day period. There is also some indication of a 30 day peak which is noticeable only after the removal of the dominating influence of the low frequency components. It is also interesting to note that the residual standard deviation of 14,500 lbs/day is still 44% of the annual mean.

Figure 2 illustrated the waste load from Plant #6 which indicated a dominant short term oscillation. Figure 7 shows the influent spectrum and primary effluent spectrum for one year of data at Plant #6. The dominant influence of the seven-day peak is obvious, indeed the peak is almost as significant as the low frequency end of the spectrum. This is in contrast to the spectra of Plant #4, shown in Figure 6. The results of the harmonic analysis of the influent and effluent data for Plant #6 are given in

Table 3. It is seen that harmonics 1-5 account for significantly less variance than the case of Plant #4 (Table 2). The dominance of the 7-day period is evident in that approximately 25% of the total variance is accounted for by this harmonic.

Figure 8 shows the 7-day and 3.5 day periodic components for the primary effluent of Plant #6, computed from the harmonic analysis. The figure indicates the interaction of these two periods. Although the 7-day oscillation has a peak positive value of almost 50,000 lbs/day, the effect of the 3.5 day oscillation is to modify the peak due to a difference in phase angles. The sum of the two periodicites shows a tendency to "flatten out" the discharge during the middle of the week at about 35,000 lbs/day.

Figure 9 is a plot of the primary effluent spectrum for Plant #6, before and after harmonic removal. The area or variance under the spectrum after removal is about 55% of the variance before removal. It is important to note that even after removal of the discrete harmonics #52, 104 and 156 that peaks still exist in these regions of the spectrum. This illustrates the difference between discrete frequency components as obtained from Fourier analysis and estimates of the variance contained in a continuous frequency band. This also illustrates the usefulness of spectral analysis since one can therefore estimate the total variance contained in a given frequency band. Thus for the frequency region surrounding the 7, 3.5 and 2.3 day peaks, it is estimated that about 30% of the total residual variance is contained in this part of the spectrum. As an order of magnitude then, about 50-60% of the total variance of the original record is due to a 5 day on and 2 day off phenomenon. Any waste sampling program or water quality analysis will obviously have to recognize this variability.

Harmonic and spectral analyses on the remaining plants generally indicated periodicities similar to those that were evident in the plants just discussed but not as dominant. Figure 10 shows the influent, primary and secondary effluent BOD mg/l spectra for Plant #3. The relative weakness of the seven day phenomenon is evident in the influent and for the secondary effluent is entirely absent. Further investigation to increase the resolution of the spectrum in the seven day region may provide further insight. Plants #5, 7 and 8 generally tended to show peaks in the same region. In all cases, the effects of seasonal influences although present were not as dominant as the case of the seasonal canning operation given by Plant #4.

## Cross - Spectral Analysis

Various analyses were carried out on the relationships between flow, influent BOD and effluent BOD as given by estimation of the cross-spectra for Plants #3, 5, 7 and 8. In general, the results indicated negligible coherency (as computed from Eq. 10) between flow and influent BOD although for Plant #3 there was a tendency at the low frequency end for flow to be inversely correlated to influent BOD. Figure 11 shows the coherencies for three pairs of records for Plant #3. With the raw influent BOD (mg/l) as the input forcing function and primary effluent BOD (mg/l) as the response function (see Figure 3), coherencies are statistically significant at the 5% level over a wide range of frequencies. This significance level is the level above which one could expect to obtain a coherence value by chance alone. Thus, there is a 5% probability that a coherence level greater than about 0.35 will be obtained by chance. With primary effluent BOD as the input and secondary effluent as the output, coherencies are significant only in two regions of the frequency domain. The most interesting region is the seven day period. Finally, the coherency between raw influent BOD and secondary effluent BOD is everywhere insignificant indicating that the secondary effluent BOD variability is not significantly influenced by the raw influent variability.

Figure 12 shows a similar plot for Plants #5 and #8 where again significant coherencies were computed between influent and primary effluent and low coherencies between influent and secondary effluent.

## DISCUSSION OF RESULTS

### Treatment and Plant Behavior

The harmonic analyses indicated the need to critically examine a plot of the time series to determine important "lines" in the spectrum. The magnitude of the 5 day on and 2 day off phenomenon, especially in Plant #6, is particularly important. It shows the quasi-deterministic nature of waste effluents which must be recognized in any sampling program. The spectrum analysis results also indicated that the variance of the effluents tends to be concentrated in the low frequency end in addition to the variance concentration at the 7, 3.5 and in some cases 2.3 day periods. The order of magnitude of the coefficients of variation is indicative of the wide variability in load. The coherency analyses indicated the relative independence of secondary effluent variability from raw influent variability. One could speculate that this independence results from introduction of the biological treatment step. This treatment process introduces a new and essentially different type of variance due to purposefully stimulated biological activity. The secondary settling step does not appear to provide the necessary additional reduction in variance similar to that provided by the primary tanks.

In order to further understand the nature of the plant behavior, a "black box" analysis can be carried out using the results of the cross-spectral analysis. Thus Eq. (12) is used where $G_{xy}(f)$ and $G_x(f)$ are known and the complex frequency transfer function H (f) is computed. The modulus of H ( f ) is the amplitude attenuation and indicates in a gross way, the manner in which a linear treatment plant operates on a given input.

The results for Plant #3 are shown in Figure 13 and for Plant #5 in Figure 14. There is a tendency for the amplitude attenuation of the frequency transfer function to behave as if "plug flow" was dominant between the raw influent and primary effluent. If complete plug flow were present, the amplitude attenuation would be a constant over all frequencies. For the treatment step from primary effluent to secondary effluent, the estimate for Plant #3 indicates a tendency for this step to reflect some mixing. One is led to this conclusion because the amplitude attenuation decreases to zero at high frequencies in systems that exhibit mixing. It is not possible with the data at hand to determine the degree of mixing.

These results indicate how one can utilize spectral analysis to verify theoretical developments of treatment plant behavior. If plug flow is assumed for the primary treatment step of a particular plant and a mathematical model is constructed on that assumption, then a good check on the validity of the assumption is to compute the amplitude attenuation from the observed spectra and cross-spectra.

Effect on Receiving Water Quality

In addition to gaining insight into plant behavior, the other motivation for examining treatment plant variability lies in the effect of load variations on the reeeiving water body. A variety of mathematical models exist for streams and estuaries which relate waste load input to a water quality output, say dissolved oxygen. The stream models as noted previously permit direct analytical computation of the modulus of the frequency transfer function. Eq. (14) gives $H(f)$ for DO deficit in a non-dispersive river. As shown by that equation, $H(f)$ ) is independent of frequency and all oscillations are transmitted downstream. The only change in the oscillations results from the biochemical and reaeration reactions. This is consequence of the "no-mixing" assumption often made for streams.

To illustrate the use of a waste effluent spectrum, a hypothetical river system was subjected to the primary effluent spectrum of Figure 7. The following conditions were assumed: Q = 1000 MGD, Kd = Kr = 0.2/day and Ka = 0.4/day. Using Eq.(11) which assumes linearity, single input and no extraneous "noise", the output DO spectrum can be computed as a function of distance downstream. Figure 4 illustrates the system. The results are shown in Figure 15.

At the outfall, Eq. (14) indicates a DO spectrum of zero. Figure 15 shows the build-up of the spectrum to a maximum at 4 days travel time. At 15 days travel time, the spectrum approaches zero. The seven day peak in the effluent spectrum is reflected in the DO oscillation. The maximum effect of the peak occurs at the location of the critical DO deficit, an unfortunate consequence. The approximate maximum amplitude of DO due to the seven day peak would be about 1.6 mg/l, a substantial amount. The spectrum itself does not provide information on the phase angle. The results of the harmonic analysis however would indicate that the minimum DO value due to the seven day effluent peak would occur on Sunday, at a location equivalent to four days travel time.

For estuaries, where tidal mixing assumes an important role, it can be shown that the modulus is a function of frequency. The consequence of this dependence is that the estuary, through tidal dispersion, will dampen out high frequency oscillations in the waste load input. The range over which high frequencies are damped depends on a number of factors, including the magnitude of the tidal dispersion. Under intensive dispersion, approaching completely mixed systems, the seven day oscillation (higher frequencies) could be damped completely.

## CONCLUSIONS

Times series analyses have been applied to several waste treatment plant records to determine the degree of variability in plant performance and to obtain further insight into treatment plant behavior. The results indicate a high degree of variability in secondary effluent BOD as measured by the coefficient of variation. Values ranged from 0.2 to 1.1. The distribution of this variance over frequency showed a strong seasonal dependence for a plant subjected to a seasonal canning load. Five day-on and two day-off oscillations were also noted in the power spectrum of several plants. For one plant, the order of magnitude of the amplitudes of this weekly oscillation was about 35,000 lbs/day or about 20% of the mean load. The variability in effluent load should be incorporated in any sampling program and efforts should be made in treatment plant design to minimize variability. As an illustration assume that the residual waste load is normally distributed and one wishes to estimate the extent of an effluent sampling program. If the C.V. is 1.0 and one wanted to be 95% certain that a measured mean would differ from a "true" mean by no more than 50%, a total of 36 samples would be necessary. At a sampling interval of one day, approximately one month of sampling by a regulatory agency would be necessary. If the C.V. is 0.5, about 9 samples would be required. This example of sampling requirements would apply only to the random component of the effluent. Total sampling reguirements would have to include incorporation of any deterministic phenomenon.

The results of the time series analyses also indicate the effect that load variability can have on the water quality of a stream. Indeed, in the final analysis, this is the primary reason for being concerned with plant variations. Dissolved oxygen variations of 1.0 mg/1 or greater appear to be possible from a mathematical model of the stream and the observed treatment plant spectrum. The theory indicates that the maximum DO variability will occur at the location of the critical DO deficit.

References


1. Delaware River Basin Commission, Basin Rules and Regulations, Water Quality, Dec. 1967, DRBC, Trenton, N.J. 32 pp.

2. Bendat, J.S. and Piersol, A.G. Measurement and Analysis of Random Data. J. Wiley, 1966, 390 pp + xv

3. Gunnerson, C.G. Optimizing Sampling Intervals in Tidal Estuaries Jour. San. Eng. Div., ASCE, Vol. 92, No. SA 2, Proc. Paper 4799, April 1966, pp 103-125

4. Thomann, R.V. Time Series Analyses of Water Quality Data. Jour. San. Eng. Div. ASCE, Vol. 93, No. SA 1, Proc. Paper 5108, Feb. 1967, pp 1-23

5. Wastler, T.A. and Walter, C.M. Statistical Approach to Estuary Behavior. Jour. San. Eng. Div. ASCE, Vol. 94, No. SA 6, Proc. Paper 6311, Dec. 1968, pp 1175 - 1194

6. Thomann, R.V. Systems Analysis and Simulation in Water Quality Management. Proceedings, IBM Scientific Computing Symp. Water and Air Res. Management, 1968, pp 223-233.

Acknowledgment

This project was supported by a research grant from the Federal Water Pollution Control Administration. Data processing and additional computational facilities were supplied by Hydroscience, Inc. Grateful appreciation is expressed to Emmanuel Mehr of New York University who assisted in the computer programming and to Donald J. O'Connor and Dominic Di Toro who provided important insight into the topic of the paper.

TABLE 1

SOME STATISTICAL PROPERTIES OF EIGHT TREATMENT PLANTS

| Plant No. | Mean Flow (MGD) | C.V.(1) Flow | Mean Inf BOD(2) (mg/l) | C.V. Inf. BOD | Mean Pri. Effl. BOD (mg/l) | C.V. Pri. Eff. BOD | Mean Sec. Eff.1 BOD (mg/l) | C.V. Sec. Eff.1 BOD (mg/l) | Remarks |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2.5 | 0.21 | 183 | .26 | 128 | .21 | 23 | 0.50 | N.J. Sec. Plt.-Little ind. waste |
| 2 | 6.5 | 0.32 | 125 | .27 | 100 | .23 | 14 | 0.82 | Conn.Sec. Plt.-Little ind. waste |
| 3 | 17.5 | 0.17 | 275 | .27 | 205 | .27 | 35 | 1.07 | Ill.Sec.Plt.-Mod. ind., operating problems |
| 4 | 60.6 | 0.19 | 453 | .26 | 319 | .39 | 59 | 0.79 | Cal.Sec.Plt.-Heavy ind., Seasonal canning load |
| 5 | 100.7 | 0.15 | 180 | .25 | 94 | .28 | - | - | Penn.Pri.Plt.-mod.to heavy ind. waste |
| 6 | 102.2 | 0.21 | 260 | .33 | 202 | .34 | - | - | Ohio Pri.Plt.-mod. to heavy ind. waste |
| 7 | 107.1 | 0.27 | 163 | .30 | 116 | .27 | - | - | Penn.Pri.Plt.-Mod.ind. waste |
| 8 | 140.2 | 0.14 | 215 | .17 | - | - | 57 | 0.22 | Penn.Sec.Plt.-Mod. to heavy ind. waste |

(1) C.V. - Coef. of variation = $s/\bar{x}$

(2) All BOD values are 5-day

TABLE 2

RESULTS OF HARMONIC ANALYSIS OF EFFLUENT BOD (LBS/DAY), PLANT #4

Mean - 33,200, Std. Dev. = 34,300

| Harmonic | Period (Days) | $A_i$ (1000 lbs/day) | $B_i$ (1000 lbs/day | $C_i$ 1000 lbs/day | % Variance |
|---|---|---|---|---|---|
| 1 | 365 | - 13.25 | - 21.56 | 25.3 | 27.2 |
| 2 | 182.5 | - 9.17 | 25.65 | 27.2 | 31.5 |
| 3 | 121.7 | 18.02 | - 1.98 | 18.1 | 13.9 |
| 4 | 91.2 | - 9.38 | - 10.21 | 13.9 | 8.1 |
| 5 | 73.0 | - 1.35 | 6.09 | 6.2 | 1.6 |
| | | | | Total Variance Due to Harmonics, 1-5 | 82.3 |

TABLE 3

## RESULTS OF HARMONIC ANALYSIS OF PLANT #6

| Harmonic | Period (Days) | Inf. BOD Amplitude (1000 lbs/day) | Inf. BOD % Variance | Pri.Effl.BOD Amplitude (10000 lbs/day) | Pri.Effl.BOD % Variance |
|---|---|---|---|---|---|
| 1 | 364 | 34.1 | 8.3 | 24.6 | 8.9 |
| 2 | 182 | 14.6 | 1.5 | 13.9 | 2.2 |
| 3 | 121.3 | 20.3 | 2.9 | 15.4 | 2.7 |
| 4 | 91.0 | 17.6 | 2.2 | 12.0 | 1.6 |
| 5 | 72.8 | 15.2 | 1.6 | 10.3 | 1.2 |
| 52 | 7.0 | 58.5 | 24.4 | 48.0 | 26.5 |
| 104 | 3.5 | 17.0 | 2.0 | 11.6 | 1.6 |
| 156 | 2.3 | 7.5 | 0.4 | 4.9 | 0.3 |
| | | Total | 43.3 | | 45.0 |
| | | C.V. Before Removal | 0.38 | | 0.38 |
| | | C.V. After Removal | 0.29 | | 0.29 |

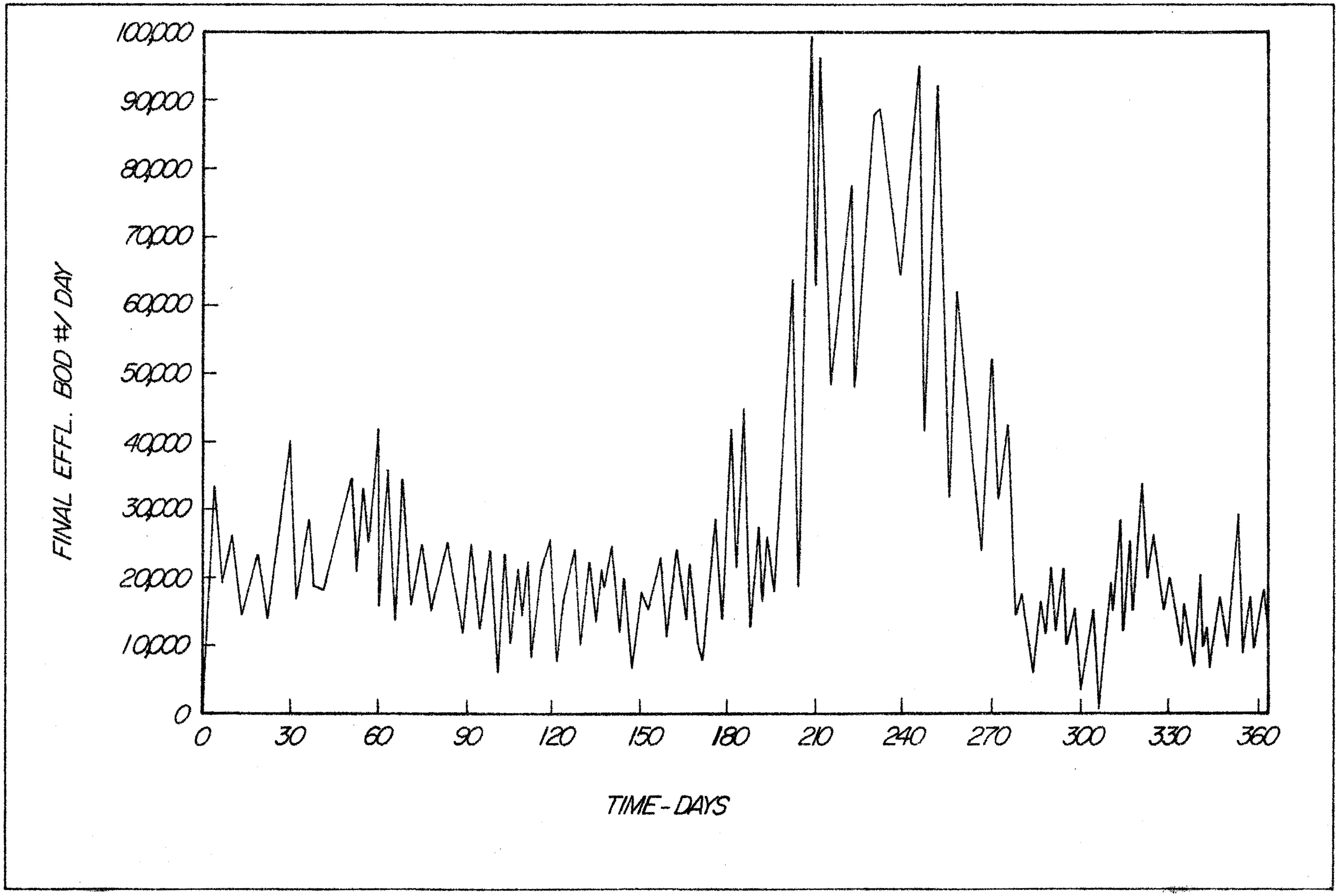



Fig. 1. Time series plot, Plant #4, activated sludge effluent seasonal load variation

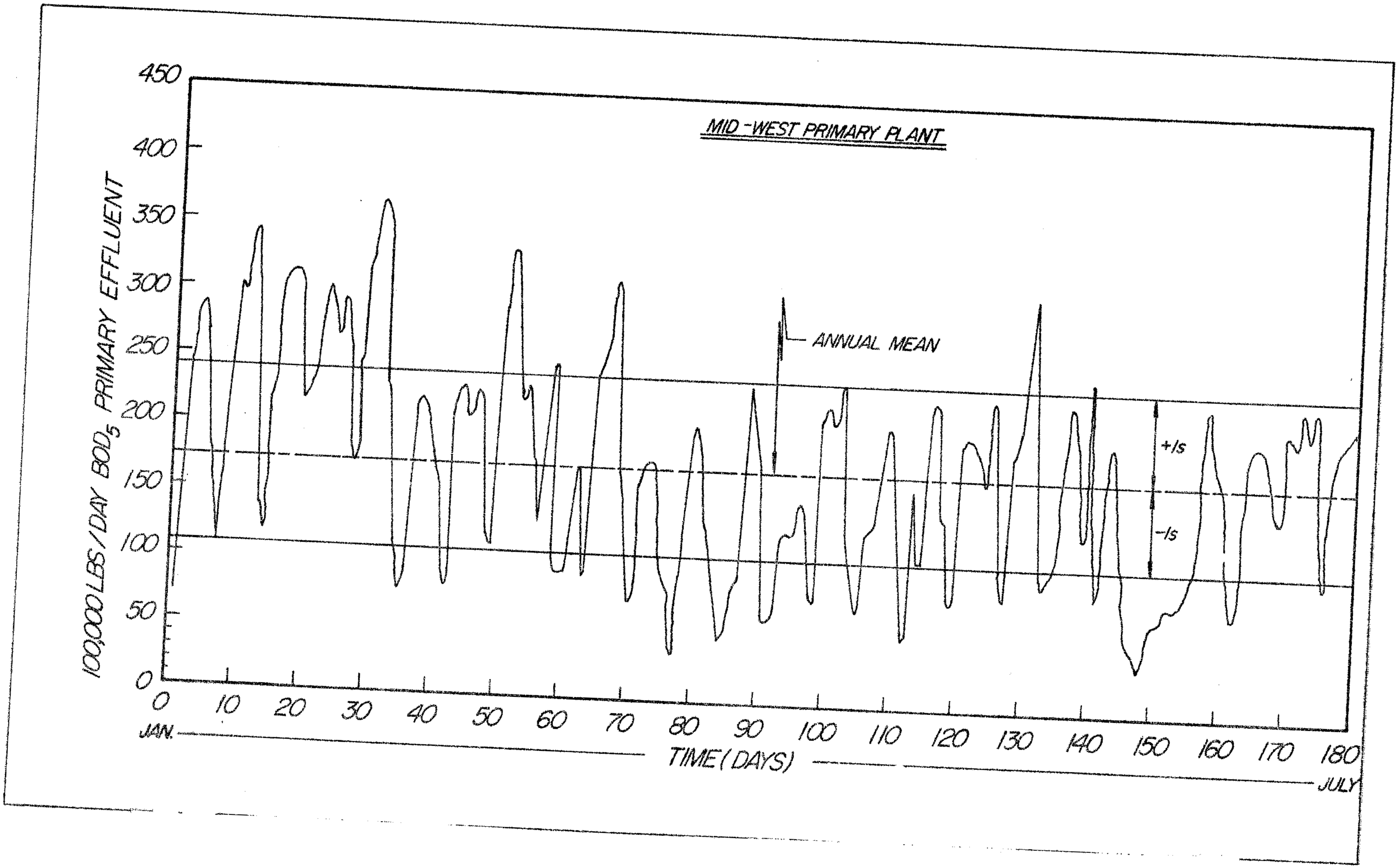



Fig. 2. Time series plot, Plant #6, primary effluent

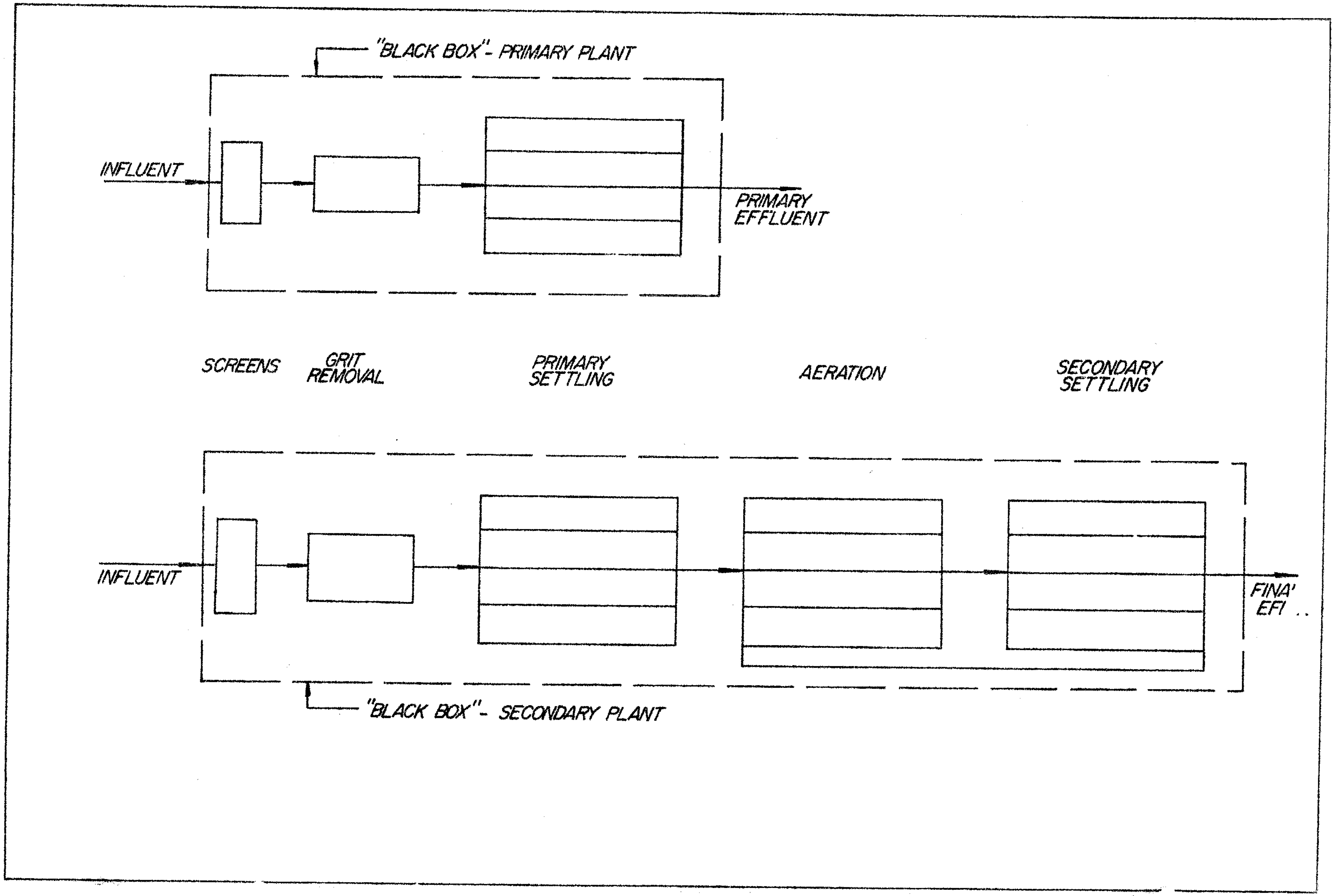



Fig. 3. "Black Box" representation of treatment plant

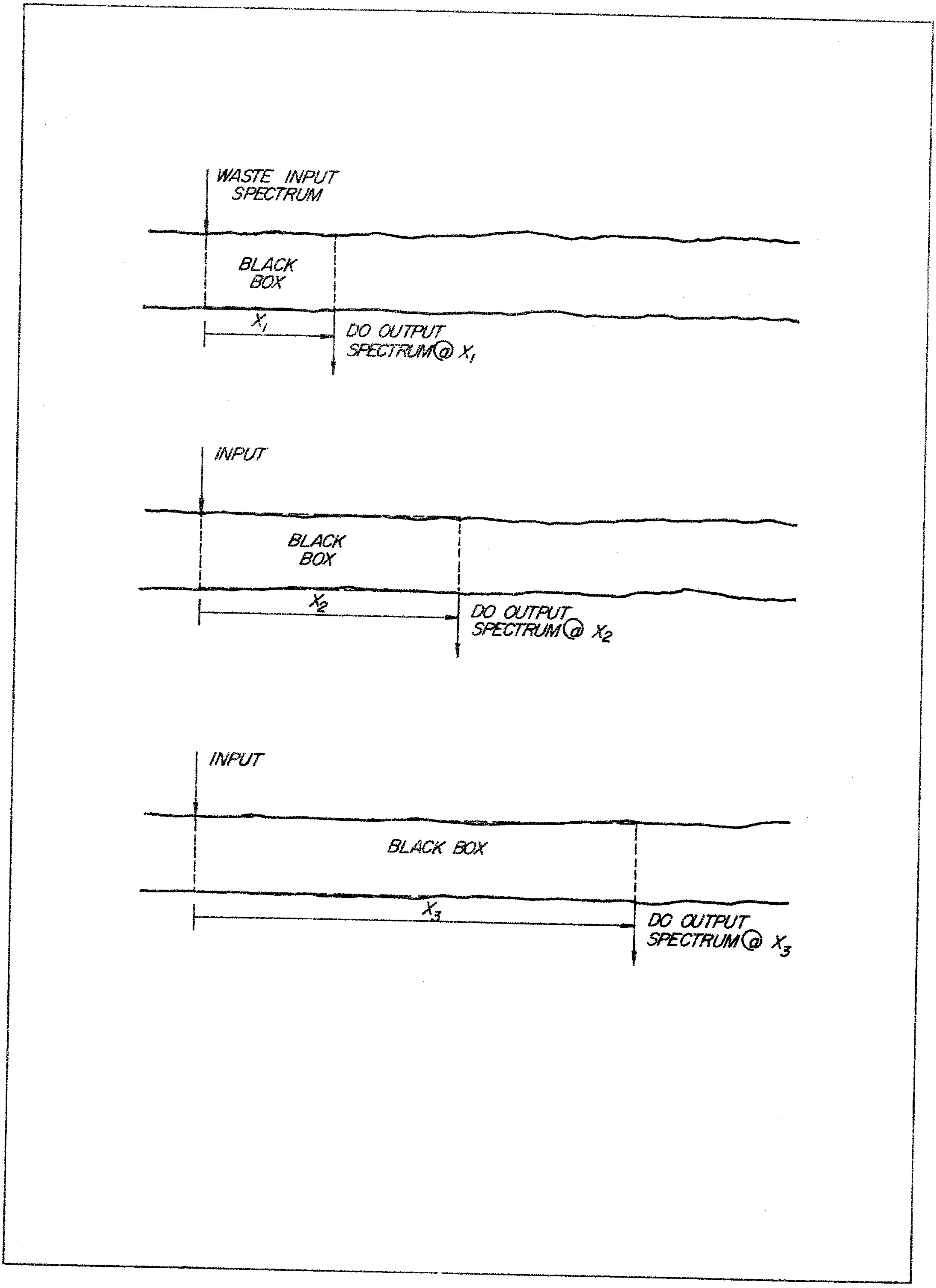



Fig. 4. "Black Box" representation of stream

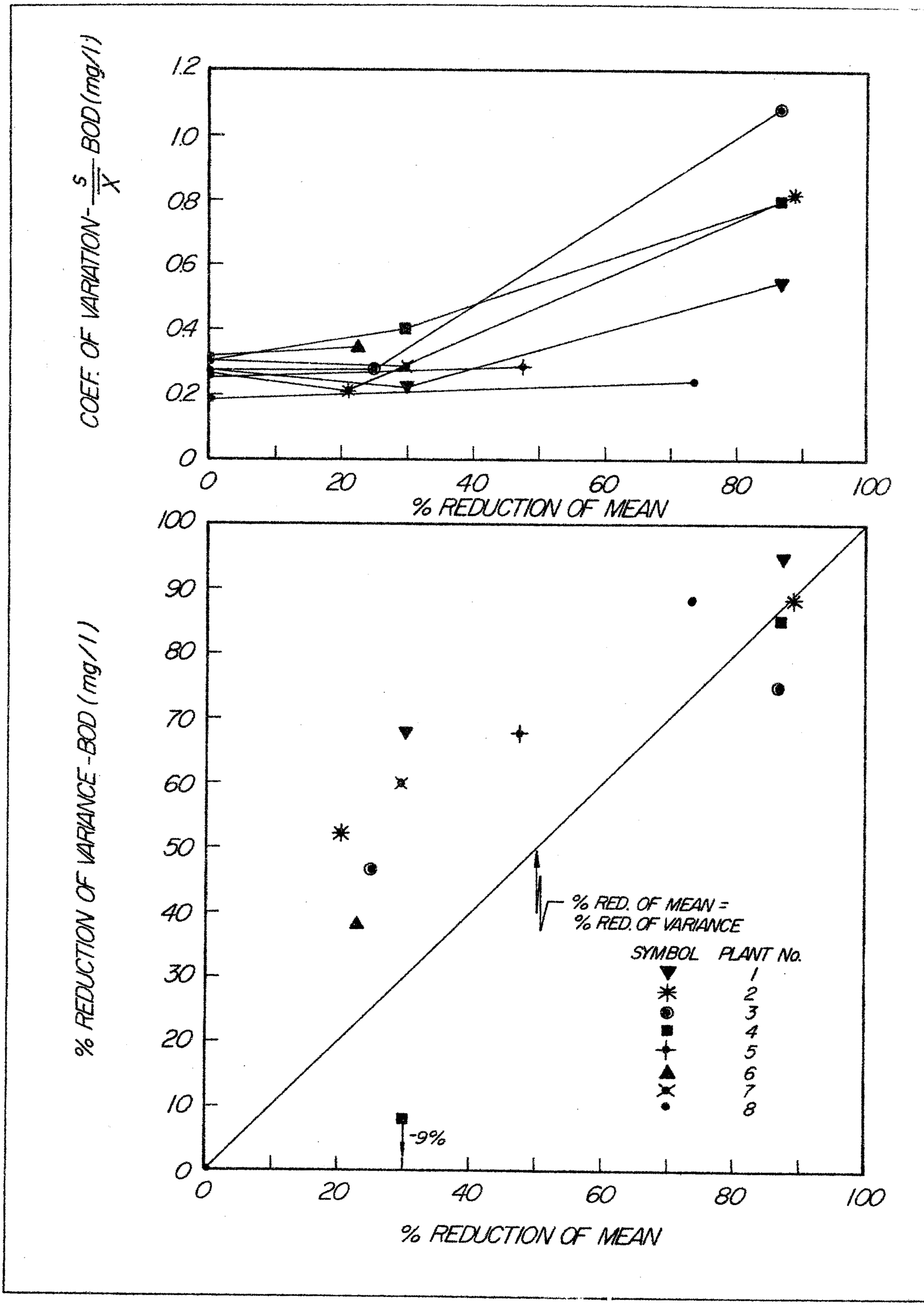



Fig. 5. Some results of statistical analyses of eight plants

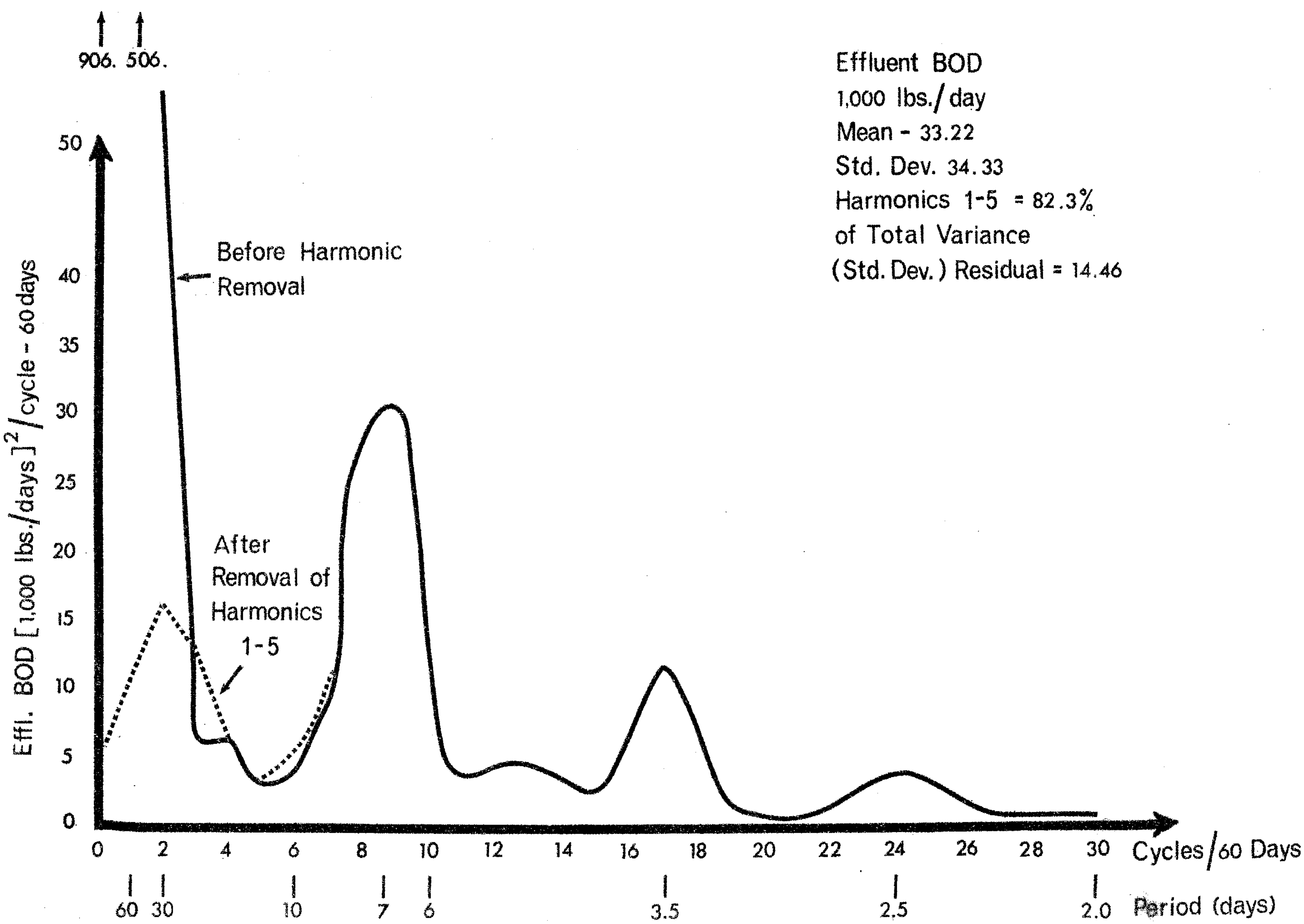



Fig. 6. Power spectrum, Plant #4, before and after harmonic removal

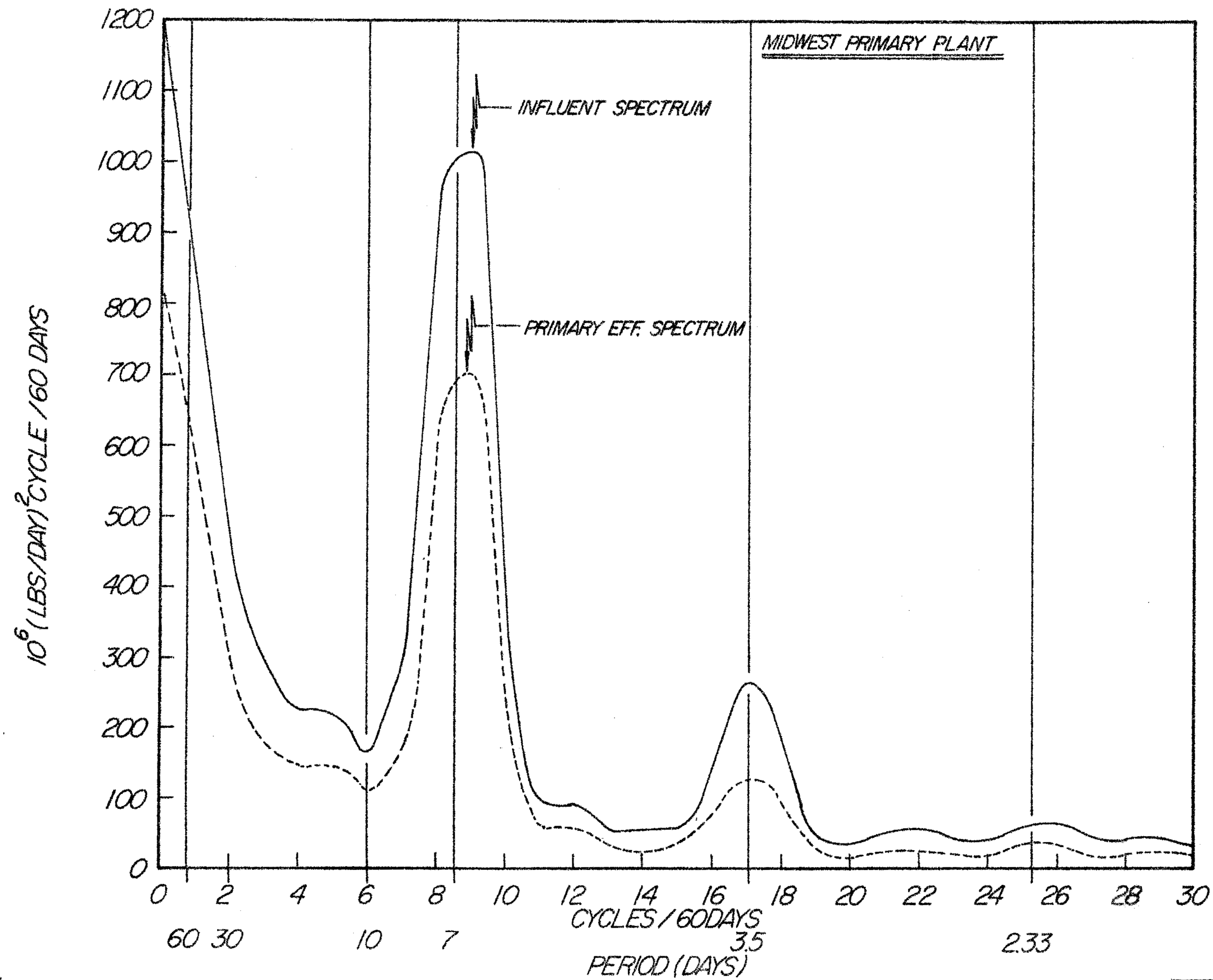
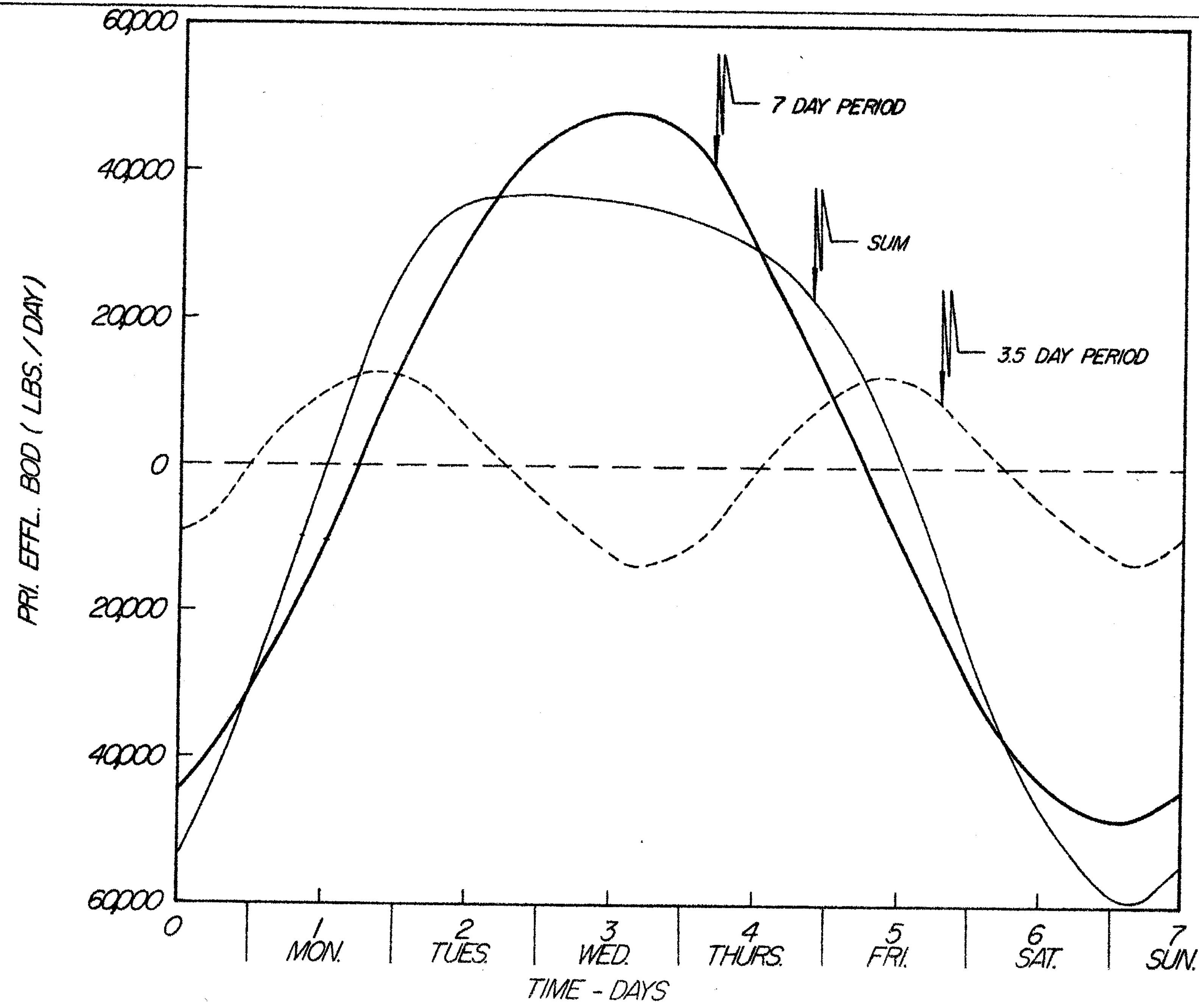



Fig. 7. Power spectrum, Plant #6, influent and primary effluent


7 DAY PERIOD
SUM
3.5 DAY PERIOD
60,000
40,000
20,000
0
20,000
40,000
60,000
PRI. EFFL. BOD ( LBS. / DAY )
0
1 MON.
2 TUES.
3 WED.
4 THURS.
5 FRI.
6 SAT.
7 SUN.
TIME - DAYS

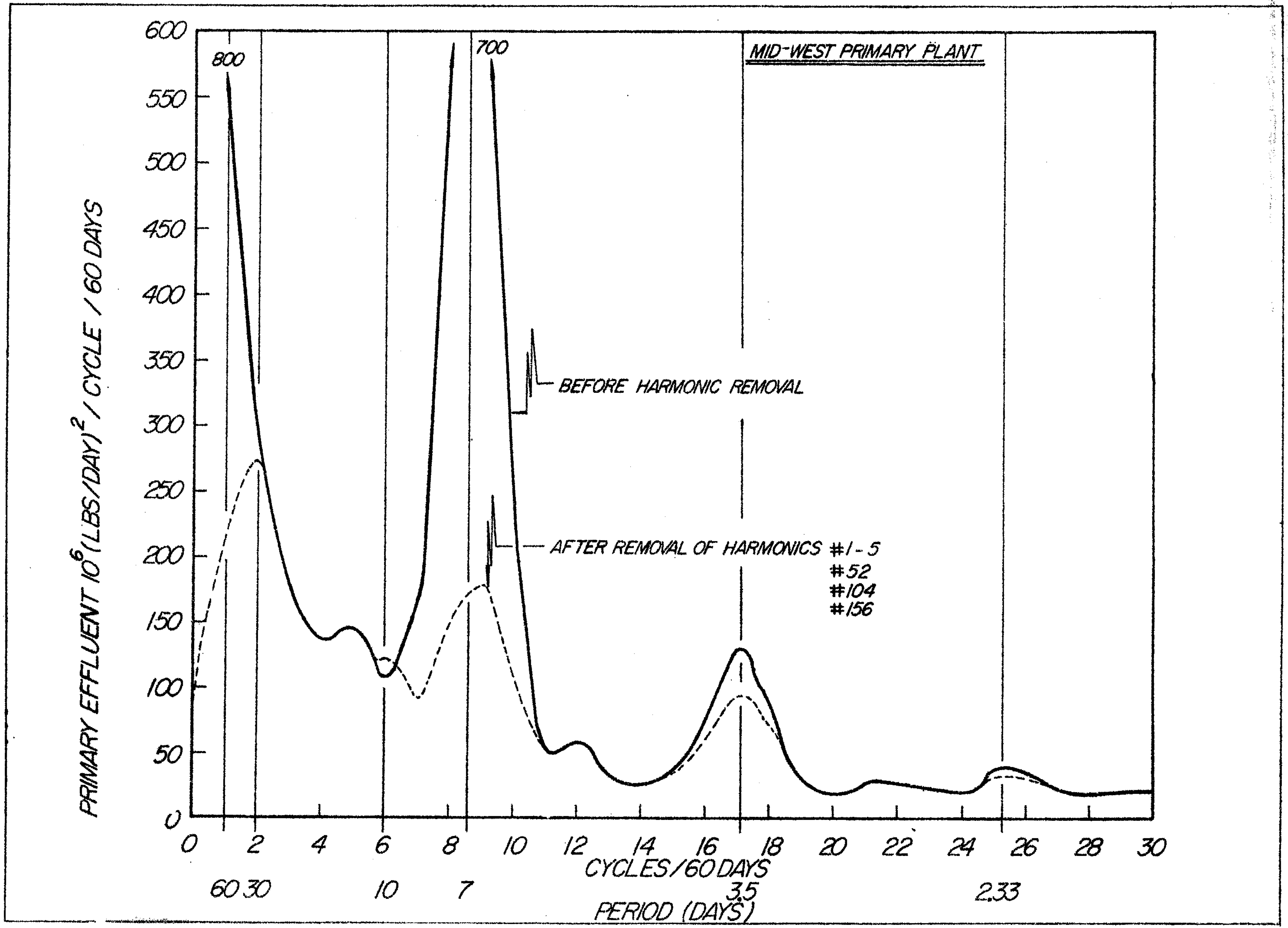



Fig. 9. Power spectrum, Plant #6, before and after harmonic removal

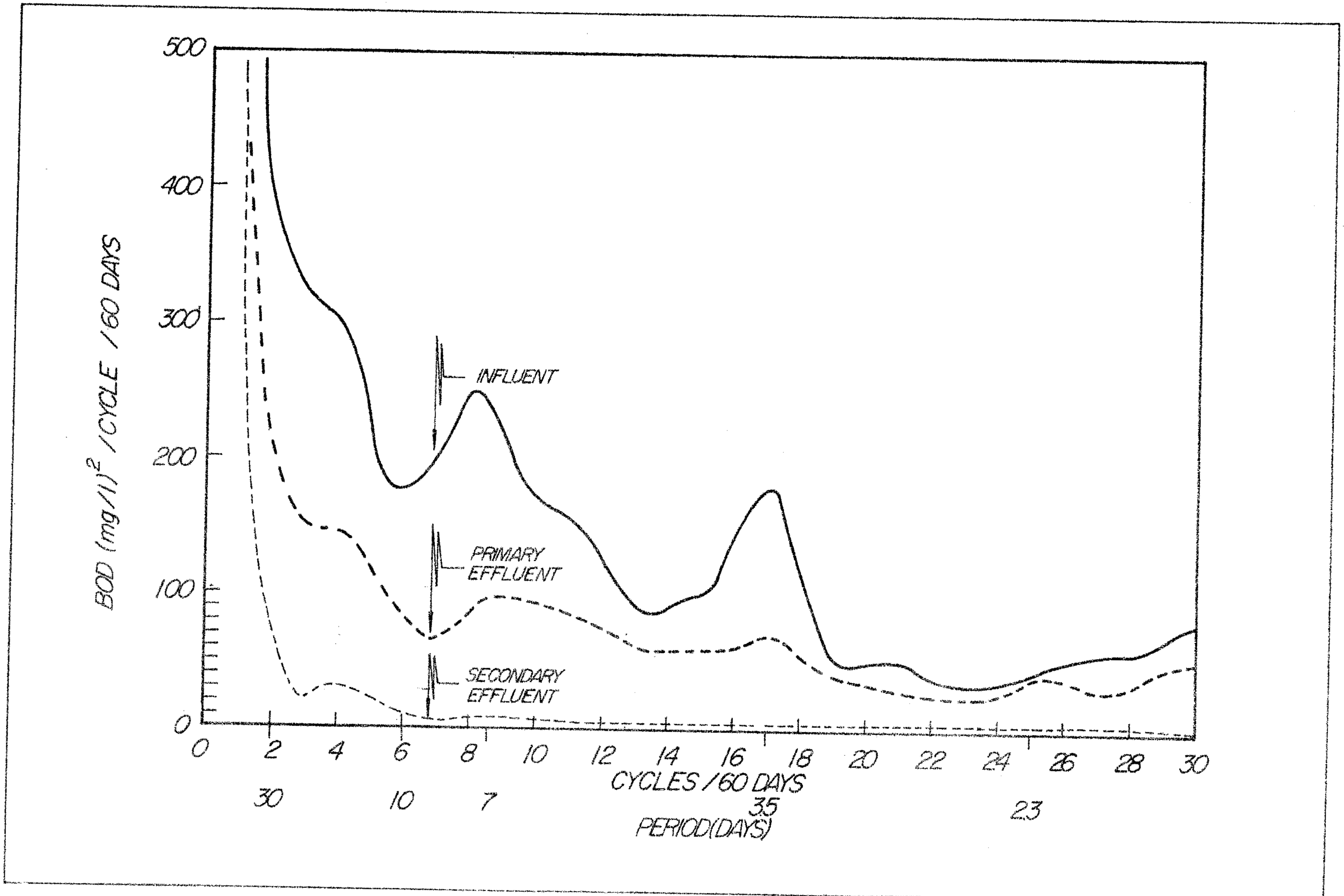



Fig. 10. Power spectrums, influent, primary and secondary effluent, Plant #3

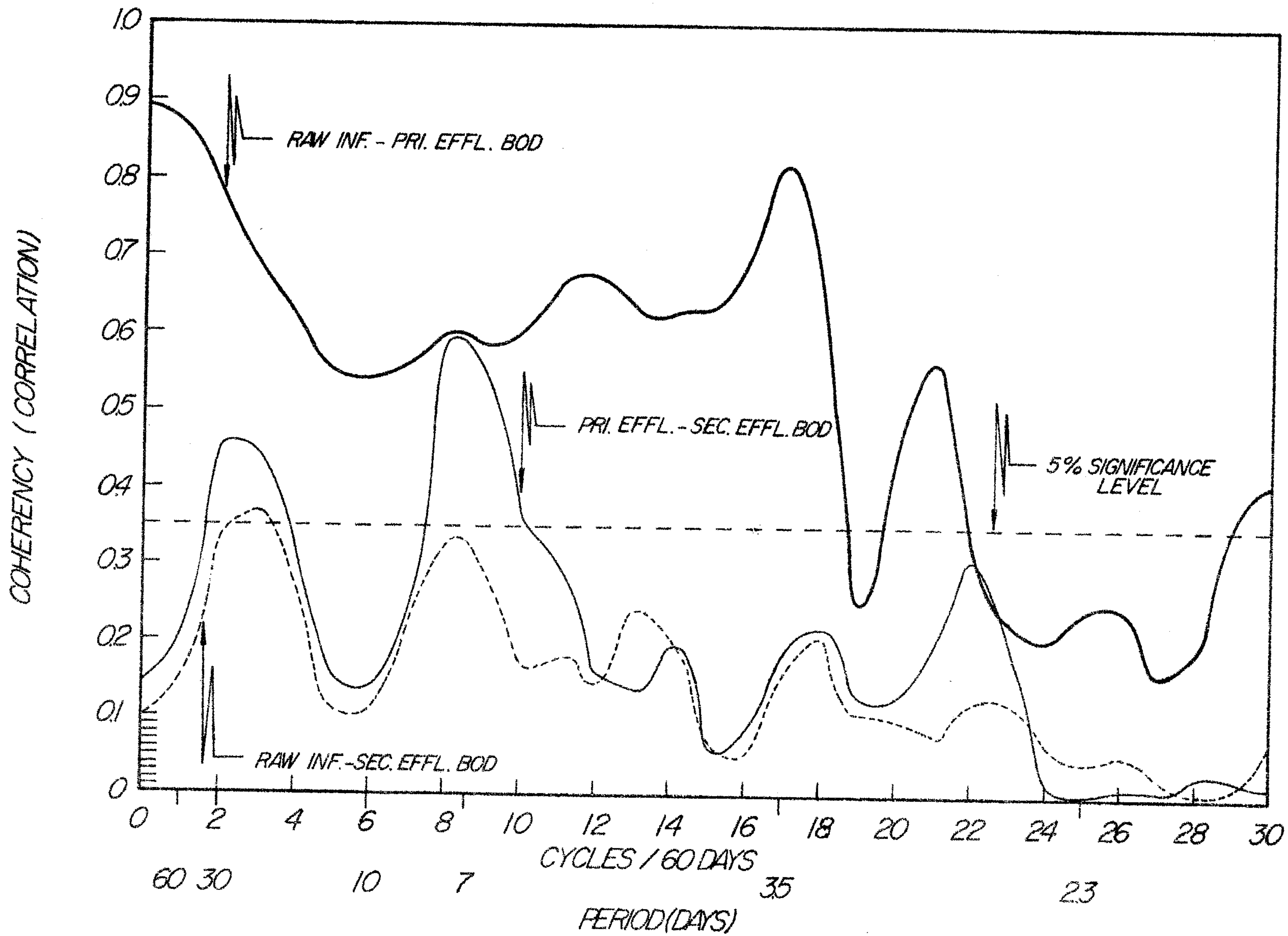



Fig. 11. Coherency plots, Plant #3

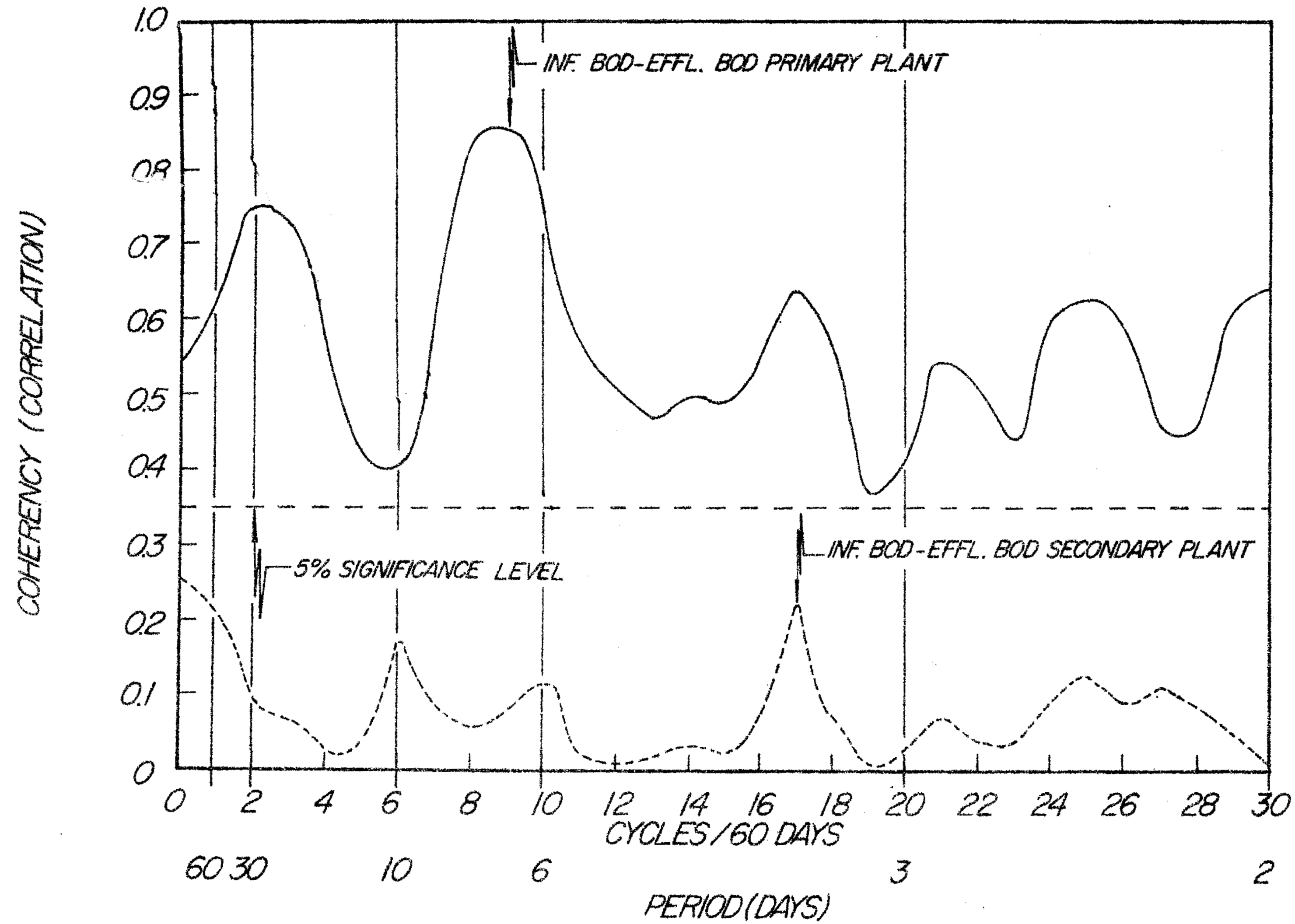



Fig. 12. Coherency plots, primary Plant #5, secondary Plant #8

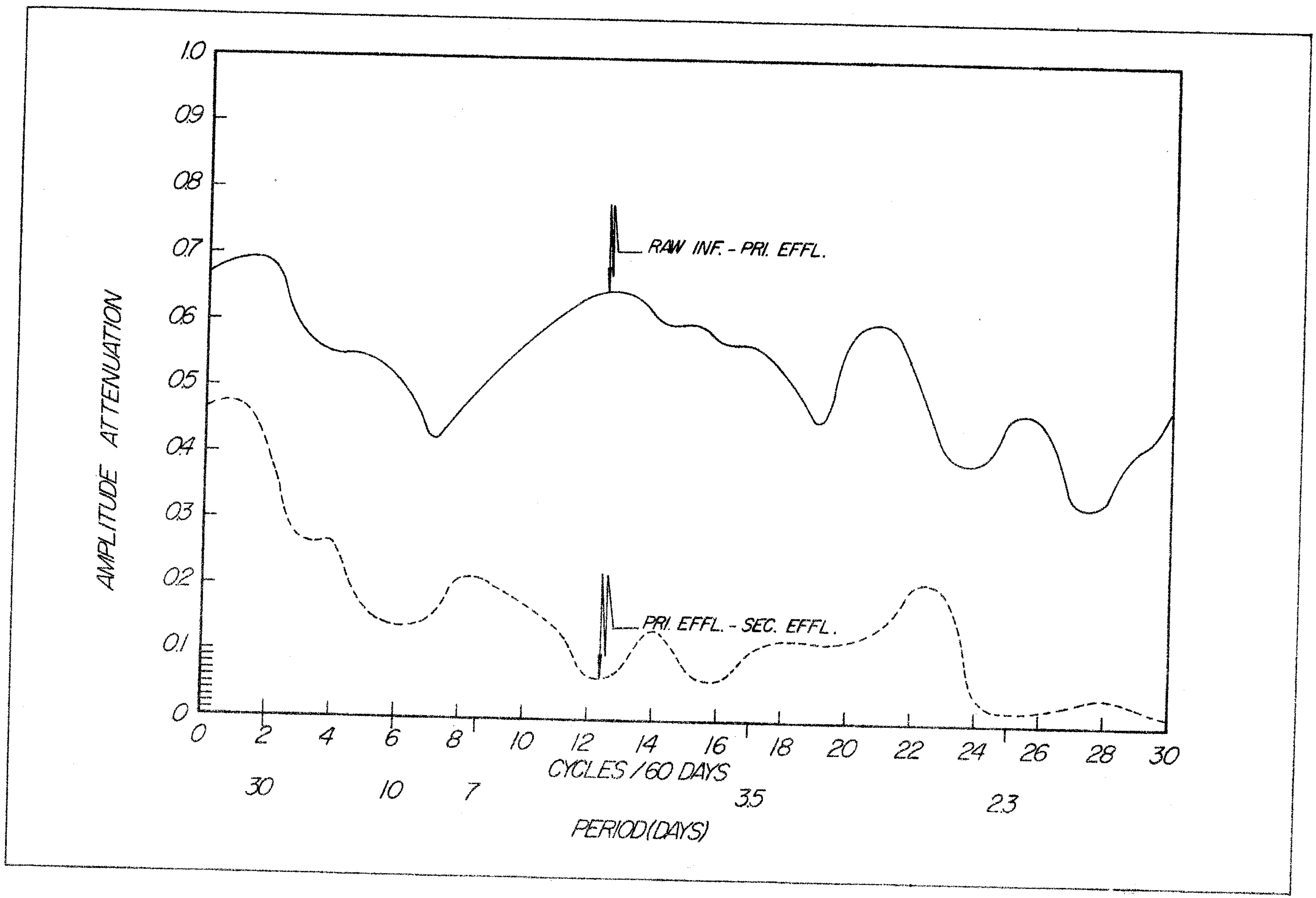



Fig. 13. Amplitude attenuation, Plant #3

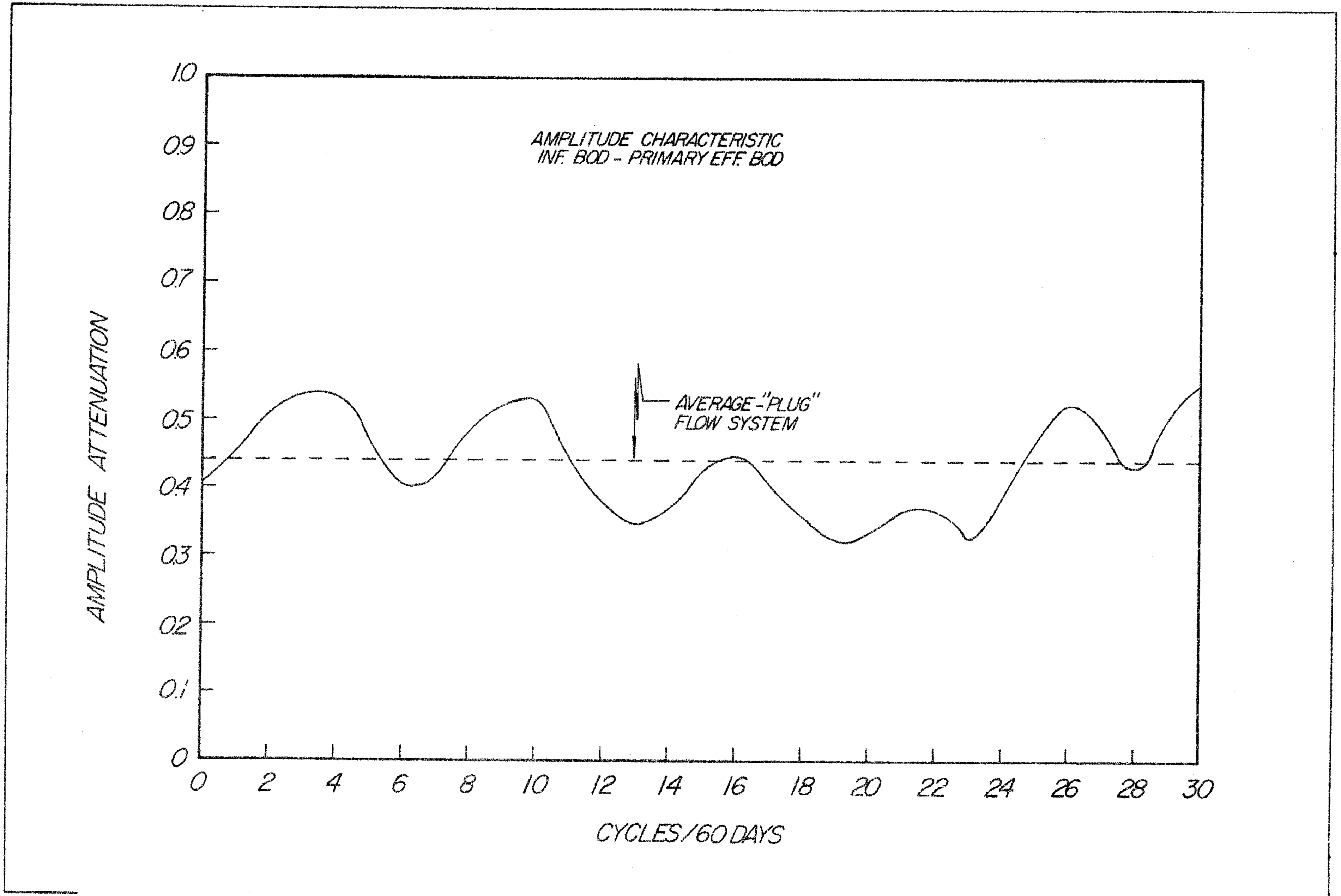



Fig. 14. Amplitude attenuation, Plant #5

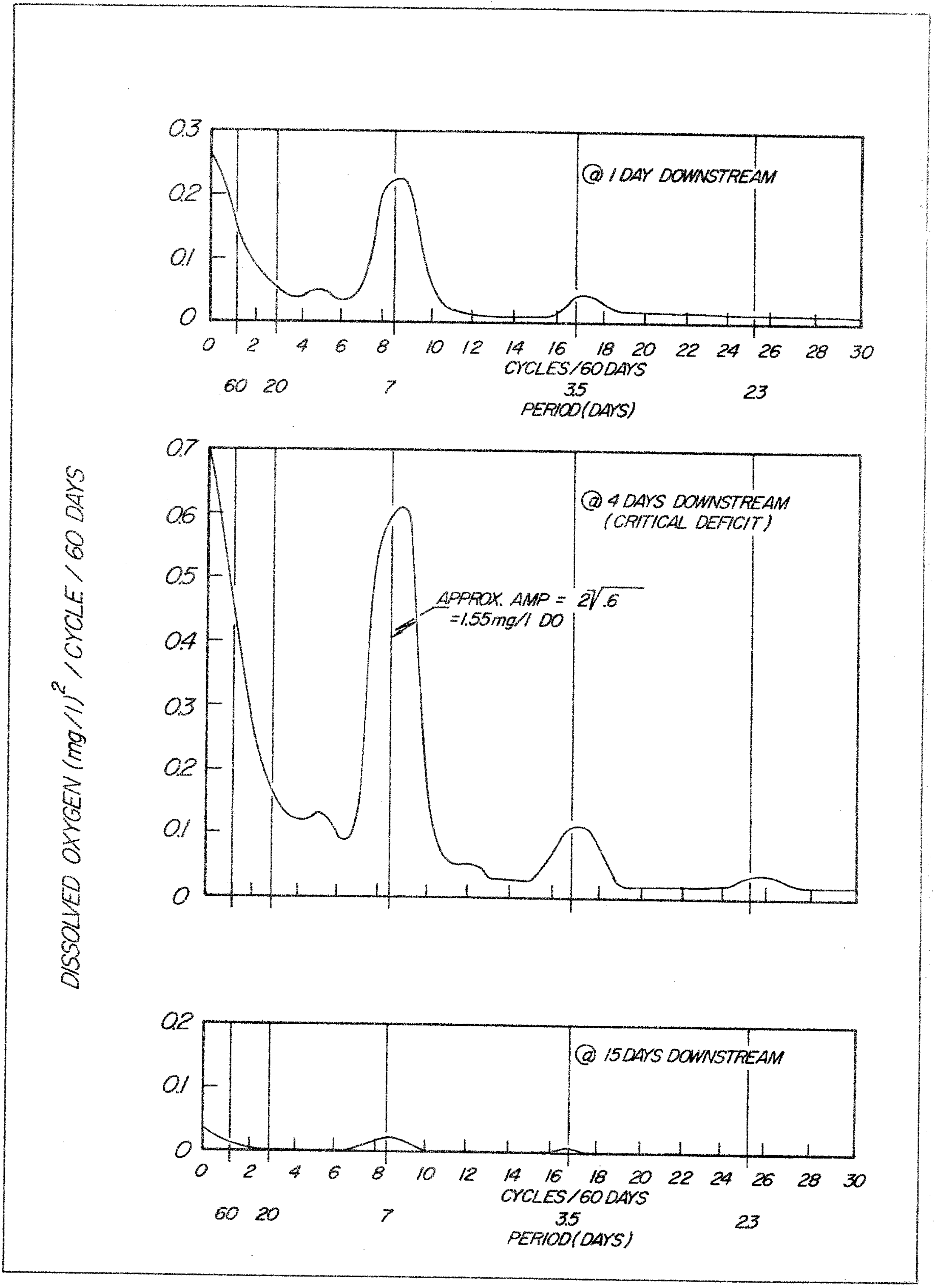



Fig. 15. Computed DO spectrums for hypothetical stream, input of Fig. 9