THE UNIVERSITY OF ALBERTA

TIME SERIES MODELS OF URBAN AIR POLLUTION

by

© EBENEZER AMOLE OWOLABI

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

IN

COMPUTING SCIENCE

DEPARTMENT OF COMPUTING SCIENCE

EDMONTON, ALBERTA

SPRING, 1973

ABSTRACT

This thesis presents two methods for modelling urban air pollution using the time series analysis technique. This technique is different from the atmospheric diffusion technique commonly applied to modelling air pollution, a brief review of which is given in Chapter I along with other works in urban air pollution. The two types of models developed here are the stochastic and the dynamic system models, which are established to describe the behavior, in Edmonton, of a particular air pollutant called oxides of nitrogen. Since models from Time Series Analysis are based on observed data, Chapter II is devoted to describing the source, precision, and organization of the data used to develop the models. In Chapter III, the stochastic models are developed, tested, and used for forecasting while Chapter IV is devoted to the development of the dynamic system models. With the oxides of nitrogen data collected in Calgary, Sarnia, Sudbury, Toronto, and Windsor, stochastic models explaining the pollutant's behavior in those cities are developed in Chapter V and compared with the corresponding model for Edmonton. This comparison resulted in a general stochastic model for the behavior of oxides of nitrogen in the urban atmosphere. Finally, in Chapter VI recommendations and suggestions concerning the use of the models, and further use of time series analysis technique in building urban air pollution models are given.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to Professor U. M. von Maydell, my supervisor, whose suggestions and advice were invaluable at all stages of this study, and to Professor K. D. Hage of the Department of Geography-Meteorology for his helpful discussions. I wish to thank the Canadian International Development Agency for financial assistance in carrying out this research, and the National Research Council for providing computer funds through NRC grant #A7915. Finally, I wish to acknowledge the assistance of the Environmental Health Services Division of the Alberta Department of Health, Edmonton, and the Ministry of the Environment, Toronto, Ontario, in making urban air pollution data available for the study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SERIES

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Constituents of air pollution include dust, fumes, gas, mist, odor, smoke, vapor and noise. The mere existence of any of these pollutants in ambient air does not necessarily constitute pollution. Air pollution occurs when the presence of one or more of them is in such quantity, characteristic, and duration capable of damaging or unreasonably interfering with comfortable enjoyment of life and property. The injurious levels of concentration of most of the pollutants as well as the sources and causes of the pollutants have been identified by researchers, several of whom are mentioned in the following sections. However, knowing the source and concentration is not enough, it is also desirable to know the fluctuation of the pollutants' concentration with time and space. That is, one would like to know in advance when a particular pollutant is likely to reach a dangerous level in the atmosphere so as to prevent it from doing so wherever possible. Such knowledge is possible to obtain through models that explain the behavior of the pollutant over time. There are several air pollution models in existence using the diffusion approach. As predictive models, they give either consistently low estimates or consistently high estimates of the pollutants as explained in the symposium report edited by Atkisson and Gaines [2].

The object of this thesis is to present a different approach to air pollution modelling. This approach employs time series analysis. The measurements of a particular pollutant observed over a long period of time will be studied and analyzed. The analysis will yield a model of the stochastic process which generates the behavior of a pollutant in an urban area. A transfer function model of a dynamic system with the measurements of the pollutant as output will also be built. The models will be used in forecasting and their forecast estimates will be compared with the observed values of the pollutant to prove the models' potential as a predictive device.

The air pollutant that will be so investigated is the gas known as oxides of nitrogen (NO_x). The behavior of this pollutant in Edmonton will be studied and compared with the behavior of the same pollutant in some other cities in Canada.

Time series analysis requires a great deal of computation including inversion of matrices, iterations involving solutions of sets of linear equations, and numerical estimation of derivatives. Since a long series of observations is required to build the models, time series analysis requires facilities for large data processing. Therefore, useful results are obtainable from this type of analysis only if a digital computer is employed. The investigator must have both practical and theoretical knowledge of the computer science discipline including numerical analysis, and infor-

mation retrieval. The problem investigated in this thesis, therefore, is in the area of the interface of computing science with one of today's main problems affecting all of humanity.

1.2 Measurement and Control of Urban Air Pollution

1.2.1 Sources

In order to measure and control air pollution, various sources of the pollutants must be known. Sources have been recognized as Natural Air Pollution Sources, and Man-made Sources as discussed in Rossano [61]. Some examples of natural sources are swamps producing gases and odors, forest fires yielding smoke and flyash, and wind blowing dust and pollen. Sources of man-made pollution cover a very wide range. Only some of the major ones affecting urban environment will be mentioned here.

One of the major man-made sources of air pollution is combustion. Combustion sources can be categorized as follows:

(a) Fuel burning in home heating units and power plants.

(b) Motor vehicles represented by autos, buses and trucks.

(c) Refuse burning in community and apartment house incinerators.

Pollutants emitted by combustion are oxides of sulfur, oxides of nitrogen, carbon monoxide, smoke, flyash, metal oxide, particles and odor.

Manufacturing processes as sources of air pollution may be classed into two categories:

(a) Metallurgical plants, represented by smelters, steel mills, and aluminium refineries, and

(b) Chemical plants, represented by petroleum refineries, pulp mills, super phosphate fertilizer plants, and cement mills.

Pollutants associated with manufacturing processes include particles, oxides of nitrogen, oxides of sulfur, hydrogen sulfide, fluorides, organic vapor, and odor.

Some other well known man-made sources include the following:

(i) Nuclear energy activities producing radioactive fall out.

(ii) Dust producing processes like crushing, grinding, demolition, and milling, which generate mineral and organic particulates.

(iii) Agricultural activities like crop spraying, field turning, and straw burning, which produce pollutants like organic phosphates, chlorinated hydrocarbons, smoke and flyash.

The number and the spatial distribution of these sources are very important in measurement and control. For this purpose the sources have been categorized into the following emission types:

(a) Point sources: these sources have a high rate of emission and can be easily recognized. Examples are power plants, petroleum refineries, and steel mills.

(b) Area-wide (or multiple) sources: these consist of a

large number of smaller sources distributed over a well-defined area like an entire residential area.

(c) Line sources: these for example include freeways, highways and arterials carrying a steady stream of moving vehicles.

1.2.2 Air Pollutant Measuring Devices and Units of Measurement

There are two main methods for the determination of airborne contaminant concentration. One is by remote sensing and the other is by removal of particulates and gases from the gas stream.

Remote sensing and analysis does the measurement on the material in situ. The advantage of this method is that it does not involve physical contact with the substance and hence no error from alteration or modification is involved. Remote infra-red sensing of sulfur dioxide in stack plumes is an example of this technique.

The second method collects particulates and gases from a gas stream, then subjects the material collected to analysis. For gases, a sample of the air stream is absorbed in specific reagent liquids before wet chemical analysis is performed. There are both manual and automatic devices using this method. The automatic devices known as Automatic Continuous Analyzers are more desirable in air pollution monitoring. They continuously perform the following operations in sequence: collect air samples, analyze them, and record

the results either in form of a curve on a strip chart or by punching the results on paper tape. These automatic devices are available for only a few of the gas pollutants. However, one of such pollutants for which they are currently available is oxides-of-nitrogen. More detailed description of these and other devices can be found in Rossano [61], while names of some of the commercially available brands can be found in Magill et al. [35].

The commonly used units of measurement in air pollution monitoring are parts per million parts of air measured by volume (ppm), and micrograms per cubic meter of air ($\mu g/m^3$). Gas pollutants are usually measured in ppm (or sometimes parts per hundred million (pphm) while particulates are measured in $\mu g/m^3$. In the OECD [50]report, these and other units of measurement are defined.

1.2.3 Control

The main objectives of measuring and analyzing air pollutants are to be able to obtain information for setting control standards and to be able to keep the pollutants within the control limits. There are tolerance thresholds with the main air pollutants. These help in setting air quality standards for the pollutants. Such standards for different pollutants are well documented in Atkisson and Gaines [2].

Once the standards are known various techniques can then be applied to ensure that these standards are not violated. One kind of control is to stop the human activity generating the pollutant. Another form of control effects

a reduction in the emission rate of the pollutant. Engineers have been able to effect emission reduction in three ways. The first is by process modification. An example of process modification is the modification of automobile engines (e.g. Chrysler "Clean Air Package" includes a modified carburetor) to prevent the generation of excessive concentrations of carbon monoxide and hydrocarbons. The second method devised by engineers is material substitution. Substitution of, for example, a low-sulfur fuel such as natural gas for a high-sulfur fuel reduces the emissions of sulfur dioxide. The third method is gas cleaning. Gas cleaning employs three methods for removing gaseous contaminants; absorbing the pollutant into a liquid, adsorbing the pollutant onto the surface of a solid, and chemically changing the pollutant into a non-polluting substance. For the removal of particulate air pollutants from their gaseous media one or a combination of mechanisms and forces are employed. Such mechanisms and forces include gravitational force, centrifugal force, magnetic force and thermal diffusion giving rise to devices like settling chambers, centrifugal collectors, wet collectors, and filters. The descriptions of some of this equipment can be found in Rossano [61].

1.3 Oxides of Nitrogen as an Air Pollutant

1.3.1 Major Air Pollutants

The known major air pollutants are the oxides of nitrogen, of sulfur, and of carbon, particulates, ozone and

oxidant, odor, and organic contaminants. Examples of some of these pollutants are listed below.

Pollutant	Constituents
Oxides of Nitrogen	Nitric Oxide (NO)
	Nitrogen Dioxide (NO_2)
Oxides of Sulfur	Sulfur Dioxide (SO_2)
	Sulfur Trioxide (SO_3)
Oxides of Carbon	Carbon Monoxide (CO)
	Carbon Dioxide (CO_2)
Organic Contaminants	Aldehydes
	Phenols

In this thesis NO_x is studied since it is one of the most 'important' air pollutants - this will become apparent in the discussions below - and because NO_x data are available over long period of time,they are suitable for Time Series Analysis.

The general term, NO_x, includes NO, NO_2, N_2O_4, and N_2O_5. Since N_2O_4 and N_2O_5 exist only in small quantities and are not known to be capable of producing any adverse effects, they are not important by themselves in air pollution. The term NO_x in air pollution often refers to only NO and NO_2 together.

Oxides of nitrogen are introduced to the atmosphere from a variety of sources. A principal source of man-made oxides of nitrogen is the combustion of fossil fuel, which is used in power plants, heating equipment, and internal combustion engines. NO_x is produced by automobile exhausts

owing to inefficient burning of fuel. Details of the emission of NO_x by automobile exhausts and steps taken to control it can be found in Agnew [1]. NO and NO_2 are known to be present in dangerous concentrations (145-1000 ppm) in cigarette smoke. Leithe [33] contains detailed information on cigarette smoke and Lawrence [32] describes efforts to remove NO_x from tobacco smoke.

Production of oxides of nitrogen during combustion can be expressed in the following chemical reaction equation:

$$N_2 + O_2 \rightarrow 2NO.$$

After the initial combustion reaction, nitric oxide further reacts with oxygen to form NO_2 and higher oxides. These reaction processes are well explained by Strauss [69]. Since NO reacts with oxygen at ambient temperature to form NO_2, NO is seldom found in appreciable concentration. Therefore NO_2 is the main component of NO_x.

1.3.2 Toxicology of NO_x

Nitrogen dioxide (NO_2) is especially poisonous and its presence, even in small quantity creates a health hazard. It can be perceived by smell in concentrations as low as 0.1 ppm, although the odor threshold may be as high as 25 ppm if one is accustomed to it. With increasing dosage of NO_2 the following sequence of effects can be observed: odor perception, nasal irritation, discomfort in breathing, acute respiratory distress, pulmonary oedema, and finally, death.

More information about these effects can be found in Strauss [69].

Nitric oxide is not toxic in the 10-50 ppm range. Since large NO concentrations are not stable in air but are converted to NO_2 little is known about damages purely due to it. Paralysis and convulsion have been reported after exposing animals to NO, but no cases of poisoning in man have been reported.

1.3.3 Effect of NO_x and Photochemical Smog on Plants

Nitrogen dioxide (NO_2) is a phototoxic substance since it can cause damage to vegetation. One of the main effects is chlorosis of leaves between the veins, where the destruction of chlorophyll results in bleaching of the leaves. The concentration level beyond which bleaching is induced is about 3 ppm. Another effect of the phototoxic NO_2 is the suppression of plant growth. Concentration of less than 1 ppm affects plant growth according to Oglesby in Rossano [61].

Although oxides of nitrogen by themselves may not reach dangerous levels in the ambient air, enough evidence shows that in the Los Angeles area the products of photochemical reactions initiated by nitrogen dioxide do reach levels which are harmful to humans, animals and plants. Middleton [42] discusses such damages. A full explanation of how photochemical smog is caused is given by Agnew [1], and Strauss [69].

1.3.4 NO_x Standards

Clean air, that is, air in areas sufficiently distant from places of human activity or other abnormal influences, is known to contain approximately 20.93% by volume of O_2; 78.1% by volume of N_2; 0.93% by volume of Argon; 0.03% by volume of CO_2; and other minor gases. Traces of so called pollutants in concentrations below 1 ppm also occur in 'clean air', where these pollutants are due to natural processes as discussed in Junge [29]. NO_x concentration in 'clean air' ranges from 0.000 to 0.030 ppm.

Air quality standards vary with organizations setting them and purposes for which they are set. Thus the World Health Organization (WHO) is always interested in the "hygienic" level of air quality or "no-effect" level, while some countries such as Canada, U.S.A., Western Germany base their standards on presently achievable goals. The three levels of air quality standards are the adverse level, the serious level, and the emergency level. Descriptions of these levels of allowable concentrations and durations of time for major pollutants can be found in Strauss [69], and Atkisson et al. [2]. In order to prevent possible risk to public health and atmospheric discoloration, California in 1969 adopted the NO_x standard of a concentration of 0.25 ppm for at most one hour's duration in ambient air. For automobile exhaust emission, California in 1971 adopted 4.0 grams of nitrogen oxides per mile as the maximum. In Alberta, the Department of Health has set 0.15 ppm as the

maximum acceptable limit of NO_x concentration in developing the Alberta Combined Air Quality Index (ACAQI). However this index, ACAQI, is still in the initial development stage. In comparison, the maximum acceptable level is the same as the serious level mentioned before. The federal Government of Canada denotes the three levels as maximum desirable level, maximum acceptable level, and maximum tolerable level where the maximum tolerable level corresponds to the emergency level mentioned earlier. More examples of NO_x concentration standards adopted in some other countries can be found in Hepple [23].

In the next section, the diffusion technique in air pollution modelling and some existing diffusion models will be discussed briefly.

1.4 Diffusion Model

1.4.1 Basis of Diffusion Models

Most efforts in modelling urban air pollution use the diffusion approach based on two similar methods. The first requires the exact solution of an equation of continuity for the pollutant where the adequate determination of some co-efficients is necessary. Details of the equation of continuity for an air pollutant are discussed in Atkisson et al. [2] while its application to atmospheric diffusion can be found in Pasquill [56]. Since the coefficients required for the solution of the equation of continuity are representatives of the product of eddy size and eddy velocity which in

turn depend on many variables, the exact solution of the equation becomes very difficult. As a result the first method is not used very often in predicting atmospheric concentration of pollutants.

The second method originally developed by Panofsky [53] enjoys more popularity because it is easy to work with. In its simple form for a continuous point source, the model for ground level concentration can be expressed as

$$C = \frac{Q}{\pi u \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2} - \frac{z^2}{2\sigma_z^2}\right) \qquad (1.4.1)$$

where C is the pollutant concentration,

u is the mean wind speed (assumed constant),

Q is the rate of pollutant emission (gram/sec.),

σ_y is the standard deviation of horizontal plume concentration,

σ_z is the standard deviation of the vertical plume concentration, and

x,y,z are the spatial coordinates representing the downwind, crosswind, and vertical distances, respectively, with regard to the origin at the point source.

The assumption inherent in this approach is that the concentration distribution from a continuous source has a Gaussian distribution relative to the centre line of the plume both in the vertical, and the perpendicular direction to the wind. Effluent from a continuous point source moves downwind,

spreading horizontally and vertically such that the distribution of the concentration of the contaminant in any cross section along either the horizontal or the vertical is Gaussian (i.e., has a normal distribution) while the mass in such a cross section is constant. The variances of the two normal distributions are functions of diffusion coefficients and of distance x, downwind. More explanation of this approach can be obtained from Moses [46].

1.4.2 Examples of Urban Air Pollution Models

Several people have used the diffusion approach to urban air pollution modelling.

Lamb [31] in 1968 used the approach to produce a most extensive atmospheric diffusion model for the Los Angeles basin. His model was based on the solution of the continuity equation. The model was used to compute carbon monoxide concentrations over Los Angeles for September 23, 1966 at 1200 grid points. The results did not agree satisfactorily with observations owing in part to some assumptions in the model and also to sources outside the Los Angeles basin which were not accounted for. Despite its limitations, Lamb's model is one of the great advancements in solving this problem.

Clarke [12]used the popular Gaussian diffusion approach to build one of the most well-known models. In his model, the receptor (or monitoring station) was located at the centre of four concentric circles having radii of 1, 4,

10, and 20 kilometers, respectively. These circles were divided into 16 equal sectors of 22.5 degrees. A source inventory was obtained for each of the 64 annular sectors. For each of the annular rings a chart was prepared relating C/Q, as defined in (1.4.1), and wind speed for various atmospheric stability classes. To refine his model, Clarke [13] later considered the contributions made by traffic flow, industry and commerce. The concentration at a point was then the sum of contributions from the point sources, traffic flow, industry, and space heating sources. He utilized his model to compute SO_2 and NO_x concentrations in Cincinnati, Ohio.

Other existing models using the diffusion approach include those by Hilst [26], Pooler [57], and Ryan [63]. Like the two discussed above the models show commendable efforts, as well as pointing to the fact that continued effort in the development of urban air pollution models is necessary. Therefore, a completely different approach to urban air pollution modelling is worth considering.

1.5 Computers and Pollution Problem

1.5.1 Data Processing

Any air pollution surveillance program has to handle a great deal of data. Efficient operation demands that the data be well organized and be available for use when it is required. For this, a good information storage and retrieval system of the quality and magnitude which only a digital com-

puter can handle is essential. An example of the task for such a system is the management of the emission inventory information. Since sources of air pollution in an urban area increase with time and the rates of emission of these sources do fluctuate, continuous information about sources is required. An information system that will handle this task should have a large storage capacity as well as an up-dating facility. The Province of Ontario has such a computerized source-inventory management system as explained by Shanks et al. [65].

The author of this thesis believes that an air pollution monitoring network can be handled better by the use of computers rather than doing it manually. The measuring instrument can be interfaced with an analog-digital converter, which in turn can be coupled with a small computer. Several small computers at different stations can then, through communication lines, at required intervals, report back to a central computer. One of the advantages of implementing this idea is that human mistakes in reading and recording data from measuring equipment will be eliminated. Since reports will be received regularly through the central computer, action to control pollution concentration could be taken more promptly. In addition the central computer can be programmed to check and report equipment break-down and other such adverse situations. The point made in this paragraph remains as a suggestion as far as this thesis is concerned, efforts in this work will concentrate on developing

models, other than the diffusion model, for NO_x.

1.5.2 Modelling and Simulation

Not only is the computer useful in the monitoring of air pollution and the storage and retrieval of the data, it also is indispensable in the analysis of the data, as for example in modelling and simulation of air pollution. As discussed by Moses [46], most of the diffusion models developed are impossible to handle without the use of the computer for the enormous computations. The indispensability of the computer in using time series analysis to develop an air pollution model will be quite obvious from Chapters III and IV of this thesis. The use of the model for predictive purposes requires computations that are better handled by the computer. If done by hand the result may not be ready in time for action.

Where models are not based on empirical data, the supply of input data has to come from simulation . For diffusion models emission data from the sources of the pollutant as well as meteorological data are required to estimate the pollutant's concentration at a point. However in order to study potential pollution distributions or ones where only sparse data is available, realistic values for the diffusion model must be generated by simulation. Simulation then could use source-inventory information combined with conjectured meteorological conditions to produce input data for the diffusion model. The Ontario computerized air pollution management system referred to earlier ,[65],

incorporates such simulation facility.

1.5.3 Models Proposed in this Thesis

Since time series analysis deals with large samples (long series) of observations the computer is an essential tool for storage and accurate retrieval of the data and for fast calculations with the data. The computer will thus be used extensively in the development of the two types of models for oxides of nitrogen proposed in this thesis. The first is the stochastic model which is developed in Chapter III while the second is the dynamic system model developed in Chapter IV. Both of the models make use of observed series of data. The stochastic model uses only the oxides of nitorgen data, whereas, for the dynamic system model the pollutant's observed values as well as the observations of factors influencing the pollutant's concentration are considered. As the model building procedures in Chapters III and IV will show, the series of data are analyzed and made to supply information on which the models are based, hence the models proposed in this thesis are bound to be representative of the pollutant's behavior.

CHAPTER II

OBSERVED NO_x DATA: SOURCE, PRECISION, AND ORGANIZATION

In stochastic modelling as opposed to deterministic modelling, past and current observations of a phenomenon are employed to establish a model that explains the behavior of the phenomenon within certain probabilistic limits. Thus stochastic model building is strictly based on observed data which must be very reliable if the model is to explain the phenomenon adequately. In addition to reliability the amount of data must be adequate. For time series analysis the data need to consist of at least fifty successive observations and preferably one hundred or more. The reasons for this data requirement will be given later in Chapter III.

2.1 Source and Precision

The NO_x data used to develop the models are those of Edmonton. Models for Calgary, Windsor, Sarnia, Toronto and Sudbury data were developed also as comparison to the Edmonton model. Both the Edmonton and the Calgary data were obtained from the Environmental Health Services Division of the Alberta Department of Health in Edmonton, while the Windsor, Sarnia, Toronto and Sudbury data came from the Ministry of Environment of the Province of Ontario. The Alberta Environmental Health Services Division has been collecting and storing air pollution data for the city of Edmonton

since July, 1964. A monthly summary of the data known as "Air Pollution Summary Edmonton" is published by the department. A similar publication exists for Calgary as well.

The main air pollutants for which data are available in Edmonton are particulate matter (recorded as a soiling index), oxides of nitrogen, nitrogen dioxide, oxidants, hydrocarbon, and carbon monoxide. Continuous monitoring of Sulfur Dioxide is just being developed. Hourly averages of NO_x and the five minute peaks are recorded at only one station located on the third floor of the Administration Building, 109 Street and 98 Avenue which is on the southward fringe of the downtown area. Owing to frequent machine malfunction there are many breaks in the record. Another cause of discontinuity of the record is the zeroing of the instrument which is done every day. This accounts for loss of two hourly observations per day, since only one machine is used. At present the zeroing takes place between 3 and 5 p.m.

The measuring instrument is a Beckman Atmosphere Analyzer Model K1004, which is manufactured by the Beckman Company for continuous measurement of oxides of nitrogen. The specifications for the Beckman Model K1004 state that the ppm NO_x in the air are measured with a precision of 3 decimal places. The instrument quantitatively introduces both the air sample and the reagent solution into a contactor where the contaminant is absorbed into the solution. After the liquid is separated from the air stream, it flows

into the detector where the chemical analysis is performed by coulometry. The signal from the detector is amplified and fed into a read-out device, in this case a chart recorder, which draws a continuous curve of the amount of NO_x in the analyzed air samples. Hourly averages of NO_x are then obtained by a technician who reads these off the chart by "eyeballing". These hourly averages are then recorded and stored as part of the data base of NO_x measurements. The principle and technique upon which the instrument is based were proposed by Saltzman [64].

Although NO_x has been monitored in Edmonton since 1964, it is not easy to obtain NO_x series long enough for time series analysis owing to the reasons given earlier in this section. Missing data which create holes in the records break the observations into very short series most of which are less than fifty data points long. Since long series of about 100 data points are desirable for time series analysis a large part of the records are not useful for the analysis. However by searching through the whole data base, series long enough to study hourly and daily behavior of NO_x are found.

The NO_x Hourly Average series which was chosen for analysis is that for the two weeks, February 22 to March 7, 1967 which contained 336 hourly measurements without any "holes" (missing observations). For the purpose of this work maximum NO_x concentration for a day is defined as the maximum of the hourly averages for the day. The daily maximum concentration of NO_x obtained as just defined constitute the

Daily Maximum NO_x series from April 1, 1971 to March 31, 1972 used to study the daily behavior. This part of the record has very few holes to be filled as compared to other parts. About twenty holes were encountered and nearly all of them consist of one missing observation while the rest have five or less observations missing. The holes that existed were filled by taking the average of one observation before the hole and one after it. In this way 366 consecutive daily-maximum observations were obtained for the analysis.

2.2 Organization

Oxides of Nitrogen data are merely a part of the pollution data collected for the City of Edmonton as the last section indicates. In addition to the pollutants' data, wind data are also collected and stored along with the air pollution data. These data, as they are collected each day, get punched on computer cards and stacked in monthly decks. The monthly decks contain an average of 400 cards, and, hence up to March 1972, data for 93 months were contined on about 37,200 computer cards. In order to make these data more amenable to computer processing in this work they are transferred to magnetic tape. The details of how the data are organized and stored on tape and how they can be retrieved is discussed in Appendix A.

CHAPTER III

STOCHASTIC MODEL BUILDING

As stated in Chapter I two types of models are developed in this thesis to represent the behavior of NO_x. The first, referred to as the stochastic model, recognizes the time-dependent nature of a phenomenon in the presence of the unknown factors influencing its behavior. The second model type is the combined transfer function-noise model which also takes into account time-dependence, and in addition recognizes the dynamic relationship between the phenomenon and the influencing factors. Hence, the transfer function-noise model is the system analysis approach to the problem, where the phenomenon observed (in this case NO_x) is the output from the system and the factors influencing the behavior (for example temperature) of the phenomenon are the inputs. The first model type will be established for NO_x in the rest of this chapter while the second will be developed in Chapter IV.

In designing and testing the models for the NO_x data the methods developed by Box and Jenkins [6] will be used extensively and, hence, the details of the theoretical development and the historical background of those techniques will not be restated here. The present analysis of the NO_x data will thus use the Box and Jenkins [6] approach as a tool in solving the problem of how the oxides of nitrogen process behaves.

3.1 Stochastic Models in NO_x Data Analysis

A stochastic process is a statistical phenomenon that evolves in time according to probabilistic laws; while a time series is a set of values observed sequentially in time and thus is one particular realization of the underlying stochastic process. Observation of a continuous time series can be made at some fixed interval giving a discrete sequence of observations. NO_x observations dealt with here constitute discrete time series since the observations are recorded for fixed time intervals. The first series consists of the NO_x hourly averages where the observations are close to the average value over a one-hour interval and hence the time scale is in hours. The second series consists of the daily maxima of NO_x where the maximum of the hourly averages is taken for each day and hence this time scale is in days.

If a time series comes from a stationary process then the underlying process from which it has been observed may be an autoregressive (AR) process, a moving average (MA) process, or a combination of both called autoregressive-moving average (ARMA) process. A detailed and rigorous definition of the stochastic stationary process is given in Box and Jenkins [6]. There are various ways in which a process can be non-stationary and so there are many ways in which a time series observed from a non-stationary process can exhibit non-stationary behavior. The number of ways non-stationary behaviors can be displayed by a time series will

not be enumerated here; they can be found in Box and Jenkins [6], but a particular kind of non-stationary behavior which the two NO_x series display is hereby mentioned. Like many empirical time series, they behave as though they have no fixed mean, but they exhibit homogeneity in the sense that, apart from a local level, or perhaps a local level and trend one part of the series behaves much like any other part. The stochastic process which underlies such homogeneous non-stationary behavior can,by differencing,be changed to a stationary process, that is the resulting process of the suitable difference may be an AR, MA, or ARMA process. Therefore, such series may be representable by the modified form of the ARMA called autoregressive integrated moving average (ARIMA) Models. Non-stationary behavior of a series can be discovered through the sample autocorrelation function as will be shown later in this chapter. The sample auto-correlation function also can be used as an aid in identifying which of the three types of models to try and fit to the series exhibiting stationary behavior.

It is necessary to give a brief definition of the terms autoregressive and moving average at this point. Let z_t, z_{t-1}, z_{t-2}, ... be the values of a process at equally spaced time points t, t-1, t-2, ...; and let y_t, y_{t-1}, y_{t-2},..... where $y_t = z_t - u$, be the deviations from the mean, u, of the stationary series. Then the model, relating the present value y_t, to the previous values, can be expressed as

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots\ldots + a_p y_{t-p} + e_t \qquad (3.1.1)$$

and is referred to as an autoregressive (AR) model of order p, because y_t is "regressed" on previous values of the same process. The coefficients a_i (i = 1,2, ..., p) are referred to as the autoregressive parameters. The series of shocks e_t, (known as "white noise") consists of a sequence of uncorrelated random variables whose distribution is usually assumed to be normal with mean zero and constant variance σ_e^2. The moving average model expresses y_t in terms of present and past random shocks,

$$y_t = e_t - b_1 e_{t-1} - b_2 e_{t-2} - \ldots - b_q e_{t-q}, \qquad (3.1.2)$$

where the coefficients b_j (j = 1, 2, ..., q) are referred to as the moving average parameters. The moving average model is equivalent to a linear filter with a finite number of non-zero weights . ARMA model is a combination of (3.1.1) and (3.1.2).

The general form of the stochastic model that will be considered here can be written as

$$a(B) \{ (1-B)^d z_t \} = b(B) e_t \qquad (3.1.3)$$

where B is the backward shift operator defined as

$$Bz_t = z_{t-1} \quad \text{and} \quad B^d z_t = z_{t-d} \ , \ \text{and}$$

$$a(B) = 1 - a_1 B - a_2 B^2 - \ldots\ldots - a_p B^p \ ,$$

$$b(B) = 1 - b_1 B - b_2 B^2 - \ldots\ldots - b_q B^q \ ,$$

z_t is the observed value (i.e. transformed as discussed in section 3.2) of NO_x at time t, $\{ (1 - B)^d z_t \}$ is the differenced series, and e_t is the random shock at time t. The integers d, p, and q in turn denote the following:

- d is the degree of differencing performed on the data,
- p is the number of autoregressive parameters (or order of the AR model part),
- q is the number of moving average parameters (or order of the MA model part).

The general model (3.1.3) is referred to as an autoregressive integrated moving average model ARIMA (p, d, q). The same model can be written as an ARMA (p, q) model

$$a(B) w_t = b(B) e_t \ , \qquad (3.1.4)$$

where $w_t = (1 - B)^d z_t$ is a stationary series whose length, $n = N - d$. In the next section the specific stochastic models that explain the behavior of the NO_x hourly averages

and the daily maxima of NO_x will be identified.

3.2 Stochastic Model - Identification

The sets of the NO_x raw data presented in Series A and B were transformed to their natural logarithms. Since the measurements constituting the data are very small numbers in the range 0.000 to 0.500 ppm, and since the natural logarithm of zero is indeterminable, 0.001 was added to each observed value of the series before logarithmic transformation. Thus the original series $\{z_t\}$ is transformed to the series $\{ \ln (z_t + 0.001) \}$. Because all subsequent analyses will be performed on the transformed series any reference to z_t from this point on is a reference to $\ln (z_t + 0.001)$ except where a distinction is made.

The transformed time series, now referred to as $\{z_t\}$, are plotted in Figures 3.1 and 3.2. Since there appears to be variation in level, no fixed mean seems to exist, however, a "seasonal" cycle is not apparent either. Series behaving in this way come from non-stationary processes. Nevertheless, appropriate differencing of the transformed series may result in a stationary series which then would be suitable for the proposed analysis.

It is conjectured here that the amount of differencing required to get a stationary series may be indicated by comparing the variance of the undifferenced series with the variances of the various differenced series. A reduction in

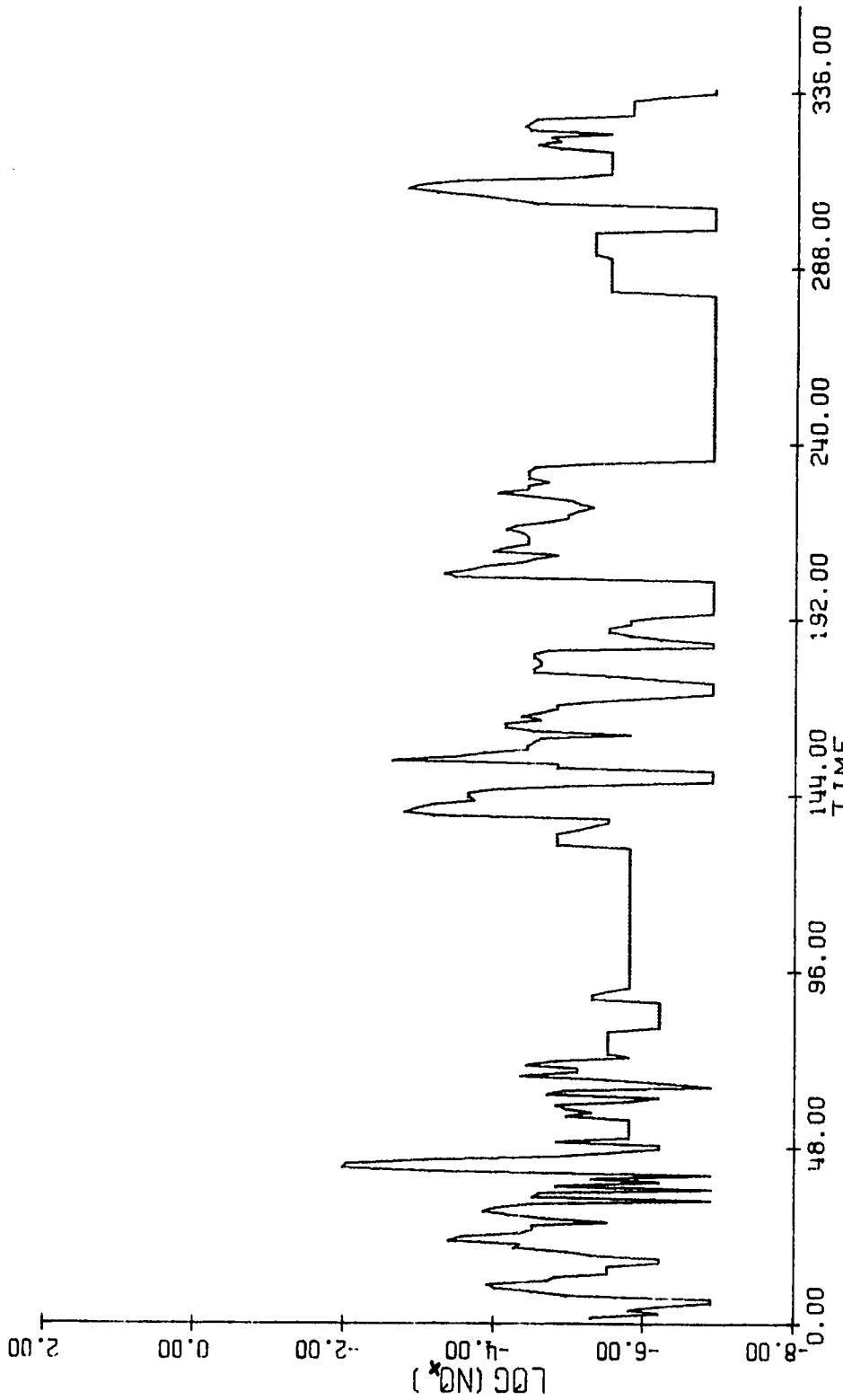

Fig. 3.1 NO_x Hourly Averages (Series A transformed)

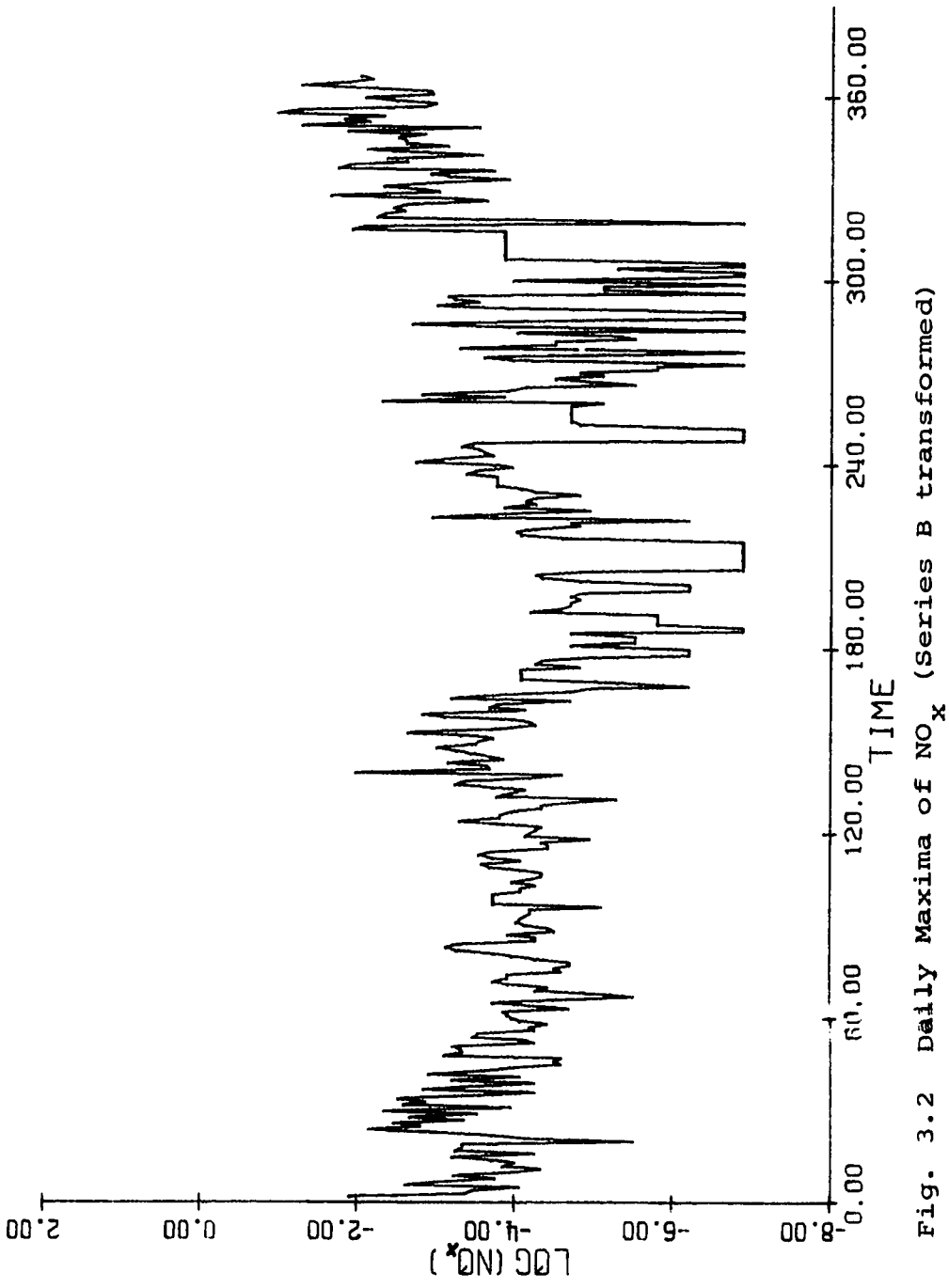

Fig. 3.2 Daily Maxima of NO_x (Series B transformed)

the observed (sample) variance seems to indicate less instability of the series. For the two series under consideration, the first differences have a lower variance than the others as shown in Tables 3.1 and 3.2.

TABLE 3.1 Differences for transformed Hourly Averages of NO_x

Difference	Length of Series	Mean	Variance
No Differencing	336	-5.495	1.202
$(1 - B)$	335	-0.004	0.444
$(1 - B)\ (1 - B)$	334	0.007	2.970
$(1 - B^{24})$	312	-0.021	2.497
$(1 - B)\ (1 - B^{24})$	311	-0.013	1.452

TABLE 3.2 Differences for transformed Daily Maxima of NO_x

Difference	Length of Series	Mean	Variance
No Differencing	366	-4.111	1.577
$(1 - B)$	365	0.000	0.981
$(1 - B)\ (1 - B)$	364	-0.008	4.587
$(1 - B^{7})$	359	0.019	1.881
$(1 - B)\ (1 - B^{7})$	358	0.003	2.115

The 24th difference of the Hourly Averages was taken to see whether or not the variation in level was due to a 24 - hour 'seasonal' pattern; while the 7th difference of the Daily Maxima was taken to investigate the 7 - day 'seasonal' pattern that could exist in the Daily Maximum of NO_x series. The results in the tables show that those "seasonal" patterns do not seem to exist in the two series, otherwise those differenced series would have had lower variances. Thus the stationary model type that should be investigated first for both series is the ARMA (p, q) model of the form (3.1.4) where d = 1.

An important and more reliable indicator of the degree of differencing needed is the autocorrelation function $\rho_{ZZ}(k)$, where

$$\rho_{ZZ}(k) = E[\ (Z_t - U)\ (Z_{t+k} - U)\ ]\ /\ \sigma_Z^2\ , \qquad (3.2.1)$$

U is the mean of $\{Z_t\}$, and k is the lag or time difference between the values of the time series considered for correlation. The failure of the autocorrelation function to rapidly decay indicates non-stationary behavior on the part of the series, hence, further differencing and or transformations must be considered and tried. For both of the observed time series (in this case their transformations) we obtain the corresponding sample autocorrelation functions $r_{zz}(k)$, referred to simply from here on as the autocorrelation function. Figures 3.3 and 3.4 show the autocorre-

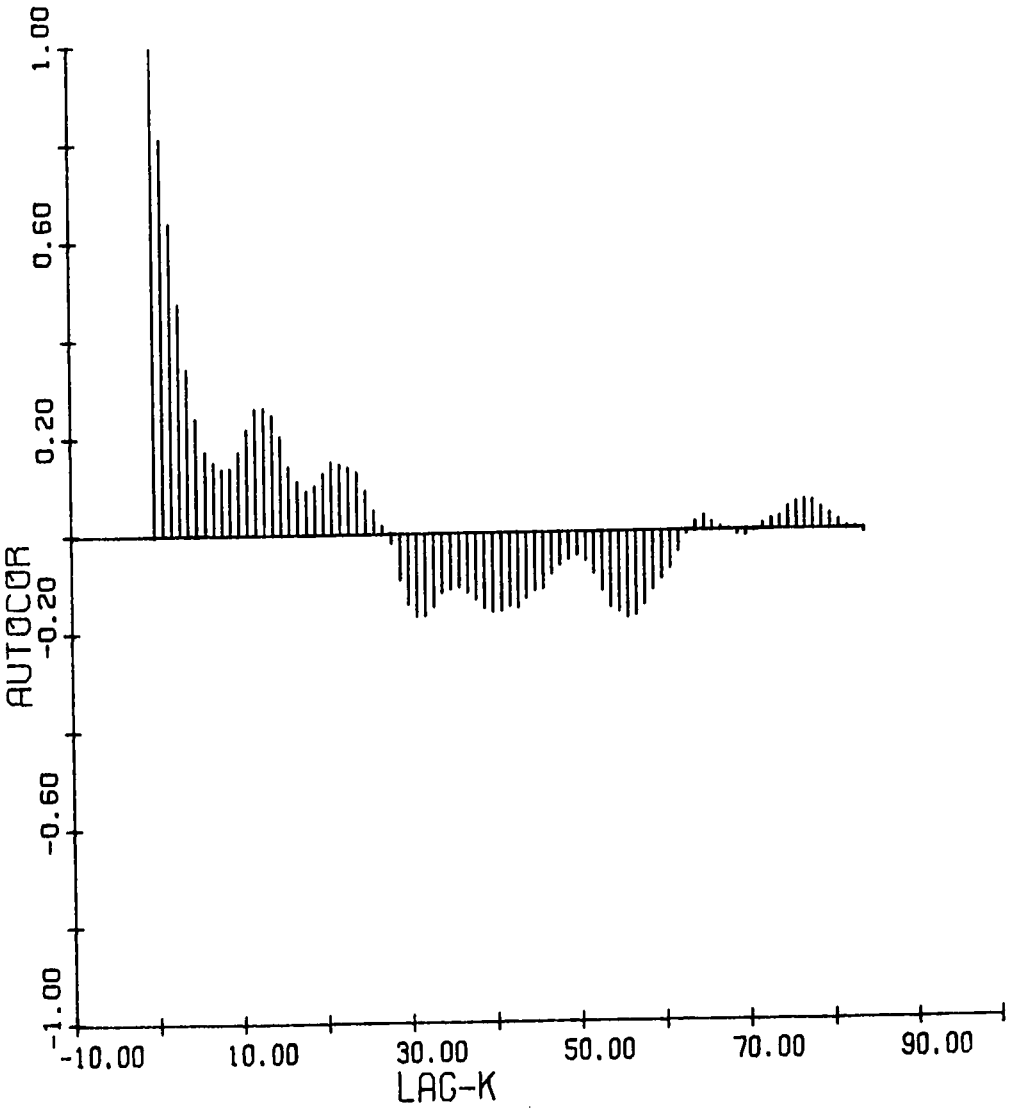

Fig. 3.3 Observed Autocorrelation Function of the transformed NO_x Hourly Averages

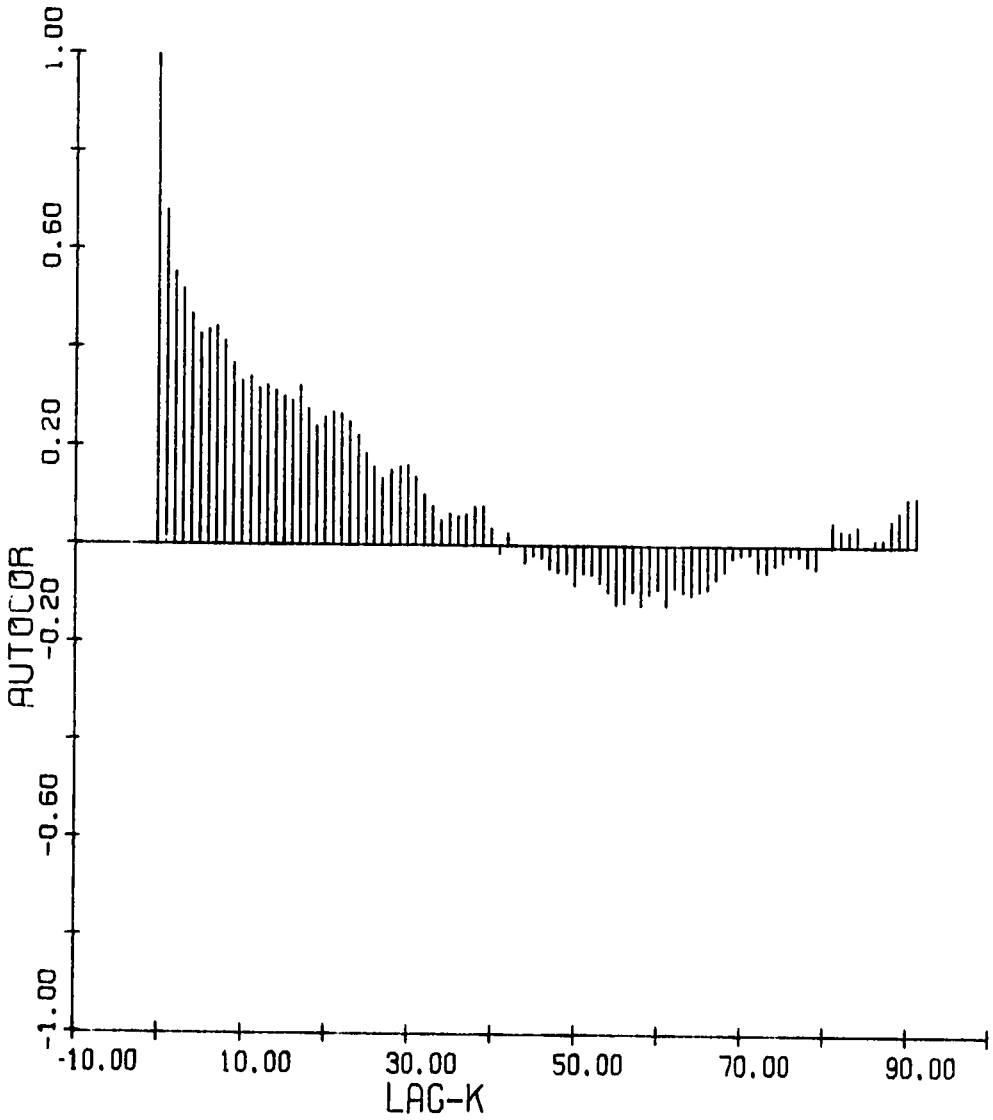

Fig. 3.4 Observed Autocorrelation Function of the transformed Daily Maxima of NO_x

lation functions of the undifferenced hourly average and daily maximum series respectively; while Figures 3.5 and 3.6 show the autocorrelation functions of the respective first difference $\{W_t\}$. The autocorrelation function of the daily maxima of NO_x plotted in Figure 3.4, certainly decays rather slowly, thus indicating non-stationary behavior. However, upon taking the first difference the corresponding autocorrelation function shown in Figure 3.6 indicates comparatively low values beyond lag 2.

Further examination of the autocorrealtion function reveals the specific types of models which may be fitted to the data. Models may be categorized according to the number of autoregressive and moving average terms to be considered, [i.e. (p, q)]. Since the theoretical autocorrelation function is known for a given model, the shapes of these functions may broadly be classed into two categories; the exponentially decaying functions and the functions displaying exponential decay mixed with damped sine functions. The examples below indicate the shapes of the theoretical autocorrelation functions for some of the model types.

(a) ARMA (1,0): $w_t = aw_{t-1} + e_t$ (3.2.2)

The autocorrelation function $\rho(k)$, decays exponentially where $\rho(1) = a$, and the residual variance

$$\sigma_e^2 = (1 - a^2)\ \sigma_w^2$$

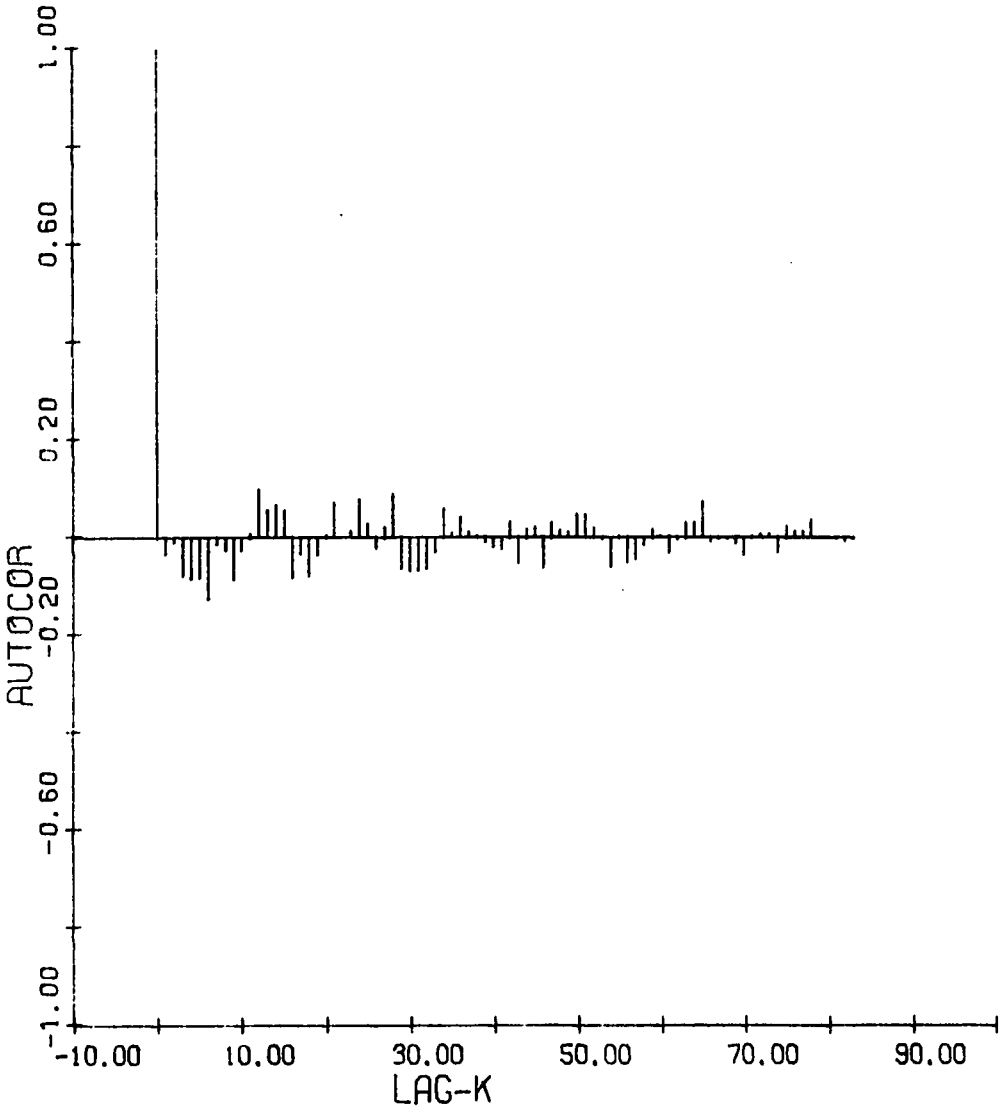

Fig. 3.5 Observed Autocorrelation Function of the first difference of the transformed NO_x Hourly Averages

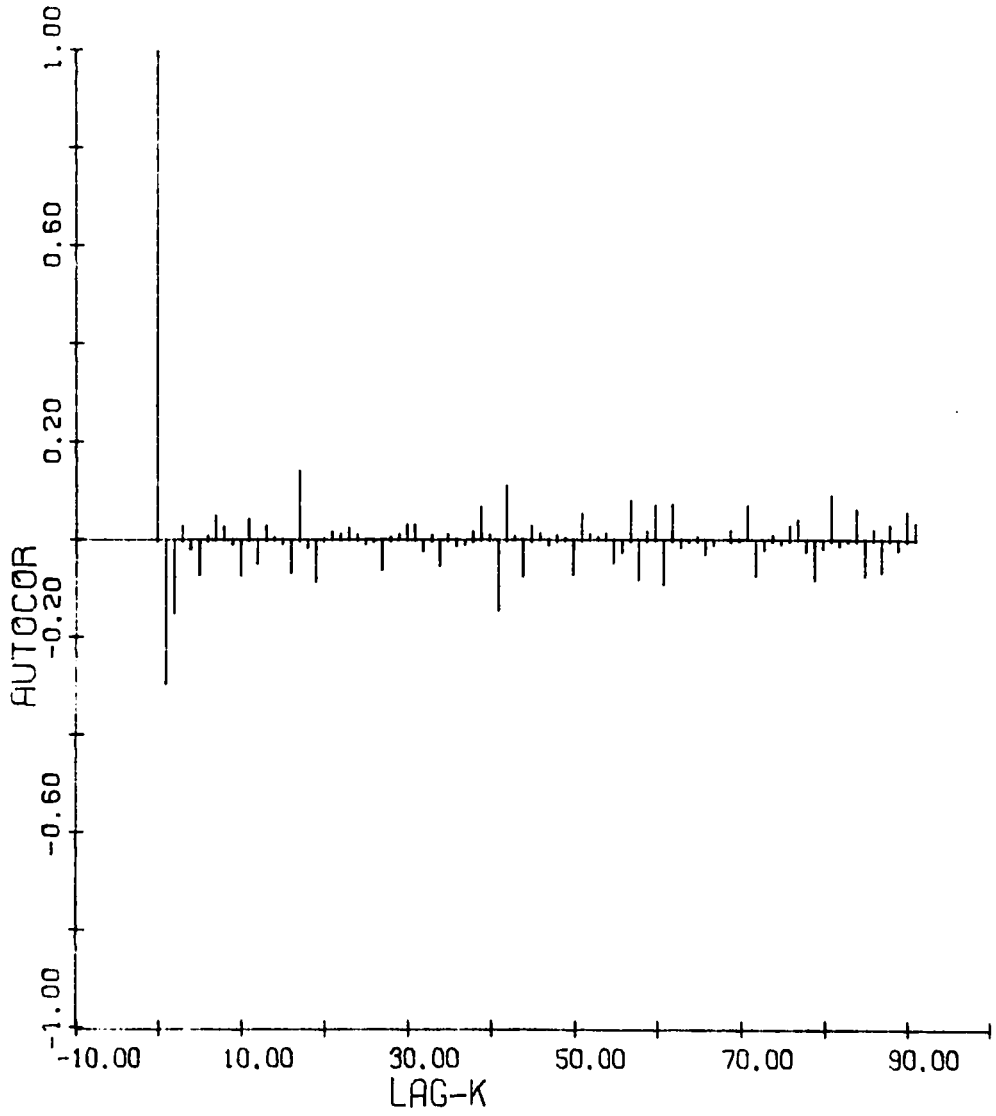

Fig. 3.6 Observed Autocorrelation Function of the first difference of the transformed Daily Maxima of NO_x

(b) ARMA(0, 1): $w_t = e_t - be_{t-1}$ (3.2.3)

The autocorrelation function has value zero beyond lag 1,

$\rho(1) = -b / (1 + b^2)$ for

$-1 < b < 1$; and,

$$\sigma_e^2 = \frac{\sigma_w^2}{1+b^2}$$

(c) ARMA(2, 0): $w_t = a_1 w_{t-1} + a_2 w_{t-2} + e_t$ (3.2.4)

The autocorrelation function is a mixture of exponential decay and a damped sine wave. The parameters a_1, a_2, are given by

$a_1 = \rho(1) [1 - \rho(2)] / [1 - \rho^2(1)]$,

$a_2 = [\rho(2) - \rho^2(1)] / [1 - \rho^2(1)]$,

where for stationarity of the model

$-1 < a_2 < 1$,

$a_2 + a_1 < 1$, and

$a_2 - a_1 < 1$.

$\sigma_e^2 = [1 - a_1\rho(1) - a_2\rho(2)] \sigma_w^2$

(d) ARMA(0, 2): $w_t = e_t - b_1 e_{t-1} - b_2 e_{t-2}$ (3.2.5)

The autocorrelation function is zero beyond lag 2. The parameters b_1, b_2, are given by

$\rho(1) = -b_1 (1 - b_2) / (1 + b_1^2 + b_2^2)$,

$\rho(2) = -b_2 / (1 + b_1^2 + b_2^2)$ where

for invertibility of the model

$-1 < b_2 < 1$,

$b_2 + b_1 < 1$, and

$b_2 - b_1 < 1$.

$$\sigma_e^2 = \frac{\sigma_w^2}{(1 + b_1^2 + b_2^2)}$$

(e) ARMA(1, 1): $w_t = aw_{t-1} + e_t - be_{t-1}$ (3.2.6)

The autocorrelation function decays exponentially as of lag 1. The parameters, a and b, can be obtained from

$\rho(1) = (1 - ab)(a - b) / (1 + b^2 - ab)$

$\rho(2) = a\rho(1)$ where

$-1 < a < 1$, and

$-1 < b < 1$ in order for the model to be stationary and invertible.

$$\sigma_e^2 = \frac{\sigma_w^2 [1 - a^2]}{(1-ab)(a-b)}$$

Since the theoretical autocorrelation function is not known, the observed sample autocorrelation function is used as its estimate. The observed autocorrelations at lag k, k = 1, 2, 3, ..., K and K = N/4 are sufficient to identify the model. The sample autocorrelation function, $r_{ww}(k)$ or

simply r(k) is obtained from the sample autocovariance function $C_{ww}(k)$ or simply C(k). Hence, the estimator of $\rho(k)$ is r(k),

$$r(k) = C(k) \; / \; C(0), \tag{3.2.7}$$

where C(k) is the estimate of the theoretical autocovariance of lag k,

$$C(k) = \frac{1}{N} \sum_{t=1}^{N-k} (w_t - \bar{w}) \; (w_{t+k} - \bar{w}), \tag{3.2.8}$$

for k = 0, 1, 2, ..., K

C(0) is the sample estimate of the theoretical variance σ_w^2 of the series $\{w_t\}$, and $\bar{w}$ is the observed average of the series. Usually $\bar{w}$ is zero since w_t is the difference of the series $\{z_t\}$. It should be noted that to obtain a useful estimate of the autocorrelation function, at least 50 observations are needed.

Figure 3.5 shows that the hourly NO_x may perhaps just be an ARMA(0, 0), namely just white noise process written as $w_t = e_t$, because r(1), r(2), ..., r(k) do not appear to be very different from zero. Figure 3.6 shows that the daily NO_x may be fitted by an ARMA(0, 2) model otherwise known as an MA model of order 2, see (3.2.5). It can be seen that after lag 2 the observed autocorrelations do not appear to be very different from zero. These conjectured models will be used as starting points in trying to find the appropriate models for the two sets of NO_x data.

3.3 Stochastic Model - Preliminary Estimation

The use of the autocorrelation function (or autocovariance function) in the initial estimation of the parameters and the residual variance can be extended to deal with models of more than two parameters. In determining initially which model to use in the preliminary stage, it must be pointed out that, although, a given ARMA model has a unique covariance structure, a particular covariance structure does not determine an ARMA model uniquely. In other words there is no one to one correspondence between autocovariance functions (or autocorrelation functions) and ARMA models.

One of the initial estimation tools is the Yule - Walker equation in Jenkins and Watts [28], which uses the observed autocorrelations to solve for parameters in autoregressive models. For instance, let the model be autoregressive and of order 5, then the five autoregressive parameters will be obtained by solving the following set of five linear equations:

$$
\begin{aligned}
a_1 r(0) + a_2 r(1) + a_3 r(2) + a_4 r(3) + a_5 r(4) &= r(1) \\
a_1 r(1) + a_2 r(0) + a_3 r(1) + a_4 r(2) + a_5 r(3) &= r(2) \\
a_1 r(2) + a_2 r(1) + a_3 r(0) + a_4 r(1) + a_5 r(2) &= r(3) \qquad (3.3.1) \\
a_1 r(3) + a_2 r(2) + a_3 r(1) + a_4 r(0) + a_5 r(1) &= r(4) \\
a_1 r(4) + a_2 r(3) + a_3 r(2) + a_4 r(1) + a_5 r(0) &= r(5)
\end{aligned}
$$

If the model is of order p, then p such equations will have to be solved for the p parameters a_i, i = 1, 2, ..., p. After obtaining these initial parameter values an estimate of the residual variance is given by

$$\sigma_e^2 \simeq C(0)\ [1 - \tilde{a}_1 r(1) - \tilde{a}_2 r(2) - \ldots\ldots - \tilde{a}_p r(p)]\ ,$$

where $\tilde{a}_i$ is the estimated parameter. For further details on this approximation of the residual variance see Bartlett [4].

A general method of using the autocovariances for obtaining initial estimates of the parameters of autoregressive - moving average processes is described in Chapter 6 of Box and Jenkins [6]. Since their method is used to obtain the initial estimates for the NO_x models, a brief summary of the algorithm is given in Appendix B.

For the problem on hand, a general Fortran IV program (see Program 2 in Owolabi [52]) following the algorithm in Appendix B is developed and implemented. Using the program the following sixteen ARMA models with the appropriate initial estimates were produced for both the hourly and daily NO_x data (i.e. for the first differences, w_t, of the transformed data):

(p,q) = (1,0), (2,0), (3,0), (4,0), (1,1), (1,2), (2,1), (2,2), (3,1), (3,2), (4,1), (4,2), (0,1), (0,2), (0,3), (0,4).

Table 3.3 shows the initial parameter estimates obtained for some of these models for hourly NO_x while Table 3.4 shows

the same thing for daily NO_x.

TABLE 3.3 Initial Parameter Estimates for transformed Hourly Averages of NO_x

Model for w_t	Parameter Estimates				Residual
	a_1	a_2	b_1	b_2	Variance
ARMA(1,0)	-0.039	-	-	-	0.444
ARMA(1,1)	0.310	-	0.349	-	0.444
ARMA(1,2)	6.769	-	0.185	-0.006	20.362
ARMA(2,0)	-0.040	-0.014	-	-	0.444
ARMA(2,1)	-0.376	2.213	0.068	-	2.725
ARMA(0,1)	-	-	0.039	-	0.444
ARMA(0,2)	-	-	0.039	0.012	0.444

By examining the initial estimates of the residual variances and of the parameters of these models, the model with the smallest variance, and with parameters enjoying the stationarity and invertibility properties is the model which promises to be the "best-fitting" one. Details regarding the invertibility and stationarity conditions are given in Box and Jenkins [6]. Another factor that needs to be considered in identifying the model that may fit best is parsimony in the use of parameters. A model with less parameters than the " optimal model " and with residual

TABLE 3.4 Initial Parameter Estimates for transformed Daily Maxima of NO_x

Model for w_t	Parameter Estimates				Residual
	a_1	a_2	b_1	b_2	Variance
ARMA(1,0)	-0.298	-	-	-	0.893
ARMA(1,1)	0.513	-	0.891	-	0.875
ARMA(1,2)	-0.203	-	0.227	0.260	0.806
ARMA(2,0)	-0.377	-0.265	-	-	0.830
ARMA(2,1)	0.768	-0.498	-0.135	-	7.450
ARMA(0,1)	-	-	0.331	-	0.884
ARMA(0,2)	-	-	0.460	0.191	0.786

variance a little bit larger than the one for the optimal may produce quite an adequate fit to the data, and may be easier to interprete in its practical context. Considering these points and looking at Table 3.4 it is easy to see that ARMA(0,2) model may be able to fit the daily data. Table 3.3 shows that any of the ARMA(1,0), ARMA(0,1) models can fit the hourly data, but they are not better than the ARMA(0,0) process conjectured earlier for the data.

The algorithm employed for the solution of the sets of linear equations in this work is the Gauss - Jordan algorithm described in many numerical analysis books such as the one by Carnahan, et al. [9].

The models decided on after the identification and initial estimation procedures must be rigorously "processed" in order to obtain efficient estimates of their parameters. This processing will be done for the two models which have been identified for the hourly and daily NO_x data in the estimation stage which is discussed in the next section.

3.4 Stochastic Model - Estimation

The initial estimates of the parameters for the models identified in section 3.3 are used here to initiate the iterative procedure which produces the values of the parameters that give the minimum residual variance (or minimum residual sum of squares).

In section 3.3 the process generating the first difference of the transformed daily maxima of NO_x (i.e. w_t) was identified as ARMA(0,2) which is

$$w_t = e_t - b_1 e_{t-1} - b_2 e_{t-2}$$

For given values of b_1, b_2 this model is used to generate recursively a series of residuals, e_t, given by

$$e_t = w_t + \tilde{b}_1 e_{t-1} + \tilde{b}_2 e_{t-2}$$

The w_t, $t = 1, \ldots\ldots, N - 1$, constitute the series to be modelled; the unknown starting values e_0, e_{-1} are equated

to the expected value of e_t which is zero. The sum of squares of the generated residual series, $\{e_t\}$, is then calculated. In this manner the sum of squares can be calculated for all possible values of the parameters b_1, b_2 and the set of values of the parameters producing the minimum sum of squares is the best set of parameter estimates for the fitted model. The algorithm for this iterative model estimation constitutes Appendix C. A Fortran IV program (see Program 3 in Owolabi [52]) is written to implement this algorithm. Using that program the most efficient set of values of the parameters of the ARMA(0,2) model for the daily NO_x data are

$\tilde{b}_1 = 0.500$ with standard error of b_1 equal to 0.045,
$\tilde{b}_2 = 0.260$ with standard error of b_2 equal to 0.051,
$\tilde{\sigma}_e^2 = 0.772$

Here the parameters are large compared to their standard errors and should be retained.

As discussed in section 3.3, w_t, the first difference of the transformed hourly NO_x data seemed to behave like a white noise process, ARMA(0,0). However, one of the sixteen initial ARMA models for the w_t series having a low residual variance estimate is ARMA(0,1). The corresponding efficient estimate of its parameter b, can be found at this stage. Then the ARMA(0,1) model can be compared to the ARMA(0,0) to see which one fits the hourly data better. The ARMA(0,1) model for the hourly data is

$$w_t = e_t - be_{t-1}$$

The series of residuals, e_t, for the model can be generated for the given value of b, according to

$$e_t = w_t + be_{t-1}$$

where again the value e_0 is set to zero. Using the model estimation algorithm of Appendix C (i.e. the same Fortran Program) as for daily data, the most efficient parameter of the ARMA(0,1) model for hourly average NO_x is found to be

$\tilde{b} = 0.054$ with standard error of $\tilde{b}$ equal to 0.064,
$\tilde{\sigma}_e^2 = 0.444$

Since the parameter is very small compared with its standard error it cannot be considered to be different from zero. Hence, the model that fits the hourly series is ARMA(0,0) and not ARMA(0,1). This implies that the first difference of the transformed hourly data behaves like a white noise process.

The model that fits the transformed NO_x Hourly Average series is

$$w_t = e_t$$

which can be written in terms of the z_t as

$$z_t = z_{t-1} + e_t$$

For the transformed daily maximum NO_x series the model that seems to have the best fit is

$$w_t = e_t - 0.5\, e_{t-1} - 0.26\, e_{t-2}$$

which can be written in terms of z_t as

$$z_t = z_{t-1} + e_t - 0.5\, e_{t-1} - 0.26\, e_{t-2}$$

Before finally accepting these models as being representative of the processes generating the corresponding series they should undergo some diagnostic checks. These diagnostic checks are presented in the next section.

3.5 Stochastic Model - Diagnostic Checks

In this section three tests are employed as diagnostic checks of the models. In effect these check the behavior of the residuals e_t, whether or not they are white noise. A good fitting model describes the data behavior (nearly) completely leaving only small independent errors in the unexplained part.

The first diagnostic check is the comparison of the estimated variance of the fitted series (i.e. w_t in this case) to the estimated residual variance. A small $\tilde{\sigma}_e^2$

implies that most of the $\tilde{\sigma}_w^2$ is explained by the model.

In the second test, the series of the residuals e_t is checked. If a model provides a good fit to the data, then the residuals remaining should constitute a series of white noise since white noise is uncorrelated (see section 3.1) the autocorrelation function $r_e(k) = 0$, for $k > 0$. This second test, therefore, involves estimating the autocorrelation function, $r_e(k)$, of the residual series e_t, $t = 1, 2, \ldots\ldots, n$, and performing a Chi - square test on these autocorrelations. The test statistic, hereby called Q, can be shown to have a Chi - Square distribution with $K - p - q$ degrees of freedom where $K = n/4$. Q is obtained by

$$Q = n \sum_{k=1}^{K} r_e^2(k)$$

If Q is less than the upper 5% critical point for Chi - Square with the appropriate degrees of freedom then the hypothesis that $r_e(k) = 0$ for $k > 0$ is accepted, and the residuals can be taken as constituting a white noise series.

The third test called the cummulative periodogram test checks if a seasonal pattern exists in the residuals which would indicate that the model does not explain the behavior of the data as well as it should. A seasonal pattern in the residuals will be indicated by periodicity in the residual series, and it can be detected by the use of periodogram. The periodogram, $I(f_i)$ of a time series e_t, $t = 1, 2, \ldots\ldots,$

n is defined as

$$I(f_i) = \frac{2}{n}\left[\left(\sum_{t=1}^{n} e_t \text{Cos } 2\pi f_i t\right)^2 + \left(\sum_{t=1}^{n} e_t \text{Sin } 2\pi f_i t\right)^2\right]$$

where $f_i = \frac{i}{n}$ is the frequency and $i = 1, 2, \ldots\ldots, \frac{n}{2}$. Since the cumulative spectrum, P(f), of white noise, if plotted against f, is a straight line running from (0,0) to $(0.5,\sigma_e^2)$, (or $P(f) / \sigma_e^2$) is a straight line running from (0,0) to (0.5,1), and since I(f) provides an unbiased estimate of the power spectrum at frequency f, the cumulative periodogram divided by the observed error (or residual) variance s_e^2 (where $s_e^2 = \tilde{\sigma}_e^2$) is an estimate of $P(f_j) / \sigma_e^2$

$$C(f_j) = \frac{\sum_{i=1}^{j} I(f_i)}{ns_e^2} \cdot$$

Detailed explanation about this can be found in Box and Jenkins [6]. $C(f_j)$ is known as the normalized cumulative periodogram. If the fitted model adequately accounts for seasonal patterns so that no seasonal pattern exists in the residuals, the plot of $C(f_j)$ against f_j will be scattered around the straight line joining the points (0,0) and (0.5,1). Using the Kolmogrow - Smirnov test, as explained in Hald [22], to draw 95% confidence limits for the plot, the probability lines should not be crossed more than 5% of the time if the residuals behave in a white noise fashion.

These three tests were applied to the residuals of the two models developed in section 3.4 and the results are the following:

(a) NO_x Hourly Averages

(i) $\tilde{\sigma}_z^2 = 1.202$

$\tilde{\sigma}_e^2 = 0.444$

Variance due to model (differencing in this case) is 0.758

(ii) Calculated Chi - Square is Q = 63.68 for n/4 = 83 degrees of freedom. The upper 5% critical point for Chi - Square with 83 degrees of freedom is 105.27.

(iii) Figure 3.7 demonstrates the result of the cumulative periodogram test. The points scattered along the theoretical line and are all within 95% confidence limits.

In test (i), $\tilde{\sigma}_e^2 = \tilde{\sigma}_w^2$ because the model is ARMA(0,0). However, the main interest is in the behavior of z_t. Since $\tilde{\sigma}_e^2 = \tilde{\sigma}_w^2 < \tilde{\sigma}_z^2$ it can be concluded that the model written in terms of z_t explains the behavior of the transformed hourly data well enough. In test (ii), the fact that the calculated Chi - Square is less than the theoretical one, implies that $r_e(k) = 0$, for $k > 0$, hence the residuals constitute a white noise process. Test (iii) also shows that the residuals constitute a white noise process, and that there is no seasonal pattern. Thus the model fitted to the hourly data in section 3.4 is finally accepted.

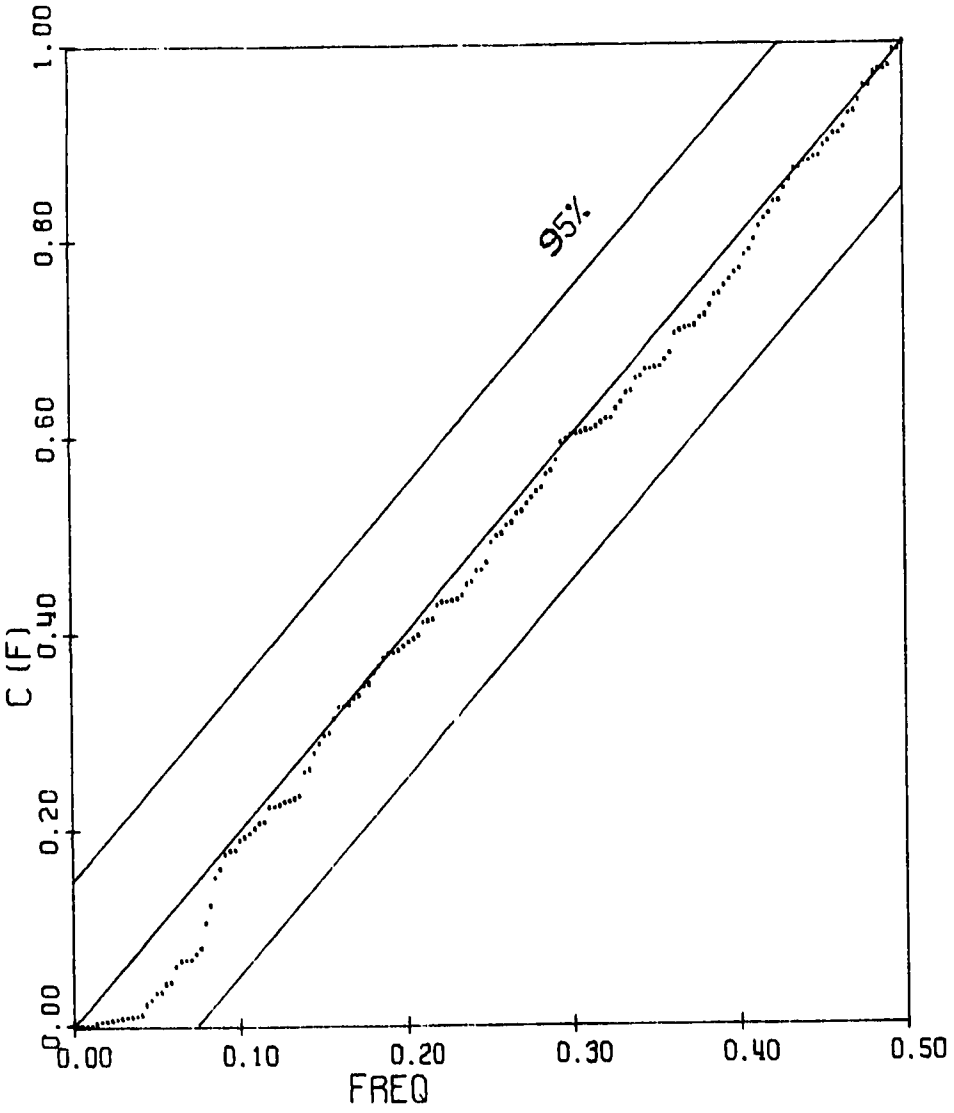

Fig. 3.7 Cumulative Periodogram for the residuals of the model fitted to the transformed NO_x Hourly Averages

(b) Daily Maxima of NO_x

(i) $\tilde{\sigma}_z^2 = 1.577$

$\tilde{\sigma}_w^2 = 0.981$

$\tilde{\sigma}_e^2 = 0.772$

Variance due to model (considering w_t) is 0.209.

Variance due to model (considering z_t) is 0.805.

(ii) Calculated Chi - Square is Q = 66.52 for [(n/4)- 2] = 89 degrees of freedom. The upper 5% critical point for Chi - Square with 89 degrees of freedom is 112.02.

(iii) Figure 3.8 is the result of the cumulative periodogram test. The plots scatter along the theoretical line and are all within the 95% confidence limits.

Since in test (i) $\tilde{\sigma}_e^2 < \tilde{\sigma}_w^2 < \tilde{\sigma}_z^2$, in test (ii) the calculated Chi - Square is less than the theoretical one, and in test (iii) the cumulative periodogram test indicates no seasonal pattern, the model fitted to the transformed daily maxima of NO_x in section 3.4 can be finally accepted.

Since the models have passed through the diagnostic checks and are found adequate for explaining the behavior of the corresponding series, they can be used for forecasting. The models will then be used to forecast NO_x concentrations

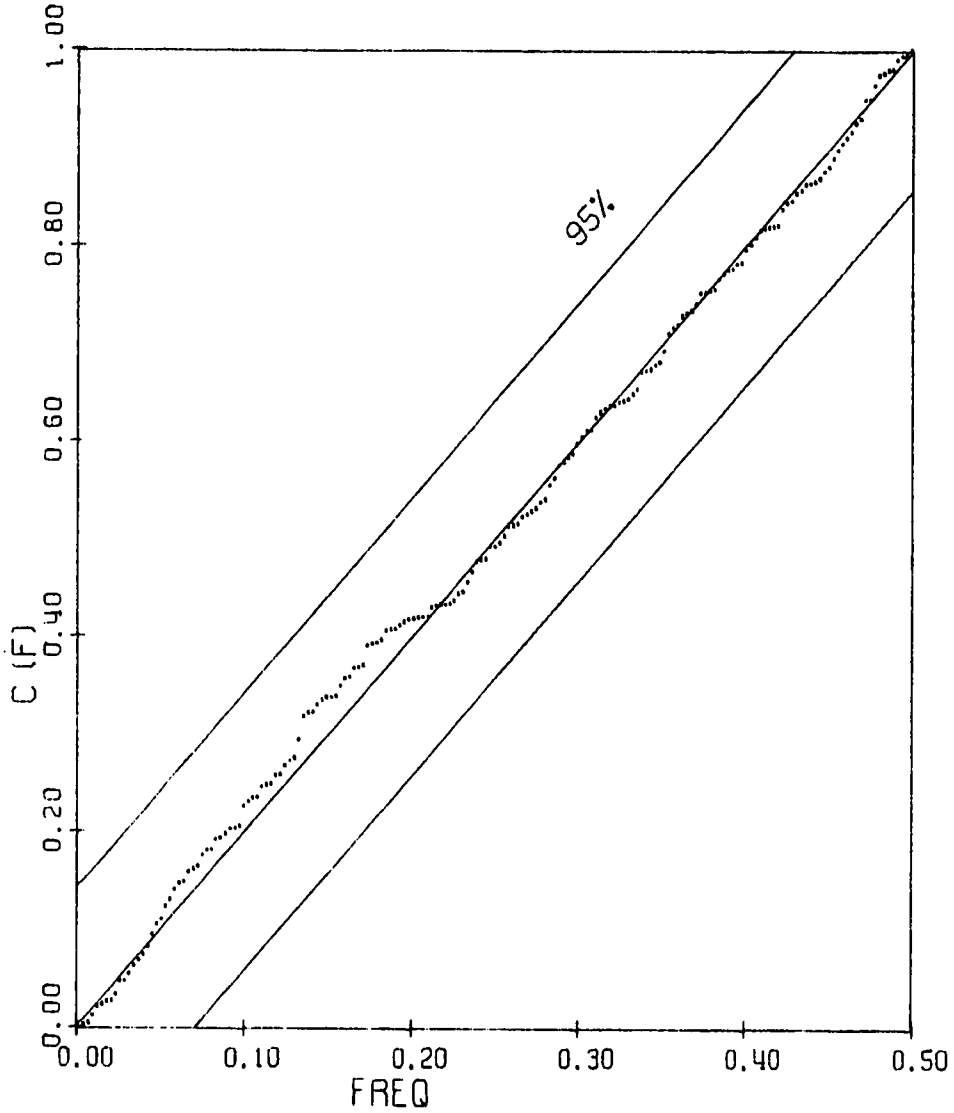

Fig. 3.8 Cumulative Periodogram for the residuals of the model fitted to the transformed Daily Maxima of NO_x

in the next section.

3.6 Forecasting with the Stochastic Models

From the models established in section 3.4 and subjected to diagnostic checks in section 3.5, forecast functions will be derived later on in this section. First, a brief description of confidence limits of forecasts will be given. Let $\hat{z}_t(L)$ be the forecast made at origin t for an observation z_{t+L}, which is L steps ahead. Then 95% confidence limits of the forecast for lead time L are

$$\hat{z}_t(L) \pm 1.96 \sqrt{V(L)} , \qquad (3.6.1)$$

where V(L) is the variance of the L steps - ahead - forecast error given by

$$V(L) = [1 + \sum_{j=1}^{L-1} \psi_j^2] \sigma_e^2$$

The weights ψ_j, j = 1, 2, ..., L - 1 can be obtained from the parameters of the fitted model by equating coefficients of powers

$$(1 - a_1B - ..... - a_pB^p)\ (1 - B)^d\ (1 + \psi_1B + \psi_2B^2 + ...)$$

$$= (1 - b_1B - b_2B^2 - ..... - b_qB^q)$$

More details about ψ_j can be obtained from Box and Jenkins [6].

The model fitted to the transformed NO_x hourly averages in section 3.4 is

$$z_t = z_{t-1} + e_t$$

and from it we can try to derive a forecast function. The forecast at origin t of an observation L steps ahead is given by

$$\hat{z}_t(L) = z_t + e_t(L)$$

Since $e_t(L)$ is unknown it is equated to its expected value zero and so the forecast function becomes

$$\hat{z}_t(L) = z_t$$

This forecast function shows that the best forecast value for all future time is the current value. Since the forecasts from origin t of the next L hourly values of NO_x supplied by this function is the value at the current origin the forecast for various L - values is simply a straight horizontal line.

The model fitted to the transformed daily maximum NO_x series in section 3.4 is

$$z_t = z_{t-1} + e_t - 0.5e_{t-1} - 0.26e_{t-2}$$

The forecast function derived from this model is

$$\hat{z}_t(L) = z_{t+L-1} + e_t - 0.5e_{t+L-1} - 0.26e_{t+L-2}$$

The terms on the right of the forecast function are treated according to the following four rules:

(i) The z_{t-j} (j=0,1,2,.....), which have already happened at origin t, are left unchanged.

(ii) The z_{t+j}(j=1,2,.....), which have not yet happened, are replaced by their forecasts $\hat{z}_t(j)$ at origin t.

(iii) The e_{t-j} (j=0,1,2,.....), which have happened, are available from $z_{t-j} - \hat{z}_{t-j-1}(1)$.

(iv) The e_{t+j} (j=1,2,.....), which have not yet happened, are replaced by zero, since zero is the unconditional expected value of e_t.

For lead 1 forecast, for example, the forecast function above can be written as

$$\hat{z}_t(1) = z_t - 0.5e_{t-1} - 0.26e_{t-2} \ .$$

Using (3.6.1) the 95% confidence interval of lead 1 forecast is found to be $\hat{z}_t(1) \pm 1.72$ for the transformed daily maxima of NO_x. Figure 3.9 shows the lead-one forecasts of the last 30 values of the transformed daily maximum NO_x series and their 95% confidence intervals. All the observed

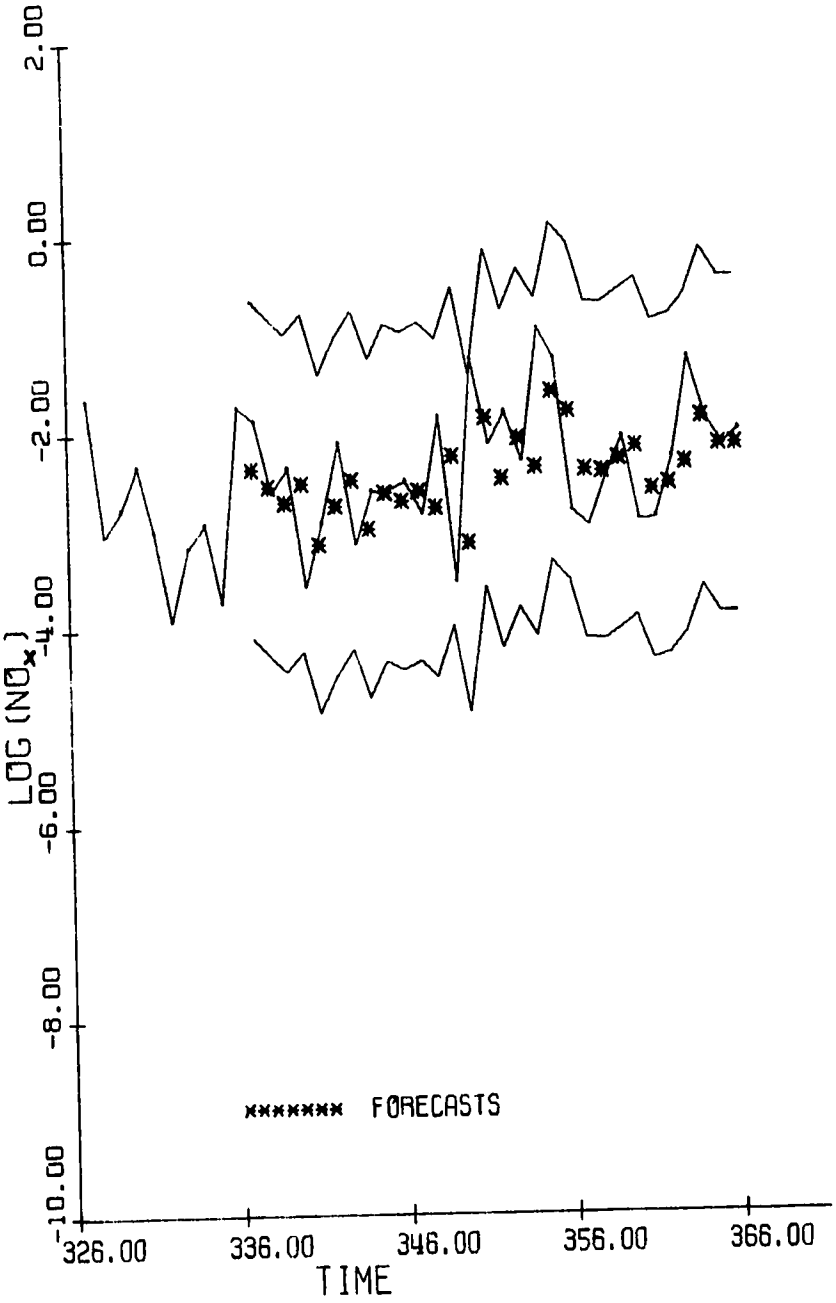

Fig. 3.9 Transformed Daily Maxima of NO_x and lead - one Forecasts for the last thirty days of the series using Stochastic Model

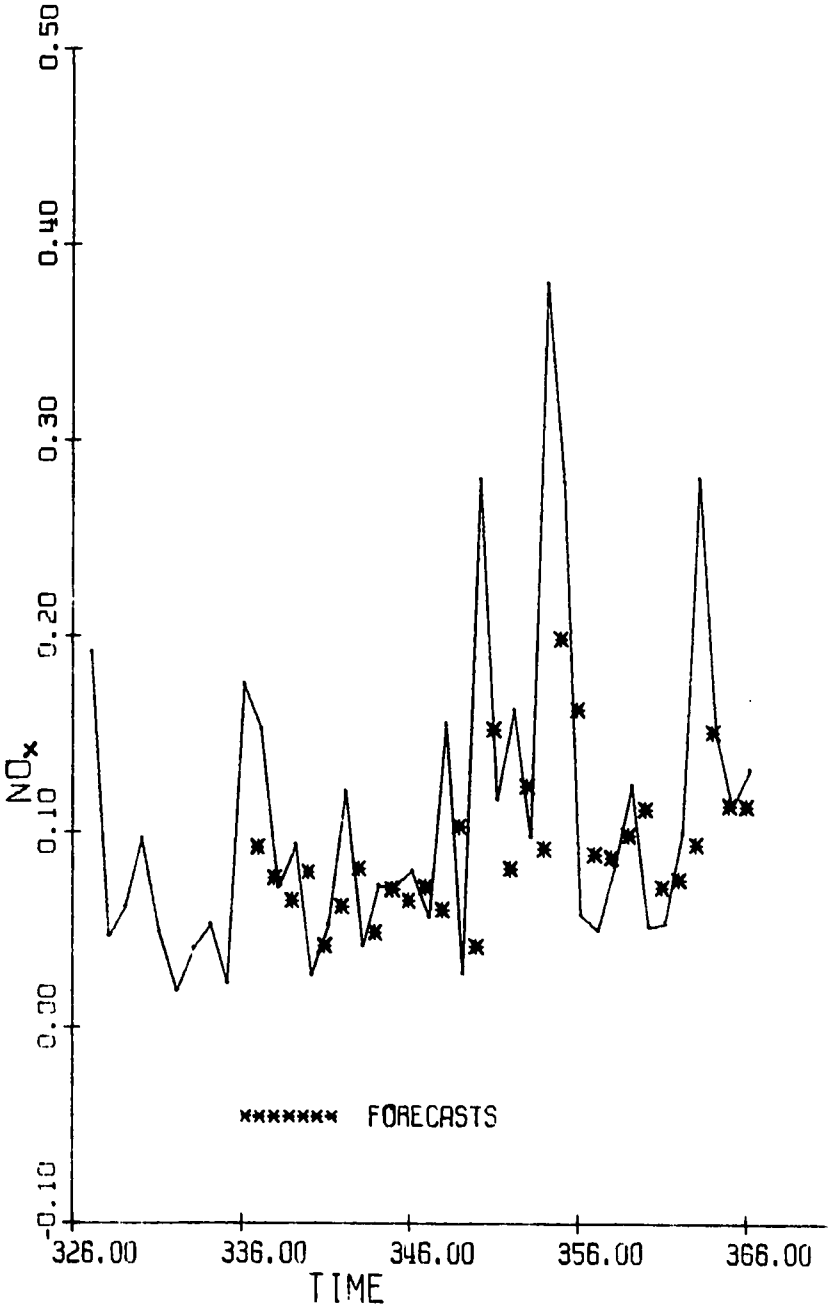

Fig. 3.10 Daily Maxima of NO_x and lead - one Forecasts for the last thirty days of the series

values lie within the corresponding confidence intervals. The forecasts obtained are transformed back to the original units by taking the antilogarithm and subtracting 0.001. These forecasts are presented in Table 3.5 and Figure 3.10 so that both the forecasts and the observed values can be compared in the original data frame. Table 3.6 contains the last 30 observations of the original series.

The forecasts of lead times 1 to 30 presented in Figure 3.11 also show that all the observed values lie within their confidence intervals. However, all forecasts from lead - 2 on have constant values. This fact is displayed on Figure 3.11 by the horizontal line traced by the lead times 2 to 30 forecast values. The lead times 1 to 30 forecasts are also transformed back to the raw data units and are presented in Table 3.7 and Figure 3.12.

TABLE 3.5 Lead-one Forecasts of the last thirty days of the Daily Maximum NO_x Series using Stochastic Model*

0.093	0.077	0.065	0.080	0.042	0.063
0.082	0.050	0.072	0.066	0.072	0.061
0.104	0.042	0.153	0.082	0.124	0.092
0.199	0.163	0.089	0.088	0.099	0.112
0.072	0.076	0.094	0.0152	0.114	0.114

* Read across the page. The forecasts are in ppm.

TABLE 3.6 The last thirty days of the Daily Maximum NO_x Series*

0.153	0.072	0.094	0.027	0.053	0.121
0.042	0.073	0.074	0.081	0.057	0.156
0.028	0.281	0.117	0.163	0.098	0.381
0.279	0.058	0.050	0.083	0.124	0.052
0.054	0.100	0.281	0.151	0.113	0.132

* Read across the page. The observations are in ppm.

TABLE 3.7 Leads one to thirty Forecasts of the last thirty days of Daily Maximum NO_x Series using Stochastic Model*

0.093	0.060	0.060	0.060	0.060	0.060
0.060	0.060	0.060	0.060	0.060	0.060
0.060	0.060	0.060	0.060	0.060	0.060
0.060	0.060	0.060	0.060	0.060	0.060
0.060	0.060	0.060	0.060	0.060	0.060

* Read across the page. The forecasts are in ppm.

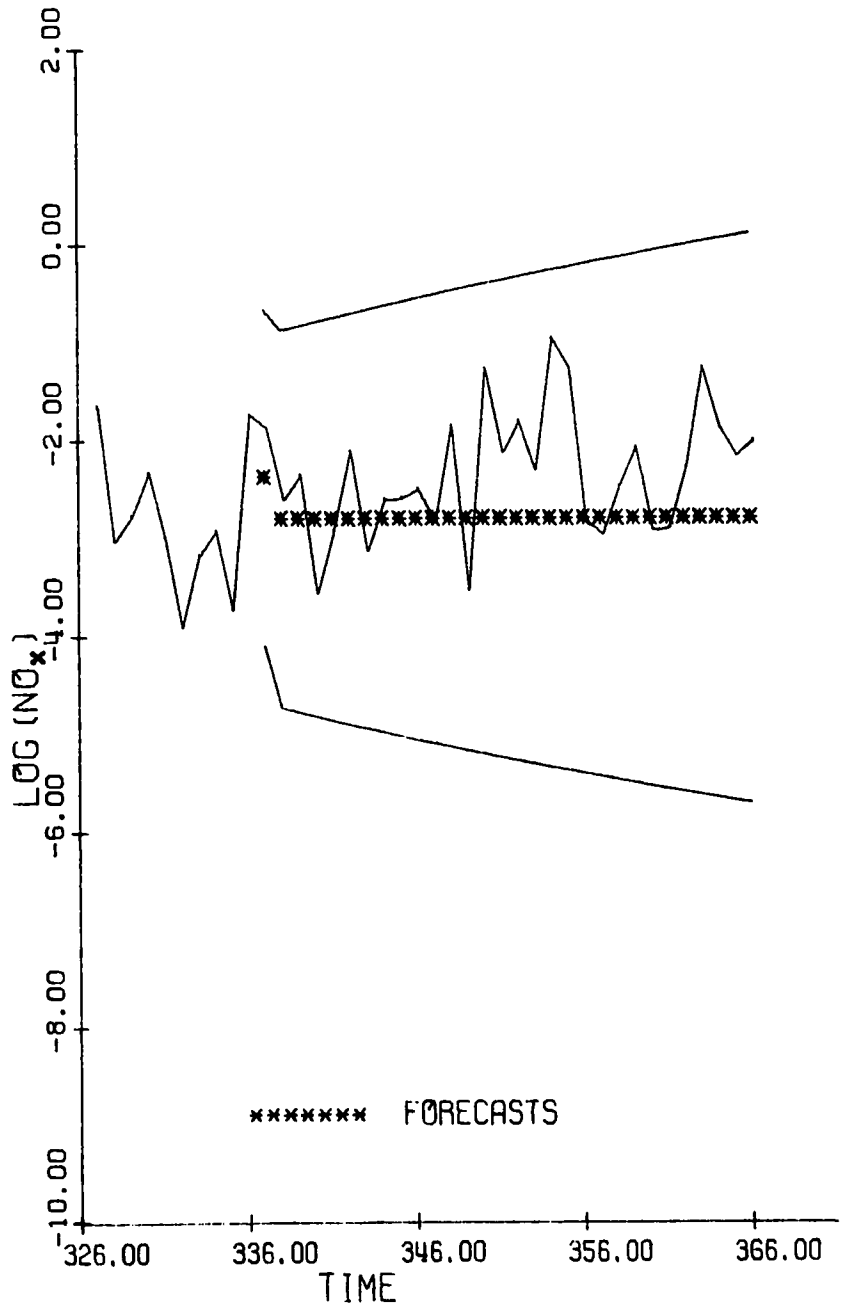

Fig. 3.11 Transformed Daily Maxima of NO_x and leads one to thirty days of the series using Stochastic Model

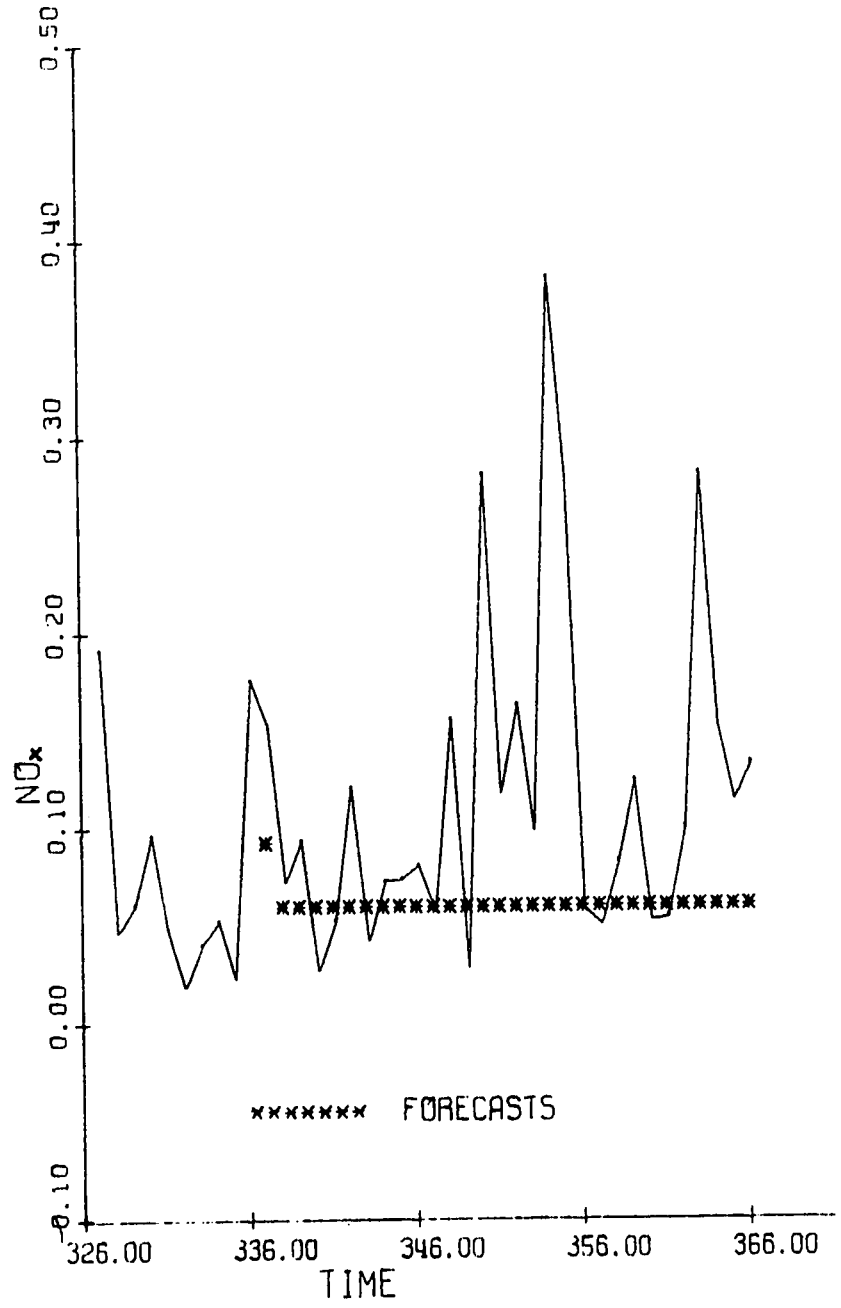

Fig. 3.12 Daily Maxima of NO_x and leads one to thirty Forecasts for the last thirty days of the series.

The forecasts presented in the tables above, especially the lead-one forecasts in Table 3.5, indicate that the stochastic model can be used to predict NO_x concentration since many of the forecasts are close to the observed values in Table 3.6. In Chapter IV, the dynamic system model will be developed and forecasts generated by it can be compared with the stochastic model forecasts.

CHAPTER IV

DYNAMIC SYSTEM MODEL

4.1 Definition

The Dynamic System Model developed for oxides of nitrogen in this chapter is known as a combined transfer function-noise model. In developing the model, the information supplied by the observed NO_x series will be combined with the information supplied by the series of factors influencing the behavior of NO_x in the urban atmosphere. The observed values of NO_x are regarded as output from a dynamic system while temperature, wind speed, and traffic flow are factors conjectured to influence NO_x in the urban atmosphere and are regarded as imputs (or leading indicators as economists know them) to the system. The dynamic system in this case is the urban atmosphere and Figure 4.1 shows a simple representation of it.

Since three measurable imputs to the system have been conjectured for this work the sort of transfer function model that can represent the system is known as a multiple input transfer function model whose general form may be written as

$$(1 - B)\, Z_t = v_1(B)\, (1 - B)\, X_{1,t} + \cdots\cdots + v_m(B)\, (1 - B)\, X_{m,t} + E_t, \qquad (4.1.1)$$

where $X_{1,t}, X_{2,t}, \cdots\cdots, X_{m,t}$ are m series of input

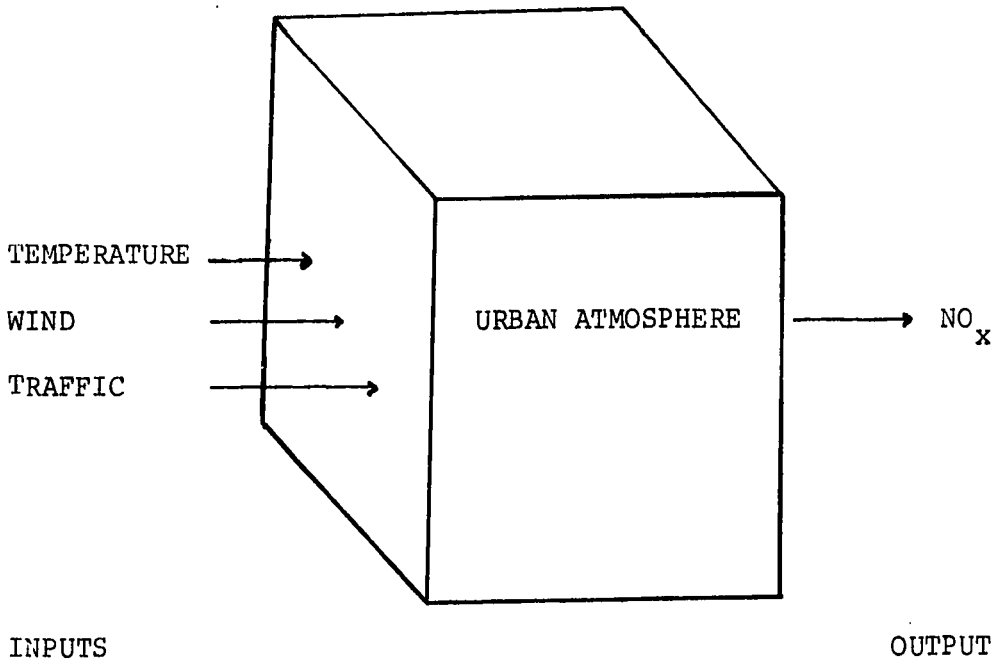

Fig. 4.1 Urban Atmosphere as a Dynamic System

factors, Z_t is the output series, and E_t is the series of noise disturbance (or imput to the system from unknown sources) not to be confused with the previously mentioned "white noise". $v_i(B)$ is an operator such that

$$v_i(B) = (v_{i0} + v_{i1}B + v_{i2}B^2 + ...)$$

and v_{i0}, v_{i1}, v_{i2}, ... are referred to as the impulse response function for imput $X_{i,t}$. B is again a backward shift operator defined as $BX_{i,t} = X_{i,t-1}$.

The model (4.1.1) relates incremental change in the output Z_t to the incremental changes in the input factors. The model as it is written involves a general assumption that each input factor requires only the first difference to make the process generating it stationary, but some may require higher differencing. Cases where the type of differencing varies among the input factors can be handled and an example of it is discussed in section 4.2. For simplicity the assumption is introduced in this general definition.

The function (4.1.1) can be expressed as follows:

$$\begin{aligned}(1 - B)\, Z_t = {} & a_1^{-1}(B)\, b_1\,(B)\,(1 - B)\, X_{1,t-d_1} + ... \\ & + a_m^{-1}(B)\, b_m\,(B)\,(1 - B)\, X_{m,t-d_m} \qquad (4.1.2) \\ & + E_t,\end{aligned}$$

where $a_j(B)$ and $b_j(B)$ are operators defined for each input series as follows:

$$a_j(B) = (1 - a_{j,1}B - \ldots\ldots - a_{j,p}B^p),$$

$$b_j(B) = (b_{j,0} - b_{j,1}B - \ldots\ldots - b_{j,s}B^s).$$

An incremental change in the level of one input may not have an immediate effect on the output level. This time delay for the j^{th} input factor known as delay response time is represented by d_j. For a system with only one input, the model can be written in the form (4.1.1) and (4.1.2) as follows:

$$\begin{aligned}(1 - B)\, Z_t &= (v_0 + v_1B + v_2B^2 + \ldots)\ (1 - B)\, X_t + E_t \\ &= v(B)\ (1 - B)\, X_t + E_t \\ &= a^{-1}(B)b(B)\ (1 - B)\, X_{t-d} + E_t,\end{aligned}$$

where parameters a_i and b_i are obtainable by equating coefficients of B in the equation

$$v(B) = a^{-1}(B)b(B),$$

$$\text{i.e. } (1 - a_1B - a_2B^2 - \ldots\ldots - a_pB^p)\ (v_0 + v_1B^2 + \ldots\ldots)$$
$$= (b_0 - b_1B - \ldots\ldots - b_sB^2)\ B^d.$$

This equation gives the following relations from which a_i and b_i could be calculated if the v_j's are known:

$$
\begin{aligned}
v_j &= 0, && j < d, \\
v_j &= b_0 && j = d, \\
v_j &= a_1 v_{j-1} + a_2 v_{j-2} + \cdots\cdots + a_p v_{j-p} - b_{j-d}, \\
& && j = d+1, d+2, \ldots\ldots, d+s \\
v_j &= a_1 v_{j-1} + a_2 v_{j-2} + \cdots\cdots + a_p v_{j-p}, && \qquad (4.1.3) \\
& && j > d+s.
\end{aligned}
$$

Assuming that the impulse response function v_j, j=0, 1,2, ... is known, then p, and s, the number of the a_i(i = 1, 2, ..., p) parameters, the b_i (i = 1, 2, ..., s) parameters, respectively can be determined from the v_j's. The delay response time d may easily be determined from the first expression in (4.1.3). Let the absolute value of v_j be a maximum at j = m, then s can be taken to be any value from 0,1, ..., m - d. If there are several consecutive maxima, m takes on the value j of the last maximum. The value of p can be determined by looking at v_j from v_m onwards. From v_m the v_j's behave like the autocorrelation function of an autoregressive process described in Chapter III; hence, their pattern can be used to identify the number of a_i parameters in the model. In practice the number of autoregressive parameters usually does not exceed two, i.e., $p \leq 2$. There are no

restrictions on the magnitudes of b_0, b_1, ..., b_s, whereas the parameters a_1, a_2, ..., a_p have to obey stationarity rules analogous to those of autoregressive parameters discussed in section 3.2. One of these rules is that $a_1 + a_2 \ldots\ldots + a_p < 1.0$, which greatly helps at the model identification stage.

The steps involved in building a dynamic system model are very similar to the ones outlined in Chapter III for the stochastic model. They are in order, the identification step, including initial estimates of the parameters, the estimation step, the diagnostic checking step and the forecasting step; and will be dealt with in sections 4.2, 4.3, 4.4, and 4.5, respectively.

Before building the model, the input factors should be described briefly. The temperature and the wind speed data are measurements taken at the Edmonton Industrial Airport. The hourly temperature is presented as Series C while the daily maximum temperature is presented as Series D. All temperature observations are given in degrees Fahrenheit. Series E contains the hourly wind speed data and Series F is the daily wind speed data, all measured in mph. The traffic flow data are the observations made on Jasper Avenue and 122 Street in Edmonton which is 3 blocks north and 13 blocks west of the NO_x monitoring station. Series G consists of the hourly data while Series H consists of the daily data of the total number of vehicles.

All the data in Series C to H correspond to the same periods relevant to the NO_x data in Series A and Series B. Thus the hourly data are observed for the period February 22 to March 7, 1967 and the daily data for the period April 1, 1971 to March 31, 1972. However the data available for the input factors leave something to be desired in the sense that they are not collected where NO_x is being monitored. Since the input factors are monitored far away from the NO_x monitoring station they may fail to give any explanation about the NO_x behavior. As the results of the transfer function identification, discussed in the next section, shows, only temperature helps to explain the behavior of NO_x in the city.

4.2 Dynamic System Model - Identification

In this section the transfer function model connecting output NO_x to each of the input factors will be identified wherever it is found that there is a dynamic relationship between the output and the particular input. Since the correlation function is used extensively, the correlation function and its sample estimator are defined below.

The crosscorrelation function $\rho_{xz}(k)$, of input X_t and output Z_t at lag k is given by

$$\rho_{xz}(k) = \frac{E[(X_t - U_x)\ (Z_{t+k} - U_z)]}{\sigma_x\ \sigma_z} \quad \text{for } k = 0, 1, 2, ..,$$

and

$$\rho_{xZ}(k) = \frac{E[(X_{t+k} - U_x)\ (Z_t - U_Z)]}{\sigma_x\ \sigma_Z}$$

$$\text{for } k = -1, -2, \ldots\ldots,$$

where U_x is the mean of the process X_t, and U_Z is the mean of the process Z_t. As the theoretical crosscorrelation is unknown the observed sample corsscorrelation function is used for identification. Usually the crosscorrelation at lags $k = 0, 1, 2, \ldots\ldots,$ is the required part of the function needed for identification of the transfer function, hence that part of the observed crosscorrelation function is given here. The observed sample crosscorrelation function $r_{xz}(k)$, is defined as

$$r_{xz}(k) = C_{xz}(k)/s_x s_z, \qquad (4.2.1)$$

where $C_{xz}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})\ (z_{t+k} - \bar{z})$, $k = 0, 1, 2, \ldots,$

$$s_x = \sqrt{C_{xx}(0)},$$

$$s_z = \sqrt{C_{zz}(0)},$$

$\bar{x}$ is the observed mean of the x_t series, and $\bar{z}$ is the observed mean of the z_t series. From now on the observed sample crosscorrelation function of x_t and z_t will be referred to simply as the crosscorrelation function and

will be written as r(k).

The impulse response function v_j, j = 0, 1, 2, ..., relating the output z_t to the input x_t can be obtained from the crosscorrelation function of x_t and z_t. One method of obtaining the impulse response function involves solving a set of linear equations for v_j. This is like solving the Yule - Walker equations (3.3.1) in section 3.3, where crosscovariances from lag zero to a conjectured lag K are substituted for autocorrelations on the right side of the equations, and autocovariances of the input series are substituted for the autocorrelations on the left side. This approach is defective in that it requires the knowledge of a lag K beyond which v_j is effectively zero. In the preferred second method which is used in the algorithm (Appendix D) for the identification of the transfer function model, this lag K problem is not encountered because the crosscorrelation function, hence v_j decays fast and so K can be set as low as 10. K is set arbitrarily equal to 14 for the algorithm in this work. A brief discussion of the second method is given in the following paragraph.

A stochastic model is fitted to the input series. The stochastic model could be any of the model types (3.1.3) discussed in section 3.1. The stochastic model is then used to prewhiten (i.e. to obtain residuals of input x_t) the input series $\{x_t\}$ such that the residuals y_t obtained are a white noise process. The same model

fitted to the input series is used to transform the output series $\{z_t\}$ such that some sort of residuals w_t, say, which are not necessarily white noise will be obtained. The crosscorrelation function r(k) of the input residual y_t and the output residual w_t obtained by this procedure is estimated according to (4.2.1) for k = 0, 1, ..., K where for practical purposes $K \leq 14$. Then the impulse response function v_k is given by

$$v_k = \frac{r(k)s_w}{s_y}, \text{ for } k = 0, 1, \ldots\ldots, K. \qquad (4.2.2)$$

After obtaining the impulse response function v_k in this way, the parameters a_i (i = 1, 2, ..., p) and b_i (i = 0, 1, ..., s) can be obtained according to (4.1.3).

Since v_k as defined in (4.2.2) is directly proportional to r(k), r(k) can be used to decide the significance of v_k. If r(k) is significantly different from zero at lag k, it can be concluded that v_k is significantly different from zero at lag k. Since r(k) has approximately a normal distribution, twice the observed standard deviation of r(k) can be used to test whether or not r(k) is significantly different from zero with approximately 95% probability. Furthermore one of the series (i.e. residuals y_t) used to estimate r(k) is known to be white noise, so that the standard deviation (STD)

of r(k) is given by

$$STD = \frac{1}{\sqrt{(n-k)}} \qquad (4.2.3)$$

More details about the distribution of r(k) are given in Bartlett [4]. Further details of the identification procedure for the transfer function model using the cross-correlation function can be found in Box and Jenkins[6] while the use of cross spectral analysis for the same purpose can be found in Jenkins and Watts [28].

In the rest of this section the results obtained by following the above identification procedure, which is the basis of the transfer function model identification algorithm in Appendix D, will be presented. The algorithm (implemented in Fortran IV; see program 4 in Owolabi [52]) is used to identify transfer function models for the NO_x hourly average series, and the daily maximum NO_x series. For the purpose of the analysis here and the rest of the chapter, the input series and the output series are transformed according to the transformations presented in Table 4.1. From now on, reference to any one of these series implies reference to its transformation.

TABLE 4.1 Raw Data Transformation

Series	Raw Data	Transformation
NO_x	z_t	$\ln(z_t + 0.001)$
Temperature	x_t	$\ln(x_t + 50.0)$
Wind Speed	x_t	$\ln(x_t)$
Traffic Flow	x_t	$\ln(x_t)$

NO_x Hourly Average

Figure 4.2 shows the plot of hourly NO_x series with the corresponding hourly temperature series, wind speed series, and traffic flow series. The plotted series are the transformations of the data in series A for NO_x hourly averages, series C for hourly temperature, series E for hourly wind speed, and series G for hourly traffic flow. The relationship of NO_x hourly averages with any of the input factors is not obvious from the plot. However, the possible transfer function relating NO_x to each of them is investigated using the procedure emphasized earlier in this section and the algorithm in Appendix D.

The stochastic model which was found to fit the hourly temperature series is a modified ARMA (2,0). Using the stochastic model fitting method of Chapter III, the hourly temperature model turns out to be

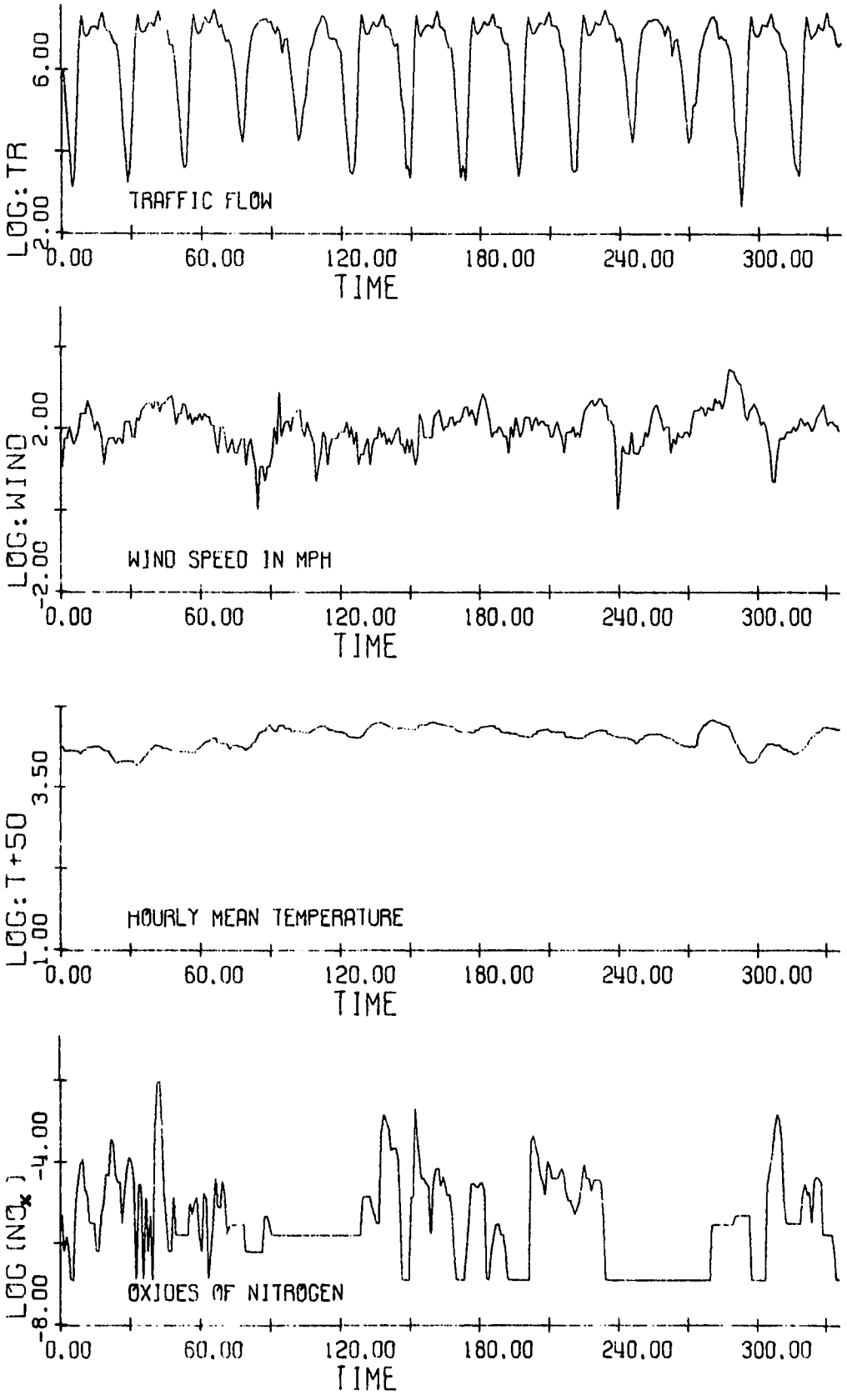

Fig. 4.2 Transformed hourly Inputs and Output

$$(1 - 0.338B - 0.211B^2)\ (1 - B)\ X_t = y_t \qquad (4.2.4)$$

This model is used to "prewhiten" the temperature series such that the remaining series of residuals y_t is merely white noise and, hence, have zero autocorrelation. The same model (4.2.4) is used to transform the z_t series, the resulting series is of course not necessarily uncorrelated. The estimated crosscorrelation function r(k), of the two new series is plotted in Figure 4.3 and its values with the standard deviation STD, are given in Table 4.2. An estimate of the impulse response function v_k is obtained from the r(k) according to (4.2.2). At this point, the significance of r(k) implies the significance of v_k. Comparing r(k), k = 0, 1, ..., 14 with twice its standard deviation [i.e. ±2 (STD)], in Table 4.2, it is found that only r(3) may be considered as significantly different from zero. r(3) = 0.107 is comparatively close to twice its standard deviation of 0.110. All other r(k), $k \neq 3$ are not as close to twice their standard deviations as r(3).
This means the effect of an incremental change in the level of hourly temperature is delayed two hours before it is apparent in the level of hourly NO_x concentration in the urban atmosphere. Since only $\hat{v}_3$ is significantly different from zero, $\hat{b}_0 = \hat{v}_3 = 2.82$ and b_0 is the only

parameter in the transfer function. Therefore, the transfer function describing hourly NO_x in terms of

TABLE 4.2 Crosscorrelation and Impulse Response Functions of prewhitened hourly Temperature and transformed NO_x

Lag k	r(k)	STD of r(k)	2 x (STD)	v_k
0	0.022	0.055	0.110	0.588
1	0.005	0.055	0.110	0.119
2	-0.049	0.055	0.110	-1.289
3	0.107	0.055	0.110	2.824
4	0.046	0.055	0.110	1.202
5	-0.048	0.055	0.110	-1.272
6	0.015	0.055	0.110	0.397
7	0.101	0.055	0.110	2.675
8	-0.015	0.055	0.110	-0.388
9	0.049	0.055	0.110	-1.300
10	-0.075	0.055	0.110	-1.988
11	0.004	0.056	0.112	0.114
12	-0.055	0.056	0.112	-1.441
13	0.016	0.056	0.112	0.435
14	-0.014	0.056	0.112	-0.370

$s_w = 0.724$

$s_y = 0.027$

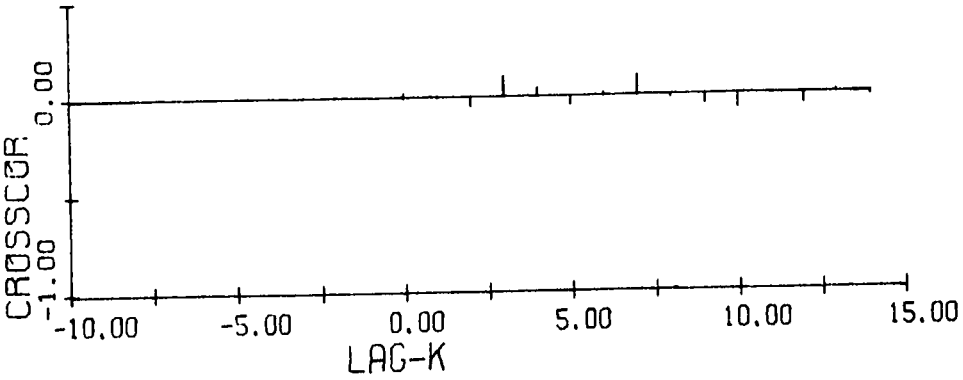

Fig. 4.3 Crosscorrelation Function of prewhitened hourly Temperature and transformed NO_x

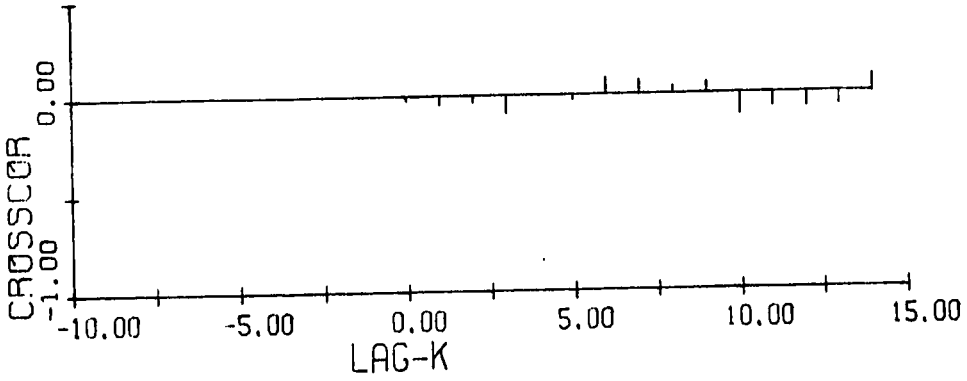

Fig. 4.4 Crosscorrelation Function of prewhitened hourly Wind Speed and transformed NO_x

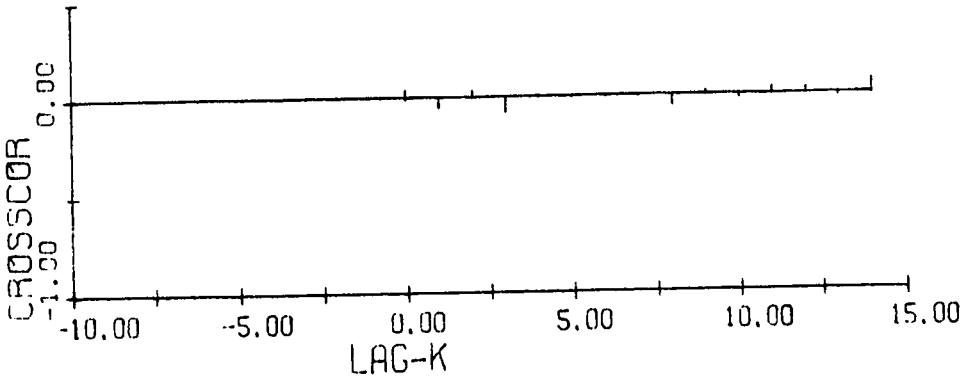

Fig. 4.5 Crosscorrelation Function of prewhitened hourly Traffic Flow and transformed NO_x

hourly temperature alone can be written as

$$(1 - B)\ Z_t = \hat{b}_0\ (1 - B)\ X_{t-3}. \qquad (4.2.5)$$

The model fitted to the hourly wind speed series is a modified ARMA (3,0) and has the following form:

$$(1 + 0.239B + 0.139B^2 + 0.165B^3)\ (1 - B)\ X_t = y_t\ . \qquad (4.2.6)$$

The hourly wind speed series is prewhitened and the hourly NO_x series transformed according to this wind speed model. The crosscorrelations of the two residuals y_t and w_t are estimated and presented in Table 4.3 and Figure 4.4. As before Table 4.3 also contains estimates of v_k and the standard deviation (STD) of r(k). From Table 4.3 it can be seen that only r(10) is significantly different from zero when compared with twice its standard deviation (i.e. r(10) = -0.116 ± 0.110). However, the fact that the significance of r(10) is a border line case coupled with the fact that lag 10 is such a large lag makes the significance of r(10) doubtful. The parameter $\hat{b}_0 = \hat{v}_{10}$ obtained by the identification procedure was subjected to the efficient estimation method of section 4.3 and $\hat{b}_0$ so produced was found to be not significantly different from zero, and, hence, was omitted from the model. This means that a change in the hourly wind speed

TABLE 4.3 Crosscorrelation and Impulse Response Functions of prewhitened hourly Wind Speed and transformed NO_x

Lag k	r(k)	STD of r(k)	2 x (STD)	v_k
0	-0.022	0.055	0.110	-0.049
1	-0.045	0.055	0.110	-0.098
2	-0.032	0.055	0.110	-0.071
3	-0.095	0.055	0.110	-0.209
4	-0.004	0.055	0.110	-0.009
5	-0.026	0.055	0.110	-0.056
6	0.090	0.055	0.110	0.198
7	0.075	0.055	0.110	0.165
8	0.041	0.055	0.110	0.090
9	0.059	0.055	0.110	0.129
10	-0.116	0.055	0.110	-0.254
11	-0.073	0.056	0.112	-0.160
12	-0.078	0.056	0.112	-0.170
13	-0.067	0.056	0.112	-0.148
14	0.088	0.056	0.112	0.194

$s_w = 0.681$

$s_y = 0.311$

does not seem to influence hourly changes in the NO_x level. The wind speed - ignoring direction - does not seem to influence NO_x since the wind direction is probably an important characteristic affecting the NO_x level. In fact it has been shown in the "Air Pollution Summary Edmonton" published by the Environmental Health Services Division of Alberta Department of Health that south and southwest winds are correlated with high NO_x concentration in the Administration Building.

The hourly traffic flow series is described by a seasonal model with a 24 - hour period. The model is

$$(1 - 0.266B)\ (1 - B^{24})\ X_t = y_t + 0.568y_{t-1}. \qquad (4.2.7)$$

As for temperature and wind this model (4.2.7) is used to prewhiten the hourly traffic flow series and to transform the NO_x hourly average series. The crosscorrelations of the two new series are given in Table 4.4 and plotted in Figure 4.5. Comparing r(k)'s with ±2(STD) in Table 4.4 it can be seen that none of the r(k)'s are significantly different from zero. This implies that an incremental change in the level of hourly traffic flow has no effect on hourly NO_x concentration. The result here is quite astonishing in view of the fact that automobile exhaust constitutes a source of NO_x in urban areas. However, two reasons may account for this apparent lack of automobile effect here. One is that most automobiles nowadays

TABLE 4.4 Crosscorrelation and Impulse Response Functions of prewhitened hourly Traffic Flow and transformed NO_x

Lag K	r(k)	STD of r(k)	2 x (STD)	v_k
0	0.044	0.057	0.114	0.145
1	-0.055	0.057	0.114	-0.181
2	0.039	0.057	0.114	0.126
3	-0.082	0.057	0.114	-0.269
4	-0.007	0.057	0.114	-0.024
5	-0.005	0.057	0.114	-0.026
6	-0.009	0.057	0.114	-0.031
7	-0.003	0.057	0.114	-0.011
8	-0.054	0.057	0.114	-0.177
9	0.021	0.057	0.114	0.069
10	-0.018	0.058	0.116	-0.059
11	0.037	0.058	0.116	-0.122
12	0.034	0.058	0.116	0.111
13	-0.016	0.058	0.116	-0.052
14	0.073	0.058	0.116	0.241

$S_w = 0.996$

$s_y = 0.304$

have pollution control devices. The other and perhaps the most important in this case is that the traffic flow data available are collected far away from the Administration Building where NO_x is monitored.

It has been found that only hourly temperature is related to hourly NO_x. In the following discussion a possible transfer function connecting daily maxima of NO_x with each of the input factors will be identified.

Daily Maxima of NO_x

The daily maximum NO_x series and the daily temperature, wind, and traffic flow series are subjected to the same type of analysis as the hourly data. Series D is the daily maximum temperature series, series F is the daily mean wind speed series, and series H is the daily traffic flow series. The daily maxima of NO_x series is of course Series B. The transformation of the daily data is identical to the transformation used for the hourly data, see Table 4.1. The transformed daily data are plotted in Figure 4.6. Again the relationship of NO_x with any of the three input factors is not obvious from Figure 4.6, however only a visual comparison is of little value here, as the hourly data indicated.

The stochastic model fitted to the daily maximum temperature is

$$(1 - 0.19B + 0.183B^2)\ (1 - B)\ X_t = y_t \qquad (4.2.8)$$

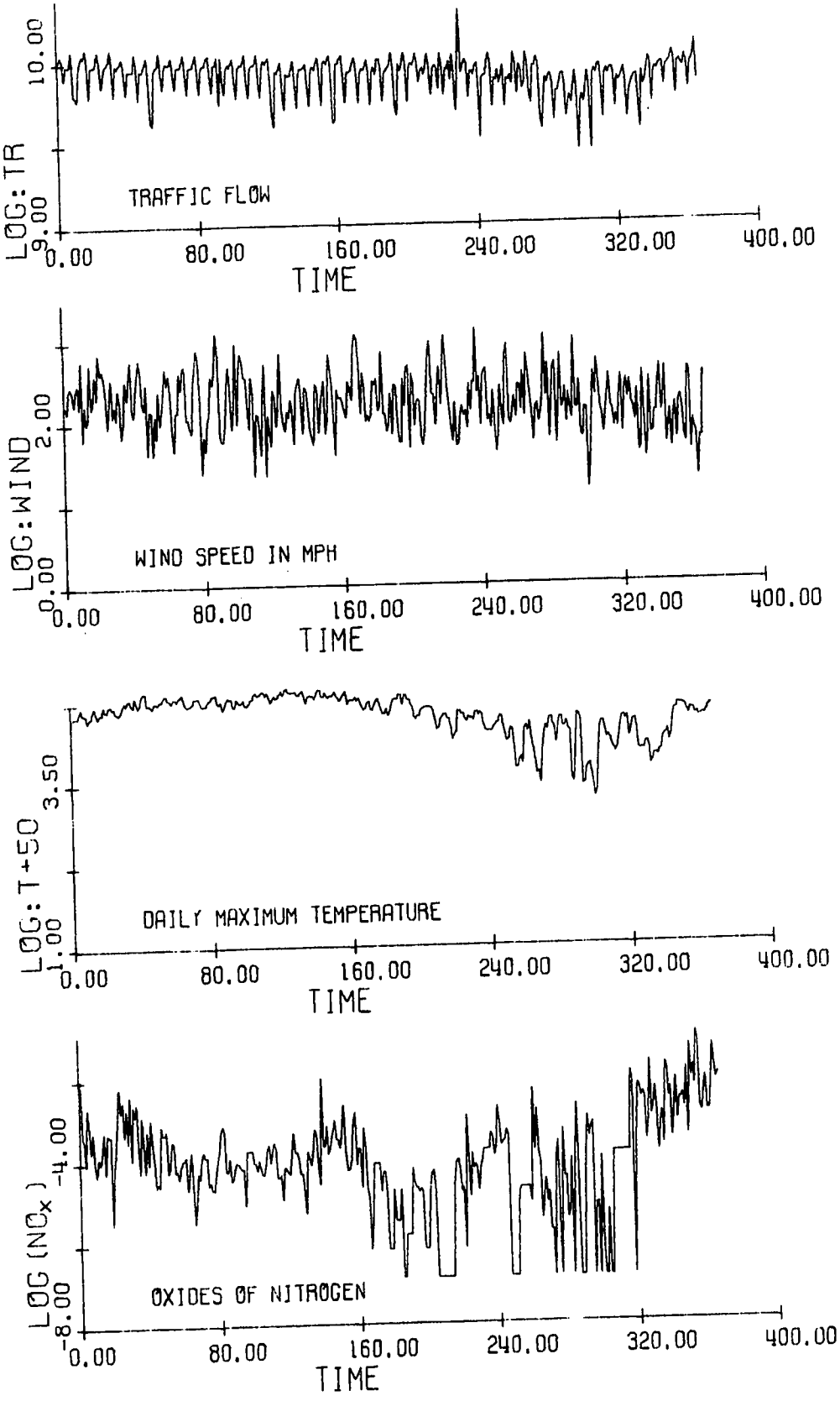

Fig. 4.6 Transformed daily Inputs and Output

After using the model to prewhiten the daily temperature series and transform the NO_x series the crosscorrelation function of the two resulting series are estimated. The estimated crosscorrelation function r(k) is plotted in Figure 4.7 and its values with the standard deviation (STD) are given in Table 4.5. An estimate of the impulse response function v_k is obtained from r(k) according to (4.2.2). Comparing r(k), k = 0, 1, ..., 14 with twice its standard deviation [i.e. ±2(STD)] in Table 4.5 it can be seen that r(0) is significantly different from zero since r(0) ±2 STD = 0.113 ±0.104 do not include zero, r(2) can be regarded at first sight as being significantly different from zero. However, results from efficient estimation of parameters in section 4.3 indicate that r(2) is not significantly different from zero. So only one impulse response weight is available and $\hat{b}_0 = \hat{v}_0 = 1.43$. The transfer function model identified for temperature alone is

$$(1 - B)\ Z_t - \hat{b}(1 - B)\ X_t \qquad (4.2.9)$$

The stochastic model fitted to the daily mean wind speed is

$$(1 + 0.436B + 0.337B^2)\ (1 - B)\ X_t = y_t. \qquad (4.2.10)$$

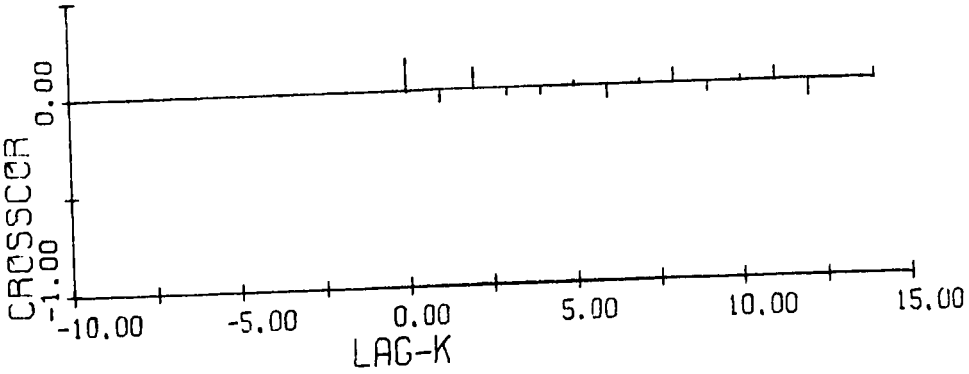

Fig. 4.7 Crosscorrelation Function of prewhitened daily Temperature and transformed NO_x

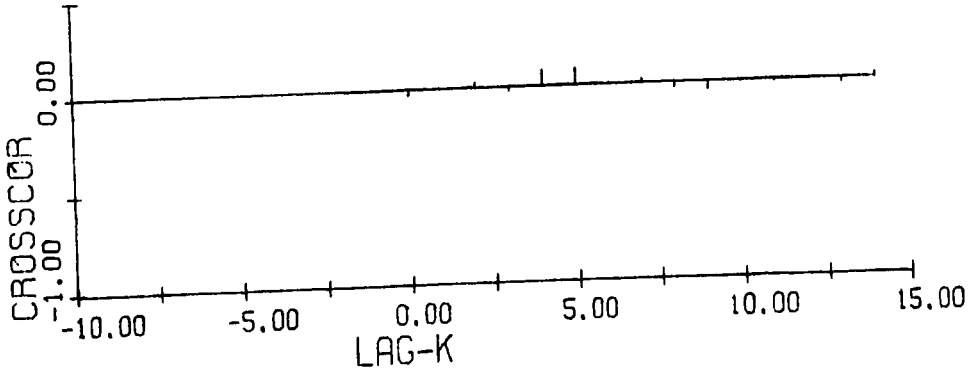

Fig. 4.8 Crosscorrelation Function of prewhitened daily Wind Speed and transformed NO_x

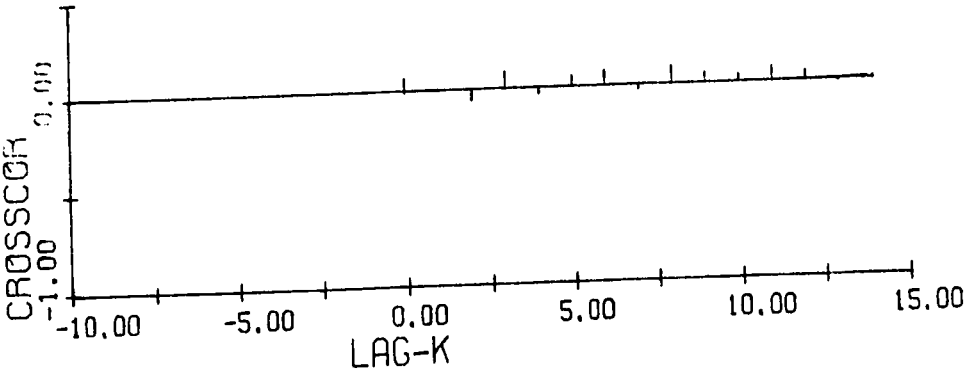

Fig. 4.9 Crosscorrelation Function of prewhitened daily Traffic Flow and transformed NO_x

TABLE 4.5 Crosscorrelation and Impulse Response Functions of prewhitened daily Temperature and transformed NO_x

Lag k	r(k)	STD of r(k)	2 x (STD)	v_k
0	0.171	0.052	0.104	1.426
1	-0.067	0.052	0.104	-0.554
2	0.113	0.052	0.104	0.941
3	-0.043	0.053	0.106	-0.355
4	-0.044	0.053	0.106	-0.362
5	0.022	0.053	0.106	0.183
6	-0.072	0.053	0.106	-0.602
7	0.022	0.053	0.106	0.179
8	0.074	0.053	0.106	0.620
9	-0.053	0.053	0.106	-0.437
10	0.029	0.053	0.106	0.241
11	0.069	0.053	0.106	0.575
12	-0.091	0.053	0.106	-0.753
13	0.006	0.053	0.106	0.050
14	0.048	0.053	0.106	0.400

$s_w = 1.058$

$s_y = 0.127$

After the usual prewhitening and transformation operations using model (4.2.10), the r(k)'s for the new series obtained are estimated along with the v_k's. Table 4.6 contains the r(k)'s with their standard deviations and the v_k's. Figure 4.8 shows the r(k)'s. From Table 4.6, it can be seen that none of the cross correlations are significantly different from zero, hence as in the hourly situation,the daily mean wind speed fails to contribute significantly to the behavior of the daily maximum level of NO_x in the city.

The daily traffic flow series is described by a seasonal model with a 7 - day period. The model is

$$(1 - 0.276B)\ (1 - B^7)\ X_t = y_t. \qquad (4.2.11)$$

As for temperature and wind this model (4.2.11) is used to prewhiten the daily traffic flow series and to transform the NO_x daily average series. The crosscorrelations of the two new series are given in Table 4.7 and plotted in Figure 4.9. Comparing r(k)'s with 2(STD) in Table 4.7 it can be seen that none of the r(k)'s are significantly different from zero. This means no transfer function connects the daily maximum NO_x with the daily traffic flow.

Since the foregoing show that there is no justification for including both wind speed and traffic flow in the transfer function model for NO_x, the combined transfer function - noise model for either hourly or daily series

TABLE 4.6 Crosscorrelation and Impulse Response Functions of prewhitened daily Wind Speed and transformed NO_x

Lag k	r(k)	STD of r(k)	2 x (STD)	v_k
0	-o.036	0.052	0.104	-0.055
1	0.000	0.052	0.104	0.001
2	0.024	0.052	0.104	0.057
3	-0.025	0.053	0.106	-0.059
4	0.081	0.053	0.106	0.193
5	0.085	0.053	0.106	0.201
6	-0.007	0.053	0.106	-0.018
7	0.016	0.053	0.106	0.039
8	-0.029	0.053	0.106	-0.069
9	-0.044	0.053	0.106	-0.104
10	-0.006	0.053	0.106	-0.015
11	-0.013	0.053	0.106	-0.031
12	-0.005	0.053	0.106	-0.011
13	-0.025	0.053	0.106	-0.060
14	0.025	0.053	0.106	0.059

$s_w = 0.911$

$s_y = 0.382$

TABLE 4.7 Crosscorrelation and Impulse Response Function of prewhitened Traffic Flow and transformed NO_x

Lag k	r(k)	STD of r(k)	2 x (STD)	v_k
0	0.066	0.053	0.106	0.900
1	0.010	0.053	0.106	0.134
2	-0.059	0.053	0.106	-0.803
3	0.087	0.053	0.106	1.183
4	-0.039	0.053	0.106	-0.525
5	0.059	0.053	0.106	0.795
6	0.079	0.053	0.106	1.071
7	-0.029	0.053	0.106	-0.396
8	0.094	0.053	0.106	1.270
9	0.054	0.053	0.106	0.727
10	0.038	0.054	0.108	0.519
11	0.078	0.054	0.108	1.054
12	0.047	0.054	0.108	0.631
13	-0.012	0.054	0.108	-0.163
14	0.021	0.054	0.108	0.279

$s_w = 1.180$

$s_y = 0.087$

will include only temperature X_t and the noise term E_t. Thus, using (4.2.5), the combined transfer function - noise model for NO_x is

$$(1 - B)\ Z_t = \hat{b}(1 - B)\ X_{t-s} + E_t\ . \qquad (4.2.12)$$

For the hourly series when the initial estimate of b is $\hat{b} = 2.82$, and from (4.2.9) it is

$$(1 - B)\ Z_t = \hat{b}(1 - B)\ X_t + E_t\ . \qquad (4.2.13)$$

For the daily series where the initial estimate of b is $\hat{b} = 1.43$.

The stochastic models identified for E_t in each of (4.1.12) and (4.2.13) are of the form:

$$E_t = e_t - h_1 e_{t-1} - h_2 e_{t-2} - h_3 e_{t-3}, \text{ and} \qquad (4.2.14)$$

$$E_t = e_t - h_1 e_{t-1} - h_2 e_{t-2}, \qquad (4.2.15)$$

respectively. The parameters $h_i (i = 1,2,.....)$ for each of these models will be estimated along with the parameter of the corresponding transfer function in the next section.

Before the parameter estimation stage it is necessary to describe briefly at this point what the combined transfer function - noise model should have looked like if any of the models (4.2.12) and (4.2.13) had at least two

input factors instead of just one. This is necessary for this approach to be generally useful for pollution models although the data does not justify more than one input in this thesis. Let the daily series,for example,have two input factors related to it instead of one. Then we would have two identified transfer functions for the output series. Let traffic flow be the second input factor so related where the relationship is expressible simply as

$$(1 - B^7)\ Z_t = b(1 - B^7)\ X_t \ . \qquad (4.2.16)$$

Then combining (4.2.9) for daily temperature with (4.2.16) for daily traffic flow, the two identified input transfer functions could have been combined as follows:

$$\Rightarrow \quad Z_t = (1 - B)^{-1} b_1(1 - B)\ X_{1,t} + (1 - B^7)^{-1}\ b_2(1 - B^7)\ X_{2,t} \qquad (4.2.17)$$

$$\Rightarrow \quad Z_t = \frac{b_1(1 - B)\ X_{1,t}}{(1 - B)} + \frac{b_2(1 - B^7)\ X_{2,t}}{(1 - B^7)}$$

$$\Rightarrow \quad (1 - B)\ (1 - B^7)\ Z_t = b_1(1 - B)\ (1 - B^7)\ X_{1,t} + b_2(1 - B)\ (1 - B^7)\ X_{2,t}$$

$$\Rightarrow \quad (1 - B - B^7 + B^8)\, Z_t = b_1(1 - B - B^7 + B^8)\, X_{1,t} + b_2(1 - B - B^7 + B^8)\, X_{2,t}$$

$$\Rightarrow \quad Z_t = Z_{t-1} + Z_{t-7} - Z_{t-8} + (1 - B - B^7 + B^8)\, X_{1,t} + b_2(1 - B - B^7 + B^8)\, X_{2,t}.$$

The function (4.2.17) can be generalized for m imput factors where $m > 2$.

Let the estimate of Z_t by (4.2.9) be

$\hat{z}_{1,t} = z_{t-1} + b_1(1 - B)\, X_{1,t}$ and estimate of Z_t by (4.2.16) be

$\hat{z}_{2,t} = z_{t-7} + b_2(1 - B^7)\, X_{2,t}$

then the series E_t can be obtained from

$$E_t = z_t - \hat{z}_{1,t} - \hat{z}_{2,t} \qquad (4.2.18)$$

and then a stochastic model can be identified for series E_t. Let the stochastic model identified for E_t be of the form

$$E_t = h(B)\, e_t, \qquad (4.2.19)$$

then combining (4.2.17) and (4.2.19) we have

$$Z_t = (1 - B)^{-1} b_1(1 - B) X_{1,t} + (1 - B^7)^{-1} b_2(1 - B^7) X_{2,t}$$

$$+ h(B) e_t, \qquad (4.2.20)$$

which is known as the combined transfer function-noise model. Once the model has been identified in this form, the parameters have to be estimated together in order that the non-significant parameters can be detected and rejected. An algorithm that can produce efficient estimates of the parameters will be discussed in the next section.

4.3 Dynamic System Model - Estimation

In this section efficient estimation of the parameters of the combined transfer function-noise models identified in section 4.2 will be made. The estimates can be obtained easily by the use of the Marquardt algorithm for nonlinear least squares. The theoretical basis of the algorithm is stated by Marquardt [36] and the practical application of it is briefly indicated in Box and Jenkins [6]. The algorithm supplies the parameter estimates that minimize the residual sum of squares, the covariances of the parameters, and the variance of the residuals generated by the optimal parameter values. The algorithm, described in detail in Appendix E, causes rapid convergence (in most cases within ten iterations, provided a minimum exists) to the optimal values even if

zeros are given as initial values of the parameters. As for the other algorithms in this work, a Fortran IV program (see program 5 in Owolabi [52]) is written to implement this algorithm. The rest of this section consists of the results obtained from using this algorithm for the hourly and daily models.

NO_x Hourly Average

The combined transfer function-noise model identified for the hourly series in section 4.2, see (4.2.12), is

$$(1 - B)\, Z_t = b(1 - B)\, X_{t-3} + E_t \ .$$

Considering E_t as an ARMA (0,3) model, the above can be written as

$$(1 - B)\, Z_t = b(1 - B)\, X_{t-3} + e_t - h_1 e_{t-1} - h_2 e_{t-2} - h_3 e_{t-3}, \qquad (4.3.1)$$

where h_1, h_2, h_3 are the moving average parameters for the stochastic model of the E_t series, and b is known from section 4.2 as the only transfer function parameter. From this model (4.3.1) the residual series, e_t, t = 1,2,....., n can be generated as

$$e_t = (1 - B)\ Z_t - b(1 - B)\ X_{t-3} + h_1 e_{t-1} + h_2 e_{t-2} + h_3 e_{t-3}$$

for different sets of values of b, h_1, h_2, h_3, in order for the algorithm to determine the optimum set of parameter values . As in Chapter III, the initial, unknown e_t's are set to zero. Since b can assume any real value as stated in section 4.1 and h_1, h_2, h_3, being moving average model parameters, have to obey the invertibility conditions discussed in section 3.2, b can quite arbitrarily be set to 1.0 and h_1, h_2, h_3 set to 0.0 as the initial parameter values in the Marquardt algorithm (i.e. the program thereof). With these initial values, the following set of optimal parameter values is determined:

$\tilde{b}$ is 2.501 with standard deviation of 1.032,
$\tilde{h}_1$ is 0.088 with standard deviation of 0.055,
$\tilde{h}_2$ is 0.050 with standard deviation of 0.055,
$\tilde{h}_3$ is 0.125 with standard deviation of 0.055,

where the corresponding variance of the residuals is 0.405.

If the optimal estimates of the parameters obtained are compared with twice their standard deviations, it can

be seen that only b and h_3 are significantly different from zero. Therefore the combined transfer function - noise model relating hourly oxides of nitrogen to temperature can be parsimoniously expressed as

$$(1 - B)\ Z_t = 2.501(1 - B)\ X_{t-3} + e_t - 0.125e_{t-3}. \quad (4.3.2)$$

Daily Maximum NO_x

Most of the comments made above in connection with the hourly model estimation apply to the daily model and so they will not be repeated. The identification procedure gives temperature as the only leading indicator; hence, the combined transfer function - noise model identified for the daily series can be written from (4.2.12) and (4.2.15) as

$$(1 - B)\ Z_t = b(1 - B)\ X_t + e_t - h_1e_{t-1} - h_2e_{t-2}. \quad (4.3.3)$$

From (4.3.3) the residual series e_t can be generated according to

$$e_t = (1 - B)\ Z_t - b(1 - B)\ X_t + h_1e_{t-1} + h_2e_{t-2}.$$

The initial values of the parameters supplied to the program of the Marquardt algorithm are arbitrarily chosen as $b = 1.0$, $h_1 = h_2 = 0.0$. The program produces the

following optimal estimates of the parameters:

$\tilde{b}$ is 0.971 with standard deviation of 0.344,
$\tilde{h}_1$ is 0.491 with standard deviation of 0.052,
$\tilde{h}_2$ is 0.268 with standard deviation of 0.052,

where the variance of the residuals is 0.742.

All three parameters are significantly different from zero when compared with twice their standard deviations. Hence, the combined transfer function - noise model relating daily maximum NO_x to daily maximum temperature is given by

$$(1 - B)\ Z_t = 0.971\ (1 - B)\ X_t + e_t - 0.491e_{t-1} - 0.268e_{t-2}. \quad (4.3.4)$$

Before the two models established here are accepted as the final dynamic system models suitable for representing the behavior of oxides of nitrogen in the urban atmosphere, they must be submitted to careful diagnostic checking. The diagnostic checks are applied in the next section.

4.4 Dynamic System - Diagnostic Checks

Comparing the residual variances obtained in the last section with the residual variances obtained through the stochastic models fitted in Chapter III and shown in

Table 4.8, it can be seen that the dynamic models are as good as the stochastic models if not better, because their residual variances are smaller than residual variances of the corresponding stochastic models.

TABLE 4.8 Variance Comparison of Stochastic and Dynamic System Models

NO_x Series	Variance of Transformed Data	Stochastic Model's Residual Variance	Dynamic System Model's Residual Variance
Hourly Averages	1.202	0.522	0.405
Daily Maxima	1.577	0.772	0.742

The first test therefore concerns the variances of the residuals which are found to be smaller than the variances of the transformed data shown in Table 4.8. Since the variances of the residuals are very small it can be concluded that the dynamic models are adequate.

One other test may be applied for the diagnostic check of the dynamic system models. This test uses the

Chi - square test as discussed in section 3.5 where the Chi - square random variable Q is a function of the autocorrelation function of the residuals according to (3.4.1). The observed Chi - square value Q of the residual autocorrelations corresponding to (4.3.2) is 73.25. The theoretical upper critical Chi - square value at 95% for 81 degrees of freedom is 103.01. Thus we accept the hypothesis that the residuals behave in a white noise fashion, implying that they are uncorrelated. For the daily maximum NO_x, the observed Chi - square value is $Q = 65.26$ while the theoretical upper critical Chi - square value at 95% for 88 degrees of freedom is 110.90. This also indicates that the residuals of (4.3.3) seem to behave like white noise. These tests show that the combined transfer function - noise models for both the hourly averages and daily maxima of NO_x as given by (4.3.2) and (4.3.3), respectively explain the NO_x behavior well. Thus the models can be used for forecasting the oxides of nitrogen concentration in the urban atmosphere.

4.5 Forecasting with the Dynamic System Models

NO_x Hourly Average

To forecast $z_t(L)$ (i.e. transformed hourly average of NO_x) at L hours ahead of a time origin t, the forecast function corresponding to the hourly model (4.3.2) is

$$\hat{z}_t(L) = z_{t+L-1} + 2.501\ (x_{t+L-3} - x_{t+L-4}) - 0.125\ e_{t+L-3} \ . \qquad (4.5.1)$$

The terms on the right hand side of (4.5.1) are given as:

$$z_{t+j} = \begin{cases} z_{t+j} & j \leq 0 \\ \hat{z}_t(j) & j > 0 \ , \end{cases}$$

$$x_{t+j} = \begin{cases} x_{t+j} & j \leq 0 \\ \hat{x}_t(j) & j > 0 \ , \end{cases} \qquad (4.5.2)$$

$$e_{t+j} = \begin{cases} e_{t+j} & j \leq 0 \\ 0 & j > 0 \ . \end{cases}$$

$\hat{x}_t(j)$ is calculated from model (4.2.4) fitted to the input series as:

$$\hat{x}_t(j) = 1.338\ x_{t+j-1} - 0.127\ x_{t+j-2} - 0.211\ x_{t+j-3} \ ,$$

and e_t is calculated from $e_t = z_t - \hat{z}_{t-1}(1)$.

Using (4.5.1) the lead - one and leads 1 - 30 forecasts of the last 30 hours for the hourly series are obtained. Figure 4.10 shows the plot of these forecasts with their 95% confidence limits and also the corresponding observed series. From Figure 4.10 it can be seen that all the observed values fall within 95% confidence limits of their lead - one forecasts which is $\hat{z}_t(1) \pm 1.247$ in the

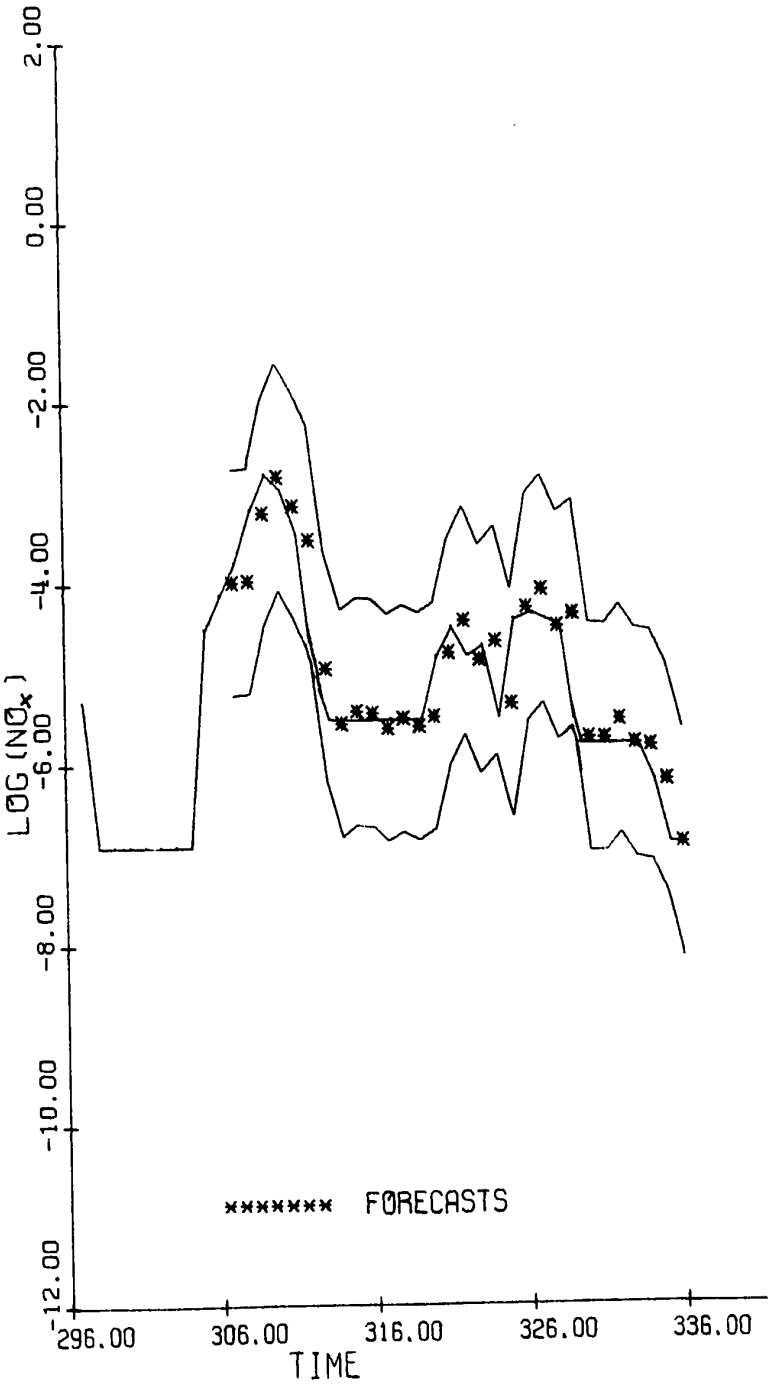

Fig. 4.10 Transformed hourly NO_x and Lead - one Forecasts using Dynamic System Model

logarithmic units. For the purpose of comparison with the raw data, the forecasts are transformed back to the raw data units by taking the antilog and subtracting 0.001. This transformed forecasts is presented in Table 4.9 while the corresponding observed raw data is presented in Table 4.10. It can be seen from the two tables that most of the lead - one forecasts are close to the observed values.

The leads 1 to 30 forecasts with their 95% confidence limits, and also the corresponding observed series are presented in Figure 4.11 in logarithmic units. Although none of the observed values fall outside the confidence limits, the confidence limit itself is so wide that none could have fallen outside it. However this model can be used to forecast hourly NO_x three hours ahead. This is an improvement over the stochastic model for the hourly series which can forecast reliably only one hour ahead. The forecasts here also are transformed back to the raw data unit and presented in Table 4.11.

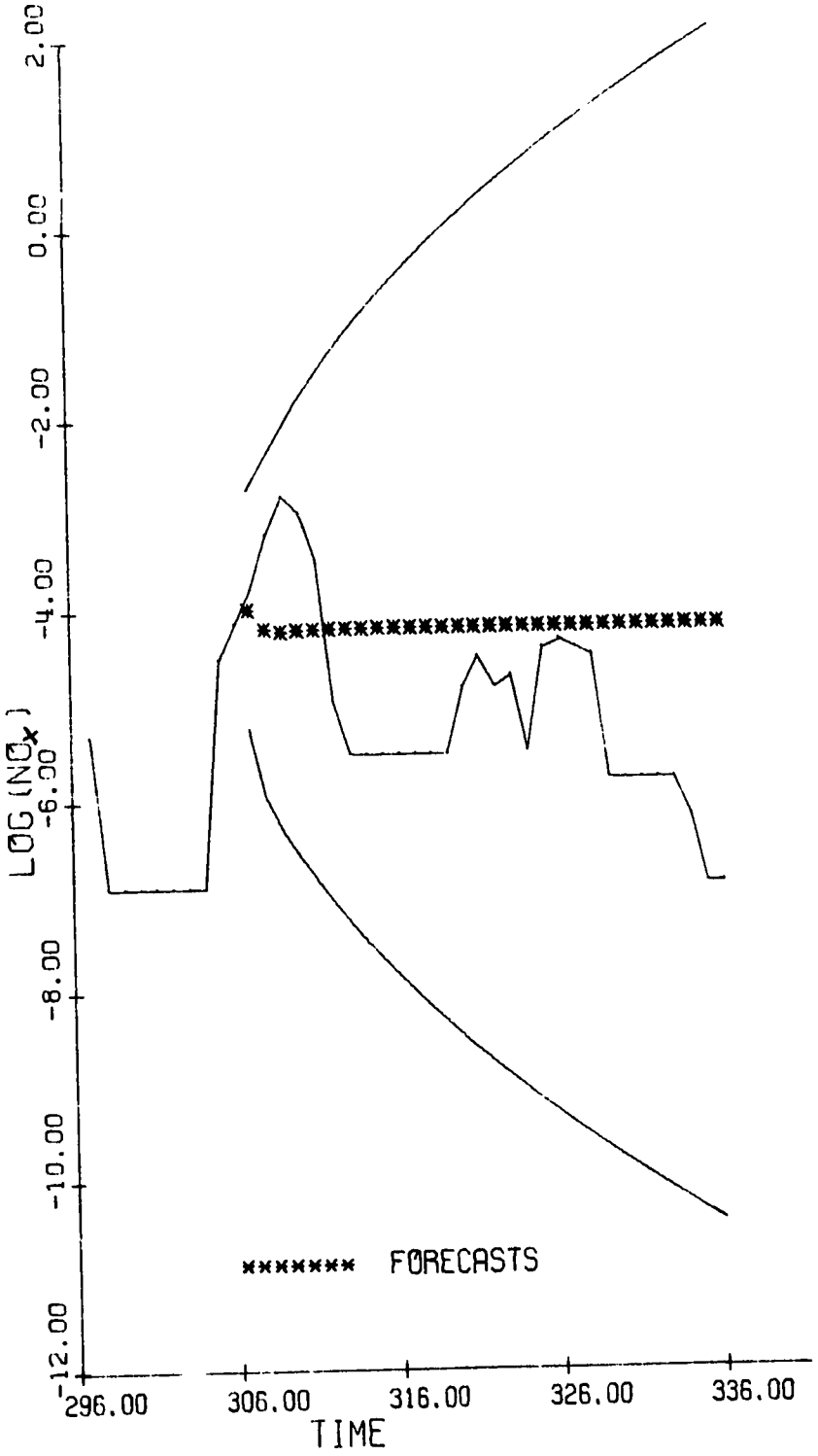

Fig. 4.11 Transformed hourly NO_x and Leads 1 to 30 Forecasts Using Dynamic System Model

TABLE 4.9 Lead - One Forecasts of the last thirty hours of NO_x Hourly Averages Series using Dynamic System Model *

0.017	0.018	0.039	0.058	0.042	0.028
0.006	0.003	0.003	0.003	0.003	0.003
0.003	0.003	0.007	0.011	0.007	0.008
0.004	0.013	0.016	0.010	0.012	0.002
0.002	0.003	0.002	0.002	0.001	0.000

* Read across the page. The unit of measurement is ppm.

TABLE 4.10 The last thirty hours of NO_x Hourly Averages Series *

0.022	0.040	0.060	0.050	0.030	0.006
0.003	0.003	0.003	0.003	0.003	0.003
0.003	0.007	0.010	0.007	0.008	0.003
0.011	0.012	0.011	0.002	0.002	0.002
0.002	0.002	0.002	0.001	0.000	0.000

* Read across the page. The unit of measurement is ppm.

TABLE 4.11 Leads 1 to 30 Forecasts of the last thirty hours of NO_x Hourly Averages Series using Dynamic System Model *

0.017	0.014	0.014	0.014	0.014	0.014
0.014	0.014	0.014	0.014	0.014	0.014
0.014	0.014	0.014	0.014	0.014	0.014
0.014	0.014	0.014	0.014	0.014	0.014
0.014	0.014	0.014	0.014	0.014	0.014

* Read across the page. The unit of measurement is ppm.

Daily Maximum NO_x

Using the model (4.3.3) for the daily maxima of NO_x, the forecast function for the daily maximum series is given by

$$\hat{z}_t(L) = z_{t+L-1} + 0.971\,(x_{t+L} - x_{t+L-1}) - 0.491\, e_{t-1} - 0.268\, e_{t-2}\,, \qquad (4.5.3)$$

where (4.5.2) can be used to obtain z_{t+j}, x_{t+j}, and e_{t+j} while $\hat{x}_t(j)$ is calculated from model (4.2.8), fitted to the input series as:

$$\hat{x}_t(j) = 1.19\, x_{t+j-1} - 0.373\, x_{t+j-2} + 0.183\, x_{t+L-3} \cdot$$

Function (4.5.3) is used to obtain lead - one and leads 1 to 30 forecasts of the last 30 days of the daily maximum series. Figure 4.12 shows the plot of these forecasts with their 95% confidence limits and also the corresponding observed series. From Figure 4.12 it can be seen that all the observed values fall within 95% confidence limits of their lead - one units. For the purpose of comparison with the raw data observed, the forecasts are transformed back to the raw data units by taking the antilog and subtracting 0.001. This transformed forecast is presented in Table 4.12 while the corresponding observed raw data are presented in Table 4.13. It can be seen from the two tables that most of the lead - one forecasts are close to the observed values.

The leads 1 to 30 forecasts with their 95% confidence limits, and also the corresponding observed series are presented in Figure 4.13 in logarithmic units. As shown in Figure 4.13 all the observations fall within the 95% confidence limits of their corresponding forecasts. The forecast function here, as in the corresponding stochastic model forecast function, can predict reliably two days ahead. However the forecast values here are closer to the raw observations than the forecast values

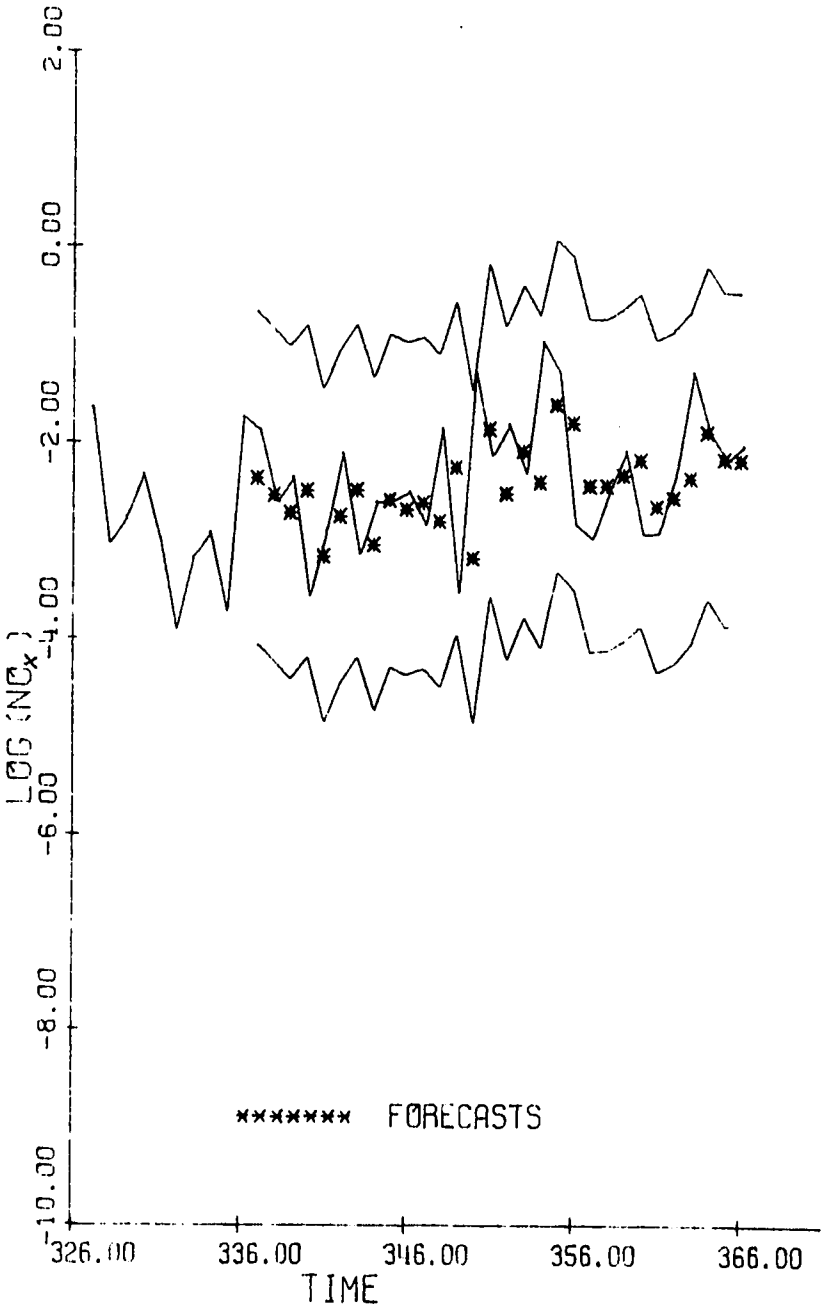

Fig. 4.12 Transformed daily NO_x and Lead - one Forecasts using Dynamic System Model

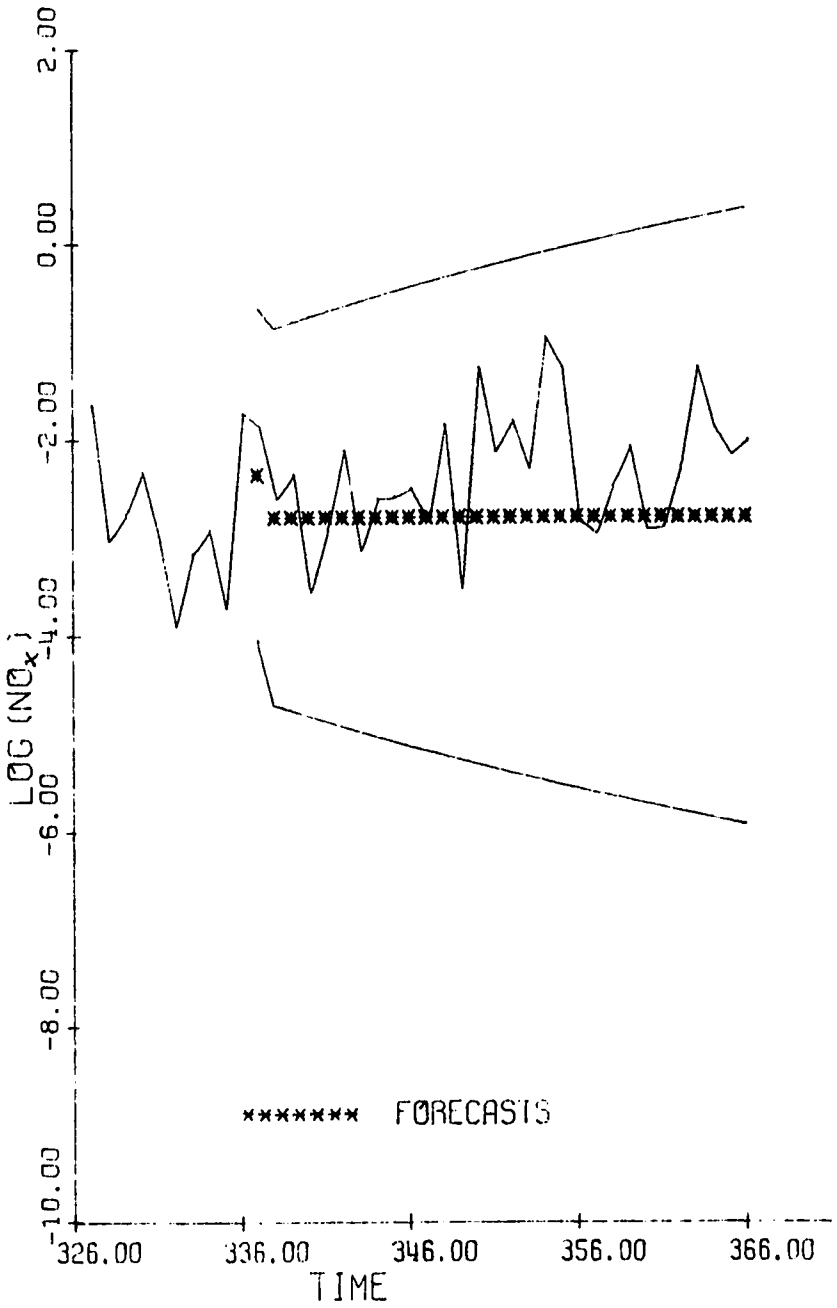

Fig. 4.13 Transformed daily NO_x and Leads 1 to 30 Forecasts using Dynamic System Model

TABLE 4.12 Lead - one Forecasts of the last thirty days of Daily Maxima of NO_x Series using Dynamic System Models *

0.094	0.079	0.065	0.082	0.042	0.063
0.083	0.047	0.075	0.068	0.073	0.060
0.105	0.043	0.156	0.080	0.125	0.090
0.201	0.167	0.087	0.088	0.097	0.114
0.071	0.077	0.095	0.152	0.115	0.114

* Read across the page. Unit of measurement is ppm.

TABLE 4.13 The last thirty days of Daily Maxima of NO_x Series *

0.153	0.072	0.094	0.027	0.053	0.121
0.042	0.073	0.074	0.081	0.057	0.156
0.028	0.281	0.117	0.163	0.098	0.381
0.279	0.058	0.050	0.083	0.124	0.052
0.054	0.100	0.281	0.151	0.113	0.132

* Read across the page. The unit of measurement is ppm.

TABLE 4.14 Lead Times 1 to 30 Forecasts of last thirty days of Daily Maxima of NO_x Series using Dynamic System Model *

0.094	0.060	0.060	0.060	0.060	0.060
0.060	0.060	0.060	0.060	0.060	0.060
0.060	0.060	0.060	0.060	0.060	0.060
0.060	0.060	0.060	0.060	0.060	0.060
0.060	0.060	0.060	0.060	0.060	0.060

* Read across the page. The unit of measurement is ppm.

There are no dynamic system models developed for Calgary, Windsor, Sarnia, Toronto, and Sudbury because the necessary additional data for wind, traffic, and temperature are not available. However, daily maximum series of oxides of nitrogen are available for these five cities and are used to develop stochastic models for oxides of nitrogen in those cities in Chapter V.

CHAPTER V

COMPARATIVE STUDY OF NO_x IN SOME OTHER CITIES

5.1 Stochastic Models for Five other Cities

In this chapter the daily maximum NO_x stochastic model as developed for Edmonton in Chapter III is compared with the same type of model established here for Calgary, Sarnia, Sudbury, Toronto, and Windsor. As will be apparent shortly the models for the six cities are all of the ARIMA (0, 1, 2) type, and hence an attempt is made to establish a general ARIMA (0, 1, 2) model which may be able to explain the behavior of the daily maximum of NO_x in any urban center. It should be noted that the stochastic model established for Edmonton's daily maximum NO_x series is also an ARIMA (0, 1, 2), but written (for the purpose of analysis)in its modified ARMA (0,2) form, because the model is expressed as

$$w_t = e_t - b_1 e_{t-1} - b_2 e_{t-2} ,$$

where $w_t = (1 - B) z_t = z_t - z_{t-1}$.

The ARIMA form will be used throughout this chapter because it shows the degree of differencing.

To save space,data employed for the development of the stochastic model for the five cities are not included in

this report. The interested reader can obtain them from the sources mentioned in Chapter II.

The various series of the original daily NO_x data are all transformed by first adding 0.001 ppm and then taking the natural logarithm, thus resulting in the transformed series z_t's. The same model building procedure as outlined in Chapter III is used for Calgary and each of the four Ontario cities. The final stochastic model and relevant parameter estimates are described briefly below.

Calgary NO_x

The daily maximum NO_x data available for Calgary consists of 730 observations (for 1970, 1971) which is quite an adequate series for stochastic modelling. The model is found to be ARIMA (0, 1, 2) and this is of the form

$$(1 - B)\ Z_t = (1 - \tilde{b}_1 B - \tilde{b}_2 B^2)\ e_t\ , \qquad (5.1)$$

where

$\tilde{b}$ = 0.564 with standard deviation of 0.031,

$\tilde{b}$ = 0.214 with standard deviation of 0.036,

with corresponding residual variance of 0.294. The variance estimate of the transformed series is 0.492.

Sarnia NO_x

The 1971 daily maximum NO_x data from Sarnia consist of 365 observations. The model that describes the process from which this series is observed is found to be an ARIMA (0, 1, 2) of the form (5.1), where

$\tilde{b}_1 = 0.563$ with standard deviation of 0.043,
$\tilde{b}_2 = 0.326$ with standard deviation of 0.050,

and the residual variance is estimated to be 0.204. The variance estimate of the transformed series is 0.329.

Sudbury NO_x

The daily maximum NO_x series from Sudbury consist of observations for only seven months, namely January to July 1971. The 212 consecutive observations again indicate an ARIMA (0, 1, 2) model as the best fitting one, where

$\tilde{b}_1 - 0.566$ with standard deviation of 0.057.
$\tilde{b}_2 = 0.414$ with standard deviation of 0.063,

and the residual variance is estimated to be 0.031. The variance estimate of the transformed series is 0.360.

Toronto NO_x

The 1971 daily maximum NO_x data available from Toronto contained many "holes" which were filled in according to the method proposed in section 2.1. After 'patching' the data the usual stochastic model building procedure is employed for the resulting 365 observations. An ARIMA (0, 1, 2) model is found to fit this Toronto series, where

$\tilde{b}_1 = 0.600$ with standard deviation of 0.042,
$\tilde{b}_2 = 0.280$ with standard deviation of 0.050,

and residual variance is estimated to be 0.296. The variance estimate of the transformed series is 0.392.

Windsor NO_x

The daily maximum NO_x data in Windsor is a series of consecutive observations for 8 months; namely January to August, 1971. An ARIMA (0, 1, 2) model is obtained for those 243 observations also where

$\tilde{b}_1 = 0.447$ with standard deviation of 0.057,
$\tilde{b}_2 = 0.323$ with standard deviation of 0.061,

and residual variance is estimated to be 0.224. The variance estimate of the transformed series is 0.342.

5.2 Result

In all the six cities considered in this study the stochastic process generating daily maximum NO_x is found to be of the ARIMA (0, 1, 2) type, that is

$$Z_t = Z_{t-1} + e_t - b_1 e_{t-1} - b_2 e_{t-1} \quad . \qquad (5.2)$$

The parameter estimates, $\tilde{b}_1$, $\tilde{b}_2$, for the different cities differ only slightly. This does not imply that the NO_x level in the different cities are the same. As a matter of fact the estimated mean of the transformed series for the different cities vary. If the models were fitted on the transformed series without differencing, the different means would have shown up different constant terms. However, since the ARIMA model is fitted on the first difference of the transformed series (for stationarity reasons) the mean of the first difference is zero as explained in Section 3.2, thus making the constant term in the model zero. Table 5.1 summarizing the comparison of the six models, shows that Toronto has a higher NO_x pollution average than any of the other cities and Edmonton NO_x average is the lowest, although its NO_x variation is more than three times as large as that of the other cities. However, it must be remembered that the series for the different cities do not belong to the same year (although all are within the 1970 - 72 period) nor do they have equal length as shown in section 5.1. Although the Sudbury

TABLE 5.1 Comparison of the stochastic models for daily maxima of NO_x

City	Mean	Variance	ARIMA(0,1,2) Parameters b_1	b_2	Residual	Variance Reduction	Population
Calgary	-2.56 (0.076)	0.492	0.564	0.214	0.294	40%	403,319
Edmonton	-4.111 (0.016)	1.577	0.500	0.260	0.772	51%	438,152
Sarnia	-3.22 (0.039)	0.329	0.563	0.326	0.204	38%	57,644
Sudbury	-3.47 (0.303)	0.360	0.566	0.414	0.301	16%	90,535
Toronto	-2.24 (0.105)	0.392	0.600	0.280	0.296	24%	712,786
Windsor	-2.44 (0.086)	0.342	0.447	0.323	0.224	34%	203,300

model explains only 16% of the observed variation it is the best fitting model for this series. Since the analysis has been performed on the logarithmic transformation of the series as indicated in Chapter III, the means and variances in Table 5.1 are in their natural logarithms and their antilogarithms have to be taken to get their values in ppm - the averages in ppm are indicated in parenthesis. For instance the average NO_x for Edmonton is 0.016 ppm. Also in Table 5.1 population figures of the cities according to the 1971 census as recorded in the 1972 issue of the Canada Yearbook are included.

The fact that the best fitting model for each city turns out to be the same type with only slight differences in their parameters is extremely surprising. Especially, since NO_x is produced primarily by motor vehicles and industry, it would stand to reason that in large industrial urban areas the behavior of daily NO_x may be different from its behavior in less densely populated urban areas. However, this study supplies evidence to the contrary. Since the ARIMA (0, 1, 2) models are so similar, one suspects that perhaps a common ARIMA model exists for all urban centers. This common model should then be the ARMA (0,2) type for the first difference of the transformed data, that is of the ARIMA (0, 1, 2) type. To try to establish such a common ARIMA (0, 1, 2) model, the first difference of the six NO_x series are pooled (that is

concatenated) together such that a first difference series, 2274 observations long, is obtained. For the pooling, variation over space is considered rather than variation over time. Thus the $(1 - B)\ z_t$ series for the different cities are regarded as samples, observed in the different cities, of the overall process $(1 - B)\ Z_t$. The ARIMA(0, 1, 2) model fitted to this overall series turns out to be

$$(1 - B)\ Z_t = (1 - 0.500B - 0.270B^2)\ e_t\ , \qquad (5.3)$$

with corresponding residual variance of 0.347. To ensure that this combined model adequately describes the NO_x behavior in the individual cities, it must be compared to the six individual models in Table 5.1. The test employed here is similar to the 'constancy' test explained by Huang [27], and the 'coincidence' test used for regressions and described by Williams [78]. This test for regression can be applied to the ARIMA(0, 1, 2) models because the first difference of the transformed z_t's are regressed on the e_t's which are white noise and hence independent. Thus this model is like a regression model which relates a dependent variable to two independent variables. The test makes use of the analysis of variance as follows.

Let there be m samples having $n_1, n_2, \ldots\ldots, n_m$

elements respectively. Let $n = n_1 + n_2 + \ldots\ldots + n_m$, and let there be k estimated parameters of the model. If SS_1 is the total residual sum of squares due to the overall model obtained by pooling all the samples, and SS_2 is the sum of the sums of squares within each sample, then $SS_3 = SS_1 - SS_2$ is the sum of sums of squares between each regression plane and the overall regression plane. It can be shown that

$\frac{SS_1}{\sigma_e}$ has a Chi - square distribution with n - k degrees of freedom,

$\frac{SS_2}{\sigma_e}$ has a Chi - square distribution with n - km degrees of freedom,

$\frac{SS_3}{\sigma_e}$ has a Chi - square with km - k degrees of freedom, and that

$$\frac{SS_3/(km - k)}{SS_2/(n - km)} \quad \text{has an F distribution with } (km - k) \text{ and } (n - km) \text{ degrees of freedom.} \tag{5.4}$$

Let $b_1, b_2, \ldots\ldots, b_k$ be the parameters of the overall model, and $b_{11}, b_{12}, \ldots\ldots, b_{1k}$; b_{21} ,

b_{22} , ..., b_{2k} ; .....; b_{m1} , b_{m2} ; be the parameters for each of the m individual models, respectively, then the hypothesis to be tested is that

$$
\begin{aligned}
&b_{11} = b_{21} = b_{31} = \ldots\ldots = b_{m1} = b_1 \\
&b_{12} = b_{22} = b_{32} = \ldots\ldots = b_{m2} = b_2 \\
&\quad .\quad .\quad .\qquad\qquad .\quad . \\
&\quad .\quad .\quad .\qquad\qquad .\quad . \\
&\quad .\quad .\quad .\qquad\qquad .\quad . \\
&b_{1k} = b_{2k} = b_{3k} = \ldots\ldots = b_{mk} = b_k \quad .
\end{aligned}
\tag{5.5}
$$

If the hypothesis is true (i.e. the overall regression hyperplane and the individual regression hyperplanes are coincident) then $SS_1 = SS_2$, and SS_3 is zero. The test of the hypothesis (5.5), therefore, can be considered as a test of significance of SS_3 . Thus under hypothesis (5.5), the observed ratio (5.4) can be compared to the upper 5% critical point of the F distribution with the appropriate degrees of freedom. If the ratio is less than the upper 5% critical point, then SS_3 is not significantly different from zero.

Assuming (5.5) for the overall and the six individual ARIMA(0, 1, 2) models respectively, the observed F - value (5.4) is 0.609 for 10 and 2263 degrees of freedom. However, the upper 5% critical point of F with 10 and 2263 degrees of freedom is 1.83, hence the hypothesis (5.5) is accepted.

It can be concluded, therefore, that the stochastic process generating daily maximum NO_x in these urban centers can be described by the ARIMA(0, 1, 2) model,

$$Z_t = Z_{t-1} + e_t - 0.500e_{t-1} - 0.270e_{t-2} . \qquad (5.6)$$

From the above evidence it seems reasonable to conclude further, that the process generating daily maximum NO_x in any urban center may be described by the same ARIMA(0, 1, 2) model.

To test how good this general model will be at forecasting, it is used to obtain lead - one forecasts of daily maximum NO_x for January 1970 in Edmonton, (where January 1970 in Edmonton is outside the period used to develop the model). Table 5.2 and Table 5.3 which contain the forecasts and the observed data respectively, indicate that the forecasts here are as good as those obtained for the corresponding model in Chapter III.

TABLE 5.2 Lead - one Forecasts of the Daily Maxima of NO_x for January 1970 in Edmonton *

0.031	0.028	0.031	0.023	0.022	0.032
0.050	0.034	0.059	0.040	0.028	0.027
0.025	0.027	0.034	0.028	0.063	0.027
0.036	0.040	0.069	0.132	0.124	0.134
0.191	0.165	0.129	0.195	0.161	0.134
0.112					

* Read across the page. The unit of measurement is ppm.

TABLE 5.3 Observed Daily Maxima of NO_x for January 1970 in Edmonton *

0.025	0.031	0.018	0.016	0.039	0.106
0.043	0.096	0.047	0.017	0.016	0.018
0.023	0.042	0.029	0.128	0.026	0.030
0.047	0.136	0.496	0.336	0.242	0.390
0.254	0.118	0.246	0.189	0.109	0.076
0.047					

* Read across the page. The unit of measurement is ppm.

Since the general model is conjectured to be adequate for describing the daily behavior of NO_x in any urban center it can be used to forecast daily maximum NO_x concentration in any city. Further discussion and recommendations for future work are given in the next chapter.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The models established for NO_x in this study generate some surprising information concerning the behavior of the pollutant in urban atmosphere. One such fact revealed by the stochastic models is that the concentration of NO_x in urban centers follows no particular pattern. For example, one would expect a 7 - day cycle in the daily behavior, such that the minimum concentration occurs during the week-end when there is less traffic flow and industrial activities. The fact that all of the particular stochastic models fitting the data from the individual cities, and the general stochastic model (5.6) for urban centers, contain no seasonal term to reflect a 7 - day cycle shows that no such cycle exists. Other information produced by stochastic models is that NO_x behaves in the same way in all urban centers. As a result it was possible to establish a general stochastic model (5.6) to explain the behavior of NO_x in urban atmosphere. The general stochastic model established is ARIMA(0,1,2) based on the first difference of the transformed data; hence, the difference in the level of concentration among the cities does not affect the parameters of the model.

The dynamic system models (combined transfer function-

noise models) show that among the three factors which were conjectured to be capable of influencing the behavior of NO_x in urban atmosphere, only temperature had such influence. The three input factors considered were temperature, wind speed, and traffic flow. Traffic flow was considered because it is known that automobile exhaust is a major source of NO_x in a city. Wind speed was considered because atmospheric mixing and transport due to wind could have an effect on the concentration of air pollutants. Temperature was considered since temperature affects the amount of energy to be generated for heating, and hence, the amount of NO_x emitted by power plants. The transfer function connecting temperature to hourly behavior of NO_x in the Edmonton atmosphere shows that an incremental change in temperature has an effect on NO_x concentration after two hours. For the daily behavior, an incremental change in the daily temperature is reflected in the NO_x maximum concentration that same day. Temperature observations in the Industrial Airport are known to be fairly representative of temperature in the city (See Hage and Longley [21]). However, the wind speed observed in the same airport and the traffic flow observed on Jasper avenue cannot represent the situation in the whole city, and, thus, cannot represent the situation in Administration Building, where NO_x is being monitored. The remoteness of the observation stations of wind speed and traffic flow from where NO_x is being monitored may account for the apparent lack of relationship between them and NO_x. It is especially surprising that no relationship is found

to exist between NO_x and traffic flow; as a result the dynamic system models do not involve traffic flow.

Comparing the model types developed in this study with diffusion models it can be seen that stochastic and dynamic system models have some advantages over diffusion models. The models give forecasts which are close to the observed values, and unlike the reported characteristic of diffusion models, the models developed here produce air pollutant concentration estimates which are neither consistently smaller nor consistently greater than the observations. In addition the confidence intervals of the forecasts generated by the stochastic and the dynamic system models could be estimated and used as a measure of the reliability of the forecasts. Thus it was found that the stochastic model for daily maxima of NO_x would produce reliable forecasts for one and two days ahead. The dynamic system model (4.3.2) for hourly NO_x would produce reliable forecasts for one, two,and three days ahead, while the dynamic system model (4.3.4) for daily maxima of NO_x would produce reliable forecasts for one and two days ahead. Although longer lead forecasts by these models are tolerable as some idea of future occurrence, their 95% confidence intervals are too wide; as a result the fact that all the observations fall within the 95% confidence intervals does not, for practical purposes, prove the reliability of longer lead forecasts. The only obvious advantage of the diffusion model over the model types proposed in this study is that the

diffusion models can supply information levels of a pollutant where no monitoring station exists, since the development of the diffusion model does not require the observed data of the pollutant at the receptor. However,the development of diffusion models require other data, like source inventories and meteorological observations. Notwithstanding, if necessary observations are available for the development of stochastic and dynamic system models, they are better as predictive models for air pollutants than diffusion models.

Comparing stochastic models with dynamic system models one would find that the dynamic system models are preferable. Dynamic system models can give information about the relationship connecting the pollutant with the influencing factor as illustrated in this study in the case of NO_x and temperature. Also the lack of relationship between the conjectured input factor and the output is detectable at the model development stage as exemplified in the case of NO_x and the two input factors, traffic flow, wind speed. In addition, dynamic system models generate better forecasts in the sense that the estimates they produce are closer to the observed values than those produced by stochastic models. Generally dynamic system models can be used as a basis for the control of the system. However, the possible dynamic system models established for NO_x in this study cannot be used for control because ambient temperature, which is the only input factor cannot be controlled. If traffic flow had an obvious connection with NO_x through a

transfer function, then a dynamic control function required to control the pollutant's level could have been obtained from the resultant combined transfer function-noise model. The control aspect, however, lies outside the scope of this thesis; further details about control functions are given in Box and Jenkins [6].

In general, the city of Edmonton still enjoys a clean atmosphere. The NO_x level is usually below the safe standard set in many places. The Alberta clean air act passed in January 1973 stipulated an annual mean of 0.03 ppm while the mean observed for daily maxima of NO_x over a year in this study is 0.016 ppm. Although some places close to power stations, refineries, and highways may experience higher concentrations, the concentration level for the city as measured at the Administration Building is generally below the safe level.

6.2 Recommendations

Precision and continuity of the measurements of concentration of air pollutants cannot be over-emphasized in any pollution surveillance program. Therefore, it is hereby recommended that monitoring stations obtain more reliable measuring instruments or that two instruments are alloted to each station, so that one may be used as a back-up device. This would prevent loss of observations which at present is common in monitoring NO_x in all the cities investigated

in this work. Wherever possible these instruments may be interfaced with computers for accurate data recording, analysis and immediate reporting.

For the City of Edmonton more monitoring stations are necessary for NO_x. A network of stations would give a better overall picture of the pollutants level in the city than only one station. In addition, input factors that could affect NO_x concentration should be monitored at or very near the NO_x monitoring stations.

To use the different forecast functions developed in Chapters III, IV, and V for the purpose of forecasting NO_x, the minimum number of past records required is the highest degree of backward shift operator B in the function (or in the model which leads to the forecast function); for example model (5.3) requires a minimum of 2 past records. However, 50 past records are recommended wherever possible. If the recursive estimation of values is started 50 points back, loss of information that may be caused by equating the initial error e_t, to zero,will be compensated by the time the required future estimate is calculated. It should be emphasized that the models here fit certain transformations of the data (see transformations in Table 4.1), hence the records used should first be appropriately transformed [$\ln (z_t + 0.001)$ in the case of NO_x] before applying the forecast function. The forecast obtained then is given as a logarithm and it should be transformed back to ppm by taking its antilogarithm and subtracting 0.001.

It may be desirable to forecast the mean concentration of a pollutant over a region including many cities (or a city including many monitoring stations). If reliable records of the pollutant exist for stations all over the region they can be used to calculate a series of means for the whole region. The series of regional means can be regarded as output while the set of local means will be inputs to the system. Then a combined transfer function - noise model connecting the inputs to the output can be developed, and this will be capable of generating forecasts of regional means. This cannot be developed for Edmonton at present because there is only one monitoring station. Also it is not tried for cities considered here because the series available are not of equal length and they do not cover the same period of time.

The time series analysis approach to pollution modelling is described generally in this study such that people interested in the field of air pollution could be aware of its potentialities. Although models in this study were developed for NO_x, it is obvious from the discussion that the same procedures, outlined in this study, could be followed to establish models that would be able to explain the behavior of any other measurable air pollutant like oxides of sulfur, oxides of carbon and particulate matter. In addition, further investigation employing dynamic system model approach could provide a powerful tool for air pollution control in urban centers.

BIBLIOGRAPHY

1. Agnew, G. 'Science and Technology in Automotive Air Pollution Research;' Research Laboratories General Motors Corporation, Research Publication GMR-743, February, 1968

2. Atkisson, A., Gaines, R.S. Development of Air Quality Standards, Charles E. Morril Publishing Company, Columbus, Ohio, 1969 .

3.* Barringer, A.R., Newbury, B.C. and Moffat, A.J. 'Surveillance of Air Pollution from Airborne and Space Platforms;' Barringer Research Limited, Ontario (after 1967).

4. Bartlett, M.S. Stochastic Processes, Cambridge University Press, Cambridge, 1955.

5.* Bowne, N.E. 'A Simulation Model for Air Pollution Over Conneticut;' Journal of Air Pollution Control Association, Vo. 19, No. 8, August, 1969.

6. Box, G.E.P., Jenkins, G.M. Time Series Analysis Forecasting and Control, Holden-Day, San Francisco, 1971.

7. Box, G.E.P., Jenkins, G.M. 'Some Statistical Aspects of Adaptive Optimization and Control,' J. Royal Stat. Soc., B24, pp 297, 1962.

8. Box, G.E.P., Cox, D.R., 'An analysis of transformations,' Jour. Royal Stat. Soc. B26, pp 211, 1964.

9. Carnaham, B., Luther, H.A., Wilkes, J.O. Applied Numerical Methods, John Wiley and Sons, New York, 1969.

10. Carpenter, S.B., Thomas, F.W. and Gartrell, F.E. 'Full-Scale Study of Plume Rise at Large Electric Generating Stations,' Tennessee Valley Authority, Division of Health and Safety, September 1968.

11.* Chatfield, C., Pepper, M.P.G., 'Time-Series Analysis An Example from Geophysical Data,' J. Royal Stat. Soc. (Series C) Applied Statistics, 20, 3, 1971.

12. Clarke, J.F. 'A Simple Diffusion Model for Calculating Point Concentrations from Multiple Sources,' Journal of the Air Pollution Control Association, Vol. 14, No. 9, November 1964.

13. Clarke, J.F. 'Forecasting Pollutant Concentrations' U.S. Weather Bureau Forecasters Training Course, Silver Spring, Maryland., 1967.

14.* Commins, B.T. and Waller, R.E. 'Observations from a Ten-Year Study of Pollution at a Site in the city of London,' Atmospheric Environment, Pergamon Press, Vol. 1, pp 49-68, 1967.

15.* Craddock, J.M. 'The Analysis of Meteorological Time Series for use in Forecasting,' Statistician, 15, 169 - 191 , 1965.

16. Danielson, J.A. (Editor) Air Pollution Manual U.S. Department of Health, Education, and Welfare, Cincinnati, Ohio, 1967.

17. Davidson, B. 'A Summary of the New York Urban Air Pollution Dynamics Research Program' Journal of the Air Pollution Control Association, Vol. 17, No. 3, 1967.

18.* Garnier, B.J. (Editor) Climatological Bulletin, Department of Geography, McGill University, No. 5, January 1969.

19. Gilpin, A. Control of Air Pollution, Butterworth & Co., London , 1963 .

20.* Glasson, W.A. and Tuesday, C.S. 'Inhibition of Atmospheric Photoxidation of Hydrocarbons by Nitric Oxide,' Environmental Science & Technology, Vol. 4, No. 1, pp 37 - 44, January 1970.

21. Hage, K.D. and Longley, R.W., 'Ventilation and Mixing in Alberta Cities,' Presented at the Banff Conference on Pollution, Banff, Alberta, May 1968.

22. Hald, A. Statistical Theory with Engineering Application, John Wiley, New York, 1952.

23. Hepple, P. (Editor) Pollution Prevention, Proceeding of the Institute of Petroleum Summer Meeting at Brighton , The Institute of Petroleum, 61 New Cavendish Street, London , 1968.

24. Hilst, G.R., Yocom, J. E. and Bowne, N.E. 'The Development of a Simulation Model for Air Pollution Over Connecticut,' Vol. 1 (Summary Report) 7233 - 7290, Connecticut Research Commission, October 1967,

25. Hilst, G.R., Badgley, F.I., Yocom, J.E., and Bowne, N.E. 'The Development of a Simulation Model for Air Pollution Over Connecticut,' Vol. 2, (Connecticut Air Pollution Simulation Model) 7233 - 7296, Connecticut October 1967.

26. Hilst, G.R., 'An Air Pollution Model of Connecticut,' Draft of a paper prepared for presentation at the IBM Scientific Symposium, October, 1967.

27. Huang, S.D. Regression and Econometric Methods, John Wiley, New York ,1964 .

28. Jenkins, G.M. and Watts, D.G. Spectral Analysis and its Applications, Holden-Day, San Francisco, U.S. , 1969.

29. Junge, Ch.E., Air Chemistry and Radioactivity, Academic Press, New York and London, 1963.

30.* Koogler, J.B., Sholtes, R.S., Danis, A.L. and Harding, C.I. A Multivariable Model for Atmospheric Dispersion Predictions, Journal of the Air Pollution Control Association, Vol. 17, No. 4, April 1967.

31. Lamb, R.G. 'An Air Pollution Model of Los Angeles,' M.S. Thesis, University of California, Los Angeles, 1968.

32. Lawrence, A.A. 'Nitrogen Oxides Emission Control,' Noyes Data Corporation, New Jersey 07656, U.S.A., 1972.

33. Leith, W. The Analysis of Air Pollutants, Ann Arbor - Humphrey Science Publishers, Ann Arbor, 1970.

34.* Lucas, D.H. The Atmospheric Pollution of Cities, Int. J. Air Pollution, Vol. 1, pp 71 - 86, Pergamon Press, 1958.

35. Magill, P.L., Holden, F.R., Sawyer, F.G. Air Pollution Handbook, McGraw-Hill, New York, 1956.

36. Marquardt, D.W. 'An algorithm for least squares estimation of non-linear paramerers,' Jour Soc. Ind. Appl. Math (SIAM), 11, 1963.

37.* Mancuso, T., and Mordell, J. 'Proposed Initial Studies of the Relationship of Community Air Pollution to Health,' Environmental Research, Vol. 2, No. 2, pp 102 - 133, April, 1969.

38.* Maurin, P.G., and Jonakin, J. 'Removing Sulfur Oxides from Stacks,' Chemical Engineering April, 1970.

39.* McGregor, J.R. 'The approximate distribution of the correlation between two stationary linear Markov series,' Biometrika, 49, 1962.

40.* McQuire, J.H. and Tamura, G.T. 'Smoke control in High-Rise Buildings,' Canadian Building Digest, Division of Building Research, N.R.C. CB0134, pp 134-1, 134-4, February 1971.

41.* Meade, P.J. and Pasquill, F. 'A Study of the Average Distribution of Pollution Around Staythorpe,' International Journal Air Pollution, Vol. 1, 1958.

42. Middleton, J.T. 'Air Quality as a Controlling Factor in Life Processes,' Biometeorology, Proceedings of the Twenty-Eighth Annual Biology Colloquium, 1967.

43.* Milford, S.N. McCoyd, G.C., Aronowitz, L., Scaulon, J.H., and Simon, C. 'Air Pollution Models of the New York/New Jersey/Connecticut Air Quality Region,' Grumman Research Department Report RE-387J, August 1970.

44. Miller, S., Lewicke, C., 'Air Pollution Over the States,' Environmental Science and Technology Vol. 6, No. 2, February 1972.

45. Miller, M.E. and Holzworth, G.C., 'An Atmospheric Diffusion Model for Metropolitan Areas,' Journal of the Air Pollution Control Association, Vol. 17, No. 1, January 1967.

46. Moses, H., Mathematical Urban Air Polution Models, Argonne National Laboratory, ANL/ES - RPY - 001, Meteorology, April 1969.

47.* Mosher, J.C., MacBeth, W.G., Leonard, M.J., Mullins, T.P., and Brunelle, M.F. 'The Distribution of Contaminants in the Los Angeles Basin Resulting from Atmospheric Reactions and Transport,' Presented at 62nd Annual Meeting Air Pollution Control Association, New York, 33 pgs., June 1969.

48.* Munn, R.E., Thomas, D.A., and Cole, A.F.W. 'A Study of Suspended Particulate and Iron Concentrations in Windsor, Canada,' Atmospheric Environment, Vol. 3, pp 1 - 10, Pergamon Press, 1969.

49.* Murgatroyd, R.J. 'The dispersion of pollutants in the free atmosphere by the large scale wind systems,' Phil. Trans. Royal Society, London, A., Vol. 265, pp 273 - 294, 1969.

50. OECD (Organization for Economic Co-operation and Development). 'Methods of Measuring Air Pollution,' Report of a working party, Paris, 1964.

51.* Oke, T.R. 'Towards a More Rational Understanding of the Urban Heat Island,' Extract from Climatological Bulletin, No. 5, 19 pages, January 1969.

52. Owolabi, E.A. Program for Time Series Models in Air Pollution Analysis, Technical Report TR73-7. Department of Computing Science, University of Alberta, 1973.

53. Panofsky, H.A. 'Air Pollution Meteorology,' American Scientist, 57, 2, 1969.

54. Panofsky, H.A. 'Meteorological Applications of Cross Spectrum Analysis,' Spectral Analysis of Time Series,(Edited by B. Harris), Wiley, New York, 1967.

55.* Panofsky, H.A. and Prasad, B. 'The Effect of Meteorological factors on Air Pollution an a Narrow Valley,' Journal of Meteorology, Vol. 6, pp 493 - 499, June 1967.

56. Pasquill, F. Atmospheric Diffusion, D. Van Nostrand, London, 1962.

57. Pooler, F. 'A Prediction Model of Mean Urban Pollution for use with Standard Wind Roses,' Journal of Air and Water Pollution, Vol. 4, Nos. 3 - 4, pp 199 - 211, Pergamon Press, 1961.

58. Pooler, F. 'A Tracer Study of Dispersion over a City,' Journal of the Air Pollution Association, Vol. 11, No. 12, December 1966.

59.* Roberts, J.J., Croke, E.J., and Kennedy, A.S. 'An Urban Atmospheric Dispersion Model, Chicago Air Pollution Systems Analysis Program,' Argonne National Laboratory, ANL - ES - CC - 005 , 1969.

60.* Rogers, L.H. 'Nitric Oxide and Nitrogen Dioxide in the Los Angeles Atmosphere,' Journal of the Air Pollution Control Association, 8, 124, 1958.

61. Rolston, J.J. A Study of Air Pollution Sources and their Significance in Edmonton, Alberta, Government of the Province of Alberta, Department of Public Health, Division of Sanitary Engeering, August 1964.

62. Rossano, A.T. Air Pollution Control Guide Book for Management, ESSC Environmental Science Service Division, E.R.A. Inc., 750 Summer Street, Stanford, Connecticut, 06901, 1969.

63. Ryan, J.F. 'A Model of Urban Air Pollution,' Masters Thesis (unpublished) Purdue University, August 1970 .

64. Saltzman, B.E. 'Colorimetric Microdetermination of Nitrogen Dioxide in the Atmosphere,' Anal. Chem. 26, 1954.

65. Shanks, S.G., et al 'A Working System for the Management of Emission Inventory Information,' Journal of the Air Pollution Control Association, Vol. 22, No. 2, February 1972.

66.* Shirvaikar, V.V., Kapoor, K., and Sharma, L.N. 'A Finite Plume Model Based on Wind Persistance for use in Environmental Dose Evaluation,' Atmospheric Environment, Vol. 3, pp 135 - 144, Pergamon Press, 1969.

67.* Slade, D.H. 'Modelling Air Pollution in the Washington, D.C., to Boston, Megalopolis,' Science, Vol. 157, pp 1304 - 1307, September 1967.

68. Slawson, P.R. 'Observations on Plume Rise from a Large Industrial Stack,' Research Report No. 1, U.S. Atomic Energy Commission Report NYO - 3685 - 7, May 1966.

69. Strauss, W. (Editor) Air Pollution Control, Part 1, Wiley - Interscience, New York, 1971 .

70. Thomas, M.K. 'Canadian Urban Climates, A brief literature survey,' Estract from Climatological Bulletin No. 5, January 1969 .

71.* TRC Service Corporation,'Highlights of a Study on Air Pollution in the Capitol Region,' December 1967.

72.* Turner, D. 'The Diurnal and Day to Day Variations of Fuel Usage for Space Heating in St. Louis, Missouri,' Atmospheric Environment, Vol. 2, pp 339 - 351, Pergamon Press,1968.

73.* Turner, D.B. Workbook for Atmospheric Dispersion Estimates, U.S. Department of Health, Education and Welfare, Public Service Publication No. 999 - Ap - 26. (PB 191 482) 3rd Printing,May 1970.

74,* Turner, D.B. 'A Diffusion Model for an Urban Area,' Journal of Applied Meteorology, Vol. 3, pp 83 - 91, February 1964.

75.* U.S. Department of Health,Education, and Welfare, Public Health Service, Air Pollution Translations: a bibliography with abstract, Vol. 1, Publication No. AP - 56, May 1969.

76.* U.S. Department of Health, Education, and Welfare, Public Health Service, Guide to Research in Air Pollution , 1969 .

77.* Waley, H. 'A Proposed Canadian Standard for Estimating Atmospheric Dispersion of Combustion Source Pollution from Chimneys,' Atmospheric Environment, Vol. 3, pp 177 - 195, Pergamon Press, 1969.

78. Williams, E.J. Regression Analysis, John Wiley and Sons, Inc., New York , 1959.

79. Wilson, G.J. 'Factorization of the generating function of a pure moving average process,' SIAM Jour Num Analysis, 6, 1, 1969.

80.* Woodwell, G. Radioactivity and fallout: the model pollution, challenge for survival: air. Edited by P. Dunsereau, pp 159 - 171, 1970.

* These are not referenced in this thesis. They are included here only as further reading for readers who want more material on the subject matter of the thesis.

Appendix A

Data Organization on the Magnetic Tape

On magnetic tape all the data for a month are stored in a file, thus there are 93 files on the tape. The first file contains data for July 1964, the second for August, 1964 and so on till the 93rd file which contains data for March 1972. As more monthly data are acquired, files can be created for them starting from file 94 for April 1972.

The files have variable size depending on the number of records each contains. A record is 80 bytes which is the size of an 80 column card. The records are blocked and the block size of 7200 bytes is fixed. Since some of the data are alphanumeric (i.e. wind speed and direction) while the others are numeric the data are stored on tape in A - format so as to make them retrievable in any suitable format.

The first 62, 60, 58, or 56 records (depending on the number of days in the month) in a file contains the daily wind records for two stations which are the Edmonton Industrial Airport and Edmonton International Airport. Each station has 12 two-hourly readings of wind speed and direction on its record per day. For all the records the first two columns contain identification code for the record, The wind record has '21' as its identification

code. The next six columns give year, month, and day, which are then followed by columns containing wind direction and speed.

The pollutants' records follow the wind data in the file. Like wind record the first 2 columns contain code that identifies the pollutant. The next 6 columns give year, month, and day. The list that follows show the code dictionary for the pollutants.

Code	Pollutant
31	Soiling Index (Smoke) for Station 1
32	Soiling Index (Smoke) for Station 2
33	Soiling Index (Smoke) for Station 3
34	Soiling Index (Smoke) for Station 4
35	Soiling Index (Smoke) for Station 5
36	Soiling Index (Smoke) for Station 6
41	NO_x hourly average
51	NO_x five minute peak
42	NO_2 hourly average
52	NO_2 five minute peak
61	Oxidant hourly average
71	Hydrocarbon hourly average
81	Carbon Monoxide hourly average

Data for soiling index of smoke or coefficient of haze (COH) are two hourly measurements, so there are 12

data points per station per day. For other pollutants there are 24 data points per day.

In each record, the actual data starts from the 9th column. Each data point for the pollutants is allocated 3 columns. Interpretation of the 3 - column data point for each pollutant is the following.

Pollutant	Value of Retrieved Measurement
COH	X.XX units
NO_x	.XXX ppm
NO_2	.XXX ppm
Oxidant	XX.X pphm
Hydrocarbon	XX.X ppm
Carbon Monoxide	XX.X ppm

If 999 is encountered in the data field, it indicates that there is no measurement for that pollutant at that particular time.

Appendix B

Initial Estimate Algorithm

(1) Perform the necessary transformation and differencing to obtain w_t from Z_t. Thus, the process to be investigated is an ARMA(p,q).

(2) Estimate the autocovariance function, C(k), of the series w_t.

(3) Estimate the autoregressive parameters a_1, a_2, ..., a_p from the autocovariances by solving the following set of equations

$$\hat{a}_1 C(q) + \hat{a}_2 C(q-1) + \ldots\ldots + \hat{a}_p C(q-p+1) = C(q+1)$$
$$\hat{a}_1 C(q+1) + \hat{a}_2 C(q) + \ldots\ldots + \hat{a}_p C(q-p+2) = C(q+2)$$
$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$
$$\hat{a}_1 C(q+p-1) + \hat{a}_2 C(q+p-2) + \ldots\ldots + \hat{a}_p C(q) = C(q+p)$$

(4) Use the estimates $\hat{a}_1$, $\hat{a}_2$, ..., $\hat{a}_p$ obtained in (3) to obtain modified covariance sequence C'(j) for j = 0, 1, ..., q as follows:

$$C'(j) = \sum_{i=0}^{p} \sum_{k=0}^{p} \hat{a}_i \hat{a}_k \, C(|j + i - k|)$$

for $p > 0$, $(\hat{a}_0 = -1)$

$C'(j) = C(j)$ for $p = 0$.

(5) Use autocovariances $C'(0), C'(1), \ldots\ldots, C'(q)$ to iteratively compute estimates of the moving average parameters $\hat{b}_1, \hat{b}_2, \ldots\ldots, \hat{b}_q$ applying the Newton - Raphson algorithm developed by Wilson [78] and summarized below

(a) Obtain initial values of X's given by

$X_0 = \sqrt{C'(0)}$

$X_j = 0$, $j = 1, \ldots\ldots, q$

(b) Start iteration

(i) Obtain F_j given by

$$F_j = \sum_{i=0}^{q-j} X_i X_{i+j} - C'(j)$$

$j = 0, 1, \ldots\ldots, q.$

(ii) Obtain Matrix T which is given by

$$T = \begin{bmatrix} X_0 & X_1 & .. & .. & X_q \\ X_1 & X_2 & .. & X_q & 0 \\ X_2 & X_3 & X_q & 0 & 0 \\ . & . & & . & . \\ . & . & & . & . \\ . & . & & . & . \\ X_q & 0 & & 0 & 0 \end{bmatrix} + \begin{bmatrix} X_0 & X_1 & .. & .. & X_q \\ 0 & X_0 & X_1 & .. & X_{q-1} \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ 0 & 0 & & & X_0 \end{bmatrix}$$

(iii) Obtain H_j , $j = 0, 1, \ldots\ldots, q$ where H_j is the correction factor for X_j by solving the set of linear equations given by

$$T\ H = F\ ,$$

where T is the $q + 1$ by $q + 1$ matrix, H and F are vectors having $q + 1$ elements each.

(iv) Obtain new values of X_j given by

$$X_j = X_j - H_j\ ; \quad \text{for } j = 0, \ldots\ldots, q$$

(v) If $|F_j| <$ epsilon or $|H_j| <$ epsilon for $j = 0, 1, \ldots\ldots, q$ where epsilon is a small number, say, 0.001 then go to (c) to obtain the estimates of b_1 , b_2 , $\ldots\ldots$, b_q , otherwise go back to (b).

(c) Obtain parameter estimates $\hat{b}_j$ given by

$$b_j = - X_j/X_0 \ , \qquad j = 1, \ldots\ldots, q \, .$$

Obtain also residual variance estimate, $\tilde{\sigma}_e^2$, given by

$$\tilde{\sigma}_e 2 = X_0^{\ 2} \ .$$

(6) Output results.

Appendix C

Model Estimation Algorithm

(1) Input the number p, and values of the initial estimates of autoregressive parameters if any. Input the number q, and values of the initial estimates of moving average parameters if any.

(2) Perform the necessary transformations and differencing on the raw data to obtain series $\{w_t\}$.

(3) Estimate series $\{e_t\}$ of the residuals using the possible model and the current estimates of the parameters.

(4) Estimate sum of squares, SS, of residuals, given by

$$SS = \sum_{i=1}^{n} e_i^2$$ where n is the length of w_t series.

(5) Repeat steps (3) and (4) for all possible values in the admissible region of the parameters. This produces a grid of residual sum of squares.

(6) Pick the set of parameters which give minimum sum of squares, SS, in the grid.

Obtain the estimate S_e^2 of the residual variance, σ_e^2, given by

$$S_e^2 = SS/(n - 1 - q - 2p) .$$

(7) Output the efficient estimates of the model's parameters and the estimate of the residual variance.

Appendix D

Transfer Function Model Identification Algorithm

(1) Plot the output and input series together to detect visually, if possible, the relationship between them.

(2) Fit stochastic model of the type discussed in Chapter 3 to the input series.

(3) Prewhiten the input and transform the output according to the model fitted to the input. This will give a series of residuals y_t , of the input which are not autocorrelated, and a series of residuals w_t , of the output which may or may not be autocorrelated.

(4) Calculate the crosscorrelation function r(k), k = 1, 2, ..., K of y_t and w_t where K = 14.

(5) Use r(k) to obtain the impulse response function v_k given by

$$v_k = r(k)\ s_w/s_y \ .$$

(6) Use r(k) to determine d,p and s as explained in section 4.1.

(7) Use v_k to estimate parameters a_i, i = 1, ..., p, and b_i , i = 0, ..., s.

(8) Carry out steps (1) to (7) for each leading indicator.

(9) Output results.

Appendix E

Marquardt Algorithm for Estimating Parameters

(1) Input the number m, of parameters to be estimated and the initial values $\hat{p}_i$, i = 1, ..., m, of the parameters. Estimate the residual sum of squares due to the initial value of the parameters.

(2) Initialize iteration counter I to zero. Set the maximum number IUP, of iterations that will be allowed, say IUP = 10. Specify R = 0.01, say, and F = 10 where R and F constrain the search. Specify also an epsilon EP = 0.001, as the convergence parameter. Specify a delta del = 0.01 that will be used to perturb the parameters one at a time when the derivatives of e_t are being calculated.

First Stage of Iteration

(3) I = I + 1

If I > IUP go to (17).

(4) Obtain the residual series, e_t , from the model using the current value of the parameters.

(5) Calculate the partial derivative of each e_t in the

series with respect to different parameters as follows:

(a) Derivative $X_{i,t}$ of $e_t(p_1 , p_2 , \ldots\ldots, p_m)$ with respect to parameter p_i is given by

$$X_{i,t} = \{e_t(p_1 , p_2 , \ldots\ldots, p_m) - e_t(p_1 , p_2 , \ldots\ldots, p_i + del, \ldots, p_m) \quad p_i + del, \ldots, p_m)\}/del.$$

(b) Calculate $X_{i,t}$ as in (a) for all e_t, $t = 1, \ldots\ldots, n$ where n is the size of the series.

(c) Carry out steps (a) and (b) for all p_i, $i = 1, \ldots\ldots, m$. Thus for m parameters there will be m vectors each containing n $X_{i,t}$ and all the m vectors constitute an n by m matrix.

(6) Form an m by m matrix

$$A = \{A_{ij}\}, \qquad \text{where}$$

$$A_{ij} = \sum_{t=1}^{n} X_{i,t}X_{j,t} .$$

(7) Form vector G having m elements where

$$G_i = \sum_{t=1}^{n} X_{i,t} e_t \ .$$

(8) Obtain a vector D, of scaling quantities having m elements where

$$D_i = \sqrt{(A_{ii})} \ .$$

Second Stage of Iteration

(9) Obtain the modified linearized equations as follows:

$$A^*_{ij} = A_{ij}/D_i D_j \ , \qquad i \neq j$$
$$A^*_{ii} = 1 + R \ ,$$
$$G^*_i = G_i/D_i \ .$$

(10) Solve the following set of linear equations for H^* :

$$A^*H^* = G^* \ .$$

(11) Scale H* back to H to give the parameter corrections thus

$$H_i = H^*_i/D_i \ .$$

(12) Obtain the new parameter values

$$p_i = p_i + H_i \ , \quad i = 1, \ldots\ldots, m.$$

(13) Estimate sum of squares, SS, of the residuals due to the new parameters. Compare this with the old sum of squares, S0, due to the old parameters.

(14) If SS $\leq$ S0 go to (16).

(15) If SS > S0 multiply R by F. Test if R is greater than its upper bound, UP, say UP = 2.
If R $\leq$ UP, go to 9; otherwise the search has failed, then go to 17.

(16) Test all H_i , i = 1, ..., m.

(a) If every H_i is less than EP, the optimal estimates of the parameters have been obtained. Obtain also the covariance matrix of the parameter estimates which is the inverse of the m by m matrix A.
Variance estimate of parameter p_i is A^{-1}_{ii} .

(b) If the condition in (a) is not satisfied, set old values of the parameters p_i to their new values. Divide R by F. Go back to (3).

(17) Output results.

SERIES A NO_x HOURLY AVERAGES FROM FEBRUARY 22 TO MARCH 7, 1967*

0.004	0.001	0.002	0.001	0.000	0.000
0.006	0.010	0.017	0.019	0.008	0.007
0.003	0.003	0.003	0.001	0.001	0.004
0.006	0.013	0.012	0.032	0.027	0.012
0.010	0.010	0.003	0.008	0.014	0.020
0.017	0.010	0.000	0.010	0.009	0.000
0.007	0.001	0.004	0.000	0.042	0.136
0.130	0.045	0.007	0.003	0.001	0.001
0,007	0.002	0.002	0.002	0.002	0.002
0.002	0.006	0.004	0.006	0.007	0.002
0.001	0.008	0.006	0.000	0.001	0.003
0.012	0.005	0.005	0.011	0.007	0.002
0.003	0.003	0.003	0.003	0.003	0.003
0.003	0.001	0.001	0.001	0.001	0.001
0.001	0.001	0.001	0.004	0.004	0.003
0.002	0.002	0.002	0.002	0.002	0.002
0.002	0.002	0.002	0.002	0.002	0.002
0.002	0.002	0.002	0.002	0.002	0.002
0.002	0.002	0.002	0.002	0.002	0.002
0.002	0.002	0.002	0.002	0.002	0.002
0.002	0.002	0.002	0.002	0.002	0.002
0.002	0.002	0.002	0.007	0.007	0.007
0.007	0.005	0.004	0.003	0.003	0.040
0.060	0.050	0.040	0.023	0.025	0.025

0.018	0.005	0.000	0.000	0.000	0.000
0.007	0.007	0.070	0.030	0.020	0.011
0.011	0.010	0.009	0.002	0.010	0.015
0.015	0.009	0.012	0.009	0.007	0.007
0,004	0.001	0.000	0.000	0.000	0.000
0.001	0.003	0.010	0.010	0.009	0.009
0.010	0.010	0.008	0.000	0.000	0.001
0.002	0.003	0.003	0.002	0.002	0.001
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.030	0.035
0.025	0.020	0.012	0.010	0.007	0.018
0.015	0.011	0.011	0.011	0.012	0.015
0.013	0.008	0.006	0.006	0.005	0.004
0.005	0.006	0.010	0.017	0.011	0.011
0.008	0.011	0.011	0.011	0.010	0.004
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.003	0.003
0.003	0.003	0.003	0.003	0.003	0.003
0.003	0.003	0.004	0.004	0.004	0.004
0.004	0.004	0.004	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.010	0.015

0.022	0.040	0.060	0.050	0.030	0.006
0.003	0.003	0.003	0.003	0.003	0.003
0.003	0.007	0.010	0.007	0.008	0.003
0.011	0.012	0.011	0.010	0.002	0.002
0.002	0.002	0.002	0.001	0.000	0.000

*The first record is the observation for midnight to 1 a.m. on February 22 and consecutive observations are listed row wise. Four rows constitute observations for one day. The observations are in ppm.

SERIES B DAILY MAXIMA OF NO_x FROM APRIL, 1971 TO MARCH, 1972*

0.147	0.033	0.030	0.016	0.072	0.041	0.022
0.038	0.017	0.012	0.020	0.017	0.023	0.039
0.013	0.037	0.034	0.034	0.003	0.015	0.023
0.049	0.116	0.059	0.083	0.033	0.067	0.028
0.095	0.018	0.073	0.055	0.079	0.027	0.013
0.057	0.027	0.013	0.039	0.016	0.053	0.026
0.017	0.009	0.010	0.009	0.044	0.034	0.035
0.039	0.013	0.016	0.030	0.028	0.013	0.014
0.011	0.016	0.018	0.019	0.020	0.008	0.014
0.023	0.009	0.003	0.006	0.013	0.011	0.019
0.023	0.019	0.019	0.009	0.010	0.008	0.008
0.013	0.018	0.023	0.037	0.043	0.037	0.013
0.013	0.019	0.010	0.011	0.015	0.017	0.016
0.015	0.014	0.014	0.005	0.023	0.023	0.023
0.023	0.016	0.016	0.013	0.018	0.014	0.012
0.012	0.016	0.024	0.027	0.016	0.023	0.028
0.024	0.011	0.011	0.012	0.006	0.015	0.014
0.013	0.012	0.020	0.036	0.021	0.021	0.018
0.012	0.012	0.008	0.004	0.022	0.018	0.015
0.023	0.038	0.033	0.014	0.009	0.139	0.024
0.025	0.042	0.020	0.024	0.031	0.037	0.049
0.029	0.028	0.023	0.035	0.071	0.021	0.013
0.014	0.018	0.044	0.059	0.015	0.024	0.022
0.008	0.040	0.023	0.008	0.006	0.001	0.004

0.016	0.016	0.016	0.016	0.007	0.013	0.012
0.007	0.001	0.001	0.001	0.008	0.003	0.003
0.003	0.008	0.000	0.000	0.002	0.002	0.002
0.002	0.014	0.009	0.008	0.008	0.007	0.008
0.007	0.001	0.001	0.001	0.007	0.012	0.013
0.007	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.009	0.016	0.017
0.013	0.007	0.008	0.001	0.052	0.017	0.006
0.020	0.013	0.015	0.013	0.007	0.013	0.016
0.022	0.022	0.022	0.022	0.033	0.029	0.018
0.023	0.064	0.041	0.023	0.026	0.028	0.035
0.030	0.000	0.000	0.000	0.000	0.000	0.007
0.008	0.008	0.008	0.008	0.008	0.008	0.005
0.099	0.020	0.059	0.019	0.015	0.003	0.006
0.010	0.005	0.007	0.002	0.002	0.000	0.018
0.026	0.003	0.000	0.036	0.010	0.010	0.003
0.004	0.017	0.000	0.019	0.067	0.007	0.000
0.000	0.000	0.019	0.049	0.028	0.038	0.042
0.000	0.005	0.005	0.000	0.018	0.001	0.000
0.000	0.004	0.000	0.000	0.020	0.020	0.020
0.020	0.020	0.020	0.020	0.020	0.020	0.020
0.145	0.122	0.000	0.029	0.107	0.096	0.074
0.086	0.073	0.025	0.043	0.192	0.047	0.062
0.097	0.049	0.019	0.041	0.053	0.023	0.176
0.153	0.072	0.094	0.027	0.053	0.121	0.042
0.073	0.074	0.081	0.057	0.156	0.028	0.281
0.117	0.163	0.098	0.381	0.279	0.058	0.050

0.083	0.124	0.052	0.054	0.100	0.281	0.151
0.113	0.132					

*The first record is the observation for April 1, 1971 and consecutive observations are listed row wise. Each row constitures observations for a week. The observations are in ppm.

SERIES C HOURLY TEMPERATURE*

12.0	8.0	7.0	9.0	8.0	8.0	7.0
7.0	5.0	9.0	10.0	12.0	12.0	12.0
13.0	13.0	12.0	10.0	8.0	7.0	7.0
3.0	1.0	-2.0	-1.0	0.0	-1.0	0.0
-1.0	-1.0	0.0	-2.0	-4.0	-2.0	0.0
3.0	5.0	8.0	10.0	12.0	14.0	14.0
12.0	11.0	10.0	10.0	9.0	8.0	8.0
8.0	8.0	7.0	8.0	7.0	7.0	6.0
6.0	7.0	10.0	11.0	14.0	16.0	18.0
19.0	21.0	21.0	20.0	15.0	15.0	15.0
15.0	14.0	9.0	14.0	13.0	14.0	12.0
12.0	10.0	9.0	11.0	13.0	16.0	18.0
23.0	28.0	30.0	31.0	34.0	38.0	32.0
29.0	28.0	37.0	36.0	35.0	32.0	32.0
31.0	27.0	30.0	28.0	29.0	28.0	26.0
26.0	28.0	30.0	32.0	33.0	35.0	36.0
36.0	35.0	32.0	30.0	30.0	28.0	27.0
28.0	28.0	26.0	24.0	22.0	21.0	21.0
20.0	21.0	22.0	24.0	28.0	32.0	35.0
38.0	40.0	40.0	41.0	40.0	37.0	36.0
34.0	33.0	31.0	33.0	33.0	34.0	33.0
31.0	31.0	32.0	30.0	31.0	31.0	33.0
37.0	37.0	38.0	38.0	40.0	41.0	41.0
39.0	37.0	36.0	36.0	35.0	34.0	34.0
34.0	32.0	29.0	27.0	27.0	27.0	27.0

28.0	26.0	28.0	30.0	31.0	34.0	35.0
35.0	36.0	36.0	36.0	34.0	32.0	31.0
27.0	26.0	27.0	27.0	25.0	25.0	25.0
24.0	23.0	22.0	22.0	21.0	21.0	23.0
25.0	27.0	29.0	31.0	31.0	32.0	31.0
28.0	27.0	27.0	27.0	27.0	27.0	26.0
22.0	22.0	22.0	21.0	20.0	21.0	20.0
21.0	23.0	25.0	27.0	28.0	29.0	30.0
31.0	29.0	30.0	28.0	25.0	24.0	23.0
22.0	21.0	20.0	22.0	21.0	21.0	19.0
19.0	17.0	15.0	17.0	20.0	20.0	24.0
26.0	26.0	26.0	26.0	26.0	26.0	24.0
23.0	21.0	20.0	20.0	19.0	16.0	15.0
13.0	13.0	12.0	12.0	11.0	13.0	11.0
14.0	27.0	31.0	37.0	38.0	41.0	44.0
44.0	42.0	40.0	40.0	37.0	36.0	35.0
34.0	27.0	20.0	15.0	12.0	6.0	4.0
2.0	-1.0	-2.0	-1.0	-1.0	1.0	4.0
7.0	10.0	13.0	15.0	15.0	15.0	13.0
13.0	1.30	11.0	9.0	9.0	7.0	7.0
5.0	6.0	7.0	9.0	10.0	12.0	15.0
19.0	23.0	24.0	28.0	29.0	31.0	34.0
34.0	33.0	32.0	32.0	31.0	30.0	31.0

*Read across the page. Measurements are in degrees Fahrenheit.

SERIES D HOURLY WIND SPEED*

3.0	6.0	6.0	8.0	6.0	5.0	6.0
7.0	11.0	11.0	11.0	15.0	12.0	10.0
7.0	9.0	8.0	6.0	3.0	5.0	6.0
6.0	6.0	5.0	6.0	6.0	5.0	9.0
9.0	9.0	6.0	6.0	10.0	9.0	12.0
13.0	14.0	14.0	14.0	12.0	15.0	14.0
11.0	14.0	14.0	15.0	16.0	17.0	12.0
8.0	11.0	11.0	14.0	13.0	9.0	11.0
8.0	10.0	9.0	11.0	9.0	11.0	10.0
10.0	8.0	8.0	5.0	4.0	8.0	8.0
6.0	4.0	5.0	6.0	4.0	4.0	5.0
6.0	6.0	3.0	5.0	6.0	4.0	3.0
1.0	3.0	3.0	2.0	3.0	3.0	4.0
8.0	5.0	19.0	6.0	8.0	9.0	9.0
7.0	10.0	10.0	12.0	12.0	7.0	6.0
9.0	7.0	6.0	5.0	2.0	3.0	4.0
7.0	6.0	3.0	5.0	6.0	6.0	6.0
7.0	7.0	7.0	9.0	6.0	8.0	5.0
7.0	3.0	4.0	4.0	5.0	6.0	3.0
5.0	6.0	6.0	8.0	6.0	7.0	5.0
6.0	5.0	6.0	7.0	8.0	5.0	5.0
4.0	6.0	4.0	6.0	4.0	3.0	4.0
11.0	9.0	6.0	6.0	6.0	6.0	10.0
11.0	12.0	9.0	7.0	8.0	8.0	10.0
10.0	8.0	9.0	9.0	10.0	13.0	13.0

10.0	12.0	10.0	8.0	12.0	15.0	18.0
15.0	12.0	9.0	6.0	7.0	7.0	8.0
7.0	6.0	6.0	4.0	8.0	6.0	10.0
8.0	6.0	7.0	10.0	10.0	10.0	7.0
9.0	10.0	8.0	9.0	9.0	8.0	8.0
6.0	6.0	8.0	9.0	7.0	6.0	4.0
7.0	7.0	7.0	8.0	6.0	7.0	6.0
9.0	12.0	13.0	14.0	15.0	13.0	14.0
16.0	15.0	13.0	13.0	9.0	9.0	8.0
5.0	1.0	2.0	5.0	4.0	4.0	4.0
7.0	4.0	4.0	4.0	5.0	5.0	7.0
6.0	8.0	11.0	13.0	14.0	11.0	9.0
7.0	8.0	8.0	4.0	6.0	5.0	7.0
6.0	6.0	7.0	7.0	6.0	9.0	12.0
12.0	12.0	13.0	12.0	10.0	13.0	12.0
12.0	16.0	14.0	15.0	12.0	15.0	21.0
33.0	31.0	31.0	25.0	23.0	22.0	16.0
10.0	9.0	12.0	13.0	14.0	12.0	10.0
8.0	9.0	8.0	6.0	4.0	2.0	2.0
4.0	6.0	6.0	6.0	7.0	7.0	7.0
8.0	7.0	8.0	10.0	9.0	8.0	9.0
8.0	7.0	8.0	10.0	11.0	12.0	14.0
10.0	8.0	9.0	9.0	8.0	8.0	7.0

*Read across the page. Measurements are in mph.

SERIES E HOURLY TRAFFIC FLOW *

403	213	94	47	22	31	139
786	1595	982	885	973	1157	1185
1151	1078	1420	1670	1122	1034	1021
837	760	731	497	236	89	52
24	41	162	798	1632	1054	1005
1048	1197	1297	1246	1176	1461	1734
1234	1262	1165	1018	718	736	481
255	130	49	36	39	161	815
1645	1086	951	1048	1262	1539	1279
1286	1513	1839	1362	1126	1265	944
793	848	736	517	340	163	95
67	124	379	641	976	1116	1255
1340	1388	1426	1446	1293	1329	1088
944	1189	954	593	860	887	600
357	208	132	68	90	151	201
382	661	753	1002	1177	1325	1552
1495	1262	883	854	869	842	698
626	279	118	64	32	31	40
137	839	1631	1103	949	1002	1179
1347	1177	1203	1389	1620	1149	1188
1111	819	705	777	371	194	137
35	39	27	141	810	1617	1045
918	1088	1186	1254	1196	1136	1440
1768	1154	119	1050	793	695	688
401	180	73	28	40	26	148
789	1592	1063	969	1062	1202	1357

1245	1237	1441	1646	1167	117	1119
847	776	758	480	207	95	47
29	47	149	795	1537	1123	977
1029	1171	1311	1269	1234	1511	1744
1185	1218	1195	1061	742	661	414
229	102	35	33	35	155	812
1654	1129	930	989	1117	1366	1242
1267	1548	1802	1412	1235	1248	1037
793	928	768	490	318	149	103
68	126	375	679	1070	1178	1288
1395	1355	1430	1425	1411	1274	1162
939	1219	1043	567	858	891	611
423	218	120	66	81	180	175
324	669	779	1022	1149	1357	1533
1534	1297	863	940	847	836	726
578	290	100	73	31	14	38
157	778	1591	1033	931	923	1071
1272	1210	1212	1424	1666	1154	1089
1019	772	704	633	374	177	78
40	35	30	147	803	1551	1062
869	986	1065	1282	1132	1091	1385
1736	1184	1194	1070	772	729	787

*Read across the page. Observations are number of vehicles.

SERIES F DAILY TEMPERATURE *

41.0	46.0	44.0	44.0	52.0	59.0	49.0
47.0	47.0	37.0	43.0	46.0	53.0	61.0
49.0	42.0	50.0	50.0	62.0	52.0	53.0
61.0	63.0	57.0	57.0	51.0	48.0	47.0
52.0	61.0	69.0	61.0	72.0	71.0	63.0
59.0	77.0	64.0	57.0	65.0	81.0	86.0
83.0	60.0	61.0	60.0	56.0	65.0	66.0
62.0	68.0	68.0	74.0	80.0	78.0	71.0
68.0	64.0	72.0	72.0	66.0	74.0	77.0
81.0	71.0	60.0	57.0	59.0	61.0	64.0
73.0	71.0	74.0	69.0	59.0	58.0	59.0
59.0	69.0	71.0	77.0	78.0	79.0	63.0
61.0	65.0	51.0	58.0	68.0	73.0	69.0
71.0	55.0	56.0	69.0	65.0	62.0	62.0
67.0	58.0	57.0	61.0	70.0	70.0	79.0
83.0	75.0	78.0	81.0	84.0	80.0	74.0
81.0	72.0	62.0	70.0	75.0	70.0	72.0
79.0	86.0	75.0	88.0	92.0	82.0	80.0
81.0	82.0	82.0	76.0	76.0	81.0	86.0
73.0	59.0	74.0	76.0	73.0	73.0	68.0
81.0	88.0	86.0	71.0	72.0	80.0	72.0
69.0	74.0	73.0	76.0	83.0	72.0	68.0
63.0	72.0	78.0	83.0	58.0	67.0	67.0
65.0	71.0	64.0	68.0	60.0	54.0	48.0
52.0	62.0	64.0	49.0	48.0	58.0	65.0
68.0	59.0	48.0	43.0	43.0	52.0	40.0

45.0	60.0	66.0	68.0	66.0	75.0	73.0
55.0	75.0	73.0	64.0	65.0	56.0	52.0
37.0	34.0	41.0	40.0	48.0	51.0	52.0
53.0	54.0	52.0	51.0	46.0	33.0	22.0
24.0	36.0	39.0	42.0	41.0	35.0	35.0
25.0	11.0	17.0	20.0	47.0	46.0	42.0
41.0	34.0	31.0	39.0	35.0	36.0	32.0
45.0	45.0	39.0	43.0	40.0	26.0	20.0
19.0	20.0	19.0	21.0	25.0	29.0	30.0
39.0	37.0	24.0	16.0	9.0	12.0	23.0
22.0	5.0	-11.0	-10.0	-5.0	-3.0	-8.0
23.0	26.0	19.0	31.0	10.0	3.0	-6.0
-14.0	-13.0	-19.0	1.0	16.0	26.0	20.0
35.0	35.0	26.0	21.0	6.0	27.0	36.0
35.0	23.0	24.0	31.0	20.0	18.0	-17.0
-19.0	2.0	42.0	37.0	30.0	-20.0	-19.0
-11.0	-9.0	-8.0	-14.0	-20.0	-25.0	-22.0
-1.0	24.0	30.0	29.0	19.0	11.0	13.0
10.0	4.0	2.0	-1.0	3.0	13.0	27.0
36.0	36.0	30.0	36.0	11.0	21.0	31.0
32.0	21.0	1.0	0.0	2.0	1.0	7.0
8.0	4.0	-10.0	-7.0	-2.0	-2.0	-4.0
0.0	2.0	11.0	14.0	16.0	19.0	7.0
14.0	33.0	44.0	46.0	53.0	51.0	51.0
49.0	48.0	41.0	35.0	45.0	43.0	43.0
35.0	38.0	33.0	34.0	37.0	35.0	41.0
47.0	50.0					

*Read across the page. Measurements are in degrees Fahrenheit.

SERIES G DAILY WIND SPEED *

9.5	8.5	11.3	11.8	10.9	9.3	11.5
12.0	9.0	16.2	6.0	9.5	7.4	8.0
15.9	8.9	9.8	13.7	10.0	17.7	13.5
14.3	12.8	10.2	7.0	9.7	13.0	8.0
9.3	7.8	11.3	9.5	6.2	7.6	12.6
9.9	12.8	16.0	8.6	8.0	11.0	13.5
16.4	13.7	9.4	8.8	12.1	5.0	8.7
8.5	4.9	8.2	6.1	11.1	7.2	8.8
10.6	14.6	11.7	13.5	10.4	6.5	5.2
8.6	8.6	13.9	11.1	14.6	15.0	7.7
7.7	6.9	6.5	12.0	14.7	18.4	15.6
9.4	3.9	6.3	5.2	7.5	8.3	13.3
11.7	14.2	22.4	16.5	6.2	5.8	5.9
7.2	11.7	14.8	10.4	6.8	9.1	19.7
7.2	14.0	17.3	15.5	12.9	12.6	9.0
6.8	13.0	10.4	3.8	8.3	8.0	6.7
4.9	10.1	15.3	3.8	8.1	6.4	5.5
11.0	10.2	7.0	9.3	17.5	8.4	7.2
6.6	8.5	7.8	9.8	8.4	5.6	10.8
11.8	12.8	9.8	6.4	11.4	10.9	8.4
6.6	6.0	5.5	11.8	11.3	6.4	9.2
10.5	12.2	6.3	7.8	17.0	13.7	11.5
7.6	4.8	11.0	9.8	9.9	9.3	8.4
7.7	12.1	10.8	9.5	18.1	21.6	20.1
14.7	8.0	13.5	11.0	7.3	8.0	7.3
8.1	12.2	11.5	12.6	8.3	8.8	17.4

8.0	7.8	6.4	7.3	10.4	6.2	11.6
11.3	7.2	5.6	5.7	12.5	8.7	14.1
14.6	5.2	13.3	11.8	6.9	9.2	9.6
9.3	6.0	5.4	6.7	10.1	16.3	19.6
10.1	9.9	6.9	10.5	14.4	8.8	13.8
21.0	14.5	10.3	8.2	6.6	5.6	9.2
5.3	5.5	7.7	8.8	7.7	8.3	6.6
11.6	8.2	11.8	11.7	22.8	10.2	12.5
6.8	12.3	13.3	13.8	7.2	7.8	6.6
10.6	6.8	4.9	6.4	11.2	7.6	6.8
16.4	18.6	7.2	6.9	8.4	9.8	7.5
11.3	11.0	8.2	13.3	7.6	16.5	12.9
10.0	8.1	5.6	8.6	8.9	7.1	8.2
10.0	21.0	9.7	14.6	5.8	9.3	13.8
14.9	5.4	8.5	16.5	8.3	8.7	10.8
7.0	10.1	8.3	7.3	20.0	8.3	8.3
11.8	6.0	5.8	8.5	8.3	3.1	4.5
7.8	13.1	9.7	15.3	11.3	8.5	7.6
6.1	8.1	9.8	12.8	8.4	8.1	6.6
7.7	6.5	10.3	12.6	6.1	10.3	12.0
9.3	11.8	8.7	8.3	7.0	10.5	4.8
5.1	13.2	7.4	4.5	6.1	12.1	6.0
6.2	8.0	8.0	10.9	14.1	7.8	9.5
13.7	5.3	6.7	8.1	5.0	7.3	8.8
5.6	9.1	9.6	8.1	5.8	5.3	13.6
12.9	4.9	7.2	6.3	6.0	3.5	5.8
5.8	12.8					

*Read across the page. Measurements are in mph.

SERIES H DAILY TRAFFIC FLOW *

22448	23241	22520	19895	21925	21694	21965
24057	18113	17912	17477	21304	22557	23327
22790	24167	21707	17798	21635	21745	21640
22259	23845	22027	18937	20917	21787	22013
22448	24034	21528	17943	21813	21264	21883
22296	22832	20698	18061	21053	21059	20938
21088	23313	21281	17564	21109	20811	21246
22089	22825	19172	15249	14937	21067	21404
21887	23522	20521	18564	21376	21904	21694
21842	22790	20559	18670	21376	21528	21512
22200	23234	20455	18059	21376	21528	21512
22200	23234	20455	18059	21376	21619	21468
22572	22947	20188	18164	21722	21768	22762
16853	22716	19180	17991	21027	21735	21236
21992	22487	19307	17513	21227	21598	21535
21909	23055	19262	17568	21227	21598	21535
21909	23055	19262	17568	21227	22190	21585
22195	22517	19304	15616	14673	20669	20458
21034	21927	18281	16318	20425	20011	20145
21019	22185	10183	16440	20410	21033	20911
21333	21591	18781	16660	20531	21270	20686
21788	22486	19989	16716	21655	22020	21413
22352	22780	19330	14950	15263	21606	21696
22197	22614	20131	16927	20535	20397	20748
21738	22483	20181	16825	20403	20841	20727
21175	22128	20230	16626	20416	21107	19997

22112	22001	21022	16953	20368	20644	20991
21422	23019	20753	16347	15560	21509	20424
21561	22959	21161	16867	20919	21523	21184
21758	22791	21441	18998	20488	20878	19966
20971	22729	20818	17587	20759	21104	20786
18224	22729	20818	17587	20759	20592	21684
19308	22638	19417	16009	20759	29791	20091
21243	21877	21237	17300	20383	20415	19842
21060	21320	18737	13479	19558	19798	19994
21376	22634	21059	16229	19448	19179	20447
18780	19805	21059	16229	19448	19179	19712
18705	22782	20979	17525	21098	20456	18185
18705	22782	20979	17525	16540	21026	20876
22108	21130	15250	14172	16799	19440	18838
19905	19763	18553	14865	17931	17883	18838
19905	19763	18648	15232	16546	17393	16577
18322	20082	17601	12475	16546	17393	16577
18322	20082	17601	12475	19802	20455	19733
20877	21000	18702	15014	19802	20455	19621
20484	20557	19923	16081	19510	18835	19273
19896	20515	18771	15065	17434	18478	19075
19537	19616	16961	14052	20100	18313	19016
20348	21848	21224	16446	19859	20478	20155
20823	21059	20903	18202	20705	20946	21198
21804	22652	21345	17193	21086	21368	20859
20502	22073	21016	17957	20934	21388	21896
24227	18916					

*Read across the page. Observations are number of vehicles.