USE OF SPECTRAL ANALYSIS AND NONLINEAR PARAMETER ESTIMATION TECHNIQUES IN ANALYZING WATER QUALITY INFORMATION SYSTEMS

by 557

JAIPRAKASH SHASTRY

B. Tech., Indian Institute of Technology, Bombay, 1968

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Chemical Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1971

Approved by:

L. T. Fan

Major Professor

TABLE OF CONTENTS

CHAPTER I INTRODUCTION 1

CHAPTER II LITERATURE REVIEW 5

INTRODUCTION 5

MODELS PROPOSED 6

STATISTICAL MODELING 11

CHAPTER III SPECTRAL ANALYSIS 17

INTRODUCTION 17

HARMONIC ANALYSIS 20

SPECTRUM ANALYSIS 23

1. Objectives 24
2. Characteristics 27
3. Autocorrelation 30
4. Autospectral Analysis 34
5. Smoothing Spectral Estimates 39
6. Crosscorrelation and Crossspectrum Analysis 46
7. Transfer Function and Coherence 50

CONCLUSION 54

CHAPTER IV APPLICATION TO WATER POLLUTION STUDIES 55

INTRODUCTION 55

DESIGN OF SPECTRAL ANALYSIS 58

COMPUTATIONAL PROCEDURE 60

ANALYSIS AND INTERPRETATION OF THE DATA 60

POSITION CORRELATION ANALYSIS FOR THE OHIO RIVER 62

Analysis of Flow Records 62

1. Autocovariance 64

2. Power Spectral Estimates 68

3. Crosscovariance 76

4. Transfer Function and Coherence 81

Analysis of Temperature Records 89

Analysis of Dissolved Oxygen Records 102

VARIABLE CORRELATION OF OHIO RIVER DATA 114

Temperature - BOD 114

Temperature - Dissolved Oxygen 115

Dissolved Oxygen - BOD 126

ANALYSIS FOR OTHER STREAMS 129

Detroit River 129

Coasa River 137

Missouri River 143

COMPARISON OF RESULTS 154

CONCLUDING REMARKS 157

CHAPTER V MODELING FOR WATER QUALITY 161

INTRODUCTION 161

RESULTS FROM SPECTRAL ANALYSIS 162

DEVELOPMENT OF MODELS 166

CHAPTER VI PARAMETER ESTIMATION 174

INTRODUCTION 174

MATHEMATICAL MODEL 174

CRITERIA FOR ESTIMATION 176

METHODS FOR PARAMETER ESTIMATION 178

BARD'S MODIFICATION 180

Choice of Direction 181

Choice of Step Size 182

RESULTS OF PARAMETER ESTIMATION 184

CHAPTER VII RECOMMENDATIONS FOR FUTURE WORK 191

NOMENCLATURE 195

ACKNOWLEGEMENTS 198

APPENDICES 199

REFERENCES 202

CHAPTER I

INTRODUCTION

The rapid expansion of population and increase in industrial development have resulted in increasingly difficult problems of water pollution control. One of the most significant problems in this area is the control of water quality in a specific body of water in order to minimize the harmful effects of pollution caused by the discharge of municipal and industrial wastes. The treated wastes discharged from municipal or industrial treatment plants contain a large variety of chemical compounds; many of these are biodegradable by aerobic microorganisms which consume oxygen. These compounds are degraded into inoffensive components by bacteria and other organisms in natural water as long as sufficient oxygen is present in the water. The source of oxygen for the stream organisms which degrade the waste materials is the dissolved oxygen in the stream. The dissolved oxygen concentration and the stream ecology are sensitively balanced, and too great a disturbance at any point may cause unpredictable and irreversible changes in the entire environment [1].

For example, if the dissolved oxygen concentration falls below a certain critical level, many of the oxygen requiring species in the stream may not survive. Microbial decomposition of the dead species further decreases the oxygen concentration until anaerobic conditions may eventually result.

The quality of any body of water can be measured in terms of the biochemical oxygen demand, the dissolved oxygen, Escherichia Coli serogroups, dissolved solids and the inorganic and organic constituents. Although many water quality standards specify minimum acceptable levels for a variety of water quality parameters, the most commonly used parameter to specify stream pollution resulting from biodegradable wastes is dissolved oxygen [2]. For the preservation of aquatic life and water reuse it is required by regulatory agencies that the dissolved oxygen concentration should not be less than 5.0 mg/l. The task of any water pollution control program, therefore, is to design a system of waste treatment plants for partially treating the waste discharges, so that the water quality criteria is not violated at any point along the stream and at the same time the cost of such a treatment system is minimized. To develop an effective water pollution control program, however, one must possess accurate knowledge and understanding of the assimilative capacity of the stream and the ability to predict the response of any body of water to specific waste inputs. This knowledge is the necessary basis for the precise determination of waste treatment requirements in any specific situation and is, therefore, the principal requirement for accurate evaluation of the costs of a pollution control program.

The purpose of developing a water pollution control program is to maximize the benefits from the available water resources [3]. This can be achieved by predetermining a desired quality

for the individual water ways and analyzing the entire system of water and pollution sources to obtain the most economical means of maintaining the desired water quality. This requires optimal design and operation of the waste treatment plants and a knowledge regarding the volume and concentration of the effluent that may be discharged into the stream. One of the important steps in this analysis is to develop techniques for predicting changes in water quality, more specifically the dissolved oxygen, due to waste discharges.

Work in this area began as early as 1925 but no precise and accurate method has been developed due to the complexity of the system. The purpose of the present investigation is to recognize the chemical, physical, biological and ecological factors that govern the water quality changes and hence, the assimilative capacity of the stream, and incorporate these factors in the mathematical formulations.

Recently, considerable effort has been expended to collect water quality data by installing automatic water quality montoring and data collection stations on a number of rivers [4]. Spectral analysis techniques are specifically designed for analyzing this type of data. This technique is used to determine the cause-effect relationships that influence the water quality. Improved mathematical models are developed based on this analysis and the values of the constants in the model are estimated by a nonlinear parameter estimation technique. The treatment is restricted to streams in the present investigation,

but the extension to estuarine systems is apparent.

CHAPTER II

LITERATURE REVIEW

INTRODUCTION

Streams have been used to carry away municipal wastes from the time people began discharging their wastes into sewers [5]. The streams have the ability to decrease their concentration of the biodegradable wastes by complex biological processes involving microorganisms and other species found in streams.

These microorganisms require oxygen for growth and the amount of oxygen required is directly related to the microbial growth rate and the number of microorganisms. The microbial growth depends on the concentration of the wastes in the stream as measured by BOD*, the temperature of the stream and other physical and biological factors. The oxygen level in the stream, therefore, is a complex function of temperature, BOD, the time of travel and other stream parameters. Because oxygen is a critical component, the quality of water is specified in terms of the dissolved oxygen and it becomes necessary to develop methods to predict and control the dissolved oxygen (DO) level in the stream. Methods have been developed to describe the change in the DO concentration in the stream resulting from a single polluter discharging at a point source [6, 7, 8].

---

* BOD is the biochemical oxygen demand which is a measure of pollution in the stream and is defined as the amount of oxygen consumed by the microorganisms in stabilizing a stream sample.

MODELS PROPOSED

A stream may be defined as a body of flowing water in which the velocity is the significant component of the flux and in which the longitudinal dispersion in the direction of the flow may be neglected.

The measure of dispersion used in these situations is the Peclet number which is defined as

$$N_{Pe} = \frac{uL_R}{D} \tag{1}$$

in which

u = average velocity of the stream

L_R = length of the reach under consideration

D = axial diffusion coefficient.

For streams, where u and L_R are very high, the Peclet number is usually very high. This high value of Peclet number corresponds to plug flow behavior [9] and hence the effect of dispersion can be neglected. Under the ideal plug flow assumption the most general form of the equation that describes the temporal and spatial distribution of conservative and nonconservative substances (water quality model) in a one dimensional stream may be written as [10]

$$\frac{\partial c}{\partial t} = \frac{1}{A(x,t)} \frac{\partial c}{\partial x} (Q(x,t),\ c) \pm S(c,x,t) \tag{2}$$

in which

c = concentration of the substance

A = cross sectional area which is a function of x and t

Q = stream flow

S = sources and sinks of the substance

x = distance along the stream

t = time of travel

Various forms of the above equations have been proposed as models to describe the change in dissolved oxygen concentration in a stream due to waste discharges. Most of these try to relate the change in dissolved oxygen to the change in BOD and are based on material balances of resources and demands. Some of these models are based on the following assumptions:

(i) The stream flow is steady and uniform.

(ii) The process for the stretch as a whole is a steady state process, the conditions at every crossection being unchanged with time.

(iii) The BOD and DO are uniformly distributed over each crossection, thus permitting the equations to be written in one-dimensional form.

The Streeter-Phelps model [11], which attempts to describe the system under steady state conditions, considers only two independent mechanisms affecting the DO. They are

a) the decrease in oxygen concentration due to bacterial oxidation which is proportional to the BOD present and

b) the increase in dissolved oxygen content due to reaeration from the atmosphere. This model is written as a set of differential equations, one each for BOD and DO.

$$\frac{dL}{dt} = -k_1 L$$

$$\frac{dC}{dt} = -k_1 L + k_2(C_s - C) \qquad (3)$$

The second equation may also be written in terms of the DO deficit, $D = C_s - C$ as

$$\frac{dD}{dt} = k_1 L - k_2 D \qquad (3a)$$

In these equations,

L = BOD level in the stream in mg/l

C = DO level in the stream in mg/l

k_1 = bacterial action rate constant in day^{-1}

k_2 = reaeration constant in day^{-1}

C_s = saturation DO level in mg/l

t = time of reaction in days

For a point source the initial conditions are

$$L = L_o \text{ and } D = D_o \text{ at } t = 0 \qquad (4)$$

Equation 3 can be solved analytically with the boundary conditions given by Equation 4 , the solution is

$$L = L_0 e^{-k_1 t}$$

$$D = \frac{k_1 L_o}{k_2 - k_1}(e^{-k_1 t} - e^{-k_2 t}) + D_o e^{-k_2 t} \qquad (5)$$

Criticisms expressed against this model for its obvious limitations led other investigators to modify this model or propose different approaches to predict the dissolved oxygen concentration in a flowing stream. Notable among these are the contributions of O'Connor [10,12], Dobbins [13], Thayer and Krutchkoff [14], Thomann [15], and Pyatt [16]. Dobbins [13] included eight additional phenomenological processes that may take place in a particular stretch of the stream; these are:

1. The removal of BOD by sedimentation or absorption.
2. The addition of BOD along the stretch by the scour of bottom deposits or by the diffusion of partially decomposed organic products from the benthal layer into the water above.
3. The addition of BOD along the stretch by the local runoff.
4. The removal of oxygen from the water by diffusion into the benthal layer to satisfy the oxygen demand in the aerobic zone of this layer.
5. The removal of oxygen from the water by purging action of gases rising from the benthal layer.

6. The addition of oxygen by the photosynthetic action of plankton and fixed plants.
7. The removal of oxygen by the respiration of plankton and fixed plants.
8. The continuous redistribution of both BOD and oxygen by the effect of longitudinal dispersion.

Dobbins [13] proceeded to show that the dispersion effect may be neglected for many streams. Incorporating these effects in the Streeter-Phelps equations, the model proposed by Dobbins is obtained as

$$\frac{dL}{dt} = -k_1 L - k_3 L + L_a$$

$$\frac{dC}{dt} = -k_1 L + k_2(C_s - C) - D_B \qquad (6)$$

or

$$\frac{dD}{dt} = k_1 L - k_2 D + D_B$$

in which

k_3 = rate constant for BOD sedimentation, assumed to be a first order rate process, in day^{-1}

L_a = BOD addition due to local runoff, assumed to be constant all along the stream, in mg/l

D_B = DO depletion due to benthal demand, again assumed to be constant, in mg/l

The model of Equation 6 can be solved analytically, with the

boundary conditions of Equation 4. The solution is

$$L = L_o \exp -(k_1+k_3)t + \frac{L_a}{k_1+k_3}[1 - \exp -(-k_1+k_3)t]$$

$$D = \frac{k_1}{k_2-(k_1+k_3)} [L_o - \frac{L_a}{k_1+k_3}][\exp -(k_1+k_3)t - \exp(-k_2t)] \quad (7)$$

$$+ D_o \exp(-k_2t) + [\frac{D_B}{k_2} + \frac{k_1 L_a}{k_2(k_1+k_3)}] \left\{ 1 - \exp(-k_2t) \right\}$$

The unsteady state BOD and DO equations were solved by Dobbins [8] for a a point input. In these calculations Dobbins included the diffusion term and the second order partial differential equation, thus obtained, was solved by a two-step numerical method. It seems at this stage that this two-step procedure can be avoided if proper boundary conditions are chosen. In the next section some stochastic approaches used for water quality modeling are presented.

STATISTICAL MODELING

Many investigators found that BOD and DO values predicted by deterministic formulations often differ from the actual observed values due to unpredictable (random) changes in the stream parameters. A probabilistic approach often becomes necessary because one has to consider random variations in rate constants, in stream flow conditions and in waste inputs. Thayer and Krutchkoff[14] analyzed Dobbins' model but used a probability approach in which the mechanisms affecting BOD and

DO concentrations were modified so as to be random in nature. By constructing a series of probability statements about BOD and DO, the joint probability function for BOD and DO distributions was developed. The differential-difference equation thus obtained was solved by a moment generating functions technique to derive expressions for the mean and variance of the BOD and DO distributions as functions of time. They examined three different initial conditions and derived expressions for the mean and variance for each of these. In this approach Thayer and Krutchkoff had to discretize the BOD and oxygen values.

Kothandaraman and Ewing [17] used a deterministic approach but accounted for the random variations in the reaction velocity constants by using Monte Carlo simulation. They accounted for the diurnal variation in the oxygen production due to photosynthesis by including in their model a periodic function expanded in terms of sines and cosines. The results obtained by using this statistical sampling technique are in good agreement with the experimental values of DO for the Ohio river system.

Thomann [15,18] used spectral analysis techniques with a systems analysis approach and obtained expressions for amplitude and phase transfer functions for the system shown in Fig 2.1. These functions were, then, used to calculate the response of the stream to deterministic and stochastic waste load inputs. In his paper Thomann has stressed the importance of the system analysis approach to this problem. The stream

is envisaged as a system represented by two blocks, the input to the system being the waste discharged and the output being the dissolved oxygen; BOD appears as an intermediate (Fig. 2.1). Relations are obtained between the power spectral densities of the input and output for the case of stochastic inputs characterized by known autocorrelation coefficients or power spectral density estimates.

In another publication Thomann and Sobel [19] assumed that the estuary can be segmented into a discrete number of sections. The DO concentration may vary from segment to segment but it is assumed to be homogeneous within each segment. The rate of change of DO concentration in the kth segment is assumed to be dependent on i) the net volume transported from the (k-1)th segment into the (k)th segment and from the (k)th to the (k+1)th segment ii) dispersion from both the (k-1)th and the (k+1)th segment into the (k)th segment and iii) sources and sinks in segment k. This model is equivalent to the n CSTR-in-series model with backflow between stages often used in chemical engineering to describe a tubular reactor system. A material balance for oxygen resources and demands gives a finite difference equation for each segment; however, these equations must be solved simultaneously.

All these models in essence describe the behavior of dissolved oxygen and BOD in the stream. In general these two water quality parameters can be represented as shown in Fig. 2.2. For a constant waste load input at $x = 0$ the biochemical oxygen

THIS BOOK CONTAINS NUMEROUS PAGES WITH DIAGRAMS THAT ARE CROOKED COMPARED TO THE REST OF THE INFORMATION ON THE PAGE.

THIS IS AS RECEIVED FROM CUSTOMER.

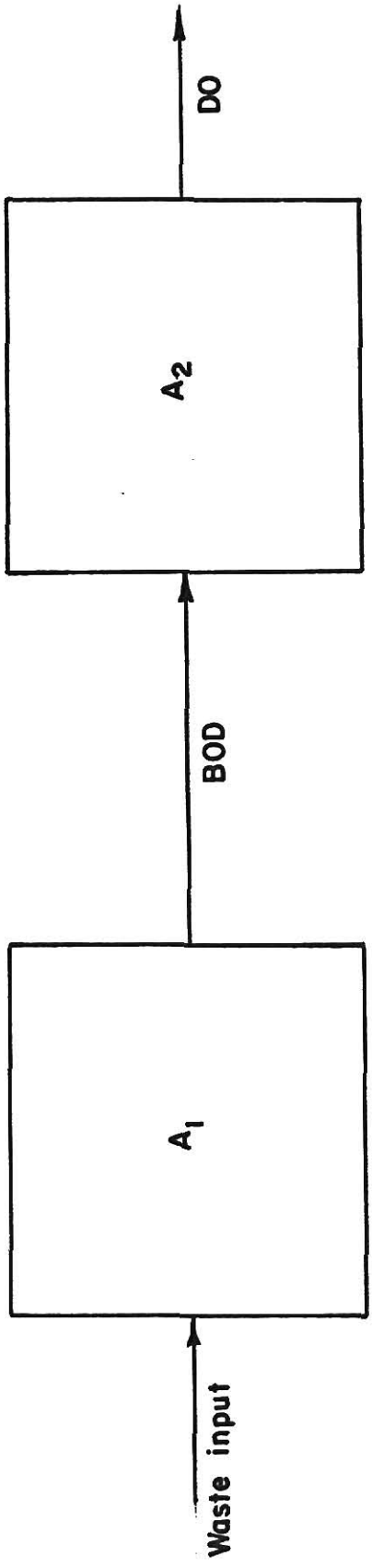


Fig2.1. Thomann's representation of the river system.[15].

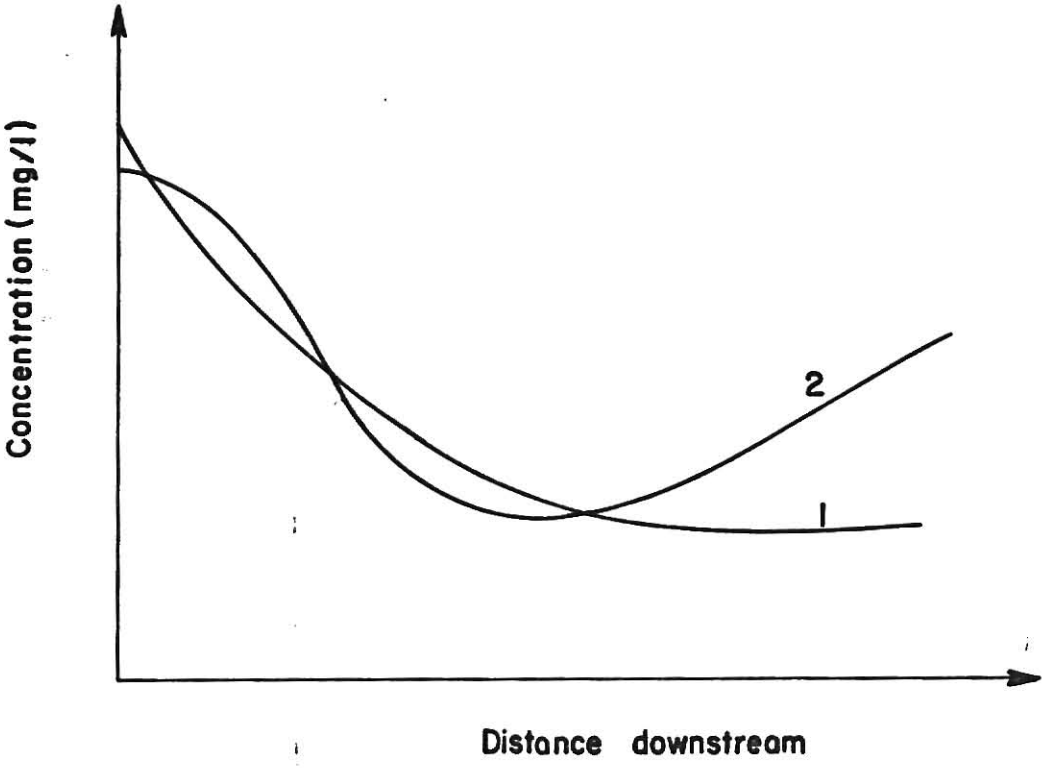


Fig.2 2. BOD (1) and DO(2) profiles in a stream[13].

demand decreases continuously with increasing distance as the microorganisms consume the wastes that are present. The dissolved oxygen concentration, however, exhibits a minima. The DO decreases initially if the biodegradable wastes consume the oxygen faster than the rate of reaeration, but after a certain stage the rate of addition of oxygen due to reaeration exceeds the rate of depletion due to consumption and oxygen concentration shows as increase. The minimum value of DO reached is the most critical one and it is this value of DO which one must take into account in designing any water pollution control program. It becomes necessary, therefore, to predict the DO values at any section along the stream. It has long been recognized that the BOD kinetic models used so far are not adequate and an effort to correlate the change in BOD and DO with the growth of microorganisms is required. An attempt has been made in this investigation to correlate all the water quality parameters by spectral analysis. Details of this data analysis technique and its application to water quality modeling are considered in later chapters.

CHAPTER III

SPECTRAL ANALYSIS

INTRODUCTION

Recently increasing attention is being paid to the use of systems analysis and simulation in the management of water quality. A water pollution control program may wish to utilize the assimilative capacity of the stream to its fullest extent. In developing such a program one needs to determine the assimilative capacity of the stream. This requires developing systematic procedures for data collection and handling which upon further analysis, can lead to problem definition and solution. The problem involves characterization of the stream system coupled with measurement of the physical, chemical, and biological parameters of interest. It has been reported that effective management of water quality is possible through the development of more sophisticated measuring techniques [20].

The wastes discharged from municipal or industrial sewage treatment plants contain a large variety of chemical compounds; usually, the most significant among these is the biodegradable and oxygen consuming portion of the wastes. The strength of these wastes is measured in terms of its biochemical oxygen demand. When this material is introduced into a body of water, if undergoes biochemical oxidation caused by microorganisms which utilize the organic matter and oxygen for energy and growth. Besides this biological phenomenon, there are other physical

and chemical processes taking place in the stream which influence the water quality. The quality of water at any section on the stream, therefore, depends on the temperature, stream flow, distance from the discharge, the count and type of plankton population in the stream and the nature of the bottom deposits. The design of the data measurement and collection system consists of developing methods to analyze each of the above parameters.

In developing any water pollution control program it is necessary to obtain water quality data. During the initial phase of this investigation water quality data were obtained from several federal, state and private agencies monitoring water quality stations. In this investigation spectral analysis techniques were used to analyze the water quality data.

Spectral analysis techniques are specifically designed for analyzing this type of data. These techniques are becoming popular due to the availability of automatically monitored data and high speed computing machines. A time series is a record of repeated observations made at a particular location; it is a momentary addition of everything that is happening to the particular parameter. Spectral analysis of time series aims at a more fundamental understanding of the underlying phenomena; it identifies the underlying generating processes which are then used for model building or prediction purposes.

Wastler [21], Gunnerson [22] and Thomann [23] have initiated the use of this technique in the area of water pollution studies.

Wastler [21] in his report described in detail a method which calculates the individual power spectra of a given time series. He then proceeded to illustrate this by computing the individual power spectra for records of BOD and DO from the field survey of the Potomac Estuary. These two spectra were examined individually for existence of 'long' and 'short' period effects. The physical interpretation of each of the harmonics present at different stations was attempted in light of the advective and diffusive processes, the diurnal variations of the waste discharge and the photosynthetic activity.

Gunnerson [22] approached the problem from the standpoint of optimizing the sampling time interval for water quality measurement. He concluded from the spectral analysis of estuarine water quality records that the sampling time interval must be shortened near the dominant source of pollution. With increasing distance downstream, mixing and stabilization processes result in increasing homogeneity and, therefore, the sampling interval may be increased. Each enviroment, according to him, has peculiar data collection requirements.

Thomann [23] analyzed the DO and temperature data taken from the Delaware estuary [24]. On the basis of this analysis he grouped the harmonics present in the record into 'low frequency harmonics' and 'high frequency harmonics'. The low frequency harmonics were attributed to the non-linear interactions among the variation of the saturation value of DO with temperature, the variability of photosynthesis with temperature

and sunlight, variable estuarine flow, and the nonlinear microbial growth process. Diurnal variations in temperature or in DO due to solar heating and photosynthesis were included in the high frequency harmonics.

The work done so far pertains to calculation and interpretation of individual power spectra for dissolved oxygen and temperature. In this chapter these techniques are reviewed and in addition, the technique of cross-spectral analysis, which is used in this investigation, is introduced.

Spectral analysis of any time series may be divided into two steps i) harmonic analysis and ii) variance (power) spectrum analysis. These techniques are discussed next.

HARMONIC ANALYSIS

This analysis is done prior to spectrum analysis to identify certain periodicities in the record [25]. The harmonics corresponding to these periodicities are, then, removed from the data and the residual is examined by spectrum analysis. The raw data can also be analyzed by spectral analysis and the results compared.

Let X(t) represent a time series with reasonably well-defined periodicitics. This can be represented by a Fourier expansion in terms of sines and cosines as

$$X(t) = \bar{X} + \sum_{k=1}^{M} \left\{ A_k \sin(kwt) + B_k \cos(kwt) \right\} + X_{RES}(t) \qquad (8)$$

in which

$w = \frac{2\pi}{T}$

T = fundamental period

t = time

k = the harmonic number

$\bar{X}$ = mean value of X(t)

M = the number of harmonics used in the series.

$X_{RES}(t)$ = residual variation not accounted for by the M harmonics

A_k, B_k = Fourier harmonic coefficients for the kth harmonic

The constants A_k and B_k are calculated by using the orthogonal properties of sine and cosine functions as

$$A_k = \frac{2}{N} \sum_{t=1}^{n} X_t \sin(kwt) \tag{9}$$

$$B_k = \frac{2}{N} \sum_{t=1}^{n} X_t \cos(kwt) \tag{10}$$

in which

X_t = discrete data at time t

N = total number of data points

After A_k and B_k have been determined from the observed data, the amplitude and phase angles for the kth harmonic may be calculated as

$$C_k = \sqrt{A_k^2 + B_k^2} \tag{11}$$

$$\theta_k = \tan^{-1} \frac{A_k}{B_k} \tag{12}$$

The variance corresponding to the kth harmonic can be computed from the equation

$$\begin{aligned} \sigma_k^2 &= C_k^2/2 \quad , \qquad k < \frac{N}{2} \\ \sigma_k^2 &= C_k^2 \quad , \qquad k = \frac{N}{2} \end{aligned} \tag{13}$$

This type of analysis is useful in physically interpreting the results and generating hypothetical or missing data. The spectral estimate calculated using the raw data shoots to a very high value for zero frequency. This makes it difficult to isolate periods with low frequency. If these low frequency harmonics are 'removed' from the data additional events with low frequencies can be identified. A computer subroutine (IBM) called RHARM is available [26] for the computation of the Fourier coefficients. This subroutine uses the fast Fourier transform method developed by Cooley and Tukey [27]. This program is used in the present work for computing the Fourier coefficients.

The next step in analyzing the time series by means of spectral analysis is to carry out a variance spectrum analysis of the residuals, $X_{RES}(t)$. This spectrum analysis can also be

applied to the raw data. The next section discusses the theoretical development necessary for this analysis.

SPECTRUM ANALYSIS

The purpose of this analysis is to describe methods to predict water quality. To control any phenomenon it is important to know the structure of the generating mechanism of the fluctuations that are present in the phenomenon. The generating mechanism is usually partitioned into the source and system [28]. The source is considered as the origin of the randomness and the system is considered to operate under this source to produce the phenomenon under consideration. The initial step in developing any control policy is process identification [29]. Process identification consists of identification of the source and identification of the system. In the identification of the source the autocorrelation functions and the power spectral density estimates are calculated for the time series representing the input. This analyzes the source from the standpoint of its frequency content. In the identification of the system the response characteristics of the system are determined. This includes the calculation of the amplitude and phase of the transfer function that relates the output to the input. In summary this is the process of system identification by spectral analysis. Once the process is identified and the underlying phenomenon understood, mathematical equations can be written down to represent the process.

For the analysis let the system under consideration be divided into two subsystems (Fig. 3.1). It is assumed that X(t), Y(t) and Z(t) can be measured as continuous functions of time. The response characteristics of systems A and B will be identified from the observations of X(t), Y(t) and Z(t). In this analysis the sets [X(t), n(t)] and [Y(t), m(t)] will be taken as second order, two dimensional stochastic processes which will be assumed stationary. Each of these sets [X(t), n(t)] and [Y(t), m(t)] are assumed to be mutually orthogonal and uncorrelated processes.

1. Objectives

The starting point in any spectral analysis is a function of time X(t) defined in the interval $0 \leq t \leq T$ relative to an arbitrary origin. Such a function X(t) is a record of repeated observations at a particular location. Each observation is a momentary summation of the effects of everthing that is happening to the particular parameter. These effects may be caused by cyclic or other phenomena affecting the system under investigation and by certain random disturbances superimposed on these phenomena.

One of the aims of spectral analysis is the determine the nature of these effects. The other objectives of this analysis are [30,31,32].

1. to determine the frequencies at which different factors cause the record to vary and to provide an estimate of the

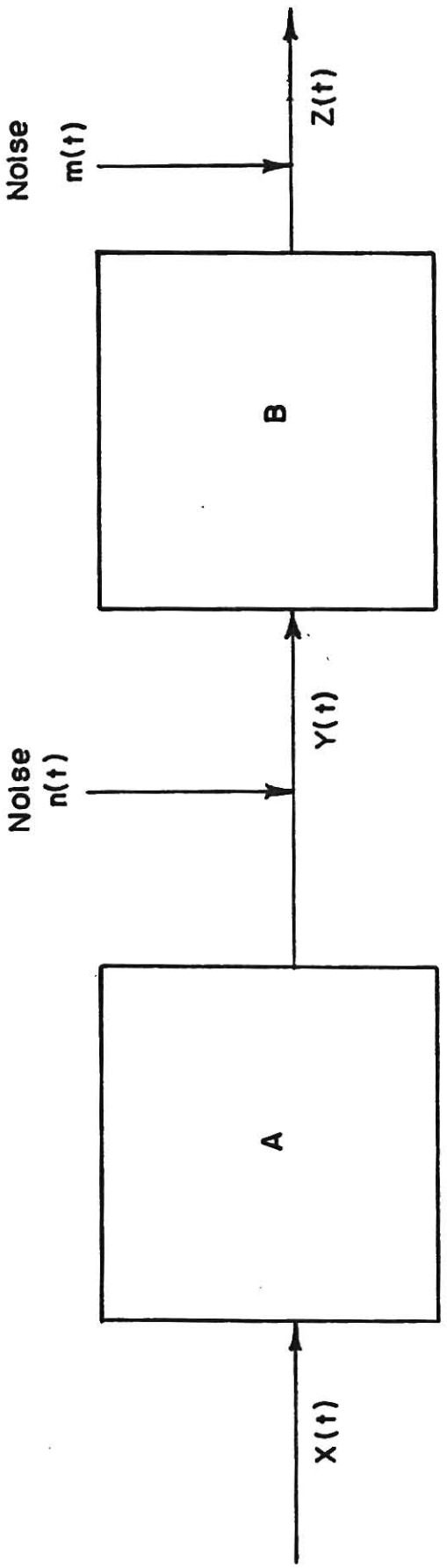


Fig. 3.1. Block diagram representation of the system under consideration.

variance that is derived from each of these factors,

2. to obtain more direct physical interpretation in the frequency domain,

3. to determine the underlying generating processes for model building and prediction purposes,

4. to determine interactions and correlations between two pairs of variables at each frequency,

5. to measure the degree of linearity of the physical processes and,

6. to estimate the amplitude and phase of the transfer function relating the input and output.

Before going into the details of spectral analysis techniques, the assumptions that need be made will be described [30]. In the spectral calculations all the random processes considered will be assumed stationary. A stochastic process is stationary if the probability distribution laws of the groups

$$X(t_1),\ X(t_2),\ \ldots,\ X(t_n)$$

and

$$X(t_1 + \tau),\ X(t_2 + \tau),\ \ldots,\ X(t_n + \tau)$$

are identical.

For a stationary process, therefore, the mean and variance are constants and the correlation between two points in the time domain is a function of time interval and not the individual time

values [33]. Symbolically,

$$\text{mean } u(t) = \text{constant} \tag{14}$$

$$\text{variance } \sigma^2(t) = \text{constant} \tag{15}$$

$$\text{cross-covariance } u_{12}(t_1, t_2) = u_{12}(t_1 - t_2) \tag{16}$$

In the next section the characteristics of the spectral analysis techniques are described. These characteristics can guide the design of spectrum calculations and the interpretation of the results which are derived.

2. Characteristics [29, 33, 34]

To understand the basic properties of time series and the variability and characteristics of its periodic and irregular oscillations, one needs to recognize the following characteristics of spectral analysis.

1. Translation of the time series on the time axis does not alter the statistical properties of a stationary process.
2. The statistical manipulation of the time series data by spectral analysis results in the computation of those parts of the variance that recur at constant time intervals as well as the part that is nonrecurring in nature.
3. The spectral analysis resolves the variance of a time series into its components. This variance is interpreted as a statistic descriptive of both the random and nonrandom character of a time series.

4. If the variance is due to a linear trend, all the variance is concentrated in the zero frequency spectral estimate. This spectral estimate includes

(i) any truly random fluctuations in the record

(ii) any linear trends in the record and

(iii) any periodic components in the record that are of so low a frequency that they appear as linear trends in the record.

5. Use of a smaller sampling interval for a given record length increases the number of measurements used in the analysis and, therefore, increases the degree of freedom upon which the estimate is based.

6. If the sampling interval is not small enough to permit resolution of the shortest periods that contribute significantly to the variance of the record, the short period variance is not lost but is reported as harmonics of true period. This is known as aliasing or folding. Aliasing occurs due to high frequency records that add variance to the record but are not "seen" by the sampling interval. Variance from this type of even is folded into the record and reappears at a lower frequency. When the period of the high frequency even is known, the period of the aliased record can readily be determined. For, any cyclic event that occurs at a period less than twice the sampling interval, the true period of the event will never be seen.

Digital analysis will have to be used if values of X(t) are read off at equidistant time intervals from a continuous record.

Suppose that as a result of reading X(t) at increments of t in a time interval $0 \leq t \leq T$ ($T = n \Delta t$), one has a sequency of n discrete observations $X_1, X_2, \ldots, X_n$. It is clear [22] that reading at this sampling interval has meant a certain loss of information. In fact all the direct information is lost for frequencies above the Nyquist frequency which is defined as [32]

$$w_N = \frac{\pi}{\Delta t} \text{ (radians/sec)} \tag{17}$$

There is no way of directly estimating the amplitude of frequencies higher than the Nyquist frequency. As $2w_N$ is indistinquishable from w_N as far as the data is concerned, what one is actually able to measure at a frequency w_N is not the spectral density function corresponding to the frequency w_N but the latter confounded with all frequencies which are indistinguishable from w_N.

7. The precision of each estimate is a function of the total number of samples and the number of lags used in the computation.
8. Occurrence of a negative spectral estimate is due to the use of a finite record.

A complete detailed discussion of spectral analysis is beyond the scope of this thesis. For a detailed description the readers are referred to references 32, 33, 34 and 35. In the next section the quantities that need to be calculated in gathering the spectral information will be described along with

the computational procedure for calculating these quantities. In the analysis that follows, X(t) is taken as a time series with finite second moments which may be either continuous or discrete, the process is assumed stationary and ergodic, and the mean of the time series is taken as zero.

3. Autocorrelation [32, 35, 36, 37]

Autocorrelation means correlation with itself. Autocorrelation coefficients are ordinary linear correlation coefficients between a time series and the same time series an interval of time later. This difference is called a lag. For a stationary time series autocorrelation depends only on the lag. In terms of expectations, autocorrelation functions are defined as

$$R_{XX}(\tau) = E[X(t) \cdot X(t + \tau)] \tag{18}$$

$$= \frac{1}{T} \int_0^T X(t)\, X(t + \tau)dt \tag{19}$$

in which

$X(t)$ = deviation from the mean value

T = total time period for which data is analyzed.

$R_{XX}(\tau)$ = autocorrelation of $X(t)$ for a lag τ.

If σ^2 is the variance of the time series, the autocorrelation coefficient can be defined as

$$\rho_{XX}(\tau) = \frac{R_{XX}(t)}{\sigma^2} \tag{20}$$

When $\tau=0$, $R_{xx}(0)$ denotes the mean square value of X(t). For a time series with zero mean, $R_{xx}(0)$ defines the variance of the series and the autocorrelation coefficient $\rho_{xx}(0)$ is unity. That is, if $E[X(t)] = 0$,

$$R_{xx}(0) = \sigma^2 \text{ and } \rho_{xx}(0) = 1 \tag{21}$$

For a nonzero τ, $R_{xx}(\tau)$ denotes the degree of correlation between the value of X at t and the value at some time later $t+\tau$.

One of the important properties of the autocorrelation function $R_{xx}(\tau)$ is that it is always less than or euqal to the variance. The autocorrelation can be regarded as a measure of the randomness of a function; if X(t) is truly random, X(t) and $X(t + \tau)$ rapidly become uncorrelated as τ increases and for large values of τ, i.e.

$$\lim_{\tau \to \infty} R_{xx}(\tau) = 0 \tag{22}$$

Autocorrelation retains all the harmonics of the given time series but discards all their phase angles. In otherwords, all the periodic functions having the same harmonic amplitudes but differing in their initial phase angles have the same autocorrelation function. Autocorrelation is an even function of τ; it is aymmetrical about $\tau=0$, i.e.

$$R_{xx}(\tau) = R_{xx}(-\tau) \tag{23}$$

For a discrete series with n data points Equation 18 can be written in terms of summation. Autocorrelation at a lag τ_r is given by

$$R_{xx}(\tau_r) = \frac{1}{n-r} \sum_{i=1}^{n-r} X_i X_{i+r}, \quad r = 0, 1, 2, \ldots, m \tag{24}$$

in which

$$m = \text{maximum number of lags} = \frac{\tau_m}{t}$$

If adjacent data points are used in these calculations the sampling interval corresponds to the increment in the lag.

$$\Delta t = \Delta \tau \tag{25}$$

Autocorrelation functions for simple well defined functions can be calculated by analytical methods. The autocorrelation of a sinusoid, for example,

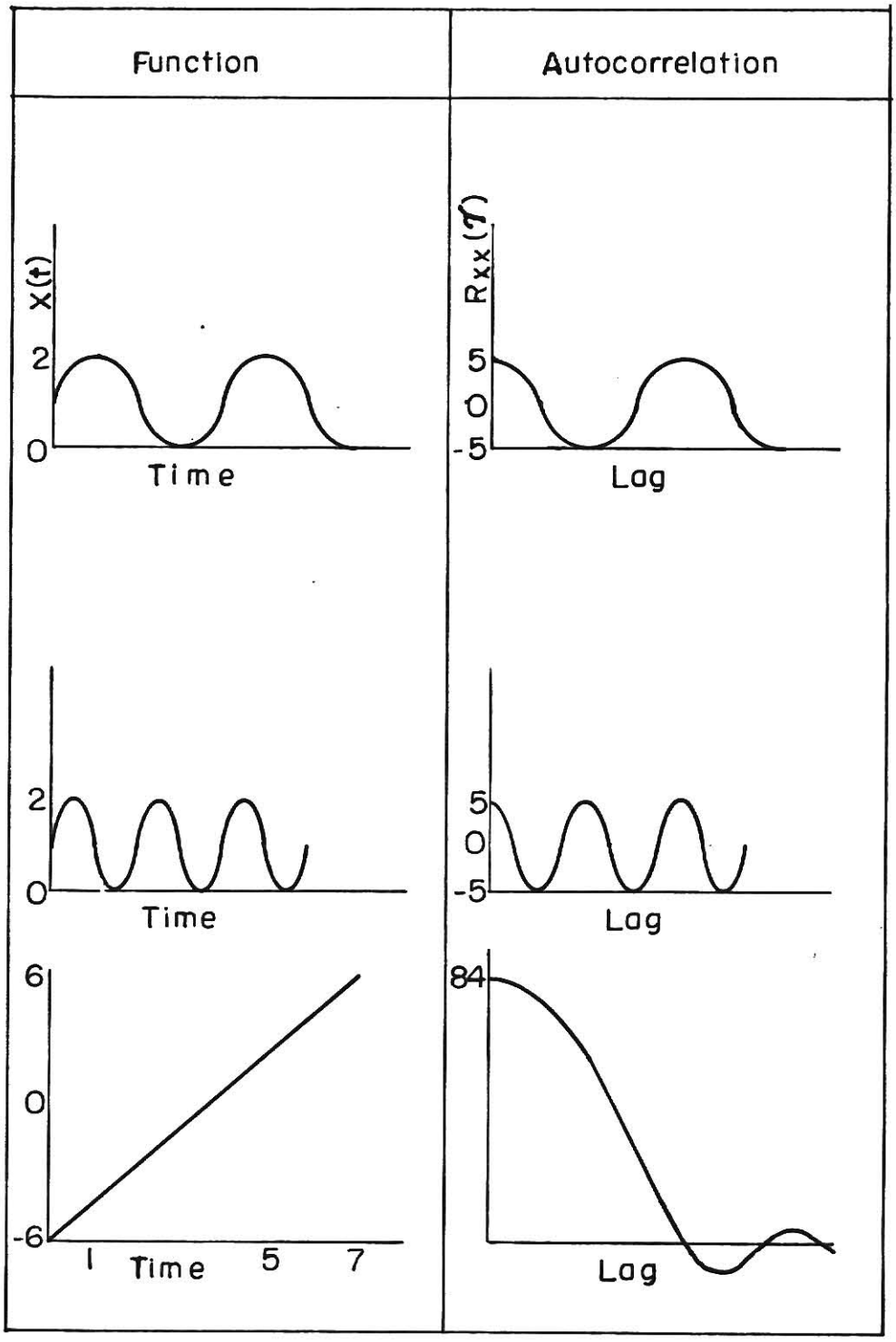
$$X(t) = A\cos(wt + \theta) \tag{26}$$

is

$$R_{xx}(\tau) = \frac{A^2}{2} \cos w\tau \tag{27}$$

These two functions as well as autocorrelations of other typical functions are shown in Fig. 3.2 [38].

Autocorrelation functions are useful in prediction and



Fig·3·2 Autocorrelation of some records [22].

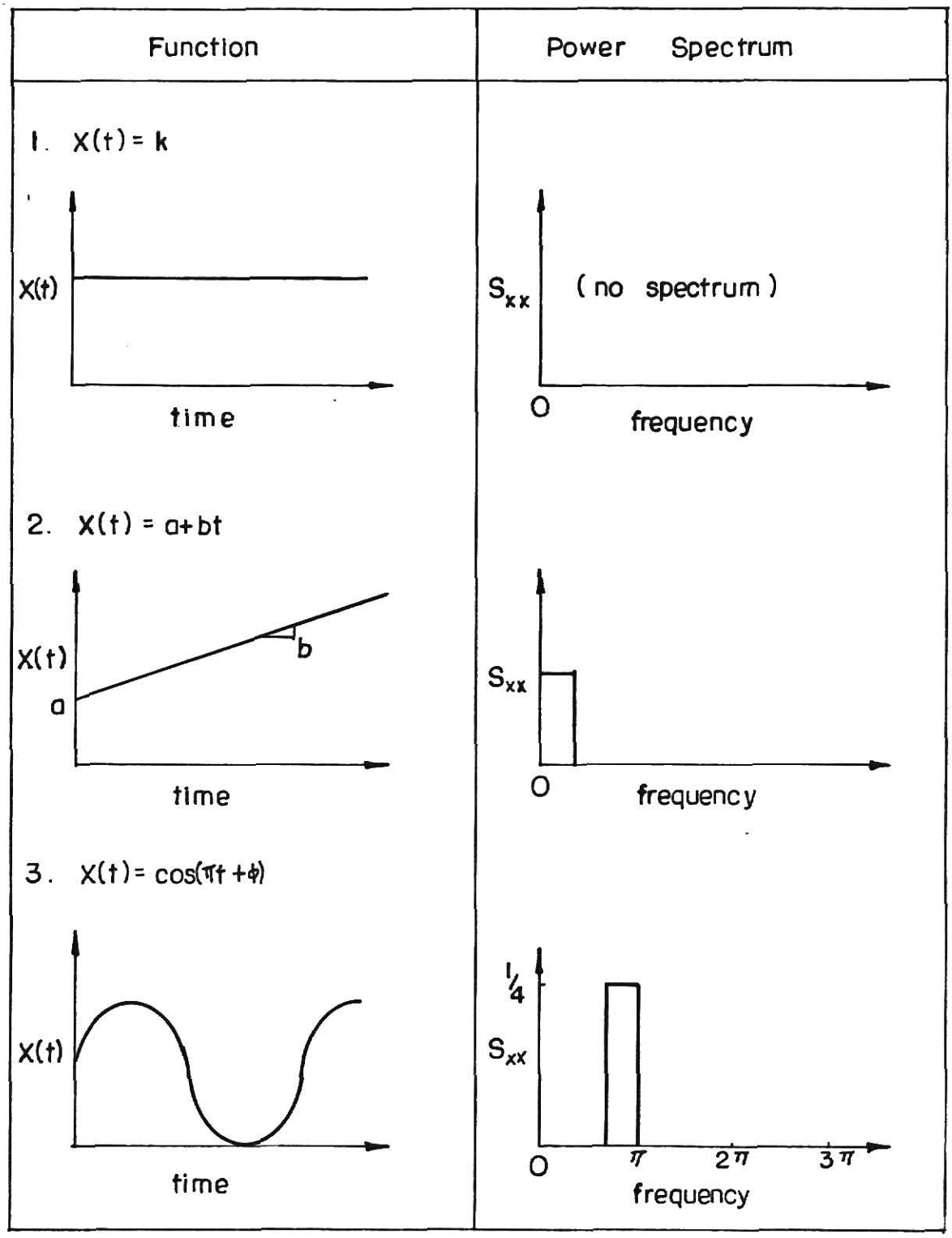
forecasting. A positive autocorrelation at a lag τ implies that if X(t) is above normal now, it will be above normal τ time intervals later. This will be useful in determining the trend in the variation of the parameter represented by the time series. This kind of information is also helpful in forecasting the temperature or DO values for a particular section of a stream. A negative autocorrelation indicates that the function will go below normal after the corresponding lag.

4. Autospectral Analysis [32, 35, 39]

To be empirically useful a stationary stochastic process must often be defined by analytic formulas other than correlation estimates of its distribution functions. Spectral analysis is an example of such an alternative. The spectrum is obtained by applying harmonic analysis to the autocorrelation of X(t). This expresses the behavior of the function in the frequency domain. Spectral analysis resolves the variance of a time series record into its components just as a prism resolves a light beam into its component colors of different intensity. Typical spectra obtained from several time dependent processes and functions are presented in Fig. 3.3 [22, 32]

The transformation from time to frequency domain is performed by the Fourier transformation, and

$$S_{XX}(w) = \int_{-\infty}^{\infty} R_{XX}(\tau) e^{-iw\tau} d\tau \qquad (28)$$



Fig·3·3. Power spectrum estimates of some records [22,32]

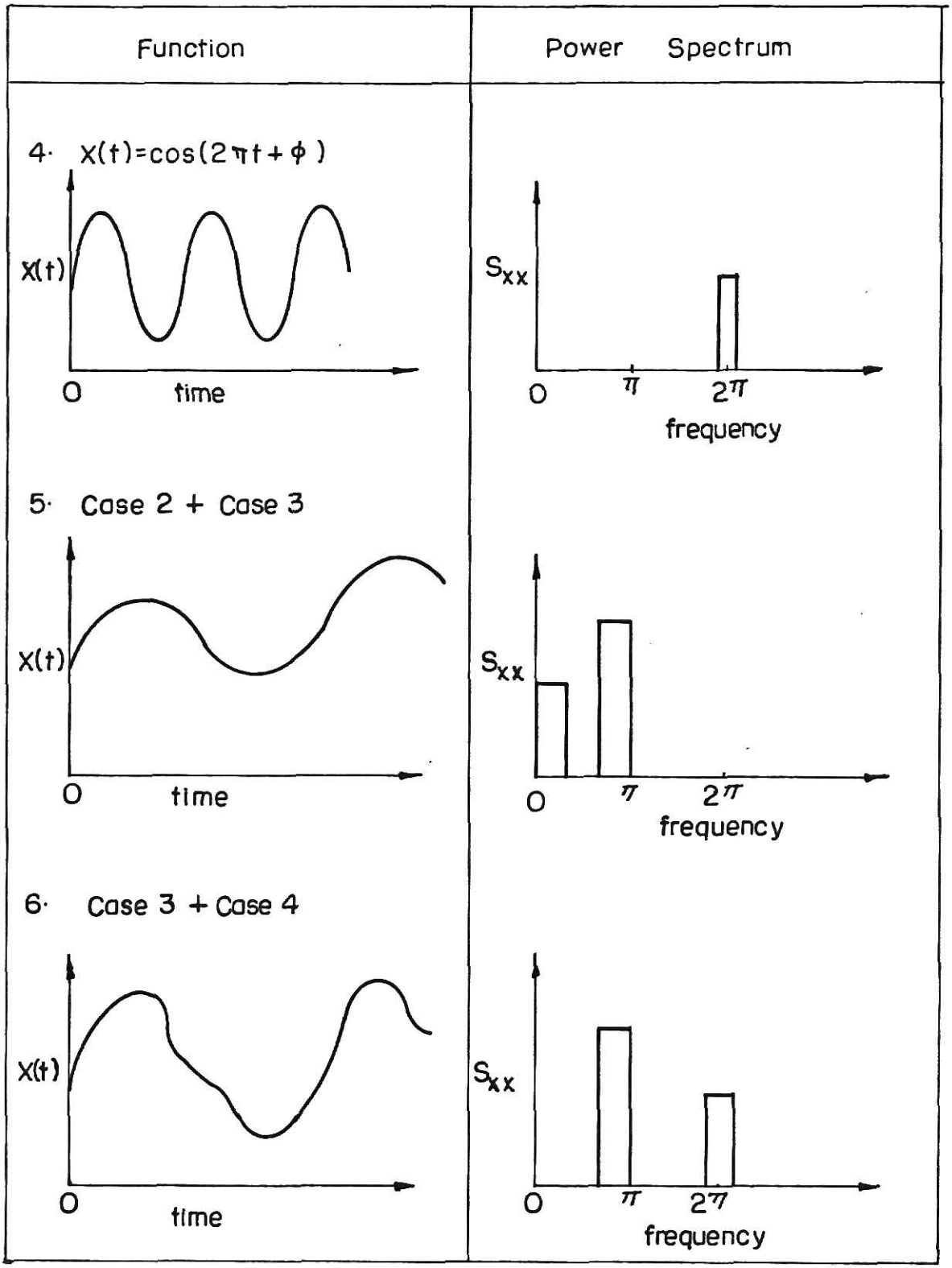

Function
Power Spectrum
4. X(t)=cos(2πt+φ)
X(t)
0
time
S_{xx}
0
π
2π
frequency
5. Case 2 + Case 3
X(t)
0
time
S_{xx}
0
π
2π
frequency
6. Case 3 + Case 4
X(t)
0
time
S_{xx}
0
π
2π
frequency

Fig 3.3. (Cont'd)

in which

$S_{xx}(w)$ = power spectral estimate of the variable X(t) at frequency w.

$R_{xx}(\tau)$ = autocorrelation at lag τ

A necessary condition for $S_{xx}(w)$ to exist is

$$\int_{-\infty}^{\infty} X(t)\ dt = k \text{ (a finite quantity)} \tag{29}$$

If $S_{xx}(w)$ exists, Equation 29 is satisfied and the inverse Fourier transform is given by

$$R_{xx}(\tau) = \frac{1}{\pi} \int_{-\infty}^{\infty} S_{xx}(w) e^{iw\tau} dw \tag{30}$$

If X(t) is assumed to be a stationary random function known for all time, the integral in Equation 29 will be infinite and the inverse will not exist. One may consider a large finite segment of the function defined by

$$X_T(t) = X(t), \quad -\frac{T}{2} \leq t \leq \frac{T}{2}$$

$$= 0, \quad t > \frac{T}{2} \tag{31}$$

where T may be chosen to be a large but finite value. The autocorrelation function for this variable is

$$R_{xx}(\tau) = \frac{1}{T} \int_{-T/2}^{T/2} X_T(t)\ X_T(t + \tau) dt \tag{32}$$

As $R_{xx}(\tau)$ contains no contribution for $t > \frac{T}{2}$, Equation 33 can be rewritten as

$$R_{xx}(\tau) = \frac{1}{T} \int_{-\infty}^{\infty} X_T(t)\, X_T(t + \tau)dt \qquad (33)$$

The power spectrum i the Fourier transform of the auto-correlation function Expanding the exponential term in Equation 28 in terms of trigonometric quantities, this equation becomes

$$S_{xx}(w) = \int_{-\infty}^{\infty} R_{xx}(\tau) \cos w\tau d\tau - i \int_{-\infty}^{\infty} R_{xx}(\tau) \sin w\tau d\tau \qquad (34)$$

As $R_{xx}(\tau)$ is an even function (Equation 23) $R_{xx}(\tau)$ sin $w\tau$ is odd and the second integral (in the right hand side) of Equation 34 vanishes over the symmetrical limits. The spectral function, therefore, is given as

$$S_{xx}(w) = \int_{-\infty}^{\infty} R_{xx}(\tau) \cos w\tau d\tau \qquad (35)$$

and the Fourier inverse formula gives

$$R_{xx}(\tau) = \frac{1}{\pi} \int_{-\infty}^{\infty} S_{xx}(w) \cos w\tau dw \qquad (36)$$

For a band limited discrete series (a series consisting of a finite number of equispaced ordinates), the spectral function

is calculated as

$$S_{xx}(p) = \frac{k}{\rho m} R_{xx}(0) + 2 \sum_{\tau=1}^{m-1} R_{xx}(\tau) \cos \frac{\pi p \tau}{m} + R_{xx}(m) \cos \pi p \qquad (37)$$

p = lag corresponding to a frequency of w, $w = \frac{\pi p}{m}$

k = 1 for $p = 1, 2, \ldots, m-1$

= 1/2 for $p = 0$ and $p = m$

$S_{xx}(p)$ = Fourier cosine transform of the autocorrelation at lag p

τ = lag number which varies from 1 tp m-1

5. Smoothing the Spectral Estimates

The spectral estimates obtained by Equation 37 are termed 'raw' and they need to be smoothed before further analysis is carried out. In this section methods are described to accomplish this.

If the sample under consideration is of a finite size, $R_{xx}(\tau)$ needs to be truncated at a certain maximum value of τ denoted by τ_m. The value of τ_m must be such that

$$\tau_m \leq \frac{T}{2} \qquad (38)$$

The effect of this truncation can be taken into account by defining a new function which is called the lag window. The simplest lag window is the rectangular window which is defined

as follows:

$$W(\tau) = 1, \qquad \tau \leq \tau_m$$
$$W(\tau) = 0, \qquad \tau > \tau \tag{39}$$

The lag window truncates the autocorrelation function. The power spectrum obtained by using this autocorrelation is called the truncated spectrum, and it is denoted by SP_{xx} and defined as

$$SP_{xx} = \int_{-\infty}^{\infty} W(\tau)\, R_{xx}(\tau)\, e^{-iw\tau}\, d\tau \tag{40}$$

$W(\tau)$ is an even function and may be regarded as a window of variables transmission which modifies the value of $R_{xx}(\tau)$ differently for different lags. This window is analogous to an electrical filter which attenuates all frequencies except those within a certain band, $f(w = 2\pi f)$. The spectral window corresponding to the lag window $W(\tau)$ is found by taking the Fourier transform, which is

$$B(w) = \int_{-\infty}^{\infty} W(\tau) \cos w\tau d\tau \tag{41}$$

It is of importance to choose the shape of the lag window $W(\tau)$ so that its Fourier transform $B(w)$ will be concentrated near $W = 0$.

In the spectral window corresponding to the lag window

defined by Equation 38 the first minor lobe is 20% as large as the major lobe (see Fig. 3.5). This will allow transmission of the spectrum generated from the frequencies outside the band width. To obtain a sharper resolution one needs to reduce the band width of the filter (window) so that transmission is concentrated at the desired frequency. Moreover, use of Equation 38 usually results in leakage from the first minor lobe; i.e. the contributions from this lobe will be negative. This shows that the rectangular lag window is a poor choice as far as the spectral window is concerned.

Various lag windows are available. Table 3.1 presents a list of the most commonly used lag windows and the corresponding spectral windows. These lag windows are plotted in Fig. 3.4 and the corresponding spectral wondows are plotted in Fig. 3.5. To concentrate the major lobe one has to make $W(\tau)$ flat and blocky and to reduce the minor lobes one has to make $W(\tau)$ smooth and gently changing. This can be easily accomplished by using either Window 4 (whose use is called 'hanning') or Window 3 (whose use is called 'hamming'). The shape of these two windows and their ability to concentrate the spectral window near zero can be seen by examining Fig., 3.4 and 3.5. The main lobes in Windows 3 and 4 are four times as wide as the side lobes, and the normal side lobe width is $1/(2\tau_m)$. These two windows, although similar, are different in the following aspects [Fig. 3.5].

1. The highest side lobe for the "hanning" window is about

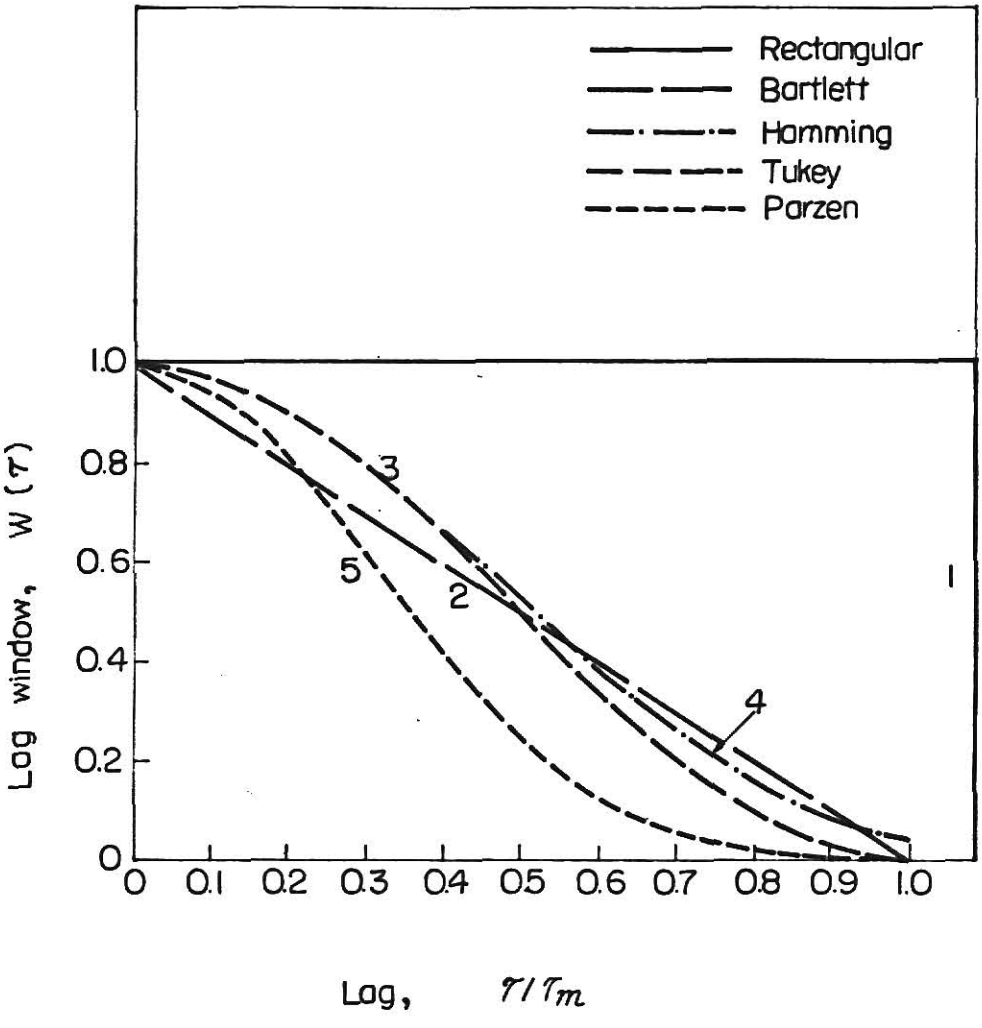


Fig. 3.4 Comparison of Rectangular(1), Bartlett(2), Tukey(3), Hamming(4) and Parzen(5) lag windows [32].

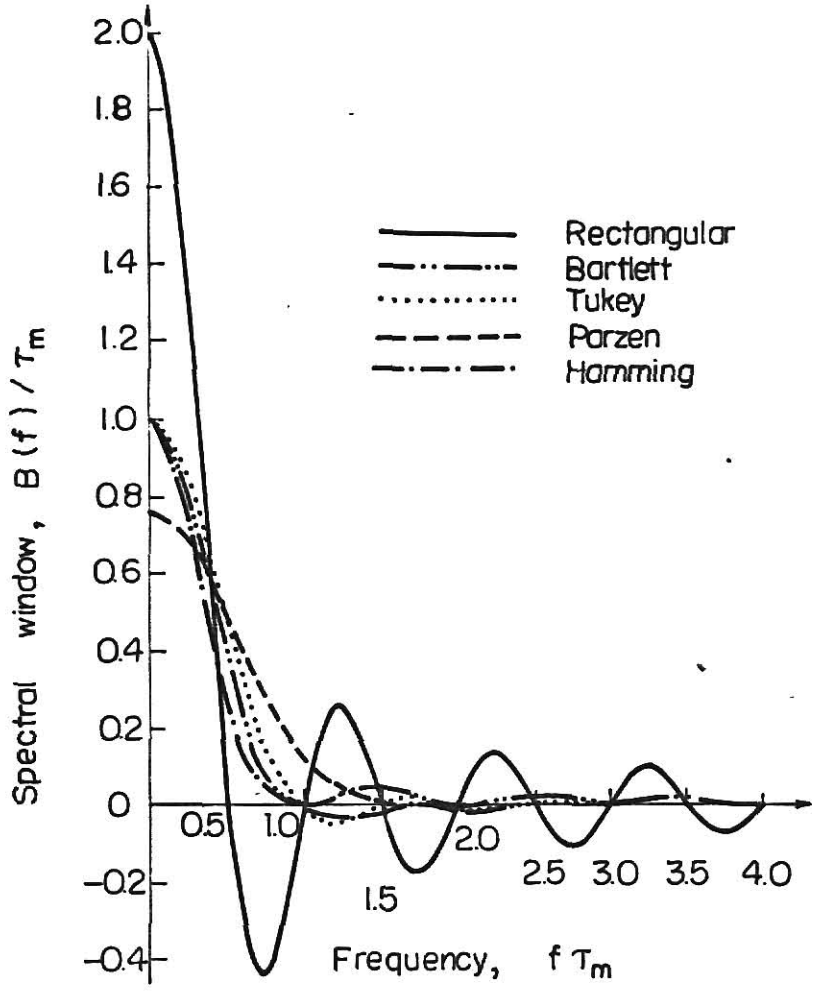


Fig. 3.5. Comparison of rectangular, Bartlett, Tukey, Hamming, and Parzen spectral windows [32, 36].

Table 3.1 Lag and Spectral Windows [32, 36]

No.	Window Type	Lag Window	Spectral Window
1	rectangular	$W(\tau) = 1, \quad \tau \leq \tau_m$ $= 0, \quad \tau > \tau_m$	$B(f) = 2\tau_m \left(\frac{\sin 2\pi d \tau_m}{2\pi 2\tau_m}\right)$
2	Bartlett	$W(\tau) = 1 - \frac{\tau}{\tau_m}, \quad \tau \leq \tau_m$ $= 0, \quad \tau > \tau_m$	$B(f) = \tau_m \left(\frac{\sin \pi f \tau_m}{\pi f \tau_m}\right)^2$
3	Tukey (Hanning)	$W(\tau) = \frac{1}{2}\left\{1 + \cos \frac{\pi\tau}{\tau_m}\right\}, \quad \tau \leq \tau_m$ $= 0, \quad \tau > \tau_m$	$B(f) = \tau_m \left(\frac{\sin 2\pi f \tau_m}{2\pi f \tau_m}\right)(1 + f)$
4	Hamming	$W(\tau) = 0.54 + 0.46 \cos \frac{\pi\tau}{\tau_m}, \quad \tau \leq \tau_m$ $= 0, \quad \tau > \tau_m$	$B(f) = 2\tau_m \left(\frac{\sin 2\pi f \tau_m}{2\pi f \tau_m}\right)$ times $(0.54 + 0.46f)$

Table 3.1 (Cont'd)

No.	Windoe Type	Lag Window	Spectral Window
5	Parzen	$W(\tau) = 1 - 6(\frac{\tau}{\tau_m})^2 + 6(\frac{\tau}{\tau_m})^3,$ $\tau \leq \frac{\tau_m}{2}$ $= 2(1 - \frac{\tau}{\tau_m})^3, \frac{\tau_m}{2} \leq \tau \leq \tau_m$ $= 0, \tau > \tau_m$	$B(f) = \frac{3}{4}\tau_m(\frac{\sin(\pi f\tau_m/2)}{\pi f\tau_m/2})^4$

3 times the height of the highest side lobe for the "hamming" window.

2. The heights of the side lobes for the "hanning" window fall off more rapidly than do those for the "hamming" window.

In this work window 4 was used for analysis. The smoothed spectral estimates for a discrete time series using "hamming" are

$$SP_{xx}(0) = 0.54\ S_{xx}(0) + 0.46\ S_{xx}(1)$$

$$SP_{xx}(p) = 0.23\ S_{xx}(p-1) + 0.54\ S_{xx}(p) + 0.23\ S_{xx}(p+1), \quad \text{(Hamming)} \tag{42}$$

$$SP_{xx}(\tau_m) = 0.54\ S_{xx}(\tau_m) + 0.46\ S_{xx}(\tau_m - 1)$$

6. Crosscorrelation and Crossspectrum Analysis [32, 39, 40]

For two stationary processes X(t) and Y(t), the theory of cross spectral analysis can be developed along the same lines as that used for autospectral analysis. This is particularly useful it one is considered as the input to the system and the other as the output.

The crosscovariance (cross correlation) of the two series X(t) and Y(t) is calculated as

$$R_{xy}(\tau) = \frac{1}{T} \int_{-\infty}^{\infty} X(t)\ Y(t + \tau)dt \tag{43}$$

For a discrete series the estimation is similar to that of autocorrelation functions. Unlike the autocorrelation, however, the crosscorrelation is not an even function, i.e.,

$$R_{xy}(\tau) \neq R_{xy}(-\tau) \tag{44}$$

One can therefore, define two crosscorrelations, one for a positive lag and one for negative lag. For a discrete record of n data points the crosscorrelation between two series X(t) and Y(t) with a positive lag is

$$R_{xy}(\tau) = \frac{1}{n-\tau} \sum_{q=1}^{n-\tau} X(q)\, Y(q + \tau) \tag{45}$$

and with a negative lag is

$$R_{xy}(-\tau) = \frac{1}{n-\tau} \sum_{q=1}^{n-\tau} X(q + \tau)\, Y(q) \tag{46}$$

The crosscovariance attempts to determine the correlation between the two series at each frequency. For a continuous time series, the crosscovariance is given by

$$R_{xy}(\tau) = \frac{1}{T} \int_{-T/2}^{T/2} X_T(t)\, Y_T(t + \tau)dt \tag{47}$$

where $X_T(t)$ and $Y_T(t)$ are defined as in Equation 31. Equations 43 and 47 are equivalent since $X_T(t)$ and $Y_T(t)$ are zero outside the range $-T/2 \leq t \leq T/2$. The cross-spectrum is defined as the

Fourier transform of the crosscovariance and written mathematically as

$$S_{xy}(w) = \int_{-\infty}^{\infty} R_{xy}(\tau)\, e^{-iw\tau}\, d\tau \qquad (48)$$

It should be noted that unlike the autospectrum the cross-spectrum is a complex quantity and, therefore, can be written as a sum of real and imaginary parts. Equations 48 can be rewritten as

$$S_{xy}(w) = C_{xy}(w) + iQ_{xy}(w) \qquad (49)$$

in which

$C_{xy}(w)$ = cospectrum of series X(t) and Y(t) at frequency w

$Q_{xy}(w)$ = quadraspectrum of series X(t) and Y(t) at frequency w

The cospectrum measures the contributions of oscillations at different frequencies to the total crosscovariance (cross-correlation) at lag zero between two time series whereas the quadrespectra measures the contributions of the different harmonics to the total crosscovariance between the series when all the harmonics of the series X(t) are delayed by a quarter period.

Assumptions similar to the ones made in the development of autospectral analysis may be made for computing the cross-

spectrum of a finite record. Under these assumptions the cospectra and quadraspectra for a discrete finite series can be calculated using the following relationships.

$$C_{xy}(p) = \frac{4t}{\pi} \sum_{\tau=0}^{m} [R_{xy}(\tau) + R_{xy}(-\tau)] \cos(p\ \tau/\tau_m) \qquad (50)$$

$$Q_{xy}(p) = \frac{4t}{\pi} \sum_{\tau=0}^{m} [R_{xy}(\tau) - R_{xy}(-\tau)] \sin(p\ \tau/\tau_m) \qquad (51)$$

in which

$$p = \frac{w\tau_m}{\pi}$$

m = number of data points in the record

$t = 1/2, \quad \tau = 0$ and τ_m

$= 1, \quad 0 < \tau < \tau_m$ (52)

These raw estimates of the cross-spectra are, then, smoothed by "hamming". The smoothed spectra are used for further analysis and interpretation.

As the cross-spectrum is a complex quantity, its amplitude can be calculated using the relation

$$AM_{xy}(w) = [SC^2_{xy}(w) + SQ^2_{xy}(w)]^{1/2} \qquad (53)$$

in which

SC_{xy} = smoothed estimate of the cospectra

SQ_{xy} = smoothed estimate of the quadraspectra

Likewise, the phase difference of the smoothed estimates of the cross-spectrum is

$$PH_{xy}(w) = ARCTAN(SQ_{xy}/SC_{xy}) \tag{54}$$

Two other important quantities the transfer function and coherence can also be calculated from the information obtained from the crossspectrum analysis. These are described in the next section.

7. Transfer Function and Coherence

These two quantities are useful in interpreting cross-spectrum results. When crosscorrelation is used, one searches for correlations for different lags in time considering the process as a whole. On the other hand when a coherence function is used one investigates if there is any correlation between the frequencies in a very small range of frequencies. The coherence function provides a direct measure of the square of the amplitude correlation at frequency w; this coherence square can be defined in terms of the smoothed cospectra, quadraspectra, and spectral estimates as follows:

$$\gamma^2_{xy}(w) = \frac{SC^2_{xy} + SQ^2_{xy}}{SP_{xx} \cdot SP_{yy}} \tag{55}$$

It can be shown that the numerical value of the coherence square function usually lies between 0 and 1.

The power spectral function is useful in computing the transfer function for the system from the measurements of input and output. If $SP_{xx}(w)$ is the power spectral function for the input and $SP_{yy}(w)$ is the power spectral function for the output (Fig. 3.6), then [41]

$$SP_{yy}(w) = SP_{xx}(w) \quad H(w)^{2} \tag{56}$$

in which H(w) is the transfer function from the input to the output and is expressed as

$$H(w) = \quad H(w) \quad e^{i\theta} \tag{57}$$

θ being the phase of the transfer function. It should be noted that only the magnitude of the transfer function appears in this relation (Equation 56) and no information regarding the phase is available. Crosscorrelation and cross spectral calculations need to be used if any information regarding the phase angle is desired. It has been shown by Solodovnikov[44] that for a linear system

$$SP_{xy}(w) = H(w)\ SP_{xx}(w) \tag{58}$$

By combining Equation 58 with Equation 59, it can be shown that for linear systems the coherence square is independent of frequency and is always equal to unity. The coherence function,

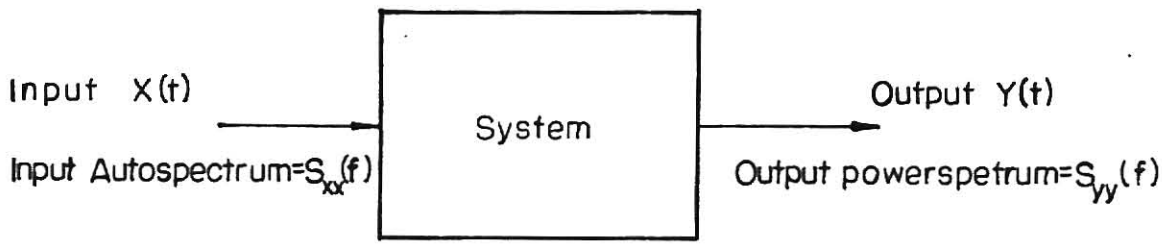


Fig 3.6. System representation for transfer function calculation.

therefore, serves as a measure of linearity of the system. Small coherence at any frequency w indicates that both C(w) and Q(w) are small. Therefore, the estimate of Q(w)/C(w) is likely to have a very large variance. Frequencies with high coherence will be more informative. A plot γ^2 vs w for $0 \leq w \leq \pi$ is called the coherence diagram.

For a system which deviates slightly from linearity, the output Y(t) can be expressed as

$$Y(t) = \int_0^\infty h(\tau)\, X(t - \tau)d\tau + \eta(t) \tag{59}$$

in which

$h(\tau)$ = unit impulse response function

$\eta(t)$ = noise

One of the problems in spectral analysis is the estimation of the noise spectra. This spectra gives an indication of periodicities in series Y(t) which are not shared by X(t). This can be computed as

$$SP_{\eta\eta}(w) = SP_{yy}(w)\left\{1 - \gamma^2_{xy}(w)\right\} \tag{60}$$

It is assumed in this analysis that the coherence between X(t) and (t) is zero for all frequencies, i.e., these two series are assumed to be orthogonal.

CONCLUSION

When all the spectral information is obtained the next step is to physically interpret the results and propose methods to predict or forecast future values of the time series. This is a black box approach in that one measures the input and output of the system and tries to find a functional relation that relates them in the frequency domain, without being concerned about what takes place in the system. These calculations yield a measure of linearity of the system and the relation between two time series considered for analysis.

CHAPTER IV

APPLICATION TO WATER POLLUTION STUDIES

INTRODUCTION

The water quality parameters of interest in this investigation are dissolved oxygen, flow rate, BOD, coliforms and temperature. In developing mathematical models for water quality, the temperature is often assumed to be constant [8, 11, 13] and the change in dissolved oxygen is often related to the corresponding change in BOD. Most of the water quality models developed so far [6, 8, 11, 12, 13] have followed this approach, but water quality prediction methods based only on dissolved oxygen and BOD have proved to be inadequate because of the inherent complexity of the system [14, 17]. Because of this complexity it may be necessary to consider temperature and the growth of microorganisms and their effect on the other water quality parameters. The nature of microbial growth and the response of these living organisms depends on many environmental factors; therefore, the response may not be identical under identical conditions of temperature, BOD and dissolved oxygen.

To include microbial growth in the model, one must understand how the bacterial count is related to other water quality parameters such as temperature, BOD and dissolved oxygen. Spectral analysis techniques can be useful for this purpose.

Five time series representing the temperature, dissolved oxygen, BOD total coliform and flow rate are used in the present

work to gather the spectral information related to the water quality parameters. This analysis enables one to forecast future values of water quality if the amount and type of the waste discharges are known. This information is also useful in determining the relationship between various water quality parameters and serves as a measure of linearity among these variables. Here the study of this spectral information is divided into two distinct topics, variable correlation and position correlation.

1. Variable Correlation: The correlation between the various variables measured at one particular location is determined. The data obtained for these calculations are for all the variables at one particular location. Some of the relationships to be investigated are represented in Fig. 4.1. The transfer function for each of these blocks is determined by calculating the input and output power spectrum. Each block is tested for linearity by calculating the coherence function between the input and output.

2. Position Correlation: In this part of the investigation correlation between the same variable measured at different locations is calculated. This information is then used to determine the transfer function relating a variable at two locations and to formulate models for the change of the water quality parameters with position. For this analysis data collected from the same stream are used. Again, coherence functions are calculated to measure the linearity of the system

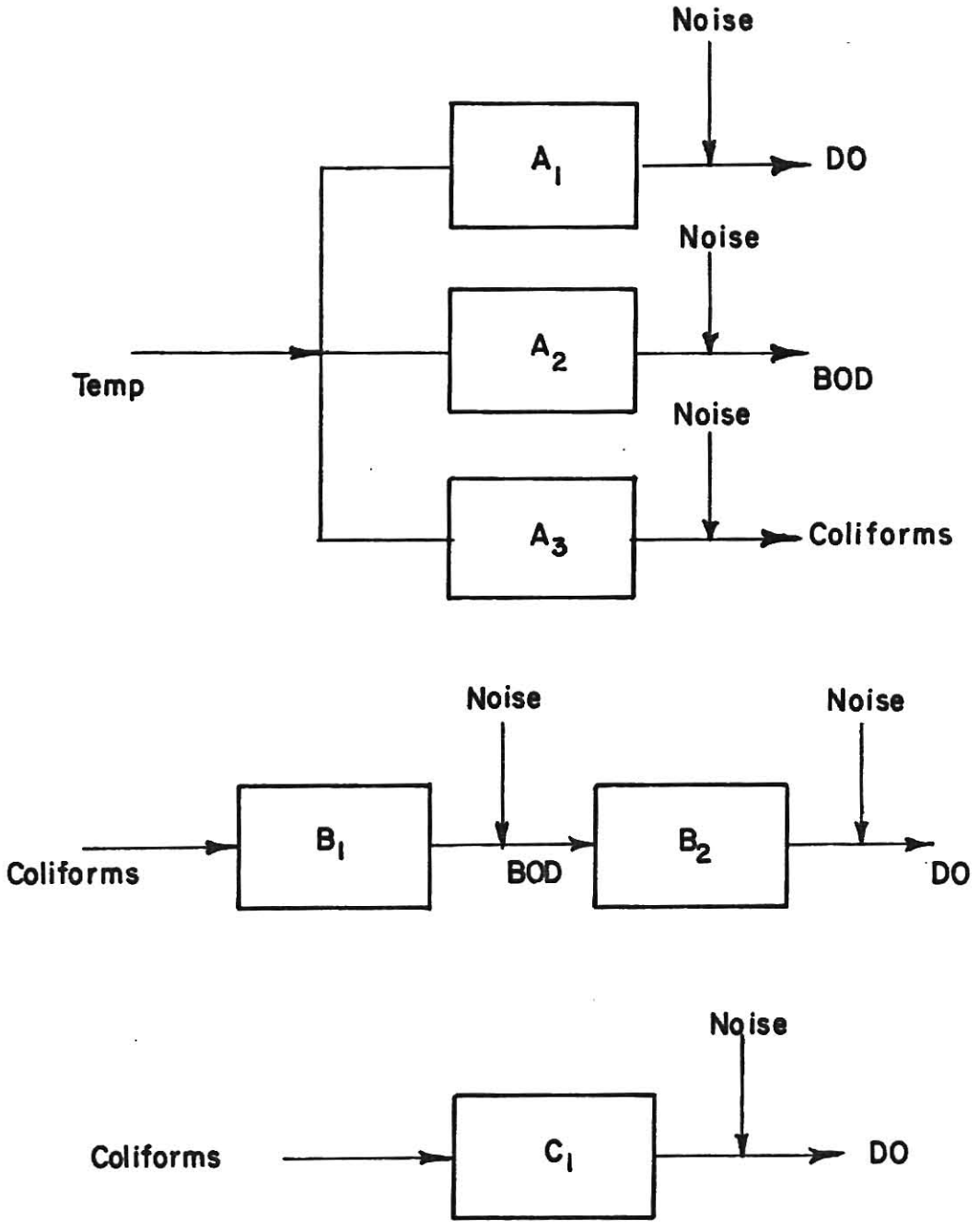


Fig. 4.1. Systems to be examined using spectral analysis.

and to formulate mathematical expressions describing the changes in water quality. The data used for this analysis were obtained from the available data on the Ohio River. Data from five stations on this river are used in this investigation.

DESIGN OF SPECTRAL ANALYSIS

Spectral analysis requires a large volume of data at a very small sampling interval. The shortest period that is necessary to resolve determines the sampling interval and the longest period necessary determines the total record length. It is assumed in this investigation that resolutions of 24 hour periods will be sufficiently small for a nontidal stream. To resolve this period a sampling interval of duration no greater than 12 hours must be used. The longest period resolved from any record depends on the number of lags used in the computation and the sampling interval. It has been shown [35] that calculation up to a lag of 10% of the total number of measurements is an optimum balance between the resolution of spectral components and the precision of individual spectral estimates. The total record length and the sampling interval employed in this investigation were limited by the amount and type of data available. As shown in Table 4.1, experimental data from several stations on several rivers were obtained for this investigation. The duration for each record is approximately one year. The information regarding these stations was obtained from Reference 45. Digitized tables were available at some of

Table 4.1 List of Stations for Which Data were Obtained

Number	Station Name	Stream or River	Data O	B	T	C	F	Duration (days)	Sampling Interval (hours)
1	South Heights	Ohio	X		X		**	366	12
2	Stratton	Ohio	X		X		**	366	12
3	Huntington	Ohio	X		X	X	**	366	12
4	Cincinnati	Ohio	X		X	X	**	366	12
5	Miami Fort	Ohio	X		X		**	366	12
6	Alabama-Georgia	Coosa	X	*	X	*	**	365	24**
7	Gibralter	Detroit	X	*	X	*		365	24**
8	St. Joseph	Missouri	X		X			303	08
9	Omaha	Missouri	X		X			303	08

0 - oxygen concentration
B - biochemical oxygen demand
T - temperature
C - coliforms
F - flow
X - indicates data is available at the sampling interval indicated
* - indicates some information is available
** - daily averages

the stations but for others (e.g. Omaha) strip charts were obtained from FWQA and digitized manually.

COMPUTATIONAL PROCEDURE

A general computer program was written during this investigation to carry out the spectral calculations on an IBM system 360/50 digital computer. This program calculates the autocorrelation functions, the estimates of power spectrum, crosscorrelation functions, the cross spectrum, transfer function and the coherence function of the time series. Three other computer programs (Table 4.2 and Appendix 4) were obtained from other sources and their performance compared with the one written during this work. It was found that BMD02T, a program written by University of California, Health Science Computing Facility [46] was very efficient and suitable for these computations. This program was used for all computations in this spectral analysis. Details of this and other programs are presented in the references [46, 47, 48] and will not be repeated here. A brief summary of each of the programs can be found in Appendix A.

ANALYSIS AND INTERPRETATION OF THE DATA

Before proceeding with the spectral calculations all the raw data were examined to detect any periodicity, recurring frequencies or missing observations. Values for missing observations were supplied using linear interpolation and the

Table 4.2 Programs From Other Sources

1) BMD02T University of California, Los Angeles [46]

2) BMDX92 University of California, Los Angeles [47]

and

3) FUTSA University of Tennessee [48]

preceeding and following data points or by examining the cycle of events graphically and determining values for the missing observations.

The complete data (no missing observations) were then subjected to harmonic analysis to determine any major periodic events. In this analysis only a few of the records were subjected to harmonic analysis, because periodicities in some of the records were similar and it was not necessary to identify them for every case. Only three of the five water quality parameters could be adequately examined because of the need for actual data at frequent intervals for harmonic analysis.

To obtain the correlation between two variables, the transfer function, crosscorelation, and coherence were evaluated directly using the complete data (no missing observations). The results of this analysis are presented in the next section along with the physical interpretation.

POSITION CORRELATION ANALYSIS FOR THE OHIO RIVER

Analysis of Flow Records

Flow records for the daily flow of the Ohio River were obtained for the five different monitoring stations listed in Table 4.1. Each was examined individually for the presence of any periodicities or recurring frequencies. These values (monthly averages) are tabulated in Table 4.3. This mean values indicate that the average flow increases as one proceeds down stream from Station 1. It was observed that Stations 4

Table 4.3 Average monthly Flow Rates (cfs) on the Ohio River (1968).

Station / Month	South Heights (Station 1)	Stratton (Station 2)	Huntington (Station 3)	Cincinnati (Station 4)
January	30,100	31,693	73,366	88,200
February	49,133	56,350	103,166	128,933
March	48,466	51,063	121,733	165,400
April	33,200	38,206	86,833	126,400
May	55,533	53,920	124,666	165,133
June	27,833	37,903	98,500	143,366
July	9,933	11,403	22,300	30,166
August	9,960	11,046	30,600	49,133
September	8,733	9,713	16,166	18,866
October	8,550	10,946	20,366	23,400
November	24,633	27,080	44,533	49,066
December	38,433	42,833	66,003	77,430
Mean	28,708	31,788	67,352	88,791

(Miami fort) and 5 (Cincinnati) are close together and, therefore, there is little change in the flow between Stations 4 and 5. Station 5 was, subsequently, omitted from the analysis of the flow behavior. The peak flow in each of the records was noted and the time of travel of this peak from one station to the other was determined.

1. Autocorrelation Function (Autocovariance): The autocorrelation values for the flow records at Station 1 (South Heights) are tabulated in Table 4.4 and are plotted in Fig. 4.2. The autocorrelation function is positive till a lag of 21 days, and then positive again between 36 and 40 days. These results indicate that large changes in stream flow usually occur over a period of time requiring many days. If the flow fluctuated daily about the mean flow, the autocorrelation would drop toward zero very rapidly. On the other hand, for a very slowly changing flow, the autocorrelation function would decrease toward zero very slowly. Since the autocorrelation function at zero lag is the variance, this value can be used as a measure of the variation that is present in the system.

A second order polynomial can be fitted to these autocorrelation values. All the autocorrelation values at this station were normalized by dividing the autocorrelation at any lag by the autocorrelation at lag zero. These values are present in Table 4.5 along with the normalized autocorrelations at other stations. These normalized autocorrelation values serve as a

Table 4.4 Autocovariance of Flow Records ($x10^{-6}$), Ohio River.

Lag (days)	Autocovariance of Station 1	Autocovariance of Station 2	Autocovariance of Station 3	Autocovariance of Station 4
0	745.618	861.248	4278.906	8511.479
1	681.294	813.883	4101.512	8293.523
2	621.915	737.117	3804.397	7851.134
3	581.428	677.119	3486.246	7306.907
4	552.545	638.358	3192.800	6736.543
5	517.490	600.226	2932.626	6167.683
6	482.313	557.944	2701.833	5624.250
7	442.830	508.777	2497.431	5119.795
8	403.111	458.941	2318.118	4659.191
9	366.481	412.252	2143.470	4227.834
10	312.849	364.574	1951.922	3805.026
11	274.032	315.761	1748.473	3366.814
12	235.530	276.463	1547.680	2927.890
13	212.051	242.303	1353.810	2471.896
14	169.493	209.767	1175.113	2032.316
15	142.246	178.343	995.956	1634.386
16	111.491	142.119	830.553	1300.916
17	87.817	108.871	680.514	1014.501
18	68.081	81.262	529.019	745.297
19	50.328	60.679	382.319	495.868
20	31.215	38.331	248.635	262.568
21	12.466	15.848	133.938	63.789
22	-4.682	-2.633	45.287	-100.104
23	-15.262	-16.278	-23.529	-252.973
24	-25.482	-30.651	-101.398	-411.021
25	-41.739	-48.890	-174.817	-562.764
26	-55.823	-64.671	-244.817	-695.290
27	-64.985	-75.252	-295.032	-807.471
28	-73.289	-83.550	-318.476	-870.407
29	-78.813	-89.757	-341.006	-872.023
30	-80.553	-93.851	-336.439	-843.010
31	-76.479	-91.538	-306.380	-801.434
32	-64.833	-83.441	-271.856	-746.596
33	-54.672	-69.423	-231.090	-662.974
34	-40.807	-52.604	-176.455	-553.054
35	-23.600	-31.450	-100.037	-415.557
36	-1.770	-8.271	4.760	-252.-54
37	19.741	18.350	128.012	-69.924
38	41.670	45.126	250.462	136.296
39	66.478	75.853	374.636	346.085
40	91.668	106.397	477.400	536.389

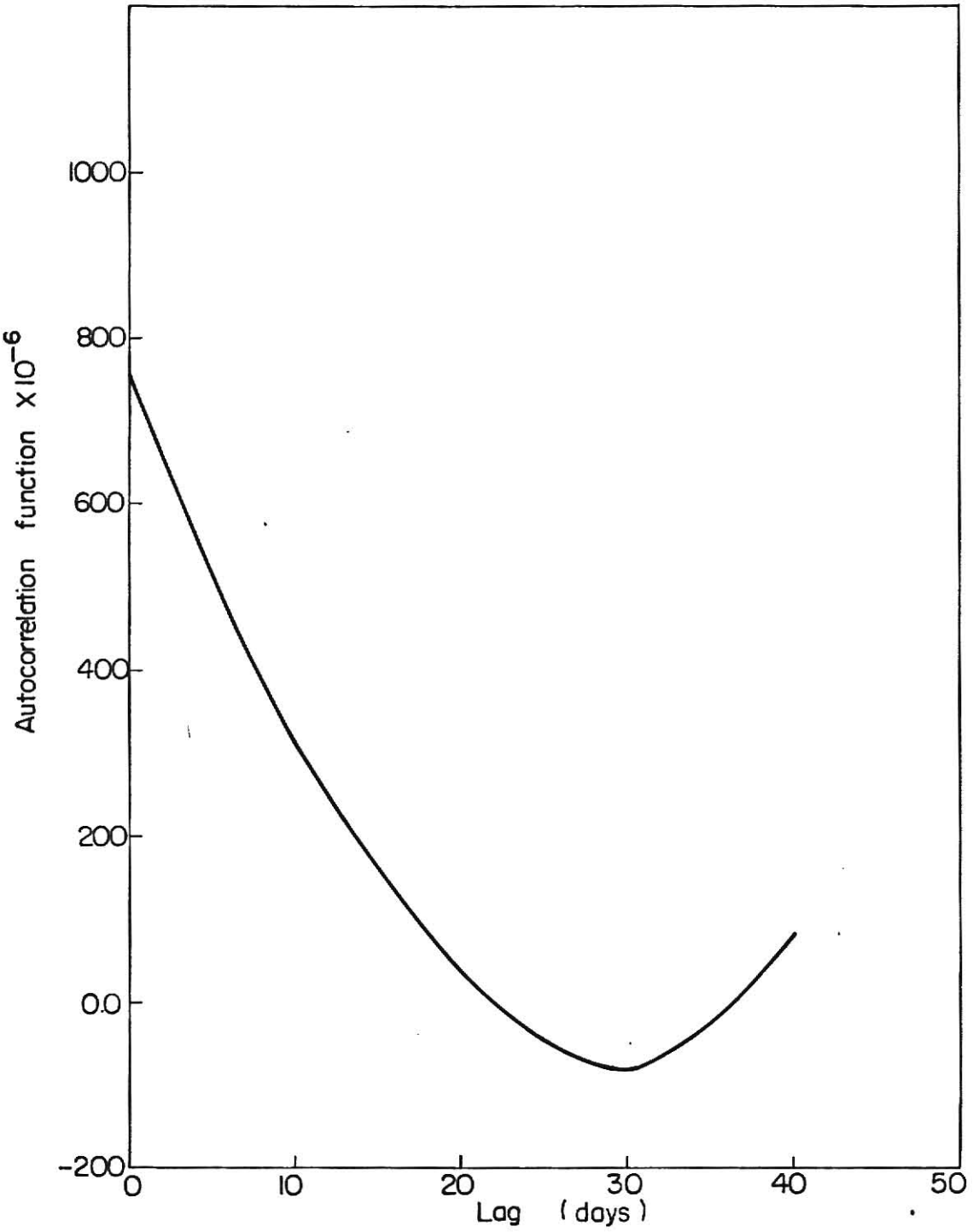


Fig. 4.2. Autocorrelation of flow values at station I (South Heights).

Table 4.5 Normalized Autocovariances for Flow Record, Ohio River (1968).

Station Lag (days)	South Heights (Station 1)	Stratton (Station 2)	Huntington (Station 3)	Cincinnati (Station 4)
0	1.000	1.000	1.000	1.000
1	0.9137	0.9450	0.9586	0.9743
2	0.8340	0.8558	0.8892	0.9224
3	0.7797	0.7862	0.8148	0.8585
4	0.7410	0.7411	0.7461	0.7914
5	0.6940	0.6969	0.6853	0.7245
6	0.6468	0.6478	0.6313	0.6607
7	0.5939	0.5906	0.5836	0.6014
8	0.5406	0.5328	0.5418	0.5474
9	0.4915	0.4725	0.5009	0.4966
10	0.4195	0.4232	0.4560	0.4470
11	0.3675	0.3677	0.4086	0.3954
12	0.3158	0.3209	0.3616	0.3439
13	0.2709	0.2813	0.3162	0.2903
14	0.2273	0.2414	0.2746	0.2387
15	0.1907	0.2070	0.2325	0.1919
16	0.1495	0.1650	0.1940	0.1527
17	0.1177	0.1263	0.1589	0.1191
18	0.0913	0.0942	0.1236	0.0875
19	0.0674	0.0703	0.0892	0.0581
20	0.0418	0.0447	0.0579	0.0308
21	0.0161	0.0183	0.0310	0.0074
22	-0.0062	-0.0030	0.0105	-0.0117
23	-0.0204	-0.0188	-0.0053	-0.0296
24	-0.0341	-0.0355	-0.0236	-0.0482
25	-0.0559	-0.0566	-0.0406	-0.0618
26	-0.0748	-0.0750	-0.0570	-0.0816
27	-0.0871	-0.0873	-0.0689	-0.948
28	-0.0982	-0.0969	-0.0743	-0.1022
29	-0.1057	-0.1041	-0.0797	-0.1024
30	-0.1080	-0.1089	-0.0785	-0.0990
31	-0.1025	-0.1062	-0.0715	-0.0941
32	-0.0869	-0.9680	-0.0633	-0.0876
33	-0.0733	-0.0805	-0.0539	-0.0777
34	-0.0547	-0.0610	-0.0411	-0.0649
35	-0.0316	-0.0364	-0.0233	-0.0487
36	-0.0023	-0.0095	0.0011	-0.0296
37	0.0264	0.0212	0.0299	-0.0081
38	0.0558	0.0523	0.0585	0.0159
39	0.0891	0.0880	0.0875	0.0406
40	0.1229	0.1234	0.1115	0.0629

criterion to compare the autocorrelation values at these stations. The monthly averages for Station 2 are also presented in Table 4.3. These values are generally larger than the flow values at South Heights (Station 1) meaning that in addition to the transfer of bulk material from Station 1 to Station 2 there is an input side stream added to the river. The graph of this autocorrelation function (Fig. 4.3) has a form very similar to the one for the South Heights station. The autocovariance (autocorrelation) is negative for all lags between 22 and 36 and is positive otherwise. The autocovariance once again reaches a minima at a lag of 30 days. The flow data at the other two stations (Stations 3 and 4) on the Ohio River are also summarized in Table 4.3. The average flow rates at Stations 3 and 4 were considerably higher than these at the first two stations. This was mainly due to additions from two streams, Muskingum and Kanawha. Autocorrelations of flow values for these two stations are plotted in Figs. 4.4 and 4.5. These values in a normalized form are also presented in Table 4.5. The autocorrelations (autocovariances) show a similar trend; all of them reach a minimum value at a lag of thirty days. The normalized autocorrelations are plotted in Fig. 4.6 for comparison. This figure shows that the autocorrelations for the same lag tend to decrease as one moves downstream from Station 1.

2. Power Spectral Estimates: The power spectrum estimates at Station 1 which are calculated according to Equation 31.are presented in Table 4.6 and in Fig. 4.7. The spectral estimates

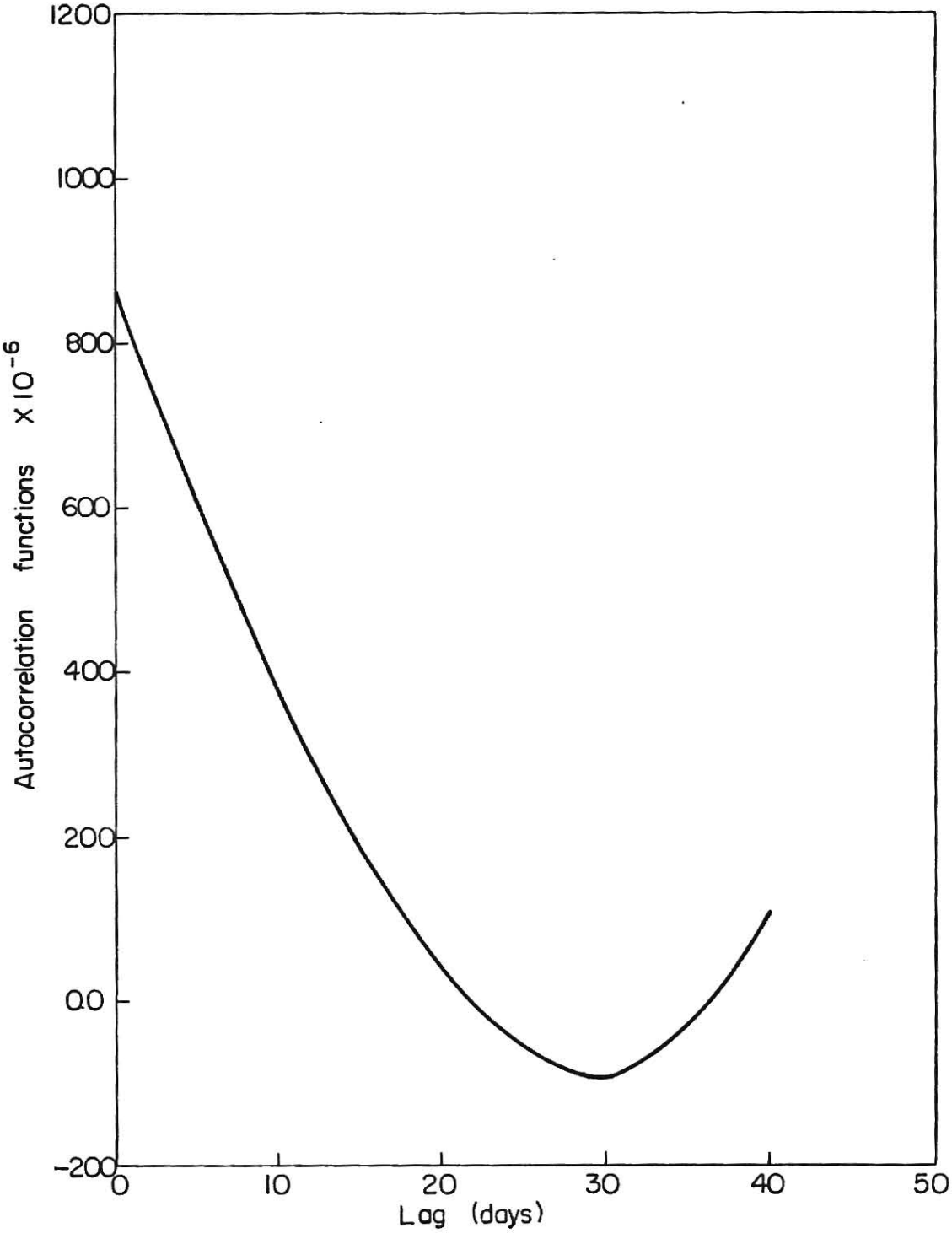


Fig. 4.3. Autocorrelation of flow values at station 2 (Stratton).

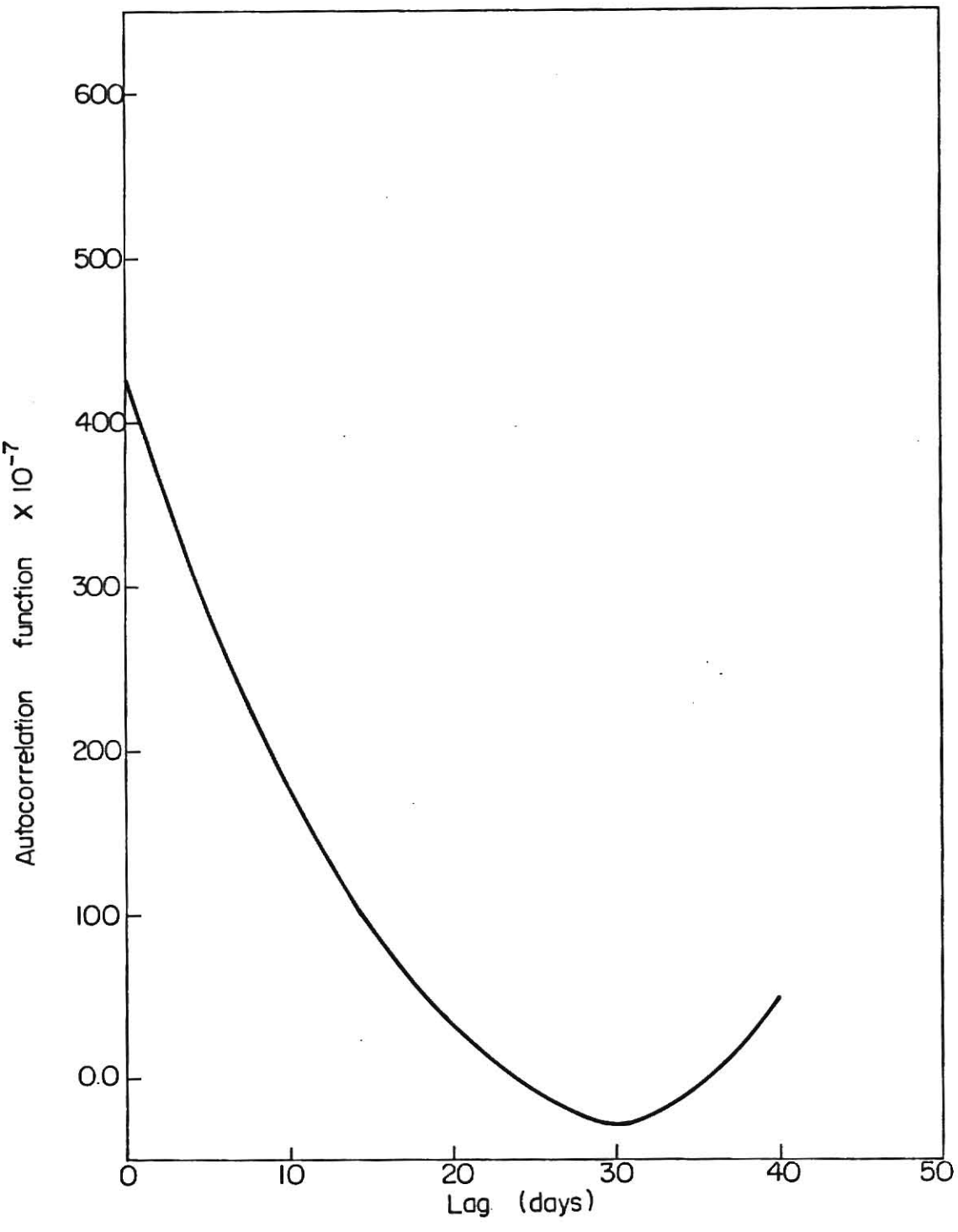


Fig. 4.4. Autocorrelation of flow values at station 3 (Huntington).

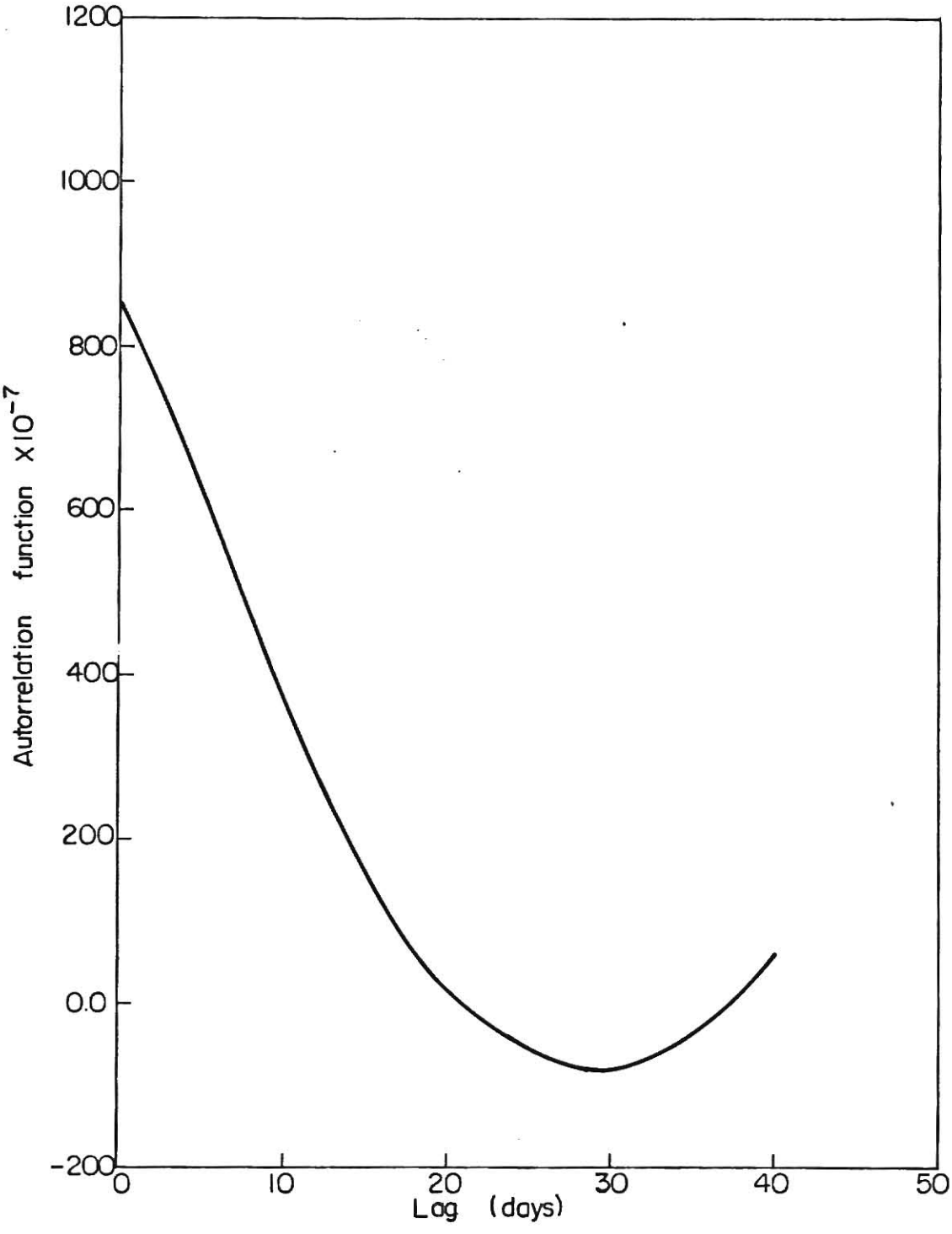


Fig. 4.5. Autocorrelation of flow values at station 4 (Cincinnati).

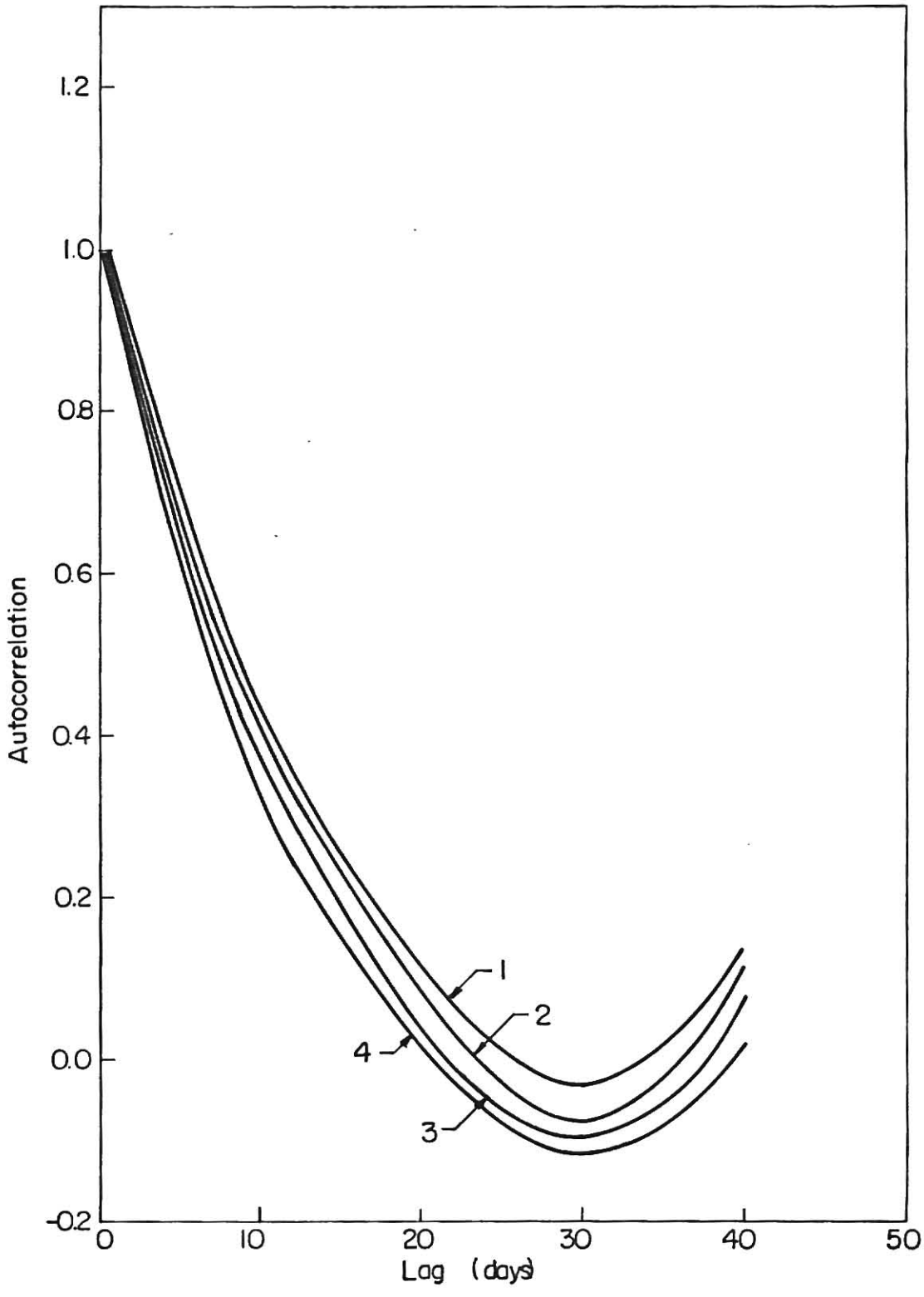


Fig. 4.6. Normalized autocorrelation for flow record, with station number as parameters.

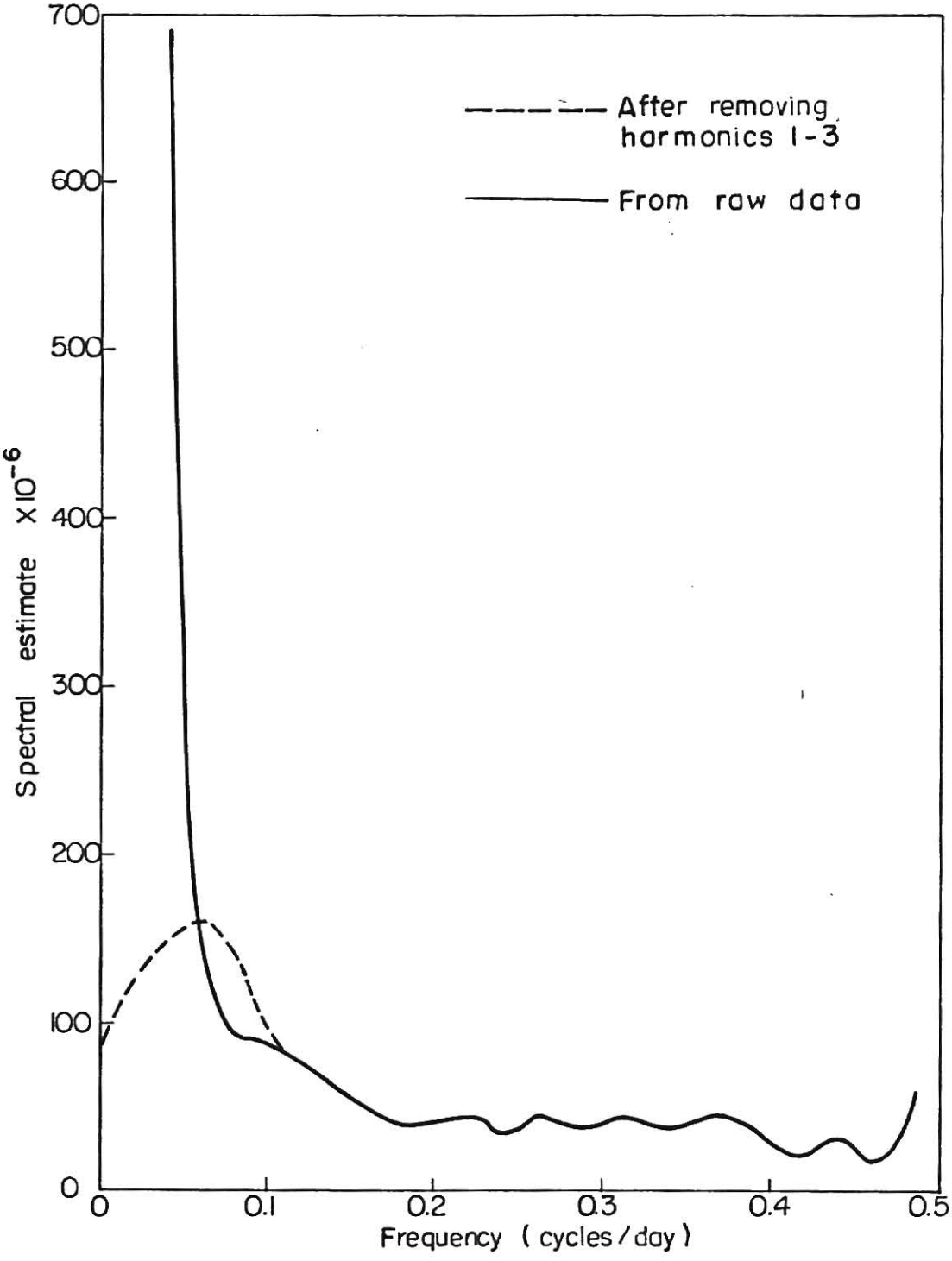


Fig. 4.7. Power spectrum estimates of flow values. at station 1. (South Height).

for smaller frequencies are exceedingly large. The period corresponding to each frequency can be calculated by noting that

$$P = \frac{1}{f} \tag{61}$$

in which

P = period in days

f = frequency in cycles/day

It should be noted that each of these periods corresponds to a lag and is related to the lag by the following relation

$$P_r = \frac{2m\Delta T}{r} \tag{62}$$

in which

P_r = period corresponding to lag r

ΔT = sampling interval in hours

m = total number of lags

r = lag number

The spectral estimates at this station are very large for low frequencies but drop suddenly at a frequency of 0.1 cycles per day corresponding to a 10 day period. After this frequency the estimate oscillates showing a number of periodic components.

Power spectral estimates at Station 2 in Fig. 4.8 show two somewhat significant periods at frequencies of 0.125 and 0.2 cycles per day, corresponding to periods of 8 and 5 days respectively. There is apparently no spectrum for frequencies

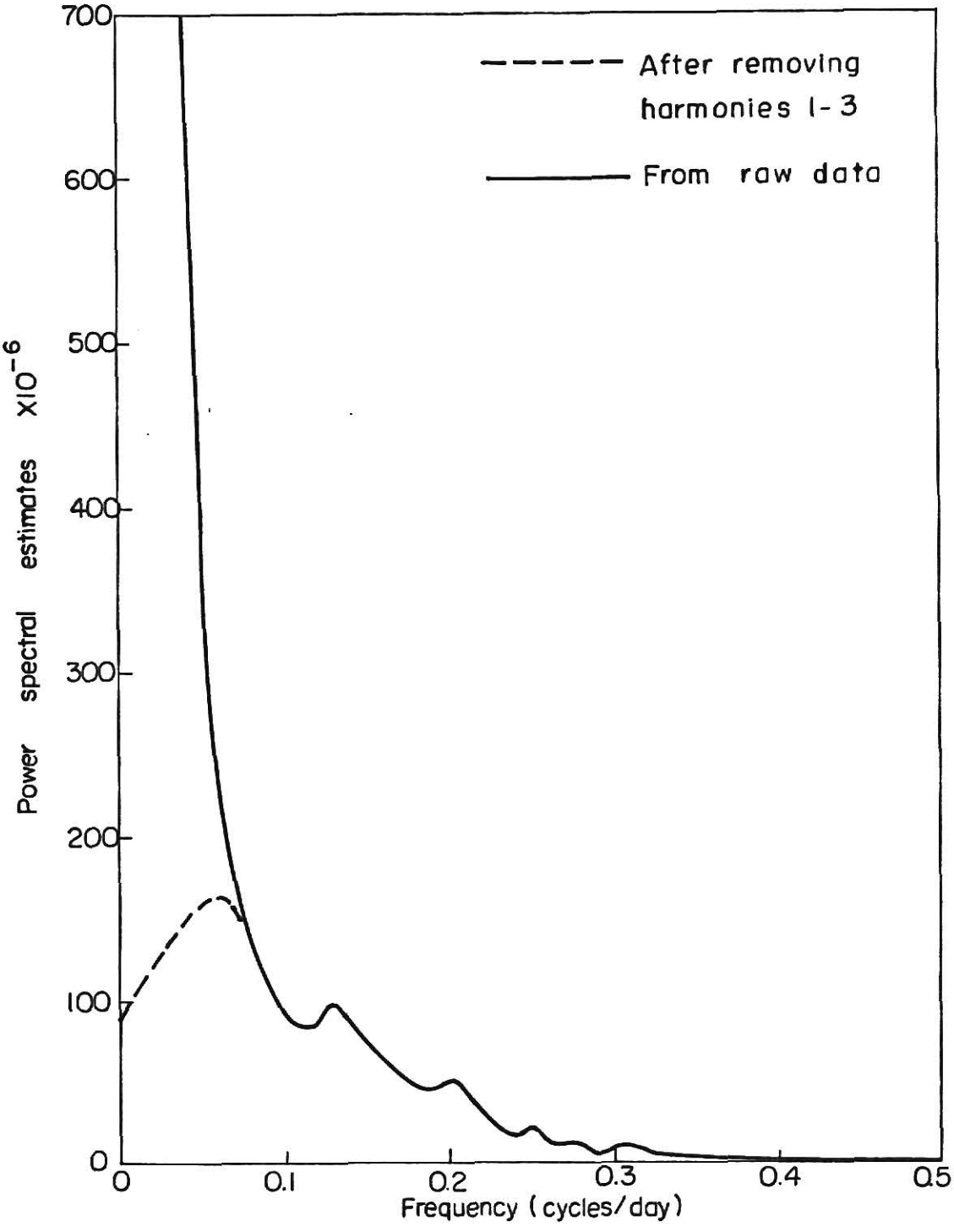


Fig. 4.8. Power spectral estimates of flow record. at station 2. (Stratton).

higher than 0.3 cycles/day. These values are also tabulated in Table 4.6.

Power spectral estimates at the other stations (Stations 3 and 4) are also presented in Table 4.6. These estimates do not show any significant periodicities. For Hungtington (Station 3), the spectral estimate drops to a low value for a frequency of 0.2 cycles/day and remains almost constant except for small periodic components. For Cincinnati (Station 4), the spectral estimates show no significant periodicities. These estimates are shown in Figs. 4.9 and 4.10. The flow records were also analyzed by harmonic analysis techniques. The residual data were calculated by removing the first, second and the third harmonics. The spectral estimates of these residuals show a periodic event with a period of about 14 days. This feature is common at all the five stations. The results of spectral analysis of the residual records are also shown in Figs. 4.7 to 4.9 .

3. Crosscorrelation function (crosscovariance): Crosscovariance (crosscorrelation function) values for a correlation monitoring stations (Stations 1 and 2) are shown in Fig. 4.11. These values are not symmeterical; the crosscovariance for negative lags is larger for very small and very large lags than the crosscovariance for positive lags. For intermediate lags, however, the crosscovariance for positive lags is larger. This can be interpreted in the following way. If the flow rate at station 1 is above normal, then the flow rate at station 2 was above normal for the previous 21 days and will be above normal for a lag

Table 4.6 Power Spectral Estimates of Flow Records, Ohio River.

Frequency (Cycles/ Day)	Power Spectral Estimates of Station 1 (x 10^{-6})	Power Spectral Estimates of Station 2 (x 10^{-6})	Power Spectral Estimates of Station 3 (x 10^{-6})	Power Spectral Estimates of Station 4 (x 10^{-6})
0.0	3819.006	4465.197	2326.6516	4564.6512
0.012	3350.475	3910.924	1980.0330	4043.9750
0.025	1969.918	2280.849	1108.4156	2399.2500
0.037	688.752	803.752	373.9676	867.1608
0.050	318.011	405.063	190.7837	396.3616
0.062	132.927	207.419	131.4337	219.8349
0.075	89.547	149.554	108.6604	163.2221
0.087	87.316	114.701	85.0515	138.4812
0.100	83.659	92.609	66.6785	101.2305
0.112	80.935	87.317	49.6595	54.1535
0.125	82.352	98.105	40.6386	35.2294
0.137	67.785	90.306	23.4193	15.0292
0.150	63.820	79.518	15.8825	13.0911
0.162	52.159	60.124	14.0573	18.1469
0.175	42.065	52.408	12.9230	17.5881
0.187	35.612	45.172	9.4057	8.8814
0.200	42.612	50.317	9.1239	7.8742
0.212	44.366	43.280	9.4123	6.2155
0.225	37.633	28.240	8.0017	7.9607
0.237	30.910	17.568	4.7270	5.4578
0.250	32.466	19.226	4.4125	4.0504
0.262	23.838	13.695	2.8633	2.8359
0.275	18.494	10.137	3.0245	2.8529
0.287	14.895	8.048	2.9246	1.4948
0.300	14.900	10.172	2.8064	3.0504
0.312	14.716	8.296	2.0799	2.3342
0.325	19.117	7.961	2.2907	2.0325
0.337	19.739	4.811	1.9675	1.3138
0.350	15.891	3.475	2.8601	1.8170
0.362	10.168	1.324	1.9515	0.9920
0.375	8.443	2.339	1.2150	1.2798
0.387	5.117	1.633	1.3680	0.8434
0.400	5.547	3.123	1.6952	1.5605
0.412	6.747	2.726	1.3318	1.5036
0.425	9.451	2.697	1.8556	1.8055
0.437	12.197	1.099	1.4942	0.6871
0.450	15.215	2.881	1.2542	0.9023
0.462	15.109	3.469	1.1314	0.5957
0.475	12.375	4.448	1.5303	1.4522
0.487	6.272	2.989	1.3740	1.2990
0.500	4.733	2.548	1.5635	1.8226

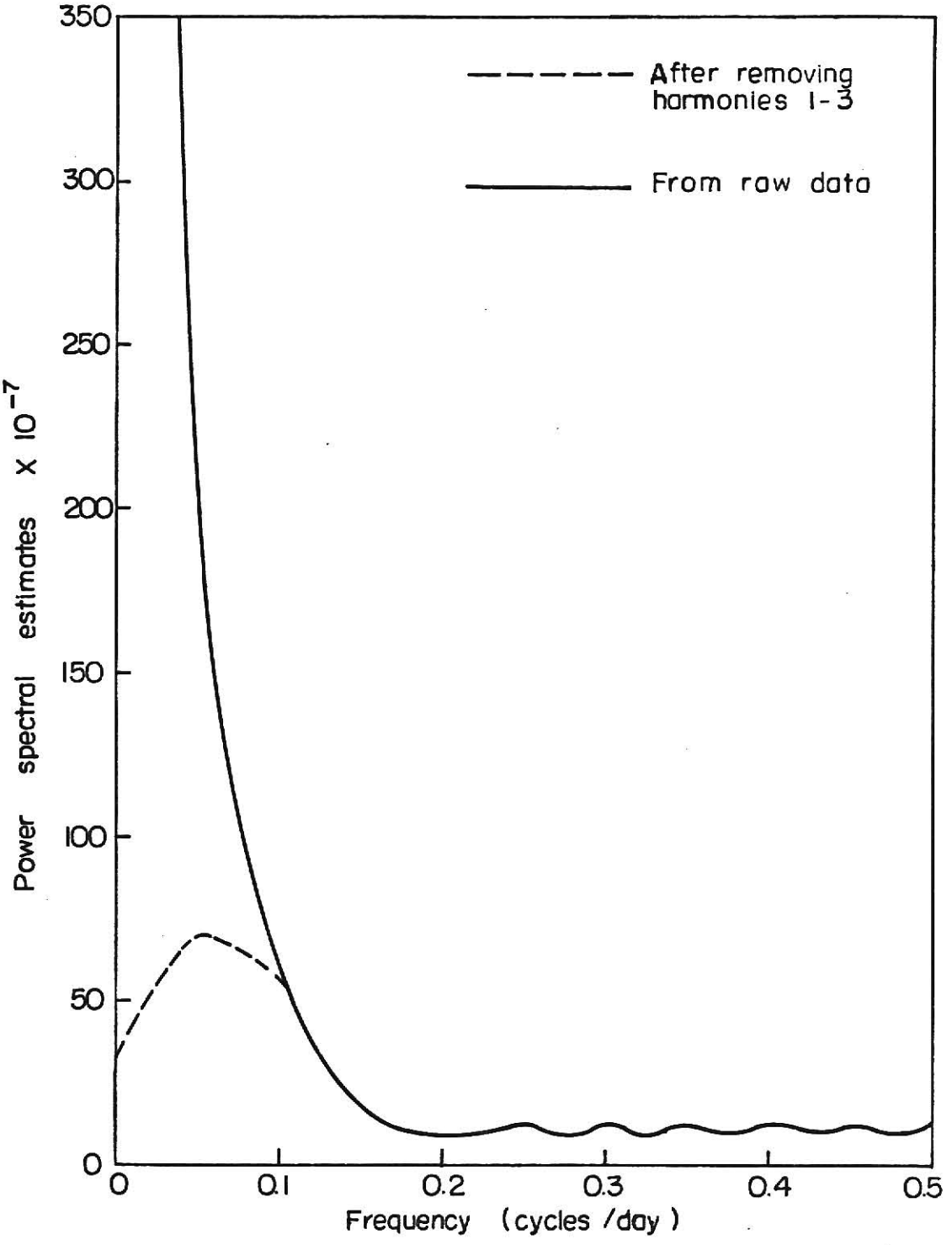


Fig. 4.9. Power spectral estimates of flow record at station 3 (Huntington).

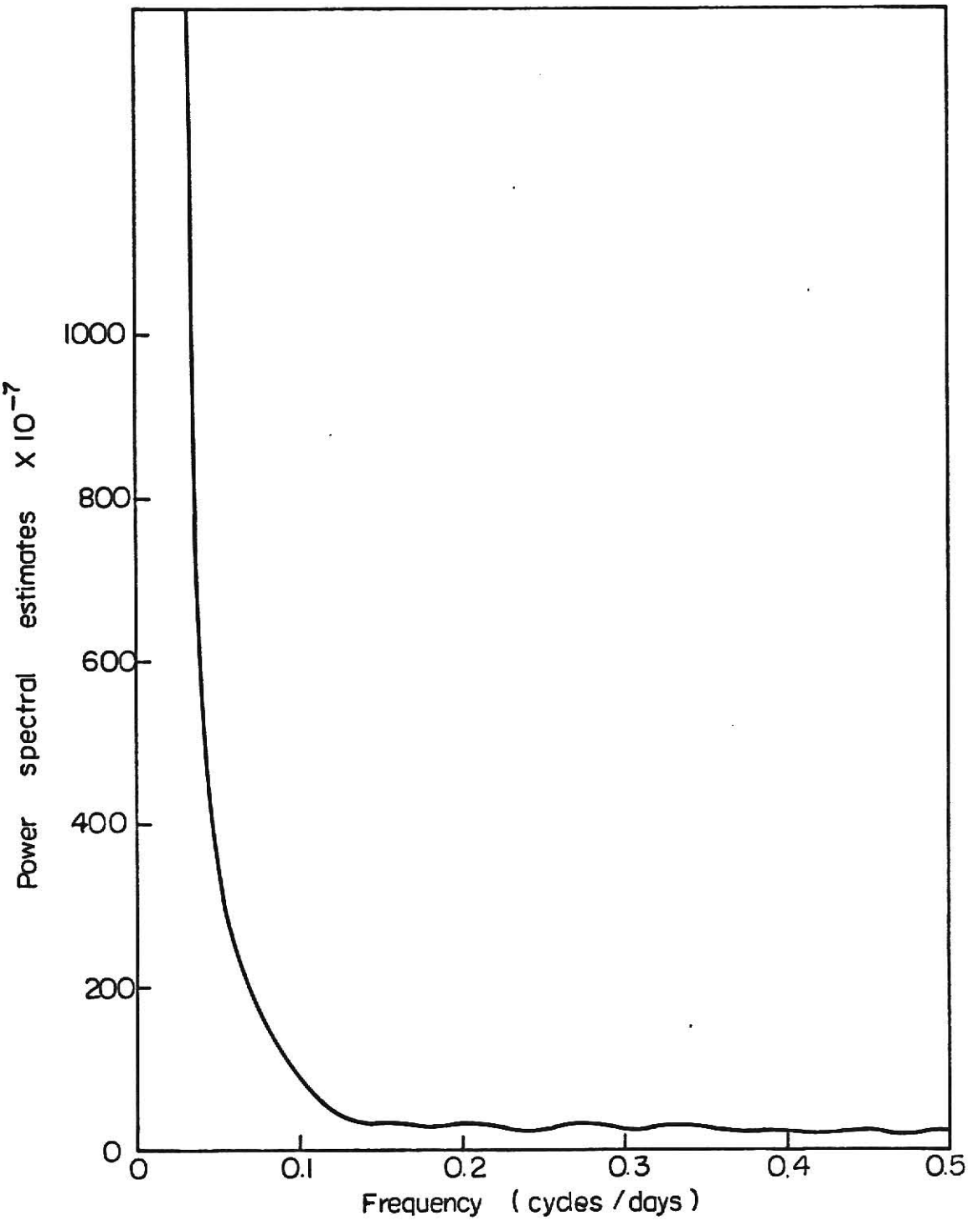


Fig. 4.10 Power spectral estimates of flow record at station 4 (Cincinnati).

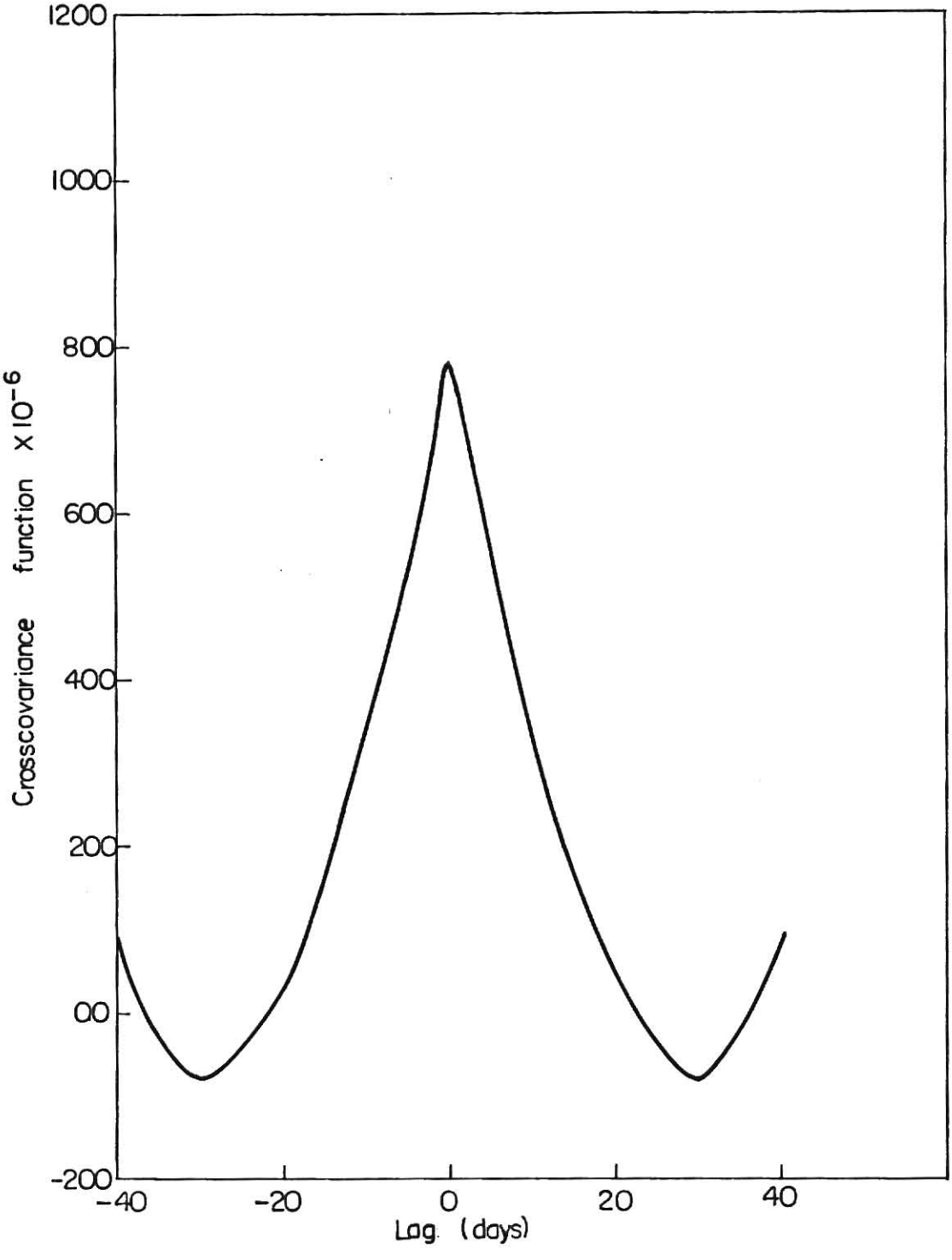


Fig. 4.11. Crosscovariance of flow records between station 1 and 2.

period of 22 days. It will be below normal or constant for lags between 23 and 36 days and also for lags from -23 to -36 days. For these two stations the crosscorrelation peaks near zero lag.

The crosscovariance between time series representing flow rates at station 1 and 3 and 1 and 4 is shown respectively in Figs. 4.12 and 4.13. The general nature of the curve is the same as that for Stations 1 and 2. The crosscovariance reaches minima at two points corresponding to lags of 30 and -30 periods. While, it cannot be easily visualized from these figures, unlike in Fig. 4.11 each of these crosscovariances reaches a maxima at a lag of 1-3 days. This is because after this period the same bulk of fluid is present at the next station, and hence there is a good correlation. These crosscovariance are tabulated in Table 4.7(a-c).

4. Transfer Function and Coherence: The amplitude and phase of the transfer function relating the flow between each of the three sets of stations may be tabulated as a function of the frequency. The amplitudes which are plotted in Figs. 4.14 and 4.15 indicate an increase in the flow rate as one proceeds downstream from Station 1. The amplitude functions fall off at high frequencies just as amplitudes in Bode diagrams do for first and second order systems. Comparison of the experimental amplitude--frequency plot with the theoretical curves for first and second order systems indicate that the system is close to a first order one with small nonlinear interactions.

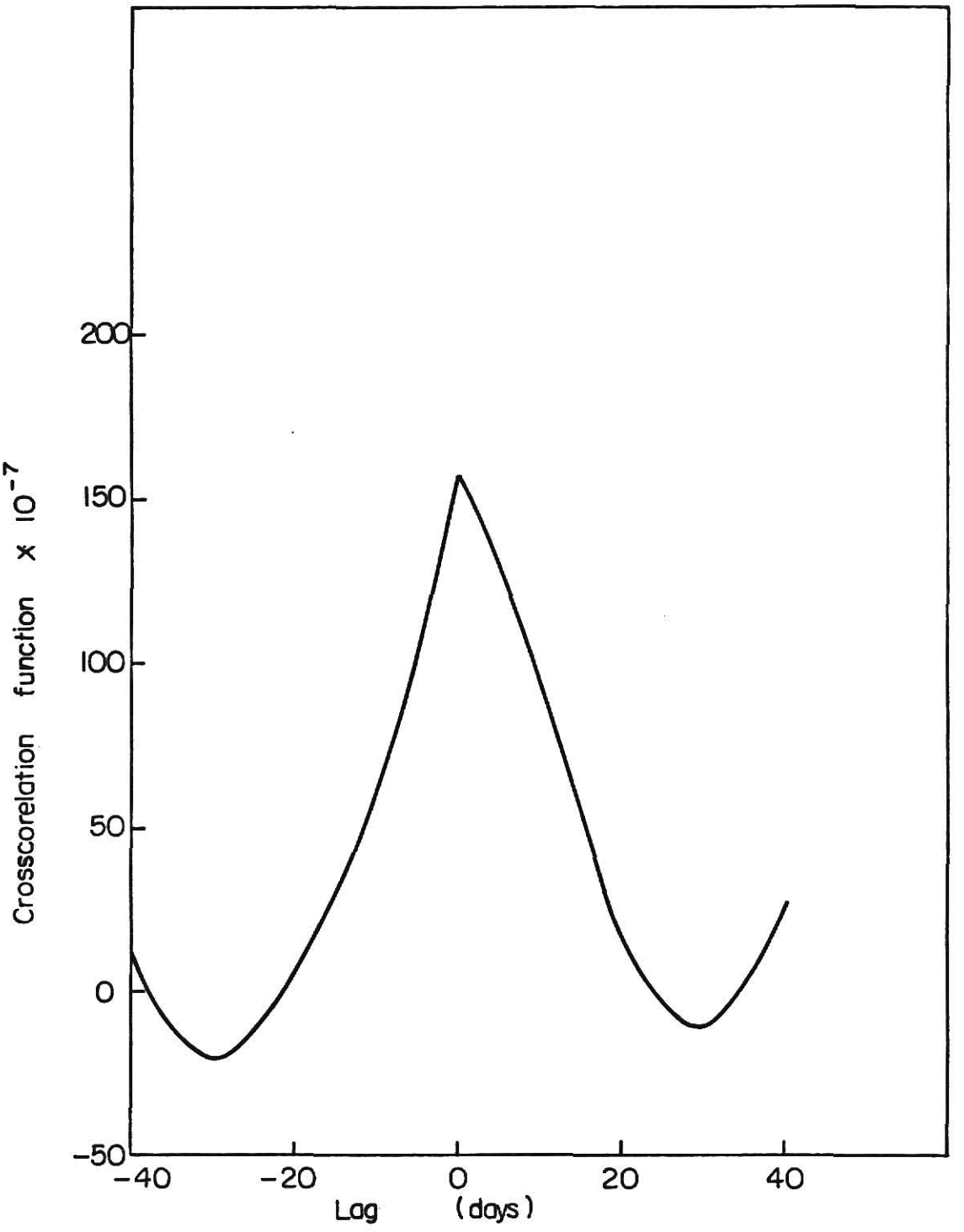


Fig. 4.12. Crosscorrelation of flow records between stations 1 and 3.

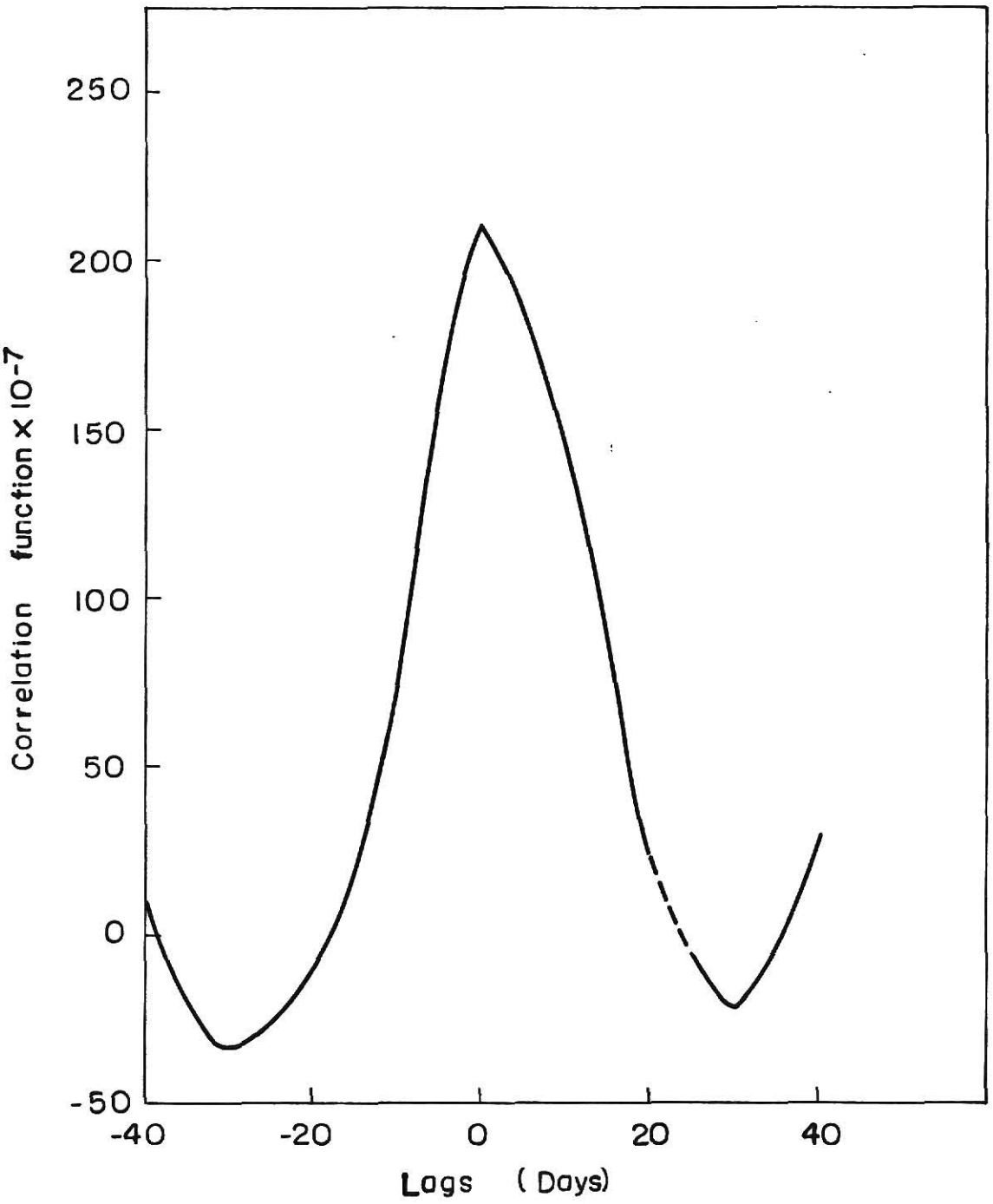


Fig.4·13. Crosscorrelation of flow records between Stations 1 and 4.

Table 4.7a Crosscovariance of Flow Record between Station 1 and 2 (x 10^{-6}).

Lag Days	Crosscovariance between Stations 1 and 2 (Positive Tau)	Crosscovariance between Stations 1 and 2 (Negative Tau)
0	780.077	780.077
1	735.927	740.743
2	672.564	676.115
3	626.443	628.352
4	591.680	593.745
5	558.379	558.357
6	521.848	520.472
7	479.144	474.905
8	437.378	429.853
9	396.532	388.715
10	345.954	343.252
11	299.742	291.012
12	262.853	249.106
13	231.968	209.432
14	201.258	174.035
15	171.483	145.669
16	136.456	114.464
17	106.693	88.622
18	82.924	64.825
19	64.170	45.703
20	43.491	25.597
21	21.871	5.578
22	3.774	-10.788
23	-10.022	-20.728
24	-23.528	-31.416
25	-41.798	-48.047
26	-57.063	-62.010
27	-67.264	-72.212
28	-77.210	-78.356
29	-84.168	-83.298
30	-86.852	-87.157
31	-81.995	-85.149
32	-70.922	-76.169
33	-58.160	-64.757
34	-45.774	-48.030
35	-28.081	-28.185
36	-4.851	-5.330
37	19.422	19.270
38	41.772	44.470
39	68.060	73.046
40	95.565	101.146

Table 4.7b Crosscovariance of Flow Record between Stations 1 and 3 ($\times 10^{-6}$)

Lag Days	Crosscovariance between Stations 1 and 3 (Positive Tau)
0	1571.601
1	1623,765
2	1569.083
3	1478.858
4	1393.004
5	1311.607
6	1225.147
7	1138.581
8	1063.954
9	995.109
10	910.190
11	818.859
12	737.303
13	646.583
14	560.549
15	492.689
16	418.506
17	345.865
18	284,225
19	219.993
20	156.334
21	103.704
22	55.174
23	12.997
24	-17.432
25	-55.837
26	-94.229
27	-116.044
28	-124.516
29	-134.967
30	-135.045
31	-123.651
32	-100.570
33	-61.821
34	-29.570
35	11.733
36	59.011
37	111.529
38	161.425
39	211.931
40	266.374

Table 4.7c Crosscovariance of Flow Record between Stations 1 and 4 ($\times 10^{-6}$).

Lag Days	Crosscovariance between Stations 1 and 4 (Positive Tau)	Crosscovariance between Stations 1 and 4 (Negative Tau)
0	2128.062	2128.062
1	2160.959	1999.896
2	2145.162	1820,389
3	2083.713	1682.772
4	1999.148	1571.745
5	1905.997	1444,096
6	1794.148	1300.929
7	1676.950	1171.734
8	1570.649	1035.488
9	1467.299	897.298
10	1349.987	761.504
11	1235.326	633.707
12	1118.371	536.587
13	986.079	440.833
14	848.084	346.397
15	725.584	246.923
16	615.442	152.212
17	513.994	68.771
18	420.941	3.418
19	334.797	-53.403
20	237.178	-108.487
21	150.028	-154.069
22	76.676	-184.563
23	13.787	-209.856
24	-42.508	-245.326
25	-96.600	-282.460
26	-144.978	-319.153
27	-185.540	-351.517
28	-205.720	-374.779
29	-220.587	-380.046
30	-227.062	-364.073
31	-213.653	-349.705
32	-186.180	-337.933
33	-137.981	-309.078
34	-85.396	-264.738
35	-31.033	-211.278
36	28.597	-148.892
37	92.924	-78.664
38	158.330	-11.347
39	224.013	61.001
40	293.068	121.554

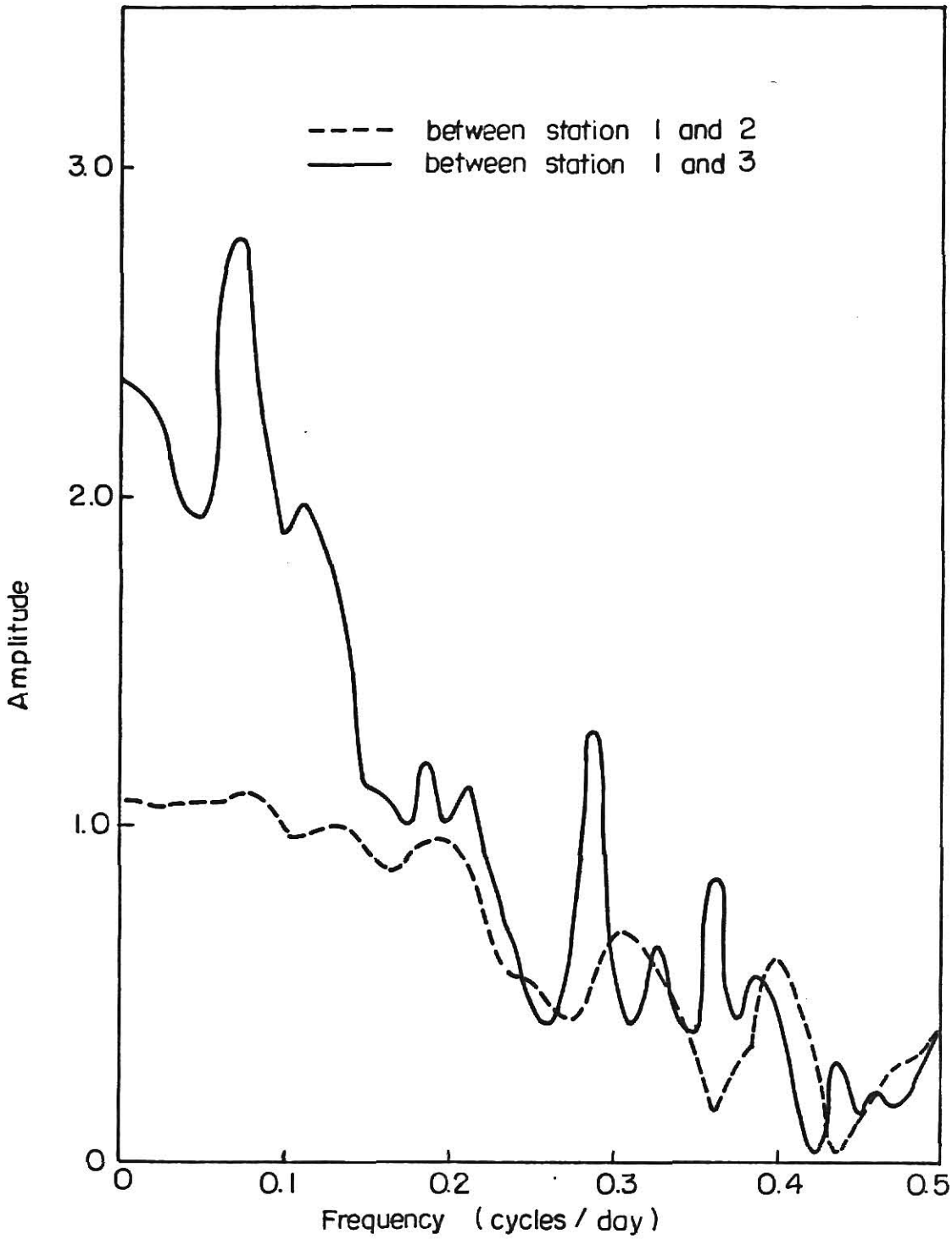


Fig. 4.14 Amplitude of transfer function for flow records, between Stations 1 and 2 and between Stations 3 and 4.

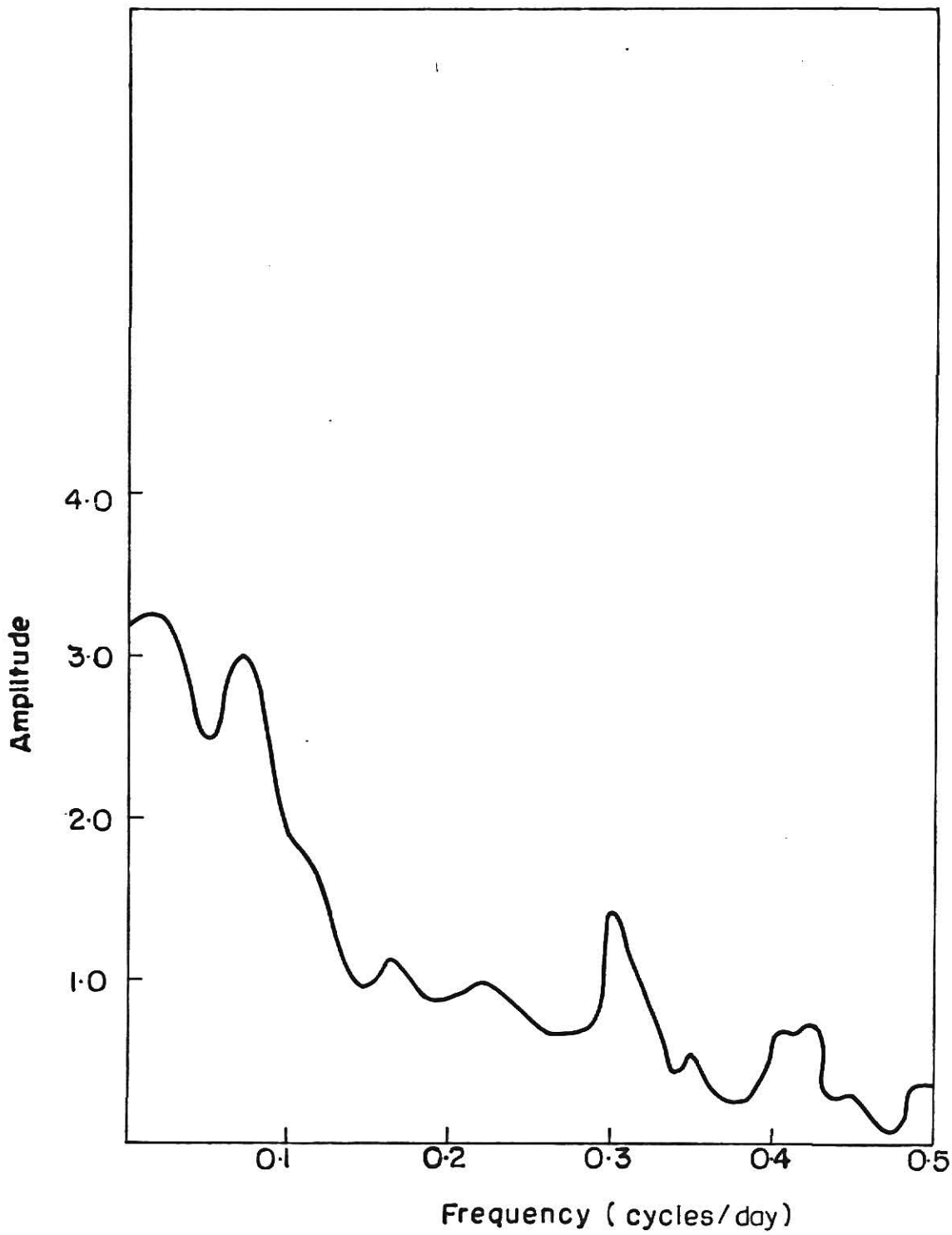


Fig. 4·15 Amplitude of transfer function for flow records between Stations 1 and 4.

The coherence squares tabulated in Table 4.8 can be used as a measure of linearity. A system is classified as linear if the coherence square between the input and output is independent of frequencies and is equal to unity for all frequencies. For the flow rate between Station 1 and 2, the coherence square ranges from 0.09974 to 0.9989. This means that the relationship between the flow rates is not perfectly linear. At small frequencies, however, the coherence square function is quite high and close to unity but this function falls down to lower values at higher frequencies. This indicates nonlinearity in the system.

Analysis of the Temperature Records

Monthly average temperature values for the five monitoring stations on the Ohio River are summarized in Table 4.9. The sampling interval for this record was 12 hours; thus, 732 data points were obtained for each of the time series. The raw data indicate that the temperature of the stream has some periodicities which can be identified using spectral analysis.

The autocovariance (autocorrelation function) of each temperature series was calculated up to a lag of 10%. This corresponds to 75 lag periods or 37 1/2 days. These results are plotted in Fig. 4.16. All autocorrelations (autocovariances) up to a lag of 75 are positive. This means if the temperatures above the annual mean temperature now, it will probably stay above the mean for all these lag periods. All autocorrelations are linear with lag for higher lags but tend to curve off for

Table 4.8 Coherence Square between Flow Records at different stations Ohio River (1968).

(Cycles/Halfday)	Coherence Square		
	Station 1 and Station 2	Station 1 and Station 3	Station 1 and Station 4
0.0	0.7479	0.8975	0.9003
0.0067	0.5768	0.7871	0.8017
0.0133	0.1271	0.0349	0.0161
0.0200	0.0020	0.0270	0.4729
0.0267	0.0314	0.3531	0.1020
0.0400	0.1284	0.2117	0.4692
0.0467	0.2453	0.3865	0.4692
0.0533	0.1296	0.2582	0.0409
0.0600	0.2199	0.2548	0.6313
0.0667	0.1708	0.0128	0.3328
0.0733	0.2330	0.3852	0.3697
0.0800	0.1411	0.0609	0.1613
0.0867	0.0142	0.0041	0.0687
0.0933	0.0027	0.1503	0.0112
0.1000	0.3743	0.2681	0.0601
0.1133	0.0881	0.2425	0.0728
0.1200	0.1991	0.0635	0.1745
0.1267	0.2684	0.0118	0.0717
0.1333	0.3705	0.1492	0.3514
0.1400	0.3996	0.1018	0.1839
0.1467	0.5730	0.1857	0.1520
0.1533	0.1642	0.3473	0.0880
0.1600	0.0474	0.3428	0.2443
0.1667	0.1007	0.1528	0.2878
0.1733	0.2850	0.2662	0.1012
0.1800	0.2027	0.2644	0.1191
0.1867	0.4740	0.0572	0.1953
0.1933	0.0659	0.0177	0.3342
0.2000	0.1089	0.0131	0.2826
0.2133	0.3119	0.0089	0.1869
0.2200	0.0478	0.0991	0.1948
0.2267	0.1172	0.0388	0.2062
0.2333	0.1736	0.0574	0.0464
0.2400	0.3417	0.0594	0.0659
0.2467	0.0484	0.0174	0.0390
0.2533	0.1193	0.4284	0.1818
0.2600	0.4637	0.0274	0.2033
0.2667	0.1209	0.3350	0.3744
0.2733	0.0624	0.1758	0.2056
0.2800	0.2415	0.1581	0.2021
0.2867	0.3565	0.1541	0.0949
0.2933	0.1537	0.1981	0.0588

Table 4.8 (Cont'd)

Frequency (Cycles/Halfday)	Coherence Square		
	Station 1 and Station 2	Station 1 and Station 3	Station 1 and Station 4
0.3000	0.1250	0.0967	0.0771
0.3067	0.1220	0.3024	0.0269
0.3133	0.4704	0.2921	0.0178
0.3200	0.2219	0.0111	0.2495
0.3267	0.1802	0.1738	0.2018
0.3333	0.1271	0.0401	0.0961
0.3400	0.3429	0.1104	0.0152
0.3467	0.2638	0.0879	0.0280
0.3533	0.2239	0.1069	0.0107
0.3600	0.1060	0.2052	0.1446
0.3667	0.1857	0.0357	0.0792
0.3733	0.2877	0.0447	0.1253
0.3800	0.1943	0.3287	0.0115
0.3867	0.1157	0.4150	0.1130
0.3933	0.3450	0.0836	0.3048
0.4000	0.3036	0.0562	0.1458
0.4067	0.0781	0.1191	0.0021
0.4133	0.1314	0.0174	0.0215
0.4200	0.2543	0.0252	0.0738
0.4267	0.2016	0.0722	0.1746
0.4333	0.0188	0.0020	0.3687
0.4400	0.1939	0.0856	0.2637
0.4467	0.1715	0.0627	0.0985
0.4533	0.1682	0.0109	0.1410
0.4600	0.2043	0.0831	0.1662
0.4667	0.3218	0.1383	0.0201
0.4733	0.3394	0.0911	0.0073
0.4800	0.2281	0.0463	0.0020
0.4867	0.1243	0.0678	0.0164
0.4933	0.0995	0.0371	0.0415
0.5000	0.0772	0.0086	0.0554

Table 4.9 Average Monthly Temperature (^{o}F) on Ohio River (1968).

Station / Month	South Heights (Station 1)	Stratton (Station 2)	Huntington (Station 3)	Cincinnati (Station 4)	Miamifort (Station 5)
January	40.76	36.44	38.47	39.09	37.34
February	36.46	35.76	41.57	38.00	36.93
March	43.45	41.79	45.71	43.47	41.70
April	56.77	55.48	58.01	55.55	54.60
May	60.56	62.54	63.55	63.23	63.62
June	68.85	68.29	69.00	67.45	66.99
July	79.47	81.55	82.30	80.79	79.72
August	80.41	81.85	82.51	82.32	82.26
September	76.11	76.43	76.49	76.09	75.52
October	66.70	68.57	70.31	68.49	69.73
November	52.81	53.48	56.63	56.28	56.43
December	39.36	40.22	43.02	44.73	43.30
Mean	58.47	58.53	60.46	59.62	59.01
Variance	98.50	93.20	78.18	79.91	82.61
Distance from Base Pt.	15.8	55.0	304.2	462.8	490.0

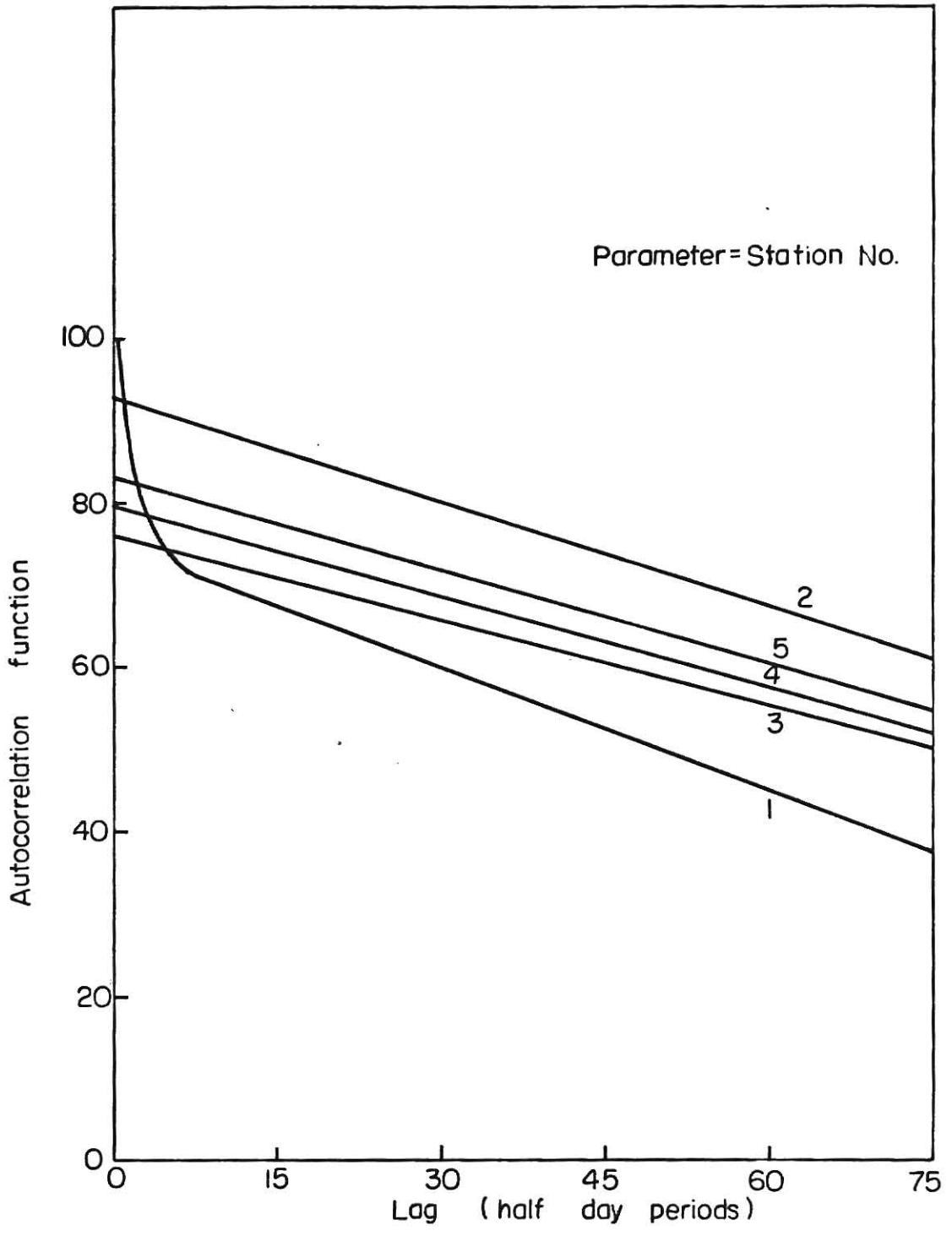


Fig. 4.16 Autocorrelation of temperature record for Stations 1-5, Ohio River, 1968.

very low lags. This long linear protion of the curve may be due to the variation in stream temperature caused by seasonal factors while the nonlinear portion of the curve at small lags may be caused by changes in stream temperature due to weather conditions, thermal discharges, and other factors. Extrapolation of these curves shows that autocorrelation goes below zero for lags beyond 125. Unlike the autocorrelation for the flow records these autocorrelations do not show any extreme, and are monotonically decreasing, they can be represented by linear equations in terms of the lag. The temperature records at all the five stations were analyzed by means of spectrum analysis. The results of this analysis are presented in Figs. 4.17 through 4.21.

The temperature series at all the five locations were also analyzed by means of harmonic analysis. For this, Fourier coefficients corresponding to each of the harmonics were computed and the significant harmonics were subtracted from the data. Table 4.10 summarized these results. In these computations any harmonic contributing less than 1% to the total variance was considered insignificant. The harmonic coefficients A and B calculated using Equations 9 and 10 (Chapter III) are also shown in this Table. The fundamental period of 366 days and other large periods (Table 4.10) were removed from the raw data by substracting off the contributions due to these harmonics. The residual record was again analyzed by spectral analysis. This time some additional periodic phenomena could be observed as

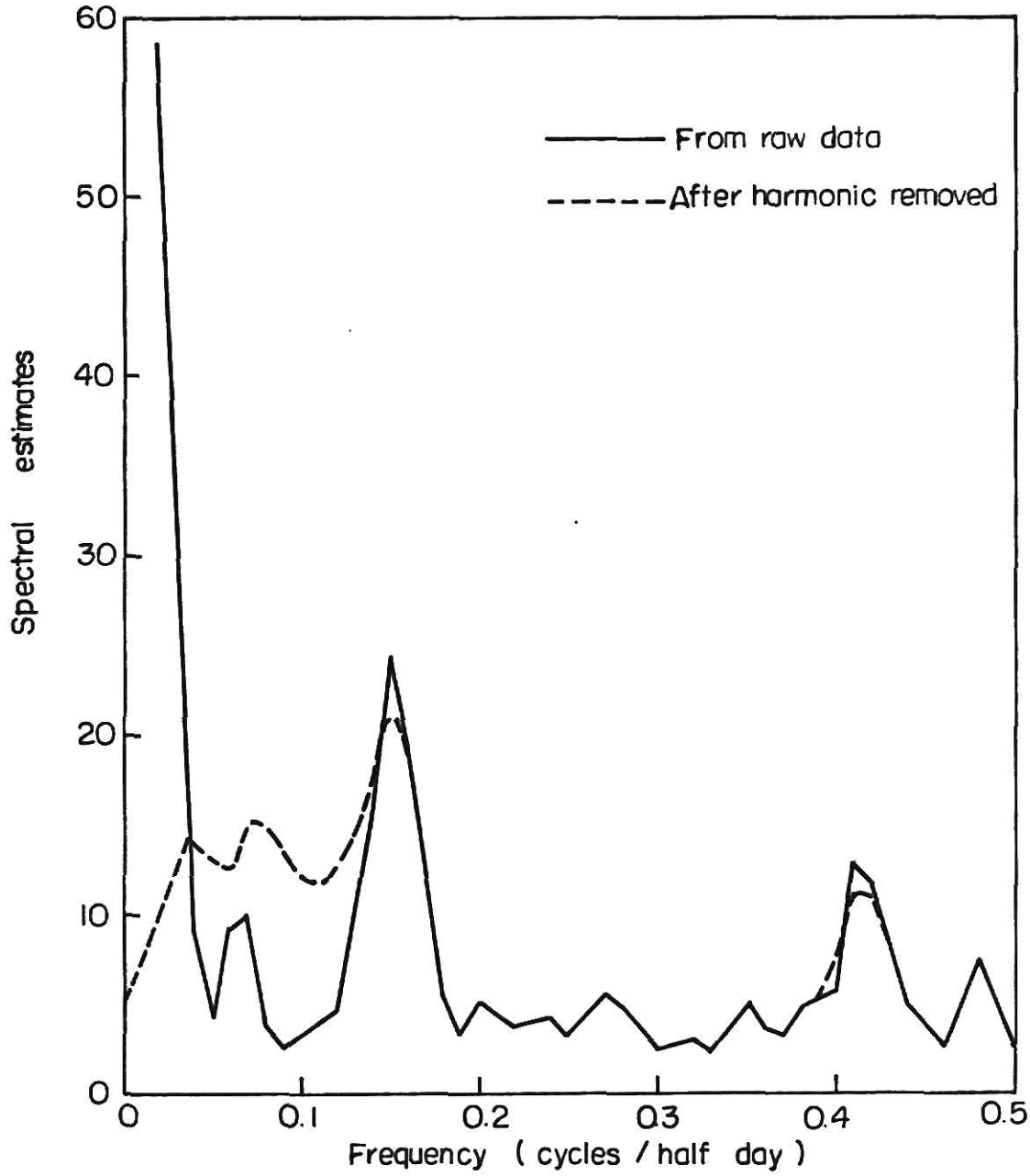


Fig. 4.17. Power spectral estimates of temperature record at Station 1, Ohio River, 1968.

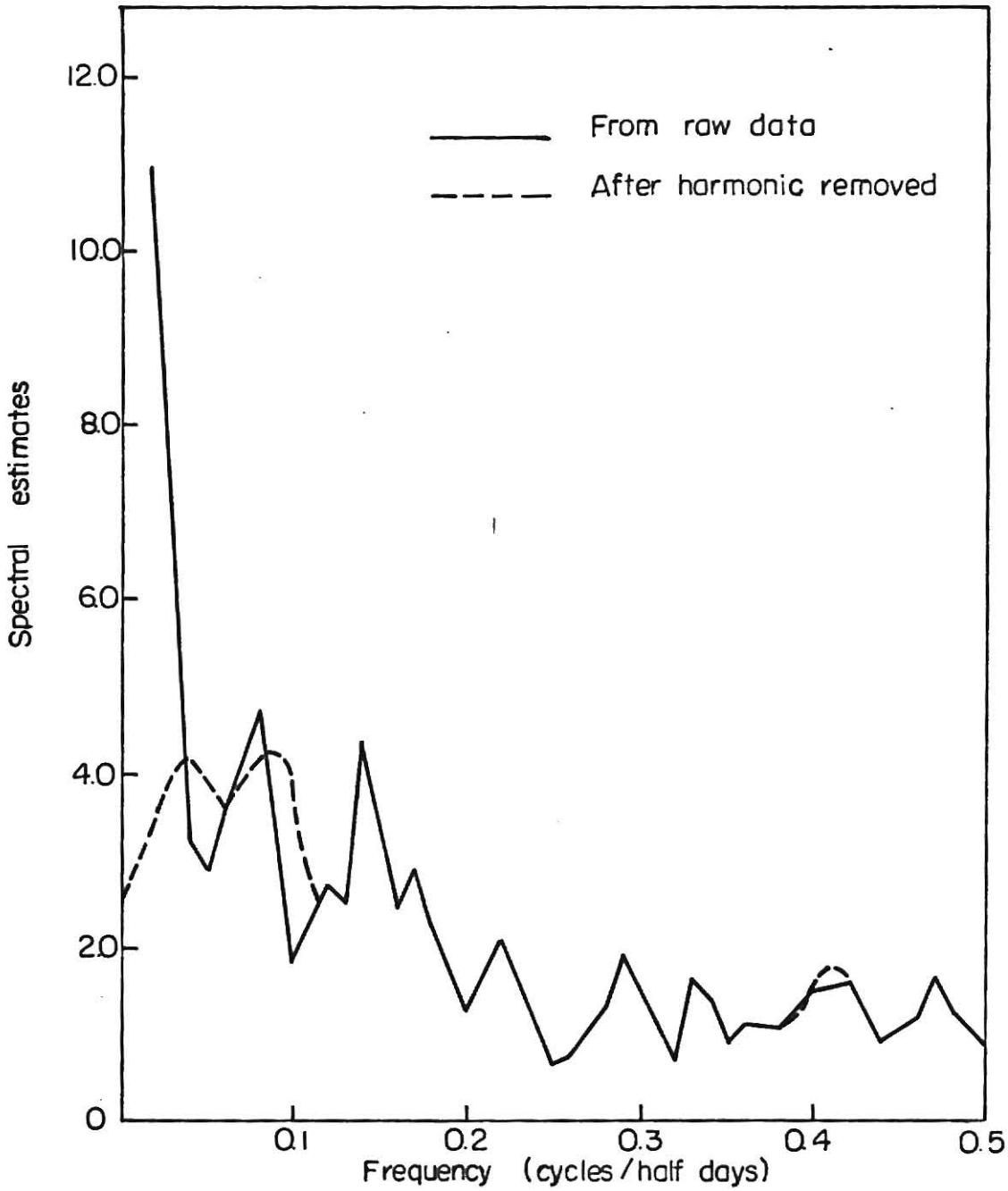


Fig. 4.18 Power spectral estimates of temperature record at Station 2, Ohio River, 1968.

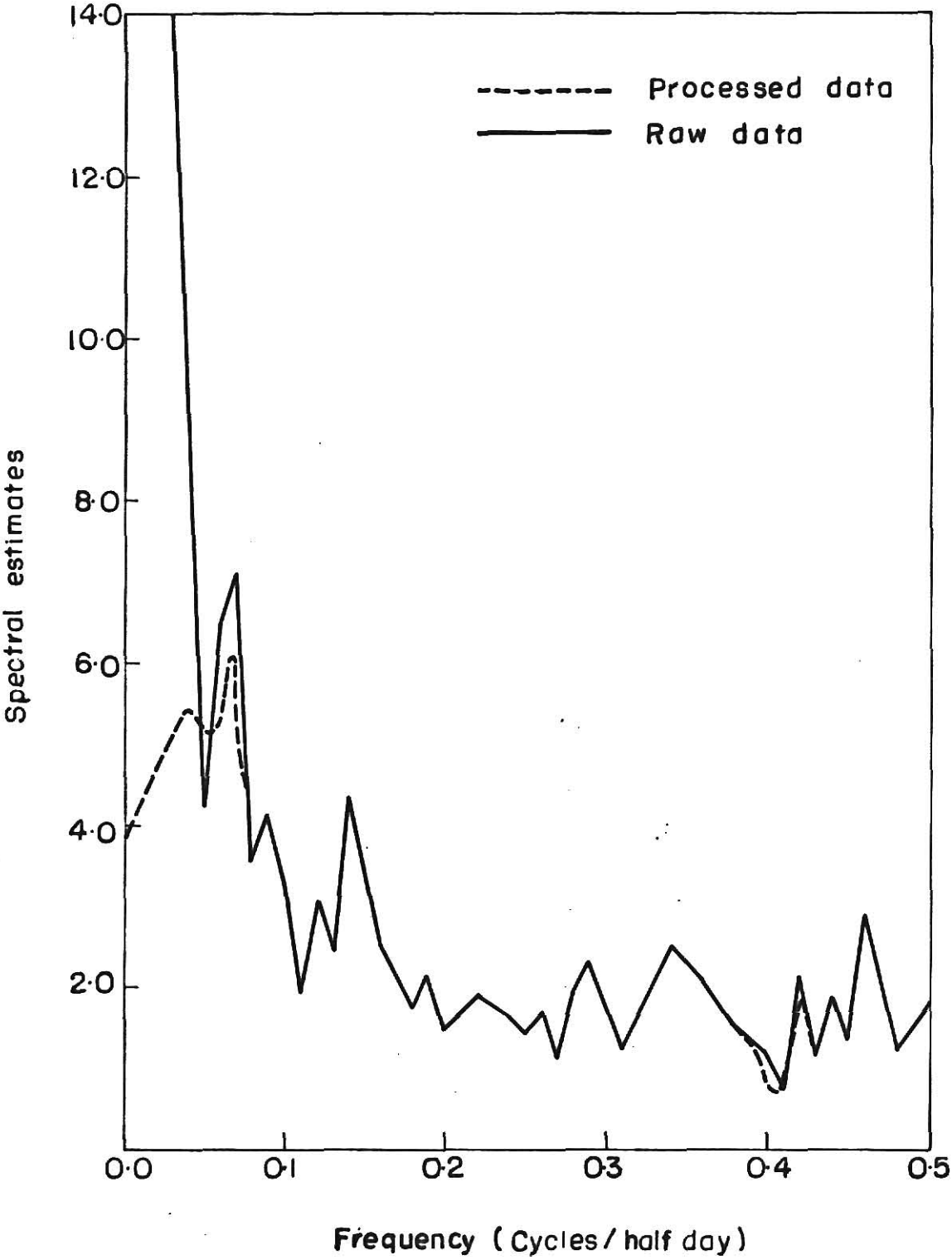


Fig. 4.19 Power spectral estimates of temperature record at Station 3, Ohio River 1968

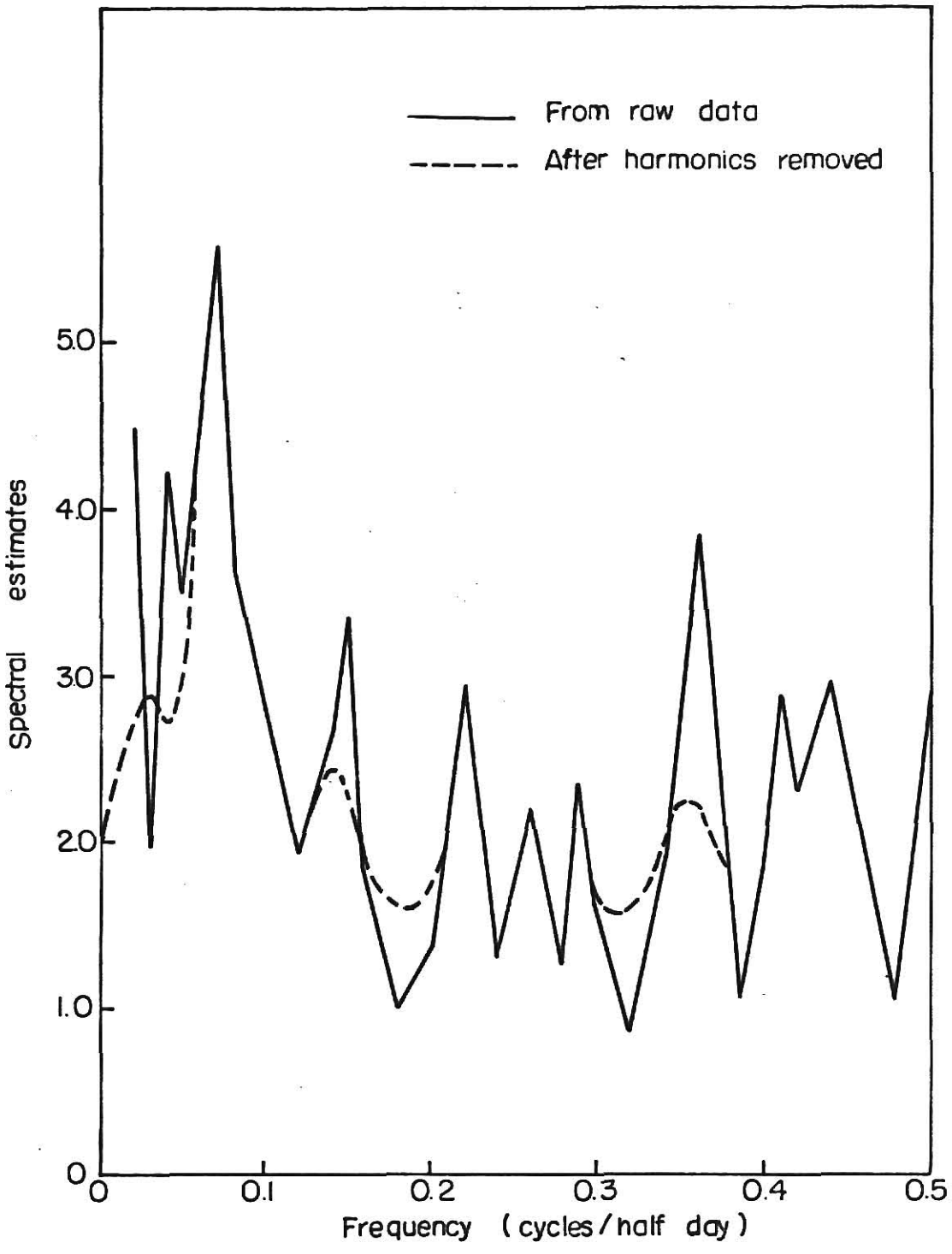


Fig. 4.20 Power spectral estimates of temperature record at Station 4, Ohio River, 1968.

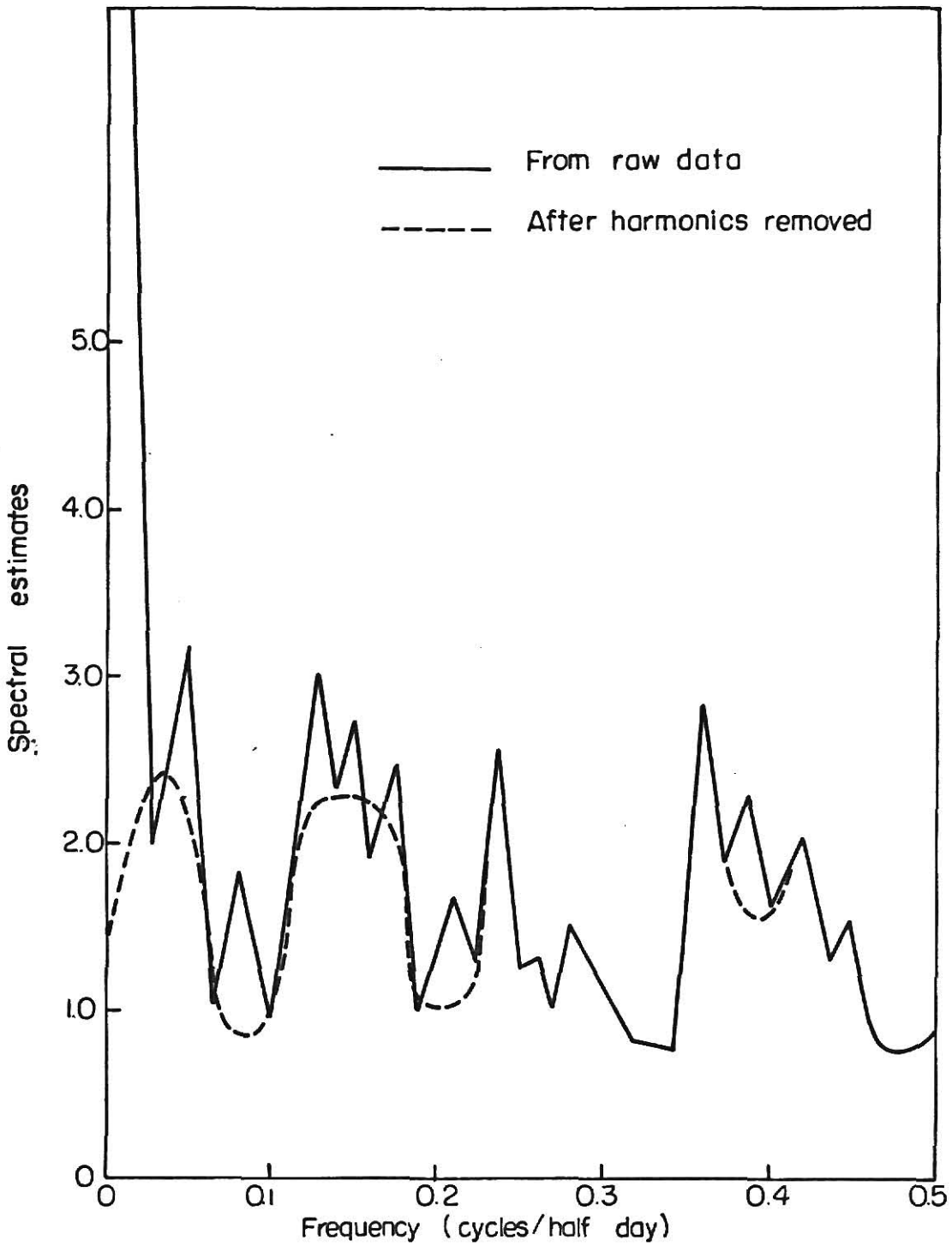


Fig.4.21 Power spectral estimates of temperature record at Station 5, Ohio River, 1968.

Table 4.10 Results of Harmonic Analysis of Temperature Records.

Station	Period (days)	Harmonic Number	A*	B*
1	366	1	23.1	24.2
	183	2	2.1	0.9
	122	3	1.0	0.28
	91.5	4	<1	<1
	73.2	5	<1	<1
2	366	1	21.6	24.5
	183	2	1.96	1.18
	122	3	0.86	1.01
	91.5	4	<1	<1
	73.2	5	<1	<1
3	366	1	25.6	19.5
	183	2	3.1	0.83
	122	3	2.6	0.08
	91.5	4	<1	<1
	73.2	5	<1	<1
4	366	1	24.1	23.8
	183	2	2.01	1.96
	122	3	0.76	0.923
	91.5	4	<1	<1
	73.2	5	<1	<1
5	366	1	26.0	26.1
	183	2	2.18	2.62
	122	3	0.77	0.613
	91.5	4	<1	<1
	73.2	5	<1	<1

* A and B represent harmonic coefficients corresponding to the harmonics given in column 3 [See Equation (8)].

shown in Figs. 4.17 through 4.21.

As can be seen from the figures the temperature record shows a large number of periodicities. This is understandable as the temperature of the stream can be influenced by periodic variations in industrial activity. From the figure certain periodic frequencies can be identified. At all the stations the raw temperature records show definite cyclic events corresponding to periods of 1.10, 3.30, 6.50 and 14 days. The daily period is due to the diurnal cycle that occurs in all streams and estuaries. Periods shorter than this could not be resolved as the sampling interval was chosen as 12 hours. The other periods may be due to some kind of regular discharge causing a thermal disturbance in the stream. Nonlinear interactions between several periodic phenomena or certain complexities inherent in the system may also cause these periodic phenomena. The 3 day period which appears to be the largest one for Stations 1 and 5 could not be identified precisely. Stations 4 and 5 show some additional rather peculiar cyclic events as shown in Figs. 4.20 and 4.21. These events range from a periodicity of 1.1 days to 14 days. A periodic event with periodicity 12.0 days was present at Station 5. Although it is often not possible to explain the cause of each periodic event by examining only the temperature record, this together with a knowledge of local weather information and industrial activity may provide good understanding of thermal conditions in the stream.

The crosscovariance between temperature records measured

at two different locations was calculated for half day lag periods from -75 to 75. All the crosscovariances from the base station to other stations were positive. As one proceeds further away from the base station, the crosscovariance decreases showing decreasing dependence on the temperature at the base station. Crosscovariance values were positive for both positive and negative lags. A Positive correlation indicates that an above normal temperature can be predicted for all stations if the temperature at the base station is above normal. The cross-covariance shows some kind of a maxima between lags of -30 and 0. This phenomena is not unnatural as the temperature at the base station may be well correlated with the temperature at other stations with negative lags. These crosscorrelations are shown in Figs. 4.22 and 4.23.

Analysis of Dissolved Oxygen Records

Data for dissolved oxygen were obtained for five stations on the Ohio River. The sampling interval was 12 hours and the duration for this record was one year. This record also shows several periodicities.

Autocorrelation values (autocovariances) for DO records for all the five stations were calculated; these results are plotted in Figs. 4.24 and 4.25. All the autocorrelation values are positive and show a similar decreasing trend and decrease with increasing lag. The slope of these curves, however, decreases with lag and approaches zero at higher lags. All autocorrela-tion values are positive indicating that if the DO is below the

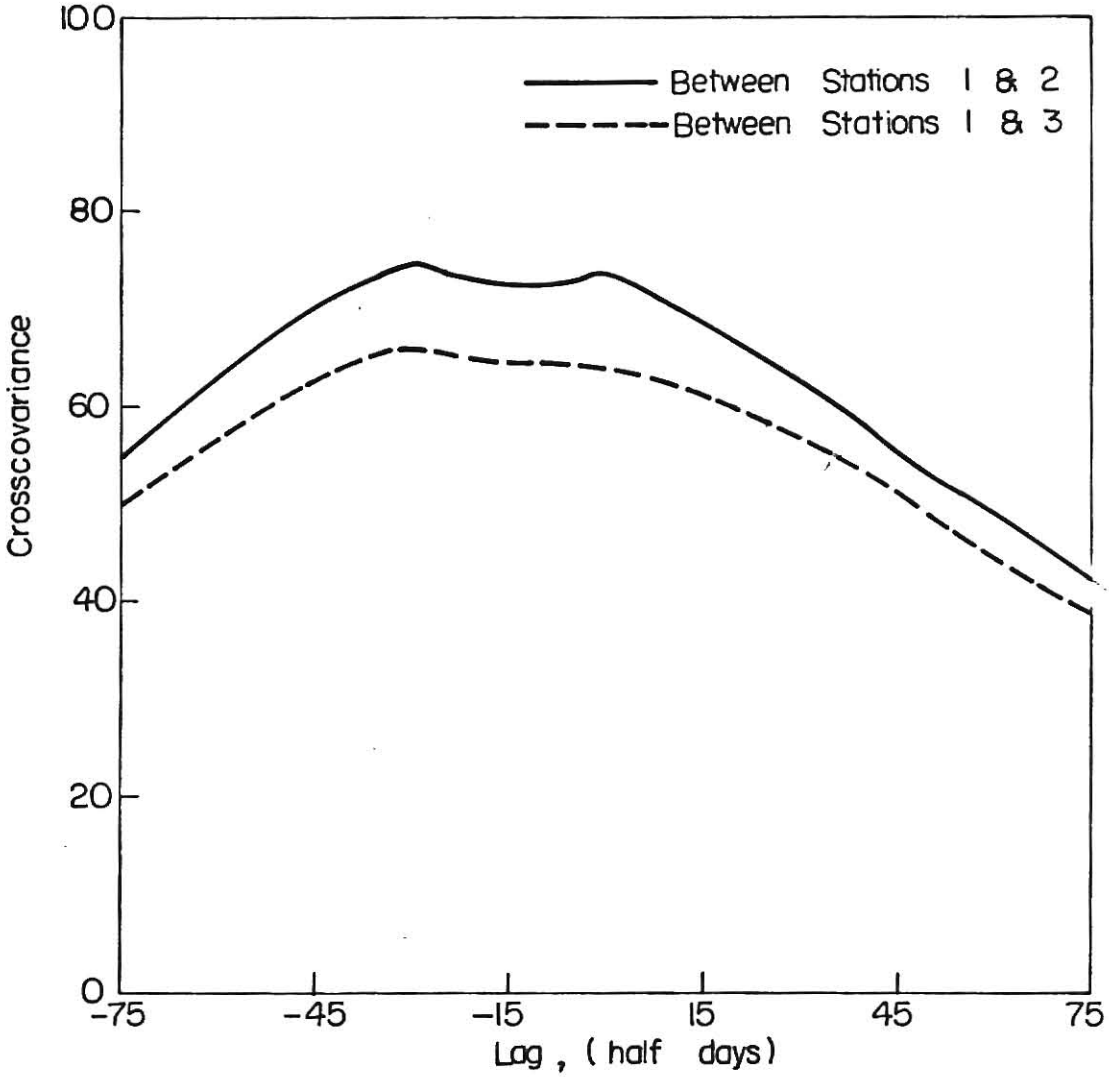


Fig.4.22 Crosscovariance of temperature records between Stations 1 and 2 and Stations 1 and 3, Ohio River, 1968.

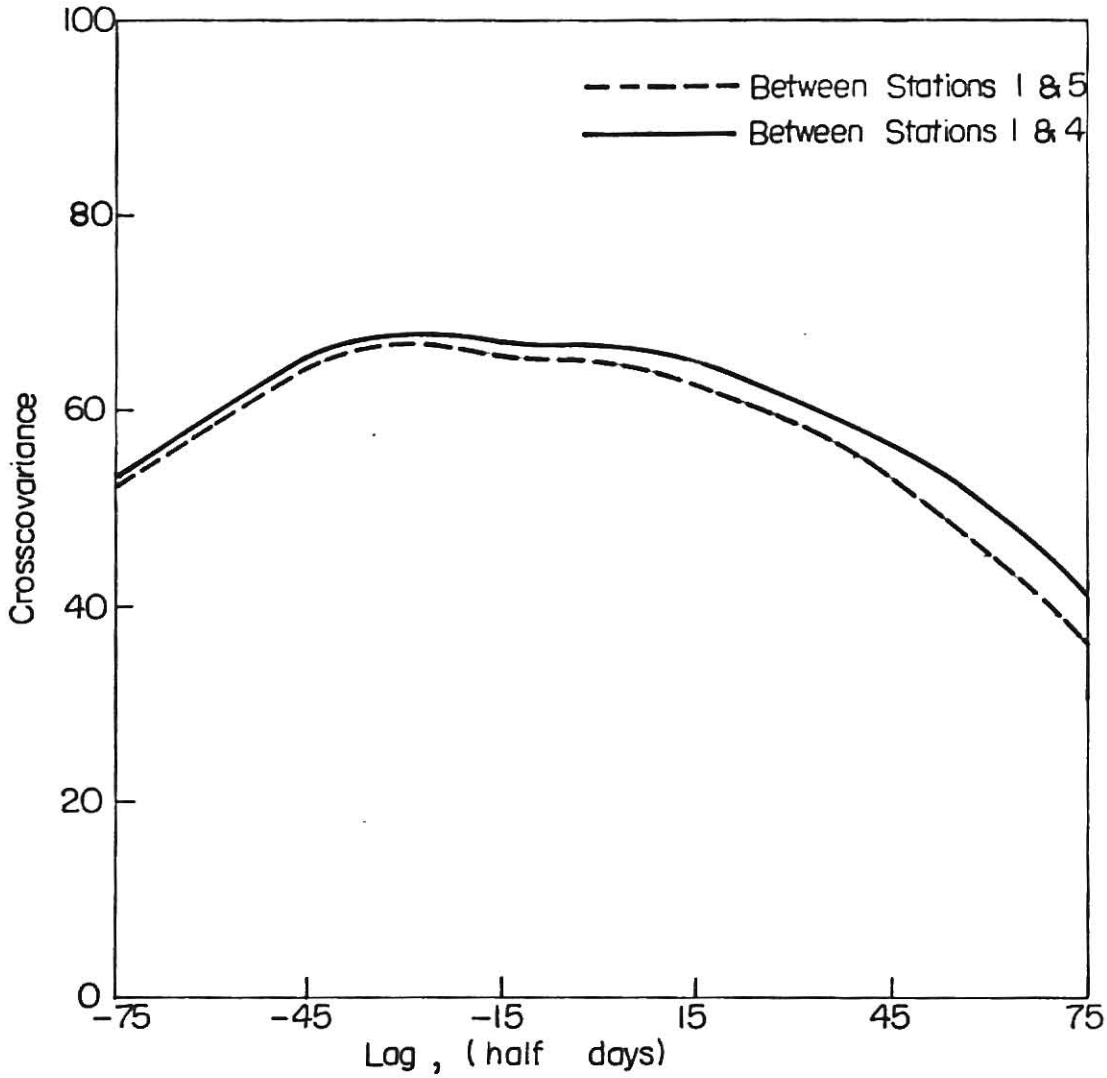
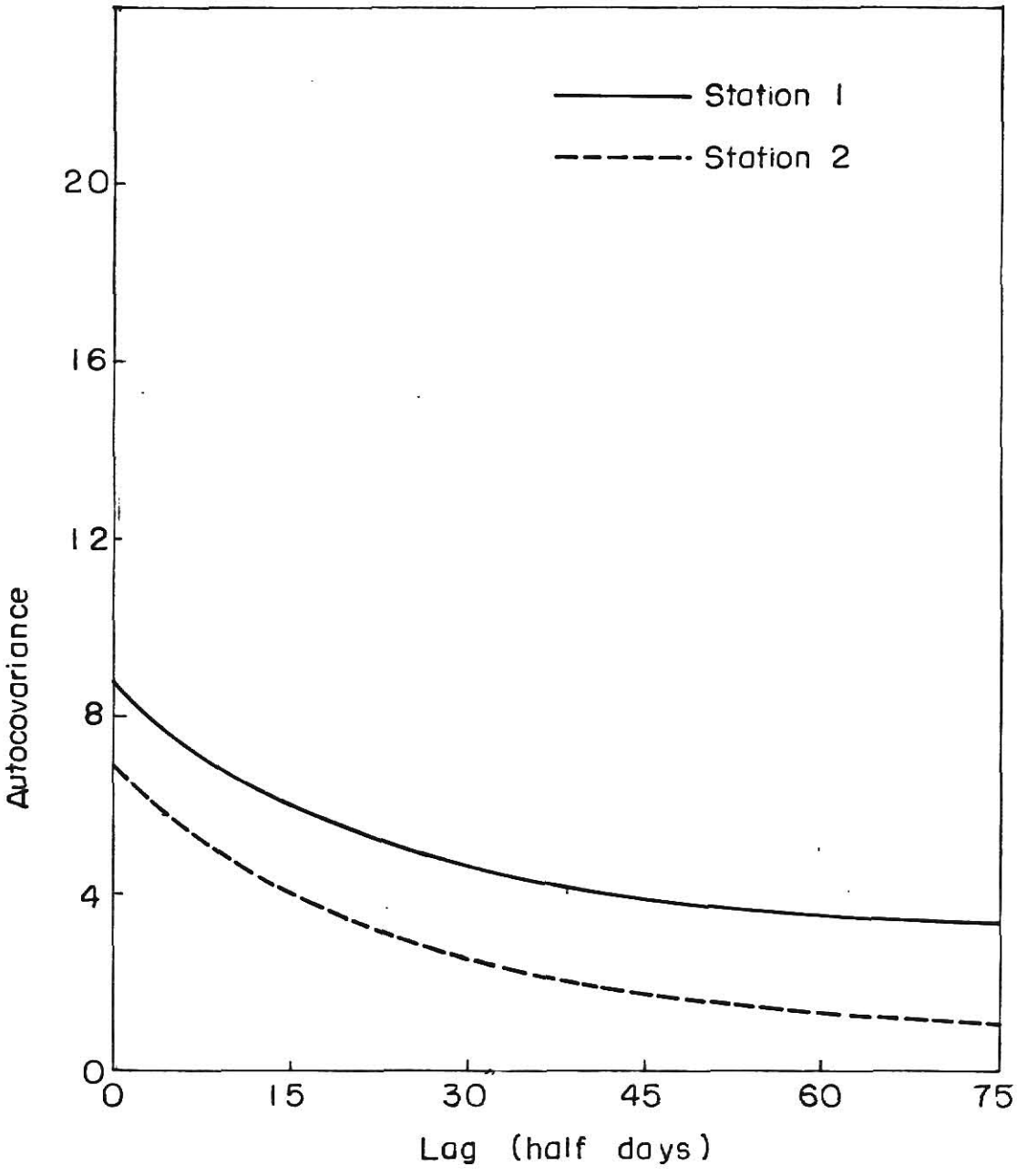


Fig.4.23 Crosscovariance of temperature records between Stations 1 and 4 and Stations 1 and 5 , Ohio River, 1968.

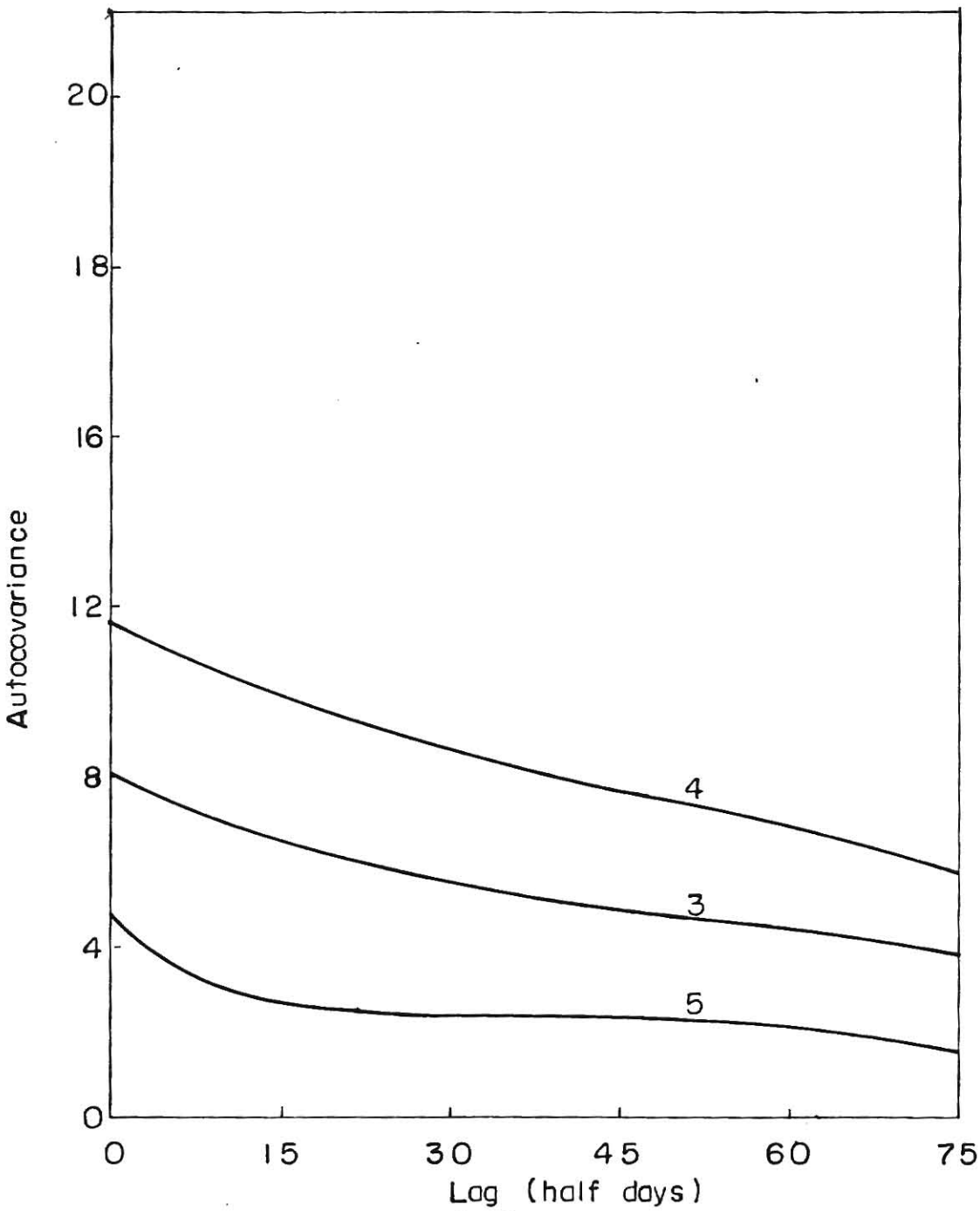
ILLEGIBLE

THE FOLLOWING DOCUMENT (S) IS ILLEGIBLE DUE TO THE PRINTING ON THE ORIGINAL BEING CUT OFF

ILLEGIBLE



g. 4.24 Autocovariance of dissolved oxygen records at Stations 1 and 2. Ohio River (1968).



g.4.25 Autocovariance of dissolved oxygen records at Stations 3, 4, and 5. Ohio River (1968).

critical value today it may remain below this value for sometime. This factor may be quite important in designing a water pollution control program.

The autocorrelation at Station 5 is considerably lower than the autocorrelation values at the other stations.

Power spectral estimates of the dissolved oxygen records are shown in Figs. 4.26 thru 4.30. The spectral estimate for Station 1, for example, shows a cyclic event with a period of 10 days (0.05 cycles/half day). This long period (10 days) may be due to interactions among the variation of the saturation value of DO with temperature periodicities, the variation in BOD due to periodic discharges, and the biological growth processes. This 10-day period may also be due to a combination of several smaller periods. At Station 2, for example, there are smaller periodicities present corresponding to the frequencies of 0.2 and 0.35 cycles/day. The five day period (0.2 cycles/day) and the smaller period (0.35 cycles/day) may be due to some industrial discharge. Station 2, also exhibits a periodic event corresponding to a frequency of 0.88 cycles/day. This period (1.15 days) is also present at Stations 4 and 5.

Several smaller periods are present at all stations; these are difficult to interpret. Some of the minor periods may be due to interactions of several phenomena.

The crosscivariance with respect to position for dissolved oxygen at different monitoring stations shows certain interesting results. As shown in Fig. 4.31 the DO records at Stations 2, 3,

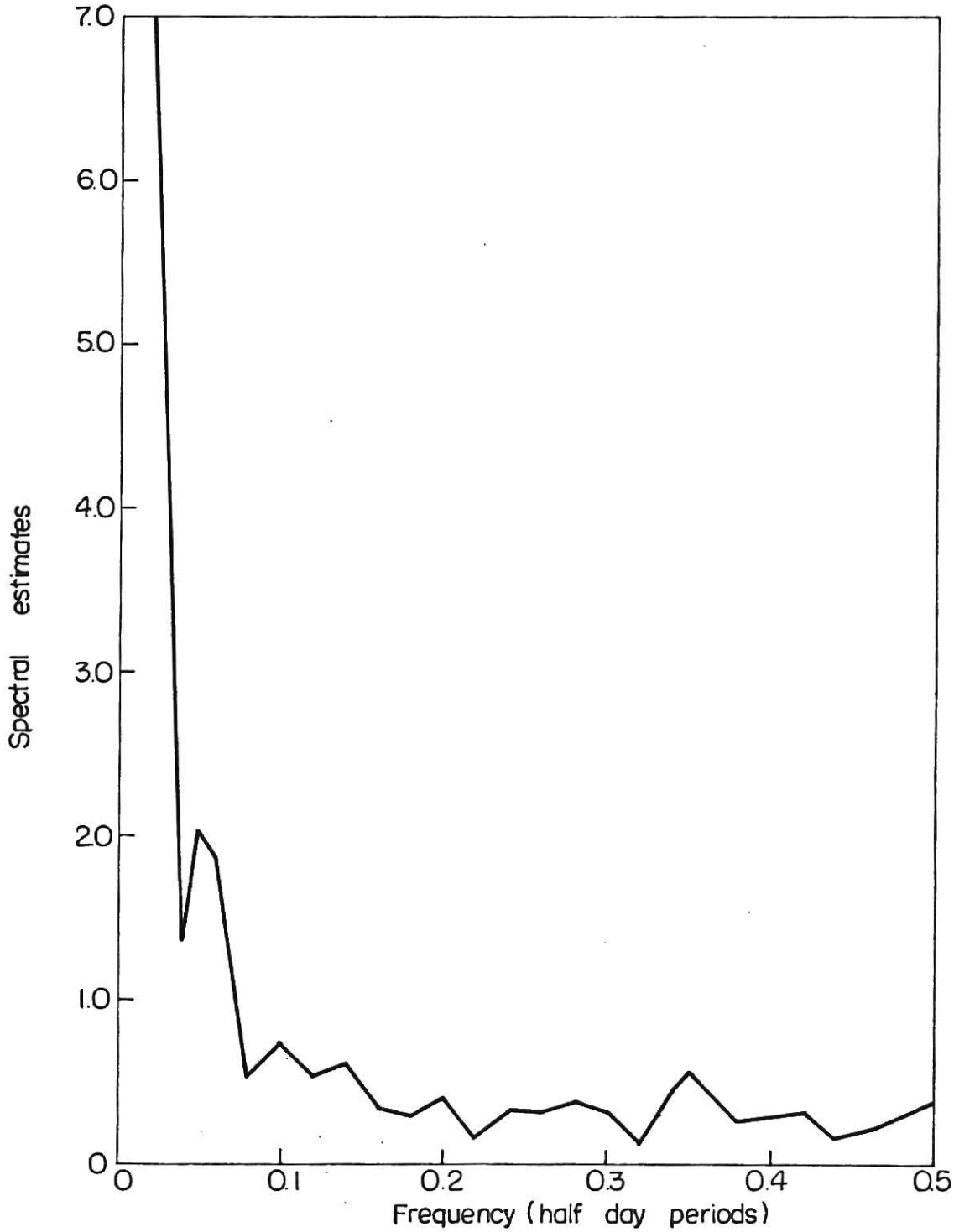


Fig. 4.26 Spectral estimates of dissolved oxygen record at Station 1, Ohio River, 1968.

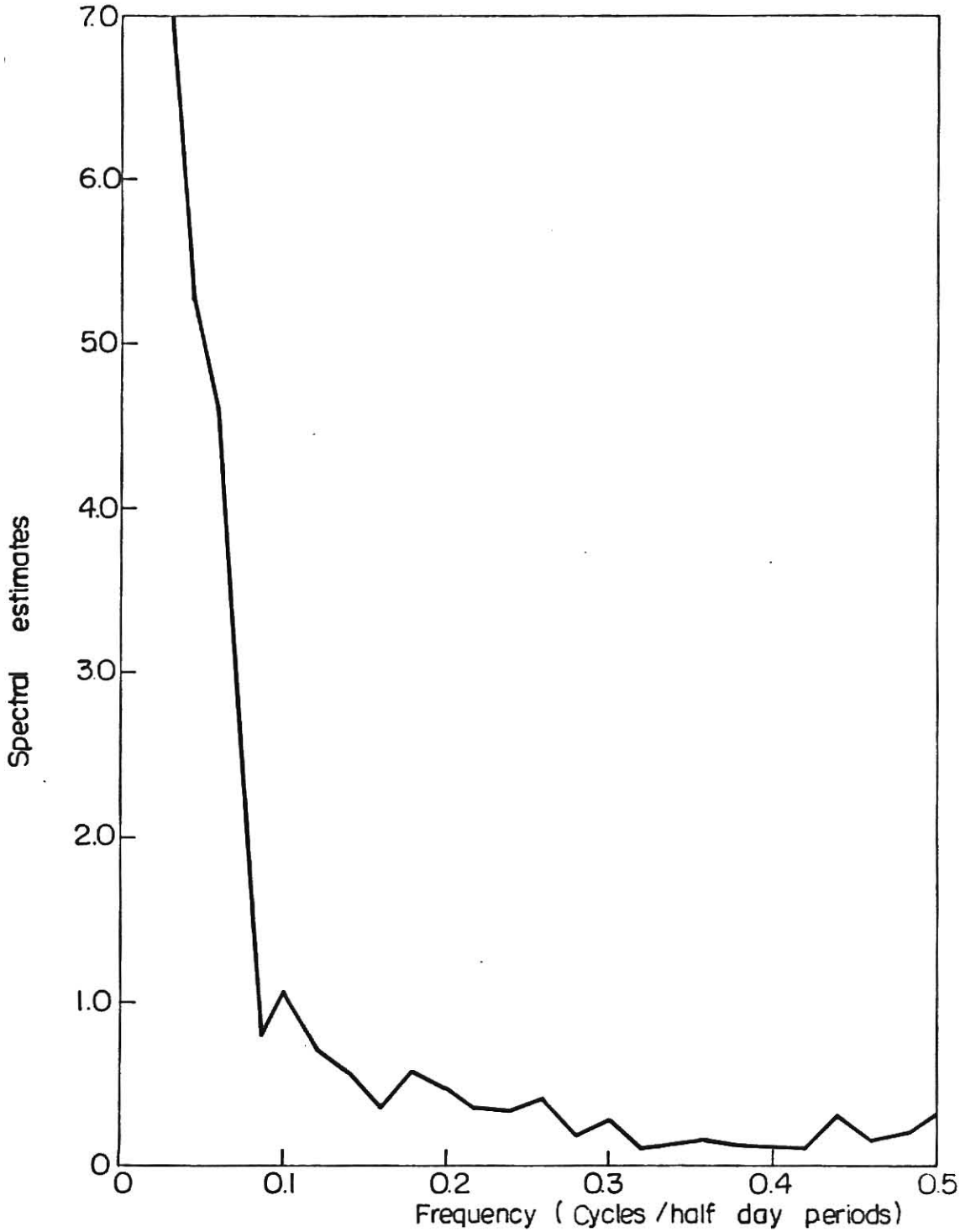


Fig. 4.27 Spectral estimates of dissolved oxygen record at Station 2, Ohio River, 1968

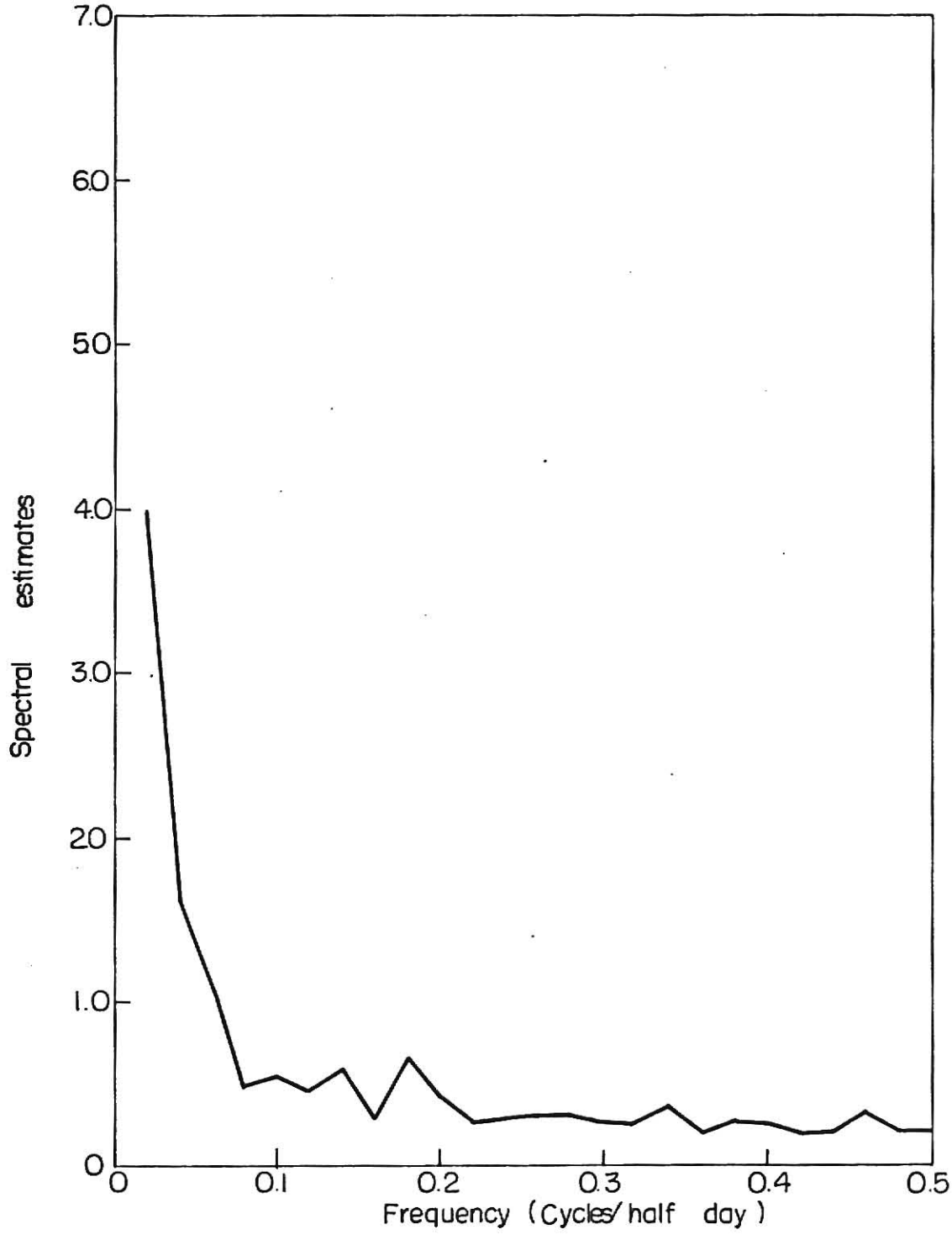


Fig. 4.28 Spectral estimates of dissolved oxygen record at Station 3, Ohio River, 1968.

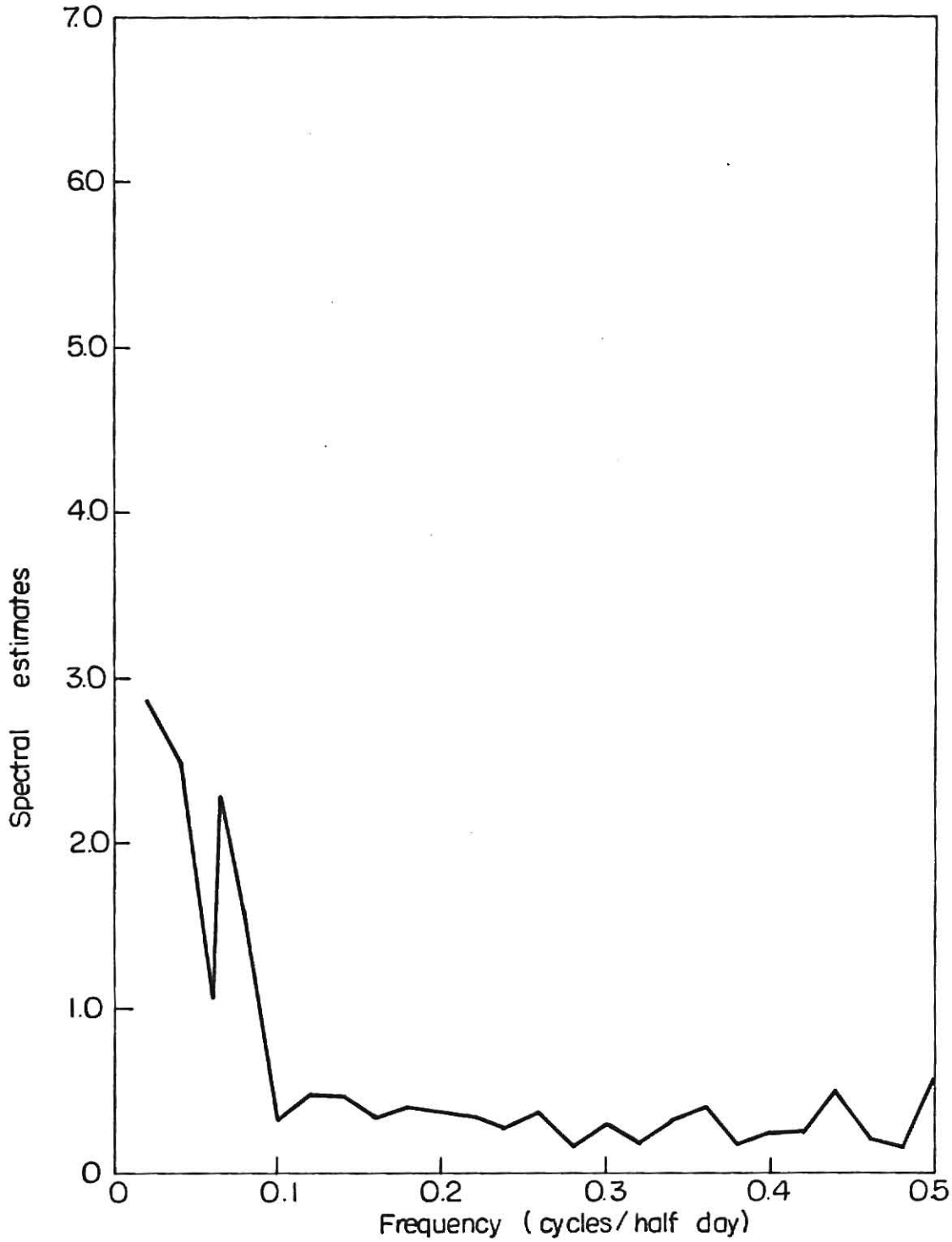


Fig. 4.29 Spectral estimates of dissolved oxygen record at Station 4, Ohio River (1968).

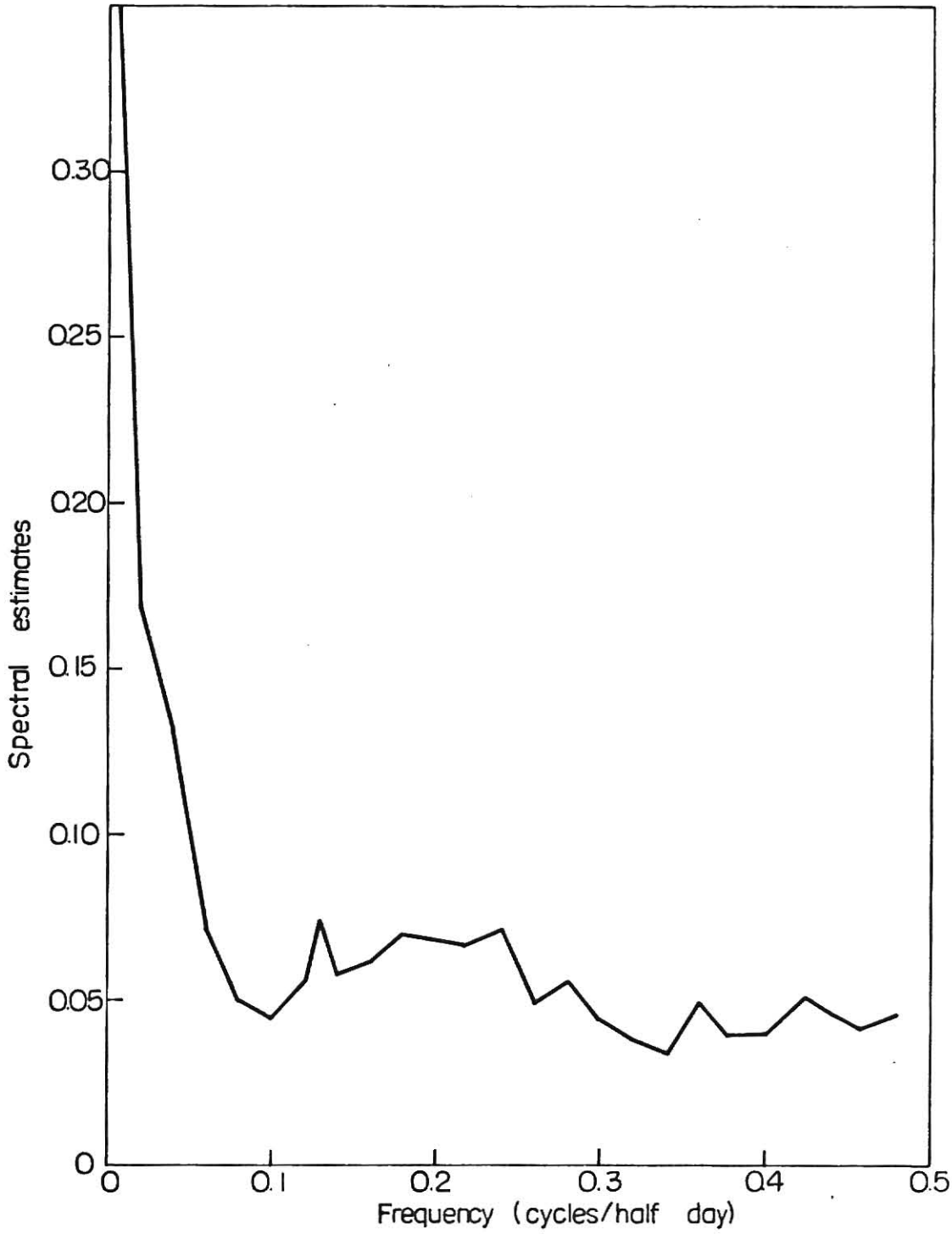


Fig. 4.30 Spectral estimates of dissolved oxygen record at Station 5, Ohio River (1968).

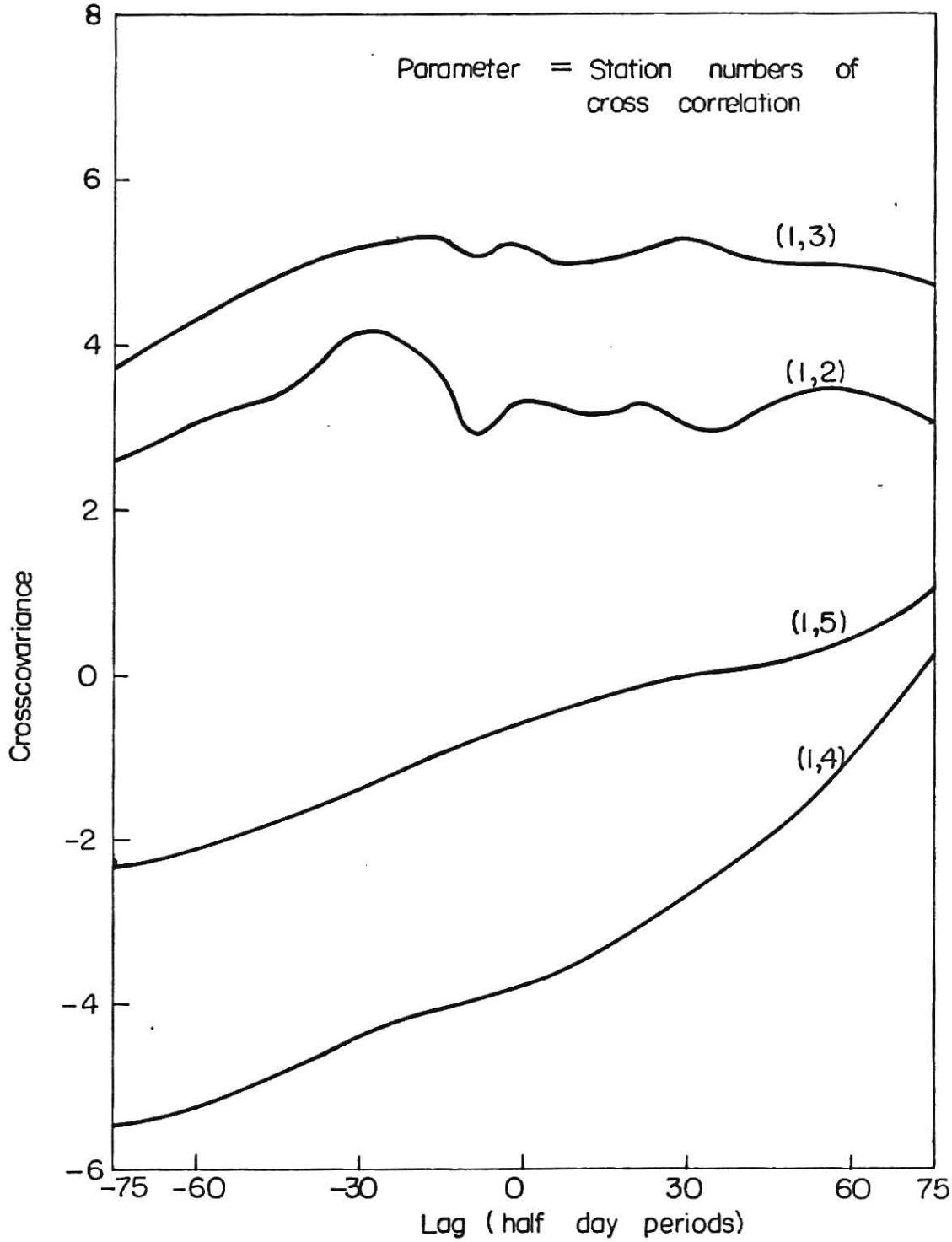


Fig. 4.31 Crosscovariance of dissolved oxygen records between Station 1 and Stations 2,3,4, and 5, respectively, Ohio River, 1968.

4 and 5 were respectively correlated with the dissolved oxygen values at Station 1. The crosscovariance between DO values at Stations 1 and 2 and 1 and 3 are positive and are more than 2.0. The crosscovariance between Stations 1 and 3 is generally higher than the crosscovariance between 1 and 2. This indicates that Stations 1 and 3 are more positively correlated than Stations 1 and 2. Stations 4 and 5 show a negative correlation with Station 1 with Station 4 being more negatively correlated than Station 5. Although one would not expect to find a negative correlation such as this, there may be good reason for it. For example, runoff from rural areas which is very dependent on local weather conditions may differ considerably from Station 1 to Station 4 in its seasonal pattern. The fact that the slope is relatively constant for Stations 4 and 5 in Fig. 4.31 suggests that seasonal changes may be at least partly responsible for these results.

VARIABLE CORRELATION OF OHIO RIVER DATA

In this section the correlation between two water quality parameters measured at the same location is considered. The results are presented in graphical form and where possible physically interpreted.

Temperature and Biochemical Oxygen Demand

It is known that an increase in temperature often results in an increase in the BOD level in the stream. To verify this Temperature and BOD were correlated at all five stations on the

stream and compared. For all stations these two water quality parameters show a positive correlation. The relative magnitudes of these crosscovariances can be obtained by comparing Figs. 4.32 and 4.33. The crosscorrelation function is maximum at Station 3 and decreases slightly as one proceeds in either direction from this station. All values are positive indicating that a higher temperature means higher BOD and, therefore, a higher state of pollution. The results for Station 5 appear to indicate that a change in temperature may also lead to a corresponding change in BOD at a later time. This type of result would be expected when temperatures decrease as the winter season approaches.

Coherence square functions for some of the stations are presented in Table 4.11. The coherence square is a periodic function of frequency and ranges from 0.075 to 0.90. This result means that the relationship between temperature and BOD is not linear.

Temperature and Dissolved Oxygen

It is common knowledge that the solubility of a gas decreases with increasing temperature. The conditions in the stream are so complex that this may not be the dominant effect. The temperature and dissolved oxygen correlations are shown in Fig. 4.34. For South Heights (Station 1), the correlation is negative as expected and temperature and DO have an inverse relationship. For all other stations, however, the cross-covariance reaches a positive value, especially for positive

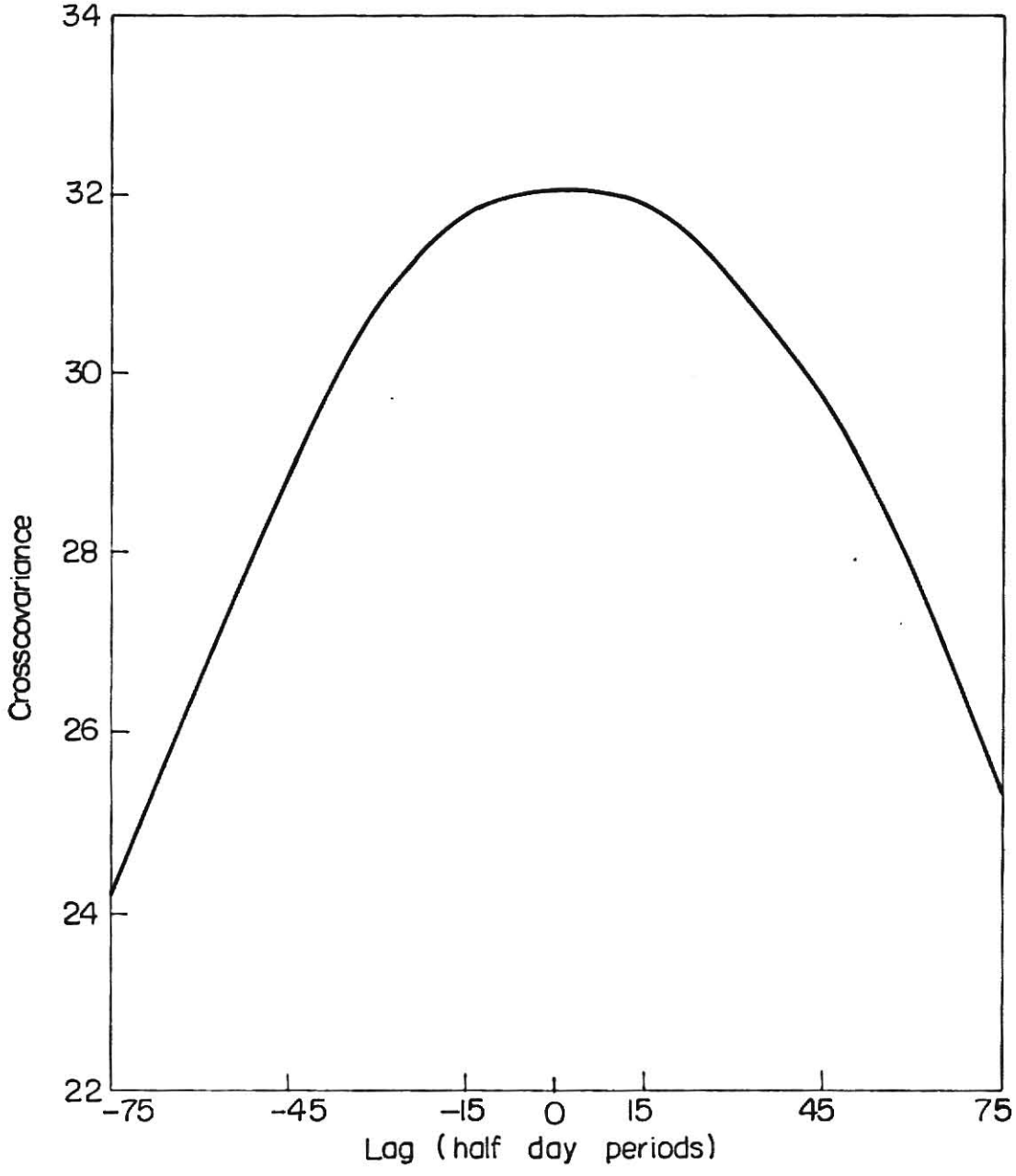


Fig. 4.32. Crosscovariance between temperature and BOD at Stations 1, Ohio River, 1968.

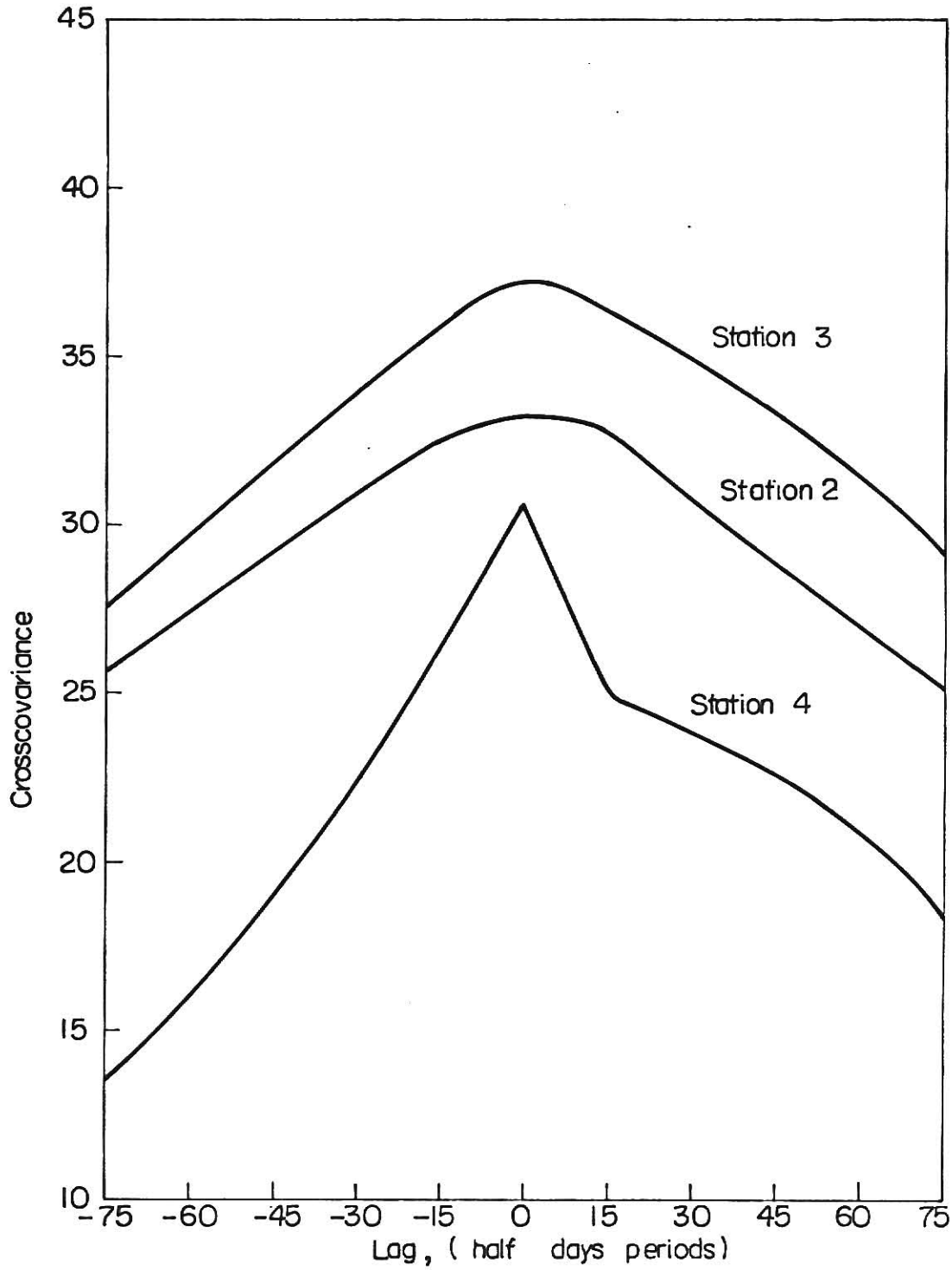


Fig. 4.32 Crosscovariance between temperature and BOD at Station 2, 3 and 4, Ohio River, 1968.

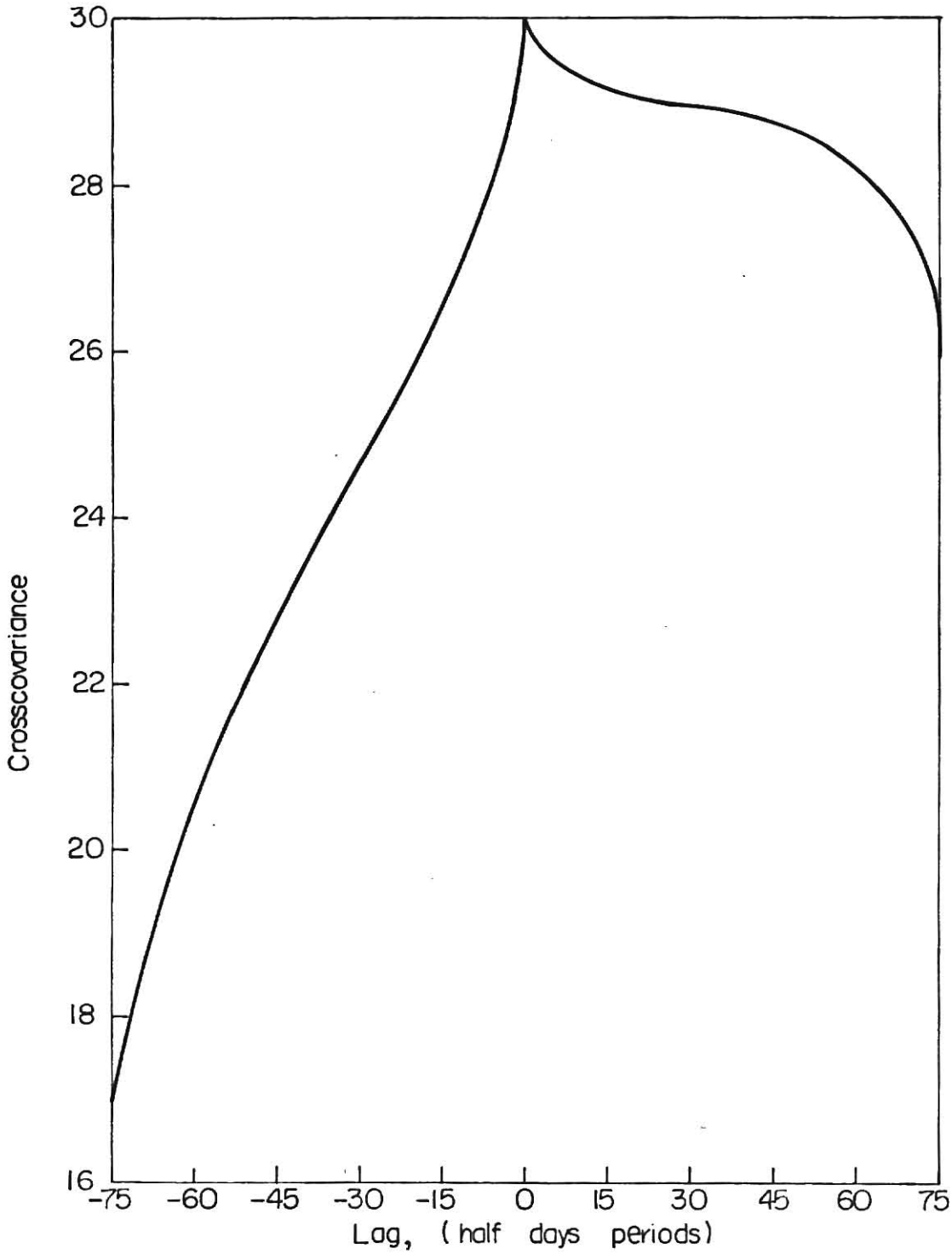


Fig. 4.33 Crosscorariance between temperature and BOD at Station 5, Ohio River, 1968.

Table 4.11 Coherence Square for Temperature and BOD Correlation.

Frequency (Cycles/ halfday)	Coherence Square		
	Station 2	Station 3	Station 5
0.0	0.4518	0.9305	0.0688
0.0067	0.4460	0.8820	0.0961
0.0133	0.2554	0.4467	0.0043
0.0200	0.4558	0.1821	0.0965
0.0267	0.2107	0.1510	0.2608
0.0333	0.9736	0.1761	0.0660
0.0400	0.7917	0.3726	0.2731
0.0467	0.6847	0.4770	0.4206
0.0533	0.6495	0.2536	0.1528
0.0600	0.9426	0.2462	0.4502
0.0667	0.5081	0.0466	0.0936
0.0733	0.7759	0.1585	0.2107
0.0800	0.9498	0.1364	0.1998
0.0867	0.6211	0.0547	0.2932
0.0933	0.0430	0.1564	0.1917
0.1000	0.5712	0.1209	0.1402
0.1067	1.1918	0.3121	0.2431
0.1133	0.1474	0.2068	0.3306
0.1200	0.7388	0.2308	0.1772
0.1267	0.2843	0.1904	0.1595
0.1333	0.4330	0.1927	0.5207
0.1400	0.3404	0.1242	0.0986
0.1467	0.3135	0.0353	0.0441
0.1533	0.9196	0.0470	0.0494
0.1600	1.0799	0.0526	0.9243
0.1667	0.3010	0.1506	0.0245
0.1733	0.8928	0.1052	0.1953
0.1800	0.3407	0.1424	0.2020
0.1867	0.5627	0.1286	0.0284
0.1933	0.2139	0.1458	0.4635
0.2000	0.3845	0.0749	0.0204
0.2067	0.3157	0.1480	0.8972
0.2133	0.1726	0.0688	0.0554
0.2200	0.3317	0.0808	0.4117
0.2267	0.1459	0.0911	0.1866
0.2333	0.0468	0.2071	0.3926
0.2400	0.4834	0.2209	0.3901
0.2467	0.3027	0.1718	0.1042
0.2533	0.3958	0.0717	0.1246
0.2600	0.3539	0.0912	0.1309
0.2667	0.5649	0.0148	0.1727
0.2733	0.5749	0.0641	0.2726
0.2800	0.3541	0.0259	0.1292

Table 4.11 (Cont'd)

Frequency (Cycles/ halfday)	Coherence Square		
	Station 2	Station 3	Station 5
0.2867	0.2534	0.1026	0.1318
0.2933	0.4174	0.1337	0.0968
0.3000	0.4107	0.0848	0.1028
0.3067	0.6678	0.0185	0.2679
0.3133	0.3553	0.0032	0.0035
0.3200	0.5581	0.0016	0.2358
0.3267	0.6042	0.0572	0.4398
0.3333	0.7584	0.1191	0.0440
0.3400	0.7740	0.0520	0.0268
0.3467	0.7854	0.0449	0.1159
0.3533	0.5732	0.2508	0.0956
0.3600	0.6325	0.2475	0.1016
0.3667	0.5069	0.1784	0.3411
0.3733	0.2316	0.1296	0.0954
0.3800	0.3067	0.2139	0.1272
0.3867	0.5088	0.2004	0.1810
0.3933	0.2594	0.0142	0.0529
0.4000	0.3673	0.0561	0.0891
0.4067	0.4630	0.2935	0.1640
0.4133	0.4222	0.3767	0.2163
0.4200	0.1193	0.1856	0.0344
0.4267	0.1088	0.1205	0.0903
0.4333	0.2612	0.1176	0.0290
0.4400	0.3190	0.0761	0.0196
0.4467	0.2618	0.0092	0.0432
0.4533	0.1132	0.0673	0.0684
0.4600	0.2560	0.0810	0.1529
0.4667	0.4457	0.0860	0.0605
0.4733	0.3471	0.0267	0.0797
0.4800	0.1592	0.1099	0.0217
0.4867	0.2906	0.1269	0.0004
0.4933	0.4710	0.0114	0.0382
0.5000	0.4868	0.0034	0.0407

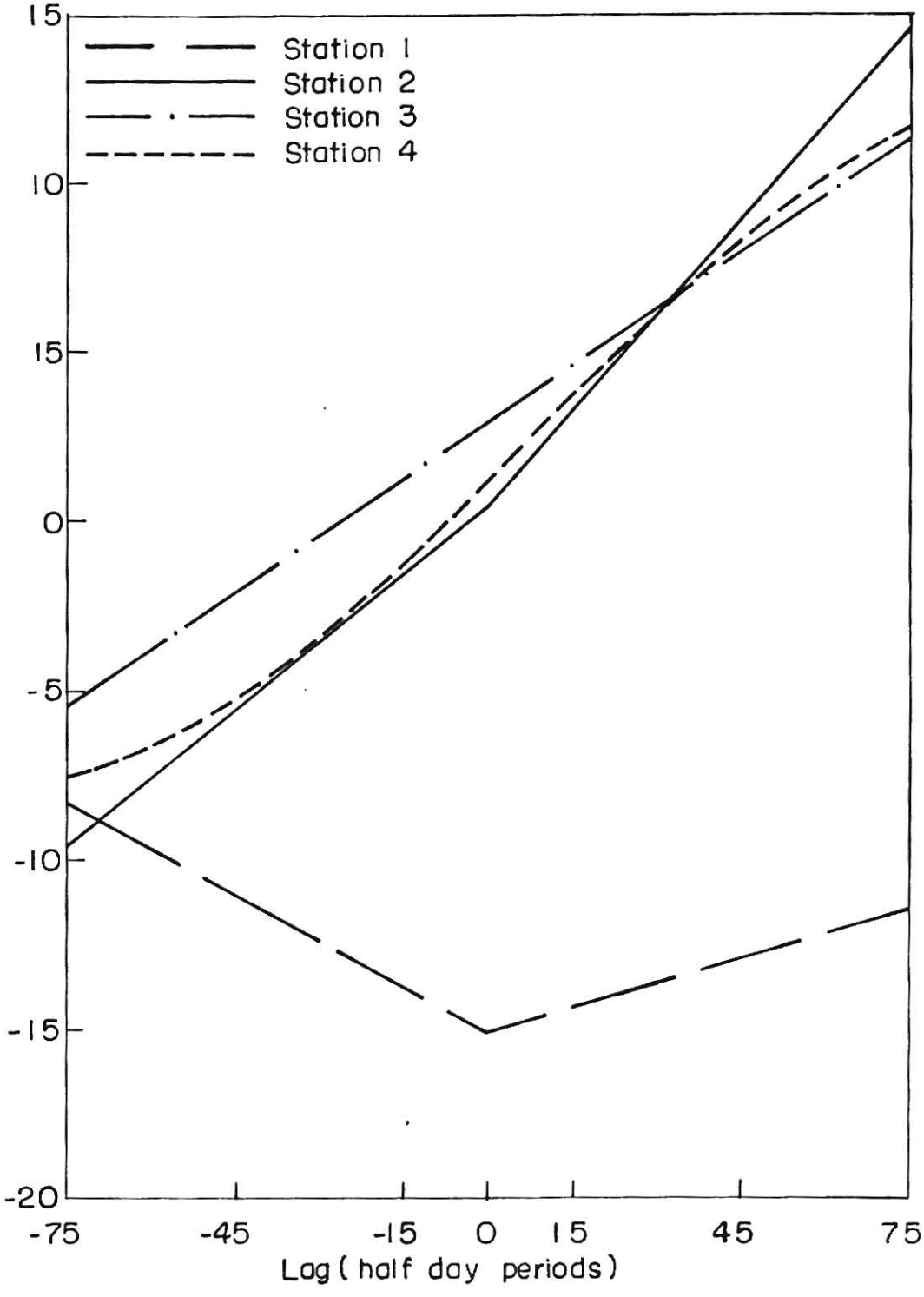


Fig. 4.34 Crosscorrelation between temperature and dissolved oxygen for Stations 1, 2, 3 and 4. Ohio River (1968)

lags. For Stations 2, 3, and 4, the slopes of the curves, which are relatively constant, may suggest a seasonal effect. Stimulation of algae growth and other photosynthetic processes by temperature could account for these results.

Coherence square functions calculated for these two water quality parameters are shown in Table 4.12. These values are scattered and thus, they are difficult to plot;however, from this table the following generalizations can be made.

i) The coherence square function does not show any definite shape.

ii) For Station 1, γ^2 is very high for low frequencies. This drops to a value of 0.142 at a frequency of 0.32 cycles/day and remains below this value except for a small number of fluctuations. This indicates a linear relation at low frequency.

iii) For Station 2 and 3 coherence value ranges from 0.001 to 0.981.

iv) A changing coherence function indicates a nonlinear relationship between dissolved oxygen and temperature.

The amplitude of the transfer function relating dissolved oxygen to temperature is always less than 1. This is a function of frequency as can be seen from Table 4.13; however the amplitude values indicate no definite form for the dissolved oxygen temperature relationship.

It is seen from these two correlations, temperature BOD and

Table 4.12 Coherence Square between DO and Temperature.

Frequency (Cycles/day)	Coherence Square	
	Station 2	Station 5
0.0	0.0340	0.8999
0.0067	0.1069	0.8202
0.0133	0.1686	0.0990
0.0200	0.0607	0.0922
0.0267	0.0594	0.3364
0.0333	0.2113	0.2757
0.0400	0.3995	0.0877
0.0467	0.1768	0.2131
0.0533	0.2290	0.2620
0.0600	0.2399	0.4830
0.0667	0.0326	0.2478
0.0733	0.2234	0.4337
0.0800	0.2061	0.1689
0.0867	0.6405	0.0112
0.0933	0.7684	0.1226
0.1000	0.2687	0.3286
0.1067	0.3244	0.1756
0.1133	0.6341	0.1678
0.1200	0.2713	0.2015
0.1267	0.3677	0.5624
0.1333	0.0578	0.1315
0.1400	0.0353	0.0530
0.1467	0.2296	0.1364
0.1533	0.0192	0.1050
0.1600	0.0172	0.1110
0.1667	0.2528	0.2169
0.1733	0.0393	0.2608
0.1800	0.0331	0.1544
0.1867	0.3464	0.0591
0.1933	0.1829	0.0565
0.2000	0.0505	0.3207
0.2067	0.1272	0.2428
0.2133	0.0973	0.2210
0.2200	0.5695	0.0583
0.2267	0.1172	0.0070
0.2333	0.2267	0.1004
0.2400	0.1004	0.0107
0.2467	0.0881	0.1007
0.2533	0.0422	0.1930
0.2600	0.0254	0.0696
0.2667	0.2187	0.0623
0.2733	0.2721	0.0134
0.2800	0.0753	0.0036
0.2867	0.1051	0.0130

Table 4.12 (Cont'd)

Frequency (Cycles/day)	Coherence Square	
	Station 2	Station 5
0.2933	0.0110	0.1244
0.3000	0.0836	0.1130
0.3067	0.2263	0.0360
0.3133	0.1081	0.0399
0.3200	0.2299	0.0262
0.3267	0.0233	0.0743
0.3333	0.0436	0.1455
0.3400	0.0362	0.1902
0.3467	0.0922	0.0662
0.3533	0.0478	0.1170
0.3600	0.1191	0.0505
0.3667	0.1424	0.0309
0.3733	0.1392	0.1041
0.3800	0.1213	0.0556
0.3867	0.1191	0.2110
0.3933	0.0544	0.1196
0.4000	0.0285	0.0927
0.4067	0.0617	0.1375
0.4133	0.0492	0.1986
0.4200	0.0132	0.2677
0.4267	0.0483	0.2151
0.4333	0.0336	0.3374
0.4400	0.0362	0.2283
0.4467	0.1928	0.2104
0.4533	0.0382	0.0176
0.4600	0.0017	0.0002
0.4667	0.0184	0.0836
0.4733	0.0934	0.0396
0.4800	0.1313	0.0140
0.4867	0.1271	0.1644
0.4933	0.1788	0.2568
0.5000	0.1773	0.3162

Table 4.13 Amplitude of Transfer Function relating DO to Temperature, Stations 1 and 2.

Frequency (Cycles/day)	Amplitude	
	Station 1	Station 2
0.0	0.0554	0.0505
0.0067	0.1025	0.0658
0.0133	0.1965	0.0295
0.0200	0.1471	0.2675
0.0267	0.1737	0.5790
0.0333	0.2311	0.2851
0.0400	0.2328	0.6642
0.0467	0.2684	0.7891
0.0533	0.4243	0.5082
0.0600	0.3295	0.7643
0.0667	0.0671	0.2956
0.0733	0.3118	0.3476
0.0800	0.1790	0.2679
0.0867	0.2601	0.3029
0.0933	0.3431	0.2879
0.1000	0.2771	0.2870
0.1067	0.3354	0.3414
0.1133	0.5429	0.3308
0.1200	0.2302	0.2181
0.1267	0.2001	0.1716
0.1333	0.2971	0.2418
0.1400	0.2175	0.1150
0.1467	0.3068	0.1005
0.1533	0.1264	0.1047
0.1600	0.3408	0.3677
0.1667	0.2059	0.0440
0.1733	0.3077	0.1564
0.1800	0.1602	0.2263
0.1867	0.2596	0.0824
0.1933	0.1759	0.3879
0.2000	0.3676	0.0898
0.2067	0.2467	0.5443
0.2133	0.2601	0.1173
0.2200	0.1313	0.2518
0.2267	0.1436	0.1570
0.2333	0.2204	0.2629
0.2400	0.2525	0.3355
0.2467	0.2260	0.2985
0.2533	0.2124	0.2879
0.2600	0.2180	0.2631
0.2667	0.2078	0.3198
0.2733	0.2049	0.2867
0.2800	0.1863	0.1384
0.2867	0.0154	0.1139

Table 4.13 (Cont'd)

Frequency (Cycles/day)	Amplitude	
	Station 1	Station 2
0.2933	0.2487	0.1186
0.3000	0.0780	0.0325
0.3067	0.0780	0.0325
0.3133	0.1434	0.2426
0.3200	0.2269	0.2830
0.3267	0.1258	0.0702
0.3333	0.0594	0.0505
0.3400	0.1116	0.1099
0.3467	0.2135	0.1257
0.3533	0.0876	0.1326
0.3600	0.1454	0.1123

temperature-dissolved oxygen, that the variable correlations at all stations are not identical. The temperature BOD correlation curve has the same general shape at all stations, but the temperature dissolved oxygen crosscorrelations show a change in sign and a change in shape indicating a change in the nature of the relationship from station to station.

Dissolved Oxygen and Biochemical Oxygen Demand

Since the oxidation of organic wastes requires oxygen, a continuously high BOD at any point along a stream should cause a reduced dissolved oxygen concentration if microbial consumption of the organic wastes is taking place. A negative crosscovariance between these two parameters is observed for all lags as shown in Fig. 4.35. At other stations the dissolved oxygen-BOD relationship is similar and, therefore, the results are not presented.

The coherence square for the dissolved oxygen-BOD transformation is very low indicating a highly nonlinear relationship. Although the transformation relating DO to BOD cannot be expressed by a linear differential equation in the time domain, the functional form resembles the form for a first order system to some extent as shown in Fig. 4.36. The dimensional amplitude of the transfer function is always less than unity and is a function of the frequency. In Fig. 4.36 time is made dimensionless by employing a time constant of 3 days and the amplitude is made dimensionless by dividing by an amplitude of 1.562.

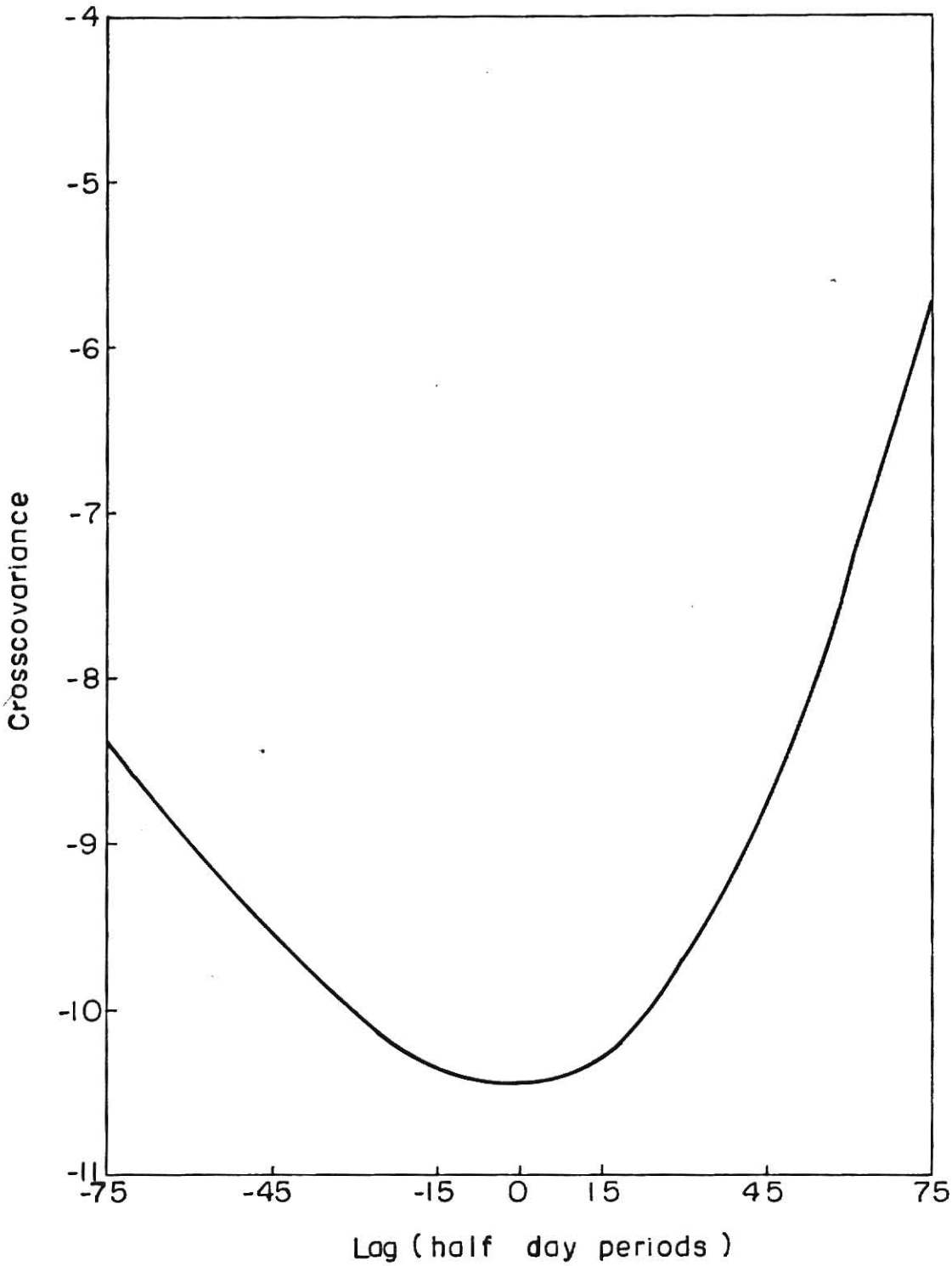
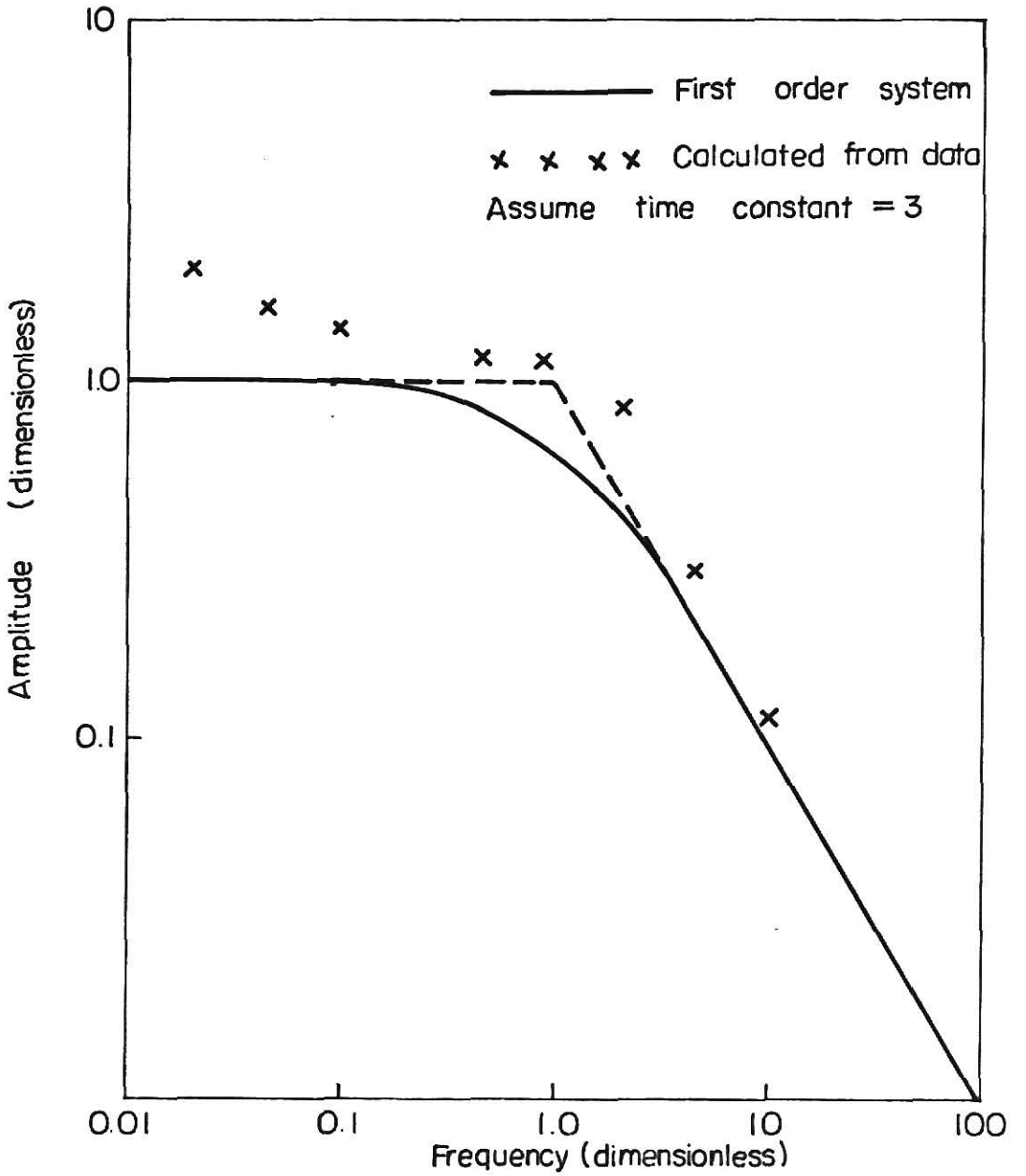


Fig 4·35 Crosscovariance between dissolved oxygen and BOD at Station 2, Ohio River (1968)



Fig·4·36 Comparison of amplitude of dissolved oxygen and BOD transfer function with that for a linear first order system, Station 2, Ohio River (1968)

Some crosscorrelation results between temperature and coliform count are shown in Fig. 4.37. These results and the crosscorrelation results relating dissolved oxygen and coliform are scattered due to the nature of the coliform time series; however, the results do indicate a positive correlation between temperature and coliform concentration.

ANALYSIS FOR OTHER STREAMS

The sampling interval for the data obtained from the Coosa and Detroit Rivers was not small enough to resolve the daily periods. These data were also analyzed using spectral analysis techniques. The results of this analysis is presented in the following sections.

Detroit River

The results for this river are presented in Figs. 4.39 thru 4.42. In Fig. 4.39, temperature, DO and BOD autocovariances are compared. For this stream all the normalized autocovariances fall in the same range. Crosscovariances are plotted in Figs. 4.40 through 4.42. For this river more information on the coliform count was available, and, therefore, the results with coliform should be more meaningful. The scatter in the cross-covariance values however, makes it impossible to plot these results which are shown in Tables 4.14 and 4.15. Crosscovariance values between BOD and coliform count are negative showing a negative relationship. The crosscovariance value tends to decrease with lags, and reaches a minimum of -2976 at a lag of

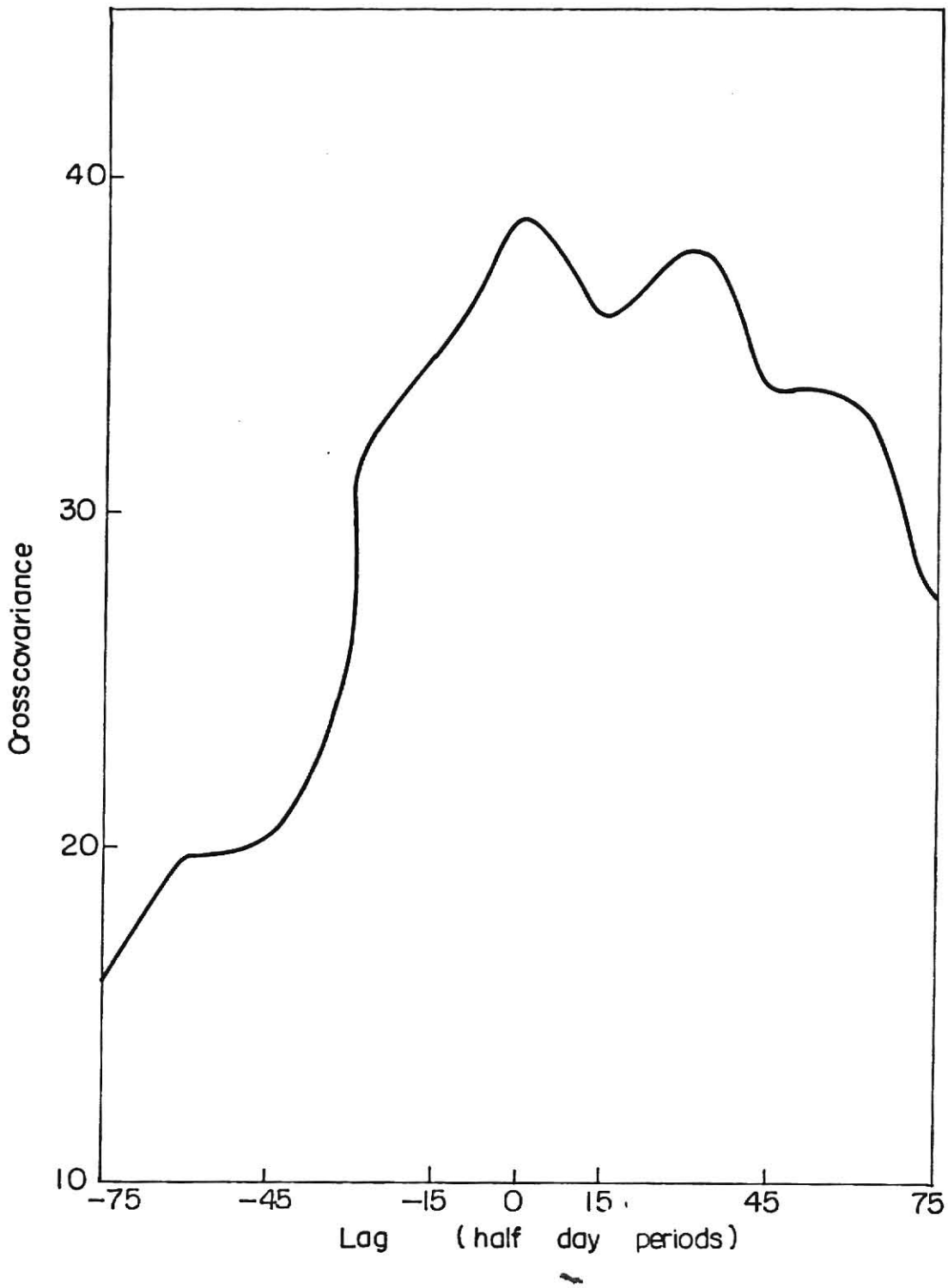
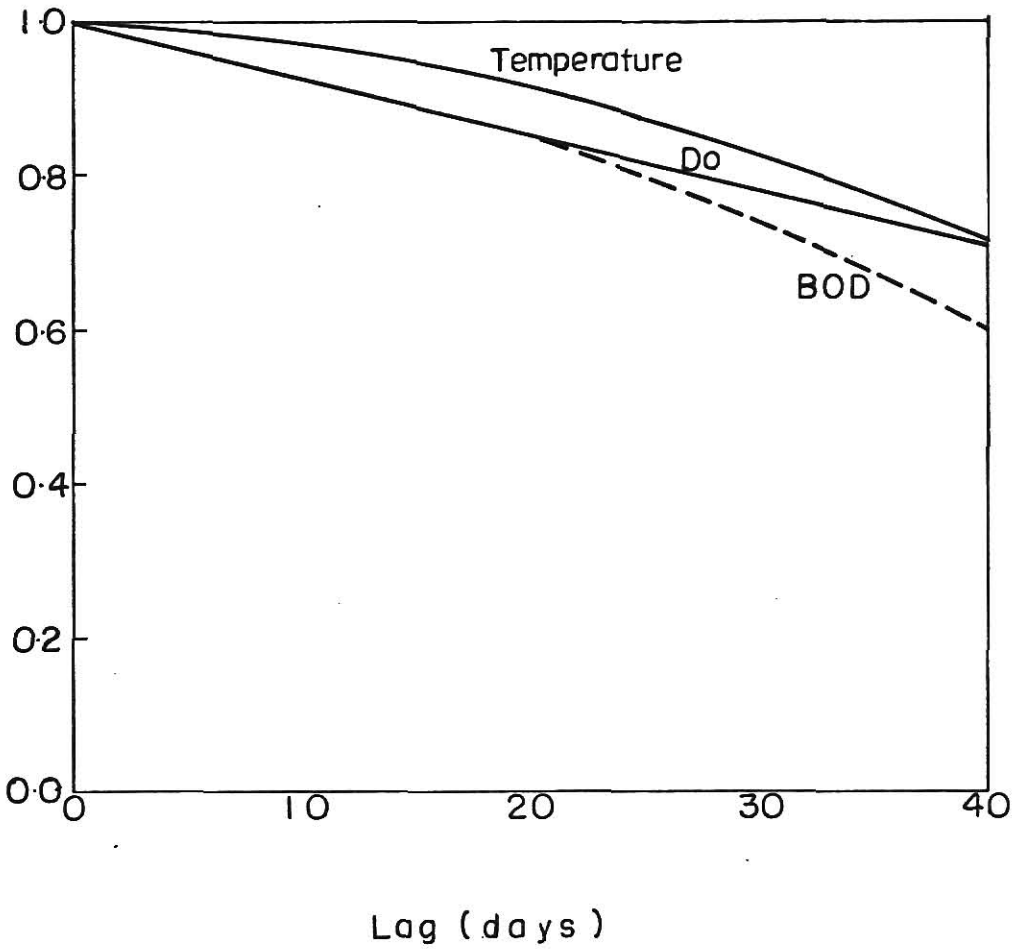
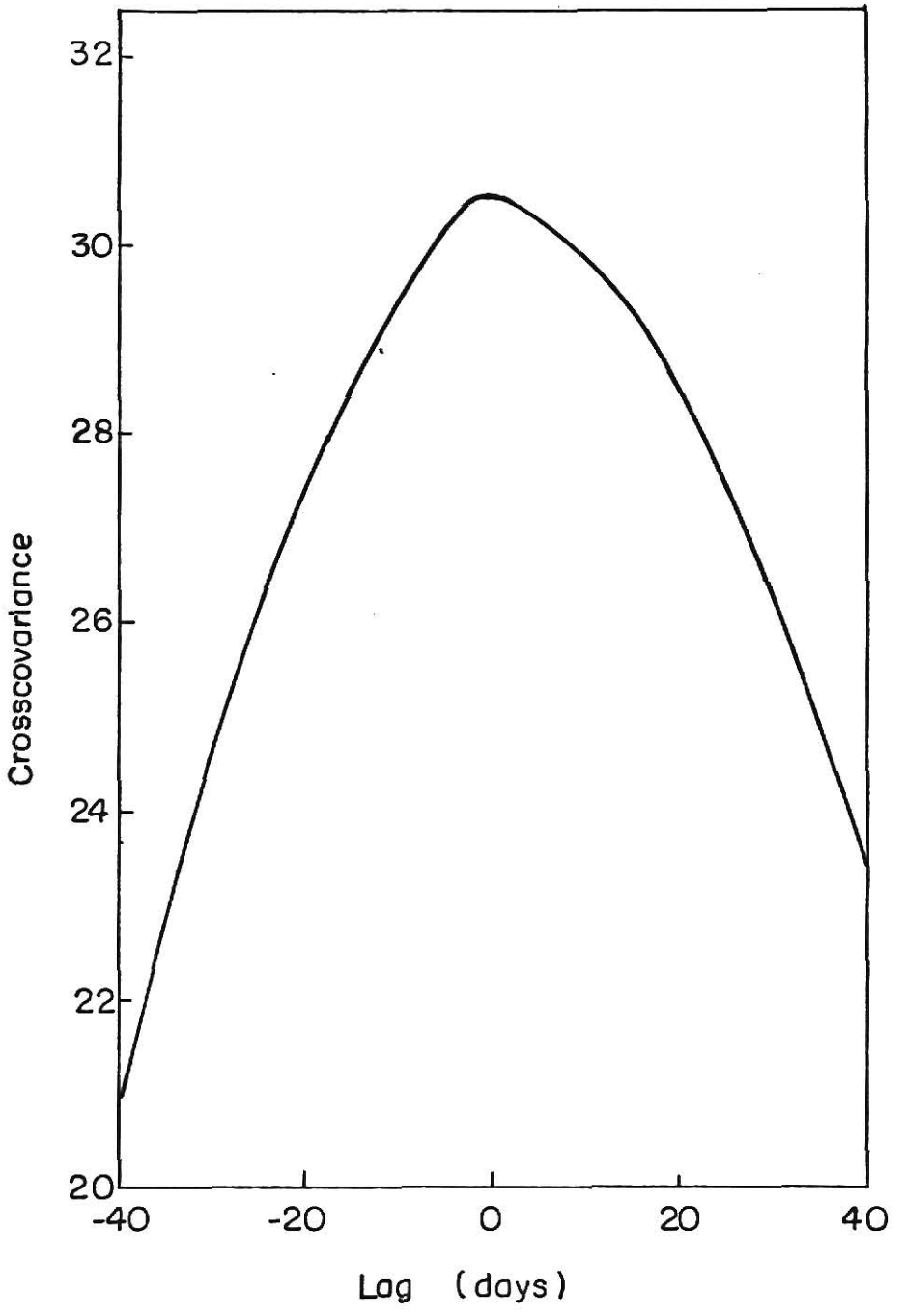


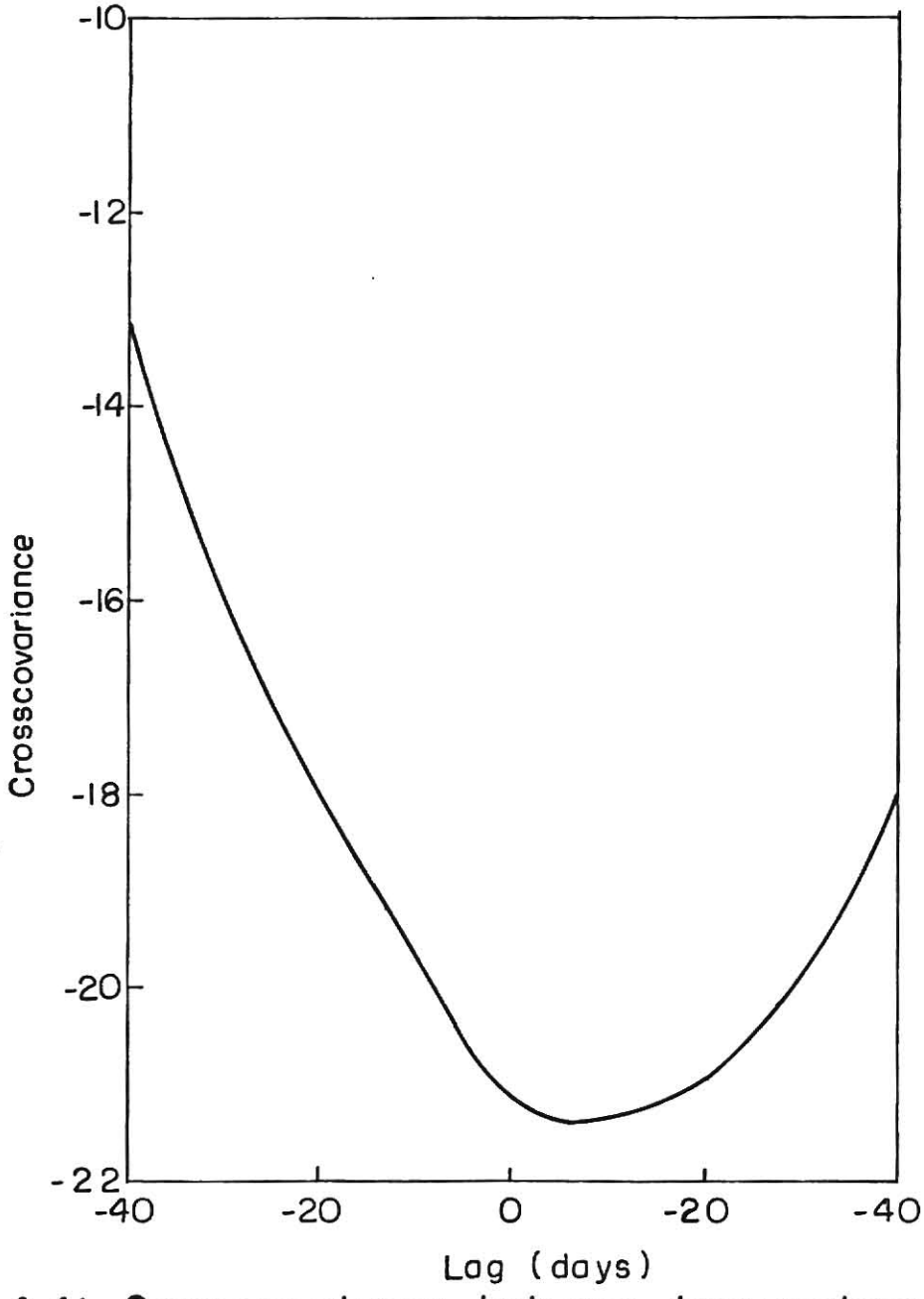
Fig. 4.37 Crosscovariance between temperature and coliform at Station 1, Ohio River (1968).



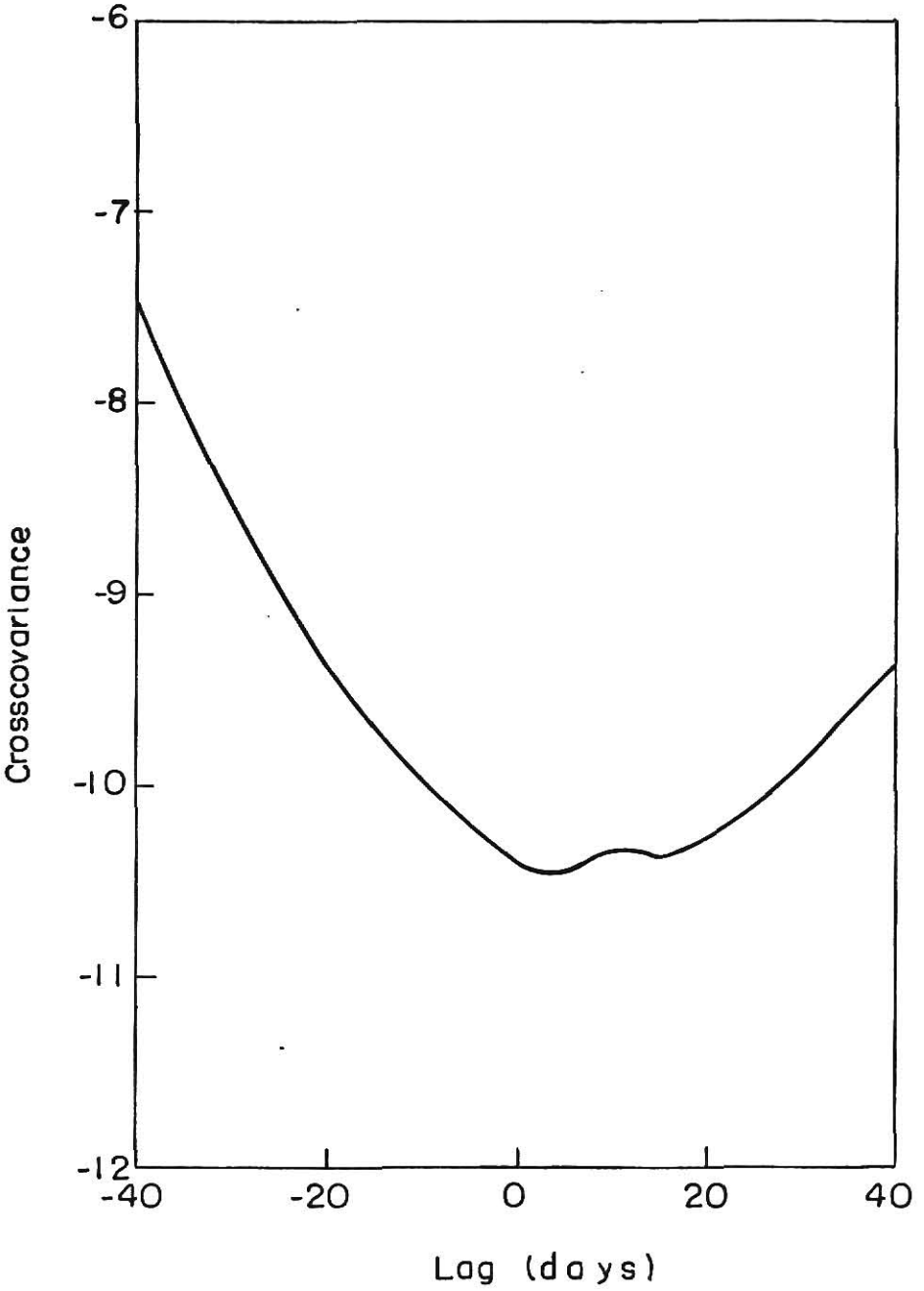
Fig·4·39 Comparison of normalized autocovariance for Detroit River (1965).



Fig·4·40 Crosscovariance between temperature and BOD, Detroit river (1965).



Fig·4·41 Crosscovariance between temperature and dissolved oxygen, Detroit river (1965)



Fig·4·42 Crosscovariance between dissolved oxygen and BOD records, Detroit River(1965)

Table 4.14 Crosscovariance between BOD and Coliforms, Detroit River.

Lags (Days)	Crosscovariance (Positive Lag)	Crosscovariance (Negative Lag)
0	-1652.746	-1652.746
1	-1645.630	-1715.417
2	-1782.123	-1745.831
3	-1742.288	-1805.902
4	-1943.806	-1851.901
5	-1955.629	-1553.277
6	-1960.218	-1838.061
7	-1722.844	-1740.201
8	-1859.088	-1863.898
9	-1971.621	-2058.644
10	-1875.666	-1736.633
11	-2181.879	-1695.930
12	-2131.885	-1529.941
13	-1911.281	-1978.516
14	-1975.816	-1717.103
15	-2117.460	-1861.124
16	-2035.323	-1869.815
17	-1898.292	-1820.069
18	-2095.375	-1775.966
19	-2366.650	-1950.637
20	-2276.294	-2053.647
21	-2219.031	-2135.403
22	-2521.761	-1942.669
23	-2637.032	-2243.978
24	-2585.198	-2178.728
25	-2425.569	-2258.572
26	-2548.340	-2429.654
27	-2608.882	-2315.233
28	-2299.920	-2299.316
29	-2515.219	-2392.449
30	-2538.885	-2497.584
31	-2315.006	-2443.651
32	-2483.529	-2355.712
33	-2658.491	-2318.524
34	-2627.669	-2343.010
35	-2919.403	-2314.225
36	-2849.830	-2286.650
37	-2914.599	-2270.414
38	-2911.360	-2536.166
39	-2973.746	-2404.014
40	-2976.631	-2390.032

Table 4.15 Coherence Square and Amplitude of Transfer Functions Relating Dissolved Oxygen and Temperature, Detroit River.

Frequency (Cycles/days)	Coherence Square	Amplitude
0.0	0.0854	98.4864
0.0125	0.0417	74.1684
0.0250	0.0353	137.7814
0.0375	0.0344	517.2260
0.0500	0.0066	226.1953
0.0625	0.1041	832.3232
0.0750	0.0442	441.9741
0.0875	0.0284	324.8149
0.1000	0.0677	654.0903
0.1125	0.0533	545.6381
0.1250	0.0535	912.5700
0.1375	0.1113	1418.9616
0.1500	0.3175	2589.4128
0.1625	0.1887	1687.5415
0.1750	0.0405	590.5170
0.1875	0.0396	494.9807
0.2000	0.0732	835.0295
0.2125	0.0751	889.5156
0.2250	0.1027	1289.2966
0.2375	0.0029	225.7072
0.2500	0.0940	1261.0937
0.2625	0.0408	683.5441
0.2750	0.1816	1618.2949
0.2875	0.2345	1765.7609
0.3000	0.0004	77.0189
0.3125	0.0722	992.3728
0.3250	0.1769	1909.0903
0.3375	0.1244	1756.8349
0.3500	0.0621	1118.2314
0.3625	0.0309	733.4743
0.3750	0.0144	366.0627
0.3875	0.0383	616.7397
0.4000	0.1995	1556.2570
0.4125	0.1952	1591.8190
0.4250	0.1440	1536.0383
0.4375	0.0264	559.9746
0.4500	0.0780	1070.4306
0.4625	0.1665	2017.3818
0.4750	0.1330	1693.2348
0.4875	0.0735	1218.9082
0.5000	0.0939	2145.4819

40 days. Coherence square between these two variables is very low; it ranges from 0 to 0.32 and is a function of frequency. This indicates a nonlinear relationship between the BOD and coliform count. The amplitude of the transfer function does not show any definite functional form; in fact, these values peak at several frequencies.

Coosa River

For the sampling station on the Coosa river some information about BOD and coliform was available. Some of the data were interpolated by a linear interpolation procedure to generate a continuous record. These data were analyzed in the same manner and the results are presented in Figs. 4.43 through 4.45. In Fig. 4.43, normalized autocorrelations for BOD, DO and temperature are plotted. Unlike the autocovariances for the Detroit River, the autocovariances for DO slope down to a very low value. Extrapolation of this curve shows that the autocovariance for DO on the Coosa River would reach zero for a lag of 52 days, whereas for DO on the Detroit River it would take more than 85 days. The results show a negative relationships between DO and BOD and between DO and Temperature.

The crosscovariance values for a correlation between DO and coliform are all positive indicating (Table 4.16) a positive relationship. This means if dissolved oxygen concentration increases from normal, the coliform count also increases. Coherence square and amplitude of the transfer function from DO to coliform are presented in Table 4.17.

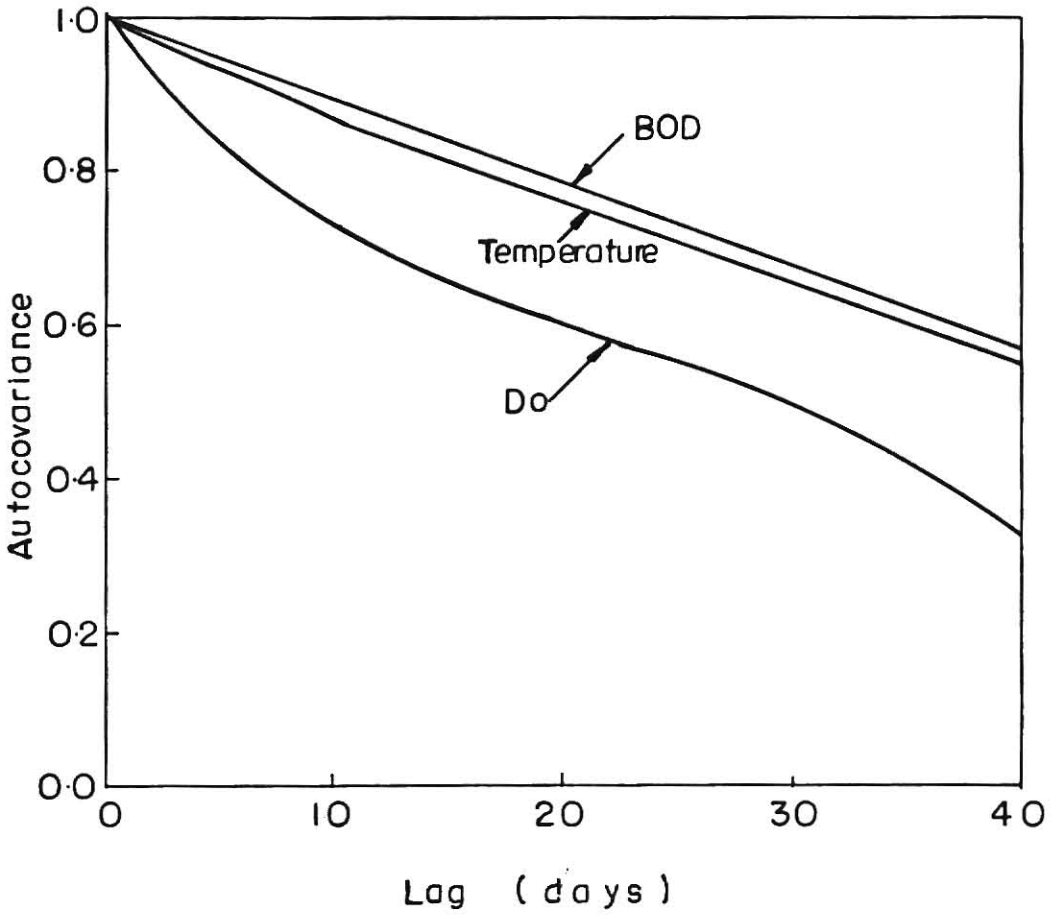
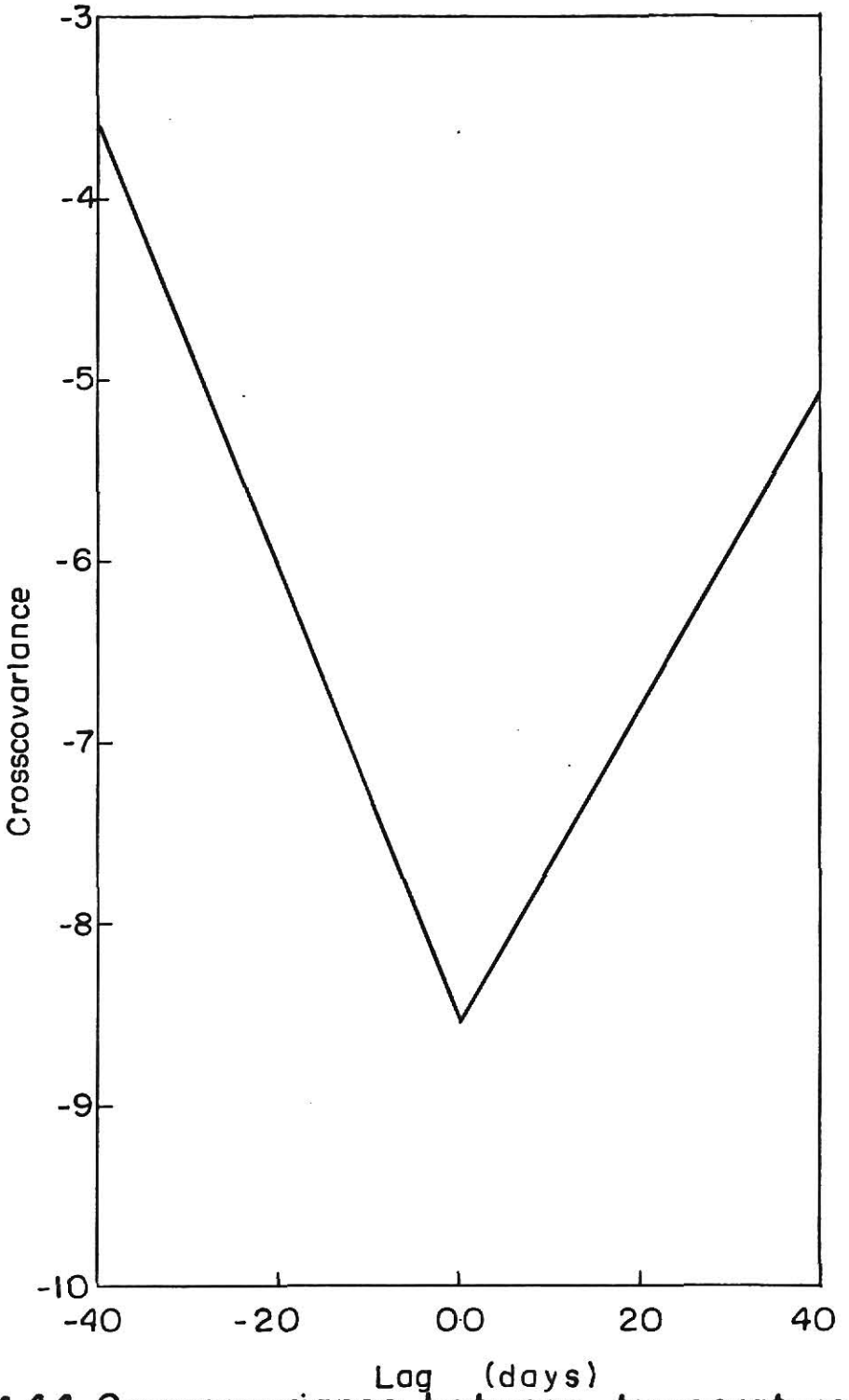


Fig. 4.43 Comparison of normalizd autocovariances for Coosa River (1965).



Fig·4.44 Crosscovariance between temperature and dissolved oxygen, Coosa River (1965)

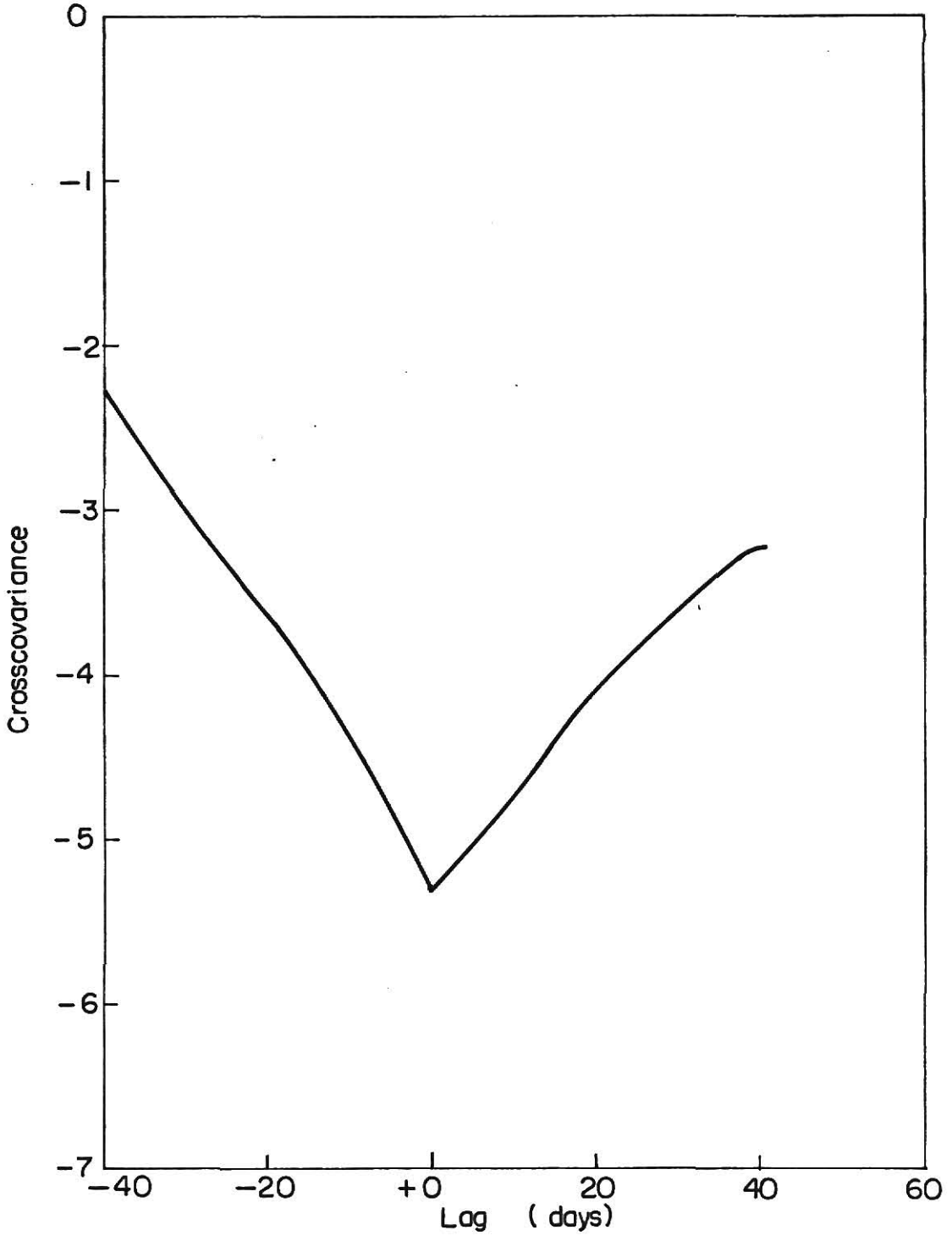


Fig. 4.45. Crosscovariance between BOD and dissolved oxygen, Coosa River (1965).

Table 4.16 Crosscovariance between DO and Coliform, Coosa River.

Lag (Day)	Crosscovariance (Positive Tau)	Crosscovariance (Negative Tau)
0	3953.17	3953.17
1	3924.51	3929.63
2	3886.04	3925.09
3	3753.22	3873.68
4	3747.94	3695.29
5	3624.17	3700.56
6	3543.91	3674.51
7	3391.38	3535.43
8	3417.60	3567.54
9	3466.04	3581.82
10	3366.38	3584.82
11	3226.68	3542.05
12	3296.94	3537.20
13	3218.89	3631.09
14	3212.89	3549.73
15	3222.75	3575.08
16	3215.60	3595.91
17	3236.45	3562.59
18	3265.17	3497.24
19	3215.89	3679.38
20	3069.20	3694.11
21	3136.80	3621.23
22	3183.92	3625.59
23	3101.72	3730.70
24	2947.94	3661.24
25	3837.51	3557.22
26	2784.50	3577.99
27	2771.32	3602.75
28	2700.11	3551.88
29	2795.47	3458.24
30	2642.80	3428.39
31	2569.02	3334.91
32	2453.41	3238.96
33	2519.07	3220.87
34	2549.00	3129.95
35	2603.85	3096.99
36	2694.80	3001.07
37	2696.35	2975.59
38	2611.57	2926.51
39	2588.54	2838.39
40	2544.25	2761.26

Table 4.17 Coherence Square and Amplitude of Transfer Function for DO and Coliform, Coosa River.

Frequency (Cycles/day)	Coherence Square	Amplitude of Transfer Function
0.0	0.6183	406.02
0.0125	0.5184	408.11
0.0250	0.1569	515.54
0.0375	0.3262	1201.22
0.0500	0.2249	1097.29
0.0625	0.1299	1121.27
0.0750	0.1702	1497.69
0.0875	0.2365	1629.67
0.1000	0.1631	1868.88
0.1125	0.0212	555.61
0.1250	0.0027	255.44
0.1375	0.0374	1680.05
0.1500	0.1022	3629.64
0.1625	0.1167	3015.94
0.1750	0.0091	909.19
0.1875	0.0060	545.89
0.2000	0.0217	1011.71
0.2125	0.1445	2359.61
0.2250	0.0546	1556.44
0.2375	0.1074	2940.01
0.2500	0.1756	4799.96
0.2625	0.0441	1786.17
0.2750	0.2438	4365.80
0.2875	0.3300	4698.14
0.3000	0.0282	1594.49
0.3125	0.1407	4221.64
0.3250	0.1121	4613.96
0.3375	0.1019	4187.70
0.3500	0.3861	10794.07
0.3625	0.0916	3778.23
0.3750	0.0689	2671.71
0.3875	0.0440	2283.33
0.4000	0.1151	3253.16
0.4125	0.0856	2519.35
0.4250	0.0243	1460.14
0.4375	0.0637	2186.68
0.4500	0.0167	1289.89
0.4625	0.0085	1082.30
0.4750	0.1427	4812.86
0.4875	0.1166	4259.02
0.5000	0.0088	1575.35

The coherence square is, generally, very low and is a function of frequency. The amplitude of the transfer function for the transformation relation DO and coliform varies considerably with frequency. This variation is quite erratic due to the nature of the coliform data and could not be put in any functional form.

Missouri River

Sampling interval for data digitized from the strip charts for Omaha Station on the Missouri River was 8 hours. With this sampling interval cyclic events with periods less than 24 hours can be resolved. The results of the analysis carried out for this station are presented in Figs. 4.46 through Fig. 4.49. Autocovariances for DO, BOD and temperature are compared in Fig. 4.46. For this river autocovariance for DO reaches zero at a lag of 43 days.

The results show a negative relationship between dissolved oxygen and Temperature but unlike the crosscorrelation for Coosa River crosscorrelations at Omaha show a minimum at zero lag: the correlation between DO and coliforms is the same as the correlation at the Coosa River.

Power Spectral estimates of Temperature and DO series are presented in Table 4.18. The estimates for temperature show cyclic events corresponding to periods of 7, 3.9, 2.3, 1.0 and 0.75 days. The DO series at the same station has periodic events corresponding to 2.9, 1.75 and 0.97 days.

Coherence square between Temperature and DO (Table 4.19)

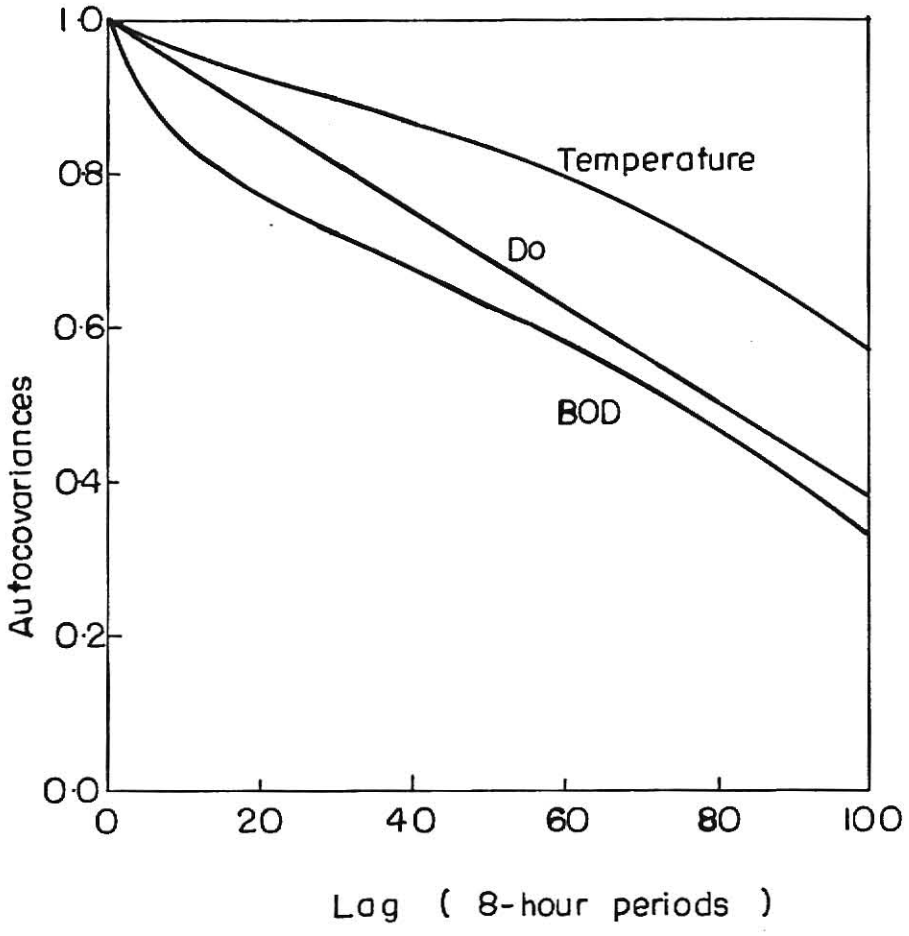


Fig. 4.46 Comparison of normalized autocovariances for Missouri River (Omaha Station)

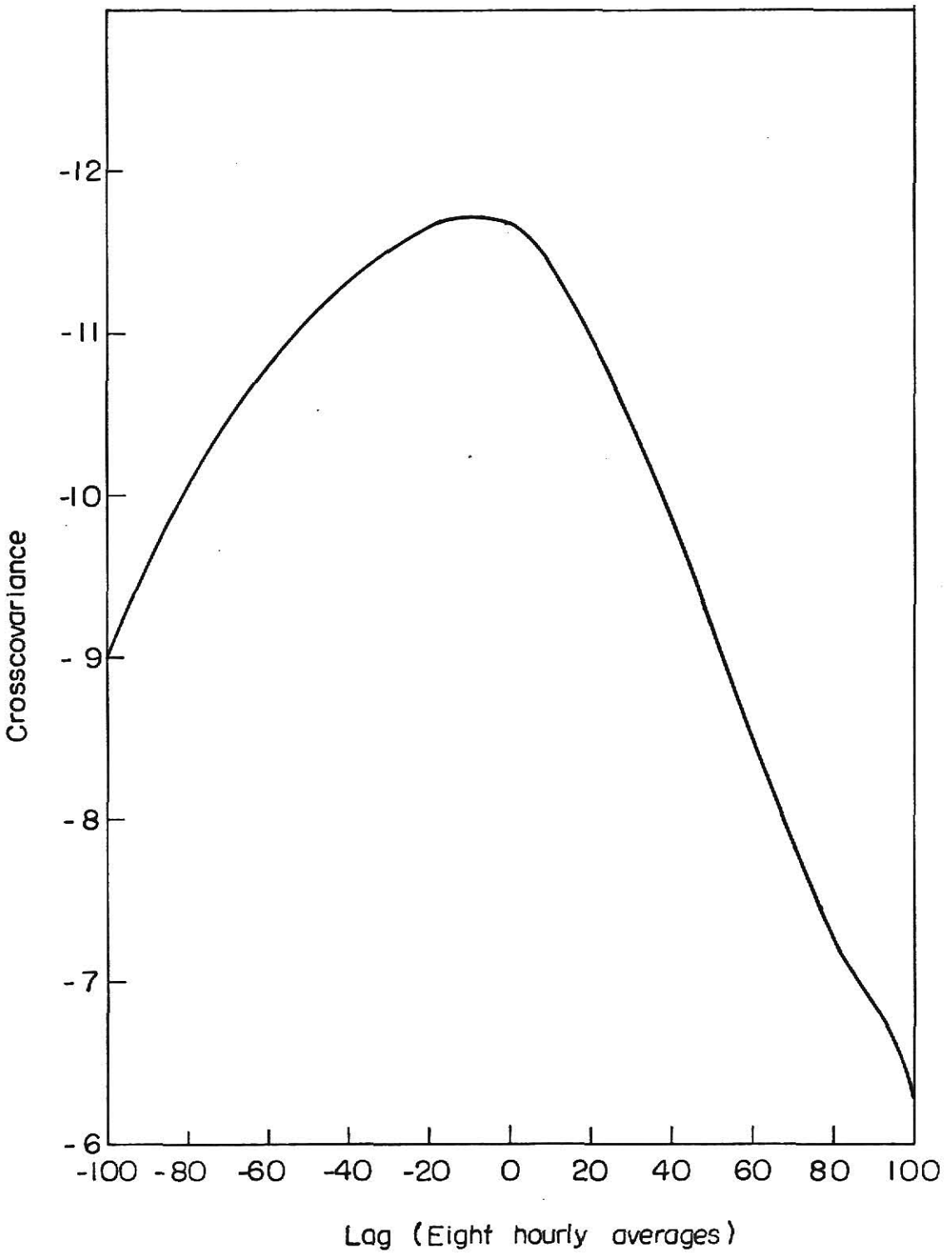
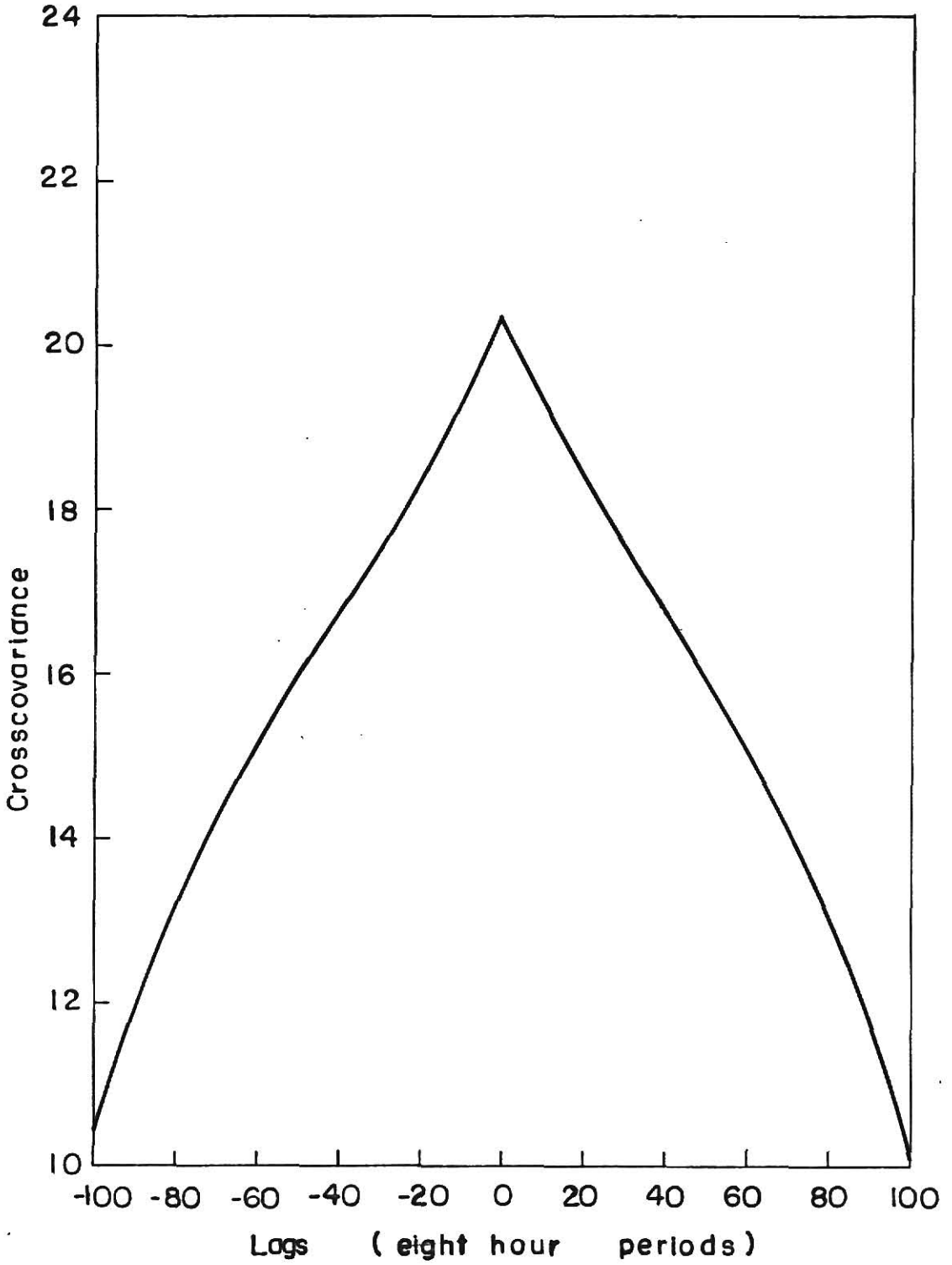


Fig. 4.47 Crosscovariance between temperature and dissolved oxygen at Omaha Station, Missouri River.



Fig·4.48 Crosscovariance between temperature and BOD at Omaha Station, Missouri River

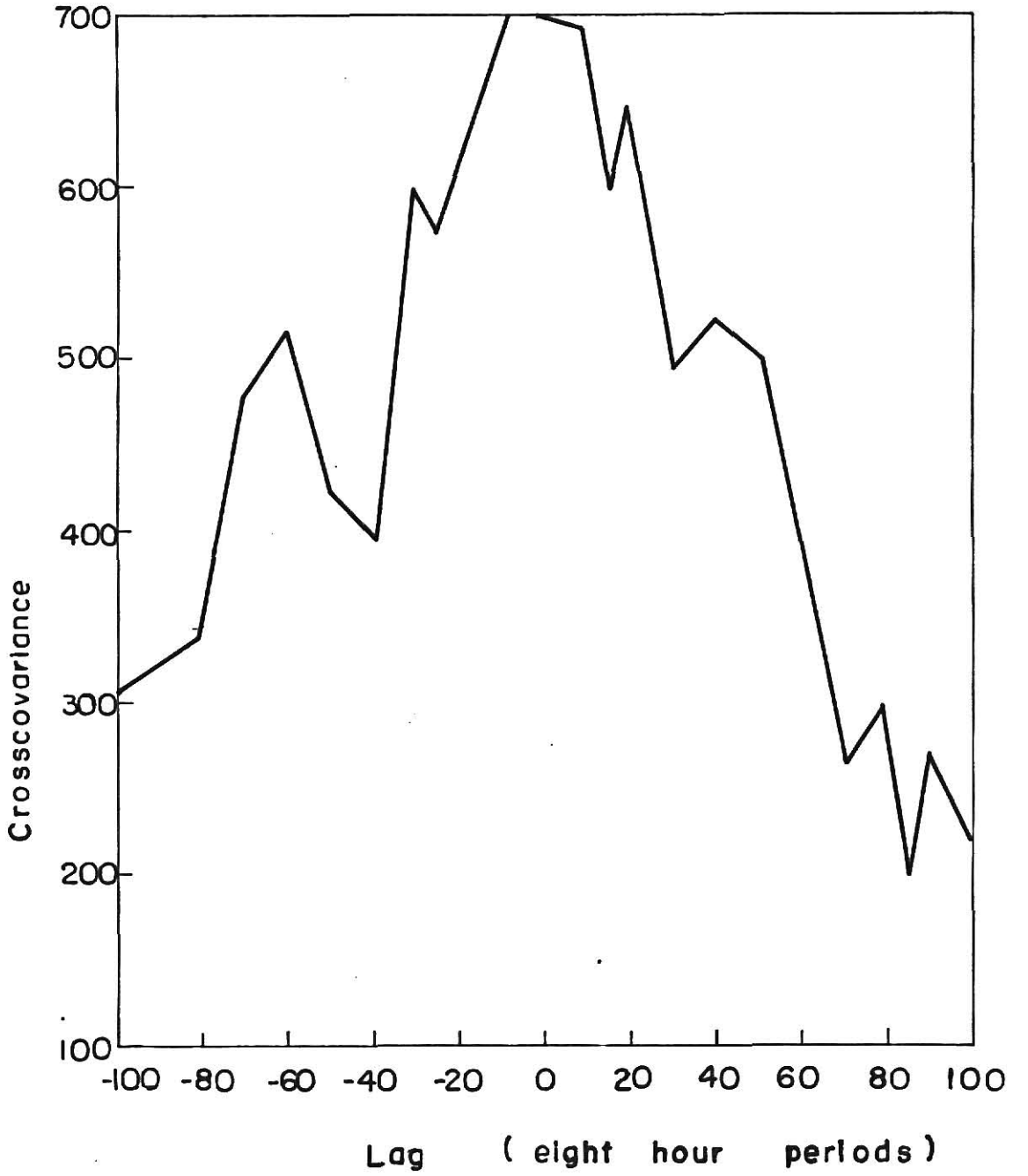


Fig.4.49 Crosscovariance between dissolved oxygen and Coliform, Omaha Station. Missouri River.

Table 4.18 Power Spectral Estimates of DO and Temperature Series, Missouri River.

Frequency (Cycles/day)	Spectral Estimates	
	Temperature	DO
0.0	2224.609	85.531
0.005	1069.850	44.157
0.010	75.194	6.725
0.015	16.225	2.995
0.020	12.357	2.156
0.025	7.533	0.777
0.030	4.951	0.313
0.035	3.693	0.273
0.040	2.631	0.291
0.045	4.809**	0.331
0.050	3.519	0.161
0.055	3.041	0.147
0.060	2.736	0.150
0.065	1.771	0.125
0.070	0.577	0.106
0.075	0.995	0.143
0.080	1.183	0.146
0.085	1.480**	0.132
0.090	0.953	0.082
0.095	0.549	0.066
0.100	0.212	0.048
0.105	0.463	0.035
0.110	0.575	0.039
0.115	0.752**	0.054
0.120	0.499	0.035
0.125	0.283	0.023
0.130	0.082	0.026
0.135	0.277	0.031
0.140	0.450	0.021
0.145	0.642	0.016
0.150	0.478	0.015
0.155	0.262	0.018
0.160	0.065	0.016
0.165	0.195	0.017
0.170	0.309	0.014
0.175	0.445	0.013
0.180	0.343	0.010
0.185	0.227	0.008
0.190	0.059	0.007
0.195	0.118	0.011
0.200	0.215	0.009
0.205	0.337	0.008
0.210	0.261	0.006
0.215	0.169	0.008
0.220	0.045	0.008

Table 4.18 (Cont'd)

Frequency (Cycles/day)	Spectral Estimates	
	Temperature	DO
0.225	0.094	0.008
0.230	0.166	0.006
0.235	0.259	0.007
0.240	0.190	0.006
0.245	0.122	0.006
0.250	0.024	0.005
0.255	0.065	0.006
0.260	0.117	0.006
0.265	0.196	0.006
0.270	0.149	0.004
0.275	0.100	0.005
0.280	0.018	0.006
0.285	0.050	0.007
0.290	0.090	0.006
0.295	0.157**	0.007
0.300	0.118	0.006
0.305	0.081	0.005
0.310	0.007	0.005
0.315	0.028	0.006
0.320	0.046	0.005
0.325	0.108	0.005
0.330	0.143**	0.007
0.335	0.132	0.010
0.340	0.011	0.005
0.345	0.017	0.005
0.350	0.031	0.004
0.355	0.079	0.005
0.360	0.058	0.004
0.365	0.041	0.004
0.370	-0.006*	0.003
0.375	0.012	0.004
0.380	0.018	0.004
0.385	0.057	0.004
0.390	0.035	0.003
0.395	0.024	0.003
0.400	-0.012*	0.003
0.405	0.007	0.003
0.410	0.011	0.003
0.415	0.042	0.004
0.420	0.022	0.004
0.425	0.020	0.004
0.430	-0.009*	0.003
0.435	0.006	0.003
0.440	0.002	0.003

Table 4.18 (Cont'd)

Frequency (Cycles/day)	Spectral Estimates	
	Temperature	DO
0.445	0.026	0.003
0.450	0.011	0.003
0.455	0.016	0.004
0.460	-0.006	0.004
0.465	0.007	0.003
0.470	-0.000	0.003
0.475	0.016	0.003
0.480	0.004	0.003
0.485	0.015	0.003
0.490	0.001	0.003
0.495	0.010	0.004
0.500	-0.000	0.005

* This estimate is negative indicating some leakage.
** This represents an important cyclic event.

Table 4.19 Coherence Square and Amplitude of Transfer Function for DO-Temperature, Missouri River.

Frequency (Cycles/day)	Coherence Square	Amplitude of Transfer Function
0.0	0.724	0.166
0.0050	0.660	0.165
0.0100	0.242	0.147
0.0150	0.052	0.098
0.0200	0.029	0.072
0.0250	0.068	0.084
0.0300	0.334	0.145
0.0350	0.105	0.088
0.0400	0.223	0.157
0.0450	0.186	0.113
0.0500	0.208	0.097
0.0550	0.084	0.063
0.0600	0.433	0.154
0.0650	0.371	0.162
0.0700	0.280	0.227
0.0750	0.336	0.220
0.0800	0.080	0.099
0.0850	0.279	0.157
0.0900	0.296	0.159
0.0950	0.216	0.162
0.1000	0.926	0.462
0.1050	0.884	0.259
0.1100	0.341	0.153
0.1150	0.466	0.183
0.1200	0.147	0.101
0.1250	2.524**	0.457
0.1300	1.048	0.577
0.1350	1.323**	0.387
0.1400	0.126	0.077
0.1450	0.890	0.152
0.1500	0.545	0.133
0.1550	0.896	0.249
0.1600	2.370**	0.769
0.1650	2.053**	0.428
0.1700	0.553	0.158
0.1750	0.606	0.136
0.1800	0.605	0.134
0.1850	1.985**	0.274
0.1900	3.924**	0.704
0.1950	3.780**	0.612
0.2000	0.231	0.098
0.2050	1.455**	0.186
0.2100	0.821	0.146
0.2150	1.463**	0.278

Table 4.19 (Cont'd)

Frequency (Cycles/day)	Coherence Square	Amplitude of Transfer Function
0.2200	5.156	0.963
0.2250	3.186	0.519
0.2300	0.752	0.170
0.2350	0.920	0.163
0.2400	0.993	0.188
0.2450	2.014**	0.335
0.2500	4.853**	1.060
0.2550	4.931**	0.724
0.2600	0.342	0.136
0.2650	1.817**	0.243
0.2700	0.943	0.175
0.2750	1.620**	0.303
0.2800	7.893**	1.646
0.2850	2.700**	0.625
0.2900	1.309**	0.298
0.2950	0.581	0.168
0.3000	0.535	0.168
0.3050	2.021**	0.379
0.3100	9.764**	2.511
0.3150	5.540**	1.094
0.3200	0.471	0.240
0.3250	1.791**	0.302
0.3300	0.507	0.163
0.3350	0.619	0.216
0.3400	8.848**	2.122
0.3450	4.313**	1.153
0.3500	2.782**	0.641
0.3550	0.571	0.202
0.3600	0.838	0.254
0.3650	1.633**	0.414
0.3700	0.0 *	0.0 *
0.3750	5.441**	1.478
0.3800	1.167**	0.531
0.3850	1.170**	0.304
0.3900	1.300**	0.356
0.3950	1.466**	0.463
0.4000	0.0 *	0.0 *
0.4050	5.993**	1.759
0.4100	2.591**	0.849
0.4150	0.490	0.221
0.4200	0.939	0.417
0.4250	1.130**	0.497
0.4300	0.0 *	0.0 *
0.4350	2.935**	1.296

Table 4.19 (Cont'd)

Frequency (Cycles/day)	Coherence Square	Amplitude of Transfer Function
0.4400	2.4913**	1.6671
0.4450	1.0169	0.3834
0.4500	1.8001**	0.7689
0.4550	0.3084	0.2994
0.4600	0.0 *	0.0 *
0.4650	1.2342**	0.7769
0.4700	0.0 *	0.0 *
0.4750	0.1500	0.1877
0.4800	0.6785	0.7451
0.4850	0.2321	0.2415
0.4900	7.9856**	4.7080
0.4950	0.1350	0.2413
0.5000	0.0 *	0.0 *

* Indicates this value is not computable due to a negative or zero power spectral estimate.

** Indicates that this value is too high due to sampling error.

exceeds 1.0 for a number of frequencies. This is due to sampling error to invalid interpolation of the missing data. Coherences at some frequencies could not be computed as temperature power spectral estimates were negative at these frequencies.

COMPARISON OF RESULTS

In this section the results obtained in the previous sections are compared. The normalized autocorrelations for three time series on the Ohio River are presented in Fig. 4.50. The autocorrelation coefficients for flow decrease fastest and reach a zero value for a lag of 22 days (45 half day periods). This indicates that flow changes from above normal to below normal in a much shorter duration of time than temperature and dissolved oxygen. The temperature autocorrelation coefficients have a very small negative slope. This shows that it takes a long time for the temperature to drop from above normal to below normal due to seasonal fluctuations. The dissolved oxygen autocorrelation coefficients lie in between these two autocorrelation coefficients (temperature and flow).

The autocorrelation coefficients for the temperature record for three rivers are presented in Fig. 4.51. Autocorrelations for Coosa River have the largest slope. But for all three cases the trend is similar and these values reach zero only for very large lags. These results in a way confirm the validity of the data at these three stations.

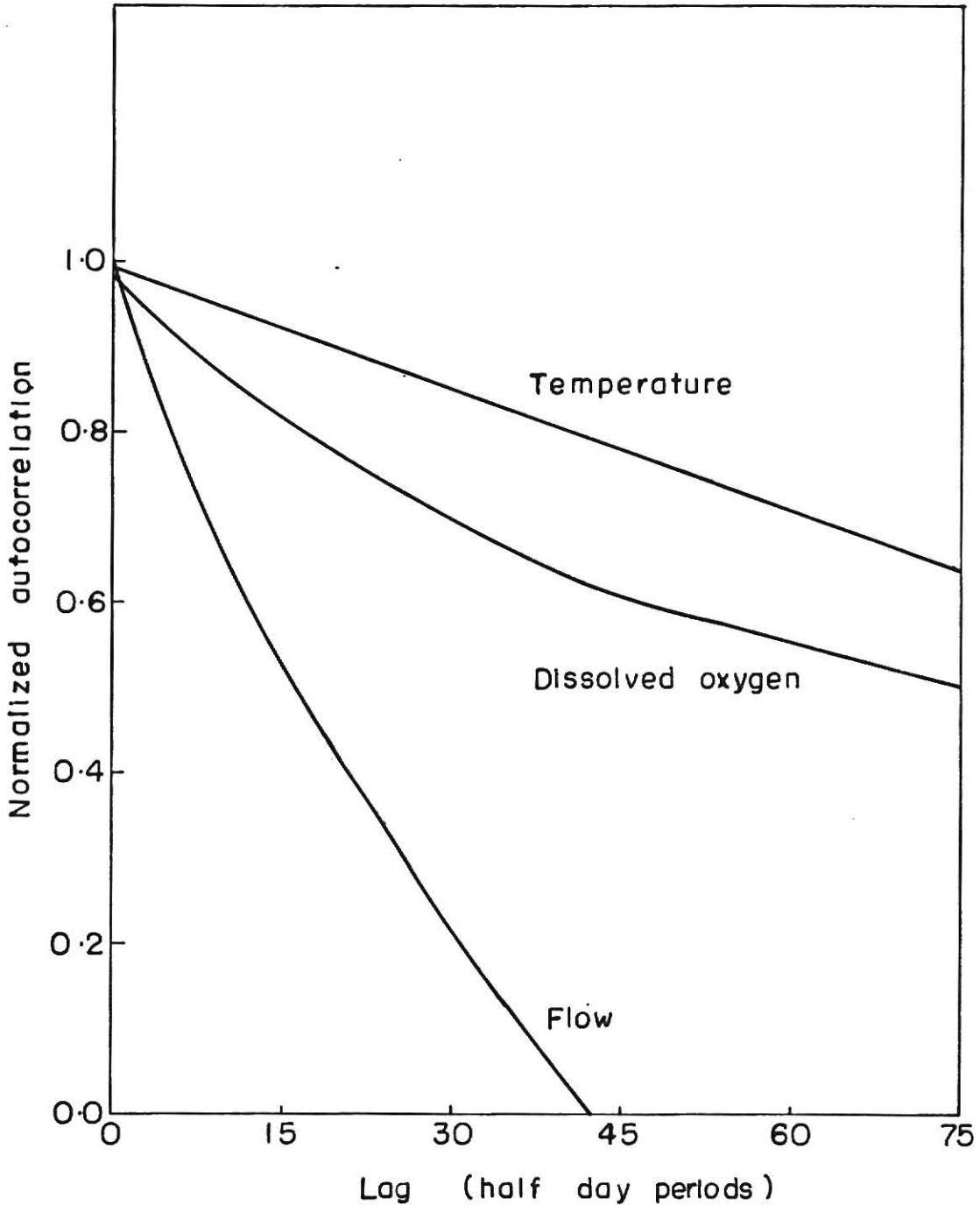
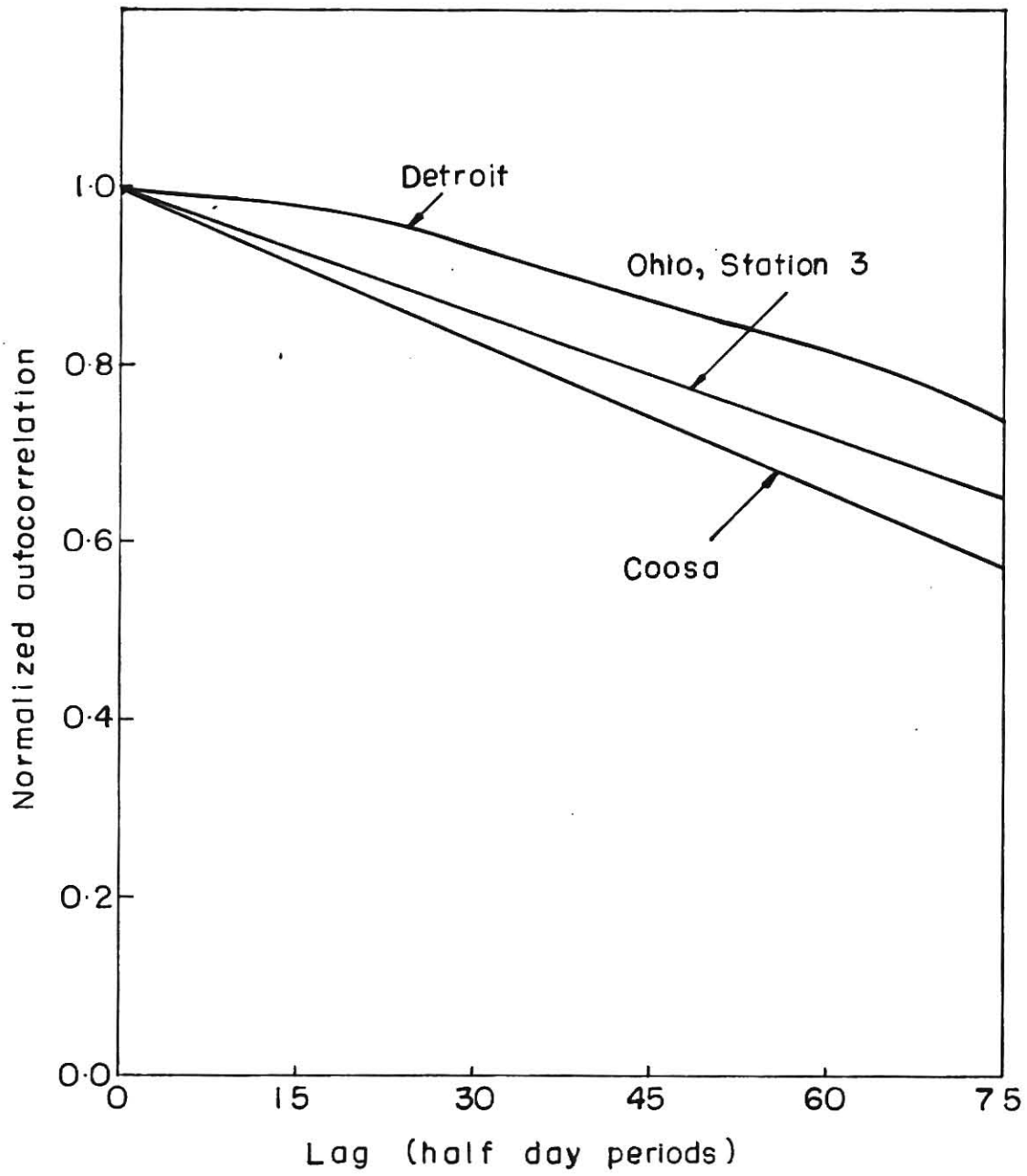


Fig. 4.50 Comparison of autocorrelations for Station 3, Ohio River (1968)



Fig·4·51 Comparison of temperature autocorrelation for three rivers.

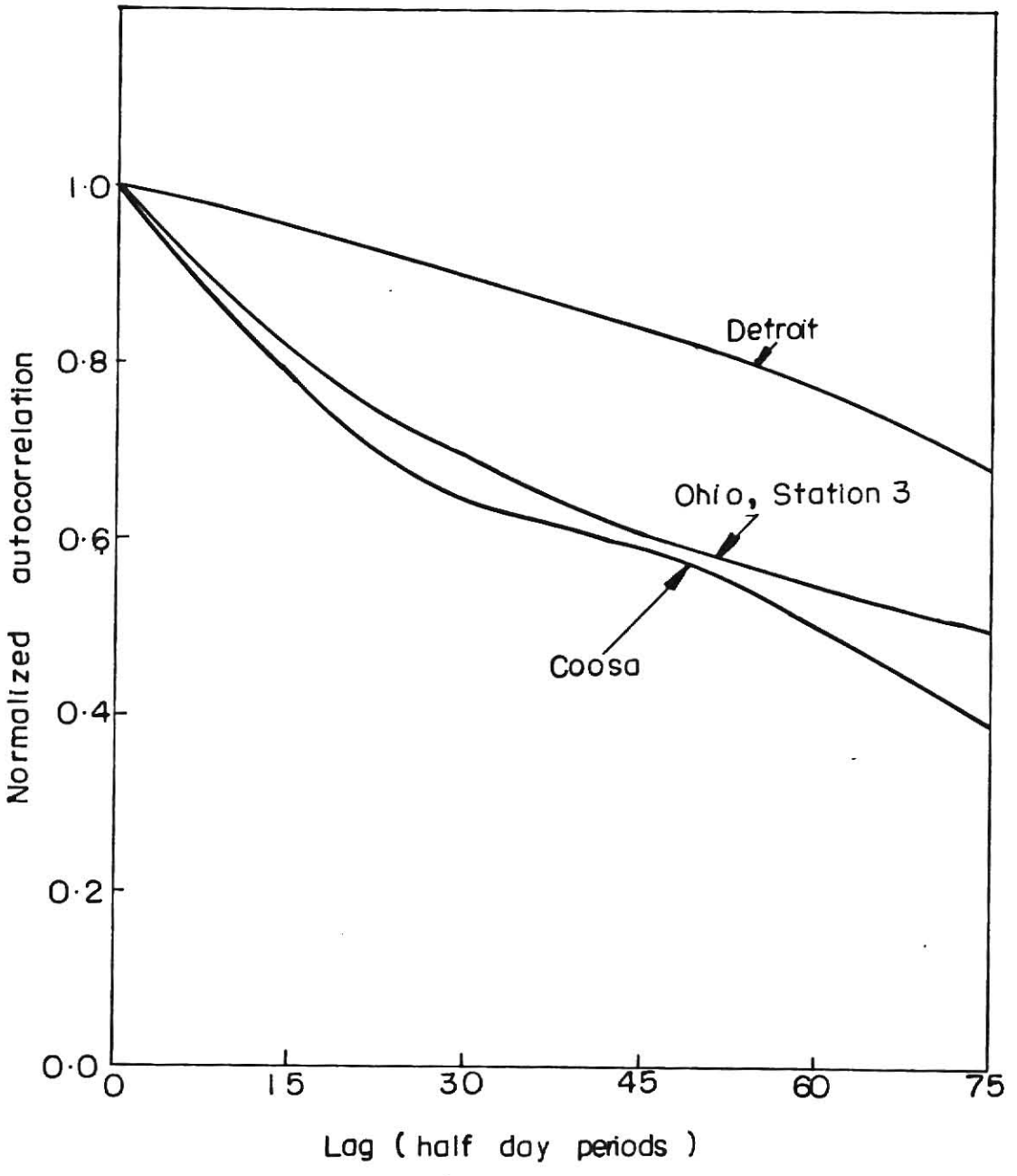
Autocorrelations for oxygen records at these three stations are shown in Fig. 4.52. These results are similar to the temperature autocorrelation coefficients in that they have the same relative magnitudes.

CONCLUDING REMARKS

It can be concluded from these calculations and physical interpretations that spectral analysis serves as a very useful means of analyzing systems with time series data. Several conclusions derived during this investigation are described in this section and in the next chapter in which these results are used.

The Ohio river system, which is assumed to consist of five time series in this work is so complex that no general conclusions true for all rivers can be derived from such a limited study. More water quality information must be collected from representative rivers of all types in various geographical locations and analyzed systematically in order to develop general mathematical formulations.

It is well known that it takes about five days to determine the BOD of a given sample. This time lag in the measurement of this very important water quality parameter is a significant factor in the development of a water pollution control program and in the analysis of the water quality data. Methods have been developed (49) to measure BOD instantaneously from a water sample. Correlation between this BOD and the standard



Fig·4·52 Comparison of dissolved oxygen auto-correlations for three rivers.

five day BOD should be attempted as all the models (Chapter 3) relate the five day BOD to dissolved oxygen. Instruments have also been developed to measure COD (Chemical Oxygen Demand) in a very short time. This parameter can be correlated with the standard five day BOD.

Another difficulty encountered during this analysis was the lack of data on coliforms of microorganisms. A new instrument which enbales one to quickly determine the bacterial bio-mass by measuring cellular ATP is now available (50). This can be correlated with the cell concentration or used as such in water quality modeling.

This analysis also indicates a sophisticated method should be developed for obtaining continuous data without any missed observations.

The results presented in this chapter show a variety of relationships in the time and frequency domains. The interpretations from these results are already presented in the previous sections. Analysis of the flow rate data indicates that there are not many significant periodicities in the power spectrum estimates. The increase in flow from one station to another (Table 4.3) due to the addition of small side streams increases the autocovariance and, therefore, the variance of flow records as one goes from station 1 to station 4. Data obtained at a smaller sampling interval would enable one to search for smaller periods.

Temperature and dissolved data were analyzed with a

sampling interval of 12 hours. With this data no cycle event with a period less than one day can be identified. The auto-correlations for these two water quality parameters provide a method for predicting the future values of the parameters.

As BOD and coliform data were hypothetically generated autocorrelation and autospectrum of these records were not calculated.

The crosscorrelation and coherence square functions of these water quality parameters yield some very interesting results. All the coherences are frequency dependent indicating a nonlinear relationship. The crosscorrelation calculations show that BOD is positively correlated with temperature and negatively correlated with DO; the oxygen is negatively correlated with temperature. No definite conclusions could be drawn from the results with coliform data.

The conclusions derived in this chapter are used in Chapter V for developing a mathematical formulation.

CHAPTER V

MODELING FOR WATER QUALITY

INTRODUCTION

In any systems engineering study, the construction of mathematical models to represent the phenomenological processes taking place in the system is important. Only phenomena pertinent to the behavior under study are allowed to enter the model. A mathematical model is often considerably more versatile and less expensive than a physical model. Once a model is obtained, either empirically or from the mechanism of the process, the next step is to obtain experimental data to test the goodness of the model. In this chapter several models will be developed for water quality prediction of streams.

In water pollution the objectives for developing mathematical models are [3, 17]

i) to develop a method for evaluating the cause and effect relationships between the external environment (waste discharges, temperature, etc.) and the quality of water in terms of its dissolved oxygen concentration.

ii) to utilize a portion of this evaluation to develop further a rational approach to the attainment of the various quality goals.

The models for water quality will be developed by using the spectral analysis results as well as other available information and experience. In the next section some general conclusions

derived from the study of spectral analysis are presented.

RESULTS FROM SPECTRAL ANALYSIS

The following general observations can be made from the results of spectral analysis presented in the preceding chapter

1. Like most other natural processes, the stream ecology is a nonlinear process. All the water quality parameters are nonlinearly related to one another. The linear models developed so far are only an approximation to the actual system. The coherence calculations indicate that the system is appreciably nonlinear and is affected by nonlinear interactions among the variables. It is necessary, therefore, to use sophiscated modeling techniques and nonlinear parameter estimation procedures to establish more realistic models.

2. As the system consists of a large number of state variables, a more complex analysis involving the calculation of multiple coherences, spectral density matrices and partial coherences is necessary. The analysis presented in this thesis considers only two variables at a time and attempts to correlate these parameters. Techniques have been developed [51] to analyze multiple order systems by spectral analysis procedures. These techniques can be used to analyze the stream with five water quality parameters, namely dissolved oxygen, biochemical oxygen demand, temperature, flow rate and coliform count simultaneously.

3. There is a definite negative relationship between

temperature and dissolved oxygen at some locations. This is understandable as the solubility of a gas in a liquid decreases with increasing temperature. As the temperature increases the rate of reaeration increases but the saturation oxygen concentration decreases. The rate of microbial decay or the rate of oxygen consumption by microorganisms increases with temperature. The net effect is that there is reduction in oxygen concentration.

This is the cause of pollution without any additional waste discharge. The dumping of high temperature water into the stream, increases the water temperature causing thermal pollution. This causes a stratified flow and may increase the rate of microbial growth and hence the rate of oxygen demand. Hence, an increase in temperature may cause the oxygen concentration to fall below a critical level.

4. The results of spectral analysis show a negative correlation between BOD and DO. This means, if at a point on the stream BOD increases with time DO would tend to decrease.

5. Although the coliform data obtained at some of the stations were not very good, the analysis at the other stations points to the following conclusions. The coliform count has a positive correlation with temperature. It is known that an increase in temperature will increase the rate of growth of microorganisms, and thus an increase in the coliform count with temperature is understandable. BOD and coliform count are negatively correlated indicating that the number of coliform

decreases with an increase in biochemical oxygen demand.

From these correlations it can be concluded that an increase in temperature may cause an increase in the number of coliform bacteria.

A more realistic water quality model should include the growth of microorganisms and their effect on the biochemical oxygen demand and dissolved oxygen concentration. This can be done in either of two ways: i) by including in the model a third differential equation based on the cell material balance or ii) by assuming a constant cell concentration and by using a Michaelis-Menten kinetic model for BOD decay in the stream. Models formulated using these approaches will be presented later in the chapter.

6. The transfer function calculated between dissolved oxygen at two points indicates that the transformation can be represented by a system somewhere in between first and second order. The first order system would represent the plug flow case where the properties would be uniform across the stream. The second order system would mean that the ideal plug flow has been superimposed by a diffusion transfer which causes mass transfer in the axial direction. It has been shown by Dobbins [13] that for nontidal streams diffusion can be neglected and first order differential equation can be used for modeling water quality. The parameter which correctly characterizes the role played by dispersion is a dimensionless group $\frac{D}{uL}$ which is the reciprocal of the axial Peclet number for mass transfer. For

stream u and L are often very large and so $\frac{D}{uL}$ is very small no matter how large D is. This low value of $\frac{D}{uL}$ corresponds to plug flow behavior [9].

This justifies the plug flow assumption used so far in the modeling of water quality. In this work first order differential equations will be used for water quality modeling.

7. The periodicities in the dissolved oxygen spectrum, especially the high frequency harmonics, necessiate the inclusion of photosynthesis terms in the model. This can be accomplished by including in the model the photosynthetic terms expanded in terms of sines and cosines. The maximum photosynthetic rate may vary with temperature and seasonal fluctuations. The duration of sunshine may change from season to season causing a shift in the photosynthetic period.

8. The most general model for a stream under steady state would include equation for the variation of temperature and temperature dependent models for the following constants.

i) reaeration constant, k_2

ii) bacterial decay constant, k_1

iii) saturation oxygen concentration, C_s

iv) the maximum photosynthesis rate, P_m

v) the maximum specific growth rate, u_{max}

Inclusion of these will enable one to analyze the effects on water quality resulting from a high temperature discharge.

DEVELOPMENT OF MODELS

The model developed to represent the system mathematically is a simplified construction of the real world that seeks to explain and predict events in this world. In setting up a model for quality prediction the following processes are assumed to be taking place.

a) The growth of microorganisms and their effect on BOD and DO level in the stream. The growth rate of microorganisms in stream can be taken as [52]

$$\frac{dN}{dt} = \frac{u_{max}\, LN}{K_L + L} - k_D\, N \qquad (74)$$

in which

N = number of microorganisms per liter

L = BOD in mg per liter

K_L = saturation concentration in mg per liter

k_D = endogenous metabolism constant, hr^{-1}

u_{max} = maximum specific growth rate, hr^{-1}

This phenomenon decreases BOD and dissolved oxygen, the rate of decrease of BOD is given by

$$-r_B = -\frac{u_{max}\, LN}{Y(K_L+L)} \qquad (75)$$

in which

Y = yield in number of cells per unit mass of substrate (BOD).

b) Increase in DO due to reaeration from the atmosphere. The rate of this increase is proportional to the difference between the saturation concentration and the actual DO present. The reaeration constant is taken as k_2. The rate of increase of oxygen concentration in the stream due to reaeration is given as

$$r_r = k_2(C_s - C) \tag{76}$$

The reaeration constant k_2 is a function of stream velocity, depth, the axial diffusion coefficient, density of the stream and the stream viscosity. Formulae for calculating k_2 from the stream parameters are presented in [53, 54]. The reaeration k_2 is also a function of temperature, the relationship being given as [53]

$$k_2 = k_{20}\ \theta^{T-20} \tag{77}$$

The saturation oxygen concentration in the stream C_s, is also temperature dependent. This parameter decreases with increasing temperature. A polynomial of order 2 has been fitted to C_s vs T data and this is in good agreement with the experimental results

$$C_s = a + bT + cT^2 \tag{78}$$

c) Decrease in the biochemical oxygen demand due to sedimentation or adsorption or both. The rate of depletion is proportional to the amount of BOD present in the system. The first order rate constant for this process is taken as k_3.

d) BOD addition due to local runoff; this is assumed constant along a stream length and denoted by L_a.

e) Addition of oxygen due to photosynthetic action of plankton and fixed plants. Results of spectral analysis show that this phenomenon needs to be considered. This addition depends on the sunlight intensity, the mass of algae, nutrients and the temperature. The rate of photosynthetic addition of oxygen is assumed to vary with the sunlight intensity reaching a maximum at noon and vanishing at sunset and sunrise. This source can be written as

$$\begin{aligned} P_t{}^1 &= P_m \sin \frac{\pi}{p} t^1 && 0 \leq t^1 \leq p \\ &= 0 && 1-p \leq t^1 \leq 1 \end{aligned} \tag{79}$$

in which

p = period of oscillation of the periodic function, days

P_m = amplitude of the periodic function

t^1 = time of flow in days of a mass of water since sunrise after the mass has entered the head end of a reach considered ($t^1 = t - \tau$)

If the period p equals half a day the function reduces to

$$P_t l = P_m \sin 2\pi(t - \tau) \qquad 0 \leq (t - \tau) \leq \frac{1}{2}$$
$$= 0 \qquad \frac{1}{2} \leq (t - \tau) \leq 1 \tag{80}$$

This function P_t is periodic in nature and can be expanded in terms of sines and cosines using Fourier analysis. The expnasion is [17]

$$P_t l = \frac{P_m}{\pi} + \frac{P_m}{2} \sin 2\pi(t - \tau) - \frac{2P_m}{3\pi} \cos 4\pi(t - \tau) \tag{81}$$

f) Oxygen consumption due to respiration and benthal demand. This is also assumed constant all along the stretch and is denoted by D_B.

It has been shown that a nontidal stream can be assumed to be in plug flow without introducing appreciable error [13]. From a differential material balance taken over an elemental length assuming a constant crosssecticnal area the following differential equations can be obtained [9]. Note that the equations derived will contain a velocity term u and a distance term x. For plug flow these can be combined ($t = \frac{x}{u}$) and replaced by a residence time term. This 't' is the time of travel along the stream measured from an arbitrary origin.

$$\frac{dL}{dt} = - \frac{u_{max} LN}{Y(K_L + L)} - k_3 L + L_a \tag{82}$$

$$\frac{dC}{dt} = -\frac{u_{max}\,LN}{Y(K_L+L)} + k_2(C_s - C) - D_B + P_m\left\{\frac{1}{\pi} + \frac{1}{2}\sin 2\pi(t-\tau) - \frac{2}{3\pi}\cos 4\pi(t-\tau)\right\} \tag{83}$$

$$\frac{dN}{dt} = \frac{u_{max}\,LN}{K_L + L} - k_D L \tag{84}$$

As a steady state for growth of microorganisms is attained after a certain time, these equations can be reduced down to two equations namely,

$$\frac{dL}{dt} = -\frac{k_1 L}{Y(K_L + L)} - k_3 L + L_a \tag{85}$$

$$\frac{dC}{dt} = -\frac{k_1 L}{Y(K_L + L)} + k_2(C_s - C) - D_B + P_m\left\{\frac{1}{\pi} + \frac{1}{2}\sin 2\pi(t-\tau) - \frac{2}{3\pi}\cos 4\pi(t-\tau)\right\} \tag{86}$$

Models of Equations 83 through 86 are examined along with the models given by the following equations:

$$\frac{dL}{dt} = -\frac{k_1 LC}{K_L + L} - k_3 L + L_a \tag{87}$$

$$\frac{dC}{dt} = -\frac{k_1 LC}{K_L + L} + k_2(C_s - C) - D_B + P_m\left\{\frac{1}{\pi} + \frac{1}{2}\sin 2\pi(t-\tau) - \frac{2}{3\pi}\cos 4\pi(t-\tau)\right\} \tag{88}$$

$$\frac{dL}{dt} = - k_1 LC - k_3 L + L_a \tag{89}$$

$$\frac{dC}{dt} = - k_1 LC + k_2(C_s - C) - D_B + P_m\left\{\frac{1}{\pi} + \frac{1}{2} \sin 2\pi(t - \tau) - \frac{2}{3\pi} \cos 4\pi(t - \tau)\right\} \tag{90}$$

$$\frac{dL}{dt} = - k_1 L - k_3 L + L_a \tag{91}$$

$$\frac{dC}{dt} = - k_1 L + k_2(C_s - C) - D_B + P_m\left\{\frac{1}{\pi} + \frac{1}{2} \sin 2\pi(t - \tau) - \frac{2}{3\pi} \cos 4\pi(t - \tau)\right\} \tag{92}$$

Table 5.1 summarizes the models and lists the number of unknown parameters in each of these models. To discriminate between these models one needs to estimate the parameters present in the models and test the validity of each of these models against the experimentally measured data. The criteria used for model discrimination are

i) value of the objective function [Chapter VI]

ii) residual at each experiment

iii) standard deviation of the parameters.

As all the proposed models are nonlinear with respect to the parameters the ordinary least squares technique cannot be used for parameter estimation. Specialized nonlinear parameter estimation techniques are available for this estimation and one of these techniques (Bard's) is used in this investigation. This technique

is described in the next chapter along the results obtained for models in Table 5.1.

Table 5.1 Models and Parameters to be Estimated.

Number	Represented by Equation (s)	Unknown Parameters	Number of Parameters
1	82, 83 and 84	max, y, k_3, L_a, k_2, C_s, D_B, P_m, τ, K_L	10
2	85 and 86	k_1, y, k_3, L_a, k_2, C_s, D_B, P_m, γ, K_L	10
3	87 and 88	k_1, k_3, L_a, K_L, k_2, D_B, C_s, P_m,	9
4	89 and 90	k_1, k_3, L_a, k_2 C_s, D_B, τ, P_m	8
5	91 and 92	k_1, k_2, k_3, L_a, D_B, C_s, P_m, γ	8

CHAPTER VI

PARAMETER ESTIMATION

INTRODUCTION

Engineers frequently propose mathematical models or equations which attempt to describe the relationships between physically measured variables. These models usually contain parameters or coefficients whose values are unknown. The object of parameter estimation is to determine the values of these unknown parameters so that the proposed model behaves as closely as possible to the actual system. In other words the proposed equations should give the best possible fit to the experimentally obtained data. The measure of 'fit' or 'closeness' are the residuals - the difference between the observed and predicted values of the dependent variables. The best fit criteria used may be least squares, weighted least squares, maximum likelihood or Bayesian fitting [55].

A method is described in this Chapter which is capable of estimating parameters in nonlinear algebraic or differential models.

MATHEMATICAL MODEL

The model which relates the observed variables to the independent variables, may usually be written in the form,

$$g(y_u, x_u, \theta) = 0, \qquad u = 1, 2, \ldots, n \tag{93}$$

in which

g = vector of functions

θ = vector of parameters, $\theta = [\theta_1, \theta_2, \ldots, \theta_u]$

x_u = the vector of independent variables, accurately known, for the uth experiment $x_u = [x_{u1}, x_{u2}, \ldots, x_{un}]$

and

y_u = the vector of dependent variables for the uth experiment

To specify the model, one must assign mathematical expressions to each of the g_k's. In explicit form Equation 93 becomes

$$y_u = f(x_u, \theta), \quad u = 1, 2, \ldots, n \tag{94}$$

If the above equation is linear in the parameters (θ) the system is referred to as a linear system and the problem of parameter estimation reduces to a linear least square problem, the solution of which can be found in any standard text on regression analysis. In this chapter a method is presented which estimates parameters in a nonlinear system. It can be shown that due to errors in experimentation and inaccuracies in the model one cannot determine parameter values that will satisfy Equation 94 exactly for each experiment. In these cases Equation 93 is rewritten as

$$u_u = f(x_u, \theta) - y_u \tag{95}$$

in which

u_u = vector of residuals which represents the difference between the predicted values and the observed values of y at the uth experiment.

CRITERIA FOR ESTIMATION

In the parameter estimation procedure a certain function $G(\theta)$ of u_u is minimized. This functional form determines the criteria used for estimation; the forms usually employed are discussed below [55]:

(i) Least squares: In this case there is only one observation per experiment and θ is determined to minimize

$$G_{LS}(\theta) = \sum_{u=1}^{n} u_u^2 = \Sigma [f(x_u, \theta) - y_u]^2 \qquad (96)$$

(ii) Weighted least squares: Here, there are k observations per experiment and one minimizes

$$G_{LS}(\theta) = \sum_{u=1}^{n} \sum_{i=1}^{k} w_i [f_i(x_u, \theta) - y_{u0}]^2 \qquad (97)$$

to obtain the values of the parameters.

(iii) Maximum likelihood: In this case u's are assumed to be random variables possessing a joint probability density function $p(u, \phi)$ and the values of θ and , which maximize the likelihood of having made the actual observations, are determined. The function to be maximized is

$$G_{ML}(\theta, \phi) = \log p(f(x_u, \theta) - y_{u0}, \phi) \qquad (98)$$

in which

ϕ = a parameter necessary to specify the probability distributions

(iv) Bayesian estimation: In maximum likelihood estimation it is assumed that the probability density of θ's as given, but in case of Bayesian fitting one assumes the probability density of y's are given and one maximizes $p(\theta/y_u)$ which is the probability density of θ's given the observations y_u

$$p(\theta/y_u) = \frac{1}{C} p(y_u/\theta)\ p_0(\theta) \qquad (99)$$

in which

C = normalizing constant

$p_0(\theta)$ = prior distribution of θ

and

$$(y_u/\theta) = p(u_u, \phi)$$

Writing in terms of logarithms, the objective function of Equation 99 becomes

$$G_B(\theta, \phi) = \log p(f(x_u, \theta) - y_u, \phi) + \log p_0(\theta) \qquad (100)$$

METHODS FOR PARAMETER ESTIMATION

A number of methods have been proposed in the literature to obtain this minimization. These range from linearization methods as proposed by Gauss [56] to steepest descent method as formulated by Davidon [57]. In the linearization method the nonlinear model of Equation 94 is linearized by a truncated Taylor series expansion. The vector $[\Delta\theta]$ is, then, calculated by solving a set of simultaneous linearized algebraic equations. The solution is [57]

$$[Q][\Delta\theta] = -[p] \qquad [101]$$

in which

Q = an ℓ x ℓ matrix

ℓ = number of independent parameters

θ = independent parameter search vector

p = vector of first partial derivatives of the sum of squares function

In case of Gauss' method the elements of the matrix Q are [56]

$$Q_{ij} = 2\frac{\partial f}{\partial \theta_i}\frac{\partial f}{\partial \theta_j} \qquad (102)$$

whereas in Newton's method

$$Q_{ij} = \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} \qquad (103)$$

In the steepest descent method the sums of squares surface is visualized to be a response surface in which the parameters are variables. It involves the calculation of a linear independent parameter search vector $\Delta\theta$ based on the first partials of the sums of squares function and, possibly, weighting factors for each parameter

$$\Delta\theta_i = - hw_i\left(\frac{\partial G}{\partial \theta_i}\right) \tag{110}$$

in which

$\Delta\theta_i$ = i^{th} element of the search vector

w_i = weighting factor for i^{th} independent parameter search vector > 0

G = quantity to be minimized

h = scalar distance factor > 0

Equation 110 defines a search direction but the distance depends on the value used for the scalar distance factor h. This equation always provides a search trajectory which possesses truncation convergence, since each element in the search vector is always of opposite sign to the corresponding partial derivative. The usual experience with this method is slow progress in the vicinity of the optimum due to what has been termed 'lack of quadratic convergence.' The steepest descent usually produces a series of ever smaller zig zag steps in the vicinity of the minimum as discussed by Buehler, Shah and Kempthorne [58].

Experience with fitting many models indicates that the steepest descent is very stable for the initial iterations while a linearization method is often more efficient for the final iterations. Compromise methods have been suggested, [55, 59] which tend to emphasize steepest descent at the outset and linearization in the final stages. In this work one such compromise method called the absolute Gauss or Bard's [55] method is used for parameter estimation. This method will be described in the next section.

BARD'S MODIFICATION [55]

Let $G(\theta)$ be a scalar function of the vector θ, possessing continuous first and bounded second derivatives. We wish to find a vector θ that maximizes G. (G may be any of the criteria functions discussed in the preceeding section.) The maximization of $G(\theta)$ proceeds in an iterative fashion. Each iteration are starts with an initial guess $\theta^{(0)}$ for the vector θ and proceeds to find a new guess $\theta^{(1)}$ such that $G(\theta^{(1)}) > G(\theta^{(0)})$. As in other gradient methods, one assumes

$$[\Delta\theta] = K[R][p] \tag{105}$$

in which

p = a vector of derivatives of G, $[\frac{\partial G}{\partial \theta}]$

K = a positive scalar which determines the step size

and

R = an n x n matrix

This method is described for maximizing G; for a minimization the negative of the objective function is maximized. In the course of each iteration two things must be determined: a direction $\Delta\theta$ to proceed in, and the length K of the step to be taken along this direction. The value of K should be admissible; that is, it should be such that $\theta^{(1)} = \theta^{(0)} + K\Delta\theta$ should satisfy all bounds and constraints on the parameters.

Choice of Direction

Let p be the vector of partial derivatives $\frac{\partial G}{\partial \theta}$ at $\theta = \theta^{(0)}$. If R is any positive definite matrix, the direction $[\Delta\theta] = [R][p]$ is easily shown to be admissible. If $\theta^{(0)}$ is sufficiently near the maximum of G, the matrix

$$Q_{ij} = - \frac{\partial^2 G}{\partial \theta_i \partial \theta_j} \tag{106}$$

is positive definite and it can be shown that Q^{-1} is the most efficient choice for [R]. Using $[R] = [Q^{-1}]$ at all points constitutes the Newton-Raphson method, but this method is likely to yield non-admissible directions in the regions where [Q] is not positive definite. Greenstadt [60] has suggested a method to overcome this difficulty. The matrix [Q] may be represented by means of its eigenvalues and eigenvectors:

$$Q_{ij} = \sum_{k=1}^{n} V_{ik} V_{jk} |u_k| \tag{107}$$

in which V_{ik} is the ith element of the kth eigenvector of Q and u_k is the kth eigenvalue of Q. We now set

$$R_{ij} = \sum_{k=1} V_{ik} V_{jk} \left| u_k \right|^{-1} \tag{108}$$

This R is always positive definite and coincides with Q where the latter itself is positive definite.

Choice of the Step Size

The value of step size K is determined by a quadratic search involving the usual interpolation-extrapolation procedure. Let K_0 be a positive number such that $\theta^{(0)} + K_0 \Delta\theta$ is in the admissible region. Initially, one sets $K_1 = \min(K_0, 1)$ and $\theta^{(1)} = \theta^{(0)} + K_1 \Delta\theta$ and evaluates $G_0 = G(\theta^{(0)})$ and $G_1 = G(\theta^{(0)} + K_1 \Delta\theta)$. Define

$$F_1(K) = G_0 + \frac{1}{K_1}[p^T][R][p]k + \frac{1}{2K}(G_1 - G_0 - [p^T][R][p])K^2 \tag{109}$$

in which

$[p^T]$ = transpose of the matrix $[p]$

This function is obtained by setting $F(K) = G(\theta^{(0)} + K[\Delta\theta])$. The function F(K) is then approximated by a parabola $a + bK + cK^2$, matching the values of the parabola at K = 0 and $K = K_1$ and its derivatives at K = 0.

$$\left.\frac{dF}{dK}\right|_{K=0} = \left.\frac{\partial G}{\partial \theta}\right|_{\theta=\theta^{(0)}} \cdot [\Delta\theta] = p \cdot \Delta\theta = \Sigma\, p_i \Delta\theta_i \tag{110}$$

Now, compute the value of K, say K_2 which maximizes this parabola. The steps in this maximization are as follows:

(a) If the $\theta^{(0)} + K_2\ \theta$ is infeasible, truncate K_2 appropriately.

(b) If $G_1 > G_0$, compute the value $K = K_2$ which maximizes $F(K)$. If $K_1 - K_2 < 0.1$ set $K = K_1$. Otherwise, compute $G_2 = G(\theta^{(0)} + K_2[R][p])$, and set $K = K_1$ or $K = K_2$ depending on whether G_1 is or is not greater than G_2. This is the <u>extrapolation</u> procedure.

(c) Interpolation

(i) If $G_1 > G_0$ again compute K_2 as before, and it is guaranteed that $0 < K_2 < K_1$. Now compute G_2 and set $K = K_2$, if $F_2 > F_0$. Otherwise, replace K_1 with K_2, define a new parabola $F_2(K)$, find its maximum, etc.

(ii) If $G_1 < G_0$, set $K_3 = \max(K_2, 1/4\ K_1)$ and compute $G_3 = G(\theta^{(0)} + K_3 Rp)$. If $G_3 > G_0$ set $K = K_3$ and proceed to the next iteration. Otherwise replace K_2 with K_3, draw a new parabola, and reproceed. This method is called the absolute Newton method.

If the model fits the data reasonably well, $y_u - f(x_u, \theta)$ will be small and hence $Q_{ij} = \dfrac{\partial^2 S}{\partial\theta_i \partial\theta_j}$ may be quite accurately approximated by the term involving only the product of first derivatives. Let the elements of this matrix be

$$QA_{ij} = 2\,\frac{\partial f_u}{\partial\theta_i}\,\frac{\partial f_u}{\partial\theta_j} \qquad (111)$$

The approach suggested by Greenstadt is again used to

insure positive definiteness of Q_A for all criteria functions. The matrix Q_A^{-1} used to calculated $\Delta\theta$ is calculated from

$$(Q_A)^{-1}_{ij} = \sum_k V_{ik} V_{jk} \left| u_k \right|^{-1} \qquad (112)$$

where V_{ik} is the ith element of the kth eigenvector of Q_A and u_k is the kth eigenvalue of Q_A. This method is called the absolute Gauss method.

A general computer program is available from the IBM contributed program library [61]. This program calculates the mean values of the parameters according to all of the criteria listed in this chapter, the standard deviations of these parameters, the residuals at each point and value of the objective function G. In estimating the parameters in the models of Table 5.1 this general program was used. The data used for estimating these parameters was obtained from the Sacremento River water quality survey [61]. The data obtained for three different runs were used for analysis.

RESULTS OF PARAMETER ESTIMATION

The parameters appearing in the three models (3, 4 and 5) presented in the chapter have been obtained using two different criteria -- least squares and maximum likelihood. Results of these estimation procedures are presented in Tables 6.1 through Table 6.3. BOD and DO values calculated using the estimated values of parameters and models 3, 4 and 5 are plotted in

Figs. 6.1 through 6.2 and compared with the experimental data.

The results indicate that the maximum likelihood gives a better estimate (lower variance) than the least squares criteria. This is because maximum likelihood estimators are calculated to maximize the probability of having actually made the observations. In other words, maximum likelihood estimates maximize the probability of having taken the data. The least squares criteria, on the other hand, minimizes the square of the error. It may be noted that all the parameters lie within the range of values given in the literature. Other models in Table 5.1 were not tested by parameter estimation techniques as no data for N, the cell number, was available. The standard deviations were used for model discrimination. It can be seen from the values of standard deviations that Model 3 fits the set of data better than the other two models.

It can be concluded from this procedure that Bard's method serves as an efficient method for parameter estimation in nonlinear dynamic systems. The method converged reasonably fast for all three models (average computation time 21.3 minutes) and it converged to almost the same values of the parameters from several different starting points.

Table 6.1 Estimated Values of Parameters for Model 3 with two Criteria.

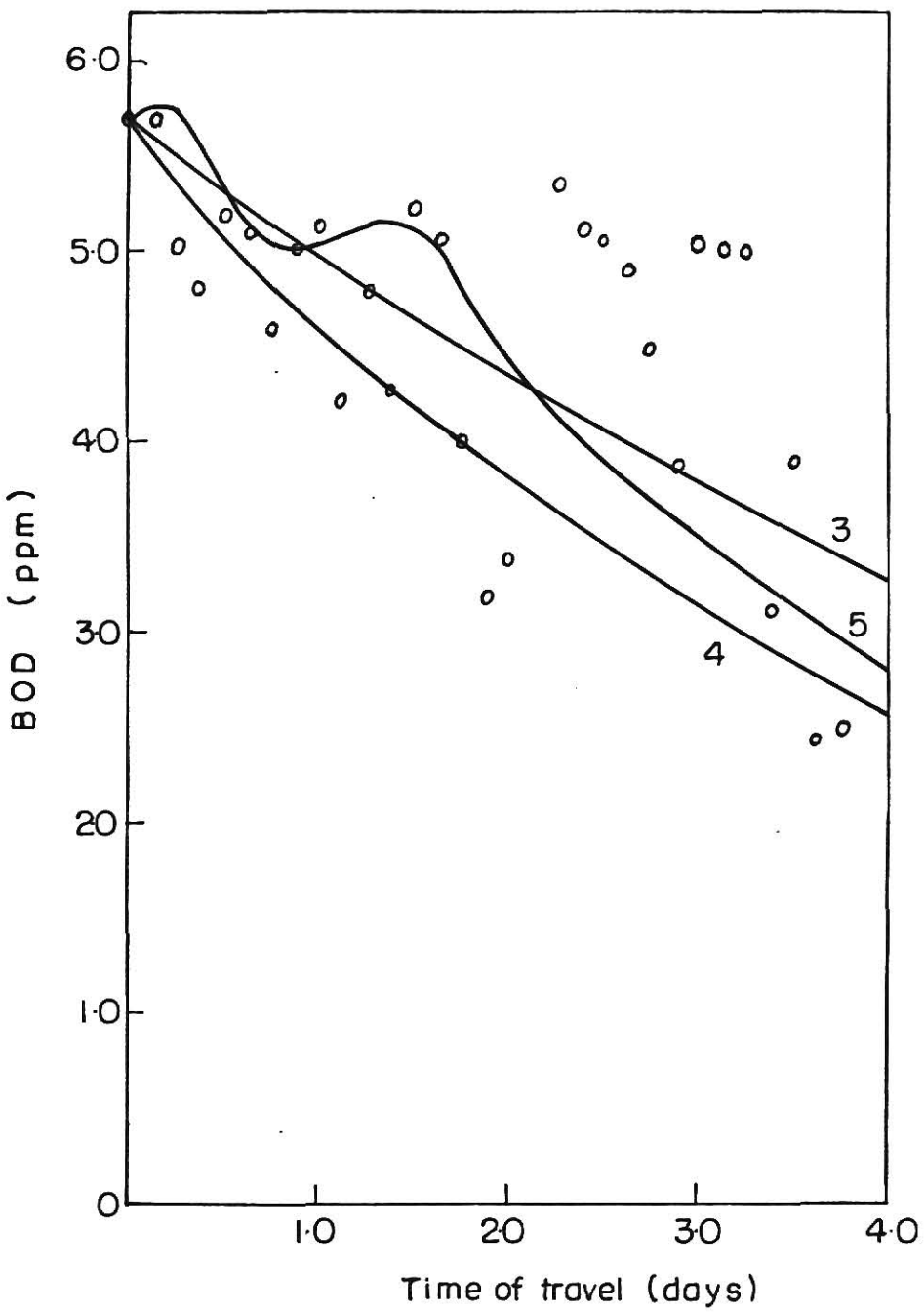
Parameter	Maximum Likelihood		Least Squares	
	Mean	Standard Deviation	Mean	Standard Deviation
k_1	0.2534×10^{-3}	0.403×10^{-5}	0.6566×10^{-4}	0.410×10^{-5}
k_2	0.4999	0.206×10^{-2}	0.4560	0.3239×10^{-1}
k_3	0.4671	0.7016×10^{-2}	0.1441	0.3020×10^{-2}
D_B	0.5000	0.4037×10^{-2}	0.4838	0.05759
L_a	0.9974	0.5162×10^{-2}	0.3873	0.004573
P_m	0.5002	0.6328×10^{-2}	1.933	0.1538
γ	0.4998	0.5079×10^{-2}	0.4907	0.06130
C_s	9.000	0.3243×10^{-2}	9.008	0.06028
K_s	0.199×10^{-3}	0.202×10^{-5}	0.219×10^{-3}	0.767×10^{-5}

Table 6.2 Estimated Values of Parameters for Model 4.

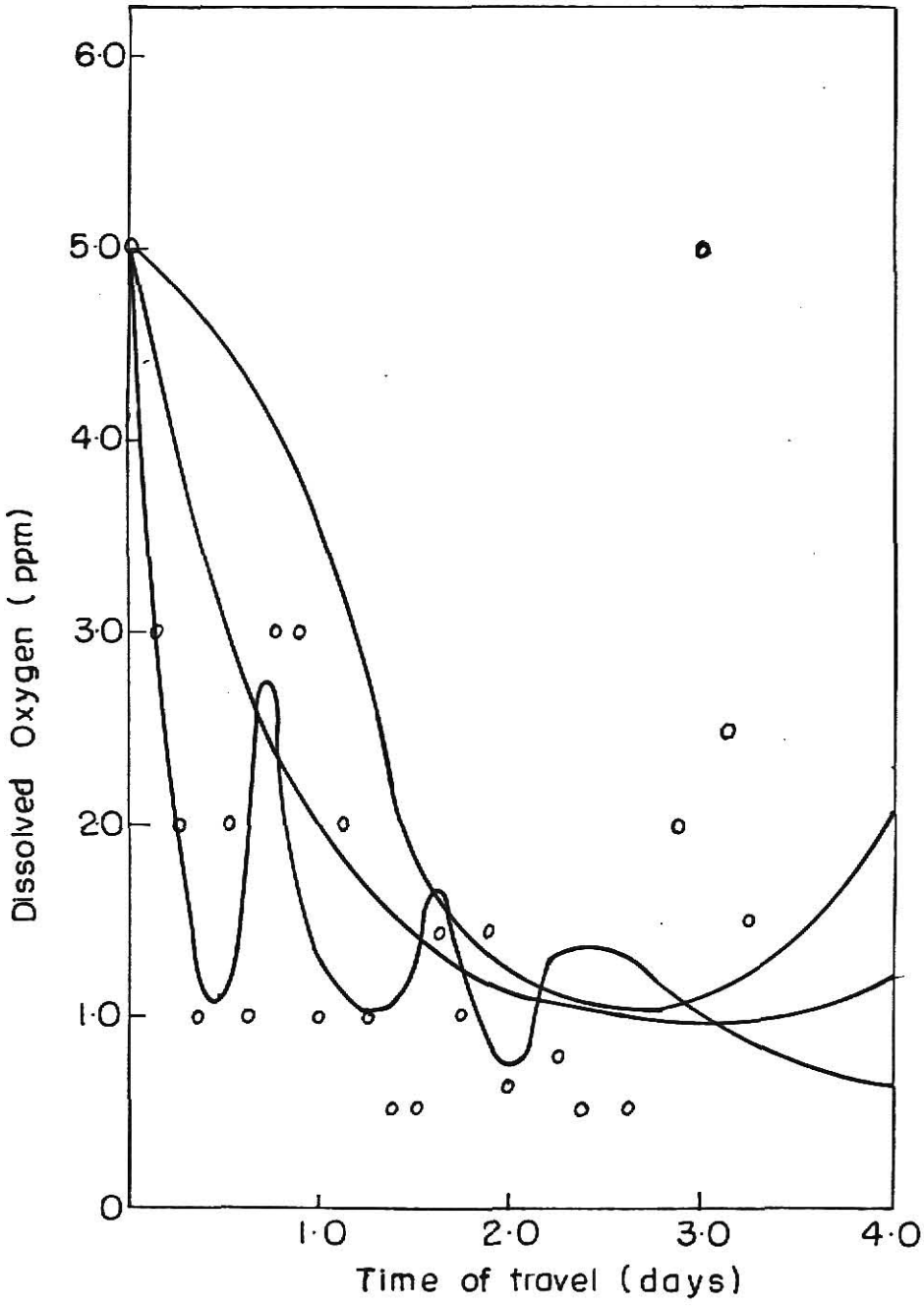
Parameter	Maximum likelihood		Least Squares	
	Mean	Standard Deviation	Mean	Standard Deviation
k_1	0.256×10^{-3}	0.39×10^{-5}	0.656×10^{-4}	0.316×10^{-5}
k_2	0.4993	0.199×10^{-2}	0.4560	0.03239
k_3	0.4671	0.7010×10^{-2}	0.4603	0.3029
D_B	0.50014	0.4002×10^{-2}	0.4315	0.06813
L_a	0.8974	0.500×10^{-2}	0.3317	0.01260
P_m	0.5002	0.6300×10^{-2}	0.99850	0.04021
γ	0.4998	0.4985×10^{-2}	0.4571	0.06035
C_s	9.007	0.003229	9.000	0.01740

Table 6.3 Estimated Values of Parameters for Model 5.

Parameters	Maximum Likelihood		Least Squares	
	Mean	Standard Deviation	Mean	Standard Deviation
k_1	0.350	4.0×10^{-5}	0.1395	0.02988
k_2	0.3859	0.0198	0.1370	0.03623
k_3	0.45448	0.06740	0.4776	0.031564
D_B	0.17553	0.00402	0.3455	0.04890
L_a	0.888909	0.01032	0.5642	0.0252211
P_m	0.25699	0.00502	1.446	0.2478
γ	0.62058	0.02817	0.4939	0.06725



Fig·6.1 Comparison of experimental and predicted values for BOD for Models 3,4 and 5, Sacremento River (1962).



Fig·6.2 Comparison of experimental and predicted values for Dissolved Oxygen for Models 3. 4 and 5, Sacremento River (1962)

CHAPTER VII

RECOMMENDATIONS FOR FUTURE WORK

Several problems arise from this study of water quality data. In this chapter some of these problems will be recognized and outlined in brief.

It was assumed in this analysis that the system variables can be treated two at a time. This is a preliminary study. All the water quality parameters can be treated simultaneously using multiple spectral analysis and multiple coherence techniques. A system-model with several inputs and outputs can be conceived for the stream (Fig. 7.1) and based on this model multiple coherence and spectral density matrix can be calculated. This analysis will yield more information regarding the basic phenomenological processes taking place in the stream.

Five water quality parameters were used for study in this investigation. Additional water quality parameters can be included in the future work. It was noted elsewhere in this thesis that continuously recorded data for BOD and coliform was not available at most of the stations. Techniques need to be developed for monitoring these two water quality parameters. An alternative would be to measure some other variable, which can be measured quickly, and correlate it with BOD.

Oxygen uptake has been used for this purpose [49] but the correlation of this variable with the standard 5 day BOD has not been confirmed. ATP can be used as a measure of biomass

or cell concentration [50]. Measurement of oxygen uptake and ATP and their correlation with BOD and biomass respectively may constitute a part of the future work.

A plug flow model was assumed for the streams under investigation. For estuaries and slow moving streams this assumption is no longer valid and axial diffusion must be included in the analysis. This approach has been attempted in the past but the analysis has been inadequate because of the use of incorrect boundary conditions. A study should be made of all the available boundary conditions for the diffusion model so that those most appropriate for each stream can be chosen for analysis.

The analysis in this study was restricted to two or three streams. More data is being made available as a large number of water quality sensing monitors are being installed. An attempt should be made to acquire additional water quality data. This should then be analyzed and compared with the results presented in this analysis. Additional data such as information in the periodic discharges to the stream should also be obtained and used in the analysis.

It is observed that there are often missing data due to difficulties in field operations. In this work these data were generated by interpolation. A technique has been developed for spectral analysis of data with regularly missed observations. This should be extended to handle irregularly missed observations so that the data may be handled without any manipulation.

New methods should be developed to generate the missing data in case the data is too irregularly missing. Harmonic analysis can be used for this purpose. The available data can be fitted to an equation consisting of the first two harmonics with a random component superimposed on them. Equation 113 is an example of such a model

$$X(t) = \bar{X}(t) + A_1 \sin wt + B_1 \cos wt + A_2 \sin 2wt + B_2 \cos 2wt + X_{RES}(t) \qquad (113)$$

The missing data can, then, be generated using this equation.

The parameter estimation results presented here were for only one set of data. This study can be extended by using data from several rivers. For some rivers, no BOD data is available. To handle such cases, the estimation technique should be modified to enable parameter estimation when only one state variable is measured. More parameter estimation should be done with the diffusion model to test the validity of this model. The parameter estimation methods should be modified for these cases.

Lastly, results of these studies should be used to develop a more economical and more precise water pollution control program.

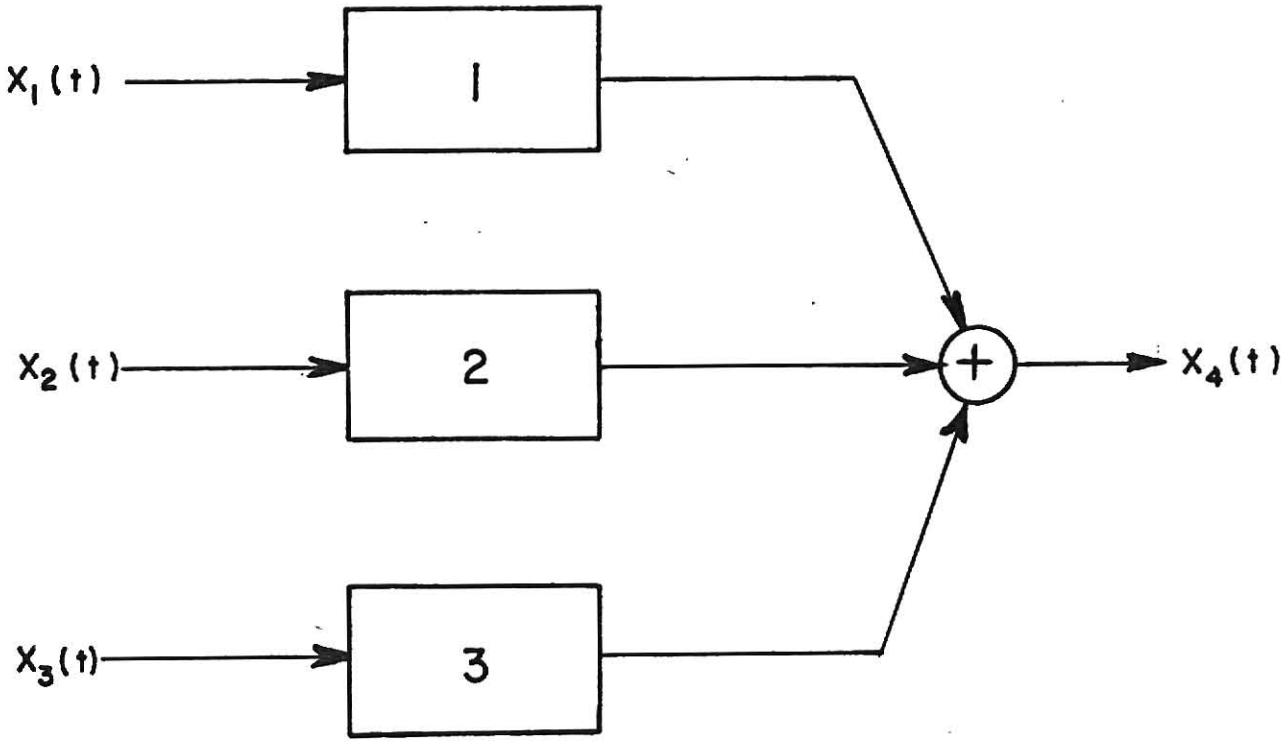


Fig. 7-1 System to be considered for multiple coherence calculations.

NOMENCLATURE

$A(x, t)$	variable stream cross-sectional area square miles
A	harmonic coefficient
B	harmonic coefficient
$B(w)$	spectral window
$C(p)$	cospectrum
C	amplitude of the harmonics
C	concentration of the substance (oxygen) in stream, mg/ℓ
C_s	saturation DO level, mg/ℓ
D	axial diffusion coefficient, $(\text{miles})^2$/day
D_B	DO depletion due to benthal demand
E	expected value of (an operator)
G	objective function for parameter estimation
$H(w)$	Fourier transform of the unit impulse function
K	step size in Band's mocification
K_L	saturation constant, mg/ℓ
k	harmonic number
k_1	bacterial action rate constant, day^{-1}
k_2	reaeration constant, day^{-1}
k_3	sedimentation rate constant, day^{-1}
L	BOD level in stream, mg/ℓ
L_R	length of the reach under consideration in miles
L_a	BOD addition due to local runoff, mg/ℓ
M	total number of harmonics

m	total number of lags
m(t)	noise
N	total number of data points
N_{Pe}	Peclet number, dimensionless
N	number of microorganisms
n(t)	noise
P	rate of photosynthetic addition of oxygen, ml/day
P	lag number
Q	Hessian matrix
Q	stream flow cubic miles/day
Q(p)	quadraspectrum
$R(\tau)$	autocorrelation function
S(w)	raw power spectral estimates
SP(w)	smoothed spectral estimate
T	temperature, ^{o}C
T	fundamental period, days
T(w)	transfer function
t	time of travel, days
u	error at each experiment
U	average stream velocity, miles
$W(\tau)$	lag window
w	frequency, radians/day
X(t)	time series
$X_{RES}(t)$	residual series
X	distance along the stream, miles
Y	yield

Y(t)	time series
Z(t)	time series

SUBSCRIPTS

i	number of data points
j	number of data points
k	harmonic number
m	maximum value
x of time series X(t)	
y of time series Y(t)	

GREEK LETTERS

θ	phase angle
θ	unknown parameter to be estimated
u	mean
u	specific growth rate
$\rho(\tau)$	autocorrelation coefficient
σ^2	variance
τ	lag for time series, days

ACKNOWLEGEMENTS

The author wishes to express his sincere gratitude to his major advisor, Dr. L. T. Fan, for his excellent advice and constant encouragement.

The helpful suggestions and continual interest of Dr. Larry E. Erickson are deeply appreciated. The author also wishes to tank the staff of the Computing Center and the secretarial staff of the Chemical Engineering Department for their assistance.

The author's work was partially supported by the Kansas Agricultural Experiment Station and Project B-021 KAN, Office of Water Resources Research and Kansas Water Resources Research.

APPENDIX

BRIEF DESCRIPTION OF THE COMPUTER PROGRAMS

1. BMD02T [46]:

This program computes the antocovariance, power spectrum, cross-covariance, cross-spectrum, transfer function and coherence function of time series. A slight modification in the original deck was necessary to run it on the KSU IBM 360/50 computer. This program allows transformation and detrending of the input data. The output from the program includes:

a) Input data printed and plotted.

b) Autocovariance printed and plotted.

c) Power spectral estimates (power spectrum) printed and plotted.

d) Cross-covariance of two time series printed and plotted.

e) Cross-spectrum of two time series printed and plotted.

f) Phase shift between two time series printed and plotted.

g) Coherence function of two time series printed and plotted.

h) Transfer function of two time series printed and plotted.

Provision is also made in the program to print out error messages whenever the power spectral estimate becomes negative or the coherence square exceeds 1.1.

2. BMDX92 [47]

This program estimates autospectra, cross-spectra and coherences for stationary time series. Each series is decomposed into frequency components by means of finite Fourier transform and the required estimates are obtained by summing products of the transformed series. The output from this program includes:

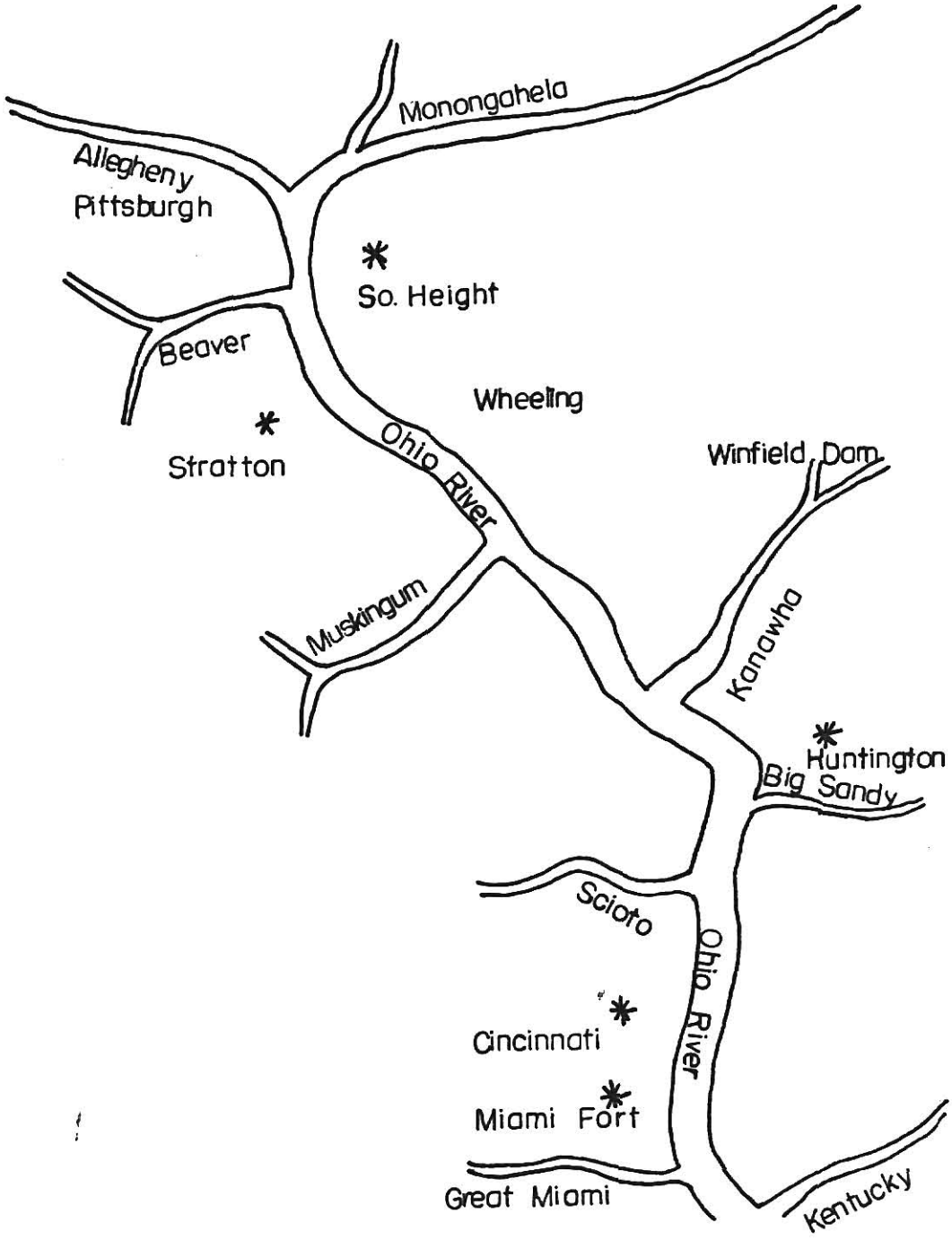
a) Graph of the frequency response function of the prefilter (optional)
b) Graph of the frequency window of the power spectrum estimates (optional)
c) Power spectrum for each series
d) Graphs of the power against frequency (optional)
e) Graphs of the logarithm of the power against frequency (optional)
f) Amplitude, phase, and coherence spectra for each pair of series
g) Filtered data (optional)

3. FUTSA (Forecasting, Using Time Series Analysis)

This program decomposes monthly time series data into components for trend, business activity level, seasonality and irregularity. Each component is described in the same units as the raw time series data. A user's manual is available. Three forecasting models can be used with this program.

APPENDIX B

Monitoring Stations Analyzed on the Ohio River



* indicates data were obtained for the station

REFERENCES

1. Klein, Louis, "River Pollution III. Control," Butterworths, London, (1966).

2. Camp, T. R., "Water and Its Impurities," Reinhold Publishing Corporation, New York, (1963).

3. Loucks, Daniel P., "A Probabilistic Analysis of Waste Treatment Systems," Water Resources Center, Cornell University, Ithaca, New York, (1965).

4. Cooke, Lioyd, M., "Cleaning our Environment - the Chemical Basis for Action," an ACS Report, C and E News, Sept., 8, 1969, pp. 58-71.

5. Brittan, M. R., "Probability Analysis Applied to the Development of Synthetic Hydrology for the Colorado River," University of Colorado, Bureau of Economic Research, Boulder, (1961).

6. Thomann, R. V., "Mathematical Model for Dissolved Oxygen," Journal of the Sanitary Engineering Division, ASCE, 89, No. SA5, Proc. Paper 3680, pp. 1-30, (1963).

7. DiToro, D. M. and D. J. O'Connor, "The Distribution of Dissolved Oxygen in a Stream with Time Varying Velocity," Water Resources Research, 4, No. 3, pp. 639-646, (1968).

8. Dresnack, Robert and W. E. Dobbins, "Numerical Analysis of BOD and DO Profiles," Journal of the Sanitary Engineering Division, ASCE, 94, No. SA5, Proc. Paper 6139, pp. 789-807, (1968).

9. Levenspiel, O., "Chemical Reaction Engineering," John Wiley and Sons, New York, (1962).

10. O'Connor, D. J. and R. V. Thomann, "Stream Pollution for Pollution Control," Proceedings of the IBM Scientific Computing Symposium on Environmental Sciences, Nov. 14-16, Thomas J. Watson Research Center, New York, (1968).

11. Streeter, H. W. and E. B. Phelps, "A Study of the Pollution and Natural Purification of the Ohio Rivers," Public Health Bulletin No. 146, U. S. Department of Health, Education and Welfare, Public Health Service, Washington, D. C., p. 95, (1925).

12. O'Connor, D. J., "Oxygen Balance of Estuary," ASCE J. Sanitary Engineering Division, 86, No. SA3, Paper 2472, pp. 35-55, (1960).

13. Dobbins, W. E., "BOD and Oxygen Relationships in Streams," Journal of the Sanitary Engineering Division, ASCE, 90, No. SA3, pp. 53-78, (1961).

14. Thayer, R. P. and R. G. Krutchkoff, "A Stochastic Model for Pollution and Dissolved Oxygen in Streams," Water Resources Research Center, Virginia Polytechnic Institute, Blackburg, Virginia, p. 130, (1966).

15. Thomann, R. V., "Systems Analysis and Simulation in Water Quality Management," Proceedings of IBM Scientific Computing Symposium on Water and Air Resources Management (1967).

16. Pyatt, E. E., "On Determining Pollutant Distribution in Tidal Estuary," U. S. Geological Survey-Water Supply, Paper 1586f (1964).

17. Kothandaraman, V. and B. Ewing, "Probabilistic Analysis of Waste Water Treatment and Disposal Systems," Research Report No. 14, Water Resources Center, University of Illinois, Urbana, Illinois, pp. 1-158, (1968).

18. Thomann, R. V., "Recent Results for Mathematical Model of Water-Pollution Control in Delaware," Water Resources Research, 1, No. 3, 3rd Quarter, pp. 349-359, (1965).

19. Thomann, R. V. and M. J. Sobel, "Estuarine Water Quality Management and Forecasting," Journal of Sanitary Engineering Division, ASCE, 90, No. SA5, Paper 4116, pp. 9-36, (1966).

20. McGanhey, P. H., "Engineering Management of Water Quality," McGraw-Hill Series in Sanitary Science and Water Resources Engineering, New York, (1968).

21. Wastler, R. A., Application of Spectral Analysis to Stream and Estuary Field Survey - I. Individual Power Spectra, Publication No. 999-WP-7, U. S. Public Health Service, Washington, C. D., pp. 1-30, (1963).

22. Gunnerson, C. Y., Optimizing Sampling Intervals in Tidal Estuaries, Journal of Sanitary Engineering Division, Proc. Am. Society of Civil Engineers, 92, No. SA2, pp. 103-125, (1966).

23. Thomann, R. V., Time Series Analysis of Water Quality Data, Journal of Sanitary Engineering Division, ASCE, 93, No. SA1, pp. 1-23, (1967).

24. McCartney, David and B. Norman, "Continuous Recording of Water Quality of the Delaware Estuary," Journal Amer. Water Works Assn., 54, No. 10, pp. 1193-1200, (1962).

25. Thomann, R. V., "Analysis of Variables in Waste Treatment Plant Performance Using Time Series Techniques," Presented at ASCE Second National Symposium on Sanitary Engineering Research, Development and Design, (July 15, 1969).

26. IBM Scientific Subroutine Package, The IBM Company, White Plains, New York, Version II.

27. Cooley, J. W. and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," Math. of Computation, 19, pp. 297-301, (1965).

28. Akaike Hirotugn, "Some Problems in the Application of the Cross Spectral Method," Spectral Analysis of Time Series, Ed. B. Harris, John Wiley and Sons, Inc., New York, (1967).

29. Chang, S. S. L., "Synthesis of Optimum Control System," McGraw-Hill Book Company, Inc., New York, (1961).

30. Grenander, U. and M. Rosenblatt, "Statistical Analysis of Stationary Time Series," John Wiley and Sons, New York, (1957).

31. Davenport, W. B. and W. L. Root, "An Introduction to the Theory of Random Signals and Noise," McGraw-Hill Book Co. Inc., New York, (1958).

32. Jenkins, G. M. and D. G. Watts, "Spectral Analysis and its Applications," Holden-day, San Francisco, California, (1969).

33. Tukey, John W., "An Introduction to the Calculations of Numerical Spectrum Analysis," Paper in "Spectral Analysis of Time Series, Ed. B. Harris, John Wiley and Sons, Inc., New York, (1967).

34. Bartlett, M. S., "An Introduction to Stochastic Processes," Cambridge University Press, London, (1961).

35. Parzen, E., "Mathematical Considerations in the Estimation of Spectra," Technometrics, 3, pp. 167-190, (1961).

36. Blackman, R. B. and J. W. Tukey, "The Measurement of Power Spectra," Dover Publications, Inc., New York, (1958).

37. Hannan, E. J., "Time Series Analysis," Methuen and Co. Ltd., London, (1963).

38. Timmons, D. L., "Noise Analysis Applied to a Liquid-Solid Fluidized Bed," M. S. Thesis, Kansas State University, (1966).

39. Lee, Y. W., "Statistical Theory of Communication," John Wiley and Sons, Inc., (1960).

40. Iturbe, I. R., "The Application of Cross Spectral Analysis to Hydrologic Time Series," Hydrology Paper 24, Colorado State University, Fort Collins, Colorado, (1967).

41. La Barriere R. Pallu De La, "Optimal Control Theory," W. B. Saunders Company, Philadelphia, (1967).

42. Granger, C. W. J. and M. Hartanka, "Spectral Analysis of Economic Time Series," Princeton University Press, New Jersey, (1964).

43. Jenkins, G. M., "Cross-spectral Analysis and the Estimation of Linear Open Loop Transfer Function in Time Series Analysis," Ed, M. Rosenblatt, pp. 267-276.

44. Solodovnikov, V. V., Introduction to Statistical Dynamics of Automatic Control Systems, Dover Publ. Inc., New York, (1960).

45. "Catalog of Information on Water Data," U. S. Department of Interior, Geological Survey, Washington, D. C., (1967).

46. BMDO2T Computer Program for "Autocovariance and Power Spectral Analysis," Health Sciences Computing Facility, University of California, Los Angeles, California.

47. BMDX92 Computer Program for "Time Series Spectrum Estimation," Health Sciences Computing Facility, University of California, Los Angeles, California.

48. Whitman, J. P., "Forecasting, Using Time Series Analysis," M. S. Thesis, Department of Chemical Engineering, the University of Tennessee, Memphis, Tennessee, (1968).

49. Arthur, R. M. and A Hursta, "Continuous Recording of Biochemical Oxygen Demand," Water and Wastes Engineering, pp. 25-26, (March, 1969).

50. E'Estachio, A. J., "Bioinstrumental Approach to Rapid Microbiology," Bulletin, E. I. duPont and de Nemours and Co., Inc., Instruments Products Division, Wilmington, Delaware, (1969).

51. Goodman, N. R., "Measurement of Matrix Frequency Response Functions and Multiple Coherence Functions," Technical Report AFFDL-TR-65-58, (June, 1965).

52. Erickson, L. E., A. E. Humphrey and A. Prokop, "Growth Models of Cultures with Two Liquid Phases," Biotechnology and Bioengineering, XI, pp. 449-466, (1969).

53. O'Connor, D. J. and W. E. Dobbins, "Mechanism of Reseration in Natural Streams," Transactions of Americal Society of Civil Engineers, 123, p. 641, (1958).

54. Thackston, E. L. and P. A. Krenkel, "Reaeration Prediction in Streams," Journal of Sanitary Engineering Division, ASCE, 95, SA1, (February, 1969).

55. Bard, Y., "A Function Maximization Method with Application to Parameter Estimation," IBM New York Scientific Center Report No. 320-2902, (1967).

56. Gauss, C. F., "Theory of Least Squares," English translation by Hale F. Trotter, Princeton University, (1957).

57. Davidon, W. C., "Variable metric Minimization for Nonlinear Problems," AEC Research and Development Report, ANL-599 (Rev.)

58. Buehler, R. J., Shah, B. V. and Oscar Kempthorne, Chemical Engineering Symposium Series No. 50, 60, 1, (1964).

59. Marquardt, D. W., "An Algorithm for Least Squares Estimation of Nonlinear Parameters," J. Soc. Ind. Appl. Math., 2, 431-441, (1963).

60. Greenstadt, J., "On the Relative Efficiencies of Gradient Methods," Math. of Comp., July, (1967).

61. Bard, Y., "IBM Contributed Program Library," 360D 13.6.003, (1967).

62. Resources Agency of California Department of Water Resources: Sacremento River Water Pollution Survey, Bulletin No. 111 and Appendix B, (August, 1962).

USE OF SPECTRAL ANALYSIS AND NONLINEAR PARAMETER ESTIMATION TECHNIQUES IN ANALYZING WATER QUALITY INFORMATION SYSTEM

by

JAIPRAKASH SHASTRY

B. Tech., Indian Institute of Technology, Bombay, 1968

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Chemical Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1971

ABSTRACT

Considerable effort has been expended to collect water quality information by installing automatic water quality monitoring and data collection stations on a number of streams and estuaries. This thesis deals with modern techniques for analyzing this data.

The technique of spectral analysis is introduced and employed to determine the cause-effect relationships that influence water quality. For this analysis monitored data available from the Coosa, Detroit, Missouri and Ohio river are used. Using the results of spectral analysis, mathematical models are developed to represent the phenomenological processes taking place in the stream. Once the structural form of a model has been obtained nonlinear parameter estimation techniques are used to determine the values of the constants in the model which are most appropriate for a particular stream or river. Specifically, Bard's modification of Gauss-Newton method is used for parameter identification,

Finally, recommendations are made for future work in this area.