ILLEGIBLE DOCUMENT

THE FOLLOWING DOCUMENT(S) IS OF POOR LEGIBILITY IN THE ORIGINAL

THIS IS THE BEST COPY AVAILABLE

TIME SERIES ANALYSIS AND PARAMETRIC ESTIMATION IN AIR POLLUTION

by

SUBHASH CHANDER SARIN

BSc. (Mech. Eng.)., University of Delhi, India, 1970

A MASTER'S THESIS

submitted in partial fulfillment of

the requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1973

Approved by

Major Professor

THIS BOOK
CONTAINS
NUMEROUS PAGES
WITH THE ORIGINAL
PRINTING BEING
SKEWED
DIFFERENTLY FROM
THE TOP OF THE
PAGE TO THE
BOTTOM.

THIS IS AS RECEIVED
FROM THE
CUSTOMER.

ACKNOWLEDGEMENT

The author wishes to express his sincere gratitude to his major professor, Dr. E. Stanley Lee for his esteemed guidance and encouragement in the preparation of master's thesis and throughout his graduate studies. The author is also grateful to Dr. L. E. Grosh for his help in programming and giving some useful suggestions.

Thanks are also due to Ms. Marie Jirak who typed the script of this thesis.

This work was partially supported by Grant 1 R01 AP-00895-01, National Air Pollution Control Administration, Environmental Protection Agency. The data used in this work were furnished by Air Pollution Control District, Los Angeles, California and Bay Area Air Pollution Control District, San Francisco, California.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Air Pollution is the problem of today. It affects every single individual on this earth and yet it is only he who is it's prime cause. Any phenomenon natural or man-made occuring in our environment has associated with it it's good and bad effects. Air pollution is the example of one of these bad effects. The natural production of carbon monoxide and nitrogen in the atmosphere are subsequently associated with other natural phenomena which mar the excessive accummulation of any kind. The man-made phenomenon like the emissions from the automobiles, industrial plants and airborne jets result in the excessive contamination of the atmosphere beyond it's recuperative power and man has to device means to avoid it as nature does.

Attempts have been made to simulate the real world situation using mathematical models. These models are then used to project the future atmospheric conditions and thus taking appropriate actions in advance to avoid accumulation. Two approaches are available in general for the development of these mathematical models. The first is based on the historic data available about the pollutant concentration and other meterological conditions and the second on the distribution of the sources and the actual transportation or diffusion of the pollutants. To date, there has been quite some study done using the second approach yet there has not been any work reported using the first approach. The earliest work reported on the air pollution modelling dates back to the

start of the last decade. The models proposed in the early and mid 60's and a few afterwards are based on the second approach [1-6]. These diffusion models were proposed individually for various urban areas. These models though performed well for the local areas were unsuccessful when applied to large metropolitan areas. In the late 60's and early 70's models based on other criterias were proposed. Takamatsu et. al. (14) proposed a computer control system for preventing air pollution. Croke and Roberts (13) proposed a model on the integration of air resource management with urban and regional planning. Marcus and Harrington (15) applied decision theory approach to develop abatement policies to control air pollution. Each of these models lack in their applicability to the complex air pollution problem.

In this paper are proposed two new approaches to develop mathematical models based on the two alternatives mentioned earlier. The first approach proposed is based on the second alternative. One metropolitan area is divided into small regions and a dynamic model is developed based on the material balance between the two adjacent regions to determine the amount of pollutant present at any time t in the region. Dividing an area into regions has the following advantages. The first is that it tends to represent the real world situation. Secondly, it helps to determine local concentrations and it's dependence on the adjacent regions. Thirdly, as the regions are small, the amount of pollutant emitted can be assumed to be distributed uniformly throughout the region. The dynamic model is developed here considering two regions and one pollutant. The model contains parameters like the volume of the region and the area

of the wind front which can not be determined directly. Quasilinearization is used for the estimation of these parameters.

The second approach is based on the statistical approach. Since the amount of pollutnat emitted at any time is a stochastic variable, this can be called as the stochastic modelling of the air pollution problem. Application of this approach has been reported in the model building of water pollution control problems [29-32]. A set of observations on the amount of pollutant taken through time at a station forms a time series. This time series contains all the information regarding the behavior of the pollutant in the region and it is analyzed to explore this information. The analysis consists in determining the cyclic variation of the pollutant concentration. The intertermporal relationship of the series which is represented by it's autocorrelation function at various time lags helps determine the cyclic variations. The Fourier transform of the autocorrelation function transforms it to the frequency domain and is called the spectral density function. It distributes the total variance over various frequency bands and is easy for physical interpretation than the autocorrelation function and therefore preferred. High peaks associated with high variance over a frequency determine the cyclic variations of the pollutant concentrations. The relationship between two or more processes forming two or more time series can similarly be carried out by studying the crosstemporal relationship from the cross correlation function and then studying the cross spectrum. The cross spectrum consists of frequency and phase. The former determines the correlation between the two processes at various frequencies and the later the corresponding phase.

In this paper are presented spectral and cross spectral analysis of the data of the San Francisco bay area and the greater Los Angeles area. Spectral analysis is done for each pollutant at each station and a model is proposed for each using regression analysis. Cross spectral analysis is done for the same pollutants at different stations taken in pair and for different pollutants at one station. The former determines how much are the amounts of the same pollutant at two adjacent stations related while the later determines how the variations of different pollutants in the same region are interrelated.

CHAPTER 2

AIR POLLUTION MODELLING AND PARAMETRIC ESTIMATION

INTRODUCTION

The complexities involved in an air pollution problem concerning an urban area are due to the diversity of the types of pollutants, emission rates and conditions of emission. Under the type of pollutant may be classified composition, radioactivity and the possibility of producing another pollutant by chemical reactions and under the condition of emission may be classified heat, volume, height of release, proximity to structures and whether the source is mobile or not. Meteorological conditions may be next to be considered as they effect the emission rates. Then there are topographical features like relief, roughness and thermal properties of the surface which influence the atmospheric processes.

An air pollution model tends to simulate such a complex existing situation. Attempts are being made to develop a better mathematical model to represent the real word situation. In this chapter some of the diffusion models developed so far are discussed. The various kinds of pollutants, their sources and effects are described. The cause of air pollution is explained and a new model is proposed. The model is developed for two regions and one pollutant. The model developed contains unknown quantities like the volume of the region, the velocity and area of the wind front or in other words the fraction of pollutant being carried from one region to another which cannot be determined directly. Quasi-linearization is used for their estimation. The computational procedure for this technique is explained.

Pollutants, Their Sources and Effects

In an effort to classify the pollutants, it is convenient to consider two general groups [8]; (a) those emitted directly from identifiable sources, and (b) those produced in the air by interaction of two or more primary pollutants, or by reaction with normal atmospheric constituents with or without photoactivation. Under the first category are included the following: fine solids (less than 100 micro inch. diameter), coarse particals (greater than 100 micro inch. diameter), sulphur compounds, organic compounds, nitrogen compounds, carbon compounds, halogen compounds and radioactive compounds. The fine aerosols include particles of metal, carbon, tar, resin, pollen, fungi, bacteria, oxides, nitrates, sulfates, chlorides, florides, silicates, and host of other species obviously overlapping most of the specific categories. Combustion of sulphur containing fuels contributes large amounts of sulphur dioxide and some sulphur trioxode, many industrial processes and waste disposal practice generate hydrogen sulphides, and the nauseous odors of mercaptons are well recognized associates of some pulp manufacturing and photochemical processes. Organic compounds released to typical community air supplies include a very large number of saturated and unsaturated aliphates, and aromatic hydrocarbons together with a variety of the oxygenerated and halogenated derivatives. The nitrogen compounds most abundantly generated and released are nitric oxides, nitrogen dioxode and ammonia. Carbon dioxide and carbon monoxide arise in huge amounts respectively from the complete incomplete combustion of carbonaceous fuels. Certain inorganic halogen compounds, among them hydrogen

floride and chloride, are produced from matallurgical and other industrial processes.

Among the second category are included the pollutants ozone, formaldehyde, organic hydro-peroxides, PAN (peroxyacyl nitrates) and other very reactive compounds. These are formed due to photochemical reactions.

The effects associated with air pollution in short are visibility reduction, material damage, physiological effects on man and domestic animals and phychological effects. Geographical location of a city can intensify these effects, because of valleys and other air movements etc. Therefore the distribution of pollutant sources taking into consideration the geography of the area is an important factor. The increasing concentration of the pollutants over the more industrialized cities threatens the very existence of human beings. Air pollution is therefore one of the most important problems facing the world today.

Inversion layer.

The accummulation of pollutants at a place occurs primarily because of the formation of inversion layer. This inversion layer is formed because of the inversion of the temperature gradient along the height at a location. If at a location the temperature gradient in the environment is more than that of the adiabatic lapse rate of any air, the air containing pollutants, as it expands in the regions of low pressure, has a tendency to move upward, still higher is the atmosphere. As the temperature gradient in the environment decreases the air containing pollutant tends to move back to its origin. But as the temperature gradient in the

environment becomes negative, that is the temperature increases with the height, the air containing pollutants becomes denser and can't rise any more. This results in the formation of a layer which resists the upward transport of pollutants thus resulting in their accummulation. An inversion layer has its base and top and there can be more than one inversion layers existing at the same time in the atmosphere. The inversion layer is varying in its height.

Mathematical Models

Many attempts have been made in recent years to develop mathematical models to represent the real world situation. Smith (1) developed a simple model of diffusion from an area source consisting of a shallow box with base at the ground, top at an assumed upper limit of vertical mixing height h and sides positioned to enclose the city as an area source. Mixing is supposed to be complete through the box and the wind passes through the sides. For a given box size, the predicted pollutant concentration is a function only of the pollutant emission rate and a suitable average wind speed. If the city source strength is Q mass per unit area per unit time, the average wind speed is V and the downward length of the rectangular base as 1 then the equilibrium concentration is given by

$$X_e = \frac{Q1}{Vh} \text{ mass per unit volume}$$

This would be 90% attained in a time given by

$$t = 2.3 \frac{1}{V}$$

Although this model is admittedly crude yet it is defensible on the grounds of simplicity, the magnitude of errors in the estimation of source strength, and the possibility of keying the model to pollutants for which both the emission rate and surface concentration are measurable over the lower surface of the box.

Pooler (2) used an empirical equation with published summaries of wind direction and speed frequencies obtained at the Nashville Weather

Bureau station to compute patterns of mean monthly relative concentrations. The patterns consisted of the relative concentration contributions from grid points at one mile intervals to the central grid point. The model used for estimating the relative concentration is

$$\frac{X}{Q} = \frac{2 \exp(-z^2/2\sigma_Z^2)}{(2\pi)^{0.5}(1/8\pi x)\ u\sigma_Z}$$

This is the relative concentration at the ground level from a continuous point source at height z, the surface layer wind velocity u, assuming a Gaussian distribution for the concentration with variance σ_Z in the vertical, reflection from the surface, and mass continuity, and assuming a uniform horizontal distribution over a sector of angle $\pi/8$. The vertical variance was assumed as,

$$\sigma_Z^2 = 2Lu^{\alpha}t^{\beta}$$

where L, α and β are constants. While the absolute magnitude of the concentration could not be verified, due to cumulative uncertainties in the source inventory the error of prediction was less than a factor of two. Predicted and observed patterns were generally similar, with a super-imposed topographic effect on observed concentration. Hillside locations facing toward the source showed generally higher than predicted values, and valley locations sheltered from the sources by intervening terrain, lower values.

Turner (3) has proposed a working model for the diffusion of gases from multiple sources in an urban area. A diffusion equation modified to use area instead of point source was used with a source inventory of

sulphur dioxide emission to calculate 24-hours concentrations at one mile intervals. Wind velocity and stability were averaged by two 2 hour periods to evaluate the diffusion equation. One effective source height was used. The model used for the determination of the relative concentration is

$$\frac{X}{Q} = \frac{\exp\ (-\ 1/2\ (y^2/\sigma_y^2 + Z^2/\sigma_Z))}{\pi u\ \sigma_y\ \sigma_Z}$$

where σ_y is the horizontal dispersion coefficient i.e. variance in the horizontal direction and y the horizontal distance travelled. Sulphur dioxide was assumed to be oxidized or adsorbed on particulates exponentially with a half life of 4 hours. This model was found to overcalculate except for local sources.

Clark (4) proposed a simple diffusion model for calculating point concentrations from multiple sources. The working equation used to calculate the probable concentration is

$$X = \frac{2Q\ \exp\ (-\ 1/2\ Z^2/\sigma_Z^2)}{\bar{U}\ (2\pi)^{0.5}\ \pi/8\ \sigma_Z}$$

where $\bar{U}$ is the mean wind velocity. The area under study was divided into sixteen sectors of 22.5 degrees each of an assumed circle with the air monitoring station at the center. The concentration resulting at the center of the circle from area sources, considering only sulphur dioxide, represented by the arcs of a sector of finite thickness is a function of distance if the stability, wind speed, and source emission height are

specified and emission rate per unit area is unity. The sectors were divided into four areas. For each of these four areas within the 16 sectors a source strength Q was determined as the sum of the proportionate sources (given by source inventory) in the area. The concentration at the monotoring station was calculated as

$$X = \sum_{1}^{4} \left(\frac{X}{Q}\right)_n Q_n$$

where n is the area number. The model is quite effective as it's results are within the precision of the source inventory. This has application as a forecasting tool and aid to evaluating urban air pollution sources and air quality measurement. The model was adaptable to other pollutants and other locations within Cincinnati.

Koogler et. al. (12) developed a dispersion model for Jacksonville, Florida, designed to account for wind speed and direction, atmospheric stability, the emission rates of point and area sources and source height. The basic dispersion equation used was

$$X = \frac{Q.D}{\pi U \sigma_y \sigma_Z} \exp\left[-1/2\ (y/\sigma_y)^2 - 1/2\ \left(\frac{Z}{\sigma_Z}\right)^2\right]$$

where D is a unitless decay factor of form D = exp (-0.693 t/T), where t is the decay time and T is the pollutant half life. The wind speed was assumed to be constant upto about 2.5 times the average building height and above that to vary according to the expression;

$$\frac{U_1}{U_2} = (Z_1/Z_2)^{\frac{n}{2} - n}$$

n is the parameter depending on the atmospheric stability. The predicted

values with this model were quite accurate being significant at .1% confidence level when tested by the chi-square test. The pollutant level used for testing the model was quite low and this poses the question of the effectiveness of the model in areas with a higher pollutant concentration.

Miller and Holzworth (6) used a semi-empirical model, adjusted by using model to find a vertual source strength. The model was developed with the assumption that the continuous pollutant sources are uniformly distributed over the urban area and the vertical diffusion occurs until the effluent from each line source reaches the top of the mixing layer, after which the effluent is uniformly distributed through the mixing layer. The model for average relative concentration is

$$\frac{\bar{X}}{Q} = \frac{U}{S}\left(\int_{t_1}^{t_L} \int_{t_1}^{t_L} \frac{2}{\sqrt{2\pi}\,\sigma_Z}\, dt\, dt + \int_{t_L}^{t_S} \int_{t_L}^{t_S} \frac{1}{L}\, dt\, dt \right)$$

where t_1 is the travel time of the nearest source, t_L the time in which the edge of the plume reaches the mixing layer of depth L, and t_S the travel time of the most distant source over the along wind length S, of the city. The first term is the concentration developed during the time the plume of pollutant reaches the mixing depth and the second term the concentration developed when the effluent is uniformly distributed through the mixing layer. Calculated average concentration of So_2 and No_x were within 2 pphm of observed average value in some cases and in some cases in overcalculates.

Then there have been models developed on criterias other than those

of atmospheric dispersion. Croke and Roberts (13) in their paper describe an approach to the development of the implementation plans for the stationary sources of sulphur dioxide and particulates based on the integration of air resource management with urban and regional planning. This was achieved through the establishment of well defined emission rights associated with each parcel of land in the region. An atmospheric dispersion model was used to design the required legislation. Within the context of this system all of the usual control techniques, such as emission control devices, fuel switching, stack height ordiances, etc., and all of the usual administrative and enforcement procedures, such as permit systems, inspection, emission inventories, etc. can be applied. The methodology ensures that the desired long range regional air quality goals are attained by implementation plans which have an inherent economic optimality, yet, which are not sensitive to detailed inputs required in a cost effectiveness analysis.

Takamatsu et. al., proposed a model concerning the establishment of a realistic control strategy by computer for preventing air pollution caused by sulphur compounds in an urban area, Osaka, with extensive consideration on mathematical modelling, analysis and optimization. The notation of adaptive control is introduced to update model parameters which constantly vary owing to fluctuations and uncertainties in meteorological factors. On the basis of these strategies all emission sources are controlled so that future pollution level is always maintained below a specified criterion.

The Proposed Model

The models which have been developed so far with various assumptions as mentioned above have been proved to be successful when applied to local areas but none of these approaches can be used easily to optimize a large metropolitan area. The complexity and nonlinearity involved in the air pollution problems can be reduced by dividing the whole area into different regions. This is the approach followed. A dynamic model is developed considering two regions and one pollutant.

The pollutant as it comes out of the source [8] 1) may be dispersed in the space around the chimney 2) may be carried away by the wind 3) may be adsorbed on the particulates or decays 4) may be converted into other products by chemical reactions 5) may settle down in the neighboring area in case of particulates. It is either one or more of these possibilities depending upon the type of pollutant and the meteorological and geographical conditions. The total pollutant in a region can be due to what is initially present plus what is emitted out of the sources in that region plus what is being brought in by the incoming air. The model proposed here is a very simplified version of the true phenomenon. It is more an examplar rather than an exhaustive model. The two regions are supposed to be boxes with the known base area. The pollutant as it is emitted gets dispersed uniformly into the region and then it is carried away by the wind into other regions. This assumption is made because the regions considered are small. It is also assumed that no volume of air is lost as it travels from one region to the other. The dynamic model is developed considering the material

balance in the two regions. The material gain or loss is considered only because of it being carried in or out by the wind. The wind velocity is considered the same over the whole of region.

Material Balance Equation

Let,

$Q_n(t)$ = Amount of pollutant after time t, in minutes, in region n, expressed in tons.

$C_{n,n+1}$ = concentration of air travelling from region n to n+1, expressed in tons per $mile^3$.

$V_{n,n+1}$ = Velocity of air travelling from region n to n+1, expressed in miles per minute.

$A_{n,n+1}$ = Area of the cross section of the region perpendicular to the wind direction, between regions n and n+1, in $mile^2$.

q_n = Amount of pollutant emitted from sources in region n, expressed in tons per minute.

v_n = Total volume of region n, expressed in $mile^3$.

Then the material balance equation can be represented by

$$\begin{pmatrix}\text{The rate of change}\\ \text{of pollutant}\end{pmatrix} = \begin{pmatrix}\text{Pollutant emitted}\\ \text{from sources}\end{pmatrix}$$

$$+ \begin{pmatrix}\text{Pollutant transferred}\\ \text{into the region}\end{pmatrix}$$

$$- \begin{pmatrix}\text{Pollutant transferred out}\\ \text{of the region}\end{pmatrix}$$

The rate of amount gained in a region consists of the rate the pollutant

emitted in the region plus the rate at which it is being transported over from the other region. Now it is assumed that no volume is lost as it is being carried over from one region to the other. Which means that the wind is travelling toward the other region. A schematic diagram of the process with regions and wind direction is shown in Fig. 1. The regions are represented by the rectangular blocks with the wind direction shown by the arrow. It is assumed that no pollutant is initially present in the regions. The rate of the amount of pollutant lost in a region is due to it's being carried away by wind and due to transformations in chemical reactions. The amount lost due to the second cause is different for different pollutants and is small compared to the first one. It isn't, therefore, considered for the sake of simplicity. For an increment of time Δt the above material balance equation can be represented mathematically as

$$Q_n(t + \Delta t) = Q_n(t) = q_n \Delta t + V_{n-1,n} A_{n-1,n} C_{n-1,n} \cdot \Delta t - V_{n,n+1} A_{n,n+1} C_{n,n+1} \cdot \Delta t$$

or,

$$\frac{Q_n(t + \Delta t) - Q_n(t)}{\Delta t} = q_n + V_{n-1,n} A_{n-1,n} C_{n-1,n} - V_{n,n+1} A_{n,n+1} C_{n,n+1}$$

and as $\Delta t \to 0$,

$$\frac{dQ_n}{dt} = q_n + V_{n-1,n} A_{n-1,n} C_{n-1,n} - V_{n,n+1} A_{n,n+1} C_{n,n+1}$$

$$= q_n + V_{n-1,n} A_{n-1,n} C_{n-1,n} - \frac{Q_n V_{n,n+1} A_{n,n+1}}{V_1}$$

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.

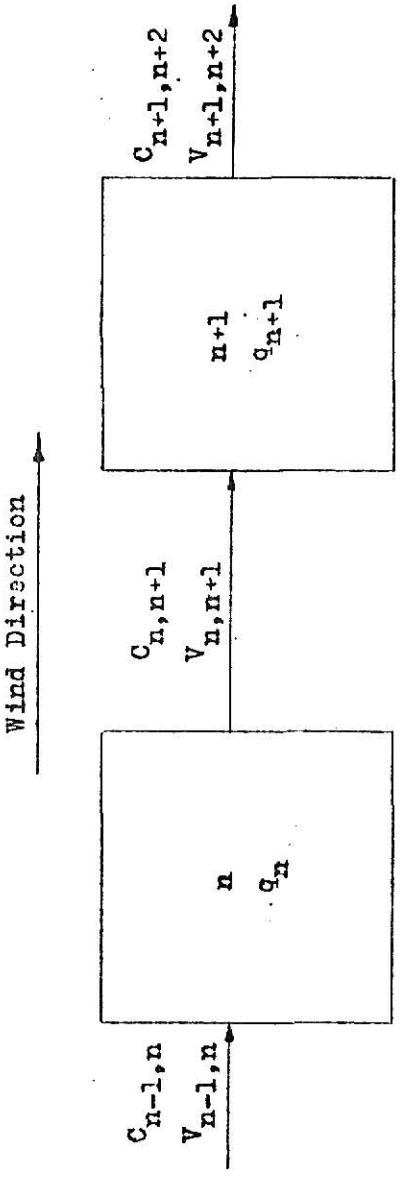

FIGURE 1. Schematic representation of the regions.

which is a differential equation in unknown Q_n.

Therefore for region 1

$$\frac{dQ_1(t)}{dt} = q_1 + C_{0,1}\, A_{0,1}\, V_{0,1} - \frac{Q_1(t)\, A_{1,2}\, V_{1,2}}{v_1} \tag{1}$$

and for region 2

$$\frac{dQ_2(t)}{dt} = q_2 + C_{1,2}\, A_{1,2}\, V_{1,2} - \frac{Q_2(t)\, A_{2,3}\, V_{2,3}}{v_2}$$

$$= q_2 + \frac{Q_1(t)\, A_{1,2}\, V_{1,2}}{v_1} - \frac{Q_2(t)\, A_{2,3}\, V_{2,3}}{v_2} \tag{2}$$

Solution of these differential equations for $Q_1(t)$ and $Q_2(t)$ will give the amount of pollutant present in regions 1 and 2 at any time t. For simplicity it is assumed that the volumes of all the regions are the same, i.e. $v_1 = v_2 = v_3 = v$ and all the areas perpendicular to the wind direction are the same i.e. $A_{0,1} = A_{1,2} = A_{2,3} = A$. Thus for the solution of these equations the values of variables $C_{0,1}$, q_1, q_2, v, A and various velocities have to be known. The values of $C_{0,1}$ can be observed. q_1 and q_2 can be determined knowning the sources in the region. v can not be obtained directly because of the change in the height of the inversion layer. A, too, cannot be determined directly because of the change in wind direction and the height of inversion layer. The velocity of wind changes with time. However for the period of time under consideration a constant value of velocity can be assumed and determined considering it as one of the parameters. The problem, therefore, reduces now to the one of parametric estimation.

The Estimation Problem

From Eqns. 1 and 2, it is seen that variables V's, A and v can be considered together. The, expression, $\gamma = \frac{VA}{v}$, represents the fraction of the pollutant containing wind being travelling from one region to the other per unit time. Thus putting this in Eqns. 1 and 2 we get

$$\frac{dQ_1(t)}{dt} = q_1 + C_{0,1}\,\gamma_1 - Q_1(t)\,\gamma_2 \tag{3}$$

$$\frac{dQ_2(t)}{dt} = q_2 + Q_1(t) \cdot \gamma_2 - Q_2(t) \cdot \gamma_3 \tag{4}$$

Thus the problem now reduces to the estimation of the parameters γ_1, γ_2 and γ_3. The estimation problem can therefore be stated as: estimate γ_1, γ_2 and γ_3 given the values of $C_{0,1}$, q_1 and q_2 and the following observed data,

$$Q_1^{obs}(t_i) = \alpha_i \qquad i = 1,2, \ldots, m_1 \tag{5}$$

$$Q_2^{obs}(t_k) = \beta_k \qquad k = 1,2, \ldots, m_2 \tag{6}$$

where $m_1 + m_2 \geq 5$, $0 \leq t_i \leq t_f$ and $0 \leq t_k \leq t_f$; given the initial amounts of pollutant at t = 0,

$$Q_1(0) = \alpha_0 \text{ and } Q_2(0) = \beta_0,$$

The quantities α_i and β_k are known. In this case they are obtained by the numerical solution of the equations 3 and 4 by assuming arbitrary values of the parameters. It must be noted that the equations 1 and 2 are simplified representation of the actual situation and these equations

can be solved in the closed form. The use of a more realistic model would make the differential equations more complicated and unsolvable analytically. The estimation of the parameters can be very difficult if the closed form solution of the process could not be obtained. The analytic solution of the equations 3 and 4, obtained by reduced equation and variation of constant [17] is

$$Q_1 = \frac{1}{\gamma_2} (q_1 + Q_0\gamma_1 - B\, e^{-\gamma_2 t})$$

and

$$Q_2 = C \cdot e^{-t\gamma_3} + \frac{(q_2 + q_1 + Q_0\gamma_1)}{\gamma_3} - \frac{B\, e^{-t\gamma_2}}{(\gamma_3 - \gamma_2)}$$

where B and C are integration constants.

The parameters appear nonlinearly in the analytic solution and their estimation by the nonlinear regression or nonlinear least squares estimation techniques would be quite involved and not very exact too. Quasilinearization, a recently developed and more powerful technique, has been used for parametric estimation.

Computational Procedure

The quasilinearization, first developed by Bellman and Kalaba [18-20], has been extensively used by Lee [16, 21-24] in solving two and multipoint boundary value problems. It has also been used in industrial management system [25,26], water resources management research [27] and others. Here the computational aspects of quasilinearization techniuqe are discussed with respect to the parametric estimation of the air pollution model. Readers are directed to references [16,18,19,21,22] for further details of theory.

The quasilinearization technique, also known as generalized Newton-Raphson method, is very similar to the Newton-Raphson root finding method; however the unknowns to be determined are functions in this case and not fixed roots as in Newton Raphson method. The first step in this technique is the linearization of the nonlinear differential equations by considering the first two terms in the Taylor's expansion of the function. Thus for a nonlinear differential equation of the form

$$\frac{d\underline{x}}{dt} = f(x,t) \tag{6}$$

where $\underline{x}$ and $\underline{f}$ represent the M - dimensional vectors with components $x_1, x_2, \ldots, x_M$ and $f_1, f_2, \ldots, f_M$ respectively, can be linearized to the form,

$$\frac{d\underline{x}_{n+1}}{dt} = \underline{f}(\underline{x}_n,t) + J(\underline{x}_n)(\underline{x}_{n+1} - \underline{x}_n) \tag{7}$$

where $\underline{x}_{n+1}$ and $\underline{x}_n$ represent M - dimensional vectors with components $x_{1,n+1}, x_{2,n+1}, \ldots, x_{M,n+1}$ and $x_{1,n}, x_{2,n} \ldots, x_{M,n}$ respectively. The Jacobian matrix $J(\underline{x}_n)$ is

$$J(\underline{x}_n) = \begin{pmatrix} \frac{\partial f_1}{\partial x_{1,n}} & \frac{\partial f_1}{\partial x_{2,n}} & \cdots & \frac{\partial f_1}{\partial x_{M,n}} \\ \frac{\partial f_2}{\partial x_{2,n}} & \frac{\partial f_2}{\partial x_{2,n}} & \cdots & \frac{\partial f_2}{\partial x_{M,n}} \\ \vdots & & & \\ \frac{\partial f_M}{\partial x_{1,n}} & \frac{\partial f_M}{\partial x_{2,n}} & \cdots & \frac{\partial f_M}{\partial x_{M,n}} \end{pmatrix} \tag{8}$$

which results in M sets of linear equations. In the system of equations represented by 3 and 4 we assume Q_1 and Q_2 also among the parameters to be estimated. It would be convenient and better to choose parameters γ_1, γ_2 and γ_3, to be in parallel with Q_1 and Q_2, as functions of time. But as these do not change with time we have,

$$\frac{d\gamma_1}{dt} = 0 \tag{9}$$

$$\frac{d\gamma_2}{dt} = 0 \tag{10}$$

$$\frac{d\gamma_3}{dt} = 0 \tag{11}$$

Using Eq. 7 the system of Eqns. 3,4 and 9-11 can then be linearized. Obtaining,

$$\frac{dQ_1^{n+1}}{dt} = -\gamma_2^n Q_1^{n+1} + Q_0 \gamma_1^{n+1} - Q_1^n \gamma_2^{n+1} + q_1 + Q_1^n \gamma_2^n \tag{12}$$

$$\frac{dQ_2^{n+1}}{dt} = \gamma_2^n Q_1^{n+1} - \gamma_3^n Q_2^{n+1} + Q_1^n \gamma_2^{n+1} - Q_1^n \gamma_3^{n+1} + q_2$$

$$- Q_1^n \gamma_2^n + Q_1^n \gamma_3^n \tag{13}$$

$$\frac{d\gamma_1^{n+1}}{dt} = 0 \tag{14}$$

$$\frac{d\gamma_2^{n+1}}{dt} = 0 \tag{15}$$

$$\frac{d\gamma_3^{n+1}}{dt} = 0 \tag{16}$$

Thus we have M = 5 sets of linear equations and five initial conditions are required for their solution. Generally the initial condition available are less than that, being available only for the state variables. In the present case the two known boundary conditions are,

$$Q_{1,n+1}(0) = \alpha_0 \tag{17}$$

$$Q_{2,n+1}(0) = \beta_0 \tag{18}$$

So the rest of the initial conditions are obtained from the least square criterion. In this approach the objective is to minimize the sum of squares of the deviations from the observed data. Thus the expression, for the state variables $Q_1(t)$ and $Q_2(t)$,

$$LS = \sum_{i=1}^{m_1} \left(Q_1(t_i) - \alpha_i\right)^2 + \sum_{j=1}^{m_2} \left(Q_2(t_j) - \beta_j\right)^2 \tag{19}$$

is minimized over the parameters γ_1, γ_2 and γ_3.

As the system of Eqs. 12 thru 16 is linear the principle of superposition [16] can be used for their solution. Their general solutions are given,

$$Q_{1,n+1}(t) = Q_{1p,n+1}(t) + \sum_{j=1}^{q} a_{j,n+1}\, Q_{1hj,n+1}(t)$$

$$Q_{2,n+1}(t) = Q_{2p,n+1}(t) + \sum_{j=1}^{q} a_{j,n+1}\, Q_{2hj,n+1}(t)$$

$$\gamma_{1,n+1}(t) = \gamma_{1p,n+1}(t) + \sum_{j=1}^{q} a_{j,n+1}\, \gamma_{1hj,n+1}(t)$$

$$\gamma_{2,n+1}(t) = \gamma_{2p,n+1}(t) + \sum_{j=1}^{a} a_{j,n+1}\, \gamma_{2hj,n+1}(t)$$

$$\gamma_{3,n+1}(t) = \gamma_{3p,n+1}(t) + \sum_{j=1}^{q} a_{j,n+1}\, \gamma_{3hj,n+1}(t) \qquad (20)$$

with q = 5. The subscript p denotes particular solution, and the subscript h_j denotes the j^{th} set of homogeneous solutions. The a's are the integration constants to be determined by the boundary conditions. As the initial conditions are chosen such that they satisfy the given initial conditions given by Eqs. 17 and 18 therefore only three sets of homogeneous solutions are needed with three integration constants, $a_{j,n+1}$, j = 1,2,3. Hence the value of q = 3.

The particular solution is obtained by integrating Eqs. 12 thru 16 numerically using Runge-Kutta method with the following initial values.

$$Q_{1p,n+1}(0) = \alpha_0$$

$$Q_{2p,n+1}(0) = \beta_0$$

$$\gamma_{1p,n+1}(0) = 0$$

$$\gamma_{2p,n+1}(0) = 0$$

$$\gamma_{3,n+1}(0) = 0 \qquad (21)$$

The homogeneous forms of Eqs. 12 thru 16 are,

$$\frac{dQ_{1,n+1}}{dt} = - \gamma_2^n Q_1^{n+1} + Q_0 \gamma_1^{n+1} - Q_1^n \gamma_2^{n+1}$$

$$\frac{dQ_{2,n+1}}{dt} = \gamma_2^n Q_1^{n+1} - \gamma_3^n Q_2^{n+1} + Q_1^n \gamma_2^{n+1} - Q_1^n \gamma_3^{n+1}$$

$$\frac{d\gamma_1^{n+1}}{dt} = 0$$

$$\frac{d\gamma_2^{n+1}}{dt} = 0$$

$$\frac{d\gamma_3^{n+1}}{dt} = 0 \qquad (22)$$

The initial conditions used for homogeneous solutions are listed in Table 1. Note that the homogeneous conditions are chosen such that they satisfy the given initial conditions given by Eqs. 17 and 18. The remaining three boundary conditions are obtained by minimizing Eq. 19. Inserting the first two equations of the equation set 20 in Eq. 19 with q = 3, we get,

$$LS = \sum_{i=1}^{m_1} \left(Q_{1p,n+1}(t_i) + \sum_{j=1}^{3} a_{j,n+1} Q_{1hj,n+1}(t_i) - \alpha_i \right)^2$$

$$+ \sum_{k=1}^{m_2} \left(Q_{2p,n+1}(t_k) + \sum_{j=1}^{3} a_{j,n+1} Q_{2hj,n+1}(t_k) - \beta_k \right)^2 \qquad (23)$$

As the particular and homogeneous solutions at various values of t are known and are obtained numerically knowing the initial conditions listed in Eq. 21 and Table 1 the only unknown variables in the above equations are $a_{j,n+1}$, j = 1,2,3. So the problem reduces to finding the values of these integration constants such that LS is minimized. This can be done by partly differentiating the above equation with respect to $a_{j,n+1}$, j = 1,2,3 in turn and equating to zero. Thus we get the following set of equations,

$$\frac{\partial LS}{\partial a_{m,n+1}} = 2 \sum_{i=1}^{m_1} Q_{1hm,n+1}(t_i) \left[Q_{1p,n+1}(t_i) + \sum_{j=1}^{3} a_{j,n+1} Q_{1hj,n+1}(t_i) - \alpha_i \right]$$

$$+ 2 \sum_{k=1}^{m_2} Q_{2hm,n+1}(t_k) \left[Q_{2p,n+1}(t_k) + \sum_{j=1}^{3} a_{j,n+1} Q_{2hj,n+1}(t_k) - \beta_k \right]$$

$$= 0, \qquad m = 1,2,3 \tag{24}$$

These form new and remaining three boundary conditions. As variables $Q_1(t_i)$ and $Q_2(t_k)$ are known and also α_i and β_k these equations can be solved for $a_{j,n+1}$, j = 1,2,3. Once these integration constants are known the general solution can be obtained.

The computation procedure can be summarized as (27),

1). Assume a reasonable set of initial functions for $Q_1(t)$, $Q_2(t)$, $\gamma_1(t)$, $\gamma_2(t)$ and $\gamma_3(t)$. Let these initial functions be

$Q_{1,n=0}(t) = 0$, $Q_{2,n=0}(t) = 0$, $\gamma_{1,n=0}(t) = 0$, $\gamma_{2,n=0}(t) = 0$, $\gamma_{3,n=0}(t) = 0$

2). Integrate Eqs. 12 thru 16 numerically using Eq. 21 as initial conditions.

3). Integrate set of equations 22 for homogeneous solution using Table 1 as initial conditions, starting with n = 0.

4). Solve set of equations 24 for the integration constants. Knowing set of integration constants and particular and homogeneous solutions from step 2 and 3 the general solution can be obtained.

5). Calculate $Q_{1,n+1}(t)$, $Q_{2,n+1}(t)$, $\gamma_{1,n+1}(t)$, $\gamma_{2,n+1}(t)$, $\gamma_{3,n+1}(t)$ using Eq. 20 with q = 3.

6). Repeat steps 2 thru 5 until there is no further improvements on the values of $Q_1(t)$, $Q_2(t)$, $\gamma_1(t)$, $\gamma_2(t)$ and $\gamma_3(t)$.

Results

The experiment data were obtained numerically using Runge-Kutta method from equations 1 and 2 using the following values of the parameters

q_1 = .00416 tons/min, q_2 = .00414 tons/min.

A = .03785 mile^2, Q_0 = 0.0, v = .762 mile^3

V_{01} = .1267 miles/min, V_{12} = .1267 miles/min.

V_{23} = .0134 miles/min.

The values of $Q_1(t)$ and $Q_2(t)$ obtained are given in Table 1.

TABLE 1

Experimental Data

t	$Q_1(t)$	$Q_2(t)$
0	0.9240	0.3380
5	0.9198	0.3623
10	0.9157	0.3865
15	0.9116	0.4107
20	0.9076	0.4347
25	0.9037	0.4586
30	0.8999	0.4824
35	0.8961	0.5061
40	0.8923	0.5298
45	0.8887	0.5533
50	0.8850	0.5767
55	0.8815	0.6000
60	0.8780	0.6233
65	0.8745	0.6464
70	0.8711	0.6695
75	0.8678	0.6924
80	0.8645	0.7153
85	0.8612	0.7381
90	0.8580	0.7608
95	0.8549	0.7834
100	0.8518	0.8059
105	0.8488	0.8283
110	0.8458	0.8507
115	0.8428	0.8729
120	0.8399	0.8951

Estimation of γ_1, γ_2 and γ_3

The system of equations is given by 3,4 and 9-11 and the linearized equations by Eqn. set 12-16. The initial conditions used for particular solutions are,

$$Q_{1p,n+1}(0) = .924, \quad Q_{2p,n+1}(0) = .338 \tag{25}$$

$$\gamma_{1p,n+1}(0) = 0, \quad \gamma_{2p,n+1}(0) = 0 \quad \text{and} \quad \gamma_{3p,n+1}(0) = 0$$

The initial condition for homogeneous solutions are given in Table 1. The initial functions used for Q_1 and Q_2 and the parameters γ_1, γ_2 and γ_3 are their true values. Thus

$$Q_{1,n=0}(t) = Q^{obs}_{1,n=0}(t) \text{ and } Q_{2,n=0}(t) = Q^{obs}_{2}(t) \tag{26}$$

$$\gamma_{1,n=0}(t) = .00625$$

$$\gamma_{2,n=0}(t) = .00625$$

and

$$\gamma_{3,n=0}(t) = .000688 \tag{27}$$

It is found that the system does not converge. The values of the parameters for ten iterations are shown in Table 3.

Estimation of γ_2 and γ_3.

The system of equations for this are represented by Eqns. 3,4,9 and 10. The linearized set of equations is

TABLE 2

Initial Conditions Used for Homogeneous Solution

Variable	j		
	1	2	3
$Q_{1hj,n+1}(0)$	0	0	0
$Q_{2hj,n+1}(0)$	0	0	0
$\gamma_{1hj,n+1}(0)$	1	0	0
$\gamma_{2hj,n+1}(0)$	0	1	0
$\gamma_{3hj,n+1}(0)$	0	0	1

TABLE 3

Variations in the Values of γ_1, γ_2 and γ_3 During Iterations

Iteration	γ_1	γ_2	γ_3
0	.00625	.00625	.000688
1	.00306	-.00597	0.0
2	.00110	.00113	0.0
3	.00125	-.01459	0.0
4	.01936	.02426	0.0
5	-.01816	-.02592	0.0
6	.03604	.03972	0.0
7	-.03659	-.04720	0.0
8	.06831	.07027	0.0
9	-.08214	-.09307	0.0
10	.13859	.13775	0.0

$$\frac{dQ_1^{n+1}}{dt} - \gamma_2^n Q_1^{n+1} - Q_1^n \gamma_2^{n+1} + Q_1^n \gamma_2^n + q_1 + Q_0 \gamma_1$$

$$\frac{dQ_2^{n+1}}{dt} = \gamma_2^n Q_1^{n+1} - \gamma_3^n Q_2^{n+1} + Q_1^n \gamma_2^{n+1} - Q_2^n \gamma_3^{n+1}$$

$$+ q_2 - Q_1^n \gamma_2^n + Q_2^n \gamma_3^n.$$

$$\frac{d\gamma_2^{n+1}}{dt} = 0.0$$

$$\frac{d\gamma_3^{n+1}}{dt} = 0.0$$

The initial conditions for particular solution are given by Eqn. 25 and

$$\gamma_{2p,n+1}(0) = 0, \qquad \gamma_{3p,n+1}(0) = 0.$$

The initial conditions for homogeneous solution are the same as, represented in Table 1, excluding for parameter γ_1. The initial functions for this problem are given by Eqn. 26 and the two sets for γ_1 and γ_2 are

$\gamma_2 = .00625$ $\qquad$ $\gamma_2 = .0125$

and . The estimated values of γ_2 and γ_3

$\gamma_3 = .000688$ $\qquad$ $\gamma_3 = .0030$

obtained in both cases are

$$\gamma_2 = .00632$$

$$\gamma_3 = .00042$$

The convergence of these parameters is obtained in the first iteration. It is seen that the estimated values of γ_2 and γ_3 are different from their actual values.

Estimation of γ_2

The system of equation for this is given by Eqns. 3,4 and 9. Linearization of these gives the following set.

$$\frac{dQ_2^{n+1}}{dt} = - \gamma_2^n Q_1^{n+1} - Q_1^n \gamma_2^{n+1} + q_1 + Q_0 \gamma_1 + Q_1^n \gamma_2^n$$

$$\frac{dQ_2^{n+1}}{dt} = \gamma_2^n Q_1^{n+1} - \gamma_3 Q_2^n + Q_1^n \gamma_2^{n+1} + q_2 - Q_1^n \gamma_2^n .$$

$$\frac{d\gamma_2^{n+1}}{dt} = 0.$$

The initial conditions for particular and homogeneous solution are given by Eqn. 25, $\gamma_{2p,n+1}(0) = 0$, and Table 1, picking only the value of the parameters under consideration. The initial functions are given by Eqn. 26 and two sets for γ_2 as,

$$\gamma_2 = .00625 \text{ and } \gamma_2 = .0125$$

In both the cases the estimated value of γ_2 is

$$\gamma_2 = .00633$$

The convergence is obtained in the first iteration.

Estimation of V_{01}

The system of equation is given by 1,2 and $\frac{dV_{01}^{n+1}}{dt} = 0$. The linearized set of equations are given by,

$$\frac{dQ_1^{n+1}}{dt} = q_1 + \frac{AV_{12}}{v_1} Q_1^{n+1} + Q_0 \frac{AV_{01}^{n+1}}{v_1}$$

$$\frac{dQ_2^{n+1}}{dt} = q_2 + \frac{AV_{12}}{v_1} Q_1^{n+1} - \frac{AV_{23}}{v_1} Q_2^{n+1}$$

$$\frac{dV_{01}^{n+1}}{dt} = 0$$

The initial conditions used for particular solution are given by Eqn. 25 and $V_{01p,n+1}(0) = 0.0$ and that for homogeneous solution are,

$$Q_{1H,n+1}(0) = 0.0 \qquad Q_{2H,n+1}(0) = 0.0 \qquad (27)$$

$$V_{01H,n+1}(0) = 1.0$$

The computations were done for the two initial values of the parameter V_{01}. These values are,

$$V_{01,n=0}(t) = .1267 \quad \text{and} \quad V_{01,n=0}(t) = .25$$

The experimental values, given in Table 1, were used as the initial functions of Q_1 and Q_2 given by Eqn. 26. It is seen that the convergence is obtained in the first iteration. For both the initial functional value of V_{01}, the estimated value is the same. This estimated value

is however different from the actual value of V_{01}.

Estimation of A

The system of Equations for this are Eqns. 1 and 2 and

$$\frac{dA^{n+1}}{dt} = 0.$$

The linearized set of equations are given as,

$$\frac{dQ_1^{n+1}}{dt} = -\frac{A^n V_{12} Q_1^{n+1}}{v_1} + A^{n+1}\left(\frac{Q_0 V_{01}}{v_1} - \frac{V_{12} Q_1^n}{v_1}\right) + q_1 - \frac{V_{01} Q_0}{v_1} A^n$$

$$\frac{dQ_2^{n+1}}{dt} = \frac{V_{12} A^n Q_1^{n+1}}{v_1} - \frac{V_{23} A^n Q_2^{n+1}}{v_1} + A^{n+1}\left(\frac{V_{12} Q_1^n}{v_1} - \frac{V_{23} Q_2^n}{v_1}\right)$$

$$+ q_2 - A^n\left(\frac{V_{12} Q_1^n}{v_1} - \frac{V_{23} Q_1^n}{v_1}\right)$$

$$\frac{dA^{n+1}}{dt} = 0$$

The initial conditions for particular and homogeneous solutions for Q_1 and Q_2 are given by Eqns. 25 and 27. For the parameter A these are,

$$A_{p,n+1}(0) = 0.0$$

$$A_{H,n+1}(0) = 1.0$$

The actual value of A is used as it's initial function. The initial functions of Q_1 and Q_2 are their experimental values given, by Eqn. 26.

$$A_{n=0}(t) = .03785$$

The system converges in seven iterations. The variables Q_1 and Q_2 and the parameter A converge to values different from their actual values. The convergence rate of A is shown in Table 4. Figs. 2 and 3 represent the convergence rates for Q_1 and Q_2 respectively. The estimated value of A is different from the actual value.

Estimation of v

The system of equations is given by 1, 2 and

$$\frac{dv^{n+1}}{dt} = 0$$

The linearized set of equations are

$$\frac{dQ_1^{n+1}}{dt} = -\frac{AV_{12}Q_1^{n+1}}{v_1^n} - \frac{v_1^{n+1}}{v_1^{2,n}}\left(Q_0 AV_{01} - Q_1^n\, AV_{12}\right)$$

$$+ q_1 + \frac{AV_{12}Q_1^n}{v_1^n}$$

$$\frac{dQ_2^{n+1}}{dt} = \frac{AV_{01}Q_1^{n+1}}{v_1^n} - \frac{AV_{23}Q_2^{n+1}}{v_1^n} - \frac{v_1^{n+1}}{v_1^{2,n}}\left(AV_{01}Q_1^n - AV_{23}Q_2^n\right)$$

$$+ q_2 + \frac{1}{v_1^n}\left(AV_{01}Q_1^n - AV_{23}Q_2^n\right)$$

$$\frac{dv^{n+1}}{dt} = 0$$

The initial conditions for particular and homogeneous solution of Q_1 and

TABLE 4

Convergence Rate of A

Iteration	A
0	.03785
1	.0201
2	.0292
3	.0241
4	.0268
5	.0253
6	.0261
7	.0251
8	.0259
9	.0258
10	.0259

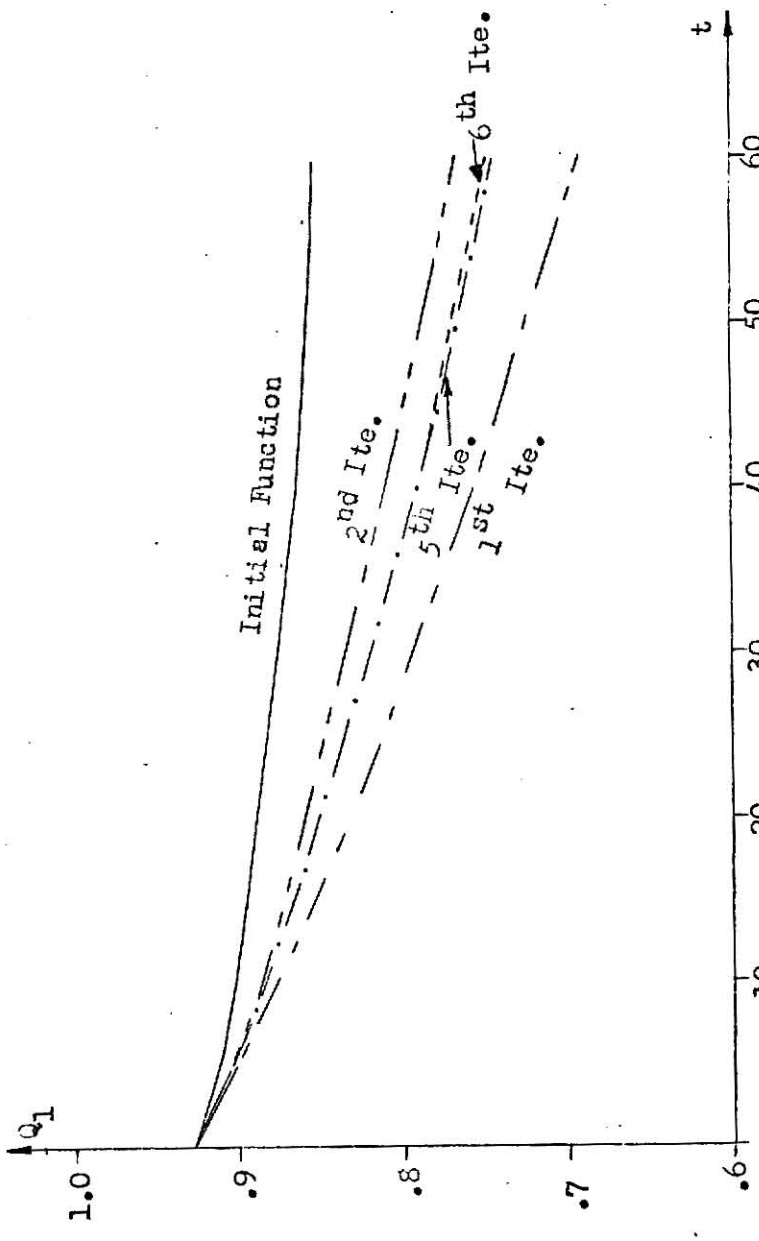

FIGURE 2. Convergence rate of Q_1.

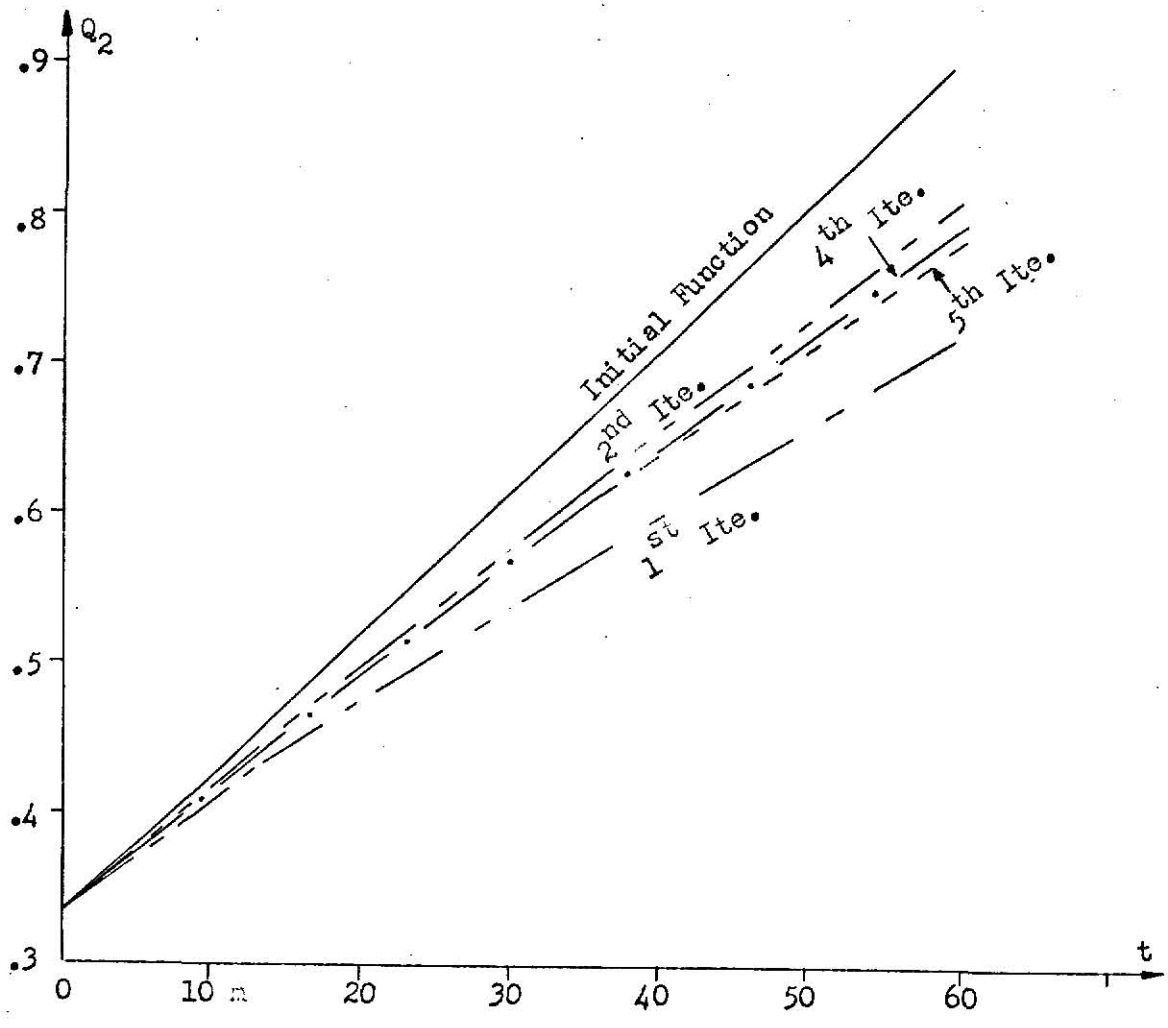

FIGURE 3. Convergence rate of Q_2.

Q_2, are given by Eqns. 25 and 27. For the variable v these ar ,

$$v_{p,n+1}(0) = 0$$

$$v_{H,n+1}(0) = 1.0$$

The actual value of A is used as it's initial function as before and also the observed values of Q_1 and Q_2 are used as the initial function for Q_1 and Q_2 given by Eqn. 25.

$$v_{n=0} = .762.$$

The convergence rates for Q_1 and Q_2 are shown in Figs. 4 and 5 and that for the parameter v in Table 5. The system converges in eight iterations. The parameter converges to a negative value.

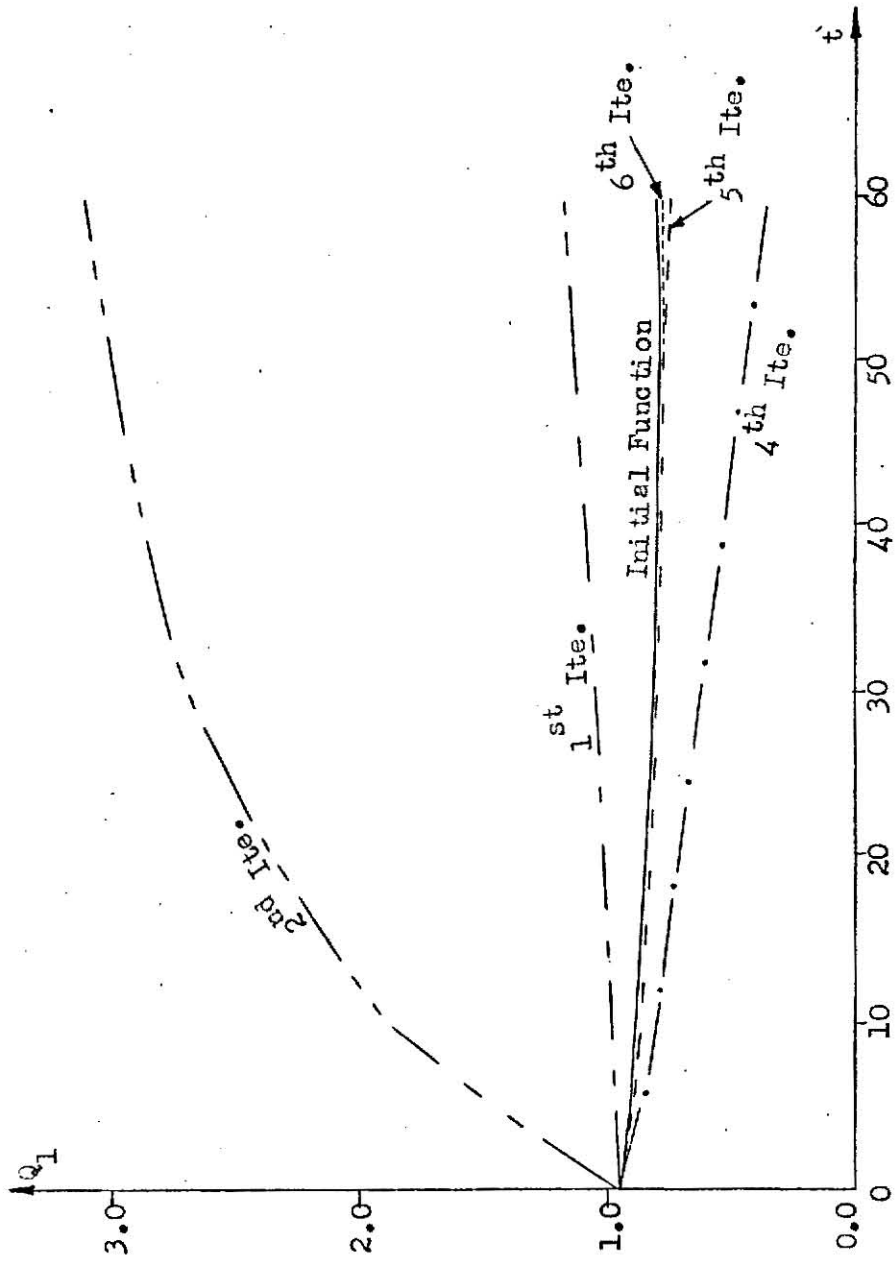

FIGURE 4. Convergence rate of Q_1.

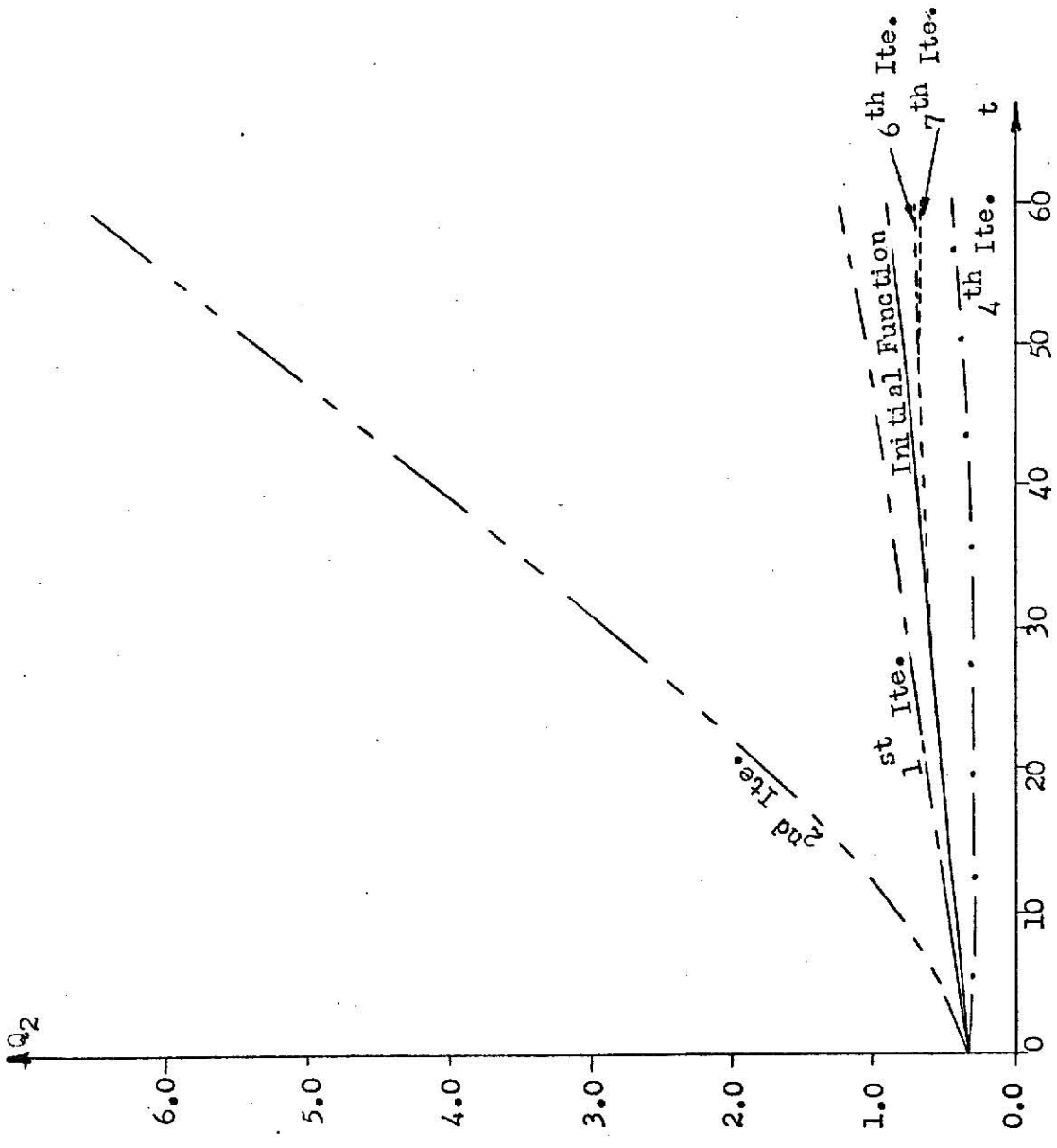

FIGURE 5. Convergence rate of Q_2.

TABLE 5

Convergence Rate of v

Iteration	v
1	.065
2	27.58
3	-.510
4	-2.72
5	-2.04
6	-1.83
7	-1.89
8	-1.89
9	-1.889
10	-1.890

CHAPTER 3

STATISTICAL MODELLING OF AIR POLLUTION IN SAN FRANCISCO BAY AREA

INTRODUCTON

In this chapter is presented the statistical modelling of the air pollution problem. Statistical methods appear to be very useful for the air pollution problem because they directly deal with the real air quality situation.

The air quality data implicitly contains all the information of the complex pollution phenomenon occuring in the air. This information is difficult to consider simultaneously in the physical model which tends to simulate the deterministic behaviour of the pollutants in the air. The analysis of the data from an area not only gives an information about the air pollution phenomenon at that location but also this information can be used for developing the prediction model for that location.

The availability of the well maintained data makes possible the application of the statistical approach to the air pollution problem. In this chapter is given the general procedure followed for the analysis of air pollution data and the development of the prediction model. This procedure is illustrated for the analysis of data obtained from the three stations Richmond, Oakland and San Jose in the San Francisco Bay area. [44] The total hydrocarbon concentration data were available from Richmond and Oakland, and the carbon monoxide concentration data were available from Oakland and San Jose. These data were available for two months duration namely, September and October, with one hour interval of observations.

ANALYSIS OF AIR POLLUTION DATA

Air pollution as a whole is a very complex phenomenon-complex in the sense that it is being contributed by diverse factors and further complicated because of the variations in these factors. These factors are;

1. source distribution
2. transporation, diffusion, or mixing of pollutants
3. chemical reaction among various pollutants
4. geographical conditions and
5. meterological conditions.

The variations in these factors are observed due to the fact that the sources may be fixed sources like industrial plants or may be mobile sources like automobiles, aeroplanes etc. The amount of pollutant transported from neighbouring subarban areas may vary because of the variation in wind speed and wind direction. Variations in the conditions of temperature, pressure, humidity may vary the pollutant concentration because of the reactions and decompositions among pollutants producing new pollutants. The geographical location of the place may or may not help in the pollutants accummulation at certain locations. Most of these factors have cyclic variations associated with them. The diverse pollutants emitted by various kinds of industrial plants in a suburban area may vary during the day due to the factory shift hours and may vary over the year because of the variation in production. The pollutants emitted by the automobiles vary during the day because of some regular social activities e.g. office or factory timings. Then there are seasonal variations due to meteorological conditions. For example, the variations in the wind direction

may cause the transportation of pollutants to some particular areas at one time and to some other areas at other times. This may result in the cyclic accummulation and ventillation of pollutants. The accummulation may further be supported by the geographical conditions. The seasonal variations in temperature cause the seasonal variations in pollutants formed from chemical reactions. Then there may be seasonal variations of the inversion layer. All this information is difficult to be considered at the same time in the physical models trying to simulate the actual phenomenon occuring in the air. On the other hand this information is implictly contained in the air pollution data and can be revealed by its analysis. The analysis of the data consists in recognizing these cyclic variations and associating them with the actual phenomenon and thereby developing the predictive model.

Description of the Data

The hydrocarbon data from Richmond and Oakland are shown in Figs. 1 and 2. It is seen that though the magnitude of the pollutants emitted in the two regions are different being higher at Richmond than at Oakland yet the overall variation is nearly the same. At both of these stations, the average amounts of hydrocarbons emitted at the end of September and at the beginning of October are higher than those emitted during the rest of the periods. The hydrocarbon accummulation seems to be building again near the end of October at both the stations. A monthly cycle seems to be indicated. There are also peaks within a month indicating high frequencies.

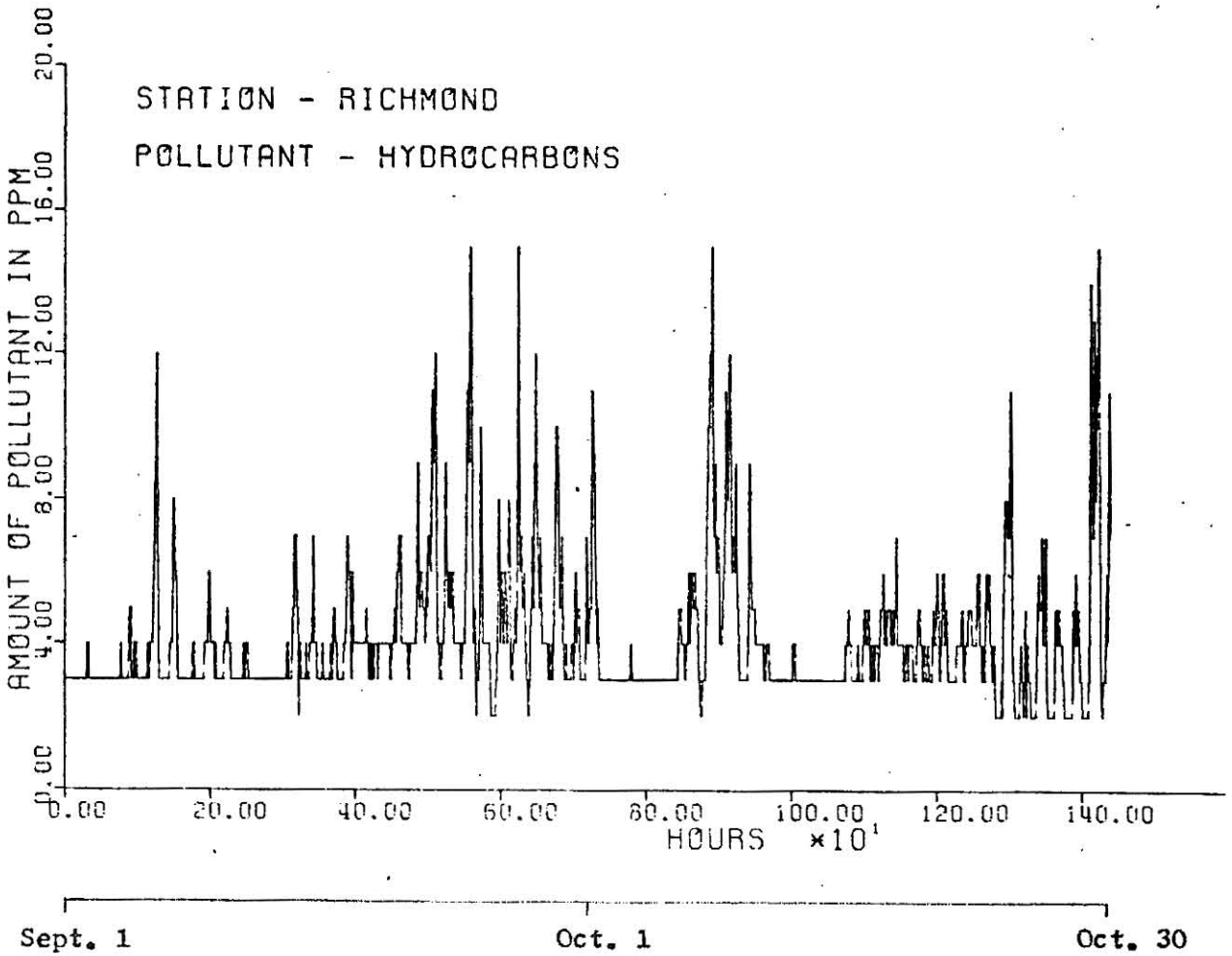

FIGURE 1. Hourly hydrocarbons data from Richmond.

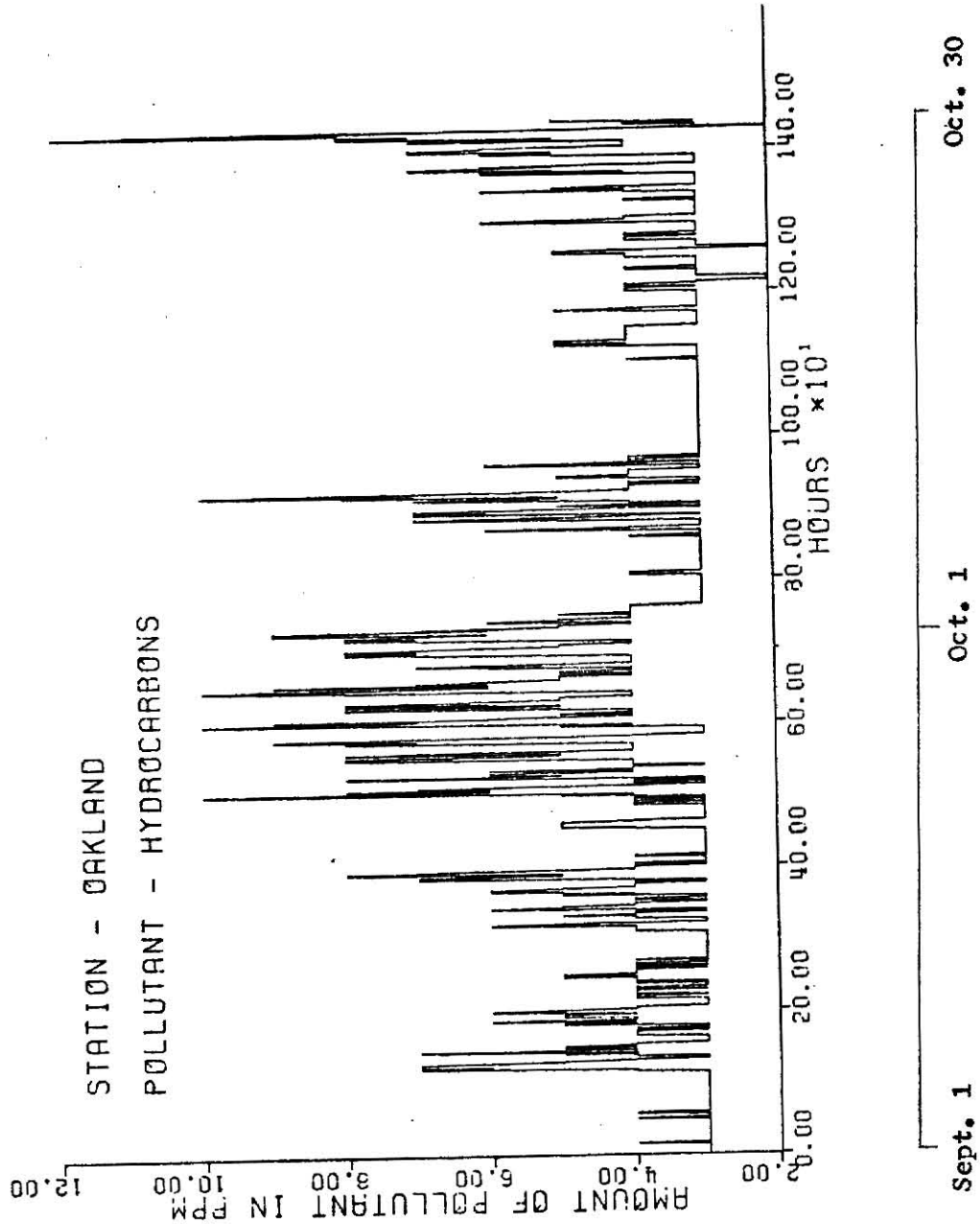

FIGURE 2. Hourly hydrocarbons data from Oakland.

The carbon monoxide data from Oakland and San Jose is shown in Figs. 3 and 4. The level of carbon monoxide emitted at both the stations is the same. The average amount of pollutant emitted at the end of the month is higher than that emitted during the rest of the months at both the stations. There appears to be cyclic variations for periods less than a month of carbon monoxide at both the stations.

Fourier Analysis

The amount of pollutant emitted at any time is a stochastic variable. The process of emission of pollutants from the sources can therefore be considered as a stochastic process. The problem of developing the statistical model from the air pollution data, therefore, can also be termed as the stochastic modelling of air pollution.

The collection of pollutant concentration observations sequentially through time forms a time series. The dependence of the series on various periodicities can first be tested using Fourier analysis. Fourier analysis, though it has its limitations because of the fact that it is based on the fixed amplitudes, frequencies and phases and does not take into consideration the random changes of amplitudes, frequencies and phases of a time series yet, it is useful enough to give the relative importance of its harmonics. It distributes the total variance of the series among its harmonics, a greater variance is associated with predominate harmonics and vice versa. This information is very useful for the further analysis of the data especially if there is no knowledge of the various phenomena contributing to the air pollution in the area under consideration and hence no knowledge of the expected cyclic variations.

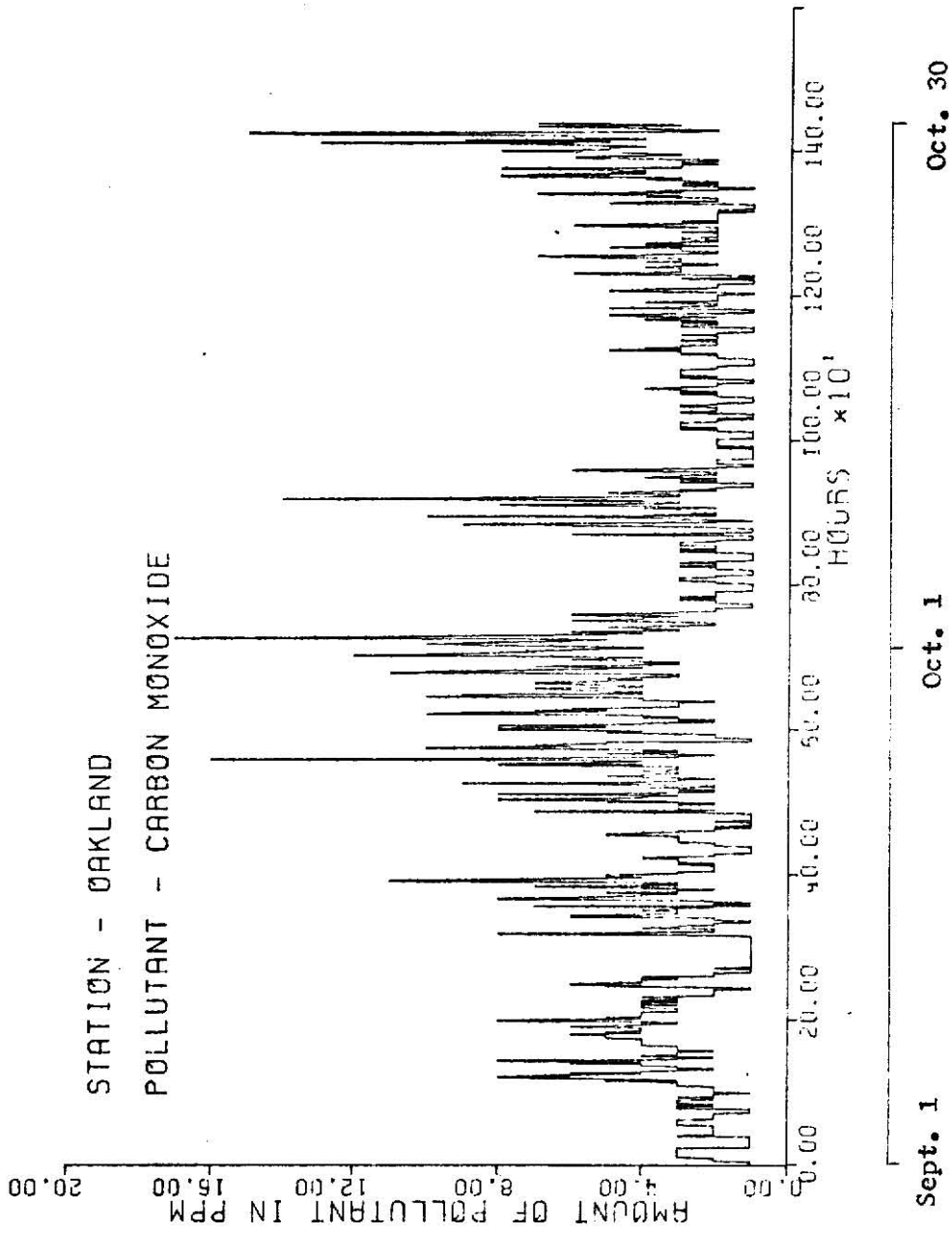

FIGURE 3. Hourly carbon monoxide data from Oakland.

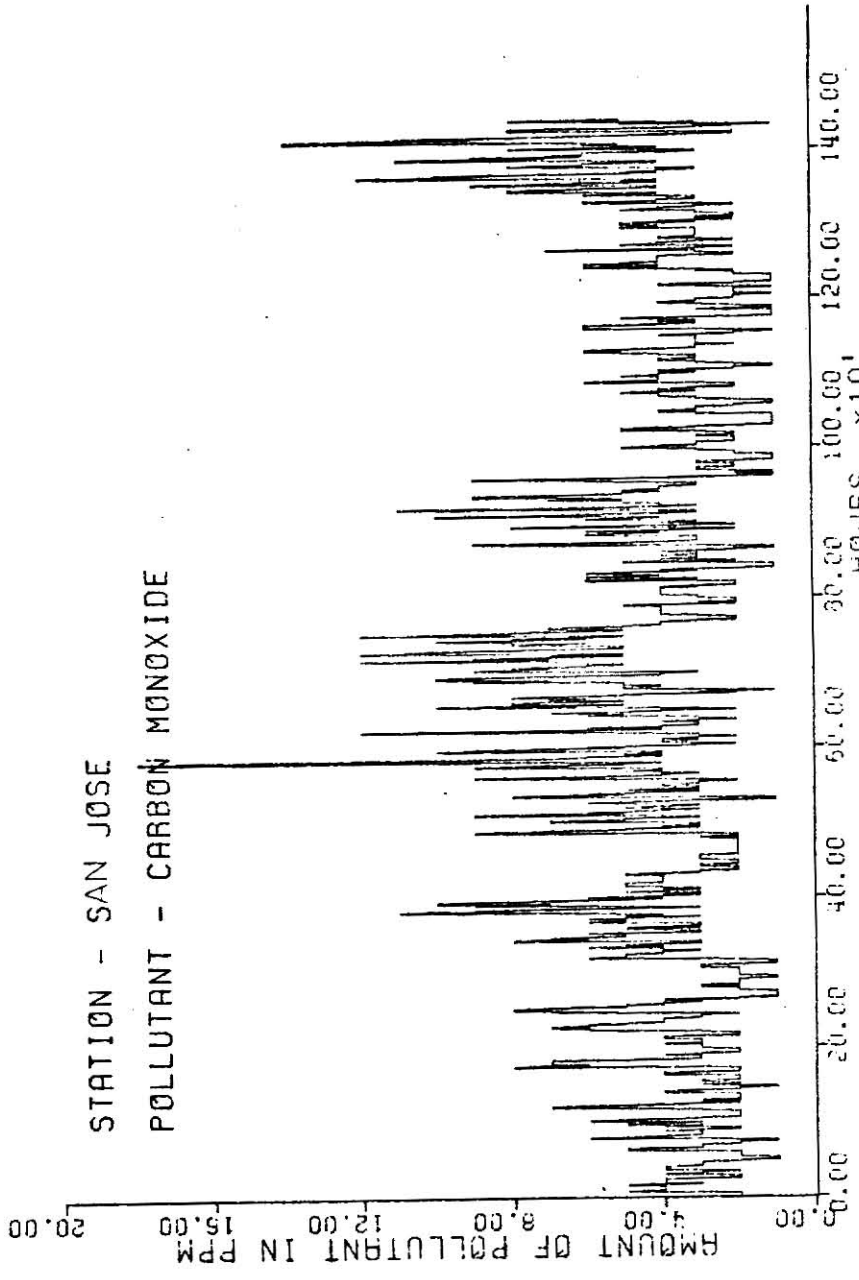

FIGURE 4. Hourly carbon monoxide data from San Jose.

This analysis gives the feel of the data to be analyzed further.

The Fourier representation of the time series is [33]:

$$x(t) = \bar{x} + 2 \sum_{k=1}^{n-1} A_k \cos(2\pi fkt - \phi_k) + A_n \cos 2\pi fnt \qquad (1)$$

where $\bar{x}$ is the average amount of pollutant emitted, A_k is the amplitude of the k^{th} harmonic, f is the fundamental frequency given by $f = 1/N\Delta$ where N is the length of the record being equal to 1440, and Δ is the interval of observation equal to one hour, n is the number of harmonics in the time series equal to N/2 and ϕ_k is the phase of the k^{th} harmonic. The fundamental frequency corresponds to a period of two months being equal to the length of the record. The function x(t) is thus composed of a sum of sine and cosine functions whose frequencies are multiples or harmonics of the fundamental frequency f. The highest frequency which can be detected is equal to $n/N = 1/2$ cycles per hour which corresponds to two sampling intervals. The Fourier coefficients of the above Fourier series for any harmonic k are,

$$\begin{aligned} C_{1k} &= A_k \cos(\phi_k) \\ C_{2k} &= A_k \sin(\phi_k) \end{aligned} \qquad (2)$$

which can further be expressed as [33]

$$\begin{aligned} C_{1k} &= \frac{1}{N} \sum_{t=-n}^{n-1} x_t \cos \frac{2\pi kt}{N} \\ C_{2k} &= \frac{1}{N} \sum_{t=-n}^{n-1} x_t \sin \frac{2\pi kt}{N} \end{aligned} \qquad (3)$$

for k = 0,1, ..., n

These Fourier coefficients can be determined from the above expression as all the quantities on the right hand side of the expression are known. x_t denotes the hourly observations of the pollutant concentration and t is the time index in hours. From Eqn. 2, the amplitude and phase at any harmonic k are given as,

$$A_k = \sqrt{C_{1k}^2 + C_{2k}^2} \tag{4}$$

$$\phi_k = \arctan\left(\frac{C_{2k}}{C_{1k}}\right) \tag{5}$$

The data have now been represented by the sine and cosine terms with their known amplitudes. But to determine their relative importance it is required to compare the variance associated with each of these harmonics. It has been shown elsewhere that the mean square value of x_t or average power dissipated by x_t can be decomposed into contributions arising from each harmonic. This, in other words, means that the contribution of average square value of x_t to one harmonic is independent of it's contribution to their harmonics. Thus considering the mean square value about the mean, $\bar{x}$, the variance can be expressed as,

$$\sigma^2 = \frac{1}{N} \sum_{t=-n}^{n-1} (x_t - \bar{x})^2$$

$$= 2 \sum_{k=1}^{n-1} A_k^2 + A_n^2 \tag{6}$$

thus the contribution to the total variance due to the harmonics is,

$$\sigma_k^2 = \begin{cases} 2A_k^2, & \text{if } k<n \\ A_k^2, & \text{if } k=n \end{cases} \qquad k = 1, \ldots, n$$

and at k=0, $\sigma_k^2 = \bar{x}^2$ (7)

The relative importance of the contribution to each harmonic is obtained by the ratio

$$\gamma_k = \frac{\sigma_k}{\sigma} \qquad k = 1, \ldots, n \qquad (8)$$

By using the above equations, harmonic analysis is carried out for the data shown in Figures 1-4. The contribution to the total variance of the hydrocarbon data from Richmond by it's mean is 82.55%. This for the hydrocarbon data from Oakland is 88.79%, for carbon monoxide data from Oakland is 69.88% and for the carbon monoxide data from San Jose is 76.66%. As this is in general very high, the values of r_k, k = 1, ...,n, would be very small. Hence to make r_k a significant figure the variance corrected for mean is used. Thus the ratio now becomes,

$$\gamma_k = \frac{\sigma k}{\sigma - \sigma_0} \qquad (9)$$

These values for the four set of data are presented in Tables 1,2,3 and 4. The values are presented for the harmonics which have significant contribution to the total variance. Other harmonics are omitted. The fraction variance taken by the zeroth harmonic is calculated by Eq. 8. Rest are all calculated using Eq. 9. It is seen that the additional contribution

Table 1

Harmonic Analysis of Hydrocarbons from Richmond

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	4.081	0.0	4.081	0.0	16.66	.8255
1	-0.233	0.051	0.238	167.60	0.114	.0324
2	0.077	-0.200	0.215	291.07	0.092	.0263
3	0.171	0.009	0.171	3.13	0.059	.0168
4	-0.227	0.022	0.228	174.46	0.104	.0296
5	0.173	0.094	0.197	28.68	0.077	.0221
6	0.012	-0.243	0.243	273.07	0.118	.0336
7	-0.189	0.106	0.151	135.68	0.046	.0131
8	0.212	-0.174	0.275	320.72	0.151	.0430
11	-0.191	0.104	0.218	151.34	0.092	.0270
25	-0.150	-0.196	0.247	232.49	0.122	.0347
33	-0.092	0.135	0.163	124.22	0.053	.0153
49	0.087	0.163	0.185	61.85	0.068	.0195
52	0.110	0.108	0.155	44.50	0.048	.0137
58	0.050	-0.172	0.179	286.45	0.064	.0183
60	0.297	0.278	0.407	43.12	0.332	.0945

Table 2

Harmonic Analysis of Hychrocarbon from Oakland

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	3.933	0.0	3.933	0.0	15.471	0.8879
1	-0.219	0.196	0.294	138.15	0.173	0.0889
2	0.230	-0.147	0.273	327.38	0.150	0.0768
3	0.0004	0.096	0.096	89.72	0.018	0.0096
4	-0.025	-0.187	0.189	262.30	0.071	0.0367
5	0.061	0.080	0.101	52.33	0.020	0.0105
6	-0.017	-0.234	0.234	265.59	0.110	0.0564
7	-0.126	-0.080	0.149	212.27	0.044	0.0230
8	0.184	-0.114	0.217	328.17	0.094	0.0484
11	-0.102	0.120	0.157	130.41	0.049	0.0255
17	0.108	-0.074	0.131	325.35	0.034	0.0177
54	-0.003	0.104	0.104	91.67	0.021	0.0111
55	0.020	-0.097	0.099	282.00	0.019	0.0100
56	0.060	0.1108	0.126	61.34	0.031	0.0160
58	0.109	-0.101	0.149	317.19	0.044	0.0220
60	0.1812	0.2524	0.3107	54.30	0.193	0.0988

Table 3

Harmonic Analysis of Carbon Monoxide from Oakland

Harmonic	C_1	C_2	Amplitude	Phase	Contribution of Variance	Fraction Variance
0	3.171	0.0	3.171	0.0	10.05	0.6988
1	-0.164	0.252	0.301	123.01	0.181	0.0419
2	0.404	-0.217	0.459	331.82	0.421	0.0973
3	-0.071	0.149	0.165	115.69	0.054	0.0126
4	-0.000	-0.216	0.216	269.95	0.094	0.0217
5	-0.062	0.097	0.115	122.53	0.026	0.0062
6	0.005	-0.393	0.393	270.79	0.309	0.0714
7	-0.090	-0.168	0.190	241.59	0.079	0.0168
8	0.374	-0.206	0.427	331.12	0.365	0.0843
9	-0.161	-0.003	0.161	181.26	0.052	0.0121
11	-0.079	0.139	0.160	119.74	0.051	0.0119
56	0.084	0.180	0.199	64.93	0.079	0.0184
120	-0.074	-0.219	-0.232	251.36	0.107	0.0240
180	0.251	-0.125	0.280	333.57	0.157	0.0363

Table 4

Harmonic Analysis of Carbon Monoxide from San Jose

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	3.941	0.0	3.941	0.0	15.536	0.7666
1	-0.251	0.185	0.312	143.67	0.195	0.0412
2	0.401	-0.207	0.452	332.75	0.408	0.0865
3	-0.0002	-0.143	0.143	269.89	0.041	0.0087
4	0.165	-0.272	0.318	301.27	0.202	0.0429
6	-0.031	-0.312	0.313	264.24	0.196	0.0416
7	-0.221	-0.161	0.273	216.12	0.150	0.0317
8	0.348	0.048	0.351	7.97	0.247	0.0522
9	-0.236	-0.005	0.236	181.21	0.055	0.0236
10	-0.000	0.168	0.168	90.05	0.056	0.0120
11	-0.001	0.178	0.178	90.63	0.064	0.0135
17	0.133	0.199	0.239	56.11	0.114	0.0240
21	0.007	-0.188	0.189	272.21	0.071	0.0151
60	-0.114	-0.143	0.184	231.41	0.067	0.0143
120	-0.134	-0.360	0.384	249.50	0.295	0.0624
180	0.135	-0.154	0.205	311.30	0.084	0.0170

to the total variance by the harmonics given in the Tables 1 thru 4 is 7.5% for hydrocarbons from Richmond, 5.5% for hydrocarbons from Oakland, 14% for carbon monoxide from Oakland and 11% for the carbon monoxide from San Jose. The remaining variance in each case is distributed among other harmonics.

The maximum contribution to the variance for the hydrocarbon data from Richmond and Oakland is by the 60^{th} harmonic being 9.45% for the former and 9.88% for the later. Out of the rest the maximum is contributed by the first eight harmonics. This contribution is about 22% for the hydrocarbons from Richmond and is about 35% for the hydrocarbons from Oakland. This gives an indication that there is probably a daily variation of hydrocarbons at these two stations and also there is a low frequency variation with the maximum detectable period of two months. Out of the low frequency harmonics the 1^{st}, 6^{th}, and 8^{th} harmonics appear to be predominant for the hydrocarbons from Richmond and the 1^{st}, 2^{nd}, 6^{th} and 8^{th} for the hydrocarbons from Oakland. The rest of the variance is distributed among other harmonics. A daily variation in the hydrocarbon concentration is expected because of daily social activities.

The carbon monoxide data from Oakland and San Jose also indicate that there is a greater contribution to the total variance by the low frequencies but in this case the fractional day cycle are also prominent. The first ten harmonics of the data from Oakland contribute about 35% to the total variance and that from San Jose contribute about 30%. The 2^{nd}, 6^{th} and 8^{th} harmonics appear to be the most prominent among them. At Oakland the daily variation is not significant while at San Jose it contributes about 1.43%. It is seen that the contribution of the 120^{th}

and 180^{th} harmonics which correspond to 12 hr. and 8 hr. cycles are quite significant at both the stations. Automobiles, coal combustion, incineration and petroleum plants are major contributors to the carbon monoxide concentration. The fractional day cycles can be due to the excessive automobile traffic at office and factory timings.

Autocorrelation

The Fourier analysis described above has been used as preliminary analysis of the air pollution data. With this analysis we can determine the discrete periodicities corresponding to length of the record while it is difficult to determine the exact periodicities present in the data and also those periodicities which have continuous frequency bands. Spectral analysis has been used to determine these hidden periodicities. Some basic and essential concepts of spectral analysis with reference to the analysis of air pollution data are described below. An exhaustive study of these can be found in references 33 through 38.

The theory of spectral analysis is applicable to the stationary stochastic processes. A stochastic process is called stationary if its mean and variance have no trend and are independent of time and also the correlation of the process at two different time depends only on the time span and not on the time itself. The total hydrocarbons and the carbon monoxide concentration obtained from the three stations are in general found to form a stationary time series. Any nonstationarity present in the series was however removed by fitting a second order polynomial regression model.

To determine the periodicities by the spectral analysis method the first step is to determine the relationship of pollutant concentration

in the series at different time spans also called lags. This relationship is expressed by the autocovariance or autocorrelation function. The autocovariance between the pollutant concentration at a time lag of k hours is given by [33].

$$C_{xx}(k) = \frac{1}{N} \sum_{t=1}^{N-k} x_t \, x_{t+k} \tag{10}$$

where x's are the deviations from the mean. The divisor used in the above equation is N and not N-k, as given in some literatures, because the value of record length N used is large. The autocorrelation is then expressed as,

$$\rho_{xx}(k) = \frac{C_{xx}(k)}{C_{xx}(0)} \tag{11}$$

The denominator in the above equation represents the variance of the time series. The autocorrelation is an even function. Thus,

$$\rho_{xx}(k) = \rho_{xx}(-k) \tag{12}$$

and it always lies between -1 and 1. A positive correlation at a lag of k hours implies that the variation of the pollutant concentration would follow the same direction after k hours and a negative correlation would imply that this variation would follow the opposite direction after k hours. Thus with the positive autocorrelation an increasing (decreasing) trend in the pollutant concentration would follow an increasing (decreasing) trend after k hours while if the correlation is negative the pollutant concentration would have a decreasing (increasing) trend after k hours. The autocorrelation retains all the harmonics of

the given time series but discards any phase difference between them. Thus, for example, all the periodic functions having the same harmonic amplitude but different phase would have the same autocorrelation.

A second order polynomial regression model was first fitted to each of the data shown in Figs. 1-4 using Eqns. 19, 22 and 23 to remove any trend present in the data. The residuals were then used to calculate the autocorrelation function. The number of lags, M, used were 140. The selection of M is described later under power spectrum. The execution time for the calculation of autocorrelation function depends upon both N and M. It increases much faster with the increase in M than with the increase in N for the present case of N = 1440 and M = 140 the execution time was 55 secs.

The autocorrelation plots are shown in Figs. 5,6,7 and 8. For both hychrocarbons and carbon monoxide at all the stations, the autocorrelation falls rapidly from unity at lag zero to less than .01 within the first thirteen lags and then it fluctuates between 0 and .03 with the dying down amplitude going negative sometimes. There is a high correlation at lags of 24 hours, 48 hours, 79 hours and 96 hours.

POWER SPECTRUM

Although autocorrelation is important in the time domain yet in practice it is always preferable to use it's Fourier transform. This converts it from the time domain to the frequency domain and corresponding quantity is called the Spectral Density Function or the power spectrum. The advantages of studying the properties of a time series in frequency domain rather than in the time domain are [36].

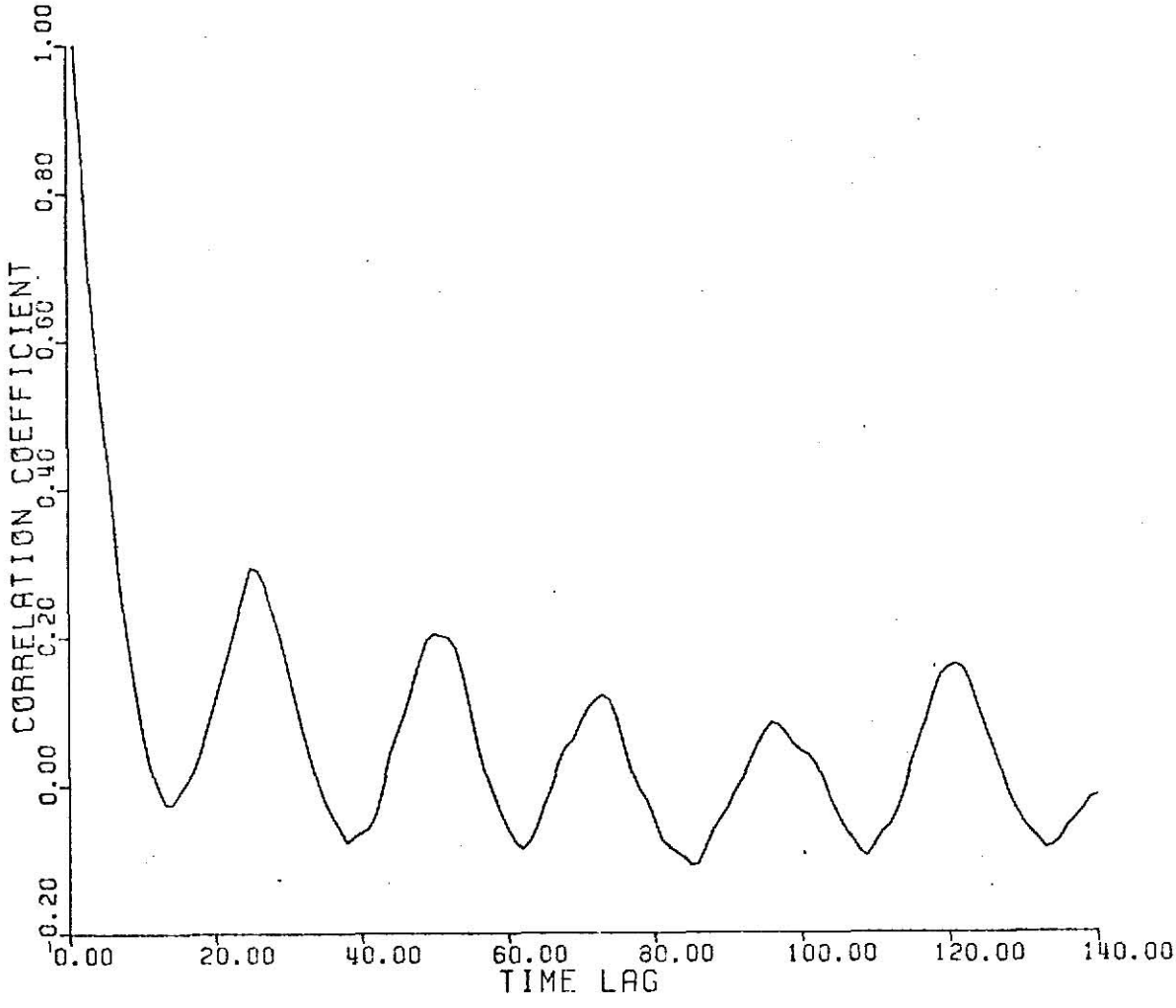

FIGURE 5. Autocorrelation function of hydrocarbons from Richmond.

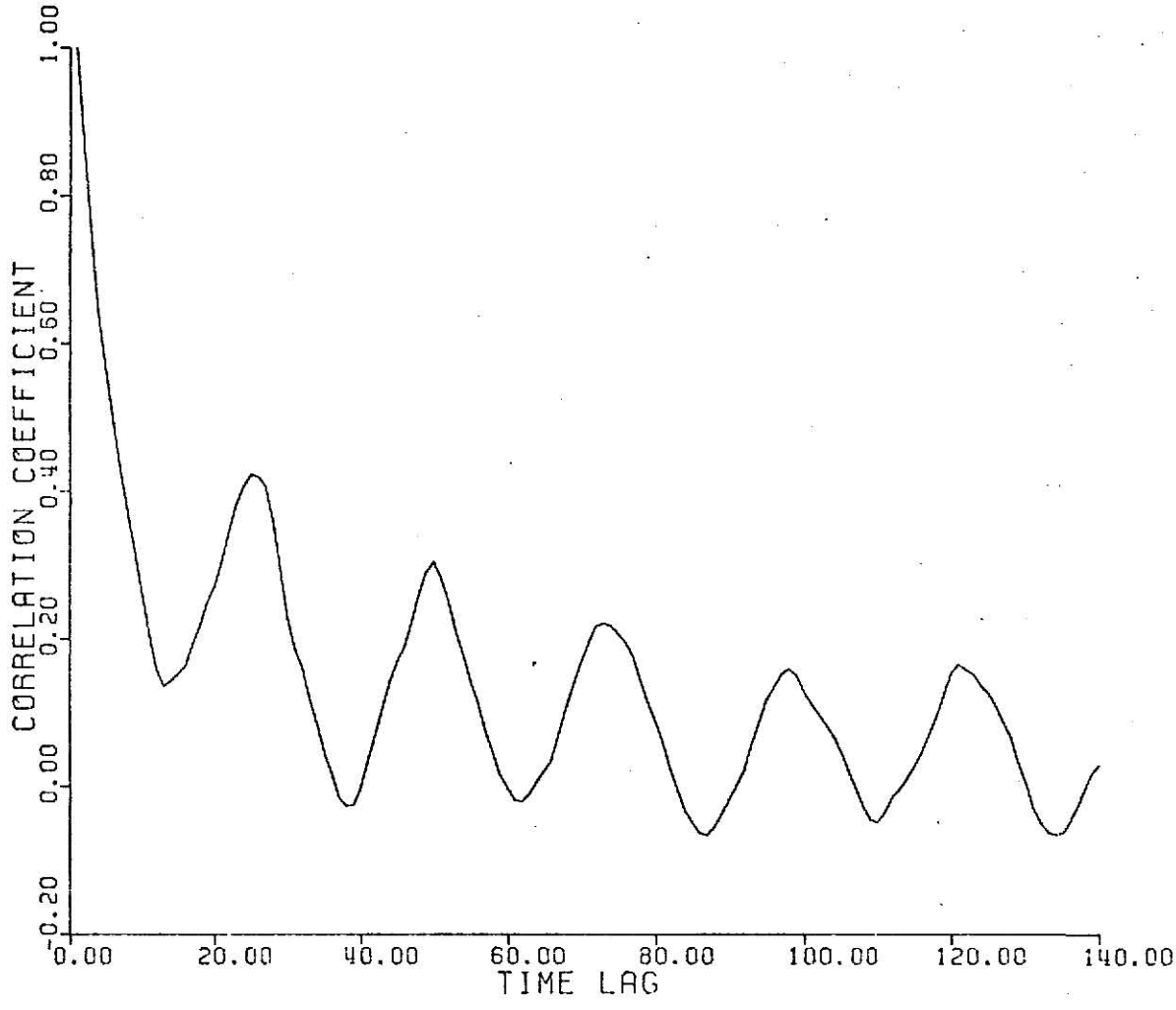

FIGURE 6. Autocorrelation function of hydrocarbons from Oakland.

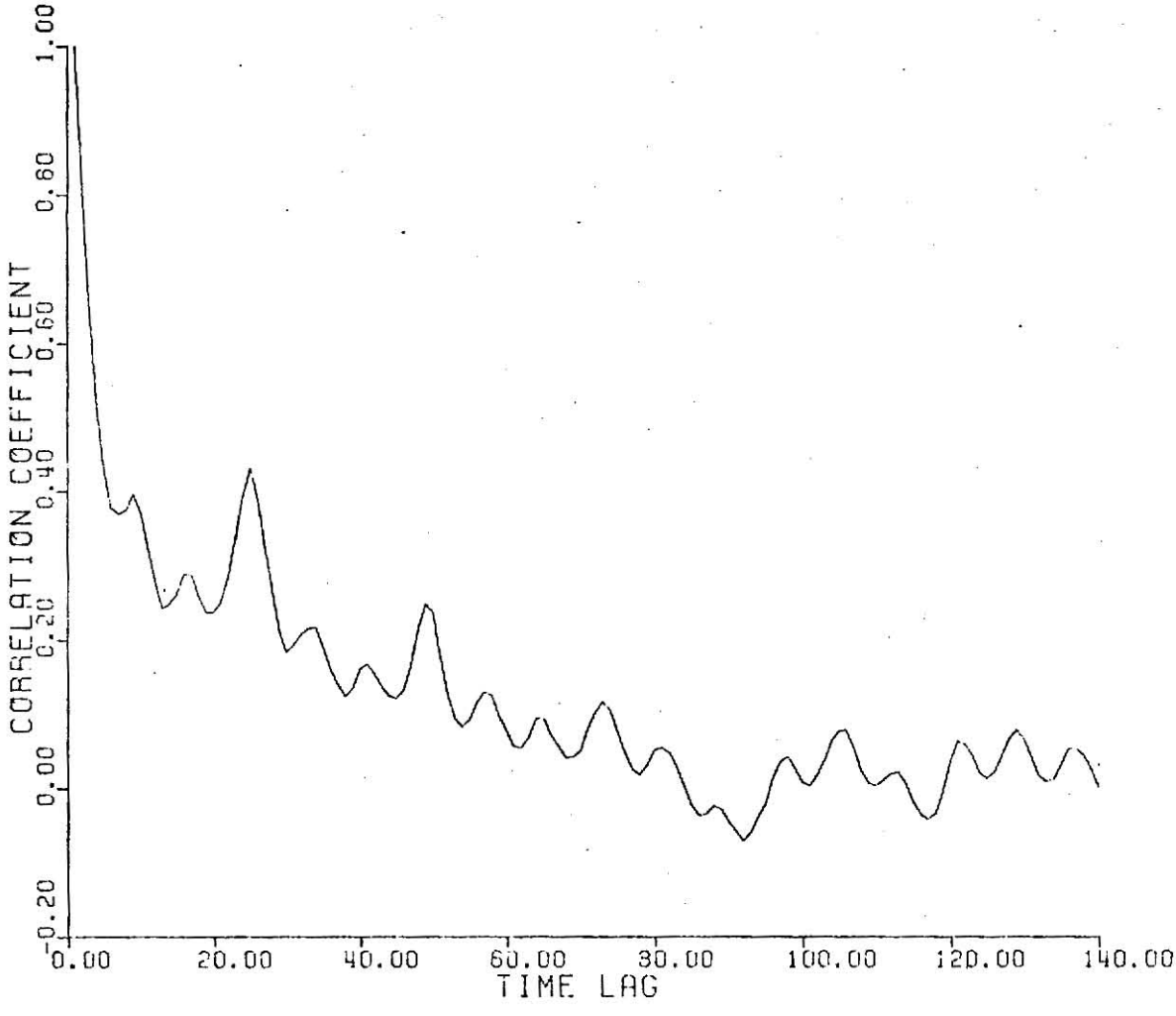

FIGURE 7. Autocorrelation function of carbon monoxide from oakland.

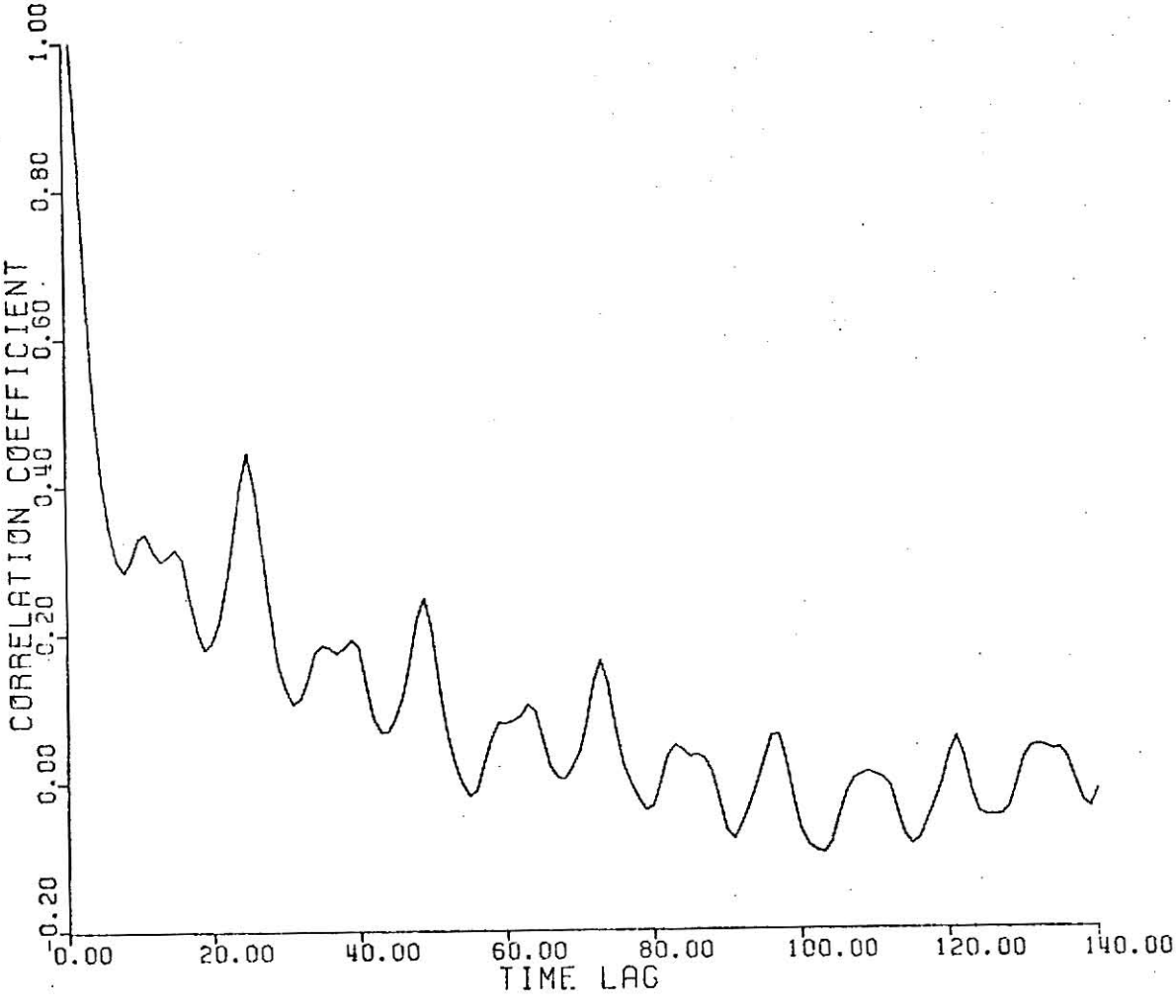

FIGURE 8. Autocorrelation function of carbon monoxide from San Jose.

1. The spectral density function conveys the information in terms of the relative square amplitude associated with oscillations at different frequencies while the autocorrelation stresses the dependence along the time axis.

2. The decomposition of the variance to the various frequency bands is independent of each other and hence permits a greater ease of analysis than does the correlation analysis where each value of the autocorrelation function is the weighted summation of the same contribution.

3. The spectral density function is easy for physical and graphical interpratation compared to autocorrelation.

Apart from these some other characteristics of the spectral density function are,

1. The variance due to the linear trend, if any in the data, is concentrated on the zero frequency spectral estimate. This zero frequency spectral estimate also includes a truly random fluctuation in the record and a low frequency trend in the data.

2. If the observations are made at intervals Δt then the maximum frequency which can be estimated is $1/2\Delta t$ with the period $2\Delta t$. If there are frequencies present with periods shorter than this then the confounding of the frequencies occurs that is the variance because of these short period frequencies is not lost but distributed in harmonics of the true period. This confounding is called 'aliasing' and creates a ambiguity in determining the origin of these variance.

3. The smaller interval used for getting the observations for a given record of time increases the number of observations and also the degree of freedom upon which the spectral estimate is based.

The Fourier transformation of the autocovariance function and it's subsequent solution results in,

$$f_x(\omega) = \frac{1}{2\pi}\left[C_{xx}(0) + \sum_{k=1}^{M-1} C_{xx}(k)\ \cos k\omega\right] \qquad (13)$$

where $f_x(\omega)$ is the spectral density function or power spectrum for time series x, ω the frequency in cycle/hr. and M the total number of lags. Because of the finite length of the records available the spectral estimates have to be smoothed or averaged out. The weighting function used for smoothing of the spectrum is called lag window in the time domain and spectral window in the frequency domain. In spectrum analysis our purpose is to estimate the spectrum within a frequency band around a given frequency. So two characteristics are required out of the spectral window. The first is that the side lobes should be smaller to reduce the possibility of distortion effect due to remote frequencies and the second is that the main lobe should concentrate around the principal frequency to give it a greater weight. According to the first requirement the lag window should be smooth and slowly changing while for the second requirement it should be flat and blocky because it's Fourier transform gives a sharp peak (33,34). Thus both the shape and width of the window are important and there has to be a compromise between them. There have been quite a number of windows proposed. Details about these can be found in references (33) and (34). Tukey-Hanning window was used for the work presented here. Smoothed estimates are calculated by the following expression

$$f_x(\omega_j) = \frac{1}{2\pi}\left[C_{xx}(0) + \sum_{k=1}^{M-1} C_{xx}(k)\, W(k) \cos \omega jk\right]$$

where,

$$W(k) = \begin{cases} \frac{1}{2}\left(1 + \frac{\cos k}{M}\right), & |k| \leq M \\ 0, & |k| > M \end{cases} \qquad (14)$$

and $\omega_j = \frac{\pi j}{NP}$, $\quad j = 0, \ldots, NP$

where NP is the number of plot points.

The log of this smoothed spectral estimate against frequency gives the spectrum plot. A large variance associated with a frequency appears as a peak on a frequency band around that frequency. The selection of the total number of lags M is an important consideration in the spectral analysis to get the real peaks. The total number of lags effect the band width of the window used and a proper value of the band width has to be used to fit the record available. Too small a value of M increases the band width of the window and thus suppresses the periodicities in the data. A too large a value of M decreases the band width of the window too much to give spurious peaks. The procedure followed to determine M is as given in Jenkins and Watts (33). To start with, a higher value of bandwidth, small M, was used. This masks some periodicities. The band width was now decreased by increasing M and the spectral plots compared till no change was observed for subsequent value of M. Three values of M were selected in the range of the values of M in which the

spectrum was found to converge and then diverge, and the subsequent spectral plots were compared. One of these was then selected for the pollution concentration data. The three values of M tried for the hydrocarbon data from Richmond were 100, 140 and 180. The value of M finally selected was 140. As the data from all the stations were very much similar, the same value of M was used for all of them.

The value of $C_{xx}(k)$ used in Eq. 14 was obtained using Eq. 11. The ω in the same expression varies from 0 to π . The desired frequency interval in the output determines the total number of plot points to be used. A smaller interval helps determine more accurate frequencies corresponding to peaks in the spectrum but at the same time may take more execution time. A large interval can overlap some periodicities. The frequency interval of .01 was used using the plot points as 300. The value of $f_x(\omega)$ can not be negative. A zero value of $f_x(\omega)$ was given a low value supplied in the program while taking its log. The execution time for the selected set of number of plot points and M was about 25 secs. The results are shown in Figures 9-12.

The spectral estimate obtained for the amount of data available is called the sample spectral estimate because the observed pollutant concentration data is only a sample of the infinite amount of data being generated from the process. The theoretical spectral estimate is based on the number of observations approaching infinity. The sample spectrum is, therefore, distributed about the theoretical estimate with a chi-square distribution with two degrees of freedom. The smooth spectral estimate is distributed as chi-square distribution with df degrees of

freedom depending on the type of window used. The degrees of freedom for the Tukey-Hanning window is given as,

$$df = \left(\frac{8N}{3M}\right) \tag{15}$$

Confidence limits were therefore drawn on the spectrum plot and peaks lying outside the confidence limits were considered the true periodicities present. At a particular frequency ω the confidence limits on the log scale are given as,

$$\text{Lower confidence limit} = L_n f_x(x) + L_n \frac{df}{x^2_{df}(1-\alpha/2)}$$

$$\text{Upper confidence limit} = L_n f_x(x) + L_n \frac{df}{x^2_{df}(\alpha/2)} \tag{16}$$

This is the confidence interval for all the frequencies and is represented on the figures by a vertical line. α in the above equations is the level of confidence. A 5% level of confidence was used for the air pollution data. The spectrum for all the pollutants from the three stations are given in Figs. 9 through 12. The hydrocarbons from Richmond have sharp peaks corresponding to the periods of one day, 16 hours and 12 hours frequencies while those from Oakland have peaks corresponding to one day, 12 hours and 7 hours frequencies. The carbon monoxide data from both Oakland and San Jose have sharp peaks outside the control limits corresponding to periods of one day, 12 hours and 8 hours frequencies. It is seen from the graphs that all the low frequencies with periods greater than one day can not be identified. One way of doing

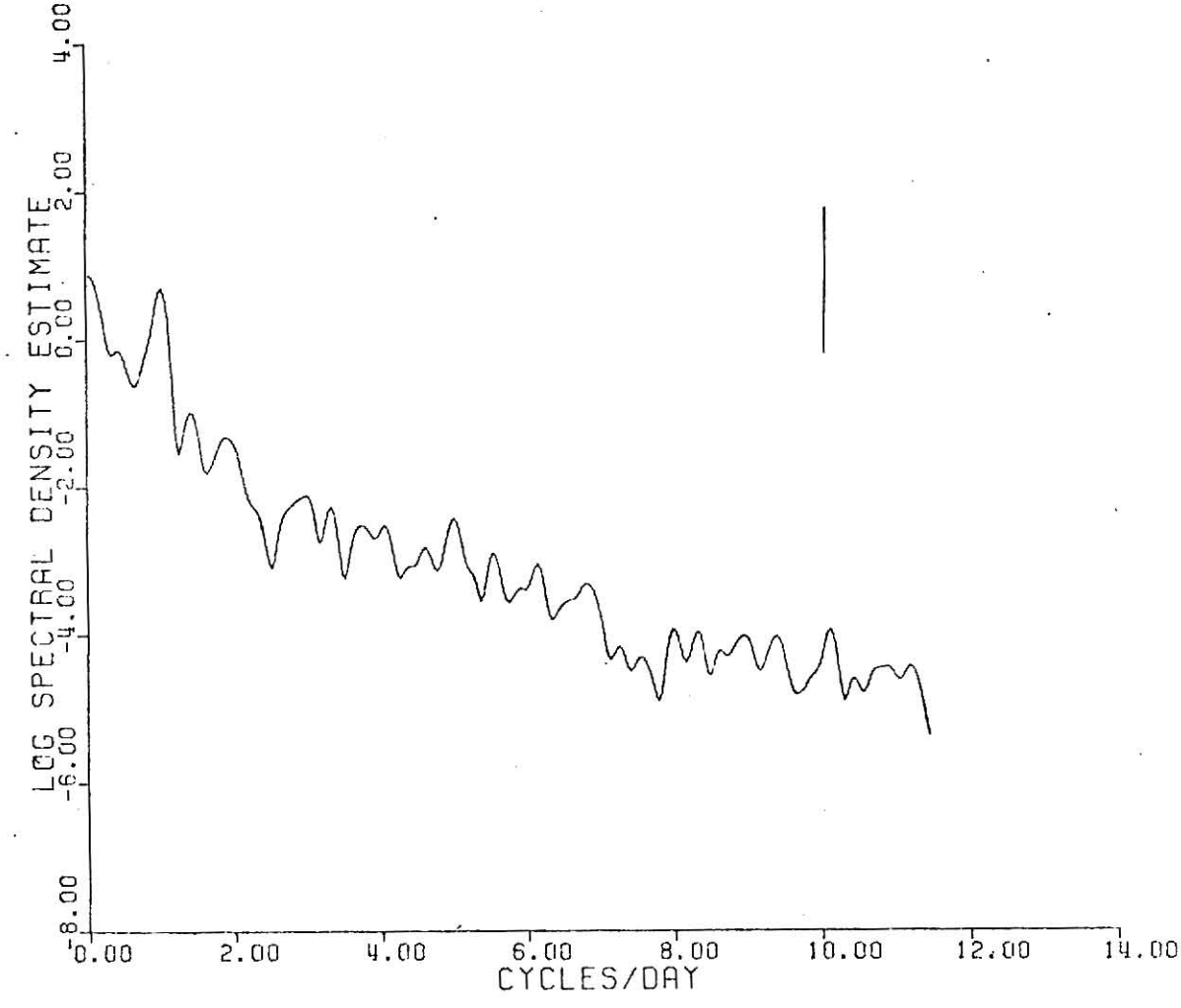

FIGURE 9. Spectral density function for hydrocarbons based on hourly data, Richmond.

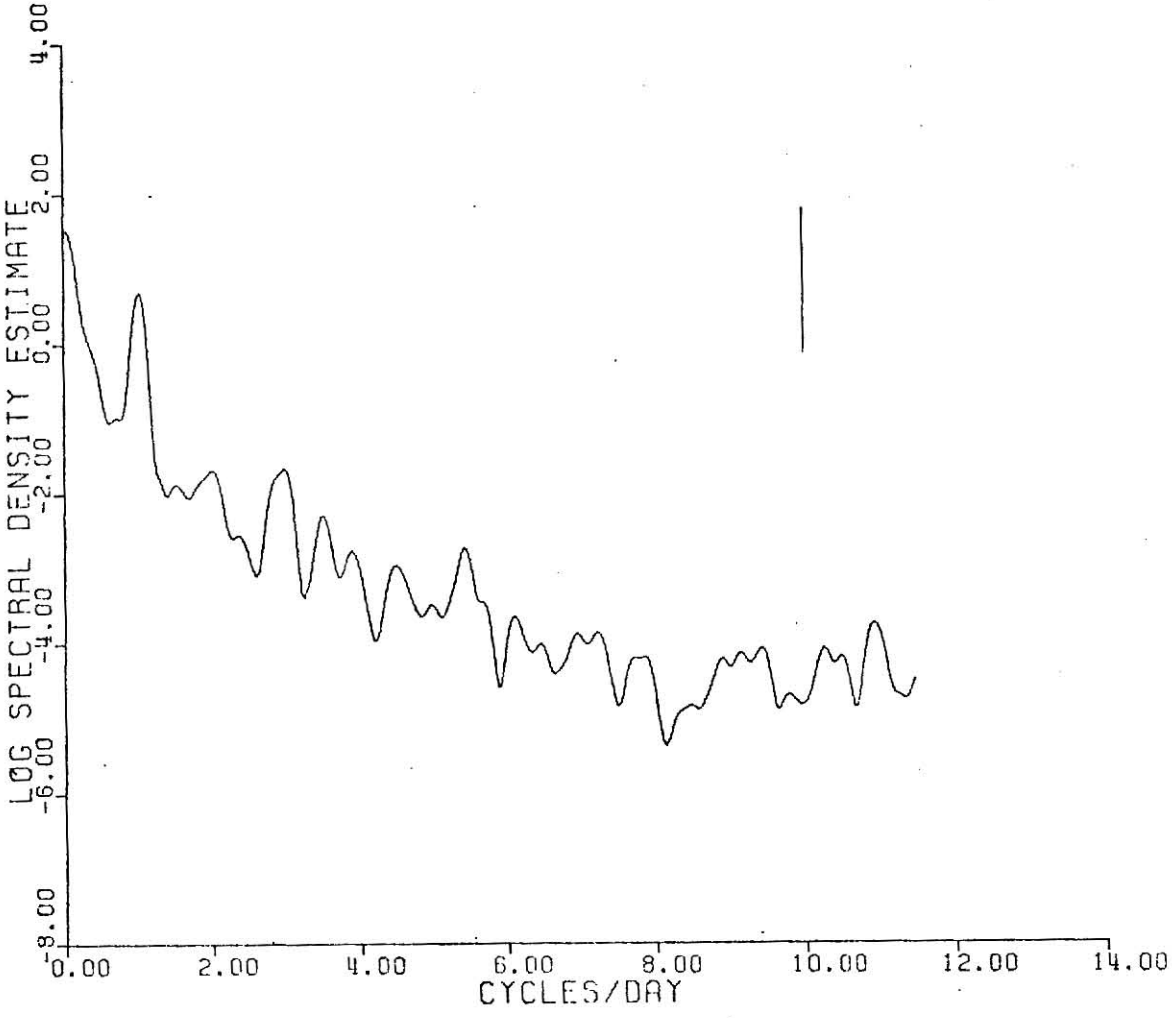

FIGURE 10. Spectral density function for hydrocarbons based on hourly data, Oakland.

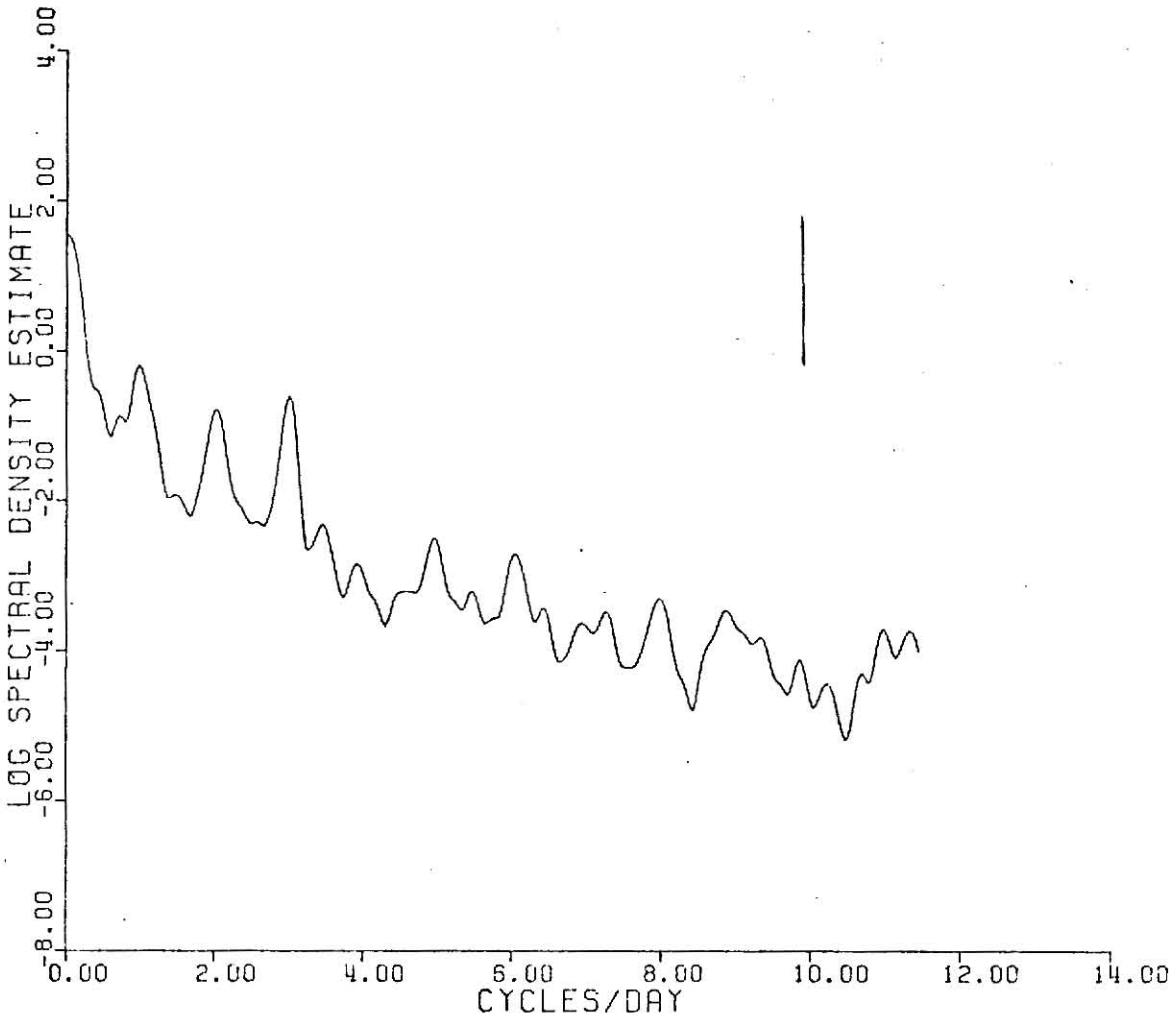

FIGURE 11. Spectral density function for carbon monoxide based on hourly data, Oakland.

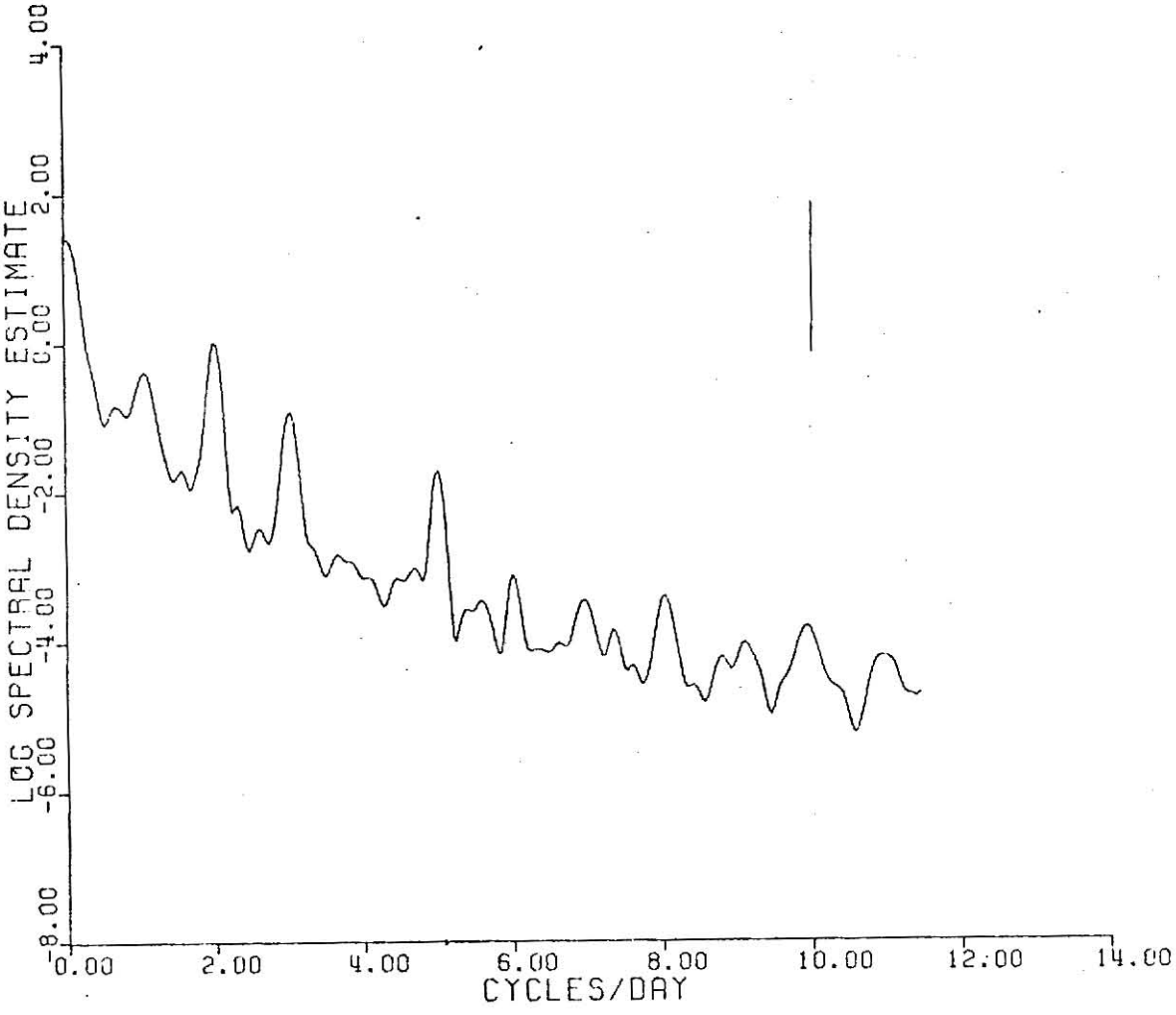

FIGURE 12. Spectral density function for carbon monoxide based on hourly data, San Jose.

this is by filtering out the high frequencies. This was done by calculating the 12 hours mean of the data. This formed another series and the spectrum plots are presented for hydrocarbon and carbon monoxide data in Figs. 13 through 16. It is seen that peaks upto one month period can be identified. Hydrocarbons appear to have 15 days and 7 or 8 days cycles at Richmond and Oakland. Carbon monoxide at Oakland appears to have 15 days and 10 days cycles while carbon monoxide at San Jose appears to have 15 days, 10 days and 7 or 8 days cycles. The frequencies thus found were removed or filtered out using regression analysis and the predictive model was developed for each using the procedure described below. The contributions due to one month and two months cycles were found to be significant in the fourier analysis. These cycles were also included in the model.

Development of Predictive Model

The regression analysis is based on the method of least squares. It fits a model to the data such that the deviation of the predicted values from the actual values is minimized. For a general regression problem the model to be fitted to the data can be given in the matrix form as, [41]

$$Y = XP + \varepsilon \tag{17}$$

where Y is the vector of observations given by,

$$Y^T = (Y_1, Y_2, \ldots, Y_N) \tag{18}$$

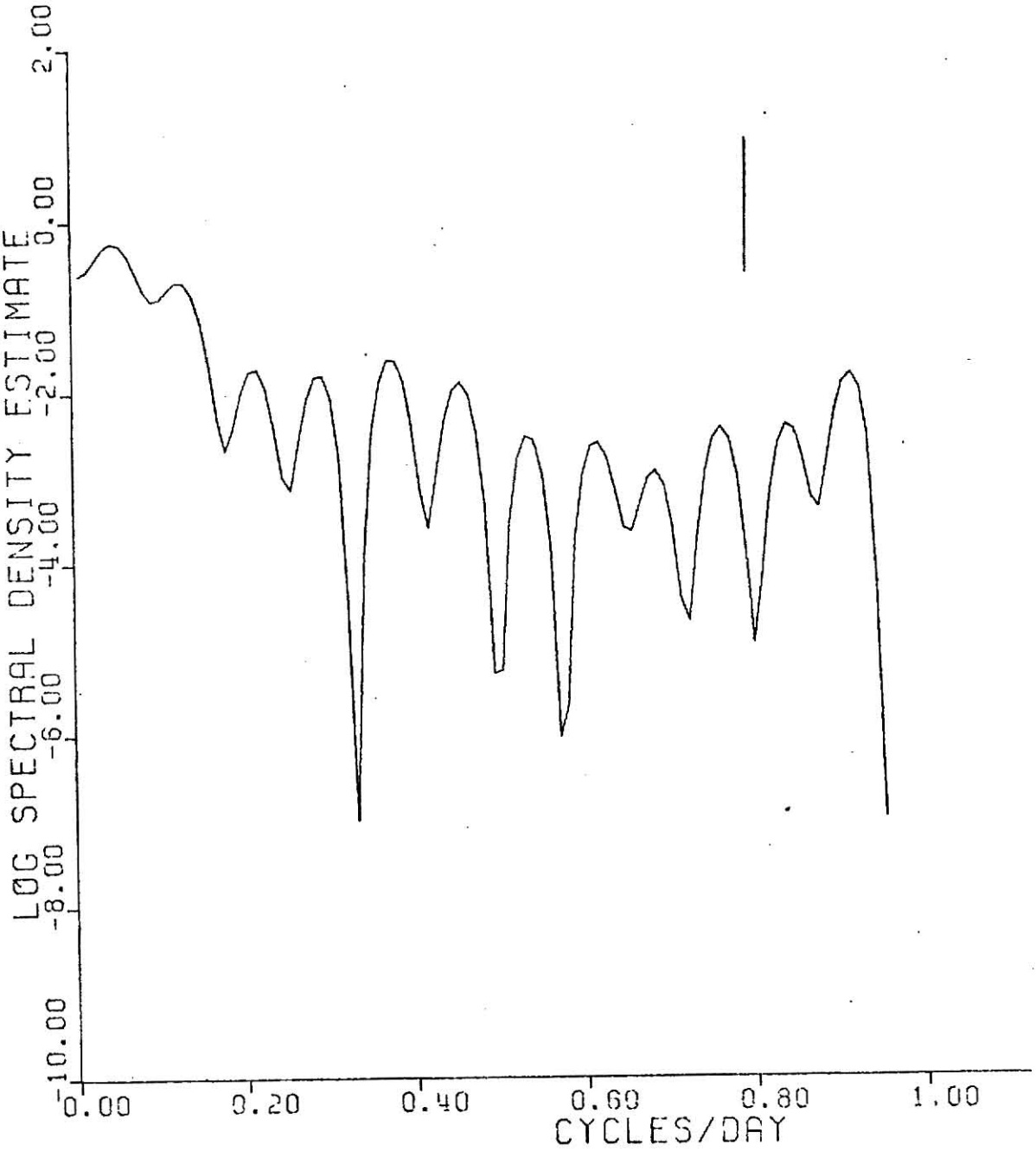

FIGURE 13. Spectral density function for hydrocarbons based on twelve hours averaged data, Richmond.

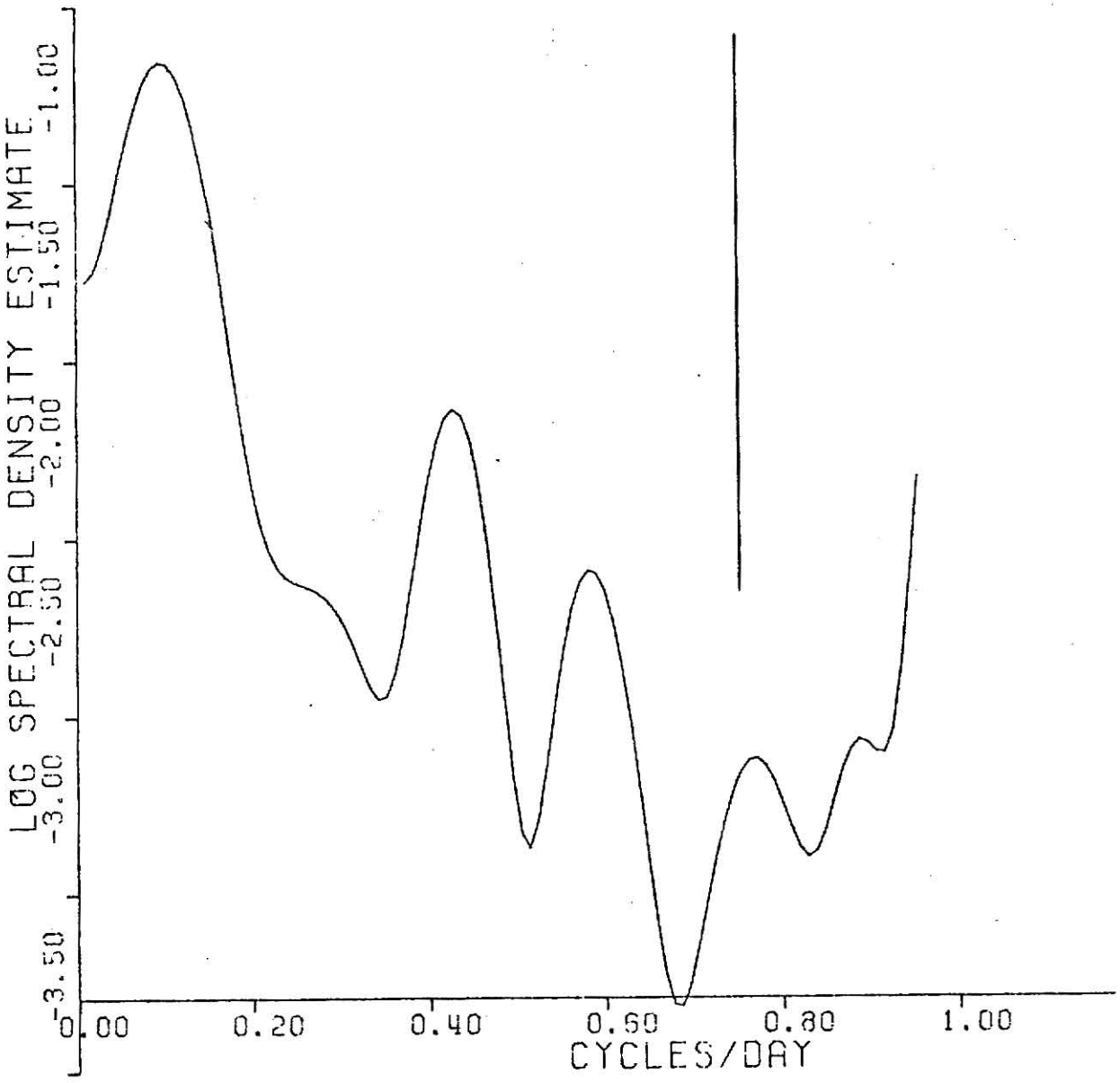

FIGURE 14. Spectral density function for hydrocarbons based on twelve hours averaged data, Oakland.

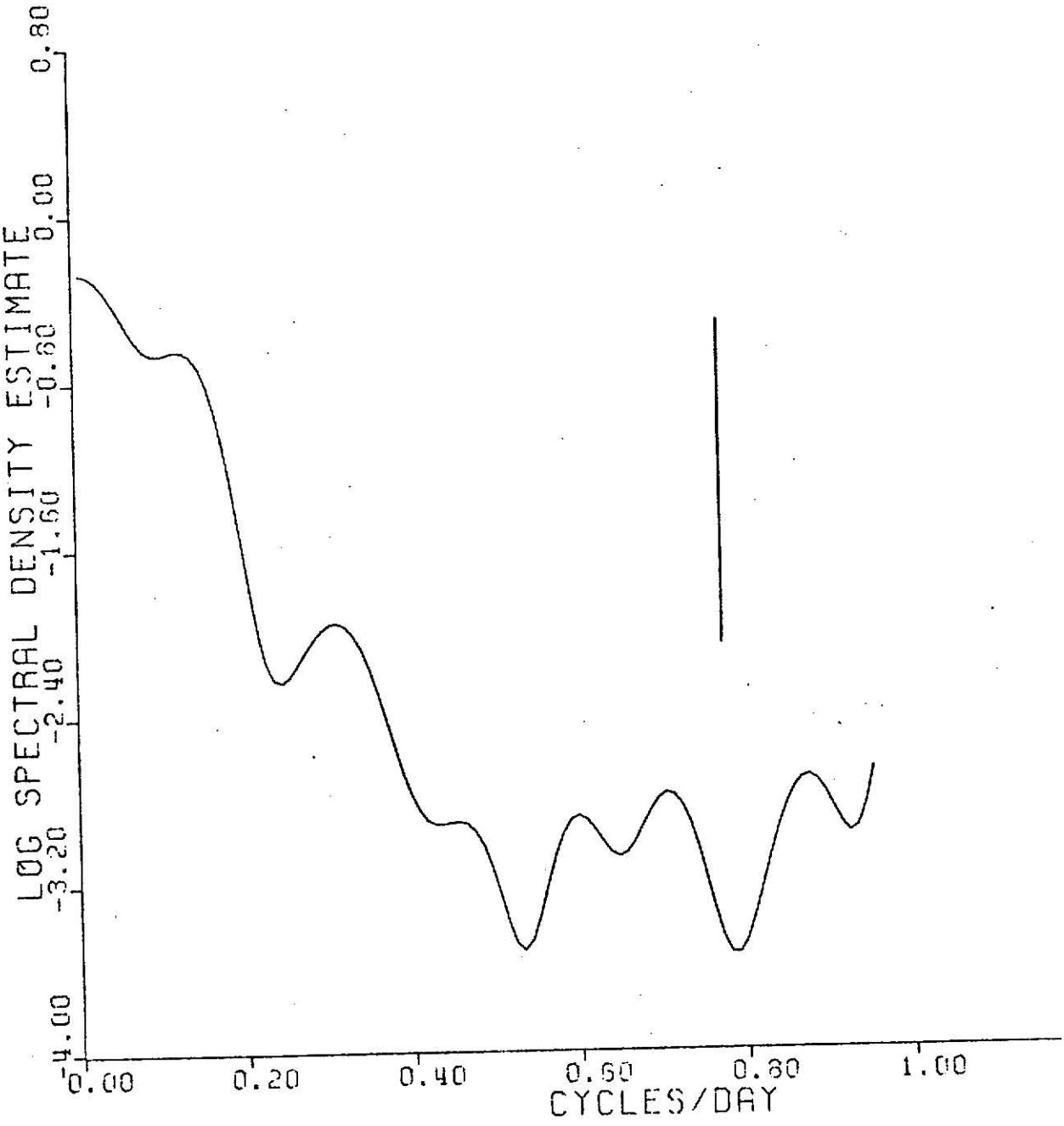

FIGURE 15. Spectral density function for carbon monoxide based on twelve hours averaged data, Oakland.

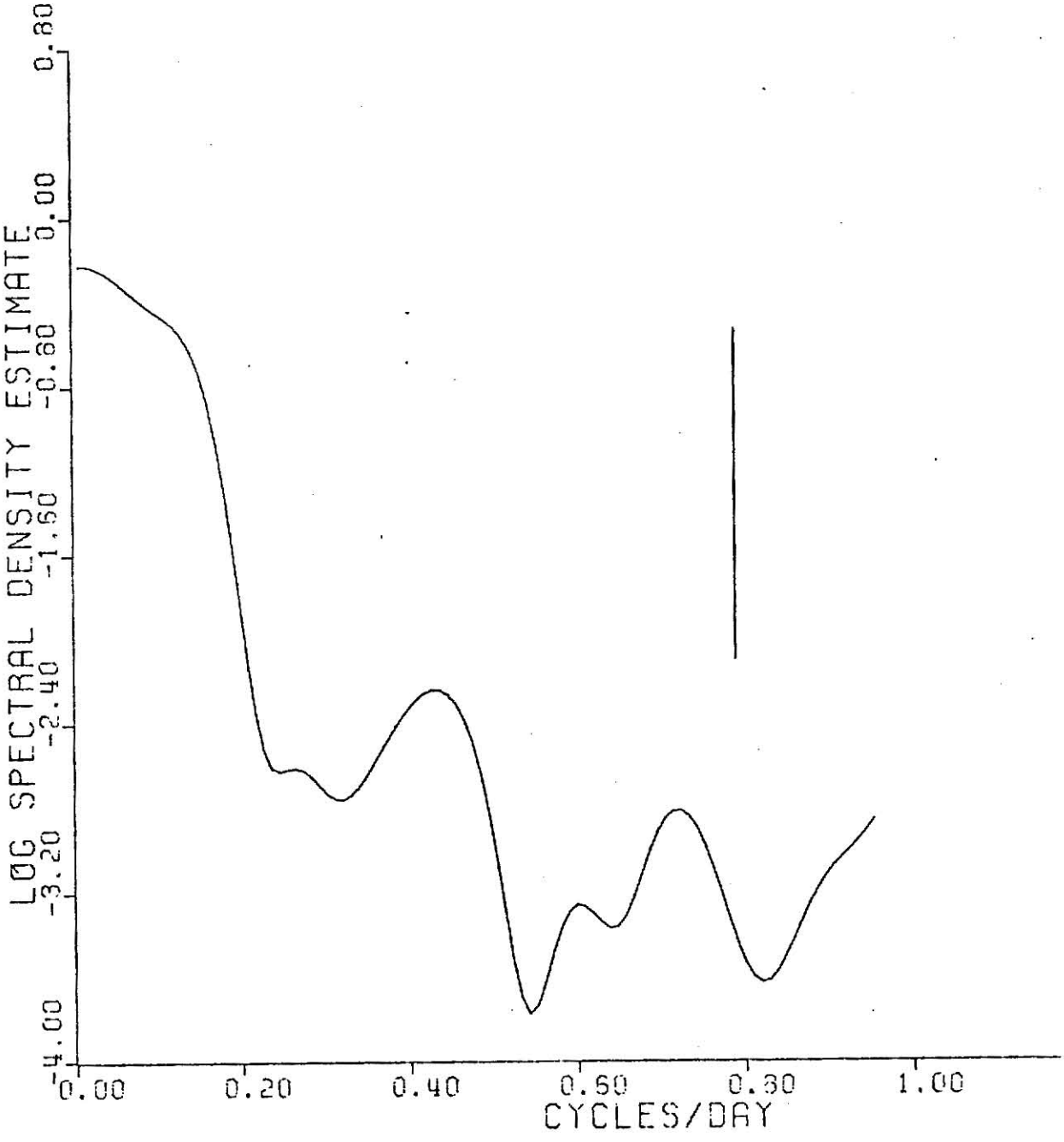

FIGURE 16. Spectral density function for carbon monoxide based on twelve hours averaged data, San Jose.

The superscript T represents the transpose. X is the matrix of the transformation of the independent variable given by,

$$X = \begin{pmatrix} 1 & Z_{11} & Z_{21} & \cdots & Z_{r1} \\ 1 & Z_{12} & Z_{22} & \cdots & Z_{r2} \\ \vdots & & & & \\ 1 & Z_{1N} & Z_{2N} & \cdots & Z_{rN} \end{pmatrix} \tag{19}$$

where Z represents the transformation, and r is the number of transformations. P is the vector of parameters given by

$$P^T = (P_0, P_1, \ldots, P_r) \tag{20}$$

and ε is the error of estimate. The least square estimate of the parameters P, namely

$$p^T = (p_0, p_1, \ldots, p_r) \tag{21}$$

is given by,

$$p = (X^T X)^{-1} X^T Y \tag{22}$$

Hence the fitted model obtained is,

$$Y' = Xp \tag{23}$$

This model is termed as the predictive model. For the air pollution data a transformation,

$$Z_1 = X \qquad \text{and} \qquad Z_2 = X^2 \tag{24}$$

was first used to remove a linear or nonlinear trend. The residuals, Y-Y', were then analyzed for the periodicities by plotting the spectrum as described before. Additional transformations were now used corresponding to the sin and cosine terms of the predominating periodicities observed in the spectrum plot. Thus,

$$Z_4 = \cos(wt) \qquad \text{and} \qquad Z_5 = \sin(wt) \tag{25}$$

where ω is the observed frequency.

The ability of the model to predict the pollutant concentration was tested using multiple correlation coefficient. It is given by the ratio,

$$R^2 = \frac{\sum_i (Y_i' - \bar{Y})^2}{\sum_i (Y_i - \bar{Y})^2} \tag{26}$$

The numerator is the sum square due to regression and the denominator the total sum square corrected for mean. R^2 is a measure of usefulness of the terms other than the mean in the model. The value of R^2 lies between 0 and 1. Higher it's value better is the prediction.

In the development of the predictive models, in general, there are two contradictory requirements. The first is that the model should include as many variables as possible to make it a reliable predictor and the second is that it should contain as less variables as possible to decrease the cost involved in obtaining the information about the variable. Stepwise regression, one of the procedures which compromise between both of these requirements, was used. The procedure is explained

below for the hydrocarbon model at Richmond.

In the first step the simple correlation matrix is determined among all the variables and the independent variable which has the maximum correlation with the dependent variable Y is included in the model. Variable $\sin(2\pi t/24)$ has the maximum correlation of .25 and thus enters the model. In the second step the partial correlation coefficients are calculated and that variable which has the maximum partial correlation with the dependable variable is included. The second variable now included is $\cos(2\pi t/180)$. In the third step the individual importance of each variable to the dependent variable is determined by the partial F-test. Though one variable at one stage might have been the best to enter the model yet at later stage it might become redundant because of the presence of other variable. Which means that both the variables are carrying the same information. Thus one variable is eliminated according to the F-value used. Similarly the next variable is selected to be included in the regression equation from the partial correlation coefficients. Each time a new variable is added the partial F-test is made to eliminate the redundant one. The procedure is thus carried out till all the variables are tried.

The lower and upper limit for the f-value used is zero. This would include all the variables. The multiple correlation coefficient at each stage indicates how much the new variable improves the prediction equation. R^2 was plotted against the number of variables in the model. The point where the value of R^2 does not improve significantly with the addition of more variables was taken as the truncation point for the number of variables

in the model. The spectrum was obtained of the residuals to find the remaining periodicities. Fig. 17 shows the variation of R^2 with the addition of each variable. For the hydrocarbons from Richmond it is seen that after the tenth variable the improvement is not very significant. The spectrum plot of residuals for the hydrocarbons from Richmond is shown in Fig. 18. It is seen that the peaks of all the fed-in frequencies have decreased but they haven't been eliminated. Some fluctuations are still present. The predictive model obtained is,

$$Y = -.4789 \cos(2\pi t/1440) - .3860 \sin(2\pi t/720)$$
$$-.4577 \cos(2\pi t/300) - .4615 \sin(2\pi t/240)$$
$$+.5219 \cos(2\pi t/180) + .4268 \cos(2\pi t/24)$$
$$+.6768 \sin(2\pi t/24) - .2057 \sin(2\pi t/12)$$
$$+.1888 \cos(2\pi t/720) - .1708 \sin(2\pi t/180)$$

with $R^2 = .5077$

Similar procedure was used to develop the models for pollutants from other stations and the results are given below. The predictive model for hydrocarbons from Oakland is

$$Y = .0004t - .4599 \cos(2\pi t/1440) + .3725 \sin(2\pi t/1440)$$
$$+ .4841 \cos(2\pi t/720) - .2515 \sin(2\pi t/720) - .3793 \sin(2\pi t/360)$$
$$- .4615 \sin(2\pi t/240) + .4248 \cos(2\pi t/180) + .2191 \cos(2\pi t/24)$$
$$+ .5791 \sin(2\pi t/24) + .1720 \cos(2\pi t/8)$$

with $R^2 = .6431$

The spectral plot of the residuals is shown in Fig. 19. The predictive model for carbon monoxide from Oakland is

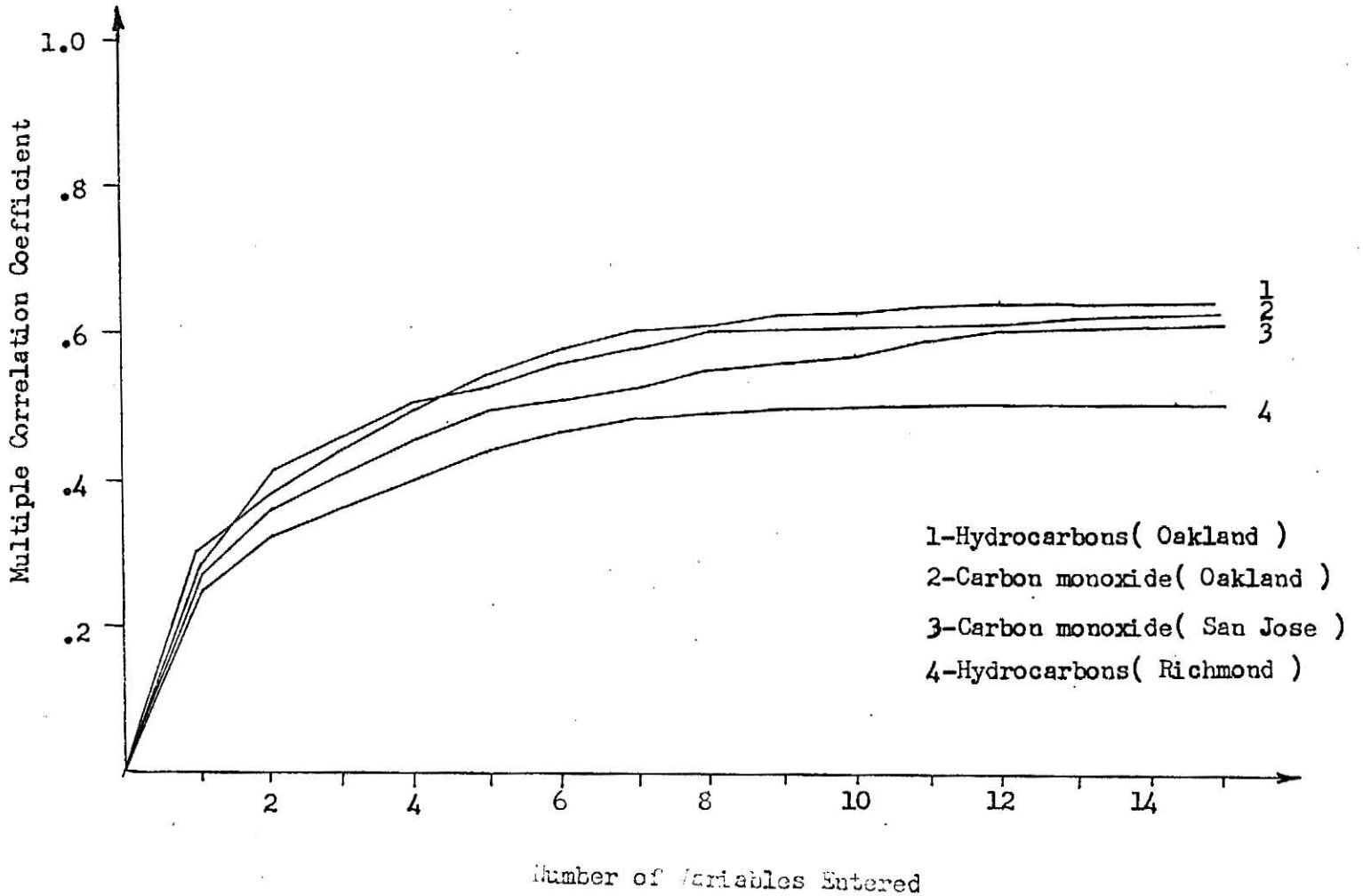

FIGURE 17. Improvement of multiple correlation coefficient against number of variables included in the model.

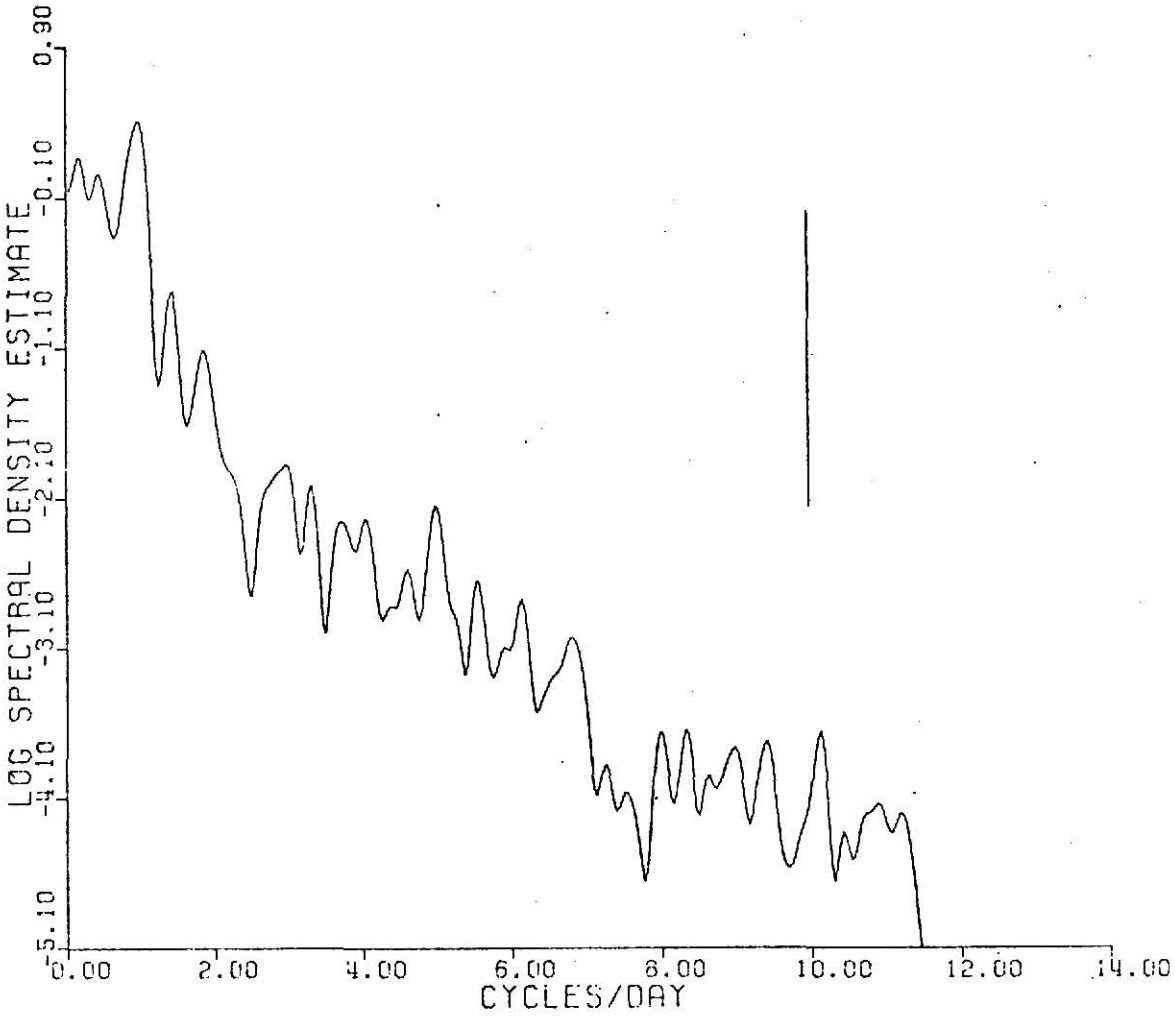

FIGURE 18. Spectrum of residuals for hourly hydrocarbons data from Richmond.

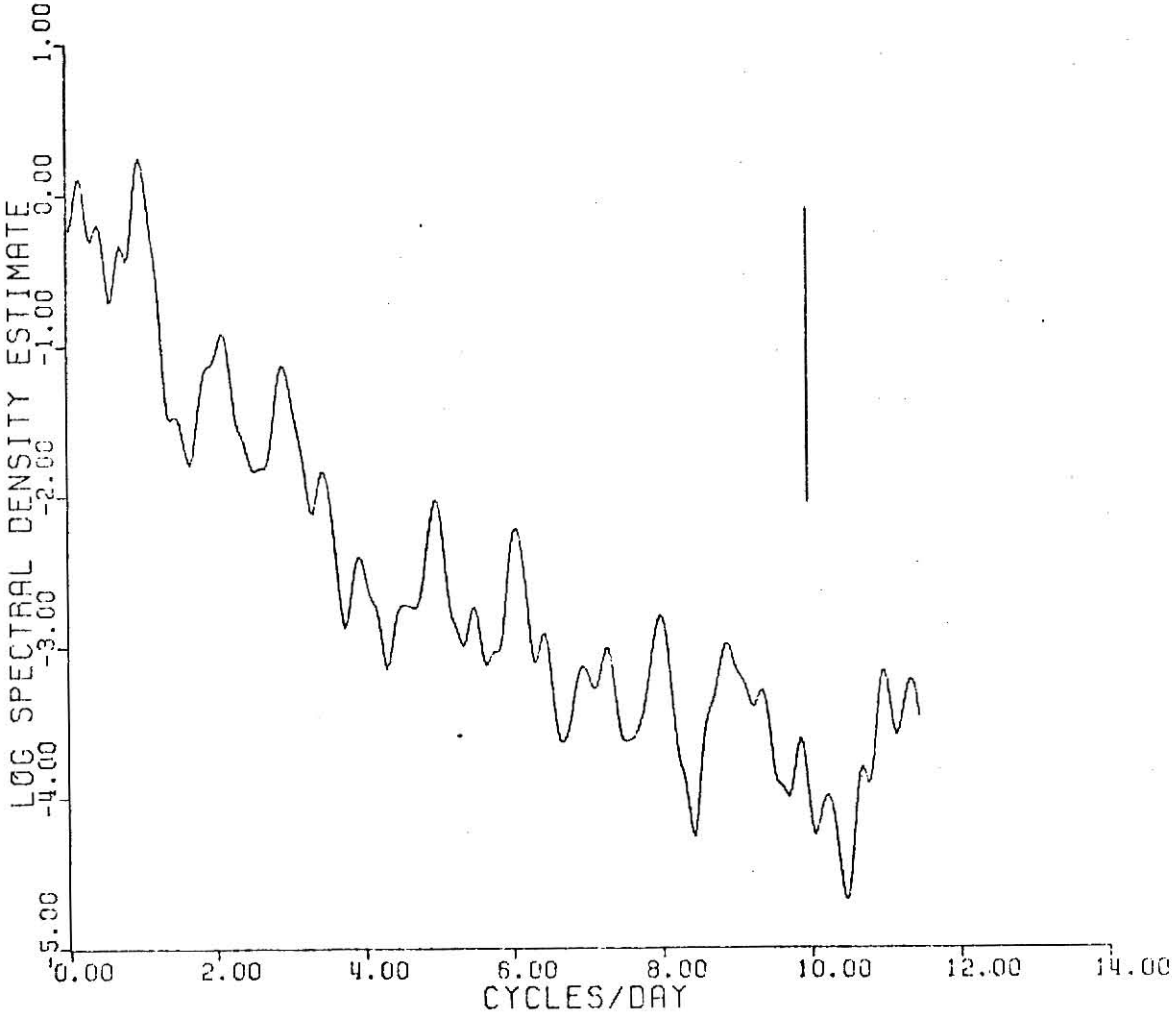

FIGURE 19. Spectrum of residuals for hourly hydrocarbons data from Oakland.

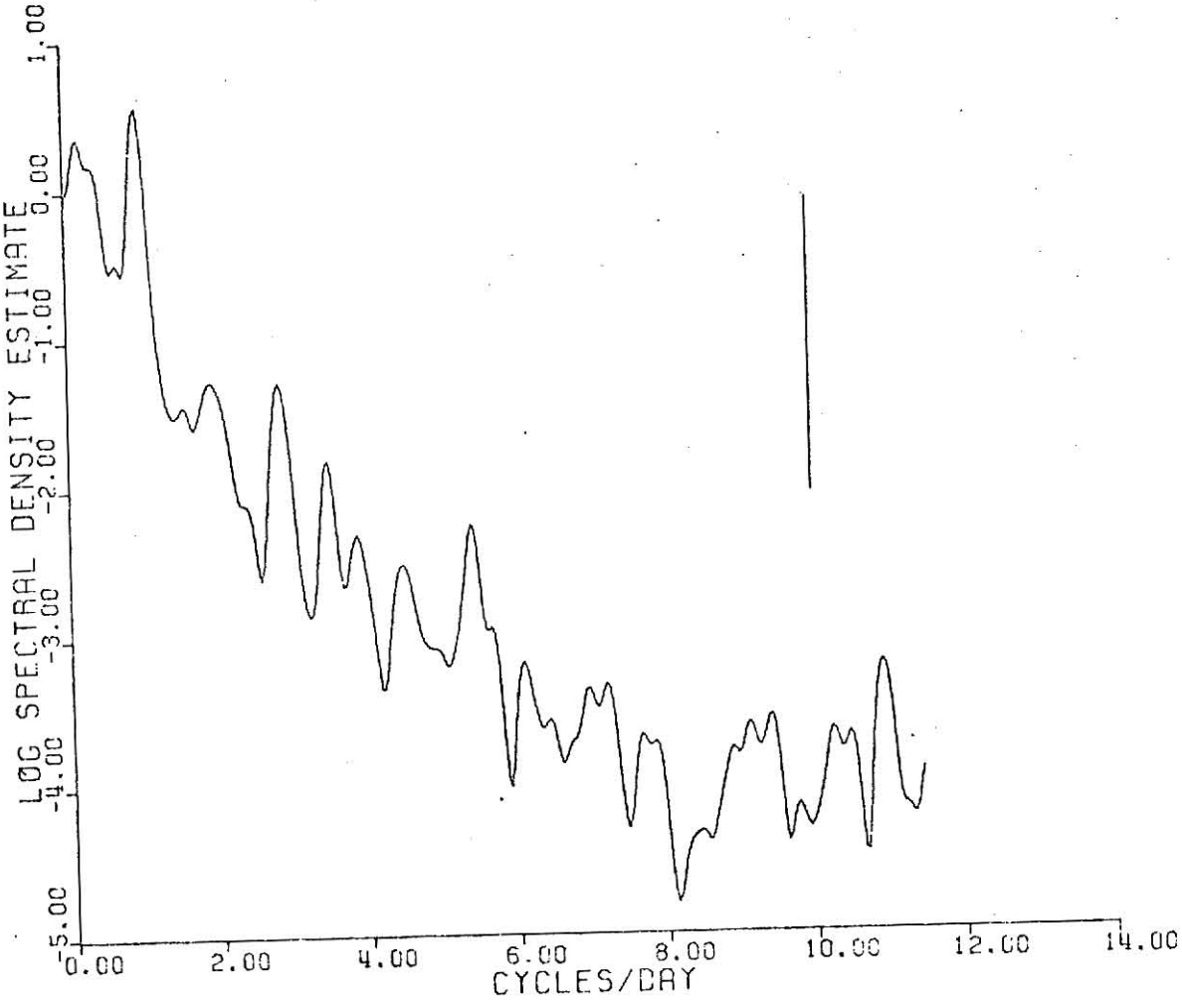

FIGURE 20. Spectrum of residuals for hourly carbon monoxide data from Oakland.

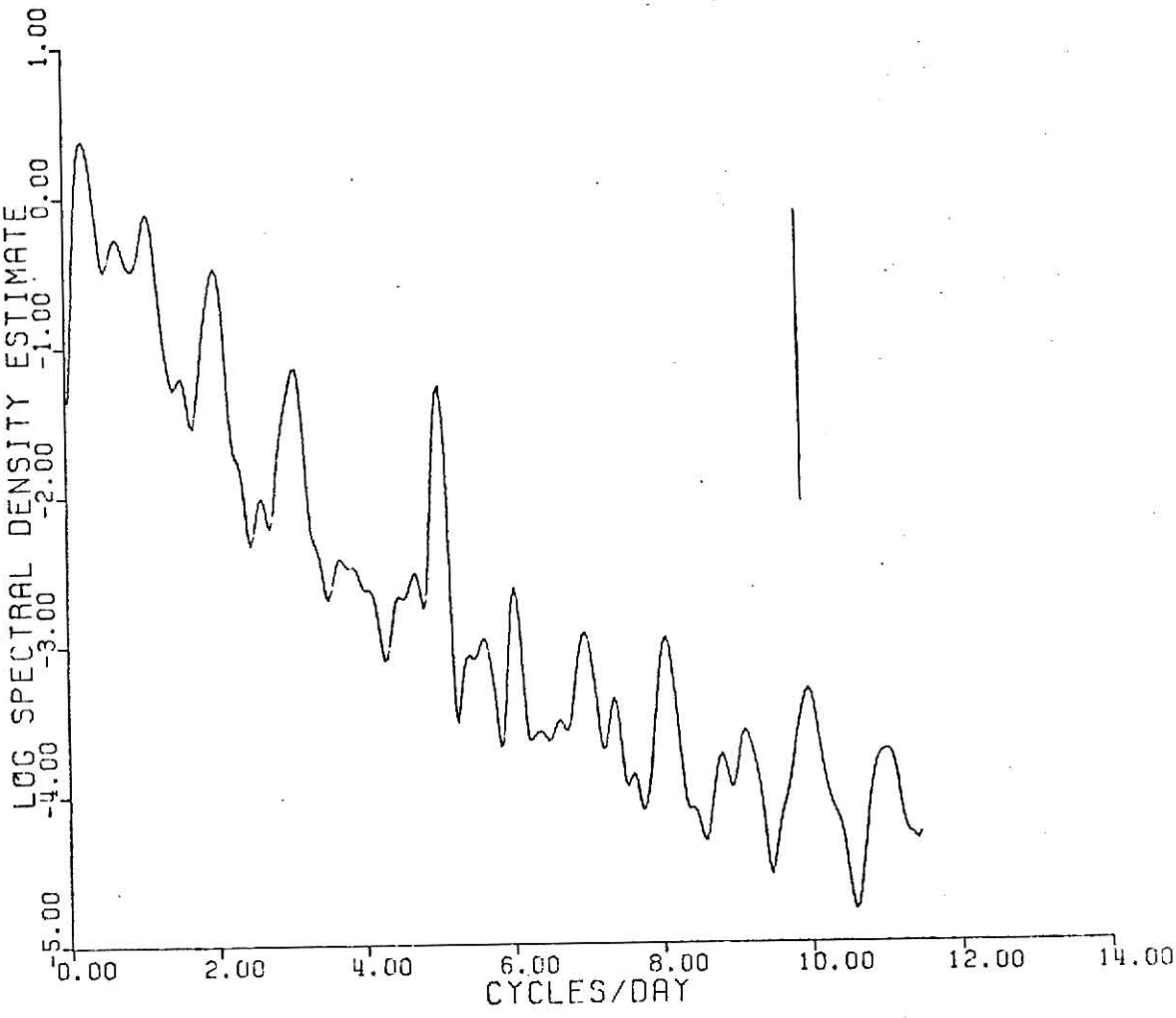

FIGURE 21. Spectrum of residuals for hourly carbon monoxide data from San Jose.

$$Y = .0037t - .00004t^2 + 1.0210 \sin(2\pi t/1440)$$
$$+ .9353 \cos(2\pi t/720) - .2949 \sin(2\pi t/300) - .6670 \sin(2\pi t/240)$$
$$+ .8488 \cos(2\pi t/240) - .4490 \sin(2\pi t/12) + .5329 \cos(2\pi t/8)$$

with $R^2 = .6131$

The spectral plot of the residuals is shown in Fig. 20. The predictive model for carbon monoxide from San Jose is

$$Y = .0017t - .5228 \cos(2\pi t/1440) + 1.1310 \sin(2\pi t/1440)$$
$$+ .8317 \cos(2\pi t/720) + .4119 \cos(2\pi t/360) - .2860 \sin(2\pi t/360)$$
$$- .4902 \sin(2\pi t/240) + .6177 \cos(2\pi t/180) + .04226 \sin(2\pi t/180)$$
$$- .3380 \sin(2\pi t/24) - .7437 \sin(2\pi t/12) + .4262 \cos(2\pi t/8)$$

with $R^2 = .6137$

The spectral plot of the residuals is shown in Fig. 21.

CHAPTER 4

ANALYSIS OF AIR POLLUTION DATA FROM LOS ANGELES COUNTY, CALIFORNIA

INTRODUCTION

In this chapter another illustration of the application of the procedure for analyzing the air pollution data, discussed in Chapter 3, is given. The data were furnished by the Air Pollution control District, county of Los Angeles, California [45]. The pollutant concentration data of four pollutants from four stations were analyzed. The four pollutants are sulphur dioxide, carbon monoxide, nitrogen oxides and ozone and the four stations are 60, 79, 1 and 71. The location of the stations is presented in Fig. 1. The interval of observations is one hour and the length of record is one year namely from January - December 1971. The missing data were linearly interpolated. The results are presented in detail for sulphur dioxide data from all the four stations while for other pollutants the results are summarized and the graphical representations attached in Appendix A. The missing data are listed in Tables P, Q, R and S in Appendix A.

Description of the Data

The plots of the daily averaged sulphur dioxide concentration from these stations is shown in Figs. 2, 3, 4 and 5. The average amount of sulphur dioxide at station 60 appears to have maximum concentration sometime during the middle of the year and minimum during the first and last quarter of the year. The average amount at station 79 appears to have maximum during the middle of the first and last quarter of the year

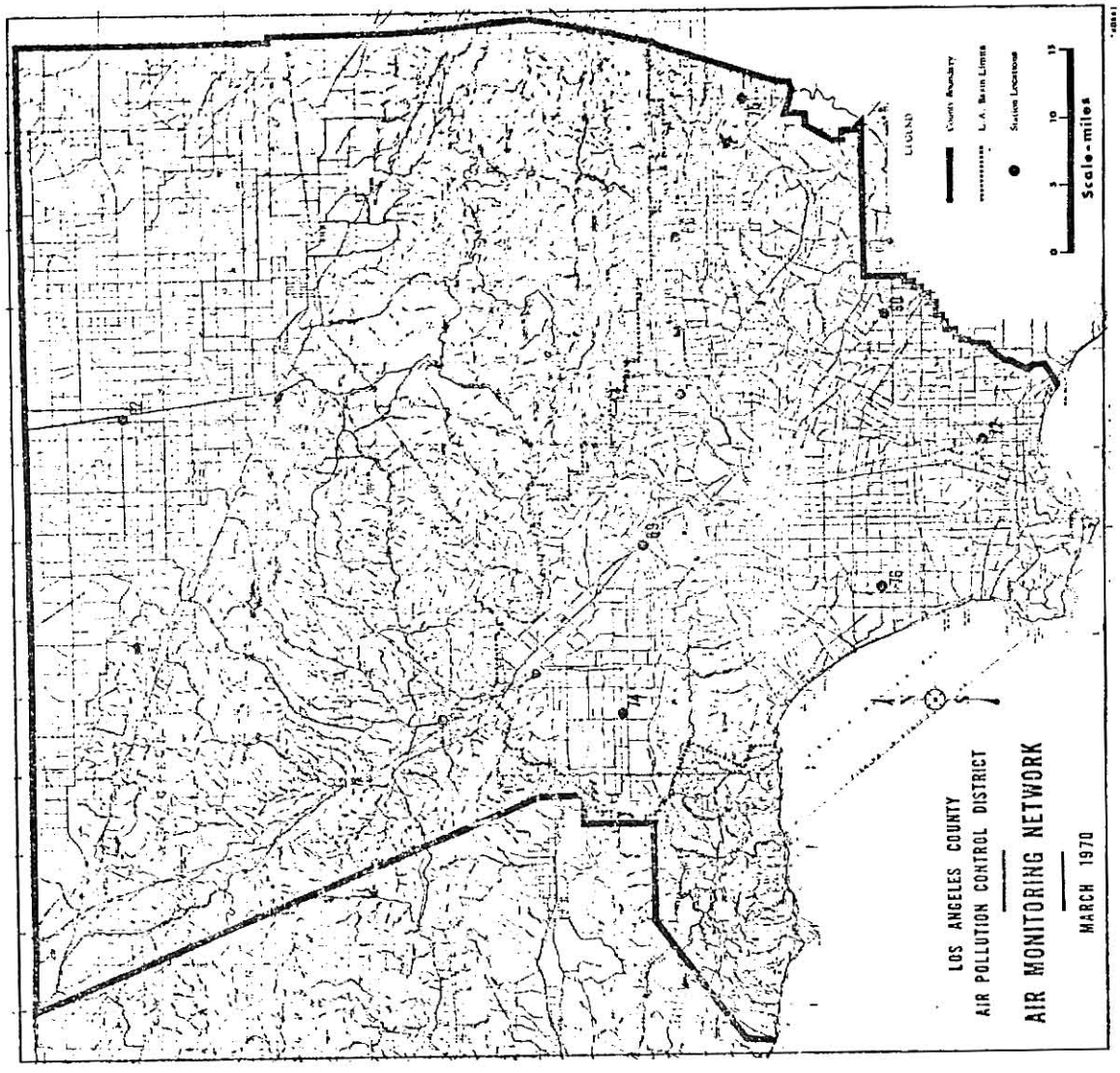

FIGURE 1. Location of stations

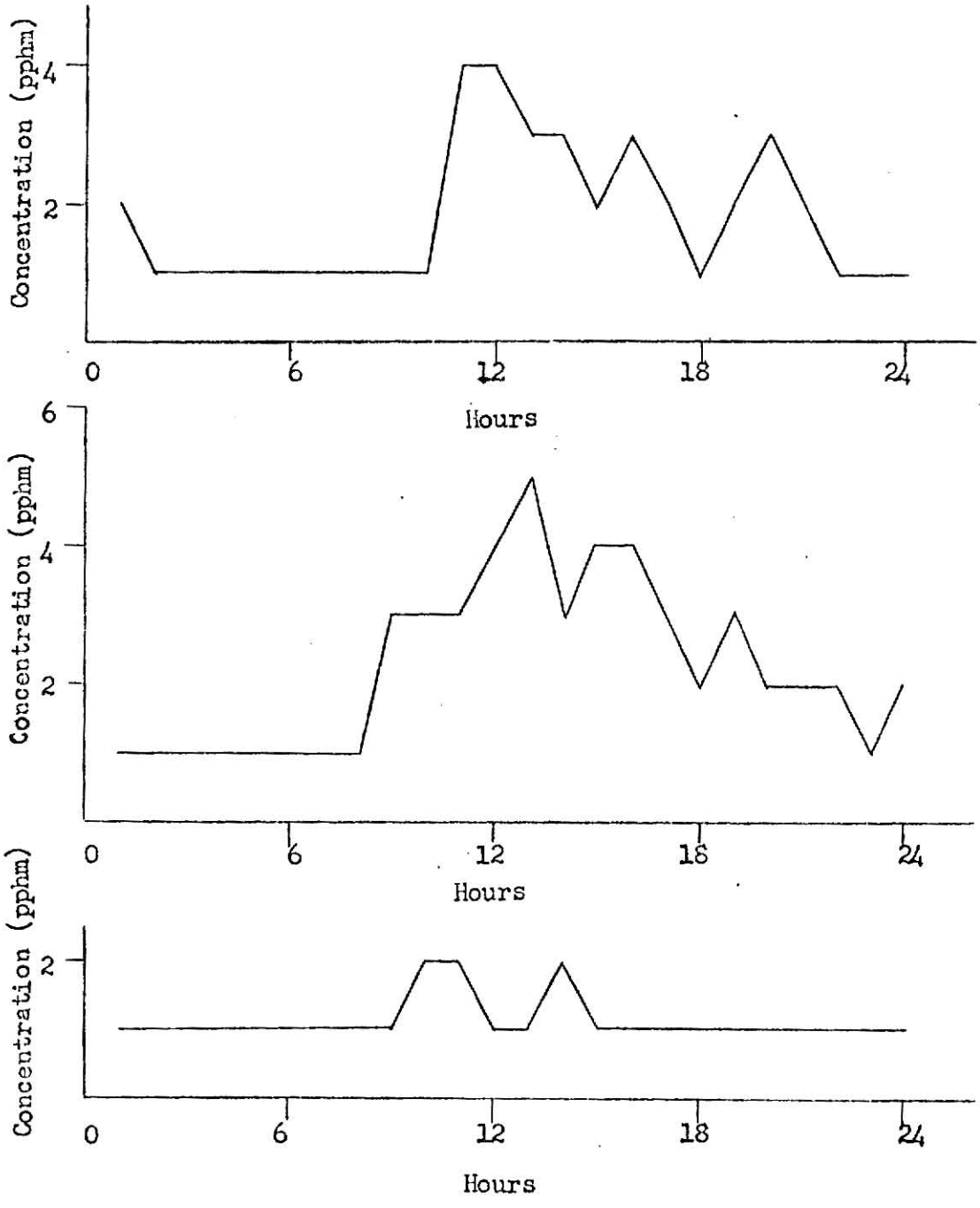

FIGURE 1A. Three typical variations in sulphur dioxide concentration during a day at station 60.

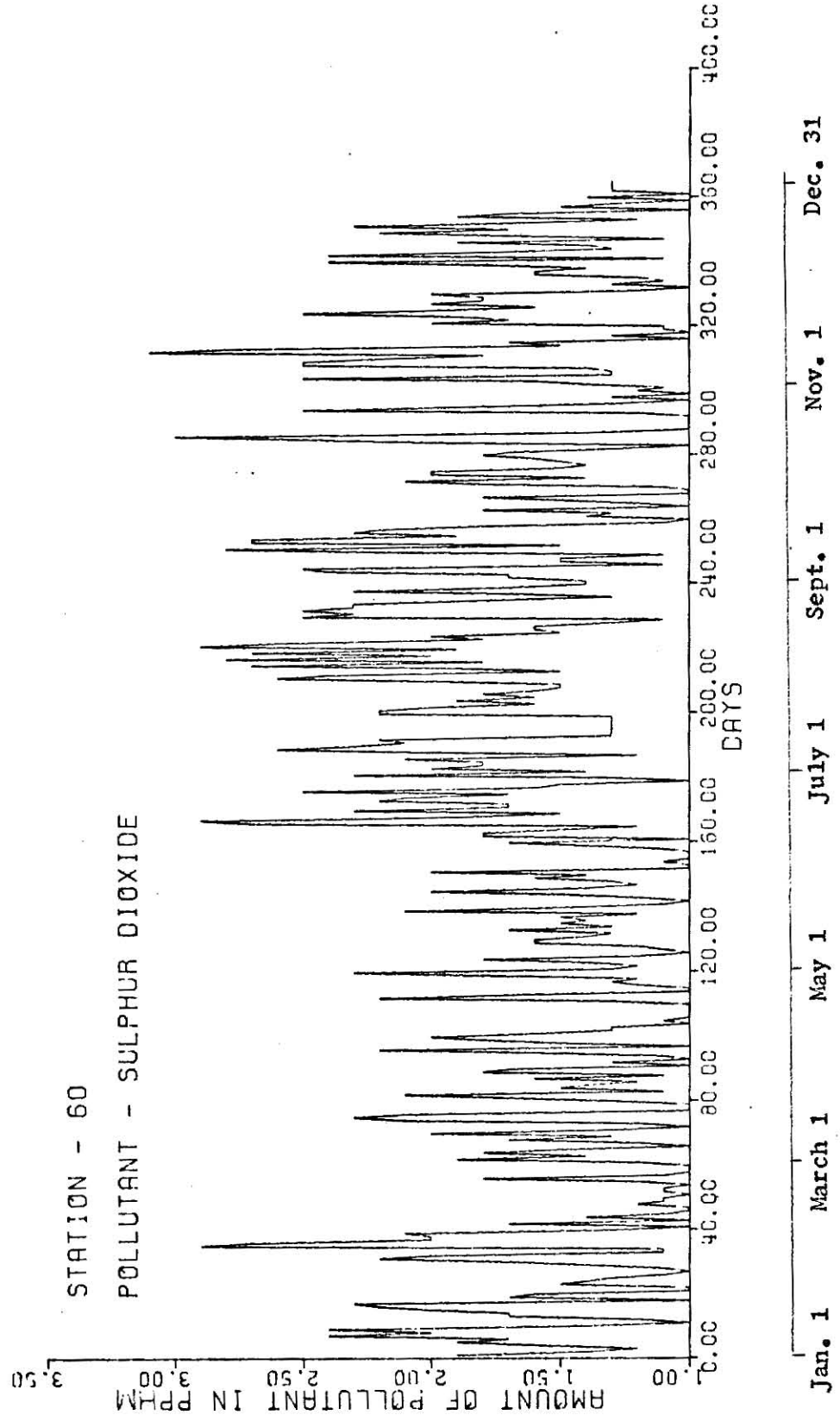

FIGURE 2. Daily averaged sulphur dioxide data from station 60.

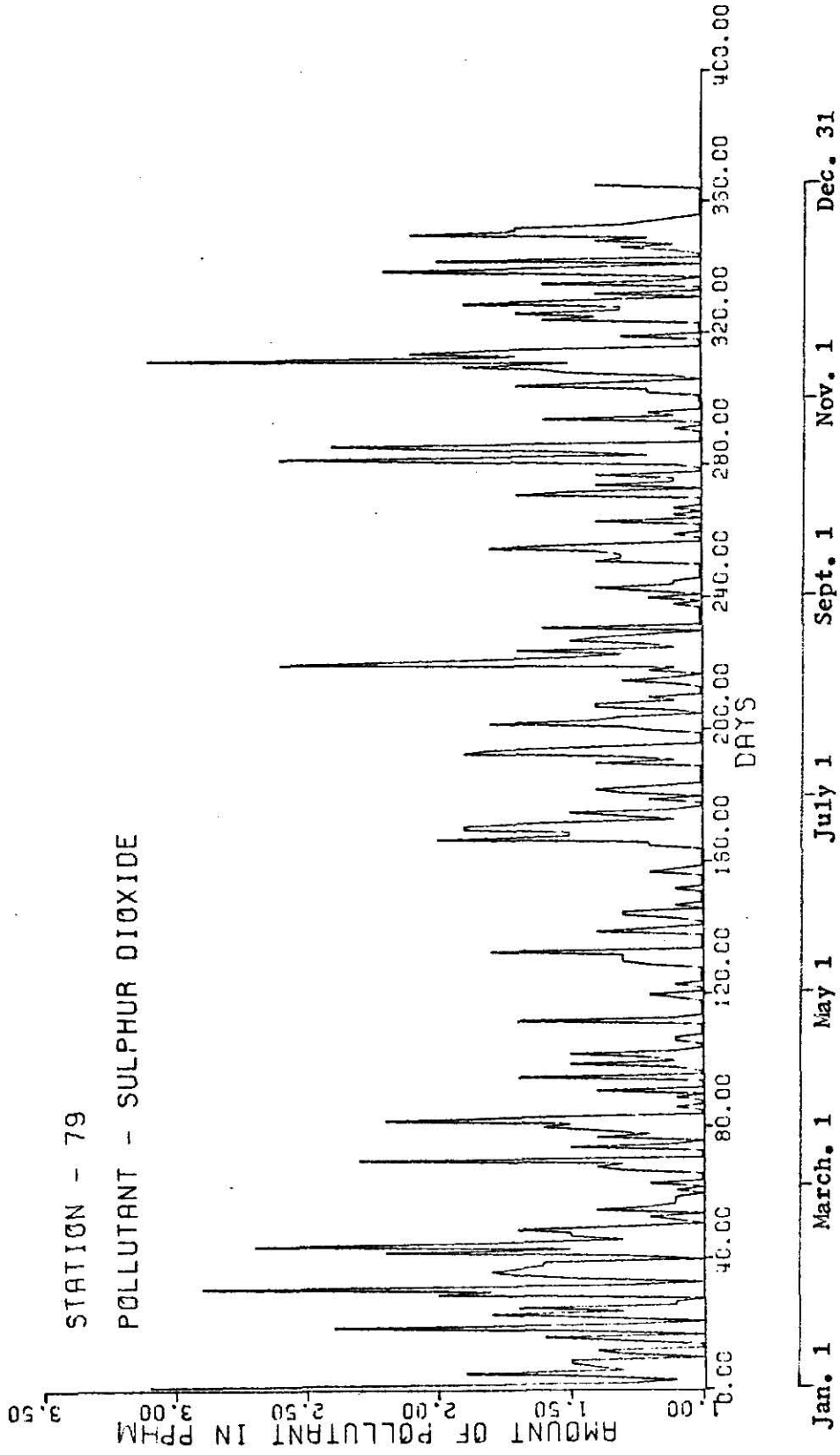

FIGURE 3. Daily averaged sulphur dioxide data from station 79

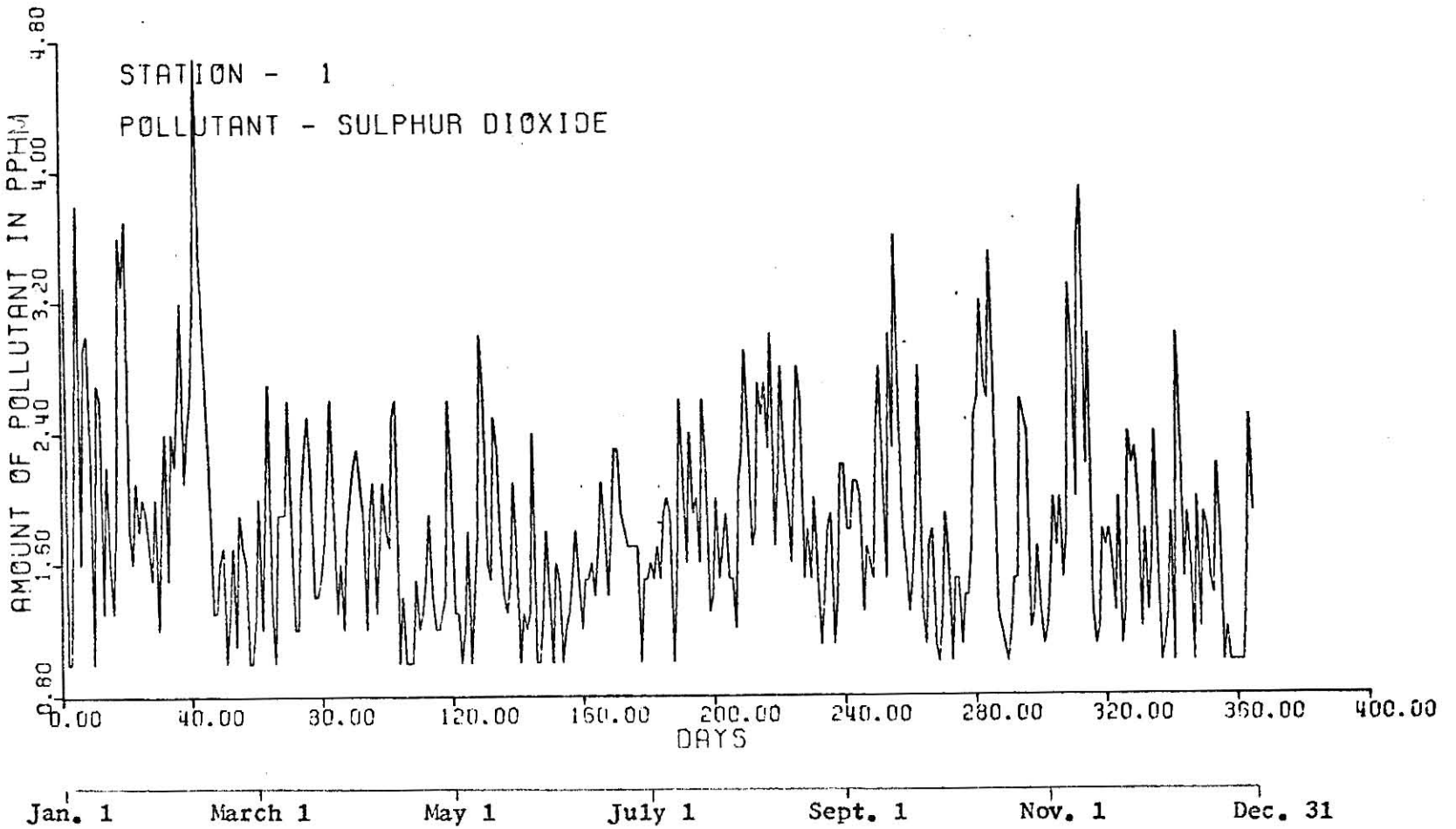

FIGURE 4. Daily averaged sulphur dioxide data from station 1.

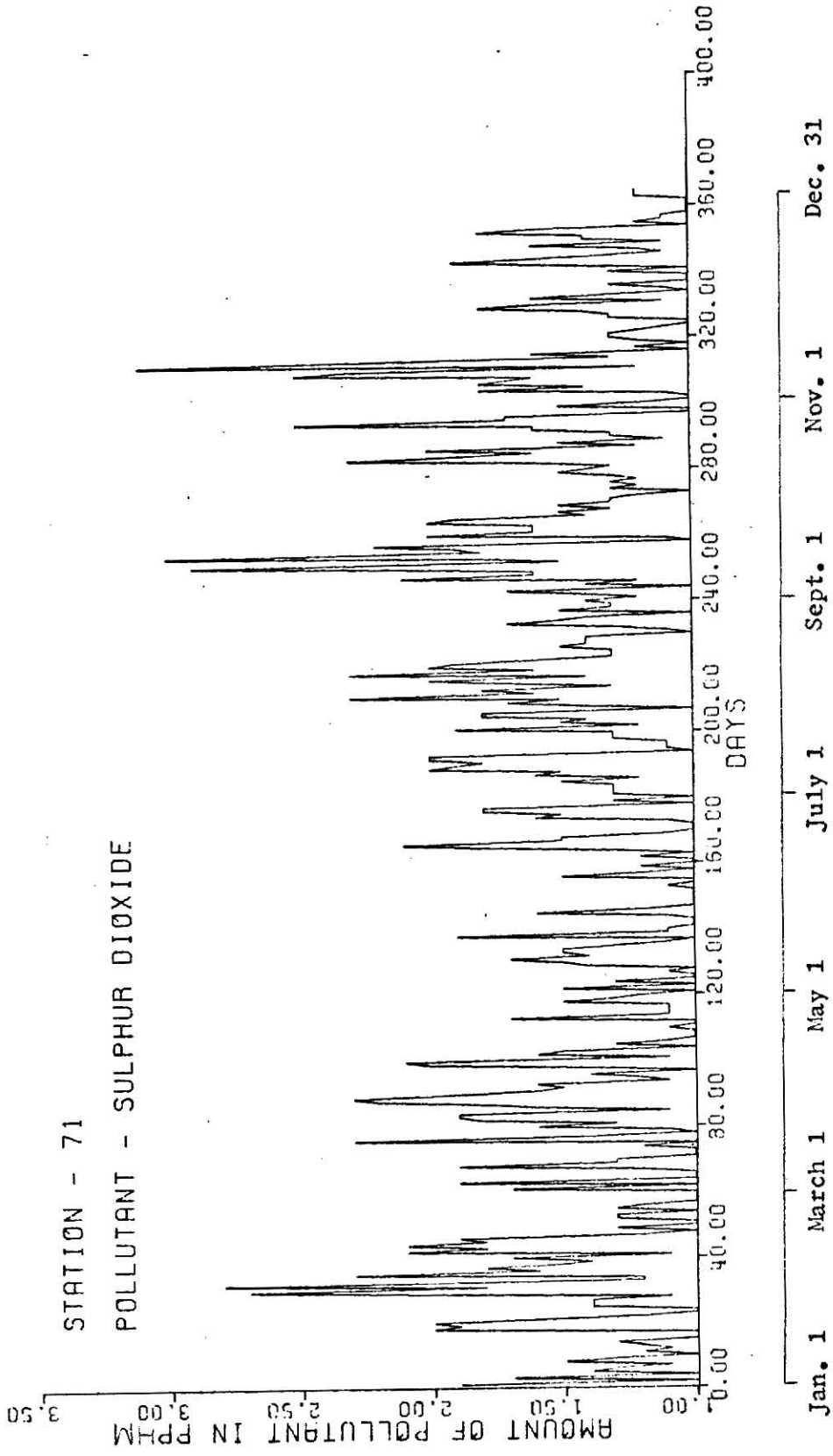

FIGURE 5. Daily averaged sulphur dioxide data from station 71.

while it is minimum sometime in the middle of the year. Basically, if looked minutely, the emission patterns at these two stations and also at station 71 appear to be similar with the one at station 60 having an additional bump around the middle of the year which may be because of an additional short term industrial process or a social activity. The average amount of sulphur dioxide at station 71 has an increasing trend till the middle of the second month from where it starts decreasing attaining the minimum sometime in the middle of the year. It again has an increasing trend till about the tenth month where it starts decreasing. The average amount of sulphur dioxide present at station 1 in the first month and in the tenth month is higher than that emitted during any other month. Superimposed on this there is a cyclic variation of amount of sulphur dioxide at station 1 which can be seen varying from second to tenth month. Overall it is seen that the sulphur dioxide concentration at station 60 fluctuates the most and that a station 1 is the most consistant throughout the year to that at any other station: The relative magnitude of emission is also highest at station 1 and lowest at station 79. The lowest average sulphur dioxide concentration at any station is one pphm while the highest is that at station 1 being 4.8 pphm. in the beginning of the second month. Some typical variations in the concentration of sulphur dioxide during a day are shown in Fig. 1A. It is seen that the concentration increases suddenly in the morning hours attaining maximum some time around 12 noon and then decreases slowly.

Fourier Analysis

The harmonic analysis was carried out for 364 points corresponding

to the daily averaged sulphur dioxide concentrations. The analysis was done for 182 harmonics. The important harmonics for the data from all the stations are presented in Tables 1 through 4. The fraction variance taken by the zeroth harmonic determines the fraction taken by the mean term. The fraction variance for the rest of the harmonics is determined using the variance corrected for the mean given by Eqn. 9 in Chapter 3. For the sulphur dioxide data from station 60, out of the total variance of 2.78, about 91% is taken by the mean. The important harmonics presented in Table 1 account for an additional variance of about 3.5%. The remaining 5.5% is distributed among other harmonics. The 1st, 2nd, 12th and 16th harmonics appear to be the most prominent contributors to the total variance corrected for mean. Other major contributors are the 24th, 27th and 52nd harmonics. For the sulphur dioxide data from station 79 the mean contributes 92% to the total variance of 1.74. The important harmonics presented in Table 2 contribute an additional of about 2.5%. The rest 5.5% is distributed among other harmonics. The 1st, 12th, and 13th appear to be the most prominent contributors to the total variance corrected for mean. Other major contributors are the 2nd, 26th, 27th, 36th and 58th harmonics. At station 1 most of the harmonics presented in Table 3 have quite a significant contribution to the total variance corrected for mean. Mean contributes about 89% to the total variance 3.70. Additional 4% is contributed by the harmonics presented in the table. Both the 2nd and 12th harmonics are major contributors revealing a monthly and half yearly variation is the sulphur dioxide concentration. Other prominent harmonics are 4th, 20th, and

Table 1

Harmonic Analysis of Sulphur Dioxide Data From Station 60

Harmonic	C_1	C_2	Amplitude	Phase	Cont. to Var.	Frac. Var.
0	1.591	0.0	1.591	0.0	2.533	0.9109
1	0.045	0.092	0.103	63.82	0.021	0.0862
2	0.049	0.040	0.063	38.93	0.008	0.0329
3	-0.002	0.040	0.040	93.26	0.003	0.0134
8	0.006	-0.046.	0.047	277.53	0.003	0.0126
12	0.003	0.067	0.067	86.70	0.009	0.0370
13	-0.022	-0.027	0.035	230.73	0.002	0.0101
14	-0.037	0.012	0.039	161.79	0.003	0.0127
16	-0.033	0.051	0.061	123.49	0.007	0.0300
20	0.046	-0.007	0.046	350.98	0.004	0.0177
21	0.017	0.043	0.046	68.51	0.004	0.0175
24	-0.034	0.033	0.047	90.27	0.004	0.0182
26	0.0002	0.043	0.043	17.04	0.003	0.0149
27	0.047	0.014	0.049	23.27	0.004	0.0195
34	-0.035	-0.015	0.038	203.27	0.003	0.0120
35	0.015	-0.041	0.044	289.82	0.003	0.0158
38	-0.040	-0.015	0.043	200.93	0.003	0.0149
52	0.022	-0.041	0.046	298.29	0.004	0.0179
80	-0.041	0.008	0.042	168.75	0.003	0.0146
97	-0.042	-0.009	0.043	192.73	0.003	0.0152

Table 2

Harmonic Analysis of Sulphur Dioxide Data From Station 79

Harmonic	C_1	C_2	Amplitude	Phase	Cont. to Var.	Frac. Var.
0	1.267	0.0	1.267	0.0	1.607	0.9234
1	-0.056	0.012	0.057	167.09	0.006	0.0498
2	0.028	0.022	0.035	38.66	0.002	0.0193
3	0.027	-0.008	0.029	342.59	0.001	0.0127
8	0.016	-0.033	0.037	296.08	0.002	0.0214
12	-0.028	0.046	0.055	121.78	0.006	0.0453
13	-0.044	-0.003	0.044	184.49	0.004	0.0302
20	0.010	0.029	0.031	70.85	0.001	0.0145
25	0.009	-0.028	0.029	288.64	0.001	0.0131
26	0.030	0.034	0.043	44.68	0.003	0.0280
27	-0.033	-0.027	0.043	219.29	0.003	0.0277
28	0.028	-0.002	0.028	355.15	0.001	0.0122
29	0.021	-0.015	0.026	323.96	0.001	0.0103
30	0.005	-0.031	0.031	279.52	0.002	0.0148
36	0.030	-0.031	0.043	313.80	0.003	0.0283
53	-0.023	-0.026	0.031	221.42	0.001	0.0145
57	0.023	0.017	0.029	35.55	0.001	0.0130
58	0.014	-0.037	0.040	291.77	0.003	0.0242
83	-0.025	0.018	0.031	144.27	0.001	0.0146
84	0.003	-0.025	0.025	277.40	0.001	0.0101
96	0.029	-0.013	0.032	334.46	0.001	0.0155
100	0.028	0.003	0.028	6.35	0.002	0.0123
101	0.031	-0.001	0.031	357.41	0.001	0.0145

Table 3

Harmonic Analysis of Sulphur Dioxide Data from Station 1

Harmonic	C_1	C_2	Amplitude	Phase	Cont. to Var.	Frac. Var.
0	1.818	0.0	1.818	0.0	3.308	0.8935
1	-0.044	0.038	0.058	319.67	0.006	0.0176
2	0.017	0.090	0.091	79.09	0.016	0.0427
4	-0.026	0.070	0.075	120.70	0.011	0.0286
9	-0.035	-0.052	0.063	235.61	0.008	0.0202
12	-0.033	0.139	0.143	103.55	0.041	0.1046
15	0.013	0.059	0.059	76.86	0.007	0.0177
19	-0.052	-0.032	0.061	211.32	0.007	0.0193
20	0.030	0.065	0.072	64.88	0.010	0.0266
24	-0.056	-0.021	0.060	200.76	0.007	0.0187
26	0.044	0.075	0.087	59.41	0.015	0.0386
36	0.053	-0.033	0.063	327.85	0.008	0.0204
52	-0.014	-0.062	0.064	257.22	0.008	0.0208
79	-0.037	0.060	0.078	121.91	.010	.0254

Table 4

Harmonic Analysis of Sulphur Dioxide Data from Station 71

Harmonic	C_1	C_2	Amplitude	Phase	Cont. to Var.	Frac. Var.
0	1.405	0.0	1.405	0.0	1.975	.9268
1	0.001	0.050	0.050	87.93	0.005	.0330
2	-0.028	0.055	0.062	117.35	0.007	.0497
4	0.006	0.031	0.032	77.47	0.002	.0132
5	0.028	-0.027	0.039	315.72	0.003	.0192
6	-0.025	0.022	0.033	138.48	0.002	.0146
7	0.007	0.049	0.049	81.65	0.005	.0319
8	-0.004	-0.061	0.061	265.86	0.007	.0484
12	0.002	0.048	0.048	86.75	0.004	.0297
18	-0.002	0.037	0.037	93.21	0.002	.0178
19	-0.025	-0.023	0.034	223.47	0.002	.0152
24	-0.033	-0.002	0.033	183.46	0.002	.0142
25	-0.039	-0.034	0.052	221.21	0.005	.0350
26	0.016	0.025	0.030	50.59	0.001	.0116
32	-0.008	-0.028	0.029	252.58	0.001	.0113
33	-0.034	0.030	0.046	138.15	0.004	.0273
40	0.032	0.010	0.033	17.57	0.002	.0145
50	0.030	0.001	0.031	3.59	0.001	.0123
58	0.011	-0.029	0.031	290.55	0.002	.0127
151	-0.031	0.003	0.032	174.09	0.002	.0132

and 26th. At station 71 the mean contributes about 92.6% to the total variance of 2.13. The harmonics presented in Table 4 contribute an additional of about 3% with about 5.4% distributed among other harmonics. The prominent harmonics are the 1st, 2nd, 7th, 8th, 12th, 25th and 33rd.

Fuel combustion and chemical plants among others are the major contributors to the sulphur dioxide concentration. In coal and oil combustion the amount of sulphur dioxide emitted increases linearly with their sulphur content. Power plants, home heating units, automobiles, buses and trucks and community and industrial incineration can be classified under the fuel combustion causing sulphur dioxide concentration. Petroleum refineries, pulp mills, super phosphate fertilizer plants can be classified under the chemical plants causing concentration of sulphur dioxide.

Spectral Analysis

Hourly observations for the whole of year form too big a series to be analyzed. To save computational time it would be very useful if we could determine the highest periodictly present in the data and then to filter the data accordingly. The harmonic analysis described above determined frequency upto a period of two days and it is seen that the higher harmonics are not very significant. However frequencies corresponding to 24 hours or less can be expected because of the daily traffic of office and factory goers, transportation of goods by trucks and the daily operation of power plants. It was thus first intended to determine the highest periodicity present. Spectral analysis was carried out for the hourly observations of the two months period namely, March and April,

at station 1 to determine the daily or fractional day frequencies. The analysis revealed 24 hours and 12 hours frequencies. Moreover a frequency for a period less than 8 hours is not expected. Hence it was decided to use a 6 hour average data. Thus a new series, x'(t), was formed from the original series, x(t), by the transformation,

$$x'(t') = \frac{1}{6} \sum_{t=i}^{j} x(t) \qquad t' = 1, \ldots, q$$

where $i = 6(t'-1)+1$, $j = 6t'$ and $q = N/6$.

This transformation was used for the sulphur dioxide data from all the stations. The new series thus formed contains 1460 points. The number of lags used were determined by the procedure explained under power spectrum in Chapter 3. The determined value of M was 150.

A linear and second order polynomial model was fitted to the data to remove any linear or nonlinear trend. The autocovariance functions calculated using Eq. 11 in Chapter 3 are presented in Figs. 6-9 for the sulphur dioxide data from all the stations. It is seen that the autocorrelation decreases rapidly within the first few lags and then it fluctuates with almost constant amplitude. The amplitude of oscillation is more in the case of sulphur dioxide from station 60 than in any other.

The spectral density function is obtained using Eqn. 14 in Chapter 3 and the plots are shown in Figs. 10-13. The vertical line on the graphs shows the confidence interval. At all the stations there is a very prominent daily variation of sulphur dioxide. At station 60, Fig. 10,

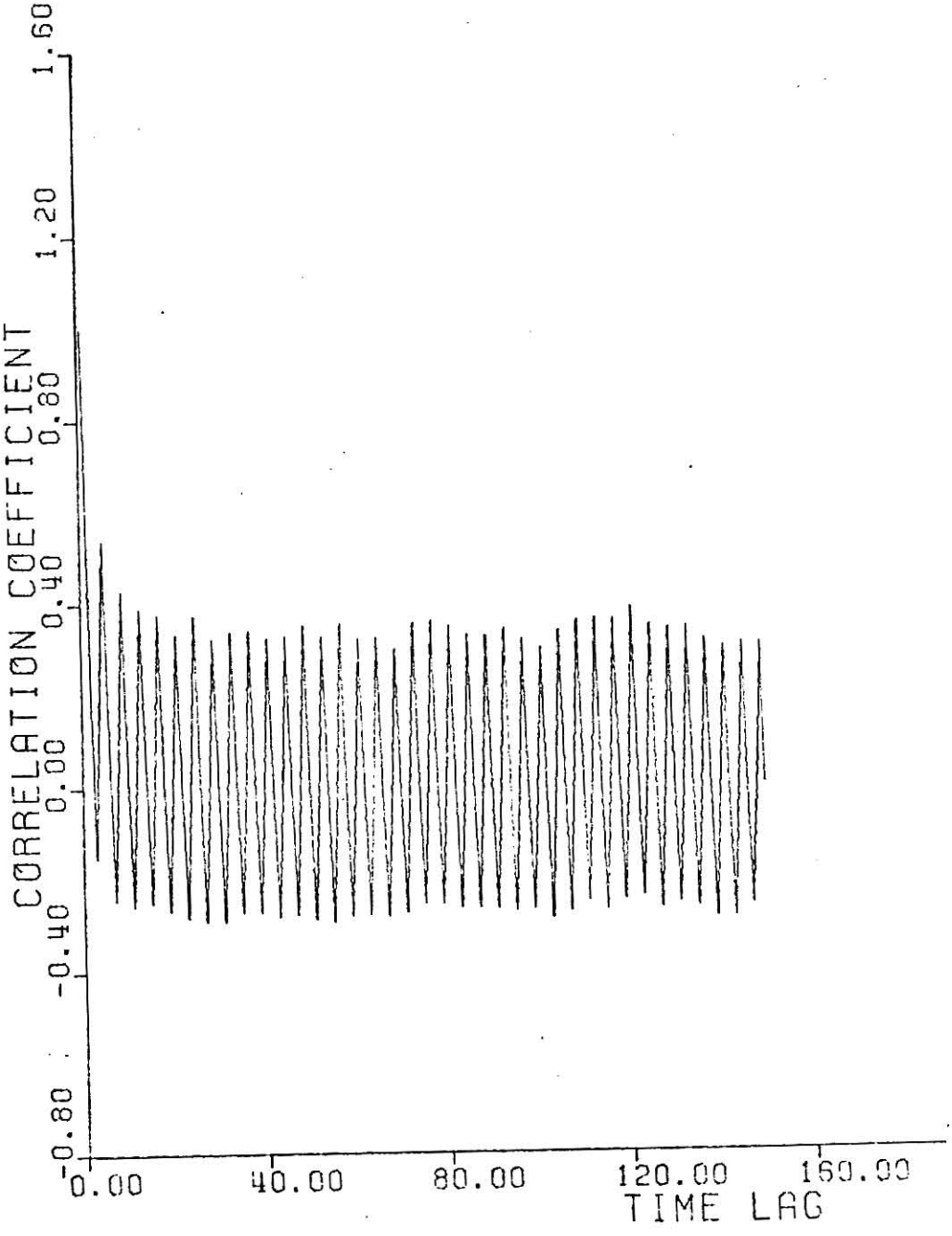

FIGURE 6. Autocorrelation function of sulphur dioxide from station 60.

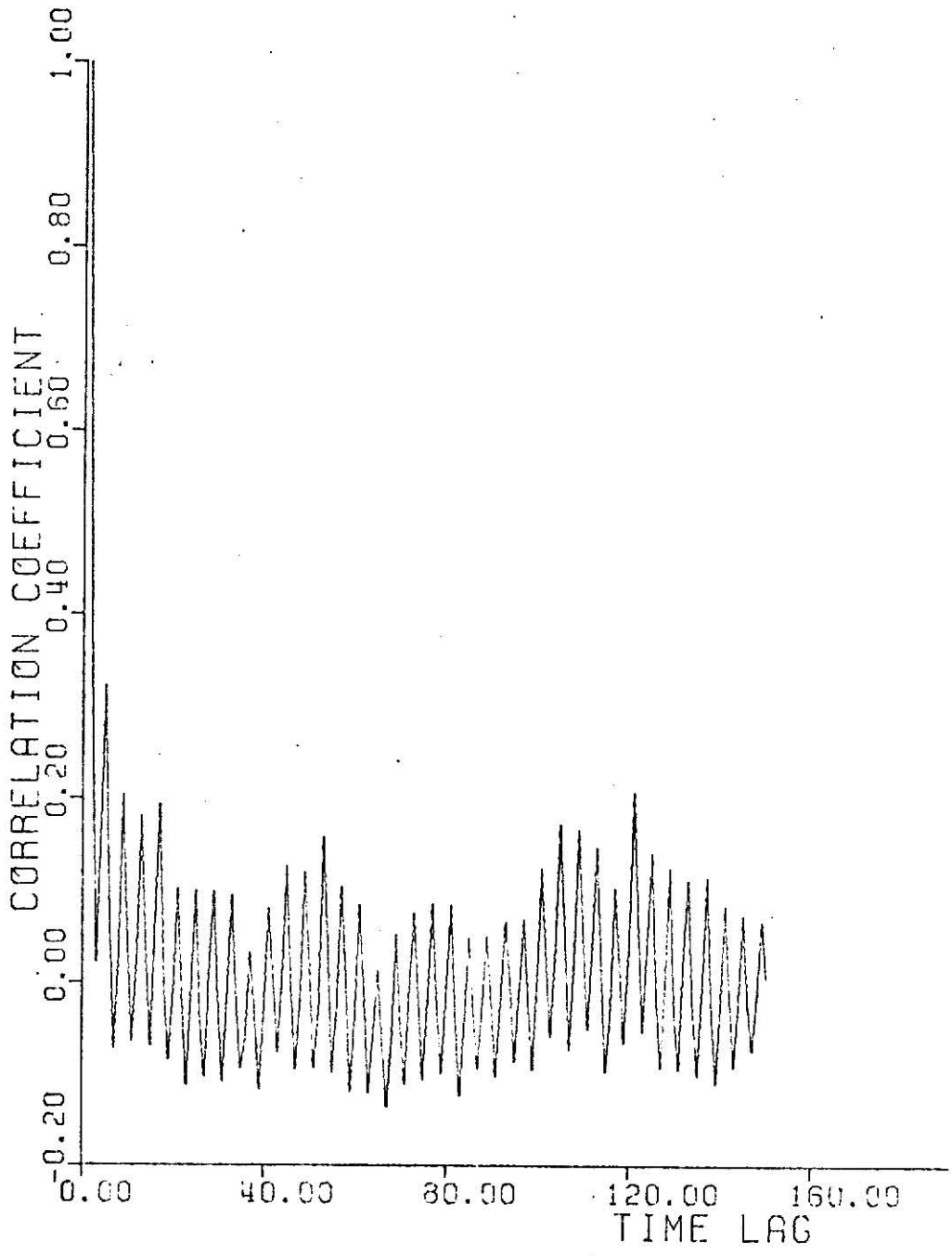

FIGURE 7. Autocorrelation function of sulphur dioxide from station 79.

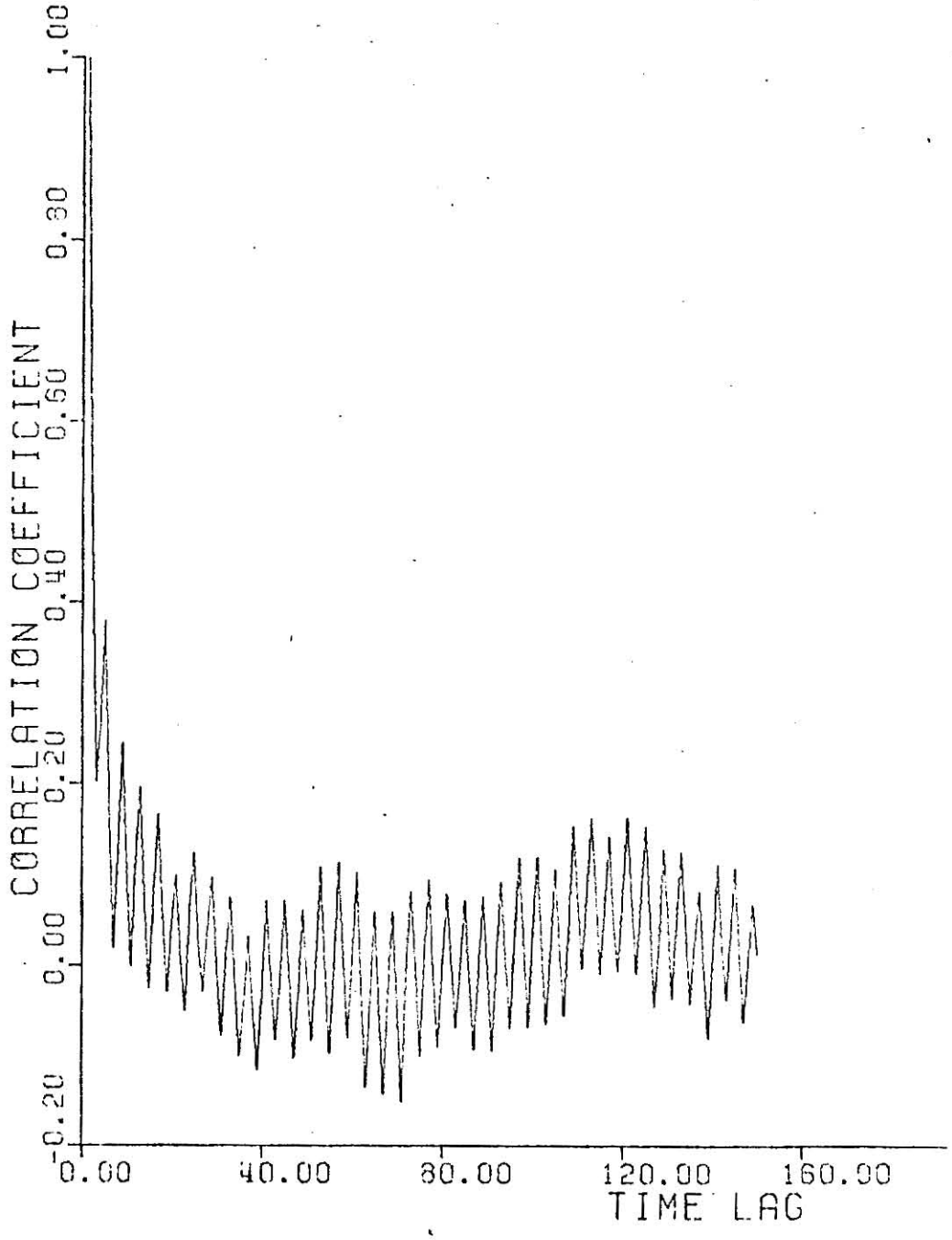

FIGURE 8. Autocorrelation function of sulphur dioxide from station 1.

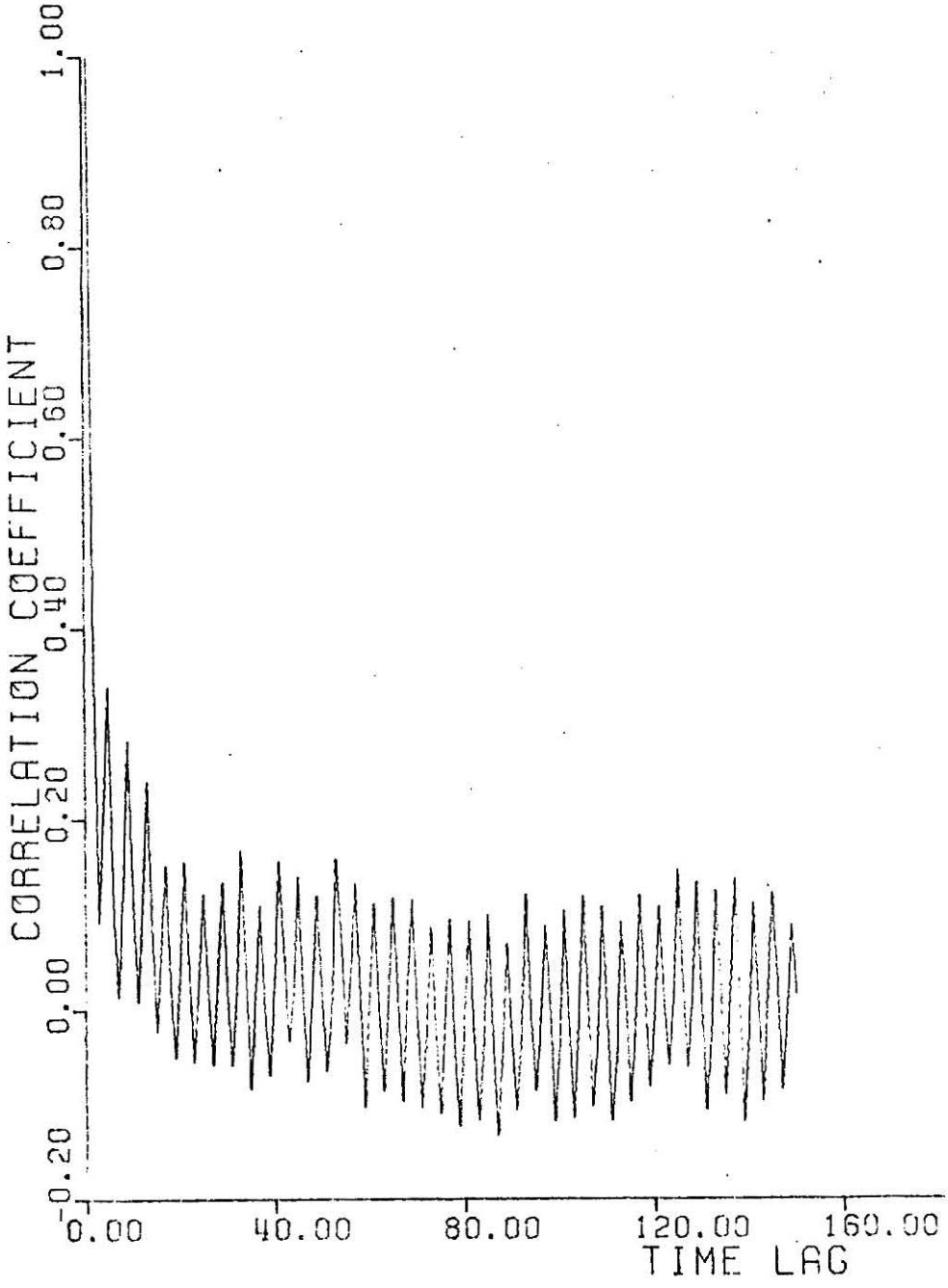

FIGURE 9. Autocorrelation function of sulphur dioxide from station 71.

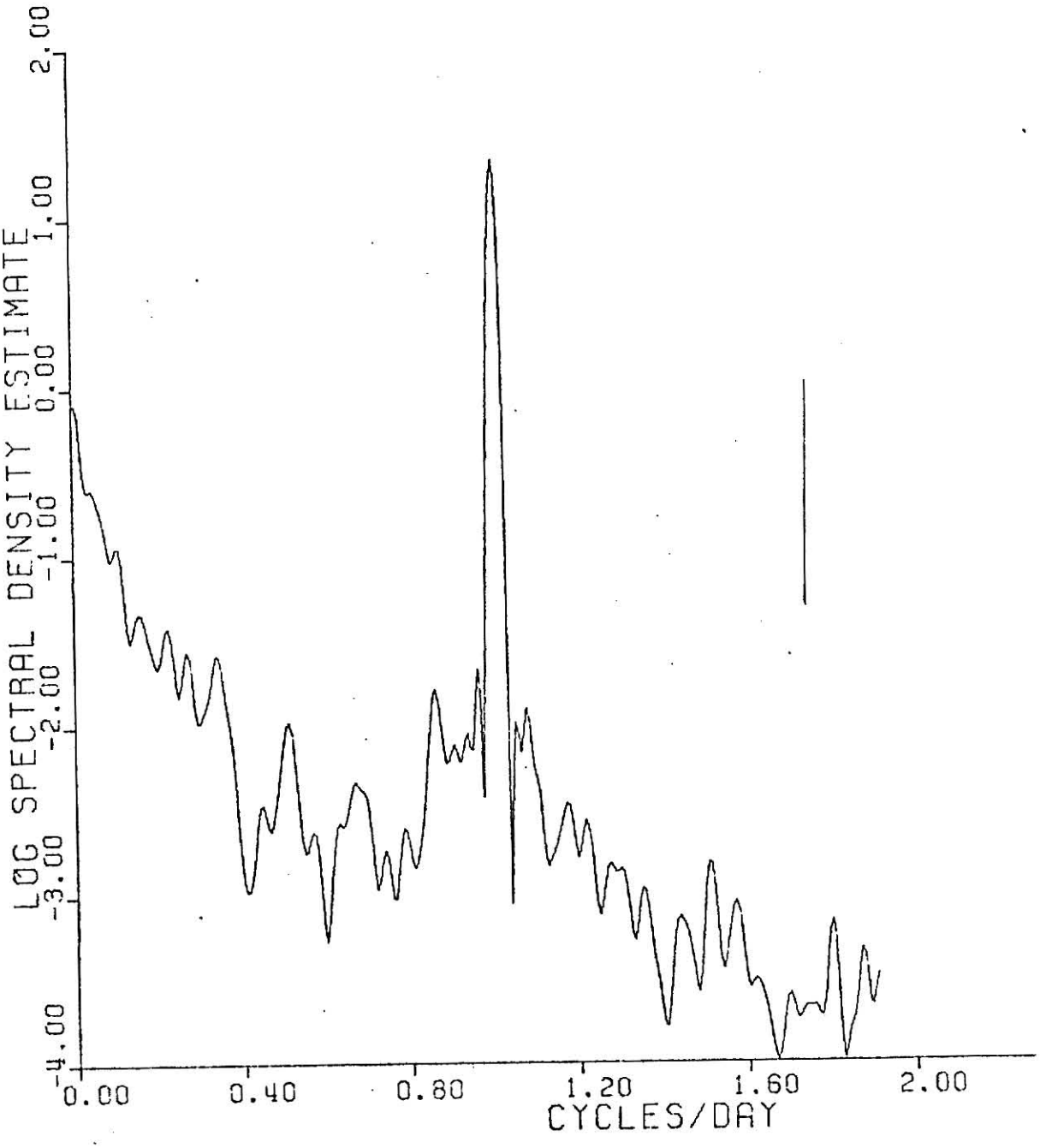

FIGURE 10. Spectral density function for sulphur dioxide from station 60 based on six hour averaged data.

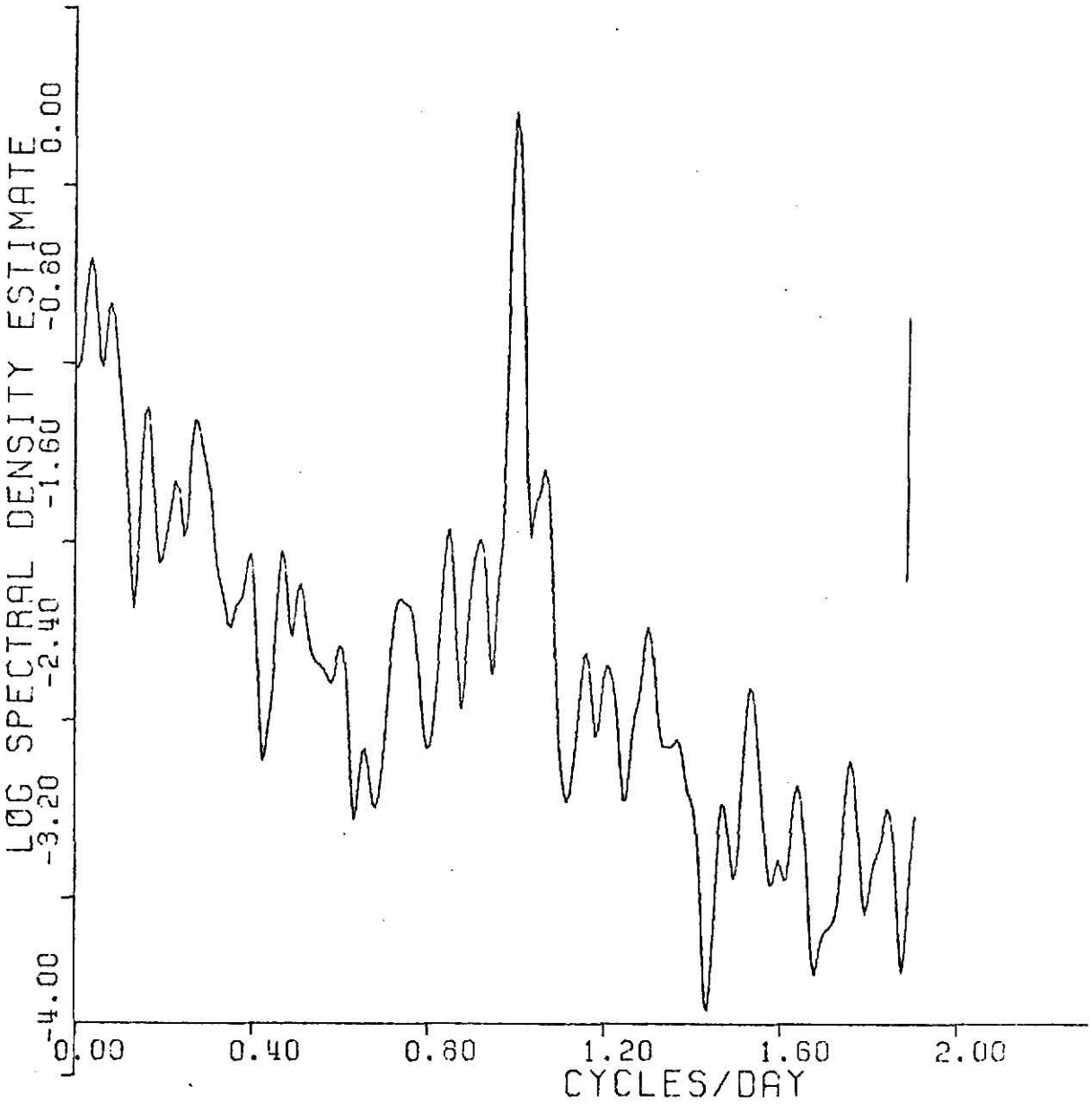

FIGURE 11. Spectral density function for sulphur dioxide from station 79 based on six hour averaged data.

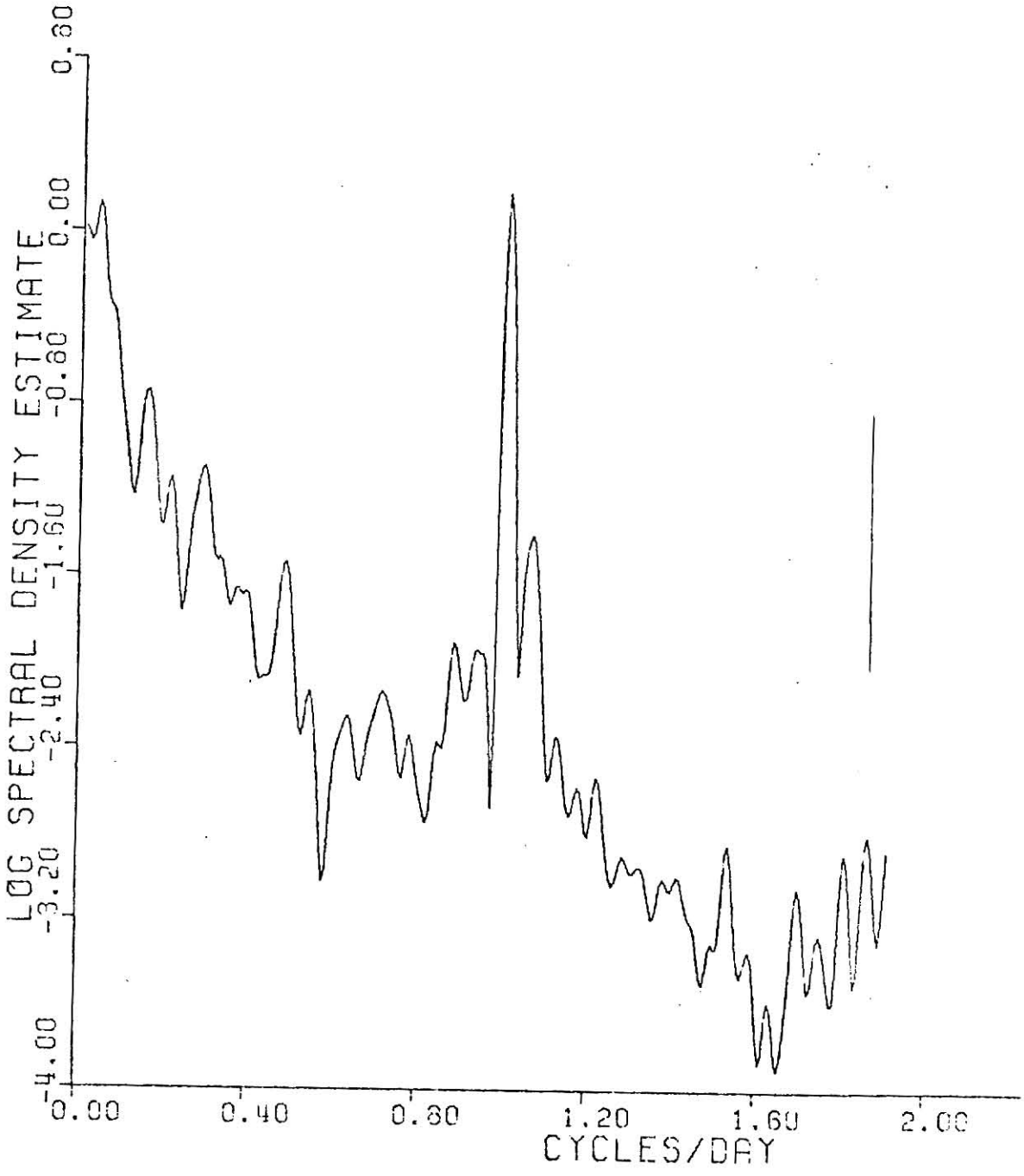

FIGURE 12. Spectral density function for sulphur dioxide from station 1 based on six hour averaged data.

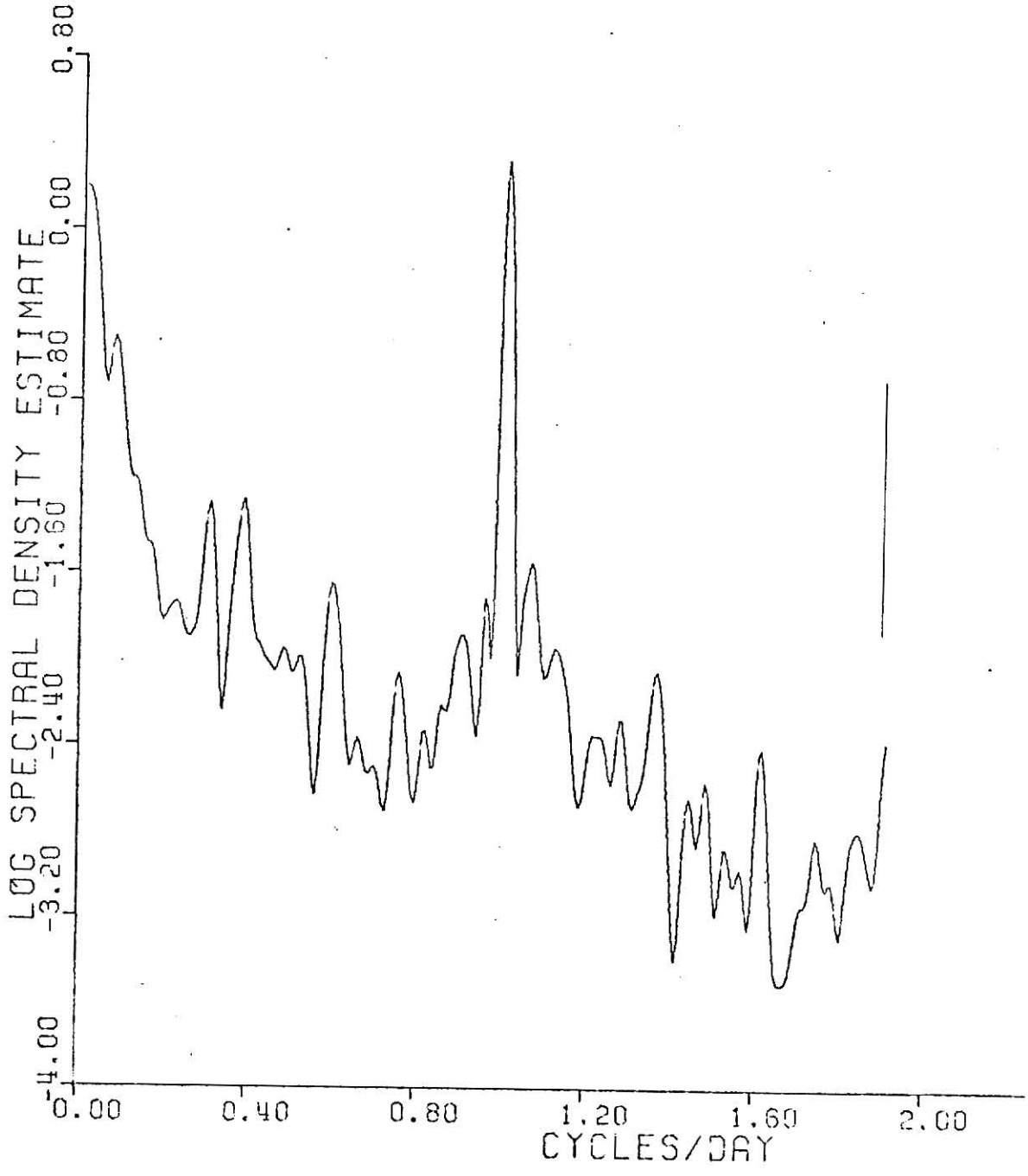

FIGURE 13. Spectral density function for sulphur dioxide from station 71 based on six hour averaged data.

the other prominent cyclic variations of sulphur dioxide are corresponding to the periods of 14 days, 10 day, 5 days, 4 days and three days. There is a large power concentration near the zone frequency. The analysis was further done using 24-hr. average data. This revealed another frequency of 30 days. The spectrals with plots for the 24 hr. averaged data are not shown. This still had a high power concentration around zero frequency. From the fourier analysis, Table 1, this power concentration at low frequencies was considered to be due to yearly, half yearly and 45 days cycles. At station 79, Fig. 11, sulphur dioxide has cyclic variation of 12 to 14 days, 6 days, and 4 days. The 24 hour averaged data also indicated a monthly cycle of 30 days. From fourier analysis the further low frequencies are considered due to yearly, half yearly and four monthly cycles. At station 1, Fig. 12, the cyclic variation sulphur dioxide are corresponding to 14 days, 5 to 7 days and 3 days. Monthly and 18 days frequencies are revealed by 24 hr. average data. From fourier analysis half yearly and yearly and 90 days frequencies seem to be prominent. At station 71, Fig. 13, 15 days and 3 days frequencies appear to be prominent. A 52 hours frequency was observed from the 24-hr data. Additional low frequencies, from the former analysis are yearly, half yearly and 72 days.

Mathematical Model

The frequencies analyzed above for each set of data were used to develop mathematical model by the procedure expalined in chapter three. This model for the sulphur dioxide data from each of the station is given below.

For sulphur dioxide from station 60,

$$Y = -.0001t^2 - 0.6006 \sin(2\pi t/4) - .1832 \sin(2\pi t/1400)$$
$$+ .1054 \sin(2\pi t/120) + .0932 \cos(2\pi t/730)$$
$$- .1041 \sin(2\pi t/180) + .0777 \sin(2\pi t/56)$$
$$+ .0752 \sin(2\pi t/20)$$

$$R^2 = .5832.$$

For the sulphur dioxide data from station 79,

$$Y = .1147 \cos(2\pi t/1400) + .0519 \cos(2\pi t/730)$$
$$+ .0462 \sin(2\pi t/730) - .1096 \cos(2\pi t/120)$$
$$+ .0482 \sin(2\pi t/72) - .0446 \cos(2\pi t/72)$$
$$+ .0810 \sin(2\pi t/56) + .0581 \cos(2\pi t/56)$$
$$+ .0640 \sin(2\pi t/40) - .2806 \sin(2\pi t/4)$$
$$- .0577 \cos(2\pi t/480)$$

$$R^2 = .4230$$

For the sulphur dioxide data from station 1

$$Y = -.2653 \cos (2\pi t/4) - .2048 \cos (2\pi t/120)$$

$$+ .1976 \sin (2\pi t/120) + .1865 \sin (2\pi t/730)$$

$$+ .1691 \sin (2\pi t/56) - .1490 \sin (2\pi t/4)$$

$$+ .1389 \sin (2\pi t/72) - .1235 \sin (2\pi t/20)$$

$$+ .0892 \cos (2\pi t/1400) + .0857 \sin (2\pi t/40)$$

$$R^2 = .4538.$$

For sulphur dioxide data from station 71

$$Y = -.0003t - .2033 \cos (2\pi t/4) - .1315 \sin (2\pi t/180)$$

$$- .1184 \sin (2\pi t/4) + .1152 \sin (2\pi t/780)$$

$$- .0985 \sin (2\pi t/1460) - .0928 \sin (2\pi t/60)$$

$$- .0896 \sin (2\pi t/208) + .0643 \sin (2\pi t/120)$$

$$- .0643 \cos (2\pi t/1400) - .0615 \cos (2\pi t/280)$$

$$R^2 = .4322$$

The plot of the residuals obtained after fitting these models are shown in Figs. 14-17. The heights of the removed peaks have reduced and in some cases it is just outside the upper control limit. This remaining fluctuation appears to be magnified in the graph due to the change of scale.

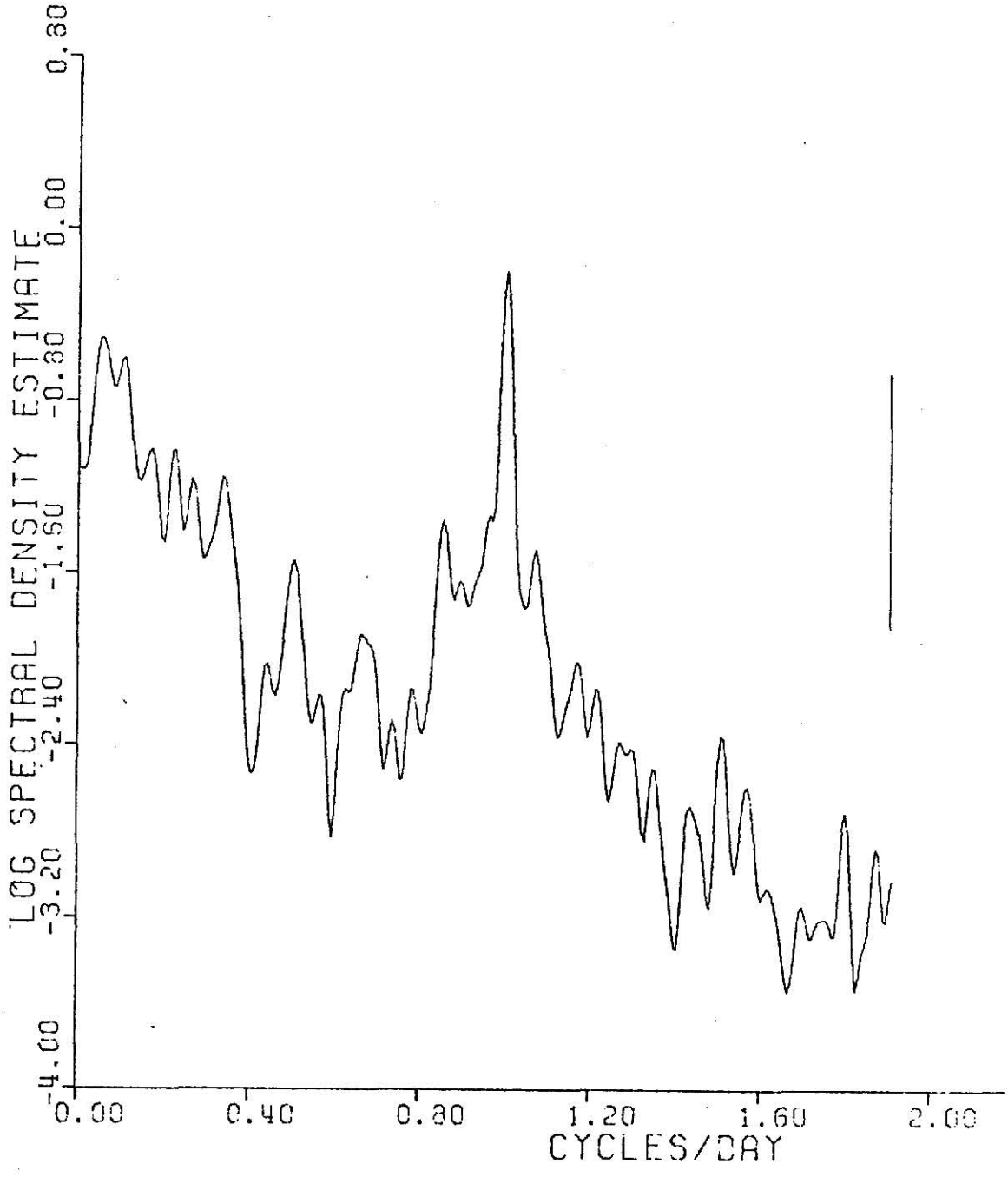

FIGURE 14. Spectral plot of residuals for six hour averaged data from station 60.

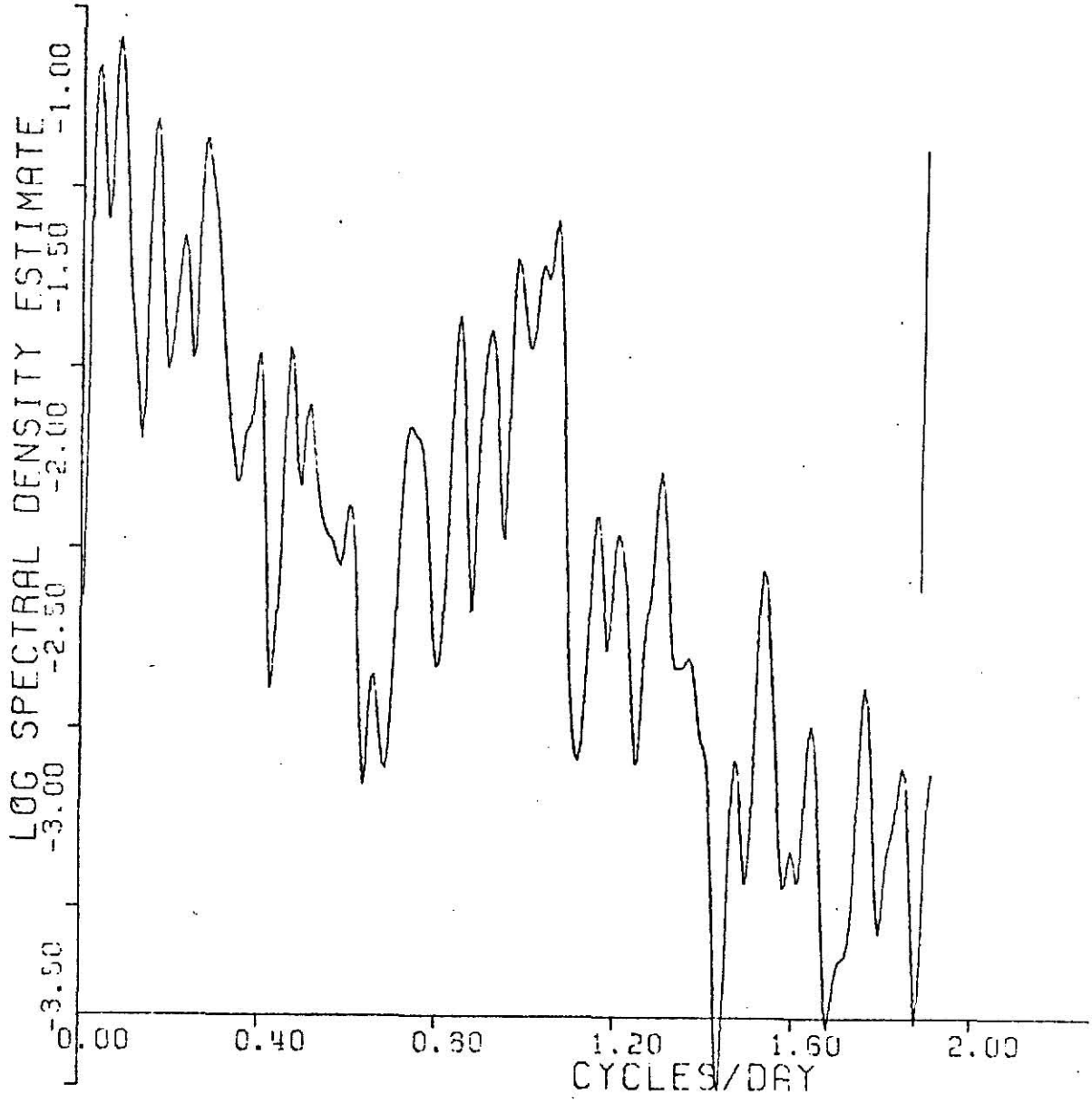

FIGURE 15. Spectral plot of residuals for six hour averaged data from station 79.

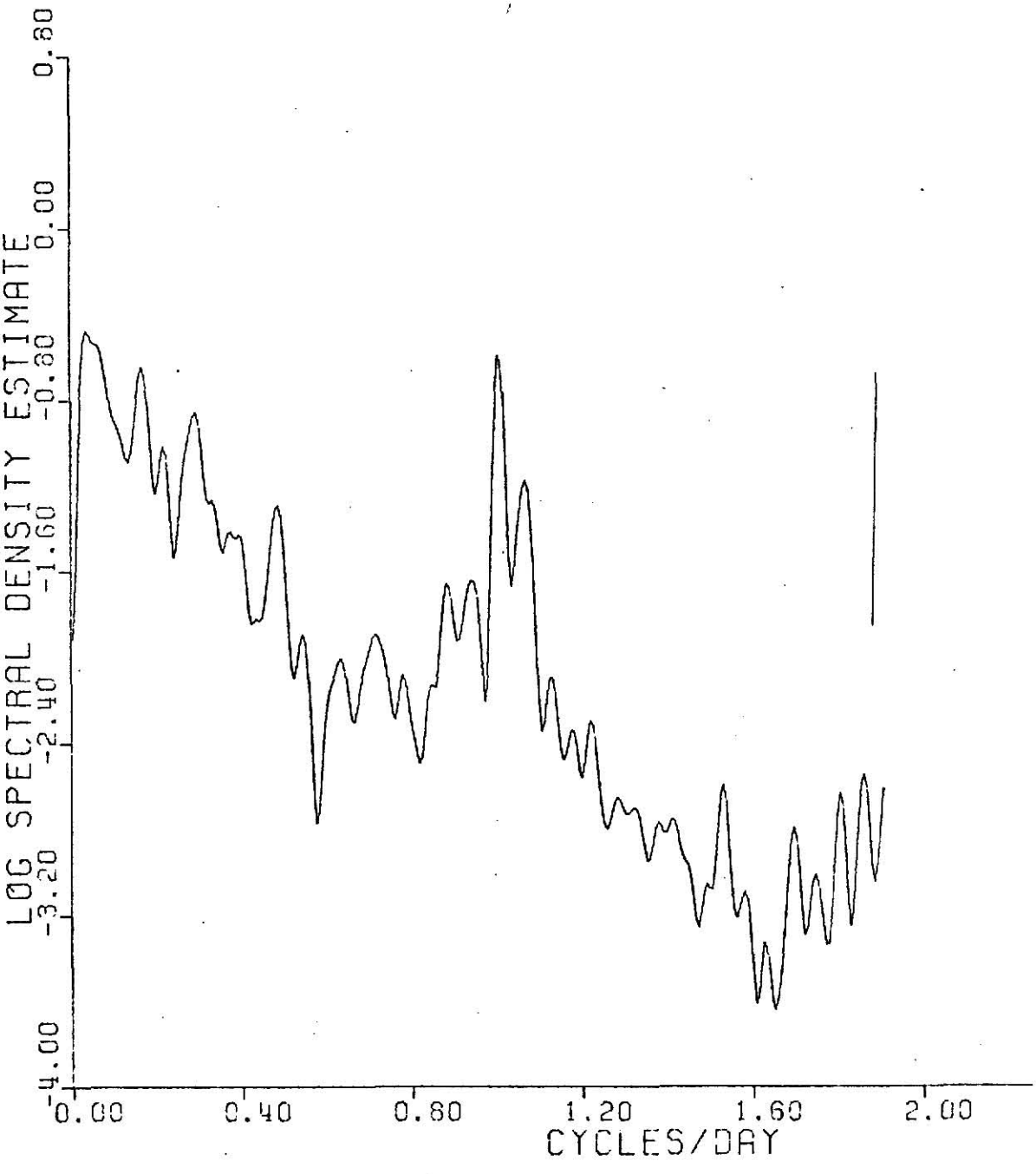

FIGURE 16. Spectral plot of residuals for six hour averaged data from station 1.

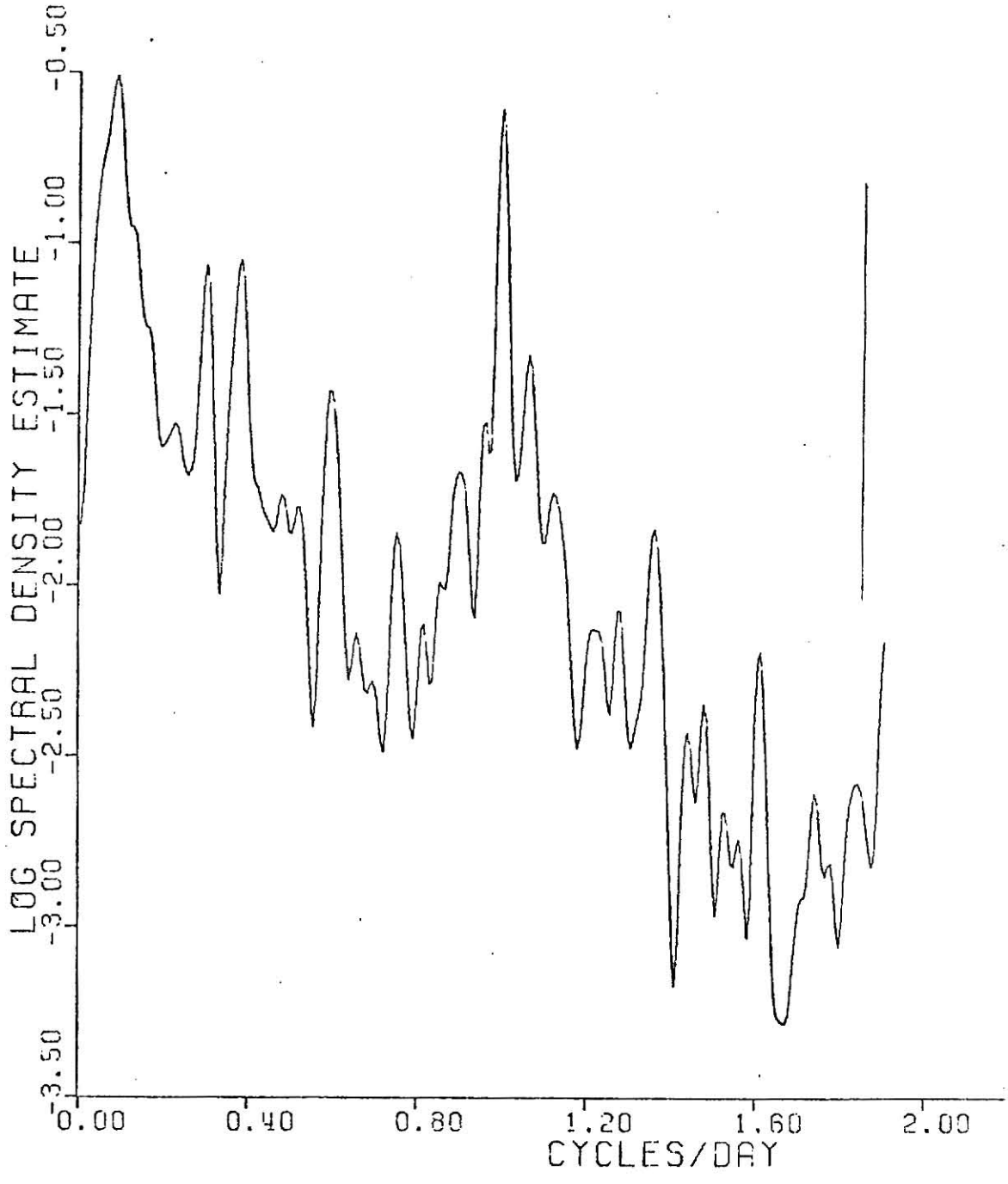

FIGURE 17. Spectral plot of residuals for six hour averaged data from station 71.

The results for the other pollutants from all the stations are summarized in Tables A, B & C. Table A gives the result summary for carbon monoxide data from all the stations, Table B gives the result summary for nitrogen oxides data from all the stations and Table C for Ozone data from all the stations. The plots of these pollutants data and the graphical representations of autocorrelation, power spectrum and the residuals power spectrum are given in Appendix A from Figs. 20-59. The haromonic analysis is given in Tables 6-17. The predictive model for all the pollutants at all the stations is given below. The first subscript of Y represents the station number and the second the pollutant. CO stands for carbon monoxide, NQ for nitrogen oxides and OZ for ozone.

Predictive model of carbon monoxide from all the stations

$$Y_{60,CO} = .38 \sin(2\pi t/240) - .38 (2\pi t/288) - .33 \sin(2\pi t/60)$$
$$+ .26 \cos(2\pi t/2) + .32 \cos(2\pi t/1480)$$
$$- .31 \cos(2\pi t/240) + .31 \sin(2\pi t/300)$$
$$+ .30 \sin(2\pi t/730) + .28 \sin(2\pi t/72)$$
$$+ .23 \sin(2\pi t/288) + .21 \cos(2\pi t/4)$$
$$- .21 \sin(2\pi t/96) + .18 \cos(2\pi t/20)$$

TABLE A

Result Summary of Carbon Monoxide Data from Stations 60, 79, 1 and 71.

Carbon Monoxide			
Station	Harmonic Analysis	Expected Cycles	R^2
60	See Table 6, Appendix A	Yearly, half yearly, 60 days, 50 days 72 days, 15 days, 7 days, 5 days, 3 days, 1 day, 12 hours	.4790
79	See Table 7, Appendix A	Yearly, half yearly, 90 days, 172 days, 60 days, 18 days, 14 days, 12 days, 6 days, 5 days, 1 day, 12 hours	.6200
1	See Table 8, Appendix A	Yearly, 72 days, 30 days, 14 days, 12 days, 7 days, 4 days, 3 days, 1 day, 20 hours, 12 hours	.4431
71	See Table 9, Appendix A	Yearly, half yearly, 120 days, 72 days, 90 days, 60 days, 14 days, 12 days, 7 days, 4 days, 1 day, 12 hours	.5724

TABLE B

Result Summary of Nitrogen Oxides Data From Stations 60, 79, 1 and 71.

Nitrogen Oxides			
Station	Harmonic Analysis	Expected Cycles	R^2
60	Table 10, Appendix A	Yearly, 120 days, 90 days, 72.5 days, 30 days, 23 days, 18 days, 15 days, 7 days, 5 days, 1 day, 12 hours.	
79	Table 11, Appendix A	Yearly, 120 days, 72.5 days, 45 days, 30 days, 23 days, 19 days, 14 days, 12 days, 7 days, 5 days, 1 day, 12 hr.	.5959
1	Table 12, Appendix A	Yearly, 120 days, 72.5 days, 60 days, 30 days, 26 days, 14 days, 12.5 days, 7 days, 1 day, 12 hrs.	.5678
71	Table 13, Appendix A	Yearly, half yearly, 120 days, 90 days, 30 days, 26 days, 14 days, 12.5 days, 10 days, 1 day, 12 hrs.	.6191

TABLE C

Result Summary of Ozone Data From Stations 60, 79, 1 and 71.

Station	Harmonic Analysis	Ozone Expected Cycles	R^2
60	Table 14, Appendix A	Yearly, half yearly, 120 days, 73 days, 40 days, 36 days, 30 days, 15 days, 9 days, 1 day, 12 hrs.	.7457
79	Table 15, Appendix A	Yearly, half yearly, 120 days, 90 days, 72.5 days, 36 days, 24 days ,15 days, 9 days, 5 days, 1 day, 12 hrs.	.7593
1	Table 16, Appendix A	Yearly, half yearly, 90 days, 72.5 days, 52 days, 36 days, 24 days, 16 days, 12.5 days, 7 days, 1 day, 12 hrs	.7002
71	Table 17, Appendix A	Yearly, half yearly, 72.5 days, 60 days, 28 days, 24 days, 19 days, 15 days, 12.5 days, 7 days, 1 day, 12 hrs.	.7030

$$Y_{79,CO} = -.0014t + 1.38 \cos (2\pi t/1400) + 1.02 \cos (2\pi t/4)$$
$$+ .62 \cos (2\pi t/2) - 1.19 \sin (2\pi t/1400)$$
$$+ .66 \cos (2\pi t/730) - .56 \cos (2\pi t/240)$$
$$- .49 \cos (2\pi t/288) + .51 \sin (2\pi t/288)$$
$$+ .43 \sin (2\pi t/50) + .42 \sin (2\pi t/72)$$
$$+ .40 \sin (2\pi t/360)$$

$$Y_{1,CO} = 1.37 \cos (2\pi t/1460) - .72 \cos (2\pi t/4)$$
$$- .66 \sin (2\pi t/1460) + .61 \cos (2\pi t/56)$$
$$+ .55 \sin (2\pi t/120) - .44 \sin (2\pi t/28)$$
$$+ .45 \cos (2\pi t/28) + .45 \cos (2\pi t/120)$$
$$- .44 \cos (2\pi t/120) + .42 \sin (2\pi t/315)$$
$$- .34 \cos (2\pi t/3.5)$$

$$Y_{71,CO} = 1.59 \cos (2\pi t/140) + .67 \cos (2\pi t/2)$$
$$+ .67 \sin (2\pi t/50) + .66 \sin (2\pi t/360)$$
$$+ .49 \sin (2\pi t/480) + .48 \cos (2\pi t/56)$$
$$+ .44 \sin (2\pi t/4) - .44 \cos (2\pi t/288)$$

$$- .41 \cos (2\pi t/240) + .39 \sin (2\pi t/730)$$

$$+ .38 \sin (2\pi t/288) - .36 \sin (2\pi t/1460)$$

Predictive models of nitrogen oxides from all the stations.

$$Y_{60,NO} = 1.09 \cos (2\pi t/20) + 1.38 \cos (2\pi t/410)$$

$$- 1.35 \cos (2\pi t/92) - 0.99 \cos (2\pi t/120)$$

$$- 0.88 \sin (2\pi t/60) - 0.85 \sin (2\pi t/1400)$$

$$- 0.83 \cos (2\pi t/290) - 0.74 \sin (2\pi t/28)$$

$$- 0.72 \cos (2\pi t/300) + 0.73 \sin (2\pi t/92)$$

$$+ 0.75 \sin (2\pi t/120) - 0.72 \sin (2\pi t/20)$$

$$+ 0.67 \cos (2\pi t/20) + 0.65 \sin (2\pi t/290)$$

$$Y_{79,NO} = 4.30 \cos (2\pi t/1460) + 3.46 \cos (2\pi t/4)$$

$$+ 1.95 \cos (2\pi t/2) + 2.18 \sin (2\pi t/290)$$

$$+ 1.48 \cos (2\pi t/28 - 1.43 \cos (2\pi t/48)$$

$$+ 1.42 \cos (2\pi t/480) - 1.39 \cos (2\pi t/120)$$

$$+ 1.31 \sin (2\pi t/76) + 1.17 \sin (2\pi t/4)$$

$$+ 1.09 \sin (2\pi t/120) + 1.05 \sin (2\pi t/56)$$

$$Y_{1,NO} = 8.76 \cos (2\pi t/1400) + 2.65 \cos (2\pi t/2)$$
$$+ 3.17 \sin (2\pi t/4) - 2.96 \cos (2\pi t/104)$$
$$+ 2.97 \sin (2\pi t/50) - 2.93 \cos (2\pi t/240)$$
$$+ 2.33 \sin (2\pi t/120) + 2.15 \cos (2\pi t/56)$$
$$+ 2.12 \cos (2\pi t/28) - 2.11 \sin (2\pi t/28)$$
$$- 2.05 \cos (2\pi t/290) + 2.03 \sin (2\pi t/290)$$

$$Y_{71,NO} = 8.29 \cos (2\pi t/1460) + 3.09 \cos (2\pi t/2)$$
$$+ 3.26 \sin (2\pi t/50) + 3.04 \sin (2\pi t/360)$$
$$+ 3.02 \sin (2\pi t/4) + 2.76 \sin (2\pi t/480)$$
$$+ 2.62 \cos (2\pi t/56) - 2.04 \cos (2\pi t/104)$$
$$- 2.14 \cos (2\pi t/120) - 1.83 \sin (2\pi t/1480)$$
$$+ 1.77 \cos (2\pi t/730) + 1.59 \sin (2\pi t/120)$$

Predictive models of ozone from all the stations.

$$Y_{60,OZ} = -3.78 \sin (2\pi/4) - 2.17 \cos (2\pi t/1400)$$
$$-1.06 \cos (2\pi t/2) - 1.26 \cos (2\pi t/4)$$
$$+ 0.89 \sin (2\pi t/730) - 0.72 \sin (2\pi t/1460)$$

$$-0.45 \cos (2\pi t/290) + 0.45 \sin (2\pi t/180)$$

$$-0.43 \sin (2\pi t/144)$$

$$Y_{79,OZ} = -3.40 \sin (2\pi t/4) - 1.99 \cos (2\pi t/1460)$$

$$-1.22 \cos (2\pi t/4 - 0.78 \cos (2\pi t/1460)$$

$$+ .75 \sin (2\pi t/730) - 0.43 \cos (2\pi t/480)$$

$$- 0.41 \cos (2\pi t/290) - 0.39 \sin (2\pi t/144)$$

$$Y_{1,OZ} = -1.67 \sin (2\pi t/4) - 0.93 \cos (2\pi t/1460)$$

$$-0.60 \cos (2\pi t/4) - 0.39 \cos (2\pi t/2)$$

$$-0.31 \sin (2\pi t/96) + 0.29 \sin (2\pi t/290)$$

$$+ 0.26 \sin (2\pi t/780) + 0.23 \sin (2\pi t/28)$$

$$- 0.23 \cos (2\pi t/290)$$

$$Y_{71,OZ} = -1.48 \sin (2\pi t/4) - 0.71 \cos (2\pi t/1460)$$

$$-0.41 \cos (2\pi t/2) - 0.29 \cos (2\pi t/4)$$

$$-0.28 \cos (2\pi t/240) - 0.28 \sin (2\pi t/96)$$

$$+ 0.26 \sin (2\pi t/730) + 0.22 \sin (2\pi t/290)$$

$$- 0.20 \cos (2\pi t/50) + 0.18 \sin (2\pi t/76)$$

$$- 0.16 \sin (2\pi t/112)$$

Discussion

It was intended to identify some of the cycles analyzed. As mentioned earlier the variations in the wind speed and in the inversion layer height are among the major contributors to the cyclic variations. Daily data were available for the inversion layer height. This data were analyzed to determine the cyclic variations in the inversion layer height. The harmonic analysis is shown in Table 18 and the autocorrelation function and power spectrum in Figs. 18 and 19. It is seen that the cyclic variations in the inversion layer height correspond to the periods of one year, half year, 72.5 days, 30 days, 24-25 days, and 15-14 days. Two adjacent peaks in the spectrum plot correspond to cycles corresponding to about 5 days and 4 days frequencies but they lie within the confidence interval. Spectral analysis was also done of the six hour averaged wind speed data from all the four stations. It was found that at all the four stations there were 24-hours and 12 hours cyclic variations in the wind speed. It is seen that these cyclic variations were observed in the analysis of the pollutant concentration data at most of the stations. The power spectrum for the wind speed data from station 71 is shown in Fig. 60. The power spectrum for wind speed at other stations is similar and is not shown. Thus apart from the cyclic variations due to social activities it is seen that the cyclic variations in the pollutant concentration are caused by the variations in the inversion layer height and wind speed.

Table 18

Harmonic Analysis of Inversion Layer Base Height at Station 1

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	49.115	0.0	49.115	0.0	2412.29	0.7127
1	-5.148	-2.724	5.824	207.87	67.85	0.0698
2	-4.213	-5.176	6.674	230.83	89.09	0.0916
5	-3.798	3.192	4.962	139.08	49.24	0.0506
8	0.110	3.427	3.427	88.12	23.51	0.0242
15	-2.075	-3.356	3.946	238.25	31.14	0.0320
16	2.009	1.380	2.438	34.47	11.88	0.0122
21	1.284	-1.967	2.349	303.16	11.03	0.0133
24	4.188	1.022	4.311	13.71	37.17	0.0382
26	0.155	-3.628	3.631	272.49	26.38	0.0271
29	-2.486	-0.481	2.532	190.95	12.82	0.0132
30	0.165	2.362	2.368	85.96	11.22	0.0115
33	2.074	-2.231	3.046	312.93	18.566	0.0191
57	-1.533	-2.250	2.722	235.20	14.83	0.0152
99	2.357	1.056	2.583	24.12	13.35	0.0132
100	-2.262	-1.443	2.683	212.51	14.40	0.0148

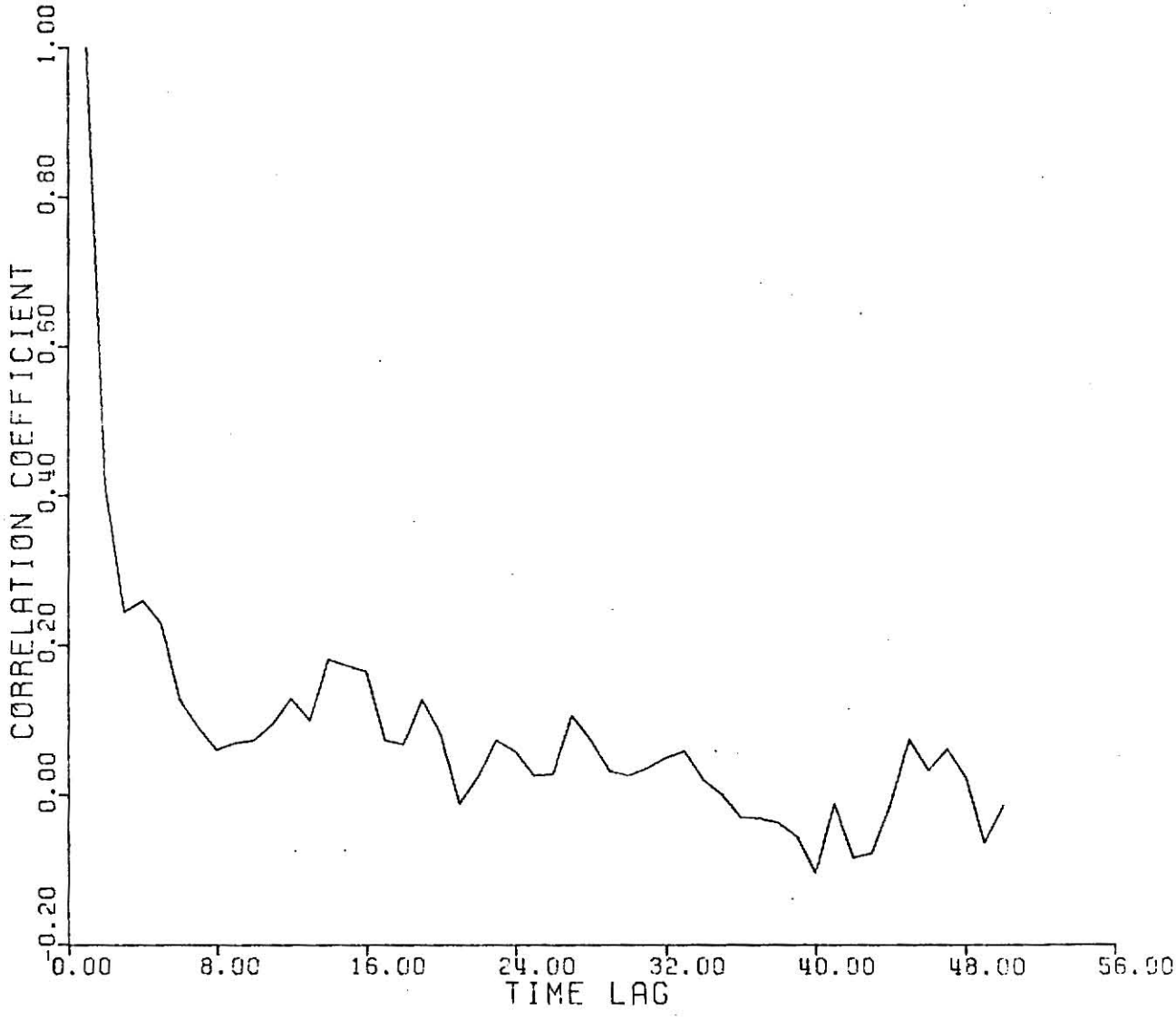

FIGURE 18. Autocorrection function of inversion layer height from station 1.

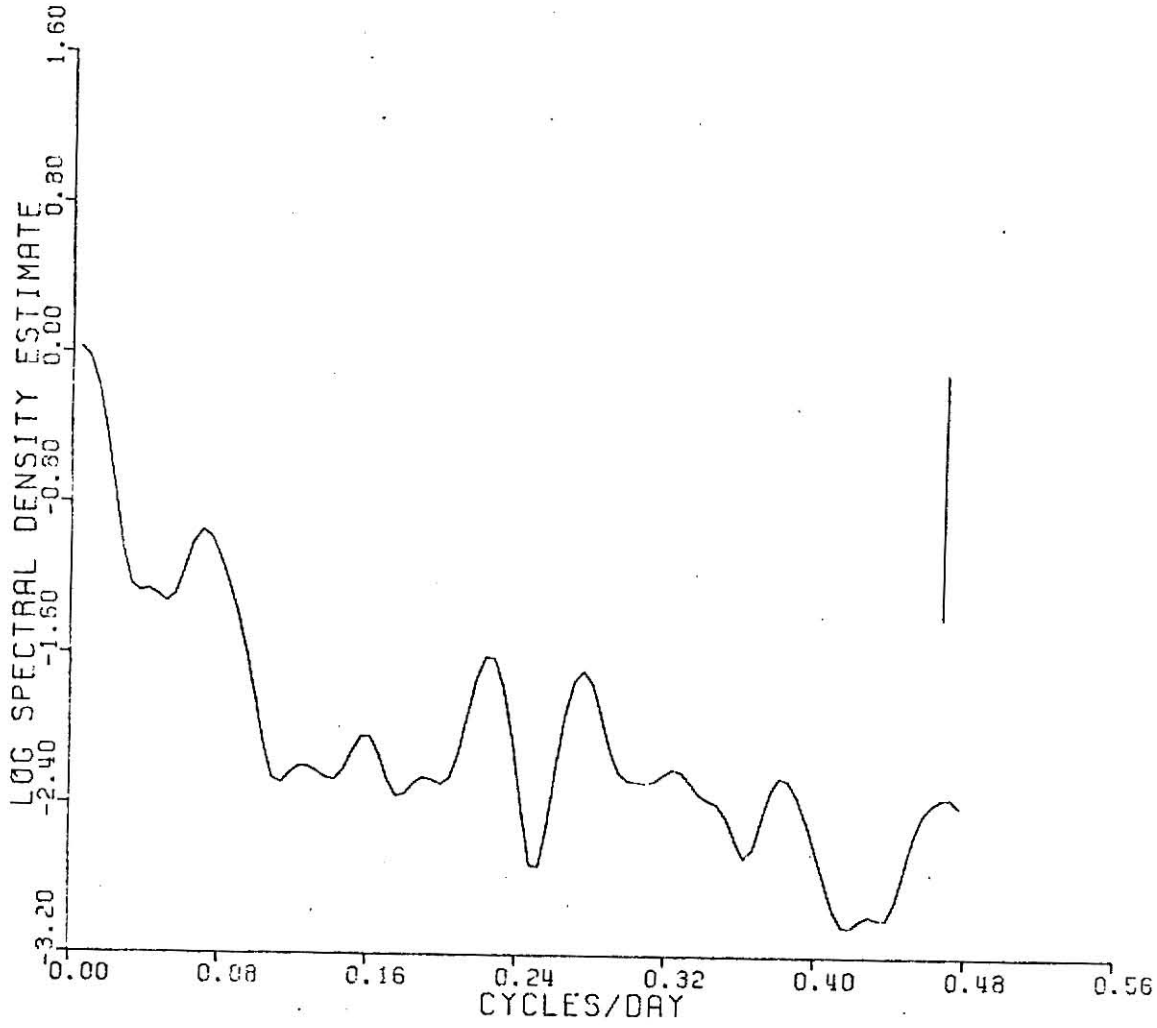

FIGURE 19. Spectral density function for inversion layer height based on daily observation, station 1.

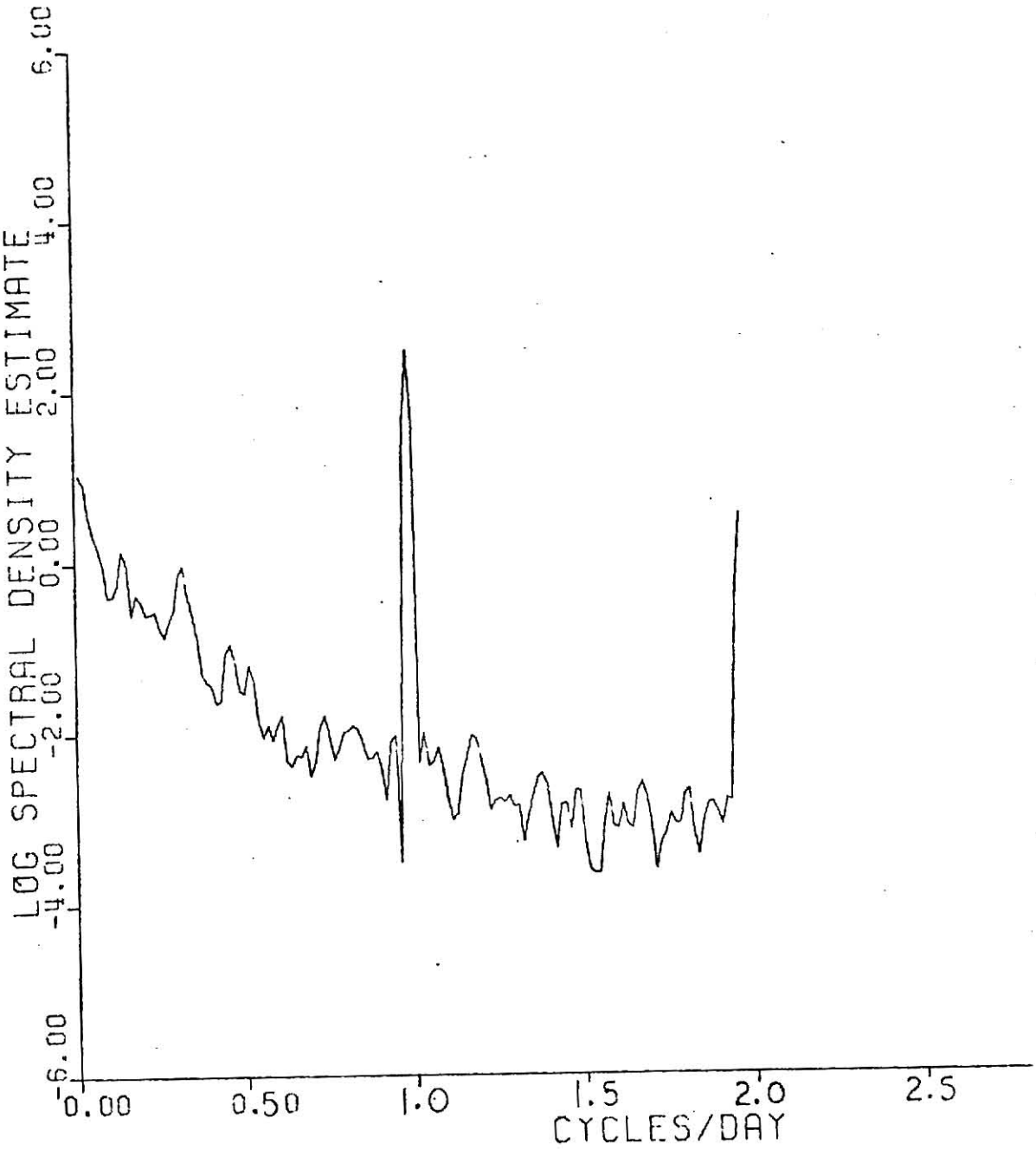

FIGURE 60. Spectral density function for wind speed based on six hours averaged data, station 71.

CHAPTER 5

CROSS SPECTRAL ANALYSIS OF THE AIR POLLUTION DATA

INTRODUCTION

In this chapter the cross spectral analysis of the air pollution data from the San Francisco bay area and the greater Los Angeles area is presented. The cross spectral analysis is done for the same pollutant at different stations and for the different pollutants at the same station. The former determines how the pollutant concentration at one station affects the pollutant concentration at the other station while the later gives the relative concentration of the pollutants at a station and in addition determines the distribution of the sources associated with the type of pollutants.

Cross Spectral Analysis

This analysis is characterized by the fact that this is carried over two or more stochastic processes rather than on one process at differnt time spans as in autospectral analysis. The theory for the cross spectral analysis can be developed working under the same lines as for spectral analysis described in Chapter 3. Both the stochastic processes are assumed to be stationary.

Let the two time series be represented by $x(t)$ and $y(t)$. As these are the pollutant concentration observations at discrete and uniform time interval they form discrete time series. To determine the cross temporal relationship between the two series, the cross covariance between $x(t)$ and $y(t)$ is determined. For a positive time log of k it, is given as,

$$C_{xy}(k) = \frac{1}{N-k} \sum_{p=1}^{N-k} x_p \, y_{p+k} \tag{1}$$

and for a negative time lag of k it is given as,

$$C_{xy}(-k) = \frac{1}{N-k} \sum_{p=1}^{N-k} y_p \, x_{p+k} \tag{2}$$

where x and y are deviations from the mean. The cross correlation function is given as,

$$\rho_{xy}(k) = \frac{C_{xy}(k)}{\sqrt{C_{xx}(0) \; C_{yy}(0)}} \tag{3}$$

The cross correlation function contrary to autocorrelation function is not an even function. Thus,

$$\rho_{xy}(k) \neq \rho_{xy}(-k) \tag{4}$$

The interpretation of the cross covariance or cross correlation function is exactly the same as for the autocorrelation function except for the fact that it is now explained concerning two series at different time lags rather than for the same series at different time lags as in the former.

The cross spectrum is a Fourier transform of the cross covariance function given by,

$$f_{xy}(\omega) = \sum_{k=-\infty}^{\infty} C_{xy}(k) \; e^{-i\omega k} \, dk \tag{4}$$

As $C_{xy}(k)$ is not an even function the expression reduces to a complex equation giving real and immaginary parts. Rewriting $C_{xy}(k)$ as the sum of an even part $\lambda_{xy}(k)$ and an odd part $\psi_{xy}(k)$ gives

$$f_{xy}(\omega) = CO_{xy}(\omega) + iQD_{xy}(\omega) \tag{5}$$

where,

$$CO_{xy}(\omega) = \sum_{k=-\infty}^{\infty} \lambda_{xy}(k) \text{ Cos } \omega k \quad dk$$

and

$$QD_{xy}(\omega) = \sum_{k=-\infty}^{\infty} \psi_{xy}(k) \text{ Sin } \omega k \quad dk$$

$CO_{xy}(\omega)$ is called the co-spectrum of the (x(t), y(t)) process at ω and $QD_{xy}(\omega)$ is called the quadrature spectrum of the process at ω. The co-spectrum determines the covariance between in-phase components i.e. it measures the covariance between two cosine components and the covariance between the two sine components. The quadrature spectrum determines the covariance between sine and cosine components, that is out of phase or quadrature components.

The estimates of the power co-spectrum and quadrature spectrum are given as,

$$CO_{xy}(\omega_j) = \frac{1}{2\pi} \left[C_{xy}(0) + \sum_{k=1}^{M-1} (C_{xy}(k) + C_{yx}(k)) \text{ Cos } \omega_j k \right.$$

$$\left. + \frac{1}{2} (C_{xy}(M) + C_{yx}(M)) \text{ Cos } \pi j \right]$$

$$QD_{xy}(\omega j) = \frac{1}{2\pi}\left[\sum_{k=1}^{M} (C_{xy}(k) - C_{yx}(k)) \text{ Sin } \omega_j k\right.$$

$$\left. + \frac{1}{2}(C_{xy}(M) - C_{yx}(M)) \text{ Sin } \pi j\right] \qquad (6)$$

where

$$\omega_j = \frac{\pi j}{M}, \qquad j = 0, \ldots, M$$

M is the total number of lags. Smoothed estimates were obtained using a Hamming window. Thus if the functions CO_{xy} and QD_{xy} are both represented by a dummy function f, the smoothed estimates can be given as,

$$Sf(0) = 0.54f(0) + 0.46\ f(1)$$

$$Sf(k) = 0.23f(k-1) + 0.54f(k) + 0.23f(k+1)$$

$$Sf(M) = 0.54f(M) + 0.46f(M-1) \qquad (7)$$

The co-spectrum and quadrature spectrum have little use as they stand therefore more important quantities namely coherence and phase are calculated.

Coherence and Phase

Coherence is defined as the correlation between the two series at a particular frequency. The coherence square is expressed as,

$$k^2_{xy}(\omega) = \frac{SCO^2(\omega) + SQD^2(\omega)}{Sf_x(\omega) \cdot Sf_y(\omega)} \qquad (8)$$

where SCO(ω) represents the smoothed co-spectrum and SQD(ω) represents the smoothed quadrature spectrum. Both are calculated using the equation set 7, as explained above. $Sf_x(\omega)$ and $Sf_y(\omega)$ represent the smoothed spectral estimates of the series x(t) and y(t) calculated using Eq. 14 in Chapter 3.

The stationarity assumption tells that if the frequency for a series is divided into various bands then the component of the variance lying in a particular band is independent of not only all other components of variance of the same series lying in other bands but also of the components of variance lying in the other frequency bands of the second series except the one corresponding to the same frequency band. Thus, we can determine the correlation between the two series at different frequencies, each is independent of the other. It is this correlation which the coherency determines. This coherency lies between 0 and 1.

The phase of the smoothed spectrum is expressed as,

$$PH_{xy}(\omega) = \text{ARCTAN } (SQD_{xy}(\omega) \; / \; SCO_{xy}(\omega)) \qquad (9)$$

This determines the phase difference between the two series at a particular frequency ω.

The coherency and phase spectrums have their own distribution. In fact the variance of their estimates is zero when the coherency is unity and the variance increases as the coherency tends to zero. Thus to determine the true coherence between the two series confidence intervals were drawn on the coherence diagram. The confidence intervals depend upon the ratio N/M and the level of confidence chosen. The confidence

intervals for various values of ratio N/M and the level of confidence are given in Granger and Hatanaka (35). These were used to draw the confidence intervals on the coherence diagram. Corresponding to the average value of the coherency and the ratio N/M the confidence intervals for phase diagram are given in Granger and Hatanaka [35]. These confidence intervals are shown on the plots with broken lines.

The Program

BMD02T program was used for the cross spectral analysis. This program computes the autocovariance, power spectrum, cross-covariance, cross spectrum, transfer function and coherence function of time series. For the details of the program the reader is referred to reference [43].

Cross Spectral Analysis of the Air Pollution Data from the San Francisco Bay Area.

The pollutant concentrations of the two pollutants formed the time series x(t) and y(t). The hourly observations were used for the two months data giving 1440 data points. The number of lags for both the hydrocarbons data from Richmond and Oakland and the carbon monoxide data from Oakland and San Jose were selected as 100. The cross spectral analysis was done for the hydrocarbons data from Richmond and Oakland, the carbon monoxide data from Oakland and San Jose and for the hydrocarbonds and carbon monoxide data from Oakland. The results are described below.

Hydrocarbons at Richmond and Oakland

The plot for the cross covariance function is shown in Fig. 1. The cross covariance is maximum at a lag of one hour. The cross covariance appears to be oscillating as the number of lags increases giving a peak

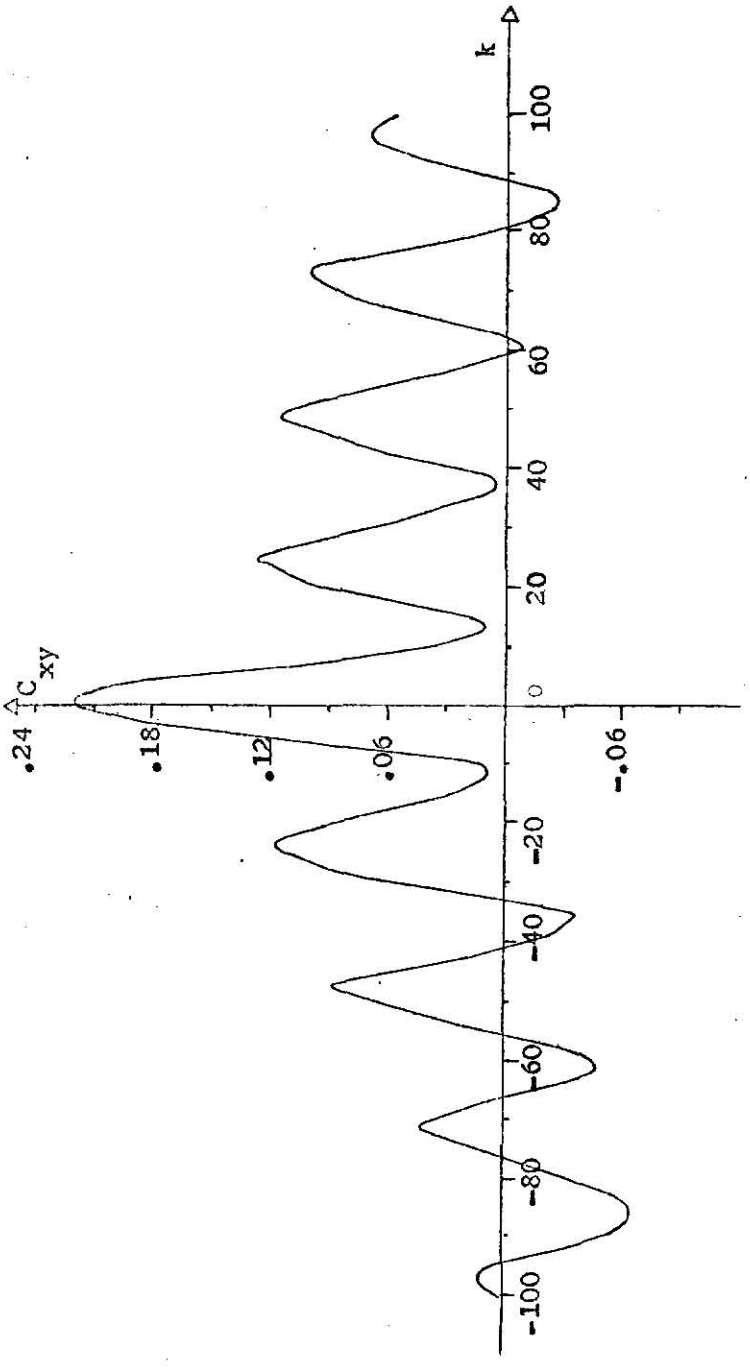

FIGURE 1. Cross covariance of hydrocarbons for Richmond and Oakland.

at a lag of every additional 24 hours. This indicates a high correlation between the hydrocarbon concentration at the two station at lags of 24 hours, 48 hours, 72 hours and 96 hours.

The coherency spectrum is presented in Fig. 2. The dashed line on the graph indicates the confidence interval. It is seen that the hydrocarbons at the two stations have quite significantly correlated cyclic variations corresponding to periods of 24 hours, 8 hours and about 4.5 hours. There is also high correlation at the low frequencies. The phase diagram, Fig. 3 reveals a significant phase relationship around frequency of 12 hours and about 5 hours. The confidence interval on this diagram is represented by the dashed line. The phase is in general clustered about the zero axis indicating not so marked phase relationship. The diagram reveals quite significant peaks at frequencies in the high range with no relationship at the low frequencies at all.

Carbon Monoxide at Oakland and San Jose

The cross covariance function plot is presented in Fig. 4. The two series have the maximum correlation at lag zero. The correlation dies down slowly indicating in general a high correlation between the carbon monoxide concentration at the two stations. There are quite significant peaks at lags of 24 hours, 48 hours, 72 hours and 96 hours. The coherence plot is shown in Fig. 5. It reveals a significant correlation between the carbon monoxide concentration at the two stations at frequencies corresponding to 12 hours, 8 hours and 5 hours with high coherence at low frequencies which can be expected to be corresponding to two months, one month, 15 days and 7 days found to be present at both the stations.

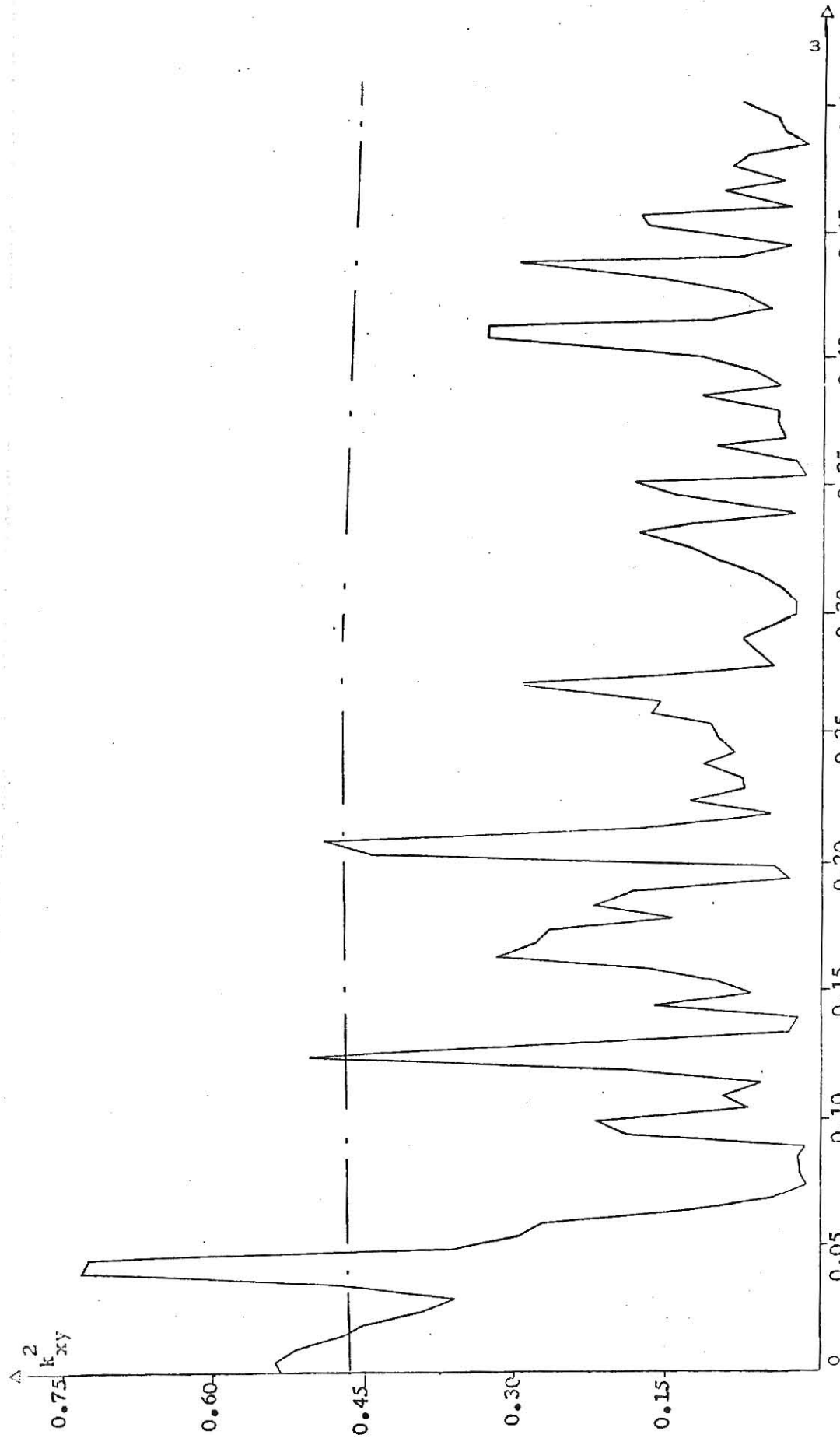

FIGURE 2. Square coherency vs. frequency, hydrocarbons from Richmond and Oakland.

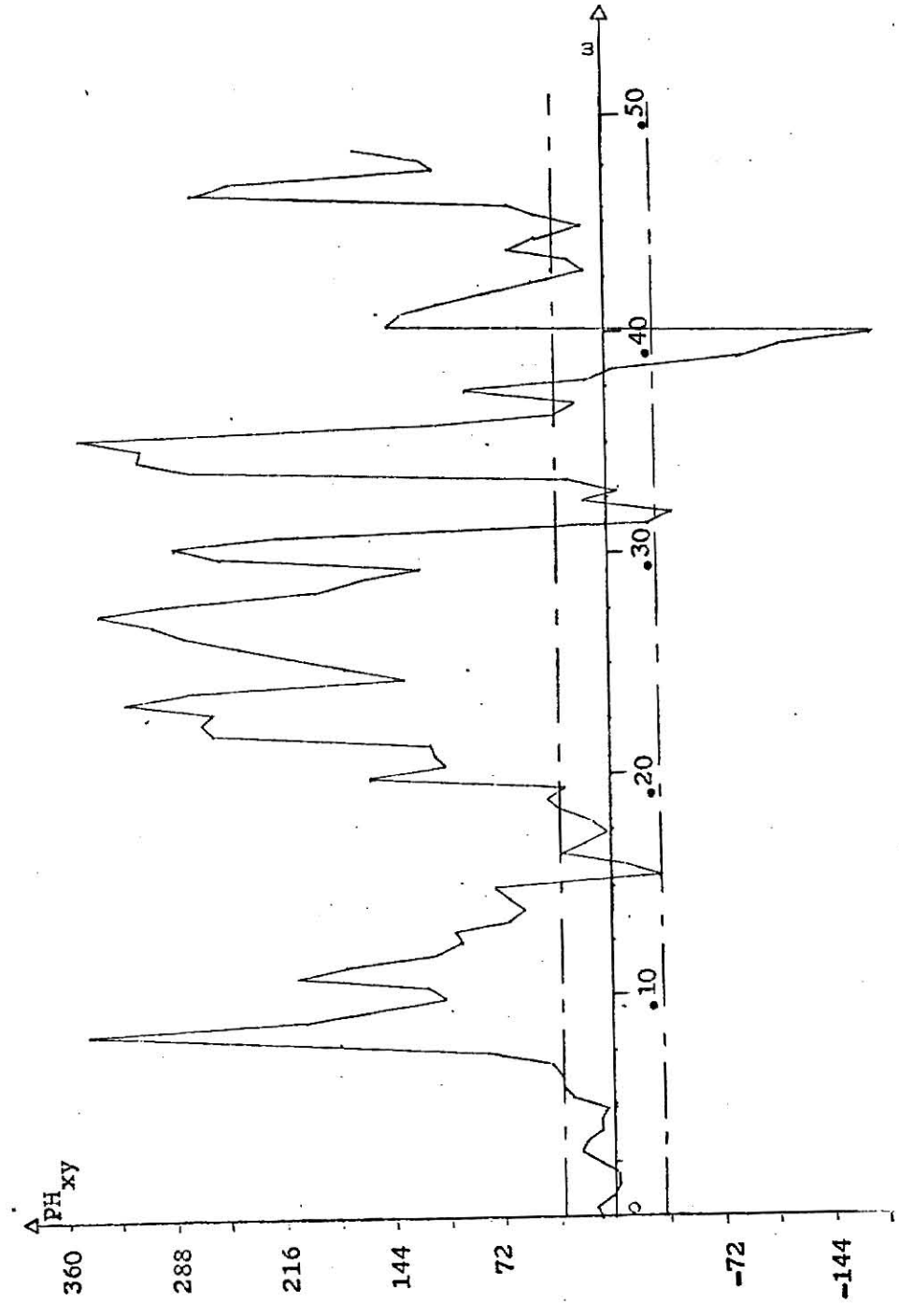

FIGURE 3. Phase vs. frequency, hydrocarbons from Richmond and Oakland.

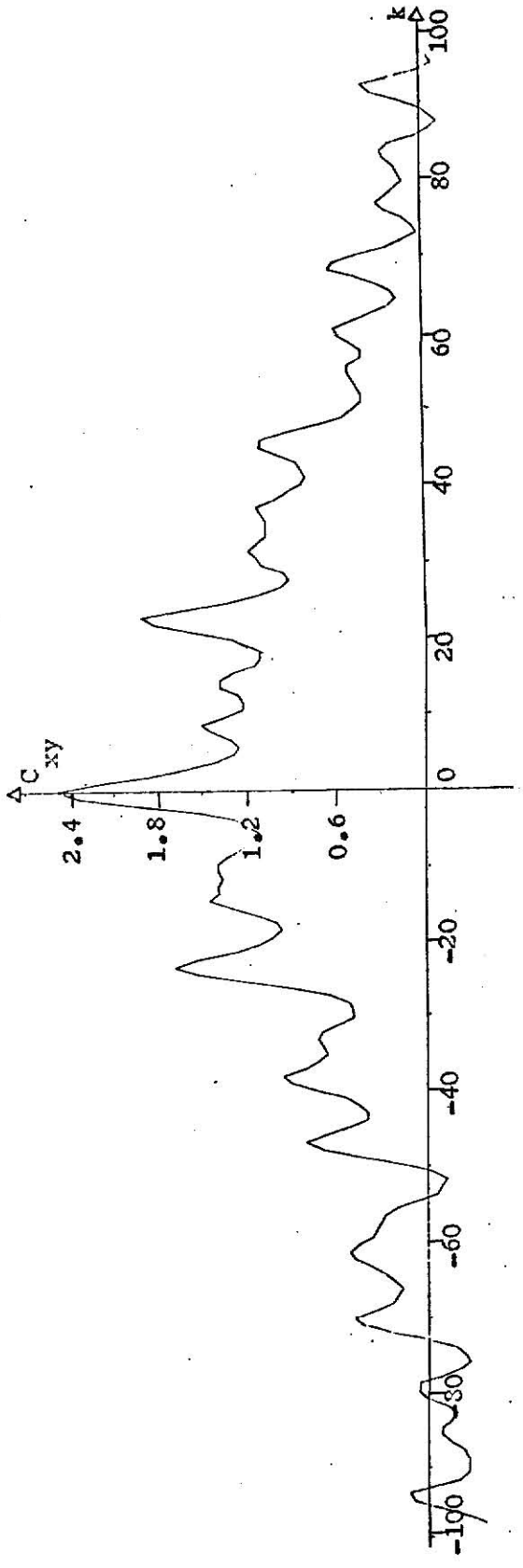

FIGURE 4. Cross covariance of carbon monoxide for Oakland and San Jose.

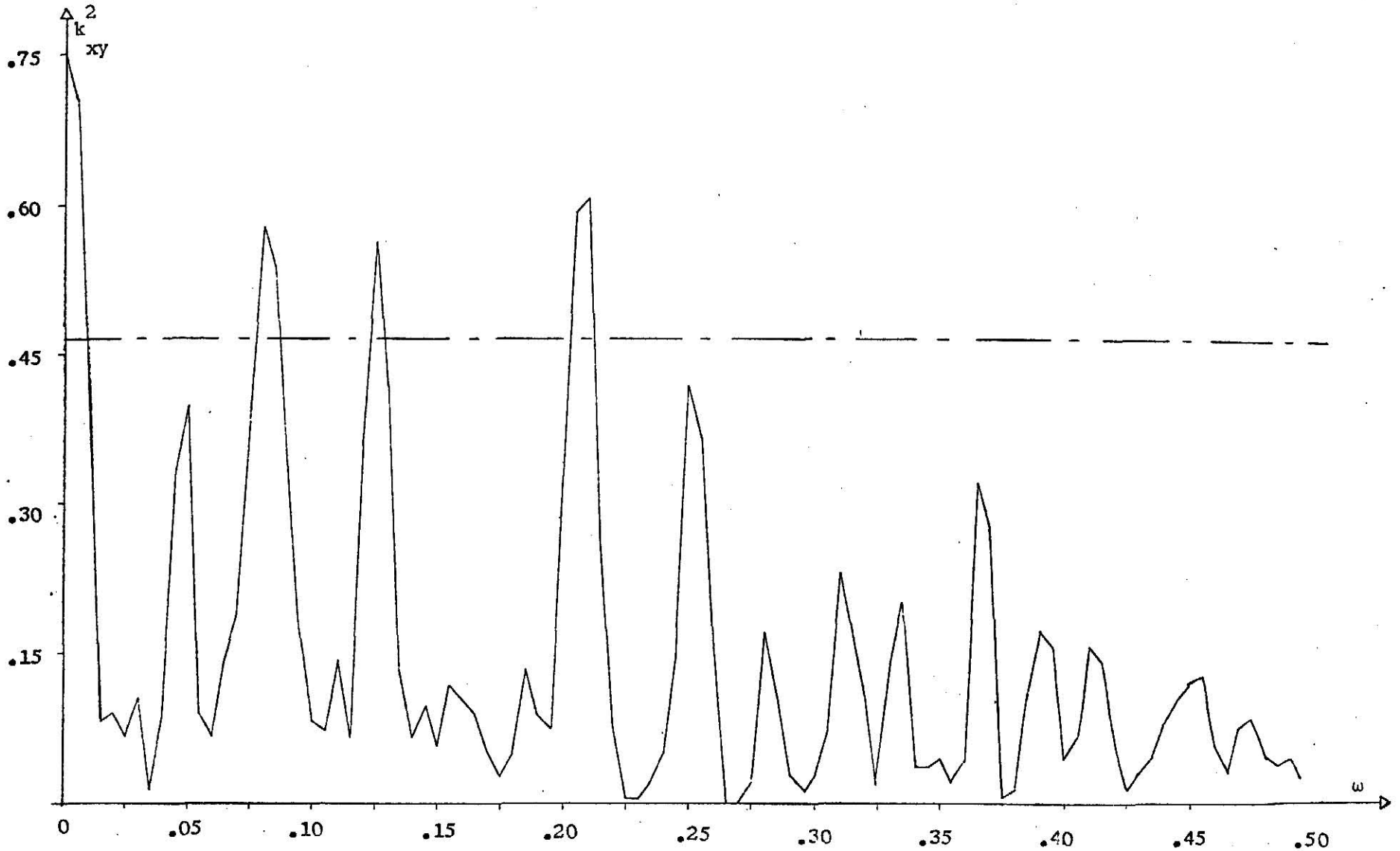

FIGURE 5. Squared coherency vs. frequency, carbon monoxide from Oakland and San Jose.

The phase diagram for this combination is shown in Fig. 6. The phase is seen to be in general clustered about the zero axis. It does not reveal any significant trend of phase relationship of the pollutant concentration at the two stations except the fact that the variation in the phase is becoming more and more pronounced in the high frequency range.

Hydrocarbons and Carbon Monoxide at Oakland

The cross covariance, coherency and phase diagrams for this combination are shown in Figs. 7-9. The covariance is maximum at lag zero dies down slowly with sharp peaks at lags of 24 hours, 48 hours and not so pronounced peaks at lags of 72 hours and 96 hours. There is also a flat peak around a lag of 8 hours. The coherence plot indicates a high coherence at all the frequencies. This is expected because hydrocarbons and carbon monoxide have many sources in common which produce them. The coherence at the low frequencies is about 0.9. A high coherence of 0.8 at 7 hours and 8 hours frequencies indicates the emission of both hydrocarbons and carbon monoxide from the automobiles. Then high coherence is also revealed at frequencies of .075 c/hour, .23 c/hour and .36 c/hour. The phase at all the frequencies is practically within the confidence interval indicating the simultaneous emission of hydrocarbons and carbon monoxide from the sources.

Cross Spectral Analysis of the Air Pollution Data From the Greater Los Angeles Area

The cross spectral analysis is presented for the sulphur dioxide concentrations at the stations 60, 79, 1 and 71 taken in pair and among

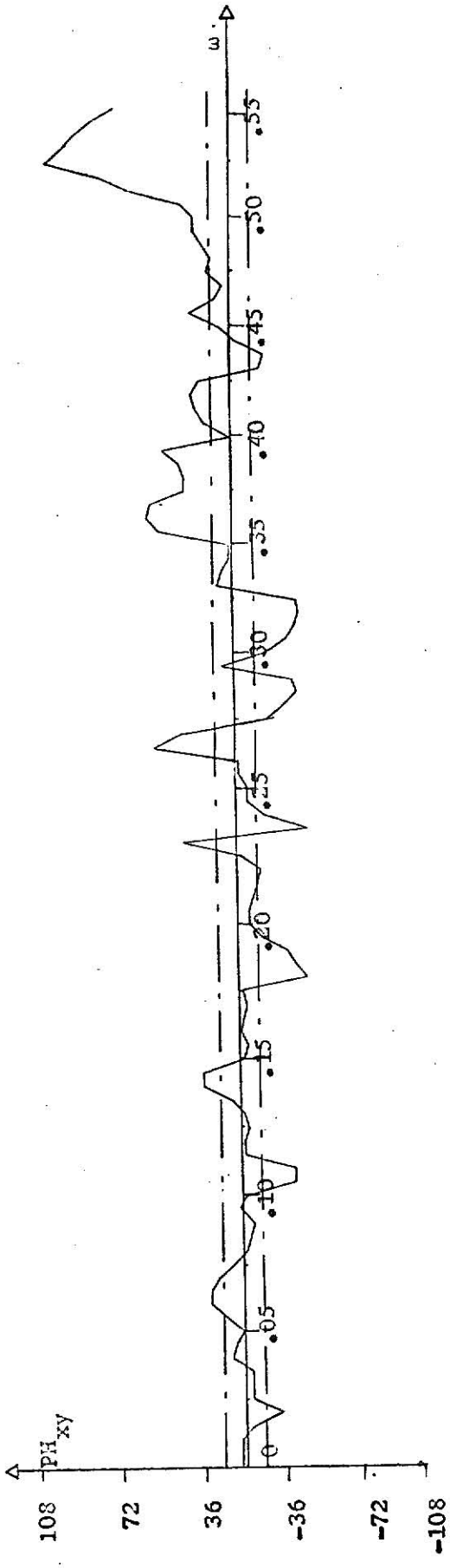

FIGURE 6. Phase vs. frequency, carbon monoxide from Oakland and San Jose.

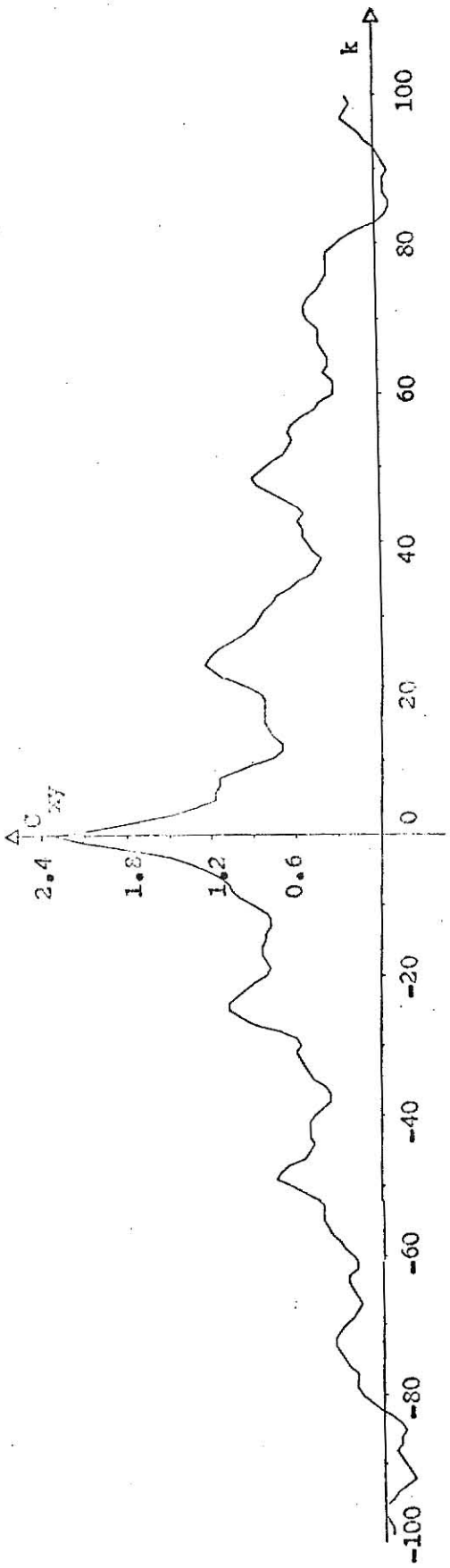

FIGURE 7. Cross covariance of hydrocarbons and carbon monoxide from Oakland.

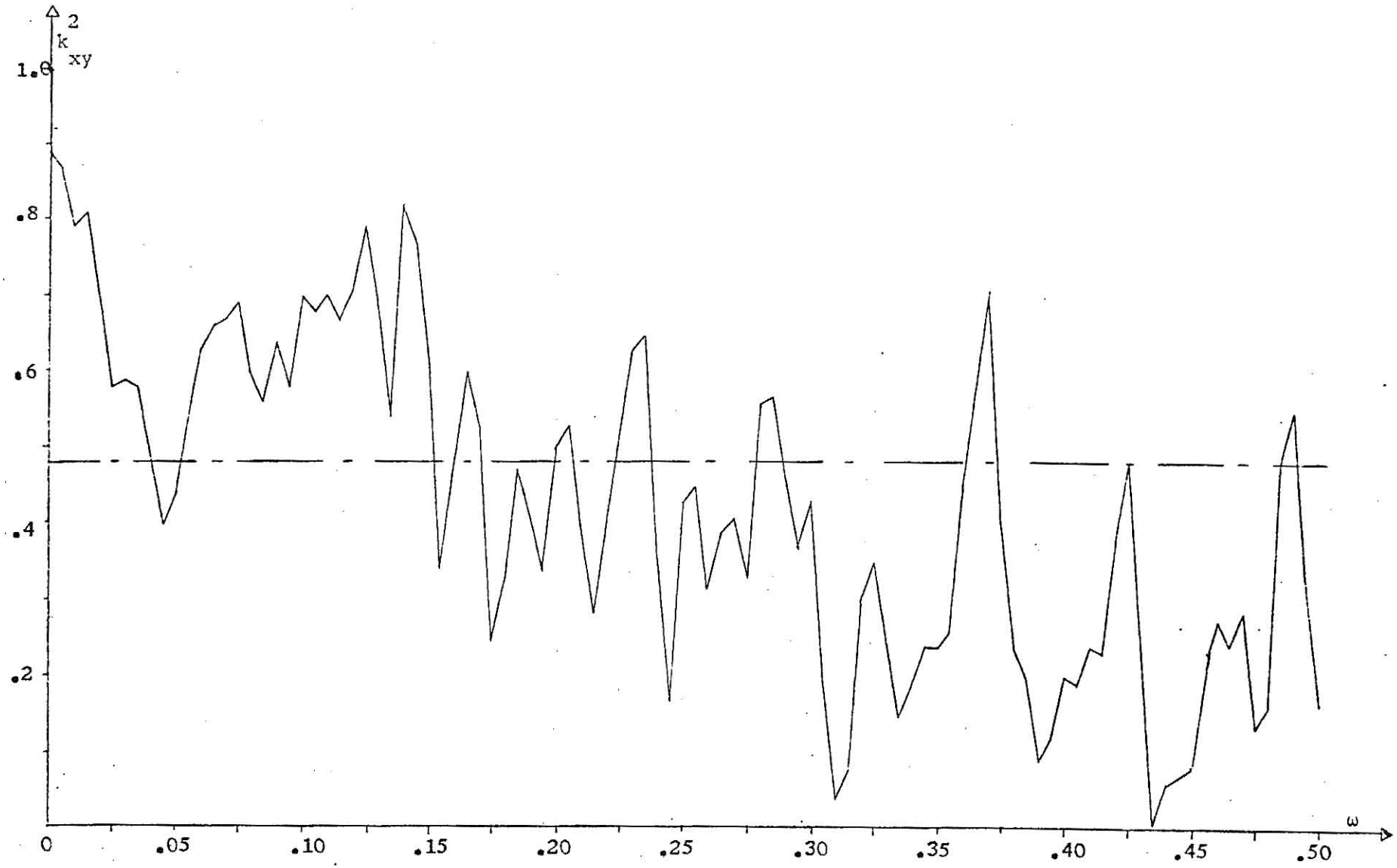

FIGURE 8. Squared coherency vs. frequency, hydrocarbons and carbon monoxide from Oakland.

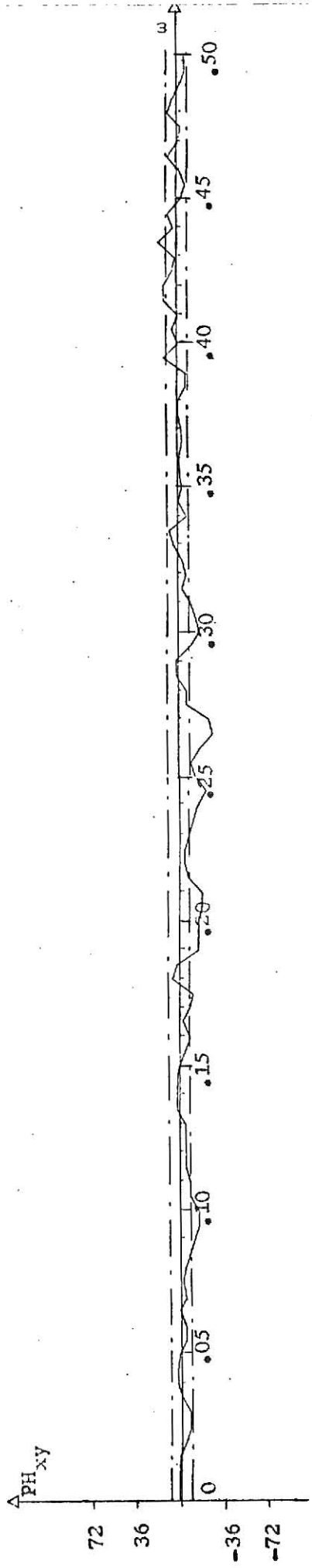

FIGURE 9. Phase vs. frequency, hydrocarbon and carbon monoxide from Oakland.

the pollutants namely sulphur dioxide, carbon monoxide, nitrogen oxides and ozone at station 1. The results are presented for the cross spectral analysis among various pollutants at station 60 and these are compared with those at station 1. As a daily cycle is present in the concentration of all the pollutants at all the stations a high coherence is expected at the daily frequency for the same pollutant at different stations and for the different pollutants at the same station. Twenty four hours averaged data were therefore considered which formed a series with 365 data points. The number of lags used were 50. The results of the analysis are given below. Also the results of cross spectral analysis between wind speed and sulphur dioxide at all the four stations are given.

Sulphur Dioxide at Stations 60 and 79

The graph of cross covariance function is presented in Fig. 10. It is seen that the sulphur dioxide concentration at the two stations has the maximum concentration at a lag of 30 days and less so at lags of 4, 12 and 18 days. The coherency plot shown in Fig. 11 reveals that the amount of sulphur dioxide at the stations 60 and 79 has higher correlation at low frequencies and it decreases with the increase in frequency. The coherence is in fact found to be insignificant as it lies within the confidence interval. The phase diagram is presented in Fig. 12. There is an upward trend in the phase plot in the range 0 to .37 c/day revealing a time lag of about a day between the two series. There is a time lag of about 15 days in the frequency band of .44 c/day to .46 c/day.

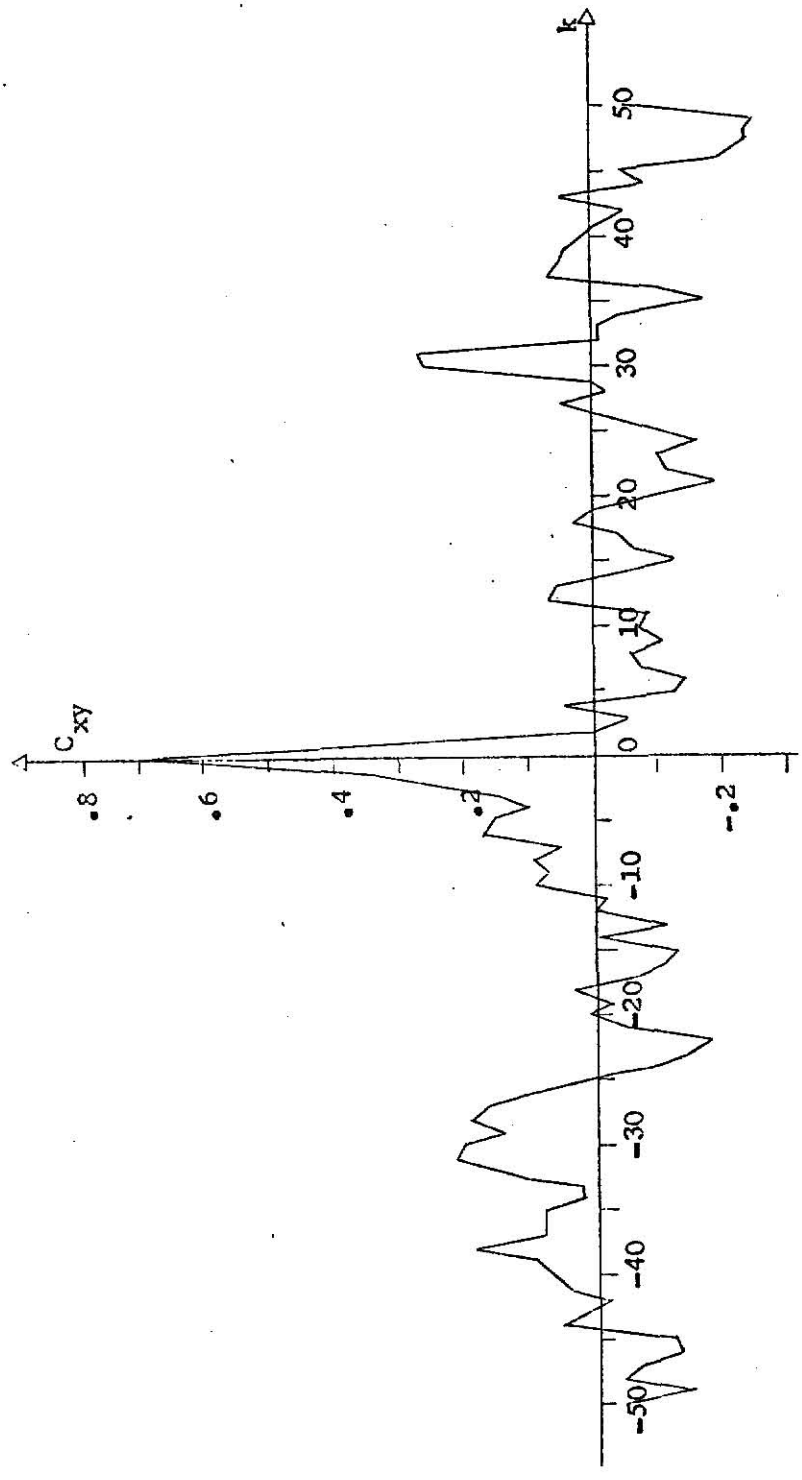

FIGURE 10. Cross covariance of sulphur dioxide for stations 79 and 60.

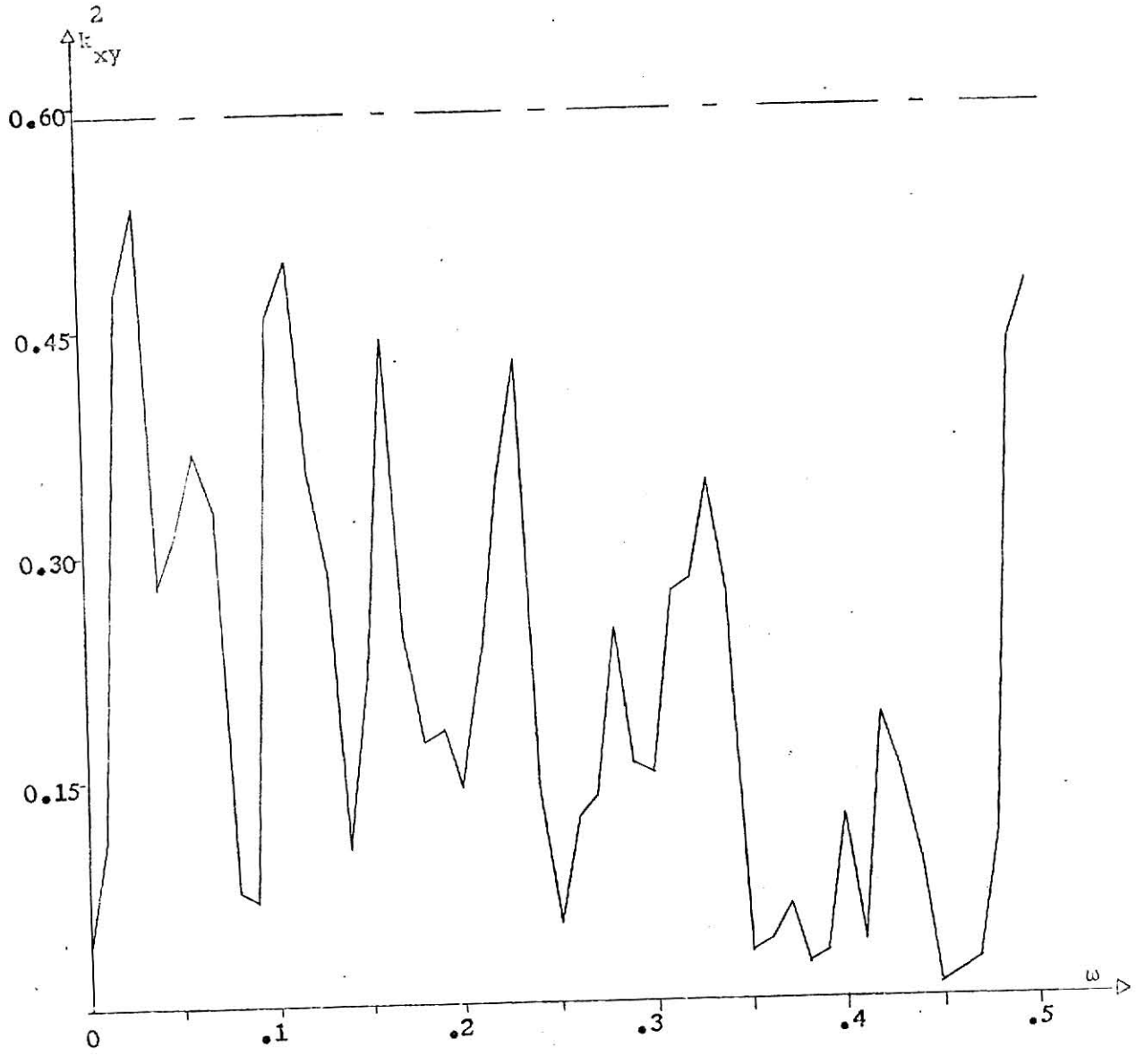

FIGURE 11. Squared coherency vs. frequency, sulphur dioxide from stations 79 and 60.

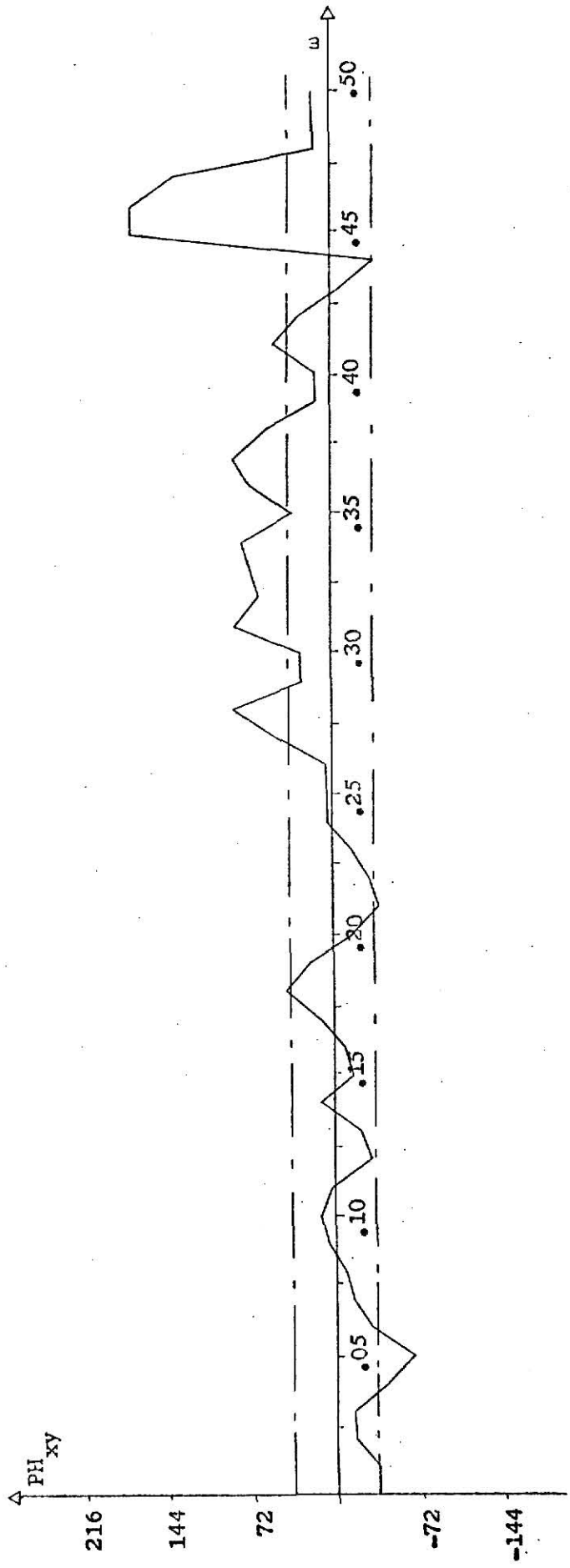

FIGURE 12. Phase vs. frequency, sulphur dioxide from stations 79 and 60.

Sulphur Dioxide at Stations 1 and 79

The cross covariance plot Fig. 13 indicates that there is a high correlation at lags 11 days and 30 days. There is rather a flat peak around the lag of 26 days. The coherence plot is shown in Fig. 14. A number of peaks in the coherence diagram indicate correlation at the cyclic frequencies. This correlation is however insignificant except at the frequencies of .36 c/day and .38 c/day. The phase diagram, Fig. 15, reveals a downward trend in the frequency range of 0 to .41 c/day indicating a time lag of one day between the series. There is a time lag of about 11 days around the frequency corresponding to 7 days.

Sulphur Dioxide at Stations 71 and 1

The cross covariance plot for the sulphur dioxide from stations 71 and 1 is shown in Fig. 16. The sulphur dioxide at these stations appears to be correlated most prominently at lags of 6 days and 30 days. The coherency plot presented in Fig. 17 reveals a high coherence at low frequencies in the range from 0 to 0.03 c/day. The coherence is comparatively high at the cyclic frequencies though it is found to be insignificant. There isn't any phase difference between the sulphur dioxide emitted at stations 71 and 1 except for a peak near the frequency of .235 c/day as shown in Fig. 18. In fact in the frequency range of .17 c/day to .23 c/day the sulphur dioxide at station 71 leads sulphur dioxide at station 1 by about 9 days.

Sulphur Dioxide and Nitrogen Oxides at Station 1

The cross covariance function, Fig. 19, has a number of peaks indicating correlations at various cycles. There are rather flat peaks

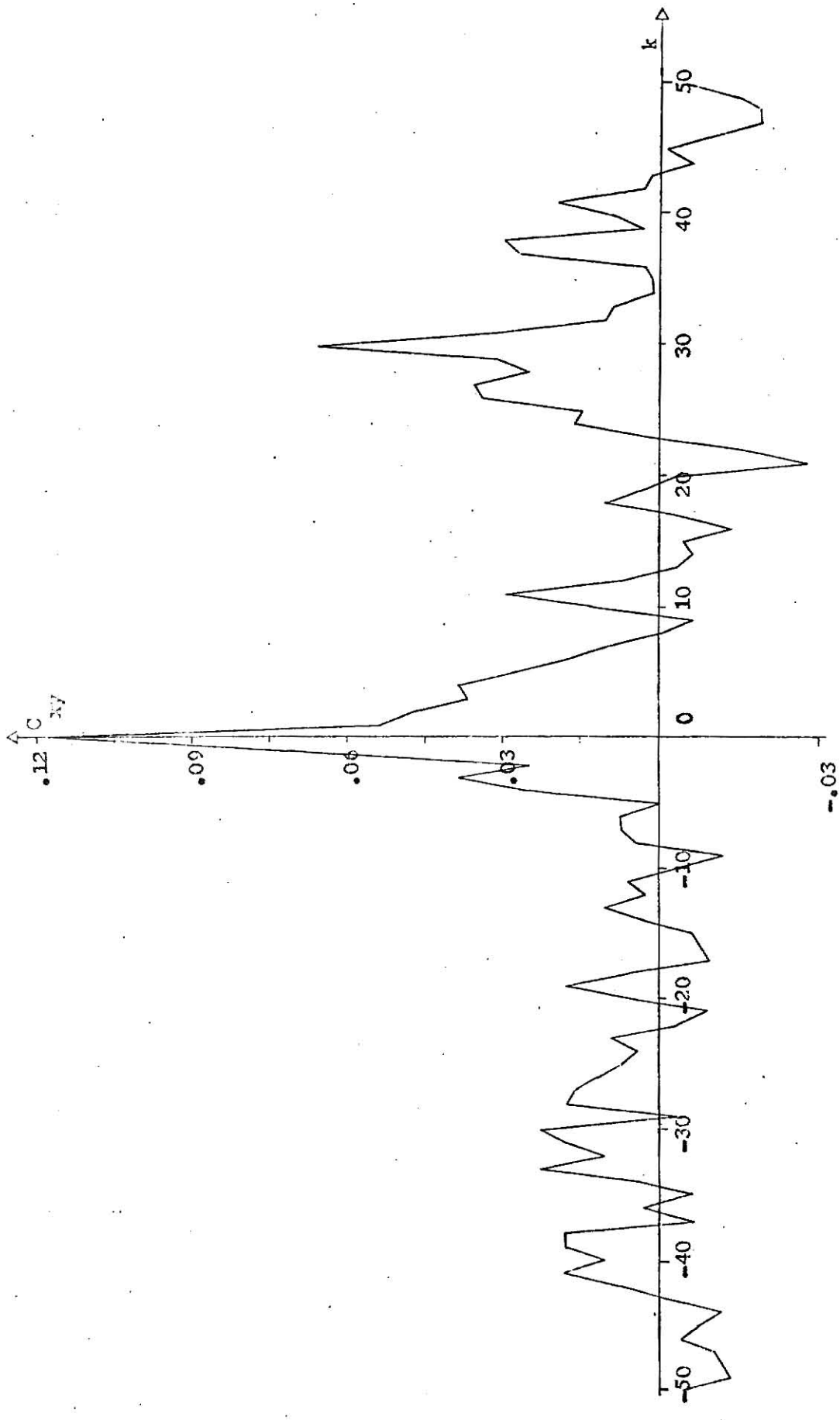

FIGURE 13. Cross covariance of sulphur dioxide for stations 1 and 79.

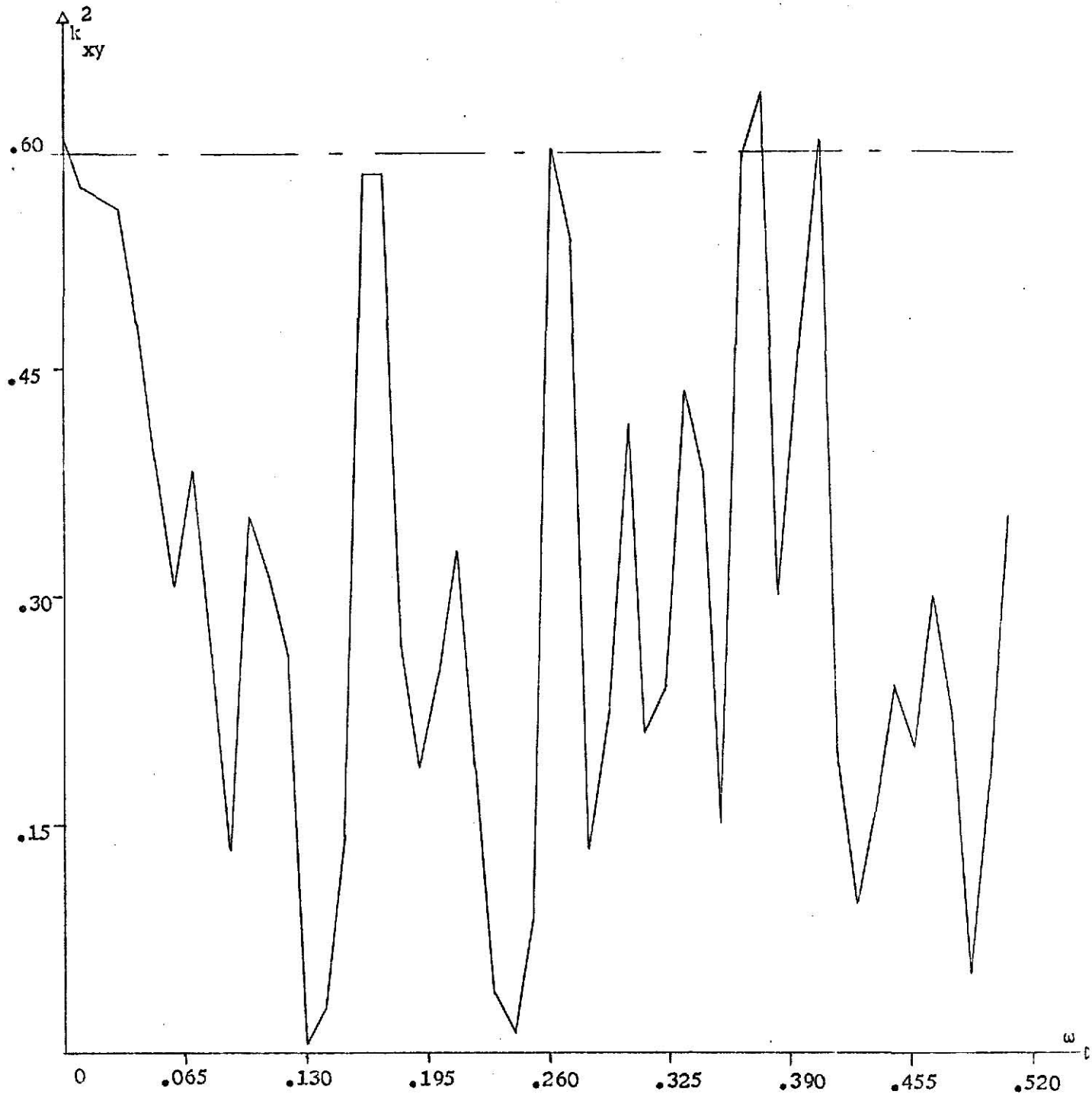

FIGURE 14. Squared coherency vs. frequency, sulphur dioxide from stations 1 and 79.

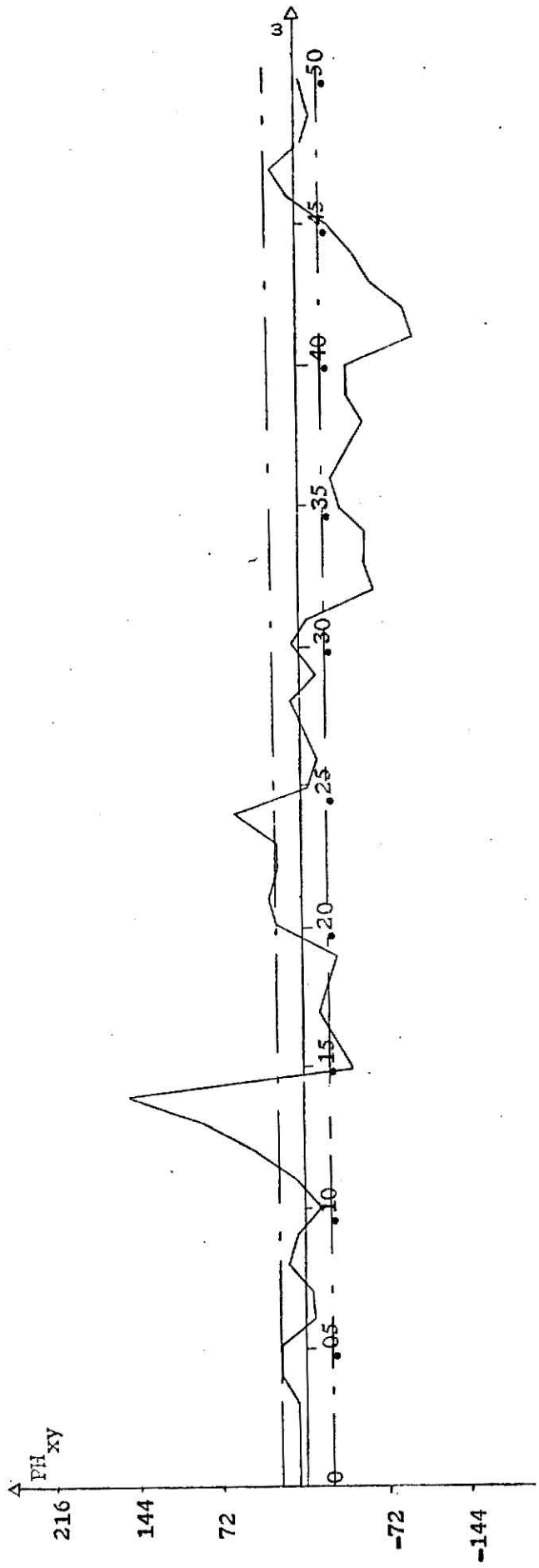

FIGURE 15. Phase vs. frequency, sulphur dioxide from stations 1 and 79.

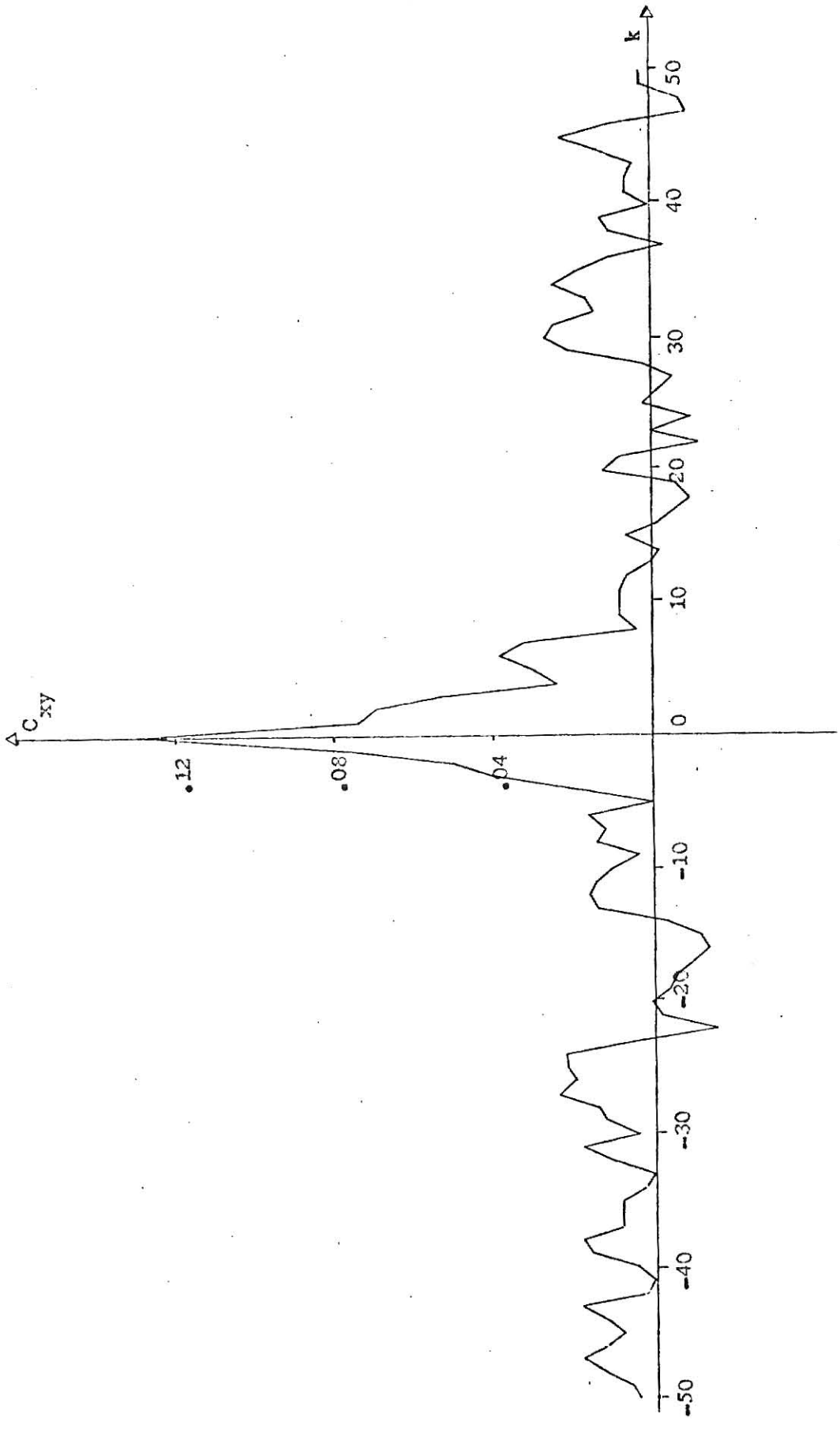

FIGURE 16. Cross covariance of sulphur dioxode for stations 71 and 1.

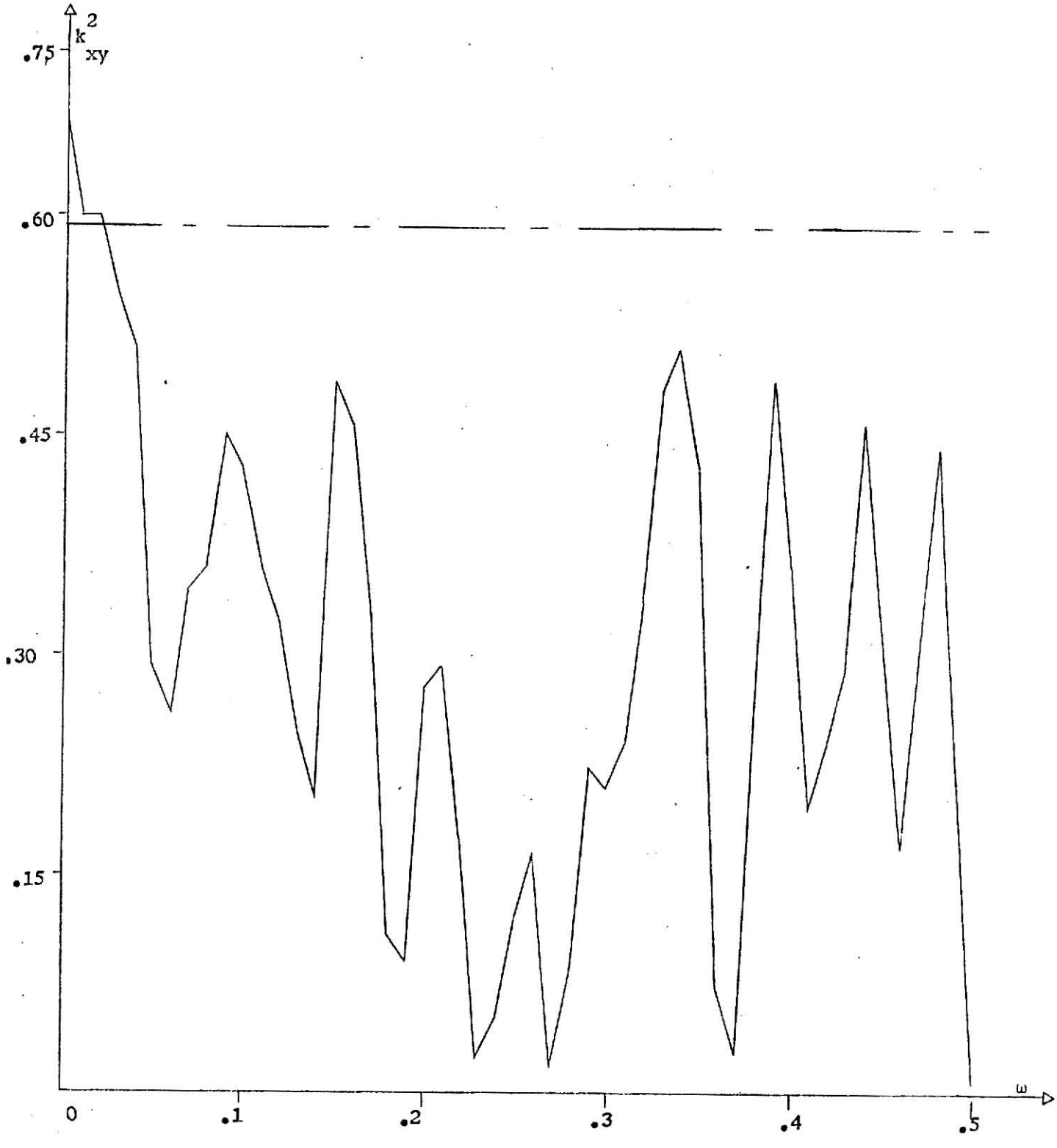

FIGURE 17. Squared coherency vs. frequency, sulphur dioxide from stations 71 and 1.

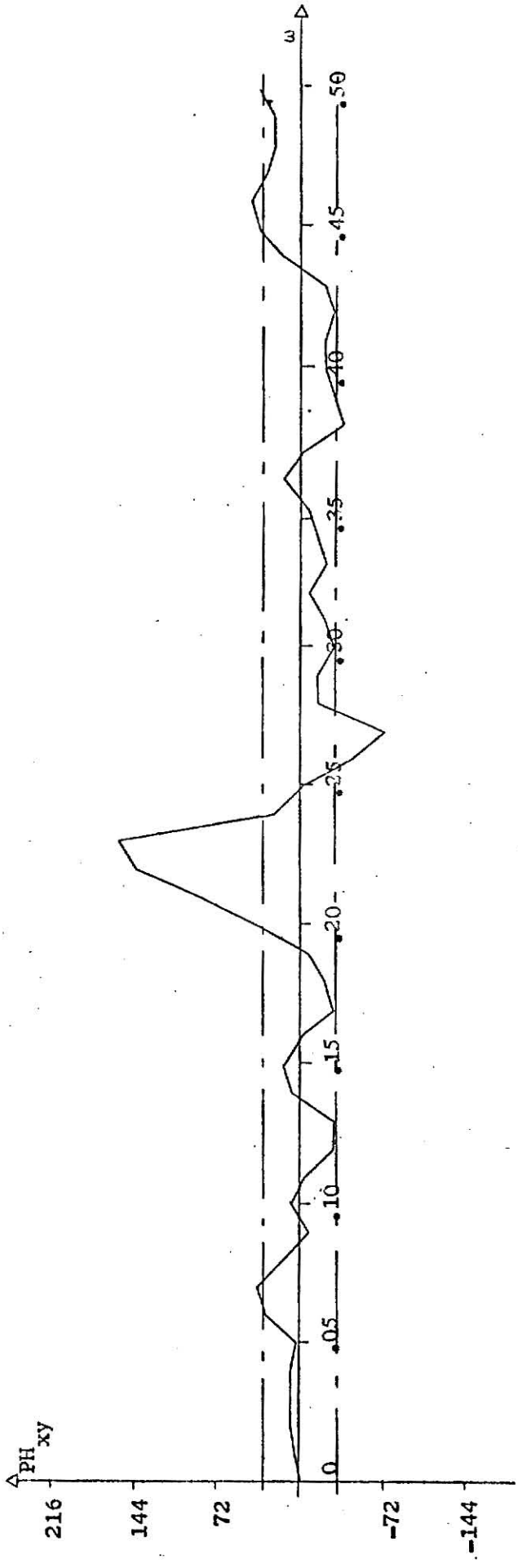

FIGURE 18. Phase vs. frequency, sulphur dioxide from stations 71 and 1.

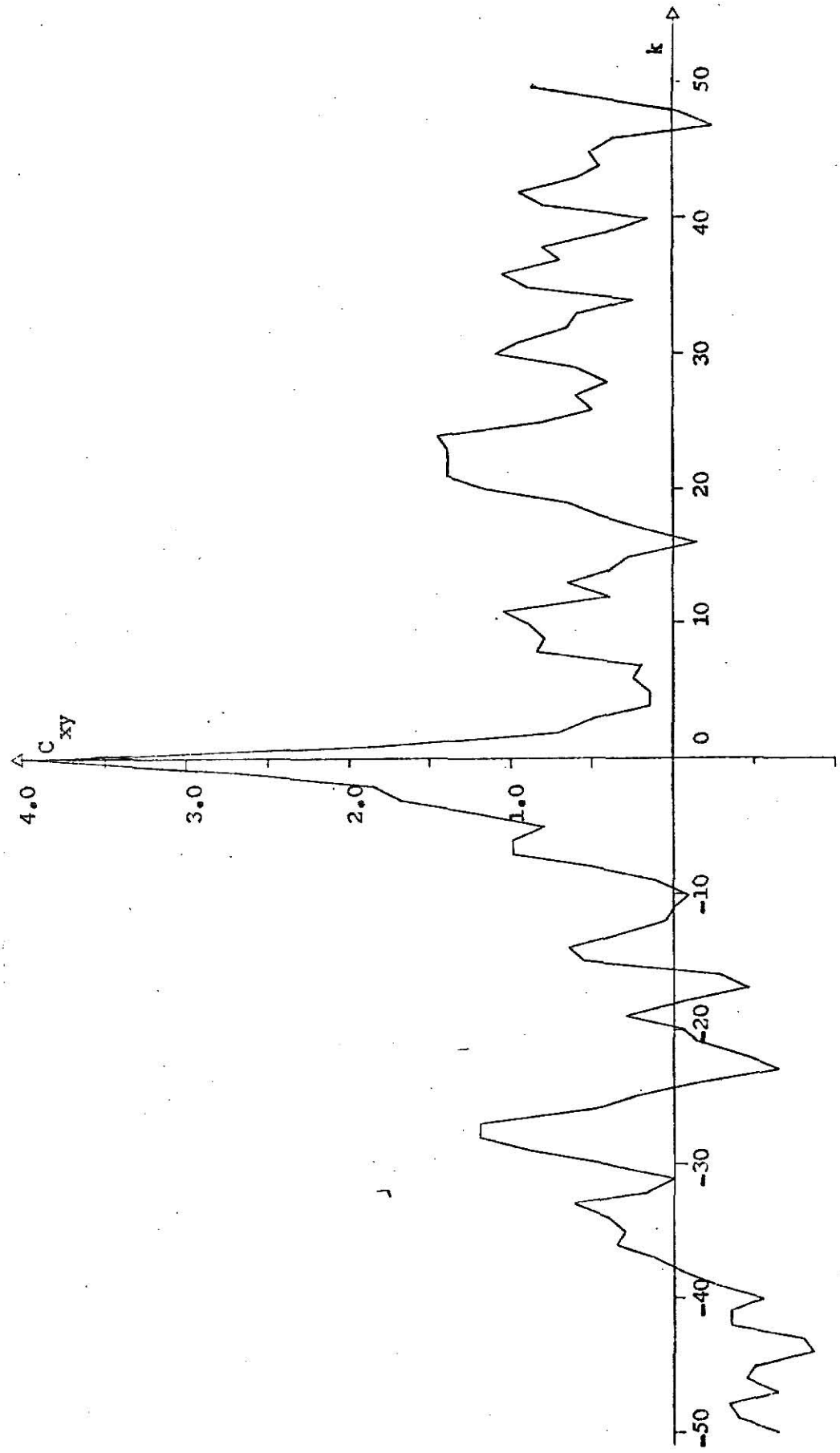

FIGURE 19. Cross covariance of sulphur dioxide and nitrogen oxides for station 1.

around lags of 10 days and 22 days and sharp peaks at lags of 30 days, 36 days and 42 days. Coherence Fig. 20, indicates high coherence in general between sulphur dioxide and nitrogen oxide which shows that they may be emitting from the common sources. The coherence between these two is 0.8 at frequency of about .08 c/day and 0.75 at frequency of about .37 c/day. The phase plot is shown in Fig. 21. There is no phase difference between these two pollutants at station 1 upto a frequency of .415 c/day. There is a sudden rising pattern after the frequency of .415 c/day indicating a time lag of about 22 days over the frequency range .42 c/day to .47 c/day.

Carbon Monoxide and Sulphur Dioxide at Station 1

The cross correlation plot for these two stations is shown in Fig. 22. Quite significant peaks are observed at lags of 7 days, 14 days and 28 days. Coherence plot, Fig. 23, indicates high coherence at frequencies of .04 c/day, .08 c/day, .11 c/day and .37 c/day. The phase, Fig. 24, is in general clustered about the zero axis with no marked trends, except some small peaks around .08 c/day, .18 c/day and .42 c/day, indicating practically no phase difference in the carbon monoxide and sulphur dioxide concentrations. This implies that there are in general common sources of their emission.

Carbon Monoxide and Nitrogen Oxides at Station 1

The cross covariance plot for this combination is shown in Fig. 25. The cross correlation dies down soon from its maximum at zero lag. There are low and flat peaks around lags of 12 days and 24 days. The coherence

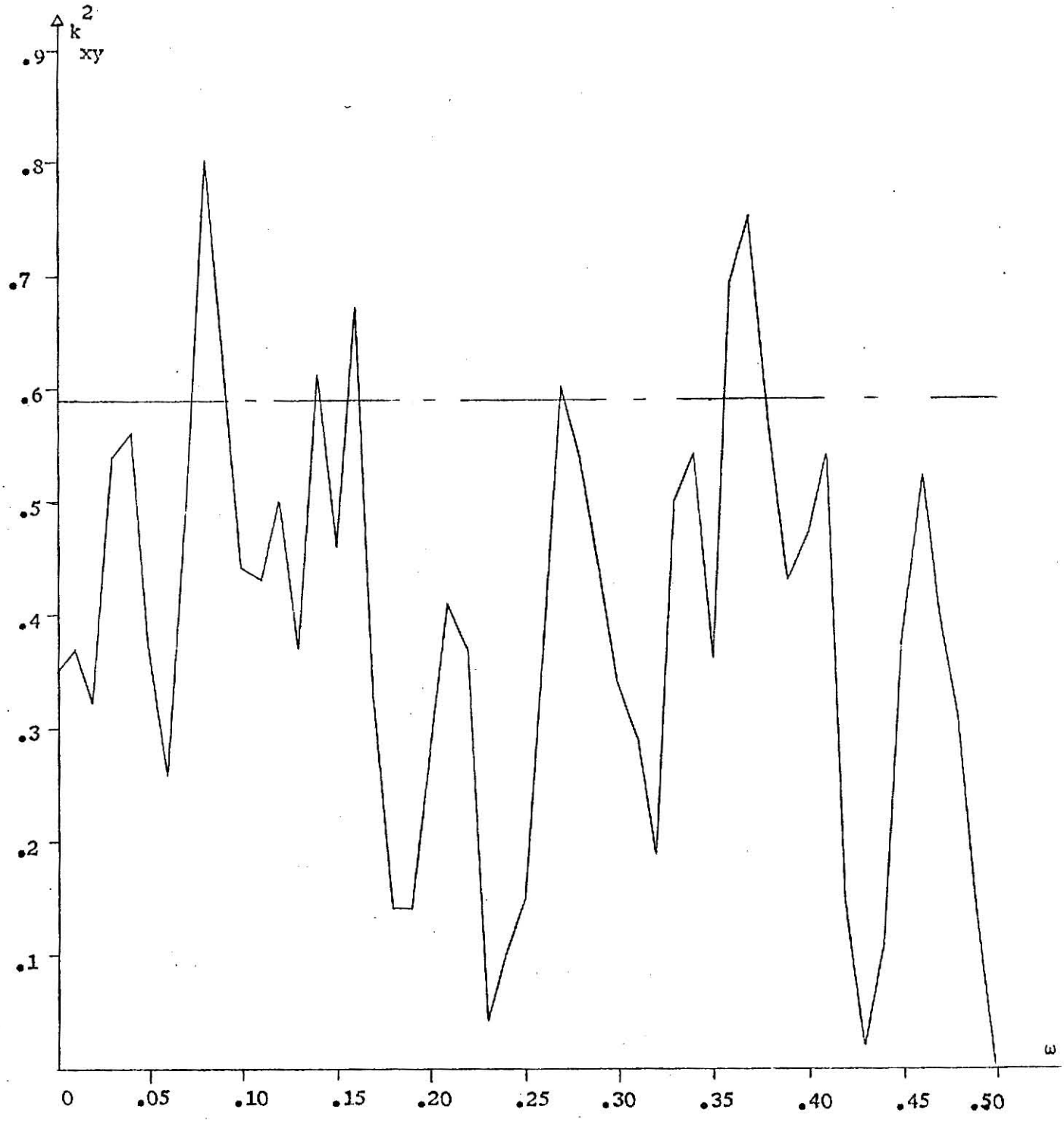

FIGURE 20. Squared coherency vs. frequency, sulphur dioxide and nitrogen oxides from station 1.

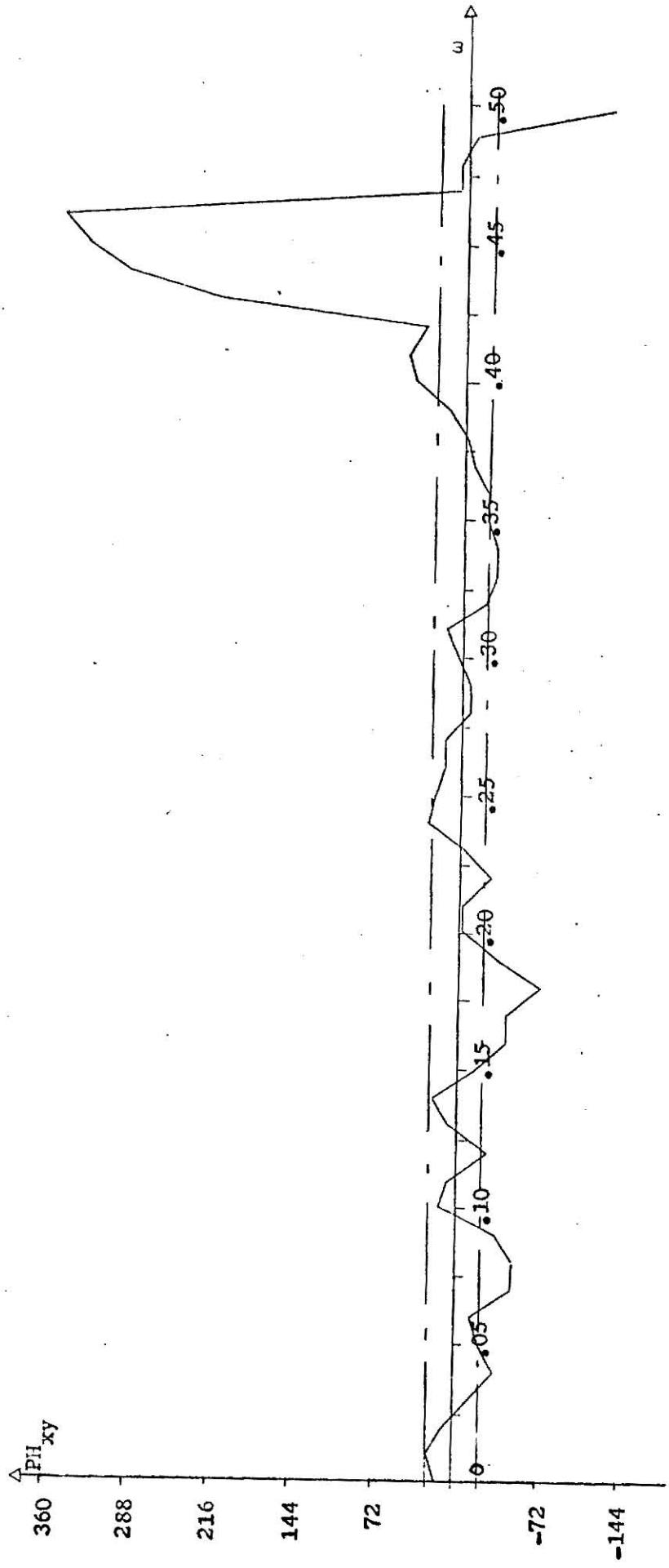

FIGURE 21. Phase vs. frequency, sulphur dioxide and nitrogen oxides from station 1.

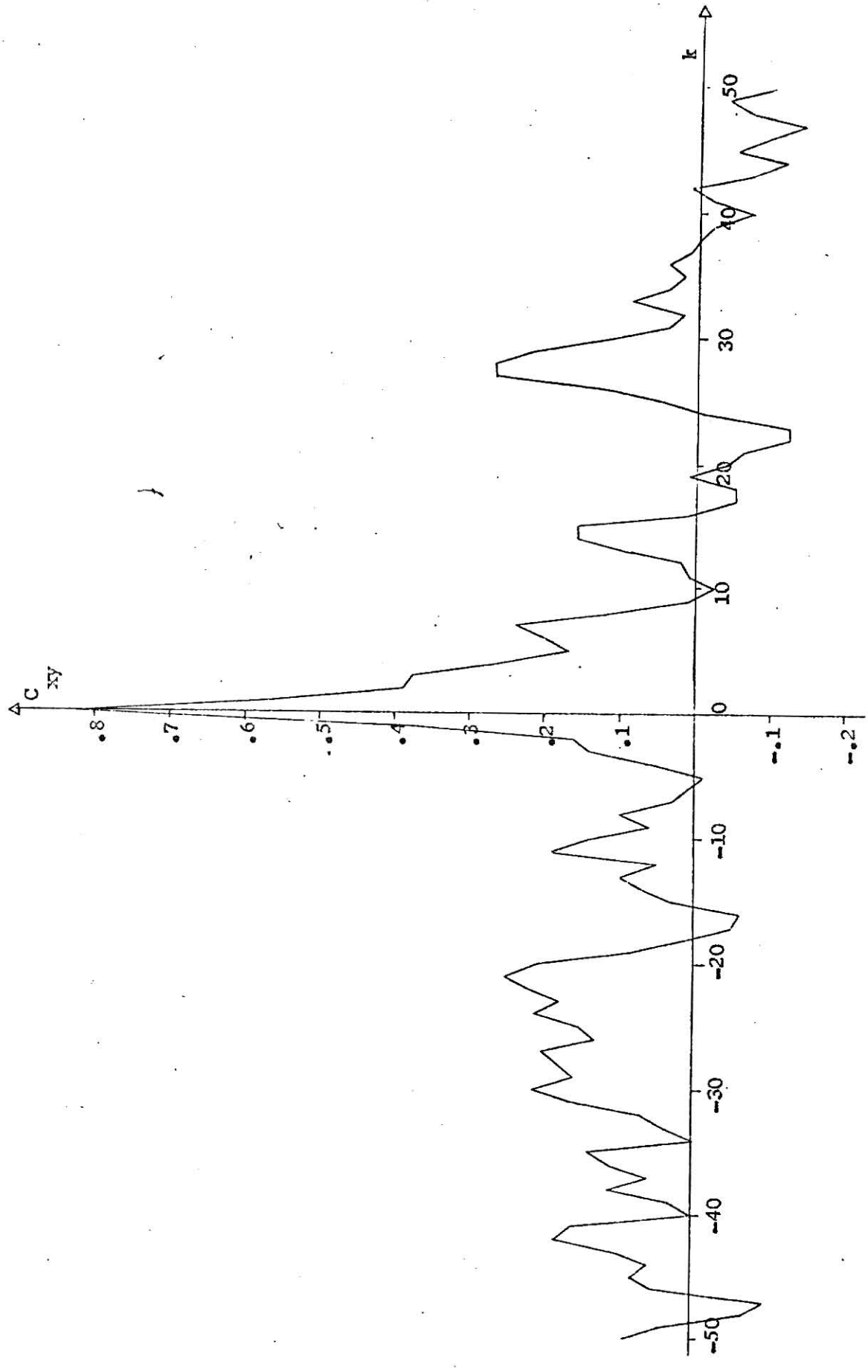

FIGURE 22. Cross covariance of carbon monoxide and sulphur dioxide for station 1.

THIS BOOK CONTAINS NUMEROUS PAGES THAT ARE CUT OFF

THIS IS AS RECEIVED FROM THE CUSTOMER

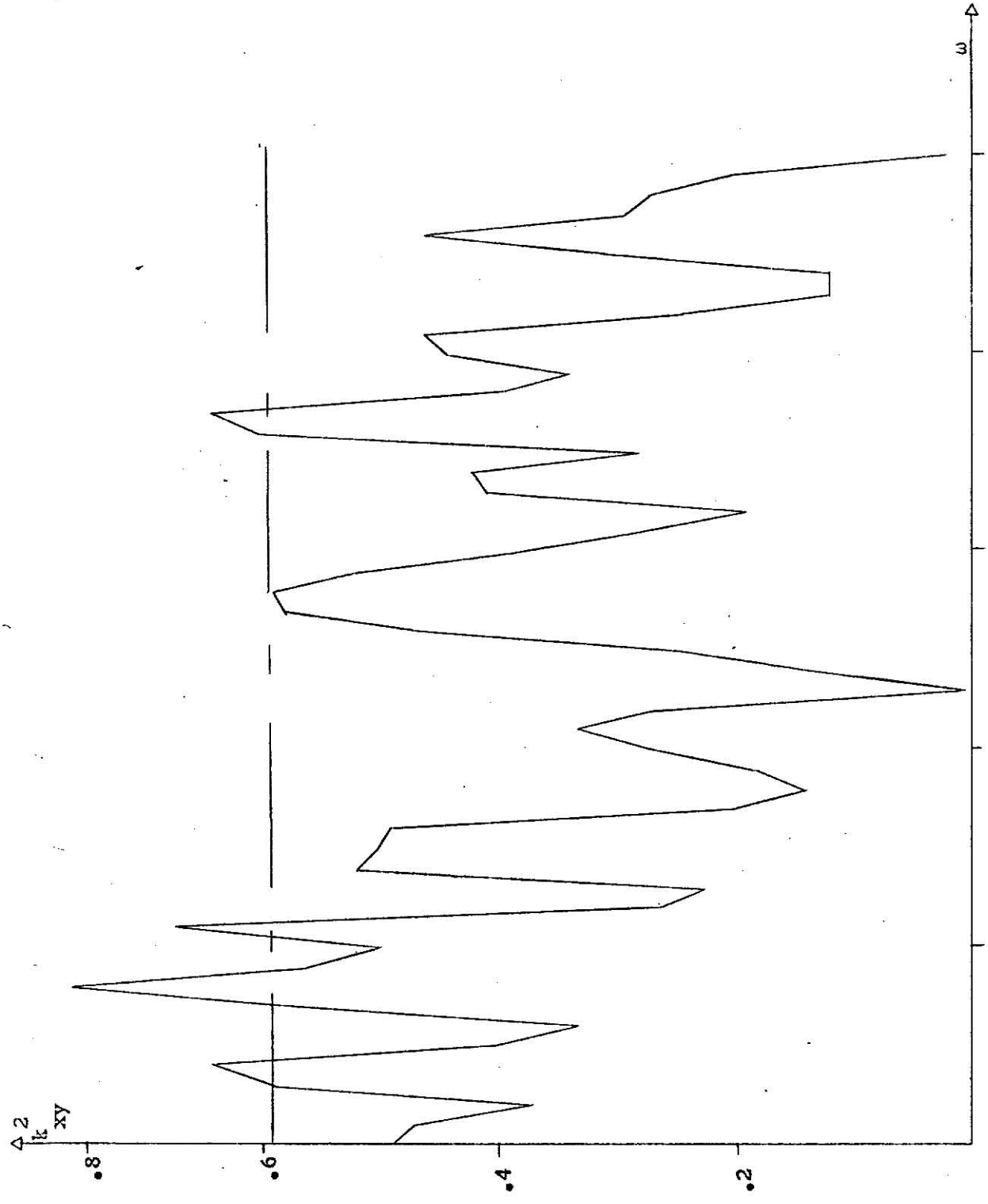

FIGURE 23. Squared coherency vs. frequency, carbon monoxide and sulphur dioxide from station 1.

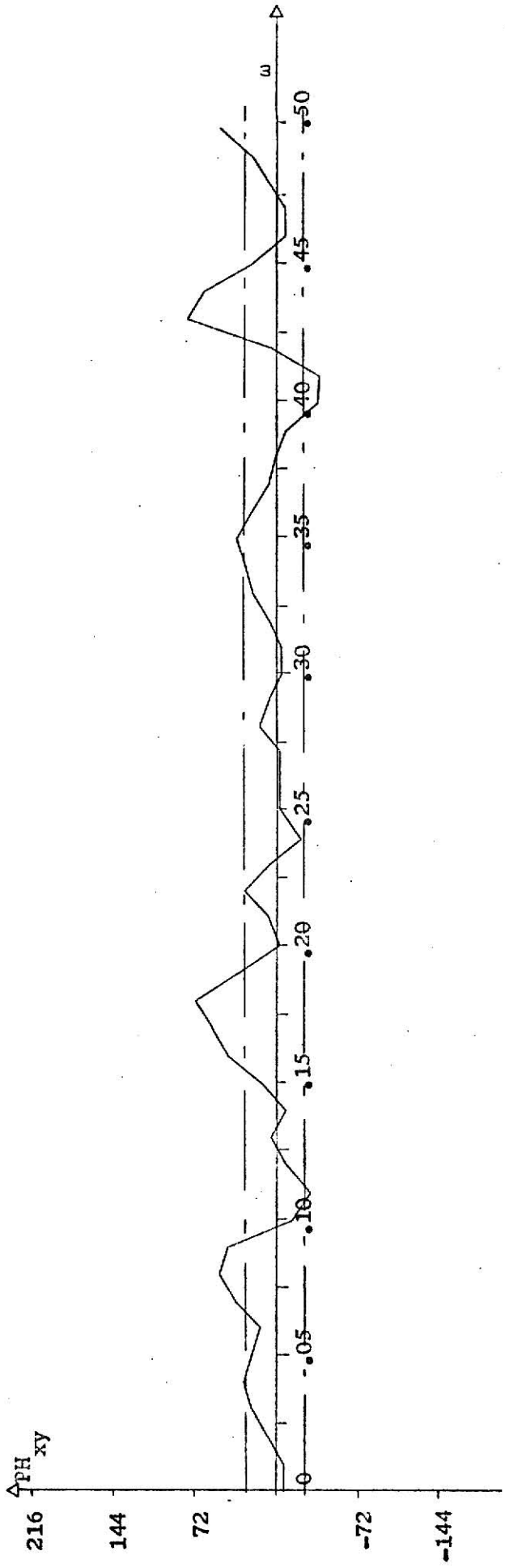

FIGURE 24. Phase vs. frequency, carbon monoxide and sulphur dioxide from station 1.

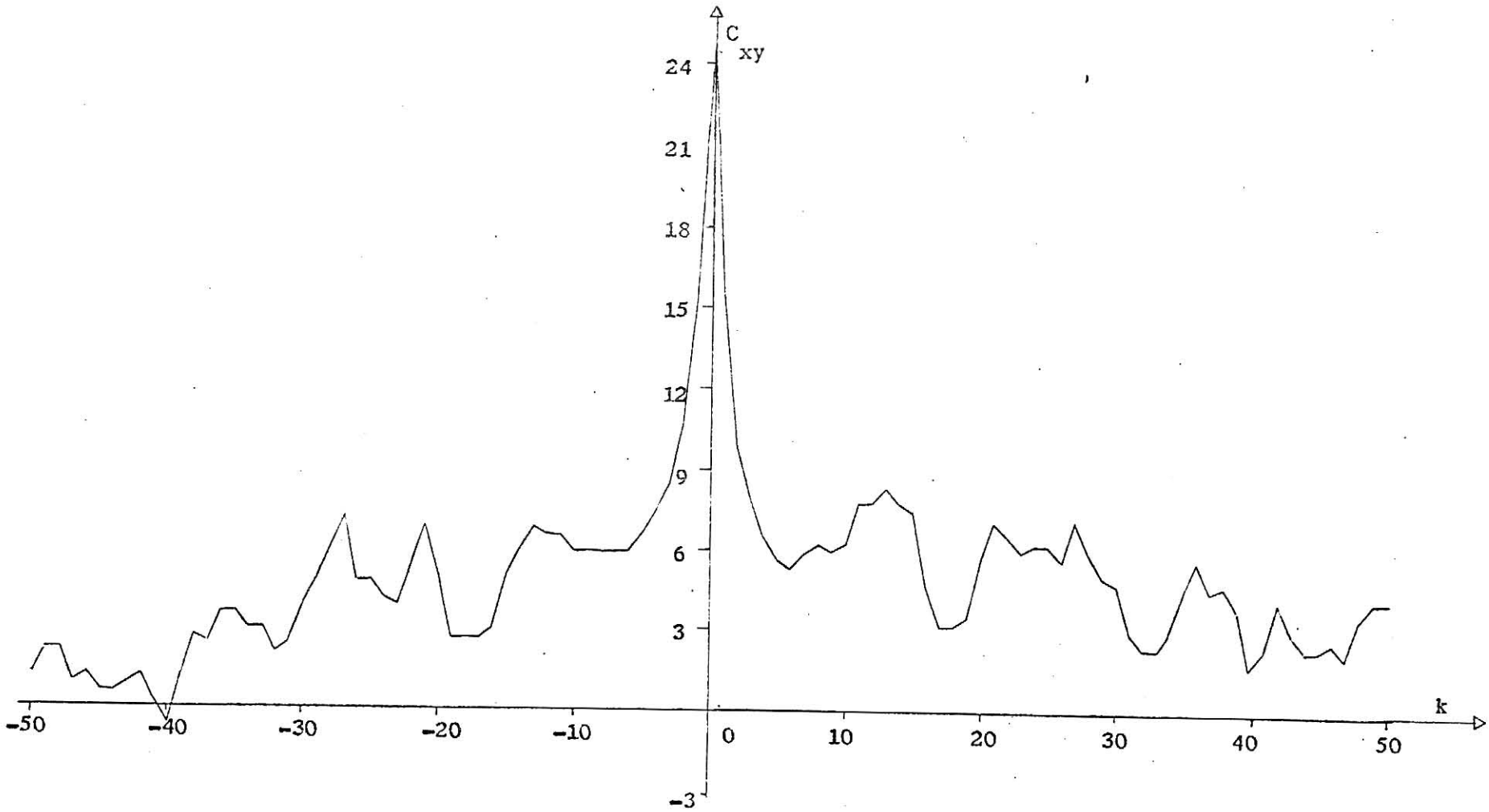

FIGURE 25. Cross covariance of carbon monoxide and nitrogen oxides from station 1.

plot, Fig. 26, reveals a very high coherence at all the frequencies between the carbon monoxide and nitrogen oxides concentration. In general the coherence is found to be above 0.8 over all the frequencies. Also the phase diagram, Fig. 27, reveals a zero phase lag between these two pollutants. This indicates that the carbon monoxide and nitrogen oxides cyclic variations are atmost the same which reflects that in general, at station 1, the source distribution is such that the carbon monoxide and nitrogen oxides are emitted simultaneously and they are equally important.

Ozone and Sulphur Dioxide at Station 1

There appears to be a significant correlation at lag ranges of 24 - 28 days, and 44-47 days and at lag of 38 days, shown in Fig. 28. Fig. 29 reveals that there is insignificant coherence between ozone and sulphur dioxide over all the frequencies. The phase diagram, Fig. 30, indicates the phase mostly within the confidence interval except at few frequencies. There is a phase difference of about 90^{o} in the frequency range of 0 - .01 c/day; of about 144^{o} at frequency of .14 c/day and 90^{o} at the frequency of .44 c/day.

Ozone and Carbon Monoxide and Nitrogen Oxides at Station 1

It is seen from Fig. 31 and Fig. 34, the cross correlation plots for ozone and carbon monoxide and ozone and nitrogen oxides respectively, that the correlation has an increasing trend with the increasing number of lags. Also the coherence plots for these two cases shown in Fig. 32 and Fig. 35, indicate insignificant coherence nearly over all the frequencies. The phase plots Fig. 33 and Fig. 36 reveal a phase difference

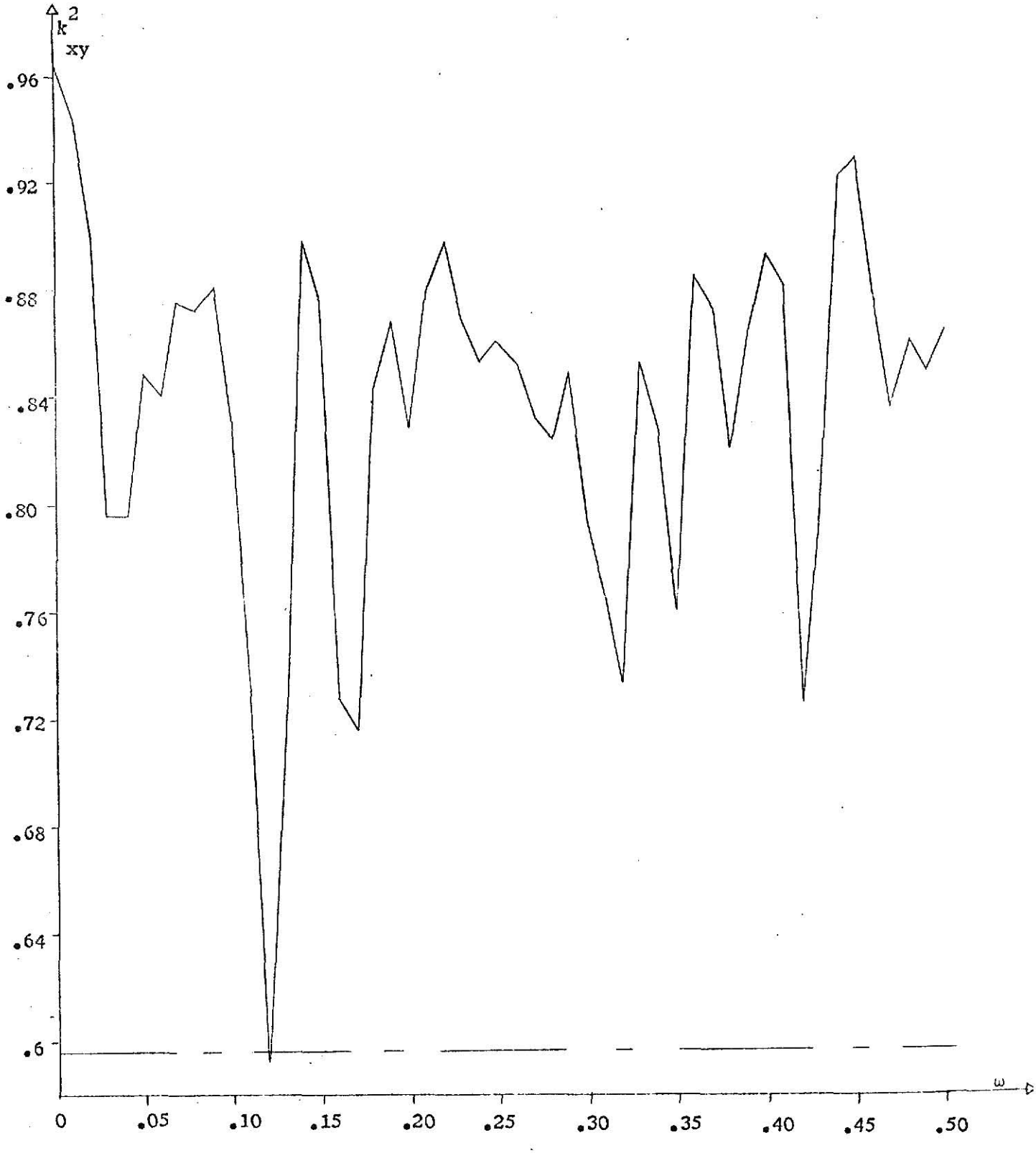

FIGURE 26. Squared coherency vs. frequency, carbon monoxide and nitrogen oxides from station 1.

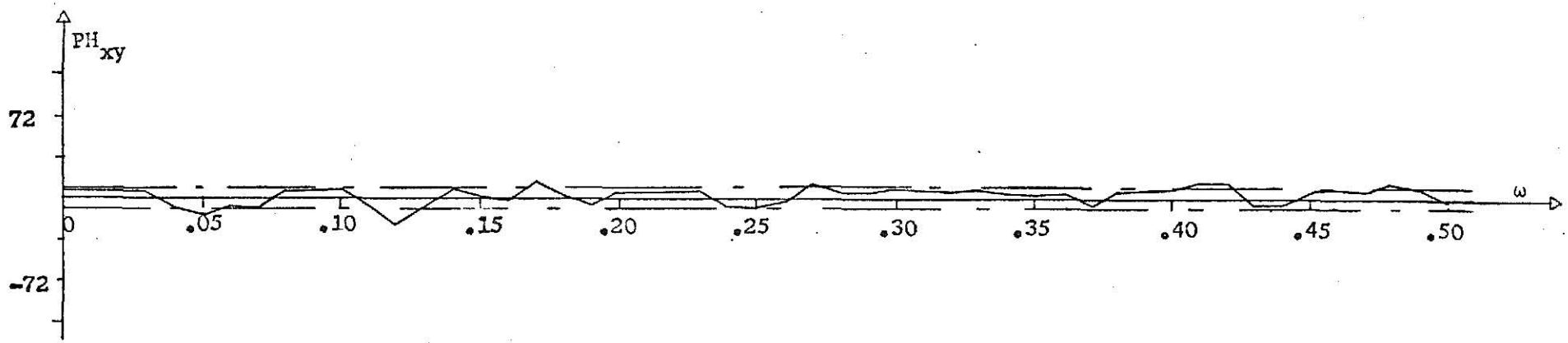

FIGURE 27. Phase vs. frequency, carbon monoxide and nitrogen oxides from station 1.

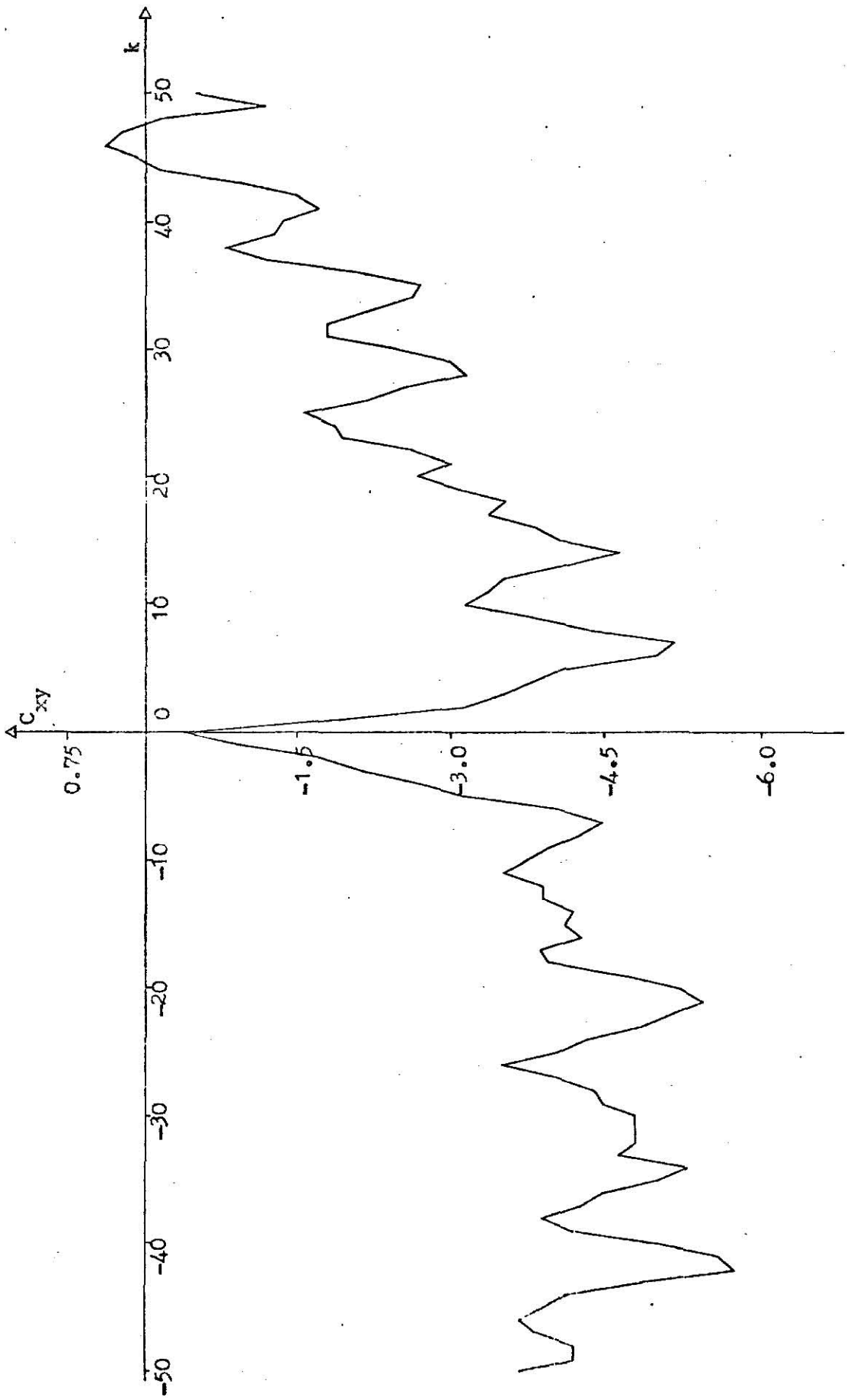

FIGURE 28. Cross covariance of ozone and nitrogen oxides for station 1.

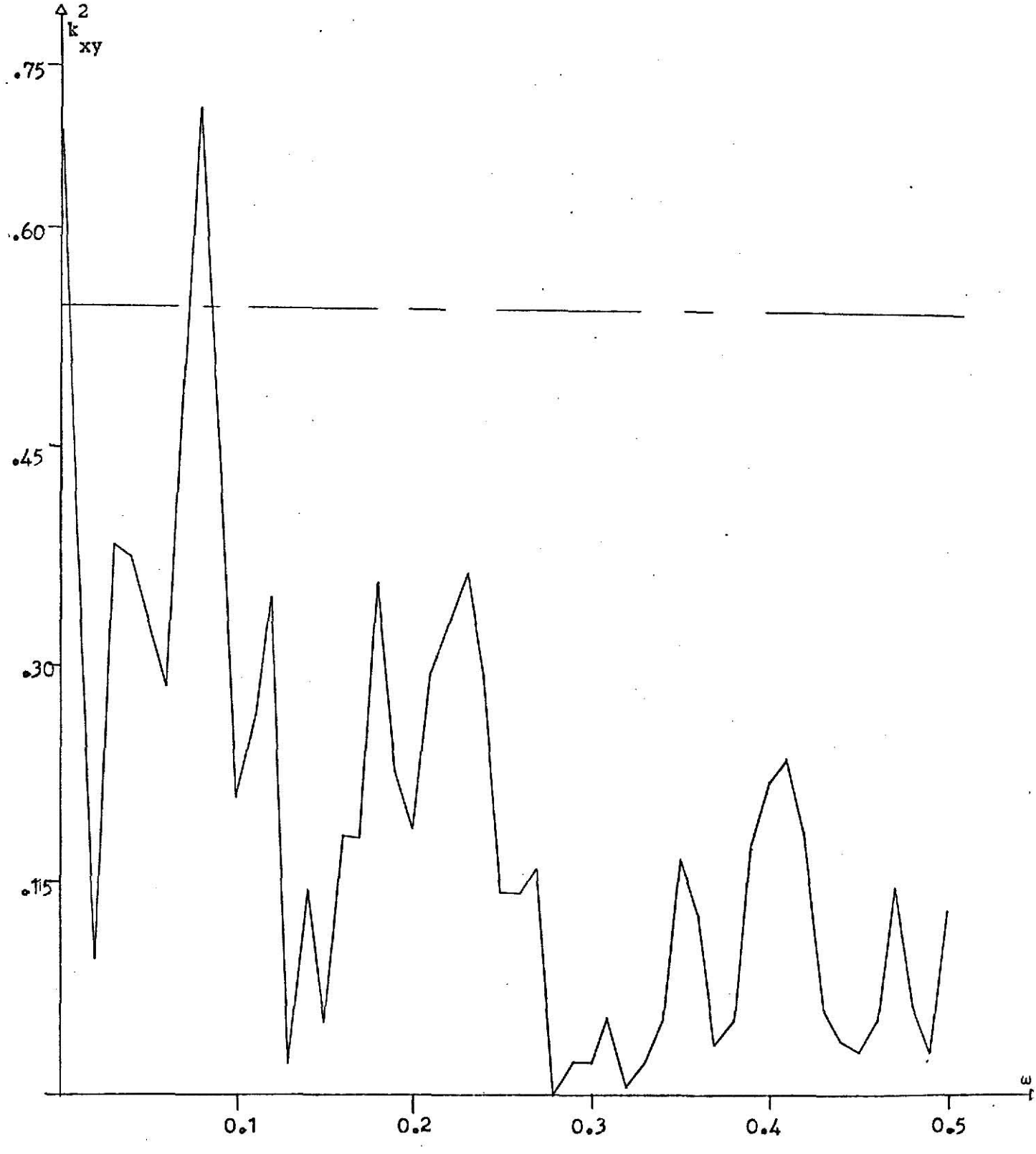

FIGURE 29. Squared coherency vs. frequency, ozone and nitrogen oxides from station 1.

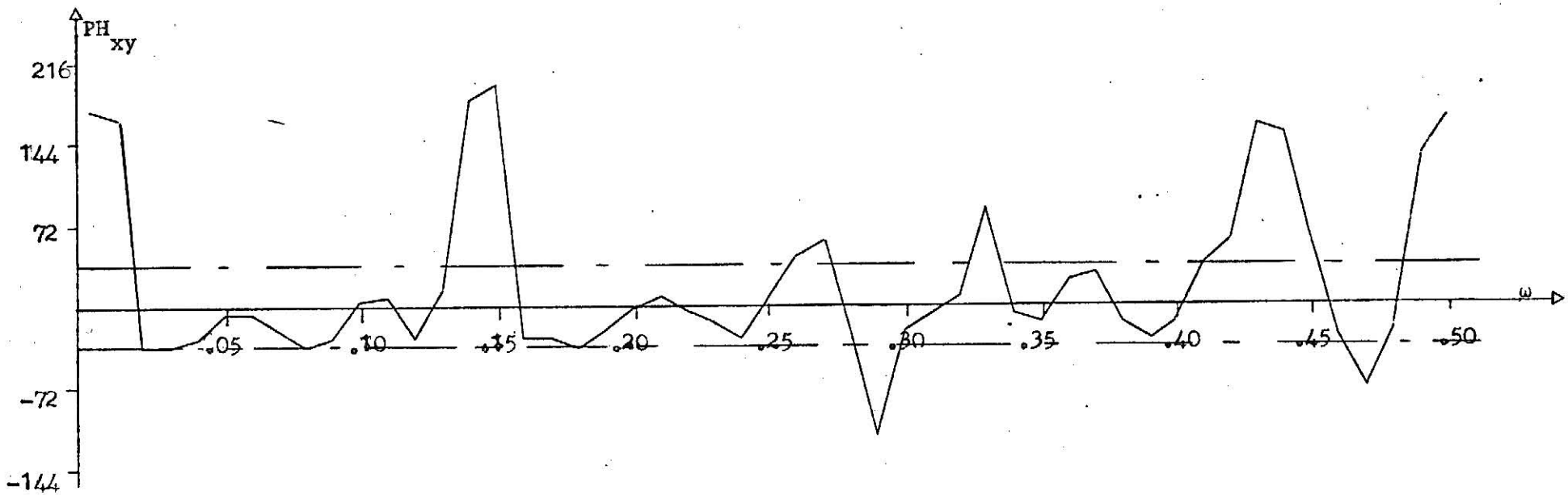

FIGURE 30. Phase vs. frequency, ozone and sulfur dioxide nitrogen oxides from station 1.

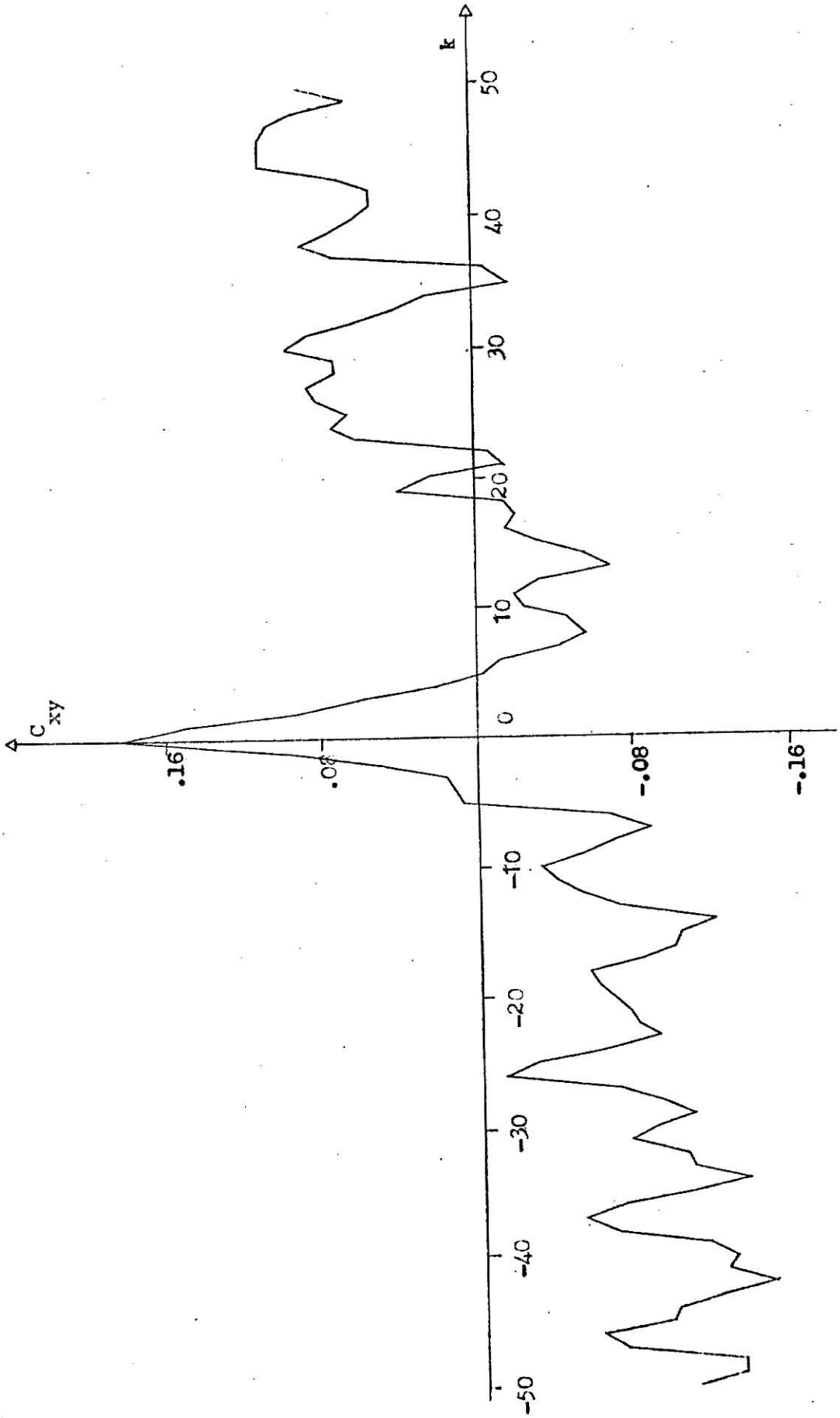

FIGURE 31. Cross covariance of ozone and sulphur dioxide for station 1.

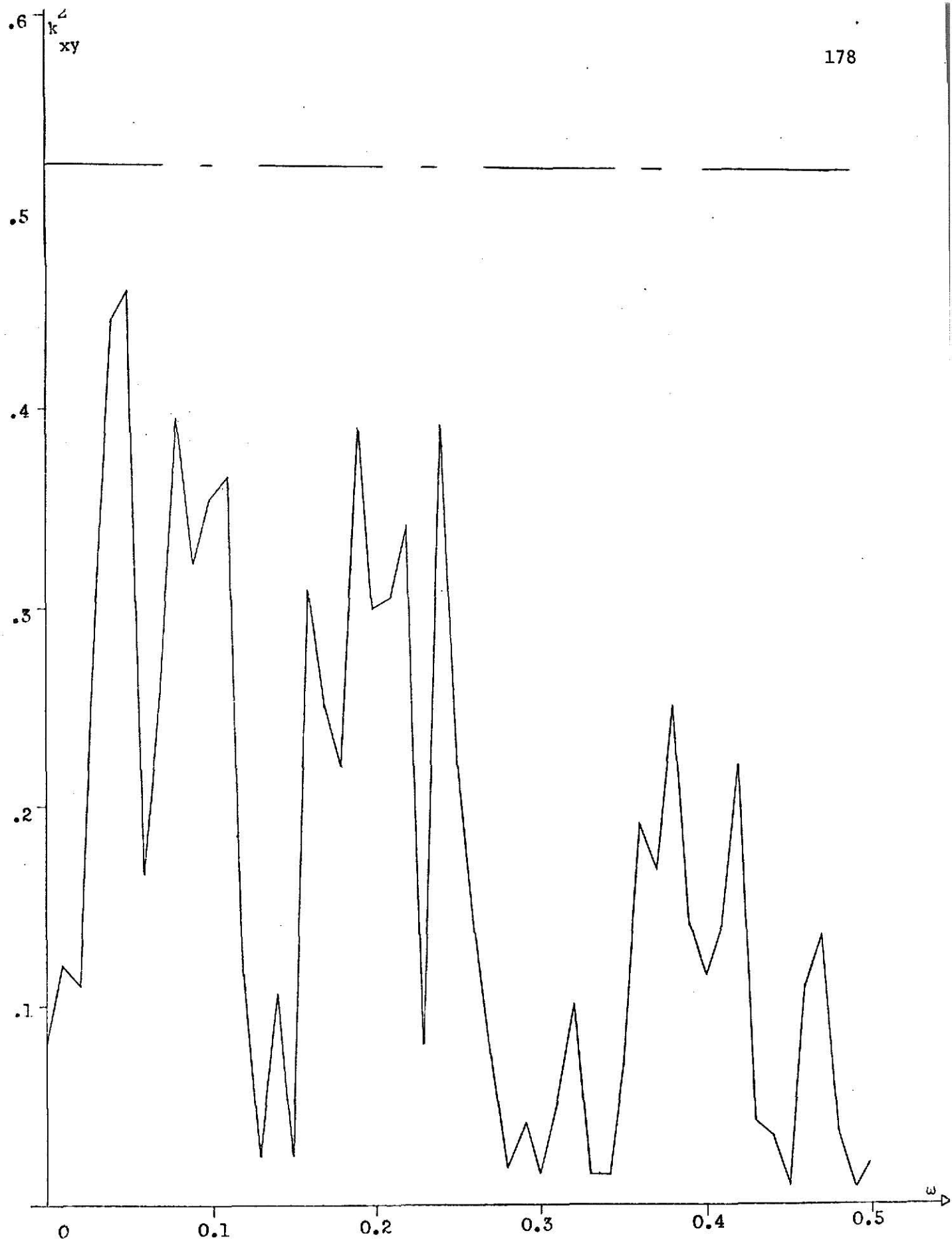

FIGURE 32. Squared coherency vs. frequency, ozone and sulphur dioxide from station 1.

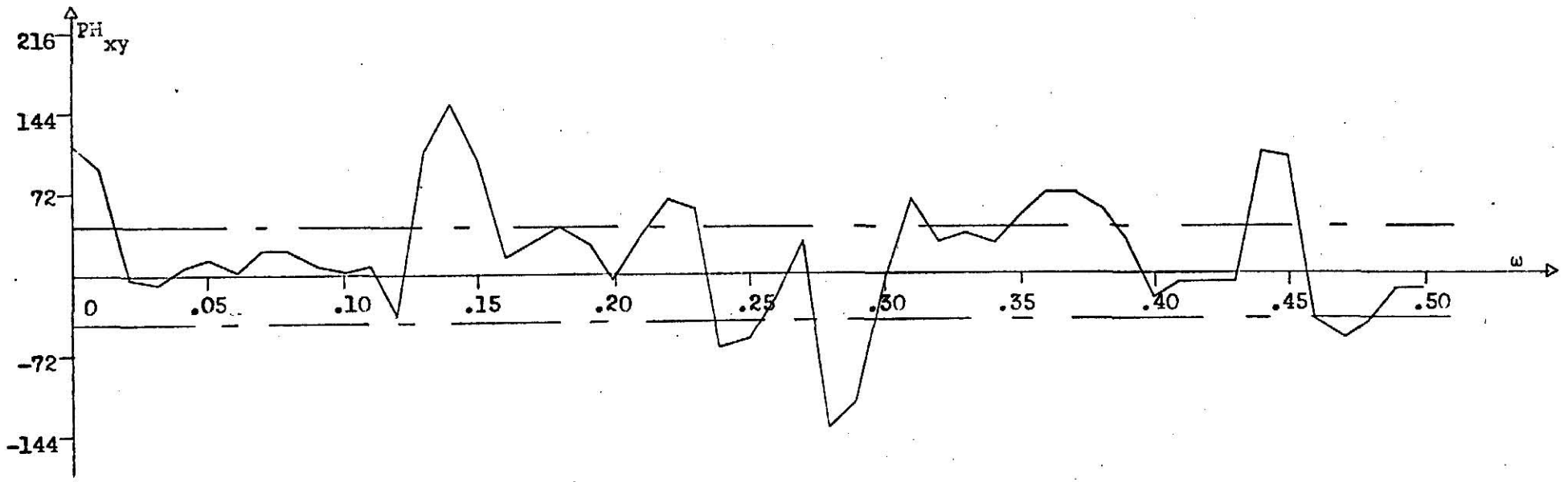

FIGURE 33. Phase vs. frequency, ozone and sulphur dicxide from station 1.

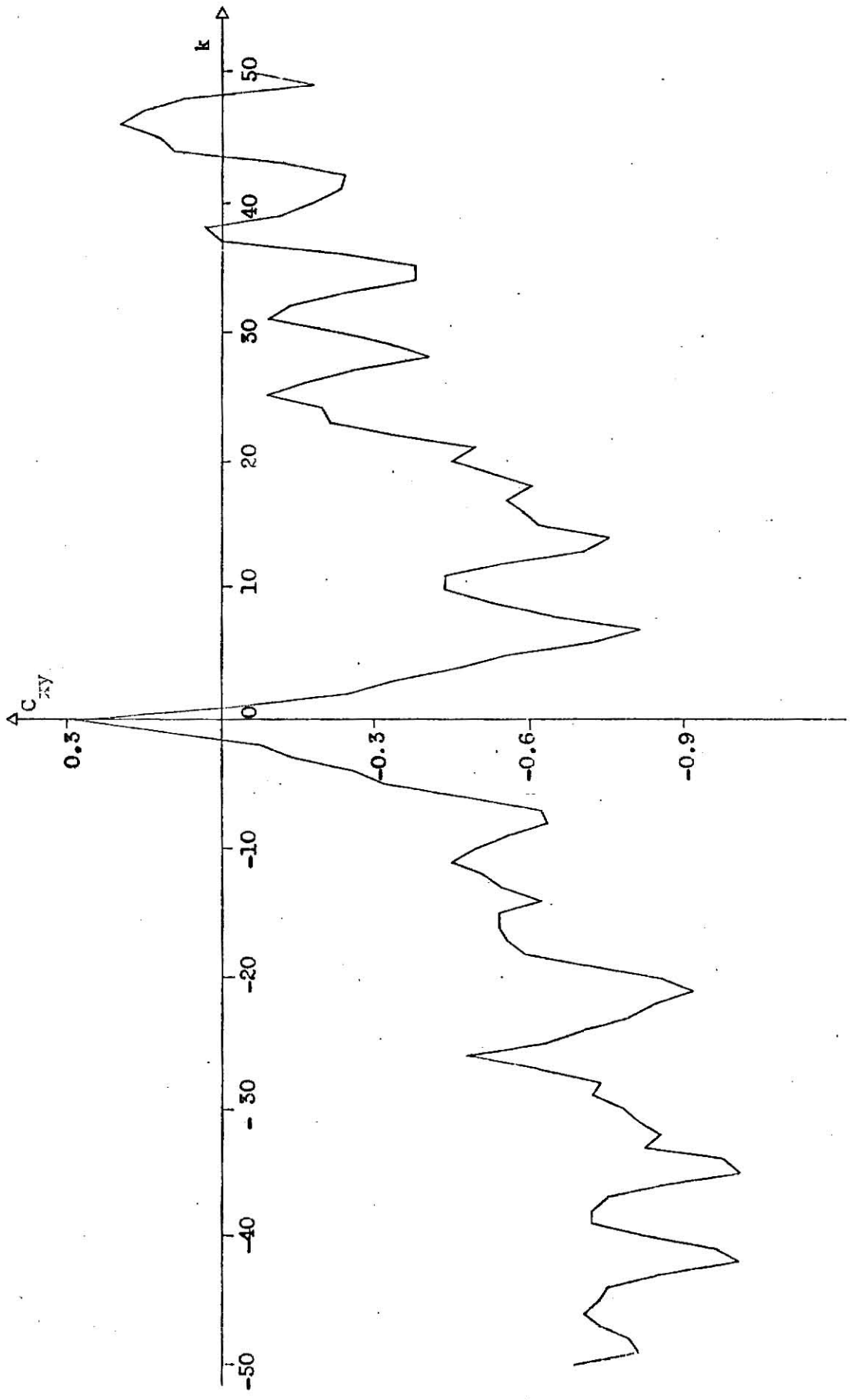

FIGURE 34. cross ccvariance of ozone and carbon mcnoxide for station 1.

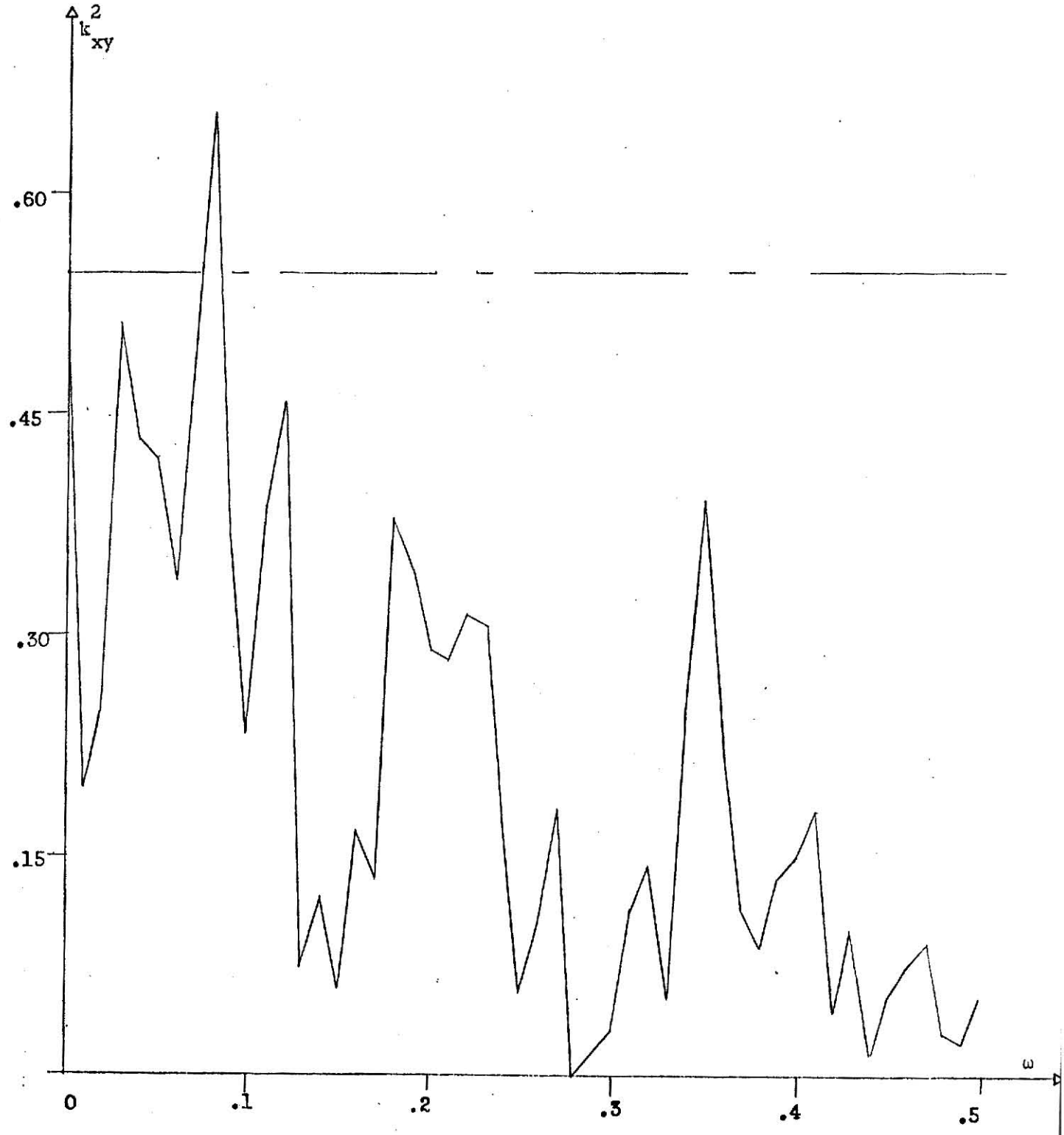

FIGURE 35. Squared coherency vs. frequency, ozone and carbon monoxide from station 1.

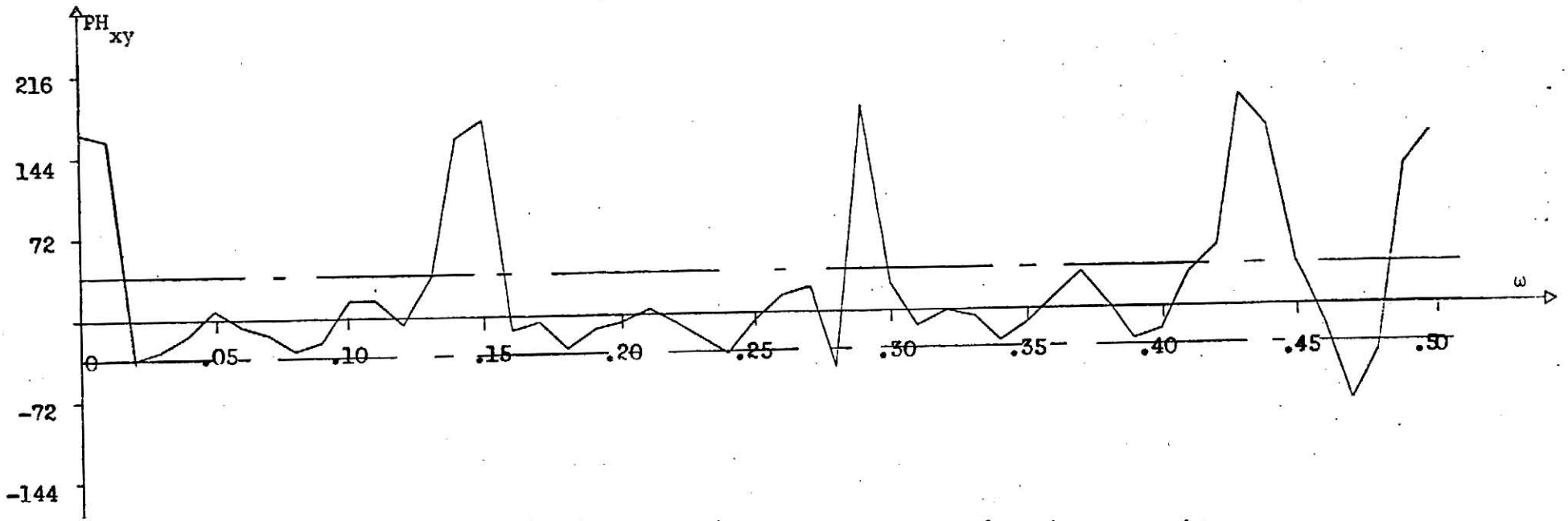

FIGURE 36. Phase vs, frequency, ozone and carbon monoxide from station 1.

of about 150^{o} in the frequency range of 0 - .01 c/day and .14 - .14 c/day and a phase of 180^{o} at .29 c/day for the former and a phase difference of about 180^{o} at .43 c/day frequency for the later. It's insignificant over all other frequencies.

Generally it is seen that the coherence between ozone concentration and other pollutants concentration is insignificant. This may be due to the fact that the ozone and other pollutants have completely different sources of emission. Ozone is formed from chemical reactions while other pollutants are emitted from physical sources.

Cross Spectral Analysis Among Different Pollutants at Station 60

The results are summarized in Table D for all the combinations of the pollutants sulphur dioxide, nitrogen oxides, carbon monoxide and ozone. The graphical representations of the cross covariance function, phase and coherence are not shown because mostly they are the same as that at station 1. The differences are discussed below.

Comparing these results to the ones obtained at station 1 it is seen that at station 60 the sulphur dioxide and nitrogen oxides are also emitted simultaneously, while the coherence shows that the frequencies simultaneous emissions of these two pollutants at these two stations are different. The coherence between carbon monoxide and nitrogen oxides is also high at station 60 but it is comparatively lesser than that at station 1 which may indicate that there may not be as many common sources of carbon monoxide and sulphur dioxide at station 60 as is found at station 1. Coherence between carbon monoxide and sulphur dioxide indicates that at station 60 there are predominantly different sources of

carbon monoxide and sulphur dioxide while it is not so at station 1. The cross spectral analysis of ozone with other pollutants also reveals that the coherence is insignificant for all thus indicating different sources of emissions of ozone and other pollutants.

Cross Spectral Analysis Between Sulphur Dioxide and Wind Speed at all the Stations

This analysis was carried out as an examplar case to show how the wind speed affects the pollutant concentration at all the four stations. Six hours averaged data were used for this analysis. Wind speed affects the concentration of the pollutant in a region irrespective of its kind. As reported in chapter 4 the wind speed at all the four stations has cyclic variations corresponding to 24 hours and 12 hours period. It is found that at all the stations there is high coherence at the frequency of 1 c/day. At stations 79 and 71 there is high coherence at the low frequency range of 0 - 0.014 c/day. At stations 60, 79 and 71 there is also high coherence at the frequency of .5 c/day. The magnitude of the coherence is given in Table E. This indicates that along with, may be some social activities, wind speed variations causes the above mentioned cyclic variations in the pollutant concentrations.

TABLE D

Result Summary of Cross Spectral Analysis Among Different Pollutants at Station 60.

Pollutants	Cross Correlation	Phase	Coherence
Sulphur dioxide and nitrogen oxides	High correlation at lags of 6, 19 and 28 days	Within control limits at all the frequencies	Significant at .04 c/day, .22 c/day and .34 c/day.
Carbon monoxide and nitrogen oxides	High correlation at lags of 14-15 days and 42-44 days	Mostly within confidence limits	Significant at frequencies of .04-.07 c/day, .2-.36 c/day and .44-.50 c/day. Average coherence is about .65
Carbon monoxide and sulphur dioxide	High correlation at lags of 12-19 days	Mostly within the control limits except in the frequency range of .42-.44 c/day	Generally insignificant. A coherence of .6 in the frequency range .05 and .48 c/day.
Ozone and nitrogen oxides	High correlation at lags of 10 days, 19 days, 28-33 days, 39 days and 44 days	Mostly within confidence limits. There is a phase lag of about 100^{o} at frequencey of .14 c/day, of about 270^{o} at .28 c/day and above 270^{o} at .38-.5 c/day.	Mostly insignificant except a coherence of .58 at the frequency of .1 c/day.
Ozone and carbon monoxide	High correlation at lags of 30 - 31 days, and 45 - 46 days	There appears to be a phase of about 108^{o} at frequencies of 0-.10 c/day, .13 c/day and .33 c/day; a phase of 90^{o} at 0 frequency of .44 c/day and of 144^{1} at the frequency of .5 c/day	Insignificant
Ozone and sulphur dioxide	High correlation at lags of 12 days, 19 days, 30 days, 38 days and 46 days	Mostly within confidence limits, except a phase of 180^{o} at .14 c/day, of 150^{o} at .33 c/day and of about 126^{o} at .5 c/day	Insignificant except a coherence of .57 in the frequency range .0-.01 c/day

Table E

Squared Coherence Associated with Cyclic Variation for Sulphur Dioxide and Wind Speed Data from the Stations 60, 79, 1 and 7.

Stations	Frequency	Squared Coherence
60	1 c/day	0.90
	.5 c/day	0.90
79	0 - 0.014 c/day	0.65
	1 c/day	0.80
	0.5 c/day	0.65
1	0.1 c/day	0.55
	1 c/day	0.75
71	0 - 0.014 c/day	0.6
	1 c/day	0.7
	0.5 c/day	0.6

References

1. M. E. Smith, Intern. Symp. chem. Reactions lower upper Atmosphere, San Francisco, 1961 Advance Papers, pp. 273-286, Stanford Research Institute.

2. F. Pooler, Jr., Intern. J. Air Water Pollution 4, 199 (1961).

3. D. B. Turner, J. Appl. Meteorol. 3, 83 (1964).

4. J. F. Clarke, J. Air Pollution Control Assoc. 14, 347 (1964).

5. R. I. Larsen, Am. Ind. Hyg. Assoc. J. 22, 97 (1961)

6. M. E. Miller and G. C. Holzworth, J. Air Pollution Control Assoc. 17, 46 (1967).

7. B. Davidson, J. Air Pollution Control Assoc. 17, 154 (1967); also, personal communication (1966).

8. Arthur C. Stern, "Air Pollution", Vol. 1, Academic Press, New York, 1968.

9. R. J. Carsen, Adv. Inviron. Health 8, 325 (1964).

10. C. F. Zimmer and R. J. Larsen, J. Air Pollution Control Assoc. 15, 565 (1965).

11. R. J. Larsen, C. E. Zimmer, D. A. Lynn, and K. G. Blernel, J. Air Pollution Control Assoc. 17, 85 (1967).

12. J. B. Koogler, R. S. Sholtes, A. L. Danis, and C. I. Harding, J. Air Pollution Control Assoc. 17, 211 (1967).

13. E. J. Croke, J. J. Roberts, Land Use as an Organizational Basis for Regional Air Resource Management, paper presented in Second International Clean Air Congress of the International Union of Air Pollution Prevention Association, December 6-11, 1970.

14. Takamatsu, T., Naito, M., Hiraoka, M., Ikeda, Y., and Kawata, K., Computer Control System for Preventing Air Pollution, paper presented in Second International Clean Air Congress of the International Union of Air Pollution Prevention Association, December 6-11, 1970.

15. S. J. Marcus, and J. J. Harrington, 'Decision Models for Short-Term Air Pollution Control Policies', Masters thesis, Harvard University, Cambridge, Massachusetts, U.S.A.

16. E. S. Lee, 'Quasilinearization and Invariant Imbedding', Academic Press, New York (1968).

17. M. G. Salvadari, R. J. Schwarz, 'Differential Equations in Engineering Problems', Prentice-Hall, Inc., New York (1954).

18. R. Bellman, and R. Kalaba, 'Dynamic Programming, Invariant Imbedding and Quasilinearization': Comparison and Interconnections. RM-4038-PR RAND Corp., Santa Monica, Calif., March, 1964.

19. R. Bellman and R. Kalaba, 'Quasilinearization and Boundary Value Problems', American Elsevier Publishing Co., New York, 1965.

20. R. Bellman, H. Kajiwada, R. Kalaba, 'Quasilinearization, System Identification, and Prediction', Rand Corp., Santa Monica, Calif., RM-3812-PR, Aug., 1963.

21. E. S. Lee, 'Quasilinearization and Estimation of Parameters in Differential Equations', I&EC Fundamentals, Vol. 7, No. 1, 132, Feb., 1968.

22. E. S. Lee, 'Quasilinearization, Nonlinear Boundary Value Problems, and Optimization," Chem. Eng. Sci., Vol. 21, 183, 1966.

23. E. S. Lee, 'Quasilinearization in Optimization. A Numerical Study," A.I.Ch.E. 59th Ann. Meeting, Detroit, Michigan, Dec. 4-8, 1966.

24. E. S. Lee, 'Reduction in Dimensionality, Dynamic Programming and Quasilinearization,' Kansas State University Bulletin, Vol. 51, No. 8, Aug., 1967.

25. R. Kumar, 'Optimization of Management Systems Using Sensitivity Analysis', Master's Thesis, Kansas State University, 1969.

26. P. D. Shah, 'Application of Quasilinearization to Industrial Management System', Master's Thesis, Kansas State University, 1969.

27. E. S. Lee and I. K. Hwang, 'Stream Quality Modeling by Quasilinearization', Journal Water Pollution Control, Federation, Feb. 1971.

28. J. B. Scaborough, Numerical Mathematicla Analysis, Johns Hopkins Press, Baltimore, Maryland, 1962.

29. T. A. Mastler, and C. M. Malter, 'Statistical Approach to Estuany Behaviour', Journal ASCE, Sanitary Engineering Division, Dec. 1968.

30. C. F. Nordin, IGNACIO-RODRIGUEZ-ITURBE., 'Some applications of Cross-Spectral analysis in Hydrology. Rainfall and Runoff', Water Resources Research, Vol. 5, June, 1969.

31. C. G. Gunnerson, 'Optimizing Sampling Intervals in Tldal Estuaries', Journal ASCE, Sanitary Engineering Division, April 1966.

32. R. V. Thomann, 'Time Series Analysis of Water Quality Dates', Journal ASCE, Sanitoary Engineering Division, Feb., 1967.

33. G. M. Jenkins and D. G. Watts, 'Spectral Analysis and its Applications', Holden-Day, August 1969.

34. R. B. Blackman and J. W. Tukey, J. W., 'The Measurement of Power Spectra', Dover Publications Inc., 1958.

35. C. W. J. Granger and M. Hatanaka, 'Spectral Analysis of Economic Time Series', Princeton University Press, Princeton, New Jersey, 1964.

36. G. S. Fishman, 'Spectral Methods in Econometrics', A report prepared for United States Air Force Project Rand, April 1968.

37. J. A. Blackburn, 'Spectral Analysis: Methods and Techniques', Marcel Dekker Inc., New York, 1970.

38. G. E. P. Box, and G. M. Jenkins, 'Time Series Analysis Forecasting and Control', Holden-Day, San Francisco, 1970.

39. N. Weiner, 'Generalized Harmonic Analysis', Acta Math., Vol. 35, 1930 pp. 117-258.

40. M. S. Bartlett, An Introduction to Stochastic Processes, Cambridge University Press, London, 1961.

41. N. R. Draper and H. Smith, 'Applied Regression Analysis', John Wiley & Sons Inc. N.Y., 1966.

42. U. Grenander, and M. Rosenblatt, 'Statistical Analysis of Stationary Series', John Wiely and Sons, New York, 1957.

43. BMD02T, Biomedical Computer Proglrams, Health Sciences Computing Facility, UCLA, Calif.

44. Air Pollution Control District, County of Los Angeles, Calif., private communication.

APPENDIX A

Tables and Graphs

Table P

Number of Missing Hourly Observations for Sulphur Dioxide

Station	Month											
	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
60	39 Mainly 20th, 21st and 22nd days' observations	9	21	20	17	12	18	11	8 Mainly 7th hour observations	11	8	7
79	18	9	15	15	10	4	13	8	6	15	12	21 Mainly 8th and 14th days' observations
1	28 Mainly 5th and 6th days' observations	144 Mainly 12th, 13th, 14th, 15th, and 16th days' observations	31 Mainly 4th and 5th days' observations	46 Mainly 5th, 6th and 29th days' observations	13	26 Mainly 15th, 16th and 17th days observations	4		0	19 Mainly 5th and 6th days observations	2	5
71	16 Mainly 11th and 12th hours' observations	15 Mainly on 10th and 11th day between the hours 9 to 14.	6	12 Mainly 13th hour observations	30 Mainly 18th and 19th days obs.	64 Mainly 20th and 21st days' obs.	12	6	14 Mainly 22nd day's obs.	14	9 Mainly 12th hour obs.	10 Mainly 12th hour obs.

Table Q

Number of Missing Hourly Observations for Carbon Monoxide

Station	Month											
	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
60	9	7	14	9	8	9	3	7	4	16 Mainly 18th and 19th day's obs.	2	8 Mainly 1st days obs.
79	13	7	18 Mainly 17th days obs.	13	7	7	42 Mainly 25th, 26th and 31st days' obs.	72 Mainly 1st, 2nd, 8th, 9th, and 23rd days' obs.	7	24 Mainly 9th day's obs.	18 Mainly 21st and 22nd day's obs.	21 Mainly 7th and 8th day's obs.
1	16	9	8 Mainly 31st days obs.	36 Mainly 6th and 7th days' obs.	8	4	4	2	6 Mainly 29th days' obs.	19 Mainly 6th and 7th days obs.	3	31 Mainly 6th and 7th days obs.
71	13	8	29 Mainly 16th and 17th days' obs.	11	8	8	7	2	3	2	0	1

Table R

Number of Missing Hourly Observations for Nitrogen Oxides

Station	Month											
	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
60	75 Mainly 12th and 13th hour observations	73 Mainly 12th and 13th hour observations	72 Mainly 12th and 13th hour observations	64 Mainly 12th hour and 18th and 19th day observations	71 Mainly 12th hour and 16th and 17th hour observations	38 Mainly 12th hour observations	71 Mainly 12th hour and 31st day observations	73 Mainly 12th hour and 1st and 2nd day observations	52 Mainly 12th hour and 14th day observations	58 Mainly 12th hour and 10th day observations	43 Mainly 12th hour observations	25 Mainly 12th hour observations
79	46 Mainly 12th hour observations	30 Mainly 12th hour observations	38 Mainly 12th hour observations	42 Mainly 12th hour and 23rd day observations	71 Mainly 8th, 9th and 10th day and 12th hour observations	47 Mainly 12th hour observations	92 Mainly 12th hour, 20th, 21st, 24th and 25th days observations	41 Mainly 12th hour observations	36 Mainly 12th hour observations	50 Mainly 12th hour, 13th and 14th days observations	114 8th to 15th hour observations on about 10 days and 23rd day's observations	101 Mainly 17th, 18th, 19th and 20th day observations
1	89 Mainly 13th hour, 9th, 10th, and 11th day observations	29 Mainly 9th and 10th hours observations	36 Mainly 10th hour observations	88 Mainly 10th 11th 12th and 26th days' observations	44 Mainly 13th hour and 22nd and 23rd days' observations	75 Mainly 13th hour and 26th and 25th days' observations	68 Mainly 13th hour and 25th day's observations	31 Mainly 13th hour observations	20 Mainly 13th hour observations	14 Mainly 13th hour observations	80 Mainly 13th, 14th, 27th, 28th, and 29th days' observations	79 Mainly 5th, 6th, 11th, 12th, and 13th days' observations
71	37 Mainly 14th hour observations	45 Mainly 13th hour 24th and 25th days observations	104 Mainly 20th, 21th and 22nd day and 13th hour observation	81 Mainly 13th hour, 11th and 13th days observations	33 Mainly 13th hour observations	88 Mainly 13th hour, 24th, 25th, 26th, 27th, and 28th days observations	136 Mainly 3rd, 4th, 5th, 6th, 20th, 21st, 22nd days' observations	47 Mainly 13th and 16th days observations	56 Mainly 9th, 10th days and 13th hours observations	50 Mainly 6th, 28th days and 13th hours observations	62 Mainly 12th, 17th, 18th days and 12th and 13th hours' observations	11

Table 5

Number of Missing Hourly Observations for Ozone*

Station	Month											
	Jan.	Feb.	March	April	May	Juno	July	Aug.	Sept.	Oct.	Nov.	Dec.
60	60 Mainly 26th and 27th days obs.	57 Mainly 1st and 2nd days obs.	43	35	34	38	37	33	28	42	32	37
79	38	33	36	36	34	39	32	38	35	33	37	58 Mainly 8th and 12th days obs.
1	57 Mainly 26th and 27th days obs.	57 Mainly 1st, 2nd, 9th, and 10th days obs.	43 61	36	34	37	38	33	38	41	33	37
71	39 Mainly 18th, 19th, 22nd and 23rd days obs.	80 Mainly on 25th, 26th, 29th, 30th, and 31st days obs. between hours 7-16.	82 Mainly on 1st, 2nd, 25th, 26th, 29th, 30th, and 31st days' obs. between hours 7-16.	124 Mainly 13th, 14th, 15th, 16th, 20th, 22nd and 23rd days obs. between hours 7-16.	68 Mainly on 21st, 24th, 25th, 26th, days obs. between hours 7-16.	48 Mainly 7th and 8th days obs. between hours 6-15.	35	32	34	36	31	35

*4th hour observations were missing at all the stations throughout the year. The numbers mentioned in the table include these missing observations.

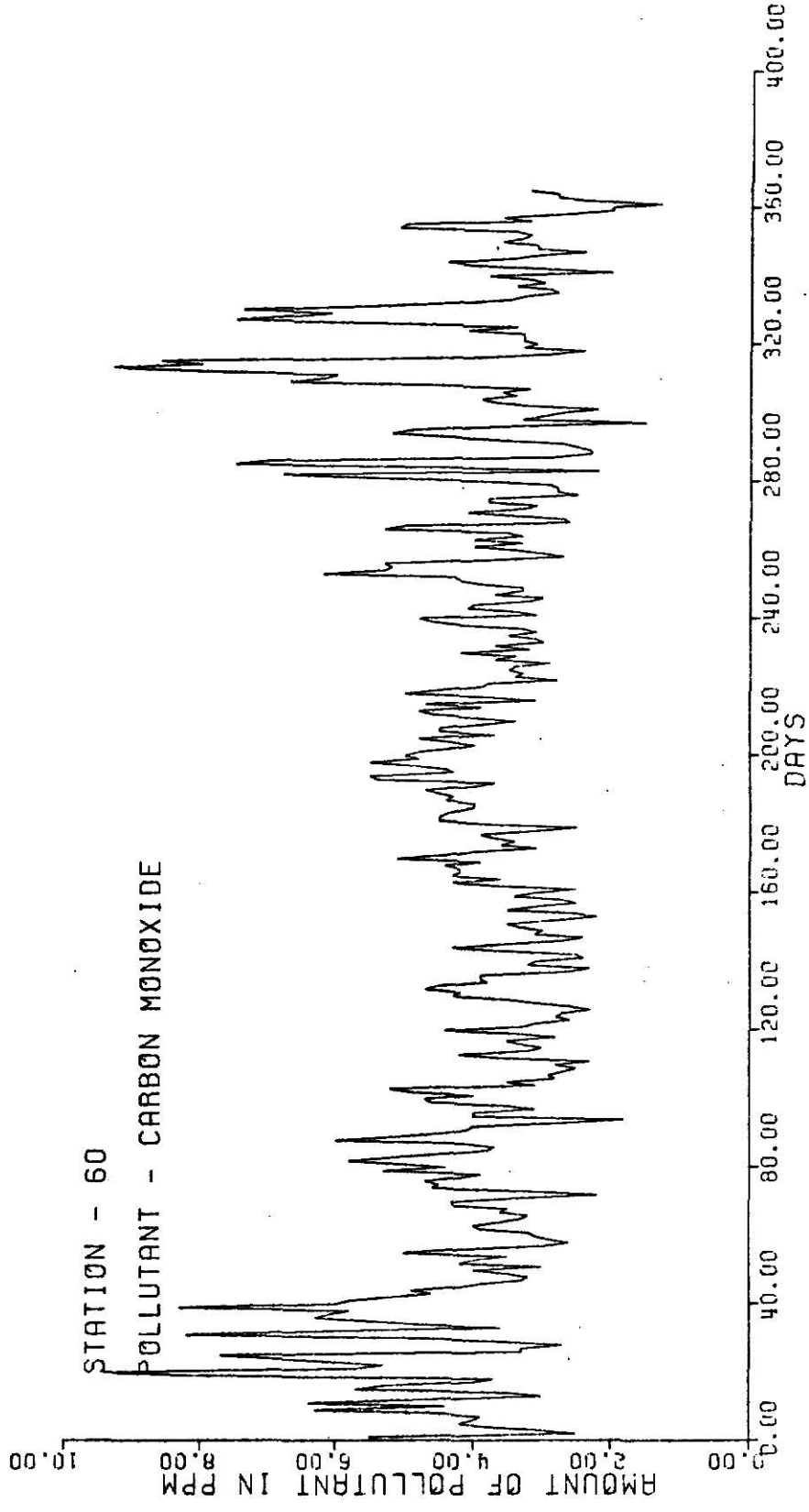

FIGURE 20. Daily averaged carbon moncxide data frcm station 60.

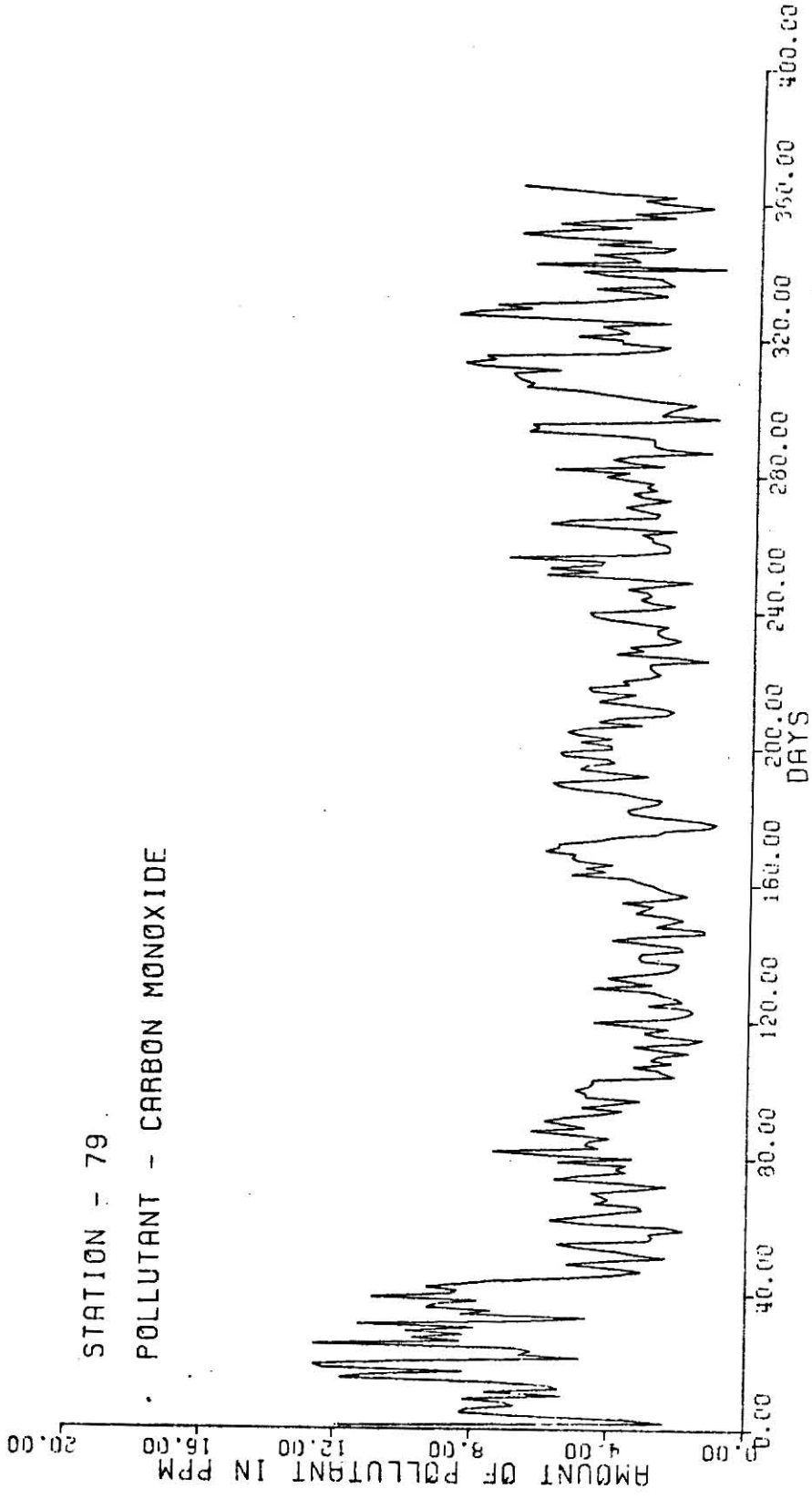

FIGURE 21. Daily averaged carbon monoxide data from station 79.

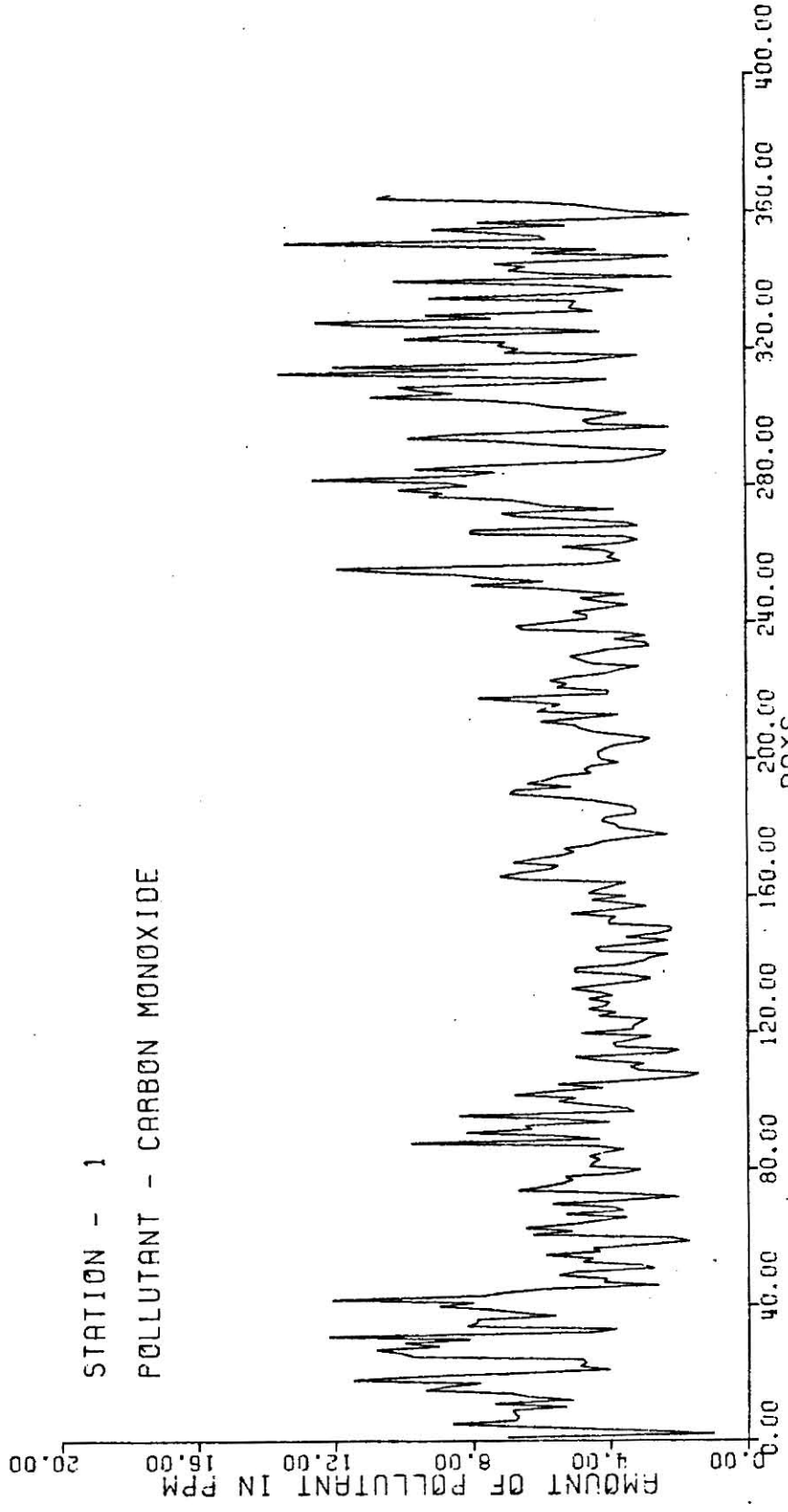

FIGURE 22. Daily averaged carbon monoxide data from station 1.

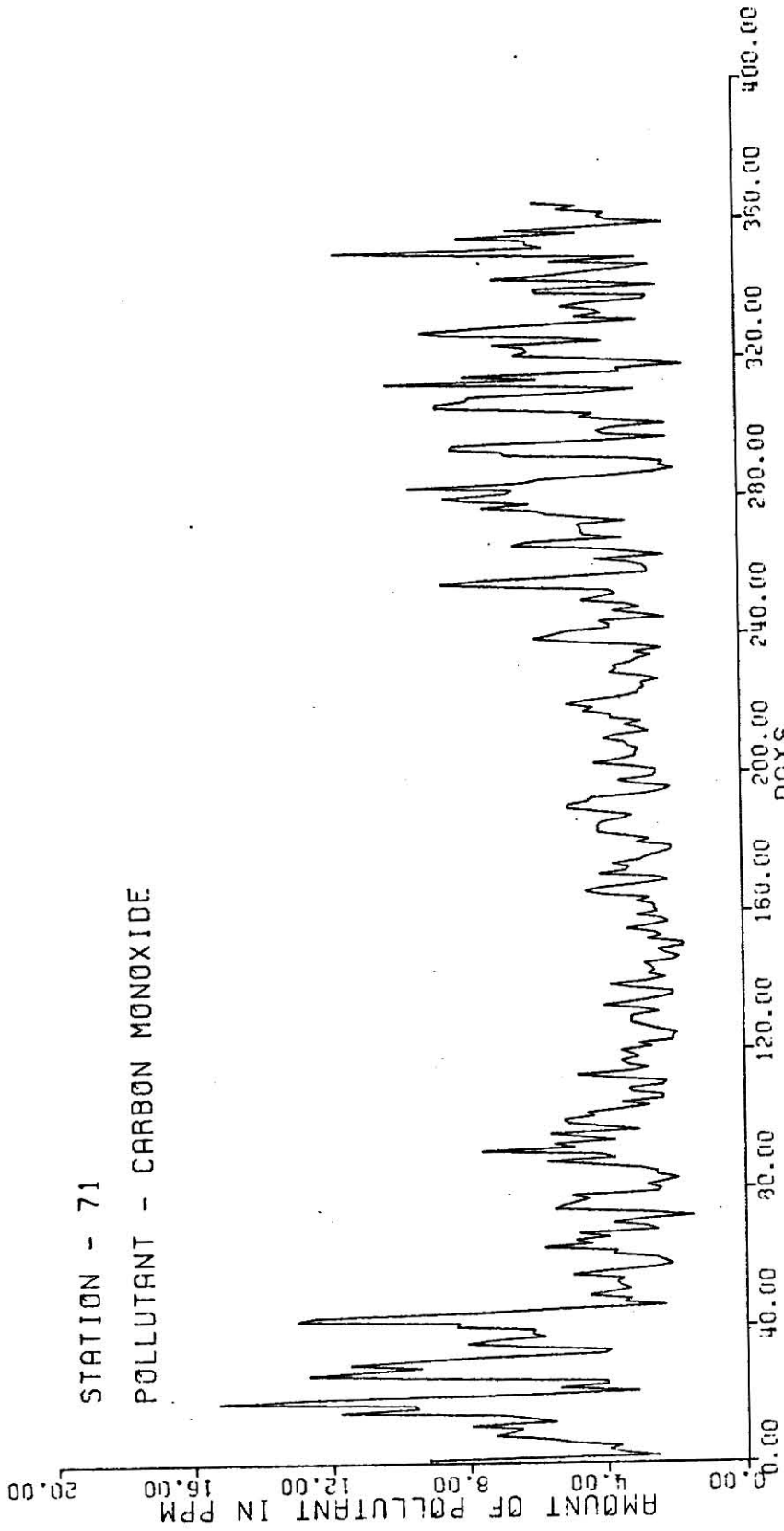

FIGURE 23. Daily averaged carbon monoxide data from station 71.

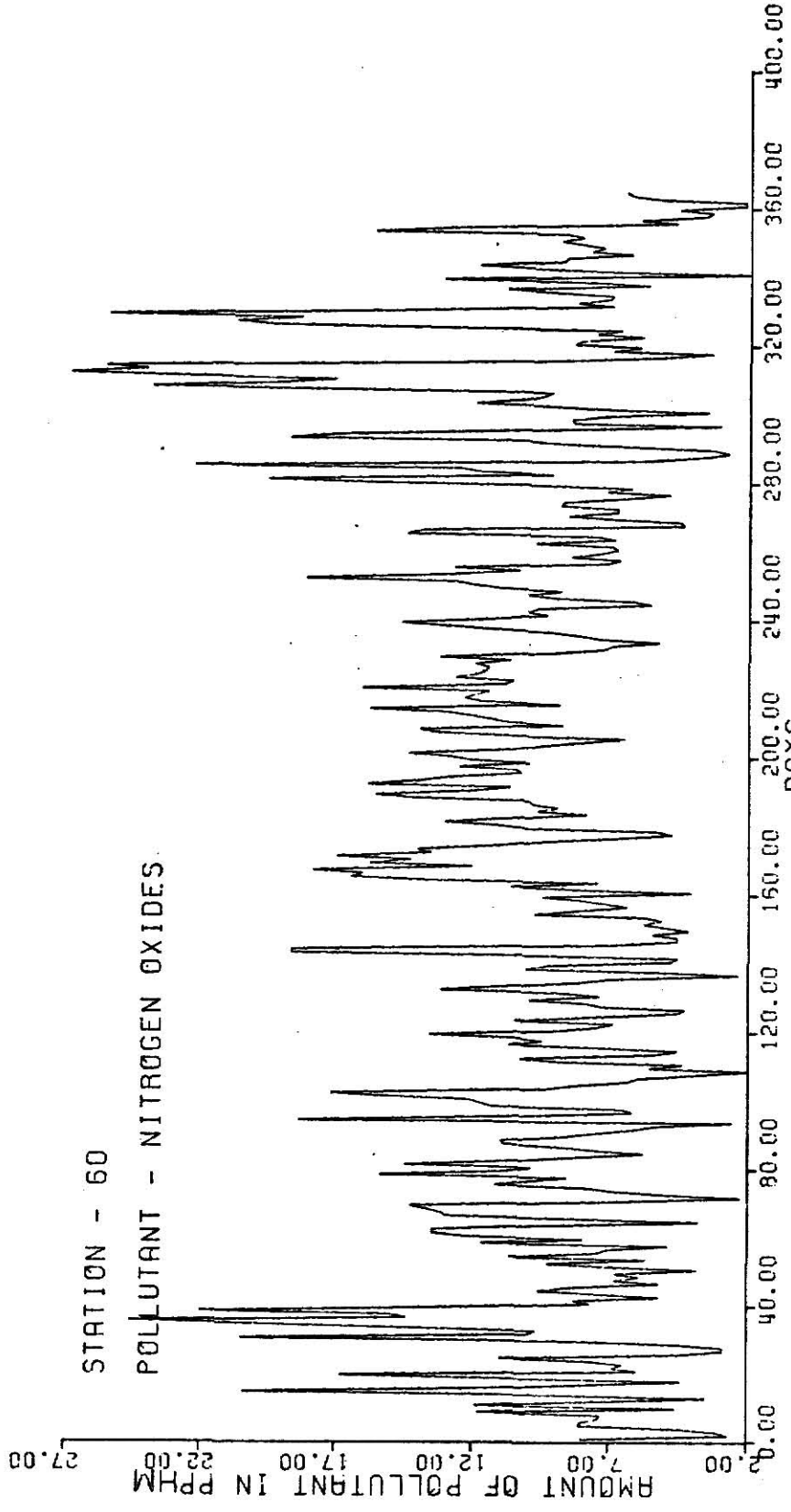

FIGURE 24. Daily averaged nitrogen oxides data from station 60.

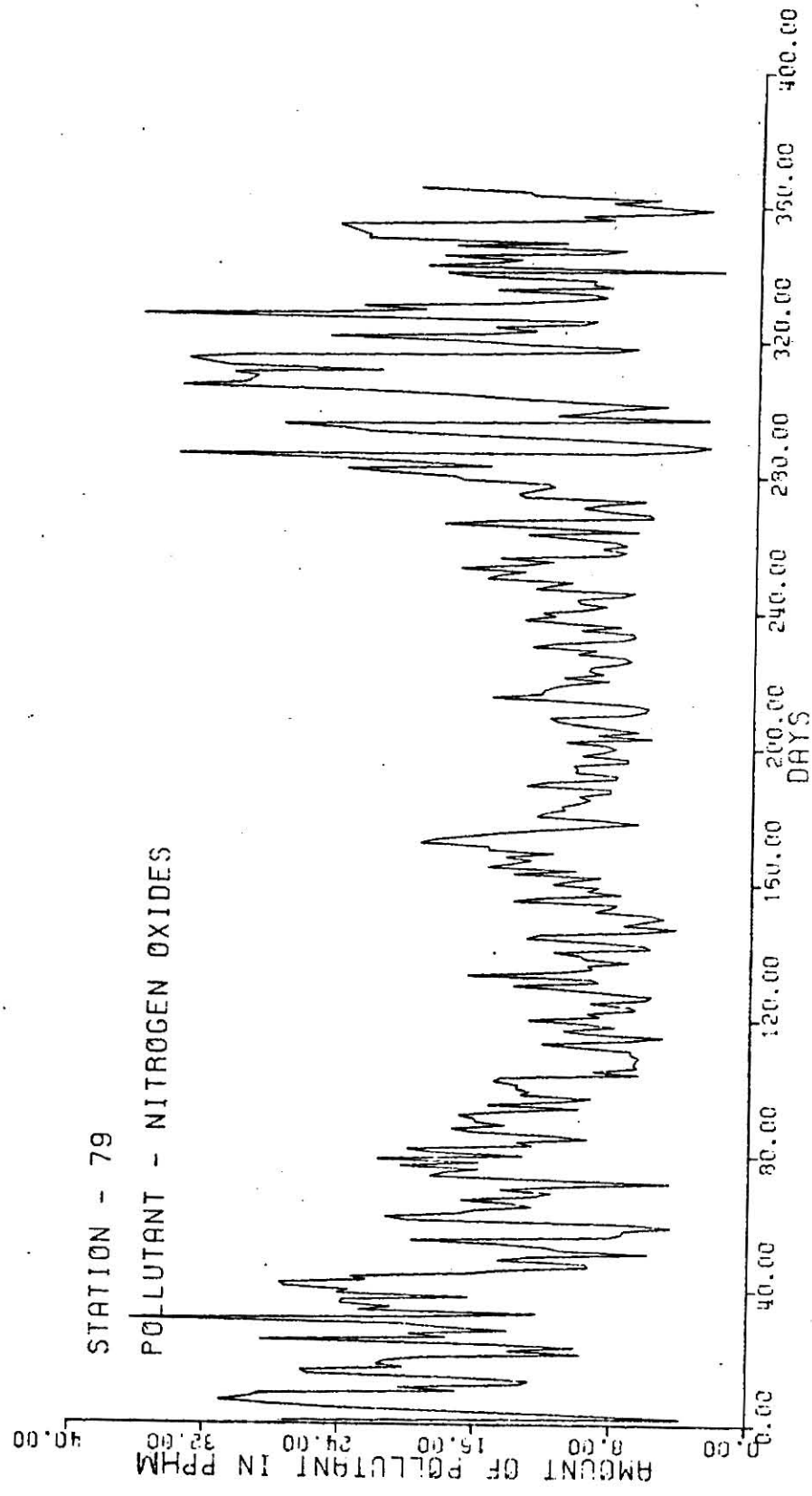

FIGURE 25. Daily averaged nitrogen oxides data from station 79.

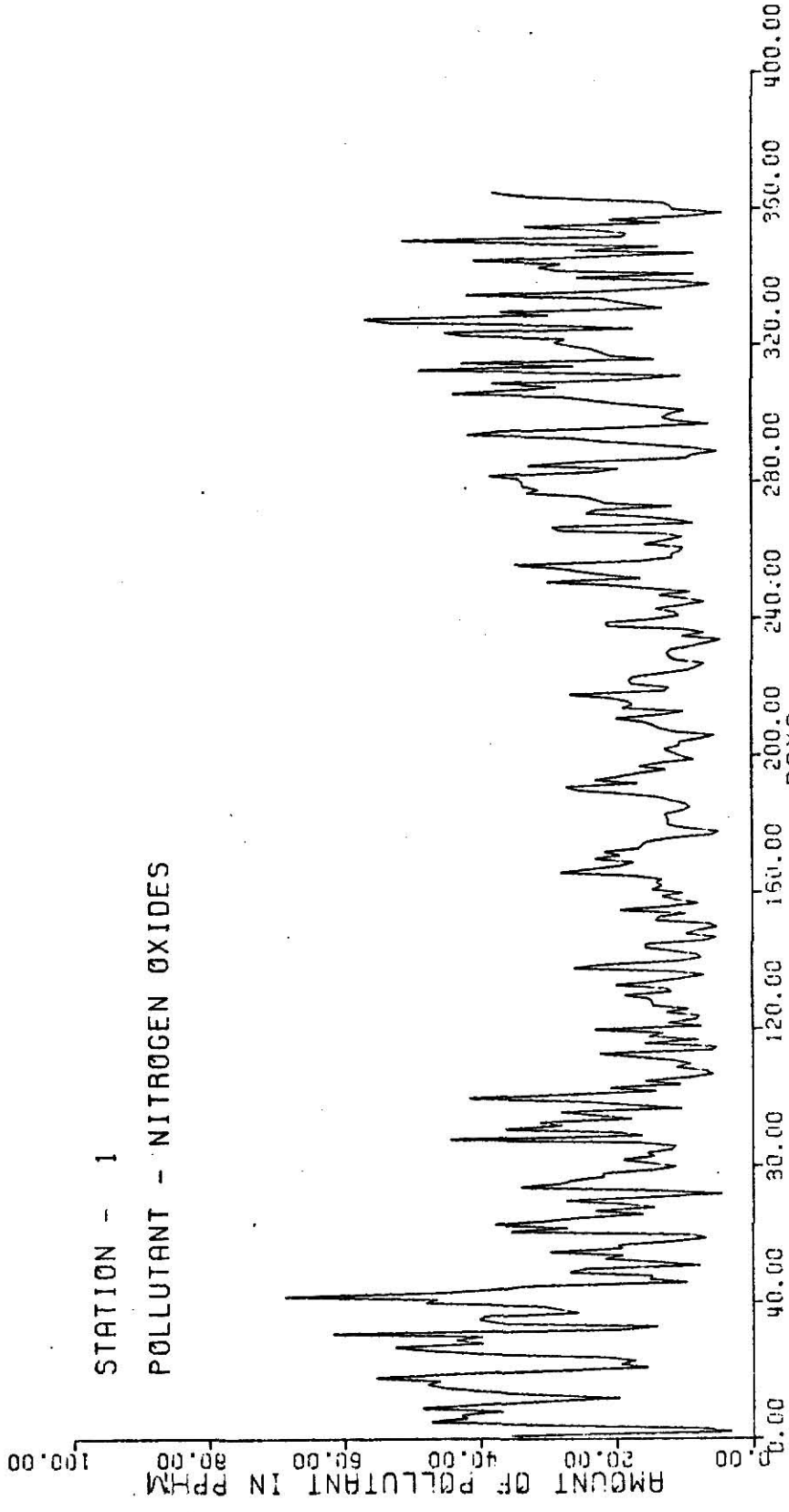

FIGURE 26. Daily averaged nitrogen oxides data from station 1.

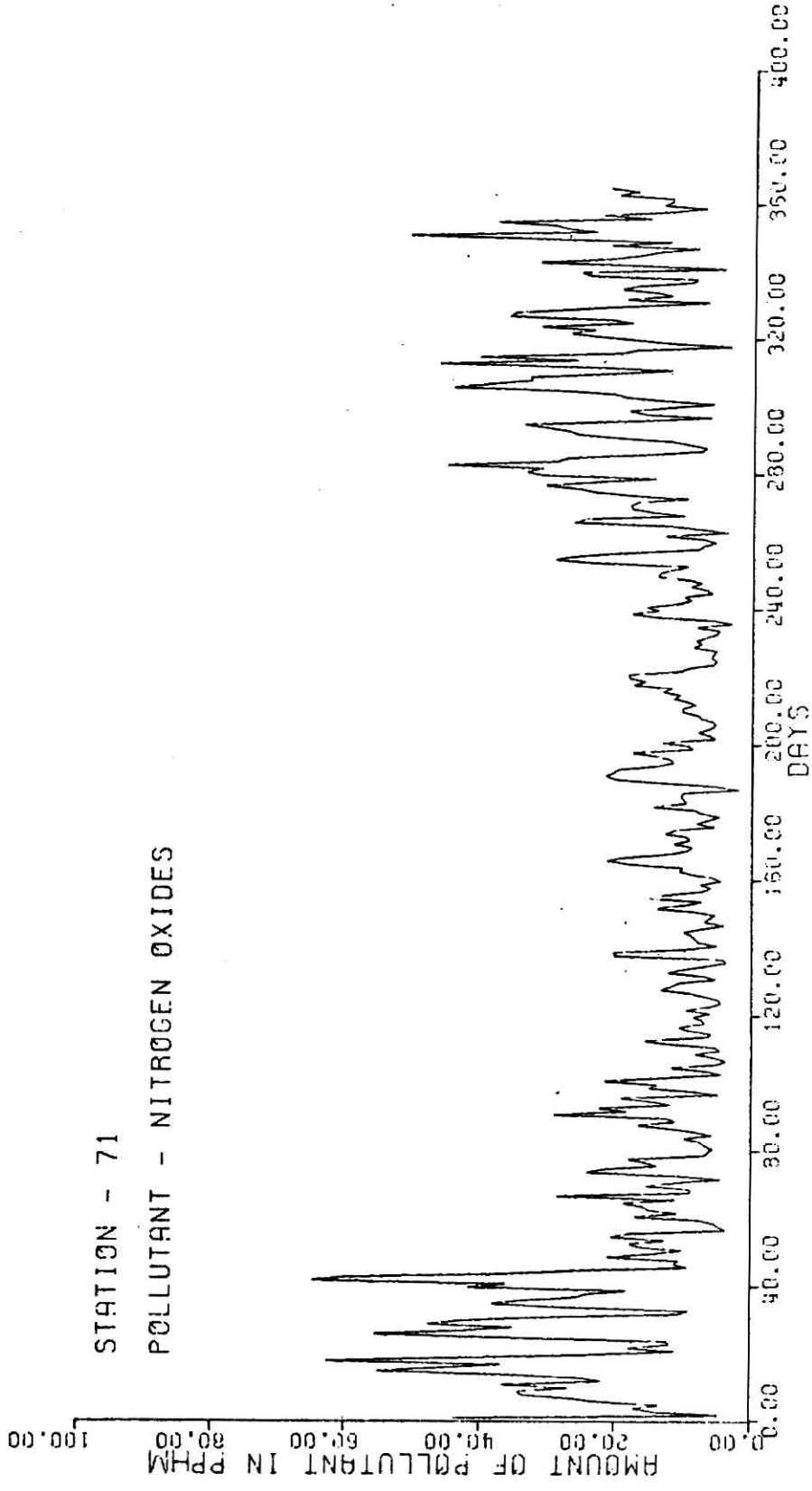

FIGURE 27. Daily averaged nitrogen oxides data from station 71.

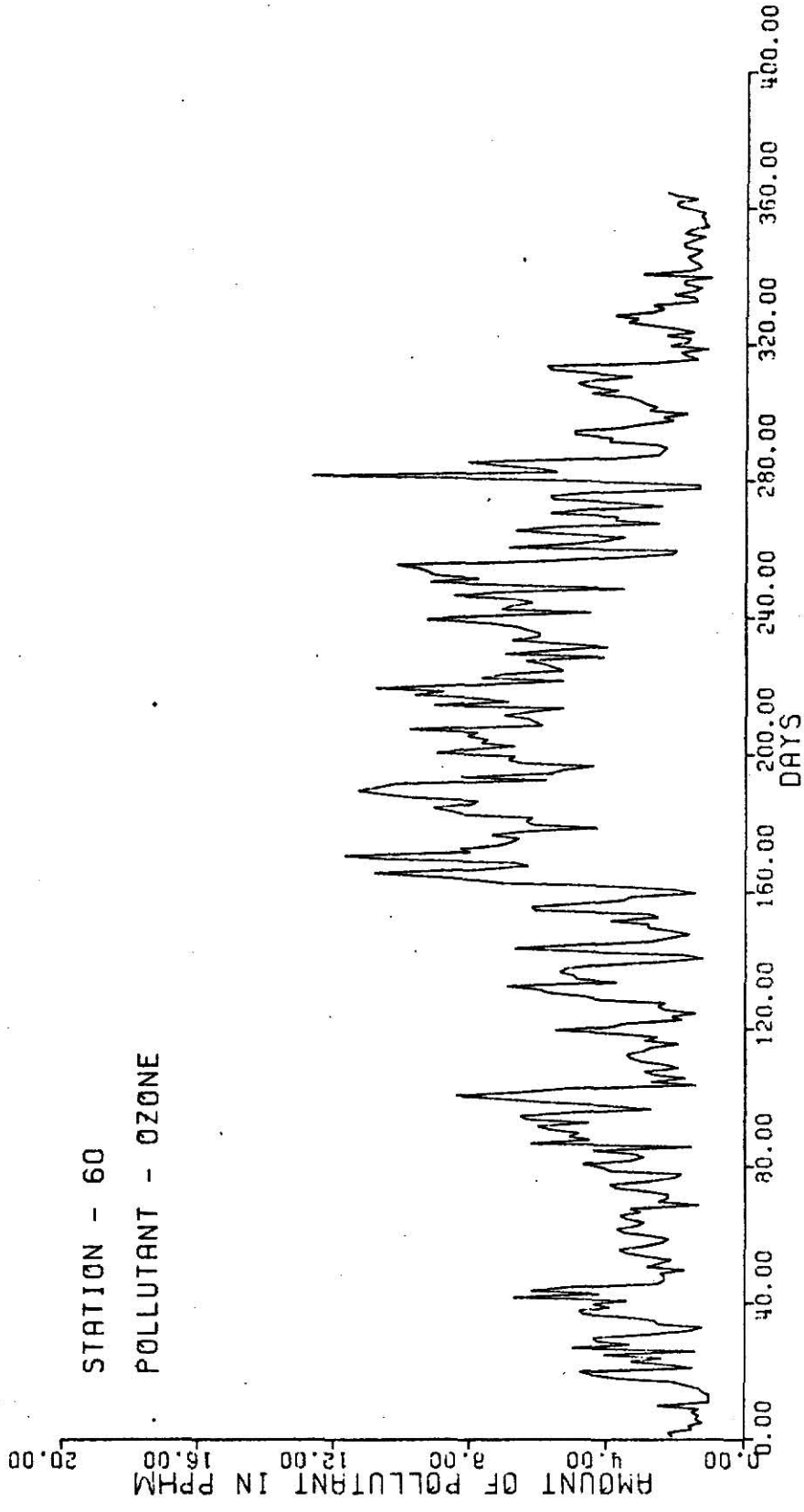

FIGURE 28. Daily averaged ozone data from station 60.

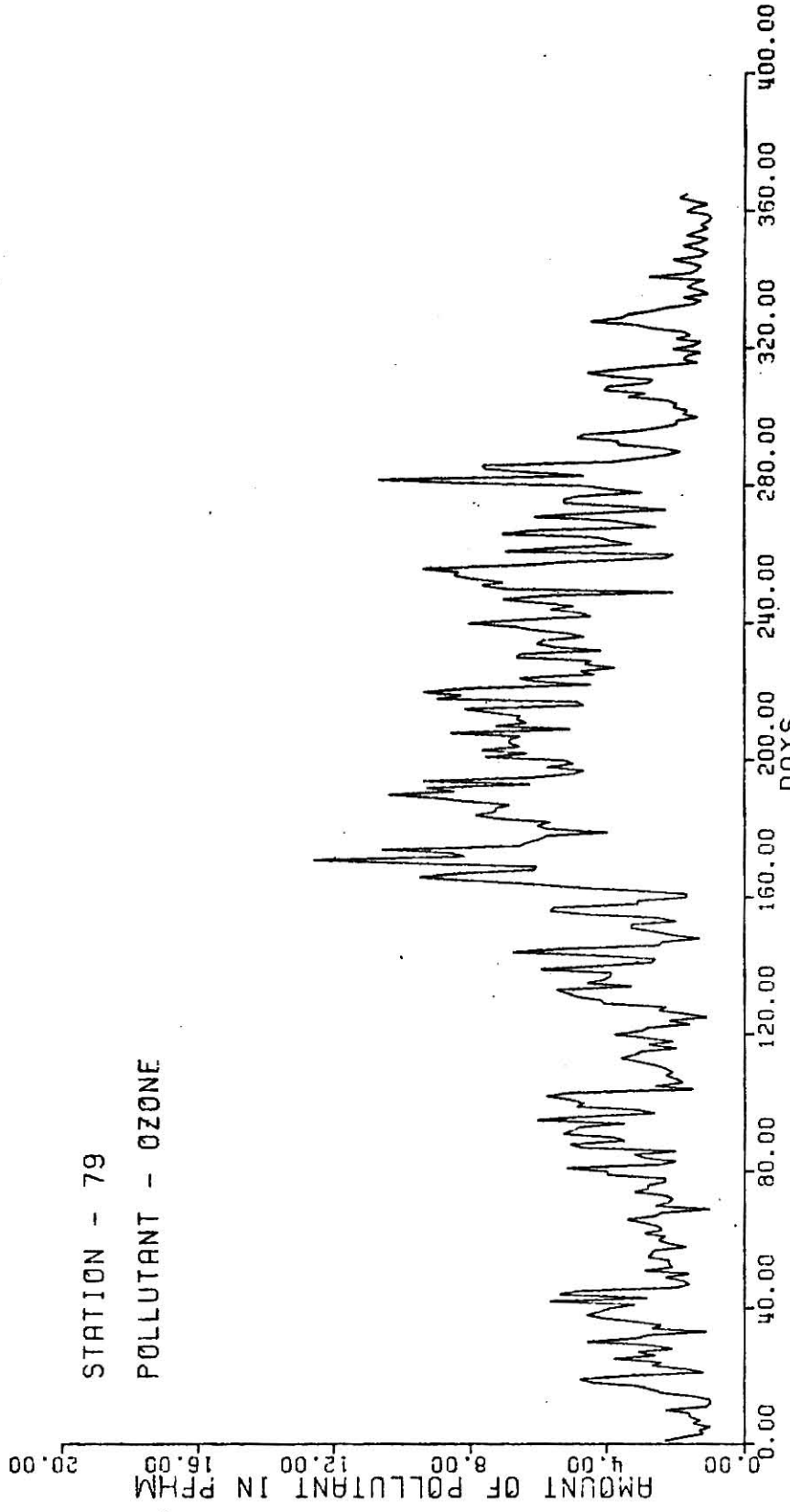

FIGURE 29. Daily averaged ozone data from station 79.

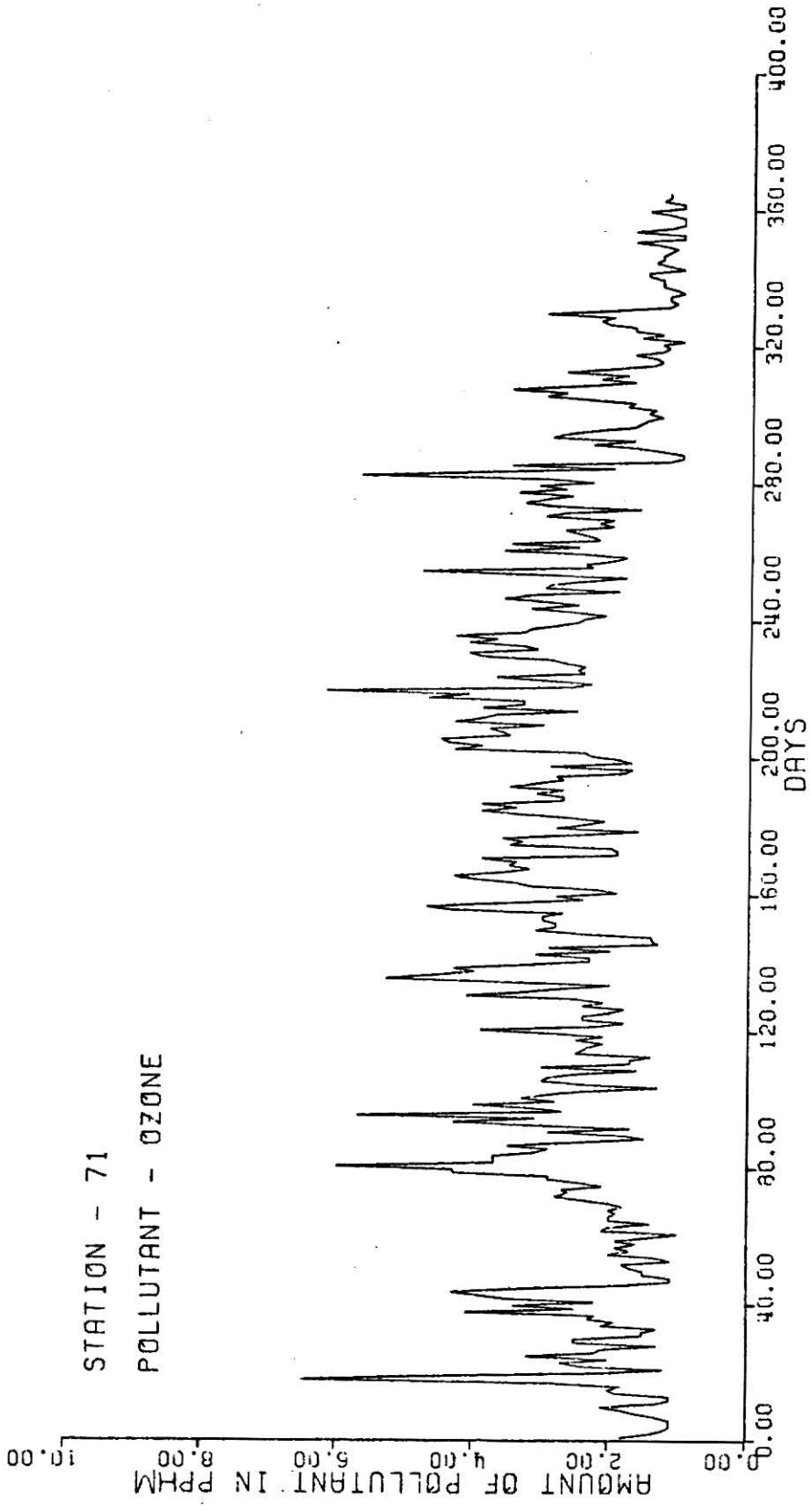

FIGURE 31. Daily averaged ozone data from station 71.

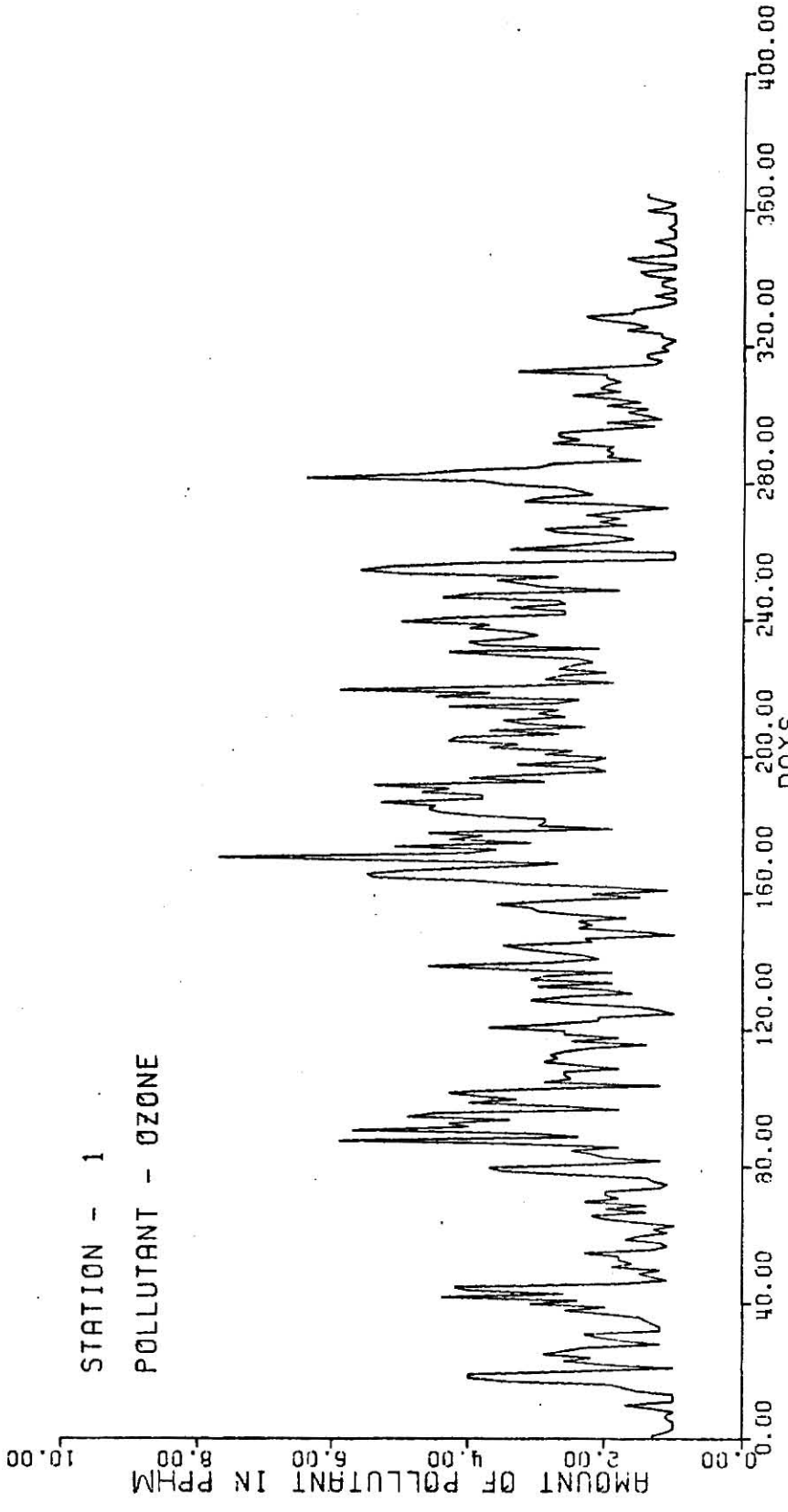

FIGURE 30. Daily averaged ozone data from station 1.

Table 6

Harmonic Analysis of Carbon monoxide from station 60

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	3.971	0.0	3.971	0.0	15.770	0.9067
1	-0.162	0.033	0.165	168.35	0.054	0.0337
2	0.082	0.150	0.171	61.12	0.058	0.0363
3	0.150	-0.013	0.150	354.74	0.045	0.0280
4	-0.025	0.170	0.172	98.39	0.059	0.0366
5	0.160	-0.160	0.227	315.78	0.103	0.0638
6	-0.098	0.209	0.231	154.23	0.107	0.0663
12	-0.092	0.098	0.135	133.34	0.036	0.0266
15	0.072	0.097	0.120	53.42	0.029	0.0180
16	-0.156	-0.034	0.160	192.25	0.051	0.0318
19	-0.095	-0.115	0.150	230.26	0.045	0.0278
20	0.128	0.091	0.157	35.59	0.049	0.0306
24	-0.191	-0.042	0.195	192.58	0.076	0.0473
25	0.141	-0.099	0.172	324.91	0.059	0.0367
26	-0.014	0.154	0.155	95.34	0.048	0.0298
36	0.142	-0.052	0.151	339.98	0.046	0.0285
73	-0.064	0.094	0.114	124.44	0.026	0.0162
74	0.060	-0.073	0.095	309.49	0.010	0.0113

Table 7

Harmonic Analysis of Carbon Monoxide from station 79

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	4.316	0.0	4.310	0.0	18.620	0.8032
1	-0.690	-0.077	0.695	186.40	0.966	0.2118
2	0.348	0.326	0.477	43.17	0.455	0.0998
3	0.104	-0.193	0.219	298.44	0.096	0.0212
4	0.047	0.332	0.336	81.86	0.225	0.0495
5	0.197	-0.402	0.448	296.16	0.402	0.0881
6	-0.208	0.302	0.367	124.53	0.269	0.0591
12	-0.092	0.163	0.188	119.54	0.070	0.0155
18	0.016	0.150	0.155	83.85	0.045	0.0100
19	-0.147	-0.129	0.196	221.13	0.076	0.0168
20	0.193	0.109	0.222	29.56	0.098	0.0216
26	0.160	0.093	0.185	30.11	0.068	0.0150
29	-0.123	-0.224	0.255	241.05	0.131	0.0287

Table 8

Harmonic Analysis of Carbon Monoxide from Station 1

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	5.396	0.0	5.39	0.0	29.126	0.8455
1	-0.635	0.369	0.736	149.91	1.085	0.2040
4	0.063	0.170	0.181	69.55	0.066	0.0124
5	0.176	-0.217	0.280	309.07	0.157	0.0296
6	-0.244	0.109	0.268	155.90	0.143	0.0270
10	-0.115	-0.199	0.230	239.91	0.105	0.0199
12	-0.013	0.306	0.307	92.57	0.188	0.0354
14	-0.200	-0.078	0.214	201.38	0.092	0.0173
16	-0.138	0.169	0.218	129.28	0.095	0.0179
26	0.280	0.110	0.301	21.47	0.182	0.0342
29	-0.176	-0.295	0.344	239.13	0.237	0.0446
52	-0.003	-0.339	0.340	269.38	0.231	0.0434

Table 9

Harmonic Analysis of Carbon Monoxide from Station 71

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	4.367	0.0	4.367	0.0	19.075	0.7915
1	-0.798	0.176	0.817	167.56	1.336	0.2660
2	0.133	0.202	0.242	56.50	0.117	0.0234
3	-0.012	-0.260	0.261	267.32	0.136	0.0271
4	0.080	0.322	0.332	75.98	0.221	0.0440
5	0.183	-0.217	0.284	310.21	0.161	0.0322
6	-0.189	0.077	0.204	157.77	0.083	0.0166
14	-0.212	-0.092	0.231	203.60	0.107	0.0214
15	0.188	0.160	0.241	41.74	0.116	0.0232
25	-0.141	-0.128	0.190	222.18	0.072	0.0145
26	0.273	0.075	0.284	15.48	0.161	0.0322
29	-0.150	-0.322	0.355	244.96	0.253	0.0504
30	0.114	0.183	0.216	237.91	0.093	0.0186
38	-0.055	-0.195	0.203	254.13	0.082	0.0165
52	0.071	-0.143	0.160	296.57	0.051	0.0102
51	-0.060	0.157	0.168	111.15	0.057	0.0113

Table 10

Harmonic Analysis of Nitrogen Oxides from Station 60

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	9.720	0.0	9.720	0.0	94.48	0.8236
1	0.200	0.443	0.486	65.68	0.473	0.0234
3	0.696	0.314	0.764	24.31	1.168	0.0577
5	0.378	-0.405	0.554	312.98	0.615	0.0304
12	-0.150	0.591	0.609	104.31	0.743	0.0367
13	-0.321	-0.215	0.386	213.74	0.299	0.0148
14	-0.408	-.003	0.408	180.47	0.333	0.0165
15	0.202	0.321	0.380	57.77	0.289	0.0143
16	-0.849	0.030	0.850	177.93	1.445	0.0714
17	0.437	0.288	0.523	33.40	0.548	0.0271
19	-0.267	-0.275	0.384	225.85	0.295	0.0146
20	0.276	0.210	0.347	37.33	0.341	0.0118
21	-0.427	0.296	0.519	145.14	0.532	0.0266
22	0.451	-0.148	0.475	341.78	0.452	0.0223
24	-0.575	0.010	0.575	178.94	0.663	0.0328
25	0.407	-0.431	0.593	313.45	0.704	0.0348
26	0.141	0.577	0.594	76.21	0.706	0.0309
34	-0.304	0.295	0.424	135.88	0.360	0.0178
45	0.386	0.165	0.420	23.20	0.353	0.0175
51	-0.171	-0.275	0.324	238.09	0.211	0.0104
52	-0.075	-0.376	0.383	258.63	0.294	0.0145
73	-0.228	0.389	0.450	120.15	0.405	0.0200

Table 11

Harmonic Analysis of Nitrogen Oxides from Station 79

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	13.507	0.0	13.507	0.0	182.44	0.8680
1	-2.181	0.284	2.199	172.59	9.676	0.2232
3	0.789	-0.372	0.873	334.74	1.525	0.0352
5	0.143	-1.097	1.106	277.51	2.442	0.0565
8	0.266	-0.653	0.705	292.24	0.996	0.0230
11	0.272	-0.515	0.582	297.89	0.679	0.0157
12	-0.163	0.949	0.963	279.80	1.854	0.0428
13	-0.225	-0.427	0.483	242.19	0.467	0.0108
15	0.497	0.164	0.523	18.25	0.548	0.0126
16	-0.688	0.246	0.731	160.39	1.068	0.0247
17	0.291	-0.415	0.507	305.07	0.515	0.0119
18	-0.009	0.473	0.474	91.21	0.449	0.0104
19	-0.441	-0.585	0.732	232.96	1.073	0.0248
24	-0.415	0.419	0.590	134.74	0.696	0.0161
25	-0.150	-0.620	0.638	256.33	0.815	0.0188
26	0.392	0.417	0.513	46.79	0.656	0.0151
32	0.055	-0.474	0.477	276.70	0.456	0.0105
33	0.194	0.662	0.688	73.60	0.952	0.0220
52	-0.160	-0.650	0.660	79.70	0.870	0.0202

Table 12

Harmonic Analysis of Nitrogen Oxides from Station 1

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	20.392	0.0	20.392	0.0	415.837	0.7469
1	-4.363	0.017	4.363	179.78	38.084	0.2703
2	0.765	0.653	1.006	40.44	2.025	0.0144
3	0.644	-0.976	1.160	303.45	2.735	0.0194
4	0.139	1.060	1.060	82.39	2.248	0.0160
5	0.898	-1.095	1.416	309.37	4.013	0.0225
6	-1.183	0.084	1.145	324.59	4.215	0.0299
12	-0.133	1.583	1.588	94.87	5.049	0.0358
14	-1.406	0.265	1.431	169.37	4.099	0.0291
16	-0.444	1.003	1.096	113.94	2.406	0.0171
26	1.096	0.169	1.109	8.80	2.463	0.0175
29	-0.619	-1.454	1.581	246.90	5.006	0.0355
30	0.219	0.900	0.926	76.30	1.718	0.0122
38	-0.430	-0.949	1.042	245.57	2.173	0.0154
39	0.756	0.869	1.151	48.95	2.653	0.0188
52	0.249	-1.486	1.506	279.56	4.54	0.0322

Table 13

Harmonic Analysis of Nitrogen Oxides from Station 71

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to variance	Fraction Variance
0	16.219	0.0	16.219	0.0	263.066	0.6777
1	-4.213	0.918	4.312	167.72	37.199	0.2973
2	0.971	0.416	1.058	23.130	2.240	0.0179
3	0.547	-1.567	1.660	289.270	5.513	0.0441
4	0.200	1.531	1.534	82.450	4.711	0.0377
5	0.461	-0.793	0.917	300.230	1.685	0.0135
6	-0.808	0.400	0.902	153.68	1.629	0.0130
12	-0.453	1.221	1.302	110.41	3.394	0.0271
14	-1.275	-0.288	1.305	192.38	3.408	0.0272
15	0.536	0.741	0.915	54.11	1.675	0.0134
26	1.411	0.080	1.414	3.26	3.999	0.0320
27	-0.891	0.158	0.905	169.95	1.638	0.0131
29	-0.772	-1.554	1.736	243.56	6.028	0.0482
30	0.412	0.820	0.918	63.30	1.687	0.0135
38	-0.406	-0.956	1.038	246.95	2.158	0.0173
39	0.727	0.667	0.987	42.55	1.949	0.0156

Table 14

Harmonic Analysis of Ozone from Station 60

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	4.314	0.0	4.314	0.0	18.616	0.7489
1	1.135	0.376	1.196	18.34	2.861	0.4585
2	0.050	0.481	0.483	83.94	0.460	0.0750
3	0.242	0.054	0.248	12.71	0.123	0.0197
5	0.232	-0.151	0.277	326.92	0.154	0.0247
9	-0.053	-0.211	0.218	255.86	0.095	0.0152
10	-0.063	-0.235	0.244	254.89	0.119	0.0191
12	0.003	0.227	0.227	89.20	0.103	0.0166
15	-0.006	0.212	0.212	91.75	0.090	0.0145
20	0.049	0.195	0.212	75.73	0.080	0.0130
21	0.085	0.183	0.202	65.02	0.081	0.0131
23	-0.001	0.180	0.180	90.40	0.065	0.0105
24	-0.177	-0.044	0.182	194.04	0.066	0.0107
40	0.273	-0.038	0.276	352.07	0.152	0.0245

Table 15

Harmonic Analysis of Ozone from Station 79

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	3.975	0.0	3.975	0.0	15.802	0.7525
1	1.050	0.382	1.117	20.01	2.499	0.4807
2	0.075	0.415	0.422	79.67	0.356	0.0685
3	0.214	-0.040	0.217	349.34	0.094	0.0183
4	0.199	0.054	0.206	15.38	0.085	0.0164
5	0.209	-0.086	0.226	337.72	0.102	0.0198
9	-0.063	-0.199	0.209	252.37	0.087	0.0168
10	-0.023	-0.235	0.236	264.34	0.112	0.0216
15	0.020	0.207	0.208	84.26	0.086	0.0167
23	-0.038	0.178	0.183	102.19	0.066	0.0128
24	-0.165	0.011	0.166	176.19	0.055	0.0106
29	0.020	-0.183	0.185	276.53	0.068	0.0132
40	0.201	0.021	0.202	6.00	0.082	0.0158

Table 16

Harmonic Analysis of Ozone from Station 1

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	2.469	0.0	2.469	0.0	6.095	0.8077
1	0.481	0.018	0.482	2.18	0.463	0.3203
2	-0.039	0.131	0.136	106.71	0.037	0.0258
3	0.060	-0.025	0.065	337.49	0.008	0.0059
4	0.125	0.056	0.138	24.28	0.038	0.0263
5	0.113	-0.159	0.196	305.53	0.076	0.0530
6	-0.047	-0.073	0.087	236.79	0.015	0.0105
7	0.089	0.054	0.104	31.26	0.021	0.0151
9	-0.028	-0.102	0.102	254.48	0.022	0.0156
10	-0.050	-0.103	0.115	244.00	0.026	0.0183
15	0.012	0.170	0.170	85.86	0.058	0.0400
23	-.004	0.127	0.127	90.66	0.032	0.0233
29	0.016	-0.134	0.135	276.92	0.036	0.0251
40	0.101	-0.004	0.101	357.37	0.020	0.0142
52	-0.021	0.145	0.145	98.45	0.043	0.0300
69	0.001	-0.094	0.094	27.98	0.017	0.0122

Table 17

Harmonic Analysis of Ozone from Station 71

Harmonic	C_1	C_2	Amplitude	Phase	Contribution to Variance	Fraction Variance
0	2.483	0.0	2.483	0.0	6.167	0.8536
1	0.367	-0.039	0.369	300.57	0.274	0.2567
2	-0.080	0.136	0.159	117.36	0.050	0.0478
5	-0.042	-0.107	0.115	248.22	0.026	0.0251
6	-0.147	0.040	0.152	164.84	0.046	0.0441
8	0.037	-0.071	0.080	297.74	0.013	0.0124
15	0.094	0.090	0.130	43.48	0.034	0.0323
18	0.074	-0.030	0.084	332.02	0.014	0.0135
19	-.120	-0.055	0.132	335.37	0.035	0.0332
20	-0.017	0.087	0.089	101.54	0.016	0.0151
23	0.030	0.072	0.078	67.17	0.012	0.0117
24	-0.042	0.074	0.085	119.73	0.014	0.0139
27	-0.073	-0.065	0.098	221.54	0.019	0.0182
29	0.077	-0.073	0.106	316.74	0.022	0.0216
52	-0.032	0.098	0.103	108.12	0.021	0.0203

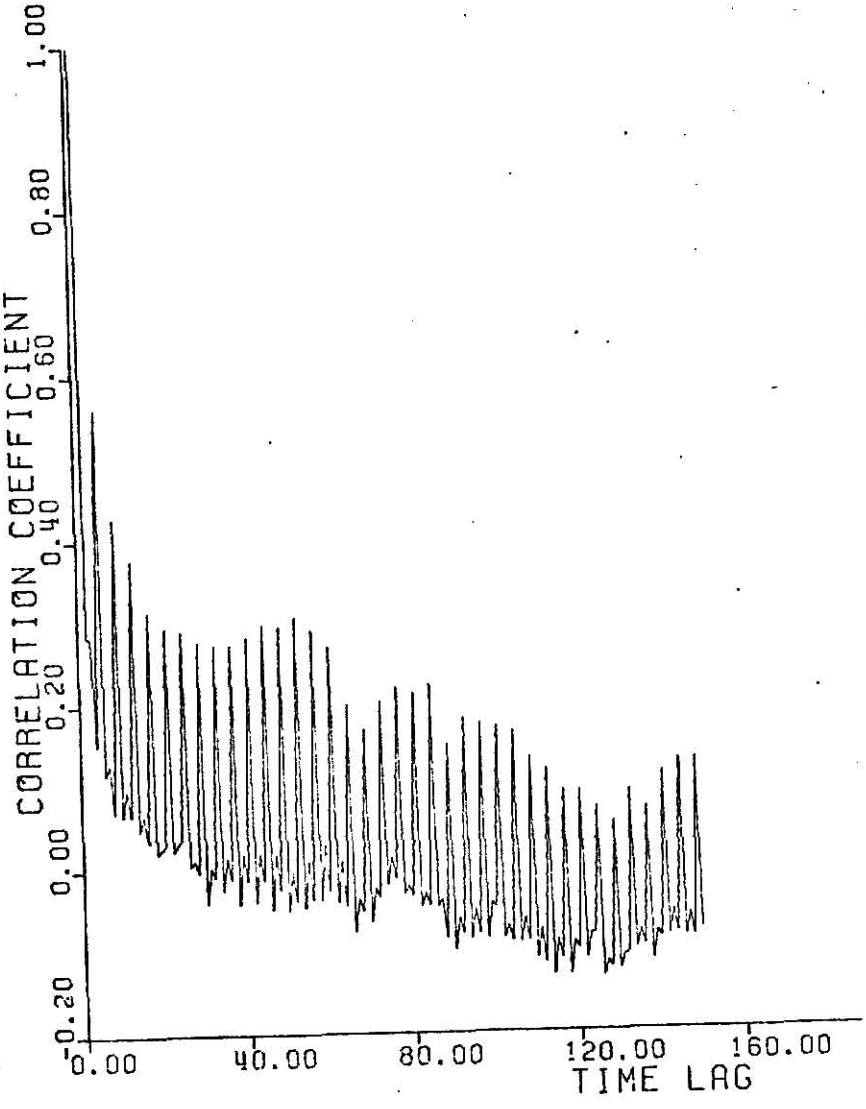

FIGURE 32. Autocerrelation function of carbon monoxide from station 60.

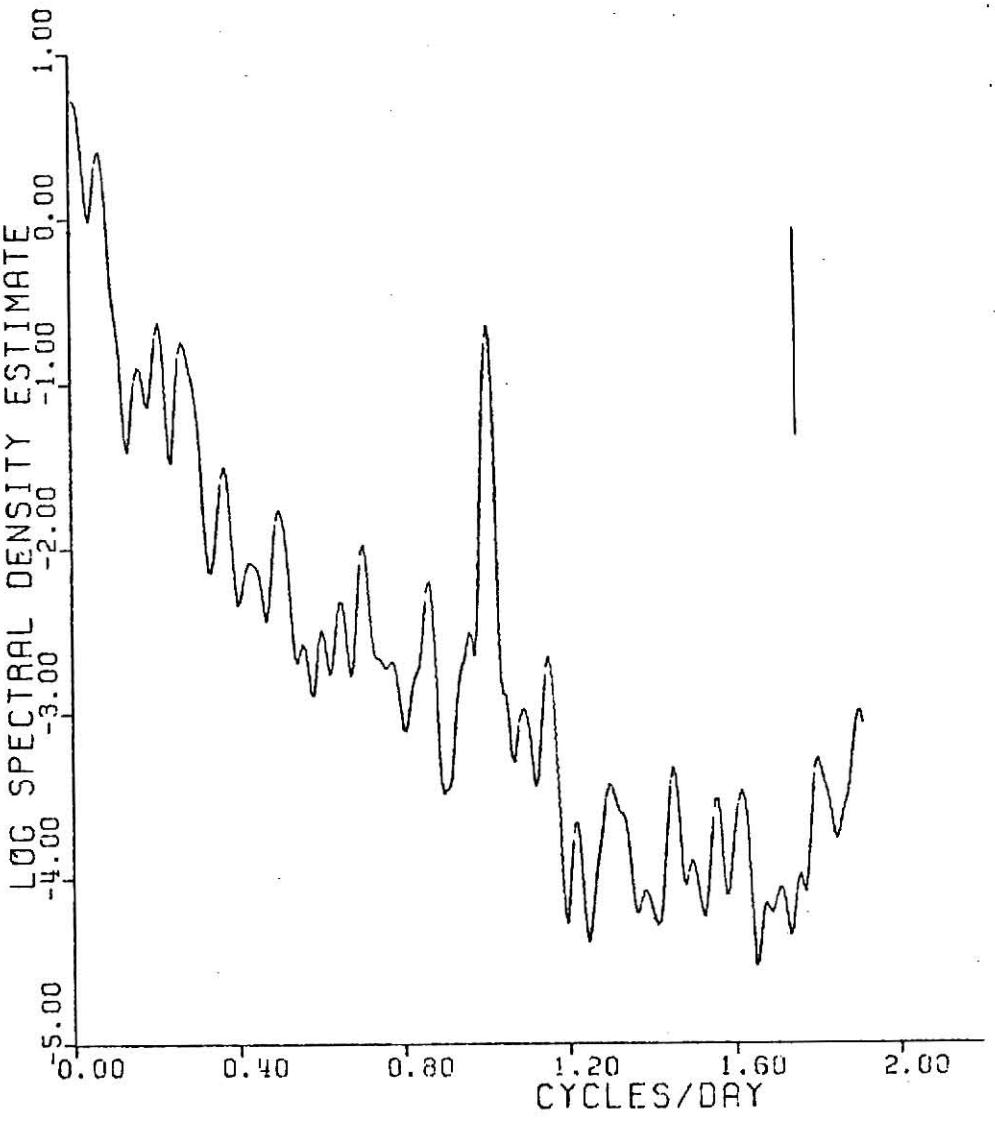

FIGURE 33. Spectral density function for carbon monoxide based on six hours averaged data, station 60.

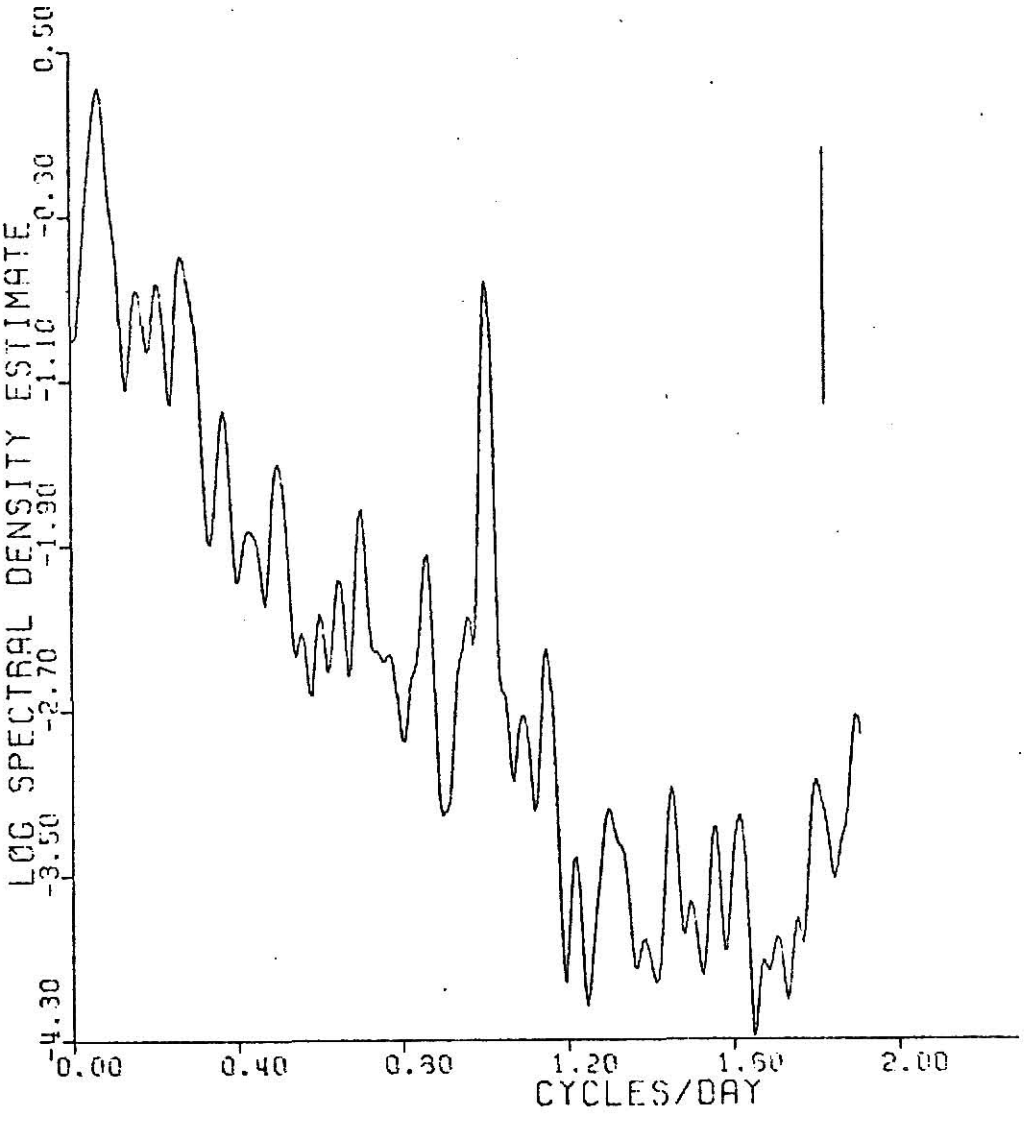

FIGURE 34. Spectral plot of residuals for six hours averaged carbon monoxide data from station 60.

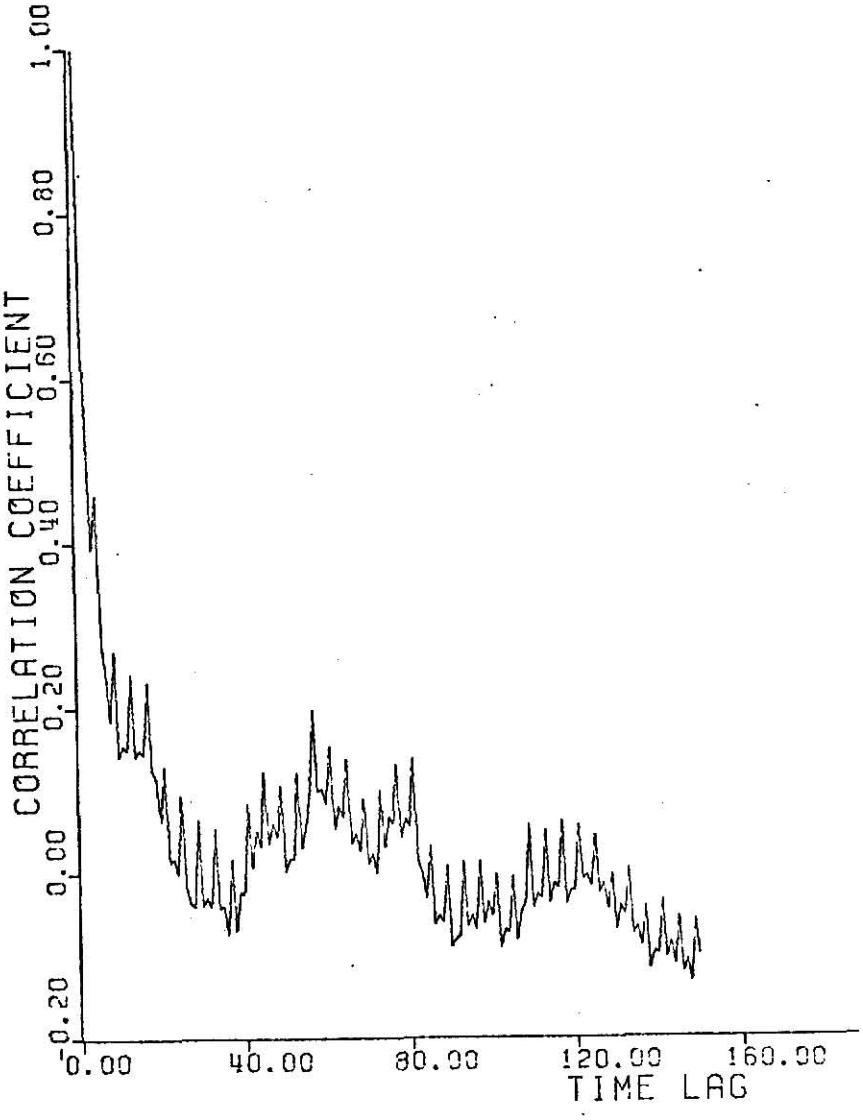

FIGURE 35. Autocorrelation function of carbon monoxide from station 79.

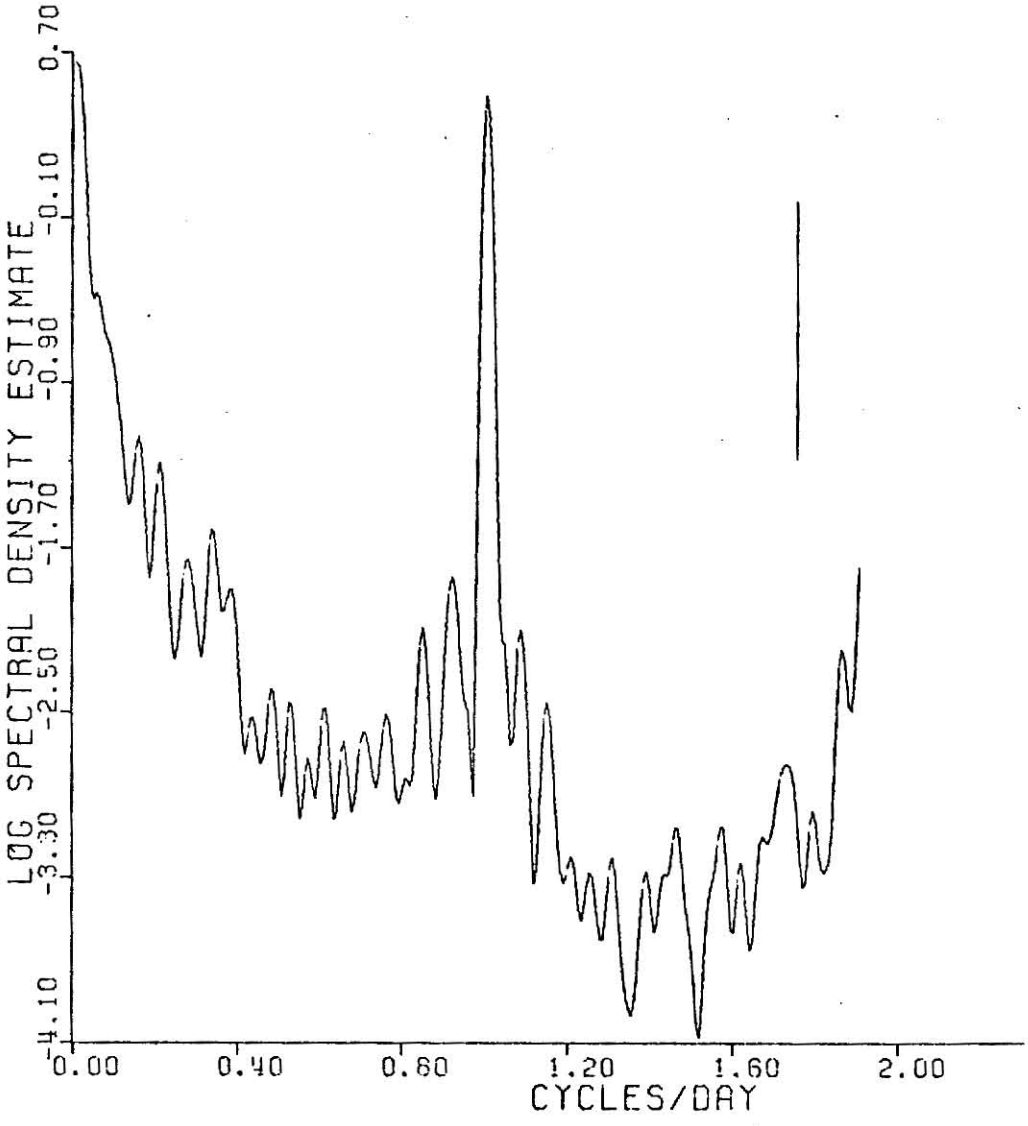

FIGURE 36. Spectral density function for carbon monoxide based on six hours averaged data, station 79.

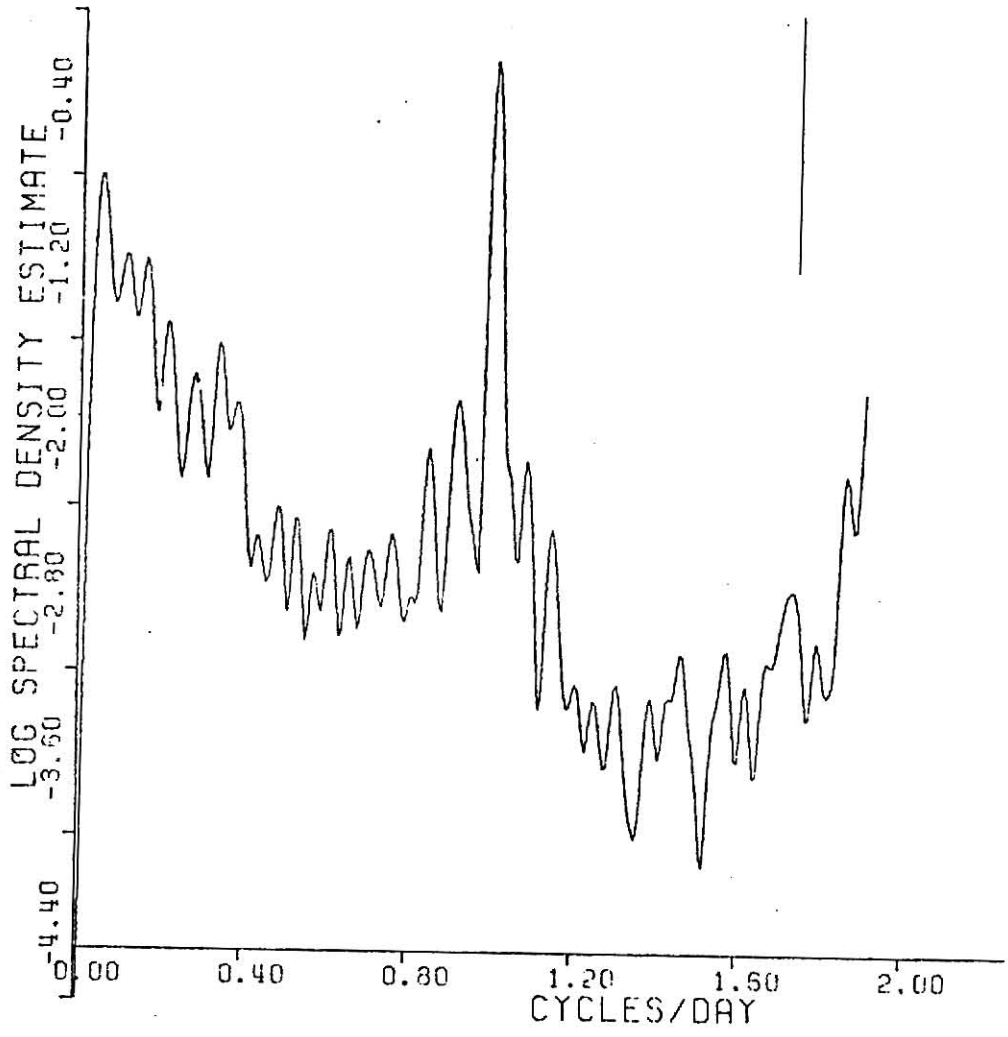

FIGURE 37. Spectral plot of residuals for six hours averaged data from station 79.

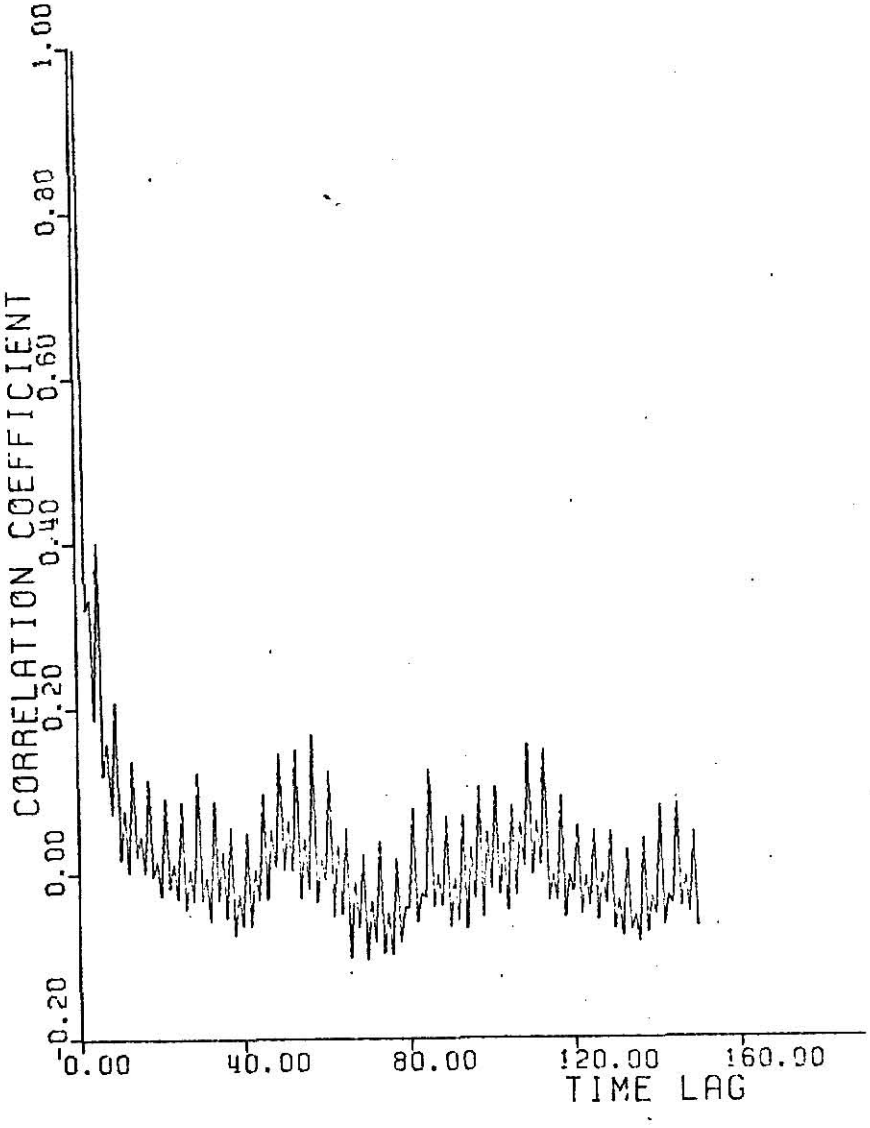

FIGURE 38. Autocorrelation function of carbon monoxide from station 1.

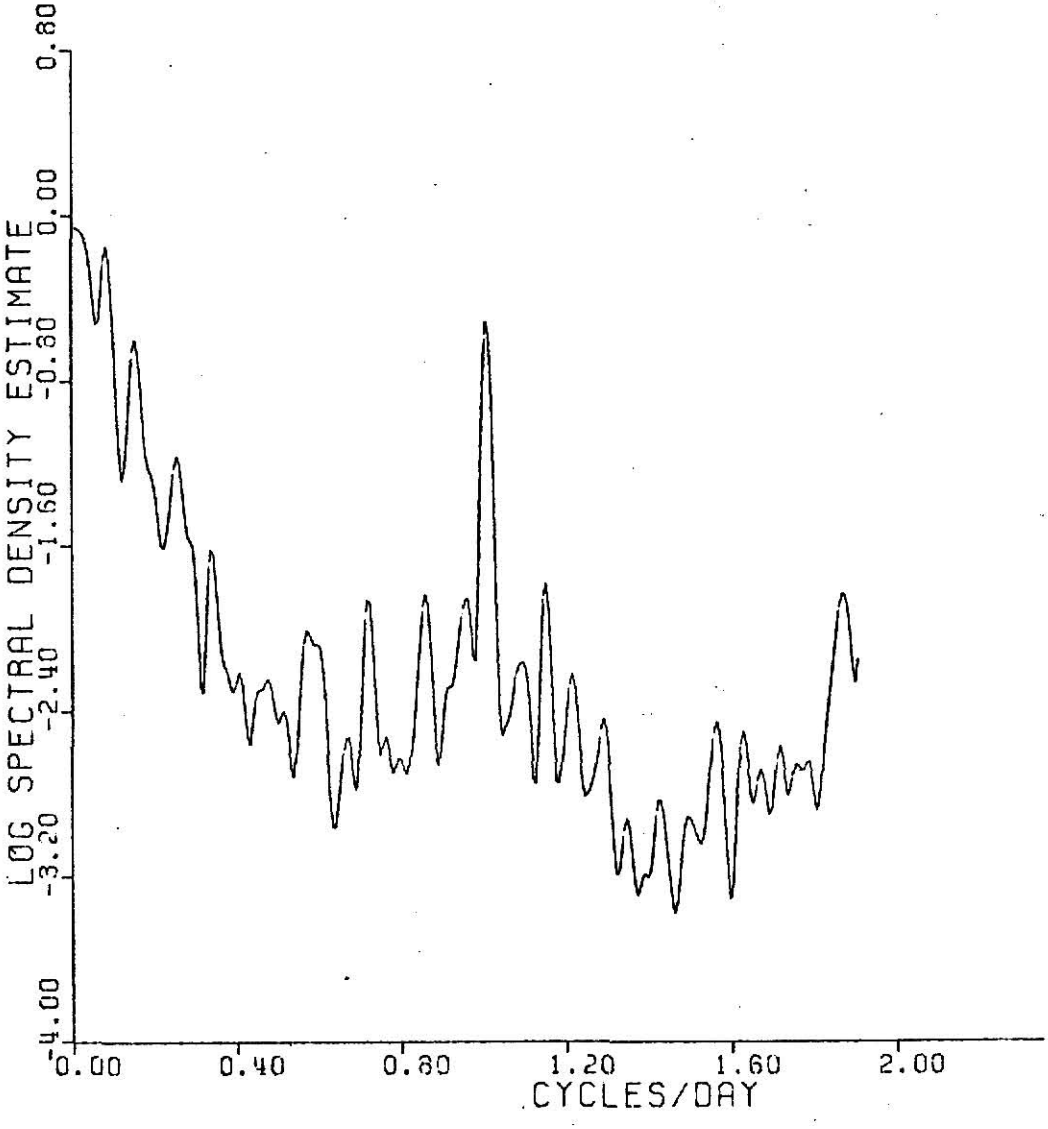

FIGURE 39. Spectral density function for carbon monoxide based on six hours averaged data, station 1.

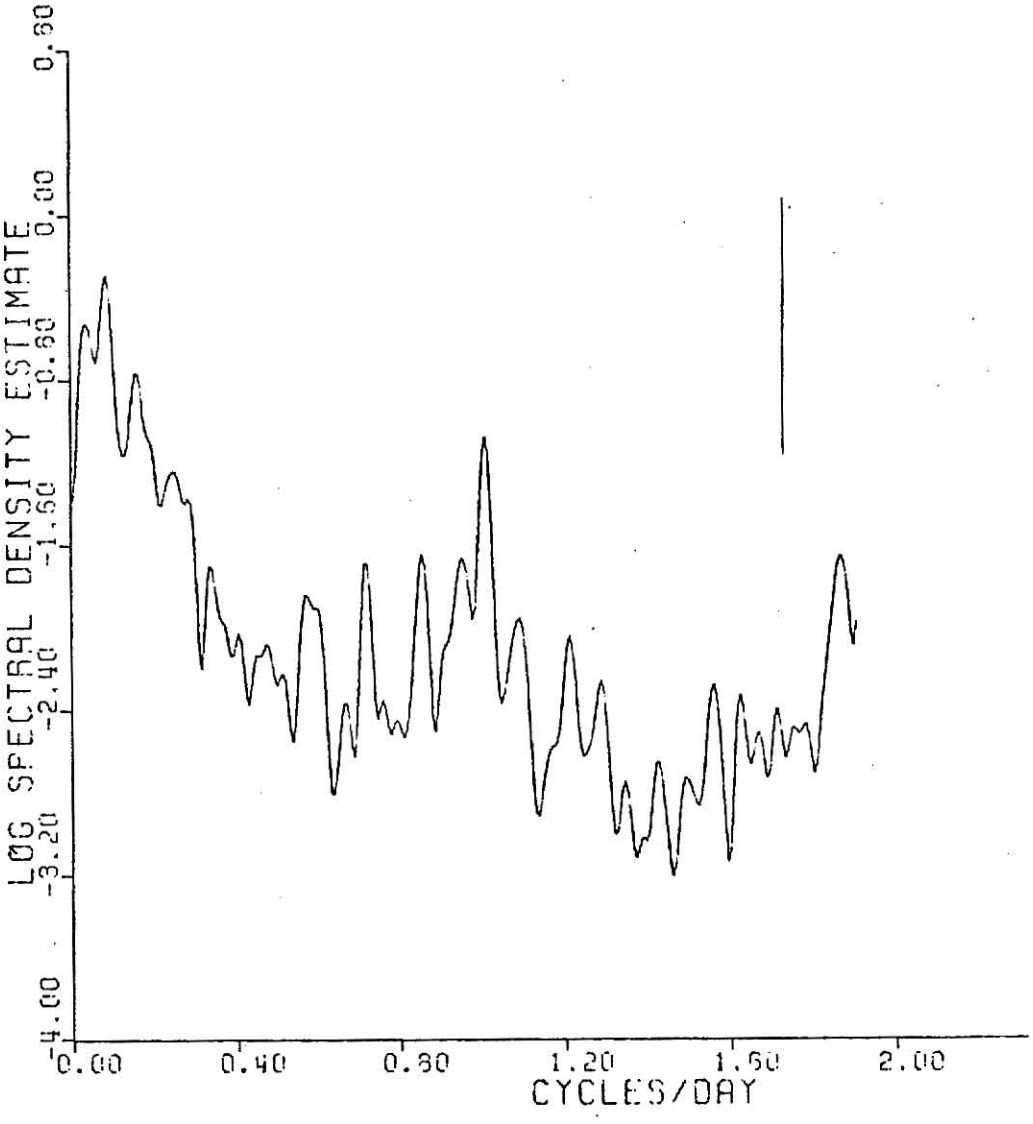

FIGURE 40. Spectral plot of residuals for six hours averaged data from station 1.

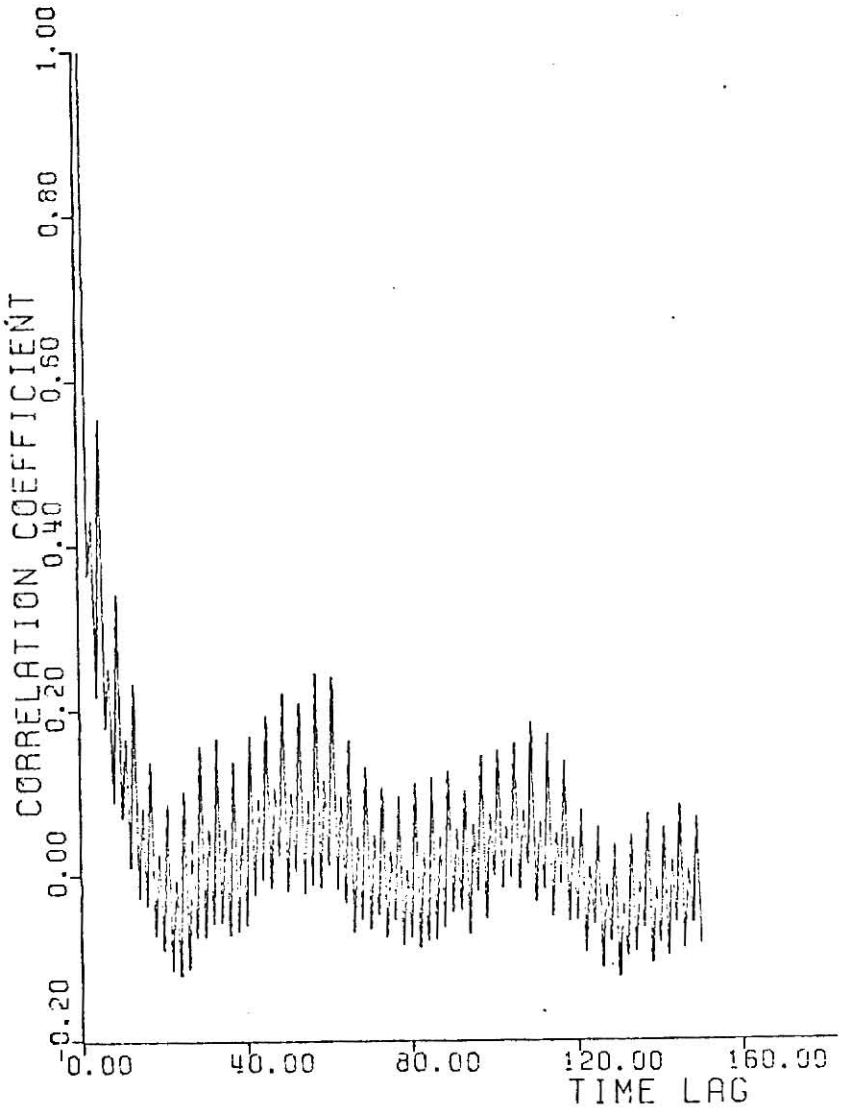

FIGURE 41. Autocorrelation function of carbon monoxide from station 71.

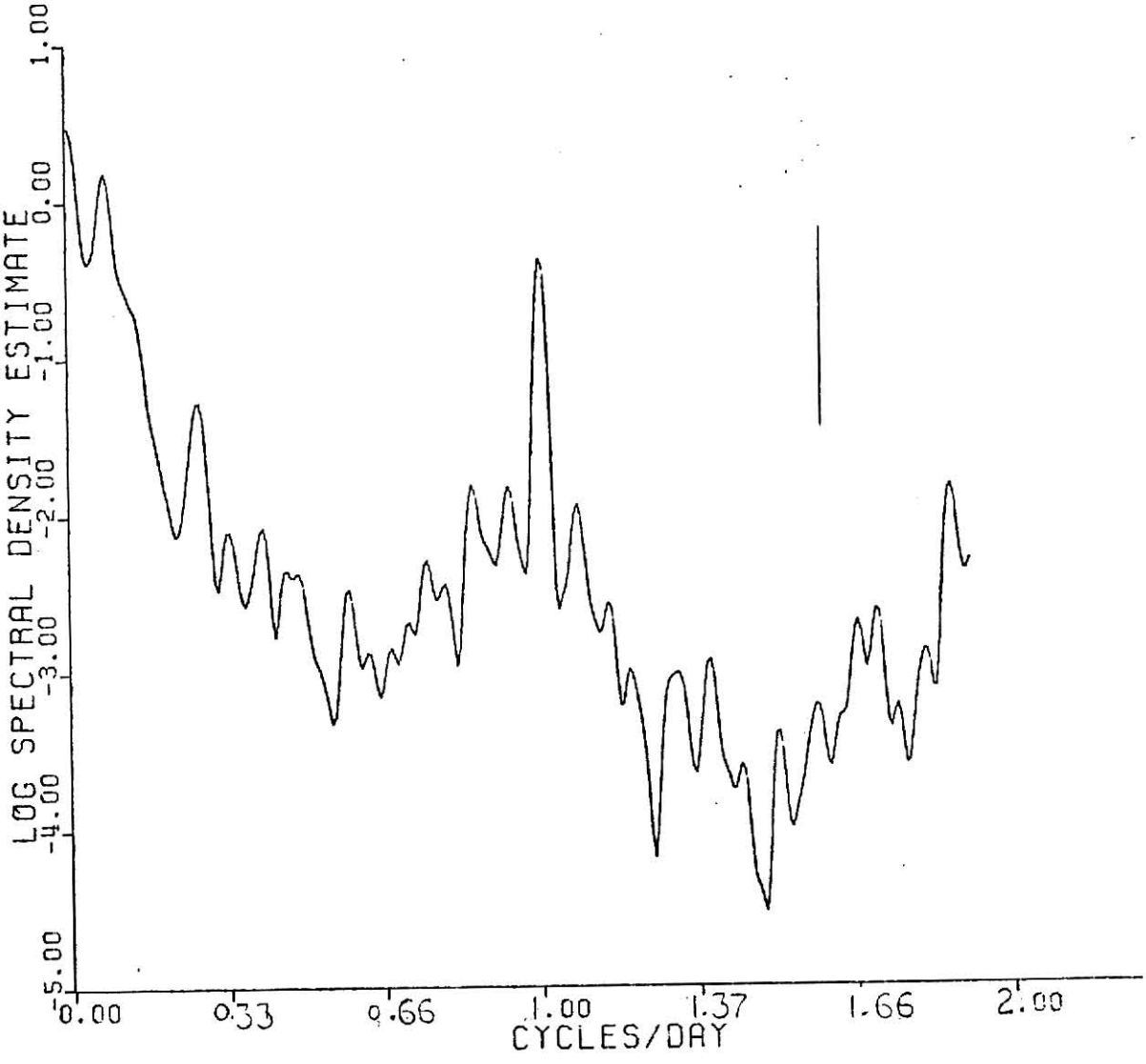

FIGURE 42. Spectral density function for carbon monoxide based on six hours averaged data , station 71.

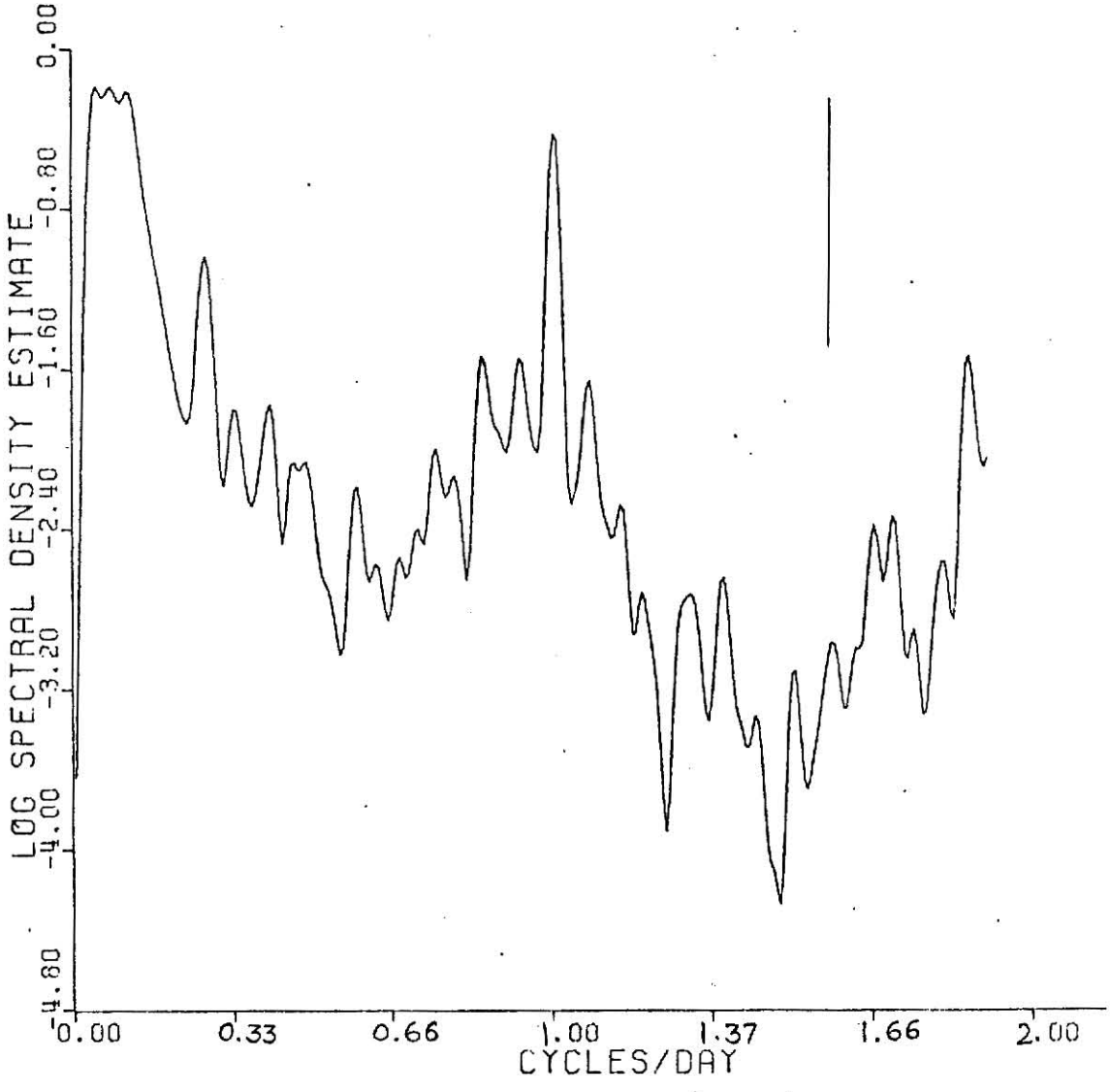

FIGURE 43. Spectral plot of residuals for six hours averaged carbon monoxide data from station 71.

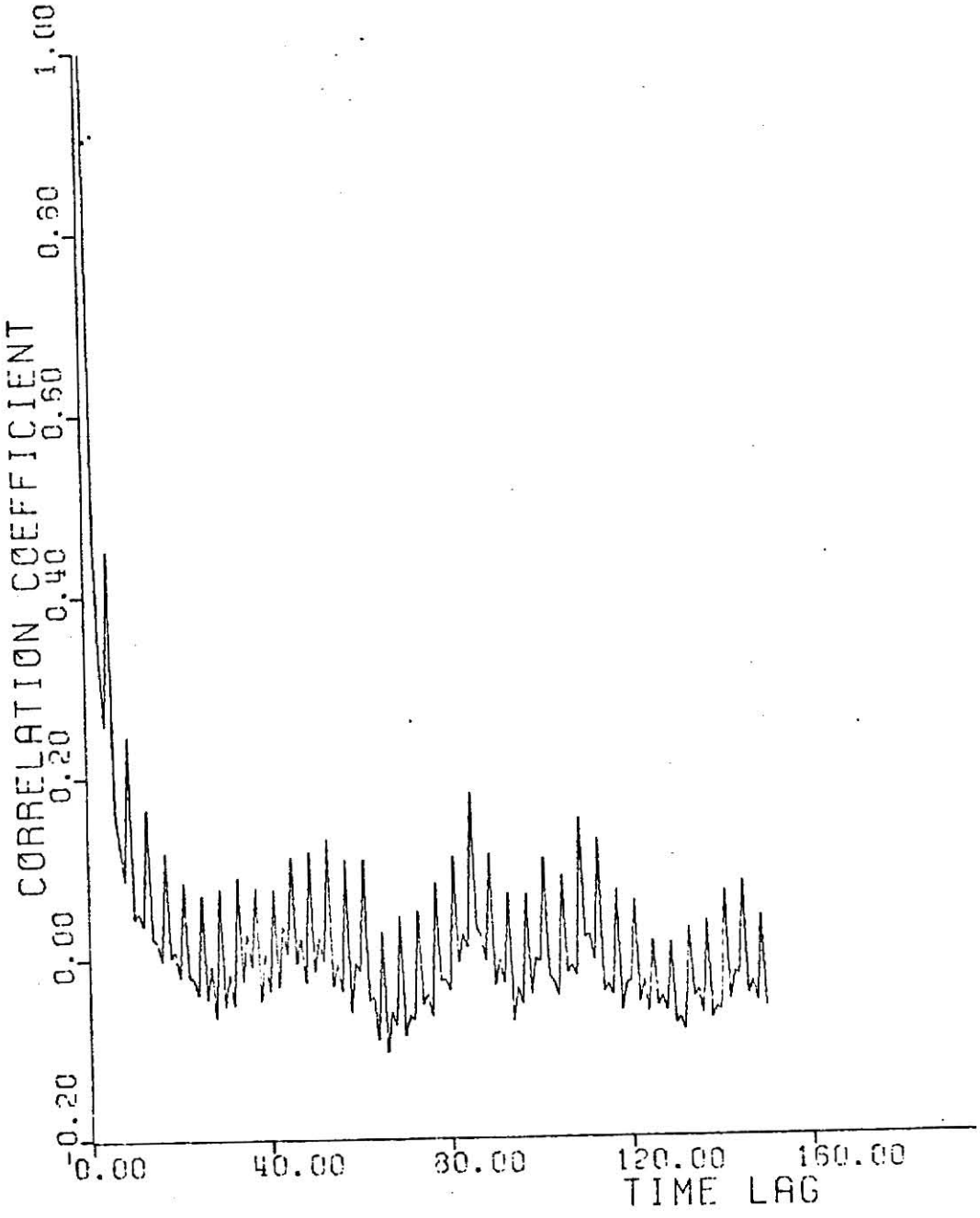

FIGURE 44. Autocorrelation function of nitrogen oxides from station 60.

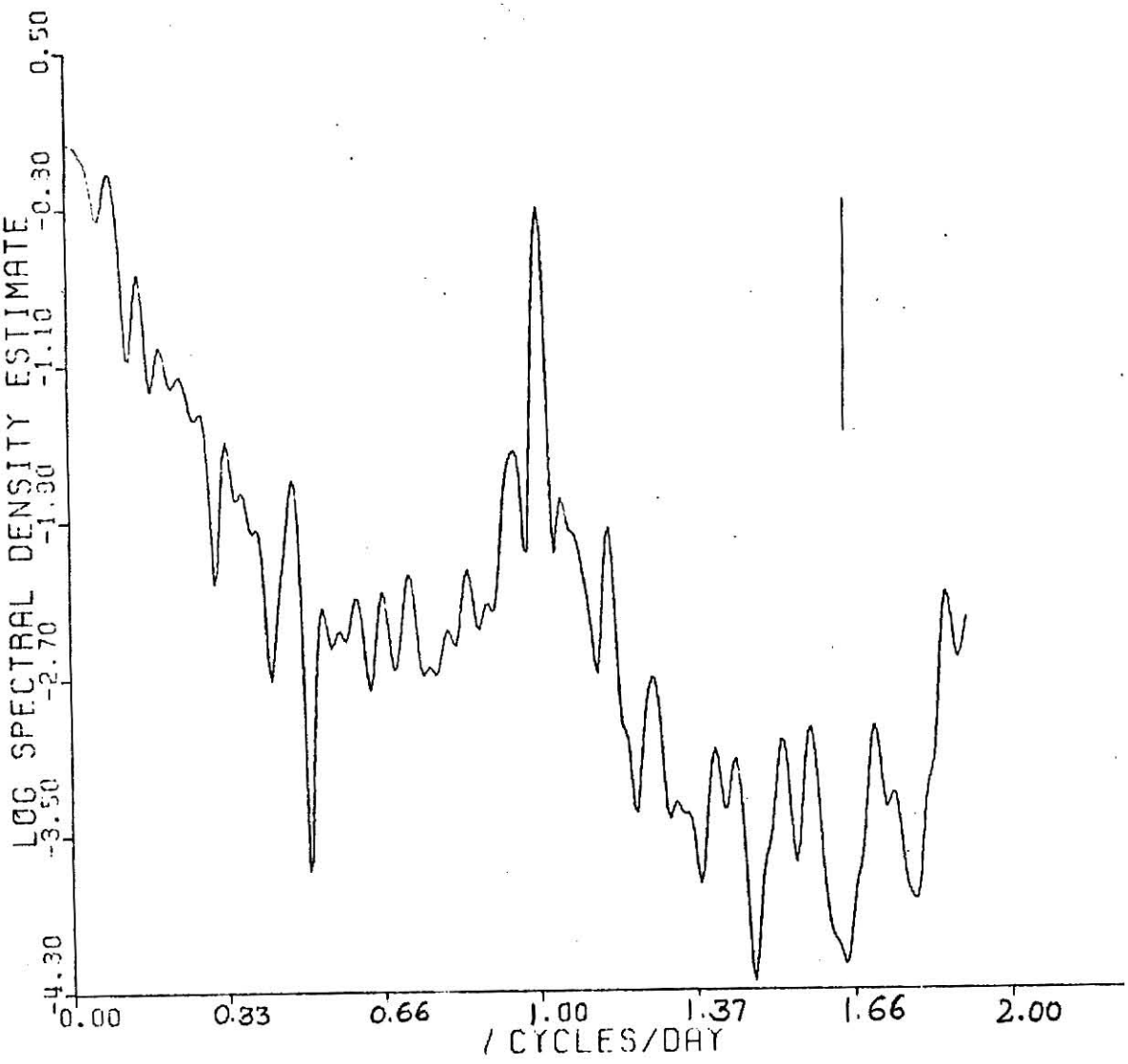

FIGURE 45. Spectral density function for nitrogen oxides based on six hours averaged data, station 60.

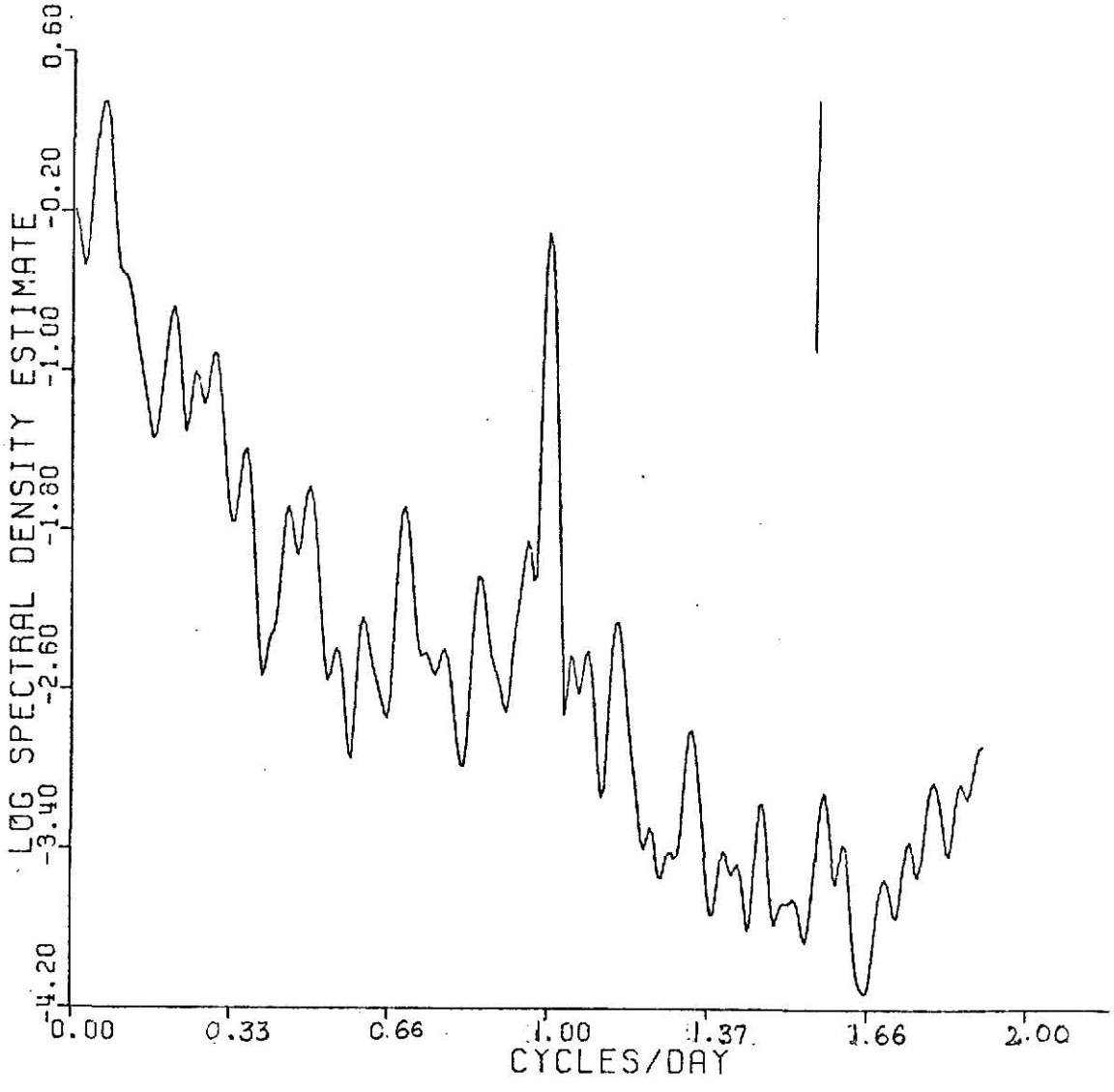

FIGURE 46. Spectral plot of residuals for six hours averaged nitrogen oxides data from station 60.

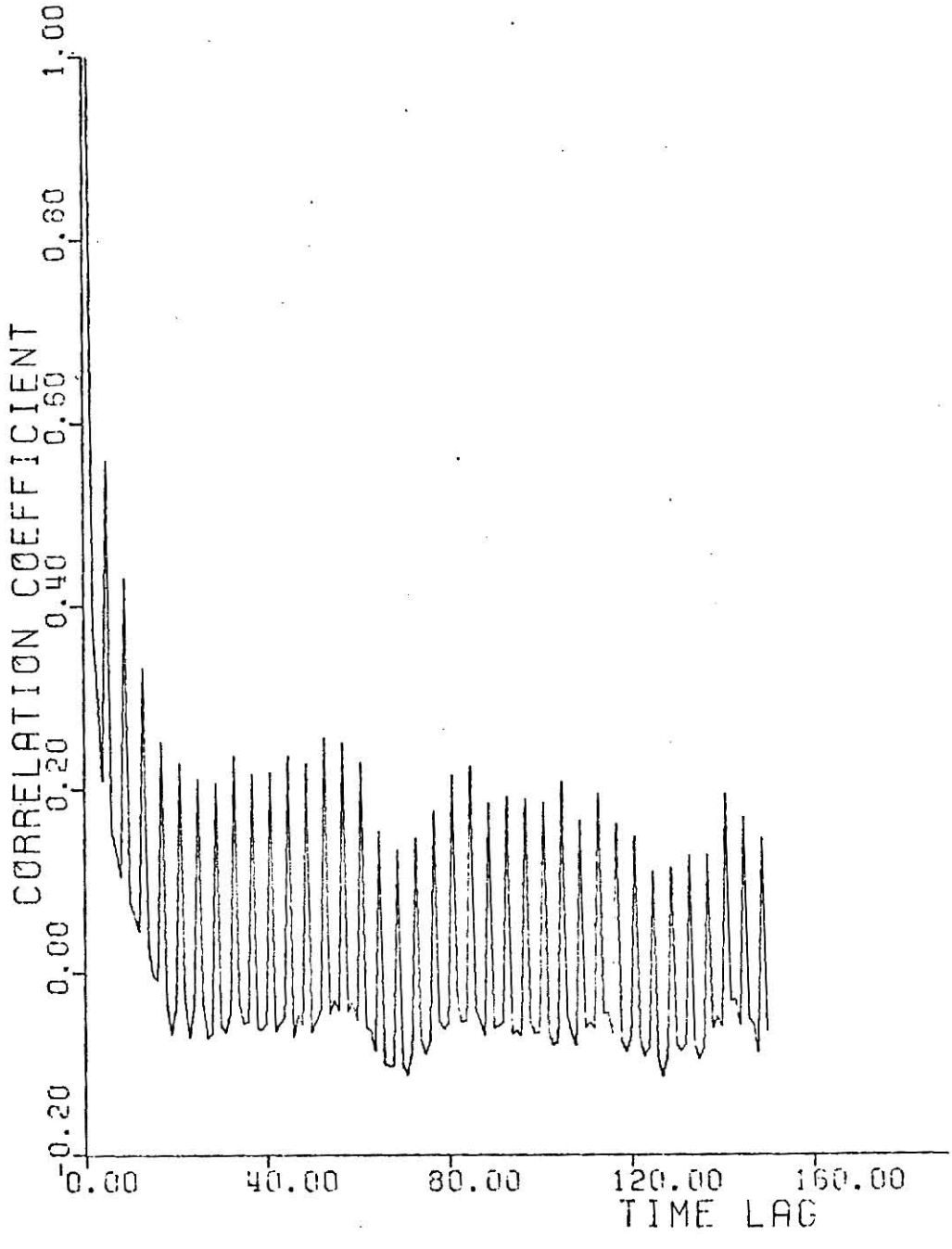

FIGURE 47. Autocorrelation function of nitrogen oxides from station 79.

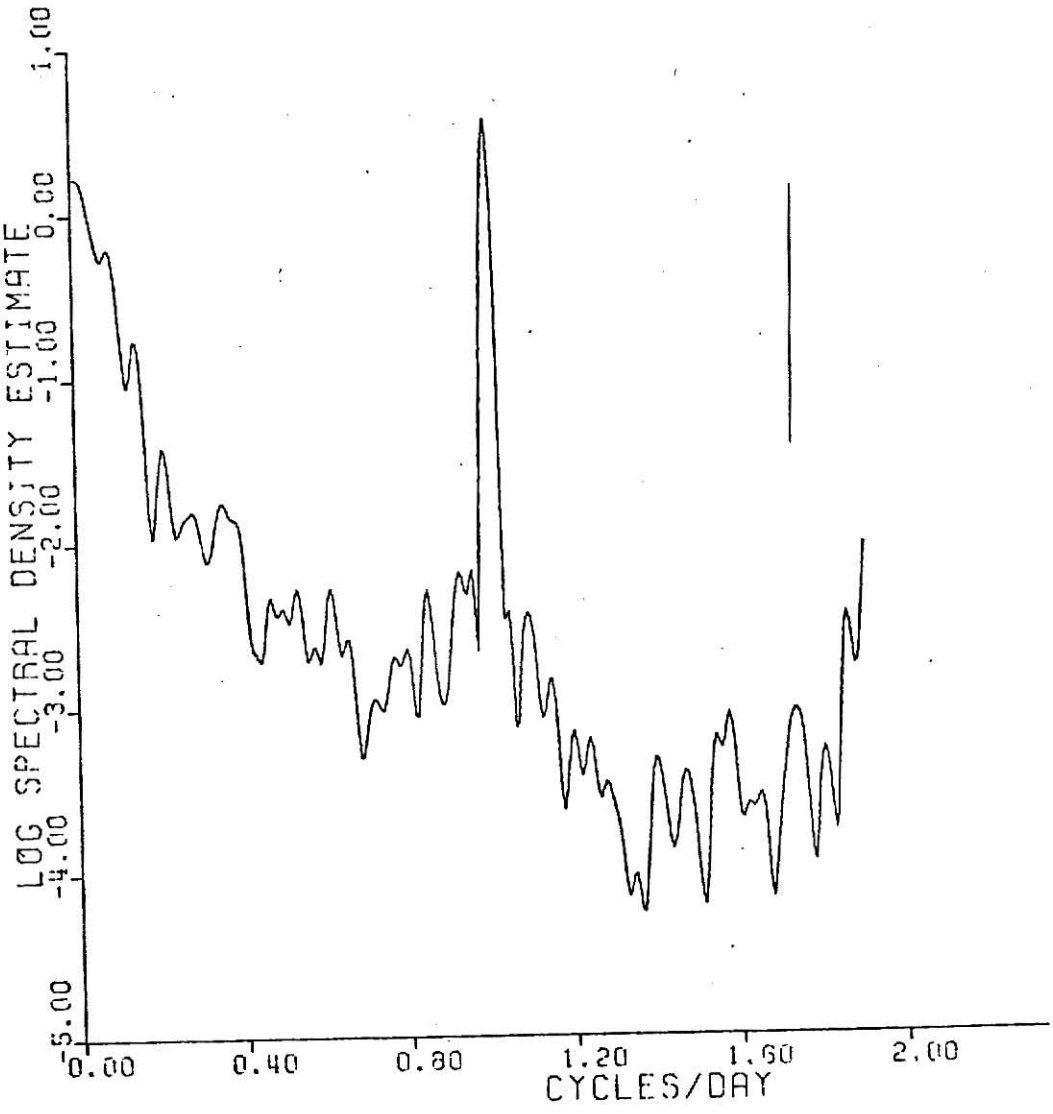

FIGURE 48. Spectral density function for nitrogen oxides based on six hours averaged data, station 79.

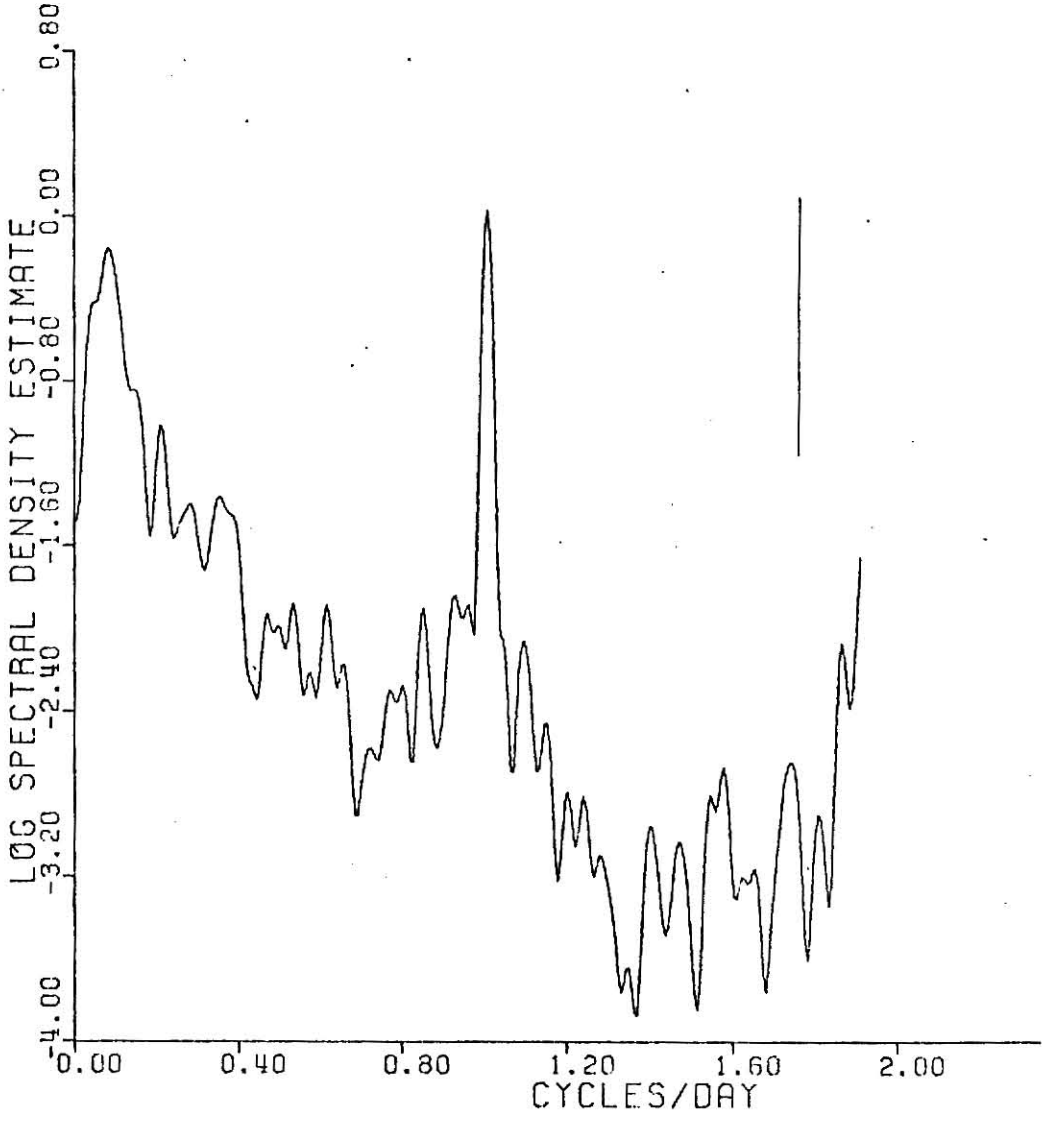

FIGURE 49. Spectral plot of residuals for six hours averaged nitrogen oxides data from station 79.

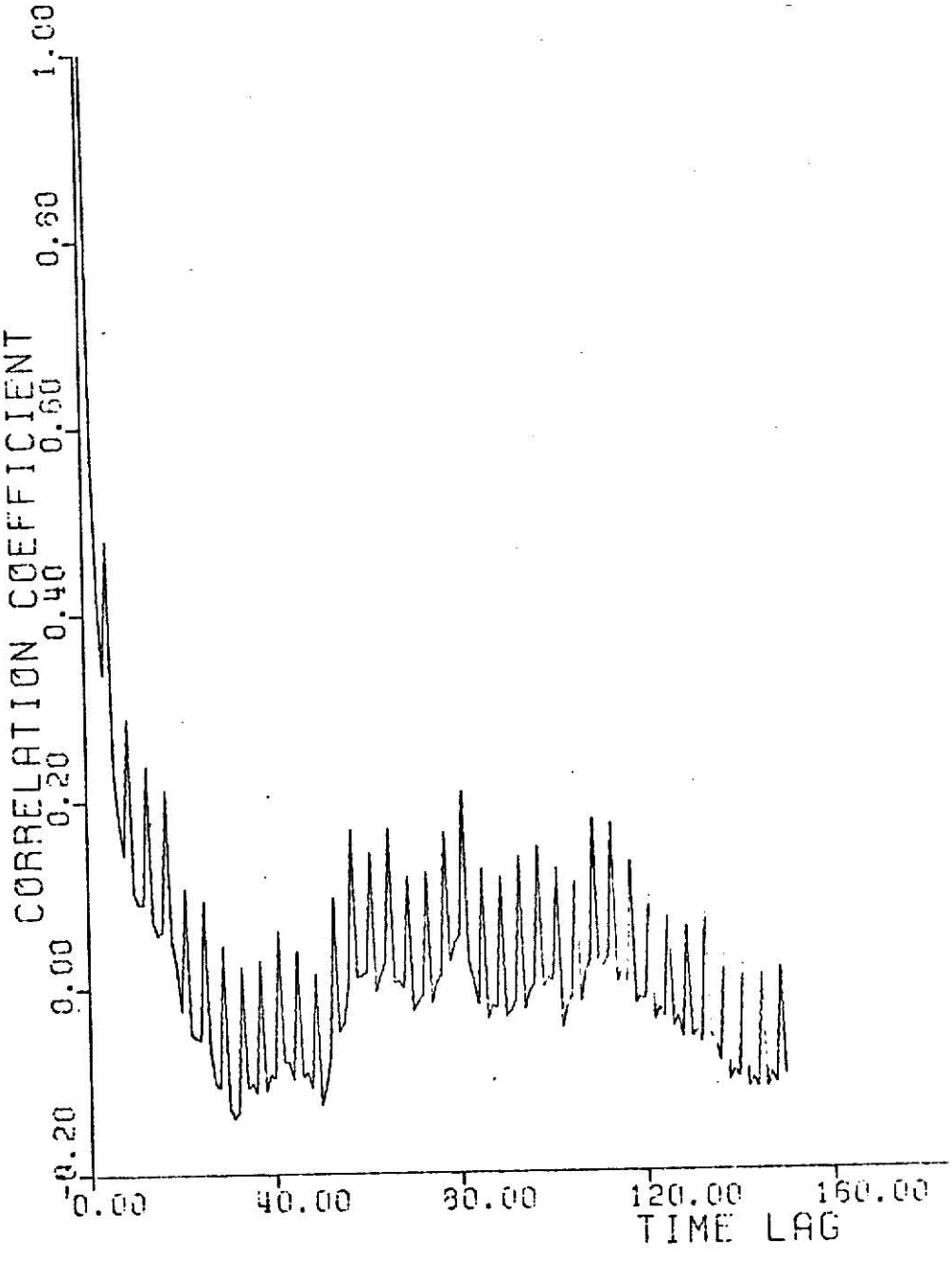

FIGURE 50. Autocorrelation function of nitrogen oxides from station 1.

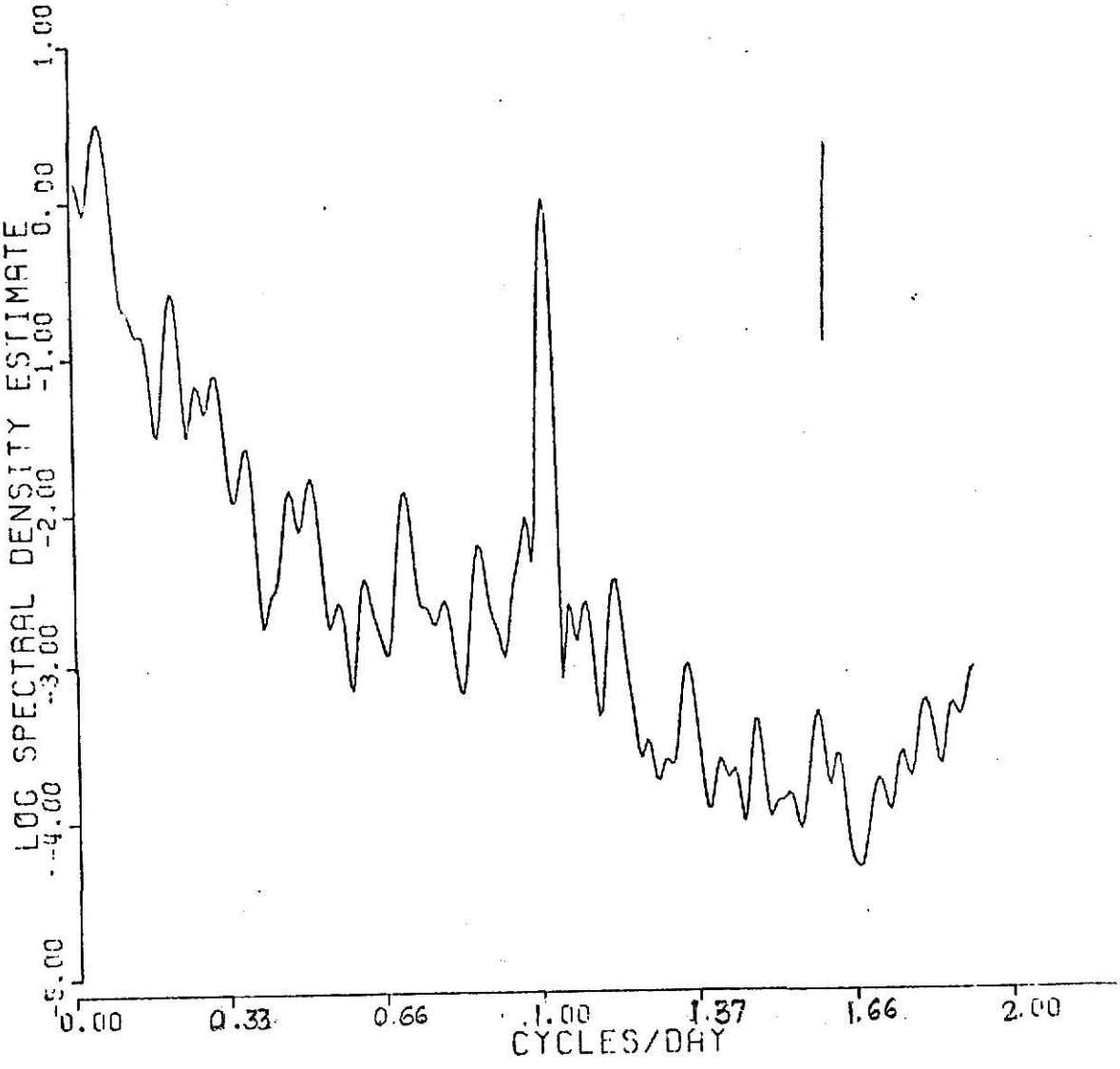

FIGURE 51. Spectral density function of nitrogen oxides based on six hours averaged data, station 1.

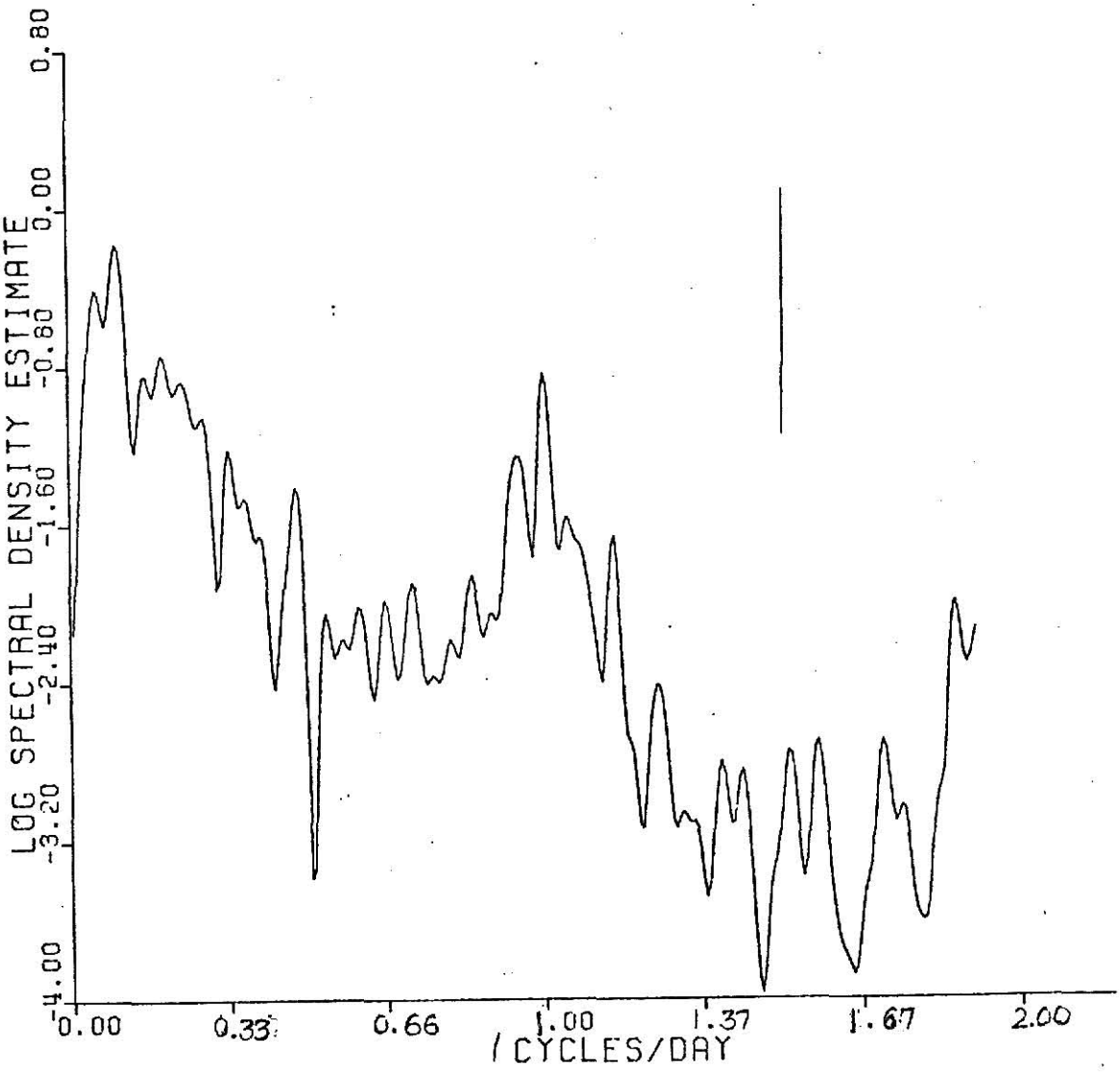

FIGURE 52. Spectral plot of residuals for six hours averaged nitrogen oxides data from station 1.

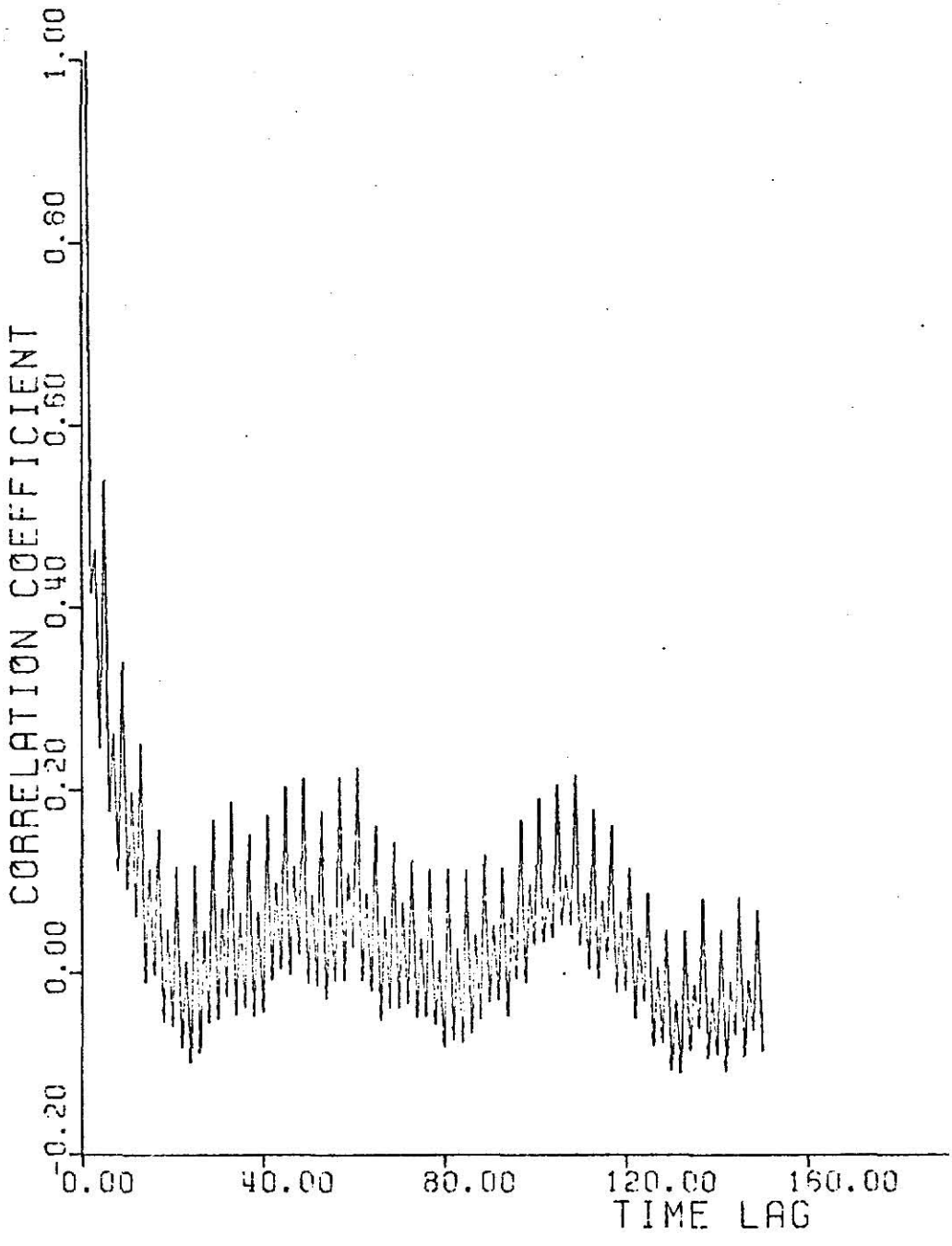

FIGURE 53. Autocorrelation function of nitrogen oxides from station 71.

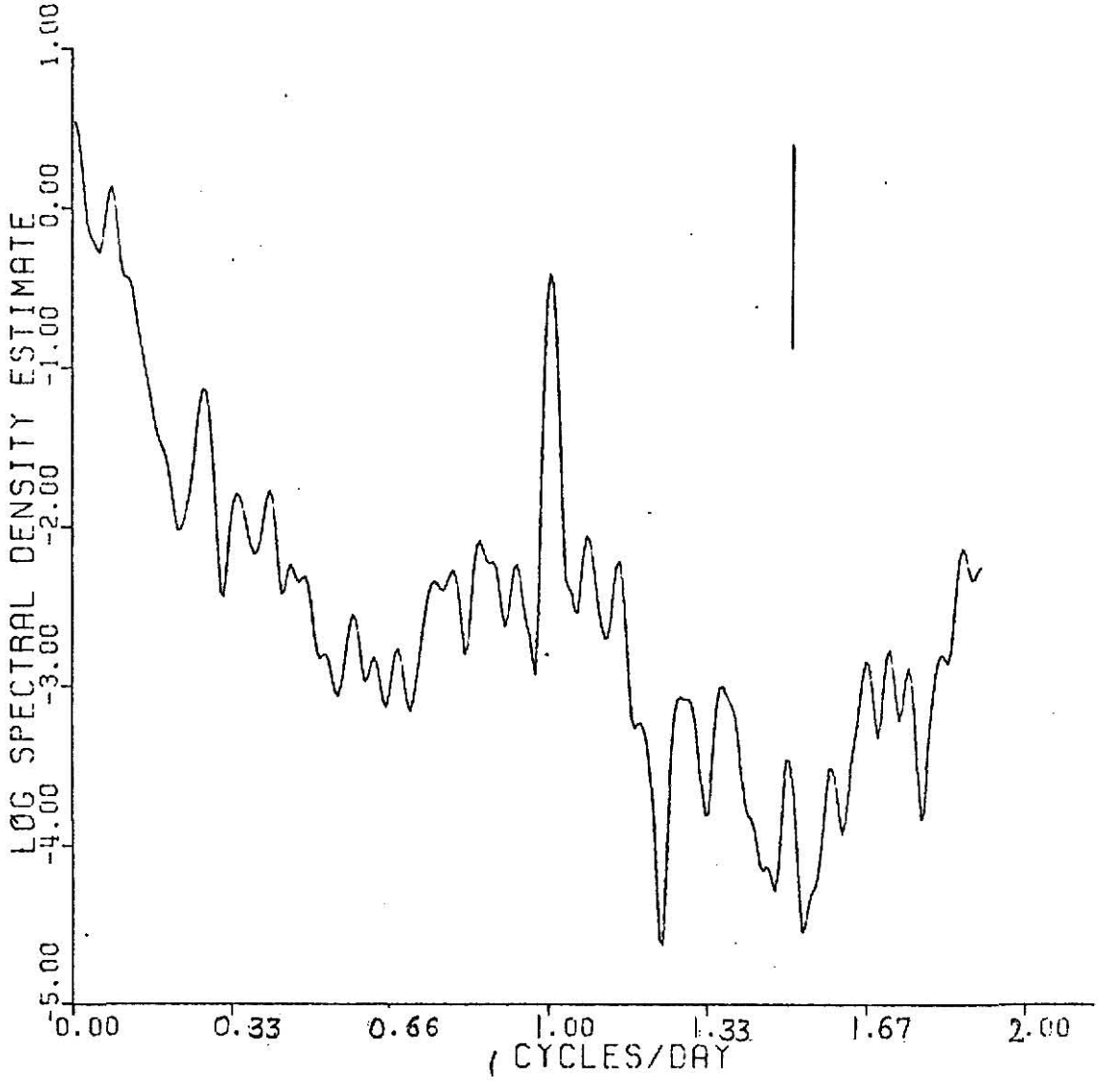

FIGURE 54. Spectral density function of nitrogen oxides based on six hours averaged data, station 71.

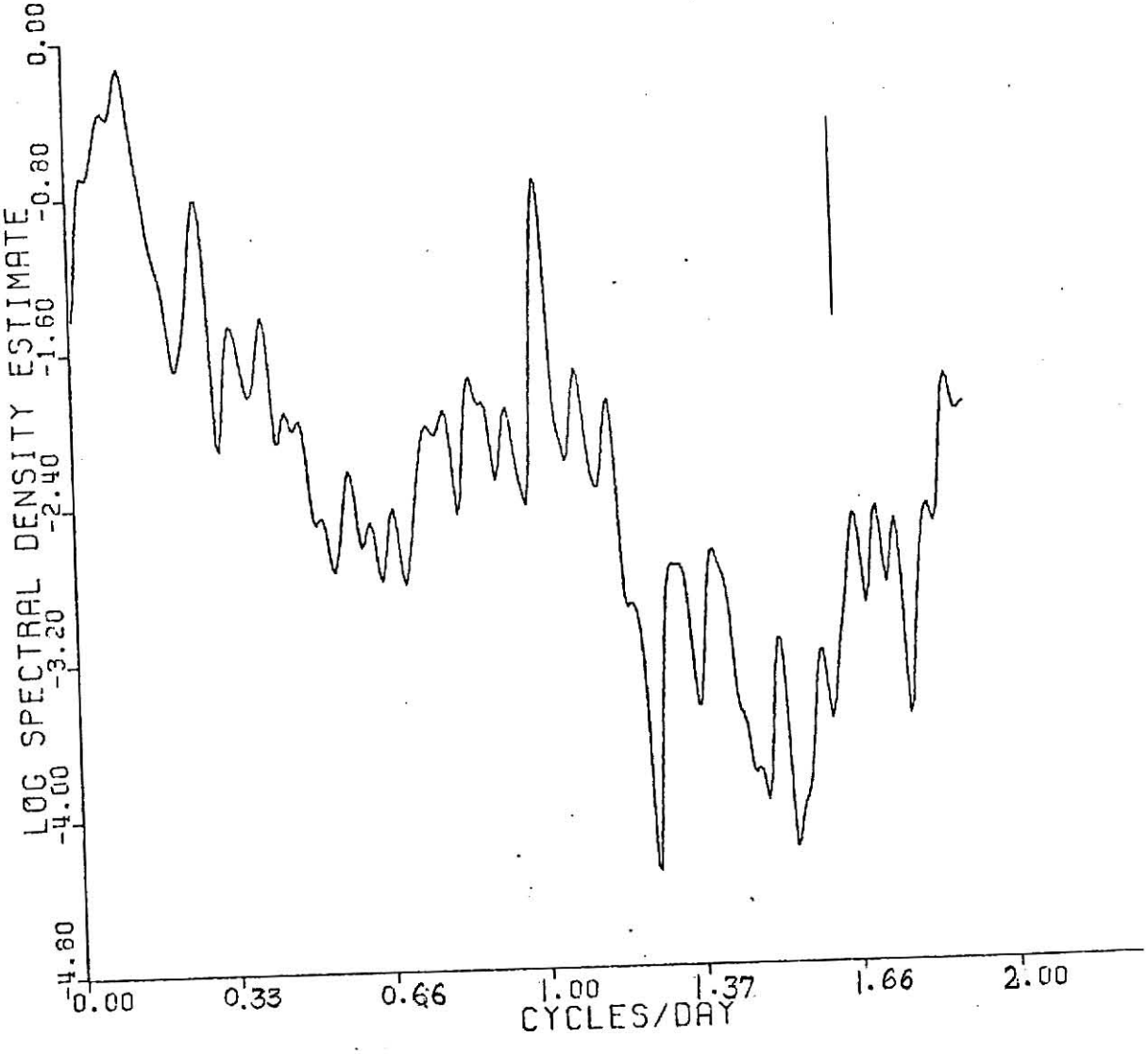

FIGURE 55. Spectral plot of residuals for six hours averaged nitrogen oxides data from station 71.

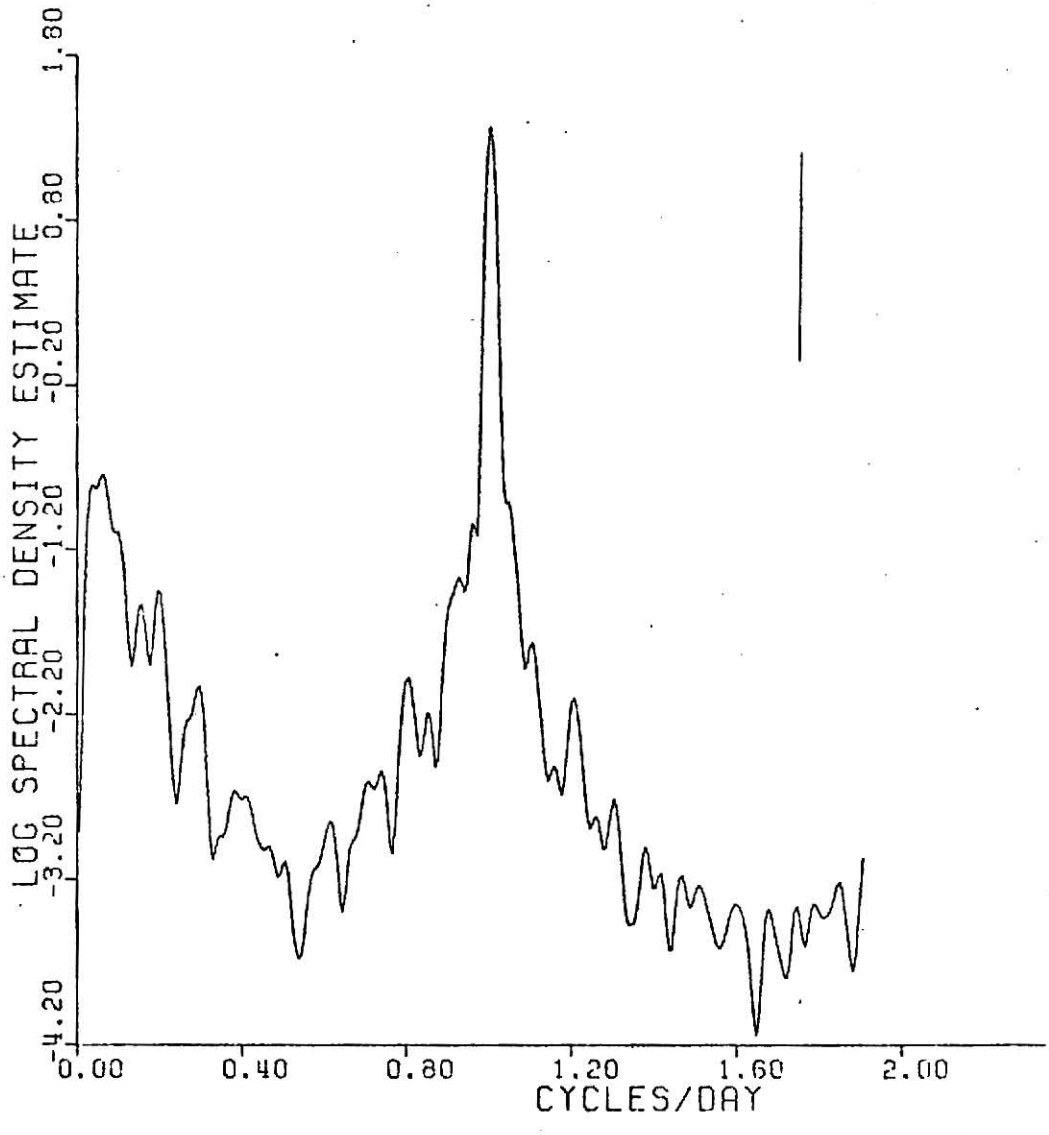

FIGURE 56. Spectral plot of residuals for six hours averaged ozone data from station 60.

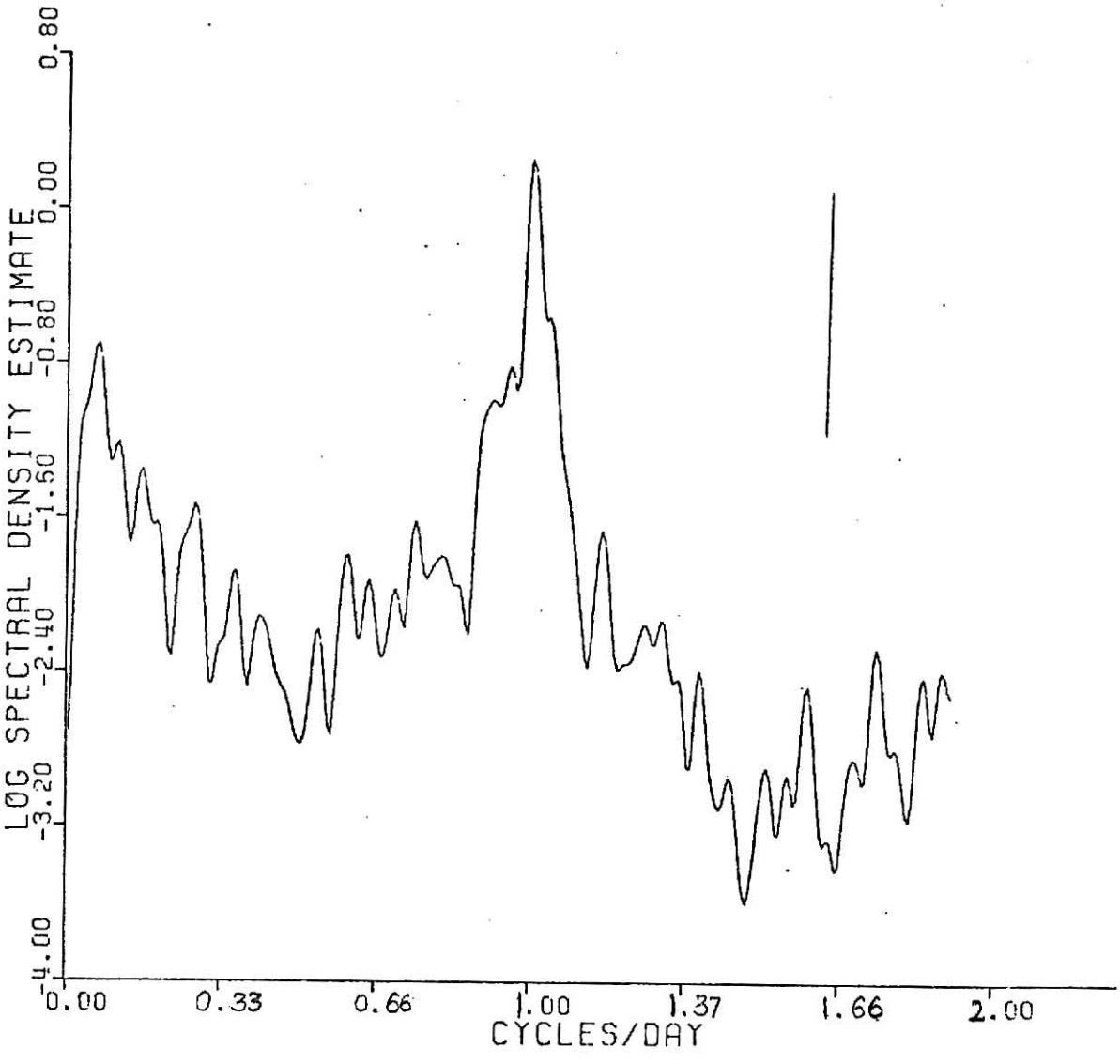

FIGURE 57. Spectral plot of residuals for six hours averaged ozone data from station 79.

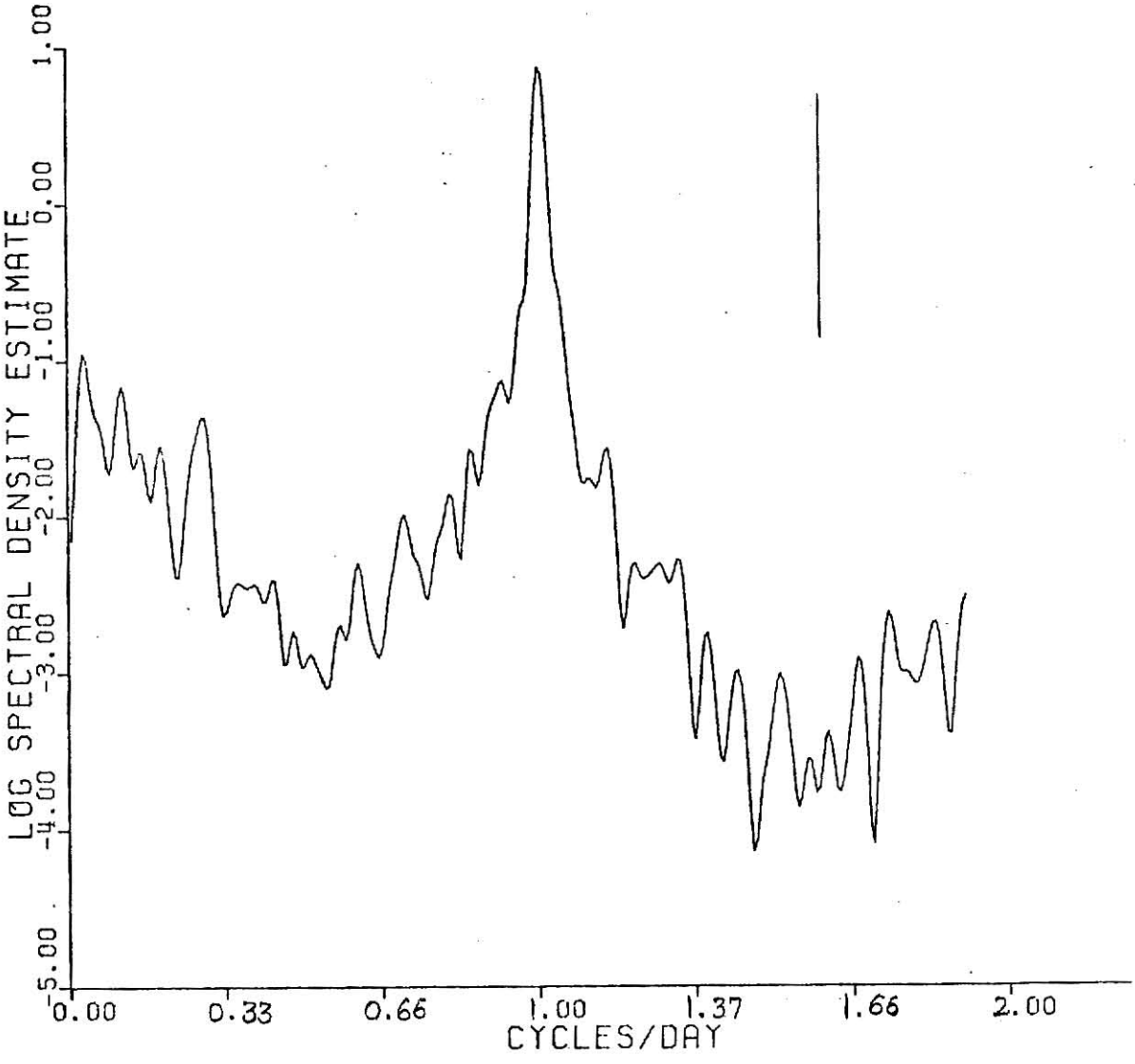

FIGURE 58. Spectral plot of residuals for six hours averaged ozone data from station 1.

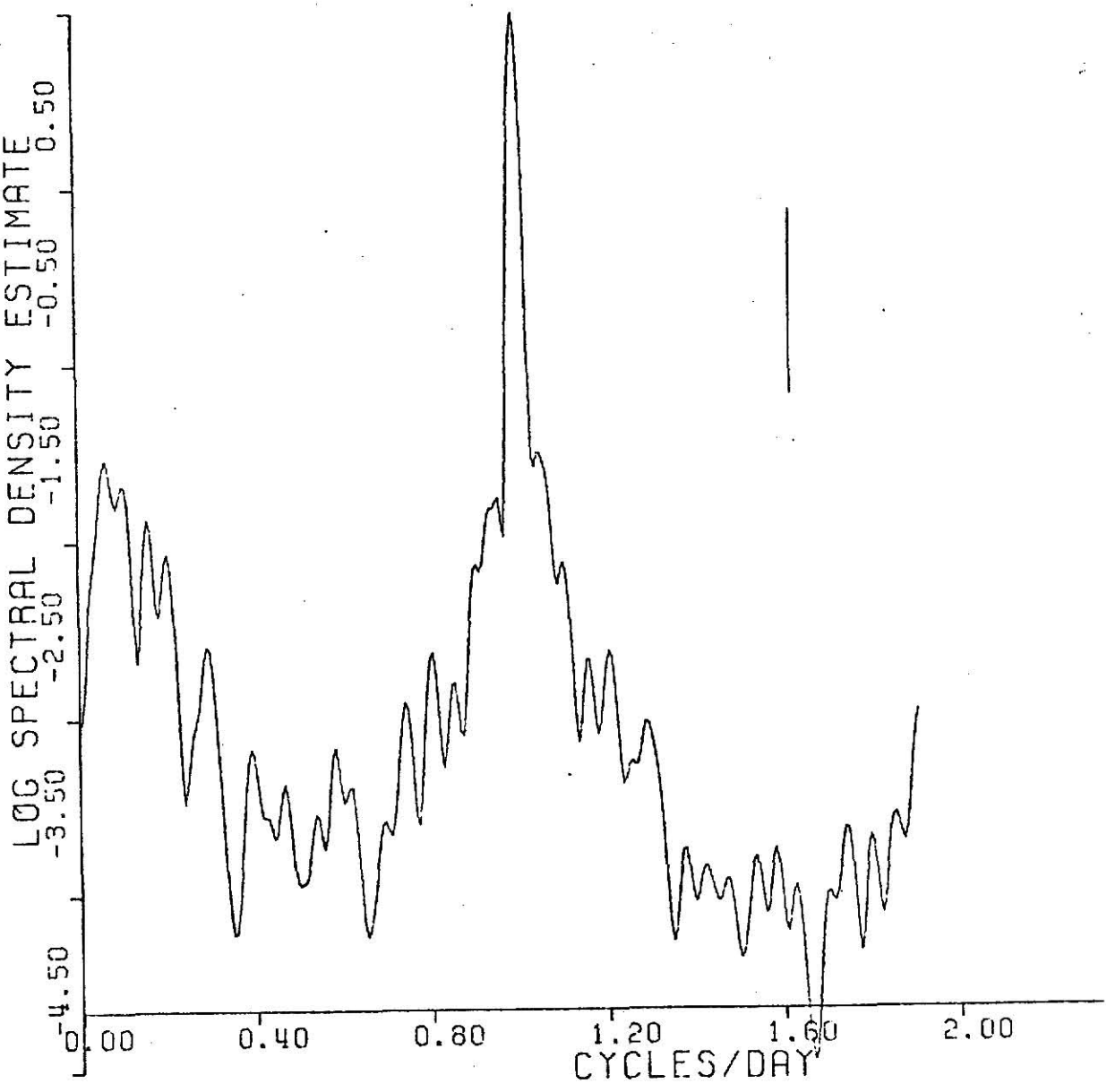

FIGURE 59. Spectral plot of residuals for six hours averaged ozone data from station 71.

TIME SERIES ANALYSIS AND PARAMETRIC ESTIMATION IN AIR POLLUTION MODELING

by

SUBHASH CHANDER SARIN

BSc. (Mech. Eng.), University of Delhi, India, 1970

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of

the requirements for the degreee

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1973

ABSTRACT

In this paper two approaches are used for the modeling of air pollution. The first approach consists of dividing a metropolitan area into different regions. A dynamic model is then developed considering the material balance around each region. This approach is illustrated by using a metropolitan area consist of two regions. It is assumed that the pollutant concentration is uniform in each region, no loss occurs as it travels from one region to another, and wind speed is assumed uniform within each region. The unknown parameters in the model are determined using quasilinearization.

The second approach is the statistical modeling of air pollution. The air pollution data from a region contains implicitly most of the information from which the characteristics of the air pollution phenomenon can be obtained. Most of the factors like social activities, change in production, change in meteorological conditions and inversion layer base height among others causing air pollution have cyclic variations. The characteristics of air pollution at a location can be determined knowing these cyclic variations. It was the purpose of this study to determine those cyclic variations and thereby developing a mathematical model. Spectral analysis was used for the determination of these cyclic variations. Using regression analysis these cyclic variations were used for the development of the mathematical model to predict the pollutant concentration at a future time t. Further analysis was done to determine how the pollutant concentration at one station is affected

by the concentrations of neighboring stations. This was done using cross spectral analysis. Also the relative distribution of the various sources at a station was determined by the cross spectral analysis for the different pollutants at the same station.

The data analyzed were furnished by the Bay Area Air Pollution control District, San Francisco and Air Pollution control District, County of Los Angeles, California. Many of the cyclic variations analyzed can be associated to the social activities. However for the Los Angeles data it has been found from the analysis of the wind speed and inversion layer base height data that many of the cyclic variations in the pollutants concentrations are caused by the variations in wind speed and inversion layer base height.