RECURSIVE ESTIMATION IN THE IDENTIFICATION OF AIR POLLUTION MODELS

by

Lloyd Paul Steele

A thesis submitted to the
Australian National University
for the degree of Doctor of Philosophy

July 1981

PREFACE

Part of the research reported in Chapter 5 was carried out jointly with Dr A.J. Jakeman and Professor P.C. Young. The results of this work have been published*, and frequently the text of this paper has been closely followed. Some sections of Chapters 4 and 6 are based on work undertaken with Dr Jakeman and partly reported in Steele and Jakeman (1980). The remainder of this thesis, except where otherwise acknowledged in the text, represents the original research of the author.

Paul Steele

* Jakeman, A.J., Steele, L.P. and Young, P.C., 'Instrumental Variable Algorithms for Multiple Input Systems Described by Multiple Transfer Functions.' IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-10, No. 10, October 1980, pp. 593-602.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Professor Peter Young and Dr Anthony Jakeman for their advice and the interest they have shown in this work.

I am also grateful to the following for their assistance: to Dr Neil Daly of the Australian National University for permission to carry out the experiments on the carbon monoxide analyser, the results of which are reported in Chapter 4, and to Mr Anthony Chetcuti for his assistance in setting up the equipment for some of these experiments; to Dr Bob Spear of the University of California at Berkeley for supplying the ozone data used in Chapter 5; to Mr Bill Danaher of the National Capital Development Commission in Canberra for supplying the traffic data used in Chapter 6; and finally to Dr Glenn Johnson of Macquarie University for permission to perform experiments on the Macquarie Urban Air Quality model, the results of which are reported in Chapter 7.

I would also like to thank Cherie Cromwell for her fast and accurate typing and Wendy Steele for her careful proof-reading of the final draft. Special thanks must go to Ian McLean for his timely encouragement and support during the writing of this thesis.

This work was made possible by the financial support of a Commonwealth Postgraduate Research Award.

ABSTRACT

This thesis is an inquiry into the use of recursive estimation procedures in both modeling and measurement of air pollution. Recursive estimation methods were selected for investigation partly because little use has been made of them in air pollution modeling, but primarily for the reason that they offered a promising additional method of time series analysis. In particular the presence of parameter variation in time series models in readily determined, thus assisting the choice of appropriate structures for air pollution models. The modeling undertaken in this thesis has as its principal objective the derivation of simple and operational models for use in air quality management. Since the management of air quality is usually based upon a model of air pollution, such management should be assisted by improved measurement of ambient pollutant levels. A possible added benefit of these improved measurements is that they may enable a better discrimination between alternative air pollution models.

The first three chapters constitute a framework within which the subsequent empirical chapters are appropriately located. A broad perspective of the modeling of complex systems in general is offered in Chapter 1. In Chapter 2 the focus is narrowed to models of air pollution - the primary concern of the thesis. The following chapter introduces the recursive estimation techniques which were employed extensively in later chapters. A series of inquiries into aspects of both air pollution monitoring and modeling are then reported. In Chapter 4 transfer function models of a continuous air pollution analyser were secured and employed in the derivation of robust input signal estimation algorithms. Then, in Chapter 5, estimates are made of missing air pollution data using simple linear dynamic models. There follows in Chapter 6 an attempt at developing a simple time series model for carbon monoxide levels in an urban area. Finally in Chapter 7, an investigation is reported into the dynamic properties of a deterministic model for simulating urban air pollutant levels.

The results of the thesis may be summarised succinctly. Recursive estimation methods have been found both appropriate and effective at several stages in the empirical sections of the thesis: in the

identification of model order and estimation of parameters in dynamic models of continuous air pollution analysers; in the estimation of 'true' input pollutant concentration; in examining the dynamic properties of a deterministic simulation model for urban air pollutant levels; and in the analysis of air pollution and meteorological data. Additionally, it is successfully demonstrated that simple linear dynamic models yield satisfactory estimates of data missing from time series of air pollutant levels. Finally, recursive algorithms are developed for estimating the parameters in a particular class of linear dynamic stochastic models, and are shown to have excellent performance.

Table of Contents

Chapter 1

MODELING COMPLEX SYSTEMS - AN OVERVIEW

1.1 Introduction

The aim of this thesis is to demonstrate that for the purposes of air quality management simple but effective dynamic air pollution models may be obtained by a particular model building procedure based upon recursive estimation methods. The models are simple in the sense that they have few parameters, and they differ from many others in that they are stochastic rather than deterministic. Often linear models are found to be adequate, but even in cases where the overall model is non-linear it is usually possible to separate it into linear and non-linear components, so that the methods of linear systems analysis may still be employed for the linear component.

The development of air pollution models was greatly spurred by the United States Clean Air Act amendments of 1970, although legislation to control air pollution had existed in several industrialised countries prior to that time (Persson, 1977). The crucial aspect of the amendments was the adoption of an air quality management approach to air pollution control (de Nevers et al., 1977), an approach which was based on the specification of a set of national ambient air quality standards. The three basic requirements of an air quality management approach are knowledge of pollutant emissions, air quality monitoring data and air pollution models. Many countries have now adopted some variation of this approach to air pollution control (see, for example, Campbell and Heath, 1977).

It has been recognised for some time (for example, Stern, 1970) that air pollution models may serve two quite diverse purposes: they assist our understanding of the physical nature of the atmosphere; alternatively, they may be used as aids in decision making in air pollution control and city and regional planning. It is the use of models for this latter purpose which is of interest in this thesis.

The development of models suitable for the purposes of air quality management has proved very difficult, and presumably as a consequence, there are a wide variety of such models in the literature

(see, for example, Hanna, 1975). Scorer (1976) concludes that the problems encountered in air pollution modeling lie primarily in the complexity of atmospheric behaviour, and also that the limitations of the models greatly restrict the usefulness of the air quality management approach to air pollution control. However, it is a central contention of this thesis that the pessimism of Scorer is not warranted. By placing air pollution modeling in the wider context of modeling complex systems it is argued here that there exists considerable scope for improving the performance of air pollution models, and hence the usefulness of the air quality management approach to controlling atmospheric pollution.

1.2 Complex Systems and 'Badly Defined' Systems

Simon (1965) defines a complex system as one 'made up of a large number of parts that interact in a non-simple way. In such systems, the whole is more than the sum of the parts, not in an ultimate, metaphysical sense, but in the important pragmatic sense that, given the properties of the parts and the laws of their interaction, it is not a trivial matter to infer the properties of the whole.' Ecosystems are complex in this sense. For example, if the respiration of each organism inhabiting an ecosystem was measured by placing it into a respirometer, the sum of the respiration of individuals would not equal that of the whole ecosystem measured by placing it in a single giant respirometer. Those characteristics which arise from such non-additivity of the component parts are called emergent properties (Perkins, 1975).

Related to the concept of complexity is that of a 'badly defined' system (Young, 1978). Poor understanding of a system may be due either to its complexity or to a number of other factors such as (i) the difficulty or impossibility of performing planned experiments, (ii) the uncertain quality of measurements drawn from the system, and (iii) the limited resources available for making comprehensive measurements of the system. For instance, macroeconomic systems may be considered to be 'badly defined' because they are complex, subject to the vagaries of human behaviour, and plagued by the three factors listed above.

We would contend that for the purposes of modeling for air quality management, urban airsheds also need to be considered as 'badly

defined' systems. Although the physical and chemical interactions that underlie the transport, diffusion and transformation of air pollutants are not conceptually complex, urban airsheds are characterised by conditions (i) to (iii). First, planned experiments in the form of atmospheric tracer experiments are possible (for example, McElroy, 1969 and Lange, 1978) but are difficult and costly. Second, measurements of air pollution levels may be uncertain because the method chosen may not be completely specific for a particular pollutant (for example, Winer et al., 1974), there may be unrecognised calibration errors (for example, see Pitts, 1976), or the location of monitoring stations may not be appropriate to the purposes of the measurements (for example, Ludwig and Shelar, 1978). Moreover, estimates of pollutant emissions are notoriously uncertain, particularly those of motor vehicle emissions (see, for example, Bullin et al., 1980). The third condition, that usually only limited resources are available for measurements of an urban airshed, is perhaps best illustrated by reference to a well known exception, namely, the ambitious St. Louis regional air pollution study (RAPS) in which an extensive network of monitoring stations was established (see, for example, Pooler, 1974). Finally, additional complexity is introduced because dispersion of pollutants in the airshed may be altered due to changes in the surface roughness and as a consequence of the development of urban heat islands (see, for example, Kopec, 1970).

1.3 Modeling 'Badly Defined' Systems

Vemuri (1978) has suggested that the range of problems encountered in the mathematical modeling of complex or 'badly defined' systems can be shown diagrammatically by means of a spectrum with 'black box' modeling at one extreme and 'white box' modeling at the other. Some additions have been made to his diagram which is shown as Figure 1.1.† In pure black box modeling there is no explicit knowledge of the processes occurring in the system and only the input(s) and output(s) may be observed. By contrast, in pure white box modeling, there is complete knowledge of the detailed workings of the system. These two extremes are useful abstractions since there inevitably exists some *a priori* knowledge of a system before attempts are made to construct a mathematical model of it.

† a version of the diagram also appears in Karplus (1976). In: "System Simulation in Water Resource Systems," G.C. Vansteenkiste (ed.), North Holland

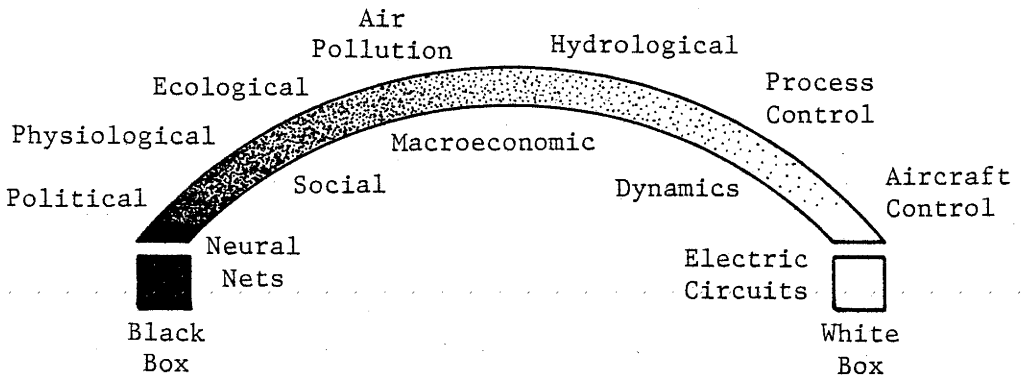


Figure 1.1 The spectrum of modeling problems

Modeling 'badly defined' systems is an inherently difficult task. To illustrate, reference will be made to certain crucial problems with the well-known global models, World2 of Forrester (1971) and World3 of Meadows et al. (1972). Thissen (1978), in concluding his comprehensive analysis of the World3 model, found that 'each specific type of behaviour of World3 appears to be primarily determined by only a fraction of all the assumptions and equations. ... The model builders... have spent needless amounts of time and energy on the formulation and quantification of equations that, in fact, do not matter at all as far as overall behaviour and conclusions are concerned.' More importantly, Thissen argues that 'model analyses ought to be performed during the model construction phase in order to know better on what to focus attention. ... Only thus can it be determined which model parts and what circumstances deserve close attention, and what can be ignored in the light of the goals set to the study as a whole.'

The importance of clearly specified goals prior to modeling 'badly defined' dynamic systems has also been stressed by Young (1978, 1980a). He discusses the relative merits of reductionist and holistic approaches to modeling and favours the latter. This does not preclude

consideration of components of the system but does impose some restraints on the way measurements of the components are made and utilised. His emphasis on dynamic systems is important because methodologies such as regression analysis, which are appropriate for the analysis of static systems, may not be applicable to dynamic systems (see Young, 1968). Some models of 'badly defined' dynamic systems are specifically designed for control and management applications. By such a specification of objectives, and by use of the concept of modal dominance in a complex system, Young (1978) argues that 'the model structure should be chosen so that it 'explains' the data in the simplest, physically meaningful manner possible, and then only to a degree which is defined carefully in relation to the uncertainty on the data.' Before turning in the next section to a detailed consideration of Young's model building procedure, some emphasis must be given to the points raised in this and previous paragraphs.

The position adopted in this thesis is in sympathy with these above views, namely that in modeling complex or 'badly defined' dynamic systems for control and management purposes, the principle of Occam's razor should apply. Thus it is argued that air pollution models for such management purposes should incorporate only as much complexity and data intensiveness as is strictly necessary to achieve consistency with the available air pollution measurements and the goals of the modeling. Similar views have been expressed by Johnson (1979a), Phadke et al., (1976) and, perhaps most convincingly, by Gifford and Hanna (1973) who concluded that 'the detailed urban diffusion models developed so far have the property that they generate much more pollution variability than is actually observed to occur. This seems to us to be a strong argument for the use of simpler models.'

1.4 The Model Building Procedure and the Model Form

The model building procedure suggested by Young (1978, 1980a), and adopted in this thesis, is basically heuristic but has the advantage of being comprehensive. It consists of four major stages -

(i) the formulation of plausible *a priori* model structures, possibly by resort to speculative simulation modeling;

(ii) the choice or identification of appropriate and suitably

parameterised model structures on the basis of the results obtained in the first stage, and in relation to the objectives of the modeling;

(iii) the estimation of those parameters which characterise the model structures eventually chosen; and

(iv) the conditional validation of the estimated models, as implied by the failure to reject the models as reasonable representations of the system's behaviour.

In the first stage, speculative simulation models may be formulated within a probabilistic context by defining the model parameters in terms of statistical probability distributions (see, for example, Hornberger and Spear, 1980; Spear and Hornberger, 1980; and Humphries et al., 1981). The resultant ensemble of models may then be investigated by use of Monte-Carlo simulation analysis in which the model equations are solved repeatedly with the parameters specified by sampling at random from their assumed parent probability distributions. Such analysis may allow determination of those parameters which are most important in giving rise to the observed behaviour in the system. It is well known that the dominant factors in air pollution episodes are the levels of pollutant emissions and certain meteorological variables such as wind speed (see, for example, Hanna, 1971). As the specific purpose of stage (i) is to determine just such basic structural features of the system, this first stage of speculative simulation modeling is not considered in this thesis. The subsequent stages of the modeling procedure are discussed in the next section, but it will first be helpful to explain briefly the philosophy underlying this approach to modeling and to describe the preferred model form.

While recognising that 'badly defined' dynamic systems will in general exhibit both non-linear and non-stationary behaviour, Young (1978) proposes that attention be limited to models designed for control and management purposes and that initially the analysis be restricted to small perturbation behaviour. A linear model is then assumed to be applicable for describing the system behaviour. Whether this assumption holds can be assessed directly because the recursive estimation procedure allows examination of the model parameters to see if they are time invariant. If this is found to be the case (as will be seen for the models of a continuous air pollution analyser in Chapter 4) the estimation of the model can proceed by using the methods

of linear systems analysis. However, if the assumption of linearity does not hold (as will be observed in Chapter 6 for a model of ambient carbon monoxide levels), the non-linear characteristics of the system will be revealed in the pattern of time variation of the model parameters. This is the case because, in general, a non-linear system may be described by a linear model with time-varying parameters. If the causes of the time variation of the parameters can be identified it may be possible to incorporate these causal factors directly into the model. Hence the final model parameters become time invariant, once again enabling them to be estimated by the methods of linear systems analysis.

Young (1978) goes on to argue that, since in practice, the observations of the system will be made in discrete time, and will be characterised by stochastic effects, it is desirable to formulate a discrete time, linear, stochastic model with parameters that may be time variable to allow for any non-stationary behaviour. Furthermore, since it is not usually possible to observe all the variables in a system, he recommends the use of the concept of an observation space which will have a dimension less than that of the full state space. The model is then defined directly in the observation space and is linear in the observations.

The discrete time series transfer function model suggested by Box and Jenkins (1970) is the preferred model form since it is a stochastic model which emphasises the separation of deterministic and stochastic components. This model form is shown diagrammatically in Figure 1.2 for the general case with multiple measured inputs (denoted by the vector $\underline{u}_k$) to the deterministic component of the model (the 'system' model), and multiple outputs $\underline{y}_k$ which can only be measured in error. The stochastic effects ξ_k are assumed to be generated by the 'noise' model, and since they are assumed to be additive to the deterministic output $\underline{x}_k$ from the system model, this model form is sometimes referred to as an errors-in-variables model (see Young, 1980b). The particular terminology used in the formulation of such transfer function models and the methods used for generating the stochastic terms ξ_k will be introduced in Chapter 3. We now turn to a description of stages (ii) to (iv) of the model building procedure.

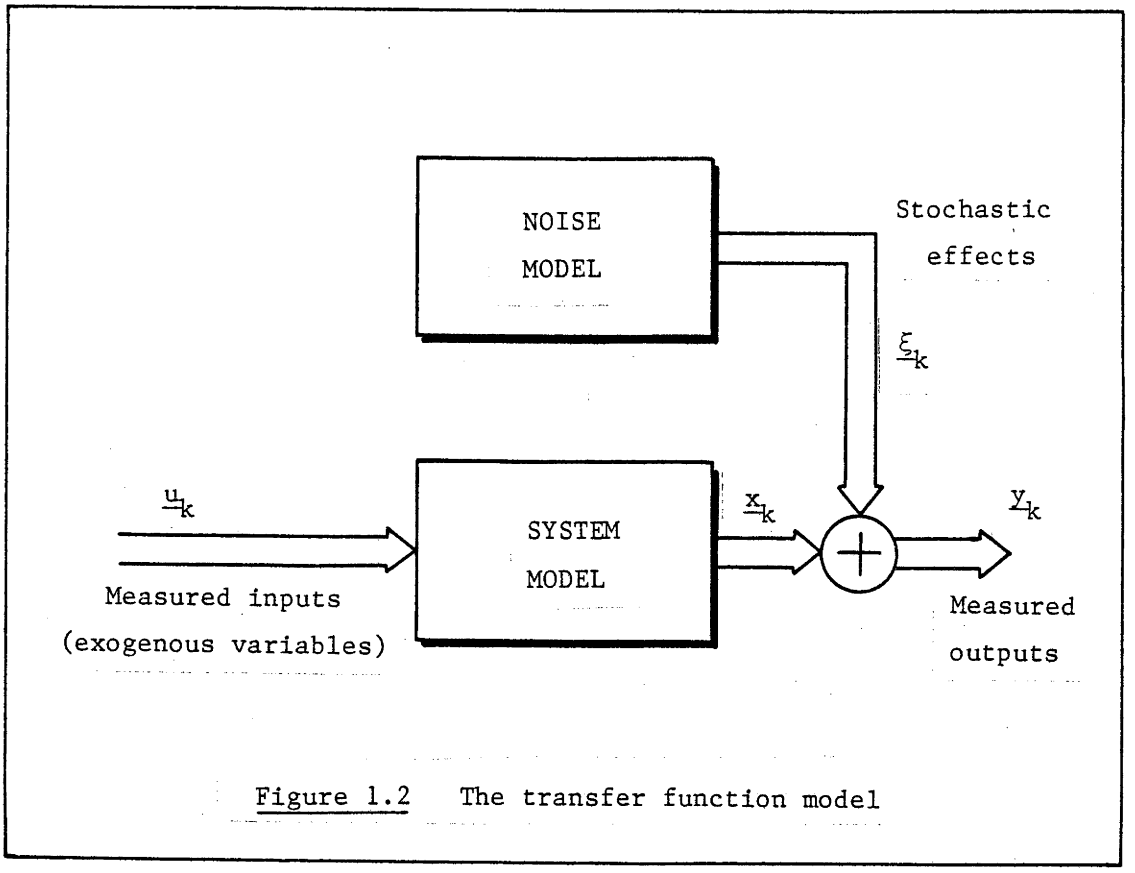


Figure 1.2 The transfer function model

1.5 Identification, Estimation and Validation

The second stage of the model building procedure is the choice or identification of appropriate model structures. The approach to this task suggested by Young (1978) is based upon the use of recursive estimation techniques (for a good introduction, see Young, 1981) in which estimates of the parameter values are obtained at each sampling instant while working serially through the data. An identified model structure is deemed acceptable only if it satisfies certain conditions, a detailed description of which is deferred until Chapter 3. For now it will be sufficient to say that an acceptable model must have a satisfactory physical interpretation and the statistical properties of the parameter estimates should not reveal any evidence of over-parameterisation. In this way the method ensures that the simplest or most parametrically efficient model which is consistent with the observations is chosen to explain the data.

The third stage of the model building procedure, namely the estimation of the parameters in the identified model, is also carried out by use of the recursive techniques utilised in the identification

stage. We also defer the description of these techniques until Chapter 3 and simply say that they are available in the CAPTAIN (Computer Aided Program for Time Series Analysis and the Identification of Noisy Systems) computer package first developed in Cambridge (see Shellswell and Young, 1973) and subsequently in the Centre for Resource and Environmental Studies at the Australian National University (see Young and Jakeman, 1979b). Extensive use of the CAPTAIN package has been made in this thesis and descriptions of it may be found in Moore and Whitehead (1975), Venn and Day (1977) and Freeman (1981).

The final stage in the modeling procedure is validation in which the forecasting ability of the model is evaluated by use of data other than those used in the identification and estimation stages. If such forecasts are found to be acceptable then the model may be considered to be conditionally validated. However, the process of validation does not end at this stage because subsequent data collected from the system may indicate a significant change in its behaviour. If this is the case then further identification, estimation and validation may be required.

In concluding this section it should be mentioned that this particular interpretation of model validation is not universally accepted in the air pollution modeling literature. For example, many claims are made of model validation which are based simply upon a comparison of the measured, and the model's estimated, pollutant concentrations (see, for example, Johnson, 1972). Turner (1979) has attempted to overcome the confusion of terminology surrounding the evaluation of air pollution models by proposing definitions which are consistent with normal dictionary usage. He employs the term model verification to describe a process in which 'some sort of mathematical analysis was performed on a set of data, usually consisting of measured air quality and modeled estimates for the same locations and times, and that the results were favourable.' He restricts the term validation to mean a second verification that substantiates the first, but recommends that any statement regarding model validation should include information on the nature of the data utilised in the model (for example, whether the pollutant concentrations are averaged over three hour or three month periods). Implicit in his definition of validation is the assumption that a model should be tested on a data set different from that used in the first verification. This is a widely accepted

definition of validation and is that which is used in this thesis. Turner also recognises that a good result in the verification of a model may be due to compensating errors in various portions of it, and his solution to this problem is to attempt independently to verify the individual components of the model. While this is an acceptable procedure as far as it goes, he omits to recommend any subsequent attempt to verify the model as a whole. Failure to do this may lead to over-parameterisation and 'surplus content' of the model, characteristics which may not be necessary to explain observed behaviour and which may not be capable of validation against the observations. It is believed that the modeling procedure adopted in this thesis overcomes these problems, and at the same time constitutes an effective and systematic approach to simple air pollution models useful in the task of air quality management.

1.6 Thesis Outline

This thesis has the following broad structure. In Chapters 2 and 3 a survey of the relevant air pollution literature is offered, and the terms and methods of recursive estimation are introduced. In subsequent chapters the principal results of the thesis are reported in the form of a series of investigations into aspects of air pollution monitoring and modeling.

In Chapter 2 the air pollution problem is outlined and important concepts that underlie the air quality management approach to air pollution control are briefly described. The procedures used in the measurement of air pollution are also sketched and the difficulties of air pollution monitoring discussed. Finally, the major types of models used in the management of air pollution are briefly reviewed.

Chapter 3 is primarily a description of the recursive estimation techniques used in the identification and estimation stages of the model building procedure. Representations of dynamic systems are also briefly described.

In Chapter 4 it is shown that a continuous air pollution analyser may be regarded as a single input-single output (SISO) dynamic system, and may be modeled by simple discrete time or continuous time linear

transfer function models. It is then shown that the discrete time models may be used in the development of robust filtering and smoothing algorithms for estimation of the true pollutant concentrations entering the analyser. The algorithms are robust in the sense that they operate successfully in the presence of noise on the measured output signal of the analyser.

The problem of estimating air pollution data missing at one location by using available data at other locations is considered in Chapter 5, and simple linear SISO models are shown to be useful in this task. In an attempt to obtain better estimates of the missing data, linear multiple input-single output (MISO) models are evaluated. A particular class of linear MISO models, in which the characteristic polynomials of the transfer functions associated with each input are not constrained to be identical, was thought to be appropriate, and algorithms are developed for estimating the parameters in such models. Stochastic Monte-Carlo simulation analysis is used to demonstrate the properties of the algorithms.

In Chapter 6 we begin with a brief review of the types of models used for the modeling of the dispersion of vehicular pollutants from roadways. Then we develop a simple time series model which describes carbon monoxide levels in the Canberra City area. We go on to report in Chapter 7 a preliminary investigation of the dynamic behaviour of a deterministic computer-based model of the Eulerian type developed for the simulation of the dispersion of an inert pollutant over an urban area. Finally, in Chapter 8 a summary of the principal conclusions of the thesis is offered.

Chapter 2

INTRODUCTION TO THE AIR POLLUTION PROBLEM AND AIR POLLUTION MODELING

2.1 A Brief History

Air pollution is not a new phenomenon. In England there were serious complaints about air pollution as early as the thirteenth century, and in 1307 Edward I issued a proclamation prohibiting the burning of sea coals in lime kilns in London because of the nuisance caused by smoke (Te Brake, 1975). The nineteenth century poet Shelley (1792-1822) wrote

'Hell is a city much like London -
A populous and smoky city'.

Most of the air pollution described in early accounts of the problem arose from the relatively inefficient burning of coal for industrial purposes and domestic heating and has been termed 'traditional' air pollution. As industrialisation progressed and the population of towns and cities grew rapidly during the nineteenth century so the problem of air pollution became more serious.

Another factor which exacerbated the air pollution problem in the emerging industrial centers was their geographical location. Nearly all were located in river valleys so that atmospheric dispersion of the pollutants was often very poor. Pittsburgh in Pennsylvania is such a center and was known as the Smoky City. There, as recently as the 1940's, it was sometimes necessary to switch on street and vehicle lights during the day since it was often difficult even to see the opposite side of the street (Faith, 1959). The three most well known air pollution disasters in this century have occurred in river valleys - the Meuse Valley in Belgium in December 1930, Donora, Pennsylvania in October 1948 and London in 1952. The great London smog of 1952 occurred over the period 5 to 9 December and caused the death of 4,000 people (UK Govt., 1953). It is generally agreed that the smog mortality was due to irritation of the respiratory tract of persons already suffering from respiratory or cardiovascular disease (Meetham, 1964). Such incidents led to legislative action and the reduction of emissions so that 'traditional' air pollution in most cities has now

been greatly reduced.

While significant advances were being made in the fight against 'traditional' air pollution, a new form was emerging. It was first experienced in Los Angeles where little coal had been used and where most of the energy requirements had been supplied by burning petroleum. This new form of air pollution is now termed photochemical smog since it has been shown to be due to a complex sequence of chemical reactions initiated by sunlight in atmospheres containing non-methane hydrocarbons and nitrogen oxides (Leighton, 1961). The first severe episode of photochemical smog occurred in Los Angeles in the late summer of 1943 'when a grey-blue pall settled over the city, burning eyes and chafing throats' (Bart, 1965). This form of air pollution is now common in large cities which have significant numbers of motor vehicles (Nieboer et al., 1976).

The effects of either traditional air pollution or that associated with motor vehicle emissions are usually restricted to urban and industrial areas. However, there is now a growing awareness that air pollutants emitted in urban areas may contribute significantly to undesirable effects over large non-urban regions. For example, there is considerable evidence that sulphur dioxide and nitrogen oxide emissions into the atmosphere may significantly increase the acidity of rainfall over wide areas. Some ecological systems are very sensitive to increases in the acidity of rainfall and may suffer serious damage as a result (Odén, 1976). The use of very tall chimney stacks as a means of minimising ground level concentrations of pollutants (such as sulphur dioxide) close to the source may simply help to convert a local problem into a regional one. The long range transport of photochemical pollutants from New York City has been reported by Cleveland et al. (1976) and from Los Angeles by Hanna (1977). In addition, photochemical smog clouds covering areas of thousands of square kilometers have been reported in Europe (Guicherit, 1976). These cases underline the fact that international cooperation may become increasingly necessary to solve many of the problems of regional air pollution.

Recently there has arisen the spectre of global air pollution (Bach, 1976). At the present time there are two areas of major concern, namely the increasing atmospheric carbon dioxide levels and

the possible effects of chlorofluoromethanes (CFM's) on the ozone layer in the stratosphere. With respect to the former, there can be no doubt that the mean concentration of carbon dioxide in the atmosphere has been rising steadily since 1958 when routine measurements began at the Mauna Loa Observatory in Hawaii and at the South Pole (Keeling, 1978). However, it is generally accepted that this has been occurring since the mid-nineteenth century when the use of fossil fuels started to become significant. Much of the observed increase in atmospheric carbon dioxide levels is attributed to the large scale burning of fossil fuels while the role played by deforestation, particularly of the tropical rainforests, is still uncertain (see, for example, Pearman, 1980). More uncertain are the climatic consequences of a continuing increase in atmospheric carbon dioxide levels, which it is believed will lead to a global warming effect, particularly in the cold and temperate latitudes (Williams, 1978).

The second major concern over global air pollution is that relating to CFM's and is of quite recent origin. The alarm was first raised by Molina and Rowland (1974) and Rowland and Molina (1975). After considerable debate the United States imposed a total ban on the use of CFM's as aerosol propellants in April 1979, but the production of CFM's for non-aerosol uses such as refrigerants continues at a high level (US Nat Acad Sciences, 1979). No other countries have yet introduced any restrictions on the use of CFM's. The political decisions to restrict CFM's in the United States have been controversial since uncertainty remains in the understanding of atmospheric behaviour, of the complex chemical interactions that take place in the stratosphere, and of the role that naturally occurring chlorine-containing gases may play in the stratospheric chemistry. These decisions have been particularly difficult to achieve since large industries are based on the manufacture and use of CFM's (Dotto and Schiff, 1978).

In the remainder of this chapter it is first proposed to introduce some of the concepts which are central to much of the current discussion about air pollution control. This is followed by a brief review of air pollution modeling in which more than the usual emphasis is given to dynamic stochastic models and time series models, since it is these types of models which will be considered in this thesis. Two additional points should be made, however, before concluding the

present section.

First, in the discussion to this point various scales of air pollution have been implied. Stern (1970) provides a useful guide to the dimensional scales of air pollution systems which is reproduced here as Table 2.1.

TABLE 2.1

Scales of Air Pollution Systems

System	Vertical Scale	Temporal Scale
global	atmosphere	decades
national	stratosphere	years
state	troposphere	months
regional	lowest mile	days
city block	heights of buildings	hours

The various scales are linked because as the horizontal scale is changed there are corresponding changes in the vertical and temporal scales. The temporal scale can be interpreted as the approximate time for the more immediate effects to manifest themselves, or the time period inherent in any immediate corrective or management action. In this thesis attention will be restricted to the two smallest of the air pollution systems in Table 2.1, namely the regional and city block systems.

Secondly, the literature relating to the whole subject of air pollution is now very large, and except for a brief review of urban air pollution modeling in Section 2.4, no attempt will be made to review this. We merely note that there are many good texts on the various aspects of the subject (for example, Magill et al., 1956; Williamson, 1973; and Seinfeld, 1975) with the most comprehensive treatment probably being the third edition of Stern (1976).

2.2 Air Quality

The growing awareness of environmental issues in the past two decades has led to the use of the now ubiquitous concept of 'quality of life'. Embodied in this concept is the realisation that the good

health and well-being of individuals and societies depends on many factors besides those which merely sustain survival. Air quality is a very important aspect of the quality of life both for reasons of public health as well as for reasons of aesthetics. However, a direct definition of air quality is not feasible. We thus choose to indicate three relevant aspects - its public good nature, air quality criteria and air quality standards.

It has been pointed out by Spofford (1975) that air quality exhibits some of the characteristics of a public good and that the air itself may be considered as a common property resource. There are two essential characteristics of the consumption of a public good -

(i) non-excludability - it cannot be provided to one person without others being able to consume it; and

(ii) non-rivalness - the consumption of it by one member of society does not reduce the quantity available for consumption by others.

The public good nature of environmental quality in general and air quality in particular has been discussed by Seneca and Taussig (1974). The air is an example of a common property resource which cannot be owned privately. Hence the market system fails to assign the full cost to users of the resource, and there is an incentive for individuals or organisations to overexploit it, imposing unsolicited costs (less frequently, benefits) on others.

Because of these properties of the atmosphere, there is a need for a collective choice mechanism to improve the utilisation of this common property resource and, hopefully, to achieve more socially acceptable levels of air quality. It is for these reasons that governments have accepted some responsibility for management of air quality. Such management has been, and continues to be, difficult because of the complexity of the problem and the lack of important knowledge about the effects on humans of long term exposure to a variety of pollutants (see, for example, Lave and Seskin, 1976 and the discussion of their paper by Whittenberger, 1976 and by Tukey, 1976).

Air quality criteria describe the state of our knowledge of the

relationship between pollutant concentrations in the air and the associated adverse effects (the dose-response relationship). These adverse effects may be on human health, plants, animals, materials, ecosystems or climate. Present knowledge of these dose-response relationships is far from complete, particularly in the area of the synergistic effects on human health of exposure to several different pollutants.

Air quality standards are legal limits placed on the average levels of ambient air pollutants. Such standards are expressions of public policy and are based upon consideration of the air quality criteria together with a broad range of economic, political, technical and social constraints. Consequently, air quality standards have evolved differently in different countries (Newill, 1977). It should be mentioned, however, that at present there are many air pollutants for which there are no air quality standards. These standards may be specified in a variety of ways, but the most common method involves the specification of an averaging time, an average concentration (over the averaging time) and a frequency of occurrence. For example, the primary ambient air quality standard for oxidants in the United States is an average concentration of oxidants (measured as ozone) of 0.~~10~~ 12 parts per million (ppm) for any one hour period which is not to be exceeded more than once per year.

Air quality standards are subject to revision as more knowledge of dose-response relationships becomes available or as better emission control techniques are developed. Thus in 1978 the United States primary standard for oxidants was relaxed from 0.08 ppm to 0.10 ppm. Also, it has been reported recently by Ott and Mage (1978) that the present United States national ambient air quality standard for carbon monoxide does not necessarily protect a non-smoking individual from levels of carboxyhaemoglobin in the blood in excess of two percent. These findings may require the standard to be specified more accurately.

2.3 Measuring and Monitoring Air Pollution

If we are to have air quality standards then there must be standards for the measurement of air pollution. Measurements of air pollutants are made with a large variety of instruments based upon many

different analytical techniques. The so-called 'wet chemical' methods are increasingly being replaced by the usually more reliable and more specific physical methods of measurement. These physical methods have been surveyed by Bryan (1976). However, the physical methods are not without their problems and Monkman (1976) has drawn attention to poor design and poor manufacture in some instruments. Often there are alternative methods available for the measurement of a particular pollutant and it is now the usual practice for governments to specify a reference method for the measurement of each particular pollutant. Alternative methods must then be demonstrated to produce results which are equivalent to those obtained from the reference method.

Most of the analysers used for the measurement of air pollution are continuous (rather than discrete) analysers and the dynamic properties of these have been described by Larsen et al. (1965), Saltzman (1970) and Schnelle and Neeley (1972). It is important for the time resolution of continuous analysers to satisfy the objectives of the measurement program, and in particular to be dynamically fast enough so that transient violations of air quality standards will be detected. The dynamic properties of continuous analysers will be considered in some detail in Chapter 4.

Most air pollution measurements are made at fixed locations. The data obtained from such a fixed monitoring station can, in the strictest sense, only be associated with that particular point. However, it is common practice to assume that data from a fixed station represent the air quality in a certain geographical area surrounding the station. Whether this assumption is reasonable depends on a number of factors such as the type of pollutant, the time period over which the pollutant concentrations are averaged and the proximity of the monitoring station to local pollution sources. In general this assumption is more reasonable for longer averaging times such as one month or one year. However, if pollutant concentrations averaged over shorter periods (such as one day or one hour) are of interest then the assumption may be invalid. Certainly the findings of Goldstein and Landovitz (1977a, 1977b) seem to indicate that in New York City the daily average measurements of sulphur dioxide and smokeshade obtained at 40 air monitoring stations do not adequately represent the areas surrounding each station.

The siting of monitoring stations for carbon monoxide is particularly difficult since Ott (1972) has shown that eight-hour carbon monoxide concentrations can vary by a factor of three between sites which are less than three kilometers apart. Some guidelines for the selection of sites for carbon monoxide monitoring have been provided by Ludwig and Kealoha (1975), although it is apparent that there is no perfect station siting plan (Bryan, 1976). Generally it is necessary to strike a balance between the density of the monitoring network and the costs of operating the monitoring program. There have been calls for development of procedures such as small and unobtrusive individual air pollution monitors in order to obtain better estimates of air pollutant exposure in urban populations (Morgan and Morris, 1977). Such devices may be able to provide data which supplement those from the fixed monitoring stations.

Fully automated air quality monitoring systems which use telecommunications to provide a direct link between the field monitoring stations and the central office of the control agency have been described by Zimmer (1976). These fully automated systems may operate in delayed batch mode or real time mode. The delayed batch mode is the more common, with data being measured and recorded at the field station and transmitted to the central office only once or twice a day. In the real time mode the data are transmitted from the field station to the central office as the data are being recorded. The real time mode permits real time modeling and instantaneously provides data for management decision making, but the high costs of a real time system do not justify its use in most situations.

Finally, it should be remembered that the data obtained from fixed monitoring stations normally provide the basis for calibration or validation of air pollution models. This fact should be recognised in the siting of monitoring stations so that maximum use can be made of the data they provide in air pollution modeling studies.

2.4 Urban Air Pollution Modeling

Air pollution modeling for entire urban areas is a relatively recent development. The first such model was suggested for the Los Angeles area by Frenkiel (1956). Since that time there has been a rapid growth in the use of mathematical models to simulate the

dispersion of airborne pollutants over cities. The subject area of urban air pollution modeling has been reviewed by several authors (for example, Fan and Horie, 1971; Johnson, 1974; Pasquill, 1974; Hanna, 1975; Eschenroeder, 1975; Johnson et al., 1976; Turner, 1979; and Johnson, 1979a) and no attempt is made here to provide a comprehensive review of the subject. Rather, the various types of models are briefly described while discussion of the modeling of dispersion of vehicular pollutants from roadways will be deferred until Chapter 6. More emphasis than usual is given to stochastic and time series models since these have received relatively little attention in the reviews.

2.4.1 Mass Conservation Models

The mass conservation approach gives rise to two main types of models, namely the Eulerian or multibox and the Lagrangian. In the Eulerian models the atmosphere over a region is divided into boxes by specifying a fixed horizontal grid and restricting attention to a layer of air between the ground and an upper boundary, usually the inversion base. The number of vertical divisions is usually kept fixed and variation of the height of the inversion base is allowed for by variation of the vertical dimension of the boxes. By contrast the Lagrangian models are characterised by a framework which moves with the air mass. However, wind shear causes distortion of the framework which poses difficulties such as the accurate introduction of source emissions to the system. This problem can be avoided by considering the movement of only a single cell, models of which are known as trajectory models.

The solution of both the Eulerian and Lagrangian formulations presents problems. For example, the accurate treatment of the advective terms is difficult in the Eulerian formulation, while in the Lagrangian formulation the difficulty encountered in specification of the turbulent diffusion term usually leads to it being ignored. The concepts of advection and turbulent diffusion are described in more detail later in this sub-section.

An alternative to either the Eulerian or Lagrangian formulations is the particle-in-cell method which Sklarew et al. (1972) have modified to include diffusive transport. This is essentially a hybrid of the Eulerian and Lagrangian formulations, in which the mass of

pollutant is divided into equal discrete elements (particles) so that each particle may be considered as a Lagrangian cell moving in relation to a fixed Eulerian framework. The number of particles in each Eulerian cell at any time is then used to indicate the mean concentration of pollutant in that cell at that time.

Johnson et al. (1976) have compared the computational requirements of Eulerian, Lagrangian and particle-in-cell models and conclude that each of the three types has computational advantages for specific applications. While consideration of the computational requirements is important, it should not be the only criterion used for the selection of a particular model since the data requirements of models may differ greatly and the data preparation may place heavy demands on computation time (Johnson, 1980a). In the following description of the mass conservation approach to air pollution modeling attention is limited to the Eulerian formulation. This is done both for simplicity of exposition and to introduce the type of model which is examined in Chapter 7.

The mathematical formulation of models based upon the mass conservation approach usually begins with the continuity equation (2.1), which describes the behaviour of N chemically reactive constituents suspended in a fluid

$$\frac{\partial c_i}{\partial t} + \frac{\partial (uc_i)}{\partial x} + \frac{\partial (vc_i)}{\partial y} + \frac{\partial (wc_i)}{\partial z} = D_i\left(\frac{\partial^2 c_i}{\partial x^2} + \frac{\partial^2 c_i}{\partial y^2} + \frac{\partial^2 c_i}{\partial z^2}\right) + R_i + S_i \qquad (2.1)$$

where c_i is the instantaneous concentration of the ith constituent at the point (x,y,z) and time t in a rectangular co-ordinate system

u, v, w are the instantaneous velocity components describing the motion of the fluid at the point (x,y,z) at time t

D_i is the molecular diffusivity of the ith constituent

R_i is the net rate of change of concentration of the species i due to chemical reaction and will be a function of the concentration terms c_1, c_2, ..., c_N and the fluid

temperature, and

S_i is the net rate of change of concentration of the species i due to sources and sinks and will in general be a function of position (x,y,z) and time t.

While equation (2.1) is often used as the fundamental equation describing the behaviour of air pollutants in the atmosphere, it is important to remember that it constitutes only one of five coupled fundamental equations describing all aspects of the interaction between chemically active constituents in a fluid (Lamb, 1971). However, since concentrations of air pollution in the atmosphere rarely exceed levels of a few parts per million by volume, the presence of air pollutants will cause only negligible changes in atmospheric density and negligible changes to the heat balance of the atmosphere. An important exception is the attenuation of solar radiation by air pollution over urban areas (Bach, 1971). Nevertheless, the reasonable assumption is made that the presence of the air pollutants in the atmosphere does not alter the meteorology and on this basis equation (2.1) becomes uncoupled from the other four fundamental equations and may be solved separately. Moreover, the fluid velocities u,v,w in (2.1) may be considered independent of the concentration terms c_i.

In dealing with turbulent flow in fluids it is customary to assume that the overall motion can be resolved into a fluctuating component superimposed on a general mean flow. Consequently, the instantaneous velocity components are usually represented in a rectangular co-ordinate system as the sum of deterministic mean velocities $(\bar{u},\bar{v},\bar{w})$ and zero mean stochastic terms (u',v',w')

$$\begin{aligned} u &= \bar{u} + u' \\ v &= \bar{v} + v' \\ w &= \bar{w} + w' \end{aligned} \qquad (2.2)$$

The instantaneous concentration terms c_i are themselves random variables which are similarly represented

$$c_i = \bar{c}_i + c_i' \qquad (2.3)$$

while for the purposes of air pollution modeling it is usual to assume that the atmosphere is incompressible, so that

$$\frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{v}}{\partial y} + \frac{\partial \bar{w}}{\partial z} = 0 \qquad (2.4)$$

Substitution of equations (2.2) and (2.3) into (2.1) followed by expansion, rearrangement, use of the assumption of incompressibility (2.4) and finally ensemble averaging leads to the following equation

$$\begin{aligned} \frac{\partial \langle \bar{c}_i \rangle}{\partial t} = & -\bar{u} \frac{\partial \langle \bar{c}_i \rangle}{\partial x} - \bar{v} \frac{\partial \langle \bar{c}_i \rangle}{\partial y} - \bar{w} \frac{\partial \langle \bar{c}_i \rangle}{\partial z} - \frac{\partial \langle \overline{u' c_i'} \rangle}{\partial x} \\ & - \frac{\partial \langle \overline{v' c_i'} \rangle}{\partial y} - \frac{\partial \langle \overline{w' c_i'} \rangle}{\partial z} + D_i \left(\frac{\partial^2 \langle \bar{c}_i \rangle}{\partial x^2} + \frac{\partial^2 \langle \bar{c}_i \rangle}{\partial y^2} \right. \\ & \left. + \frac{\partial^2 \langle \bar{c}_i \rangle}{\partial z^2} \right) + \langle R_i \rangle + S_i \end{aligned} \qquad (2.5)$$

where $\langle\ \rangle$ denotes the ensemble average (see, for example, Batchelor, 1960).

The equation (2.5) may be interpreted as meaning that the time rate of change of mean pollutant concentration is caused by advection or transportation of the mean concentration by the mean wind (the first three terms on the right-hand side), transportation of the concentration fluctuations by the correlated velocity fluctuations (the fourth, fifth and sixth terms on the right-hand side), molecular diffusion, the chemical reactions, and the rate of change due to sources and sinks.

The terms such as $\partial \langle \overline{u' c_i'} \rangle / \partial x$ mean that there is a closure problem and analytic solutions of equation (2.5) cannot be found (Lamb, 1971). Since there is generally little knowledge of the stochastic fluctuation terms it is usual to hypothesise that terms such as $\langle \overline{u' c_i'} \rangle$ can be approximated by a diffusion of a mean concentration

$$\overline{\langle u'c_i'\rangle} = -K_x \frac{\partial \langle \bar{c}_i \rangle}{\partial x}$$

$$\overline{\langle v'c_i'\rangle} = -K_y \frac{\partial \langle \bar{c}_i \rangle}{\partial y} \qquad (2.6)$$

$$\overline{\langle w'c_i'\rangle} = -K_z \frac{\partial \langle \bar{c}_i \rangle}{\partial z}$$

where K_x, K_y and K_z are the eddy or turbulent diffusivities and are generally functions of space and time. The approximations in (2.6) can be based upon the mixing length theory of the gradient transport description of diffusion (Pasquill, 1974). It should be noted that it is the terms such as $\overline{\langle u'c_i'\rangle}$ which lead to the notion of turbulent diffusion. Turbulent diffusion is thus not an inherent phenomenon of fluid motion but simply an artifact of the lack of complete knowledge of the true velocity field (Lamb, 1971). It is clear also that the conceptual separation of dispersion into the components of advection and turbulent diffusion arises from the assumptions (2.2) made about the fluid velocity components. This is not the only way of providing a mathematical treatment of dispersion and alternative treatments have been considered, particularly in the engineering literature (for example, Buffham and Gibilaro, 1970).

In the usual co-ordinate system adopted for air pollution modeling the z-axis is chosen to be normal to the plane of the ground and the eddy diffusivities are then written as $K_z \equiv K_V$ (vertical) and $K_x \equiv K_y \equiv K_H$ (horizontal). Then substitution of equation (2.6) into (2.5) and making the assumption that molecular diffusion is negligible compared to turbulent diffusion leads to the following equation which is known as the advection - diffusion equation

$$\frac{\partial \langle \bar{c}_i \rangle}{\partial t} = -\bar{u}\frac{\partial \langle \bar{c}_i \rangle}{\partial x} - \bar{v}\frac{\partial \langle \bar{c}_i \rangle}{\partial y} - \bar{w}\frac{\partial \langle \bar{c}_i \rangle}{\partial z} + \frac{\partial\left(K_H \frac{\partial \langle \bar{c}_i \rangle}{\partial x}\right)}{\partial x} \qquad (2.7)$$

$$+ \frac{\partial\left(K_H \frac{\partial \langle \bar{c}_i \rangle}{\partial y}\right)}{\partial y} + \frac{\partial\left(K_V \frac{\partial \langle \bar{c}_i \rangle}{\partial z}\right)}{\partial z} + \langle R_i \rangle + S_i$$

In equation (2.7) the term $\langle R_i \rangle$ will contain terms of the form $\langle c_i' c_j' \rangle$ which arise when chemical reactions occur in turbulent flow. It is very difficult to assess the importance of such terms (Donaldson and Hilst, 1972) and it is usual to assume that $\langle R_i \rangle$ is a function of mean concentrations and temperature only.

Of course, assumptions such as the mixing length theory which leads to equation (2.7) cannot in general be true in an atmospheric boundary layer. As a consequence the diffusion equation (2.7) cannot be generally valid for the atmosphere (Calder, 1965). Nevertheless, equation (2.7) is still very useful and the conditions under which it is valid for describing the transport and reactions of air pollutants in the atmosphere have been examined by Lamb (1973) and Lamb and Seinfeld (1973). Reynolds et al. (1973) found that the maximum temporal and horizontal spatial resolution in the source emission function must be 1000 s and 2000 m respectively, so that emissions must be averaged over relatively large distances to conform to the resolution of equation (2.7). This poses difficulties for the accurate treatment of large point sources of pollution, particularly if the pollutants are very reactive.

Air pollution models based upon the advection - diffusion equation (2.7) are implemented by employing a numerical integration procedure to solve the partial differential equation. Finite -difference methods used to solve the equation may cause a pseudo or numerical diffusion which is due to errors in approximating the advective terms (Molenkamp, 1968). Various procedures have been suggested to overcome or minimise this numerical diffusion. Sklarew et al. (1972) use the particle-in-cell method modified to include diffusive transport. However, a large number of particles is necessary to obtain acceptable accuracy. Egan and Mahoney (1972) use a method of moments which requires computation of the zeroth, first and second moments of the concentration distribution within each grid element. Boris and Book (1973) utilise a technique called 'flux correction' while Christensen and Prahm (1976) use a pseudospectral method.

If in equation (2.7) the eddy diffusivities are constant and independent of x, y or z and $\langle R_i \rangle = S_i = 0$ then the simplified equation will describe Fickian diffusion. Analytic solution of the equation is then possible and, in the absence of boundaries, the

solution is Gaussian. For example, in the case of a 'puff' of pollution generated by an instantaneous release from a point source, the distribution of pollutant with respect to distance from the center of the 'puff' is Gaussian. It is this result which has given rise to the so called Gaussian diffusion models.

2.4.2 Gaussian Diffusion Models

The Gaussian diffusion models are the most widely used models for describing the dispersion of primary pollutants. In fact, all eleven dispersion models currently available on the United States Environmental Protection Agency's User's Network for Applied Modeling of Air Pollution (UNAMAP) are Gaussian models (Turner, 1979). Lamb (1971) has shown that the so-called Gaussian puff and Gaussian plume models may be derived as special solutions of equation (2.5) but since the latter are much more common than the former they will be described first. Their essential nature can be best appreciated if they are derived from first principles as shown by Johnson (1979a).

The derivation of the Gaussian plume models is based upon the assumptions that:

(i) both the meteorological conditions and the pollutant emission rates are steady, so that an equilibrium situation has been achieved;

(ii) a single horizontal wind speed u and single wind direction, valid for the whole region, may be specified;

(iii) the effect of the turbulent motion of the atmosphere is to produce cross-wind and vertical concentrations of pollutant which are normally distributed and independent; and

(iv) the pollutant is unreactive.

Assumption (iii) has been partially substantiated by field experiments such as those described by Barad (1958). As is conventional, a rectangular co-ordinate system is chosen with the x-axis aligned with the wind direction and the z-axis normal to the ground so that cross-wind and vertical standard deviations of pollutant concentration can be represented by σ_y and σ_z respectively. If the conservation of mass is invoked and a mixing layer of infinite depth is assumed, then the Gaussian plume models for single point, line and area

sources may be derived by suitable integration (Johnson, 1979a). Since the ground is a barrier to the dispersion of pollutant, it is usually assumed that it behaves as a perfect reflector. Thus for a continuous point source at the point (0,0,H) emitting pollutant at a steady rate of Q (in units of mass per unit time), reflection of the pollutant at the ground is modeled by means of an identical imaginary source at the point (0,0,-H). Remembering that u is the horizontal wind speed, the Gaussian plume formula for such a point source gives the concentration c at any point downwind of the source as

$$c(x,y,z) = \frac{Q}{u\pi\sigma_y\sigma_z} \exp\left(\frac{-y^2}{2\sigma_y^2} - \frac{(z-H)^2}{2\sigma_z^2}\right) \qquad (2.8)$$

The standard deviation terms σ_y and σ_z are usually expressed as empirical functions of both distance downwind from the source and atmospheric stability. The wide range of atmospheric stability typically is represented by a relatively small number of categories (usually six). The choice of these and the specification of the terms σ_y and σ_z have been reviewed by Gifford (1976).

There is a wide variety of Gaussian plume models in use, some of which have been modified to allow for an increase of wind speed with height (Turner, 1979). While these models have the advantage of simplicity they do have limitations such as an inability to perform satisfactorily when wind speed is very low and its direction is poorly specified (Johnson et al., 1976). Gifford (1968) gives a good account of Gaussian plume models while Turner (1979) considers various sources of errors in these models and makes suggestions for improving their performance. Perhaps the most controversial aspect of the use of Gaussian plume models is their application in areas of rough terrain because tracer experiments over such terrain have not been performed. Thus the validity of the assumptions in the Gaussian plume model cannot be tested adequately (Budiansky, 1980).

The Gaussian puff model was developed to overcome some of the limitations of the Gaussian plume models, particularly the steady state assumption and the poor performance in light winds. The Gaussian puff model treats emissions as a series of puffs which may follow different trajectories and have different rates of diffusion. This type of model

appears to perform very well (Roberts et al., 1970) but it has not been widely used. This is somewhat surprising because although the Gaussian puff model requires extensive calculations these can readily be performed on a computer. In fact, the data requirements and computational load of the Gaussian puff model are probably similar to those of an Eulerian model simulation for the same urban region. However, the author is not aware of any comparison of these two types of models.

2.4.3 Physically Based Stochastic Dispersion Models

The Gaussian diffusion models may be considered deterministic, and the models based upon the mass conservation equation may also be considered deterministic even though the fluid velocity components are treated as random variables. Implicit in the formulation of these models is the assumption that variables such as mean pollutant concentration and emission rates can be measured exactly. In view of the discussion in Chapter 1 it would seem highly desirable to make some explicit allowance for uncertainties in the knowledge of the airshed system, and several authors have attempted this.

Desalu et al. (1974) begin with the advection-diffusion equation (2.7) and reduce it to a linear discrete difference equation in which the input terms are pollutant sources, transportation of pollutants by the wind (the advection term) and the boundary conditions. They then develop a stochastic model using a state-space representation of the system in which allowance is made for stochastic inputs and noisy observations of pollutant concentration. Since the discretisation implies a subdivision of the region of interest into grid cells, the states of the system are taken to be the average value of the distribution of pollutant concentration in each cell at each time instant. The state estimates are generated by use of a modified Kalman filter method. The Kalman filter may be thought of as an estimation technique which utilises knowledge of the observations, the process dynamics and the noise statistics. It combines all the information available up to and including the time of the latest measurement to generate the state estimates at that time.

Desalu et al. (1974) go on to describe a simple simulation example of dispersion of sulphur dioxide from a single elevated

continuous source, and compare the downwind ground-level pollutant concentrations estimated from their dynamic model with those obtained from the steady state Gaussian plume model of equation (2.8). They use the conventional co-ordinate system, assume an inversion layer of constant height as the upper boundary, and divide a region downwind of the source into 500 grid cells of fixed dimensions. The estimation of the resultant state vector of dimension 500 creates a very large computational burden for a normal Kalman filter technique. They therefore use a modified technique which bypasses direct calculation of the error covariances. While they find that their model gives reasonably similar results to the Gaussian plume model they realise that such a comparison does not illustrate the full capability of their model because in such a comparison their model is restricted to a steady input and the stochastic terms are set to zero.

Bankoff and Hanzevack (1975) have also developed a stochastic air pollution model based upon the advection-diffusion equation (2.7). These authors address the problem of short term or incident control of urban air pollution (where real-time prediction of pollution levels is essential) and introduce an adaptive, or learning, algorithm to the overall model so as to allow for on-line corrections (based on incoming measurements) to predicted concentrations and selected model parameters. The on-line correction procedure is based upon the Kalman filter technique. They apply their stochastic model to the problem of the estimation of sulphur dioxide concentrations at eight monitoring stations in the Chicago area during a short term pollution incident. Comparison of the pollutant concentrations over periods ranging from one to four hours predicted by their model and by a Gaussian puff model (using the same data) showed their model to be significantly better.

More recently, several authors have modeled sulphur dioxide pollution in the Venetian Lagoon area. Runca et al. (1976) show that a Gaussian plume model is satisfactory for describing long term (three month or twelve month) average pollutant concentrations in the region but is unable to provide an adequate description over shorter periods. For the latter task, and especially for forecasting air pollution episodes (that is, periods of hours or days of greatly elevated pollutant concentrations), Runca et al. (1979) and Fronza et al. (1979) have developed both deterministic and stochastic models based upon the advection-diffusion equation (2.7). The deterministic model

formulation is solved by means of a fractional step algorithm and they use a non-uniform grid spacing in all directions. This allows for fine grid spacing in areas of particular interest and the location of grid points at monitoring stations, thus avoiding the need for interpolation when comparing model output and measured data. The deterministic model performs adequately under 'normal' pollution conditions but it is unsatisfactory for the forecasting of pollution episodes. They overcome this problem by means of a stochastic model which is derived from the deterministic model by the incorporation of suitably defined noise terms. Application of the stochastic model is restricted to small grid subregions so that relatively few state variables require calculation. These authors examined three separate subregions with 18, 18 and 16 state variables respectively. The stochastic model provides forecasts by means of a Kalman filter technique and the four-hour ahead forecasts during pollution episodes show a dramatic improvement over those obtained from the deterministic model.

The stochastic model formulations described in this section can be seen as belonging to the category of models mentioned in Chapter 1, namely those which consist of a deterministic system component and a stochastic noise component. This category of models will be considered in more detail in Chapter 3. Although the results described in this section are not conclusive, it seems clear that stochastic formulations of air pollution models based on the advection-diffusion equation (2.7) are generally superior to purely deterministic formulations, at least for real-time prediction of air pollution levels. At first sight this may appear a little odd, but it is suggested that these findings support the hypothesis that models of badly defined systems will in general give better performance if they make some explicit allowance for the uncertainties in the system.

2.4.4 Statistical Models and Time Series Models

For control and management of air pollution there are alternatives to the types of models described so far. These are the statistical models which are based upon data analysis. Of course, analysis requires the existence of appropriate data and/or the planning of experiments for the collection of such data. The linear rollback models are of this type and are based upon the simple concept that pollutant concentrations are proportional to emissions. This concept

cannot be useful for secondary pollutants such as photochemical oxidants, and a modified rollback procedure has been suggested for such cases (see, for example, de Nevers and Morris, 1975). Other types of statistical models relate ambient pollutant concentrations to meteorological factors and pollutant emissions (for example, Hanna, 1971) or some surrogate measure of emissions (see, for example, Smith and Jeffrey, 1972). Several other statistical models have been described in the reviews by Hanna (1975) and Johnson (1979a).

A class of statistical air pollution models which is of particular interest here is the time series models. The methods of analysis used in the construction of such linear time series models are well described in the pioneering publication of Box and Jenkins (1970). A Box-Jenkins model form will be considered in some detail in Chapter 3, but for now it should be emphasised that normally such models are not mechanistic and only provide an empirical description of the data. A guiding principle in the choice of such time series models is that of parsimony, that is, a model is sought which has few parameters and at the same time provides a good description of the data. This is also the essence of the modeling methodology proposed by Young (1978) and described in Chapter 1. The presence of over-parameterisation in models may make them unreliable if they are used for prediction. Box-Jenkins techniques are also able easily to handle time series which have high serial correlations (as do most air pollution and meteorological time series) or an autoregressive and/or moving average character. By contrast, ordinary regression techniques cannot handle time series with a moving average character.

The linear time series models in the literature can be divided into two categories. Firstly, there are the purely stochastic models in which some variable or transformation of it is assumed stationary in a probabilistic sense and in which statistical patterns are deduced from the data. The models generally utilised are autoregressive (AR), moving average (MA) or autoregressive moving average (ARMA). Time series which exhibit non-stationary behaviour (as do many air pollution time series) may often be transformed by suitable differencing of the original series to produce a behaviour which may be considered stationary. Such differencing procedures lead to the ARIMA (autoregressive integrated moving average) model, but use of these differencing procedures can be dangerous since amplification of the

noise may occur. The second category of time series models is the transfer function models which are composed of a deterministic system component (with input variable(s)) and a stochastic noise component. These will be considered in detail in Chapter 3.

Most of the time series air pollution models described in the literature are of the purely stochastic type. Thus Merz et al. (1972) described univariate ARIMA models for several air pollutants in Los Angeles and were able to discern trends which were qualitatively in agreement with the expected changes in air quality resulting from programs to control emission of air pollutants. Similar types of models for several air pollutants in Riverside, California, have been described by Chock et al. (1975) who find an anomaly in the trend of carbon monoxide levels which they are unable to explain. This illustrates the value of simple time series models for purposes of data verification whereby errors such as those due to calibration faults would readily become apparent. For their modeling studies both Merz et al. (1972) and Chock et al. (1975) use time series which consist of the logarithms of the pollutant concentrations rather than the pollutant concentrations themselves.

McCollister and Wilson (1975) describe two purely stochastic univariate time series models used to forecast concentration of carbon monoxide and photochemical oxidant in Los Angeles. One model utilises past daily maximum concentrations to forecast daily maxima while the other model utilises hourly concentrations to provide forecasts of hourly concentrations. Both models were found to provide forecasts which were significantly more accurate than forecasts based upon persistence (that is, forecasting the concentration at the next time instant(s) to be the same as that measured at the present time instant). Similar findings are reported by Sawaragi et al. (1979) who compare the performance of four non-physical models for predicting air pollutant concentration at times of one, two and three hours in advance. The four types of models which they consider are a persistence model, an autoregressive model, an ARIMA model and a multiple linear regression model. The first three of these models are purely stochastic in nature whereas the fourth incorporates meteorological variables. The prediction accuracy of the models was measured by means of three performance indices, the simplest of which is the sum of the squares of the prediction errors. The persistence

model consistently shows the poorest performance of the four models, with the autoregressive model the best for prediction at one hour in advance and the multiple linear regression model the best for prediction at two or three hours in advance. These findings of Sawaragi et al. (1979) are not necessarily inconsistent with the comment in Chapter 1 that linear regression analysis may not be applicable for modeling dynamic systems, since regression models of dynamic systems may yield unbiased predictions of the dependent variable (see Johnston, 1972). If however, the parameter estimates found by regression methods are used themselves to make any inferences about the physical nature of the dynamic system being modeled, then errors may result.

The reliability of air pollution forecasts provided by ARIMA models has been improved by Finzi et al. (1979) through the explicit incorporation of meteorological inputs, yielding the so-called ARMAX model (ARMA with exogenous inputs) frequently used by economists and statisticians (see, for example, Rowe, 1970). Their ARMAX model, an example of a transfer function model, is used to obtain reasonably accurate real-time forecasts of daily and hourly concentrations of sulphur dioxide in the Venetian Lagoon area.

While the purely stochastic time series models are useful for providing real-time forecasts of air pollution concentrations they suffer from the drawback that they do not explicitly relate emission to the air pollution concentrations as do the Gaussian models or the models based upon the advection-diffusion equation (2.7). Thus if a violation of an air quality standard is predicted to occur by a purely stochastic model, the type or degree of emission control required to avert the pollution episode cannot be specified by the model. This drawback of the purely stochastic models can be overcome by the development of input-output stochastic models (the transfer function and ARMAX models mentioned earlier) which do describe the relationship between inputs such as emissions and/or meteorological variables and outputs which are usually ambient pollutant concentrations. For example, Chock et al. (1975) describe a transfer function model which relates photochemical oxidant (output) to three meteorological variables (inputs) but their model does not include pollutant emission explicitly.

More recently Tiao and Hillmer (1978) have described an interesting approach to the modeling of carbon monoxide and lead levels at several closely related locations in Los Angeles. While their models are not of the linear transfer function type, their approach has similarities to that adopted in this thesis. Firstly, they rely heavily upon analysis of data and then use statistical procedures to estimate parameter values in their proposed models. Secondly, they make some explicit allowances for uncertainty in the system by incorporation in their model of an additive noise or error term. They then assume that the noise sequence is normally distributed with zero mean. This approach of Tiao and Hillmer (1978) will be examined in more detail in Chapter 6 where another approach to the modeling of carbon monoxide levels is described.

In summary, the applicability of transfer function models in air pollution modeling has so far received scant attention. Since suitable transfer function models offer the advantages of being stochastic as well as incorporating the relevant emission and meteorological variables as inputs, they are proposed as an alternative to the other types of models described in this chapter. Further, if the parameters in the transfer function models are estimated by the recursive procedures described briefly in the next chapter it is believed that further flexibility is achieved in this approach to modeling, since time variation of the model parameters can be readily investigated.

Chapter 3

RECURSIVE METHODS FOR THE IDENTIFICATION AND ESTIMATION OF TIME SERIES MODELS

3.1 Introduction

In Chapter 1 the four stages of the model building procedure adopted in this thesis were briefly described and the general form of the preferred time series model was introduced. In addition the concept of recursive procedures for the identification and estimation of time series models was outlined. Part of the identification procedure is the estimation of parameters which are allowed to be time varying. This is accomplished by a minor modification to the usual recursive estimation algorithm. In this chapter it is proposed to amplify these points so that the concepts and methods frequently used in subsequent chapters may be referred to conveniently.

3.2 Recursive Estimation

It may be recalled from Chapter 1 that recursive estimation is a procedure in which estimates (such as parameter or state estimates) are obtained at each sampling instant by working serially through a data set a sample at a time. Thus recursive estimation procedures can be contrasted with the more familiar en bloc methods in which the whole data set is processed in a single operation. The notable feature of recursive estimation algorithms is that they provide an estimate at any particular sampling instant simply by updating the estimate at the previous sample by means of a correction term which itself is composed of terms relevant to either the previous or present sample. It does not seem to be widely known, but the familiar least squares procedure for estimating the single unknown parameter 'a' in the regression relationship

$$y_i = ax_i + e_{yi} \tag{3.1}$$

may be formulated as a recursive procedure. In (3.1) it is assumed that the regression variable x_i is exactly known for all samples (i = 1, 2, ..., k) and that e_{yi} is the error or noise associated with the measurement of y_i. If we denote the least squares estimate of the

parameter 'a' by $\hat{a}$, then $\hat{a}$ is obtained by minimising the least squares cost function J

$$J = \sum_{i=1}^{k} (y_i - x_i\hat{a})^2 = \sum_{i=1}^{k} e_i^2 \tag{3.2}$$

with respect to $\hat{a}$. Differentiation of J with respect to $\hat{a}$ and setting the result to zero yields the familiar least squares estimate $\hat{a}_k$ for k samples as

$$\hat{a}_k = [c_k]^{-1} b_k = p_k b_k$$

$$\text{where} \quad c_k = \sum_{i=1}^{k} x_i^2 \quad \text{and} \quad b_k = \sum_{i=1}^{k} x_i y_i \tag{3.3}$$

and the weighting factor of gain p_k is a strictly decreasing function of k.

The recursive form of (3.3) may be derived quite simply as shown by Young (1981). Firstly it is clear that

$$\sum_{i=1}^{k} x_i^2 = p_k^{-1} = p_{k-1}^{-1} + x_k^2 \tag{3.4}$$

and

$$\sum_{i=1}^{k} x_i y_i = b_k = b_{k-1} + x_k y_k \tag{3.5}$$

Multiplying (3.4) by $p_k p_{k-1}$ yields

$$p_{k-1} = p_k + p_k x_k^2 p_{k-1} \tag{3.6}$$

and then further multiplying through by x_k gives us

$$\begin{aligned} p_{k-1}x_k &= p_k x_k + p_k x_k^3 p_{k-1} \\ &= p_k x_k (1 + p_{k-1} x_k^2) \end{aligned} \tag{3.7}$$

so that division by the term in brackets gives the expression

$$p_k x_k = p_{k-1} x_k (1 + p_{k-1} x_k^2)^{-1} \tag{3.8}$$

Multiplying (3.8) by $p_{k-1} x_k$ and using (3.6) gives

$$\begin{aligned} p_k x_k^2 p_{k-1} &= p_{k-1}^2 x_k^2 (1 + p_{k-1} x_k^2)^{-1} \\ &= p_{k-1} - p_k \end{aligned} \tag{3.9}$$

Thus by rearrangement

$$p_k = p_{k-1} - p_{k-1}^2 x_k^2 (1 + p_{k-1} x_k^2)^{-1} \tag{3.10}$$

Substitution of this result in (3.3), use of (3.5) and noting that $\hat{a}_{k-1} = p_{k-1} b_{k-1}$ we obtain the expression

$$\hat{a}_k = \hat{a}_{k-1} - k_k (x_k \hat{a}_{k-1} - y_k) \tag{3.11}$$

where

$$k_k = p_{k-1} x_k (1 + p_{k-1} x_k^2)^{-1} \tag{3.12}$$

An alternative expression for k_k can be obtained by writing

$$k_k = (p_k p_k^{-1}) p_{k-1} x_k (1 + p_{k-1} x_k^2)^{-1}$$

and substituting the expression for p_k^{-1} from (3.4)

$$\begin{aligned} k_k &= p_k (p_{k-1}^{-1} + x_k^2) p_{k-1} x_k (1 + p_{k-1} x_k^2)^{-1} \\ &= p_k x_k (1 + p_{k-1} x_k^2)(1 + p_{k-1} x_k^2)^{-1} \\ &= p_k x_k \end{aligned} \tag{3.13}$$

The recursive least squares algorithm for the estimation of the single unknown parameter 'a' in the regression relationship (3.1) is thus given by equations (3.10), (3.11) and (3.12) or (3.13). It is clear that the estimate $\hat{a}$ at the kth instant is a linear sum of the estimate at the (k-1)th instant and a correction term based upon the data at the kth instant and the estimate $\hat{a}$ at the (k-1)th instant. In order to use the recursive algorithm (3.10) to (3.12) it must be initiated, and this may be achieved in two ways. Firstly, $\hat{a}_1$ and p_1 can be calculated from the non-recursive least squares algorithm (3.3) so enabling the recursive algorithm to be used for all subsequent samples. Secondly, the value of $\hat{a}_o$ can be set to some arbitrary finite value such as zero while p_o is set to a large number such as 10^3. It can be shown (see Young, 1981) that the recursive least squares algorithm (3.10) to (3.12) provides estimates that are equivalent to those obtained by the en-bloc method (3.3) if the value of p_o is chosen to be large enough.

It should be noted that equation (3.11) can be seen as a discrete step gradient algorithm (see Wilde, 1964) in which the estimate $\hat{a}_{k-1}$ is updated at each sample. The direction of the update is determined by the gradient of the instantaneous cost, and its magnitude by the time variable gain p_k which reduces monotonically as k tends to infinity. Thus the gradient measure has a steadily decreasing effect as more samples are taken, so that the time varying gain p_k can be seen as smoothing out the inaccuracies introduced by the noise terms e_{yi}.

A recursive least squares algorithm for the more general problem of the estimation of the n unknown parameters a_j (j = 1, ..., n) in the linear regression relationship (3.14) also may be readily obtained.

$$x_o = a_1x_1 + a_2x_2 + \ldots + a_nx_n \qquad (3.14)$$

As in the case of a single unknown parameter, the observation y of x_o is made in error, and if vector-matrix notation is used, this relationship may be expressed as

$$y_i = \underline{x}_i^T \underline{a} + e_{yi} \tag{3.15}$$

where $\underline{a}^T = (a_1 \quad a_2 \quad \ldots \quad a_n)$ is the parameter vector, the superscript T denotes the vector transpose, while $\underline{x}_i^T = (x_{1i} \quad x_{2i} \quad \ldots \quad x_{ni})$ for $i = 1,2, \ldots,k$ denotes the vector of exactly known, linearly independent variables. It is further assumed that the elements x_{ji} that comprise the vector $\underline{x}_i^T$ are statistically independent of the measurement noise terms e_{yi}. The recursive least squares algorithm can then be written as

$$P_k = P_{k-1} - P_{k-1}\underline{x}_k(1 + \underline{x}_k^T P_{k-1}\underline{x}_k)^{-1}\underline{x}_k^T P_{k-1} \tag{3.16}$$

$$\underline{\hat{a}}_k = \underline{\hat{a}}_{k-1} - K_k(\underline{x}_k^T \underline{\hat{a}}_{k-1} - y_k) \tag{3.17}$$

where

$$K_k = P_{k-1}\underline{x}_k(1 + \underline{x}_k^T P_{k-1}\underline{x}_k)^{-1} \tag{3.18}$$

or

$$K_k = P_k \underline{x}_k \tag{3.19}$$

and may also be interpreted as a gradient procedure. The weighting factor or gain in this case is the time variable P_k matrix which contains elements which decrease monotonically as k tends to infinity, so that it exerts a smoothing effect which is analogous to that of P_k in the single parameter case. For the non-recursive least squares solution, the P_k matrix has the form

$$P_k = \left(\sum_{i=1}^{k} \underline{x}_i \underline{x}_i^T \right)^{-1} \tag{3.20}$$

As before it is necessary to specify initial values for the recursive algorithm (3.16) to (3.18). By setting $\underline{\hat{a}}_o$ to zero and P_o with large diagonal elements (such as 10^6) and zeros elsewhere, the recursive algorithm will provide performance equivalent to that obtained by a stage-wise solution of the non-recursive least squares equation.

The recursive least squares algorithm (3.16) to (3.18) may be thought of as a deterministic estimation procedure since it is based on

few assumptions about the nature of the regressors or the noise and it provides no statistical information on the parameter estimates. However, this aspect of the algorithm can be improved by the following two mild assumptions about the nature of the noise terms e_{yi} in equation (3.15), remembering that it has already been assumed that the e_{yi} are independent of the variables x_{ji} that comprise the vector $\underline{x}_i$:

(i) the e_{yi} are a zero mean sequence of random variables; and

(ii) the e_{yi} are serially uncorrelated and have a constant variance σ^2.

These assumptions may be summarised as follows

$$E(e_{yi}) = 0 \; ; \quad E(e_{yi}e_{yj}) = \sigma^2\delta_{ij} \; ; \quad E(\underline{x}_i e_{yi}) = 0 \; , \quad \text{for all } i,j$$

where $E(\;\;)$ denotes the expectation operator and δ_{ij} is the Kronecker delta function

$$\delta_{ij} = \begin{cases} 1 \; ; & i = j \\ 0 \; ; & i \neq j \end{cases}$$

If the estimation error after k samples is $\underline{\tilde{a}}_k = \underline{\hat{a}}_k - \underline{a}$ then the covariance matrix of the estimation errors is defined as

$$P_k^* = E(\underline{\tilde{a}}_k \underline{\tilde{a}}_k^T) = \begin{bmatrix} E(\tilde{a}_1^2) & E(\tilde{a}_1\tilde{a}_2) & \cdots & E(\tilde{a}_1\tilde{a}_n) \\ E(\tilde{a}_2\tilde{a}_1) & E(\tilde{a}_2^2) & \cdots & E(\tilde{a}_2\tilde{a}_n) \\ \vdots & \vdots & & \vdots \\ E(\tilde{a}_n\tilde{a}_1) & E(\tilde{a}_n\tilde{a}_2) & \cdots & E(\tilde{a}_n^2) \end{bmatrix}$$

but it can be shown (see Young, 1981) that

$$P_k^* = \sigma^2 P_k \tag{3.21}$$

where P_k is identical to that defined in equation (3.20), and the recursive least squares regression algorithm can be obtained quite simply by substituting $P_k^*/\hat{\sigma}^2$ for P_k in the basic recursive least squares algorithm (3.16) to (3.18). The term $\hat{\sigma}^2$ may be either an a priori estimate of the noise variance σ^2 or an estimate obtained by use of the recursive algorithm.

$$\hat{\sigma}_k^2 = \hat{\sigma}_{k-1}^2 - \frac{1}{k}\left(\hat{\sigma}_{k-1}^2 - e_k^2\right) \tag{3.22}$$

It should be noted, however, that the estimates obtained from (3.22) will be biased for low values of k because it does not include any adjustment for degrees of freedom.

While the recursive least squares algorithm provides the basis for the estimation procedures used in this thesis, it cannot be used directly because it is not suitable for the estimation of parameters in some dynamic models. In the next section we describe dynamic time series models and the means by which unbiased recursive estimates of their parameters are obtained.

3.3 Dynamic Time Series Models

Dynamic systems with a deterministic input (as distinct from purely stochastic dynamic systems) may be represented in the most general way either by a differential equation

$$\begin{aligned} g_q \frac{d^q u}{dt^q} + g_{q-1} \frac{d^{q-1} u}{dt^{q-1}} + \ldots + g_o u \\ = f_p \frac{d^p x}{dt^p} + f_{p-1} \frac{d^{p-1} x}{dt^{p-1}} + \ldots + x \end{aligned} \tag{3.23}$$

or a convolution integral equation

$$x(t) = \int_o^t r(t-w)u(w)dw \qquad (3.24)$$

where u(t) is the input, x(t) is the output and w is a dummy variable introduced for the purpose of the integration. The convolution integral equation (3.24) may also be used in a multivariable form (see Jakeman and Young, 1980a) to describe dynamic systems with multiple inputs and outputs, but for ease of presentation we will restrict attention here to single input-single output (SISO) dynamic systems. If in (3.23) the coefficients g_i (i = 0, 1, ..., q) and p_i (i = 1, 2, ..., p) are functions of time then it may be used as a description of a non-linear dynamic system. If, however, the coefficients g_i and p_i are constant then the equation (3.23) will describe a linear dynamic system. For the moment we will restrict attention to linear dynamic systems and will use the convolution integral (3.24) as the basis for deriving equivalent but more useful representations of dynamic systems. Firstly, we assume that the output x(t) in (3.24) is measured exactly. Then taking the Laplace transform L{ } of (3.24) yields

$$L\{x(t)\} = L\{r(t)\}\, L\{u(t)\} \qquad (3.25)$$

It is well known (see Takahashi et al., 1972) that for linear systems it is possible to replace L{ r(t) } by a ratio of usually finite order polynomials (known as the rational transfer function) in the Laplace operator s, so that

$$L\{x(t)\} = [G_q(s) / F_p(s)]\, L\{u(t)\} \qquad (3.26)$$

where $F_p(s) = 1 + f_1 s + f_2 s^2 + \ldots + f_p s^p$

$$G_q(s) = g_o + g_1 s + g_2 s^2 + \ldots + g_q s^q$$

The benefit in this approach is that (3.26) provides an efficient parameterisation of the convolution integral equation (3.24) which is nominally infinite dimensional. By taking inverse Laplace transforms of (3.26) and returning to the time domain, the differential equation model (3.27) is obtained

$$F_p(D)x(t) = G_q(D)u(t) \qquad (3.27)$$

where D denotes the derivative operator. This is, of course the operator representation of the differential equation (3.23).

The representation of dynamic systems considered so far have been in continuous time. This presents no difficulties for modeling if the systems can be studied intensively, but in studying environmental systems it is much more likely that observations will be made in discrete time and that they will comprise a finite time series. With this in mind we seek discrete time representations of dynamic systems that are also efficient in the sense of having few parameters. This is straightforward since it can be shown (see Phadke and Wu, 1974) that a differential equation model such as (3.27) can be transformed into the discrete time transfer function model

$$x_k = \frac{B(z^{-1})}{A(z^{-1})} u_k \tag{3.28}$$

where x_k and u_k are the values of x(t) and u(t) at the kth sampling instant. $A(z^{-1})$ and $B(z^{-1})$ are polynomials in the backward shift operator z^{-1} (that is, $z^{-1}x_k = x_{k-1}$) of the form

$$A(z^{-1}) = 1 + a_1 z^{-1} + \ldots + a_n z^{-n}$$

$$B(z^{-1}) = b_o + b_1 z^{-1} + \ldots + b_m z^{-m}$$

where n and m are positive integers. The values of n and m are not necessarily equal to p and q in (3.26) but the sampling interval may be chosen so that n = p, that is, the dynamic orders of the continuous and discrete time models are equal. The z^{-1} notation is chosen for the backward shift operator in preference to the B notation used by Box and Jenkins (1970) because we wish to stress the relationship between the forward shift operator z and the z-transform operator (see, for example, Takahashi et al., 1972).

In considering dynamic systems we have assumed so far that the inputs u(t) or u_k and the outputs x(t) or x_k have been measured

exactly. Since this is unlikely to be the case when taking measurements of environmental systems, it seems reasonable to allow explicitly for error or noise of a general nature on the observation of the output signal x_k. This noise term can be considered to account for both measurement error and any other stochastic disturbances affecting the system. By assuming that in general the noise will be 'coloured' and have rational spectral density, the noise ξ_k can be considered to be generated from a discrete 'white' noise sequence e_k with zero mean and constant variance σ^2 by a rational transfer function model such as

$$\xi_k = \frac{D(z^{-1})}{C(z^{-1})} e_k \tag{3.29}$$

Thus the representation of linear dynamic systems that is preferred in this thesis may be written as

$$y_k = \frac{B(z^{-1})}{A(z^{-1})} u_k + \frac{D(z^{-1})}{C(-^{-1})} e_k \tag{3.30}$$

where the subscript k denotes samples taken at equally spaced time intervals. The terms $A(z^{-1})$, $B(z^{-1})$, $C(z^{-1})$ and $D(z^{-1})$ are polynomials in z^{-1} of the following form

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + \ldots + a_n z^{-n} \\ B(z^{-1}) &= b_o + b_1 z^{-1} + \ldots + b_m z^{-m} \\ C(z^{-1}) &= 1 + c_1 z^{-1} + \ldots + c_s z^{-s} \\ D(z^{-1}) &= 1 + d_1 z^{-1} + \ldots + d_r z^{-r} \end{aligned} \tag{3.31}$$

The model form (3.30) is the discrete time transfer function model of Box and Jenkins (1970) and may also be considered to be a difference equation or time series representation of a time invariant parameter system with additive stochastic disturbances. It is shown in block diagram form in Figure 3.1 where it is clear that y_k is the observed

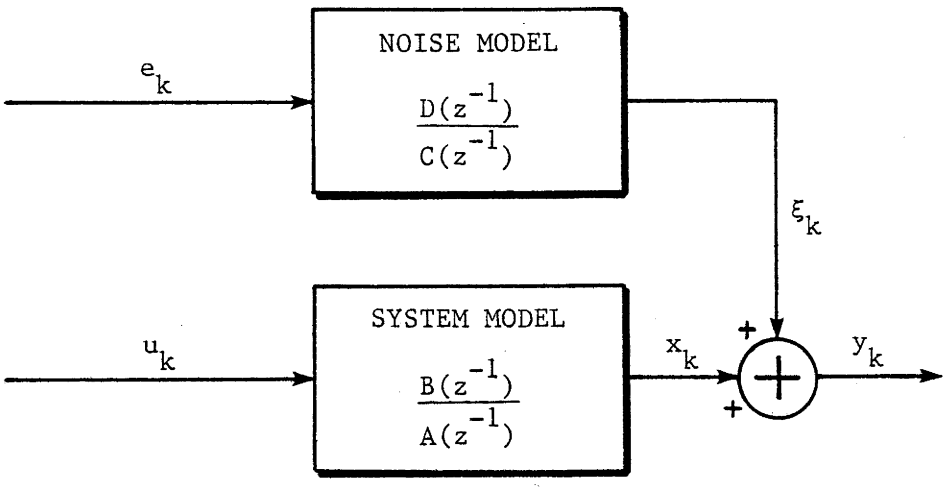


Figure 3.1 Block diagram of discrete time transfer function model

'noisy' output of the 'system' model which is perturbed by the deterministic (measurable) input u_k. The superposition property of linear systems (see Naslin, 1965) allows us to write equation (3.30) as the following connected equations

$$\begin{aligned} &\text{Deterministic component:} && A(z^{-1})x_k = B(z^{-1})u_k && \text{(i)} \\ &\text{Stochastic component} && : C(z^{-1})\xi_k = D(z^{-1})e_k && \text{(ii)} \\ &\text{Observation equation} && : y_k = x_k + \xi_k && \text{(iii)} \end{aligned} \qquad (3.32)$$

where y_k is obtained as the sum of a hypothetical 'noise free' output of the system, x_k, and the noise signal ξ_k. It is usually assumed that the e_k are statistically independent of the input terms u_k so that the properties of the white noise may be written as

$$E(e_k) = 0 \; ; \; E(e_j e_k) = \sigma^2 \delta_{jk} \; ; \; E(e_j u_k) = 0 \; ; \; \text{for all } j,k$$

where E( ) is the expected value operator and δ_{jk} is the Kronecker delta function.

Several other assumptions about the model form (3.30) and (3.32) need to be made (see Young et al., 1971) and may be summarised as follows:

(a) The process (3.32)(i) should be stable; that is, the roots of $A(z^{-1})$ lie outside the unit circle in the complex plane.

(b) The coefficients b_j ($j = 0, 1, \ldots, m$) of the polynomial $B(z^{-1})$ should not all be zero, since otherwise the output will not be activated by the input.

(c) The polynomials $A(z^{-1})$ and $B(z^{-1})$ should have no common factors.

(d) The input u_k should be persistently exciting so that adequate excitation of the system takes place to allow estimation of all the unknown parameters (see Astrom and Bohlin, 1966).

(e) The noise model (3.32)(ii) should be stable and minimum phase; that is, all the roots of both $C(z^{-1})$ and $D(z^{-1})$ should lie outside the unit circle in the complex plane.

(f) The time series data are stationary in a statistical sense (see Box and Jenkins, 1970) either because they are naturally so or because such stationarity has been induced by prior data processing.

These assumptions will ensure that asymptotically efficient estimates are obtained. However, even if some of the assumptions are not satisfied, reasonable (if not asymptotically efficient) estimates may still be obtained.

We now turn to the problem of estimating the parameters in the time series model (3.30). For the moment let us restrict attention to the estimation of the coefficients in the polynomials $A(z^{-1})$ and $B(z^{-1})$ and assume that the model structure has been identified; that is the number of coefficients required in the polynomials has been determined. Least squares estimation procedures will not in general provide acceptable estimates of the parameters because (3.30) is a structural rather than a regression model (see Kendall and Stuart, 1961). Hence least squares estimates may not be statistically consistent and may be biased. This can be seen most readily by writing (3.30) in the form of the vector relationship

$$y_k = \underline{z}_k^T \underline{a} + \eta_k \tag{3.33}$$

where

$$\underline{z}_k^T = (-y_{k-1} \quad \cdots \quad -y_{k-n} \; u_k \quad \cdots \quad u_{k-m})$$

$$\underline{a} = (a_1 \quad \cdots \quad a_n \; b_o \quad \cdots \quad b_m)$$

$$\eta_k = \xi_k + a_1 \xi_{k-1} + \cdots + a_n \xi_{k-n}$$

Clearly the residuals η_k in (3.33) are, in general, both serially dependent and correlated with elements of the vector $\underline{z}_k$.

This problem of estimating the parameters in time series models of the form (3.30) can be approached in a number of ways and the available techniques have been surveyed by Astrom and Eykhoff (1971). One technique which overcomes the problem of bias is the maximum likelihood method suggested by Box and Jenkins (1970) and Astrom and Bohlin (1966), in which all the parameters in the polynomials $A(z^{-1})$, $B(z^{-1})$, $C(z^{-1})$ and $D(z^{-1})$ are estimated simultaneously. However, this procedure requires non-linear estimation to determine the parameters that maximise the likelihood function and, more importantly, it does not readily permit the estimation of time variable parameters. This latter point is of concern to us here because although the model form (3.30) was derived as a representation of a linear dynamic system with time invariant parameters, it can be argued that (3.30) provides a general description of non-linear systems if the parameters are allowed to be time varying (see Young, 1978).

Thus we seek an estimation procedure that readily permits an investigation of any time variation of the parameters in models of the form (3.30). One approach that will allow such an investigation is to formulate the problem in state-space terms and apply the extended Kalman filter (EKF) methods (see Jazwinski, 1970). But the method preferred in this thesis is the alternative recursive instrumental variable-approximate maximum likelihood (IV-AML) procedure suggested by Young (1974). This approach conveniently breaks the estimation problem into two separate parts. First, the coefficients of the polynomials $A(z^{-1})$ and $B(z^{-1})$ are estimated by a recursive IV algorithm, and

second, if they are desired the noise model parameters are estimated by a recursive AML algorithm.

The use of instrumental variables for the estimation of parameters in structural models such as (3.30) overcomes the problems of lack of consistency and bias in the estimates and involves only a simple modification to the recursive least squares regression algorithm described in the previous section. The recursive IV algorithm may be written as follows

$$\underline{\hat{a}}_k = \underline{\hat{a}}_{k-1} - \underline{\hat{k}}_k (\underline{z}_k^T \underline{\hat{a}}_{k-1} - y_k) \tag{3.34}$$

$$\underline{\hat{k}}_k = \hat{P}_{k-1} \underline{\hat{x}}_k (1 + \underline{z}_k^T \hat{P}_{k-1} \underline{\hat{x}}_k)^{-1} \tag{3.35}$$

$$\hat{P}_k = \hat{P}_{k-1} - \hat{P}_{k-1} \underline{\hat{x}}_k (1 + \underline{z}_k^T \hat{P}_{k-1} \underline{\hat{x}}_k)^{-1} \underline{z}_k^T \hat{P}_{k-1} \tag{3.36}$$

where $\underline{z}_k^T$ is as defined in (3.33) and $\underline{\hat{x}}_k$ is an IV vector of the form

$$\underline{\hat{x}}_k = (\hat{x}_{k-1} \ \cdots \ \hat{x}_{k-n} \ u_k \ \cdots \ u_{k-m})^T \tag{3.37}$$

The methods used for the construction of the IV vector $\underline{\hat{x}}_k$ have been well described by Young (1974) and it will be sufficient to mention here that it involves the construction of an 'auxiliary' model of the system which has the same form as the system component of (3.30), that is, (3.32)(i). The algorithm (3.34) to (3.36) provides the basis for the IV estimation procedures used in this thesis.

Once the system model parameters have been estimated with the IV algorithm (3.34) to (3.36) it remains to estimate the noise model parameters, that is, to find the vector $\underline{c}$ in equation (3.38)

$$\xi_k = \underline{n}_k^T \underline{c} + e_k \tag{3.38}$$

where $\underline{c} = (c_1 \ \cdots \ c_s \ d_1 \ \cdots \ d_r)$. But if the IV algorithm

provides a good estimate $\hat{x}_k$ of the hypothetical noise-free output x_k, then a good estimate of the noise ξ_k is provided simply by

$$\hat{\xi}_k = y_k - \hat{x}_k \tag{3.39}$$

The AML algorithm for estimating $\underline{c}$ (see Young, 1974) may be written as

$$\underline{\hat{c}}_k = \underline{\hat{c}}_{k-1} - \underline{k}^n_k (\underline{\hat{n}}^T_k \underline{\hat{c}}_{k-1} - \hat{\xi}_k) \tag{3.40}$$

where $$\underline{k}^n_k = P^n_{k-1} \underline{\hat{n}}_k (1 + \underline{\hat{n}}^T_k P^n_{k-1} \underline{\hat{n}}_k)^{-1} \tag{3.41}$$

and $$P^n_k = P^n_{k-1} - P^n_{k-1} \underline{\hat{n}}_k (1 + \underline{\hat{n}}^T_k P^n_{k-1} \underline{\hat{n}}_k)^{-1} \underline{\hat{n}}^T_k P^n_{k-1} \tag{3.42}$$

with $\underline{\hat{n}}_k$ defined as

$$\underline{\hat{n}}_k = (-\hat{\xi}_{k-1} \cdots -\hat{\xi}_{k-s} \; \hat{e}_{k-1} \cdots \hat{e}_{k-r}) \tag{3.43}$$

The estimates $\hat{e}_k$ are based on the current estimates of $\underline{\hat{c}}_k$ and are provided by using equation (3.38), that is

$$\hat{e}_k = \hat{\xi}_k - \underline{\hat{n}}^T_k \underline{\hat{c}}_k \tag{3.44}$$

The basic IVAML algorithms outlined here have been developed further and a 'refined' IVAML procedure which involves 'prefiltering' of the variables in the basic IV solution has been described by Young (1976). Refined IVAML algorithms have been implemented and evaluated for both SISO and multivariable systems (see Young and Jakeman, 1979c; Jakeman and Young, 1979; and Young and Jakeman, 1980). The development and evaluation of the refined IVAML algorithms for a particular class of multiple input linear systems has been part of the present research programme and will be described in Chapter 5.

3.4 Estimation of Time Varying Parameters

The derivation of the recursive least squares regression

algorithm which was described in Section 3.2 was based upon the assumption that the parameters in the regression model were constant over the observation period, that is

$$\underline{a}_k = \underline{a}_{k-1} \qquad \text{for all } k$$

If this assumption is not correct, then use of the recursive IVAML algorithms described in the previous section for time series analysis may not yield satisfactory results. A simple solution to this problem has been suggested by Young (1974) and involves only a minor modification to the recursive least squares algorithm.

It is first assumed that the regression model parameters in equation (3.15) are time varying and that their variation can be described by a Gauss-Markov stochastic matrix difference equation

$$\underline{a}_k = \Phi \underline{a}_{k-1} + \Gamma \underline{q}_{k-1} \tag{3.45}$$

where Φ is a transition matrix, Γ is an input matrix and $\underline{q}_{k-1}$ is a vector of serially independent, random variables with zero mean and covariance matrix Q_p, so that

$$E(q_k) = 0 \qquad \text{and} \qquad E(q_j q_k^T) = Q_p \delta_{jk}$$

The simplest example of such parameter variation is the random walk model

$$\underline{a}_k = \underline{a}_{k-1} + \underline{q}_{k-1} \tag{3.46}$$

In the case of the recursive least squares algorithm, the additional information provided by equation (3.45) allows 'predictions', or updates (denoted by $\hat{a}_{k/k-1}$ and $P^*_{k/k-1}$) to $\hat{a}_{k-1}$ and P^*_{k-1} obtained at the (k-1)th sample, and these 'prediction' equations may be written

$$\underline{\hat{a}}_{k/k-1} = \Phi\underline{\hat{a}}_{k-1} \tag{3.47}$$

$$P^*_{k/k-1} = \Phi P^*_{k-1}\Phi^T + \Gamma Q_p \Gamma^T \tag{3.48}$$

If the random walk model (3.46) is used to describe the parameter variations then $\Phi = \Gamma = I$ (where I is the identity matrix) and the 'prediction' equations simply become

$$\underline{\hat{a}}_{k/k-1} = \underline{\hat{a}}_{k-1} \tag{3.49}$$

$$P^*_{k/k-1} = P^*_{k-1} + Q_p \tag{3.50}$$

The estimation of the time varying parameters $\underline{a}_k$ is completed by means of the 'correction' equations

$$\underline{\hat{a}}_k = \underline{\hat{a}}_{k/k-1} - P^*_{k/k-1}\underline{x}_k(\hat{\sigma}^2 + \underline{x}_k^T P^*_{k/k-1}\underline{x}_k)^{-1}(\underline{x}_k^T\underline{\hat{a}}_{k/k-1} - y_k) \tag{3.51}$$

$$P^*_k = P^*_{k/k-1} - P^*_{k/k-1}\underline{x}_k(\hat{\sigma}^2 + \underline{x}_k^T P^*_{k/k-1}\underline{x}_k)^{-1}\underline{x}_k^T P^*_{k/k-1} \tag{3.52}$$

which are applied after receipt of the samples at the kth instant. It can be noted at this stage that the prediction-correction algorithm described by equations (3.47), (3.48), (3.51) and (3.52) can be considered as a 'dynamic' least squares procedure and is identical in form to the optimal filter-estimation algorithm first derived by Kalman (1960) for estimating the state vector of a linear discrete time stochastic dynamic system. This latter algorithm is now known as the Kalman filter.

These modifications to allow for the estimation of time variation in the parameters can also be applied to the basic and refined IV algorithms. In the basic IV case, they are completely heuristic in the sense that the modifications are incorporated for purely algorithmic reasons to allow the algorithm to 'track' any changes in the parameters. In effect, the introduction of the prediction equations prevents the elements of the P^*_k matrix from reducing to zero as

k tends to infinity, as we have seen happen in the normal 'constant parameter' recursive algorithms of the least squares type. Consequently the algorithm is always able to monitor changes in the gradient caused by variations in the parameters and so it will adjust the parameter estimates accordingly.

In the refined IV case, this modification can be fully justified on theoretical grounds (Young and Jakeman, 1980) since the P_k^* matrix generated by the algorithm, like the P_k^* matrix in the recursive least squares algorithm, can be considered as a good estimate of the error covariance matrix associated with the recursive estimates.

If the random walk model (3.46) is used, least squares prediction-correction algorithms such as (3.49) to (3.52) or their IV relations can be implemented simply by specifying the covariance matrix Q_p of the parameter variations between samples, and this provides a reasonable approach to the estimation of parameters which are varying slowly, that is, those whose percentage change at each sampling instant is reasonably constant. Also for simplicity it is usually assumed that the patterns of variation of the different parameters are not correlated so that the Q_p matrix can be chosen to be of diagonal form, where the magnitude of each diagonal element may be specified separately to allow for different rates of variation of each of the parameters. In practice, it is unlikely that a priori information which allows the user to specify Q_p is available. As a result the Q_p elements can be considered as 'program parameters' and adjusted on a trial-and-error basis to investigate any parameter variations. This is usually quite acceptable because, as we will see, such time-variable parameter estimation algorithms are used mainly as a device to investigate the presence of non-linearities in the time series data rather than explicitly to obtain 'good' estimates of the parameters. In other words it is the relative changes in the parameters which are important to the analysis rather than the exact numerical changes. In addition, small changes in magnitude of Q_p result in small changes in magnitude of variation. Also, for relatively smooth input data, the values of Q_p which yield a better time varying model fit to the output data are preferred. Usually only one parameter is varied at a time so there is no sympathetic, ambiguous parametric variation. In this way, only one diagonal element of Q_p is non-zero and need be specified.

3.5 Model Order Identification

In the description of the recursive estimation algorithms introduced in previous sections, it was assumed that the model structure had been 'identified' in the sense that the need for the presence of each of the parameters in the model had already been determined. For the time series model form (3.30) this identification stage consists of choosing the order of the polynomials (3.31). Several model order identification procedures have been proposed (see, for example, Box and Jenkins, 1970; and Akaike, 1970) but the method preferred in this thesis is one proposed by Young et al. (1980) which is based on the IV estimation algorithm described in Section 3.3. Here the order identification is based on test statistics which are computed directly from the IV estimation procedure. The aim of the method is to determine that model structure (if any) which provides both a good explanation of the data and a small estimation error for the model parameters. As was mentioned briefly in Chapter 1, the emphasis is placed on obtaining a model structure which avoids over-parameterisation and the ambiguity that can result from the poor definition of parameter estimates. Such over-parameterisation is indicated by monitoring an 'error variance norm' associated with the covariance matrix of the parametric estimation errors P_k^*.

The primary test statistic used in the identification procedure of Young et al. (1980) is an estimation error variance norm (EVN) associated with the $\hat{P}_k^*$ matrix obtained from the refined IV algorithm, and is defined as the arithmetic mean of the diagonal elements $\hat{p}_{ii}^*$ of $\hat{P}_k^*$, that is

$$\text{EVN}\ (\hat{n},\hat{m}) = \frac{1}{\hat{n} + \hat{m} + 1} \left(\sum_{i=1}^{\hat{n}} \hat{p}_{ii}^* + \sum_{i=0}^{\hat{m}} \hat{p}_{\hat{n}+i+1,\ \hat{n}+i+1}^* \right) \quad (3.53)$$

Other norms can be defined (see Young et al., 1980) but the one which is of particular use in this thesis is the normalised EVN (NEVN) defined as

$$\text{NEVN}\ (\hat{n},\hat{m}) = \frac{1}{\hat{n} + \hat{m} + 1} \left(\sum_{i=1}^{\hat{n}} \frac{\hat{p}_{ii}^*}{|\hat{a}_i|} + \sum_{i=0}^{\hat{m}} \frac{\hat{p}_{\hat{n}+i+1,\ \hat{n}+i+1}^*}{|\hat{b}_i|} \right) \quad (3.54)$$

which offers some advantage when the magnitudes of the estimated parameters differ widely. Since the recursive refined IV algorithm (see Young and Jakeman, 1979c) generates an estimate $\hat{P}_k^*$ of P_k^* at each recursive step by

$$\hat{P}_k^* = \hat{\sigma}_k^2 \hat{P}_k \tag{3.55}$$

where $\hat{\sigma}_k^2$ is the estimate (after k samples) of the variance of the e_k sequence in (3.30), the $\hat{P}_k$ matrix may itself be used to assess whether the covariance matrix of the estimation errors has become large. The norms can also be based on the $\hat{P}_k$ matrix calculated in the basic IV algorithm and, although somewhat inferior to those obtained from the refined IV methods, they are usually adequate in most cases. Finally, since the norms can vary over a large range it is convenient to use the natural logarithms of the norms rather than the norms themselves.

Another of the test statistics used in the identification procedure is the total correlation coefficient (or coefficient of determination) R_T^2 defined as

$$R_T^2 = 1 - \frac{\sum_{k=1}^{N} \hat{\xi}_k^2}{\sum_{k=1}^{N} (y_k - \bar{y}_k)^2} \tag{3.56}$$

where N is the total number of samples in the data set and $\bar{y}_k$ is the mean value of the observations y_k. Thus R_T^2 can be seen as a normalised measure of the degree to which the model explains the data, so that if $R_T^2 = 1.0$ the data are explained perfectly by the model, while if $R_T^2 = 0$ there is no explanation at all.

The remaining test statistics are the multiple correlation coefficients defined as

$$R^2_{a_i} = 1 - \frac{1}{\hat{p}^*_{ii} \sum_{k=1}^{N} y_k^2} \qquad (3.57)$$

$$R^2_{b_{j-1}} = 1 - \frac{1}{\hat{p}^*_{j+n,j+n} \sum_{k=1}^{N} u_k^2} \qquad (3.58)$$

where $R^2_{a_i}$ ($i = 1,2, \ldots, \hat{n}$) is the multiple correlation coefficient (MCC) for the ith parameter in the polynomial $A(z^{-1})$, and $R^2_{b_j}$ ($j = 0,1, \ldots ,\hat{m}$) is the MCC for the jth parameter in $B(z^{-1})$. If any of the MCC's tend to unity, it is indicated that a lower order model can adequately represent the data. While the multiple correlation analysis is useful when used with a refined IV algorithm it tends to be less so with basic IV estimation where the MCC's are biased towards unity by noise on the data and can be affected very badly at higher noise levels.

The selection procedure based on this IV method suggested by Young et al. (1980) may be summarised as follows:

(a) Look for that model structure which causes the EVN (or NEVN) to attain, or be close to, its minimum value, so that further increases in $\hat{n}$ lead to substantial increases in the EVN.

(b) Look for a value of R^2_T which is consistent with the degree of model fit expected, where this degree of fit will be dependent on the levels of noise on the data. The value of R^2_T should usually reach a plateau level at a particular value of $\hat{n}$, where further increases in $\hat{n}$ should not substantially increase the value of R^2_T.

(c) Look for values of the MCC that are well removed from unity, but remember that they may be biased towards unity by noise effects.

(d) Check for any time variation of the parameters in the identified model using the IV algorithm described in the previous section. If some meaningful parametric variation is indicated, further model modification

and identification may be necessary.

(e) Check that the estimated $\hat{\xi}_k$ terms are purely stochastic and that they are statistically independent of the deterministic input u_k. This is to ensure that all the systematic components present in the data have been adequately modeled.

(f) Check that when using the refined IV algorithm, the estimated residual error sequence $\hat{e}_k$ has the required zero mean, serially uncorrelated 'white noise' properties.

Chapter 4

APPLICATION OF RECURSIVE TECHNIQUES TO MODELING A CONTINUOUS AIR POLLUTION ANALYSER

4.1 Introduction

Most of the continuous air pollution analysers in use today may be seen as simple linear SISO dynamic systems. These devices usually operate by continuously pumping an air sample at a constant rate through a detection cell where some property of the pollutant gas is measured. The concentration of pollutant in the air sample is thus the input to the SISO dynamic system while the device's recording of the pollutant concentration is the output. The measured property of the pollutant gas may be the absorption of infrared radiation (for carbon monoxide detection), the absorption of ultraviolet radiation (for ozone detection), the fluorescence of the pollutant gas after absorption of light of a suitable wavelength (for sulphur dioxide detection), or the emission of light (chemiluminescence) when the pollutant gas chemically reacts with some reagent gas (for nitric oxide and ozone detection).

It has long been appreciated that such analysers do not always produce an exact recording of the input pollutant concentrations because of their dynamic characteristics (Larsen et al., 1965; Saltzman, 1970; Mage and Noghrey, 1972). A lack of appreciation of the dynamic properties of these devices may lead to their failure to register violations of air quality standards (Schnelle and Neeley, 1972). Larsen et al. (1965) use a first-order differential equation to describe the dynamic response of continuous air pollution analysers and, in order to minimise the noise on the measured output, make recommendations for the selection of flow rate through the analyser. Further, they demonstrate that the first-order differential equation may be solved continuously so as to provide estimates of the true pollutant concentrations.

Both Mage and Noghrey (1972) and Schnelle and Neeley (1972) characterised continuous air pollution analysers by transfer function models. The former used a single step input signal to such a device and differentiated the recorded step response to obtain the discrete time impulse response. This discrete approximation of the impulse

response was then used to estimate true input pollutant concentrations by employing a finite difference method for the solution of the convolution integral equation. Schnelle and Neeley (1972) used a pulse testing technique, followed by data reduction to the frequency domain and a non-linear least squares method to determine the parameters of the transfer function model. A particular feature of the work of Larsen et al. (1965) and Mage and Noghrey (1972) is that neither addressed the problem of noisy output data when devising methods to estimate true input pollutant concentrations.

In this chapter experiments on a continuous air pollution analyser are reported and the applicability of recursive IV algorithms to the identification and estimation of both discrete time and continuous time transfer function models of the analyser is demonstrated. More importantly, the discrete time transfer function model is subsequently used in a stochastic state space formulation which allows the estimation of the true input pollutant concentrations entering the analyser even though the output signal of the analyser is contaminated by noise.

4.2 Experimental Procedure

The experiments were carried out on a Mine Safety Appliances (MSA) carbon monoxide analyser. The particular model was a Model 202 Luft-type infrared analyser (LIRA) which was designed to measure concentrations of carbon monoxide in the concentration range 0 to 200 parts per million (ppm) by volume. This particular device was chosen partly because the generation of input signals was relatively easy and partly because previous studies (Schnelle and Neeley, 1972, and Mage and Noghrey, 1972) have examined the dynamic properties of MSA LIRA carbon monoxide analysers. A schematic diagram of the MSA LIRA Model 202 is shown in Figure 4.1. The principle of operation of this device can be described briefly. Similar beams of infrared radiation pass through the parallel sample and comparison cells and thence into a single detector cell, which in this case contains four per cent carbon monoxide in argon. Between the infrared sources and the cells is a semi-circular beam chopper which operates at a frequency of 2.0 hertz. The chopper alternately blocks the infrared radiation to the sample cell and to the comparison cell. Absorption of infrared radiation by the gas in the detector cell causes an increase in its temperature and

pressure which moves the membrane of a condenser microphone.

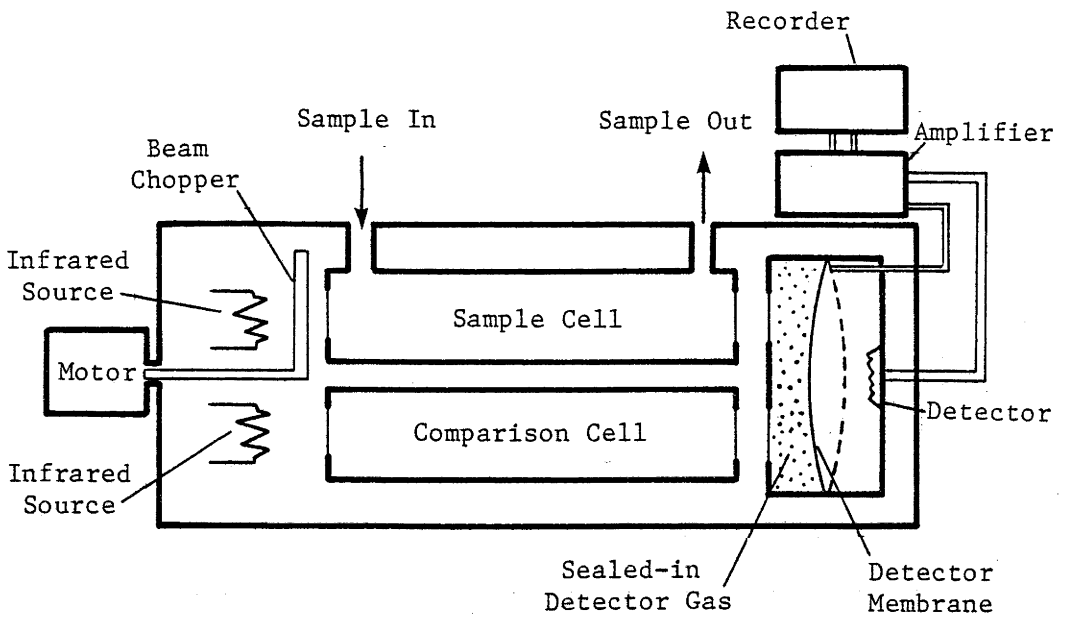


Figure 4.1 Schematic diagram of MSA LIRA Model 202 carbon monoxide analyser

This movement is converted to an electrical signal and amplified by a phase sensitive amplifier which amplifies only those variations which are occurring at the chopped frequency. The amplified signal is then fed to a meter and a chart recorder. The chart recorder has a suitably high impedance so that no significant errors are introduced into the measurement of the output of the device.

A block diagram of the experimental apparatus is shown in Figure 4.2. The plastic bags had a capacity of approximately 40 litres and the gas samples were prepared in the bags immediately prior to the experiments. One bag was filled with dry nitrogen gas from a compressed gas cylinder while in the other bag was prepared a mixture of carbon monoxide in dry nitrogen gas. Thorough mixing of the carbon monoxide in the dry nitrogen was achieved by placing a 20 cm diameter plastic ball inside the plastic bag and shaking vigorously for several minutes. It was verified that such gas mixtures prepared in these plastic bags maintain a stable concentration for periods of several hours. The tap was of a type that, depending on its position,

permitted gas from either bag to be pumped through the sample conditioning system (silica gel filter to remove moisture, a particulate filter to remove dust, plus the associated tubing) and then the analyser. The position of this tap could be changed rapidly so that abrupt changes of the type of gas passing to the analyser could easily be effected. All experiments began with the tap in such a position that allowed pure nitrogen to pass through the device. Experiments were then carried out by changing the position of the tap at known times to produce step-up and step-down changes in the input gas stream. Such changes in the position of the tap were always carried out at times which were integer multiples of the sampling interval which was chosen as 5 s.

The analyser was calibrated in the normal way prior to the experiments. This calibration procedure was a steady state one and first involved the passing of pure dry nitrogen through the analyser for a sufficient time so that the device's output reached a steady state. When this was achieved the device's output was set to zero. In a similar fashion, a known standard concentration of carbon monoxide in nitrogen was passed through the device until a steady state output was achieved and the machine output was altered to match the known concentration. The steady state calibration curve for the analyser is shown in Figure 4.3 and it can be seen that the response is not perfectly linear. This does not affect the investigation of the analyser's dynamic properties because we express all inputs and outputs as a percentage of the full scale deflection (FSD) of the instrument, where this occurred with a concentration of 200 ppm of carbon monoxide. The output from the instrument was recorded in continuous analogue form by a chart recorder which was adjusted to a chart speed of 0.1 cm s^{-1} where the chart paper was marked in divisions of 0.5 cm along the time axis. The chart records were then digitised with a value of the recorder trace being taken at the chosen sampling interval of 5 s.

Five experiments were performed and their basic details are shown in Table 4.1. The experiments were designed to investigate the linearity of the analyser in a dynamic sense (experiments 1 and 2), to examine the efficacy of input signals of differing complexity (1 and 3), and to demonstrate the effect of sample gas flow rate on the dynamic response of the analyser (1, 4 and 5).

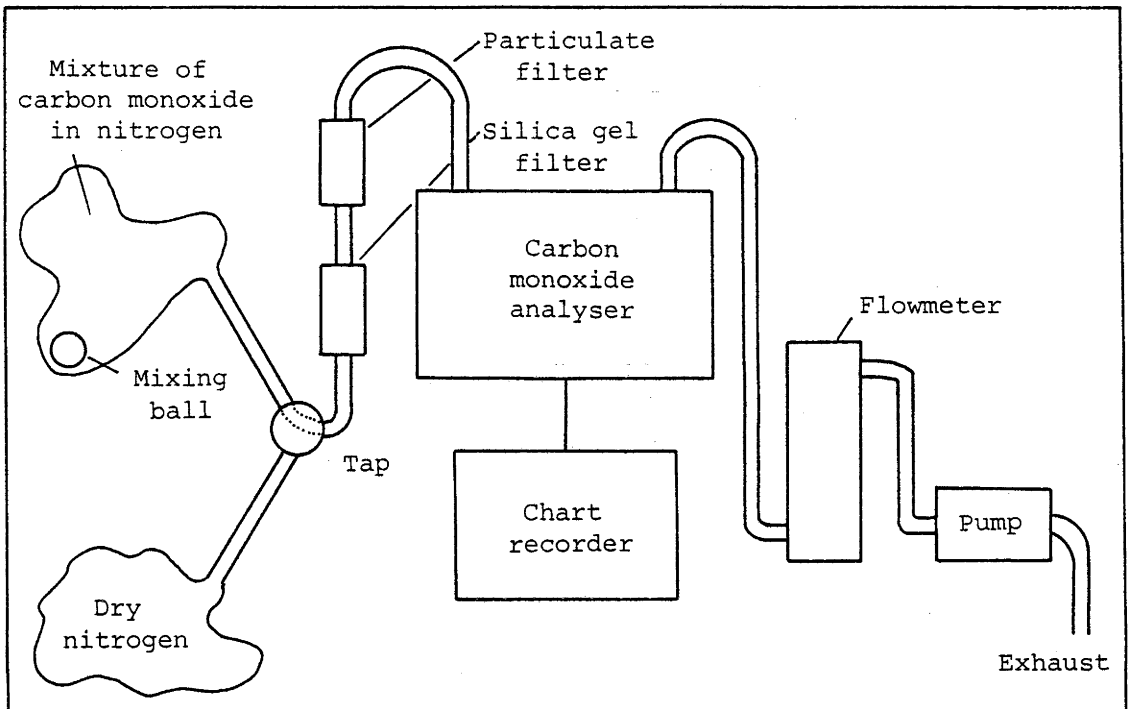


Figure 4.2 Block diagram of apparatus used for performing experiments on carbon monoxide analyser.

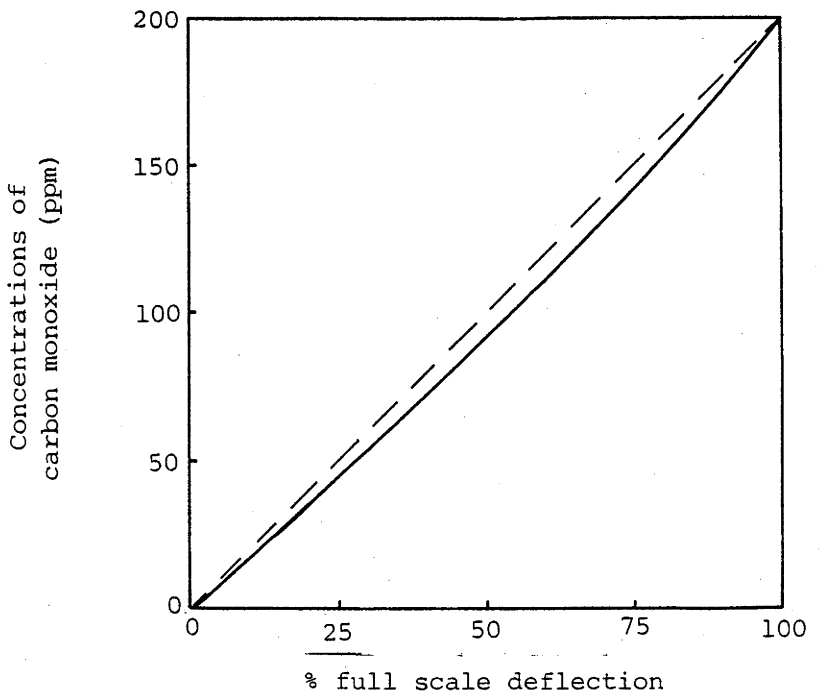


Figure 4.3

Steady state calibration curve (—) of the carbon monoxide analyser with a hypothetical perfectly linear response (--) for comparison.

Before going on to describe the findings of these experiments it is necessary to discuss the identification of model order and the estimation of model parameters, and this is the purpose of the next section. To avoid unnecessary repetition, we will illustrate the general procedures used in all experiments by reference to experiment 1.

Table 4.1

Conditions for Experiments on the Carbon Monoxide Analyser

Experiment Number	1	2	3	4	5
Flow rate of sample gas through analyser[a]	1.0	1.0	1.0	2.0	0.7
Carbon monoxide concentration used for input signal (ppm)	151	11.2	151	151	125
Carbon monoxide concentration for input signal (per cent of FSD)[b]	78.0	5.5	78.0	78.0	66.0
Type of input signal[c]	PRBS	PRBS	SS	PRBS	PRBS

a In litres min^{-1}

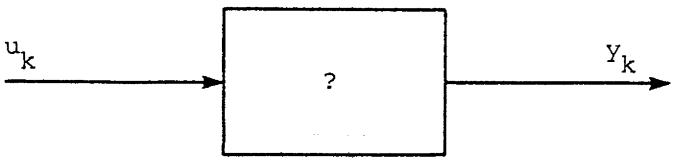
b FSD is full scale deflection

c PRBS is psuedo random binary signal, and SS is single step

4.3 Identification of Model Order and Estimation of Model Parameters

The problem of model order identification can be represented

diagrammatically as follows



where we utilise knowledge of the input u_k and the recorded output y_k to deduce the dynamic structure of the analyser. If both u_k and y_k are measured exactly then the identification problem is relatively straightforward. But we know that the output of continuous air pollution analysers is noisy and the noise level usually is quoted by manufacturers as being of the order of one per cent of the full scale deflection. This level of noise is not particularly serious. In a previously mentioned study Schnelle and Neeley (1972) were able to ignore the noise on the output signal and simply to deduce that the characteristic 'S' shaped curve of the single step response indicated a second order system with dead time.

However, we propose a discrete time linear dynamic stochastic model for the analyser so that the noise on the output can be accounted for in an explicit way. We express this model as

$$y_k = x_k + \xi_k$$

$$x_k = \frac{B_m(z^{-1})}{A_n(z^{-1})} u_{k-\tau} = \frac{b_o + b_1 z^{-1} + \ldots + b_m z^{-m}}{1 + a_1 z^{-1} + \ldots + a_n z^{-n}} u_{k-\tau} \qquad (4.1)$$

where y_k is the noisy output at time k, ξ_k is some stochastic disturbance, x_k is the 'noise free' output, $u_{k-\tau}$ is the input, and τ is a time delay integer. The symbol τ is introduced for convenience since the analyser exhibits a pure time delay between input and observable response. The identification problem is then simply one of determining those values of the integers n, m and τ in equation (4.1) which 'best' describe the dynamic properties of the analyser.

For the purpose of determining these integer values we propose to

use the instrumental variable (IV) method of Young et al. (1980) which was introduced in Chapter 3. In this method a range of plausible model structures is investigated by evaluation of a number of test statistics which arise naturally as part of the estimation process. Basic IV procedures are shown to be adequate for the model identification but refined IV methods are used for the final estimation of the parameters in the estimated models, since the refined IV estimates have superior statistical properties.

The primary statistics we use here for model order identification, each of which has been defined in Chapter 3, are the coefficient of determination R_T^2, and the normalised error variance norm (NEVN) associated with the P matrix obtained from the basic IV estimation. We use the normalised version of the EVN in conjunction with the usual EVN because the magnitude of the 'b' parameters is found in some models to be significantly different from that of the 'a' parameters. In such cases the weighting of the average parameter variances provided by the normalised EVN gives a more discerning criterion than the usual EVN. However, care must be exercised in the use of the normalised EVN. For example, if the incorrect time delay has been chosen then the minimum normalised EVN may indicate a model in which the fit to the output is poor. However, this will normally be detected by the poor R_T^2.

The test statistics for a range of model structures are shown in Table 4.2 for experiment 1. In searching for the most suitable model we look for that one which yields the best combination of NEVN and R_T^2. The NEVN should attain or be close to its minimum value and should increase substantially for further increase in model order. The value of R_T^2 should be consistent with the degree of model fit expected. In the present case we would also expect R_T^2 to be close to unity, partly because of the low level of noise on the observed data, and partly because the analyser is engineered to have 'nice' properties and we are able to perform well-controlled experiments on it. Furthermore, the value of R_T^2 should reach a 'plateau' level and thereafter not increase significantly with further increases in model order.

Young et al. (1980) have suggested further stages in their model identification procedure. In the present case we could bypass these

because we have been able to find models which provide very good explanations of the output data, and also because we have a priori information about the analyser which leads us to expect it to exhibit both linear and stationary behaviour. For the sake of completeness, however, we will describe these further stages, and in so doing we are able to assist the discrimination between the alternative models in Table 4.2. First, we need to ensure that the recursive estimates of the model parameters converge in a well-defined and non-ambiguous manner. Second, it is necessary that the estimated noise terms $\hat{\xi}_k$ are purely stochastic and are statistically independent of the input terms u_k. That is, the $\hat{\xi}_k$ should not be correlated with the estimated noise free output $\hat{x}_k$ or with the input u_k.

Table 4.2

Model Order Identification Criteria for Experiment 1

Model (n,m,τ)	R_T^2	ln(NEVN)	ln(EVN)
(1,4,6)	0.9877	-6.4379	-9.4655
(1,6,6)	0.9970	-7.0839	-10.5210
(1,7,6)	0.9978	-6.9645	-10.6012
(1,8,6)	0.9984	-6.8701	-10.7145
(1,9,6)	0.9990	-7.1640	-10.9443
(1,10,6)	0.9990	-6.8400	-10.6701
(2,1,6)	0.9831	-6.0625	-6.1453
(2,2,6)	0.9969	-7.0465	-7.4960
(2,3,6)	0.9963	-5.6637	-6.3374
(2,4.6)	0.9970	-5.3328	-5.8034
(2,5,6)	0.9979	-4.8891	-5.3829
(2,6,6)	0.9979	-4.0648	-4.7066
(1,10,5)	0.9992	-7.2634	-11.1252
(2,3,5)	0.9940	-6.9357	-7.7724

It can be seen in Table 4.2 that several models provide a good description of the analyser's output through having values of R_T^2 at the 'plateau' level, while at the same time having NEVN values close to the minimum. Since the levels of measurement noise are quite low, any one of these models would be adequate for our ultimate purpose of estimating the true input pollutant concentrations. However, consideration of the nature of the recursive parameter estimates, the time variation of the parameter estimates and the cross-correlation between the $\hat{\xi}_k$ and u_k leads us to conclude that the (2,2,6) model is the most acceptable. Thus the (1,9,6) model is rejected in favour of the (2,2,6) model even though the former has slightly better parameter definition. This is satisfying since a simpler model has obvious advantages in analytical terms. The recursive estimates of the parameters in the (2,2,6) model are shown in Figure 4.4 and can be seen to converge satisfactorily. The time variation of the parameter estimates is shown in Figure 4.5 and it is clear that the model parameters are time invariant as anticipated.† The cross correlation function for $\hat{\xi}_k$ and u_k is shown in Figure 4.6. Clearly the $\hat{\xi}_k$ satisfy requirements with the cross correlation always smaller than the confidence bands on the correlation function.

Finally, we estimate the parameters of the identified (2,2,6) model by use of refined IV procedures and these are shown in Table 4.3. The standard errors of the parameter estimates are acceptably low and the steady state gain is found to be 1.0400. This is a necessary condition for a satisfactory model of the analyser since we know from physical considerations that all carbon monoxide molecules that pass into the analyser must eventually emerge. The model fit to the observed output for experiment 1 is shown in Figure 4.7 together with the estimated terms $\hat{\xi}_k$.

4.4 Dynamic Check of the Linearity of the Analyser

The parameters of the identified (2,2,6) model in experiment 1 were time invariant, and this would normally be adequate to satisfy any doubt that the analyser could be modeled as a linear dynamic system. However, in order to provide a dynamic check of the linearity of the analyser, experiment 2 was performed. This experiment was identical in all respects to experiment 1 except that the magnitude of the input signal was reduced from 78.0 per cent of full scale deflection to 5.5

† the results in Fig 4.4 are just the recursive parameter estimates obtained for one iteration with suitable initialisation using equations (3.34) - (3.36). Those in Fig 4.5 are those estimates when the added assumption is made that the parameters can vary with time as

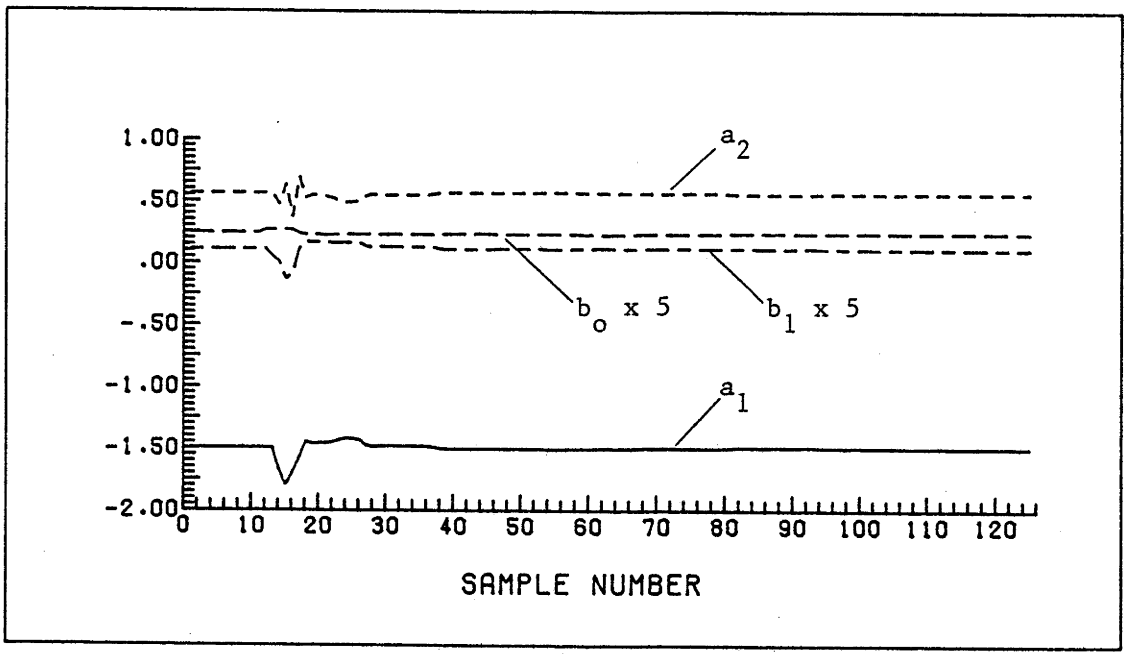


Figure 4.4 Recursive estimates of the parameters in the (2,2,6) model of the carbon monoxide analyser in experiment 1.

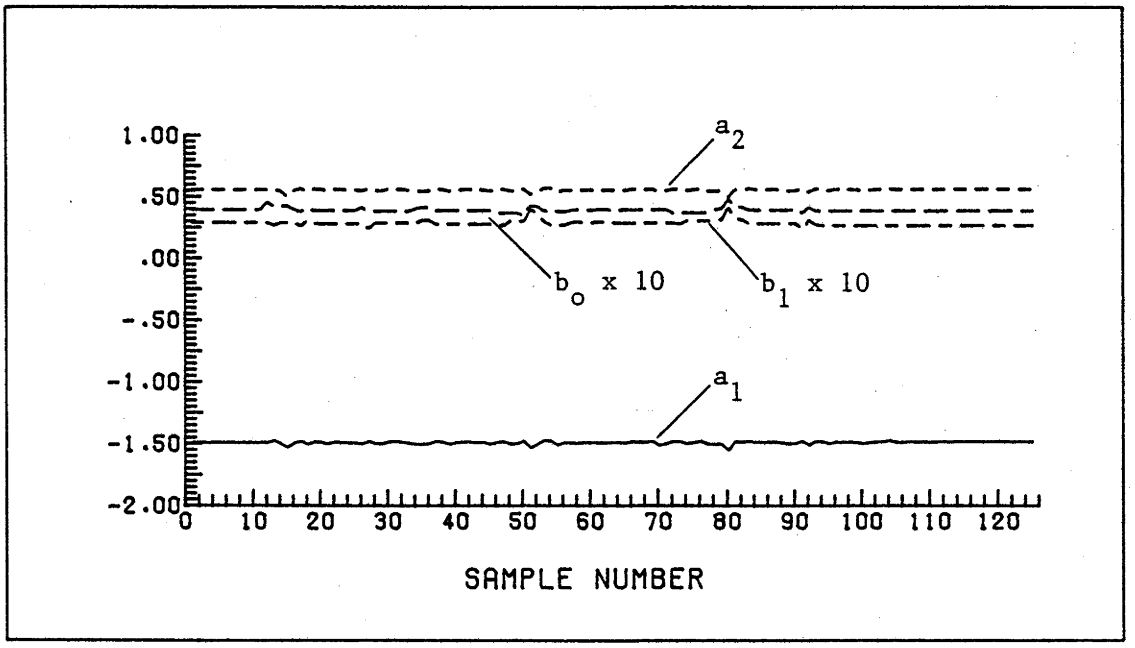


Figure 4.5 Estimated time variation of the parameters in the (2,2,6) model of the carbon monoxide analyser in experiment 1.

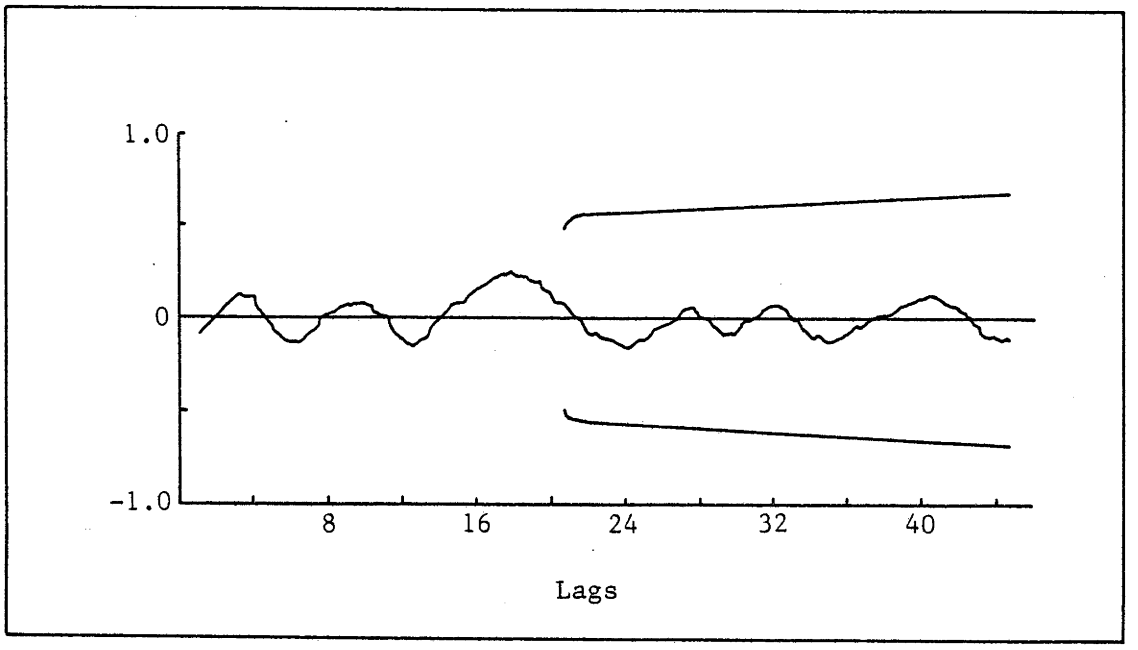


Figure 4.6 Cross-correlation function for the estimated residuals $\hat{\xi}_k$ and the input u_k from the (2,2,6) model in experiment 1.

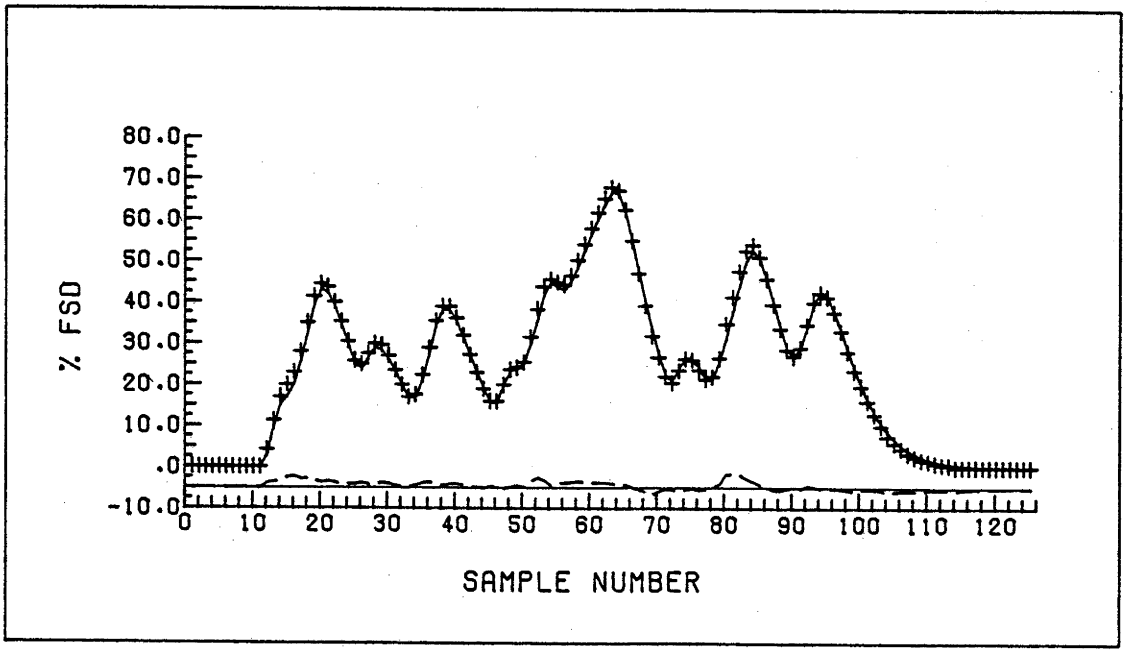


Figure 4.7 Model fit (—) to output measurements (+) provided by the (2,2,6) model of the carbon monoxide analyser in experiment 1. Estimated residuals $\hat{\xi}_k$ (---) are plotted underneath.

Table 4.3

Refined IV Parameter Estimates and Standard Errors for the Identified (2,2,6) Model for Experiment 1

Parameters In Identified Model	Estimated Parameter Values	Standard Errors of Parameter Estimates
a_1	-1.49619	0.00640
a_2	0.56415	0.00558
b_0	0.04872	0.00118
b_1	0.02196	0.00191

per cent. The model identification procedure described in the previous section was applied to experiment 2 and the primary test statistics for a range of plausible models are shown in Table 4.4. From inspection of the results we conclude that the (2,2,6) model has the most acceptable model structure. This is the same structure identified for experiment 1 and thus provides further evidence that the analyser behaves as a linear dynamic system. Comparison of the test statistics in Tables 4.2 and 4.4 shows that neither the fit to the data nor the parameter definition are quite as good as in experiment 1. This is consistent with the knowledge that the level of measurement noise was considerably larger in experiment 2. As with experiment 1, the parameters in the identified (2,2,6) model for experiment 2 are estimated finally by refined IV methods, the estimates and their standard errors being shown in Table 4.5. Comparison of Tables 4.3 and 4.5 shows that for experiment 2 the standard errors are three to four times larger than those found for experiment 1.

The model fit to the observed data in experiment 2 is shown in Figure 4.8. As a final check on the suitability of the model

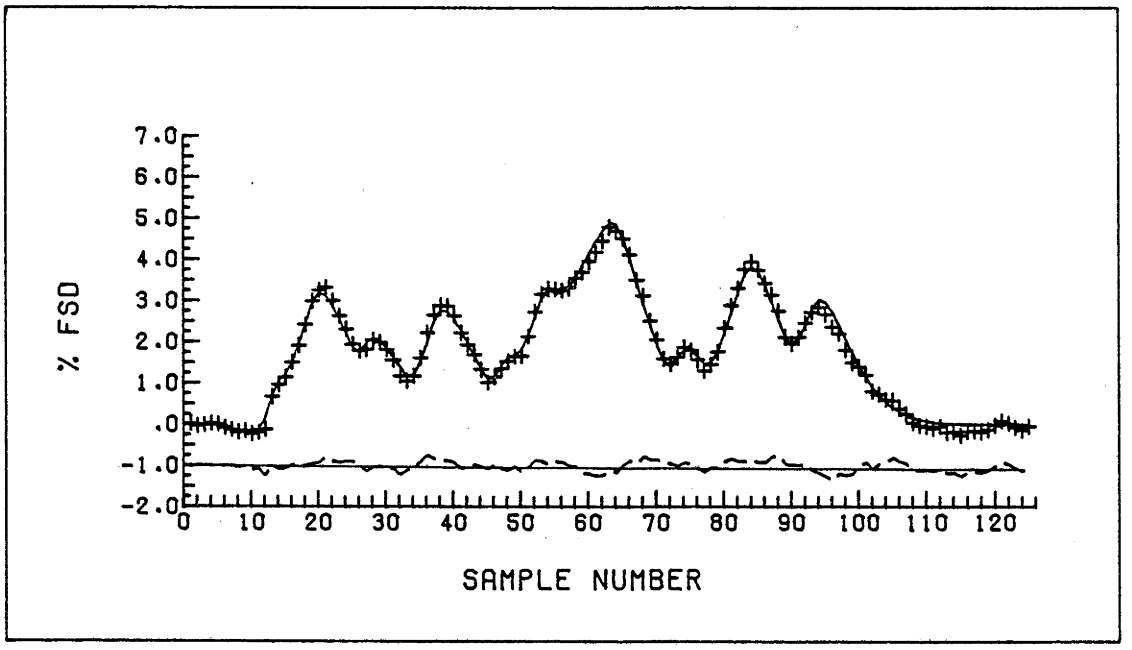


Figure 4.8 Model fit (——) to output measurements (+) provided by the (2,2,6) model of the carbon monoxide analyser in experiment 2. Estimated residuals $\hat{\xi}_k$ are plotted underneath.

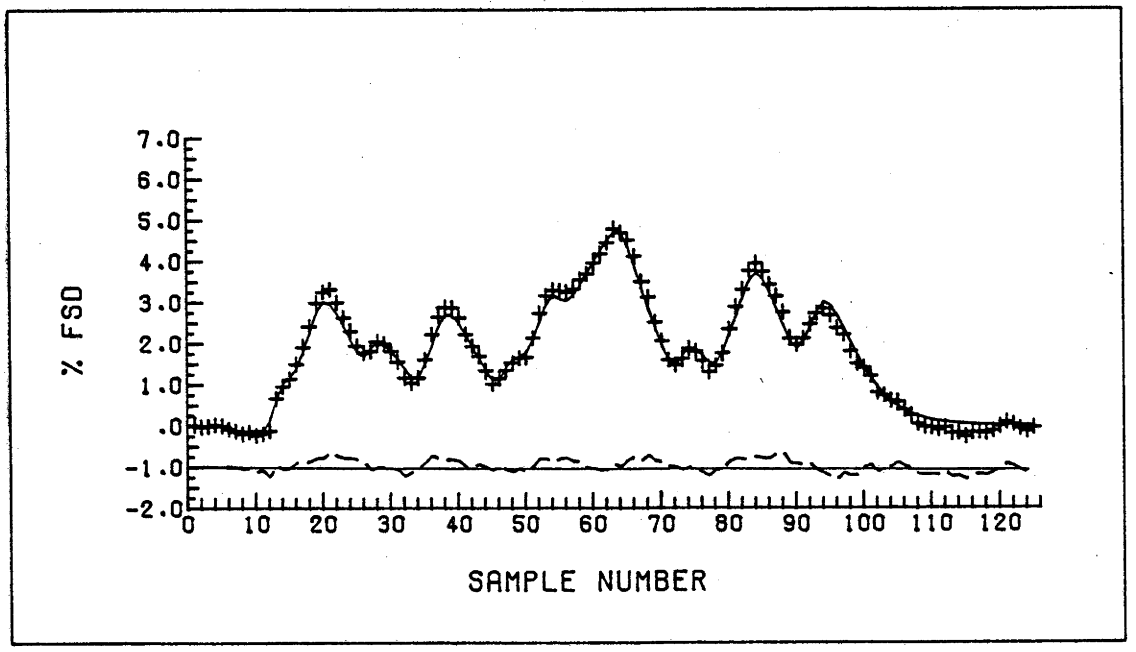


Figure 4.9 Comparison of observed output (+) in experiment 2 with the noise free output (——) calculated from the (2,2,6) model estimated in experiment 1.

parameters found in experiment 1 we calculate the noise free output by passing the input for experiment 2 into the model obtained in experiment 1, then compare this calculated output with the observed output in experiment 2. This comparison is shown in Figure 4.9, clearly indicating that the model estimated in experiment 1 provides an adequate description of the data in experiment 2. This final step in the analysis can be considered as an evaluation of the validity of the model of the analyser since it shows the ability of the model to explain data other than that on which its parameter estimation was based.

Table 4.4

Model Order Identification Criteria for Experiment 2

Model (n,m,τ)	R_T^2	ln(NEVN)	ln(EVN)
(1,6,6)	0.9859	-5.7149	-9.1061
(1,7,6)	0.9878	-5.6101	-9.0710
(1,8,6)	0.9899	-5.2724	-8.9882
(1,9,6)	0.9909	-5.5537	-8.8533
(2,1,6)	0.9841	-5.9308	-6.0134
(2,2,6)	0.9918	-5.8646	-6.3365
(2,3,6)	0.9891	-4.7767	-5.4324
(2,4,6)	0.9893	-4.0325	-4.6004
(2,5,6)	0.9918	-4.2484	-4.7488

4.5 Choice of Input Signals for Optimal Estimation

We would expect that relatively simple input signals would provide adequate information in the output signal for identification and estimation of the model parameters because the dynamic behaviour of the carbon monoxide analyser is not particularly complex. We therefore compare results obtained in experiments 1 and 3 as the only difference

Table 4.5

Refined IV Parameter Estimates and Standard Errors for the Identified (2,2,6) Model for Experiment 2

Parameters In Identified Model	Estimated Parameter Values	Standard Errors of Parameter Estimates
a_1	-1.57088	0.01742
a_2	0.63630	0.01503
b_0	0.04581	0.00454
b_1	0.02312	0.00670

between the two was the nature of the input signal used, namely, a random binary signal (RBS) in experiment 1 and a single step signal in experiment 3. These two input signals are shown in Figure 4.10. It is expected that use of the random binary input signal will lead to better definition of the parameter estimates. This is because adequate excitation of the system exposes, via the output response, all of the important modes of the system behaviour (Goodwin and Payne, 1977).

The primary test statistics for various plausible model structures for experiment 3 are shown in Table 4.6, and inspection of these results indicates the (2,1,6) model as the most suitable. Such a structure is only slightly different from that identified for experiment 1, since it is the 'a' parameters that are the major determinants of the dynamic behaviour of the model. The model fit to the step response of experiment 3 is shown in Figure 4.11. The refined IV estimates of the parameters in the (2,1,6) model are shown in Table 4.7 and comparison of the standard errors with those in Table 4.3 shows that a slight improvement may be obtained by use of the RBS input. For this reason we choose a RBS input in all other experiments.

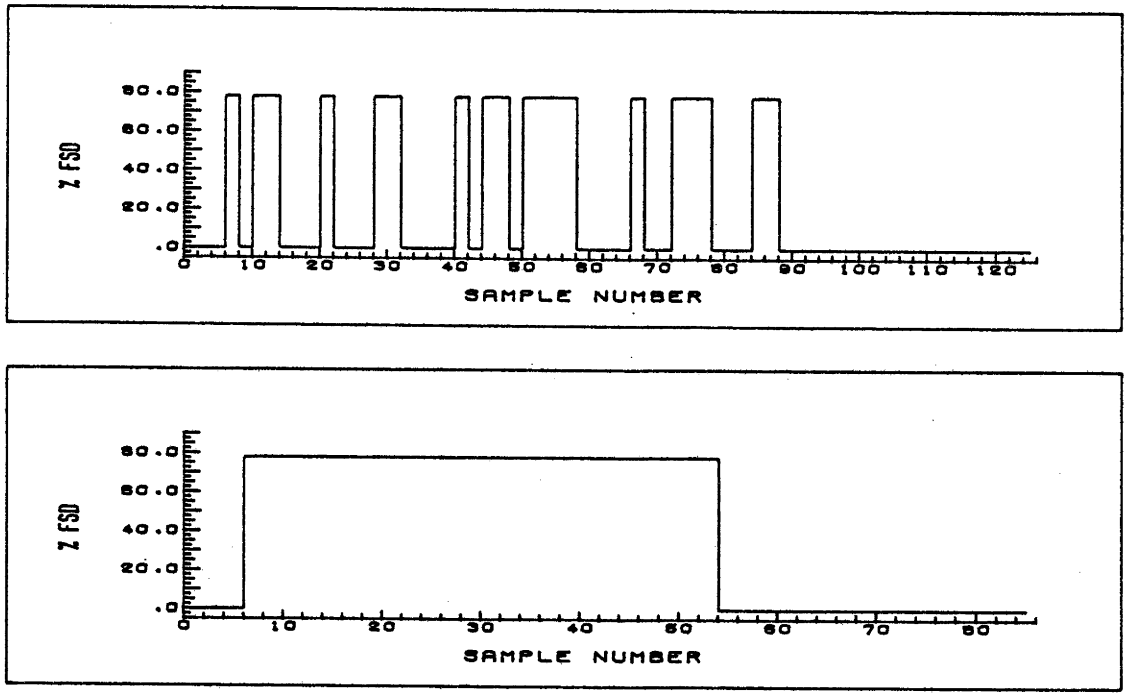


Figure 4.10 Pseudo random binary signal (PRBS) and single step inputs used for experiments on the carbon monoxide analyser.

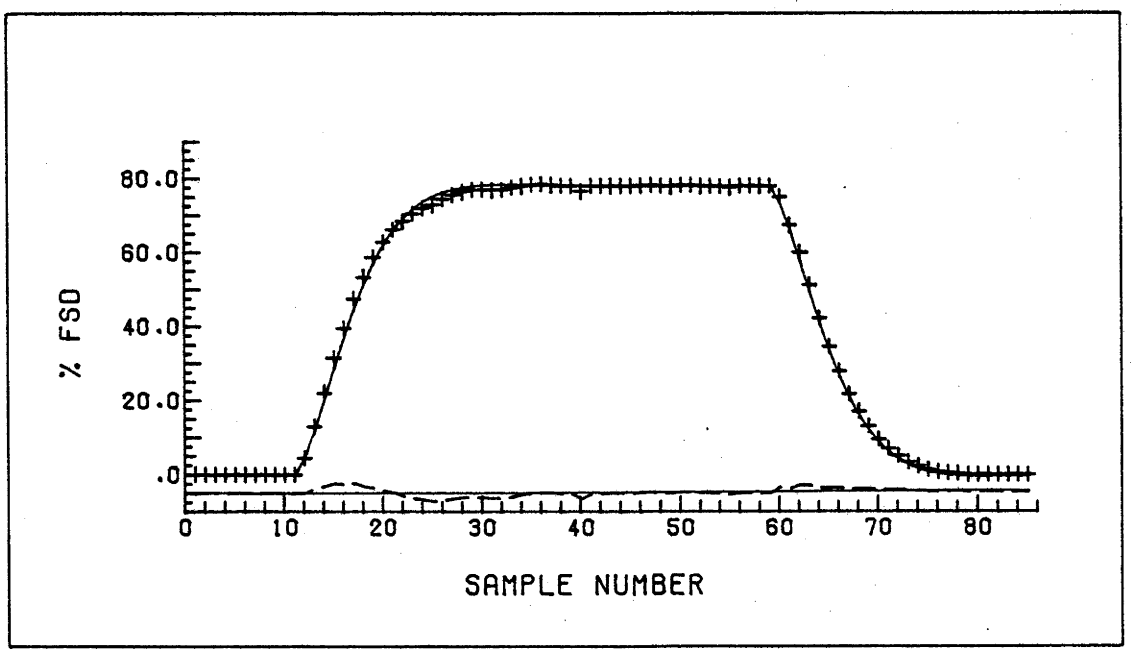


Figure 4.11 Model fit (——) to output measurements (+) provided by the (2,1,6) model of the carbon monoxide analyser in experiment 3. Estimated residuals $\hat{\xi}_k$ (---) are plotted underneath.

Table 4.6

Model Order Identification Criteria for Experiment 3

Model (n,m,τ)	R_T^2	ln(NEVN)	ln(EVN)
(1,1,6)	0.9882	-6.7063	-8.1189
(1,2,6)	0.9957	-4.7398	-7.4314
(1,3,6)	0.9974	-4.7096	-7.5127
(1,4,6)	0.9986	-4.5396	-7.9216
(1,5,6)	0.9987	-4.5469	-7.9348
(1,6,6)	0.9989	-4.5739	-8.0429
(2,1,6)	0.9991	-5.7377	-5.9266
(2,2,6)	0.9964	-3.6795	-4.2918
(2,3,6)	0.9954	-2.7637	-3.3841
(2,4,6)	0.9984	-3.3611	-4.0216

Table 4.7

Refined IV Parameter Estimates and Standard Errors for the Identified (2,1,6) Model for Experiment 3

Parameters In Identified Model	Estimated Parameter Values	Standard Errors of Parameter Estimates
a_1	-1.53634	0.00770
a_2	0.59824	0.00677
b_0	0.06207	0.00102

4.6 Effect of Sample Gas Flow Rate on the Dynamics of the Analyser

Both Larsen et al. (1965) and Schnelle and Neeley (1972) showed that the dynamic response of continuous air pollution analysers could be altered by changing the flow rate of the sample gas through the detection cell. In this section we describe experiments 4 and 5 in which the sample gas flow rate was altered from the recommended 1.0 litre min^{-1} used in experiments 1 to 3. In experiment 4 the flow rate was increased to 2.0 litres min^{-1} while in experiment 5 it was decreased to 0.7 litres min^{-1}. The primary test statistics for model order identification are shown in Table 4.8 for experiment 4 and Table 4.9 for experiment 5. Inspection of these shows a (2,2,4) model for experiment 4 and a (2,3,8) model for experiment 5 as the preferred models. As we would expect from physical considerations, the pure time delay between input stimulus and observable response is related to the sample gas flow rate, with shorter delay times at higher flow rates. The refined IV parameter estimates for experiments 4 and 5 are shown in Table 4.10. In neither experiment 4 nor 5 are the standard errors of the parameters as low as those obtained in experiment 1. It is not apparent why this has occurred, but it may be due to slight ambiguities in the specification of the pure time delay. Thus the true time delay may lie between samples rather than being an exact integer multiple of the sampling interval. We will consider the results of experiments 4 and 5 in more detail in the next section where we describe continuous time transfer function models and the frequency response characteristics of the analyser.

4.7 Continuous Time Models and the Analyser's Frequency Response

Continuous air pollution analysers may produce a continuous chart recording of pollutant concentration, although if they are used in conjunction with modern data logging equipment the continuous output is usually sampled at some fixed frequency. Since these analysers may generate a continuous output it seems appropriate to find continuous time models to describe their dynamic behaviour. To do this it is convenient to use sampled data to identify discrete time models and from them to obtain continuous time models. This is done simply by factoring the polynomials in z^{-1} of the discrete time models into factors of the form $(1 + \alpha z^{-1})$. These factors then transform to

Table 4.8

Model Order Identification Criteria for Experiment 4

Model (n,m,τ)	R_T^2	ln(NEVN)	ln(EVN)
(0,9,4)	0.9920	-7.0012	-9.3703
(0,10,4)	0.9948	-7.0834	-9.5887
(1,3,4)	0.9900	-7.3588	-9.6283
(1,4,4)	0.9951	-7.3557	-9.9662
(1,5,4)	0.9967	-7.2366	-10.1721
(1,6,4)	0.9975	-6.9872	-10.1073
(1,7,4)	0.9975	-5.2338	-9.6584
(1,8,4)	0.9977	-5.5685	-9.2465
(1,9,4)	0.9976	-5.4159	-8.5081
(1,10,4)	0.9976	-3.5174	-7.8475
(2,2,4)	0.9976	-7.3904	-8.0770
(2,3,4)	0.9971	-5.2002	-6.5896
(2,4,4)	0.9972	-3.0254	-5.3729
(2,3,3)	0.9971	-6.9780	-8.2409

Table 4.9

Model Order Identification Criteria for Experiment 5

Model (n,m,τ)	R_T^2	ln(NEVN)	ln(EVN)
(1,5,8)	0.9927	-7.1862	-10.2802
(1,6,8)	0.9967	-7.2513	-10.5231
(1,7,8)	0.9981	-7.4674	-10.8743
(1,8,8)	0.9985	-7.3433	-10.9292
(1,10,8)	0.9987	-5.8523	-10.5284
(2,3,8)	0.9978	-7.1826	-7.8867
(2,4,8)	0.9983	-6.7277	-7.4523
(2,5,8)	0.9973	-5.3846	-6.2356
(2,7,8)	0.9984	-3.9623	-4.7530

Table 4.10

Refined IV Parameter Estimates and Standard Errors for the Identified (2,2,4) Model for Experiment 4 and the (2,3,8) Model for Experiment 5

Parameters In Identified Model	Estimated Parameter Values	Standard Errors of Parameter Estimates
	Experiment 4	
a_1	-1.28312	0.01313
a_2	0.41644	0.01084
b_0	0.06749	0.00260
b_1	0.06795	0.00465
	Experiment 5	
a_1	-1.45502	0.00893
a_2	0.53241	0.00767
b_0	0.02847	0.00198
b_1	0.01148	0.00405
b_2	0.03899	0.00329

$(1 - \alpha)(s + 2/T((1 + \alpha)/(1 - \alpha)))$ where $s = \sigma + jw$ and T is the sampling interval (Caprihan and Neto, 1977). This transformation provides the initial estimates for the parameters in the continuous time models and the final estimates are obtained by use of the

continuous time modeling capacity of the CAPTAIN package (see Young and Jakeman, 1979b).

In this way continuous time models are obtained for experiments 1, 4 and 5 and the results are summarised in Table 4.11. In all three experiments the continuous time models provide good explanations of the output data and are second order with the form

$$\frac{g_o}{1 + f_1 s + f_2 s^2}$$

where g_o is the steady state gain K, f_1 is $2\zeta/\omega_n$ and f_2 is $1/\omega_n^2$. The term ζ is the damping ratio and ω_n is the natural frequency of oscillation. The continuous time model fit obtained for experiment 1 is shown in Figure 4.12 where for clarity of presentation the measured output and the model fit are separated vertically by two units. It is evident that the continuous time model provides as good an explanation of the data as the discrete time model (see Figure 4.7).

We have seen that a continuous air pollution analyser can be well described by either discrete time or continuous time linear dynamic models. Given that it is possible to summarise all the information about the dynamic behaviour of a linear dynamic system in a Bode diagram (see, for example, Naslin, 1965), it is surprising that manufacturers of continuous air pollution analysers do not utilise such diagrams. This is unfortunate because Bode diagrams are easy to construct once an analyser has been characterised by a transfer function and would thus enable easy comparison of the dynamic properties of different analysers. Schnelle and Neeley (1972) determine the frequency response characteristics of some continuous air pollution analysers and, predictably, found that the gain decreased as the frequency of the input signal increased. A desirable property of such analysers would be a 'flat' frequency response (that is, gain $\equiv$ 0) over the range of input frequencies most likely to be met in practice. In the next section it is shown how this can be achieved by means of a robust input estimation algorithm.

Table 4.11

Summary of Results for Continuous Time Models

Expt. No.	Flow Rate[a]	CO Conc.[b]	Parameters[c]	Estimates[d]	S.E.[e]	R_T^2	SSG[f]	ζ[g]	ω_n[h]
1	1.0	78.0	f_1	6.3078	0.0539	0.9996	1.0408	0.99	0.314
			f_2	10.1445	0.2412				
			g_o	1.0408	0.0032				
4	2.0	78.0	f_1	4.3135	0.0331	0.9989	1.0162	0.94	0.437
			f_2	5.2302	0.1177				
			g_o	1.0162	0.0032				
5	0.7	66.0	f_1	6.3238	0.0650	0.9979	1.0177	0.84	0.267
			f_2	14.0297	0.3062				
			g_o	1.0177	0.0039				

(a) Flow rate of sample gas through analyser, in litre min^{-1}. (b) Carbon monoxide concentration used for input signal; expressed as a percentage of full scale deflection. (c) Parameters in identified model. (d) Estimated parameter values. (e) Standard errors of parameter estimates. (f) Steady state gain. (g) Damping ratio. (h) Natural frequency in radians/time unit where time unit is 5s.

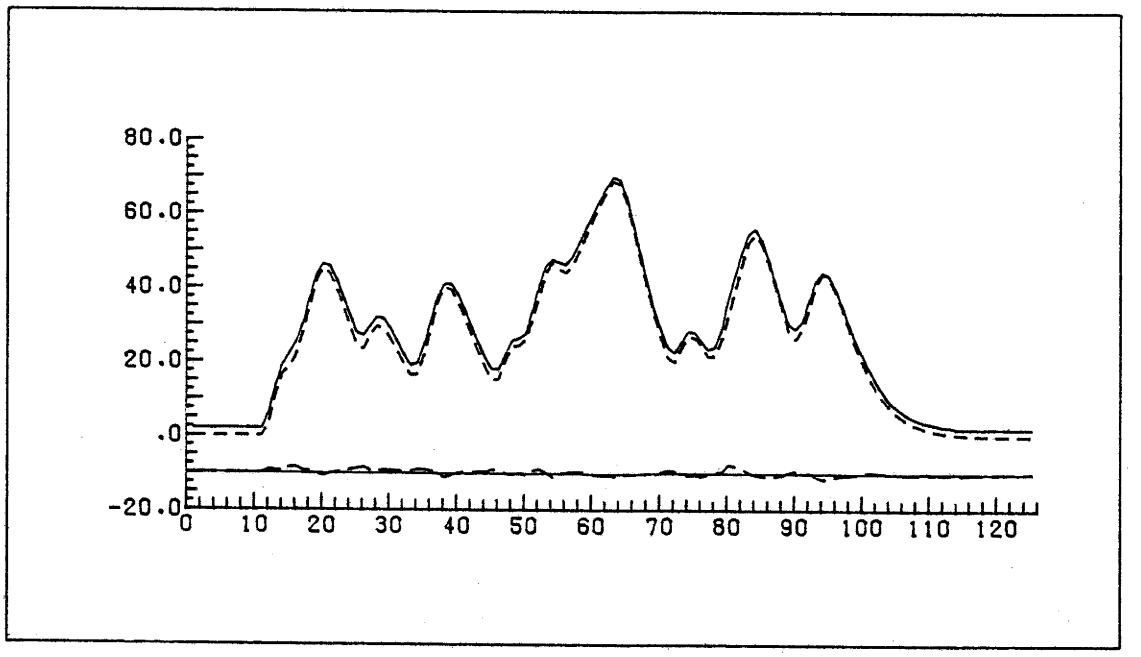


Figure 4.12 Continuous time model fit (---) to the observed output (——) in experiment 1. The observed output has been translated by +2 units for clarity.

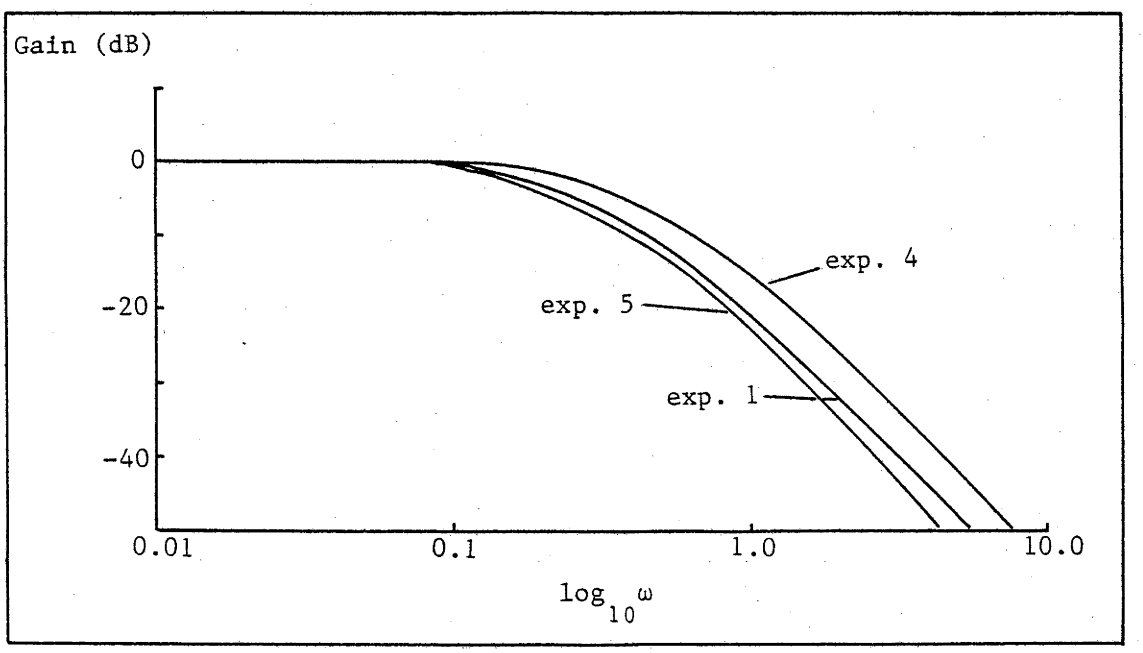
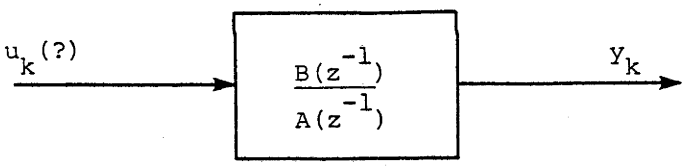


Figure 4.13 Gain frequency response of the carbon monoxide analyser in three experiments using different sample gas flow rates.

Before proceeding to describe this, the frequency response characteristics of the analyser are determined for experiments 1, 4 and 5. This is done simply by using the values found for the natural frequency of oscillation ω_n, the damping ratio ζ and the steady state gain K. Then substitution of $s = j\omega$ in the continuous time transfer function models allows the gain and the phase to be calculated for a range of frequencies ω (see Truxal, 1955). The former (determined from the values for natural frequency, damping ratio and steady state gain shown in Table 4.11) is shown for experiments 1, 4 and 5 in Figure 4.13. The results illustrated in this Figure are in accord with the physical nature of the analyser, where an increase in sample gas flow rate tends to extend the bandwith of the analyser a little. It can also be noted that the analyser has critically damped response (ζ = 1.0) at the 'design' flow rate of 1.0 litre min^{-1}.

4.8 Input Estimation

In general the input estimation problem can be represented diagrammatically as

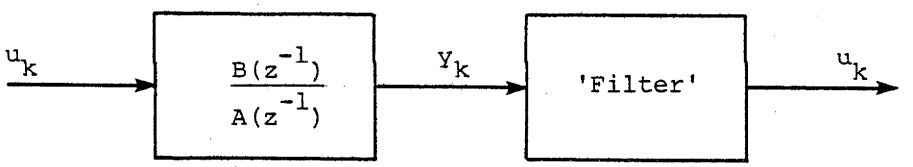


where the aim is to use knowledge of the output y_k and the model structure $B(z^{-1})/A(z^{-1})$ to estimate the input u_k. Finite difference methods may successfully be used to do this if the output is measured exactly but may be unsatisfactory if y_k is measured in error. This is demonstrated in the Appendix where the discrete time model of the carbon monoxide analyser estimated in experiment 1 is used as the description of the analyser.

Another method for input estimation is that described by Young and Jakeman (1979a). This first requires the transformation of the transfer function model of the dynamic system so that $x_k = (B(z^{-1})/A(z^{-1}))u_k$ is inverted to give $u_k = (A(z^{-1})/B(z^{-1}))x_k$. This method proves useful when there is little noise on the data and when $B(z^{-1})$ is invertible (minimum phase). The latter condition requires that the roots of the polynomial $B(z^{-1})$, in terms of z^{-1}, lie outside the unit circle so that the

convergence of $(A(z^{-1})/B(z^{-1}))^{-1}x_k$ is assured. If this condition is not satisfied the inverse operation is unstable and the solution u_k is unbounded (Jakeman and Young, 1980a). Even if $B(z^{-1})$ is not invertible it may be possible to approximate the non-invertible linear operator by an invertible one with similar dynamic properties (Jakeman and Young, 1980a). A method of input estimation can, however, be chosen that bypasses the required invertibility of $B(z^{-1})$ and which is robust to errors on the output y_k. It is this method that is now described.

This method is essentially equivalent to synthesising a filter for the output data, where the frequency response characteristics of the filter are chosen such that the series connection of system model and the filter approximates a 'unity' transfer function; in other words, even though no actual filter is formulated, the method can be interpreted in terms of such a filter which may be thought of as an approximation to the inverse of the transfer function model of the analyser, with due allowance made for the attenuation of noise effects. This can be represented diagrammatically as



where if the 'filter' can be chosen properly, the output from the filter should be virtually equivalent to the input u_k.

The decision to model the input u_k as a first-order Gauss-Markov process is the basis of the method used here (Young and Jakeman, 1979a). The process may be written as

$$u_k = \Phi u_{k-1} + \Gamma v_k \tag{4.2}$$

where u_k may be a vector consisting of u_k appended with other states, Φ is a transition matrix, Γ is a partitioned matrix comprising null and identity submatrices, v_k ($k = 1, \ldots, N$) is an independent and identically distributed sequence of random vectors with zero mean and covariance matrix Q, and N is the number of samples. The

simplest example of such a process is the simple random walk where Φ in (4.2) is the identity matrix of order unity, so that we have

$$u_k = u_{k-1} + v_k \tag{4.3}$$

and the covariance matrix Q becomes a scalar denoted by q. The input variation may be described in other simple ways and one that is used later in this chapter is the integrated random walk

$$\begin{aligned} u_k &= u_{k-1} + s_{k-1} \\ s_k &= s_{k-1} + v_{k-1} \end{aligned} \tag{4.4}$$

which is also a special form of the general linear Gauss-Markov form (4.2).

However, to demonstrate the simplicity of the input estimation algorithms we will derive them for the case of the simple random walk model (4.3) of the input variation. Then by using a stochastic observation equation encompassing equation (4.1) which is placed in a special transfer function form, the input estimation problem can be formulated as one of estimating the single state u_k of a stochastic dynamic system described by

$$\begin{aligned} u_k &= u_{k-1} + v_k \\ y'_k &= \underline{b}^T \underline{u}_k + e_k \end{aligned} \tag{4.5}$$

where e_k denotes a white measurement error sequence (with constant variance σ^2) and the following definitions apply

$$\underline{u}_k \triangleq (u_k, \ldots, u_{k-m})^T$$

$$y'_k \triangleq y_k - a_1 y_{k-1} - \ldots - a_n y_{k-n}$$

$$\underline{b}^T \triangleq (b_o \; b_1 \ldots b_m)$$

The equations (4.5) may be solved for u_k by either a filtering or a smoothing algorithm, where the former provides an estimate of the state u_k at the time of measurement of the kth sample, and the latter generates estimates of u_k at <u>each</u> sampling instant which are based on <u>all</u> N samples (see Gelb, 1974). The filtering algorithm corresponds directly to the Kalman filter of state estimation theory (see Gelb, 1974) and may be written as

$$\begin{aligned} \hat{u}_k &= \hat{u}_{k-1} + p_k b_o (y_k^* - \underline{b}^T \hat{\underline{u}}_{k-1}^T) \\ p_k &= (p_{k-1} + q)\{ 1 - \frac{b_o^2 (p_{k-1} + q)}{\sigma^2 + b_o^2 (p_{k-1} + q)} \} \end{aligned} \tag{4.6}$$

A slightly modified form of the algorithm (4.6) which avoids the necessity of providing an estimate of σ^2 may be derived simply by dividing through the equation for p_k by σ^2 and rearranging the term within the curly brackets to give

$$\bar{p}_k = (\bar{p}_{k-1} + \bar{q})\{ 1 - \frac{b_o^2 (\bar{p}_{k-1} + \bar{q})}{1 + b_o^2 (\bar{p}_{k-1} + \bar{q})} \} \tag{4.7}$$

where $\bar{p}_k = p_k / \sigma^2$ and $\bar{q} = q / \sigma^2$. In the next section such a simplified form of the filtering algorithm is used. The filtering algorithm offers on-line potential and, as will be shown, it provides input estimates which are sufficiently accurate in most cases. Implementation of the algorithm requires u_o to be set to zero or to some <u>a priori</u> value, and p_o (or $\bar{p}_o$) to a high value such as 10^6. It is not possible to specify a simple equation for estimation of the value of q which corresponds to the level of variation in the input signal. While Kaldor (1978) gives a clear review of available heuristic procedures for estimating q from the data, it is not considered necessary to utilise these procedures in this case. It is clear from the properties of the simple random walk that a value of the variance q which is too large will force the estimates of the input u_k to be very 'spiky', while a value of q which is too small will not allow the u_k to change quickly enough to the correct level. Usually q lies between 10^{-4} and 10^2 and experience has shown that an order

of magnitude estimate of q is usually sufficient to provide good input estimates; in other words q (or $\bar{q}$) is used as a 'program parameter' to be adjusted by a trial and error procedure.

Unlike the filtering solution for equations of the form (4.5), the smoothing solution may have many forms, some of which are unstable (see Norton, 1975). The solution preferred here is that derived by Bryson and Ho (1975) since it provides smoothed estimates by utilising the filtered estimates and then working recursively backwards through the data with a recursion that is stable. It may be written as

$$\hat{a}_{k/N} = \hat{a}_{k+1/N} + q\lambda_k$$

$$\text{where} \quad \lambda_N = 0 \tag{4.8}$$

$$\text{and} \quad \lambda_k = \left(1 - \frac{p_{k+1} b_o^2}{\sigma^2}\right)\left(\lambda_{k+1} - \frac{b_o}{\sigma^2}(y_k^* - \underline{b}^T \underline{\hat{a}}_k)\right)$$

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

where the variance σ^2 of the white noise sequence may be estimated from the filtered estimates obtained at each sampling instant. Alternatively, estimation of σ^2 may be avoided entirely by a scaling similar to that described for the filtering algorithm. In the next section we demonstrate both filtering and smoothing algorithms on two sets of data, for which it was found that the integrated random walk model of the input variation was more suitable than the simple random walk.

4.9 Demonstration of the Input Estimation Algorithms

Both sets of data used to demonstrate the performance of the input estimation algorithms consist of 720 data points, that is, one hour of continuous measurements sampled every five seconds. The first set of data was taken from a chart recording of carbon monoxide concentrations in Canberra City and the second was a synthetic data set constructed to have high frequency components. The former is believed to be representative of air quality data normally met in practice, whereas the latter was constructed primarily to provide a difficult

test for the input estimation algorithms. The former data set will be considered first.

Even though the data taken from the chart recording are not the true input pollutant concentrations, it will be sufficient for our purposes to assume that they represent the true input signal. The input is then passed through the discrete time model of the carbon monoxide analyser estimated in experiment 1, so that the noise free output is obtained. White noise of a level normally found in such analysers (about one percent of full scale deflection) is then added to this noise free output to give a noisy output. This is shown in Figure 4.14 where for clarity of presentation the noise free output is displaced vertically by five units and the noise is plotted directly underneath. A plot of the known input and the noisy output is shown in Figure 4.15, where again for clarity of presentation the known input is displaced vertically by 15 units, and their difference is plotted underneath.

The filtering algorithm is initiated in the manner described in the previous section and various values of q tried in order to determine that most suitable. The estimated input using $\bar{q} = 0.1$ is shown with the known input in Figure 4.16, where two features are apparent. First, the estimated input closely 'tracks' the known input although the former is more 'spiky' than the latter. Second, there is a phase lag between the estimated and known inputs. The 'spiky' nature of the input estimates may be attributed to the value of q chosen. Smaller values of $\bar{q}$ would produce less 'spiky' input estimates but only at the cost of poorer tracking of the true input variation. The phase lag between the estimated and known inputs arises because of the dynamic lag between the input and output of the carbon monoxide analyser. The filtering algorithm does not correct for this lag, but this is not a serious drawback in estimating true pollutant concentrations, since it will mean only a slight error in the estimated time of occurrence of the maximum instantaneous pollutant concentration.

The smoothing algorithm is applied to the same noisy data as above and the results obtained using $\bar{q} = 50$ are shown in Figure 4.17. The known input is displaced vertically by two units and the difference between the known and the estimated smoothed input is

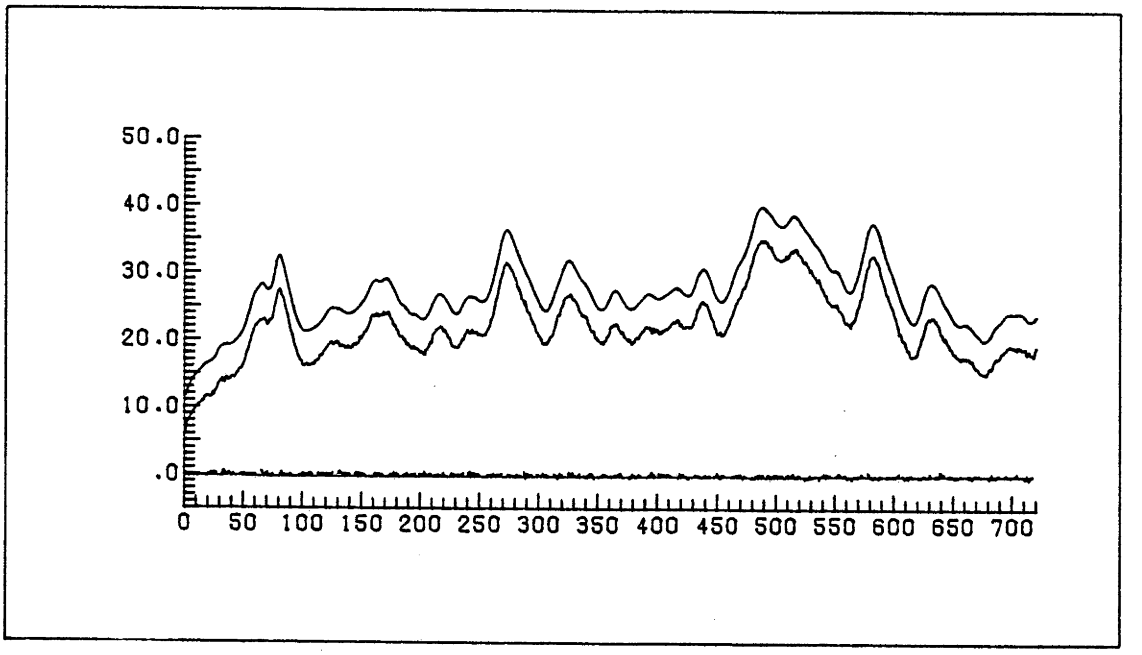


Figure 4.14 The noise free output of the carbon monoxide analyser (upper graph) resulting from the first known input. The noisy output obtained by adding white noise to the noise free output is shown for comparison and is translated vertically by -10 units for clarity.

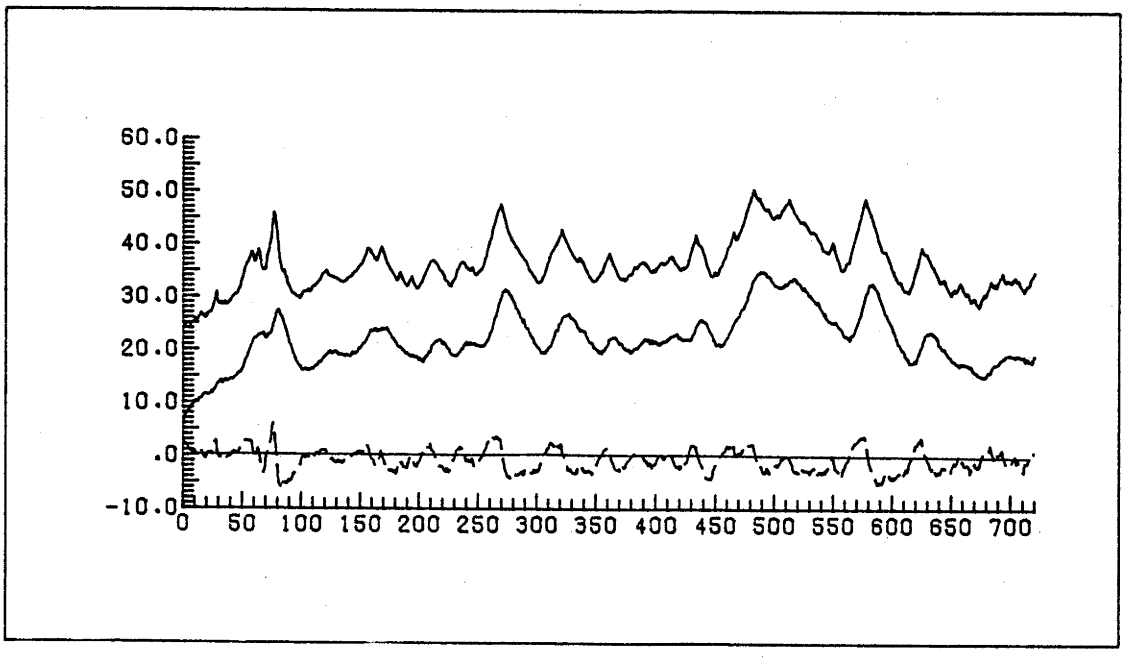


Figure 4.15 Comparison of the first known input (upper graph) with the noisy output of the analyser. The known input is translated vertically by +15 units for clarity, and the difference between the input and noisy output is plotted underneath.

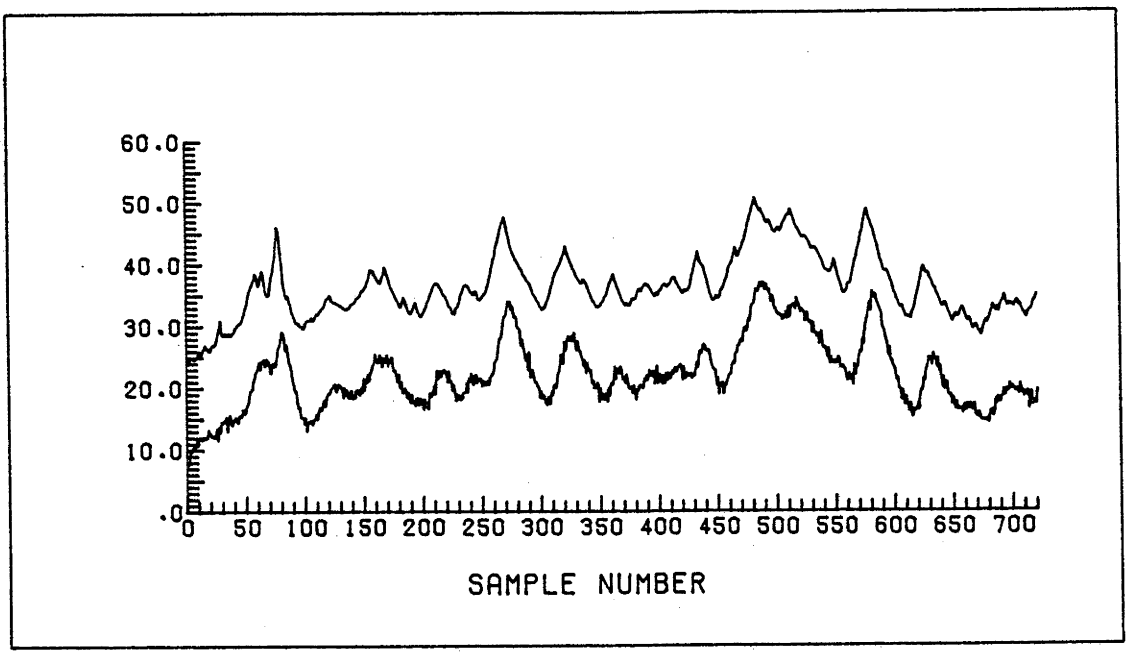


Figure 4.16 Comparison of the known first input (upper graph) and the estimate of the input provided by the filtering algorithm with $\bar{q} = 0.10$. The known input is translated vertically by +15 units for clarity.

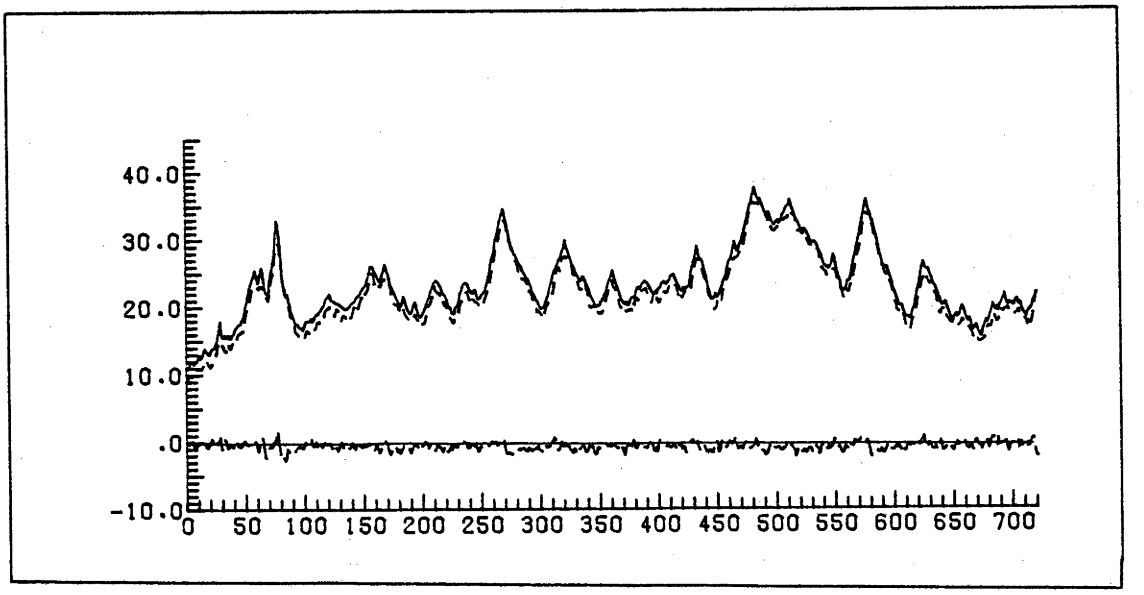


Figure 4.17 Comparison of the first known input (——) and the estimate (---) of the input provided by the smoothing algorithm with $\bar{q} = 50$. The known input is translated vertically by +2 units for clarity and the difference between known and estimated inputs is plotted underneath.

plotted underneath. Unlike the filtered estimates, the smooth estimates are not 'spiky' and there is no lag between the estimated and known inputs. Quite obviously if on-line estimates of the input are not required, the smoothing algorithm may be used and should provide input estimates that are superior to those obtained from the filtering algorithm.

For the second data set we proceed as before and pass the assumed known input through the same transfer function model of the analyser and add a suitable level of noise to the noise free output. The noise free and noisy outputs are shown in Figure 4.18. The noisy output and the known input are shown in Figure 4.19, and the attenuation of the input signal by the analyser is much more pronounced than before. It was expected that the value of $\bar{q}$ necessary to obtain good estimates in this case would be significantly larger than that used in the previous example. A value of $\bar{q} = 500$ was required and the estimated smoothed input is shown together with the known input in Figure 4.20.

A visual comparison of the known and the estimated inputs in this Figure indicates that the input estimation algorithms work very satisfactorily despite the considerable complexity of the noisy output signal. However, to provide some quantitative measure of the performance of the algorithms, a selection of statistics appropriate to the input and output signals were computed and are shown in Table 4.12. As would be expected from the dynamic properties of the analyser, the noise free and noisy output signals have a variance considerably less than that of the known input, but both the filtering and smoothing algorithms produce input estimates with variances that are restored to values close to that of the known input. Similarly the input estimation procedures yield reasonable estimates of the maximum and minimum values of the known input. The discrepancy between the mean values of the known input and the output signals is probably due to the fact that the steady state gain of the analyser was estimated in experiment number one to be 1.04. Thus it has not been possible to determine the extent to which the frequency response characteristics of the carbon monoxide analyser will cause errors in the measurement of mean pollutant concentrations which are averaged over periods of one hour or longer.

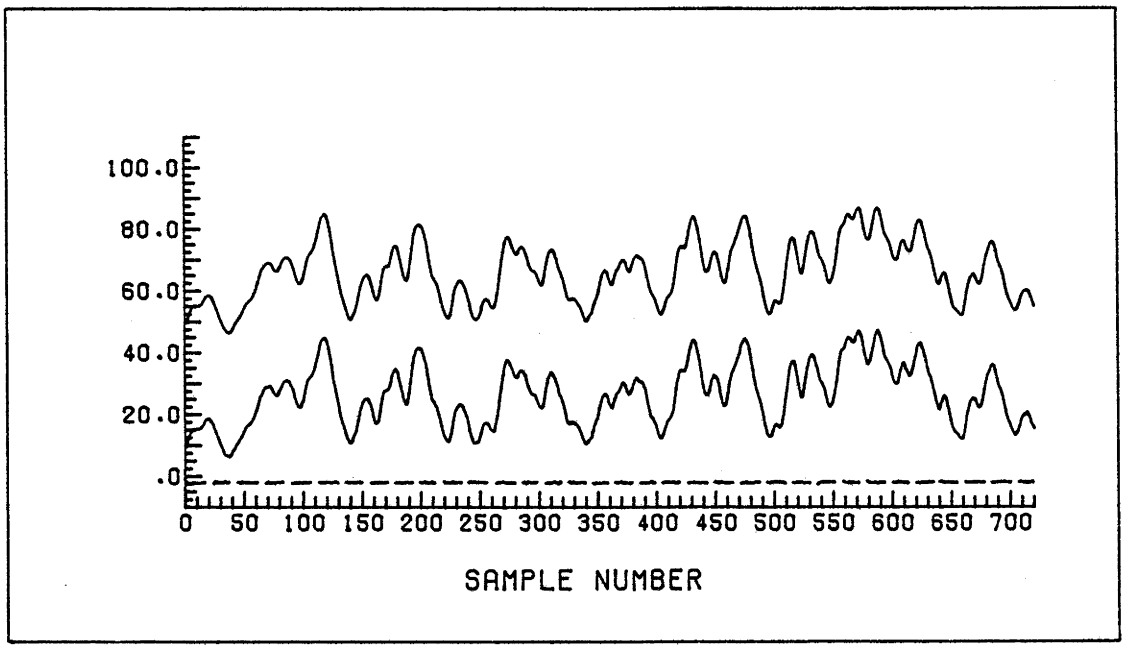


Figure 4.18 The noise free output (upper graph) of the model of the carbon monoxide analyser resulting from the second known input. The noisy output obtained by adding white noise to the noise free output is also shown and the noise free output is translated vertically by +40 units for clarity.

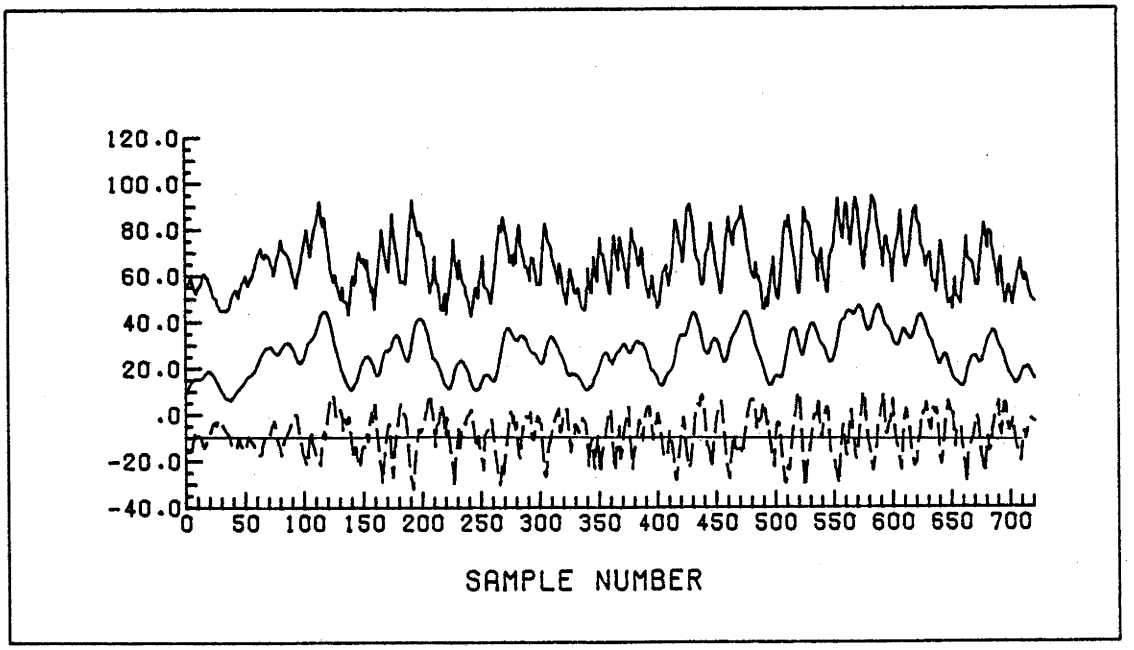


Figure 4.19 Comparison of the second known input (upper graph) with the noisy output of the model of the analyser. Known input is translated vertically by +40 units for clarity.

Table 4.12

Comparison of Some Statistics for the Second Data Set of 720 Samples

Signal	Mean	Variance	Maximum Value	Minimum Value
Known input	25.11	135.68	54.30	2.30
Noise free output	26.10	90.60	46.91	6.37
Noisy output	26.11	90.93	47.24	6.14
Estimated (filtered) input	26.00	138.09	52.85	0.68
Estimated (smoothed) input	25.97	127.78	53.12	1.20

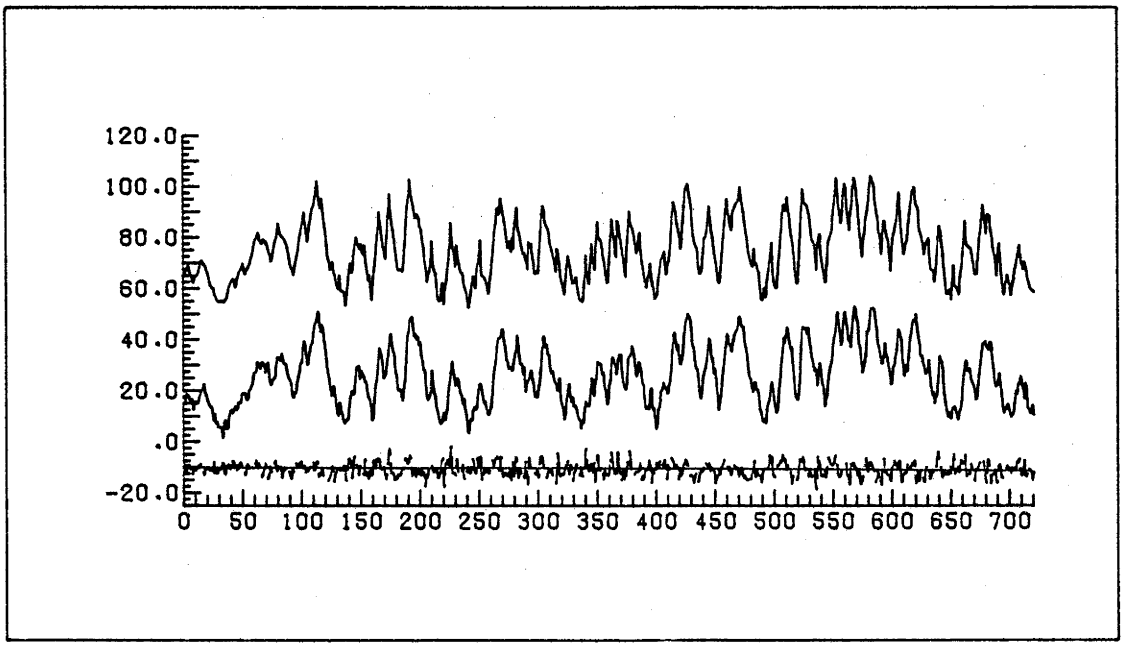


Figure 4.20 Comparison of second known input (upper graph - translated vertically by +50 units) and smoothed estimate of input using q = 500.

4.10 Conclusion

In this chapter we have shown that a continuous air pollution analyser may be described by either discrete time or continuous time transfer function models. Recursive instrumental variable techniques were seen to be useful in the identification of model order and in the estimation of the model parameters. Input estimation algorithms which are robust in the presence of noise on the observed output were presented and evaluated on two sets of data, such algorithms being applicable to any continuous analyser which may be modeled as a linear dynamic system. Precisely because such analysers behave as linear dynamic systems they have frequency response characteristics which result in the attenuation of the input signal, with high frequency components being affected more seriously than low frequency ones. Consequently the statistics of the true input concentrations may not be reproduced exactly. This discrepancy may or may not be serious but as the input estimation procedure presented here is so simple, its adoption would seem advisable since superior data will be obtained.

Chapter 5

SIMPLE LINEAR DYNAMIC MODELS FOR ESTIMATING MISSING AIR POLLUTION DATA

5.1 Introduction

In this chapter we investigate the problem of estimating missing air pollution time series. Many air pollution measurements of interest consist of daily average levels which are computed from the relevant hourly data. Most analyses of such air pollution data simply delete from consideration the days with incomplete records. This may be inefficient if the data set is thereby significantly reduced, and may lead to biased estimates if the missing observations are unrepresentative. Wyzga (1973) draws attention to these possibilities and describes an iterative regression procedure which he employs to obtain estimates of the unobserved data. Use is made of pollutant measurements obtained at the location from which observations are missing (measurements relating both to the required pollutant on days adjacent to those for which the observations are lacking, and to surrogate pollutants on the same day), as well as measurements obtained at other locations in the vicinity (from which he secured same-day observations of the required pollutant). Although Wyzga (1973) does not report the regression model parameters, it is likely they were biased. As has been noted in previous chapters, however, regression models may produce unbiased predictions even if the model parameters are biased, and this probably accounts for the relative success of his procedure.

An alternative method of filling gaps in air pollution time series has been used by Chock et al. (1975). The data used in their study were obtained from a single location. Their approach was to apply a univariate Box Jenkins model to the known data prior to the gap in order to forecast the first half of the missing values, and apply a similar model to the subsequent known data in order to 'back-forecast' the remainder of the missing observations.

The missing air pollution data that we try to estimate in this chapter relate to ozone. Measurements were obtained at locations in the San Joaquin valley in California in connection with a problem of poisoning of agricultural field-workers exposed to organophosphate

pesticide residues (Spear, 1978). Although the daily average level of ozone is usually of less interest than the maximum hourly concentration, it was believed in this case that the former exerted a significant influence on the rate of chemical conversion of the pesticide parathion to its oxygen analogue paraoxon, which was suspected to be the chemical agent primarily responsible for the poisonings (Spear et al., 1978).

The complex and non-linear chemical reactions typical of photochemical smog formation (see Guicherit, 1976) are not modeled explicitly in the present attempt to estimate missing ozone measurements because this is not believed to be necessary. This is not a novel approach since it was used by Hanna (1977) who outlined a useful but simple model of photochemical smog movement along the Los Angeles-Palm Springs trajectory. Reference to this work by Hanna is pertinent for the additional reason that the linear dynamic models used here to estimate missing ozone data are, in effect, similarly describing ozone transport between different locations.

In the course of investigating model structures which were believed suitable for the estimation of the missing ozone data, instrumental variable (IV) algorithms were developed (Sections 5.2 and 5.3) and evaluated (Sections 5.4 and 5.5). These algorithms were used to determine the parameter values in a particular class of multiple input-single output (MISO) models, in which the characteristic polynomials of the transfer functions associated with each input are not constrained to be identical. In the remainder of this chapter we describe the application of both single input-single output (SISO) and MISO models to the estimation of the missing observations in the San Joaquin valley data set referred to earlier.

5.2 Derivation of the Instrumental Variable Methods

Estimation procedures for linear MISO systems have usually been developed for the following multiple input autoregressive moving average, exogenous variables (ARMAX) model

$$y_k = \sum_{i=1}^{l} \frac{B_i(z^{-1})}{A(z^{-1})} u_{ik} + \frac{D(z^{-1})}{A(z^{-1})} e_k \qquad (5.1)$$

where $A(z^{-1})$, $B_i(z^{-1})$, and $D(z^{-1})$ are appropriate polynomials in the backward shift operator z^{-1}; y_k is the single measured output of the system; u_{ik}, $i = 1,2, \ldots, \ell$, are the multiple inputs; and e_k is a zero mean white noise input with variance σ^2. In this chapter, however, an alternative time series model representation is considered and has the form

$$x_{ik} = \frac{B_i(z^{-1})}{A_i(z^{-1})} u_{ik}$$

$$\triangleq \frac{b_{io} + b_{il}z^{-1} + \ldots + b_{im_i}z^{-m_i}}{1 + a_{il}z^{-1} + \ldots + a_{in_i}z^{-n_i}} \qquad (5.2)$$

$$y_k = \sum_{i=1}^{\ell} x_{ik} + \xi_k$$

where ξ_k denotes output noise of a general nature. In this manner direct allowance is made for ℓ transfer functions between each of the multiple inputs and single output which have different denominator polynomials. This model, which can be considered the dynamic equivalent of the linear regression model with regression coefficients replaced by transfer functions, can be contrasted with (5.1) where the denominator polynomials are assumed identical. In general, this multiple input transfer function (MITF) model should ensure a parametrically more efficient representation. It has been considered briefly by Box and Jenkins (1970), and in more detail by Barrett-Lennard (1978) who used it to model a chemical process in an aluminium plant. It should also be noted that the formulation (5.2) allows for a mixed transfer function-regression model simply by setting the appropriate parameters $a_{il}, \ldots, a_{in_i}$ to zero for one or more of the inputs u_{ik}.

With no assumptions about ξ_k we will see that basic instrumental variable procedures can be used to obtain consistent estimates of the parameter vectors $\underline{a}_i = (a_i \ \ldots \ a_{in_i} \ b_{io} \ \ldots \ b_{im_i})^T$, $i = 1,2, \ldots, \ell$ in (5.2). However, if it is assumed

further that ξ_k can be generated from a white noise process via an auto-regressive moving average model of the form

$$\xi_k = \frac{D(z^{-1})}{C(z^{-1})} e_k \triangleq \frac{1 + d_1 z^{-1} + \ldots + d_r z^{-r}}{1 + c_1 z^{-1} + \ldots + c_s z^{-s}} e_k \qquad (5.3)$$

then the noise parameter vector $\underline{c} = (c_1 \ \ldots \ c_s d_1 \ \ldots \ d_r)^T$ and the variance $\sigma^2 = \text{var}\ (e_k)$ can be estimated by a procedure such as approximate maximum likelihood (AML)(see Young, 1974).

In addition we will see that these assumptions facilitate the derivation of a refined IVAML procedure, which has the advantage of asymptotic efficiency. As might be expected from the results of Young and Jakeman (1979c) and Jakeman and Young (1979) refined IVAML proves to be asymptotically efficient while basic IVAML and basic IV yield consistent estimates and seem statistically efficient enough for most practical purposes. Also we will see that the IV model order identification procedure detailed in Young et al. (1980) is reasonably successful for the structure (5.2) although the identification of model order in the simulation experiments is not as clearly defined as in the equivalent SISO situation.

Recursive IV methods for the estimation of the system model in (5.2) and AML methods for the noise model (5.3) can be derived by approaches based either on the concept of generalised equation error (GEE) minimisation or that of prediction error minimisation. In the present case estimates are chosen to minimize a cost function J of the form

$$J = \Sigma e_k^2 \qquad (5.4)$$

where $$e_k = \frac{C}{D}\left(y_k - \sum_{i=1}^{l} \frac{B_i}{A_i} u_{ik} \right)$$

In seeking to minimise (5.4) the following definitions are useful:

$$\underline{x}_{ik} = (-x_{ik-1}, \ldots, -x_{ik-n_i}, u_{ik}, \ldots, u_{ik-m_i})^T$$

$$\underline{a}_i = (a_{i1}, \ldots, a_{in_i}, b_{io}, \ldots b_{im_i})^T$$

$$x_{ik} = \underline{x}_{ik}^T \underline{a}_i \qquad (5.5)$$

$$\eta_{ik} = y_k - \sum_{j \neq i} \underline{x}_{jk}^T \underline{a}_j$$

and $$\underline{\eta}_{ik} = (-\eta_{ik-1}, \ldots, -\eta_{ik-n_i}, u_{ik}, \ldots, u_{ik-m_i})^T$$

Also

$$x^*_{ik} = \frac{C}{A_i D} x_{ik}$$

$$u^*_{ik} = \frac{C}{A_i D} u_{ik} \qquad (5.6)$$

$$\eta^*_{ik} = \frac{C}{A_i D} \eta_{ik}$$

In (5.6) the star superscripts are used to denote those variables which are <u>prefiltered</u> by C/A_iD and which are associated with the i-th transfer function B_i/A_i. Such prefiltering is an essential part of refined IV estimation (see Young, 1976).

With these definitions, another useful expression for e_k is the following

$$e_k = \frac{C}{D} (\eta_{ik} - \frac{B_i}{A_i} u_{ik}) = \frac{C}{A_i D} (A_i \eta_{ik} - B_i u_{ik})$$

$$= \frac{C}{A_i D} (\eta_{ik} - \underline{\eta}_{ik}^T \underline{a}_i)$$

$$\triangleq \eta^*_{ik} - \underline{\eta}^{*T}_{ik}\,\underline{a}_i$$

Now by setting the gradient $\nabla_{\underline{a}_i} J = \Sigma e_k(\partial e_k/\partial \underline{a}_i)$ to zero we obtain the following 'normal' equations

$$\sum_{k=1}^{N} (\eta^*_{ik} - \underline{\eta}^{*T}_{ik}\,\hat{\underline{a}}_i)\,\underline{x}^{*T}_{ik} = \underline{0}^T \qquad (5.7)$$

where it will be noted that evaluation of this expression nominally requires knowledge of the noise model parameters, as well as the parameters of the other transfer functions.

The linearised system of estimation equations (5.7) for finding $\hat{\underline{a}}_i$ can be placed quite straightforwardly in recursive form. In a similar manner to that described for the SISO case (see Young and Jakeman, 1979c), the symmetric matrix gain version of the recursive algorithm is given by equation (5.8) for $i = 1, 2, \ldots, \ell$

$$\begin{aligned}
\hat{\underline{a}}_{ik} &= \hat{\underline{a}}_{ik-1} - \hat{P}_{ik-1}\hat{\underline{x}}^*_{ik}(1 + \hat{\underline{x}}^{*T}_{ik}\,\hat{P}_{ik-1}\,\hat{\underline{x}}^*_{ik})^{-1}(\hat{\underline{\eta}}^{*T}_{ik}\hat{\underline{a}}_{ik-1} - \hat{\eta}^*_{ik}) \\
\hat{P}_{ik} &= \hat{P}_{ik-1} - \hat{P}_{ik-1}\hat{\underline{x}}^*_{ik}(1 + \hat{\underline{x}}^{*T}_{ik}\,\hat{P}_{ik-1}\,\hat{\underline{x}}^*_{ik})^{-1}\hat{\underline{x}}^{*T}_{ik}\,\hat{P}_{ik-1}
\end{aligned} \qquad (5.8)$$

Since we do not have knowledge of all the variables in (5.8), those which cannot be obtained exactly have been given hat superscripts to denote replacement by their estimated values; we will discuss subsequently both how these estimates ($\hat{\underline{x}}^*_{ik}$ and $\hat{\underline{\eta}}^*_{ik}$) are generated, and the initialisation of $\hat{\underline{a}}_{io}$ and $\hat{P}_{io}$.

If it is assumed for the moment that the estimates $\hat{x}_{ik}$, $i = 1,2,\ldots, \ell$ are available, an estimate $\hat{\xi}_k$ of ξ_k may be obtained from

$$\hat{\xi}_k = y_k - \hat{x}_{1k} - \hat{x}_{2k} - \ldots - \hat{x}_{\ell k},$$

and then the refined AML algorithm is found by setting the derivative of J (with respect to $\underline{c}$) to zero, and placing the result in recursive

form as

$$\hat{\underline{c}}_k = \hat{\underline{c}}_{k-1} - \hat{P}^n_{k-1}\hat{\underline{n}}^*_k \left(1 + \hat{\underline{n}}^{*T}_k \hat{P}^n_{k-1} \hat{\underline{n}}^*_k \right)^{-1} (\hat{\underline{n}}^T_k \hat{\underline{c}}_{k-1} - \hat{\xi}_k)$$

$$\hat{P}^n_k = \hat{P}^n_{k-1} - \hat{P}^n_{k-1}\hat{\underline{n}}^*_k \left(1 + \hat{\underline{n}}^{*T}_k \hat{P}^n_{k-1} \hat{\underline{n}}^*_k \right)^{-1} \hat{\underline{n}}^{*T}_k \hat{P}^n_{k-1} \qquad (5.9)$$

with $\hat{\underline{n}}^T_k = (-\hat{\xi}_{k-1} \cdots -\hat{\xi}_{k-s} \hat{e}_{k-1} \cdots \hat{e}_{k-r})$

and $\hat{\underline{n}}^*_k = (-\hat{\xi}^*_{k-1} \cdots -\hat{\xi}^*_{k-s} \hat{e}^*_{k-1} \cdots \hat{e}^*_{k-r})^T$

obtained from the prefiltering operations

$$\hat{\xi}^*_k = \frac{1}{\hat{D}} \hat{\xi}_k \quad \text{and} \quad \hat{e}^*_k = \frac{1}{\hat{D}} \hat{e}_k \qquad (5.10)$$

where $\hat{D}$ is the estimate of the polynomial $D(z^{-1})$. As in the case of $\hat{x}_{ik}$ and $\hat{n}_{ik}$, generation of the estimates $\hat{\xi}_k$ and $\hat{e}_k$ will be considered in the next section.

It should be noted that certain non-symmetric gain algorithms can be obtained as alternatives to (5.8) and (5.9) by using, for example, the GEE approach. In the case of (5.9), the non-symmetric algorithm offers no advantages but we will see that for (5.8) the non-symmetric form possesses superior robustness and convergence characteristics, particularly when there is a paucity of data and low signal to noise ratios (see Young and Jakeman, 1979c). In the non-symmetric gain form of (5.8), the transposed vector $\hat{\underline{x}}^{*T}_{ik}$ is replaced by $\hat{\underline{n}}^{*T}_{ik}$ wherever it appears.

5.3 Implementation of the Algorithms

In this section we discuss the implementation of the refined IVAML algorithm for the multiple transfer function formulation (5.2) and (5.3). The implementation of the basic IV (AML) algorithm follows

easily from this since the only differences are the following:

(a) the prefiltering operations given by (5.6) and (5.10) are not utilised and recursive estimation proceeds according to (5.8) and (5.9) with star superscripts removed from all variable definitions.

(b) if a noise model is required, estimation of the parmeters in (5.3) is best performed subsequent to all system models being satisfactorily estimated. Alternatively it can be omitted entirely either if it is not required or does not correspond to the form (5.3).

Figure 5.1 is a summary of the refined IVAML algorithm for a system with $\ell = 2$ inputs. It shows the decomposition into three co-ordinated sub-system estimators, the refined IV for each of the two system transfer functions, and the refined AML sub-algorithm for the noise model. As described for the SISO system (see Young and Jakeman, 1979c), iterative and fully recursive versions of the algorithms can be specified. In both cases, recursive updating of the parameter estimates is utilised but, in the iterative case, the parameters $\hat{\underline{a}}_i$ ($i = 1, 2$) of the three prefilters and two auxiliary models are not updated recursively. In fact, these parameters are kept constant during each iteration (pass) through the entire data and are updated only at the completion of the iteration. Henceforth we will restrict attention to the iterative case for convenience of presentation. The fully recursive version is conceptually simpler except that devices must be employed to facilitate convergence. The interested reader is referred to Soderstrom, Ljung and Gustavsson (1974) for discussion of this topic.

In order to consider the details of the refined IV algorithm in the general ℓ input case, consider the estimation of parameters $\underline{a}_i$ for a particular transfer function operating on the i-th input. At the k-th recursive step, the new variable required is $\hat{\eta}^*_{ik}$ and to obtain this we note the following relationships:

$$\hat{\eta}_{ik} = y_k - \sum_{j \neq i} \hat{x}_{jk}$$

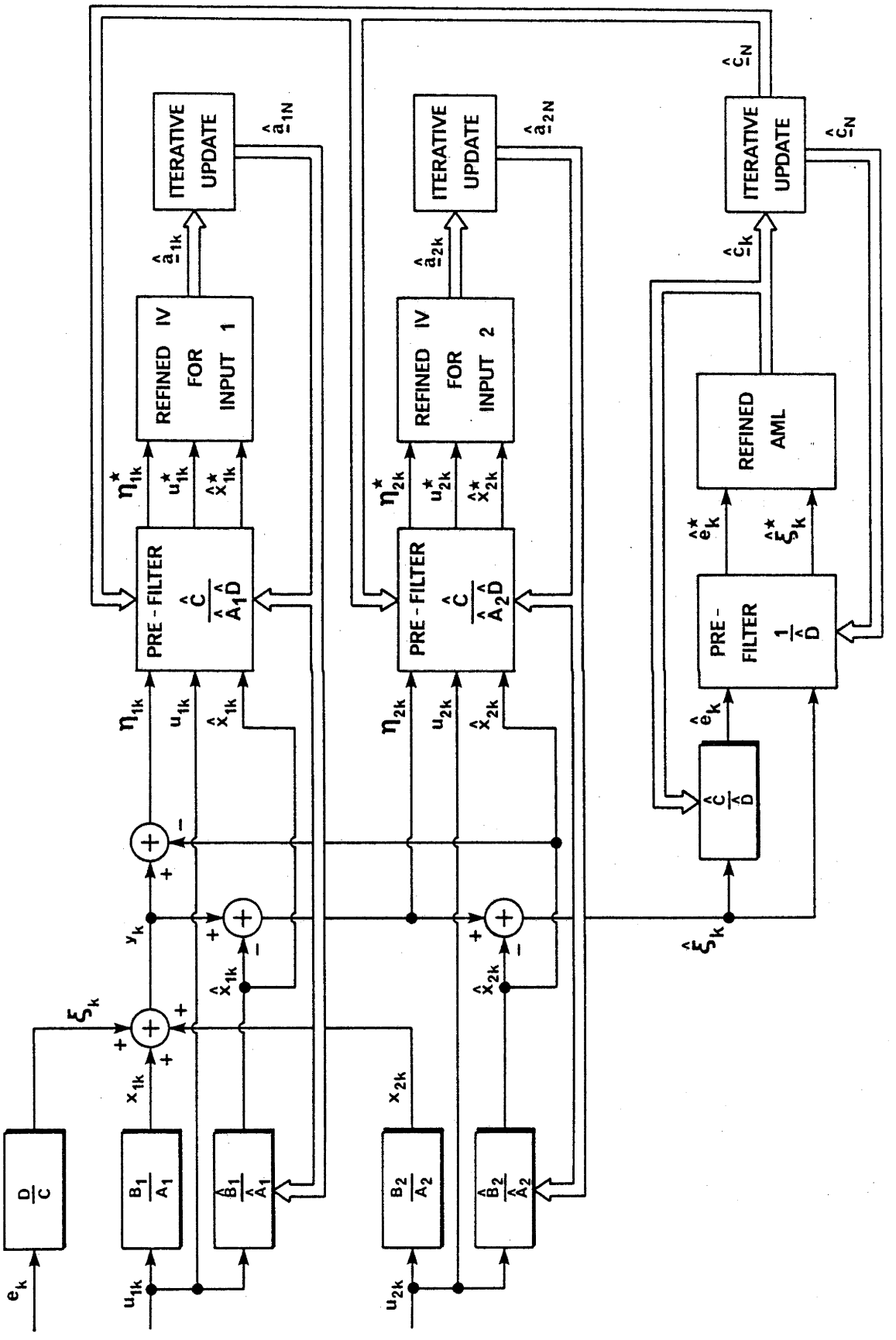


Figure 5.1 Summary of the refined IVAML algorithm for a MITF system with two inputs

$$\hat{\eta}^*_{ik} = \frac{C}{A_i D}\hat{\eta}_{ik}$$

To make use of these, we require (i) the auxiliary model outputs $\hat{x}_{jk}$

$$\hat{x}_{jk} = \underline{\hat{x}}^T_{jk}\,\underline{\hat{a}}_j$$

for the other $\ell - 1$ transfer functions; (ii) an estimate of A_i for the current transfer function being considered; and (iii) estimates of C and D from the noise model.

For the basic IV algorithm on first iteration, initial estimates for a particular transfer function j can be obtained using either SISO modeling of y_k against each input u_{jk} (k = 1, 2, ..., N) in turn or, preferably, the converged basic IV estimates obtained from an ARMAX model of the form (5.1). For the noise model, we usually set C = D = I initially. Such initial estimates also may be used for the refined INVAML algorithm (5.8) and (5.9) but it is preferable that the converged parameter estimates from the basic IVAML form of the algorithm (5.8) and (5.9) be used.

Clearly, an estimate of ξ_k can be obtained either after an iteration on just one transfer function or after an iteration on them all. We chose to implement AML each time an iteration was completed for the estimation of a single transfer function. If time is not critical, this is a good strategy since it ensures that the prefiltering used for the estimation of the next transfer function involves the best current estimates of the noise model.

The remaining details of the algorithm are are as given in Young and Jakeman (1979c) for the SISO case. Thus, the matrices $\hat{P}_{io}$ (i = 1, 2, ..., ℓ) and $\hat{P}^n_o$ are assumed to be diagonal on the first iteration, with all elements set to a high level like 10^6; and for each iteration after the first, $\hat{P}_{io}$ and $\hat{P}^n_o$ are given the values of $\hat{P}_{iq}$ and $\hat{P}^n_q$ (q $\cong$ N/10) at the previous iteration, in order to inject (heuristically) into the algorithm our increasing confidence in the post-iteration 1 estimates. Also the recursive algorithms are not called until the initialisation of the vectors $\underline{\hat{n}}^*_{ik}$, $\underline{\hat{x}}^*_{jk}$, $\underline{\hat{n}}^*_k$ and $\underline{n}_k$ is

complete. In this way, $\underline{\hat{a}}_{io}$ is not updated until the recursive step max. $(n_i + 1, m_i + 1)$ and $\underline{\hat{c}}_o$ until recursive step max. $(s + 1, r + 1)$. Also, while $\hat{\xi}_k$ is obtained from the auxiliary model (that is, via the estimates of A_i and B_i at the previous iteration), $\hat{e}_k$ is always obtained recursively using the latest recursive estimate of C and D. Iteration in this manner continues until convergence is achieved.

Finally, by simple extension of the results of Pierce (1972) the $\hat{P}_{jk}$ and $\hat{P}_k^n$ matrices in (5.8) and (5.9) can be used to provide estimates of the variance-covariance matrix of the parameter estimation errors. This is best achieved by performing a dummy iteration on the $\ell + 1$ subsystems with the auxiliary model, the $\underline{\hat{a}}_i$ and $\underline{\hat{c}}$ parameters and prefilters set to their converged values, and the $\hat{P}$ matrices set to their initial values with large diagonal elements. We will see how this strategy can be used to obtain a very satisfactory estimate of the error covariance matrix. The exact error covariance matrix can be obtained in a similar manner by replacing the converged estimates in the dummy iteration by the equivalent true parameter values. If the non-symmetric gain algorithm is utilised, as seems preferable in practice, the symmetric gain form is still utilised in these dummy iterations since the symmetric gains $\hat{P}_{iN}$ and $\hat{P}_N^n$ provide a superior basis for the estimation of the error covariance matrix.

5.4 Simulation Studies

The major Monte-Carlo simulation was performed on 2, 3 and 4 input systems of second order using both refined and basic IVAML. The models employed were as follows:

Model 1

$$y_k = \frac{1.0 + 0.5z^{-1}}{1 - 1.5z^{-1} + 0.7z^{-2}} u_{1k} + \frac{0.4 - 1.8z^{-1}}{1 - 0.5z^{-1} - 0.3z^{-2}} u_{2k} + \xi_k$$

Model 2

$$y_k = \frac{1.0 + 0.5z^{-1}}{1 - 1.5z^{-1} + 0.7z^{-2}} u_{1k} + \frac{0.4 - 1.8z^{-1}}{1 - 0.5z^{-1} - 0.3z^{-2}} u_{2k}$$

$$+ \frac{-0.4 + 1.7z^{-1}}{1 - 0.5z^{-1} + 0.2z^{-2}} u_{3k} + \xi_k$$

Model 3

$$y_k = \frac{1.0 + 0.5z^{-1}}{1 - 1.5z^{-1} + 0.7z^{-2}} u_{1k} + \frac{0.4 - 1.8z^{-1}}{1 - 0.5z^{-1} - 0.3z^{-2}} u_{2k}$$

$$+ \frac{-0.4 + 1.7z^{-1}}{1 - 0.5z^{-1} + 0.2z^{-2}} u_{3k} + \frac{1.0 + 1.0z^{-1}}{1 + 0.9z^{-1} + 0.95z^{-2}} u_{4k} + \xi_k$$

with

$$\xi_k = \frac{1 - 0.2z^{-1}}{1 - 0.5z^{-1}} e_k$$

in each case. The system models were adapted from Barrett-Lennard (1978).

Results were also obtained for the two input Model 4 given below which has the same characteristic polynomial in each transfer function. This experiment investigates whether the MITF algorithms are capable of handling such a model, even if they are not informed that there is a common denominator polynomial.

Model 4

$$y_k = \frac{1.0 + 0.5z^{-1}}{1 - 1.5z^{-1} + 0.7z^{-2}} u_{1k} + \frac{1.0z^{-1}}{1 - 1.5z^{-1} + 0.7z^{-2}} u_{2k} + \xi_k$$

In all the simulation exercises, 30 Monte Carlo experiments were

performed using the same psuedo-random binary input signals and with the white noise series (e_k) for each experiment randomly sampled from a normal distribution with variance appropriately scaled to yield selected signal to noise ratio S, where S is defined as

$$S \triangleq \sum_{k=1}^{N} \left(\sum_{i=1}^{\ell} x_{ik} \right)^2 / \sum_{k=1}^{N} \xi_k^2$$

The values of S and N were chosen by analogy with previous experimentation. For example, the lowest values that Soderstrom, Ljung and Gustavsson (1974) use for second order SISO models are S = 1 and N = 100. Since we have ℓ second order system models to estimate, we chose N = 100 x ℓ for the lowest sample size and S = 1 x ℓ for the lowest signal to noise ratio. For each Model, three different sample sizes and two different signal to noise ratios were chosen. In Figures 5.2 and 5.3 examples of the signals used in the simulation experiments are shown. The input signals, noise free output signals and typical noise and noisy output signals are shown for the smallest sample size and both signal to noise ratios chosen for Models 1 and 4 respectively.

Satisfactory estimation results were obtained for all the Models 1 to 4 and the detailed results are shown in Tables 5.1, 5.2 (Model 1), Tables 5.3, 5.4 (Model 2), Tables 5.5, 5.6 (Model 3) and Tables 5.7, 5.8 (Model 4). For each sample size the first column shows the mean value of the parameter estimates while the second and third display the standard error of the estimate from its true value and mean value respectively. The fourth column (in parentheses) contains the standard errors of the parameters estimated from the algorithm. For the refined IVAML results these standard errors for the system parameters are given by the square root of the diagonal elements of $\hat{\sigma}^2 \hat{P}_{iN}$, i = 1, 2, ..., ℓ, the estimated covariance matrices. Similarly, for the noise model parameters the standard errors are given by the square root of the diagonal elements of $\hat{\sigma}^2 \hat{P}_N^n$. The estimate of the white noise variance σ^2 is obtained as follows subsequent to convergence

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{k=1}^{N} \hat{e}_k^2$$

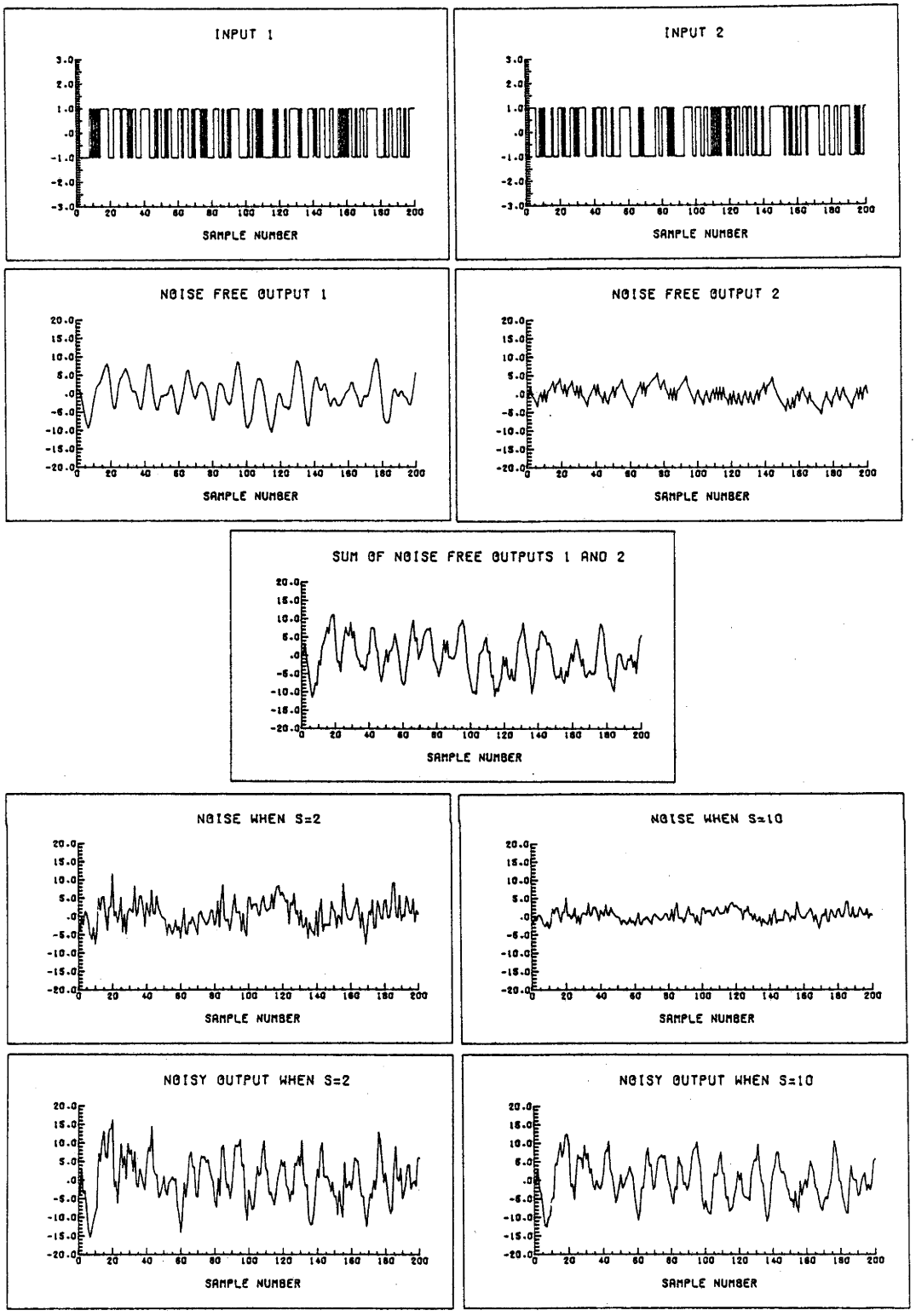


Figure 5.2 Plot of the input signals, noise free output signals and typical noise and noisy output signals for Model 1

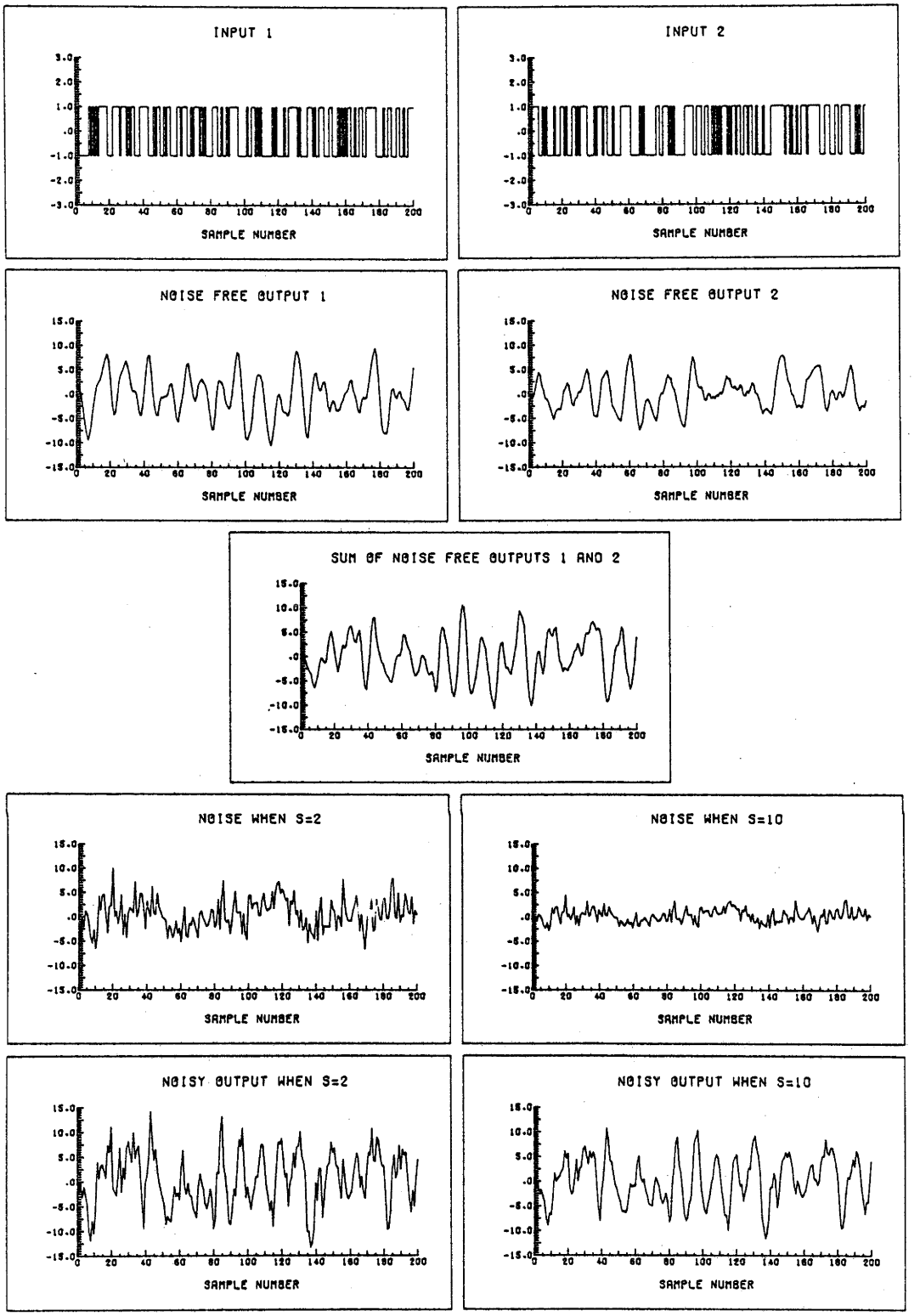


Figure 5.3 Plot of the input signals, noise free output signals and typical noise and noisy output signals for Model 4

When basic IV is used, it is not possible to procure $\hat{\sigma}^2$ in the above manner since $\hat{e}_k$ is not available. As a crude approach, however, $\hat{\sigma}^2$ can be replaced by $\hat{\sigma}^2_\xi$ where

$$\hat{\sigma}^2_\xi = \frac{1}{N} \sum_{k=1}^{N} \hat{\xi}^2_k$$

is an estimate of the variance of the $\hat{\xi}_k$ sequence. The resulting covariance matrices will not provide good estimates of error estimation statistics but they will be useful in indicating the relative errors in different estimates and they will usually be conservative. For interest, these calculations were used in the basic IV simulations: this allows us to investigate quantitatively how the crude approximation to the covariance matrix using $\hat{\sigma}^2_\xi$ compares with the empirically computed results. Thus for the basic IVAML results in Tables 5.1 to 5.8 the standard errors in column 4 are determined in a fashion similar to that described for refined IVAML except that for the system parameters $\hat{\sigma}^2_\xi$ is used instead of $\hat{\sigma}^2$.

The standard errors associated with the minimum variance estimates† are given (in parentheses) in the fifth column. Finally, in all the Tables 5.1 to 5.8 the coefficient of determination R^2_T, which was defined in Chapter 3, is also quoted. Here the value of R^2_T is the average value over 30 experiments.

In connection with the results of Tables 5.1 to 5.8, the following additional points should be mentioned. When the symmetric refined IVAML algorithm was implemented on data generated with the lower S ratios and lowest samples sizes, some of the experiments failed to converge. It was found, however, that the non-symmetric refined IV did not suffer from this problem.

These observations are consistent with the results obtained by Young and Jakeman (1979c) for SISO models and Jakeman and Young (1979) for multiple input-multiple output models, and they provide further evidence for the superiority of non-symmetric gain IV algorithms in low signal/noise, low sample size situations. In this connection, it is interesting to note that the equivalent algorithms to (5.8) and (5.9) generated by GEE minimisation procedures are naturally in the

† see p.103 which explains how these are calculated using the results of Pierce (1972)

non-symmetric form.

For higher values of S and N, the performance of the symmetric and non-symmetric gain algorithms were virtually identical but, to be consistent, only the non-symmetric gain results are quoted in Tables 5.1 to 5.8. On the other hand, the error covariance matrices were generated in all cases from the 'dummy' run mentioned earlier, using the symmetric gain algorithms.

One final but important simulation study investigated the accuracy of the IV model structure identification procedure of Young et al. (1980) (outlined in Chapter 3) when applied to the MITF model form (5.2). Using Models 1 and 2 an attempt was made to identify the model order ($n_i = m_i = 2$) $i = 1, 2, \ldots, \ell$, by reference to the IV structure identification statistics obtained from SISO modeling. In this particular case the IV structure identification procedure makes full use of the robustness of the IV algorithm in the presence of noise by determining separately the model order of each transfer function using SISO modeling. Thus for a particular SISO sub-system, contributions to the output from all other transfer functions are regarded as 'noise' additional to the ξ_k term. The results are given in Tables 5.9 and in 5.10 for each transfer function of Models 1 and 2 respectively.

5.5 Discussion of the Simulation Results

The results in Tables 5.1 to 5.8 confirm the robustness of the IV algorithms proposed in this chapter. As expected the parameter estimates and R_T^2 values are superior and the standard errors lower for refined IVAML than for basic IVAML. The results also confirm that the standard errors obtained directly from the refined algorithm (column 4) agree well with those calculated empirically (columns 2 and 3). Even for the basic IV algorithm, there is a very useful sympathetic relationship between the standard errors calculated empirically and the standard errors obtained from the algorithm in which $\hat{\sigma}^2_\xi$ is used. As expected, the crude standard errors in this case are always safely conservative, being larger than the standard errors obtained when the white noise variance $\hat{\sigma}^2$ is used to estimate the standard errors.

Table 5.1

Basic IVAML and Refined IVAML for Model 1 with S = 2.0

Method	Parameter	True Value	Number of Samples: 200 (R_T^2=0.64)	500 (R_T^2=0.65)	2000 (R_T^2=0.67)
Basic IVAML	a_{11}	-1.5	-1.5022 ± .1164 ± .1164 (.1619) (.0548)	-1.5071 ± .0845 ± .0842 (.0956) (.0337)	-1.5070 ± .0436 ± .0430 (.0479) (.0168)
	a_{12}	0.7	0.7036 ± .1173 ± .1173 (.1620) (.0489)	0.7069 ± .0772 ± .0769 (.0938) (.0292)	0.7058 ± .0369 ± .0364 (.0465) (.0144)
	b_{10}	1.0	0.8810 ± .4665 ± .4511 (.2750) (.2383)	0.9657 ± .2056 ± .2027 (.1655) (.1453)	1.0234 ± .1068 ± .1043 (.0788) (.0702)
	b_{11}	0.5	0.4624 ± .4451 ± .4435 (.3279) (.3208)	0.4870 ± .3186 ± .3183 (.1953) (.1955)	0.4644 ± .1674 ± .1636 (.0934) (.0943)
	a_{21}	-0.5	-0.5378 ± .1702 ± .1659 (.1483) (.1292)	-0.4977 ± .0754 ± .0754 (.0868) (.0782)	-0.5061 ± .0440 ± .0436 (.0404) (.0367)
	a_{22}	-0.3	-0.2821 ± .1466 ± .1455 (.1449) (.1270)	-0.3050 ± .0797 ± .0796 (.0858) (.0768)	-0.2936 ± .0407 ± .0402 (.0398) (.0362)
	b_{20}	0.4	0.4017 ± .2984 ± .2984 (.2732) (.2524)	0.4443 ± .1477 ± .1409 (.1653) (.1536)	0.4165 ± .0808 ± .0791 (.0787) (.0739)
	b_{21}	-1.8	-1.8115 ± .2896 ± .2893 (.2801) (.2499)	-1.7730 ± .1788 ± .1768 (.1683) (.1530)	-1.8052 ± .0740 ± .0738 (.0807) (.0734)
	c_1	-0.5	-0.4883 ± .1378 ± .1373 (.1900) (.1765)	-0.4700 ± .1120 ± .1079 (.1242) (.1110)	-0.4916 ± .0612 ± .0606 (.0629) (.0553)
	d_1	-0.2	-0.1460 ± .1555 ± .1459 (.2038) (.1998)	-0.1472 ± .1324 ± .1214 (.1322) (.1257)	-0.1817 ± .0684 ± .0659 (.0668) (.0626)
			200 (R_T^2=0.68)	500 (R_T^2=0.67)	2000 (R_T^2=0.67)
Refined IVAML	a_{11}	-1.5	-1.5081 ± .0605 ± .0599 (.0517) (.0548)	-1.5095 ± .0332 ± .0318 (.0322) (.0337)	-1.4974 ± .0172 ± .0170 (.0169) (.0168)
	a_{12}	0.7	0.7074 ± .0510 ± .0504 (.0458) (.0489)	0.7104 ± .0319 ± .0301 (.0279) (.0292)	0.6978 ± .0136 ± .0134 (.0145) (.0144)
	b_{10}	1.0	0.9635 ± .2630 ± .2605 (.2282) (.2383)	0.9454 ± .1344 ± .1228 (.1435) (.1453)	1.0109 ± .0702 ± .0693 (.0701) (.0702)
	b_{11}	0.5	0.4829 ± .3107 ± .3102 (.3052) (.3208)	0.5120 ± .1674 ± .1670 (.1906) (.1955)	0.5021 ± .1008 ± .1008 (.0945) (.0943)
	a_{21}	-0.5	-0.5139 ± .1635 ± .1629 (.1242) (.1292)	-0.4978 ± .0712 ± .0711 (.0771) (.0782)	-0.5095 ± .0421 ± .0410 (.0365) (.0367)
	a_{22}	-0.3	-0.3040 ± .1556 ± .1556 (.1215) (.1270)	-0.2994 ± .0745 ± .0745 (.0758) (.0768)	-0.2938 ± .0420 ± .0416 (.0360) (.0362)
	b_{20}	0.4	0.3818 ± .2372 ± .2365 (.2452) (.2524)	0.4225 ± .1277 ± .1257 (.1519) (.1536)	0.4147 ± .0745 ± .0731 (.0737) (.0739)
	b_{21}	-1.8	-1.7914 ± .2598 ± .2597 (.2463) (.2499)	-1.7932 ± .1605 ± .1604 (.1514) (.1530)	-1.8128 ± .0691 ± .0679 (.0733) (.0734)
	c_1	-0.5	-0.4141 ± .2074 ± .1888 (.2123) (.1765)	-0.4242 ± .1422 ± .1204 (.1192) (.1110)	-0.4843 ± .0671 ± .0652 (.0573) (.0553)
	d_1	-0.2	-0.1328 ± .2385 ± .2289 (.2323) (.1998)	-0.1193 ± .1491 ± .1254 (.1332) (.1257)	-0.1796 ± .0773 ± .0746 (.0649) (.0626)

Table 5.2

Basic IVAML and Refined IVAML for Model 1 with S = 10.0

Method	Parameter	True Value	Number of Samples: 200 (R_T^2=0.90)	500 (R_T^2=0.91)	2000 (R_T^2=0.91)
Basic IVAML	a_{11}	-1.5	-1.5063 ± .0468 ± .0463 (.0699) (.0246)	-1.5053 ± .0375 ± .0371 (.0425) (.0151)	-1.5035 ± .0196 ± .0192 (.0214) (.0075)
	a_{12}	0.7	0.7074 ± .0473 ± .0467 (.0700) (.0220)	0.7056 ± .0341 ± .0337 (.0417) (.0131)	0.7030 ± .0167 ± .0164 (.0208) (.0064)
	b_{10}	1.0	0.9425 ± .2087 ± .2006 (.1236) (.1071)	0.9837 ± .0924 ± .0910 (.0743) (.0652)	1.0106 ± .0476 ± .0464 (.0352) (.0314)
	b_{11}	0.5	0.4751 ± .2006 ± .1990 (.1460) (.1442)	0.4912 ± .1430 ± .1428 (.0876) (.0877)	0.4839 ± .0746 ± .0729 (.0416) (.0422)
	a_{21}	-0.5	-0.5206 ± .0712 ± .0682 (.0651) (.0580)	-0.4988 ± .0334 ± .0333 (.0385) (.0351)	-0.5029 ± .0197 ± .0195 (.0181) (.0164)
	a_{22}	-0.3	-0.2985 ± .0608 ± .0608 (.0644) (.0571)	-0.3027 ± .0346 ± .0345 (.0382) (.0344)	-0.2976 ± .0182 ± .0180 (.0178) (.0162)
	b_{20}	0.4	0.4049 ± .1290 ± .1290 (.1229) (.1134)	0.4201 ± .0662 ± .0630 (.0742) (.0689)	0.4075 ± .0360 ± .0352 (.0352) (.0330)
	b_{21}	-1.8	-1.8065 ± .1316 ± .1315 (.1247) (.1123)	-1.7877 ± .0787 ± .0777 (.0752) (.0686)	-1.8025 ± .0334 ± .0333 (.0360) (.0328)
	c_1	-0.5	-0.4854 ± .1423 ± .1415 (.1861) (.1746)	-0.4869 ± .0975 ± .0966 (.1228) (.1104)	-0.4973 ± .0592 ± .0591 (.0628) (.0552)
	d_1	-0.2	-0.1401 ± .1580 ± .1462 (.2000) (.1986)	-0.1645 ± .1177 ± .1122 (.1309) (.1253)	-0.1880 ± .0637 ± .0626 (.0667) (.0626)
			200 (R_T^2=0.91)	500 (R_T^2=0.91)	2000 (R_T^2=0.91)
Refined IVAML	a_{11}	-1.5	-1.5038 ± .0265 ± .0263 (.0236) (.0246)	-1.5042 ± .0151 ± .0145 (.0147) (.0151)	-1.4988 ± .0077 ± .0076 (.0075) (.0075)
	a_{12}	0.7	0.7038 ± .0228 ± .0225 (.0210) (.0220)	0.7048 ± .0144 ± .0135 (.0128) (.0131)	0.6989 ± .0061 ± .0060 (.0065) (.0064)
	b_{10}	1.0	0.9786 ± .1186 ± .1166 (.1039) (.1071)	0.9729 ± .0619 ± .0557 (.0645) (.0652)	1.0053 ± .0315 ± .0311 (.0314) (.0314)
	b_{11}	0.5	0.4936 ± .1405 ± .1403 (.1394) (.1442)	0.5072 ± .0763 ± .0760 (.0863) (.0877)	0.5008 ± .0451 ± .0451 (.0422) (.0422)
	a_{21}	-0.5	-0.5111 ± .0628 ± .0618 (.0570) (.0580)	-0.4991 ± .0313 ± .0312 (.0347) (.0351)	-0.5041 ± .0188 ± .0184 (.0163) (.0164)
	a_{22}	-0.3	-0.3009 ± .0598 ± .0598 (.0559) (.0571)	-0.3000 ± .0315 ± .0315 (.0341) (.0344)	-0.2974 ± .0187 ± .0185 (.0161) (.0162)
	b_{20}	0.4	0.3993 ± .1013 ± .1013 (.1106) (.1134)	0.4102 ± .0574 ± .0564 (.0683) (.0689)	0.4063 ± .0335 ± .0329 (.0330) (.0330)
	b_{21}	-1.8	-1.7960 ± .1142 ± .1142 (.1098) (.1123)	-1.7967 ± .0711 ± .0711 (.0680) (.0686)	-1.8062 ± .0309 ± .0303 (.0328) (.0328)
	c_1	-0.5	-0.4084 ± .1862 ± .1622 (.1969) (.1746)	-0.4569 ± .1283 ± .1208 (.1157) (.1104)	-0.4765 ± .0718 ± .0678 (.0572) (.0552)
	d_1	-0.2	-0.1158 ± .1895 ± .1698 (.2175) (.1986)	-0.1525 ± .1328 ± .1241 (.1300) (.1253)	-0.1704 ± .0843 ± .0789 (.0648) (.0626)

Table 5.3

Basic IVAML and Refined IVAML for Model 2 with S = 3.0

Method	Parameter	True Value	Number of Samples: 300 (R_T^2=0.72)	750 (R_T^2=0.75)	2000 (R_T^2=0.75)
Basic IVAML	a_{11}	-1.5	-1.4960 ± .1026 ± .1025 (.1115) (.0391)	-1.4941 ± .0457 ± .0453 (.0643) (.0232)	-1.5005 ± .0354 ± .0353 (.0418) (.0145)
	a_{12}	0.7	0.6925 ± .0951 ± .0948 (.1106) (.0344)	0.6969 ± .0403 ± .0401 (.0630) (.0201)	0.6981 ± .0305 ± .0304 (.0406) (.0124)
	b_{10}	1.0	0.9233 ± .3069 ± .2972 (.2031) (.1735)	0.9576 ± .1539 ± .1479 (.1106) (.0985)	1.0116 ± .1104 ± .1098 (.0679) (.0605)
	b_{11}	0.5	0.5615 ± .3731 ± .3680 (.2411) (.2349)	0.5541 ± .1893 ± .1814 (.1287) (.1326)	0.4746 ± .1421 ± .1398 (.0806) (.0814)
	a_{21}	-0.5	-0.4744 ± .0922 ± .0886 (.1059) (.0917)	-0.4915 ± .0609 ± .0603 (.0574) (.0530)	-0.5073 ± .0302 ± .0293 (.0355) (.0317)
	a_{22}	-0.3	-0.3165 ± .0755 ± .0737 (.1042) (.0900)	-0.3023 ± .0544 ± .0543 (.0568) (.0520)	-0.2960 ± .0297 ± .0294 (.0350) (.0312)
	b_{20}	0.4	0.4288 ± .1904 ± .1883 (.2010) (.1824)	0.4038 ± .1071 ± .1071 (.1106) (.1031)	0.3930 ± .0746 ± .0743 (.0679) (.0637)
	b_{21}	-1.8	-1.7461 ± .1753 ± .1668 (.2059) (.1815)	-1.8116 ± .1040 ± .1033 (.1123) (.1022)	-1.7714 ± .0623 ± .0554 (.0694) (.0633)
	a_{31}	-0.5	-0.5232 ± .1085 ± .1059 (.1239) (.1026)	-0.4969 ± .0690 ± .0689 (.0678) (.0584)	-0.5017 ± .0348 ± .0348 (.0406) (.0355)
	a_{32}	0.2	0.1841 ± .1001 ± .0989 (.1181) (.1009)	0.2053 ± .0576 ± .0574 (.0653) (.0581)	0.2077 ± .0395 ± .0387 (.0397) (.0354)
	b_{30}	-0.4	-0.3824 ± .2101 ± .2094 (.2019((.1848)	-0.4429 ± .1131 ± .1047 (.1105) (.1043)	-0.3900 ± .0705 ± .0698 (.0679) (.0639)
	b_{31}	1.7	1.6689 ± .2201 ± .2179 (.2110) (.1862)	1.6833 ± .1053 ± .1040 (.1149) (.1046)	1.6926 ± .0704 ± .0700 (.0696) (.0640)
	c_1	-0.5	-0.4996 ± .1060 ± .1060 (.1417) (.1330)	-0.4792 ± .0919 ± .0895 (.1015) (.0863)	-0.4978 ± .0704 ± .0704 (.0631) (.0529)
	d_1	-0.2	-0.1350 ± .1420 ± .1263 (.1536) (.1525)	-0.1633 ± .1080 ± .1016 (.1079) (.0978)	-0.1896 ± .0733 ± .0725 (.0669) (.0599)
			300 (R_T^2=0.75)	750 (R_T^2=0.75)	2000 (R_T^2=0.75)
Refined IVAML	a_{11}	-1.5	-1.5047 ± .0395 ± .0393 (.0380) (.0391)	-1.4839 ± .0272 ± .0220 (.0236) (.0232)	-1.5024 ± .0138 ± .0136 (.0145) (.0145)
	a_{12}	0.7	0.7036 ± .0352 ± .0350 (.0334) (.0344)	0.6876 ± .0230 ± .0193 (.0204) (.0201)	0.7003 ± .0119 ± .0119 (.0125) (.0124)
	b_{10}	1.0	0.9913 ± .1428 ± .1425 (.1709) (.1735)	0.9680 ± .1063 ± .1013 (.0987) (.0985)	1.0142 ± .0696 ± .0682 (.0604) (.0605)
	b_{11}	0.5	0.4987 ± .1838 ± .1838 (.2310) (.2349)	0.5878 ± .1539 ± .1263 (.1333) (.1326)	0.4762 ± .0924 ± .0893 (.0814) (.0814)
	a_{21}	-0.5	-0.4800 ± .0802 ± .0776 (.0900) (.0917)	-0.4943 ± .0595 ± .0592 (.0518) (.0530)	-0.5068 ± .0290 ± .0281 (.0322) (.0317)
	a_{22}	-0.3	-0.3182 ± .0697 ± .0673 (.0887) (.0900)	-0.3021 ± .0495 ± .0494 (.0508) (.0520)	-0.2958 ± .0262 ± .0259 (.0317) (.0312)
	b_{20}	0.4	0.4242 ± .1665 ± .1647 (.1783) (.1824)	0.4028 ± .0982 ± .0982 (.1023) (.1031)	0.3925 ± .0649 ± .0645 (.0635) (.0637)
	b_{21}	-1.8	-1.7496 ± .1558 ± .1474 (.1777) (.1815)	-1.8193 ± .0965 ± .0946 (.1015) (.1022)	-1.7753 ± .0645 ± .0596 (.0631) (.0633)
	a_{31}	-0.5	-0.5238 ± .1230 ± .1207 (.1021) (.1026)	-0.5006 ± .0697 ± .0697 (.0585) (.0584)	-0.5002 ± .0340 ± .0340 (.0354) (.0355)
	a_{32}	0.2	0.2086 ± .1094 ± .1091 (.0989) (.1009)	0.2106 ± .0583 ± .0574 (.0576) (.0581)	0.2043 ± .0377 ± .0375 (.0352) (.0354)
	b_{30}	-0.4	-0.3772 ± .1924 ± .1910 (.1805) (.1848)	-0.4301 ± .1201 ± .1163 (.1034) (.1043)	-0.3914 ± .0664 ± .0658 (.0637) (.0639)
	b_{31}	1.7	1.6779 ± .2171 ± .2159 (.1858) (.1862)	1.6815 ± .0936 ± .0917 (.1048) (.1046)	1.6957 ± .0679 ± .0678 (.0639) (.0640)
	c_1	-0.5	-0.3784 ± .1701 ± .1188 (.1395) (.1330)	-0.4241 ± .1253 ± .0998 (.0935) (.0863)	-0.4794 ± .0856 ± .0831 (.0573) (.0529)
	d_1	-0.2	-0.0583 ± .1756 ± .1036 (.1562) (.1525)	-0.1145 ± .1402 ± .1111 (.1057) (.0978)	-0.1748 ± .0916 ± .0881 (.0649) (.0599)

Table 5.4

Basic IVAML and Refined IVAML for Model 2 with S = 15.0

Method	Parameter	True Value	Number of Samples: 300 (R_T^2=0.92)	750 (R_T^2=0.94)	2000 (R_T^2=0.94)
Basic IVAML	a_{11}	-1.5	-1.5032 ± .0454 ± .0453 (.0520) (.0181)	-1.4979 ± .0209 ± .0208 (.0293) (.0104)	-1.5006 ± .0156 ± .0156 (.0185) (.0065)
	a_{12}	0.7	0.7009 ± .0423 ± .0423 (.0515) (.0160)	0.6994 ± .0185 ± .0185 (.0287) (.0090)	0.6995 ± .0134 ± .0134 (.0180) (.0056)
	b_{10}	1.0	0.9713 ± .1329 ± .1298 (.0957) (.0804)	0.9813 ± .0690 ± .0664 (.0500) (.0444)	1.0053 ± .0494 ± .0491 (.0304) (.0271)
	b_{11}	0.5	0.5176 ± .1733 ± .1724 (.1144) (.1069)	0.5236 ± .0854 ± .0820 (.0586) (.0597)	0.4885 ± .0629 ± .0619 (.0359) (.0364)
	a_{21}	-0.5	-0.4868 ± .0399 ± .0377 (.0491) (.0425)	-0.4974 ± .0265 ± .0264 (.0260) (.0239)	-0.5032 ± .0132 ± .0128 (.0157) (.0142)
	a_{22}	-0.3	-0.3118 ± .0327 ± .0305 (.0486) (.0417)	-0.3004 ± .0236 ± .0236 (.0257) (.0234)	-0.2985 ± .0130 ± .0129 (.0155) (.0140)
	b_{20}	0.4	0.4042 ± .0883 ± .0882 (.0948) (.0846)	0.4004 ± .0475 ± .0475 (.0500) (.0464)	0.3969 ± .0334 ± .0332 (.0304) (.0285)
	b_{21}	-1.8	-1.7674 ± .0819 ± .0751 (.0965) (.0841)	-1.8082 ± .0469 ± .0462 (.0507) (.0460)	-1.7872 ± .0278 ± .0247 (.0310) (.0283)
	a_{31}	-0.5	-0.5203 ± .0514 ± .0472 (.0565) (.0476)	-0.4971 ± .0304 ± .0302 (.0303) (.0263)	-0.5008 ± .0155 ± .0155 (.0181) (.0159)
	a_{32}	0.2	0.1863 ± .0451 ± .0430 (.0543) (.0468)	0.2005 ± .0257 ± .0257 (.0294) (.0261)	0.2033 ± .0178 ± .0175 (.0175) (.0158)
	b_{30}	-0.4	-0.3838 ± .0939 ± .0925 (.0952) (.0857)	-0.4156 ± .0491 ± .0466 (.0500) (.0470)	-0.3956 ± .0313 ± .0310 (.0304) (.0286)
	b_{31}	1.7	1.7034 ± .0946 ± .0945 (.0985) (.0863)	1.6916 ± .0454 ± .0447 (.0516) (.0471)	1.6968 ± .0315 ± .0314 (.0311) (.0286)
	c_1	-0.5	-0.4755 ± .0897 ± .0863 (.1255) (.1209)	-0.5204 ± .0924 ± .0902 (.0970) (.0839)	-0.4928 ± .0702 ± .0698 (.0630) (.0528)
	d_1	-0.2	-0.0596 ± .1639 ± .0846 (.1387) (.1434)	-0.1998 ± .0983 ± .0983 (.1036) (.0963)	-0.1835 ± .0780 ± .0763 (.0669) (.0599)
			300 (R_T^2=0.93)	750 (R_T^2=0.94)	2000 (R_T^2=0.94)
Refined IVAML	a_{11}	-1.5	-1.5018 ± .0177 ± .0176 (.0185) (.0181)	-1.4928 ± .0121 ± .0097 (.0105) (.0104)	-1.5010 ± .0063 ± .0062 (.0065) (.0065)
	a_{12}	0.7	0.7014 ± .0160 ± .0159 (.0164) (.0160)	0.6945 ± .0101 ± .0084 (.0092) (.0090)	0.7001 ± .0054 ± .0054 (.0056) (.0056)
	b_{10}	1.0	0.9928 ± .0647 ± .0643 (.0807) (.0804)	0.9834 ± .0485 ± .0456 (.0443) (.0444)	1.0067 ± .0314 ± .0307 (.0270) (.0271)
	b_{11}	0.5	0.5034 ± .0789 ± .0788 (.1090) (.1089)	0.5408 ± .0708 ± .0579 (.0596) (.0597)	0.4892 ± .0419 ± .0405 (.0364) (.0364)
	a_{21}	-0.5	-0.4878 ± .0366 ± .0345 (.0408) (.0425)	-0.4983 ± .0259 ± .0259 (.0234) (.0239)	-0.5031 ± .0126 ± .0122 (.0142) (.0142)
	a_{22}	-0.3	-0.3134 ± .0319 ± .0290 (.0405) (.0417)	-0.3001 ± .0217 ± .0217 (.0230) (.0234)	-0.2983 ± .0115 ± .0114 (.0140) (.0140)
	b_{20}	0.4	0.4062 ± .0746 ± .0744 (.0829) (.0846)	0.4018 ± .0437 ± .0437 (.0461) (.0464)	0.3967 ± .0289 ± .0287 (.0284) (.0285)
	b_{21}	-1.8	-1.7707 ± .0711 ± .0648 (.0827) (.0841)	-1.8114 ± .0435 ± .0420 (.0457) (.0460)	-1.7888 ± .0288 ± .0266 (.0283) (.0283)
	a_{31}	-0.5	-0.5193 ± .0538 ± .0502 (.0474) (.0476)	-0.4983 ± .0300 ± .0299 (.0262) (.0263)	-0.5002 ± .0151 ± .0151 (.0158) (.0159)
	a_{32}	0.2	0.1974 ± .0416 ± .0415 (.0459) (.0468)	0.2041 ± .0254 ± .0251 (.0259) (.0261)	0.2017 ± .0171 ± .0170 (.0158) (.0158)
	b_{30}	-0.4	-0.3773 ± .0886 ± .0856 (.0840) (.0857)	-0.4120 ± .0529 ± .0515 (.0465) (.0470)	-0.3962 ± .0299 ± .0297 (.0285) (.0286)
	b_{31}	1.7	1.7019 ± .0919 ± .0919 (.0857) (.0863)	1.6906 ± .0405 ± .0394 (.0469) (.0471)	1.6978 ± .0304 ± .0304 (.0286) (.0286)
	c_1	-0.5	-0.3893 ± .1277 ± .0636 (.1181) (.1209)	-0.4720 ± .0994 ± .0954 (.0845) (.0839)	-0.4786 ± .0840 ± .0812 (.0571) (.0528)
	d_1	-0.2	-0.0074 ± .1939 ± .0220 (.1355) (.1434)	-0.1551 ± .1128 ± .1035 (.0973) (.0963)	-0.1734 ± .0901 ± .0860 (.0647) (.0599)

Table 5.5

Basic IVAML and Refined IVAML for Model 3 with S = 4.0

Method	Parameter	True Value	Number of Samples: 400 (R_T^2=0.70)	1000 (R_T^2=0.78)	2000 (R_T^2=0.79)
Basic IVAML	a_{11}	-1.5	-1.4805 ± .0981 ± .0962 (.1167) (.0387)	-1.5121 ± .0503 ± .0488 (.0631) (.0216)	-1.4901 ± .0476 ± .0466 (.0437) (.0152)
	a_{12}	0.7	0.6780 ± .0926 ± .0900 (.1144) (.0335)	0.7088 ± .0427 ± .0418 (.0615) (.0186)	0.6903 ± .0407 ± .0396 (.0425) (.0130)
	b_{10}	1.0	0.9595 ± .4159 ± .4139 (.2057) (.1686)	1.0202 ± .1605 ± .1592 (.1044) (.0908)	1.0221 ± .1227 ± .1207 (.0719) (.0633)
	b_{11}	0.5	0.6003 ± .3249 ± .3090 (.2464) (.2274)	0.4475 ± .2281 ± .2220 (.1248) (.1222)	0.5073 ± .1632 ± .1631 (.0852) (.0851)
	a_{21}	-0.5	-0.6007 ± .1418 ± .0999 (.1150) (.0898)	-0.4811 ± .0511 ± .0475 (.0549) (.0487)	-0.4942 ± .0488 ± .0485 (.0370) (.0331)
	a_{22}	-0.3	-0.2110 ± .1431 ± .1120 (.1124) (.0889)	-0.3138 ± .0572 ± .0555 (.0542) (.0478)	-0.2949 ± .0420 ± .0417 (.0366) (.0327)
	b_{20}	0.4	0.4521 ± .1948 ± .1877 (.2043) (.1809)	0.4081 ± .0919 ± .0915 (.1044) (.0946)	0.3894 ± .0855 ± .0849 (.0719) (.0667)
	b_{21}	-1.8	-1.6957 ± .2117 ± .1842 (.2133) (.1807)	-1.7873 ± .1252 ± .1246 (.1059) (.0941)	-1.8021 ± .0714 ± .0713 (.0735) (.0663)
	a_{31}	-0.5	-0.4626 ± .1141 ± .1078 (.1241) (.0998)	-0.5066 ± .0557 ± .0553 (.0629) (.0541)	-0.4829 ± .0450 ± .0416 (.0431) (.0371)
	a_{32}	0.2	0.2296 ± .1181 ± .1143 (.1195) (.0999)	0.2075 ± .0674 ± .0669 (.0610) (.0537)	0.1929 ± .0350 ± .0343 (.0419) (.0370)
	b_{30}	-0.4	-0.5450 ± .2201 ± .1655 (.2048) (.1820)	-0.3957 ± .1166 ± .1165 (.1044) (.0961)	-0.4208 ± .0852 ± .0826 (.0719) (.0668)
	b_{31}	1.7	1.7049 ± .1995 ± .1994 (.2149) (.1820)	1.7056 ± .1012 ± .1011 (.1083) (.0966)	1.6985 ± .0707 ± .0707 (.0740) (.0670)
	a_{41}	0.9	0.8965 ± .0150 ± .0146 (.0569) (.0038)	0.9012 ± .0070 ± .0069 (.0342) (.0023)	0.9003 ± .0040 ± .0040 (.0232) (.0016)
	a_{42}	0.95	0.9533 ± .0185 ± .0182 (.0544) (.0038)	0.9505 ± .0104 ± .0104 (.0329) (.0022)	0.9484 ± .0080 ± .0079 (.0223) (.0015)
	b_{40}	1.0	0.9504 ± .3087 ± .3047 (.2056) (.0626)	0.9543 ± .1980 ± .1926 (.1043) (.0335)	1.0003 ± .1468 ± .1468 (.0720) (.0253)
	b_{41}	1.0	0.8598 ± .3631 ± .3349 (.2174) (.0635)	0.9543 ± .1893 ± .1837 (.1096) (.0341)	1.0187 ± .1258 ± .1244 (.0756) (.0256)
	c_1	-0.5	0.2483 ± .8105 ± .3115 (.3304) (.1148)	-0.4664 ± .1218 ± .1171 (.0984) (.0702)	-0.4715 ± .0726 ± .0667 (.0675) (.0508)
	d_1	-0.2	0.4205 ± .6887 ± .2989 (.3361) (.1186)	-0.1808 ± .1382 ± .1368 (.1034) (.0801)	-0.1783 ± .0703 ± .0668 (.0712) (.0576)

Table 5.5 (continued)

Basic IVAML and Refined IVAML for Model 3 with S = 4.0

Method	Parameter	True Value	Number of Samples: 400 (R_T^2=0.81)	1000 (R_T^2=0.80)	2000 (R_T^2=0.80)
Refined IVAML	a_{11}	-1.5	-1.4987 ± .0407 ± .0407 (.0345) (.0387)	-1.5001 ± .0246 ± .0246 (.0216) (.0216)	-1.5000 ± .0123 ± .0123 (.0153) (.0152)
	a_{12}	0.7	0.6980 ± .0341 ± .0340 (.0293) (.0335)	0.6997 ± .0206 ± .0206 (.0186) (.0186)	0.6990 ± .0098 ± .0097 (.0131) (.0130)
	b_{10}	1.0	0.9898 ± .1407 ± .1403 (.1583) (.1686)	1.0077 ± .0926 ± .0923 (.0906) (.0908)	1.0066 ± .0609 ± .0605 (.0633) (.0633)
	b_{11}	0.5	0.5030 ± .2200 ± .2200 (.2148) (.2274)	0.4876 ± .1501 ± .1496 (.1223) (.1222)	0.4879 ± .0747 ± .0737 (.0853) (.0851)
	a_{21}	-0.5	-0.5737 ± .1236 ± .0992 (.1046) (.0898)	-0.4879 ± .0432 ± .0415 (.0476) (.0487)	-0.4970 ± .0473 ± .0472 (.0329) (.0331)
	a_{22}	-0.3	-0.2389 ± .1207 ± .1041 (.1007) (.0889)	-0.3086 ± .0513 ± .0506 (.0468) (.0478)	-0.2987 ± .0422 ± .0422 (.0325) (.0327)
	b_{20}	0.4	0.3800 ± .1428 ± .1413 (.1805) (.1809)	0.4061 ± .0859 ± .0857 (.0937) (.0946)	0.3963 ± .0812 ± .0811 (.0664) (.0667)
	b_{21}	-1.8	-1.7167 ± .1708 ± .1490 (.1807) (.1807)	-1.8026 ± .1086 ± .1085 (.0933) (.0941)	-1.7994 ± .0665 ± .0665 (.0661) (.0663)
	a_{31}	-0.5	-0.4671 ± .0985 ± .0929 (.0991) (.0998)	-0.5054 ± .0523 ± .0521 (.0528) (.0541)	-0.4831 ± .0430 ± .0395 (.0373) (.0371)
	a_{32}	0.2	0.2146 ± .0754 ± .0740 (.0985) (.0999)	0.2073 ± .0693 ± .0689 (.0522) (.0537)	0.1919 ± .0359 ± .0350 (.0369) (.0370)
	b_{30}	-0.4	-0.5552 ± .2097 ± .1411 (.1785) (.1820)	-0.3924 ± .1097 ± .1094 (.0952) (.0961)	-0.4224 ± .0750 ± .0716 (.0666) (.0668)
	b_{31}	1.7	1.7292 ± .1607 ± .1580 (.1845) (.1820)	1.7147 ± .0986 ± .0975 (.0962) (.0966)	1.7025 ± .0667 ± .0667 (.0671) (.0670)
	a_{41}	0.9	0.9002 ± .0030 ± .0030 (.0041) (.0038)	0.9006 ± .0022 ± .0021 (.0023) (.0023)	0.8998 ± .0015 ± .0015 (.0015) (.0016)
	a_{42}	0.95	0.9499 ± .0028 ± .0028 (.0040) (.0038)	0.9494 ± .0019 ± .0018 (.0022) (.0022)	0.9500 ± .0016 ± .0016 (.0015) (.0015)
	b_{40}	1.0	1.0060 ± .0524 ± .0521 (.0661) (.0626)	1.0009 ± .0351 ± .0351 (.0332) (.0335)	1.0042 ± .0230 ± .0226 (.0252) (.0253)
	b_{41}	1.0	1.0006 ± .0660 ± .0660 (.0671) (.0635)	1.0117 ± .0257 ± .0228 (.0337) (.0341)	0.9968 ± .0221 ± .0219 (.0255) (.0256)
	c_1	-0.5	-0.3684 ± .3234 ± .2954 (.2537) (.1148)	-0.4208 ± .1265 ± .0987 (.0783) (.0702)	-0.4516 ± .1052 ± .0934 (.0573) (.0508)
	d_1	-0.2	-0.1861 ± .2958 ± .2955 (.2592) (.1186)	-0.1044 ± .1423 ± .1054 (.0891) (.0801)	-0.1440 ± .1104 ± .0952 (.0649) (.0576)

Table 5.6

Basic IVAML and Refined IVAML for Model 3 with S = 20.0

Method	Parameter	True Value	Number of Samples: 400 (R_T^2 = 0.88)	1000 (R_T^2 = 0.94)	2000 (R_T^2 = 0.95)
Basic IVAML	a_{11}	-1.5	-1.4558 ± .0593 ± .0395 (.0649) (.0239)	-1.5062 ± .0221 ± .0212 (.0292) (.0100)	-1.4972 ± .0206 ± .0204 (.0195) (.0068)
	a_{12}	0.7	0.6553 ± .0585 ± .0377 (.0636) (.0206)	0.7045 ± .0186 ± .0180 (.0285) (.0086)	0.6971 ± .0176 ± .0173 (.0190) (.0058)
	b_{10}	1.0	1.0417 ± .1977 ± .1933 (.1199) (.1040)	1.0094 ± .0727 ± .0721 (.0489) (.0420)	1.0085 ± .0545 ± .0538 (.0322) (.0284)
	b_{11}	0.5	0.6850 ± .2252 ± .1283 (.1431) (.1403)	0.4832 ± .1005 ± .0991 (.0579) (.0566)	0.5003 ± .0724 ± .0724 (.0380) (.0382)
	a_{21}	-0.5	-0.5900 ± .1010 ± .0457 (.0671) (.0554)	-0.4843 ± .0264 ± .0213 (.0255) (.0226)	-0.4977 ± .0219 ± .0217 (.0165) (.0149)
	a_{22}	-0.3	-0.2286 ± .0874 ± .0505 (.0662) (.0548)	-0.3127 ± .0275 ± .0244 (.0252) (.0221)	-0.2975 ± .0190 ± .0188 (.0163) (.0147)
	b_{20}	0.4	0.4212 ± .0857 ± .0831 (.1193) (.1116)	0.4015 ± .0407 ± .0407 (.0489) (.0438)	0.3972 ± .0382 ± .0381 (.0322) (.0299)
	b_{21}	-1.8	-1.6666 ± .1585 ± .0855 (.1235) (.1115)	-1.7939 ± .0558 ± .0555 (.0495) (.0436)	-1.8019 ± .0318 ± .0317 (.0329) (.0297)
	a_{31}	-0.5	-0.4956 ± .0426 ± .0424 (.0677) (.0616)	-0.5074 ± .0256 ± .0245 (.0294) (.0251)	-0.4935 ± .0199 ± .0188 (.0193) (.0167)
	a_{32}	0.2	0.2608 ± .0752 ± .0444 (.0658) (.0616)	0.2023 ± .0297 ± .0296 (.0285((.0249)	0.1981 ± .0156 ± .0154 (.0188) (.0166)
	b_{30}	-0.4	-0.5540 ± .1714 ± .0755 (.1196) (.1123)	-0.4033 ± .0519 ± .0518 (.0489) (.0445)	-0.4122 ± .0391 ± .0371 (.0322) (.0300)
	b_{31}	1.7	1.7896 ± .1246 ± .0866 (.1243) (.1123)	1.7097 ± .0461 ± .0450 (.0506) (.0448)	1.7002 ± .0321 ± .0321 (.0331) (.0300)
	a_{41}	0.9	0.8990 ± .0059 ± .0058 (.0317) (.0024)	0.9017 ± .0034 ± .0029 (.0158) (.0010)	0.9003 ± .0017 ± .0017 (.0103) (.0007)
	a_{42}	0.95	0.9555 ± .0094 ± .0076 (.0303) (.0023)	0.9516 ± .0049 ± .0046 (.0152) (.0010)	0.9495 ± .0034 ± .0034 (.0100) (.0007)
	b_{40}	1.0	0.9703 ± .1514 ± .1485 (.1201) (.0386)	0.9627 ± .0938 ± .0861 (.0488) (.0155)	1.0016 ± .0654 ± .0654 (.0322) (.0113)
	b_{41}	1.0	0.8928 ± .1819 ± .1469 (.1269) (.0392)	0.9711 ± .0870 ± .0821 (.0513) (.0158)	1.0099 ± .0565 ± .0556 (.0338) (.0115)
	c_1	-0.5	0.2138 ± .7272 ± .1386 (1.1881) (.1258)	-0.4830 ± .1034 ± .1020 (.0908) (.0638)	-0.4405 ± .1106 ± .0932 (.0681) (.0507)
	d_1	-0.2	0.1809 ± .4322 ± .2042 (1.1902) (.1174)	-0.1782 ± .1256 ± .1237 (.0963) (.0745)	-0.1457 ± .1076 ± .0939 (.0717) (.0574)

Table 5.6 (continued)

Basic IVAML and Refined IVAML for Model 3 with S = 20.0

Method	Parameter	True Value	Number of Samples: 400 (R_T^2=0.95)	1000 (R_T^2=0.95)	2000 (R_T^2=0.95)
Refined IVAML	a_{11}	-1.5	-1.4993 ± .0177 ± .0177 (.0166) (.0239)	-1.5000 ± .0109 ± .0109 (.0102) (.0100)	-1.5001 ± .0053 ± .0053 (.0068)
	a_{12}	0.7	0.6994 ± .0145 ± .0145 (.0136) (.0206)	0.7001 ± .0091 ± .0091 (.0088) (.0086)	0.6996 ± .0043 ± .0042 (.0058)
	b_{10}	1.0	0.9823 ± .0725 ± .0703 (.0847) (.1040)	1.0009 ± .0411 ± .0411 (.0424) (.0420)	1.0035 ± .0272 ± .0270 (.0284)
	b_{11}	0.5	0.5116 ± .1098 ± .1092 (.1156) (.1403)	0.4960 ± .0674 ± .0673 (.0572) (.0566)	0.4939 ± .0330 ± .0325 (.0383)
	a_{21}	-0.5	-0.5715 ± .0862 ± .0480 (.0659) (.0554)	-0.4901 ± .0215 ± .0191 (.0219) (.0226)	-0.4991 ± .0213 ± .0213 (.0148)
	a_{22}	-0.3	-0.2391 ± .0784 ± .0495 (.0616) (.0548)	-0.3072 ± .0237 ± .0226 (.0216) (.0221)	-0.2992 ± .0191 ± .0191 (.0146)
	b_{20}	0.4	0.3601 ± .0825 ± .0722 (.1042) (.1116)	0.3996 ± .0390 ± .0390 (.0435) (.0438)	0.3998 ± .0362 ± .0362 (.0298)
	b_{21}	-1.8	-1.6984 ± .1238 ± .0708 (.1116) (.1115)	-1.8052 ± .0493 ± .0490 (.0433) (.0436)	-1.8009 ± .0297 ± .0297 (.0296)
	a_{31}	-0.5	-0.4749 ± .0442 ± .0364 (.0554) (.0616)	-0.5050 ± .0234 ± .0229 (.0247) (.0251)	-0.4930 ± .0190 ± .0177 (.0166)
	a_{32}	0.2	0.2019 ± .0397 ± .0396 (.0544) (.0616)	0.1993 ± .0302 ± .0302 (.0243) (.0249)	0.1969 ± .0158 ± .0155 (.0166)
	b_{30}	-0.4	-0.5305 ± .1489 ± .0717 (.1043) (.1123)	-0.4013 ± .0486 ± .0486 (.0442) (.0445)	-0.4117 ± .0344 ± .0323 (.0299)
	b_{31}	1.7	1.7703 ± .0995 ± .0704 (.1067) (.1123)	1.7174 ± .0464 ± .0430 (.0447) (.0448)	1.7014 ± .0300 ± .0300 (.0300)
	a_{41}	0.9	0.9000 ± .0013 ± .0013 (.0027) (.0024)	0.9003 ± .0010 ± .0009 (.0010) (.0010)	0.8999 ± .0007 ± .0006 (.0007)
	a_{42}	0.95	0.9500 ± .0013 ± .0013 (.0026) (.0023)	0.9497 ± .0008 ± .0008 (.0010) (.0010)	0.9500 ± .0007 ± .0007 (.0007)
	b_{40}	1.0	1.0039 ± .0236 ± .0233 (.0433) (.0386)	1.0003 ± .0156 ± .0156 (.0151) (.0155)	1.0019 ± .0103 ± .0101 (.0113)
	b_{41}	1.0	0.9964 ± .0299 ± .0296 (.0440) (.0392)	1.0050 ± .0113 ± .0102 (.0153) (.0158)	0.9986 ± .0099 ± .0098 (.0115)
	c_1	-0.5	-0.1119 ± .3912 ± .0493 (1.6537) (.1258)	-0.3831 ± .1290 ± .0546 (.0689) (.0638)	-0.4264 ± .1243 ± .1002 (.0579)
	d_1	-0.2	-0.1812 ± .0641 ± .0613 (1.6476) (.1174)	-0.0328 ± .1770 ± .0580 (.0799) (.0745)	-0.1179 ± .1307 ± .1017 (.0654)

Table 5.7

Basic IVAML and Refined IVAML for Model 4 with S = 2.0

Method	Parameter	True Value	Number of Samples: 200 (R_T^2 = 0.61)	500 (R_T^2 = 0.65)	2000 (R_T^2 = 0.66)
Basic IVAML	a_{11}	-1.5	-1.4956 ± .1089 ± .1088 (.1456) (.0477)	-1.5060 ± .0889 ± .0887 (.1001) (.0349)	-1.5072 ± .0464 ± .0458 (.0520) (.0181)
	a_{12}	0.7	0.6965 ± .1079 ± .1079 (.1453) (.0425)	0.7046 ± .0804 ± .0803 (.0981) (.0303)	0.7061 ± .0393 ± .0388 (.0506) (.0156)
	b_{10}	1.0	0.8883 ± .4038 ± .3881 (.2440) (.2072)	0.9781 ± .2096 ± .2085 (.1733) (.1507)	1.0253 ± .1145 ± .1116 (.0857) (.0758)
	b_{11}	0.5	0.4726 ± .3709 ± .3699 (.2903) (.2789)	0.4892 ± .3357 ± .3355 (.2050) (.2028)	0.4618 ± .1796 ± .1755 (.1017) (.1019)
	a_{21}	-1.5	-1.4418 ± .1832 ± .1737 (.1718) (.0545)	-1.4798 ± .0832 ± .0807 (.1308) (.0435)	-1.5023 ± .0450 ± .0450 (.0604) (.0201)
	a_{22}	0.7	0.6502 ± .1810 ± .1740 (.1718) (.0516)	0.6770 ± .0818 ± .0785 (.1309) (.0414)	0.7042 ± .0471 ± .0469 (.0604) (.0196)
	b_{21}	1.0	1.0088 ± .2723 ± .2721 (.2425) (.1368)	0.9970 ± .2044 ± .2044 (.1729) (.1013)	0.9666 ± .1382 ± .1341 (.0857) (.0507)
	c_1	-0.5	-0.4973 ± .1435 ± .1435 (.1734) (.1715)	-0.4719 ± .0907 ± .0862 (.1200) (.1106)	-0.4944 ± .0719 ± .0717 (.0612) (.0549)
	d_1	-0.2	-0.1292 ± .1809 ± .1664 (.1890) (.1967)	-0.1371 ± .1166 ± .0982 (.1283) (.1254)	-0.1772 ± .0726 ± .0689 (.0652) (.0624)
			200 (R_T^2 = 0.68)	500 (R_T^2 = 0.67)	2000 (R_T^2 = 0.67)
Refined IVAML	a_{11}	-1.5	-1.5018 ± .0551 ± .0551 (.0469) (.0477)	-1.5089 ± .0344 ± .0332 (.0335) (.0349)	-1.4979 ± .0186 ± .0184 (.0183) (.0181)
	a_{12}	0.7	0.7042 ± .0486 ± .0484 (.0415) (.0425)	0.7099 ± .0327 ± .0312 (.0291) (.0303)	0.6980 ± .0146 ± .0144 (.0157) (.0156)
	b_{10}	1.0	0.9804 ± .2232 ± .2223 (.2006) (.2072)	0.9586 ± .1332 ± .1266 (.1492) (.1507)	1.0081 ± .0752 ± .0747 (.0761) (.0758)
	b_{11}	0.5	0.4657 ± .2611 ± .2589 (.2699) (.2789)	0.5027 ± .1751 ± .1751 (.1985) (.2028)	0.5043 ± .1097 ± .1096 (.1025) (.1019)
	a_{21}	-1.5	-1.4894 ± .0691 ± .0683 (.0522) (.0545)	-1.5009 ± .0416 ± .0416 (.1432) (.0435)	-1.5031 ± .0164 ± .0161 (.0198) (.0201)
	a_{22}	0.7	0.6993 ± .0639 ± .0639 (.0496) (.0516)	0.6990 ± .0381 ± .0381 (.0410) (.0414)	0.7050 ± .0166 ± .0158 (.0194) (.0196)
	b_{21}	1.0	1.0117 ± .1719 ± .1715 (.1320) (.1368)	1.0044 ± .1157 ± .1156 (.1004) (.1013)	0.9936 ± .0458 ± .0453 (.0507) (.0507)
	c_1	-0.5	-0.3818 ± .2012 ± .1629 (.1747) (.1715)	-0.4199 ± .1256 ± .0967 (.1169) (.1106)	-0.4539 ± .0899 ± .0771 (.0557) (.0549)
	d_1	-0.2	-0.0802 ± .2022 ± .1630 (.1965) (.1967)	-0.1118 ± .1379 ± .1061 (.1313) (.1254)	-0.1410 ± .0999 ± .0806 (.0634) (.0624)

Table 5.8

Basic IVAML and Refined IVAML for Model 4 with S = 10.0

Method	Parameter	True Value	Number of Samples: 200 (R_T^2 = 0.89)	500 (R_T^2 = 0.90)	2000 (R_T^2 = 0.90)
Basic IVAML	a_{11}	-1.5	-1.5015 ± .0418 ± .0417 (.0645) (.0219)	-1.5052 ± .0399 ± .0396 (.0450) (.0157)	-1.5037 ± .0209 ± .0205 (.0235) (.0082)
	a_{12}	0.7	0.7014 ± .0411 ± .0411 (.0644) (.0195)	0.7050 ± .0361 ± .0358 (.0441) (.0136)	0.7033 ± .0178 ± .0175 (.0229) (.0070)
	b_{10}	1.0	0.9507 ± .1808 ± .1739 (.1129) (.0950)	0.9864 ± .0943 ± .0933 (.0784) (.0679)	1.0113 ± .0509 ± .0496 (.0389) (.0342)
	b_{11}	0.5	0.4854 ± .1667 ± .1661 (.1336) (.1279)	0.4941 ± .1510 ± .1509 (.0926) (.0913)	0.4825 ± .0801 ± .0782 (.0459) (.0460)
	a_{21}	-1.5	-1.4822 ± .0569 ± .0540 (.0753) (.0250)	-1.4945 ± .0369 ± .0365 (.0579) (.0196)	-1.5019 ± .0207 ± .0206 (.0270) (.0091)
	a_{22}	0.7	0.6855 ± .0575 ± .0556 (.0753) (.0237)	0.6924 ± .0352 ± .0343 (.0579) (.0186)	0.7028 ± .0216 ± .0214 (.0271) (.0088)
	b_{21}	1.0	1.0054 ± .1237 ± .1236 (.1123) (.0627)	0.9932 ± .0979 ± .0976 (.0783) (.0456)	0.9854 ± .0610 ± .0592 (.0389) (.0229)
	c_1	-0.5	-0.5496 ± .1403 ± .1312 (.1566) (.1627)	-0.5058 ± .0859 ± .0857 (.1156) (.1087)	-0.5463 ± .0634 ± .0434 (.0579) (.0537)
	d_1	-0.2	-0.1637 ± .1615 ± .1574 (.1725) (.1913)	-0.1665 ± .1015 ± .0958 (.1241) (.1243)	-0.2231 ± .0497 ± .0440 (.0621) (.0617)
			200 (R_T^2 = 0.91)	500 (R_T^2 = 0.91)	2000 (R_T^2 = 0.91)
Refined IVAML	a_{11}	-1.5	-1.4992 ± .0242 ± .0242 (.0220) (.0219)	-1.5036 ± .0157 ± .0153 (.0155) (.0157)	-1.4993 ± .0083 ± .0082 (.0083) (.0082)
	a_{12}	0.7	0.7013 ± .0219 ± .0218 (.0196) (.0195)	0.7044 ± .0148 ± .0141 (.0135) (.0136)	0.6992 ± .0065 ± .0065 (.0072) (.0070)
	b_{10}	1.0	0.9999 ± .0988 ± .0988 (.0931) (.0950)	0.9868 ± .0594 ± .0579 (.0677) (.0679)	1.0016 ± .0336 ± .0335 (.0344) (.0342)
	b_{11}	0.5	0.4740 ± .1194 ± .1165 (.1253) (.1279)	0.4960 ± .0802 ± .0801 (.0907) (.0913)	0.5037 ± .0491 ± .0489 (.0462) (.0460)
	a_{21}	-1.5	-1.4936 ± .0296 ± .0289 (.0246) (.0250)	-1.5000 ± .0179 ± .0179 (.0196) (.0196)	-1.5014 ± .0075 ± .0073 (.0091) (.0091)
	a_{22}	0.7	0.6991 ± .0286 ± .0285 (.0235) (.0237)	0.6991 ± .0161 ± .0161 (.0187) (.0186)	0.7022 ± .0074 ± .0071 (.0090) (.0088)
	b_{21}	1.0	1.0083 ± .0753 ± .0749 (.0625) (.0627)	1.0020 ± .0512 ± .0512 (.0458) (.0456)	0.9966 ± .0205 ± .0203 (.0233) (.0229)
	c_1	-0.5	-0.4355 ± .1722 ± .1596 (.1517) (.1627)	-0.4288 ± .1244 ± .1019 (.1094) (.1087)	-0.4886 ± .0671 ± .0661 (.0506) (.0537)
	d_1	-0.2	-0.0954 ± .1925 ± .1616 (.1748) (.1913)	-0.1084 ± .1399 ± .1057 (.1243) (.1243)	-0.1640 ± .0772 ± .0682 (.0587) (.0617)

Comparison of the Pierce minimum standard errors with the empirical quantitities in columns 2, 3 and 4 of Tables 5.1 to 5.8 shows that the refined IVAML algorithm is certainly asymptotically efficient and seems to obtain minimum variance with respect to the system parameters even for low sample size. However, as in the case of SISO and multiple input-multiple output (MIMO) models (Young and Jakeman, 1979c; Jakeman and Young, 1979) the algorithm is, at best, only asymptotically efficient in the case of the noise parameters. Here there is clearly an improvement in the comparison between the estimated (column 4) and Pierce standard errors (column 5) as sample size increases, but a very large number of samples is obviously needed to show good agreement.

The parameter values and their standard errors (columns 2 and 3) given by basic IVAML were quite insensitive to initialisation. Most of the time, almost identical results were obtained whether the poor SISO estimates or the exact values were used initially. Similarly, good convergence was obtained in the case of refined IVAML whether initialisation was from basic IVAML results or exact values. The only initialisation problem came with Model 3 for basic IVAML using S = 4 and 20 with N = 400. To obtain convergence, the SISO estimate of the noise-free output corresponding to the transfer function with the largest steady state gain was subtracted from the noisy output, so that SISO estimation could proceed on the other two inputs in the presence of less 'noise'. In these difficult cases, the multiple auxiliary models concept of Finigan (1978) may be of use for improving convergence. He considers these more complex updating schemes justifiable for small signal to noise ratios.

The following observations can be made in connection with the modeling of MIMO systems described by Jakeman and Young (1979). Firstly, in order to obtain sensible parameter estimates in the case of Model 3 one needs $400/12 \cong 30$ samples per parameter. This is quite high as Jakeman and Young (1979) obtain sensible estimates for 20 system parameters with only 150 samples for a 2 input - 2 output system. The relatively high number of samples required for the successful estimation of the parameters in MITF models may be due to the fact that so much information is convoluted into a single output series. Secondly, the simulation results from Model 4 indicate that the algorithms presented in this chapter perform very well in those

cases where each transfer function has the same characteristic polynomial, even where the algorithm is not supplied with this information.

Investigations into the application of successive SISO model order detection for the identification of MITF model structures were carried out for Models 1 and 2 and the results are displayed in Tables 5.9 and 5.10 respectively. In both cases the most difficult test was chosen, namely that one with the lowest sample size and the smaller signal to noise ratio. While the procedure correctly identifies the order of both transfer functions in Model 1, the results in Table 5.10 show that it is not as successful for Model 2 since there is some ambiguity about which is the correct model order for inputs 1 and 3. In the SISO situation, the choice of best model is usually clear cut since the correct model's EVN is singularly low in relation to higher order models (Young et al., 1980). However, in the cases mentioned above, some of the models within one order of the true model possess very similar criteria. In such cases, the intuitive procedure is to choose the model with the lowest number of parameters. From Table 5.10, this would mean that the model chosen would be incorrect on the first transfer function.

It appears that this difficulty in precisely identifying the correct model may be a consequence of the input perturbations selected in the Monte Carlo experiments. For example, simply changing one of the inputs to a sinusoidal type made the EVN criterion more discerning. It should be noted, however, that when some slightly incorrect model orders were chosen for Models 1 and 2 then the standard errors of the parameters in these cases were as low as those obtained when the correct model order was chosen. Also, some incorrect model orders gave values of R_T^2 which were almost the same as those obtained by selection of the correct model order. In other words, the 'wrong' models with the equally attractive EVN and R_T^2 criteria obtained by SISO identification may well be satisfactory for most purposes. In addition, this ambiguity has not been noticed in the analysis of real data. For example, in the next section a two input process for modeling atmospheric ozone levels is discussed and the model order seems well defined in this case.

One final point should be made in relation to Table 5.10: for

models of higher order than the correct one, the R_T^2 values do improve a little more than has been experienced in the case of the SISO systems. This is due to the large amount of 'noise' involved when each SISO sub-system of the decomposed MISO system is considered individually. Redundant parameters will always try to explain more of the output and, in these circumstances, there is much more 'noise' to be explained than that indicated by the S values.

Table 5.9

Identification of Model Order for Model 1 with 200 Samples and S=2

Non-Zero Model Parameters	Input 1		Input 2	
	ln(EVN)	R_T^2	ln(EVN)	R_T^2
b_{io} a_{i1}	-2.255	0.2103	*	*
b_{io} a_{i1} a_{i2}	-0.702	0.2821	*	*
b_{io} b_{i1} a_{i1} a_{i2}	-2.543	0.5364	-2.167	0.2094
b_{io} b_{i1} b_{i2} a_{i1} a_{i2}	-2.444	0.5471	-2.036	0.2150
b_{io} b_{i1} a_{i1} a_{i2} a_{i3}	+	0.5463	-2.270	0.2084
b_{io} b_{i1} b_{i2} a_{i1} a_{i2} a_{i3}	*	*	*	*

+ denotes no satisfactory convergence

* denotes total lack of convergence

Table 5.10

Identification of Model Order for Model 2 with 300 Samples and S=3

Assumed Model Structure	Input 1		Input 2		Input 3	
	ln(EVN)		ln(EVN)		ln(EVN)	
(1,1)	-3.010	0.1877	+	-0.0204	-1.106	-0.0234
(2,1)	-3.503	0.5020	3.071	-0.0217	-0.716	-0.0186
(2,2)	-3.088	0.5133	-2.639	0.0726	-2.517	0.0518
(2,3)	-3.053	0.5096	-2.556	0.0739	-2.452	0.0542
(3,2)	+	0.5246	-2.191	0.0737	-2.576	0.0577
(4,2)	*	*	*	*	*	*

\+ denotes no satisfactory convergence

* denotes total lack of convergence

5.6 Estimating Missing Ozone Data

In the introduction to this chapter reference was made to the problem of poisoning of agricultural field-workers by pesticide residues, and to the suspicion that atmospheric ozone levels may have played a significant role in the conversion of the pesticide parathion to its oxygen analogue paraoxon (Spear et al., 1978). As part of a field study by Spear (1978) to investigate this problem a continuous ozone analyser was set up in a citrus grove in which parathion was routinely applied. The citrus grove was located near Lindcove in the San Joaquin valley and hourly average ozone measurements were taken over the period 19 July to 30 September 1977. Breakdowns of the ozone analyser resulted, however, in several periods for which measurements for the citrus grove were not available, although routine measurements of ozone levels at two nearby locations over the same period of time were made by State or municipal bodies. These were in the city of Visalia (population approximately 40,000) and the Kearney station which was situated at the rural location of Reedley. The position of the three sites is shown on the location map in Figure 5.4 and plots of the hourly average ozone measurements taken at each are shown in Figure

5.5, where periods of missing data are indicated by arrows. It is clear from Figure 5.5 that the ozone levels measured in Visalia are significantly lower than those at Kearney or Lindcove. This can probably be explained in terms of Visalia being the only urban location of the three sites. The greater density of motor vehicles there will lead, in general, to higher atmospheric levels of exhaust pollutants than in rural locations. Nitric oxide, a component of motor vehicle exhaust gases, is well known to react very rapidly with ozone, resulting in its destruction (see Leighton, 1961). It should also be noted that the ozone measurements at Visalia and Kearney are reported to the nearest ten parts per billion by volume (ppb) while those at Lindcove are reported to the nearest one ppb, where 1 billion = 10^9.

For several reasons it was believed that the ozone measurements at Visalia and Kearney provided the most suitable basis for estimating the missing data at Lindcove. The prevailing winds in the region are from the west and Visalia and Kearney were the nearest locations west of Lindcove at which routine ozone measurements were available. Furthermore it was thought that the levels of ozone experienced at Lindcove were strongly related to the emission of photochemical smog

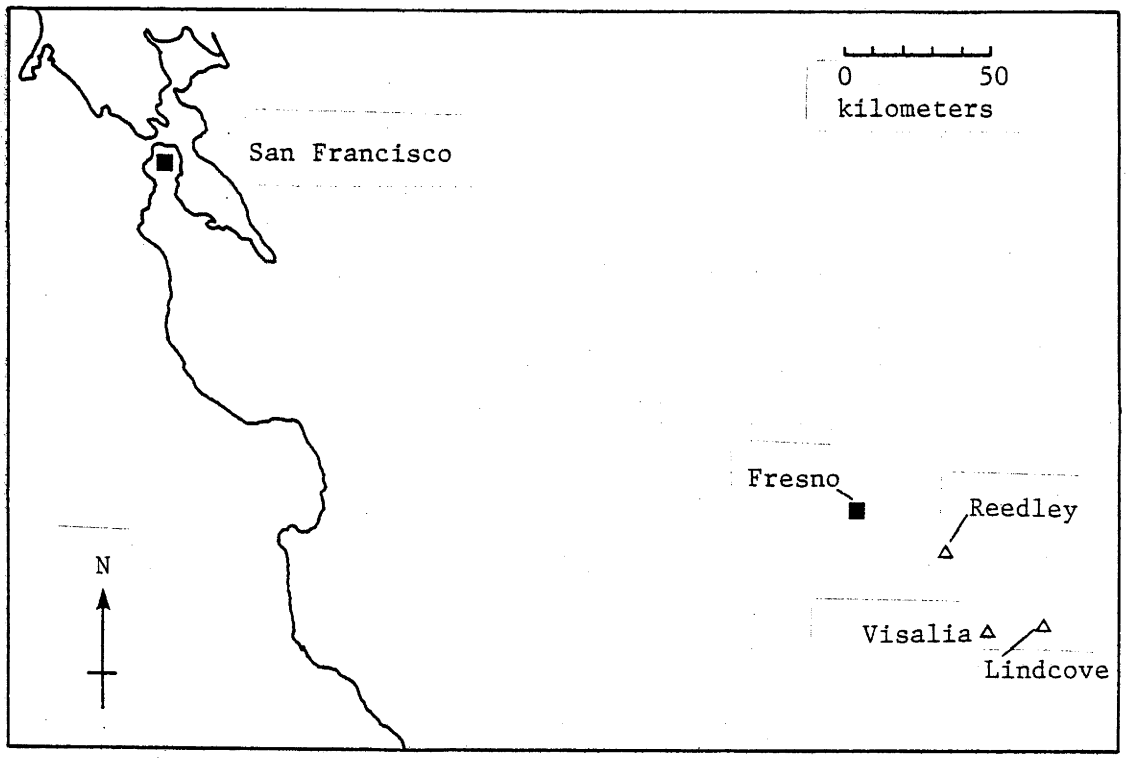


Figure 5.4 Location of the three sites in the San Joaquin valley at which ozone measurements were taken.

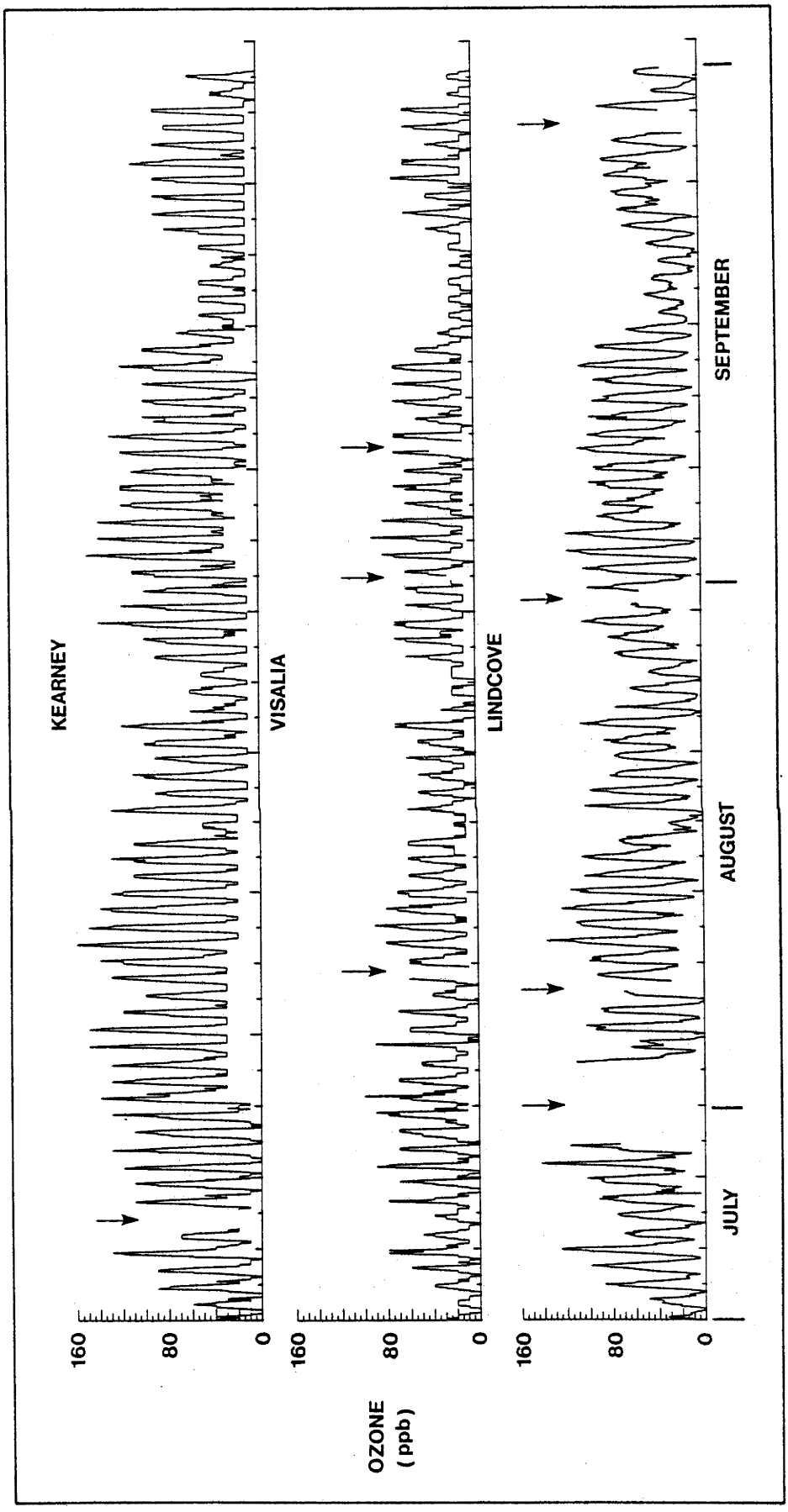


Figure 5.5 Plots of the hourly average ozone measurements (in parts per billion) made at three sites in the San Joaquin valley over the period 19 July to 30 September 1977. Arrows indicate periods of missing data.

precursors (nitrogen oxides and non-methane hydrocarbons) in the heavily populated regions to the west.

In the rest of this section we first investigate whether simple linear dynamic SISO models provide a useful description of the relationship between ozone levels at the different sites. Because the objective of such modeling is the estimation of missing Lindcove data, and owing to the direction of the prevailing winds, we choose the time series at Lindcove as the output. We then investigate various MISO models to see if better descriptions of the data at Lindcove may be obtained. Finally, the usefulness of such models for estimating missing ozone data is demonstrated by a comparison of known and predicted ozone levels over a 24 hour period.

5.6.1 Simple SISO Models

Because of the wish subsequently to investigate the applicability of MISO models, and to make direct comparisons with the SISO models, it was necessary to restrict attention to those periods for which data are available at all three sites and not simply at Lindcove. These periods are shown in Table 5.11, and it can be seen that nearly all observations have been utilised.

Table 5.11

Periods Selected for Ozone Modeling

Period Number	Number of Hourly Samples	Period in 1977 Covered by Data
1	128	1700 hours 19 July - 2400 hours 24 July
2	93	0000 hours 26 July - 2100 hours 29 July
3	101	1600 hours 3 Aug. - 2000 hours 7 Aug.
4	515	0700 hours 9 Aug. - 1700 hours 30 Aug.
5	431	0700 hours 9 Sept. - 0500 hours 27 Sept.
6	61	1200 hours 28 Sept. - 2400 hours 30 Sept.

SISO models applied to the data from (subsequently abbreviated to 'SISO models between') Kearney and Lindcove and Visalia and Lindcove were identified using the basic IV procedures described in Chapter 3. For illustration the primary test statistics are shown in Table 5.12 for a range of model structures considered for the first period of data. It can be seen from the Table that a pure time delay of zero seems most appropriate in the model between Visalia and Lindcove, whereas a delay of one hour is most suitable between Kearney and Lindcove. This was also found to be the case in the identification of models for the remaining five periods. This pattern of time delays is consistent with the geographical relationship between the sites, and with the hypothesis that much of the ozone measured in the study region is due to advection of polluted air from the populated regions to the west. In addition, it is evident from Table 5.12 that the identified model structure in each case is a two parameter model comprising one 'a' and one 'b' parameter. Similar model structures were also identified for the other five periods.

In general, the identified models provide a reasonable description of the data. This is indicated by the values of R_T^2 shown in Table 5.13, and illustrated in Figure 5.6 for the model between Kearney and Lindcove over the first period. A minor problem is that these values vary between different time periods, although this was expected given the known complexity of meteorological systems and given that this complexity has not been incorporated explicitly in the models. More serious is the significant time variation of the parameters in the identified models, which indicates that they are not perfectly acceptable linear models. This time variation of the parameters is illustrated in Figure 5.7 for the identified model between Kearney and Lindcove for the first period. In theory it may be possible to account for this variation by consideration of known influences on photochemical smog formation. While this may be desirable as leading to models possessing more satisfactory properties it is probably not justified here given the stated requirements of the models.

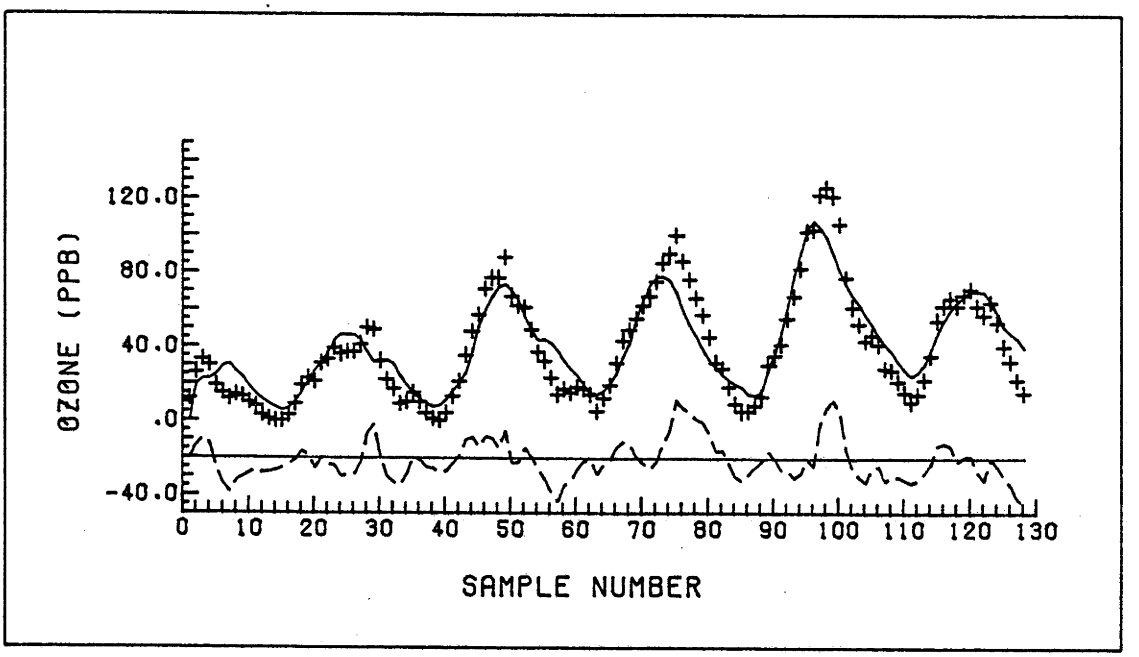


Figure 5.6 SISO model fit (—) to ozone data (+) at Lindcove over the first period, using data at Kearney as the input.

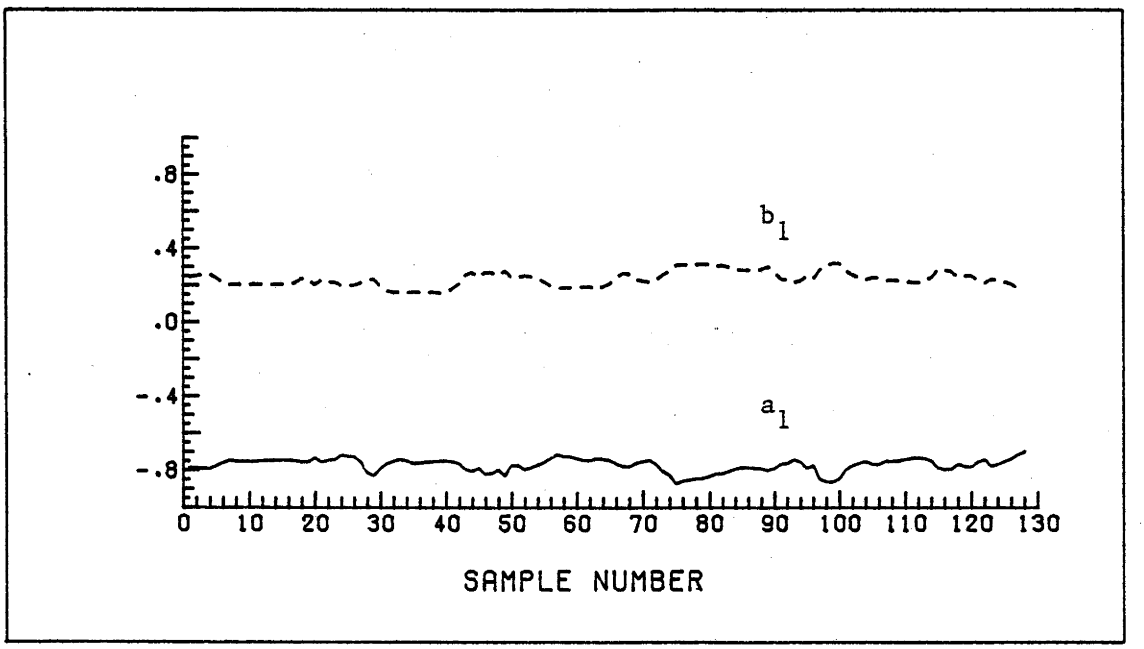


Figure 5.7 Estimated time variation of the parameters in the first-order SISO model relating ozone data at Kearney and Lindcove over the first period.

Table 5.12

Identification of Model Order for SISO Models of Ozone Levels for the First Period

Model Structure	R_T^2	ln(NEVN)	ln(EVN)
A: Input Site 1 (Kearney); Output Site Lindcove			
(0,1,0)	0.1292	-5.8695	-6.0158
(0,2,0)	0.4369	-3.3729	-3.6199
(0,1,1)	0.4045	-6.2934	-6.3739
(0,2,1)	0.6391	-3.2717	-4.0317
(1,1,0)	0.8033	-5.2515	-6.4179
*(1,1,1)	0.8582	-5.4745	-6.4750
(1,1,2)	0.8683	-5.3505	-6.1972
(1,2,1)	0.8740	-3.1143	-4.9644
(1,3,1)	0.8821	-2.6962	-4.5342
(2,1,1)	0.8734	-3.1216	-3.3531
(2,2,1)	0.8734	2.1698	-2.9472
(3,1,1)	0.8748	1.6817	0.0400
B: Input Site 2 (Visalia); Output Site Lindcove			
(0,1,0)	0.4754	-5.8918	-5.4040
(0,2,0)	0.6679	-2.4206	-3.4974
(0,1,1)	0.6655	-6.3716	-5.8444
(0,2,1)	0.7724	-3.6326	-3.8602
*(1,1,0)	0.8248	-4.7488	-5.4247
(1,1,1)	0.8258	-4.6044	-5.0936
(1,1,2)	0.7749	-4.0234	-4.3551
(1,2,0)	0.8466	-3.0564	-4.2389
(1,3,0)	0.8563	-2.3407	-3.8862
(2,1,0)	0.8401	-3.0658	-3.3805
(2,2,0)	0.8492	-1.6233	-2.9446
(3,1,0)	0.8439	-0.5561	-0.8672

*denotes identified model

Table 5.13
Comparison of R_T^2 for all First Order SISO Models

Period Number	(1,1,0) Model for Visalia - Lindcove	(1,1,1) Model for Kearney - Lindcove
1	0.8248	0.8582
2	0.7326	0.7653
3	0.6957	0.7013
4	0.8060	0.8210
5	0.7140	0.6494
6	0.7637	0.7987

The parameter values in all the identified SISO models were estimated by basic IV procedures since the time variation of the model parameters did not warrant the use of refined IVAML methods. They are shown in Table 5.14 with the conservative measures of the standard errors of the parameters provided by the basic IV algorithm. It is clear that while the system model parameters have similar values in several of the models, it is not possible to specify a single model for the whole period of approximately ten weeks which would provide a fit to the data that was as good as that provided by the separate models. In the next sub-section we look at the usefulness of MISO models for describing the relationship between the ozone levels measured at the three sites.

5.6.2 MISO Models

Multiple (two in this case) input models were investigated to see whether they 'explained' the ozone levels at the output site of Lindcove better than the SISO models described previously. The expectation was that MISO models would show improved performance (higher R_T^2) as they make use of all available data in contrast to the SISO models which could not simultaneously utilise the data from both input sites. The MISO models considered were of the forms (5.1)

and (5.2) described earlier in the chapter. In both, first order transfer functions were believed to be appropriate operators on each input because such transfer functions with appropriate pure time delays were identified in the previously described SISO models. In deriving the IV algorithms earlier in the chapter it was seen that the estimation procedure for models of the form (5.2) requires explicit provision of initial parameter estimates, whereas the IV estimation procedure for models of the form (5.1) generates initial estimates by a linear least squares method. The initial estimates for the form (5.2) were obtained either as the converged parameter values from the previously described SISO models or as converged parameter values from models of the form (5.1).

In order to illustrate the results obtained from these MISO models, the estimated parameters for the first period, together with a measure of their standard errors provided by basic IV procedures, are shown in Table 5.15. For the purposes of comparison, the previously reported parameters of the corresponding SISO models (see Table 5.14) are also shown, together with values of R_T^2. It is apparent from Table 5.15 that the MISO models provide better descriptions of the data than either of the SISO models. However, this is achieved only with associated significant increases in the estimated standard errors of the model parameters. This is often an indication of over-parameterisation but, in this case, it is more likely to be indicative of some correlation between the ozone measurements at the two input sites. This is analagous to the multicollinearity problem in regression analysis where partial linear dependence between the regressors leads to high estimation error variance even though the 'fit' to the data may appear to be very good (see Brownlee, 1965). It should also be noted that smaller standard errors were obtained from use of the models of the form (5.2) than models of the form (5.1), a result probably due to the allowance in the former for different denominator polynomials in the transfer functions operating on the different inputs. Nevertheless, the standard errors obtained from the model form (5.2) were still much larger than those obtained in the SISO models. In an attempt to improve on these results, in the next sub-section we overcome the problem of correlated inputs by creating a single input series in the form of a weighted average of the original two input series. Such a simple procedure simultaneously permits the attainment of two desirable objectives - the use of all the data and

the retention of the preferred, simpler SISO model form, with its better defined parameter estimates.

Table 5.14

Summary of Estimated Parameters[a] in SISO Models for the Six Periods

Model For	Period No	'a_1' parameter (and standard error[b])	'b' parameter (and standard error[b])
Kearney to Lindcove	1	-0.7859 (.0391)	0.2444 (.0395)
	2	-0.8518 (.0425)	0.1976 (.0459)
	3	-0.7167 (.0929)	0.1975 (.0627)
	4	-0.7759 (.0224)	0.2154 (.0197)
	5	-0.8170 (.0303)	0.1995 (.0294)
	6	-0.8250 (.0566)	0.2296 (.0665)
Visalia to Lindcove	1	-0.7633 (.0469)	0.4547 (.0813)
	2	-0.8665 (.0426)	0.2898 (.0737)
	3	-0.7980 (.0651)	0.3473 (.1017)
	4	-0.7846 (.0243)	0.3879 (.0401)
	5	-0.8256 (.0266)	0.3436 (.0464)
	6	-0.8458 (.0602)	0.3723 (.1246)

[a]Estimated using basic IV procedures

[b] Standard errors based on estimated variance of $\underset{\sim}{a}_k$ terms

5.6.3 SISO Models Using 'Averaged' Inputs

It was not clear *a priori* what weights should be applied to the input series from Kearney and Visalia in order to create a single input. This problem was approached by means of a sensitivity analysis in which the weights were systematically varied to produce a number of single input series. These were evaluated by identifying the model structure of the SISO models and comparing the test statistics arising from the basic IV estimation procedure. These are shown in Table 5.16,

Table 5.15

Comparison of SISO and MISO Models for the First Period

Model	Parameter Estimates[a]	Standard Error Estimates[a]	R_T^2
SISO (Kearney input)	a_1 = -0.7859	0.0393	0.8582
	b_1 = 0.2444	0.0393	
SISO (Visalia input)	a_1 = -0.7633	0.0469	0.8248
	b_0 = 0.4547	0.0813	
MISO[b]	a_1 = -0.7082	0.1889	
	b_1 = 0.0272	0.5014	
	b_2 = 0.0419	0.8537	0.8638
	b_3 = 0.1185	0.6597	
	b_4 = 0.2707	0.8275	
MISO[c]	a_{11} = -0.7485	0.0978	
	a_{12} = -0.6851	0.0785	
	b_{10} = 0.0240	0.0837	0.8643
	b_{11} = 0.1023	0.1025	
	b_{20} = 0.0573	0.1100	
	b_{21} = 0.2775	0.1375	

[a] Found from basic IV procedures

[b] Model of the ARMAX form (5.1)

[c] Model of the MITF form (5.2)

where it can be seen that any of the selected weights leads to models which have larger values of R_T^2 than either of the SISO models which use the single unweighted input series. While no single set of weights leads to a model form which maximizes the value of R_T^2 for all six periods, the 'optimum' weighting appears to be 0.6 x Visalia + 0.4 x Kearney. The differences in the values of ln(NEVN) are minor, indicating that the weighting procedure has not resulted in any significant increase in the standard errors of the model parameters.

It is evident from a comparison of Tables 5.15 and 5.16 that the

SISO models which arise from use of the 'optimum' weighting of the two inputs provide descriptions of the data relating to the first period that are superior to those provided by either of the MISO models. This is also found to be the case for the other five periods. It is believed these results provide a good example of the value of simple model forms, and illustrate how attempts to incorporate too much detail in more complex forms may be counter-productive. The 'optimum' SISO models still exhibit some time-variation of the parameters and so are not entirely acceptable linear models. Nevertheless, this time variation does not appear critical for two reasons. First, since the objective is to obtain estimates of missing data, it is anticipated that models which provide a better description of the existing data (that is, resulting in a higher R_T^2) will be more useful. Second, smaller standard errors of the model parameters will lead to forecasts which have more acceptable (narrower) confidence bands. While evaluation of the estimates for the actual missing data was not possible, in the next subsection the 'optimum' SISO models are used to generate estimates for a period in which output data were available, thus permitting a comparison of forecast with original observations.

5.6.4 Direct Evaluation of the SISO Model Forecasts

The strategy adopted to evaluate directly the usefulness of the 'optimum' SISO models for providing forecasts of ozone levels was to take the first 400 points of the fourth period, and estimate the model parameters for that portion of the period. Then, those model parameters were used to provide deterministic forecasts of the hourly ozone levels for the remainder of the period. These forecasts then could be compared with the observations made at Lindcove over the same period. The model parameters estimated by basic IV methods for the period on which the forecast was based were $a_1 = -0.7195$ and $b_0 = 0.3720$. A pure time delay of one hour was employed, and a R_T^2 of 0.8492 was recorded. The forecast of the hourly ozone levels was highly successful, and a comparison with the known levels is shown in Figure 5.8. The success of the forecasting is reflected in the value of R_T^2 of 0.8629, which was obtained as the correlation between known and forecast hourly ozone levels. If a noise model had been estimated stochastic forecasts could have been made. However, given the objective of obtaining forecasts of daily average ozone readings, the additional difficulty of providing stochastic forecasts did not seem warranted.

Table 5.16

Comparison of R_T^2 and ln(NEVN) for a Range of SISO Models which use 'Averaged' Inputs

Period	ln(NEVN)	ln(NEVN)	ln(NEVN)	ln(NEVN)
	(1,1,0) Model	(1,1,1)Model	(1,1,1)Model	(1,1,1) Model
	1.0V + 0.0K	0.9V + 0.1K	0.8V + 0.2K	0.7V + 0.3K
1	0.8248 -4.7488	0.8543 -4.8167	0.8691 -4.9753	0.8758 -5.0987
2	0.7326 -4.4386	0.7688 -4.8037	0.7974 -4.9385	0.8113 -5.0238
3	0.6957 -3.5198	0.7596 -4.0879	0.7686 -4.1293	0.7675 -4.1426
4	0.8060 -5.9981	0.8281 -6.0816	0.8432 -6.2518	0.8494 -6.3761
5	0.7140 -5.5785	0.6966 -5.5540	0.7128 -5.6589	0.7178 -5.7326
6	0.7637 -3.8665	0.8229 -4.0454	0.8360 -4.2034	0.8413 -4.2536
	(1,1,1) Model	(1,1,1) Model	(1,1,1) Model	(1,1,1) Model
	0.6V + 0.4K	0.5V + 0.5K	0.4V + 0.6K	0.3V + 0.7K
1	0.8777 -5.1967	0.8768 -5.2656	0.8743 -5.3281	0.8709 -5.3710
2	0.8154 -5.0699	0.8132 -5.0900	0.8069 -5.0928	0.7982 -5.0812
3	0.7611 -4.1411	0.7520 -4.1382	0.7418 -4.1369	0.7313 -4.1337
4	0.8504 -6.4668	0.8482 -6.5341	0.8442 -6.5854	0.8390 -6.6260
5	0.7156 -5.7852	0.7088 -5.8205	0.6992 -5.8474	0.6877 -5.8681
6	0.8415 -4.3399	0.8380 -4.3983	0.8322 -2.0937	0.8249 -4.4311
	(1,1,1) Model	(1,1,1) Model	(1,1,1) Model	
	0.2V + 0.8K	0.1V + 0.9K	0.0V + 1.0K	
1	0.8669 -5.4033	0.8626 -5.4413	0.8582 -5.4745	
2	0.7879 -5.0758	0.7768 -5.0638	0.7653 -5.0510	
3	0.7209 -4.1380	0.7108 -4.1419	0.7013 -4.1373	
4	0.8333 -6.6593	0.8272 -6.6873	0.8210 -6.7115	
5	0.6753 -5.8854	0.6624 -5.8990	0.6494 -5.9129	
6	0.8166 -4.4924	0.8078 -4.4341	0.7987 -4.4711	

5.7 Conclusions

In this chapter linear dynamic models have been shown to provide a simple method for estimating missing air pollution data at one location by utilising data at other locations. Such models are particularly useful for estimating missing ozone data because, apart from discrepancies caused by local effects, ozone levels tend to be

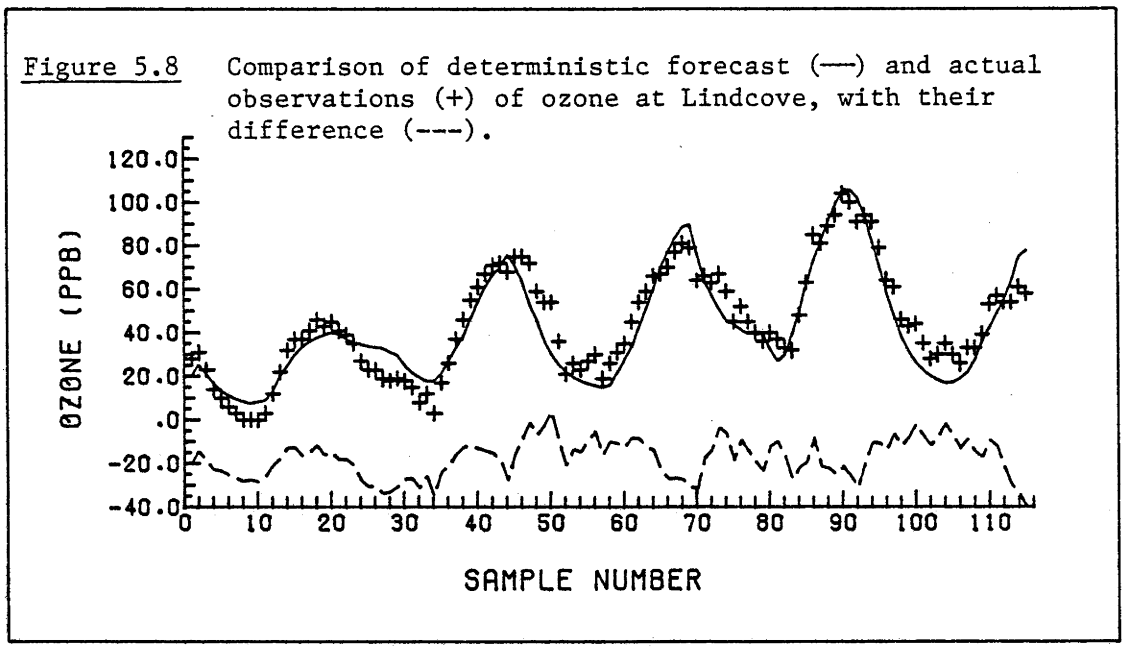


Figure 5.8 Comparison of deterministic forecast (—) and actual observations (+) of ozone at Lindcove, with their difference (---).

correlated spatially over relatively large regions. However, forecasts made in this way will be most reliable if they are restricted to relatively short periods of missing data.

In the course of investigating different types of linear dynamic models that may have been useful for estimating missing air pollution data, recursive refined IVAML algorithms were developed for estimating the parameters in a multiple input transfer function model of stochastic dynamic systems. Such a model can be considered as the dynamic equivalent of the regression model, with the regression coefficients replaced by the unknown transfer functions. The excellent performance of the algorithms has been demonstrated by a comprehensive stochastic (Monte Carlo) simulation exercise.

Chapter 6

RECURSIVE METHODS FOR MODELING CARBON MONOXIDE IN AN URBAN AREA

6.1 Introduction

Linear models were shown in Chapters 4 and 5 to be appropriate in applications to aspects of air pollution measurement, and significant temporal variation of the parameters was not encountered. In modeling carbon monoxide levels in urban areas, however, we know from previous research that these levels are a non-linear function of emission rates and wind speed, thought to be the major determinants of carbon monoxide levels in urban areas (Hanna, 1971). In this chapter the recursive techniques will be used to aid the initial data analysis, and in addition, an attempt is made to identify any non-linearities in time series models of carbon monoxide levels by making full use of the ability of the recursive estimation methods to reveal temporal changes in model parameters.

Special problems are encountered in modeling carbon monoxide levels in urban areas due to the fact that the major proportion of carbon monoxide emissions are from motor vehicle exhausts - in Washington D.C. and Canberra Australia (both non-industrial cities) it is estimated to be as high as 99 per cent (US Dept. HEW, 1970; Daly and Steele, 1975). Unlike stationary sources of pollution, the wide spatial distribution of motor vehicles is constantly changing. As a consequence the spatial and temporal variations of carbon monoxide levels in an urban area are large. For example, Ott (1972) has shown that eight-hour carbon monoxide concentrations can vary by a factor of three between sites which are less than three kilometers apart. It follows that areas of concern about pollution tend to be at locations where these concentrations are very high, such as areas with heavy traffic, or in locations with poor dispersion characteristics such as street canyons, both now frequently referred to as 'hot spots'. One consequence is that models based upon the advection-diffusion equation are not readily applicable because of their limited spatial resolution (see Section 2.4.1). These difficulties have encouraged alternative approaches to modeling the vehicular pollutant levels in such 'hot spots'.

Perhaps the greatest difficulty faced in the modeling of carbon monoxide levels is that of accurately specifying the emissions. It is well established that many factors greatly influence the emission rates of carbon monoxide from vehicles (US EPA 1978). Among these are the average speed of the vehicle (see Rashidi and Massoudi, 1980), its age (which may determine whether the vehicle is fitted with emission control devices), the engine size, the type of transmission, the type of fuel (petrol, diesel or LPG), the ambient temperature, and whether or not the vehicle has started from cold. While considerable effort has been expended in the search for methods that provide better estimates of vehicle emission rates from roadways (for example, Bullin and Polasek, 1978, and Bullin et al., 1980), uncertainties are likely to persist because of the factors enumerated above and the limited resources available for air pollution analysis.

An additional problem in modeling pollution in 'hot spots' is that the motion of air over such locations is made more complex by irregular arrays of buildings and by the existence of urban heat islands. Whereas the former may lead to channeling and vertical circulations in street canyons, and the latter to less stabilization of the surface air on clear nights, both may cause an increased mixing action of the air as compared to rural sites (see, for example, Bornstein and Johnson, 1977). Under some meteorological conditions these complex air movements may actually assist the dispersion of airborne pollutants, but under others, such as in light winds, street canyons can act as pollution traps.

In the remainder of this chapter we first briefly describe the types of models that have been developed for predicting dispersion of vehicular pollutants from roadways. Then the recursive methods of time series analysis are used in an attempt to develop a model of carbon monoxide levels in the Canberra City area. This is done initially for weekdays with the objective of proceeding to its validation by use of data from weekend days. Discrepancies between the modeled and observed concentrations during afternoon peak hours on weekdays are discussed and suggested causes are advanced.

6.2 Review of Models for Describing Vehicle Pollution From Roadways

The need to assess the environmental effects of proposed roadways

has led to the development of models for the specific purpose of determining the impact of traffic on air quality. Many of these models (see, for example, Turner, 1970; Calder, 1973; and Zimmerman and Thompson, 1974) are based upon a simple Gaussian plume formulation in which no allowance is made for site topography, and wind speed is assumed to be uniform with height. The application of such simple formulations appears satisfactory where there exist long, straight sections of roadway at ground level with few nearby obstructions to wind flow. However, as Maldonado and Bullin (1977) have shown by a comparison of the performance of several of these models, they usually do not perform well at sites located in urban areas.

These authors described a model which also was based upon the Gaussian plume formulation except that they explicitly allowed for the incorporation of knowledge of the surface roughness of the site, as well as for variation of wind speed and eddy diffusivity with height. Their model performed quite well when tested on several independent data sets and was a significant improvement on several previously reported models which also used a Gaussian plume formulation. Like the previously mentioned models, theirs was limited to the consideration of single locations, rather than entire urban areas, in the sense of being unsuitable for simultaneously estimating pollutant levels at other but nearby locations. They approached the task of modeling carbon monoxide pollution in two stages since they adopted the view that the initial dispersion over the roadway was due to mechanical turbulence created by the movement of vehicles, whereas the dispersion downwind of the roadway was the result of normal atmospheric turbulence. The inputs required for their model were the width of the roadway, the wind velocity at some reference height, the mean height of buildings in the vicinity, and the estimated mass of pollutant emitted from the vehicles per unit time. For the first stage they used correlation methods to estimate the vertical profile of pollutant concentrations at the edge of the roadway. The second stage required that this profile be matched by a Gaussian plume model which was subsequently used to predict pollutant concentrations downwind from the roadway.

A more ambitious attempt to model dispersion of vehicle pollutants from roadways is that described by Ludwig et al. (1970) and Johnson et al. (1973). Whereas the model referred to in the previous paragraph was limited to the sequential consideration of single

locations, these authors endeavoured to model simultaneously the pollutant levels for an entire urban area. The inputs to the model of Ludwig et al. (1970) were traffic volumes on major streets and highways in the urban area, wind speed and direction, atmospheric stability and mixing depth. The last two of these were estimated from twice daily radiosonde data and hourly meteorological measurements taken at the city airport. The meteorological measurements were assumed to be uniform over the urban area. During calms at the airport a small finite value of 1 ms^{-1} was used for wind speed together with the last observed wind direction. The model was capable of producing outputs in the form of concentration isopleths, sequences of hourly concentrations at a single site, or cumulative frequency distributions of concentrations. However, the agreement between the calculated and the observed concentrations was not good, with the latter usually being higher.

These discrepancies, which were believed to be due partly to the inability of the urban scale model to account for local effects, were reduced by subsequent incorporation of an empirical sub-model to account in particular for the street canyon effects (Johnson et al., 1973). These in turn were thought to result from cross-street circulation, a representation of which is shown in Figure 6.1. The sub-model estimated concentration components on the leeward and windward sides of streetside buildings, and the total concentration was then found by adding the component from the sub-model to the urban 'background' concentration calculated by the urban scale model. The performance of the modified model was reasonable with 80 per cent of calculated hourly values within three ppm of observed values. Supporting the view that street canyon effects may be ubiquitous, an Australian study by Bailey (1976) found that significant differences existed in the levels of carbon monoxide on the leeward and windward sides of a Sydney street.

A common feature of the models described so far is their total or partial reliance on a Gaussian plume formulation. Other approaches to modeling vehicular pollutants near roadways which avoid the assumptions inherent in this formulation have been suggested. Thus Danard (1972) employed a two-dimensional diffusion equation (see Section 2.4.1) to describe carbon monoxide concentrations in the vicinity of a plane highway and a depressed highway. A numerical integration procedure was

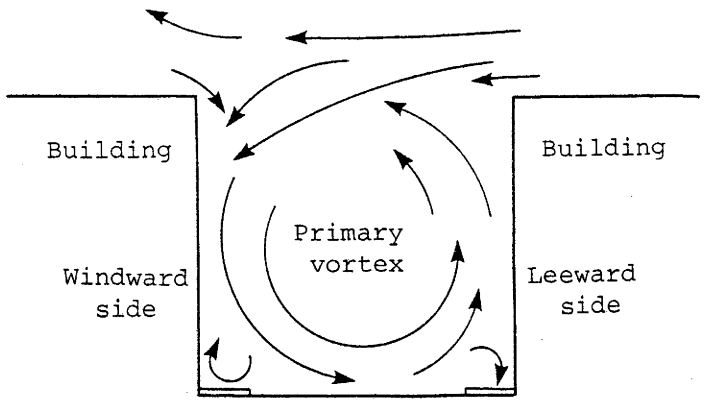


Figure 6.1 Representation of cross-street air circulation in a street canyon surrounded by buildings.

then used to solve the two dimensional diffusion equation. In his method the vertical and horizontal eddy diffusivities were variable to allow both for mechanical mixing of the air over the roadway caused by the traffic movement, and for the increased mixing action of the air if cold air advection was occurring. He also allowed wind speed to vary logarithmically in the lowest ten meters and linearly above that. Simple correlation coefficients of 0.82 to 0.94 were reported for the observed and modeled concentrations.

The use of a Gaussian plume assumption has also been avoided by the development of statistical models for vehicular pollutant concentrations near roadways (for example, Phadke et al., 1976; Tiao and Hillmer, 1978). Such models are based upon statistical analysis of data and do not require any prior assumptions about atmospheric behaviour. In their statistical models, Tiao and Hillmer (1978) have stressed the importance of efficiency (that is, few parameters with low variance) and have explicitly allowed for 'noise' or error terms which they assumed to be independently and normally distributed with zero mean. They estimated the model parameters by employing a non-linear least squares procedure in which the sum of squares of the error terms

was minimized. Some features of their approach are similar to the one which is described later in this chapter.

6.3 Description of the System

The data used in the modeling in this chapter were taken from a study of carbon monoxide levels in Canberra City during 1973-74 (Daly and Steele, 1975). The principal series available were (i) five minute average concentrations of carbon monoxide measured at one location; (ii) half-hourly average wind speed and direction also measured at one location and at a height of 10 meters above the ground with a Woelfle anemometer; and (iii) hourly traffic counts on the major roadway nearest the carbon monoxide monitoring location. Some additional comments on the methods by which these data were obtained, and on adjustments made to them prior to their use in the modeling, seems warranted, and is the purpose of this section.

The primary objective in selecting a monitoring site for carbon monoxide was to obtain estimates of the carbon monoxide concentrations to which people in the Canberra City area were exposed. Consequently a site in West Row, one block west of the major north-south roadway (Commonwealth/Northbourne Avenue) passing through the city center, was selected as more appropriate than a site adjacent to this roadway itself. The streetside sampling point was three meters above the pavement and five meters back from the kerb. A Mine Safety Appliances (MSA) LIRA Model 202 carbon monoxide analyser of the non-dispersive infrared type was employed to make the measurements. Water vapour in the air sample was removed by silica gel filters prior to its entry to the analyser. Particulate matter was also removed by passing the air sample through a filter with a pore size of 0.5 micron.

It can be seen from the map showing the location of the various sites at which data were collected (Figure 6.2) that the wind and carbon monoxide measurements were made some distance from each other. There are two reasons why this is a not unusual feature of air pollution modeling studies. One is that measurements of local winds in the city center may not be representative of regional average winds because the location of anemometers near obstacles such as buildings may distort the wind flow. An ideal location for obtaining measurements of regional average winds is thus a large, open and relatively flat space free of tall trees and buildings, where any grass

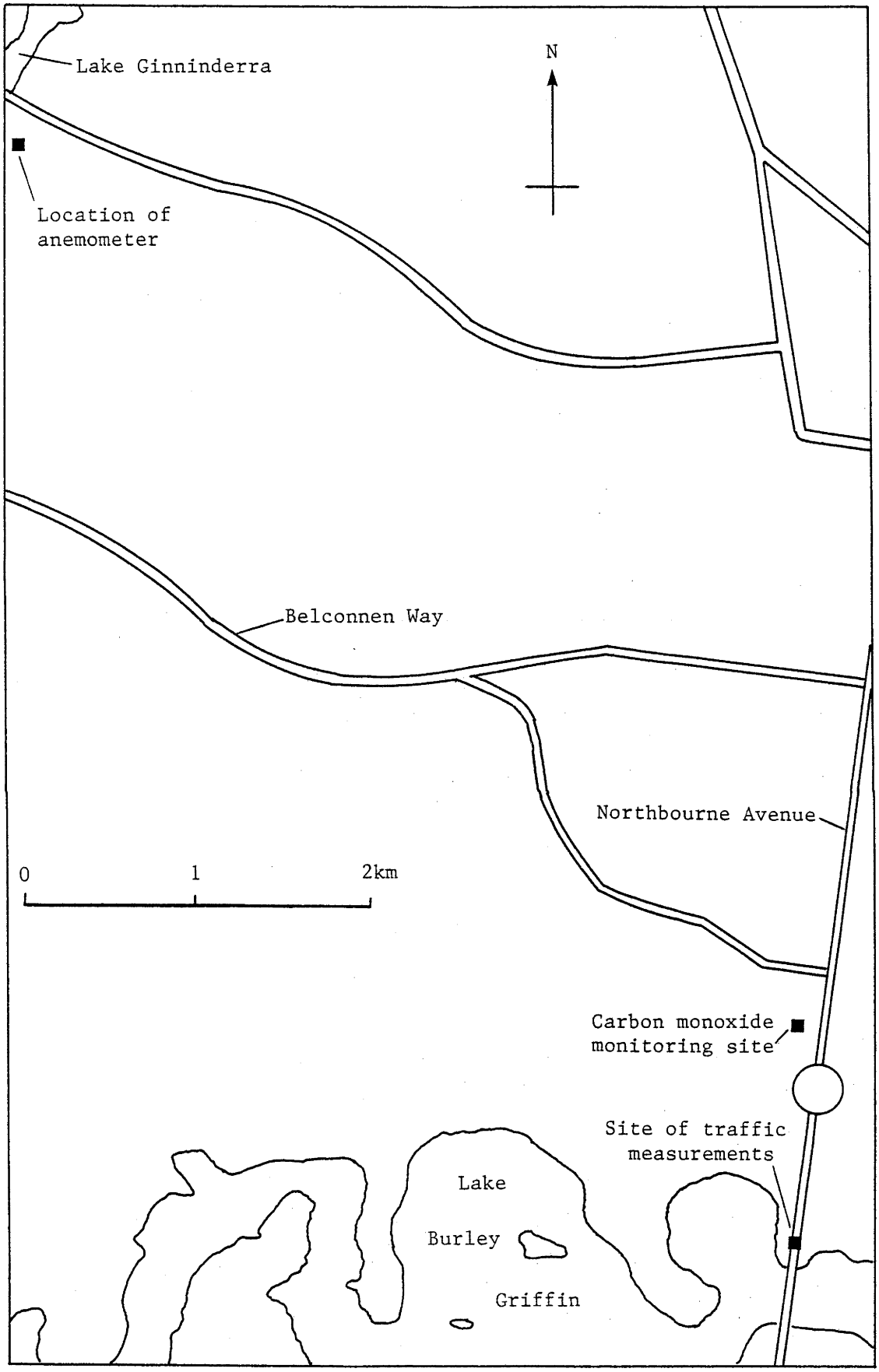


Figure 6.1 Map showing the locations at which traffic counts and wind and carbon monoxide measurements were made.

is kept cut close to the ground. Hence, it is not surprising that airports are favoured sites for taking meteorological observations. A second reason is that anemometer data may already have been collected, and funds to set up anemometers specifically for the air pollution modeling study may not be available. Measurements of both local and regional average winds would be very desirable for any air pollution modeling, but as this was not possible in the present case, measurements of the regional average winds had to suffice. This use of less than ideal wind measurements is an example of the poorly defined nature of the system. In defence of our use of regional average rather than local winds it should be noted that Gifford (1974) found that the prediction of carbon monoxide levels in Los Angeles (by the use of Hanna's (1973) model) was considerably improved if regional average, rather than local, winds were used. Of the two available records of regional winds, the measurements recorded at Belconnen were preferred to those made at Canberra airport because the Woelfle anemometer used at the former site was specially designed for micrometeorological studies and had the low starting speed of 0.5 ms^{-1}.

The traffic counts of vehicles passing along Commonwealth Avenue in both directions were obtained by use of induction loop counters installed in each lane of the roadway, the location of which - and relation to the site at which the carbon monoxide measurements were recorded - can also be seen in Figure 6.2. The siting of the traffic measurements may be less than ideal, but other available traffic flow data for the area related to comparatively minor city streets, and the measurements were not thought to constitute a reliable indication of traffic passing through Canberra City. This is because the principal traffic movements are north-south and constrained to pass the traffic monitoring point in order to gain access to one of only two roadways crossing the nearby Lake Burley Griffin. As in the case of the wind data, the reliance on existing traffic flow counts made for other purposes as a surrogate for the desired traffic measurements - a characteristic of studies of this type - only serves to emphasize further the poorly defined nature of the system.

A period of 20 weeks was chosen on the basis of the completeness of the carbon monoxide, wind speed and wind direction data sets for each week. These 20 weeks were spread between November 1973 and August 1974. It should be noted that none contained public holidays or

covered periods of petrol shortages. The hourly traffic counts were supplied by the National Capital Development Commission and covered the week 11 to 17 February 1974. Unfortunately, estimates of average traffic speed were not available, so that the importance of 'traffic density' as defined by Tiao and Hillmer (1978) in determining carbon monoxide levels could not be evaluated here. However, total daily traffic counts at the same location were measured for many of the 20 weeks selected. These revealed significant weekly variation, raising the problem of defining an average time resolved traffic count over the 20 week period. To meet this, we have assumed the traffic data for the week in February to be highly representative of all the 20 weeks considered. However, it seems likely that significant hourly variations in traffic flow occurred since, as mentioned above, such is the case with the daily traffic counts if a comparison is made of the week in February and other weeks in the study period. This feature of the data can be cited as a third example of the poorly defined nature of this system.

For all the modeling described in this chapter, hourly data were used. These were derived by averaging the five-minute mean carbon monoxide measurements and the half-hourly mean wind speeds. The hourly observations so obtained were, in turn, averaged over the 20 week period yielding a 'typical' seven-day week. In Figure 6.3 the resultant hourly carbon monoxide and wind speed estimates are shown together with the hourly traffic counts for the week in February. Superficial inspection reveals that whereas the weekday morning peak is greater than the afternoon peak in the traffic flow, the reverse pattern is evident for the carbon monoxide concentrations. This implies the possibility that the emissions per vehicle are larger in the afternoon peak period on weekdays.

6.4 Initial Modeling

It is appropriate, methodologically, to begin by considering the most simple linear model form. Then, additional variables required in the model may be suggested by inspection of any patterns of time variation in the parameters of the initial model as revealed by the recursive estimation methods. The objective of this approach is to secure an overall non-linear model which could be separated into a component which is linear in the parameters, and a non-linear

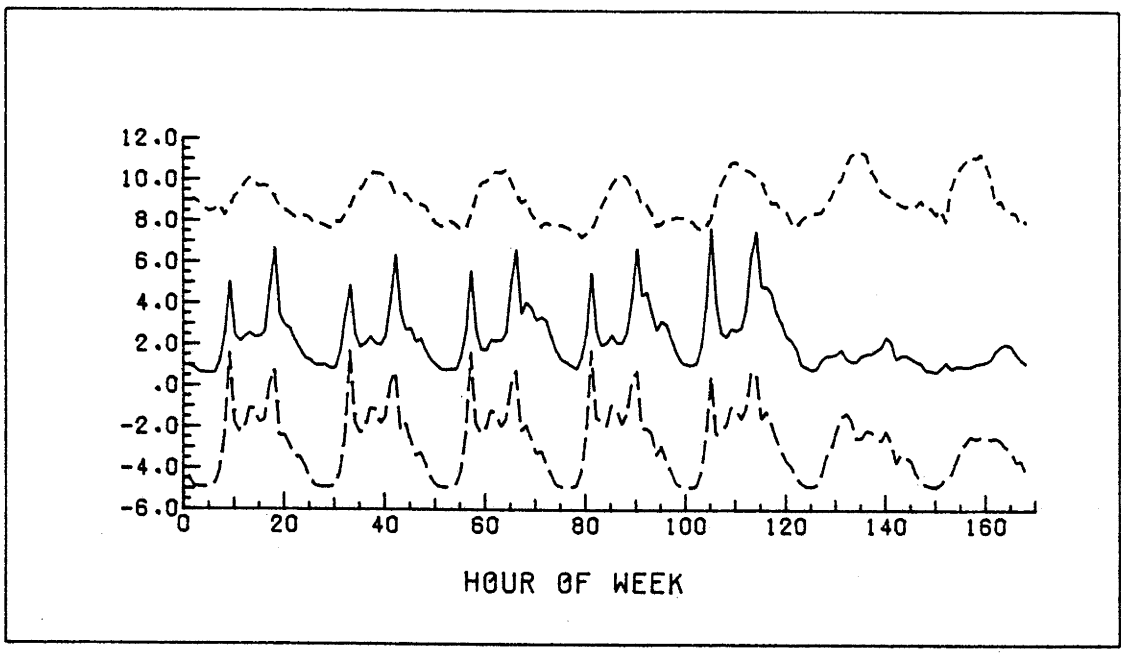


Figure 6.3 Average levels of carbon monoxide in ppm (——), wind speed in ms^{-1} (---), and traffic count ÷ 1000 (— —). Wind speed translated +5.0 and traffic count -5.0 on vertical axis for clarity.

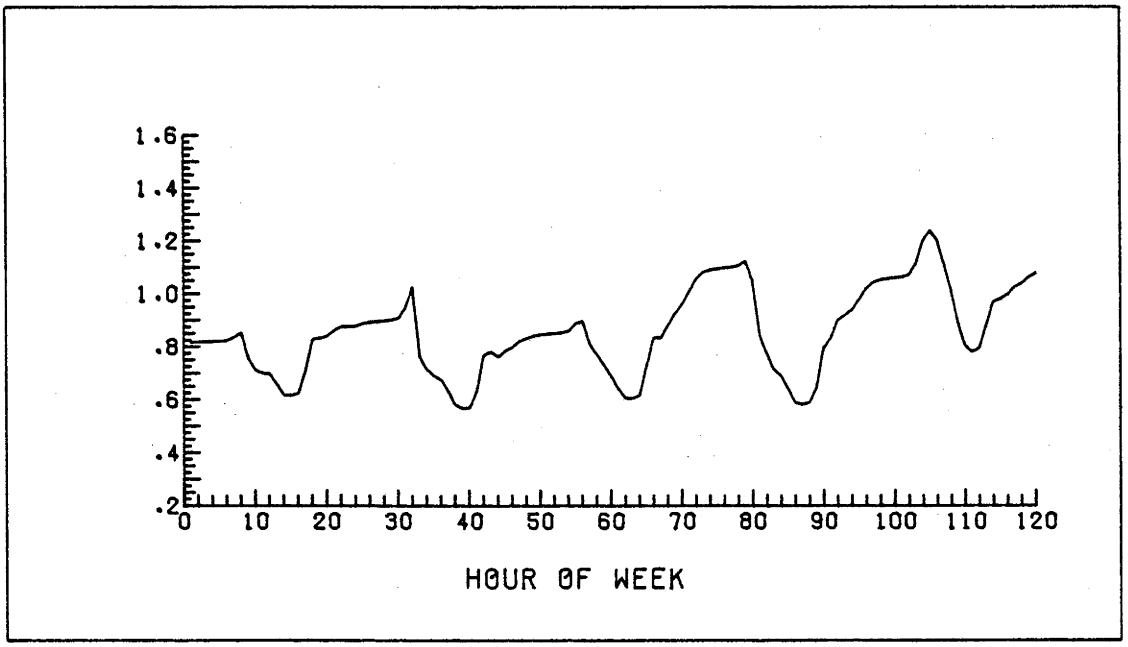


Figure 6.4 Estimated time varying b_k parameter in the initial two parameter model relating carbon monoxide levels C_k to traffic counts T_k.

component.

Knowing that the major source of carbon monoxide in urban areas is motor vehicles, it seems defensible to start with a simple SISO model with counts of traffic flow T_k as the input and carbon monoxide concentrations C_k as the output. In choosing between alternative model structures, it is pertinent to indicate a difference in procedure from that followed in Chapters 4 and 5. In the present context the aim is not necessarily to select that model which simultaneously minimizes the EVN or NEVN criterion and provides a value of R_T^2 at the plateau level. Rather, a model is sought which is both plausible in its physical properties (for example, does not produce negative concentrations of carbon monoxide) and at the same time is efficient in terms of containing as few parameters as possible.

Since the immediate aim is to investigate any patterns of temporal variation in the model parameters, the most simple model consistent with both autoregressive and moving average behaviour has been selected, that is, a model with one 'a' and one 'b' parameter. This can be written as

$$C_k = -a_1 C_{k-1} + b_k T_k + \xi_k$$

where k = 1,2,..., N (in this case N, the number of samples, being 120), and ξ_k are the 'noise' terms. In the absence of prior information as to the time variation of the model parameters, they are allowed to vary as a random walk, for example, in the case of b_k

$$b_k = b_{k-1} + v_k$$

where v_k is serially uncorrelated white noise.

To obtain an indication of whether factors other than the input traffic term may be influencing carbon monoxide levels at all significantly, a close examination is required of the estimated temporal variation of the 'b' parameter. This is because the non-linearities expected, and found, in this case related to the amplitude of the carbon monoxide signal, rather than to the dynamic

lags in the system. This pattern of variation, ~~together with wind speed~~, is shown in Figure 6.4. A significant inverse relationship between the two is clearly evident. Given the findings of other researchers (for example, Hanna, 1971), this result is not surprising. Whether wind direction as well as speed should be incorporated explicitly into the model is not clear at this stage. For the present an attempt to include just wind speed will be made, with the intention of using evidence of any remaining patterns of time variation of the model parameters as a basis for deciding whether yet further variables need to be added to the model. Thus in the next section consideration is given the alternative means by which wind speed may be incorporated, together with the question of wind direction.

6.5 Modeling the Wind Effects

The incorporation of wind effects into the model, and the results obtained from this extension, will be described in the following sequence. An analysis of the additional data is first presented, in which the diurnal and weekly patterns of measured wind speed and direction are carefully scrutinised and compared with the pattern of recorded carbon monoxide concentrations. The concept of an 'effective traffic' flow, incorporating some allowance for wind effects, is also introduced. Using the present data this permits a brief evaluation of Hanna's (1971) relatively simple model for air pollution levels arising from urban area sources. The principal findings of this section (and chapter) are then presented. These comprise a comparison of results obtained from several models in which the wind effects are differently specified, as no single model unambiguously out-performed all others when applied to the weekday averaged data. These findings raise a number of issues regarding the modeling methodology adopted as well as the quality and suitability of the data employed, issues fully appraised in section 6.6.

6.5.1 Analysis of Wind Speed and Direction

The basic information on wind speed was shown earlier, together with concentrations of carbon monoxide and traffic flow, in Figure 6.3. We now wish to examine this information in more detail in order to ascertain the most appropriate form in which to incorporate wind effects into the model.

To readily comprehend the relationship between wind speed and carbon monoxide concentrations, particularly given their wide scatter when cross-tabulated, it has been necessary to group the former observations in a limited number of categories. These categories were selected so that each spanned a range of 0.5. Thus the first category included winds $0 < W_{jk} \leq 0.5$, the second $0.5 < W_{jk} \leq 1.0$, and so on where $j = 1, \ldots, 20$ denotes the week and $k = 1, \ldots, 120$ denotes the hour of the week. For all values of W_{jk} falling within a particular category, the corresponding values of C_{jk} were accumulated so that a mean concentration for that category was determined. The results are shown in Figure 6.5 and a relationship of the form

$$\ln(\text{carbon monoxide}) = 1.34 - 0.12 \times \text{wind speed}$$

provided a good fit to the data. This outcome is intuitively reasonable, simply indicating a decline in pollution concentrations with increased wind speed, even though no adjustment has been made to account for variations in the level of traffic flow.

The preferred means by which wind direction together with speed may jointly be incorporated into the model is to utilise the wind components. The wind vector was decomposed into easterly and southerly components following the results obtained by Tiao and Hillmer (1978) which indicated that that wind component perpendicular to the direction of the street in which the pollution readings were taken was the most important in relating wind to pollution concentrations. Since in the present study the monitoring device was located in a street running north-south, the easterly component is perpendicular and the southerly component parallel to the street direction. Inherent is the notion of positive, negative and zero wind components. Since the major north-south roadway in the vicinity was to the east of the measuring station, this determined the assignment of positive and negative wind components; that is, easterly and southerly components were of positive and westerly and northerly components were of negative sign.

Wind categories were then chosen as before (except that now of course the positive and negative components must be allowed for). A mean carbon monoxide level was then calculated for each category and the results are shown in Figures 6.6 (east wind components) and 6.7

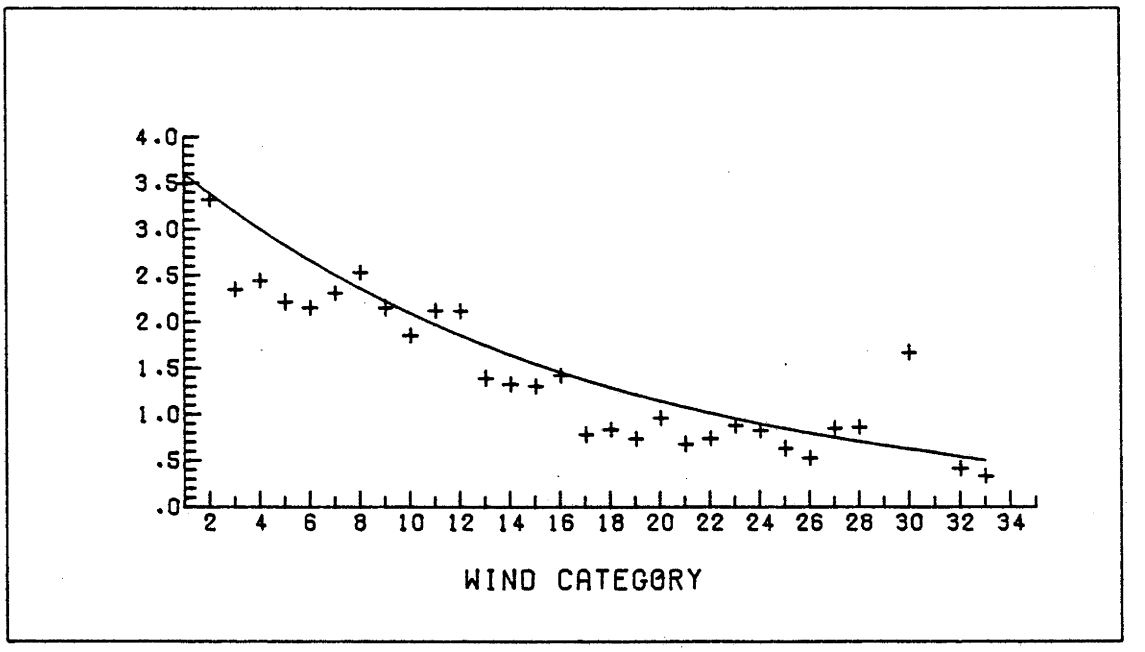


Figure 6.5 Mean carbon monoxide level (ppm) associated with wind speed categories. The number of each category is twice the wind speed in ms^{-1}.

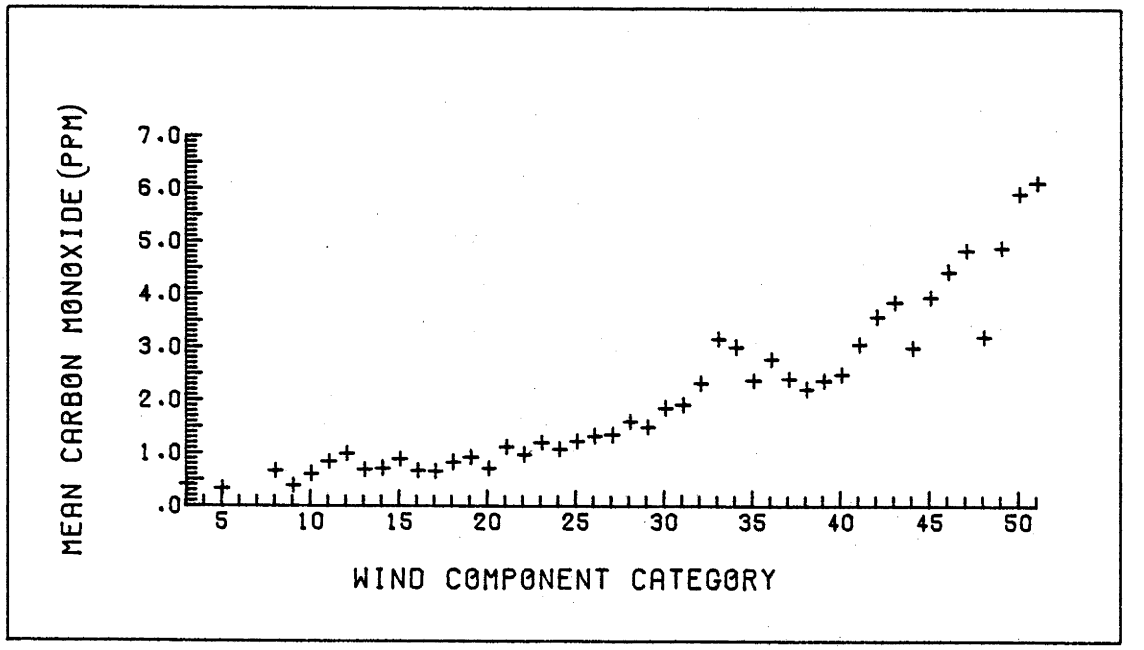


Figure 6.6 Mean carbon monoxide level versus east wind component categories. Category 35 corresponds to zero east wind component. Higher (lower) values indicate increasing magnitude of winds from the east (west).

(south wind components).

The carbon monoxide concentrations associated with the southerly wind components are arrayed symmetrically around the value of zero. That is, there is little evidence of a directional influence on the mean carbon monoxide levels, with the larger values of carbon monoxide being associated with winds of low magnitude. However, with the easterly wind components there is a significant trend. The results indicate that the carbon monoxide levels increase steadily with the increasing magnitude of winds from the east. To eliminate the possibility that this result may be an artifact of the way the mean carbon monoxide values were calculated, a subset of the total data set was selected on the basis of traffic counts; specifically, only those hours in which the traffic count lay in the range 3375 to 4050. Then the same procedure was followed and the result is shown in Figure 6.8. This quite clearly replicates the pattern revealed in Figure 6.6, a pattern which was found also to be the case for each of the 10 separate categories of traffic counts.

This result may indicate that a significant quantity of the carbon monoxide measured at the West Row monitoring site is transported by the wind from a source or sources to the east. The traffic in Commonwealth/Northbourne Avenue is the most likely source. Reinforcing rather than offsetting these results, some cross-street circulation of carbon monoxide along the lines indicated earlier in this chapter in Figure 6.1 may be occurring. The topography of the West Row site is not strictly equivalent to the street canyon situation as depicted in that figure, since there is a small park and then a carpark opposite the monitoring site, not another building. Nevertheless, some cross-street circulation cannot be ruled out and further measurements would be needed to resolve these doubts.

A quite different feature of the wind component data which is also of relevance is the diurnal patterns averaged over the 20 weeks for which the carbon monoxide concentrations were recorded. These are shown in Figures 6.9 and 6.10 for the east and south wind components respectively. It is noticeable that the times of morning and afternoon peak traffic during weekdays fairly closely coincide with times at which the east wind component is close to zero.

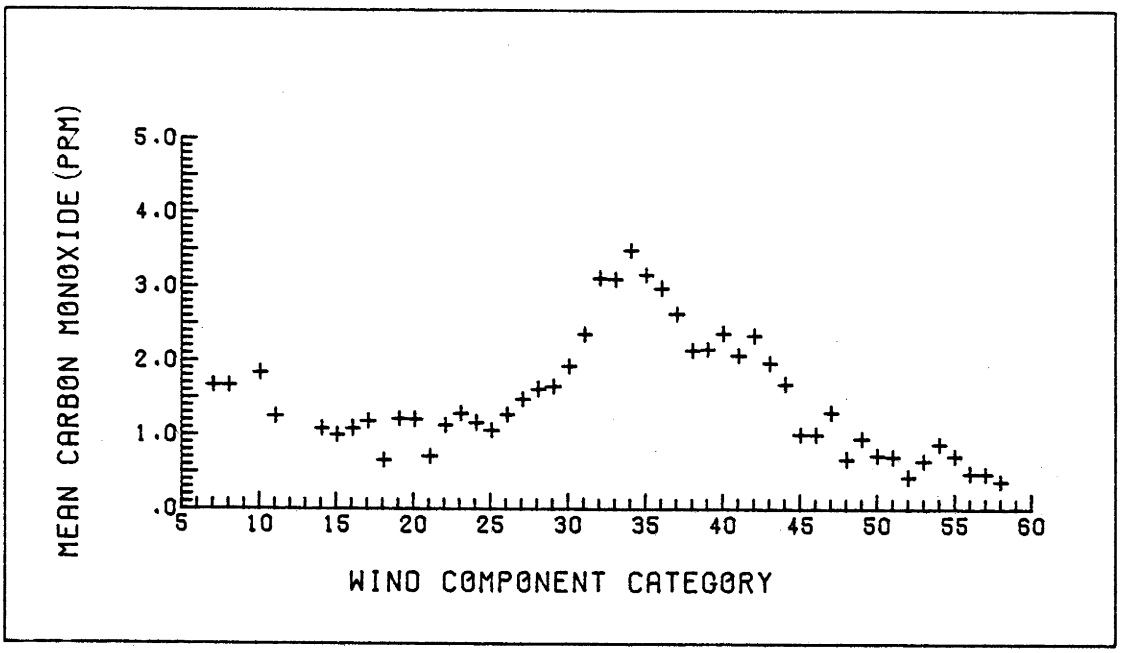


Figure 6.7 Mean carbon monoxide level versus south wind component categories, where category 35 corresponds to zero south wind. Higher (lower) values indicate increasing magnitude of winds from the south (north).

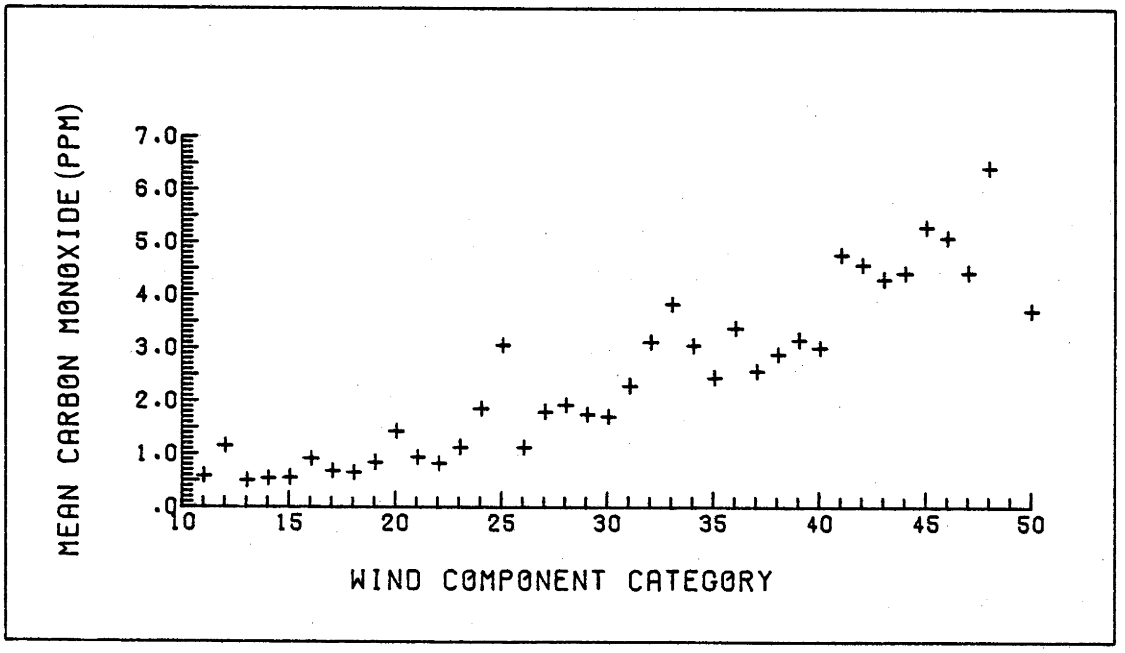


Figure 6.8 Mean carbon monoxide level versus east wind component categories for those hours in which traffic count lay in the range 3375 to 4050.

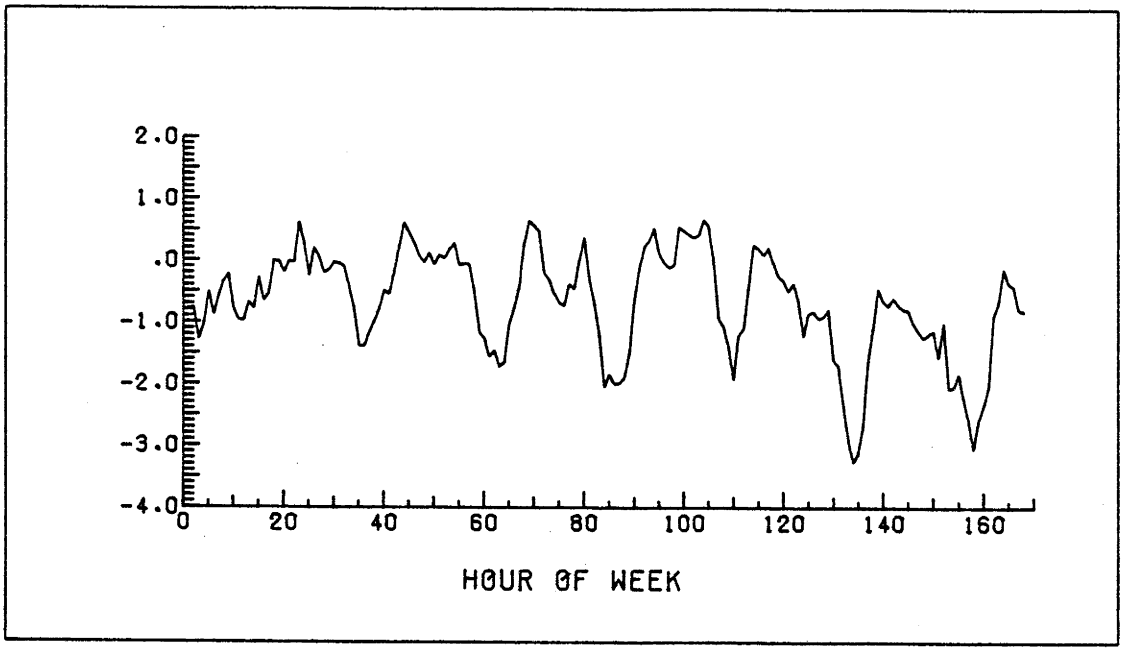


Figure 6.9 Average diurnal variation of the east wind component over the 20 weeks.

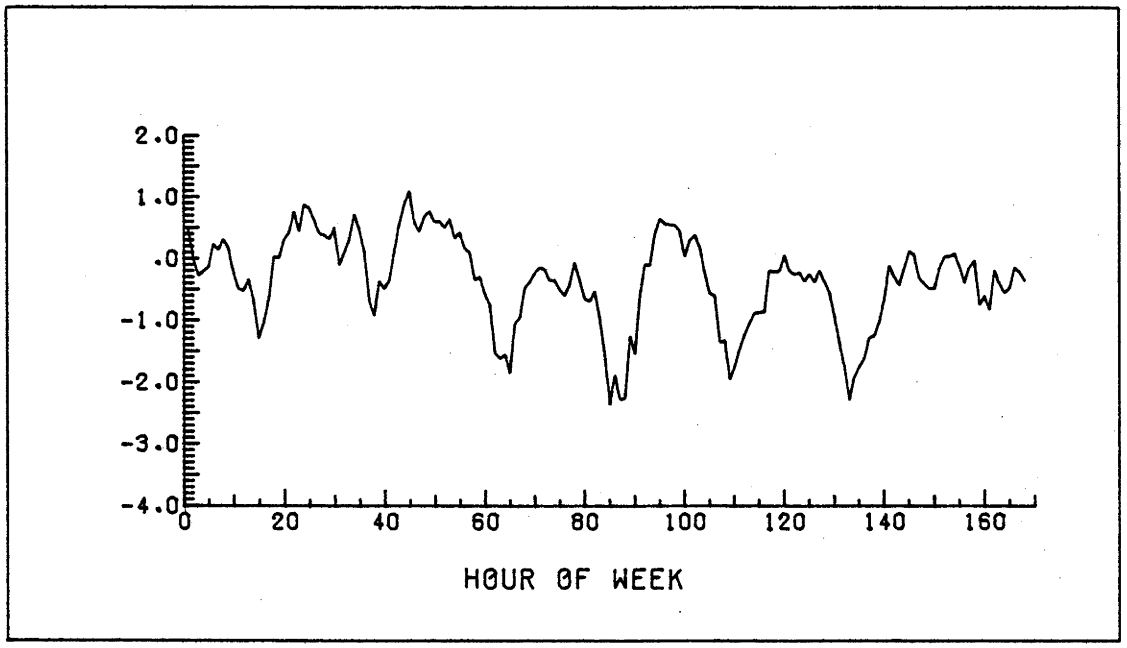


Figure 6.10 Average diurnal variation of the south wind component over the 20 weeks.

6.5.2 Evaluation of Hanna's Model

An appropriate starting point for modeling carbon monoxide levels in Canberra City is the simple model structure employed by Hanna (1971) for urban area sources of pollution. This was obtained by him through reasoning that the area source in the immediate vicinity of any monitoring site is usually the most important in determining the air pollution levels measured there, and that this assumption permits a simple model specification to be used which incorporates wind speed but not wind direction explicitly. The model is specified

$$\chi = C \frac{Q}{U}$$

where χ is the level of air pollution, Q is the source strength of the grid square in which the pollutant is measured, U is the wind speed, and C is a constant reflecting atmospheric stability, the size of the grid squares and the number of grid squares upwind. Like the Gaussian plume model for a point source of pollution (see equation 2.8), this model becomes degenerate for wind speeds of zero.

Since in the present study the pollution source is motor vehicles, we may define an 'effective traffic' for any hourly period in the 20 weeks as T_{jk}/W_{jk}, where T_{jk} is the actual traffic count and W_{jk} the wind speed, both referring to the jth week and the kth hour of the week. This concept is somewhat analagous to that of effective rainfall in hydrological modeling studies (see, for example, Whitehead et al., 1979). Since we are using data 'averaged' over 20 weeks, the 'effective traffic' E_k^1 is calculated by defining

$$E_k^1 = \frac{\sum_{j=1}^{20} T_{jk} W_{jk}^{-1}}{20} \qquad (k = 1, 2, \ldots, 120)$$

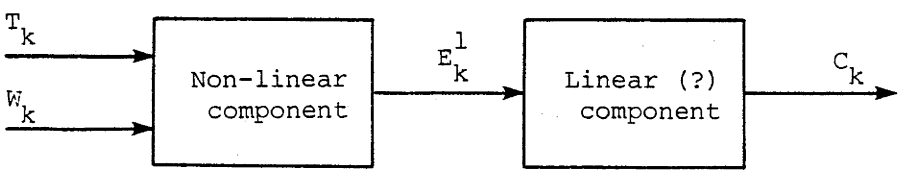
as the average 'effective traffic' at each hour of the average week. In rainfall-runoff modeling 'normalisation' is typical, where normalisation is an amplitude scaling so that the effective rainfall series possesses a maximum value identical to that of the original rainfall series. In the present case this was not considered necessary

because the recursive estimation algorithm automatically accomodates any such scaling factor by incorporating it into the 'b' parameters of the identified model.

Although this definition has the theoretical limit property that 'effective traffic' tends to infinity as the wind speed tends to zero, it can be used in practice because 'real' wind speeds of zero are rare occurrences in the open atmosphere. Normal anemometers will register 'calms' when the wind speed drops below their starting speed, and so accurate measures of very low wind speeds are not normally available. When 'calms' occurred in the data set used in the present study, the wind speed was set to the value of the starting speed of the anemometer ($0.5\ ms^{-1}$), so overcoming any likelihood of the 'effective traffic' becoming infinite.

Following from the initial model sketched above in Section 6.4 where the unadjusted traffic count T_k was the input to the model (where it was found that the pattern of time variation of the parameters indicated the need to incorporate wind), and at the same time drawing on Hanna's model to create a measure of 'effective traffic', we are now in a position to evaluate a more realistic model structure. By specifying a non-linear component in this way it will be possible to use the 'effective traffic' (E_k^1) as the input to a transfer function model with carbon monoxide as the output. Examination of the time variation of the parameters in this part of the model will reveal whether the wind effects have been incorporated appropriately.

The overall model may be represented in the form of a block diagram



where time invariant parameters in the second component of the model will indicate a successful separation into linear and non-linear components. The IV model identification procedure was then applied to

this second component of the model, and a (1,2) model was selected as that structure with the largest value of R_T^2 consistent with physical plausibility. The model fit to the data is shown in Figure 6.11 where two clear patterns are apparent. Firstly, the model consistently predicts concentrations of carbon monoxide during the afternoon peak hours which are lower than those observed; and secondly, the model overestimates the carbon monoxide levels during the time between morning and afternoon peak hours. Once again the 'b_0' parameter is estimated as a time varying parameter because the model fit shown in Figure 6.11 illustrates that the major deficiency of the model lies in its inability to reflect fully the magnitude of the peaks. The result is shown in Figure 6.12. The pattern of variation shows some correlation with the wind speed, and this may indicate either that the effectiveness of wind in dispersing carbon monoxide is not fully captured in the model structure or that there exists an omitted variable such as atmospheric stability.

These findings by themselves do not necessarily imply that basic weaknesses exist in Hanna's model. The traffic counts used here as a surrogate measure of the pollution source strength may not be a particularly good indicator, and in addition, Hanna makes it clear that the model (6.2) will be most appropriate for pollutant concentrations averaged over long time periods.

In order to investigate whether wind speed may be incorporated more appropriately into the model structure (6.3) we now look at alternative definitions of 'effective traffic'.

6.5.3 Modeling Wind Speed and Direction

Another definition of 'effective traffic' that simply combines traffic counts and wind speed in some non-linear fashion is suggested by the results of Tiao and Hillmer (1978) and also by the analysis of the relationship between the mean carbon monoxide concentrations and wind speed category reported earlier in this section. Tiao and Hillmer used a diffusion factor of the form $\exp(-b(WS_\perp - W_0)^2)$ to model the dispersion of carbon monoxide from traffic in city streets. The term $WS_\perp$ denotes the wind component perpendicular to the street direction, and b and W_0 are empirically determined parameters. So by analogy with the expression for E_k^1 we define

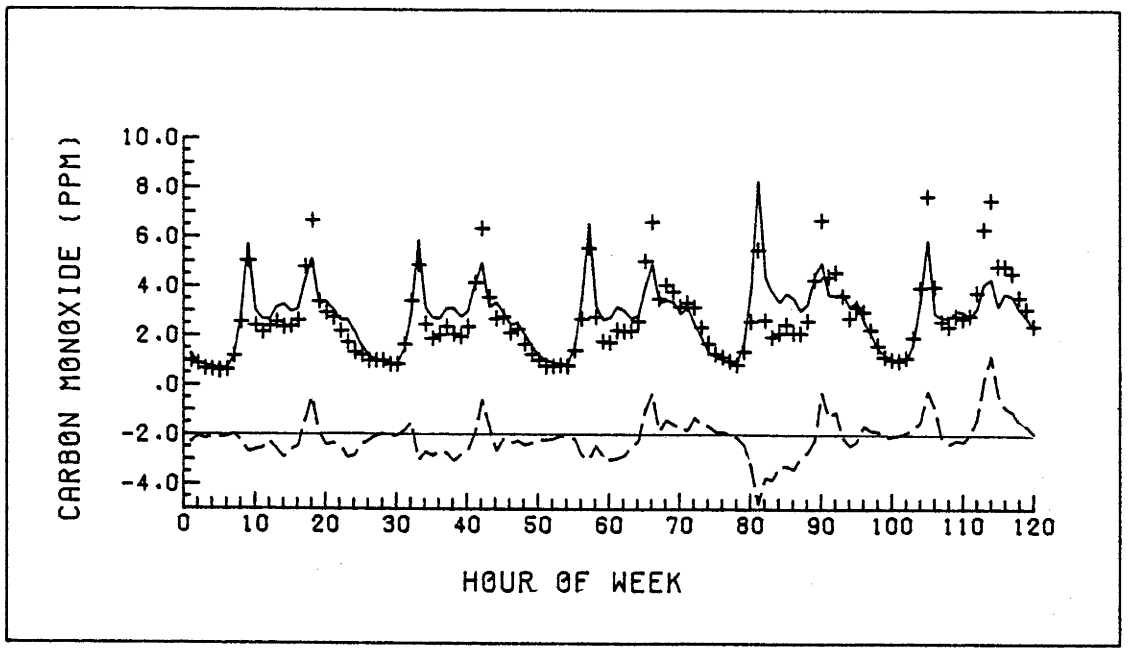


Figure 6.11 Model fit (——) to observed carbon monoxide levels (+) given by the (1,2) model with E_k^1 as the effective traffic input. Estimated residuals are plotted directly underneath.

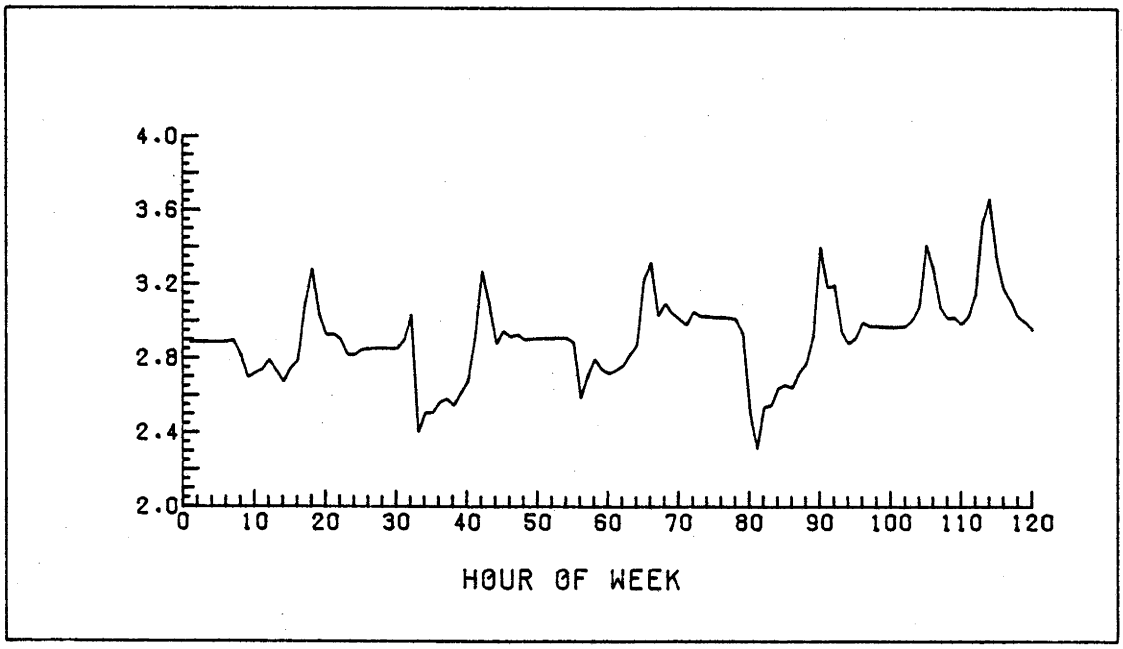


Figure 6.12 Estimated time varying b_o parameter in the (1,2) model with E_k^1 as the effective traffic input.

$$E_k^2 = \frac{\sum_{j=1}^{20} T_{jk} \exp (\beta W_{jk})}{20} \qquad (k = 1,2, \ldots, 120)$$

where the constant parameter β has negative sign, and its value is found by a process of trial and error. (Alternatively β could have been estimated by an overall non-linear optimisation procedure.) The definition for E_k^2 overcomes the objection raised against E_k^1 since E_k^2 has the limit property that it tends to T_k as W_{jk} tends to zero. This property is intuitively reasonable since any wind speeds greater than zero will produce greater dispersion of pollutants than would occur with wind speeds of zero, and hence 'effective traffic' should be expected to be less at higher wind speeds.

Before any assessment of the performance of the model incorporating the redefined measure of 'effective traffic' (E_k^2) is undertaken, however, it is worthwhile considering an additional point: namely, can wind direction as well as wind speed may be combined with the traffic count into a new measure of 'effective traffic' by use of the concept of wind components previously discussed? Such an approach offers a comparatively straightforward method for including wind direction in the model. We have chosen to use only the easterly component and ignore the southerly, because as previously noted, the symmetry of mean carbon monoxide concentrations associated with the southerly components implies that little improvement in model performance over the inclusion of wind speed can be anticipated, whereas this implication cannot be drawn with respect to the easterly components (because of the asymmetrical distributions of the concentrations with which they are associated (see Figure 6.6)). Our new measure of 'effective traffic' thus becomes

$$E_k^3 = \frac{\sum_{j=1}^{20} T_{jk} \exp (\beta' \times EC_{jk})}{20}$$

where $j = 1,\ldots,20$ and $k = 1,\ldots,120$. The value of β' is found by

trial and error, that value being sought which gives parameters in the linear component of the model which show no correlation with the easterly wind component.

6.6 Discussion of Results

From the immediately preceding discussion, two models associated with the alternative measures of 'effective traffic' E_k^2 and E_k^3 were chosen for comparison with the model form suggested by Hanna (1971) and decribed in Section 6.5.2. The first two models were obtained after specifying values of $\beta = -0.5$ and $\beta' = 0.4$ to be used in the derivation of E_k^2 and E_k^3 respectively. At the outset it may be indicated that none of the models performed as satisfactorily as anticipated, nor is any one clearly superior where the several criteria for evaluating model 'performance' are considered. To introduce the comparative evaluation, a succinct description of the three models and their respective summary test statistics are shown in Table 6.1. We observe that in all cases the identified model structure is a (1,2) model with a zero pure time delay. It is noticeable that the R_T^2 associated with the Hanna model is greater than that for either of the other two, but this result is in a sense deceptive. Inspection of the model fit for the Hanna model (Figure 6.11) indicates that although its overall conformity with the actual pollutant levels throughout the day is relatively more successful than either of the other two models, it fails fully to capture either the mid-day troughs or afternoon peaks in measured concentrations as successfully as the morning peaks. We may also observe from Table 6.2 that the parameter definition of the E_k^3 model as indicated by ln(NEVN) is slightly better than that obtained for the other two models. However, this ranking has to be seen in the light of additional aspects of model performance.

One such aspect is the pattern of time variation of parameters. It was observed earlier that with Hanna's (1971) model there was evidence of some correlation between this pattern and wind speed (Figure 6.12). When the parameter variation of the other two models incorporating 'effective traffic' are inspected (Figures 6.13 and 6.14), there is little evidence to suggest any correlation with either wind speed or the easterly wind component. This is not surprising as the values of β and β' were chosen in order to reduce such correlation.

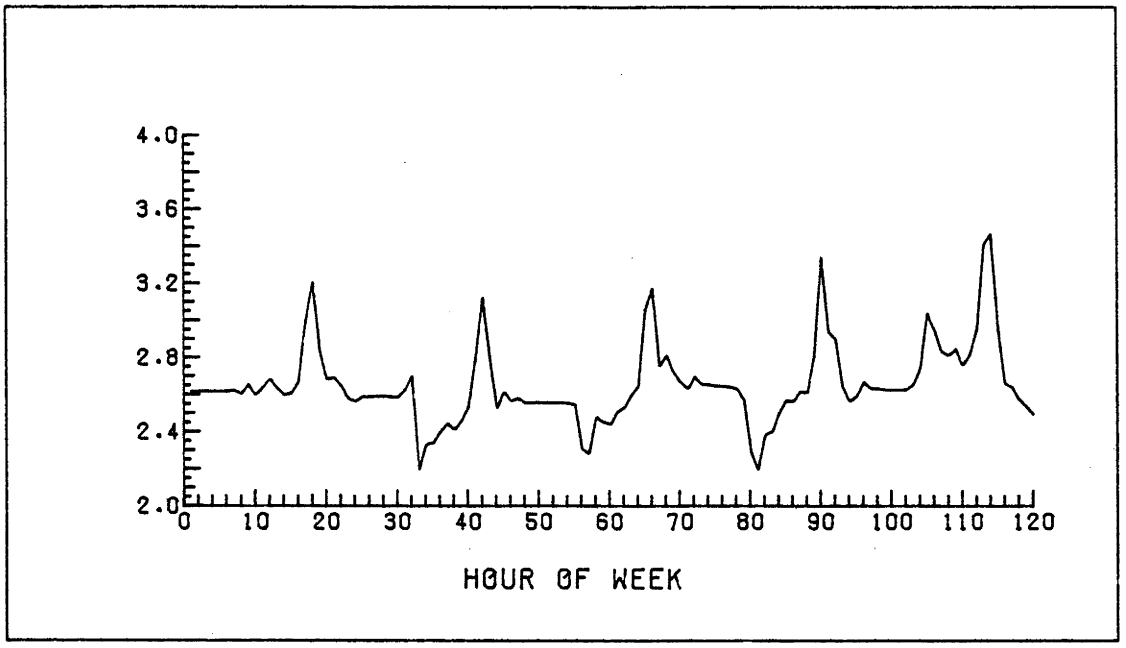


Figure 6.13 Estimated time varying b_o parameter in the (1,2) model with E_k^2 as the effective traffic input.

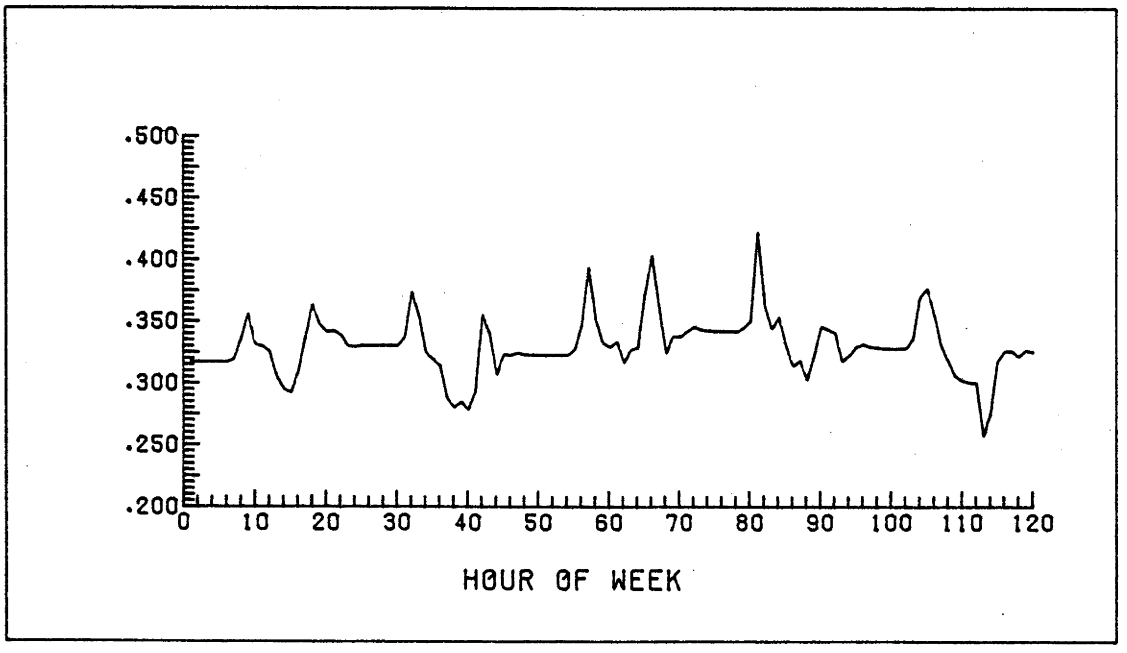


Figure 6.14 Estimated time varying b_o parameter in the (1,2) model with E_k^3 as the effective traffic input.

However, the reduction of this source of parameter variation has not eliminated all the variation, rather it has emphasised the existence of other sources.

Clearly, a successful decomposition of the model into linear and non-linear components has not been achieved in the above exercises. Particularly evident in Figure 6.13 is the coincidence between the regular pattern of (residual) parameter variation and the times of afternoon peak traffic flow, a coincidence strongly suggesting some remaining deficiency in the measure of 'effective traffic' (E_k^2) that has been used. The incorporation of the easterly wind component, and hence derivation of E_k^3, has reduced the dominance of the afternoon peak in the pattern of parameter variation (Figure 6.14), but some difficulty remains in closely modeling the concentrations during times of both morning and afternoon peak traffic flows.

Another perspective on the performance of the alternative models is offered by the plots of the model fits (Figures 6.15 and 6.16). Interpreted in conjunction with the evidence of the test statistics and parameter variation, it is clear that the results of the modeling exercises undertaken in this chapter are somewhat inconclusive, and raise a number of substantive issues. In the remainder of this section some effort is expended in probing and cautiously speculating on the reasons for these results.

Table 6.1

Comparison of Primary Test Statistics for Identified Models using 'Effective Traffic' Inputs

Definition of E_k	Model	R_T^2	ln(NEVN)
E_k^1	(1,2)	0.7290	-3.1882
E_k^2	(1,2)	0.6694	-3.3047
E_k^3	(1,2)	0.6628	-5.1528

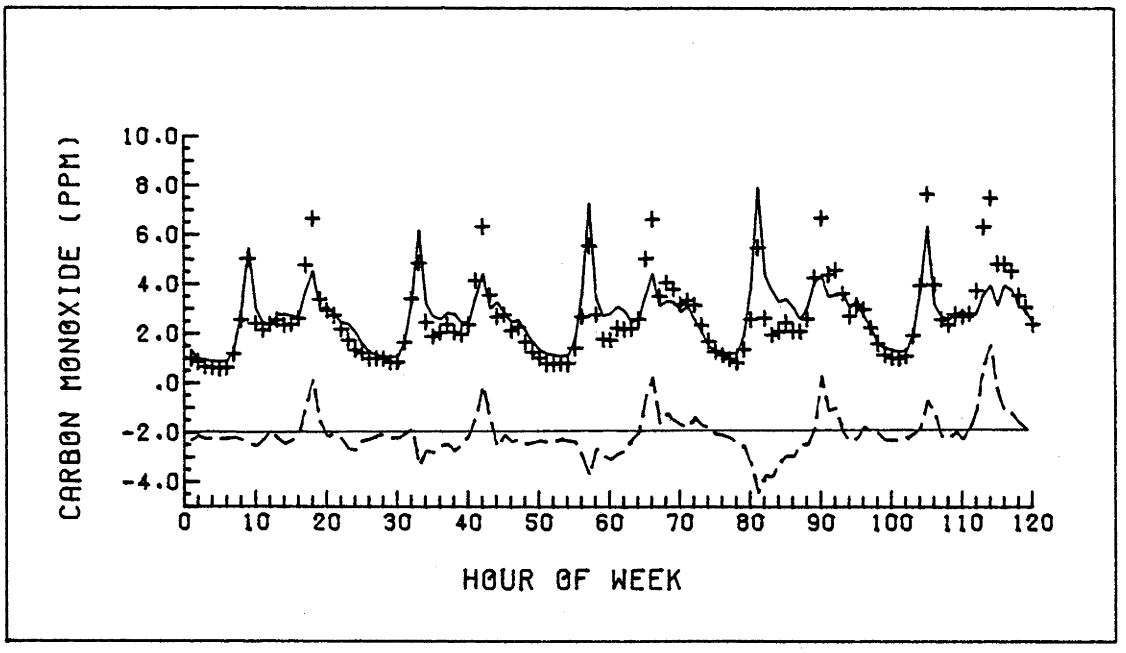


Figure 6.15 Model fit (——) to observed carbon monoxide levels (+) given by the (1,2) model with E_k^2 as the effective traffic input. Estimated residuals are plotted directly underneath.

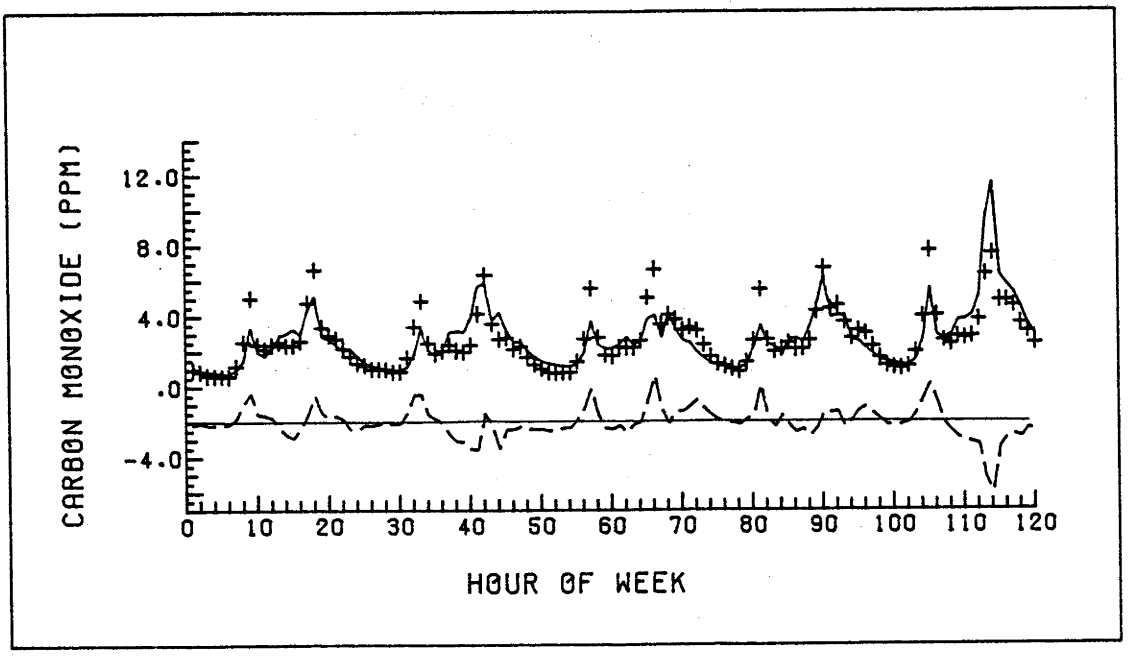


Figure 6.16 Model fit (——) to observed carbon monoxide levels (+) given by the (1,2) model with E_k^3 as the input.

As the objective of the modeling procedure we have adopted here is to separate the overall model into linear and non-linear components and to secure time invariant parameters in the linear component, a case might be made that the model in which 'effective traffic' was defined as E_k^2 came closest to meeting this objective. Although as previously indicated some parameter variation remains (Figure 6.13), a plausible explanation for this is that it is strictly associated with the afternoon peak traffic flow. Specifically we require an explanation for the higher carbon monoxide concentrations per vehicle during the afternoon peak period relative to the remainder of the day. This suggests that the model may be failing to allow adequately for the higher level of carbon monoxide emissions from vehicles which are started from cold (see, for example, Roth et al., 1974). In the context of Canberra City's diurnal traffic movements, the afternoon peak flow almost certainly consists of a higher proportion of vehicles passing the vicinity of the monitoring station shortly after leaving nearby carparks, and hence with relatively cold engines, than is true of the morning peak when the traffic flow through the area more largely consists of vehicles nearing the completion of their journey. This tentative hypothesis may be amplified by a number of supporting observations. The monitoring station is adjacent to a number of large carparks, in which parking was largely unrestricted and hence used by all-day parkers working in the city area. The majority of vehicles leaving the carpark during the afternoon peak period would thus have cold engines, generating higher carbon monoxide emissions. An associated feature of the parking arrangements in the area is that vehicles leaving carparks must obey a 'Give Way' to all traffic rule, which leads to engine idling, and thus also to higher carbon monoxide emissions. Finally, we should note that during these afternoon peak periods the carparks empty relatively quickly compared to the remainder of the day when a smaller turnover of vehicles occurs.

One means by which the 'cold start' hypothesis might be subjected to assessment within the context of the existing models is to attempt a validation of the model estimated on the weekday data by forecasting weekend day pollutant concentrations. The pattern of weekend traffic flow is very different with no morning or afternoon peak, and the utilisation of carpark capacity was markedly lower than during

weekdays. Hence the failure of the model to account fully for the weekday afternoon peak concentrations should be no barrier to its usefulness as a description of the weekend day concentration levels if the 'cold start' hypothesis is valid and given that there exists no comparable pattern of traffic movement on weekend days. The results of this out-of-sample validation of the model (and associated test of the 'cold start' hypothesis) is shown in Figure 6.17. The model is relatively successful in capturing fluctuations in carbon monoxide concentrations, but fairly consistently overestimates the concentration levels throughout the weekend. A corollary of the former finding is that the validation test lends support to the 'cold start' hypothesis in that it establishes the cause of the model's underestimation of carbon monoxide levels during weekday afternoon peak hours as deriving from some feature of the system behaviour on weekdays which is not present during the weekend.

Taking a broader perspective of all the modeling results, some comment on the methodological approach seems warranted. In general, simply to obtain an overall model consisting of separated linear and non-linear components in which the linear component has time invariant parameters does not necessarily imply that the model structure is the best possible representation of the underlying relationships. This is because the evaluation of the model takes the form of a goodness-of-fit test, the absence of any evidence of over-parameterisation, and the successful validation of the model for a period over which data had not been used for the identification and estimation. However, in the present case, if a satisfactory model had been obtained, then the 'correctness' of the model structure would not be critical because the objective of the modeling exercise is the provision of simple models which can be used as aids in the management of air quality. The fact that we did not achieve a satisfactory model even by resorting to 'averaged' data indicates that the system behaviour is more complex than was originally anticipated. It is likely, therefore, that further improvement in the model will only be possible by resorting to the first stage of the model building procedure outlined in Chapter 1, namely, speculative simulation modeling.

Such an exercise in simulation modeling would permit a systematic evaluation of the importance of the variables that have already been considered for inclusion in the model structure and plausible

candidates for variables not yet included. The speculative simulation modeling might have as its basis relevant knowledge from meteorology and traffic engineering, and pinpoint additional data required for a thorough evaluation of the model structure. We are thus in the unsatisfactory position that either suggesting other omitted influences which might account for the mixed results so far obtained, or offering guidance on further amendments to the model structure, must of necessity be somewhat speculative.

First, some further brief observations on the data and results should be made. It is noticeable that the E_k^2 model consistently over-estimates the concentration levels at times of low traffic flow, not just during weekends (as previously noted) but also during weekdays. It is also of interest to note that there exists a strong coincidence between the periods of peak traffic flow during weekday mornings and afternoons on the one hand and, on the other, periods during which the wind direction is changing (from east to west or vice versa). Together with this coincidence we note that the E_k^3 model, which incorporates the easterly component, fails adequately to estimate the peak period concentrations (although the evidence for the 'cold start' hypothesis in this model is somewhat reduced). It is unclear, however, whether or not this failure resides in poor incorporation of wind direction in the model.

Second, it is worth pointing out that an attempt was also made to validate the E_k^3 model against the weekend observations, but with similar results to those obtained (and previously discussed) when the E_k^2 model was employed. A further example of testing the results so far reported to different assumptions is that we established the relative insensitivity of the performance of the E_k^2 and E_k^3 models to changes in the value of β and β' respectively.

Third, inversion height was not included in any of the models because the data were unavailable. This variable has been included in some air pollution models and found to be of importance in determining urban carbon monoxide levels (Phadke et al., 1976, and Remsberg et al., 1979). As the Canberra region experiences a high frequency of radiation inversions (Daw and O'Loughlin, 1972), we may venture the view that the omission of inversion height may be significant in any explanation of the results we have obtained.

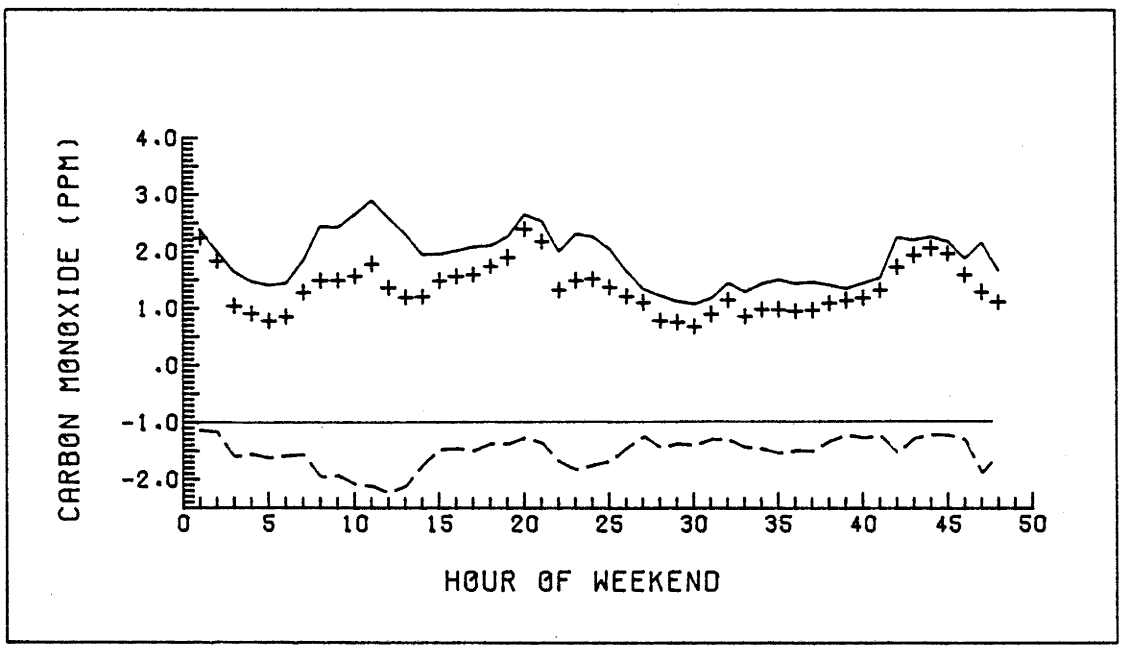


Figure 6.17 Comparison of observed carbon monoxide levels on weekends (+) with predictions (——) made using the (1,2) model with E_k^2 input and estimated on weekday data.

Finally, we conclude by re-emphasising that the principal goal of these modeling exercises is to attain models of urban air pollution suitable for air quality management. It seems necessary, therefore, that any further research should attempt to meet this objective more fully. For example, it would be desirable to design models that performed well on concentration data that had not been averaged. Since a commonly adopted air quality standard for carbon monoxide is specified in terms of the frequency of occurrence of eight-hour average concentrations, it is particularly important that unaveraged data provide the basis for such models. However, this goal has proved very elusive, not just in this study, but in several other attempts to model carbon monoxide levels. Thus, Tiao et al. (1975), Hanna (1978) and Tiao and Hillmer (1978) reported models which were based only on averaged data, while attempts to model unaveraged data have not been very successful (see Reynolds et al., 1974). A related suggestion is that the seasonal variations be incorporated explicitly into any extended model. For example, it is known that the incidence of radiation inversion in Canberra is higher in autumn and winter (Daw and O"Loughlin, 1972). It is also known that easterly winds dominate during summer but that during the remainder of the year north-westerly winds are predominant (Kalma et al., 1974).

Chapter 7

INVESTIGATION OF THE DYNAMICS OF A DETERMINISTIC AIR POLLUTION MODEL

7.1 Introduction

In this chapter a preliminary investigation of the dynamics of the Macquarie Urban Air Quality (MUAQ) model is reported. In its present version this model simulates the dispersion of an inert air pollutant in the atmosphere over an urban area, and is based upon a model known as the SAI (System Applications, Inc.) model which was originally developed by Reynolds et al. (1973, 1974) and Roth et al. (1974) to simulate the dispersion of photochemical smog in Los Angeles. The SAI model is an example of a model which, having originally been designed for one particular urban area, has subsequently been applied to others. Thus it has been refined (Reynolds et al., 1976) and applied to the Denver air basin (Anderson et al., 1977). In Australia application of the SAI model has been made to the Sydney air basin (Johnson, 1980a) while the present version of the MUAQ model has been applied to that of Canberra (Johnson, 1980b).

It is beyond the scope of this thesis to attempt a complete validation of either the SAI or the MUAQ model. Any such exercise should include an examination of their dynamic behaviour, since unacceptable dynamics, such as physically implausible properties, may bring seriously into question the overall performance of the model. These types of models may be considered as systems whose internal dynamics may be examined by means of their input-output relationships. As far as the author is aware an examination of the dynamics of such models has not been reported in the literature.

In the investigation of the MUAQ model reported in this chapter an input signal representing air pollutant emissions is introduced and the modeled atmospheric concentrations of air pollutant are considered to be the output signal. The methods of model identification described and demonstrated in previous chapters are then used to identify linear rational transfer function representations of the MUAQ model. The recursive procedures readily permit the investigation of any parametric variation in the identified models. Since the MUAQ model is based substantially on the SAI model, it is convenient first to describe the

latter and then subsequently outline how the MUAQ model differs from it.

7.2 The SAI Model

The original version of the SAI model has been described in detail by Reynolds et al. (1973, 1974) and Roth et al. (1974) and only a brief outline will be given here. The SAI model is of the Eulerian multi-box type and is based upon the advection-diffusion equations described previously (Section 2.4). The air basin is divided into cells by laying down a fixed uniform horizontal grid with each grid square measuring 2 x 2 miles. The horizontal boundaries of the modeling region are taken to be the ground and the base of the inversion layer, since it is assumed that the inversion base is the upper limit for vertical mixing or transport. The vertical scale is divided into five layers of equal depth (H-h)/5 where h(x,y) is the ground elevation above sea level at (x,y) and H(x,y,t) is the ground elevation above sea level of the inversion base at (x,y) and time t. Due to the variations in h and H with x and y and the variation of H with time, the modeling region has an irregular 'floor' and 'ceiling'. The numerical methods used to solve the partial differential equations of the model cannot easily cope with these irregularities, and a change of variables is performed to remove them. The northern, southern, eastern and western boundaries of the modeling region are usually chosen so as to include all of the significant sources of air pollution in the air basin as well as the receptor areas likely to be affected by the transport of pollutants. The center of each cell is the node and is the point to which values of all variables are assigned or referenced.

The SAI model can be visualised as consisting of three sub-models, namely a meteorological sub-model, an emissions sub-model and a sub-model to describe the chemistry of the photochemical smog formation process. This last-mentioned sub-model may be 'switched off' if only inert pollutants are being modeled. The major task of the meteorological sub-model is the construction of a wind field in which components of the wind velocity are specified at each node and each time step. The wind field is constructed on the basis of observations of wind speed and direction in the modeling region. Such observations normally consist of data from a network of ground based anemometers

with perhaps a few measurements of the winds aloft. Due to the paucity or even complete lack of measurements of the latter it was assumed in the original SAI model that the wind velocity components were identical for all levels between the ground and the inversion base. The absence of measurements of the winds aloft also means that the turbulent diffusivities cannot be estimated directly. Reynolds et al. (1973) assumed that for Los Angeles the horizontal diffusivity was constant at 2980 m^2 min^{-1} over the whole air basin, and used a relationship derived from Eschenroeder and Martinez (1969) to describe the variation of the vertical diffusivity with height. Subsequent improvements to the meteorological sub-model have allowed the treatment of wind-shear phenomena and the generation of a wind field that is mass consistent (Seinfeld and Wilson, 1977).

The emissions sub-model is described in detail by Roth et al. (1974) and is essentially a record of the spatial and temporal nature of pollutant emissions. Emissions from elevated sources enter the appropriate cell directly and are treated separately from emissions from ground level sources. The latter are treated as surface fluxes defined at the base of each ground level cell and are specified in the boundary conditions.

The sub-model describing the chemistry of smog formation in the original SAI model consists of a compact generalised reaction mechanism for the photo-oxidation of a complex mixture of hydrocarbons. All the reactive hydrocarbons are lumped into two fictitious species, and all of the peroxy radicals (those capable of oxidising nitric oxide to nitrogen dioxide) are lumped into a single fictitious species. The reactive radical species are assumed to be in the pseudo-steady state, while carbon monoxide and water vapour are treated as having constant concentration since they are present in excess. Thus there are five differential equations describing the change with time in concentration of the species nitric oxide, nitrogen dioxide, ozone and the two lumped hydrocarbon species. A good discussion of the rationale behind the choice of a reaction mechanism for inclusion in urban airshed models is given by Lamb and Seinfeld (1973). An improved reaction mechanism, known as the carbon-bond mechanism, has been developed for describing oxidant formation (Whitten et al., 1980) and has been incorporated in more recent versions of the SAI model (see Anderson et al., 1977).

The overall model is implemented by specifying initial and boundary conditions and then solving the coupled advection-diffusion equations which are in four-dimensional space. The numerical solution of the equations is obtained by use of the method of fractional steps described by Yanenko (1971) where each four-dimensional equation is split into three two-dimensional equations. Then each of the two-dimensional equations is integrated in succession over one time step, and upon completion, the integration of the full four-dimensional equation over one time step has been approximated. Later versions of the SAI model (see Anderson et al., 1977) obtain the numerical solution of the model equations by employing the method of Boris and Book (1973) to treat the advection term and a Crank-Nicholson method for the vertical diffusion and chemistry.

7.3 The Macquarie Urban Air Quality (MUAQ) Model

While the MUAQ model derives from the SAI model, it differs in the major respect that the former simulates only the dispersion of an inert pollutant whereas the latter may be used also for modeling photochemical smog. That is, there is no chemistry sub-model within the present version of the MUAQ model. In this section the description of the MUAQ model will focus on its additional significant differences from the SAI model.

Whereas the modeling region for the original SAI model was irregularly shaped to cover the Los Angeles basin, that for the MUAQ model, which was developed for the Canberra area, was rectangular. Within the region the grid squares of the SAI model measured 2 x 2 miles, while those in the map reference co-ordinate system of the MUAQ model were 1 x 1 kilometers. Each square in this system could be unambiguously identified by the co-ordinates (in grid units) of its top right hand corner, referenced to the origin located at the lower left hand corner of the modeling region. Furthermore, in the MUAQ model the simulation area may be selected as some subset of the area defined by the map reference co-ordinate system (25 x 43 kilometers). The definition of the boundaries of this simulation area can be achieved simply by specifying its origin within the grid. An additional property of the MUAQ model which adds flexibility is that the grid system for the simulation area can be specified as either coarse (1 x 1 kilometer) or fine (0.5 x 0.5 kilometers), but constrained to be an

area of 25 x 25 squares whichever specification is chosen.

The structure of the MUAQ model and the information flow within it is shown in Figure 7.1. It can be seen that there are three basic parts: the meteorological data preparation program; the emissions data preparation system; and the atmospheric pollution simulation program.

The meteorological program takes raw meteorological data, generates from them a wind field and mixing depths for each hour of the day over the simulation area, and creates a meteorological data file to be used subsequently by the atmospheric pollution simulation program. Similarly, the emissions system takes emissions data, evaluates an average area source emission rate for each grid square of the area and for each hour of the day, and creates an emissions data file for later use by the atmospheric pollution simulation program.

Apart from the meteorological and emissions data files, the atmospheric pollution simulation program requires specification of the initial concentrations of the modeled pollutant over the simulation area, the background concentrations of the pollutant both aloft and on its northern, southern, eastern and western borders during the simulation period, and the emission rates from large point sources. The output of the atmospheric pollution simulation program comprises both the instantaneous concentrations in the ground-level cells and hourly averaged concentrations in all cells, specified in units of parts per million (ppm). The time interval (in minutes) between print-outs of the instantaneous concentrations must be specified in the atmospheric pollution program. The specification does not necessarily guarantee that the print-outs are produced at exactly this time interval. This is due to the numerical integration procedure automatically altering the time step in a fashion which leads to values below the maximum value of four minutes if this is necessary to achieve the specified accuracy in the numerical integration. Instructions for the use of the MUAQ model have been prepared by Johnson (1979b, 1979c, 1979d, 1980c).

7.4 Experimental Procedure

The simple test of the MUAQ model can be thought of as a tracer experiment carried out to investigate its behaviour - an approach

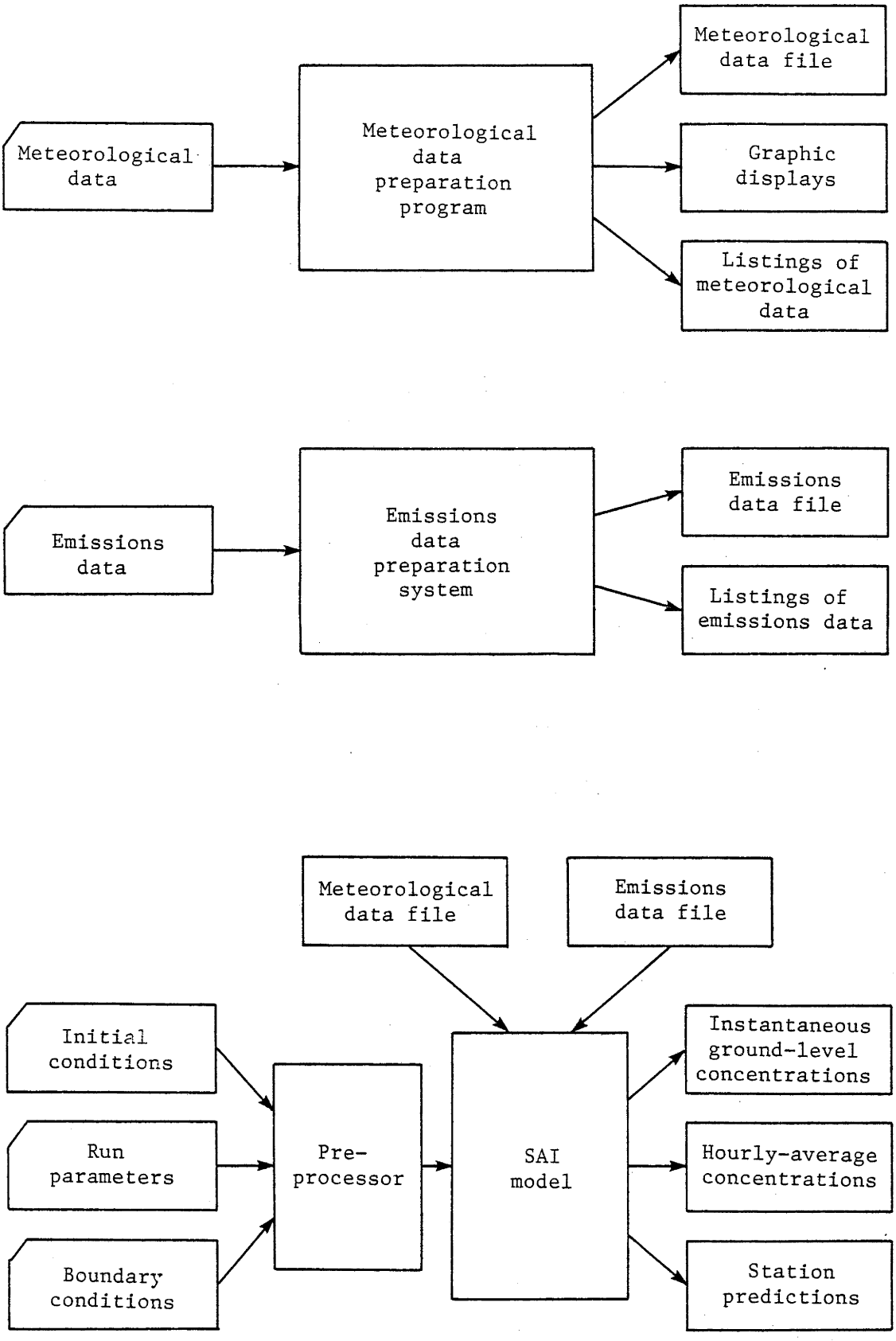


Figure 7.1 Flow of information in the Macquarie Urban Air Quality model. The preprocessor and the SAI model form the atmospheric pollution simulation program.

similar to performing such an experiment in the atmosphere itself. Sulphur dioxide was chosen as the pollutant and for the purposes of the experiment it was assumed to be inert. The temperature and atmospheric pressure were set at values of 20°C and 1000 millibars respectively and were not varied during the experiment. The one kilometer grid was selected with its origin at (0,0) and the single area source of sulphur dioxide was located in the grid square (1,12). The meteorological data for the experiments were supplied to the model via a single imaginary monitoring station at the point (12,12) for recording mixing depth, wind speed and wind direction. The mixing depth was chosen as 500 m while the wind was specified to be from due west, and also to have a speed of 1.5 ms^{-1}. All three were held constant at these values for the duration of the simulation experiment (that is, 0100 to 2200 hours). In addition it was assumed that over the entire simulation area no variation occurred in wind speed or direction from those set for the monitoring station. It should be emphasised, parenthetically, that the restrictive assumptions of a uniform wind and of a single area source of sulphur dioxide have been made the more readily to obtain time series representations of the MUAQ model. The horizontal and vertical diffusivities were specified in the same way as that described by Reynolds et al. (1973) for the original SAI model. Employing these conditions, and specifying a print-time interval of four minutes, led to the generation of instantaneous pollutant concentrations at precise four minute intervals, and thus led to 316 equally spaced observations in each ground-level grid cell during the simulation period (0100 to 2200 hours). However, it was found that simply changing the wind speed from 1.5 to 3.0ms^{-1} led to observations that were not equally spaced in time and hence were less amenable to the methods of time series analysis used here. Consequently, only those experiments which utilised a constant wind speed of 1.5ms^{-1} were selected for analysis.

The nominal emission rate of sulphur dioxide in the grid square (1,12) was chosen as $0.5 \times 10^{6.5}$ gm km^{-2} $hour^{-1}$ so that sufficiently high pollutant levels to permit input-output modeling were experienced at the maximum downwind distance of 25 kilometers. The actual emission rate for each hour of the day was varied by specification of a factor for each hour and then simply multiplying the nominal emission rate by the variation factor. The factors for each hour were randomly chosen to be either zero or unity in order to generate a random binary input signal. The emissions data preparation

system does not utilise this random binary signal directly, but linearly interpolates between the mid-points of the constant hourly values of the emission rates to produce a modified signal pattern which is shown, together with that of the original, in Figure 7.2.

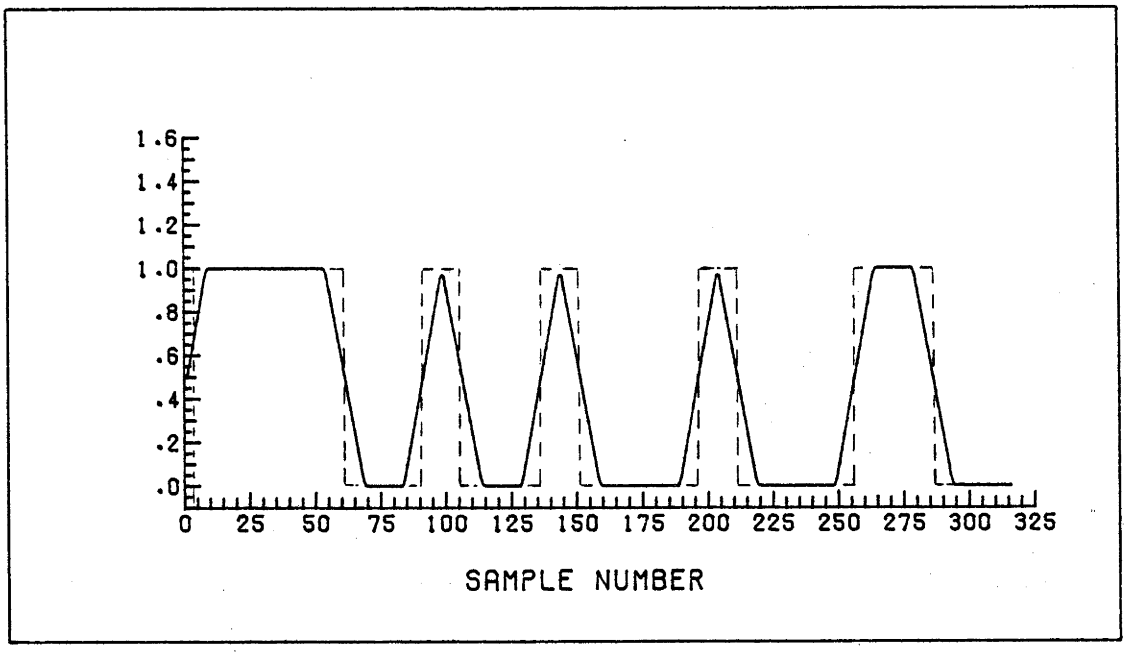


Figure 7.2 Original random binary input signal (---) and the interpolated signal (——) created by the emissions data preparation system.

Use of the previously enumerated conditions, together with initial and all background concentrations set to zero, led to premature termination of the program. This was found to be due to the generation of 'negative' pollutant concentrations, also a characteristic of the SAI model (Johnson, 1980a), and illustrates the point that care must be exercised in any modeling study to ensure that the model results are physically plausible. In the present experiment this problem of 'negative' concentrations was overcome by specifying initial and background concentrations of 0.1 ppm sulphur dioxide. This strategy resulted in pollutant concentrations which were sometimes less than the background level, but since none fell below zero the program continued to normal completion, thus enabling an investigation of the conditions which led to this undesirable behaviour in the first place.

Although this strategy was successful for the previously mentioned nominal emission rate of $0.5 \times 10^{6.5}$ gm km^{-2} $hour^{-1}$, it was found in subsequent experiments that increasing the emission rate

to $0.5 \times 10^{7.0}$ gm km^{-2} $hour^{-1}$ led to the recurrence of 'negative' pollutant concentrations. It is believed that these may have been avoided by increasing further the level of initial and background concentrations but this was not verified by further experiments. However, for the analyses reported in this chapter the nominal emission rate was maintained at the former level of $0.5 \times 10^{6.5}$ gm km^{-2} $hour^{-1}$, and at the completion of the experiment the constant background concentration of 0.1 ppm was subtracted from all the recorded output concentrations.

The results which are analysed and discussed in subsequent sections were therefore obtained by specifying for the simulation period a set of experimental conditions which may be summarised as follows:-

(i) a simulation period of 0100 to 2200 hours;
(ii) a one kilometer grid with its origin at (0,0);
(iii) a constant temperature of 20°C;
(iv) a constant pressure at 1000 millibars;
(v) a wind from due west at a constant speed of 1.5 ms^{-1} throughout the simulation area;
(vi) sulphur dioxide as the chosen pollutant (and assumed inert) from an area source in grid square (1,12) with a nominal emission rate of $0.5 \times 10^{6.5}$ gm km^{-2} $hour^{-1}$;
(vii) the initial and background concentrations set as 0.1 ppm sulphur dioxide; and
(viii) the instantaneous pollutant concentrations being printed every four minutes, yielding 316 observations in each ground-level grid cell.

7.5 Initial Results

Brief inspection of the MUAQ model's simulated pollutant concentrations reveals that its behaviour is broadly consistent with expectations, since the pollutant is seen to travel downwind and to be dispersed vertically and across-wind. To evaluate the performance of the model more carefully, it is convenient to focus separately on its cross-wind dispersion properties and its advection and longitudinal dispersion characteristics. This task is assisted by defining the (straight) line of ground level cells due east of the pollutant source (that is, cells (2,12) to (25,12)) as the reference cells. Then, the

advection property of the model was examined by use of the time series of instantaneous pollutant concentrations in the reference cells (5,12), (10,12), (15,12), (20,12) and (25,12). These time series are shown in Figure 7.3. By contrast, the cross-wind dispersion was examined by taking, in effect, cross-section 'snapshots' of the plume, thus enabling observation of the distribution of pollutant concentration around (that is, north and south of) the reference cells.

Several interesting points are revealed in the graphs in Figure 7.3. Firstly, the systematic occurrence of the 'negative' concentrations as the initial response to each 'step' in the input signal indicates that the MUAQ model has non-minimum phase characteristics (Truxal, 1955). Secondly, the magnitude of the 'negative' concentrations associated with the response to the first 'step' in the input signal is always larger than that of the 'negative' concentrations associated with the subsequent input 'steps'. This is almost certainly an artifact of the choice of 0100 hours rather than 0000 hours as the start of the simulation period. More explicitly, the emissions data preparation system creates a file containing emission data for each hour of the day, and also linearly interpolates between the mid-points of the constant hourly values of the emission rates. Hence, the emission rate at the specified start of the simulation period was 0.5 units (where one unit corresponds to an emission rate of $0.5 \times 10^{6.5}$ gm km^{-2} $hour^{-1}$) rather than zero units which would have applied had the start of the simulation period been chosen as 0000 hours. Thirdly, there is a suggestion that the output signal of the model has oscillatory characteristics, particularly in those reference cells furthest from the pollutant source. Finally, as the distance downwind of the source increases, the 'negative' concentrations associated with the response to all 'steps' in the input signal other than the first, are found to increase in magnitude whereas the peak concentrations decrease. To illustrate, in the reference cell (5,12), the peak concentration experienced was 0.4284 ppm, while the 'negative' concentration (of largest magnitude) was -0.0156 ppm. Comparison with the reference cell (25,12) yields figures of 0.0880 ppm and -0.0346 ppm respectively. It is suspected that the 'negative' concentrations arise as a result of the numerical integration procedure and this, together with the points noted above, will be discussed further in Section 7.7.

To complete this section we consider the cross-wind dispersion

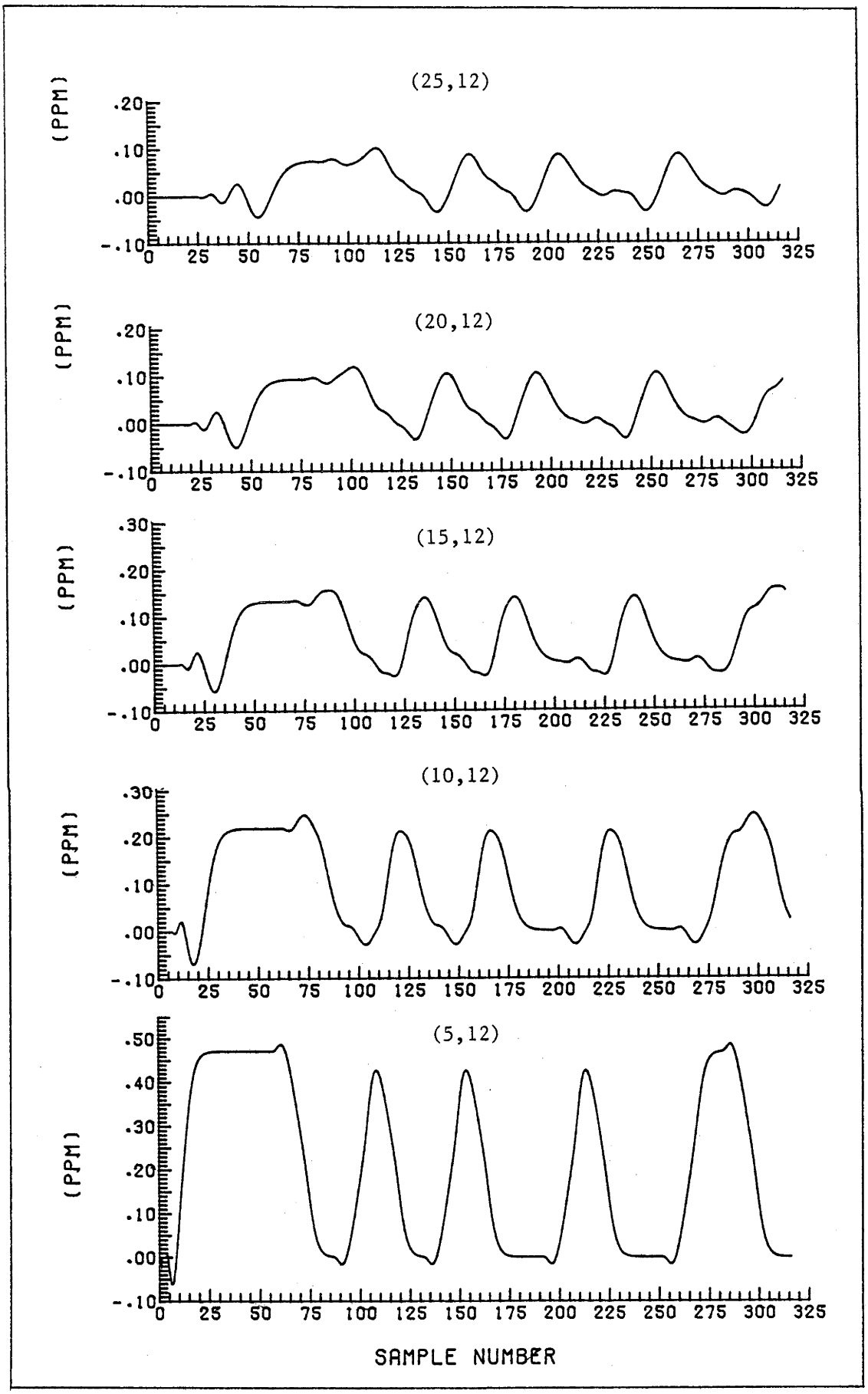


Figure 7.3 Time series of instantaneous pollutant concentrations (ppm) in each of five reference cells.

characteristics of the MUAQ model. As mentioned previously this is done simply by taking 'snapshots' of the plume to enable observation of the distribution of pollutant concentration around the reference cells. Since we have chosen to examine the advection characteristics of the model by analysis of the instantaneous concentrations in the five evenly spaced reference cells (5,12) to (25,12), we will use the same reference cells when considering the cross-wind dispersion. The times at which the 'snapshots' were taken were selected so that a peak concentration of pollutant, resulting from the same 'step' in the input signal, was occurring at the five specified downwind distances.

The selected cross-section pollutant concentrations are shown in Table 7.1 and plotted in Figure 7.4. Clearly, cross-wind dispersion is taking place but the pollutant concentrations are not normally distributed around the reference cells as expected. A consistent pattern emerges in which the pollutant concentrations in cells one kilometer north of the reference cells are approximately double those in the cells one kilometer south, for all distances downwind. A clear explanation of this unanticipated asymmetry of the pollutant plume is not immediately apparent, but further examination of this problem lies beyond the scope of this chapter, even though it indicates some deficiencies in the performance of the MUAQ model. It is tentatively suggested that the asymmetry of the plume may result from errors in the specification of wind direction within the model.

7.6 Input-Output Modeling Results

Input-output modeling clearly requires input and output series and some method of determining their relationship. Both the input and the output series used here have been described in previous sections. The input was shown in Figure 7.2 as the modified signal produced as a result of the emissions data preparation system linearly interpolating between the mid-points of the constant hourly values of the emission rates. The output signals were shown in Figure 7.3 as the time series for the reference cells (5,12), (10,12), (15,12), (20,12) and (25,12). In order to relate the input to the output series, models of the linear rational transfer function form were sought. The structure of these models was identified using the IV model identification procedure described in Chapter 3.

The principal objective of this modeling is to determine whether the MUAQ model possessed any dynamic characteristics, and if so, their nature. To do this, we attempt to find a transfer function representation of the input-output relationships of the model.

Table 7.1

Cross-Wind Dispersion of Pollutant by Distance Downwind and Time Elapsed

(ppm)

Cells	5 km (0808)	10 km (0900)	15 km (0956)	20 km (1048)	25 km (1136)
4 North	-0.0001	-0.0008	-0.0014	-0.0002	-0.0003
3 North	-0.0021	-0.0027	0.0004	0.0048	0.0097
2 North	-0.0021	0.0148	0.0273	0.0356	0.0415
1 North	0.1013	0.1108	0.1005	0.0906	0.0838
Reference	0.4284	0.2137	0.1401	0.1056	0.0880
1 South	0.0499	0.0557	0.0515	0.0472	0.0442
2 South	0.0027	0.0072	0.0102	0.0122	0.0137
3 South	0.0001	0.0006	0.0013	0.0022	0.0030
4 South	0.0000	0.0000	0.0001	0.0003	0.0005

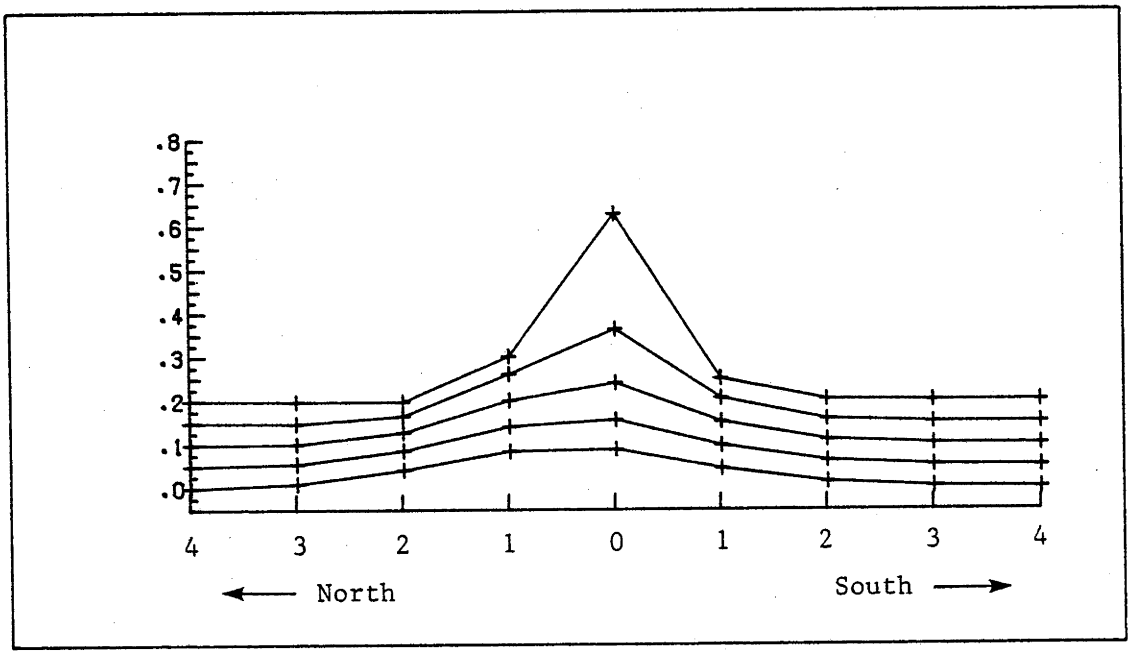


Figure 7.4 Instantaneous cross-section pollutant concentrations (ppm) in cells north and south of the reference cells. Plots are separated vertically by 0.05 ppm for clarity.

If successful, the dynamics of the MUAQ model will be captured in the 'a' parameters of the transfer function model. For example, a representation with 'b' parameters but no 'a' parameters would indicate that the MUAQ model possessed no inherent 'long term' dynamics, merely a dynamic behaviour where the dynamic 'memory' was limited to a period of time corresponding to that number of sampling intervals equal to the number of 'b' coefficients. By contrast, in representations containing 'a' parameters the 'memory' will extend into the infinite past.

When modeling real world situations it is usual to give particular emphasis to the basic requirement that models satisfy the condition of physical plausibility. However, in the present case our objective is simply to find a transfer function representation of the MUAQ model. Hence, if the MUAQ model exhibits physically implausible behaviour by generating 'negative' pollutant concentrations, it is desirable to capture this in the transfer function representation.

The extent to which the data are 'explained' is measured by R_T^2, one of the primary test statistics used in the model identification procedure. Together with other statistics, the R_T^2 values are shown in Table 7.2 for selected model structures. In all cases the modeling is from the area source in grid square (1,12) to the specified reference cells. The selection of models is only made, however, after reference to additional model attributes. Since two of these attributes are good parameter definition and no sign of over-parameterisation, we seek models that have NEVN values that are both low and close to the minimum. Those models which simultaneously possess attractive primary test statistics are then examined to see if their parameters are time invariant, and if there is correlation between the model input and the estimated residuals $\hat{\xi}_k$.

Inspection of Table 7.2 shows that for the reference cell (5,12) there are two models with very favourable primary test statistics, namely the (1,1,9) and (0,1,10) models. While the former offers a slightly better fit to the data, the latter is more efficient (only one parameter) and has better parameter definition. Both are physically plausible in the sense that they do not generate 'negative' concentrations from non-negative (by definition) pollutant emissions.

Furthermore, either is a satisfactory representation of the relationship between the pollutant emissions at the source and the concentrations in the cell (5,12). For the remaining four reference cells the primary test statistics unambiguously indicate a (0,2) model structure as the most acceptable, with R_T^2 at the plateau level, and NEVN values close to their minimum. These (0,2) models, like the MUAQ model, may predict 'negative' concentrations, because the 'b_0'

Table 7.2

Model Order Identification Criteria for Input-Output Representations of the MUAQ Model

Grid Cell where Output Measured	Model (n,m,τ)	R_T^2	ln(NEVN)	ln(EVN)
(5,12)	(0,1,8)	.9281	-9.8621	-10.6372
	(0,1,9)	.9727	-10.8412	-11.6043
	(0,1,10)	.9946	-12.4552	-13.2133
	(0,1,11)	.9939	-12.3351	-13.0931
	(0,2,9)	.9944	-6.7074	-7.9847
	(0,3,9)	.9951	-4.1062	-4.9573
	(1,1,8)	.9978	-9.5970	-10.5001
	* (1,1,9)	.9988	-9.5130	-10.3542
	(1,1,10)	.9978	-7.6765	-8.7363
	(1,2,9)	.9989	-5.7116	-7.6460
	(1,3,9)	.9987	-4.4216	-6.1575
(10,12)	(0,1,24)	.9811	-11.8702	-13.3520
	(0,1,25)	.9803	-6.7470	-8.0810
	(0,1,26)	.9460	-10.7753	-12.2691
	(0,2,24)	.9812	-4.8623	-8.1521
	* (0,2,25)	.9803	-6.7470	-8.0810
	(0,2,26)	.9765	-7.1482	-7.8768
	(0,3,25)	.9803	-3.3229	-4.9076
	(0,4,25)	.9808	-3.4281	-3.9717
	(1,1,25)	unstable		
	(1,2,25)	.9755	-3.6302	-3.8286
	(1,3,25)	.9730	-2.1215	-3.9300

Table 7.2 (continued)

Grid Cell where Output Measured	Model (n,m,τ)	R_T^2	ln(NEVN)	ln(EVN)
(15,12)	(0,1,39)	.9385	-11.0002	-12.9601
	(0,1,40)	.9038	-10.5202	-12.4870
	(0,1,41)	.8552	-10.0823	-12.0624
	(0,2,39)	.9595	-6.8438	-8.1276
	* (0,2,40)	.9546	-7.0921	-7.9945
	(0,2,41)	.9431	-7.0966	-7.7573
	(0,3,40)	.9548	-3.5090	-4.8827
	(0,4,40)	.9552	-3.1418	-3.9270
	(1,1,40)	unstable		
	(1,2,40)	unstable		
(20,12)	(0,1,52)	.9108	-10.8311	-13.1252
	(0,1,53)	.8771	-10.4754	-12.7781
	(0,1,54)	.8293	-10.1122	-12.4290
	(0,2,52)	.9304	-6.6049	-8.2528
	* (0,2,53)	.9264	-6.9321	-8.1804
	(0,2,54)	.9159	-7.0309	-8.0269
	(0,3,53)	.9265	-3.2025	-5.0276
	(0,4,53)	.9268	-2.7660	-4.0647
	(1,1,53)	unstable		
	(1,2,53)	unstable		
(25,12)	(0,1,65)	.8792	-10.6251	-13.1384
	(0,1,66)	.8481	-10.3692	-12.8964
	(0,1,67)	.8060	-10.0950	-12.6421
	(0,2,65)	.8998	-6.3898	-8.1662
	* (0,2,66)	.8949	-6.6797	-8.0968
	(0,2,67)	.8827	-6.7709	-7.9660
	(0,3,66)	.8952	-3.4820	-5.0545
	(0,4,66)	.8954	-2.5592	-4.0705
	(1,1,66)	unstable		
	(1,2,66)	unstable		

* denotes identified model

parameter is positive and the 'b_1' parameter is negative in each case. These 'negative' concentrations will occur under conditions where the input emissions are zero at one sample, but greater than zero at the previous sample. It should be noted that these (0,2) models are not strictly non-minimum phase, in that the initial model response to the input signal results in an output signal which initially moves in the same direction. This initial conformity between the direction of movement of the two signals is, however, an outcome of the choice of the pure time delays being made in order to maximise the value of R_T^2 for each reference cell.

The parameter values in the identified models are estimated by use of refined IVAML procedures. The system and noise model parameter estimates are shown in Table 7.3 with their standard errors, together with the value of R_T^2 for each identified model. It is evident that as the distance downwind of the source increases the values of R_T^2 and the steady state gain both decline. This was anticipated. The decline in the values of R_T^2 is due primarily to the worsening of the 'negative' concentrations as distance downwind increases, while the decline in the steady state gain values is consistent with increasing pollutant dispersion associated with distance from the pollutant source. The model fit to the pollutant concentrations in each of the five reference cells is shown in Figure 7.5.

An additional point relating to the identified models is warranted. The most appropriate pure time delays associated with each of these models closely conform to those time intervals that would be required for a uniform wind of speed 1.5 ms^{-1} (5.4 km hr^{-1}) to traverse the distance between the pollutant source at (1,12) and the relevant reference cells downwind. Since the pollution will have to travel only four kilometers to reach grid square (5,12), and recalling that the time series sampling interval is four minutes, the expected time of travel may be computed and compared with the observed pure time delay. Such a comparison is made for all the reference cells, and is shown in Table 7.4. These satisfactory results confirm that the MUAQ model provides an accurate description of the mean speed of pollutant transport.

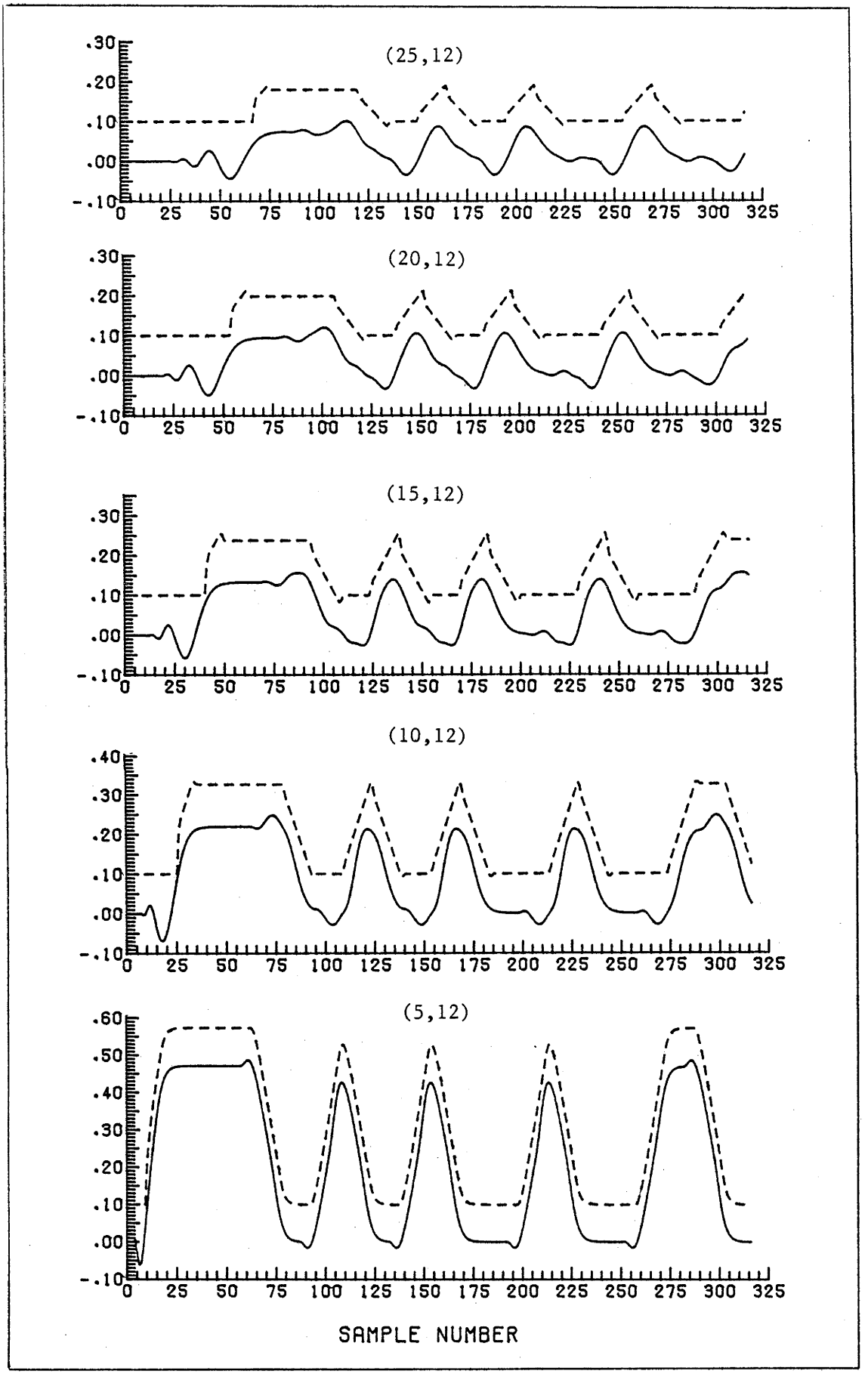


Figure 7.5 Time series model fit (---) to the output signal (—) of the MUAQ model in each of five reference cells. The model fit has been translated vertically by 0.10 ppm in each case for clarity.

Table 7.3

Summary of the Identified Input-Output Model Representations of the MUAQ Model

Output from Grid Cell	Model Parameters	Parameter Estimates[a]	Standard Error[a]	SSG[b]	R_T^2	Time Delay
(5,12)	a_1	-0.6032	.0042	.4740	.9988	9
	b_0	0.1881	.0020			
	d_1	0.3461	.0139			
(10,12)	b_0	0.3121	.0212	.2263	.9803	25
	b_1	-0.0858	.0212			
	d_1	0.6556	.0272			
(15,12)	b_0	0.4125	.0526	.1383	.9546	40
	b_1	-0.2742	.0526			
	c_1	-0.4952	.1230			
	d_1	-0.0998	.1464			
(20,12)	b_0	0.3233	.0329	.0985	.9264	53
	b_1	-0.2248	.0330			
	d_1	0.3396	.0539			
(25,12)	b_0	0.2302	.0238	.0797	.8949	66
	b_1	-0.1505	.0237			
	d_1	0.5322	.0555			

[a] Obtained by using refined IVAML procedures.

[b] Steady state gain.

7.7 Discussion and Conclusions

There are three major findings from this preliminary investigation of the MUAQ model. Firstly, the model exhibits non-minimum phase behaviour which may result in the generation of physically implausible 'negative' pollutant concentrations, the magnitude of which increases with increasing distance from the pollutant source. The generation of 'negative' pollutant

Table 7.4

Comparison of Pure Time Delays and Expected Times of Travel

Reference Cell	Distance of Reference Cell Source (km)	Expected Time of Travel from Source to Reference Cell[a,b]	Pure Time Delay[b] Integer found from Model Identification
(5,12)	4	11.1	9 or 10[c]
(10,12)	9	25.0	25
(15,12)	14	38.9	40
(20,12)	19	52.8	53
(25,12)	24	66.6	66

[a] Wind speed constant at 1.5 ms^{-1}.

[b] Times in units of four minutes.

[c] Depending on whether the (1,1,9) or (0,1,10) model is chosen.

concentrations is obviously an undesirable characteristic of the model. In the case of inert pollutants, serious errors may not necessarily arise in practice. This assessment is derived from the knowledge that, on the one hand, the 'negative' concentrations that the MUAQ model generates are of smaller magnitude for sites close to the pollutant source, and on the other, that in reality pollutant sources are spread throughout urban areas so that the most significant contributions to pollution levels at any given site will generally be made by nearby, upwind emission sources. In the case of secondary pollutants such as photochemical smog, however, this is not necessarily true, since transport of the precursor pollutants over considerable distances will often occur before the maximum concentrations of ozone are generated. Hence, it seems essential to examine both the SAI model, and the later versions of the MUAQ model (which will incorporate a capacity to simulate photochemical smog formation), to determine whether 'negative' concentrations of secondary pollutants are a serious problem with either.

Some further, brief comments concerning the 'negative' concentrations may be warranted. These concentrations are not inherent in the model formulation but almost certainly arise from the methods employed to solve the model's partial differential equations. Also, there appears to exist a strong resemblance between the initial model response (seen most clearly in reference cells (15,12), (20,12) and (25,12)) and the Padé approximation to a pure time delay (Korn and Korn, 1964).

Secondly, the results of our investigation into the MUAQ model reveal that negligible longitudinal dispersion occurs despite an expectation that clear evidence of such dispersion would emerge. This expectation is derived from knowledge that the model is based on the advection-diffusion equation and explicitly incorporates horizontal and vertical diffusivity terms K_H and K_V respectively (see Section 2.4). Some tentative explanation for the absence of dispersion may be offered. It may simply be an artifact of the conditions chosen for the particular experiment described here. This possibility can only be assessed by varying these conditions to ascertain whether the absence of longitudinal dispersion is a general feature of the MUAQ model. Such an exercise (which lies beyond our essentially preliminary inquiry into the MUAQ model) should include an investigation of whether the value of K_H (2980 m^2 min^{-1}) used in the experiment reported here may have been too small to allow a realistic simulation of atmospheric dispersion. The absence of longitudinal dispersion in the results is consistent with the previously mentioned finding that the model response strongly resembles a Padé approximation to a pure time delay. In other words, restricting attention to the relationships between the pollutant emissions at the source and pollutant concentrations at any of the downwind reference cells indicates that the behaviour of the MUAQ model appears to be well described in terms of a simple translation of the pollutant with a suitable reduction in its magnitude, that is

$$y_k = b_o u_{k-\tau} + b_1 u_{k-1-\tau} + \xi_k$$

One interpretation of this behaviour which seems most likely in practice is that, in the model's partial differential equation (see Section 2.4), the second order dispersion term for describing the

longitudinal dispersion is having a negligible effect compared to the advection term. The non-minimum phase oscillatory behaviour then could be interpreted as arising from the numerical solution of the resulting (largely first-order) partial differential equation. This first-order equation, of course, represents only pure translation. Such a numerical solution could well exhibit a behaviour like that of a Padé approximation to a pure time delay.

A third significant finding was that the MUAQ model displays a crosswind dispersion of pollutant resulting in concentrations which were not normally distributed around the reference cells. We have earlier advanced the tentative view that this asymmetry of the plume may be due to problems in the model's specification of wind direction.

It may be appropriate to conclude the preliminary investigation of the MUAQ model undertaken in this chapter with some broader comments pertaining to its suitability for air pollution modeling tasks. The MUAQ model is designed to facilitate the treatment of chemical interactions between pollutants. Since these interactions are non-linear it is not straightforward to use a simple additive procedure such as multiple source Gaussian plume models for chemically reactive pollutants. Hence we should expect that the computer simulation time for the experiment described here will be greatly in excess of that required for a Gaussian plume or Gaussian puff model simulation of it. Nevertheless, we should require that the MUAQ model at least provides acceptable estimates of pollutant concentrations for a single source of inert pollutant. If it does not meet this simple requirement, then its ability to handle multiple sources of pollutants must be questioned, not just where these pollutants are inert, but particularly in the more difficult case where they are chemically reactive.

Our final comment takes the form of a reiteration. The investigation of the MUAQ model reported in this chapter was restricted to the simplest case of a single pollutant source and uniform wind precisely in order that time series representations of it could easily be obtained. Although this strategy resulted in a useful elucidation of the limitations of the MUAQ model (and led us to recommend that some attention be given to their rectification), we do not wish to imply that it can be supplanted by the time series models. This is because the MUAQ model has the distinct advantage of being an efficient

formalisation of basically simple rules for describing the dispersion (and chemical interaction) of air pollutants from many scattered sources over large urban regions. The fact that a simple time series representation of the MUAQ model could easily be obtained was methodologically desirable, however, in that it not only highlighted the model's essential properties but in addition greatly facilitated the evaluation of its dynamic characteristics and more readily permitted recommendations to be made for its further development and improvement. Specifically, it is recommended that the diagnostic methods employed in this chapter should be applied to the refined version of the SAI model (Reynolds et al., 1976) to investigate whether the improved numerical integration procedures incorporated therein have overcome any or all of the problems we have found in the MUAQ model.

Chapter 8

SUMMARY AND CONCLUSIONS

The aim of this thesis has been to investigate the usefulness of recursive estimation methods in the construction of air pollution models for the purposes of air quality management and in the provision of improved measurements of air pollution. The position which has been adopted in this thesis is that air pollution models designed for control and management purposes should be simple and incorporate only as much complexity and data intensiveness as is strictly necessary to achieve consistency with the available air pollution measurements and the goals of the modeling. In particular, avoidance of over-parameterisation or 'surplus content' has been emphasised. Namely, such models should not possess characteristics superfluous to the explanation of observed phenomena or incapable of empirical validation.

The thesis essentially comprised two broad sections. In Chapters 2 and 3 the air quality management approach to the control of air pollution was described, the major types of models employed in the management of air pollution were reviewed, and a description was offered of the recursive estimation techniques extensively used in this thesis. These chapters thus provided a necessary context for the empirical investigations made into aspects of air pollution modeling and monitoring and reported in the remaining chapters of the thesis.

Transfer function models of a continuous air pollution analyser were obtained and used in the construction of robust input estimation algorithms. Then data missing from air pollution time series were estimated by use of simple linear dynamic models which related air pollution measurements made at different geographical locations. Next an attempt was made to develop a simple time series model of carbon monoxide levels in an urban area. Finally, a preliminary investigation was undertaken into the dynamic properties of an Eulerian model used for the simulation of urban air pollution levels.

The principal findings of this thesis may be summarised concisely. Firstly, recursive IV techniques were found to be useful in the

identification of model order and in the estimation of parameters in transfer function models of a continuous air pollution analyser. These models provided the basis for the construction of robust input estimation algorithms, which in turn have the potential to yield more accurate estimates of air pollution levels (Chapter 4). Second, recursive estimation methods were also found to provide a convenient means for examining the dynamic behaviour of a deterministic simulation model for urban air pollution. A simple time series model representation of the MUAQ model's input-output relationships was readily obtained, the analytical significance of which was that the former enabled the inherent dynamic properties of the latter to be more easily comprehended (Chapter 7). Third, recursive methods were again employed as an aid in the analysis of air pollution and meteorological data with the final objective of developing for air quality management purposes a simple time series model of carbon monoxide levels in an urban environment. Regrettably, somewhat inconclusive results were obtained from these modeling exercises. In part this was because a successful decomposition of the model into linear and non-linear components was not achieved, and partly due to the fact that, despite use of 'averaged' data, the system behaviour was found to be more complex than originally anticipated (Chapter 6). Finally, linear dynamic models were shown to provide a simple method for estimating missing air pollution data. In the course of this exercise recursive refined IVAML algorithms were developed for estimating the parameters in a multiple input transfer function (MITF) model of linear stochastic systems. This MITF model can be regarded as the dynamic equivalent of a linear regression model. By use of a comprehensive stochastic Monte Carlo simulation exercise, the IVAML algorithms were seen to perform very satisfactorily (Chapter 5).

This thesis has demonstrated that, given the important goal of improved air quality management, recursive estimation techniques are a powerful analytical tool both in the construction of air pollution models and for providing better measurements of atmospheric pollutant levels. It is not claimed, however, that the inquiries reported here fully explore the possible applications of recursive estimation methods in the analysis of air pollution relevant to the design of policies for air quality management. Specifically, further research in the following two directions would seem potentially most fruitful. First, application

to other dynamic models of the methodology adopted in Chapter 7 seems warranted. That is, using recursive methods to secure simple representations of more complex models, which yet retain all their essential or 'dominant mode' properties, provides an additional important technique for assessing the performance of deterministic simulation models. Second, attention might profitably be given both to speculative simulation modeling of levels of vehicular pollutants in urban areas, with the objective of identifying variables which (in addition to those we have incorporated, namely, traffic flow and wind components) significantly influence measured pollutant levels, and of securing satisfactory models without recourse to 'averaging' of the underlying pollution data being necessary. If these can be achieved, recursive estimation methods can then be brought to bear in the pursuit of simple models suitable for management purposes.

APPENDIX

As was seen in Chapter 4 Mage and Noghrey (1972) described continuous air pollution analysers in terms of the convolution integral equation, and obtained estimates of the true input pollutant concentration by employing a finite difference method to solve this equation. In this appendix it is demonstrated that finite difference methods will not always provide a satisfactory solution to the convolution integral equation.

The continuous time convolution integral equation may be written as

$$x(t) = \int_0^t r(t-w)u(w)dw \qquad 0 \leq w \leq t \leq T \tag{A.1}$$

where x is the output, u is the input, r is the impulse response, t denotes time, w is a dummy variable introduced for the purpose of integration, and attention is restricted to the time interval (0,T). A finite difference approximation for this (or any other) formulation involves a two step procedure. Firstly the formulation is discretised on an (abscissa) grid which we choose to be of uniform spacing Δ so that the k-th grid point has a data value x_k for x(t). If N is the number of samples, then $\Delta = T/N-1$ and equation (A.1) becomes

$$x_k = \int_0^{k\Delta} r(k\Delta-w)u(w)dw$$

$$\text{i.e.} \quad x_k = \sum_{i=0}^{k-1} \int_{i\Delta}^{(i+1)\Delta} r(k\Delta-w)u(w)dw \qquad (k = 1,2, \ldots ,N) \tag{A.2}$$

Secondly in order to approximate the integrals in (A.2), a quadrature rule is applied so that, in general, equation (A.2) can be written as

$$x_k = \sum_{i=0}^{k-1} w_{i+1}\, r_{k-i}\, u_{i+1} \qquad (k = 1,2, \ldots ,N) \tag{A.3}$$

where the w_i are the weights corresponding to the quadrature rule

used and the r_k and u_i are appropriate discretisations. In an Euler (simple rectangular) quadrate, for example, w_i is Δ for all i. If we designate that here we are solving for the impulse response function r then the following arguments also apply to the solution for the input function u because r and u are interchangeable under the convolution operation.

The equations (A.3) form a triangular system and the solution may be obtained directly as

$$r_1 = x_1/(w_1 u_1)$$

$$r_k = (x_k - \sum_{i=1}^{k-1} w_{i+1}\, r_{k-i}\, u_{i+1})/(w_1 u_1) \qquad \text{for } k = (2, \ldots, N) \tag{A.4}$$

Unfortunately it does not seem to be well known by applied scientists that unconstrained finite difference algorithms like (A.4) for the solution of equation (A.1) are unsatisfactory in the presence of even moderate noise levels on the x_k, especially if the formulation (A.1) is ill-posed (Tihonov, 1963). One example of ill-posedness occurs when the response $r(t-s)$ is given by $(t-s)^{-1/2}$ in which case equation (A.1) is called an Abel equation (Jakeman and Anderssen, 1975). In such an equation small errors in the data yield much larger error perturbations in the solution. This can be demonstrated very simply by considering that there is noise δx_k on just <u>one</u> output measurement x_k. Then from (A.4), at the k^{th} grid point the estimate for r_k is given by

$$r_k + \delta r_k = (x_k + \delta x_k - \sum_{i=1}^{k-1} w_{i+1}\, r_{k-i}\, u_{i+1})/(w_1 u_1)$$

where clearly $\delta r_k = \delta x_k/(w_1 u_1)$. In general, w_1 decreases as Δ decreases so that the error δr_k in the solution increases as the discretisation becomes finer. In other words, the more dense the data, the larger is the quadrature error. However, the use of a coarser grid spacing increases the risk of sampling or discretisation error. Thus it is necessary to compromise to find an optimal Δ which, when used in

(A.4), may still yield an unacceptable solution. It should be noted that in this example we have considered error in just one data point x_k and that this error is propagated and compounded into the estimates for r_{k+1}, ..., r_N. If error is introduced on all data points x_k (k=1, ..., N) then the situation becomes considerably worse and in order to find an acceptable solution it is necessary to play around with the values of u_k and x_k. This amounts to an intuitive smoothing of the data.

REFERENCES

AKAIKE, H. (1974). A new look at statistical model identification. IEEE Trans. Autom. Control Vol. AC-19, No.6, Dec., pp.716-723.

ANDERSON, G.E., HAYES, S.R., HILLYER, M.J., KILLUS, J.P. and MUNDKUR, P.V. (1977). Air Quality in the Denver Metropolitan Region 1974-2000. US Environmental Protection Agency, Report No. EPA-908/1-77-002, Systems Applications, Inc., San Rafael, California. (NTIS PB-271 894)

ASTROM, K.J. and BOHLIN, T. (1966). Numerical identification of linear dynamic systems from normal operating records. In: Theory of Self Adaptive Systems. P.H. Hammond (ed.), Plenum Press, New York, pp.96-111.

ASTROM, K.J. and EYKHOFF, P. (1971). System identification - a survey. Automatica Vol.7, pp.123-162.

BACH, W. (1971). Variation of solar attenuation with height over an urbanized area. J. Air Pollut. Control Assoc. Vol.21, No.10, Oct., pp.621-628.

BACH, W. (1976). Global air pollution and climatic change. Reviews Geophys. Space Phys. Vol.14, No.3, Aug., pp.429-474.

BAILEY, M. (1976). Carbon monoxide measurements in a Sydney street. Proceedings of Symposium on Air Pollution Diffusion Modeling, 18-20 Aug., Australian Environment Council, Canberra, Australia.

BANKOFF, S.G. and HANZEVACK, E.L. (1975). The adaptive-filtering transport model for prediction and control of pollutant concentration in an urban airshed. Atmos. Environ. Vol.9, pp.793-808.

BARAD, M.L. ed. (1958). Project prairie grass, a field program in diffusion. Vol.1. Geophysical Research Papers No.59, Geophysics Research Directorate, US Air Force Cambridge Research Center, Bedford, Massachusetts.

BARRETT-LENNARD, I. (1978). Some Estimation Examples Using a Method of Instrumental Variables. Thesis (M. Eng. Sc.) Univ. of Western Australia.

BART, P. (1965). Saturday Review (22 May) p.39.

BATCHELOR, G.K. (1960). The Theory of Homogeneous Turbulence. Student's edition, Cambridge University Press, Cambridge, England.

BENNETT, R.J. (1979). Spatial Time Series - Analysis, Forecasting, Control. Pion Limited, London.

BORIS, J.P. and BOOK, D.L. (1973). Flux-corrected transport. I. SHASTA, a fluid transport algorithm that works. J. Computational Phys. Vol.11, pp.38-69.

BORNSTEIN, R.D. and JOHNSON, D.S. (1977). Urban-rural wind velocity differences. Atmos. Environ. Vol.11, pp.597-604.

BOX, G.E.P. and JENKINS, G.M. (1970). Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

BROWNLEE, K.A. (1965). Statistical Theory and Methodology in Science and Engineering. John Wiley, New York.

BRYAN, R.J. (1976). Ambient air quality surveillance. In: Air Pollution, Third Edition, Vol.III. Measuring, Monitoring and Surveillance of Air Pollution, A.C. Stern (ed.), Academic Press, New York, pp.343-392.

BRYSON, A.E. Jr and HO, Y-C. (1975). Applied Optimal Control. Hemisphere Publishing Corp., Washington, D.C.

BUDIANSKY, S.P. (1980). ES & T outlook - dispersion modeling. Environ. Sci. Technol. Vol.14, No.4, April, pp.370-374.

BUFFHAM, B.A. and GIBILARO, L.G. (1970). A unified time delay model for dispersion in flowing media. The Chemical Engineering Journal Vol.1, pp.31-36.

BULLIN, J.A., GREEN, N.J. and POLASEK, J.C. (1980). Determination of vehicle emission rates from roadways by mass balance techniques. Environ. Sci. Technol. Vol.14, No.6, June, pp.700-705.

BULLIN, J.A. and POLASEK, J.C. (1978). Traffic measurement for roadway pollution studies by radar methods. J. Air Pollut. Control Assoc. Vol.28, No.2, Feb., pp.158-160.

CALDER, K.L. (1965). On the equation of atmospheric diffusion. Quarterly J. Royal Meteorological Society Vol.91, pp.514-517.

CALDER, K.L. (1973). On estimating air pollution concentrations from a highway in an oblique wind. Atmos. Environ. Vol.7, pp.863-868.

CAMPBELL, W.A. and HEATH, M.S. Jr. (1977). Air pollution legislation and regulations. In: Air Pollution, Third Edition, Vol.V. Air Quality Management, A.C. Stern (ed.), Academic Press, New York, pp.355-379.

CAPRIHAN, A. and NETO, A.G. (1977). Optimum modeling of vascular beds using indicator-dilution measurements. IEEE Trans. Biomedical Engineering Vol.BME-24, No.3, May, pp.219-226.

CHOCK, D.P., TERRELL, T.R. and LEVITT, S.B. (1975). Time-series analysis of Riverside, California air quality data. Atmos. Environ. Vol.9, pp.978-989.

CHRISTENSEN, O. and PRAHM, L.P. (1976). A pseudospectral model for dispersion of atmospheric pollutants. J. Appl. Meteorology Vol.15, Dec., pp.1284-1294.

CLEVELAND, W.S., KLEINER, B., McRAE, J.E. and WARNER, J.L. (1976). The analysis of ground-level ozone data from New Jersey, New York, Connecticut and Massachusetts: Transport from the New York City metropolitan area. Proceedings of the 4th Symposium on Statistics and the Environment, March 3-5, Washington, D.C., American Statistical Association, pp.70-90.

DABBERDT, W.F., LUDWIG, F.L. and JOHNSON, W.B. (1973). Validation and applications of an urban diffusion model for vehicular pollutants. Atmos. Environ. Vol.7, pp.603-618.

DALY, N.J. and STEELE, L.P. (1975). Air Quality in Canberra. Report to Department of the Capital Territory, Canberra, Australia.

DANARD, M.B. (1972). Numerical modeling of carbon monoxide concentrations near highways. J. Appl. Meteorology Vol.11, Sept., pp.947-957.

DAW, F.A. and O'LOUGHLIN, K.J. (1972). Canberra, A.C.T. - dispersion climatology and meteorological effects on the distribution of air pollution. Proceedings of the International Clean Air Conference, Melbourne, Australia, May 15-18, Clean Air Society of Australia and New Zealand, pp.39-46.

DE NEVERS, N. and MORRIS, J.R. (1975). Rollback modeling: basic and modified. J. Air Pollut. Control Assoc. Vol.25, No.9, Sep., pp.943-947.

DE NEVERS, N.H., NELIGAN, R.E. and SLATER, H.H. (1977). Air quality management, pollution control strategies, modeling and evaluation. In: Air Pollution, Third Edition, Vol.V Air Quality Management, A.C. Stern (ed.), Academic Press, New York, pp.3-40.

DESALU, A.A., GOULD, L.A. and SCHWEPPE, F.C. (1974). Dynamic estimation of air pollution. IEEE Trans. Autom. Control Vol. AC-19, No.6, Dec., pp.904-910.

DONALDSON, C. du P. and HILST, G.R. (1972). Effect of inhomogeneous mixing on atmospheric photochemical reactions. Environ. Sci. Technol. Vol.6, No.9, Sep., pp.812-816.

DOTTO, L. and SCHIFF, H. (1978). Ozone War. Doubleday, New York.

EGAN, B.A. and MAHONEY, J.R. (1972). Numerical modeling of advection and diffusion of urban area source pollutants. J. Appl. Meteorology Vol.11, March, pp.312-322.

ESCHENROEDER, A.Q. (1975). An Assessment of Models for Predicting Air Quality. ERT DOC. ERTW-75-03, Environmental Research and Technology, Inc., Santa Barbara, California.

ESCHENROEDER, A.Q. and MARTINEZ, J.R. (1969). Mathematical Modeling of Photochemical Smog. IMR 1210, General Research Corp., Santa Barbara, California.

ESCHENROEDER, A.Q., MARTINEZ, J.R. and NORDSIECK, R.A. (1972). Evaluation of a Diffusion Model for Photochemical Smog Simulation. US Environmental Protection Agency, Report No. EPA-R4-73-012, General Research Corp., Santa Barbara, California. (NTIS PB-220 242)

EYKHOFF, P. (1974). System Identification. Wiley-Interscience, London.

FAITH, W.L. (1959). Air Pollution Control. John Wiley and Sons, New York.

FAN, L.T. and HORIE, Y. (1971). Review of atmospheric disperson and urban air pollution models. CRC Critical Reviews in Environ. Control Vol.2, Oct., pp.431-457.

FINIGAN, B.M. (1978). Generation of an asymptotically optimal instrumental variable estimation process using a time-varying auxiliary linear system model. Proceedings of the 4th IFAC Symposium - Identification and System Parameter Estimation, Tbilisi, USSR, 21-27 Sept. 1976, edited by Rajbman, North-Holland Publishing Co., pp.1949-1958.

FINZI, G., ZANETTI, P., FRONZA, G. and RINALDI, S. (1979). Real time prediction of SO_2 concentration in the Venetian Lagoon area. Atmos. Environ. Vol.13, pp.1249-1255.

FORRESTER, J.W. (1971). World Dynamics. Wright-Allen, Cambridge, Massachusetts.

FREEMAN, T.G. (1981). Introduction to the Use of CAPTAIN for Time - Series Analysis. Computing Note 43, CSIRO Division of Computing Research, Canberra, Australia.

FRENKIEL, F.N. (1956). Atmospheric pollution in growing communities. In: Annual Report of the Smithsonian Institution 1956. US Govt. Printing Office, Washington D.C. pp.269-299.

FRONZA, G., SPIRITO, A. and TONIELLI, A. (1979). Real-time forecast of air pollution episodes in the Venetian region. Part 2: The Kalman predictor. Appl. Math. Modeling Vol.3, Dec., pp.409-415.

GAMEL, J., ROUSSEAU, W.F., KATHOLI, C.R. and MESEL, E. (1973). Pitfalls in digital computation of the impulse response of vascular beds from indicator-dilution curves. Circulation Research Vol. 32, April, pp.516-523.

GELB, A. (1974). Applied Optimal Estimation.(ed.) The M.I.T. Press, Cambridge, Massachusetts.

GIFFORD, F.A. (1968). An outline of theories of diffusion in the lower layers of the atmosphere. In: Meteorology and Atomic Energy 1968, D.H. Slade (ed.), US Atomic Energy Commission, Report No. TID-24190, Chap.3, pp.65-116.

GIFFORD, F.A. (1974). Further comparison of air pollution models. Proceedings Fifth Meeting of the Expert Panel on Air Pollution Modeling. Roskilde, Denmark, June 4-6, NATO Committee on the Challenges of Modern Society, Brussels, Belgium, pp.38.1-38.9. (NTIS PB-240 582)

GIFFORD, F.A. Jr. (1976). Turbulent diffusion-typing schemes: a review. Nuclear Safety Vol.17, No.1, Jan.-Feb., pp.68-86.

GIFFORD, F.A. and HANNA, S.R. (1973). Technical note: modeling urban air pollution. Atmos. Environ. Vol.7, pp.131-136.

GOLDSTEIN, I.F. and LANDOVITZ, L. (1977a). Analysis of air pollution patterns in New York City - I. Can one station represent the large metropolitan area? Atmos. Environ. Vol.11, pp.47-52.

GOLDSTEIN, I.F. and LANDOVITZ, L. (1977b). Analysis of air pollution patterns in New York City - II. Can one aerometric station represent the area surrounding it? Atmos. Environ. Vol.11, pp.53-57.

GOODWIN, G.C. and PAYNE, R.L. (1977). System Identification: Experiment Design and Data Analysis. Academic Press, New York.

GRAYBILL, F.A. (1961). An Introduction to Linear Statistical Models Volume 1. McGraw-Hill, New York.

GUICHERIT, R. (1976). The occurrence of photochemical oxidants and their precursors in Western Europe. Proceedings of Symposium on Occurrence and Control of Photochemical Pollution (Smog '76), Clean Air Society of Australia and New Zealand, February, Sydney, Australia, paper III.

HANNA, S.R. (1971). A simple method for calculating dispersion from urban area sources. J. Air Pollut. Control Assoc. Vol.21, No.12, Dec., pp.774-777.

HANNA, S.R. (1973a). Application of a simple model of photochemical smog. Proceedings Third International Clean Air Congress, Dusseldorf, Germany, pp.B72-B74.

HANNA, S.R. (1973b). Description of ATDL computer model for dispersion from multiple sources. In: Industrial Air Pollution Control. K. Noll and J. Duncan (eds.), Ann Arbor Science Publishers, Ann Arbor, Michigan, pp.23-32.

HANNA, S.R. (1975). Urban diffusion problems. In: Lectures on Air Pollution and Environmental Impact Analyses, sponsored by Amer. Meteor. Soc., 29 Sep.-3 Oct., Boston, Massachusetts, pp.209-227.

HANNA, S.R. (1977). Modeling smog along the Los Angeles-Palm Springs trajectory. In: Advances in Environ. Sci and Technol. Vol.8, I.H. Suffet (ed.), John Wiley, New York, pp.269-295.

HANNA, S.R. (1978). Diurnal variation of the stabililty factor in the simple ATDL urban dispersion model. J. Air Pollut. Control Assoc. Vol.28, No.2, Feb., pp.147-150.

HORNBERGER, G.M. and SPEAR, R.C. (1980). Eutrophication in Peel Inlet - I. The problem-defining behaviour and a mathematical model for the phosphorus scenario. Water Research Vol. 14, pp. 29-42.

HUMPHRIES, R.B., YOUNG, P.C. and BEER, T. (1981). Systems Analysis of an Estuary. The CRES Contribution to the Peel-Harvey Estuarine System Study. Report No. AS/R43, Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia.

JAKEMAN, A.J. and ANDERSSEN, R.S. (1975). Abel type integral equations in stereology. I. General discussion. J. Microscopy Vol. 105, Part 2, Nov., pp. 121-133

JAKEMAN, A.J. and YOUNG, P.C. (1979). Refined instrumental variable methods of time series analysis. Part II: multivariable systems. Int.J. Control Vol.29, pp.621-644.

JAKEMAN, A.J. and YOUNG, P.C. (1980a). Systems identification and estimation for convolution integral equations. In: The Application and Numerical Solution of Integral Equations. R.S. Anderssen et al. (eds.) Sijthoff and Noordhoff, pp.235-255.

JAKEMAN, A.J. and YOUNG, P.C. (1980b). Towards optimal modeling of translocation data from tracer studies. Proceedings Fourth Biennial Conference, Simulation Society of Australia, Brisbane, pp.248-253.

JAZWINSKI, A.H. (1970). Stochastic Processes and Filtering Theory. Academic Press, New York.

JOHNSON, G.T. (1974). Mathematical Modeling of the Dispersion of Air Pollutants over Cities. Report prepared for Australian Environment Council, Canberra, Australia.

JOHNSON, G.T. (1979a). Mathematical modeling of pollution dispersion. In: Air Quality Technology - Short Course. 16-18 May, Sydney, N.S.W., Dept. of Science and the Environment, pp.27-46.

JOHNSON, G.T. (1979b). Emissions Data Preparation Handbook. Technical Report No. 79-003 (Computing), School of Mathematics and Physics, Macquarie University, North Ryde, N.S.W.

JOHNSON, G.T. (1979c). Meteorological Data Preparation Handbook. Technical Report No. 79-004 (Computing), School of Mathematics and Physics, Macquarie University, North Ryde, N.S.W.

JOHNSON, G.T. (1979d). Atmospheric Pollution Simulation Handbook. Technical Report No. 79-005 (Computing), School of Mathematics and Physics, Macquarie University, North Ryde, N.S.W.

JOHNSON, G.T. (1980a). Sydney Air Quality Modeling Project - First Report. Technical Report No. 80-004 (Computing), School of Mathematics and Physics, Macquarie University, North Ryde, N.S.W.

JOHNSON, G.T. (1980b). Mathematical Modeling of the Dispersion of Air Pollutants over Urban Areas. Thesis (Ph.D.), Macquarie University, N.S.W.

JOHNSON, G.T. (1980c). User's Guide to the Macquarie Urban Air Quality Model. Technical Report No. 80-003 (Computing), School of Mathematics and Physics, Macquarie University, North Ryde, N.S.W.

JOHNSON, W.B. (1972). Validation of air quality simulation models. Proceedings Third Meeting of the Expert Panel on Air Pollution Modeling, Paris, France, Oct. 2-3, NATO Committee on the Challenges of Modern Society, Brussels, Belgium, pp.VI.1 - VI.9. (NTIS PB-240 574)

JOHNSON, W.B., LUDWIG, F.L., DABBERDT, W.F. and ALLEN, R.J. (1973). An urban diffusion simulation model for carbon monoxide. J. Air Pollut. Control Assoc. Vol.23, No.6, June, pp.490-498.

JOHNSON, W.B., SKLAREW, R.C. and TURNER, D.B. (1976). Urban air quality simulation modeling. In: Air Pollution, Third Edition, Vol.I. Air Pollutants, Their Transformation and Transport. A.C. Stern (ed.), Academic Press, New York, pp. 503-562.

JOHNSTON, J. (1972). Econometric Methods. Second edition. McGraw-Hill Kogakusha, Tokyo.

KALDOR, J.M. (1978). The Estimation of Parametric Change in Time Series Models. Thesis (M.A.) Australian National University.

KALMA, J.D., ANDERSON, M.C. and BYRNE, G.F. (1974). Surface Winds and Wind Exposure in Gungahlin New Town, A.C.T. Technical Memorandum 74/3, CSIRO Division of Land Use Research, Canberra, Australia.

KALMAN, R.E. (1960). A new approach to linear filtering and prediction problems. J. Basic Engineering (ASME) Vol.82D, March, pp.35-45.

KALMAN, R.E. and BUCY, R.S. (1961). New results in linear filtering and prediction theory. J. Basic Engineering (ASME) Vol.83D, March, pp.95-108.

KEELING, C.D. (1978). The influence of Mauna Loa observatory on the development of atmospheric CO_2 research. In: Mauna Loa Observatory, a 20th Anniversary Report. NOAA Special Report, J. Miller (ed.), US Dept. Commerce, Environment Research Laboratories, Boulder, Colorado, pp.36-54.

KENDALL, M.G. and STUART, A. (1961). The Advanced Theory of Statistics: Volume 2: Inference and Relationship. Charles Griffin and Co., London.

KOPEC, R.J. (1970). Further observations of the urban heat island in a small city. Bull. Amer. Meteor. Soc. Vol.51, No.7, July, pp.602-606.

KORN, G.A. and KORN, T.M. (1964). Electronic Analog and Hybrid Computers. McGraw-Hill, New York.

LAMB, R.G. (1971). Numerical Modeling or Urban Air Pollution. Thesis (Ph.D.) Univ. of California, Los Angeles.

LAMB, R.G. (1973). Note on the application of K-theory to diffusion problems involving non-linear chemical reactions. Atmos. Environ. Vol.7, pp.257-263.

LAMB, R.G. and SEINFELD, J.H. (1973). Mathematical modeling of urban air pollution. General theory. Environ. Sci. Technol. Vol.7, No.3, March, pp.253-261.

LANGE, R. (1978). ADPIC - a three dimensional particle-in-cell model for the dispersal of atmospheric pollutants and its comparison to regional tracer studies. J. Appl. Meteorology Vol. 17, March, pp.320-329.

LARSEN, R.I., BENSON, F.B. and JUTZE, G.A. (1965). Improving the dynamic response of continuous air pollutant measurements with a computer. J. Air Pollut. Control Assoc. Vol.15, No.1, Jan., pp.19-22.

LAVE, L.B. and SESKIN, E.P. (1976). Does air pollution cause mortality? Proceedings of the 4th Symposium on Statistics and the Environment, March 3-5, Washington, D.C., American Statistical Association, pp.25-35.

LEIGHTON, P.A. (1961). Photochemistry of Air Pollution. Academic Press, New York.

LJUNG, L. (1979). Convergence of recursive estimators. Proceedings of 5th IFAC Symposium on Identification and System Parameter Estimation, Darmstadt, Federal Republic of Germany, 24-28 Sept. 1979, Vol.1, R. Iserman (ed.), Pergamon Press, pp.131-144.

LUDWIG, F.L., JOHNSON, W.B., MOON, A.E. and MANCUSO, R.L. (1970). A Practical Multipurpose Urban Diffusion Model for Carbon Monoxide. National Air pollution Control Administration, Report No. APTD-0604, Stanford Research Institute, Menlo Park, California. (NTIS PB -196 003)

LUDWIG, F.L. and KEALOHA, J.H.S. (1975). Selecting Sites for Carbon Monoxide Monitoring. US Environmental Protection Agency, Report No. EPA-450/3-75-077, Office of Air and Waste Management, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina. (NTIS PB-249 700)

LUDWIG, F.L. and SHELAR, E. (1978). Site Selection for the Monitoring of Photochemical Air Pollutants. US Environmental Protection Agency, Report No. EPA-450/3-78-013, Office of Air and Waste Management, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina. (NTIS PB-283 404)

MAGE, D.T. and NOGHREY, J. (1972). True atmospheric pollutant levels by use of a transfer function for an analyser system. J. Air Pollut. Control Assoc. Vol.22, No.2, Feb., pp.115-118.

MAGILL, P.L., HOLDEN, F.R. and ACKLEY, C. (1956). Air Pollution Handbook. (eds.) McGraw-Hill, New York.

MALDONADO, C. and BULLIN, J.A. (1977). Modeling carbon monoxide dispersion from roadways. Environ. Sci. Technol. Vol.11, No.12, Nov., pp.1071-1076.

MEADOWS, D.H. MEADOWS, D.L., RANDERS, J. and BEHRENS, W.W. (1972). The Limits to Growth. Potomac Associates/Universe Books, New York.

MEETHAM, A.R. (1964). Atmospheric Pollution, its Origins and Preventions. 3rd revised edition, Macmillan, New York.

MERZ, P.H., PAINTER, L.J. and RYASON, P.R. (1972). Aerometric data analysis - time series analysis and forecast and an atmospheric smog diagram. Atmos. Environ. Vol.6, pp.319-342.

MOLENKAMP, C.R. (1968). Accuracy of finite-difference methods applied to the advection equation. J. Appl. Meteorology Vol.7, April, pp.160-167.

MOLINA, M.J. and ROWLAND, F.S. (1974). Stratospheric sink for chlorofluoromethanes: chlorine atom-catalysed destruction of ozone. Nature Vol.249, June 28, pp.810-812.

MONKMAN, J.L. (1976). Evidence of electronic malpractice in air pollution instrumentation. Proceedings of 12th International Colloquium on Atmospheric Pollution, Paris, France, May 5-7, Elsevier, Amsterdam, pp.297-308.

MOORE, R. and WHITEHEAD, P. (1975). The CAPTAIN Package - A Program Manual. Water Research Centre, Medmenham, England.

MORGAN, M.G. and MORRIS, S.C. (1977). Needed: a national R & D effort to develop individual air pollution monitor instrumentation. J. Air Pollut. Control Assoc. Vol.27, No.7, July, pp.670-673.

McCLENNY, W.A. and CHANEY, L.W. (1978). Pollutant variability in the regional air pollution study. J. Air Pollut. Control Assoc. Vol.28, No.7, July, pp.693-696.

McCOLLISTER, G.M. and WILSON, K.R. (1975). Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants. Atmos. Environ. Vol.9, pp.417-423.

McELROY, J.L. (1969). A comparative study of urban and rural dispersion. J. Appl. Meteorology Vol.8, Feb., pp.19-31.

NASLIN, P. (1965). The Dynamics of Linear and Non-Linear Systems. Blackie and Son Limited, London.

NEWILL, V.A. (1977). Air quality standards. In: Air Pollution, Third Edition, Vol. V. Air Quality Management, A.C. Stern (ed.), Academic Press, New York, pp.445-504.

NIEBOER, H., CARTER, W.P., LLOYD, A.C. and PITTS, J.N. Jr. (1976). The effect of latitude on the potential for formation of photochemical smog. Atmos. Environ. Vol.10, pp.731-734.

NORTON, J.P. (1975). Optimal smoothing in the identification of linear time-varying systems. Proc.IEE Vol.122, No.6, June, pp.663-668.

ODÉN, S. (1976). The acidity problem - an outline of concepts. Water, Air and Soil Pollut. Vol.6, Sep. - Oct. - Nov., pp.137-166.

OTT, W.R. (1972). An Urban Survey Technique for Measuring the Spatial Variation of Carbon Monoxide Concentrations in Cities. Thesis (Ph.D.) Stanford University.

OTT, W.R. and MAGE, D.T. (1978). Interpreting urban carbon monoxide concentrations by means of a computerised blood COHb model. J. Air Pollut. Control Assoc. Vol.28, No.9, Sep., pp.911-916.

PANKOV, A.R. and SKURIDIN, A.M. (1979). Recursive estimation of linear model parameters from several groups of measurements. Automation and Remote Control Vol.40, No.9, Part 1, Sep., pp.1323-1331.

PASQUILL, F. (1974). Atmospheric Diffusion. 2nd ed., Ellis Horwood Limited, Chichester, Sussex.

PEARMAN, G.I. (1980). Carbon Dioxide and Climate: Australian Research. (ed.) Australian Academy of Science, Canberra, Australia.

PERKINS, R.J. (1975). The environment from a systems viewpoint. Intern. J. Environ. Studies Vol.8, pp.59-63.

PERSSON, G. (1977). Organisation and operation of national air pollution control programs. In: Air Pollution, Third Edition, Vol.V. Air Quality Management, A.C. Stern (ed.), Academic Press, New York, pp.381-413.

PHADKE, M.S., BOX, G.E.P. and TIAO, G.C. (1976). Empirical-mechanistic modeling of air pollution. Proceedings of the 4th Symposium on Statistics and the Environment, March 3-5, Washington, D.C., American Statistical Association, pp.91-101.

PHADKE, M.S. and WU, S.M. (1974). Modeling of continuous stochastic processes from discrete observations with application to sunspots data. J. Amer. Statist. Assoc. Vol. 69, No. 346, June, pp. 325-329.

PIERCE, D.A. (1972). Least squares estimation in dynamic-disturbance time series models. Biometrika Vol.59, No.1, pp.73-78.

PITTS, J.N. Jr. (1976) California's experience in the control of photochemical smog. Proceedings of Symposium on Occurrence and Control of Photochemical Pollution (Smog '76), Clean Air Society of Australia and New Zealand, February, Sydney, Australia, paper I.

POOLER, F. Jr. (1974). Network requirements for the St.Louis regional air pollution study. J. Air Pollut. Control Assoc. Vol.24, No.3, March, pp.228-231.

RASHIDI, M. and MASSOUDI, M.S. (1980). A study of the relationship of street level carbon monoxide concentrations to traffic parameters. Atmos. Environ. Vol.14, pp.27-32.

REMSBERG, E.E., BUGLIA, J.J. and WOODBURY, G.E. (1979). The nocturnal inversion and its effect on the dispersion of carbon monoxide at ground level in Hampton, Virginia. Atmos. Environ. Vol.13, pp.443-447.

REYNOLDS, S.D., AMES, J., HECHT, T.A., MEYER, J.P., WHITNEY, D.C. and YOCKE, M.A. (1976). Continued Research in Mesoscale Air Pollution Simulation Modeling. Vol.II. Refinements in the Treatment of Chemistry, Meteorology and Numerical Integration Procedures. US Environmental Protection Agency, Report No. EPA-600/4-76-016b, Systems Applications, Inc., San Rafael, California. (NTIS PB-257 527)

REYNOLDS, S.D., LIU, M-K., HECHT, T.A., ROTH, P.M. and SEINFELD, J.H. (1974). Mathematical modeling of photochemical air pollution. III. Evaluation of the model. Atmos. Environ. Vol.8, pp. 563-596.

REYNOLDS, S.D., ROTH, P.M. and SEINFELD, J.H. (1973). Mathematical modeling of photochemical air pollution. I. Formulation of the model. Atmos. Environ. Vol.7, pp.1033-1061.

RHODES, I.B. (1971). A tutorial introduction to estimation and filtering. IEEE Trans. Autom. Control Vol. AC-16, Dec., pp.688-706.

ROBERTS, J.J., CROKE, E.S. and KENNEDY, A.S. (1970). An urban atmospheric dispersion model. Proceedings of Symposium on Multiple-Source Urban Diffusion Models. A.C. Stern (ed.), Air Pollut. Control Office Publ. No.AP-86, US Environmental Protection Agency, Research Triangle Park, North Carolina, pp.6.1-6.72.

ROSE, A.H. Jr., SMITH, R., McMICHAEL, W.F. and KRUSE, R.E. (1965). Comparison of auto exhaust emissions in two major cities. J. Air Pollut. Control Assoc. Vol.15, No.8, Aug., pp.362-366.

ROTH, P.M., ROBERTS, P.J.W., LIU, M-K., REYNOLDS, S.D. and SEINFELD, J.H. (1974). Mathematical modeling of photochemical air pollution. II. A model and inventory of pollutant emissions. Atmos.Environ. Vol.8, pp.97-130.

ROWE, I.H. (1970). A bootstrap method for the statistical estimation of model parameters. Int.J. Control Vol.12, No.5, pp.721-738.

ROWLAND, F.S. and MOLINA, M.J. (1975). Chlorofluoromethanes in the environment. Reviews Geophys. Space Phys. Vol.13, No.1, Feb., pp.1-35.

RUNCA, E., MELLI, P. and SPIRITO, A. (1979). Real-time forecast of air pollution episodes in the Venetian region. Part 1: the advection-diffusion model. Appl. Math. Modelling Vol.3, Dec., pp.402-408.

RUNCA, E., MELLI, P. and ZANETTI, P. (1976). Computation of long-term average SO_2 concentration in the Venetian area. Appl. Math. Modelling Vol.1, June, pp.9-15.

SALTZMAN, B.E. (1970). Significance of sampling time in air monitoring. J. Air Pollut. Control Assoc. Vol.20, No.10, Oct., pp.660-665.

SAWARAGI, Y., SOEDA, T., TAMURA, H., YOSHIMURA, T., OHE, S., CHUJO, Y. and ISHIHARA, H. (1979). Statistical prediction of air pollution levels at a short range by non-physical models. In: Planning, Design and Control of Environmental Systems. Special Project Research on Detection and Control of Environmental Pollution Vol.4, Ministry of Education, Japan, pp.100-107.

SCHNELLE, K.B. Jr and NEELEY, R.D. (1972). Transient and frequency response of air quality monitors. J. Air Pollut. Control Assoc. Vol.22, No.7, July, pp.551-555.

SCORER, R.S. (1976). Inherent errors and dangers in environmental monitoring and modeling. Proceedings of Symposium on Systems and Models in Air and Water Pollution. Institute of Measurement and Control, London, 22-24 Sept., pp.19.1-19.13.

SEINFELD, J.H. (1975). Air Pollution - Physical and Chemical Fundamentals. McGraw-Hill, New York.

SEINFELD, J.H. and WILSON, K.R. (1977). The issue of air quality simulation model utility. International Conference on Oxidants, 1976- Analysis of Evidence and Viewpoints, Part VI. US Environmental Protection Agency, Report No. EPA-600/3-77-118, Environmental Sciences Research Lab., Office of Research and Development, Research Triangle Park, North Carolina. (NTIS PB-277 463)

SENECA, J.J. and TAUSSIG, M.K. (1974). Environmental Economics. Prentice-Hall, Englewood Cliffs, New Jersey.

SHELLSWELL, S.H. and YOUNG, P.C. (1973). CAPTAIN - a computer aided program for time-series analysis and the identification of noisy systems. Proceedings IEE Conference on Computer Aided Control System Design, IEE, London.

SIMON, H.A. (1965). The architecture of complexity. General Systems: The Yearbook of the Society for General Systems Research. Vol.10, pp.63-76.

SKLAREW, R.C., FABRICK, A.J. and PRAGER, J.E. (1972). Mathematical modeling of photochemical smog using the PICK method. J. Air Pollut. Control Assoc. Vol.22, No.11, Nov., pp.865-869.

SMITH, F.B. and JEFFREY, G.H. (1972). The prediction of high concentrations of sulphur dioxide in London and Manchester air. Proceedings Third Meeting of the Expert Panel on Air Pollution Modeling, Paris, France, Oct. 2-3, NATO Committee on the Challenges of Modern Society, Brussels, Belgium, pp.XVIII.1 - XVIII.6. (NTIS PB-240 574)

SODERSTROM, T., LJUNG, L. and GUSTAVSSON, I. (1974). A Comparative Study of Recursive Identification Methods. Report No. 7427, Dept. of Automatic Control, Lund Institute of Technology, Lund, Sweden.

SORENSEN, H.W. (1970). Least squares estimation: from Gauss to Kalman. IEEE Spectrum Vol.7, July, pp.63-68.

SPEAR, R.C. (1978). Organophosphate Residue Poisoning among Agricultural Fieldworkers in the Western United States: Towards a Strategy for a Long-Term Solution. Report No. AS/R18, Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia.

SPEAR, R.C. and HORNBERGER, G.M. (1980). Eutrophication in Peel Inlet - II. Identification of critical uncertainties via generalised sensitivity analysis. Water Research Vol. 14, pp. 43-49.

SPEAR, R.C., LEE, Y., LEFFINGWELL, J.T. and JENKINS, D. (1978). Conversion of parathion to paraoxon in foliar residues: effects of dust level and ozone concentration. J. Agric. Food Chem. Vol. 26, No.2, pp.434-436.

SPOFFORD, W.O. Jr. (1975). Ecological modeling in a resource management framework: an introduction. Proceedings of Symposium on Ecological Modeling in a Resource Management Framework. C.S. Russell (ed.), Symposium sponsored by NOAA and Resources for the Future, Inc., Washington, D.C., pp.13-48.

STEELE, L.P. and JAKEMAN, A.J. (1980). Recursive time-series analysis for modeling atmospheric environments. Proceedings Fourth Biennial Conference, Simulation Society of Australia, Brisbane, pp.75-81.

STERN, A.C. (1970). Utilisation of air pollution models. Proceedings of Symposium on Multiple-Source Urban Diffusion Models. A.C. Stern (ed.) Air Pollut. Control Office Publ. No. AP-86, US Environmental Protection Agency, Research Triangle Park, North Carolina, pp.13.1-13.8.

STERN, A.C. (1976). Air Pollution, Third Edition (5 volumes). (ed.) Academic Press, New York.

TAKAHASHI, Y., RABINS, M.J. and AUSLANDER, D.M. (1972). Control and Dynamic Systems. Addison-Wesley, Reading, Massachusetts.

TE BRAKE, W.H. (1975). Air pollution and fuel crises in preindustrial London, 1250-1650. Technol. and Culture Vol.16, No.3, July, pp.337-359.

THISSEN, W. (1978). Investigtions into the World3 model: lessons for understanding complicated models. IEEE Trans. Systems, Man, and Cybernetics Vol. SMC-8, No.3, March, pp.183-193.

TIAO, G.C., BOX, G.E.P. and HAMMING, W.J. (1975). A statistical analysis of the Los Angeles ambient carbon monoxide data 1955-1972. J. Air Pollut. Control Assoc. Vol.25, No.11, Nov., pp.1129-1136.

TIAO, G.C. and HILLMER, S.C. (1978). Statistical models for ambient concentrations of carbon monoxide, lead and sulphate based on the LACS data. Environ. Sci. Technol. Vol.12, No.7, July, pp.820-828.

TIHONOV, A.N. (1963). Solution of incorrectly formulated problems and the regularisation method. Soviet Math. Dokl. Vol. 4, pp. 1035-1038.

WILLIAMSON, S.J. (1973). Fundamentals of Air Pollution. Addison-Wesley, Reading, Massachusetts.

WINER, A.M., PETERS, J.W., SMITH, J.P. and PITTS, J.N. Jr. (1974). Response of commercial chemiluminescent NO-NO_2 analysers to other nitrogen-containing compounds. Environ. Sci. Technol. Vol.8, No.13, Dec., pp.1118-1121.

WONG, K.Y. and POLAK, E. (1967). Identification of linear discrete time systems using the instrumental variable method. IEEE Trans. Autom. Control Vol. AC-12, No.6, Dec., pp.707-718.

WYZGA, R.E. (1973). Note on a method to estimate missing air pollution data. J. Air Pollut. Control Assoc. Vol.23, No.3, March, pp.207-208.

YANENKO, N.N. (1971). The Method of Fractional Steps. Springer-Verlag, Berlin.

YOUNG, P.C. (1965a). Process parameter estimation and self adaptive control. In: Theory of Self Adaptive Control Systems, P.H. Hammond, (ed.), Plenum Press, New York, pp.118-140.

YOUNG, P.C. (1965b). On a Weighted Steepest Descent Method of Parameter Estimation. Tech. Rep. No. PCY/Camb/1, Engineering Department, University of Cambridge.

YOUNG, P.C. (1968). Regression analysis and process parameter estimation: a cautionary message. Simulation Vol.10, No.3, March, pp.125-128.

YOUNG, P.C. (1969). The Differential Equation Error Method of Real Time Process Identification. Thesis (Ph.D.) University of Cambridge.

YOUNG, P.C. (1974). Recursive approaches to time series analysis. Bull. Inst. Maths and its Applic. Vol.10, pp.209-224.

YOUNG, P.C. (1976). Some observations on instrumental variable methods of time series analysis. Int. J. Control Vol.23, No.5, pp.593-612.

YOUNG, P.C. (1978). A general theory of modeling for badly defined systems. In: Modeling, Identification and Control in Environmental Systems. G.C. Vansteenkiste (ed.) IFIP, North-Holland Publishing Company, pp.103-135.

YOUNG, P.C. (1979). Self-adaptive Kalman filter. Electronics Letters Vol.15, No.12, June, pp.358-360.

YOUNG, P.C. (1980a). The Validity and Credibility of Models for Badly Defined Systems. Report No. AS/R40, Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia.

YOUNG, P.C. (1980b). Recursive Methods of Time Series Analysis: the Choice Between Alternative Models and Techniques. Report No. AS/R39, Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia.

YOUNG, P.C. (1981). An Introduction to Recursive Estimation. Springer-Verlag, Berlin.

YOUNG, P.C. and JAKEMAN, A.J. (1979a). An Inverse Problem: the Estimation of Input Variables in Stochastic Dynamic Systems. Report No. AS/R28, Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia.

YOUNG, P.C. and JAKEMAN, A.J. (1979b). The development of CAPTAIN: a computer aided program for time series analysis and the identification of noisy systems. Proceedings of IFAC Symposium Computer Aided Design of Control Systems, Zurich, Switzerland, 29-31 August, Pergamon Press, Oxford, pp.391-400.

YOUNG, P.C. and JAKEMAN, A.J. (1979c). Refined instrumental variable methods of recursive time series analysis. Part I: single input-single output systems. Int. J. Control Vol.29, pp.1-30.

YOUNG, P.C. and JAKEMAN, A.J. (1980). Refined instrumental variable methods of recursive time series analysis. Part III. Extensions. Int. J. Control Vol.31, No.4, pp.741-764.

YOUNG, P.C., JAKEMAN, A.J. and McMURTRIE, R.E. (1980). An instrumental variable method for model order identification. Automatica Vol.16, No.3, May, pp.281-294.

YOUNG, P.C., SHELLSWELL, S.H. and NEETHLING, C.G. (1971). A Recursive Approach to Time Series Analysis. Report No. CUED/B-Control/TR16, Control and Systems Group, Dept. of Engineering, University of Cambridge.

YOUNG, P.C. and WHITEHEAD, P.G. (1977). A recursive approach to time series analysis for multivariable systems. Int. J. Control Vol.25, No.3, pp.457-482.

ZIMMER, C.E. (1976). Air quality data handling and analysis. In: Air Pollution, Third Edition, Vol.III. Measuring, Monitoring and Surveillance of Air Pollution. A.C. Stern (ed.), Academic Press, New York, pp.453-484.

ZIMMERMAN, J.L. and THOMPSON, R.S. (1974). HIWAY: A Highway Air Pollution Model. National Environment Research Centre, Research Triangle Park, North Carolina.