

INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.
2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.
4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.
5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.

University
Microfilms
International

300 N. ZEEB ROAD, ANN ARBOR, MI 48106
18 BEDFORD ROW, LONDON WC1R 4EJ, ENGLAND

7911140

AHMED, MOHAMED SAMIR
ANALYSIS OF FREEWAY TRAFFIC TIME SERIES DATA
AND THEIR APPLICATION TO INCIDENT DETECTION.

THE UNIVERSITY OF OKLAHOMA, PH.D., 1979

University
Microfilms
International 300 N. ZEEB ROAD, ANN ARBOR, MI 48106

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ANALYSIS OF FREEWAY TRAFFIC TIME SERIES DATA AND
THEIR APPLICATION TO INCIDENT DETECTION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

MOHAMED SAMIR AHMED

Norman, Oklahoma

1979

ANALYSIS OF FREEWAY TRAFFIC TIME SERIES DATA AND
THEIR APPLICATION TO INCIDENT DETECTION

APPROVED BY

Allen R. Cook

Allen R. Cook, Chairman

Charles E. Barb, Jr.

Charles E. Barb, Jr.

Adel A. Aly

Adel A. Aly

Ed F. Crim

Ed F. Crim

DISSERTATION COMMITTEE

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to the members of my doctoral committee; Dr. Charles E. Barb, Dr. Adel A. Aly, Dr. Ed F. Crim, and, in particular, Dr. Allen R. Cook; for their encouragement, assistance, and suggestions during the preparation of this dissertation. My thanks are also extended to all who have helped me throughout this research effort.

This research was sponsored in part by the Office of Research Administration of the University of Oklahoma. The School of Civil Engineering and Environmental Science at O.U. is acknowledged for supplying the computer time.

Finally, I wish to thank my family for their encouragement and patience during the course of my study.

ABSTRACT

This study investigates the stochastic and dynamic nature of freeway traffic time series data using the analysis techniques described by Box and Jenkins. The objectives of this investigation are to develop an improved filtering and prediction model for use with computer-supervised freeway surveillance and control systems, and to propose computer algorithms for detecting accidents and other capacity-reducing incidents which are typical occurrences on urban freeways. The analysis was based on surveillance data recorded at the Los Angeles, Minneapolis, and Detroit freeway systems during the afternoon peak periods.

An autoregressive integrated moving average model of the form ARIMA (0,1,3) was found to be representative of 166 time series of traffic volume and occupancy data (more than 27,000 minutes of observations) with varying detector configurations and data aggregation time intervals. The moving average parameters of the model, however, vary from detector station to another and over time. In operational forecasting, parameter updating over time can be done occasionally, for example, at the beginning of peak and off-peak periods. With the increasing trend toward decentralized control and distributing the computational capabilities, parameter updating can be efficiently performed using microcomputers. In terms of mean absolute error and mean square error, the ARIMA (0,1,3) model was found to be superior to moving average, double exponential smoothing, and Trigg and Leach adaptive models.

Using the ARIMA (0,1,3) model, freeway capacity-reducing incidents were detected by the sudden and pulsed changes they generate in traffic stream time series data. Eight traffic features sensitive to incident situations were compiled during a sample of 50 on-freeway incidents (1694 minutes of observations). The form of the ARIMA (0,1,3) model was found representative of the time series of these features. An incident was detected if the observed feature value laid outside the probability limits constructed two standard deviations away from the corresponding point forecasts. This approach eliminated the need for threshold calibration and allowed for the uncertainty associated with the forecasts to be included in the detection decisions. The ARIMA station occupancy algorithm dominated the California algorithm, five of the TTI algorithms, and the exponential station occupancy algorithm in terms of higher detection rate, lower false-alarms, and faster response.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vii
LIST OF FIGURES	viii
 <u>Chapter</u>	
I. INTRODUCTION	1
1.1 Computer-Supervised Freeway Surveillance and Control Systems	1
1.2 A Capsule History, and Most Likely Projections . .	7
1.3 The Purpose and Utility of This Research	12
1.4 Overview of the Next Chapters	15
II. NATURE OF FREEWAY TRAFFIC STREAM MEASUREMENTS	17
2.1 Introduction	17
2.2 Traffic Stream Variables: Definitions and Measurements	17
2.3 Volume, Density, and Speed Interrelationship . . .	21
2.4 Microscopic Traffic Flow Models	26
2.5 Macroscopic Traffic Flow Models	28
2.6 Propagation of Disturbances Caused by Freeway Incidents	32
2.7 Summary	39
III. EXISTING FORECASTING TECHNIQUES APPLIED TO TRAFFIC SYSTEMS	40
3.1 Introduction	40
3.2 Basic Definitions and Concepts	41
3.3 A Typology of Forecasting Techniques	43
3.3.1 Deterministic Models	43
3.3.2 Functional and Structural Models	45
3.3.3 Ad hoc Models	46
3.3.4 Time Series Analysis Models	46
3.4 Ad hoc Forecasting Models	47
3.4.1 Moving Average Model	48
3.4.2 Exponential Smoothing Models	49
3.4.3 Exponential Smoothing With Adaptive Response	52
3.4.4 Summary of Ad hoc Models	53
3.5 Box and Jenkins ARIMA Models for Time Series Analysis	54
3.5.1 Models for Stationary Time Series	54
3.5.2 Stationarity and Invertibility Requirements	60

	Page
3.5.3 Models for Nonstationary Time Series	60
3.5.4 Model Fitting Procedure	61
3.5.5 Forecasting and Simulation	65
IV. MODELING FREEWAY TRAFFIC TIME SERIES DATA	69
4.1 Introduction	69
4.2 Summary of Previous Work on Traffic Time Series Analysis	70
4.3 Freeway Traffic Surveillance Data Base	73
4.4 ARIMA Model Development	79
4.5 ARIMA Model Utility in Forecasting and Simulation .	90
4.6 Comparative Evaluation of Forecasting Performance .	99
4.7 Summary	104
V. APPLICATION OF TIME SERIES ANALYSIS TO URBAN FREEWAY INCIDENT DETECTION	105
5.1 Introduction	105
5.2 Computer Incident Detection Algorithms	106
5.2.1 Measures of Effectiveness for Detection Algorithms	109
5.2.2 A Typology of Detection Algorithms	112
5.2.3 Review of the Structures of Current Detec- tion Algorithms	114
5.3 Incident Data Base	123
5.4 ARIMA Model Application to Incident Detection . . .	125
5.5 Summary and Implications for Microcomputers	135
VI. CONCLUSIONS AND SUGGESTED RESEARCH	139
6.1 Conclusions	139
6.2 Prospects for Future Research	144
REFERENCES	145

LIST OF TABLES

Table	Page
4.1 Data Sources and Types	75
4.2 Moving Average Parameters of Volume Series Aggregated Over Lanes	88
4.3 Moving Average Parameters of Occupancy Series Aggregated Over Lanes	89
4.4 Moving Average Parameters of 20- and 60-Second Lane Volume Series, Harbor Freeway, Los Angeles	91
4.5 Moving Average Parameters of 20- and 60-Second Lane Occupancy Series, Harbor Freeway, Los Angeles	92
5.1 Incident Detection Algorithms	115
5.2 Moving Average Parameters of Traffic Features	129
5.3 Variation in Moving Average Parameters Over Time at Seward Station	130
5.4 Mean Time Lags to Detection	136

LIST OF FIGURES

Figure	Page
1.1 Automated Freeway Surveillance and Control Concepts	3
1.2 General Levels of Freeway Surveillance and Control Technology Evolution	8
1.3 Role of Traffic Prediction in Freeway Surveillance and Control	14
2.1 Vehicle Trajectories and Measurement Methods	19
2.2 The Fundamental Diagram of Traffic Flow	23
2.3 Volume Versus Occupancy at the Chicago Detector Station . .	24
2.4 Volume Versus Occupancy at the Glendale Detector Station .	25
2.5 Vehicle Space-Time Trajectories, and Rectangular Region for Derivation of the Principle of Conservation of Vehicles	29
2.6 Flow Regimes and Shock Waves Generated by a Freeway Incident	33
2.7 Spatial Distributions of Speed, Density and Volume During an Incident	35
2.8 Time-Space Diagram of Freeway Flow Regimes and Shock Waves During and After an Incident	37
3.1 Freeway Traffic Volume and Occupancy Series	42
3.2 First Differences of Volume and Occupancy Series	44
3.3 Exponential Smoothing of Linear Trend	50
3.4 Representation of Correlated Noise as Output from Linear Filter	56
3.5 Representation of ARIMA Models as Three Filters	56
3.6 Theoretical Autocorrelation Functions of Basic Processes .	63
4.1 Schematic Representation of the 42-mile Freeway Loop, Los Angeles, California	76

	Page
4.2 Interstate-35 Freeway Corridor, Minneapolis, Minnesota . .	78
4.3 John C. Lodge Freeway Corridor, Detroit, Michigan	78
4.4 Sample Autocorrelations and Partial Autocorrelations, Volume Series, Minneapolis, I-35, Station 1	81
4.5 Sample Autocorrelations and Partial Autocorrelations, Occupancy Series, Minneapolis, I-35, Station 1	82
4.6 Residual Plot and Sample Autocorrelation Function, Volume Series, Minneapolis, I-35, Station 1	85
4.7 Residual Plot and Sample Autocorrelation Function, Occu- pancy Series, Minneapolis, I-35, Station 1	86
4.8 Sensitivity of ARIMA (0,1,3) Parameters Over Time	93
4.9 Observed Volume and Occupancy Series with Forecasts . . .	97
4.10 Simulated Traffic Volume and Occupancy Series from Fitted ARIMA (0,1,3) model	98
4.11 Ratio-to-Box-Jenkins for Mean Absolute Error	102
4.12 Ratio-to-Box-Jenkins for Mean Square Error	103
5.1 Volume and Occupancy Changes During a Capacity-Reducing Incident	108
5.2 Outcomes When Executing an Incident Detection Algorithm .	110
5.3 Performance of Some Traffic Features Under an Incident . .	132
5.4 Operating Characteristic Curves for Incident Detection Algorithms	134

ANALYSIS OF FREEWAY TRAFFIC TIME SERIES DATA AND
THEIR APPLICATION TO INCIDENT DETECTION

CHAPTER I

INTRODUCTION

Short-term prediction is at the heart of most computer-supervised traffic surveillance and control systems. These systems are planned, designed, and implemented to increase the operational efficiency of urban streets and freeways. In particular, this dissertation addresses the stochastic and dynamic nature of freeway traffic time series data, and the development of improved filtering and prediction models for traffic stream measurements. It also proposes and explores the development of incident detection algorithms based on time series analysis techniques.

1.1 Computer-Supervised Freeway Surveillance and Control Systems

A computer-supervised freeway surveillance and control system is defined by Weinberg [85] as:

A closed loop system ... that provides for surveillance of traffic operations, acquiring data on those operations which can be processed by a computational logic in real-time, testing observed conditions against a set of decision rules, selecting commands in light of the results of the test, activating appropriate controls and/or communicating with drivers to improve traffic movement when necessary, and then reassessing the traffic behavior to determine if further corrections are to be made.

In general, one can distinguish between two approaches of automated traffic surveillance and control. The first approach requires monitoring and controlling each individual vehicle, and is particularly appropriate for automated guideways or high speed train networks. The second approach is an aggregate type of control applied to an entire stream of vehicles rather than individual ones, and is suitable to many traffic systems like freeway corridors, tunnels, and bridges. In the latter approach traffic is typically monitored by means of vehicle detectors and/or television cameras located at strategic positions along the corridor.

The primary functions of a computer-supervised freeway surveillance and control system are to maintain an acceptable level of service for freeway operations, reduce the extent and duration of traffic congestion, minimize the adverse effects of accidents and other incidents, and lower the accident experience. From the overall community values standpoint, two categories of benefits can be recognized, direct benefits to the motorists and indirect benefits to the residents near the corridor. Savings in travel time, enhancing safety, reducing energy consumption, and minimizing discomfort of drivers are examples of the former category, while reducing air pollution and some of the effects of accidents are examples of the latter.

Figure 1.1 illustrates the concepts of automated freeway surveillance and control. In general, three functionally interrelated control and surveillance techniques can be characterized: ramp metering, variable message signs, and incident detection. When freeway demand exceeds capacity at some location as a result of traffic bottleneck or incident, the flow of additional vehicles onto the freeway upstream is

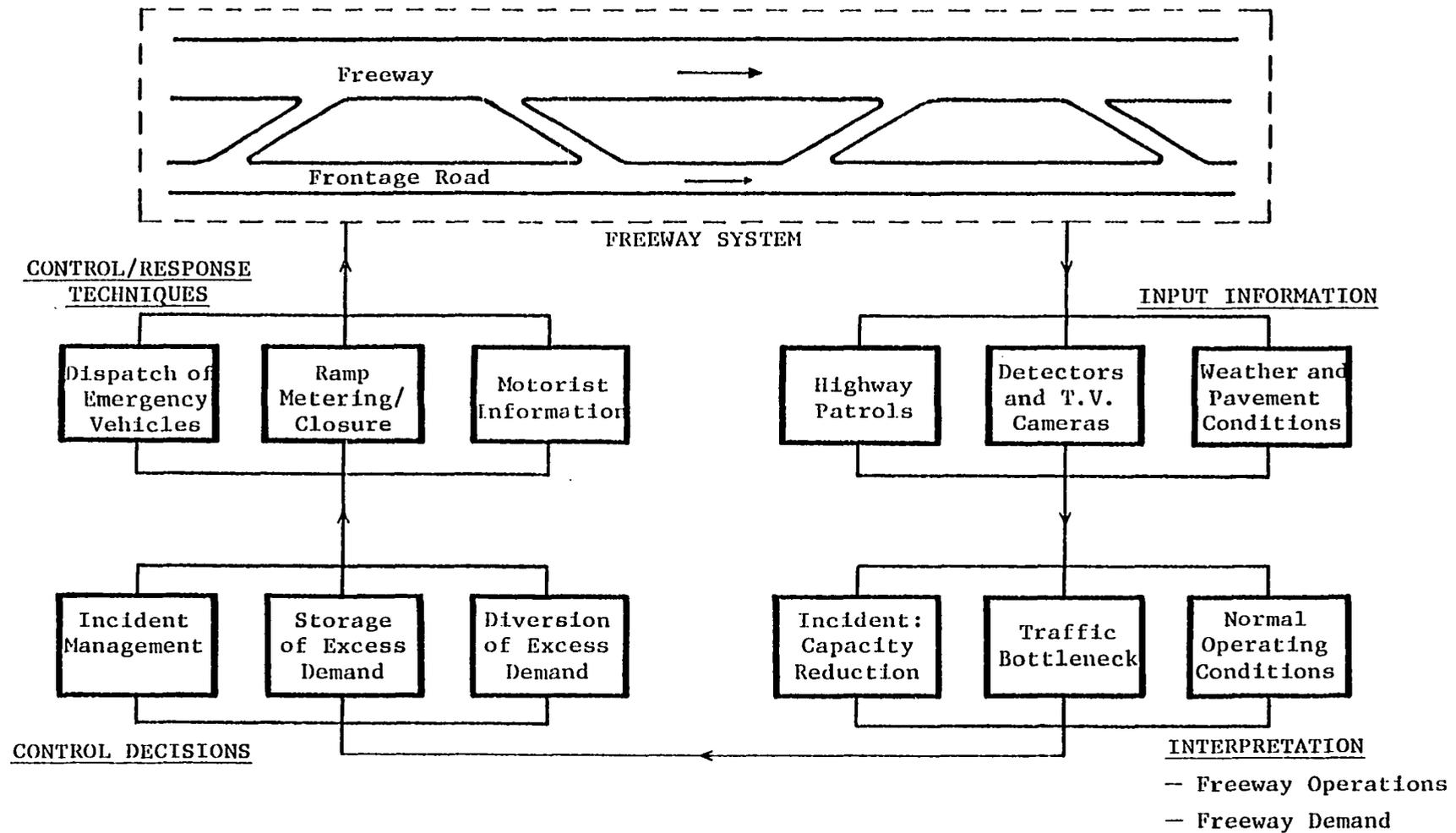


FIGURE 1.1

AUTOMATED FREEWAY SURVEILLANCE AND CONTROL CONCEPTS

restricted by means of traffic signals installed at entrance ramps. The queued vehicles stored at an entrance ramp are permitted to smoothly enter the freeway at a particular rate known as metering rate, which is determined by the difference between upstream demand and downstream capacity. In the extreme case, if upstream demand exceeds downstream capacity, the ramp is completely closed. To assure safe merging operations for drivers entering the freeway, another type of ramp control known as the moving-merge system is used. This system displays the right-lane gaps to the ramp drivers so that they can easily merge into the freeway even when the view of the right-lane is restricted. Demonstration projects in Houston, Chicago, Detroit, and Atlanta have shown that ramp control has the potential of significantly reducing the extent and duration of peak-period congestion and the number of rear-end collisions that occur on the freeway [55,77,83].

In addition to ramp metering control, variable message signs are used to provide the drivers with real-time information concerning traffic conditions on the freeway and the surrounding surface streets. The philosophy behind these messages is that traffic demand can be best distributed among the freeway corridor, frontage road, and surface street links if the drivers have a prior knowledge about the location, extent and duration of congestion. However, minimizing the overall travel time by redistributing traffic demand does not necessarily imply that the travel time of each individual motorist is also minimized. Different types of variable message signs exist on the freeways and alternate surface streets of many major cities in the United States, Europe, and Japan. In Detroit, for example, variable message signs were installed

on the entrance ramps to the John C. Lodge Freeway corridor. Should congestion exist on the freeway downstream of an entrance ramp, drivers who are going to enter the freeway using this ramp are advised to continue along an alternate route or to use other entrance ramps [82]. For motorists already on the freeway corridor, variable message signs are installed at strategic positions to advise drivers of traffic conditions ahead, and in some cases recommend alternate routes to drivers upstream of the congested sections [82]. In addition, other variable message signs are provided on the freeway near the sections where sufficient sight distances are not available, to assist drivers in formulating their expectations of the actual downstream traffic conditions. In Houston, for example, warning signs of this type have been installed on the Gulf Freeway near crest vertical curves to inform approaching drivers about traffic stoppages downstream resulting from accidents, stalled vehicles, or traffic bottlenecks. This warning system has shown significant reduction in the accident experience and overall delay on the Gulf Freeway [21].

Finally, the desired redistribution of traffic demand can be potentially facilitated if the drivers are made aware of congestion particularly caused by an accident or capacity-reducing incident. Freeway accident or incident information, and traffic conditions upstream can be carried to the motorists by variable message signs and commercial radio. In such case, it becomes more acceptable to the drivers who intend to enter the freeway from upstream of the troublesome location if metering rates are reduced or ramps are completely closed. Concurrently, freeway drivers upstream can be diverted by advising them to proceed along alternate routes, knowing that downstream capacity will be available

by the time they reach the downstream entrance ramps. This, of course, can be accomplished only if the incident is promptly detected both in time and location, and if immediate response and effective management are made by the surveillance and control system. While different methods of incident detection have been established including highway patrols, emergency roadside telephones, citizen-band radio, and television surveillance, demonstration projects of automated incident detection have shown substantial reduction in detection time due to eliminating the time lag of human interpretation [82]. Furthermore, computer-supervised incident detection has the capability of assessing the magnitude of capacity reduction on a quantitative basis, information that could only be guessed at by a human observer. With this quantitative information, entrance ramp metering rates can be automatically adjusted by the computer, and estimates of the likely delay to be encountered by freeway traffic can be made and relayed to the drivers or used to control variable message signs. Also, the computer has the ability for surveillance of the system hardware components, and for making automatic response when a hardware failure takes place. However, computer-supervised incident detection cannot exactly identify the nature of a detected incident or the kind of help needed unless it is coupled with closed-circuit television surveillance. Previous research on the Lodge Freeway indicated that computer detection can be useful in supplementing a television system by directing the attention of observers to incident situations [14]. Computer algorithms for incident detection are mostly in the research and development stage. Basically, these algorithms attempt to identify changes in traffic characteristics either temporally at a given

point on the freeway, spatially at a particular point in time or in combinations. False incident alarms generated by the computer represent a major operational problem in the field of freeway incident detection and management, particularly if highway patrols are to be dispatched every time. Generally, the rate of false incident alarms can be reduced only at the expense of the fraction of incidents detected. Research work is needed to develop improved methods for detecting and managing incident situations. Only after the development and testing of such methods will it be possible to operationally incorporate an incident detection capability into automated freeway surveillance and control systems.

1.2 A Capsule History, and Most Likely Projections

The evolution of freeway surveillance and control technology in the U.S. can be described by four technological levels as depicted in Figure 1.2. The bar graph gives the approximate timing of each of these levels. The first level includes the periods of scientific and technological resources development which generally precede the time frame shown in Figure 1.2. Scientific resources include the basic traffic flow and control theories, the empirical relationships, and the controlled laboratory and field experiments. Technological resources, on the other hand, are primarily the computers but they also include traffic signal, television, and detector technologies.

Level II, the period of first-generation freeway surveillance and control systems, can be dated back to the 1960's. During this period, research and demonstration projects financed by Federal, state, county, and city funds, were designed in Detroit, Chicago, and Houston [82]. In

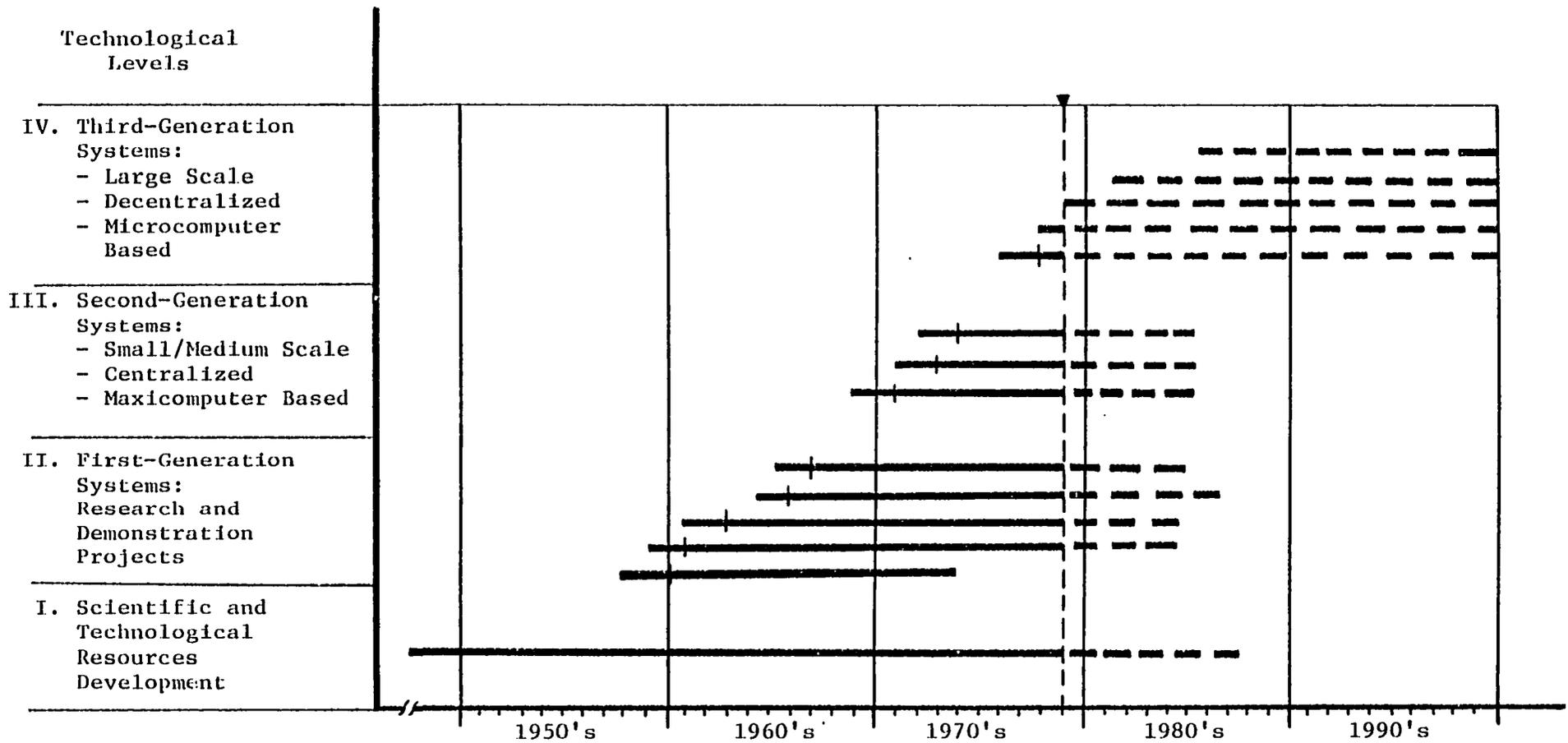


FIGURE 1.2

GENERAL LEVELS OF FREEWAY SURVEILLANCE AND CONTROL TECHNOLOGY EVOLUTION

1960, the use of closed-circuit television as a freeway surveillance tool was pioneered on the John C. Lodge Freeway in Detroit. Manually operated lane control signals, variable speed signs, and ramp closure signs were the major forms of control and driver information. Ramp metering control started later in 1966 utilizing a closed network of ultrasonic vehicle detectors connected to an IBM 1800 computer for data logging. The Chicago Area Expressway Surveillance Project was initiated in 1961 pioneering the instrumentation of automatic detection systems. Vehicle presence detectors connected to a central computer were installed on the Eisenhower Freeway for ramp control purposes. In 1963, the Gulf Freeway Surveillance and Control Project in Houston, Texas, combined television surveillance and ramp control emphasizing gap acceptance concept. Using the same concept, the moving-merge type of control was first experimented in 1967 in Woburn, Massachusetts. Finally, on-line computer surveillance and control of tunnels was pioneered in the Lincoln Tunnel during a joint study conducted by the Port of New York Authority and IBM which ran from 1966 until 1969.

Level III is the period of second-generation freeway surveillance and control systems. Starting about 1970, several surveillance and control projects with different features have been designed and implemented on the freeways of some cities. Notable in this regard is the Los Angeles Area Freeway Surveillance and Control Project, LAAFSCP, which was put into use in 1971. It has since evolved a fully operational system of automatic monitoring. Ramp metering control has been coordinated with freeway improvements and incident detection. On the Northern Long Island Corridor, New York, an integrated motorist information system was established in

1973 utilizing current motorist aid and information technology and real-time traffic management. The Dallas North Central Corridor Project, initiated in 1974, is noteworthy because it combines arterial street signal control with freeway ramp control and selective television monitoring. Other surveillance and control systems have been developed and implemented at various levels in Seattle, Minneapolis, Atlanta, and Phoenix in attempt to relieve existing congestion problems [82]. The Minneapolis system, initiated in 1973, is of particular interest because it provides preferential treatment to express buses at the entrance ramps to the I-35 W. Corridor.

In summary, today there are about ten operational surveillance and control systems covering only about one percent of the urban freeway mileage in the U.S. [82]. The centralized type of surveillance and control is the one in predominant use. It is characterized by having all the decision-making and computational capabilities on one level located at one geographic location. Most of the installed computers are of the conventional medium type (minicomputers).

Using the scenario technique described by Kahn and Weiner [44], the basic underlying trends which bear on the future of freeway surveillance and control technology can be summarized in the following six points:

1. Growing automobile travel and the associated congestion, accidents, and environmental problems.
2. Increasing recognition among transportation engineers and planners that the traditional solution to prevent congestion by constructing new freeways has become both economically and politically more unfeasible.
3. The joint policy promulgated by the Federal Highway Administration and the Urban Mass Transportation Administration emphasizing transportation system management as a short-term planning concept. This concept requires making best

use of existing highways and vehicles to achieve greater urban mobility [82].

4. Increasing concern about the reliability and efficiency of existing traffic control and surveillance systems. Many of the concepts presented in section 1.1 are being further developed and refined to increase their functionality and effectiveness [82].
5. Rapid technological advances in solid-state technology which led to the relatively recent development of microcomputers at reduced cost and size [1,80]. These developments are observed by the industry with an eye toward incorporating them into future surveillance and control hardware [82].
6. Noticeable shift toward the decentralized and multi-level schemes of surveillance and control in which the decision-making and computational capabilities are distributed from a geographic viewpoint and placed at various levels in a hierarchical organization [71,82].

From the standpoint of looking toward the future, the above trends seem likely to continue at least for the next two decades. Examination of these trends indicates that rapid changes and advances in freeway surveillance and control systems are anticipated in the near future. This is represented by technological level IV in Figure 1.2. It is the successful application of technology, in particular microelectronics, which will lead to hardware advances. Other advances will be made possible as a result of current research and development efforts coupled with the experience gained over the past years. Significant developments are expected to take place in traffic responsive ramp control strategies, automatic incident detection and management, driver information systems, and merge control. Concurrently, the increasing cost of communication links and equipment will accelerate the shift toward decentralized schemes of surveillance and control. The technological developments and reduced cost of microcomputers will provide another technological impetus for

decentralization and distributed computations. Last but not least, preferential treatment of high occupancy vehicles is expected to receive more attention in future freeway control systems.

1.3 The Purpose and Utility of This Research

The basic premise of any computer-supervised freeway surveillance and control system is its adaptability to the rapidly changing traffic conditions in time and space. However, in complex traffic situations where optimum utilization of corridor capacity is needed, it is not sufficient to determine the control strategies on the basis of traffic conditions which exist at the exact time of implementing these strategies. To help illustrate, consider a variable message sign warning the approaching freeway drivers of traffic congestion downstream, and suggesting alternate routes based on traffic conditions observed at the same moment of the message. Keeping in mind that congested traffic operations are usually unstable, that is, congestion appears first at one point then at another, and that the warning message itself will probably change the situation, it is very likely that when the drivers reach the trouble-spot, they may encounter quite different conditions from those previously displayed. As a result, the message information may seem to have been misleading, and perhaps over the long run drivers may completely ignore such information. In addition, the message may cause significant degradation to the performance of the overall control system. It appears important, therefore, that variable message signs should allow for the effects they are likely to cause in the immediate future through incorporating short-term forecasts of traffic performance into the control logic. Similarly, in both centralized and decentralized schemes of real-time ramp control, the

optimal control strategies are based upon short-term forecasts of traffic demand conditions. These forecasts are usually required for 5 to 10 minutes ahead in the future in order to allow enough response time to the implemented control. Figure 1.3 illustrates these concepts.

Another important application of traffic prediction in freeway surveillance is that of incident detection. Recent research [13] has concluded favorable results when statistical forecasting is utilized to detect traffic disturbances caused by accidents or capacity-reducing incidents. In particular, a forecasting model is employed to provide real-time estimates of the expected values of future traffic performance measurements. Due to the rapid changes in freeway traffic conditions, these estimates are usually made for one to five minutes ahead in time. An accident is detected if the difference between an actual observation and its estimate significantly exceeds the expected difference due to random variations in traffic behavior.

Forecasting techniques of real-time traffic systems are still in their early development stage. With the increasing complexity of these systems, and the growing concern about their operational reliability, considerable interest has been recently devoted to the analysis and forecasting of their dynamic behavior. Ad hoc models have been proposed in many situations, however, a consensus as to which model is appropriate has not been reached yet. The major weakness of these models is that there is no sound basis for their form or their transferability from one freeway system to another. A more general and powerful approach for modeling and forecasting a traffic system is to collect data describing how this system did behave over time, and then to construct a model

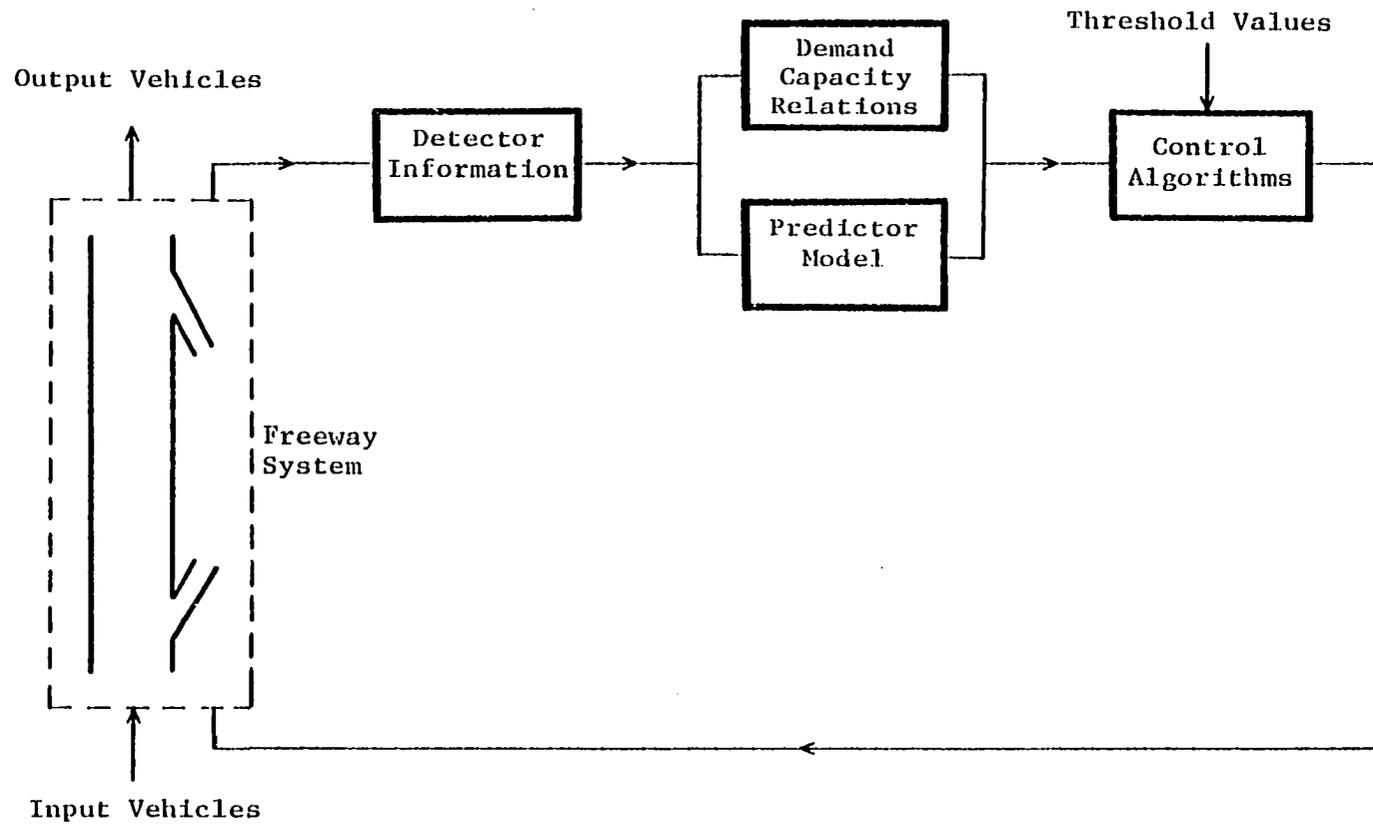


FIGURE 1.3

ROLE OF TRAFFIC PREDICTION IN FREEWAY
SURVEILLANCE AND CONTROL

based on the observed system behavior. This approach is known as "time series analysis" where very general types of models exist for describing the stochastic properties of dynamic systems over time. The purpose of this dissertation is to investigate these issues. In particular, this study explores: (1) the dynamic nature of freeway traffic measurements over time, (2) the alternative techniques of forecasting real-time traffic systems, (3) the development of a time series forecasting and simulation model for these systems, and (4) the formulation of incident detection algorithms based on the developed forecasting model.

1.4 Overview of the Next Chapters

In Chapter II, the nature of freeway traffic measurements is discussed in terms of their inherent characteristics, their measurement techniques, and their interrelationship. Traffic flow models are briefly reviewed both at the microscopic and macroscopic levels, with emphasis upon the propagation of disturbances caused by incidents. Chapter III covers the alternative techniques of forecasting real-time traffic control and surveillance systems. There is a discussion of three ad hoc forecasting models, as well as the time series analysis models developed by G. E. Box and G. M. Jenkins [7]. Chapter IV describes the freeway traffic data base used in the analysis, and the previous work on modeling traffic time series data. Also, it describes in detail the development of a time series forecasting model and its utility in operational forecasting and simulation. The developed model is then compared on a comparative basis with the other ad hoc models. Chapter V discusses the application of the developed model in freeway incident detection. It represents a review of incident detection algorithms and discusses the formulation of new

algorithms based on time series analysis of traffic measurements. Enough theory is provided throughout Chapters II-V so as to make the discussion meaningful, but not so much as to obscure the underlying practical applications. Finally, Chapter VI covers the conclusions of this study and some prospects for future research.

CHAPTER II

NATURE OF FREEWAY TRAFFIC STREAM MEASUREMENTS

2.1 Introduction

The study of the true nature of freeway traffic measurements has been pursued extensively by investigators in the transportation field. Much of this interest has evolved from the difficulties inherent in modeling the behavior and interactions of a complex made up of drivers, vehicles, and a roadway. This chapter begins with the definitions of basic traffic variables and their measurement techniques. Then, there is a review of various traffic flow models, specifically microscopic car-following models and macroscopic continuum models. Particular emphasis is given to the propagation of disturbances caused by freeway incidents which provides the necessary background required for Chapter V. No attempt will be made in this chapter to review the stochastic properties of traffic variables.

2.2 Traffic Stream Variables: Definitions and Measurements

Streams of traffic are by their nature physical successions of discrete occurrences of vehicles, however, they are sometimes treated as a continuous fluid. The major macroscopic variables which characterize a stream of vehicles are the flow rate or volume q (vehicles per unit time), the concentration or density K (vehicles per unit length of

roadway), and the average speed of a group of vehicles in space or space mean speed \bar{u}_s (miles per hour or kilometers per hour) [19]. Volume is a characteristic of traffic passing a point, while density and space mean speed are spatial characteristics at a given instant of time [41]. These three variables are of particular interest to the traffic engineer, since volume describes how many vehicles are moving (demand), density and space mean speed together describe the quality of service experienced by motorists [89].

In studying the behavior of a physical stream of traffic, it is interesting to consider the projection of the stream onto a space-time plane. A plot of the trajectories of each vehicle in both time and longitudinal distance might appear as in Figure 2.1, with each trajectory describing the movement of a single vehicle as a function of time. The slope of the trajectory at any point represents the vehicle's current speed. Vehicle trajectories can be constructed by continuous monitoring of point detectors over some period of time, or by sequential aerial photos of a stretch of roadway.

Measurements of traffic stream variables are made either by observing vehicles as they pass a particular point in space, or by observing their location at a particular point in time, as depicted in Figure 2.1. The variables q , K , and \bar{u}_s are meaningful only as averages over small band of width dx or dt since they count the number of vehicles in a unit of time or unit of space. Edie [26] provides operational definitions of the variables q , K , and \bar{u}_s based on the methods used in measuring these variables. He also discusses the different averaging schemes to be employed in each situation.

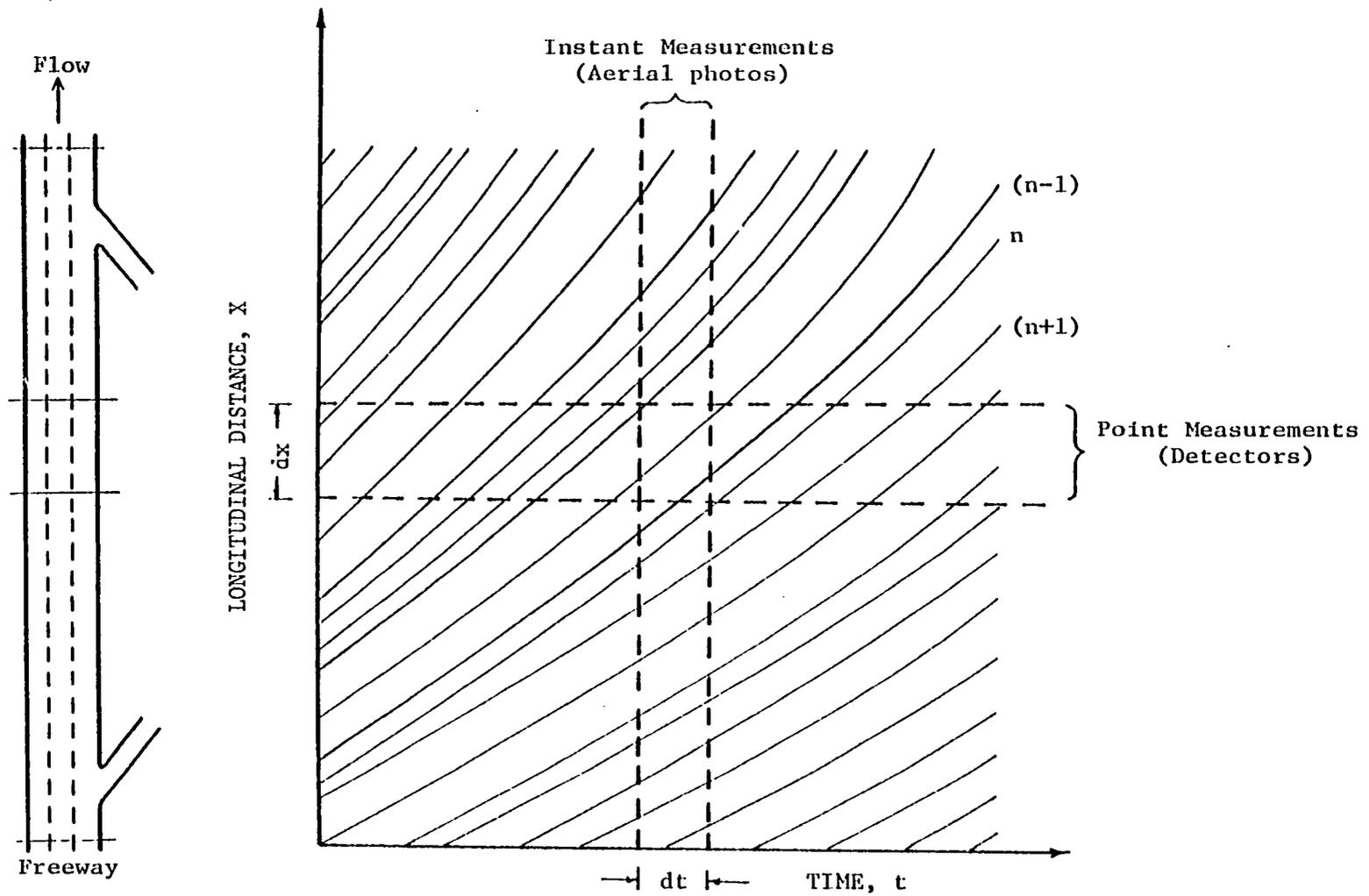


FIGURE 2.1

VEHICLE TRAJECTORIES AND MEASUREMENT METHODS

Every computer-supervised traffic surveillance and control system uses detectors to obtain information concerning traffic conditions. Magnetic loop detectors are the most widely used devices to sense vehicle presence over successive detection points. The principle of magnetic loop detectors is that a vehicle resting in, or passing over the loop causes unbalance in a tuned circuit and sends an impulse signal to a computer [82]. Current freeway surveillance systems (notably those in Los Angeles, Chicago, and Minneapolis) use a central computer for signal processing. As mentioned in Chapter I, evolving technology may make it possible to do some or all of the required processing by microprocessors [1,76]. The processed signal gives direct measure of traffic volume and time headway between vehicles [2]. Vehicle speeds can be derived from a combination of detectors forming a speed trap, or from a single detector under the assumption of a standard length of each vehicle [31]. The average speed of a group of vehicles passing a point during a time interval is called the time mean speed and is different from the space mean speed [41]. Traffic density at a freeway section bounded by two detectors is equivalent to the number of vehicles stored between these two detectors. Although this method has been used successfully to measure traffic density in the Lincoln Tunnel in New York, it requires precise measurement of car lengths and presumes relatively few changes of lane by individual vehicles [31,32]. Therefore, the application of this method to freeways is rather questionable [31]. A surrogate for traffic density that can be measured at a point from detector information is occupancy, the percentage duration of activation of a presence detector [2]. It is a

nondimensional variable giving the proportion of time a single point on a roadway is occupied by vehicles. Athol [2], found that density and occupancy are linearly related up to densities of about 80 vehicles per mile per lane depending on the average vehicle length.

Detector failures represent a major hardware problem in freeway surveillance systems today. Dudek, et al. [22], reported that 87 percent of the hardware problems experienced on the Gulf Freeway warning system in Houston were due to detector failures, while eleven percent were related to computer hardware. They also found that detectors failed at a rate of 3.78×10^{-4} failures per detector hour. The Traffic Control Systems Handbook [82] describes new innovations in detector technology which may relieve the problem of detector failures.

2.3 Volume, Density, and Speed Interrelationship

The definitions of macroscopic traffic stream variables indicate that they are all interrelated, and by dimensional analysis this interrelationship can be expressed as:

$$q = K \cdot \bar{u}_s \quad (2.1)$$

where:

q = volume (vehicle/hour),

K = density (vehicle/mile),

\bar{u}_s = space mean speed (miles/hour)

The above relation is known as the "equation of state" of traffic flow [24]. In addition, the definitions require volume to be zero when density is zero (no vehicles on the road), and require volume to be zero at maximum density (jammed conditions). Further, volume should increase to a maximum level (known as capacity) and decrease before

density reaches a maximum value. Therefore, the relationship between traffic volume and density must have a general form similar to that depicted in Figure 2.2. This is called the fundamental diagram of traffic flow [24]. The slope of the line extending from the origin to any (q,k) point on the volume-density curve represents the space mean speed corresponding to flow state (q,k) .

Several volume-density models have been proposed or derived from empirical studies or theoretical considerations [24]. Most traffic engineering applications, however, adapted the Greenshield's linear relationship between density and space mean speed to develop workable approximations [89]. This relationship can be expressed as:

$$\bar{u}_s = \bar{u}_f(1 - K/K_j) \quad (2.2)$$

where:

\bar{u}_f = space mean free speed

K_j = jam density

The corresponding volume-density model is parabolic and can be derived by substituting q/K for \bar{u}_s in equation 2.2.

$$q = \bar{u}_f K(1 - K/K_j) \quad (2.3)$$

The main advantage of such traffic model is that it gives useful averages of the overall statistical properties of a traffic stream. However, Haight [37] reported that there is no unique volume-density relationship which applies in all situations, and that a volume-density curve is a characteristic of a particular location at a specific time with a given population of drivers. Figures 2.3 and 2.4 are plots of volume versus occupancy (surrogate for density) for the afternoon peak period from 2:40 p.m. to 7:00 p.m. at the Chicago and Glendale detector

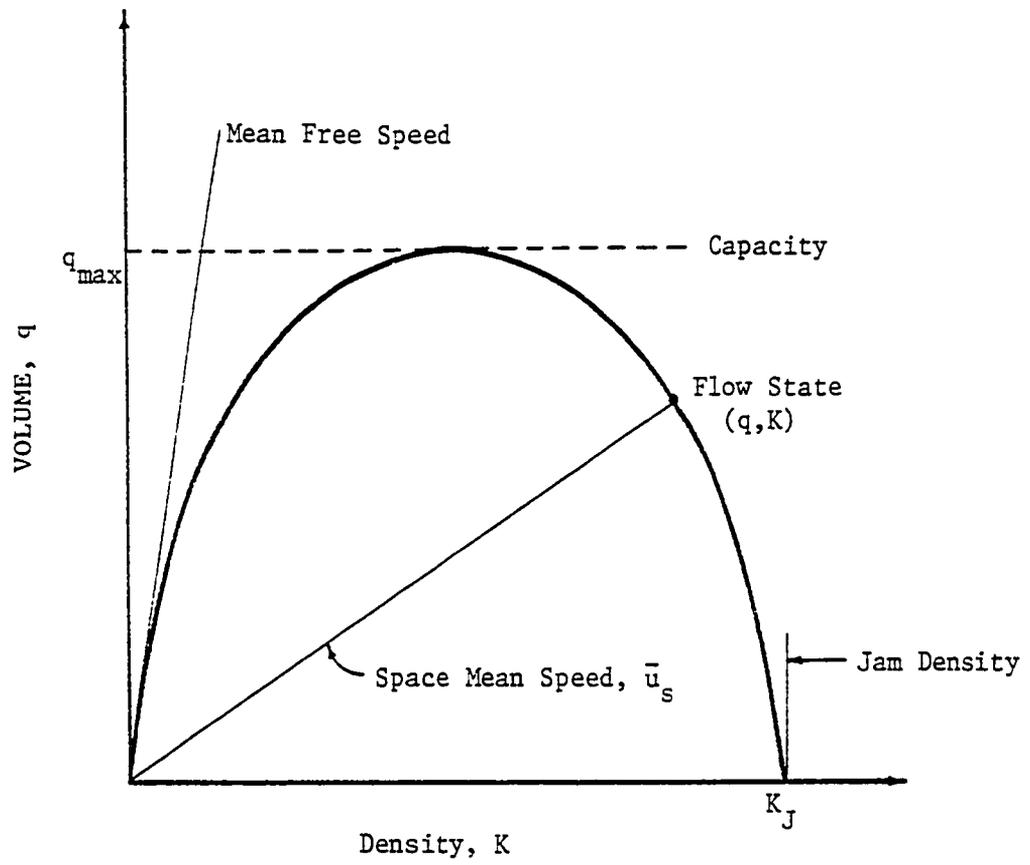


FIGURE 2.2

THE FUNDAMENTAL DIAGRAM OF TRAFFIC FLOW

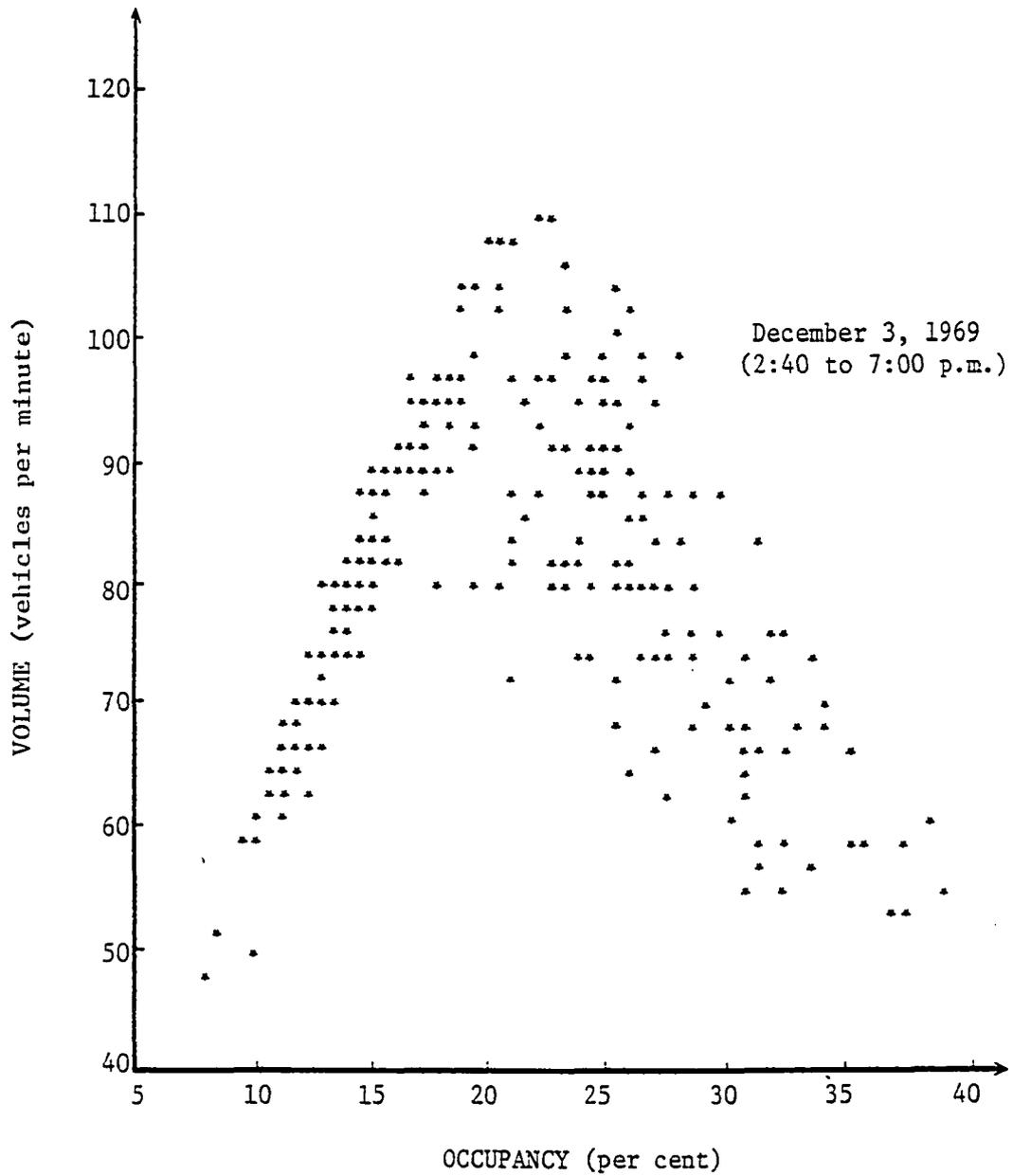
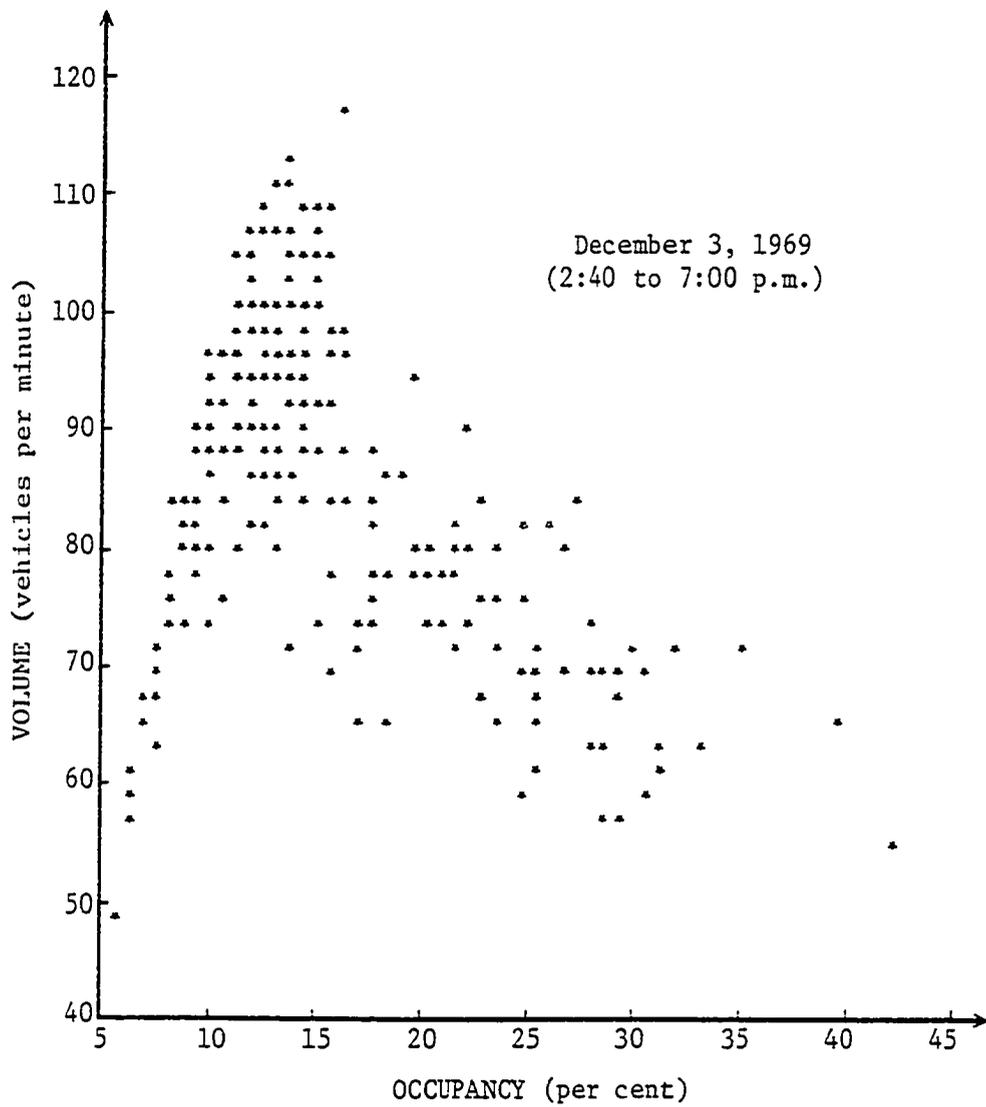


FIGURE 2.3

VOLUME VERSUS OCCUPANCY AT THE
CHICAGO DETECTOR STATION



stations on the Lodge Freeway in Detroit. The two plots differ both in shape and range of values although the two stations are separated by only one mile. Probably, these differences arise from variations in flow characteristics, types of vehicles, driving situations, and geometrics. The increased scatter of data points at occupancies greater than about 15 percent has been confirmed by Mika, et al. [54], who differentiated between two modes of traffic stream behavior, one relatively stable over time, and the other oscillatory when occupancies exceeded 13 to 16 percent.

2.4 Microscopic Traffic Flow Models

A more theoretical approach for studying the microscopic behavior of a traffic stream was first proposed by Clayton [12] who related a driver's headway to his ability to avoid a collision when the vehicle in front brakes suddenly. Chandler, et al. [10], incorporated this concept in what is known as the car-following model, which can be expressed in the form:

$$\text{response} = \text{sensitivity} \times \text{stimulus}$$

A driver probably reacts to the relative speed between his car and the car in front ($\dot{x}_n - \dot{x}_{n+1}$) which is assumed to be the stimulus. His response is constrained by the vehicles surrounding him and is taken as his acceleration or deceleration rate \ddot{x}_{n+1} . If λ represents the driver's sensitivity coefficient (intensity of reaction to stimulus from the vehicle in front), and τ the average driver reaction time (time between receiving a stimulus and implementing a response), the car-following model at time t can be expressed as

$$\ddot{x}_{n+1}(t + \tau) = \lambda[\dot{x}_n(t) - \dot{x}_{n+1}(t)] \quad (2.4)$$

Gazis, et al. [34], found experimentally that the driver's sensitivity coefficient λ is inversely proportional to the space headway, that is,

$$\ddot{x}_{n+1}(t + \tau) = \frac{\lambda_0}{x_n(t) - x_{n+1}(t)} [\dot{x}_n(t) - \dot{x}_{n+1}(t)] \quad (2.5)$$

where λ_0 is the constant of proportionality. They also found that by integrating equation 2.5, the steady-state relationship between speed and density is given by

$$\bar{u}_s = \lambda_0 \log_e (K_j/K) \quad (2.6)$$

A traffic stream is in steady-state condition when the joint distribution of vehicle speeds and lengths is the same at all points along the roadway [24]. An expression identical to that in equation 2.6 was derived by Greenberg [36] using a different approach based on the fluid flow analogy which is discussed in the next section. Greenberg fitted his model to speed and density observations from the north tube of the Lincoln Tunnel in New York, and found values of λ_0 equals 17.2 and K_j equals 227 vehicles per mile. One of the limitations of the model described by equation 2.6 is its violation of the boundary condition at zero volume and density where there is no upper bound on stream velocity. In order to overcome this discrepancy, several investigators [33,52] proposed modifications of the driver's sensitivity coefficient λ to include various powers of vehicle speed and reciprocal headway, that is,

$$\lambda = \frac{\lambda_0 \dot{x}_{n+1}^m(t)}{[x_n(t) - x_{n+1}(t)]^l} \quad (2.7)$$

Particular solutions of the modified car-following model include the case where $m = 0$ and $\lambda = 2$ which yields the Greenshield's model discussed earlier. Gazis, et al. [33], fitted car-following data from the Lincoln Tunnel and found the best correlation for values of $m = 1$ and $\lambda = 2$. May and Keller [52] proposed values of $m = 0.8$ and $\lambda = 2.8$ for freeways, and $m = 0.6$ and $\lambda = 2.1$ for tunnels.

2.5 Macroscopic Traffic Flow Models

In contrast to the car-following models which keep track of individual car movements at the micro-level, the hydrodynamic theory of traffic flow describes the movement of a sizable aggregate of vehicles forming a traffic stream at the macro-level. Lighthill and Whitham [50] and Richards [67] introduced analogies between the flow of fluids and the movement of vehicular traffic. These analogies only hold for high traffic densities, the case where important traffic control problems exist. Individual vehicles forming a traffic stream are replaced with a one-dimensional compressible fluid which has a certain density $K(x,t)$ and a flow rate $q(x,t)$, where x is a linear coordinate measured in the direction of flow, and t is the time. Figure 2.5, adapted from Edie [24], shows the trajectories of vehicles in both time and longitudinal distance. If measurements of the flow rate q and density K are made on a section of highway Δx long during a time period Δt , as illustrated in Figure 2.5, the principle of conservation of vehicles results in the differential equation

$$\frac{\partial K}{\partial t} + \frac{\partial q}{\partial x} = 0 \quad (2.8)$$

Equation 2.8 is recognized as the continuity equation of a compressible fluid.

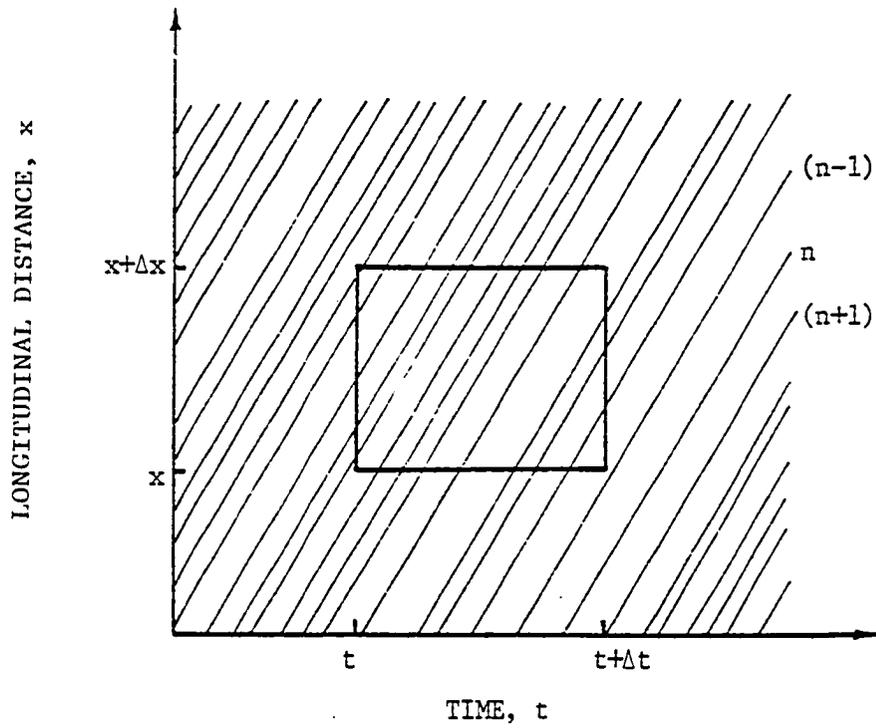


FIGURE 2.5

VEHICLE SPACE-TIME TRAJECTORIES, AND RECTANGULAR REGION
FOR DERIVATION OF THE PRINCIPLE OF CONSERVATION OF VEHICLES
(Adapted from Reference 24)

The fundamental hypothesis of Lighthill and Whitham is that at any point along a roadway, the flow rate q depends explicitly on the density of traffic K , and on the position x , so that

$$q = q(K, x) \quad (2.9)$$

If the spatial distribution of density is homogeneous over a particular roadway, then equation 2.9 reduces to

$$q = q(K) \quad (2.10)$$

Equation 2.10 is the "equation of state" discussed earlier, and by partial differentiation with respect to x the result is

$$\frac{\partial q}{\partial x} = \frac{\partial q}{\partial K} \cdot \frac{\partial K}{\partial x} \quad (2.11)$$

For small variations of density about an average value, the partial derivative of the flow rate with respect to density can be assumed as a constant C , and consequently equation 2.11 will have the form

$$\frac{\partial q}{\partial x} = C \frac{\partial K}{\partial x} \quad (2.12)$$

By substituting equation 2.12 into the equation of continuity 2.8, this results the differential equation

$$\frac{\partial K}{\partial t} + C \frac{\partial K}{\partial x} = 0 \quad (2.13)$$

which has the general solution

$$K = f(x - Ct) \quad (2.14)$$

where f is an arbitrary function. This solution represents a wave of density K moving in the direction of increasing x with velocity $C = \partial q / \partial K$. The conceptual implication of equation 2.14 is that a small disturbance in traffic density propagates as a kinematic wave with velocity C given by the tangent to the volume-density curve at the corresponding point. For densities below the density at capacity-flow,

the slopes of the tangents are everywhere positive, and the disturbances propagate forward relative to the roadway. This is the case where the interaction between vehicles is small. For densities above the density at capacity-flow, the slopes of the tangents are negative so that disturbances propagate backwards relative to the roadway. At conditions of capacity-flow the wave is stationary.

Based upon the relationship $q = K \cdot \bar{u}_s$ of equation 2.1, the velocity of wave propagation C is given by

$$\begin{aligned} C &= \frac{\partial}{\partial K} (K \cdot \bar{u}_s) \\ &= \bar{u}_s + K \frac{\partial \bar{u}_s}{\partial K} \end{aligned} \tag{2.15}$$

Herman and Rothery [39] found experimentally that the term $\partial \bar{u}_s / \partial K$ is always negative, and therefore C is always less than the average speed of the traffic stream. The quantity ($\lambda = \bar{u}_s - C$) was defined by Herman and Rothery as the driver's sensitivity coefficient, and is equal to the velocity of propagation of a disturbance back through a chain of vehicles relative to the moving traffic stream. They also observed an increasing linear relationship between λ and average speed up to 60 mph. In the car-following model of equation 2.5, Herman, et al. [40], found that instability in flow conditions arises when $\lambda_0 \tau$ exceeds $\pi/2$, and hence the spacing between successive vehicles oscillates with increasing amplitude. The hydrodynamic analogy of this situation is the case when drivers adjust spacing between their cars to changing traffic conditions, and therefore restrict the freedom of individual motion [66]. To allow for this possible instability, Lighthill and Whitham suggested an extension to their hydrodynamic theory to include a "diffusion effect" and an "inertia effect." The diffusion effect accounts for the fact that each driver adjusts his speed to flow changes slightly ahead of

him, while the inertia effect allows for a time-dependence in the behavior of individual vehicles [30]. Therefore equation 2.10 is modified so that the flow rate q becomes a function of the density K , the density gradient $\partial K/\partial x$, and the time rate of concentration $\partial K/\partial t$, that is,

$$q = q(K, \partial K/\partial x, \partial K/\partial t) \quad (2.16)$$

Franklin [30] showed how this modification makes it possible that flow becomes unstable for certain ranges of density.

2.6 Propagation of Disturbances Caused by Freeway Incidents

Since Kinematic waves propagate with velocities which are functions of the vehicular density, it is possible to find different waves moving through a traffic stream. A particular situation of interest is that where a capacity reducing incident occurs on a section of freeway as shown in Figure 2.6. For simplicity, Greenshield's parabolic volume-density curve is used to demonstrate the different flow states associated with the incident situation. It is assumed that the flow state before the incident is below capacity as represented by point one. Point two is the congested regime of traffic operations upstream the incident, and point three represents the flow state downstream the incident. Kinematic waves in the high-density traffic (point two) propagate upstream of the incident site, while kinematic waves in the low-density traffic (point three) travel downstream at a higher velocity than the waves in the normal flow (point one). When the two latter waves meet, a new wave will form. Lighthill and Whitham [50] referred to this new wave as "shock wave" and its velocity C_{31} is

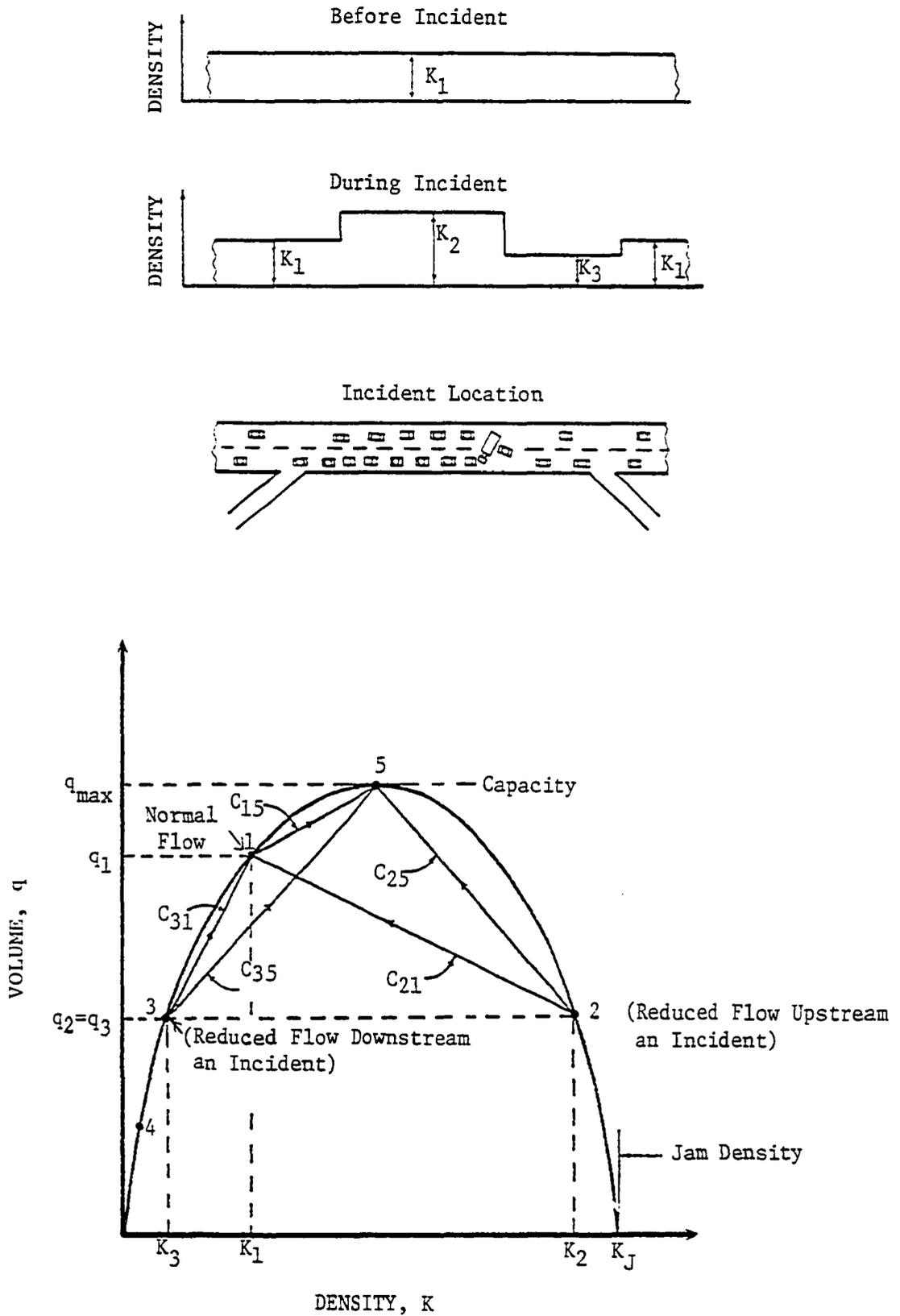


FIGURE 2.6

FLOW REGIMES AND SHOCK WAVES GENERATED BY A FREEWAY INCIDENT

given by the slope of the chord between points one and three on the fundamental diagram.

In addition to the shock wave of expansion which travels downstream of the incident site, other shock wave of congestion growth may propagate upstream depending on the slope of the chord connecting the prevailing flow state before the incident with point two. In Figure 2.6 the slope of the chord connecting flow states one and two is negative indicating that the disturbance will propagate upstream and generate state two of congested operations. This is the case when a capacity-reducing incident occurs under medium or heavy flow conditions (peak period operations). If light flow conditions prevail before the incident as represented by point four, then the slope between states two and four is positive. This indicates that the demand downstream the incident can still be accommodated at an increased level of traffic density.

Figure 2.7 demonstrates the same sequence of events by means of spatial distributions of average speed, density and volume at some fixed moment. Regions one and four represent prevailing normal traffic conditions before the incident. Region two immediately upstream of the incident represents congested operations where a queue of vehicles builds up if demand exceeds bottleneck capacity. This is indicated by the lower than normal speed and higher than normal density. In region three immediately downstream of the incident, the traffic flow is function of the bottleneck capacity which is less than the freeway capacity in region three. Consequently, region three is characterized by higher than normal speed and lower than normal density. The boundary between

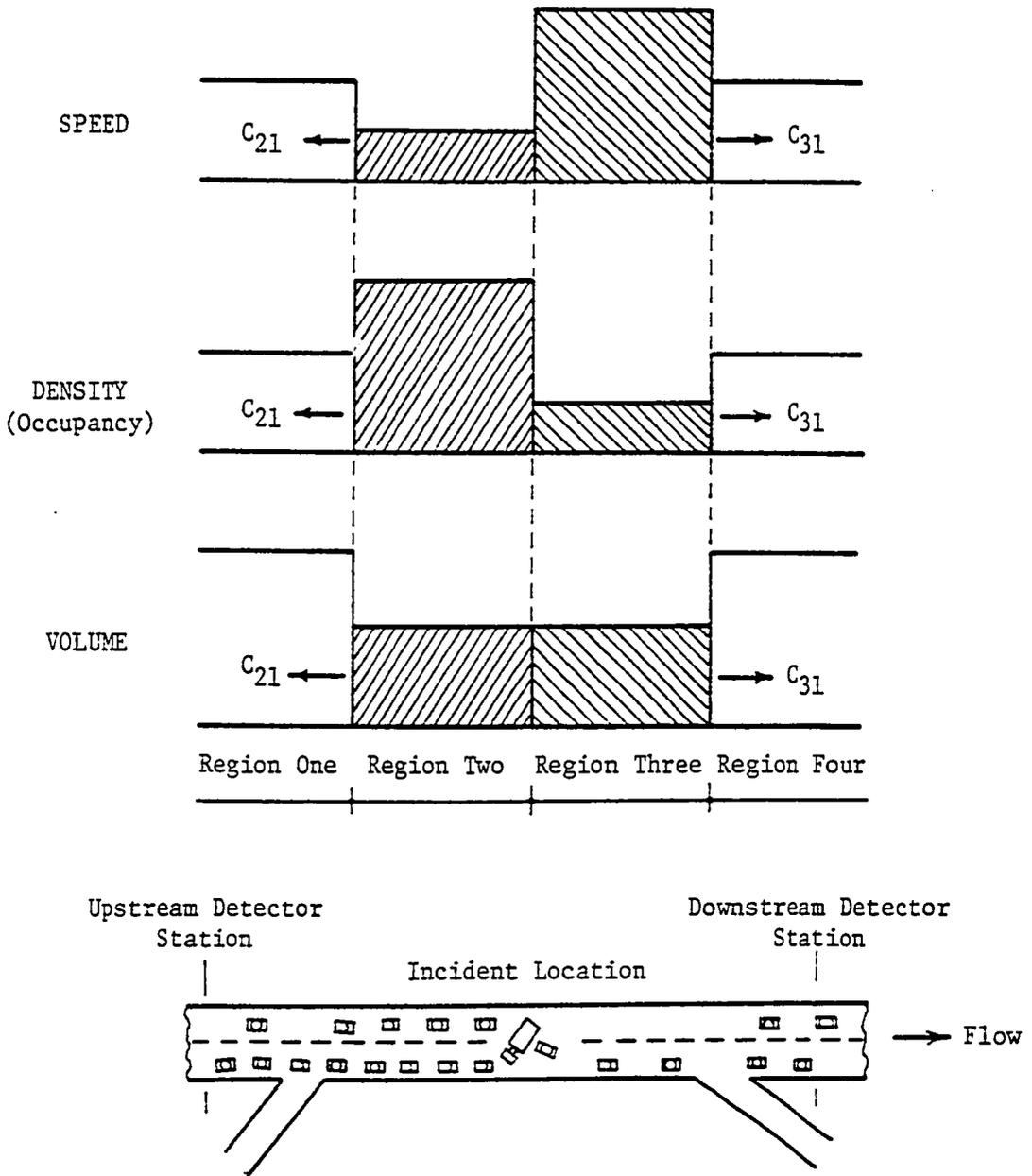


FIGURE 2.7

SPATIAL DISTRIBUTIONS OF SPEED, DENSITY AND VOLUME DURING AN INCIDENT

regions one and two moves upstream, while the boundary between regions three and four moves downstream. The movement of these boundaries is what has been referred to as shock waves by Lighthill and Whitham. This type of diagram provides the conceptual ideas necessary for developing incident detection models which are discussed in Chapter V. The models should detect the passage of either or both of the waves at the upstream and downstream detector stations.

The removal of an incident after some elapsed time T since its occurrence allows the queue of vehicles stored upstream of the incident site to start moving downstream. Theoretically, the flow rate should rise to capacity of the freeway at the incident site. The application of kinematic wave theory to the recovery process results in two shock waves with velocities C_{25} and C_{35} as shown in Figure 2.6. A shock wave of recovery propagates upstream of the incident site with velocity C_{25} indicating the clearing up of the queue of vehicles stored during time T . Also, a shock wave of release which defines the boundary between capacity flow and the reduced flow during the incident travels downstream of the incident site at velocity C_{35} . Figure 2.8, adapted from Messer, et al. [53], is a space-time diagram showing the different shock waves which exist when an incident occurs and after it is removed. Point A represents the beginning of the incident where the two shock waves of congestion and expansion proceed upstream and downstream with velocities C_{21} and C_{31} respectively as discussed before. Point B represents the incident removal, and it is noticed that some time, T , has elapsed. The shock waves of recovery and release propagate upstream and downstream from point B with velocities C_{25} and C_{35} , respectively.

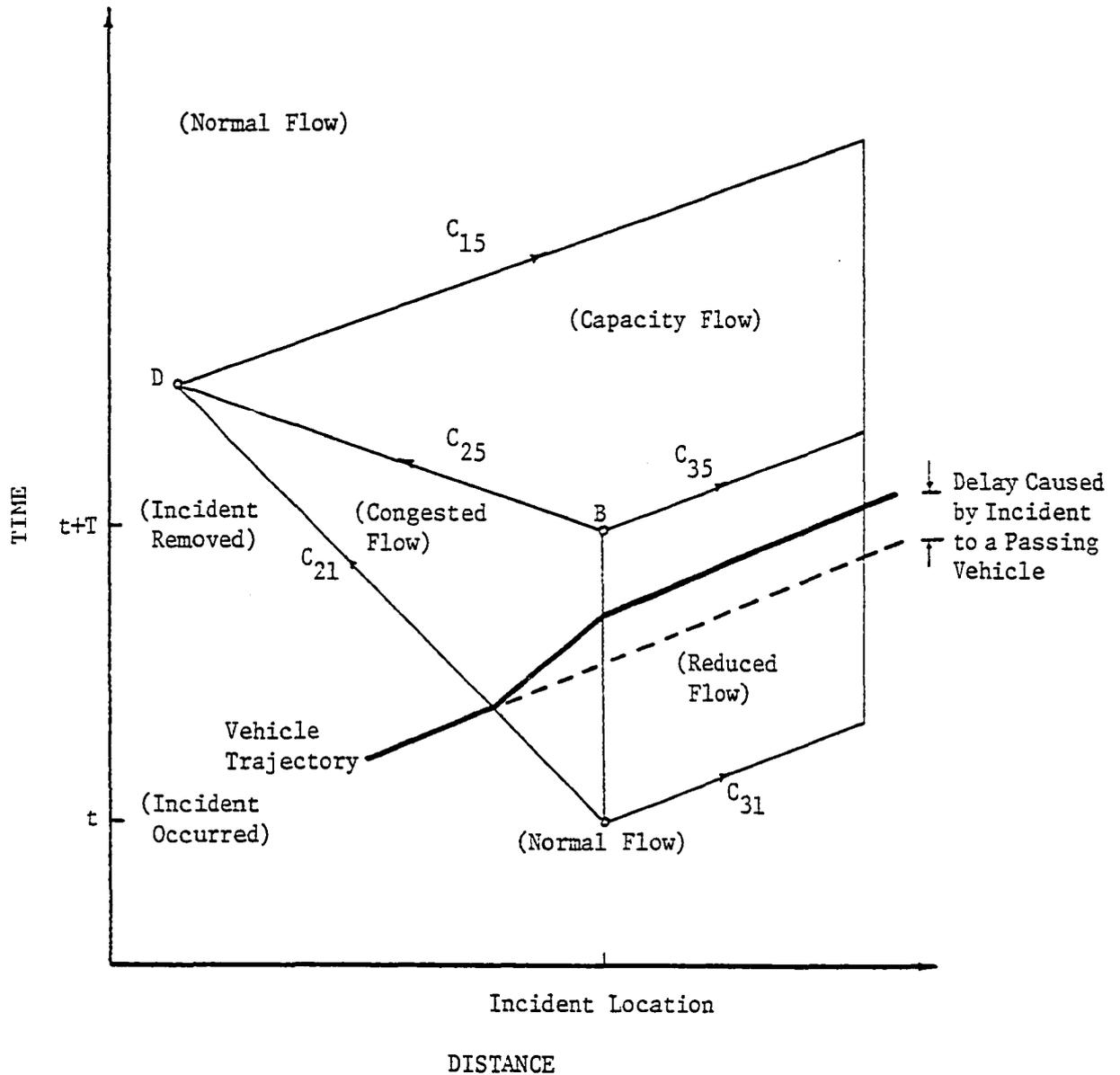


FIGURE 2.8

TIME-SPACE DIAGRAM OF FREEWAY FLOW REGIMES AND SHOCK WAVES DURING AND AFTER AN INCIDENT

As depicted in Figure 2.8, one remaining shock wave should propagate upstream before all traces of the disturbances generated by the incident disappear. Sometime after the incident is removed, the shock wave of recovery C_{25} meets the shock wave of congestion C_{21} at point D which designates the complete dissipation of the queued vehicles. The final clearing shock wave proceeds downstream from point D with velocity C_{15} defining the boundary between capacity flow and normal traffic flow before the incident.

Messer, et al. [53], used the time-space diagram of Figure 2.8 to estimate the freeway travel time during incident conditions. The computational procedure used was based on Greenshield's linear speed-density model of equation 2.2. In a similar manner, Dudek, et al. [23], discussed the use of the same kind of diagram to determine the maximum detector spacing for an automated freeway incident detection system. They found that the maximum spacing between detectors is a function of the prevailing traffic conditions before the incident as well as the elapsed time between incident occurrence and removal.

Several studies have been made to test and apply the conceptual ideas of kinematic wave theory. Edie and Baverez [25] investigated the generation of shock waves in tunnels by means of contour maps of flow, density, and speed at traffic bottlenecks, and they found wave velocities to vary between 9.5 and 14 mph. These results were obtained for a single lane traffic with no passing allowed. Herman and Rothery [39] measured the speed of propagation of disturbances through an eleven-car platoon moving on a test track and found it to be 12.3 mph. By means of aerial photographs of individual lane traffic on the Lodge Freeway

in Detroit, Forbes and Simpson [29] found that acceleration waves propagate faster upstream than deceleration waves. Their results reflected the interaction between driver response time, speed and headway.

2.7 Summary

A considerable amount of effort has been expended over the years in studying the true nature of freeway traffic stream measurements. It is interesting that traffic flow theory still receives considerable emphasis in current research activities. However, the concern is growing towards reducing the gap between theory and practice by implementing the body of knowledge so far developed to solve urban traffic problems. To some extent, this chapter has attempted to discuss the opposite ends of traffic flow behavior spectrum. The car-following models represent movements of individual vehicles, whereas the fluid flow models describe movement of a population of vehicles. Importantly, the two approaches are deterministic flow behavior models. In some occasions, however, it is necessary to have a probabilistic model of traffic behavior. For example, in ramp metering studies it is important to know the probability distribution of time headways in the freeway traffic stream. References [19,89,24,37] discuss the probabilistic aspects of traffic behavior.

CHAPTER III

EXISTING FORECASTING TECHNIQUES

APPLIED TO TRAFFIC SYSTEMS

3.1 Introduction

The design and operation of real-time traffic control and surveillance systems depends heavily on developing efficient and accurate methods for making short-term forecasts of traffic demand. Whereas ad hoc forecasting models have been used in the past without solid theoretical justification, the growing concern about control and surveillance reliability demands more strenuous attempts to find a complete and general approach for forecasting the dynamic aspects of traffic systems.

This chapter begins with some basic definitions and concepts in time series, followed by a typological breakdown of forecasting techniques into four fairly distinct categories. There is then a review of three widely used ad hoc forecasting models: the moving average model, the exponential smoothing model, and the Trigg and Leach adaptive exponential model. Particular emphasis is given to the contribution that time series analysis can make to operational forecasting, specifically the autoregressive integrated moving average models developed by Box and Jenkins.

3.2 Basic Definitions and Concepts

A discrete time series is a sequence of observations on a variable process taken at equally spaced points in time. For illustration, Figure 3.1 shows representative plots of traffic volume and occupancy series observed at detector station 1 of the I-35 Freeway in Minneapolis. It is common to represent a time series by mathematical model expressing the process values as function of time, that is,

$$x_t = F_t + \varepsilon_t \quad (3.1)$$

where F_t is the value of a deterministic component at time t , and ε_t is a stochastic noise component. In general, the noise components are serially correlated, and therefore induce correlation in the x_t series. It is sufficient in many processes to assume that the series fluctuates about some fixed level μ , that is, F_t equals μ . However, trends over time can be represented by polynomial and/or cyclical terms.

An important concept in the analysis of time series is that of stationarity. Mathematically, a stationary time series is one for which the probability distribution of any $(K + 1)$ observations (x_t, \dots, x_{t-K}) is invariant with respect to time t . Consequently, any set of observations from a stationary time series will have the same mean value μ . Traffic time series as well as many other series encountered in practice are of the nonstationary type. In Figure 3.1 it is hard without additional prior information to confirm that any of the two series is stationary. Also, it is interesting to notice that the kind of nonstationarity displayed is homogeneous, in a sense that the behavior of each series at different periods in time is essentially the same. A homogeneous nonstationary time series can be reduced to a stationary

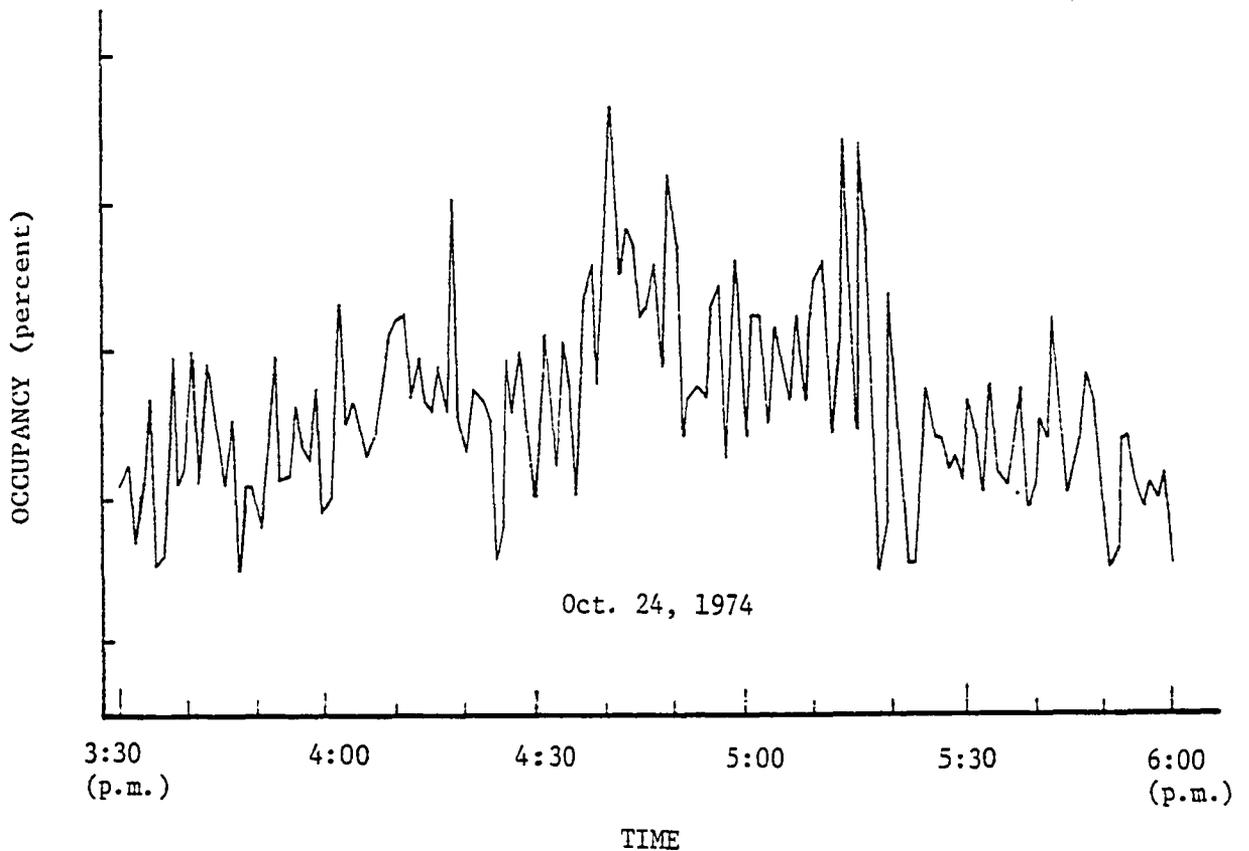
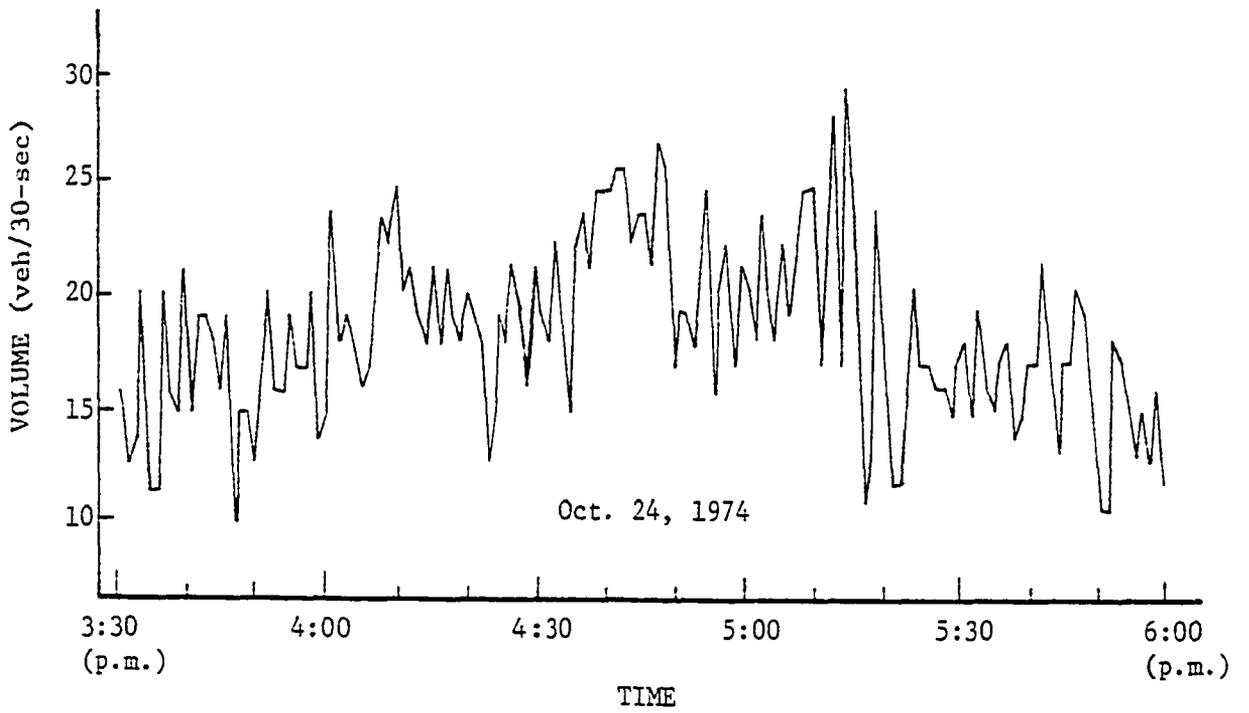


FIGURE 3.1

FREEWAY TRAFFIC VOLUME AND OCCUPANCY SERIES
(Minneapolis, I-35, Station 1)

form by means of differencing some finite number of times d . To help clarify the concept of differencing, it is useful to introduce the backward shift operator B which is explicitly defined by the relationship

$$B^d x_t = x_{t-d} \quad (3.2)$$

Another important operator is the backward difference operator $(1-B)$ defined by

$$(1-B)^d x_t = x_t - x_{t-d} \quad (3.3)$$

Figure 3.2 shows plots of the first differences (d equals one) of the volume and occupancy series presented in Figure 3.1. As one may readily notice, the temporal changes in the local level have been eliminated. In most real processes, stationarity can be achieved by one of the first three consecutive differences.

3.3 A Typology of Forecasting Techniques

The alternative approaches to constructing models for forecasting traffic systems can be categorized into four reasonably distinct categories. These are deterministic models, functional and structural models, ad hoc models, and time series analysis models. The following paragraphs briefly describe the characteristics of each of these categories.

3.3.1. Deterministic Models

In this class of models, the variable of interest is treated as a deterministic function of time so that there is only one possible outcome at each point in time. The forecasts are based on extrapolation, assuming that whatever factors produced the trend in the past will

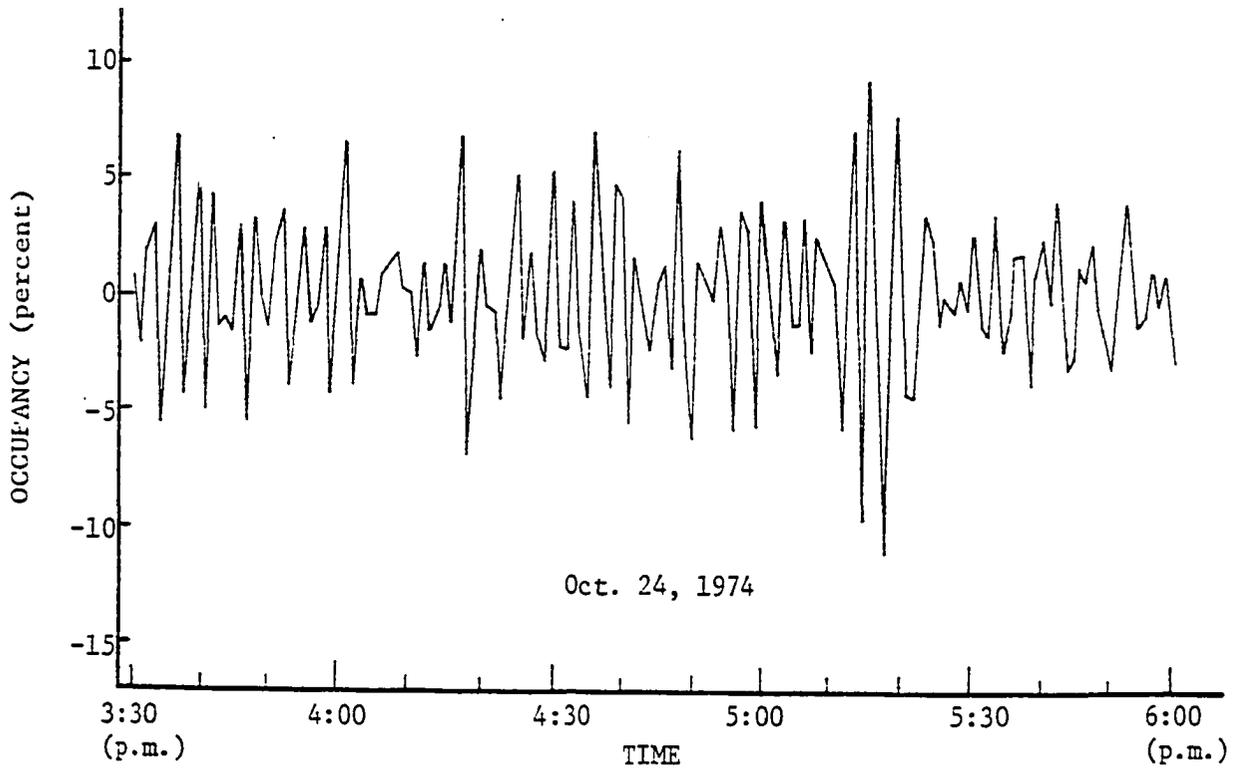
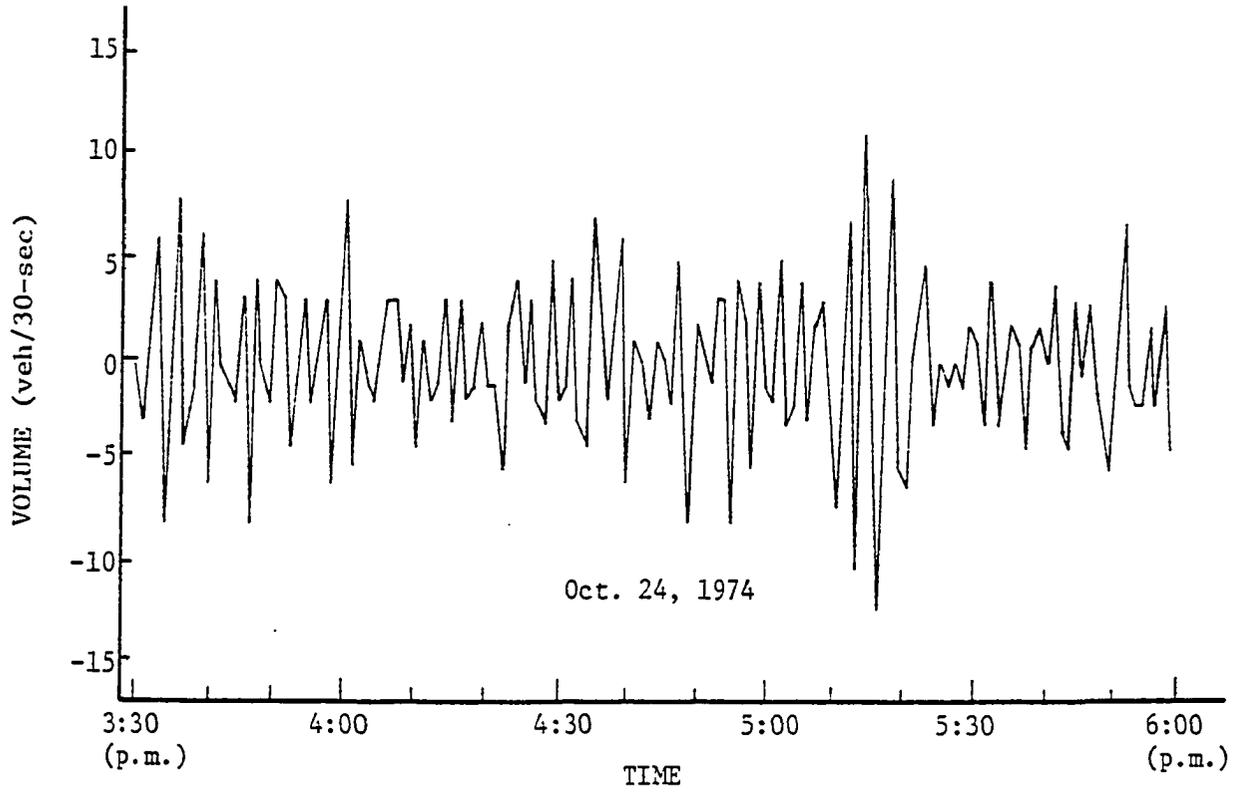


FIGURE 3.2

FIRST DIFFERENCES OF VOLUME AND OCCUPANCY SERIES
(Minneapolis, I-35, Station 1)

continue to produce the same trend in the future. The general form of a deterministic model is

$$x_t = f(t) \quad (3.4)$$

where $f(t)$ is some function of time. The major limitation of any deterministic model is the implication that the structure of the time series is perfectly systematic, and therefore completely predictable. Given the variety of factors which contribute to the behavior of any traffic system, it is almost impossible to postulate a sound deterministic model.

3.3.2 Functional and Structural Models

These models can be constructed by applying mathematical and statistical methods to a set of historical observations. The first stage in building such models is the specification of the model, that is, deciding upon the endogenous and exogenous variables describing the subject of the forecast. If x_t is the value of the endogenous variable at time t , and y_t , x_{t-1} , and ωt , are the exogenous variables, then the specified model has the general form

$$x_t = f(y_t, x_{t-1}, \cos \omega t) \quad (3.5)$$

The choice of the specific form of the function f is also part of the model specification. The next stage is the parameter inference or estimating the unknown coefficients from historical observations.

The theory of econometrics provides different estimation techniques when the classical assumptions of the linear regression model are violated. Consistent estimators can usually be found, but they are likely to have large variances, and consequently the precision

of the forecasts falls. In addition, the model itself may be subject to specification error, that is, the structure of the model is deficient in some respect. For example, the form of the model might not be linear as it is assumed.

3.3.3 Ad hoc Models

This class of models utilizes the past history of the variable of interest to provide forecasts of its future values. The general form of any ad hoc model is

$$\hat{x}_{t+l} = f_l(x_t, x_{t-1}, \dots) \quad (3.6)$$

where:

\hat{x}_{t+l} = forecast made at time t for l points ahead in time,

f_l = function of the past observations on x , and depends only on the lead time l .

There exist many forecasting models which can be characterized as ad hoc. The basic difference between these models is the weighting pattern assigned to current and past observations. Three major weighting schemes have been proposed in the literature: uniform, exponential, and adaptive exponential. The primary virtue of all ad hoc models is their ease of implementation and computational convenience. However, their major weakness stems from the inherent lack of a general approach for choosing among alternative schemes, and therefore they are characterized as ad hoc.

3.3.4 Time Series Analysis Models

The distinguishing feature of this final class of models is that the past history of a given process is viewed as a realization of jointly

distributed random variables, that is, any sequence of observations x_1, \dots, x_t is thought of as being drawn from a particular probability density function $p(x_1, \dots, x_t)$. The form of such probability function is determined by studying the structure of correlation among x 's displayed over time. As an example, if it is postulated that successive changes in a process value are generated independently from some probability distribution with mean zero, hence the process evolves according to the relationship

$$x_t = x_{t-1} + a_t \quad (3.7)$$

where a_t is a random variable with mean zero and constant variance. If x_0 is the starting value of the process, then the succeeding values are given by

$$\begin{aligned} x_1 &= x_0 + a_1 \\ x_2 &= x_0 + a_1 + a_2 \\ &\dots\dots\dots \\ x_t &= x_0 + a_1 + \dots + a_t \end{aligned} \quad (3.8)$$

The process described by equation 3.7 is known as the random-walk process.

Information about future values of a process can be obtained from the probability distribution which describes the process evolution. In particular, the first and second moments of the distribution of x_{t+l} given its history up to time t yield the mean and variance of the forecast made at time t for l points ahead in time.

3.4 Ad hoc Forecasting Models

This section presents a discussion of three ad hoc forecasting techniques: the moving average, the exponential smoothing, and the

Trigg and Leach models. These models will be used in the next chapter as a comparative basis for assessing the forecasting performance of time series analysis models.

3.4.1 Moving Average Model

The moving average at time t defined over the N previous observations is given by

$$m(t,N) = (1/N) \sum_{K=1}^N x_{t-K} \quad (3.9)$$

This model weighs each of the previous N observations by $1/N$, while the earlier observations have zero weight. The forecast of x_t is

$$\hat{x}_t = m(t,N) \quad (3.10)$$

Whitson, et al. [86], proposed the use of five-minute moving average of traffic volume data, (N equals five), with upper and lower threshold limits (twice the standard deviation of the five-minute observations) in incident detection on freeways. In a similar approach, Dudek, et al. [23] incorporated the moving average concept in their standard normal deviate model (SND) for incident detection. The SND is defined as

$$SND(t,N) = \frac{x_t - m(t,N)}{S(t,N)} \quad (3.11)$$

where $S(t,N)$ is the moving standard deviation at time t , and is given by

$$S(t,N) = \sqrt{\frac{1}{N} \sum_{K=1}^N x_{t-K}^2 - [m(t,N)]^2} \quad (3.12)$$

The value of SND measures the relative change in a traffic variable compared to the average trends which existed during previous time intervals.

3.4.2 Exponential Smoothing Models

The two basic exponential smoothing versions are the work of Winters [87] and Brown [9]. Single exponential smoothing as proposed by Brown assumes that F_t in equation 3.1 represents some equilibrium level, and the corresponding smoothing function is given by

$$S_1(t) = \alpha \cdot x_t + (1-\alpha) \cdot S_1(t-1) \quad (3.13)$$

where:

$S_1(t)$ = smoothed value of x at time t ,

α = smoothing constant, $0 < \alpha < 1$

The function $S_1(t)$ is a linear combination of all previous observations weighed by damped exponential weights. The forecast of x_t is

$$\hat{x}_t = S_1(t) \quad (3.14)$$

As an extension, the double exponential smoothing which is simply an exponential smoothing of S_1 assumes that F_t in equation 3.1 can be described by a linear trend, and the corresponding smoothing function is

$$S_2(t) = \alpha \cdot S_1(t) + (1-\alpha) \cdot S_2(t-1) \quad (3.15)$$

Brown [9] demonstrated that the steady-state response of exponential smoothing to a linear trend has a constant lag of $(1-\alpha)/\alpha$. Hence, the lag between the line through the observations and the line through S_1 is equal to the lag between the lines through S_1 and S_2 , and is equal to $b(1-\alpha)/\alpha$, as depicted in Figure 3.3. A forecast of the future observation at time $(t+l)$ would therefore be

$$\hat{x}_{t+l} = [2S_1(t) - S_2(t)] + [\alpha/(1-\alpha)][S_1(t) - S_2(t)] \cdot l \quad (3.16)$$

In operational forecasting it is evident from equation 3.16 that the only two pieces of information required to be stored are the values of S_1 and S_2 .

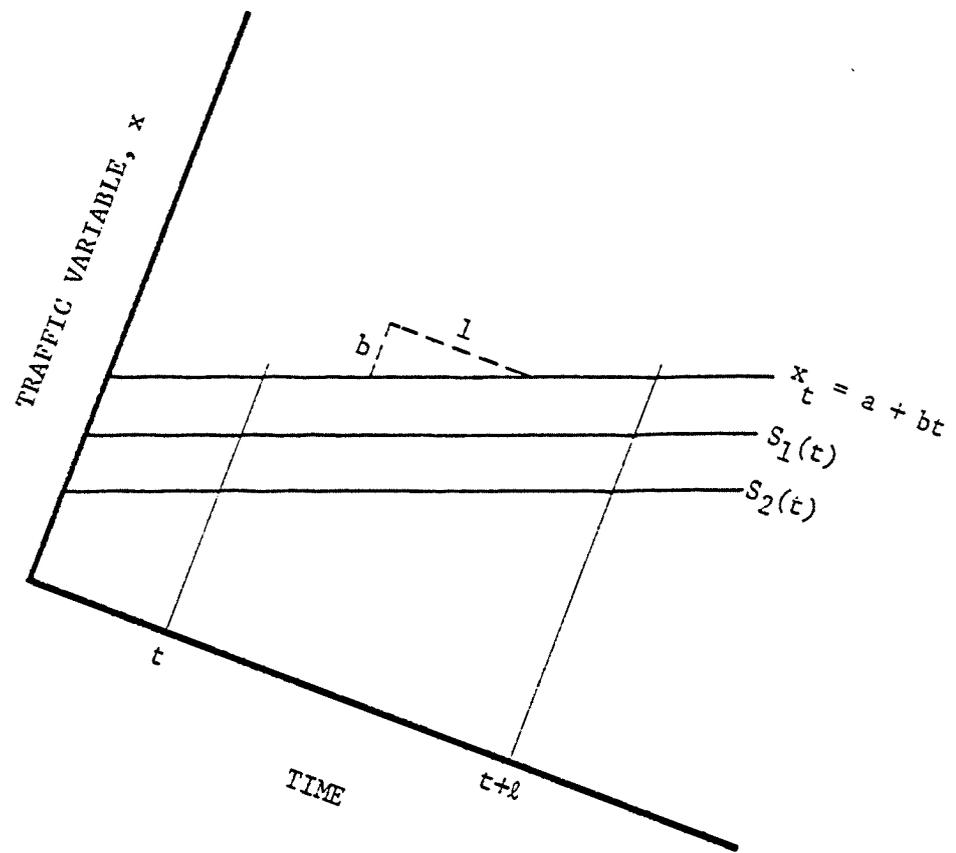


FIGURE 3.3
EXPONENTIAL SMOOTHING OF LINEAR TREND

To detect the presence of large forecast errors ($x_t - \hat{x}_t$), Brown introduced a "tracking signal" which reflects the magnitude of past cumulative errors. The conceptual idea is that if the sum of forecast errors departs significantly from what would be expected to occur at random, the tracking signal should indicate a change in the input observations not properly represented by the forecasting model. The tracking signal is defined by the relationship

$$TS(t) = \frac{Y(t)}{MAD(t)} \quad (3.17)$$

where:

$TS(t)$ = tracking signal at time t ,

$Y(t)$ = cumulative sum of errors at time t ,

$MAD(t)$ = mean absolute deviation of forecast errors at time t

An important property of the mean absolute deviation is that it is proportional to the standard deviation of the forecast errors. If these errors are normally distributed, and the noise components ϵ_t of equation 3.1 are serially independent, then the constant of proportionality is approximately 0.8.

Many applications of exponential smoothing have been proposed in the field of real-time traffic surveillance and control. Gazis and Knapp [31] applied the single exponential smoothing, with α equals 0.7 to speed measurements from three half-mile sections in the Lincoln Tunnel. Smoothed speeds were then used to predict travel time and traffic density. Most extensively tested are the predictor algorithms used in the first, second, and third generations of the Urban Traffic Control System programs (UTCS). Single exponential smoothing has been incorporated into these predictor algorithms to account for temporal

and spatial trends in traffic volume and occupancy in an urban street network [46,47,75]. Values of the smoothing constant α equal 0.5, 0.1, and 0.05 are used in the first-, second-, and third-generation UTCS predictors respectively. Tarnoff [75] discussed the inherent limitations in these predictor models and showed that the magnitude of the resulting errors had the potential for degrading the control system operations.

Cook and Cleveland [13] applied the double exponential smoothing with α equals 0.3 to thirteen traffic variables as a means of incident detection on freeways. They used the tracking signal of equation 3.17 to detect sudden changes in behavior of traffic time series data caused by incidents. Tignor [79] recommended a single exponential smoothing for incident detection under low volume conditions.

3.4.3 Exponential Smoothing with Adaptive Response

One major problem with exponential smoothing models is that of choosing a proper value of the smoothing constant α which remains unchanged over time. The role of the smoothing constant can be understood by studying how it affects the resulting forecasts. In the limiting case with α equals one, all the weight is given to current observation and the forecast would vary as widely as the input data. At the other extreme with α near zero, this will produce very stable forecasts regardless of any recent change in input data. Some studies have attempted to determine optimal smoothing constants based on minimizing the mean square error of the forecasts [9,58]. These studies,

however, assumed certain properties for the time series like stationarity or constant level which are unrealistic in practice.

Adaptive approaches have been suggested by many authors including Chow [11], Roberts and Reed [68], and Trigg and Leach [81]. The following is the adaptive approach proposed by Trigg and Leach:

$$\begin{aligned}
 TS(t) &= \frac{SE(t)}{SAE(t)}, \quad -1 < TS < 1 \\
 SE(t) &= \gamma \cdot e_t + (1-\gamma) \cdot SE(t-1) \\
 SAE(t) &= \gamma \cdot |e_t| + (1-\gamma) \cdot SAE(t-1) \\
 e_t &= x_t - \hat{x}_t
 \end{aligned}
 \tag{3.18}$$

where:

$$\begin{aligned}
 TS(t) &= \text{tracking signal at time } t, \\
 SE(t) &= \text{smoothed error at time } t, \\
 SAE(t) &= \text{smoothed absolute error at time } t, \\
 e_t &= \text{forecast error at time } t, \text{ and} \\
 \gamma &= \text{smoothing constant, } 0 < \gamma < 1
 \end{aligned}$$

Adaptive response is achieved by setting α equal to the absolute value of the tracking signal.

3.4.4 Summary of Ad hoc Models

Many models have been proposed for making short-term forecasts of traffic variables, however, a consensus as to which model is most appropriate has not emerged yet. Most of these models have been successfully applied to certain situations, but the results from one study are frequently contradicted by another. To make matters worse, the theoretical developments of these models have no real means of determining

which is best. The previous sections have attempted to summarize the types of ad hoc models which are commonly used, and to discuss their general properties. In using these models to forecast the performance of a traffic system, much is left to the personal judgement of the investigator or engineer, who must assume any special knowledge of such system.

3.5 Box and Jenkins ARIMA Models for Time Series Analysis

The use of stochastic linear models to describe the dynamic behavior of discrete time series dates back to the early twentieth century [91,92,72]. Yule [92] in his analysis of sunspot series introduced the idea of random shocks which is basic to the theory of stochastic processes. A rigorous probabilistic foundation for stochastic time series properties was provided by Kolmogorov [45], while a common terminology is due to Wold [88].

In the late 1960's, Box and Jenkins [6,7] formulated a general approach for time series analysis, forecasting, and control. The approach proposes a class of autoregressive integrated moving average, ARIMA, models. It also suggests a strategy for identification and estimation of an appropriate model for a particular series. In the following sections, the ARIMA models are discussed with their identification and estimation techniques. The utility of the fitted models in forecasting, control, and simulation is also described.

3.5.1 Models for Stationary Time Series

The basic premise in building a model for the stochastic behavior of a time series is that ϵ_t of equation 3.1 can be represented

by a linear combination of a white noise sequence $a_t, a_{t-1}, a_{t-2}, \dots$, where the a 's are independent variables with mean zero and variance σ_a^2 . As illustrated in Figure 3.4 the white noise can be thought of as an input to a linear filtering operation which transforms the a 's to an output series ε_t whose successive values are correlated. The observation x_t in equation 3.1 can therefore be written as

$$\begin{aligned} x_t &= \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \\ &= \mu + \psi(B) a_t \end{aligned} \quad (3.19)$$

where F_t is replaced by μ since the series x_t is stationary, and $\psi(B)$ is the transfer function of the filter relating a 's to ε 's. It turns out that the major task in modeling a given time series is to choose an appropriate linear filter which generates a correlation sequence similar to that observed in the series. In particular, the class of ARIMA models utilizes three linear filtering operations as depicted in Figure 3.5. The transfer functions of these filters can be expressed as the ratio of finite polynomials in B of the general form

$$\psi(B) = \theta_q(B) / \phi_p(B) \quad (3.20)$$

where $\theta_q(B)$ is the transfer function of the moving average component of the model and $\phi_p(B)$ is the transfer function of the autoregressive component.

Moving Average Processes. A finite moving average process of order q or just MA(q) expresses ε_t of equation 3.1 as a linear combination of $a_t, a_{t-1}, \dots, a_{t-q}$, that is,

$$\psi(B) = \theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (3.21)$$

Hence, the process value at time t can be written as

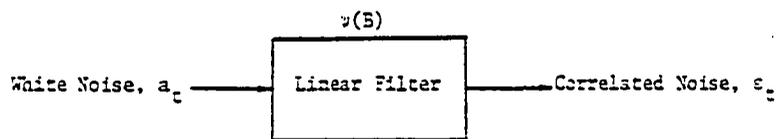


FIGURE 3.4

REPRESENTATION OF CORRELATED NOISE
AS OUTPUT FROM LINEAR FILTER

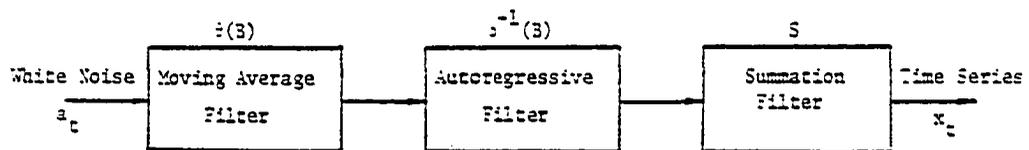


FIGURE 3.5

REPRESENTATION OF ARIMA MODELS AS THREE FILTERS

$$x_t = \mu + a_t^{-\theta_1} a_{t-1}^{-\theta_2} a_{t-2}^{-\dots-\theta_q} a_{t-q} \quad (3.22)$$

The name "moving average" is somewhat misleading, because the weights $1, -\theta_1, -\theta_2, \dots, -\theta_q$ need not sum to unity nor need they be positive. However, this nomenclature is of common use, and therefore it is pursued in the rest of this dissertation.

An MA(q) process is fully described by (q+2) parameters; $\sigma_a^2, \mu, \theta_1, \dots, \theta_q$, which in practice need to be estimated from the observed series. The mean and variance of MA(q) process are given by

$$E(x_t) = \mu$$

$$\gamma_0 = \sigma_a^2 \sum_{i=0}^q \theta_i^2, \quad \theta_0 = 1$$

Also, by multiplying equation 3.22 by x_{t-K} and taking expected value, the autocovariance function is

$$\gamma_K = \sigma_a^2 (-\theta_K + \theta_1 \theta_{K+1} + \dots + \theta_{q-K} \theta_q)$$

for $K = 1, 2, \dots, q$ and zero for $K > q$.

The autocorrelation function is obtained by dividing γ_K by γ_0 , that is,

$$\rho_K = \begin{cases} (-\theta_K + \theta_1 \theta_{K+1} + \dots + \theta_{q-K} \theta_q) / \sum_{i=0}^q \theta_i^2 & , K = 1, 2, \dots, q \\ 0 & , K > q \end{cases}$$

It follows that the autocorrelation function of an MA(q) process consists of q spikes at lags $1, \dots, q$, and zero thereafter.

Autoregressive Processes. A finite autoregressive process of order p or just AR(p) expresses ε_t of equation 3.1 as a linear combination of $a_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-p}$. The inverted transfer function has the form

$$\psi^{-1}(B) = \phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (3.23)$$

Hence, the process value at time t can be written as

$$x_t = \mu + a_t + \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \dots + \phi_p(x_{t-p} - \mu)$$

or

$$x_t = \delta + a_t + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} \quad (3.24)$$

where δ is a constant whose value is a function of μ and the ϕ_i weights.

The name "autoregressive" comes from the fact that the model of equation 3.24 is essentially a regression model in which x_t is regressed on its own past values as independent variables.

An AR(p) process contains ($p+2$) parameters; σ_a^2 , μ , ϕ_1, \dots, ϕ_p which in practice need to be estimated from the observed series. The autocorrelation function of an AR(p) process is given by the difference equation

$$\rho_K = \phi_1 \rho_{K-1} + \phi_2 \rho_{K-2} + \dots + \phi_p \rho_{K-p}, \quad K > 0$$

which has a general solution of the form

$$\rho_K = C_1 G_1^{-K} + C_2 G_2^{-K} + \dots + C_p G_p^{-K}$$

where the C_i 's are constants, and G_1, G_2, \dots, G_p are the p roots of the characteristic equation $\phi_p(B) = 0$. For real values of ϕ_i each root G_i will be either real or one of a complex conjugate pair. Real roots contribute a damped exponential or geometrical decay to the autocorrelation function as the lag K increases, whereas a complex pair of roots contribute a damped sine wave. It follows that the autocorrelation function of an AR(p) process consists of a mixture of damped exponentials and/or damped sine waves. Unlike the moving average process,

there is no cut-off at a certain lag, a fact which is important in identifying an unknown process.

Mixed Autoregressive-Moving Average Processes. In some cases, a process is neither strictly of the moving average or autoregressive type, and the model must include both AR and MA terms. A mixed autoregressive-moving average model of order (p,q) or just ARMA (p,q) expresses ϵ_t of equation 3.1 as a linear combination of $a_t, a_{t-1}, \dots, a_{t-q}, \epsilon_{t-1}, \dots, \epsilon_{t-p}$, and the transfer function is a ratio of two polynomials of the form

$$\psi(B) = (1 - \theta_1 B - \dots - \theta_q B^q) / (1 - \phi_1 B - \dots - \phi_p B^p) \quad (3.25)$$

Hence, the process value at time t can be written as

$$x_t = \delta + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (3.26)$$

An ARMA (p,q) process has $(p+q+2)$ parameters, $\sigma_a^2, \mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ which in practice need to be estimated from the observed series. The autocorrelation function of an ARMA (p,q) process satisfies the difference equation

$$\rho_K = \phi_1 \rho_{K-1} + \phi_2 \rho_{K-2} + \dots + \phi_p \rho_{K-p}$$

except for the first q lags which depend directly on the moving average parameters and provide the necessary starting values. It follows that if p is greater than q , then the autocorrelation function of an ARMA (p,q) will consist of a mixture of damped exponentials and/or damped sine waves whose nature is dictated by the polynomial $\phi(B)$ and the starting values. Whereas, if q is greater than p , then there will be $(q-p-1)$ initial autocorrelations which do not follow this pattern.

3.5.2 Stationarity and Invertibility Requirements

A finite moving average process of order q can be expressed as an infinite autoregressive process. Also, a finite autoregressive process of order p can be expressed as an infinite moving average process. In general, this duality can be represented as

$$\phi_p(B) = \theta_q^{-1}(B) \quad (3.27)$$

which implies that the two transfer functions of AR(p) and MA(q) processes should be invertible. Box and Jenkins [7] discussed the stationarity and invertibility requirements of linear stochastic processes in relation to the convergence of the two polynomials in B

$$\begin{aligned} \phi_p(B) &= 0, \\ \theta_q(B) &= 0 \end{aligned}$$

which is satisfied if B lies on or within the unit circle, that is, $|B| \leq 1$.

3.5.3 Models for Nonstationary Time Series

The application of a finite difference filter to homogeneous nonstationary processes reduces them to a stationary form. The transfer function of a difference filter of order d is

$$D(B) = (1-B)^d$$

It follows that the class of models discussed for stationary processes can be utilized to represent differenced nonstationary processes.

A general model for expressing stationary and nonstationary time series is called an autoregressive integrated moving average process

of order (p,d,q) or just $ARIMA(p,d,q)$. The difference form of this model is

$$\phi_p(B)(1-B)^d(x_t - \mu) = \theta_q(B)a_t \quad (3.29)$$

which expresses the process value at time t in terms of previous process values and current and past white noise variables. Equation 3.29 is known as the basic Box-Jenkins model for non-seasonal time series.

Box and Jenkins demonstrated that the ad hoc forecasting models are derivatives of the general ARIMA model, and that they are appropriate for particular types of nonstationary processes. For example, the single exponential smoothing model of equation 3.13 is equivalent to an ARIMA $(0,1,1)$ process with the smoothing constant α set equal to θ_1 .

3.5.4 Model Fitting Procedure

ARIMA models are fitted to a particular time series data by a three-stage iterative procedure, preliminary identification, estimation, and diagnostic checking. The various steps of the procedure are outlined below.

Model Identification. The values of p,d , and q of equation 3.29 are determined by inspecting the autocorrelation and partial autocorrelation functions of the observed series and/or its differences, and comparing them with those of some basic stochastic processes. The estimated autocorrelation function of a sample x_1, x_2, \dots, x_n is given by

$$r_K = \frac{\sum_{t=1}^{n-K} (x_t - \bar{x})(x_{t+K} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}, \quad K = 1, 2, \dots \quad (3.30)$$

where:

\bar{x} = sample mean,

n = number of observations

The autocorrelation function provides a measure of how long a disturbance to a process at some point in time affects the state of this process in the future. The variance of the sample autocorrelation estimates is approximately

$$\text{Var}(r_K) \approx (1/n) \left[1 + 2 \sum_{i=1}^{K-1} r_i^2 \right] \quad (3.31)$$

Figure 3.6 illustrates the theoretical autocorrelation functions of typical first and second order MA and AR processes. In general, the autocorrelation function of an MA(q) process has significant spikes only at the first q lags (memory of lag q). Conversely, if the true process is AR(p), the autocorrelation function tails off in the form of damped exponentials and/or damped sine waves. Failure of the autocorrelation function to die out rapidly suggests that differencing is needed ($d > 0$).

The sample partial autocorrelation function provides another identification tool, particularly for the order of an AR process. The value of this function at lag K , ϕ_{KK} , measures the dependence of x_t on x_{t-K} given $x_{t-1}, x_{t-2}, \dots, x_{t-K+1}$, and can be thought of as the last coefficient in an autoregressive model of order K fitted to the time series. In practice, the partial autocorrelation function is estimated by fitting successively AR processes of order $1, 2, \dots$ by least squares and picking out the estimates of the last coefficient at each stage. If the true process is AR(p), then estimated partial autocorrelations beyond lag p are approximately independent with mean zero and variance $1/n$, an

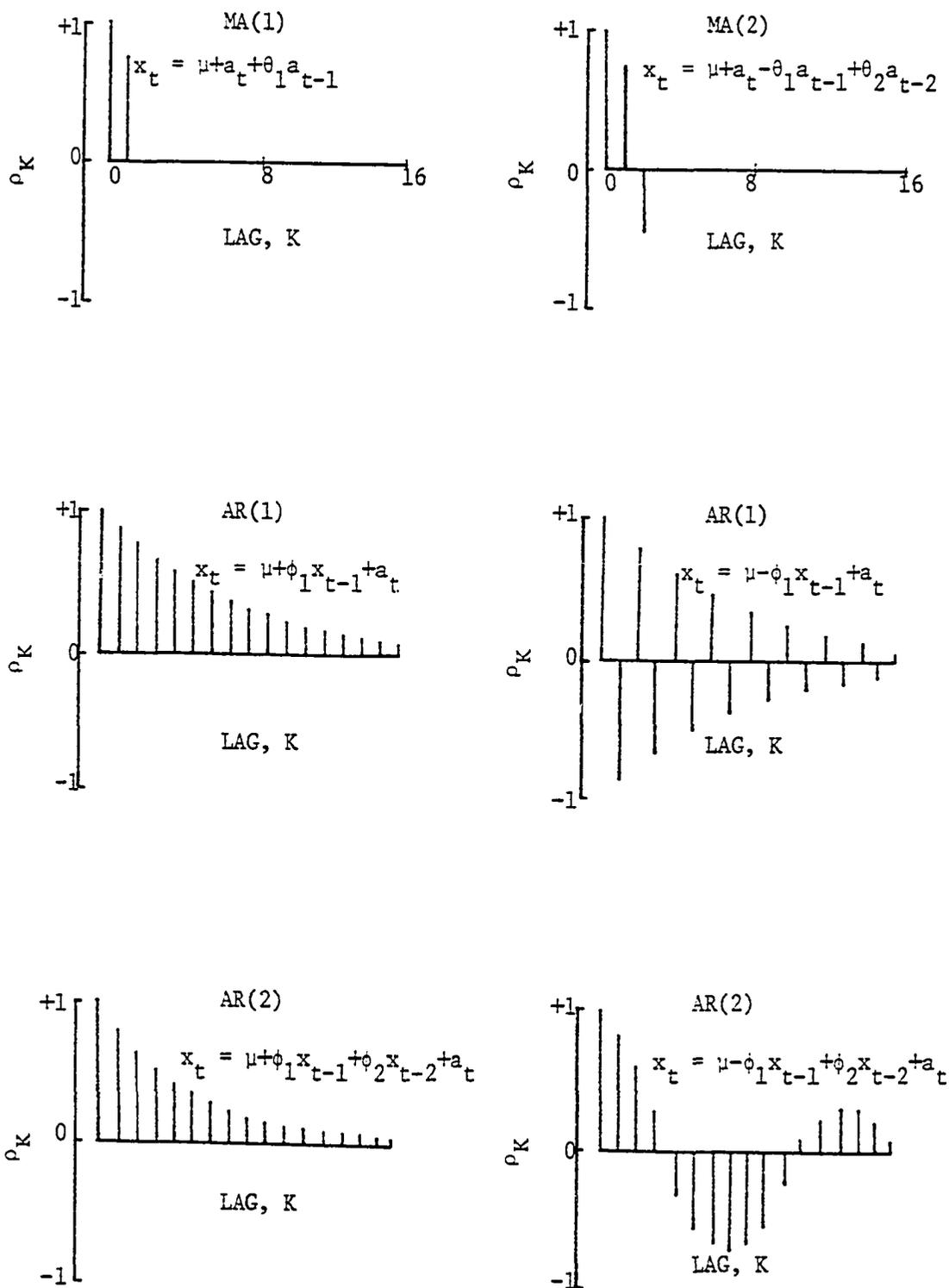


FIGURE 3.6

THEORETICAL AUTOCORRELATION FUNCTIONS OF BASIC PROCESSES

important clue for identifying p . For moving average processes, ϕ_{KK} decays as a mixture of damped exponentials and damped sine waves.

Parameter Estimation. Box and Jenkins [7] demonstrated that a close approximation to the maximum likelihood estimates of the parameters $\phi_1, \phi_2, \dots, \phi_p$ and $\theta_1, \theta_2, \dots, \theta_q$, assuming that a 's are normally distributed can be obtained by minimizing the sum of squared residuals

$$SS(\hat{\phi}, \hat{\theta}) = \sum_t a_t^2(\hat{\phi}, \hat{\theta}) \quad (3.32)$$

In practice, iterative non-linear least squares subroutines are used to obtain estimates of the parameters, their approximate standard errors, and an estimate of σ_a^2 . These subroutines require only the provision of initial guess of the parameter values.

Diagnostic Checking. Several tests of model adequacy have been proposed [7], and they are based upon detecting departures from randomness among residuals. If the form of the fitted model is correct, and if the parameter estimates are close to their true values, then the estimated residuals \hat{a}_t should be uncorrelated random deviates. A simple test of model adequacy is to examine the autocorrelation function of the residual series which should not display any noticeable structure.

Rather than considering the individual autocorrelations of residuals, Box and Pierce [8] developed an overall test for a whole set of residual autocorrelations for lags one through K . They showed that the variable

$$Q = n \sum_{i=1}^K r_i^2(\hat{a}) \quad (3.33)$$

where:

n = number of observations minus the order of differencing in the postulated model,

$\hat{r}_i(a)$ = residual autocorrelation for lag i

is approximately distributed as $\chi^2(K-p-q)$.

Other tests to determine the normality of residuals must also be carried out. Non-normality of residuals may require transformation of the original series to natural logs or the like. If the postulated model proves inadequate, then a new model is proposed and fitted. A model cannot be useful for forecasting or simulation before diagnostic checks are satisfied.

3.5.5 Forecasting and Simulation

If an ARIMA model passes the diagnostic checks, then it can be used to forecast future values of the fitted time series. Box and Jenkins [7] demonstrated that the minimum mean square error forecast of x_{t+l} made at time t for l points ahead in time (written as $\hat{x}_t(l)$) is given by

$$\hat{x}_t(l) = E(x_{t+l}/H_t) \quad (3.34)$$

where $E(x_{t+l}/H_t)$ is the expected value of x_{t+l} given the past history H_t of the series up to time t . Conditional expected values are calculated from the fitted model by replacing unknown a 's by their expected value which is zero. For demonstration, consider the general linear process of equation 3.19. The one-step ahead forecast made at time $(t-1)$ can be written as

$$\begin{aligned} \hat{x}_{t-1}(1) &= E(x_t/H_{t-1}) \\ &= \mu + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \end{aligned} \quad (3.35)$$

Hence, the one-step ahead forecast error at time $(t-1)$ is

$$e_{t-1}(1) = x_t - \hat{x}_{t-1}(1) = a_t \quad (3.36)$$

which means that the white noise variables generating the process turn out to be the one-step ahead forecast errors.

The extension to forecasting l units ahead in time can be defined in terms of the transfer function $\psi(B)$ as

$$\begin{aligned} \hat{x}_t(l) &= B^l \psi(B) a_t \\ &= \psi_l a_t + \psi_{l+1} a_{t-1} + \dots \end{aligned} \quad (3.37)$$

The forecast error for l units ahead in time is

$$e_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \dots + \psi_{l-1} a_{t+1} \quad (3.38)$$

which has expected value zero and variance

$$\text{Var}[e_t(l)] = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{l-1}^2) \sigma_a^2 \quad (3.39)$$

Confidence intervals for the forecasts can be constructed assuming that the forecast errors are normally distributed. For a specified probability level α , the confidence limits are given by

$$\hat{x}_t(l) \pm Z_{\alpha/2} (1 + \sum_{i=1}^{l-1} \psi_i^2)^{1/2} \cdot S_a \quad (3.40)$$

where S_a is an estimate of the standard deviation of the white noise variables, and $Z_{\alpha/2}$ is the deviate exceeded by a proportion $\alpha/2$ of the area under the standardized normal distribution.

Forecasts of any ARIMA process can be simply and efficiently updated over time. Box and Jenkins [7] showed that the eventual shape of the forecasting function of equation 3.37 can be written in the form

$$\begin{aligned}\hat{x}_t(\lambda) &= b_0^{(t)} \cdot f_0(\lambda) + b_1^{(t)} \cdot f_1(\lambda) + \dots + b_{p+d-1}^{(t)} \cdot f_{p+d-1}(\lambda) \\ &= \sum_{i=0}^{p+d-1} b_i^{(t)} \cdot f_i(\lambda) \quad , \lambda > (q-p-d)\end{aligned}\quad (3.41)$$

where the terms $f_i(\lambda)$ are functions of the lead time λ , and $b_i^{(t)}$ are updating coefficients whose values change from point in time to the next. For values of λ less than or equal to $(q-p-d)$, and q greater than $(p+d)$, the forecasting function will have additional terms containing a 's, so that

$$\hat{x}_t(\lambda) = \sum_{i=0}^{p+d-1} b_i^{(t)} \cdot f_i(\lambda) + \sum_{i=0}^j c_{\lambda i} \cdot a_{t-i} \quad , \lambda \leq (q-p-d) \quad (3.42)$$

where j equals $(q-p-d-\lambda)$, and the c 's are coefficients obtained by substituting equation 3.42 in the difference equation 3.29.

A fitted ARIMA model can also be utilized to generate a simulated time series with the same stochastic properties as the observed series. The only requirement is to generate a stream of random normal deviates with zero mean and appropriate variance and passing it through the estimated linear filter. Simulated time series of traffic variables have many potential applications in evaluating the performance of complex traffic systems without the need to observe these systems for long periods of time. This reduces the cost and manpower required to conduct traffic

studies and enables running through a variety of operational designs in a very small fraction of the time required to try each on the physical system.

CHAPTER IV

MODELING FREEWAY TRAFFIC TIME SERIES DATA

4.1 Introduction

The analysis and modeling of time series of traffic variables has been the subject of much conjectural and experimental research work in recent years. This has been motivated by the need for better understanding of the dynamic behavior of traffic systems, and by practical problems of prediction and control. However, the literature which have come forth on this topic indicates that many theoretical developments are yet to be accomplished.

In this chapter, the Box and Jenkins approach is applied to traffic time series data obtained from freeway surveillance systems in Los Angeles, Minneapolis, and Detroit. Basically, a forecasting model is developed for traffic volumes and occupancies. The chapter begins with a brief summary of previous work on modeling freeway traffic time series data. There is then a description of the data base used in model development, followed by a detailed discussion of the modeling procedure. In addition, the model utility in operational forecasting and simulation is discussed. Finally, the developed model is evaluated on a comparative basis with the ad hoc forecasting models presented in chapter three.

4.2 Summary of Previous Work on Traffic Time Series Analysis

In transportation literature, two analysis techniques have been commonly applied to modeling time series of traffic variables. These are spectral analysis and discrete time series analysis. Spectral analysis as discussed by Jenkins and Watts [43] is a descriptive tool which examines the data stream in the frequency domain, whereas discrete time series analysis examines the data stream in the time domain. Mathematically, the two techniques are equivalent since the spectral density function and the autocorrelation function are the Fourier transforms of each other. However, spectral analysis is particularly relevant to modeling time series which exhibit periodic behavior and relatively long-term trends.

One of the earliest works on spectral analysis of traffic time series is that done by Bartlett [4] who analyzed time-headway observations from a rural two-lane road in Sweden. He recognized the tendency of vehicles to form platoons when traveling at constrained speeds. Cox and Lewis [16] used the same data set to fit branching renewal models to time-headway series. A renewal process is a series of events in which the times between events are independently and identically distributed. In a more recent paper, Bartlett [5] continued his line of thought discussing the cross-spectral analysis of volume and speed series on a three-lane roadway.

Probably, most of the research work on traffic time series analysis was conducted at the National Proving Ground for Freeway Surveillance, Control, and Electronic Traffic Aids in the John C. Lodge Freeway Corridor in Detroit. Mika et al. [54], analyzed one-minute average speed observa-

tions from four locations at the center lane of this facility. They distinguished between two modes of traffic flow, an oscillation-free mode and an oscillatory mode with the transition between these two modes near the maximum flow value. By means of spectral analysis technique they calculated the frequency of oscillation which was found to be 1/4 cycle per minute. Furthermore, they estimated the propagation velocity of speed shock waves using cross-autocorrelation between pairs of time series, and found it to be approximately 16 mph. They concluded that this is the same velocity that could be obtained using Lighthill and Whitham's kinematic wave theory [50]. Herman and Rothery [38] investigated speed records of two instrumented vehicles moving with traffic in the center lane of the same corridor. They found that high-frequency speed oscillations were more readily absorbed in the platoons of vehicles than low-frequency oscillations. More recently, Lam and Rothery [48] performed an analysis of average speed time series from four adjacent locations in the center lane of the same freeway. Their results revealed periodic flow patterns with dominant frequencies of 1/6 to 1/3 cycles per minute. They also estimated the propagation velocity of speed shock waves and found it to be 12 mph.

In addition, Darroch and Rothery [17] applied the cross-spectral analysis to car-following data in order to explain the dynamic behavior of a freeway traffic stream. Also, Nicholson and Swann [59] employed the same technique to make short-term forecasts of traffic volumes in tunnels. Their analysis was based on traffic volume observations from the Mersey Tunnel in Liverpool, England. This application is particularly important to scheduling tunnel ventilation systems and improving traffic operations in tunnels.

The second analysis technique, that is, discrete time series analysis of traffic variables has been conducted on many occasions. Welding [84], in discussing the influence of delays on regularity of bus operations in London noted that traffic conditions had little effect on bus services because of the degree of autocorrelation in journey times, and partly because of the efforts of drivers in correcting irregularities. Of particular interest is his model for time series of journey times in which measurements are assumed to vary about constant mean levels but subject to random shocks with mean zero. Wright [90], postulated a first-order autoregressive (Markovian) model for traffic volumes and densities in discussing volume-density relationships. He recognized the existence of bias in regression coefficients when time-aggregated variables are used. Hillegas et al. [42], examined time series of traffic volume and occupancy from the John F. Kennedy freeway in Illinois. They distinguished between two regimes of traffic stream behavior and proposed a stationary first-order autoregressive model. Their results indicated random behavior for light traffic (occupancy less than 15 percent), and strong autoregressive behavior for heavy traffic. However, they did not estimate the parameters of any model or perform statistical tests. Furthermore, they concluded that "a higher order autoregressive process or integrated autoregressive moving average process may be occurring that cannot be analyzed using only autocovariance techniques."

More elaborate models for investigating the dynamic nature of traffic time series data have been discussed by Polhemus [64], Der [18], and Eldor [27] based on the techniques formulated by Box and Jenkins. Polhemus explored the application of discrete finite parameter time

series models to describe the fluctuations in air traffic flow operations. As a case study, he analyzed an eight-hour sample of air traffic in the local control sector at the New York's LaGuardia Airport. He found that a second-order autoregressive model is adequate in describing the dynamic behavior of aircraft landing series after separating the local fluctuations. Der investigated traffic occupancy time series from the Dan Ryan Freeway in Chicago, and suggested an ARIMA (1, 0, 1) model. However, he reported that a higher order ARIMA process may be a better candidate model. The problem with an ARIMA (1, 0, 1) model is that it assumes stationarity of traffic time series which is not always true. Finally, Eldor analyzed 5-minute aggregations of volume time series data which were collected during the morning peak-period at the Santa Monica Freeway in Los Angeles. He proposed an ARIMA (0, 1, 1) model as a predictor for traffic volumes and suggested its application in determining traffic responsive control strategies for freeway entrance ramps. Importantly, the ARIMA (0, 1, 1) model is equivalent to Brown's single exponential smoothing model [7]. He also evaluated the ARIMA (0, 1, 1) model on a comparative basis with the predictor models used with the second and third generations of the Urban Traffic Control System programs. It should be pointed out, however, that the data used by Eldor in identifying the form of his ARIMA (0, 1, 1) model and in estimating the model parameters was not the same data which he used in evaluating the forecasting performance of the ARIMA (0, 1, 1) model.

4.3 Freeway Traffic Surveillance Data Base

The data used in this research work consist of freeway traffic volume and occupancy compilations recorded at electronic surveillance

systems in Los Angeles, Minneapolis, and Detroit. In these systems, point detectors communicate vehicle presence information to a central computer facility in the form of electrical signals. These signals are electronically processed and aggregated in real-time so as to provide averages over a specific time period, for example, twenty-second average volume. Table 4.1 summarizes the data sources and types. In all, a total of 166 time series data sets representing more than 27,000 minutes of observations were used in the analysis.

The Los Angeles and Minneapolis data were collected by the California Department of Transportation and the Minnesota Highway Department, respectively. In particular, the collection period lasted from January 1974 to February 1976. These data are furnished by the National Technical Information Service in the form of three magnetic tapes. The contents and organization of these tapes are described in [62]. Generally, the Los Angeles data are twenty-second averages of volume and occupancy per lane updated every twenty seconds, while the Minneapolis data are thirty-second averages of volume and occupancy aggregated over lanes and updated every thirty seconds.

Figure 4.1 shows the Los Angeles data collection site which contains a 42-mile freeway loop of segments of the Santa Monica Freeway, the San Diego Freeway and the Harbor Freeway in Metropolitan Los Angeles. These three freeways are all heavily traveled and they encounter substantial variations in traffic volume and occupancy. The detection system consists of induction loop detectors instrumented at one-half mile intervals in lane one (median lane), and lane three. In addition, all lanes are instrumented with detectors at three-mile intervals. For

TABLE 4.1
DATA SOURCES AND TYPES

Freeway System	Detection Hardware	Data Description				
		Type	Aggregation Interval (Sec.)	Total Intervals per Set	No. of Data Sets	Observation Date
Los Angeles Area Freeway Surveillance and Control Project (LAAFSCP)	Induction Loop	Volume (agg. over lanes)	60	175	10	April 23, 74
		Volume (per lane)	60	175	30	April 23, 74
		Volume (per lane)	20	525	30	April 23, 74
		Occupancy (agg. over lanes)	60	175	10	April 23, 74
		Occupancy (per lane)	60	175	30	April 23, 74
		Occupancy (per lane)	20	525	30	April 23, 74
Minneapolis (Interstate-35 S.)	Induction Loop	Volume (agg. over lanes)	30	150	10	Oct. 24, 74
		Occupancy (agg. over lanes)	30	150	10	Oct. 24, 74
Detroit (Lodge Freeway)	Ultra-Sonic	Volume (agg. over lanes)	60	260	2	Dec. 3, 69
		Occupancy (agg. over lanes)	60	121 260	2 2	Dec. 23, 68 Dec. 3, 69

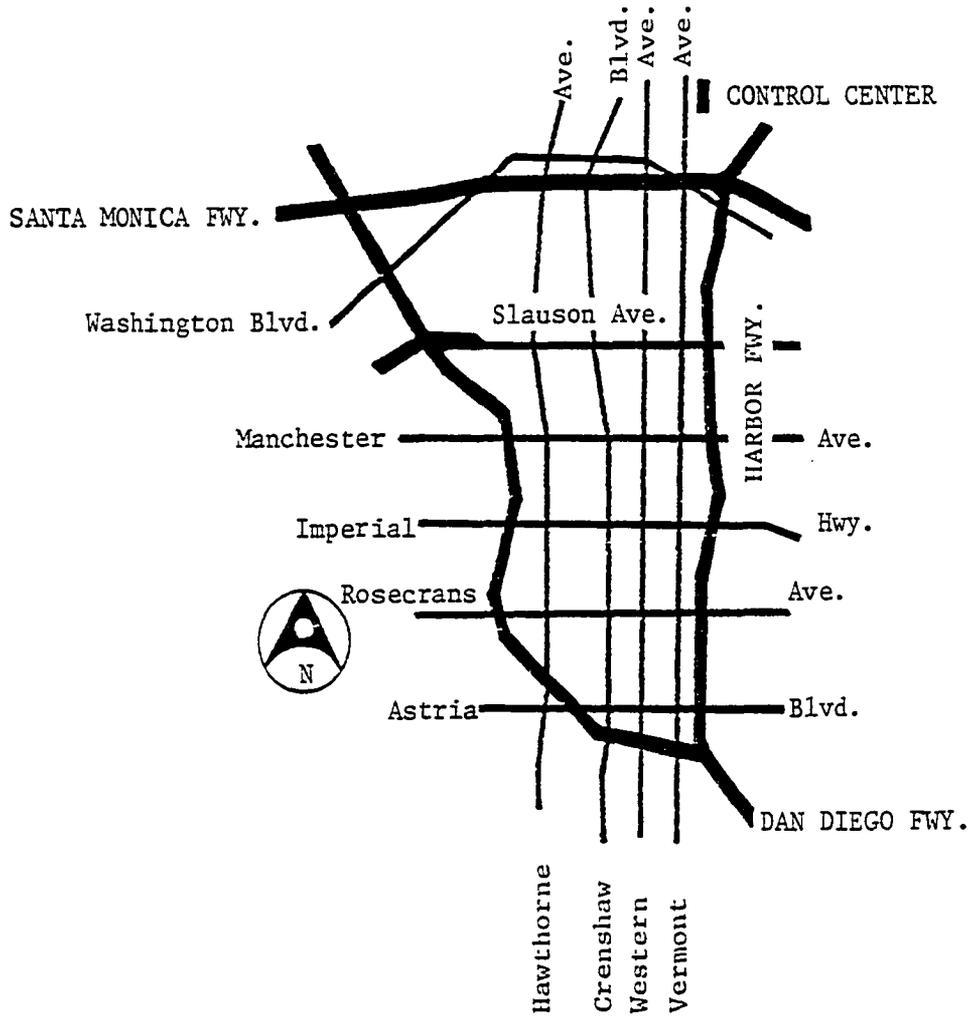


FIGURE 4.1

SCHEMATIC REPRESENTATION OF THE 42-MILE FREEWAY LOOP
(Los Angeles, California)

ramp metering purposes, detectors are also instrumented at exit and entrance ramps. The Minneapolis data were collected from a 28-mile section of the I-35 freeway, shown in Figure 4.2, which represents a portion of the Minneapolis freeway surveillance system. In this freeway section, induction loop detectors are instrumented at one-half mile intervals in all lanes, and at entrance and exit ramps.

Other data from Detroit were collected from an eight-mile research section of the John C. Lodge Freeway Corridor, shown in Figure 4.3, and are described in [13]. The collection period ranged from December 1968 to December 1969 when this section was under continuous television surveillance. In addition to fourteen television cameras, the detection system contained a closed network of ultrasonic vehicle presence detectors placed at variable spacings on the freeway and at all entrance and exit ramps. These detectors represented the source of traffic data which were recorded in the form of one-minute averages of volume and occupancy aggregated over all lanes and updated every one minute.

In summary, it is believed that the data utilized in this research work represent an extensive source for modeling and analyzing time series of traffic variables. However, one point to be made regarding the limitations of the data base is that the collection period covered only few hours in one single day. Some minor problems were encountered with the California and Minneapolis data which were available in binary character. This was due to the inadequate preparation of the magnetic tapes and the confusing description of the data records. Upon completion of the data preparation effort, further aggregation over different time intervals and/or over lanes was accomplished.

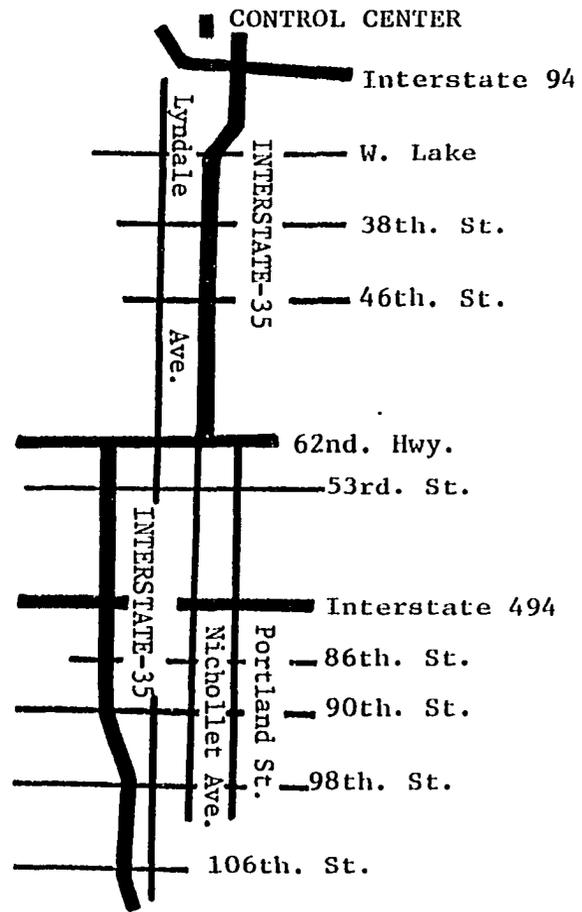


FIGURE 4.2

INTERSTATE-35 FREEWAY CORRIDOR
(Minneapolis, Minnesota)

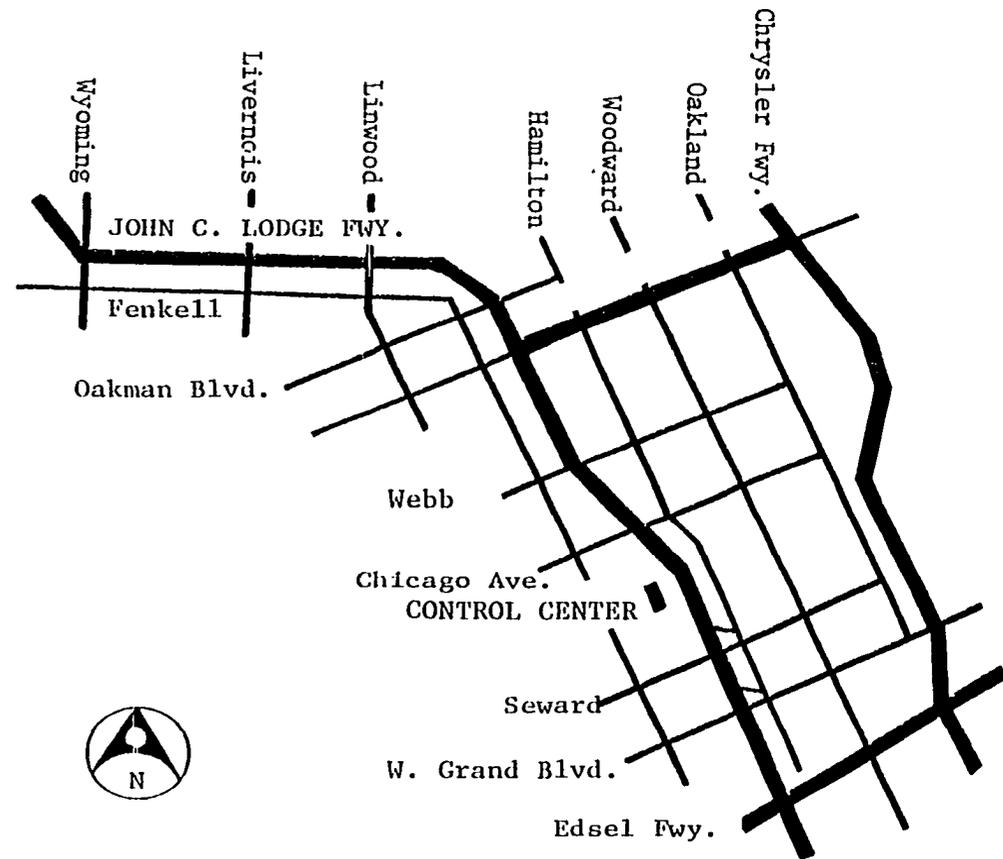


FIGURE 4.3

JOHN C. LODGE FREEWAY CORRIDOR
(Detroit, Michigan)

4.4 ARIMA Model Development

Three computer programs entitled PDQ, ESTIMATE and FORECAST were used in this research to perform the computations required by the Box-Jenkins technique. These programs are written in FORTRAN IV for the IBM 360 computer and are described in [57]. Program PDQ provides estimates of the sample statistics required for identification of a tentative model or models, and requires a priori specification of orders of consecutive differencing, d . Storage requirement for this program is 90K for a maximum of 1,000 observations. In program ESTIMATE, values of the autoregressive parameters and moving average parameters are estimated utilizing nonlinear iterative least squares technique. The estimation logic is based on Marquardt's iterative algorithm [51] which is a compromise between the Gauss-Newton and Steepest descent methods. It is necessary to specify the values of p (number of autoregressive parameters), d (order of consecutive differencing), and q (number of moving average parameters) to be included in the ARIMA model. Output from program ESTIMATE includes the variance and autocorrelations of residuals, and values of the Q statistic of equation 3.33 for lags K equal 8, 12, 24 and 36. Storage requirement is 130K under the constraint: $(\text{No. of observations} + 104) \times (\text{No. of parameters to be estimated}) \leq 9,624$. Finally, program FORECAST computes point forecasts and their 95 percent confidence intervals for the specified lead times. This program requires 100K storage capacity for a maximum series length of 1,500.

Application of the Box-Jenkins techniques to all of the time series listed in Table 4.1 resulted the same ARIMA model, albeit with different parameter values. To help demonstrate the model fitting

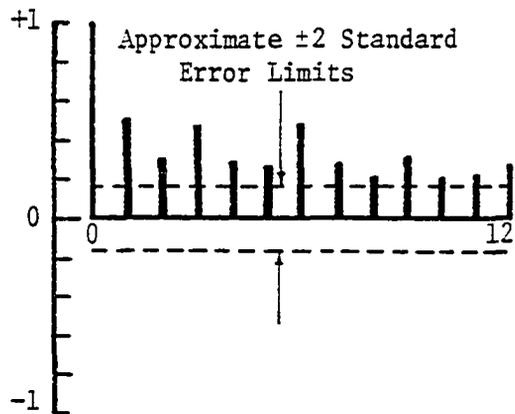
procedure, the representative volume and occupancy series shown in Figure 3.1 will serve as a case study. Sample autocorrelations and partial autocorrelations of the raw series and of their first differences are illustrated in Figures 4.4 and 4.5. It appears that sample autocorrelations of the raw series damp off very slowly as lag increases, suggesting that differencing is needed. Meanwhile, however, sample autocorrelations of first differences indicate that only the spikes at lags one, two, and three are large relative to their standard errors. Also, partial autocorrelations of these first differences gradually tail off. From the results of section 3.5.4, it might be reasonably postulated that the stochastic process generating the data stream is ARIMA (0, 1, 3), that is, the first differences of traffic data can be represented by the third-order moving average model

$$(1-B)(x_t - \mu) = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)a_t, \quad |\theta| < 1 \quad (4.1)$$

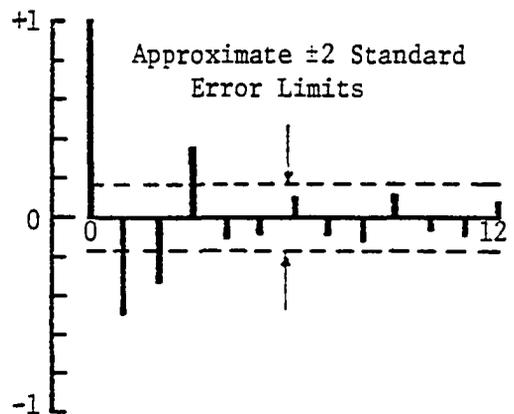
or simply,

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} \quad (4.2)$$

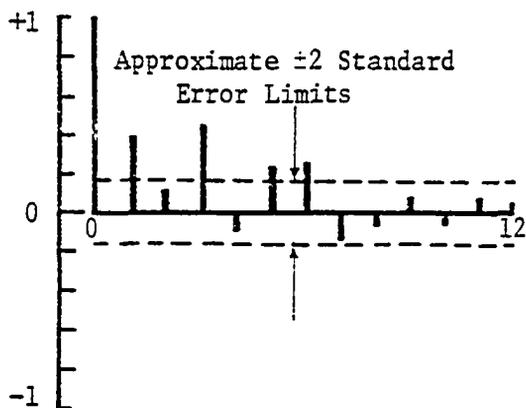
where $Z_t = x_t - x_{t-1}$. The above model states that the series of differences $Z_1, Z_2, \dots, Z_t, \dots$ is a series of moving linear combinations of $(a_0, a_1, a_2, a_3), (a_1, a_2, a_3, a_4), \dots, (a_{t-3}, a_{t-2}, a_{t-1}, a_t), \dots$ with weight functions $(-\theta_3, -\theta_2, -\theta_1, 1)$. Alternatively, it is perhaps more meaningful to view the model as showing that the shock a_t , coming into the system at time t , will persist over times $t, (t+1), (t+2)$ and $(t+3)$ in proportion to $(1, -\theta_1, -\theta_2, -\theta_3)$ before dissipation. The vector $(1, -\theta_1, -\theta_2, -\theta_3)$ which is the mirror image of the weight function $(-\theta_3, -\theta_2, -\theta_1, 1)$ is called the shock effect function.



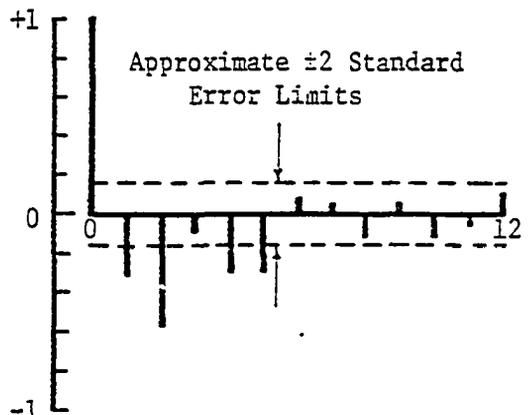
Sample Autocorrelation Function
of Raw Data



Sample Autocorrelation Function
of First Differences



Sample Partial Autocorrelation
Function of Raw Data



Sample Partial Autocorrelation
Function of First Differences

FIGURE 4.4

SAMPLE AUTOCORRELATIONS AND PARTIAL AUTOCORRELATIONS
VOLUME SERIES, MINNEAPOLIS, I-35, STATION 1

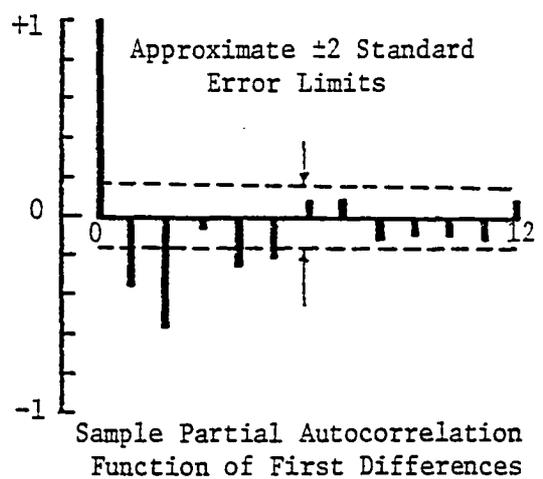
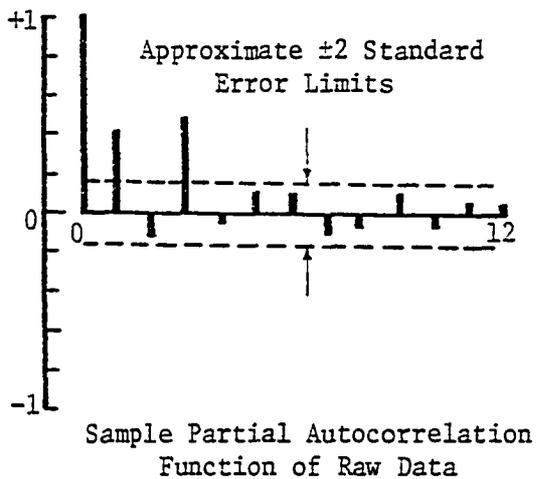
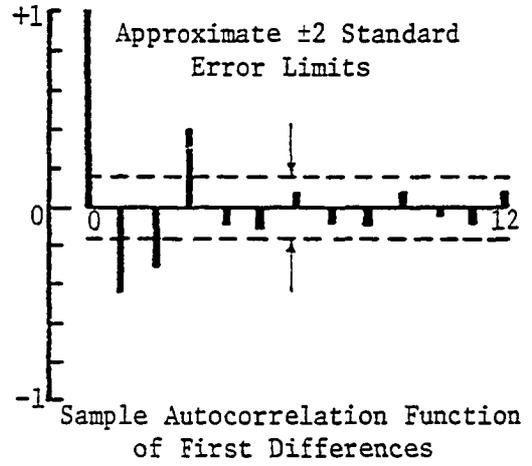
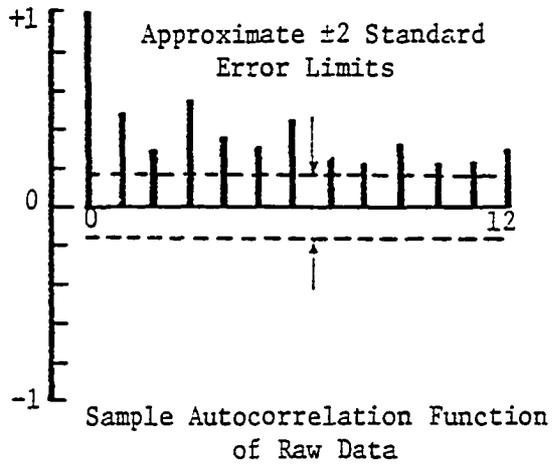


FIGURE 4.5

SAMPLE AUTOCORRELATIONS AND PARTIAL AUTOCORRELATIONS
OCCUPANCY SERIES, MINNEAPOLIS, I-35, Station 1

The estimated moving average parameters of the volume and occupancy series shown in Figure 3.1 are

Volume Data

$$\theta_1 = 0.6178 \text{ (Standard error} = 0.0795)$$

$$\theta_2 = 0.3730 \text{ (Standard error} = 0.0906)$$

$$\theta_3 = -0.0297 \text{ (Standard error} = 0.0804)$$

Occupancy Data

$$\theta_1 = 0.6039 \text{ (Standard error} = 0.0792)$$

$$\theta_2 = 0.3819 \text{ (Standard error} = 0.0885)$$

$$\theta_3 = -0.3097 \text{ (Standard error} = 0.0795)$$

As it is typical in many time series of physical processes, the parameter estimates are highly correlated. The estimated correlation matrices are

Volume Data

$$\begin{array}{c} \theta_1 \\ \theta_2 \\ \theta_3 \end{array} \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \\ 1 & & \\ -0.5016 & 1 & \\ -0.1996 & -0.5046 & 1 \end{bmatrix}$$

Occupancy Data

$$\begin{array}{c} \theta_1 \\ \theta_2 \\ \theta_3 \end{array} \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \\ 1 & & \\ -0.4851 & 1 & \\ -0.2204 & -0.4836 & 1 \end{bmatrix}$$

This makes marginal inference for individual parameters difficult and not particularly informative. In other words, when testing whether any of the θ parameters is different from zero using the t-test, one may incorrectly drop this parameter while the truth may be that the parameter should be retained.

Diagnostic checking was carried out by inspecting the residuals, \hat{a}_t , from the estimated model. Figures 4.6 and 4.7 show plots of the residuals and their autocorrelation functions. The autocorrelations exhibit no significant structure remaining in the data, and they are all quite small in magnitude. Also, the average of residuals, \bar{a} , from the volume series is -0.038 , and the estimated standard error of \bar{a} is 0.248 , strongly supporting that the a 's have zero mean. Similarly, the average of residuals from the occupancy series is -0.019 , with an estimated standard error of 0.214 , supporting the same conclusion. The lack-of-fit statistic, Q , which is described by equation 3.33 was computed from the first 24 residual autocorrelations. The values of $Q(24)$ are 27.4 and 23.5 for the volume and occupancy series, respectively. When these values of Q are compared with tabulated chi-square values with 21 degrees of freedom, they indicate that the residuals are white noise at the five percent level of significance.

In addition, the normality of residuals was checked using the chi-square goodness-of-fit test. Basically, this test makes use of a statistic which is approximately distributed as a chi-square variable to indicate the degree of discrepancy between an observed probability density function, and the hypothesized theoretical density function.

The test statistic

$$x^2 = \sum_{i=1}^K \frac{(f_o - f_t)^2}{f_t} \quad (4.3)$$

where:

K = number of class intervals,

f_o = observed frequency in class interval i ,

f_t = theoretical frequency in class interval i .

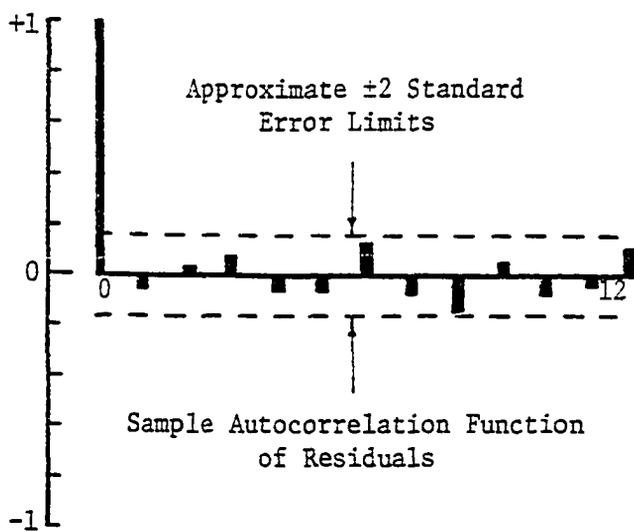
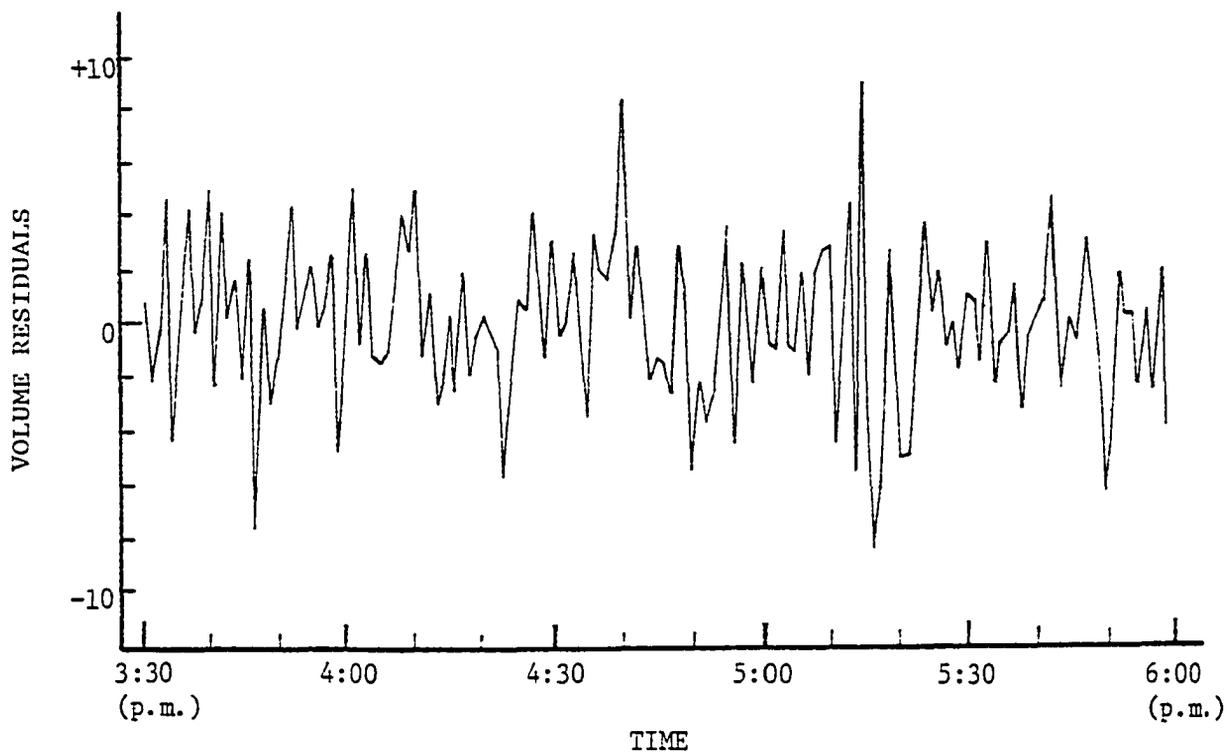


FIGURE 4.6

RESIDUAL PLOT AND SAMPLE AUTOCORRELATION FUNCTION
VOLUME SERIES, MINNEAPOLIS, I-35, STATION 1

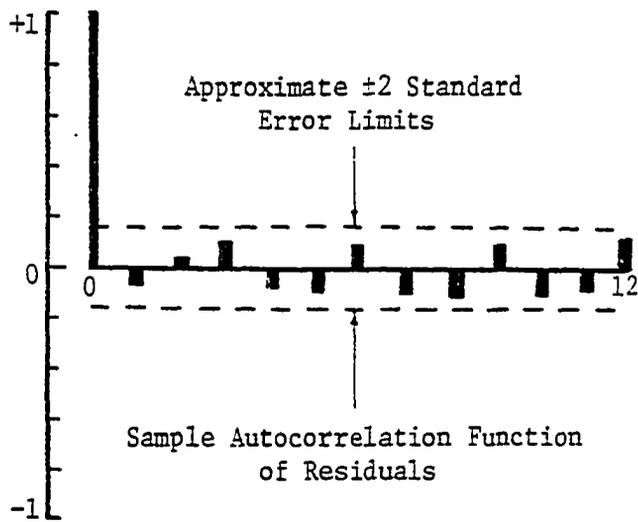
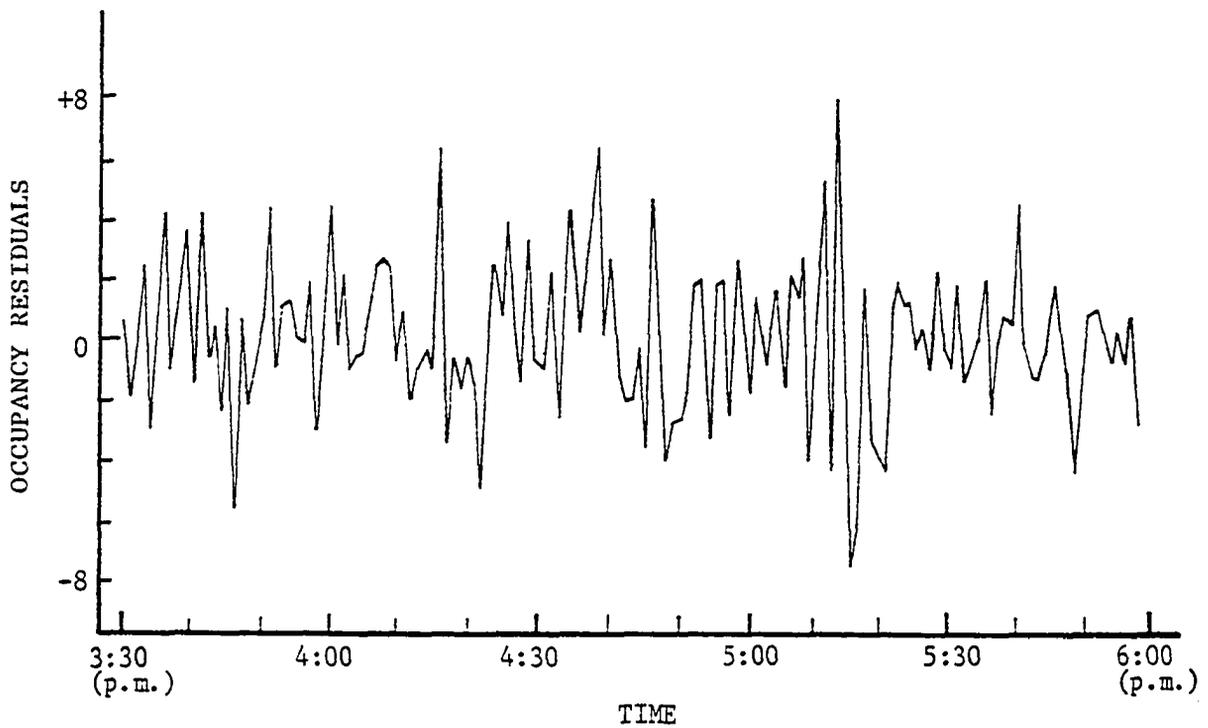


FIGURE 4.7

RESIDUAL PLOT AND SAMPLE AUTOCORRELATION FUNCTION
OCCUPANCY SERIES, MINNEAPOLIS, I-35, STATION 1

asymptotically has a chi-square distribution with $(K-1)$ degrees of freedom. This fact holds only when the parameters of the theoretical distribution under consideration, (mean and variance of the normal distribution), are known. When the parameters are unknown, it turns out that x^2 becomes asymptotically distributed as a chi-square variable with $(K-S-1)$ degrees of freedom, where S is the number of parameters of the theoretical distribution. It is recommended that f_0 of equation 4.3 should be greater than five for acceptable results [60]. The computed x^2 values are 36.03 and 30.79 with 27 degrees of freedom for the volume and occupancy series respectively. This indicates that the hypothesis of normality of residuals cannot be rejected at the five percent level of significance.

As noted earlier, freeway surveillance data are generally aggregated in two forms, over a specific time interval which is usually 20, 30, or 60 seconds, and/or over lanes. The transferability of the ARIMA (0,1,3) model under these conditions was explored by applying the model to different time series from the three freeway systems in Table 4.1. The range of values of the moving average parameters for 46 time series of traffic volume and occupancy aggregated over lanes is shown in Tables 4.2 and 4.3. It is of crucial importance that although there are some differences in the parameter estimates between and/or within the different freeway systems, the same form of the ARIMA model was found adequately representative. Probably, these differences arise from variations in driving situations, geometrics and other environmental factors.

In addition, the Los Angeles data which consist of 20-second compilations of volume and occupancy per lane, provided an opportunity to model traffic time series of individual lanes. The ARIMA (0,1,3)

TABLE 4.2

MOVING AVERAGE PARAMETERS OF VOLUME SERIES
AGGREGATED OVER LANES

Freeway System	No. of Data Sets	Total No. of Observations	Moving Average Parameters		
			θ_1	θ_2	θ_3
Harbor Fwy. (Los Angeles)	10	1750	0.7301 \pm 0.1885	0.1777 \pm 0.5765	0.0391 \pm 0.1398
I-35 Fwy. (Minneapolis)	10	1500	0.7553 \pm 0.1375	0.1519 \pm 0.5249	-0.1530 \pm 0.1438
Lodge Fwy. (Detroit)	2	381	0.7420 \pm 0.0732	0.0403 \pm 0.0671	0.0012 \pm 0.0143

TABLE 4.3

MOVING AVERAGE PARAMETERS OF OCCUPANCY SERIES
AGGREGATED OVER LANES

Freeway System	No. of Data Sets	Total No. of Observations	Moving Average Parameters		
			θ_1	θ_2	θ_3
Harbor Fwy. (Los Angeles)	10	1750	0.5611 \pm 0.3541	0.1145 \pm 0.2711	0.2596 \pm 0.3507
I-35 Fwy. (Minneapolis)	10	1500	0.4710 \pm 0.2160	0.1307 \pm 0.2515	-0.0848 \pm 0.2249
Lodge Fwy. (Detroit)	4	762	0.6121 \pm 0.1649	0.0859 \pm 0.1398	0.0704 \pm 0.1678

model was applied to 60 series of 20-second lane volumes and occupancies, and was found representative in all these cases. Also, the effect of sampling interval was investigated by aggregating the 20-second observations to 60-second observations with similar confirmation of the model. Tables 4.4 and 4.5 give the range of values of moving average parameters for the 20-second and 60-second lane volumes and occupancies.

The sensitivity of the performance of the ARIMA (0,1,3) model to variations in the moving average parameters over time was explored to a limited extent using eight volume and occupancy time series, each 150 time intervals (of 60 seconds) long, from the I-35 freeway in Minneapolis. Each series was broken into three 50-interval segments, and the ARIMA (0,1,3) was applied separately to each segment. The variations in the estimated moving average parameters θ_1 , θ_2 , and θ_3 for both sets of series are depicted in Figure 4.8. The horizontal scatter of points indicates that the parameters vary over time but no consistent pattern in this variation was noted. However, due to the limited number of observations used in estimating the parameters for each 50-interval segment, the conclusion that these parameters vary by time cannot be accurately drawn. Importantly, the same form of the ARIMA (0,1,3) model which represented the 150-observation series did as well representing the 50-observation segments.

4.5 ARIMA Model Utility in Forecasting and Simulation

In forecasting traffic control systems, a crucial requirement for the employed forecasting model is that it should be operational, that is, the model can be used in real-time to provide forecasts on an on-going basis. To appreciate the operational value of the ARIMA (0,1,3)

TABLE 4.4

MOVING AVERAGE PARAMETERS OF 20- AND 60-SECOND
LANE VOLUME SERIES, HARBOR FREEWAY, LOS ANGELES

Lane No. *	Moving Average Parameters					
	20-Second Series (15,750 Observations)			60-Second Series (5,250 Observations)		
	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
1	0.8081 ± 0.1263	0.0752 ± 0.1145	0.0426 ± 0.0742	0.8280 ± 0.1720	0.0311 ± 0.1786	0.0085 ± 0.1129
2	0.8701 ± 0.1245	0.0404 ± 0.1133	0.0401 ± 0.0738	0.7860 ± 0.1127	0.0056 ± 0.1788	0.0229 ± 0.1222
3	0.8131 ± 0.1582	0.0296 ± 0.1253	0.0074 ± 0.0933	0.8180 ± 0.1414	-0.0389 ± 0.3364	0.0412 ± 0.1466
4	0.8811 ± 0.1130	0.0569 ± 0.0573	0.0278 ± 0.0622	0.4526 ± 0.3510	0.1250 ± 0.2358	0.0437 ± 0.0904

*Numbering begins with the lane closest to the median and goes toward the right shoulder.

TABLE 4.5

MOVING AVERAGE PARAMETERS OF 20- AND 60-SECOND
LANE OCCUPANCY SERIES, HARBOR FREEWAY, LOS ANGELES

Lane No.*	Moving Average Parameters					
	20-Second Series (15,750 Observations)			60-Second Series (5,250 Observations)		
	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
1	0.6196 ± 0.2786	0.1971 ± 0.1988	0.0814 ± 0.1261	0.7057 ± 0.2856	0.1666 ± 0.2337	-0.0581 ± 0.0565
2	0.7096 ± 0.2353	0.1037 ± 0.1234	0.0674 ± 0.0776	0.6284 ± 0.2547	0.1330 ± 0.2681	0.0134 ± 0.1046
3	0.6672 ± 0.2394	0.1658 ± 0.1700	0.0211 ± 0.1116	0.6888 ± 0.3111	-0.0261 ± 0.2991	0.0388 ± 0.1448
4	0.6539 ± 0.1802	0.0400 ± 0.0857	0.1094 ± 0.1203	0.5617 ± 0.3655	0.1855 ± 0.1277	0.0431 ± 0.1687

*Numbering begins with the lane closest to the median and goes toward the right shoulder.

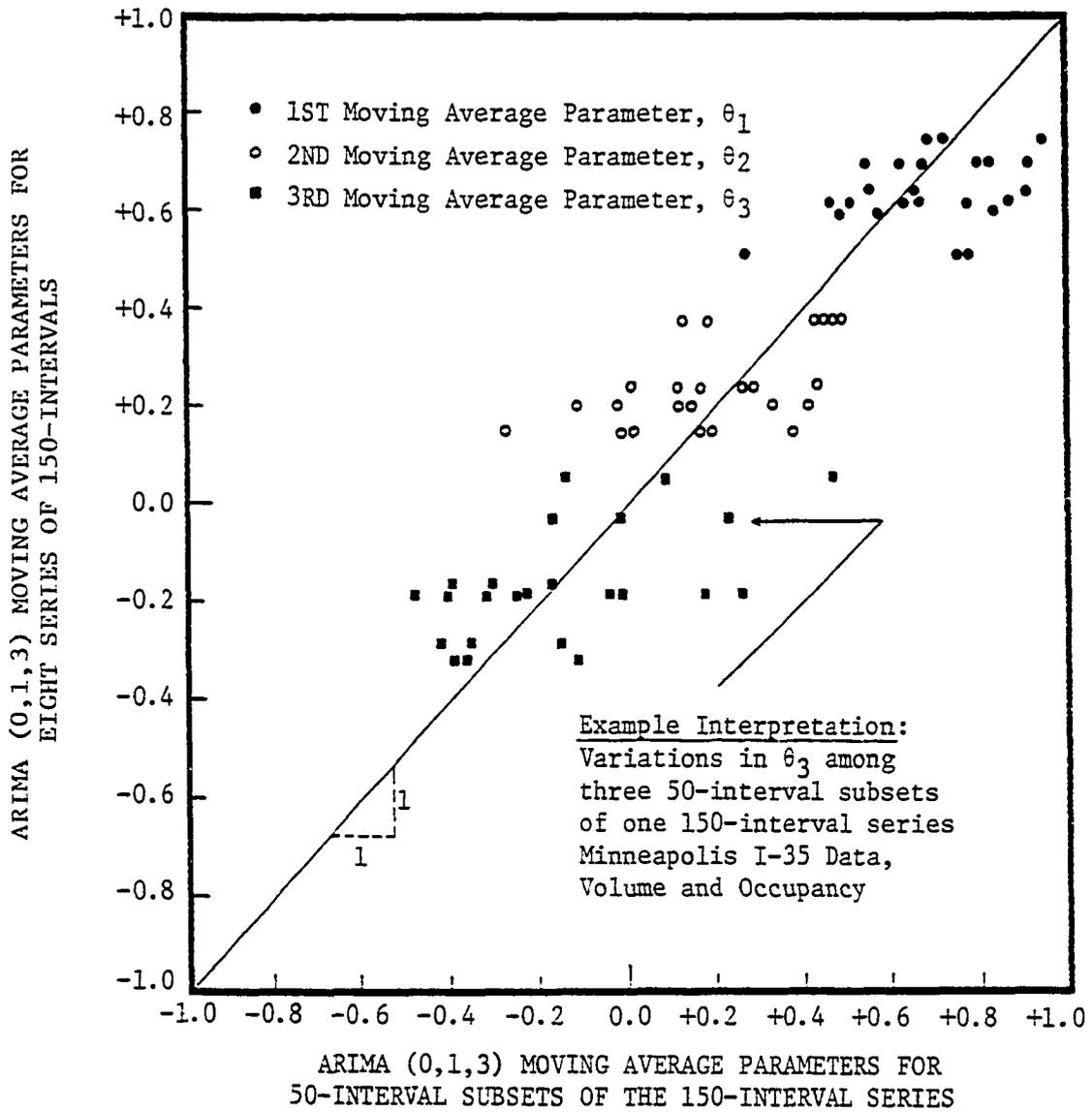


FIGURE 4.8

SENSITIVITY OF ARIMA (0,1,3) PARAMETERS OVER TIME

model, it is instructive to show how the model can be efficiently applied in making short-term forecasts. Given that the observations x_t , x_{t-1} , ... on a traffic variable have been made at time t , suppose that it is required to forecast x_{t+1} which when observed will be given by

$$x_{t+1} = x_t + a_{t+1} - \theta_1 a_t - \theta_2 a_{t-1} - \theta_3 a_{t-2} \quad (4.4)$$

The above equation can be readily obtained by replacing t by $(t+1)$ in equation 4.2. Since at time t , the future random shock a_{t+1} cannot be forecasted, it would be rationale to replace it by its expected value which is zero. Also, from the results of section 3.5.5, the disturbances a_t , a_{t-1} , a_{t-2} are simply previous one-step-ahead forecast errors. Hence, the forecast made at time t for x_{t+1} would be

$$\hat{x}_t(1) = x_t - \theta_1 \cdot e_{t-1}(1) - \theta_2 \cdot e_{t-2}(1) - \theta_3 \cdot e_{t-3}(1) \quad (4.5)$$

At the first beginning, values of the one-step-ahead forecast errors are unknown, and therefore they are replaced with zeros.

The problem of forecasting future traffic observations becomes considerably easier if the values of the moving average parameters θ_1 , θ_2 , and θ_3 are constant over time. In such case, equation 4.5 provides an operational expression for forecasting. The computational utility of this expression stems from the fact that its application requires storage only of the latest three forecast errors, and the current observation. Values of the moving average parameters, however, may vary by time period as discussed in the previous section. In this case, the simplest approach is to use fixed average values of the parameters, thus, ignoring any probable variations in the parameter estimates over time. This approach was used by Der [18], and Eldor [27].

The other extreme approach is to update the parameter estimates in real-time. It is believed, however, that a rapid adjustment in the parameter estimates, each observation interval, for example, may degrade the overall forecasting performance of the ARIMA model. Past experience with adaptive exponential smoothing models, particularly the Trigg and Leach model, has shown that successively changing the smoothing constant values over time yielded potentially larger forecast errors than those resulting from Brown's original models [73]. Another important factor which should be taken into consideration when thinking of real-time updating of the model parameters is that of computational requirements. One intermediate approach between these two extremes is to update the parameter estimates only occasionally, for example, at the beginning of peak and off-peak periods. Parameter updating was not explored in this research, in part because the available data sets consisted of afternoon peak-period time series for one single day only.

So far, the question of how long a time series should be in order to estimate the moving average parameters and forecast future values of the series has not been discussed yet. Certainly, the total amount of information included in a series affects the accuracy of estimation and consequently the computed forecasts. Most of the results presented in Chapter III on estimating the autocorrelation functions and moving average parameters, and on diagnostic checking of the ARIMA model adequacy are based on large sample theory. Nelson [56] conducted Monte Carlo simulation studies on the reliability of identification and estimation procedures of moving average models and related

test statistics. He found that series of length 100 observations are adequate for most practical studies.

Using the above approach, one-step-ahead forecasts for the different time series in Table 4.1 were computed from the fitted ARIMA (0,1,3) models utilizing program FORECAST. Figure 4.9 shows forecasts of the representative volume and occupancy series presented in Figure 3.1 superimposed on the observed series. The forecasts seem to be fairly close to the actual observations except in some few cases. Measurement errors from detectors are probably the reason for these exceptions. As a measure of uncertainty of the forecasts, 95-percent probability limits were constructed using expression 3.40, and are indicated in Figure 4.9 by broken lines.

To illustrate the utility of the ARIMA (0,1,3) model in simulation, a computer program was written in PL/1 to generate compilations of traffic volume and occupancy series. The program consists of main procedure TRAFIK and subprocedure RNNORM. Each call to RNNORM returns a pseudorandom number which is normally distributed with specified mean and variance. These pseudorandom numbers are approximately white noise, and they are input to procedure TRAFIK. In particular, if the mean and variance of residuals from a fitted ARIMA model are set equal to the mean and variance in RNNORM, and for given values of moving average parameters, a time series of traffic volumes or occupancies can be generated. The starting value of the series must be specified. To serve as an example, results of the fitted ARIMA (0,1,3) models for the representative volume and occupancy series of Figure 3.1 were used in the simulation program. Figure 4.10 displays profiles of the simulated

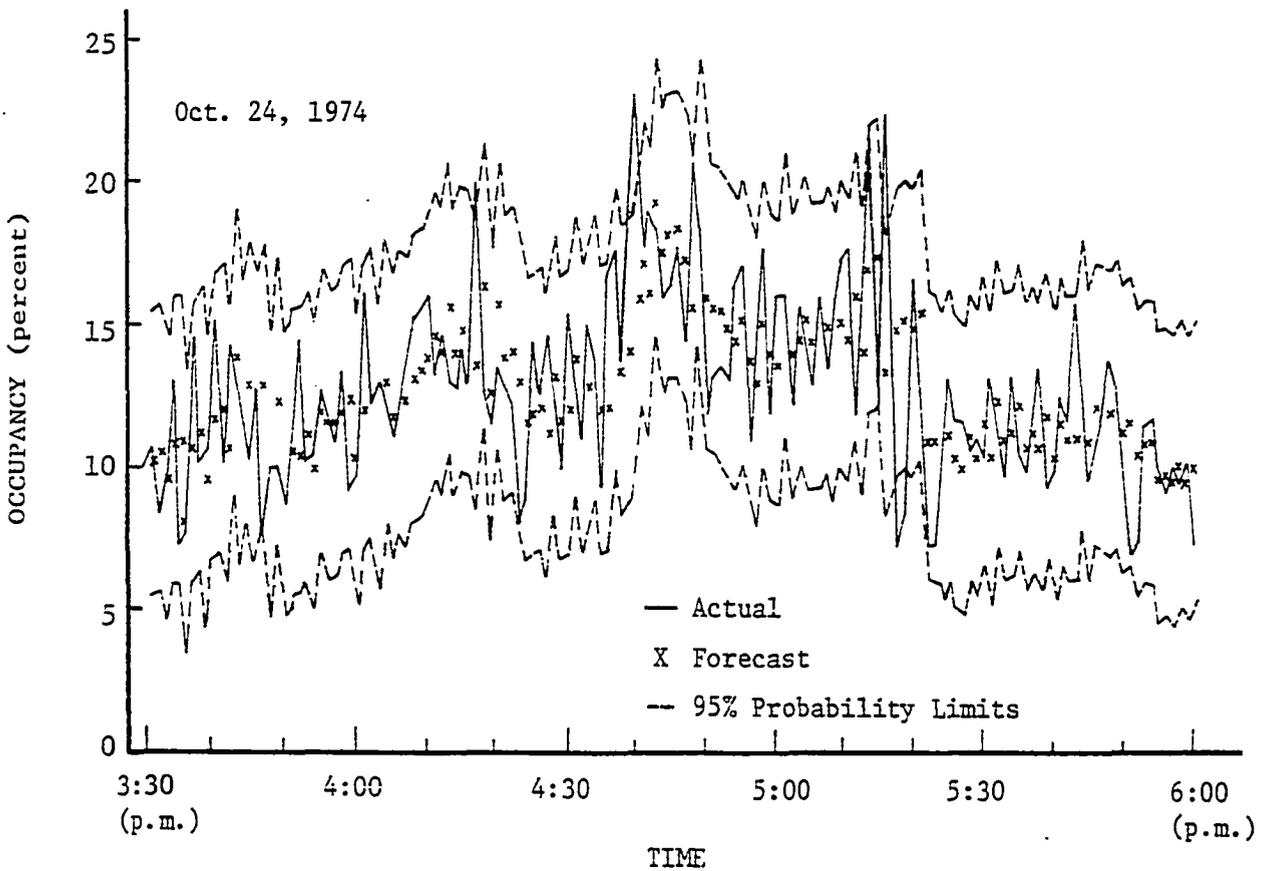
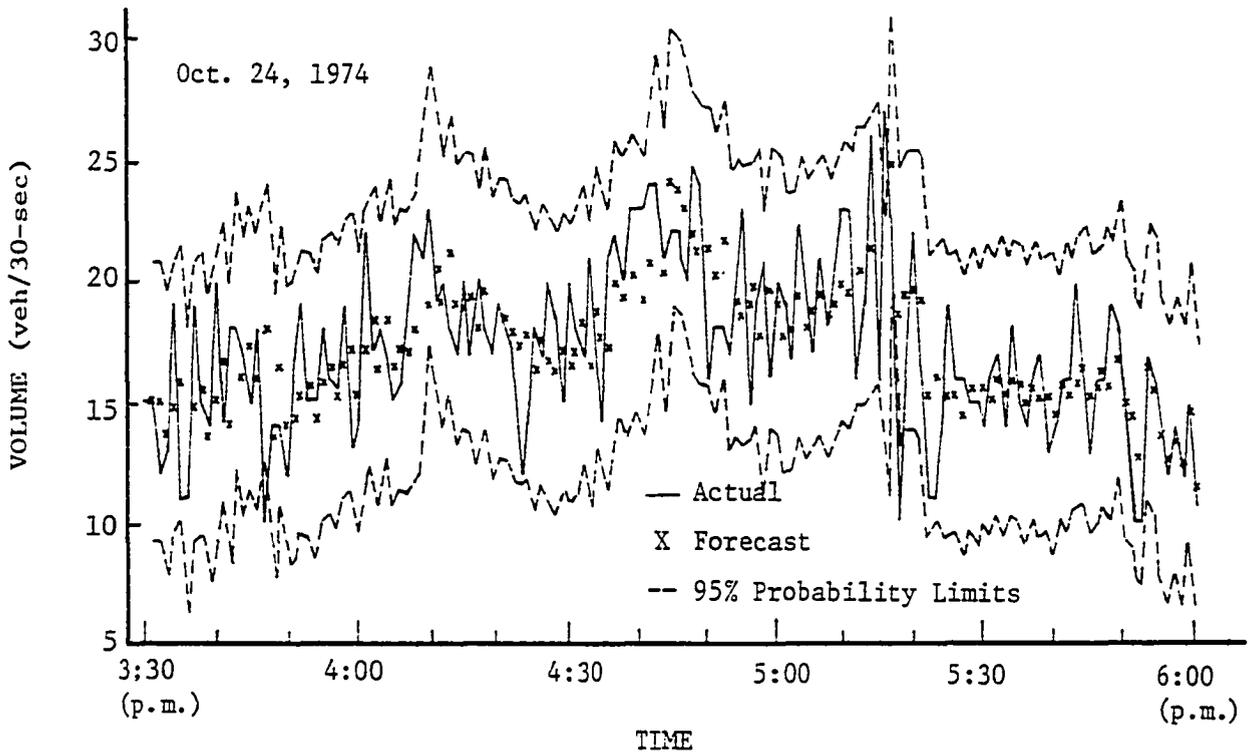


FIGURE 4.9

OBSERVED VOLUME AND OCCUPANCY SERIES WITH FORECASTS
(Minneapolis, I-35, Station 1)

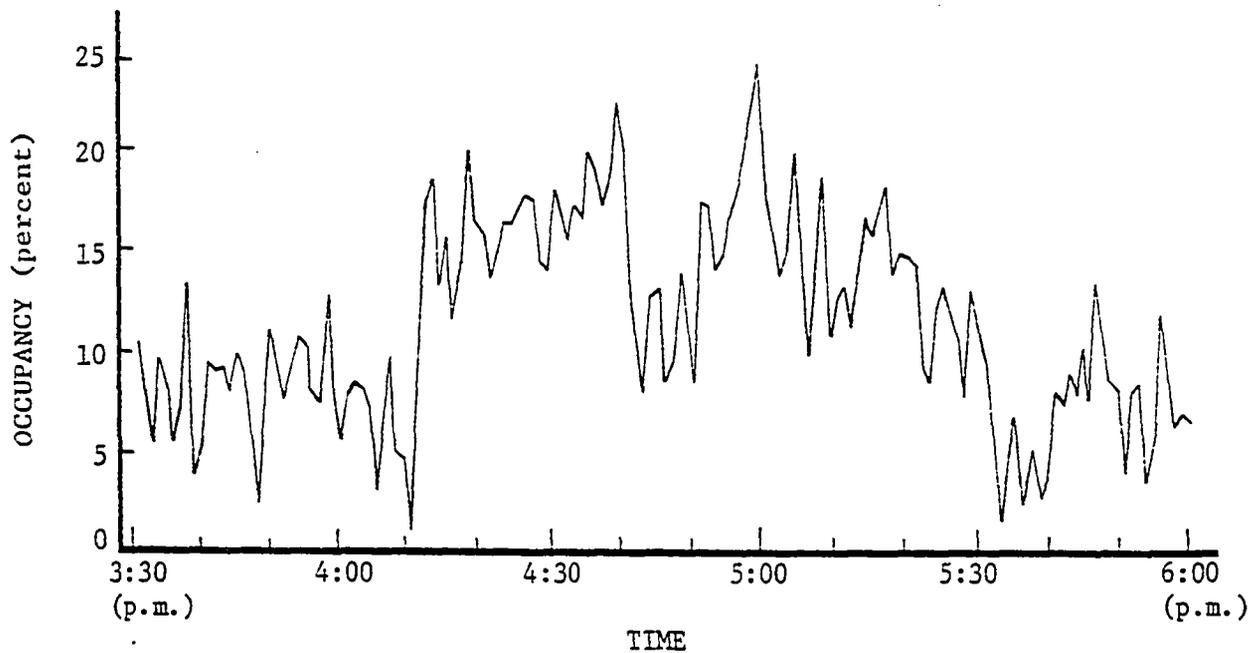
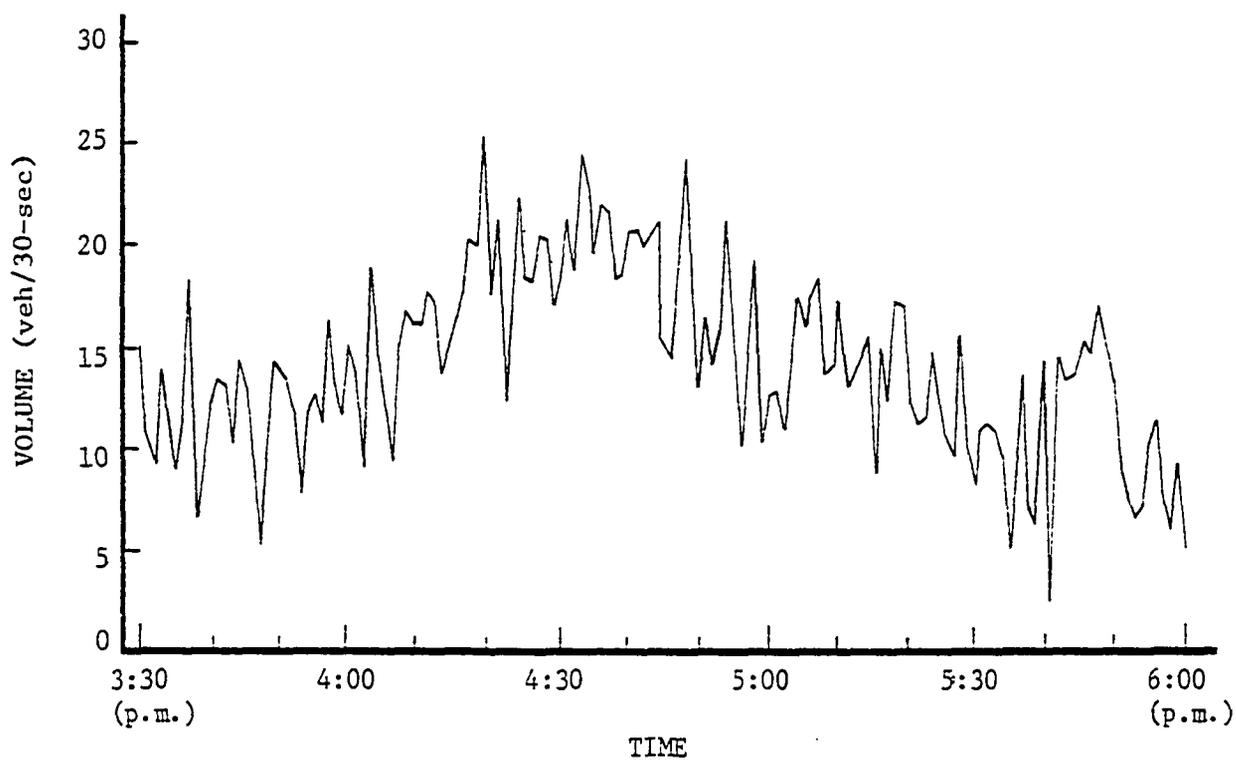


FIGURE 4.10

SIMULATED TRAFFIC VOLUME AND OCCUPANCY SERIES
FROM FITTED ARIMA (0,1,3) MODEL

traffic volume and occupancy series. The dynamic behavior of the simulated series and of the actual series looks very similar. This is because the simulation logic is not solely based on theoretical considerations about how traffic might behave under idealized conditions, but on an observed set of data describing how it did behave during the observation period.

The simulation program described above can be of potentially broad usage in traffic studies. It can be utilized separately as a cheap and convenient source of traffic data which may be stored on a magnetic tape or in the form of computer cards for further analysis. Alternatively, it can be used with other programs to simulate the dynamic behavior of traffic systems presently in existence or anticipated for the future.

4.6 Comparative Evaluation of Forecasting Performance

To assess the forecasting performance of the ARIMA (0,1,3) model on a comparative basis with the ad hoc models described in Chapter III, the moving average model, the double exponential smoothing model, and the Trigg and Leach model were applied to the traffic volume and occupancy data base of Table 4.1. Computer programs were written in FORTRAN IV to perform the computations required by these models. Two performance criteria were used in the evaluation process. These are the mean absolute error, MAE, and mean square error, MSE, which are defined as

$$MAE = (1/n) \sum_{t=1}^n |x_t - \hat{x}_t| \quad (4.6)$$

and

$$\text{MSE} = (1/n) \sum_{t=1}^n (x_t - \hat{x}_t)^2 \quad (4.7)$$

where:

x_t = observed value at time t ,

\hat{x}_t = predicted value at time t ,

n = number of observations

The mean absolute error indicates the expected error which may take place in each individual forecast, while the mean square error detects the presence of frequent large forecast errors. Values of MAE and MSE for the ARIMA (0,1,3) models ranged from 1.30 to 6.50 and from 2.80 to 91.41, respectively. For the purpose of comparison, these values of MAE and MSE of the ARIMA (0,1,3) model were chosen as a basis, and ratios-to-Box-Jenkins for MAE and MSE were calculated for the other models as follows

$$\text{Ratio-to-Box-Jenkins for MAE} = \frac{\text{MAE for a given model}}{\text{MAE for ARIMA (0,1,3) model}} \quad (4.8)$$

$$\text{Ratio-to-Box-Jenkins for MSE} = \frac{\text{MSE for a given model}}{\text{MSE for ARIMA (0,1,3) model}} \quad (4.9)$$

In evaluating the moving average model of equation 3.10, five values of N , (N equals 5,10,20,50 and 100), were used in the analysis. Results of this model indicated that both MAE and MSE increase with increasing N . When N equaled five, the ratio-to-Box-Jenkins varied between 1.00 and 1.27 for MAE and between 1.00 and 1.45 for MSE.

Values of the smoothing constant α included in assessing the double exponential smoothing model of equation 3.15 were 0.1 through 0.9 in increments of 0.1. As expected, the best results of this model

were associated with small values of α . In particular, smoothing constants between 0.1 and 0.3 resulted values of the ratio-to-Box-Jenkins in the range from 1.00 to 1.64 for MAE, and from 1.00 to 1.43 for MSE. For larger values of α , (0.4 to 0.9), the ratio-to-Box-Jenkins varied between 1.10 and 2.31 for MAE, and between 1.22 and 3.80 for MSE.

The Trigg and Leach model of equation 3.18 was tested using nine values of the smoothing constant α between 0.1 and 0.9 in increments of 0.1, and three values of the smoothing constant γ equals 0.1, 0.2 and 0.3. With large initial values of the smoothing constant α between 0.6 and 0.9 and a smoothing constant γ of 0.1, (which gave the best results of this model), the ratio-to-Box-Jenkins varied between 1.45 and 8.20 for MAE, and between 2.08 and 44.34 for MSE. The reason for the poor performance of this model could be the abrupt successive changes in the smoothing constant α [73]. This tends to confirm the belief that real-time updating of the moving average parameters in the ARIMA (0,1,3) model may not be warranted.

Figures 4.11 and 4.12 display the ranges of the best values of the ratio-to-Box-Jenkins for MAE and MSE for the different models. It is readily seen that the ARIMA (0,1,3) model is superior to the other ad hoc forecasting models. This is due to its more accurate representation of the stochastic process generating freeway traffic data. Also, the magnitude of reduction in forecast errors has the potential for significantly improving the operations of surveillance and control systems.

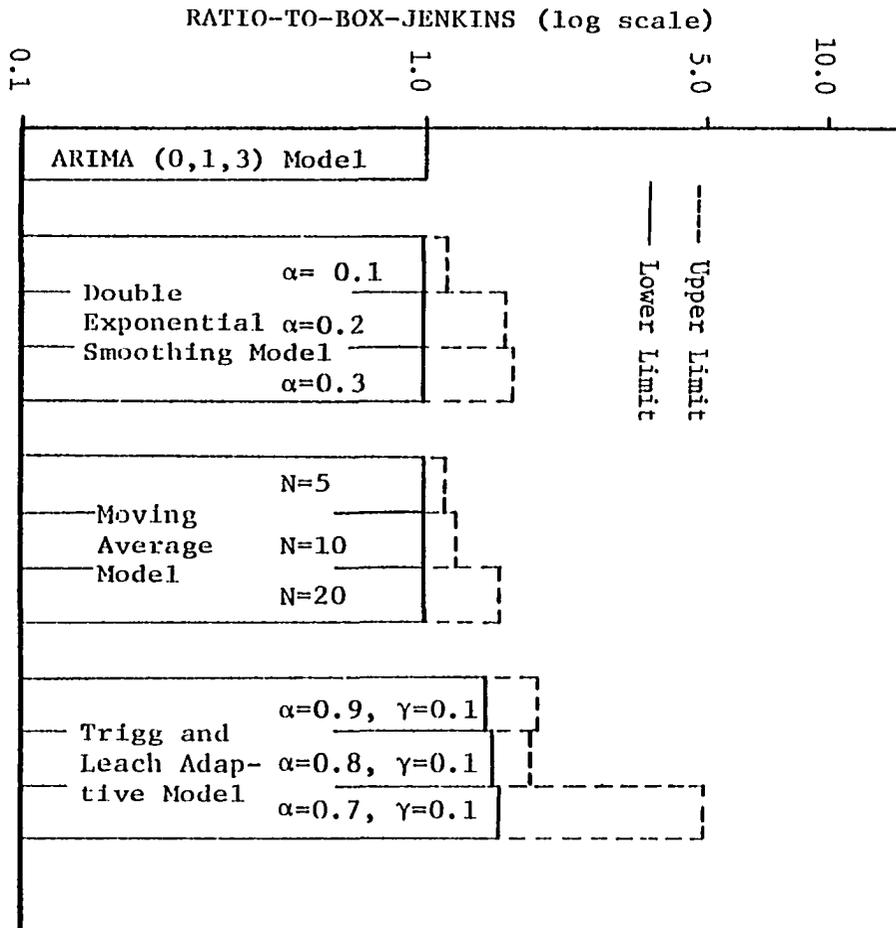
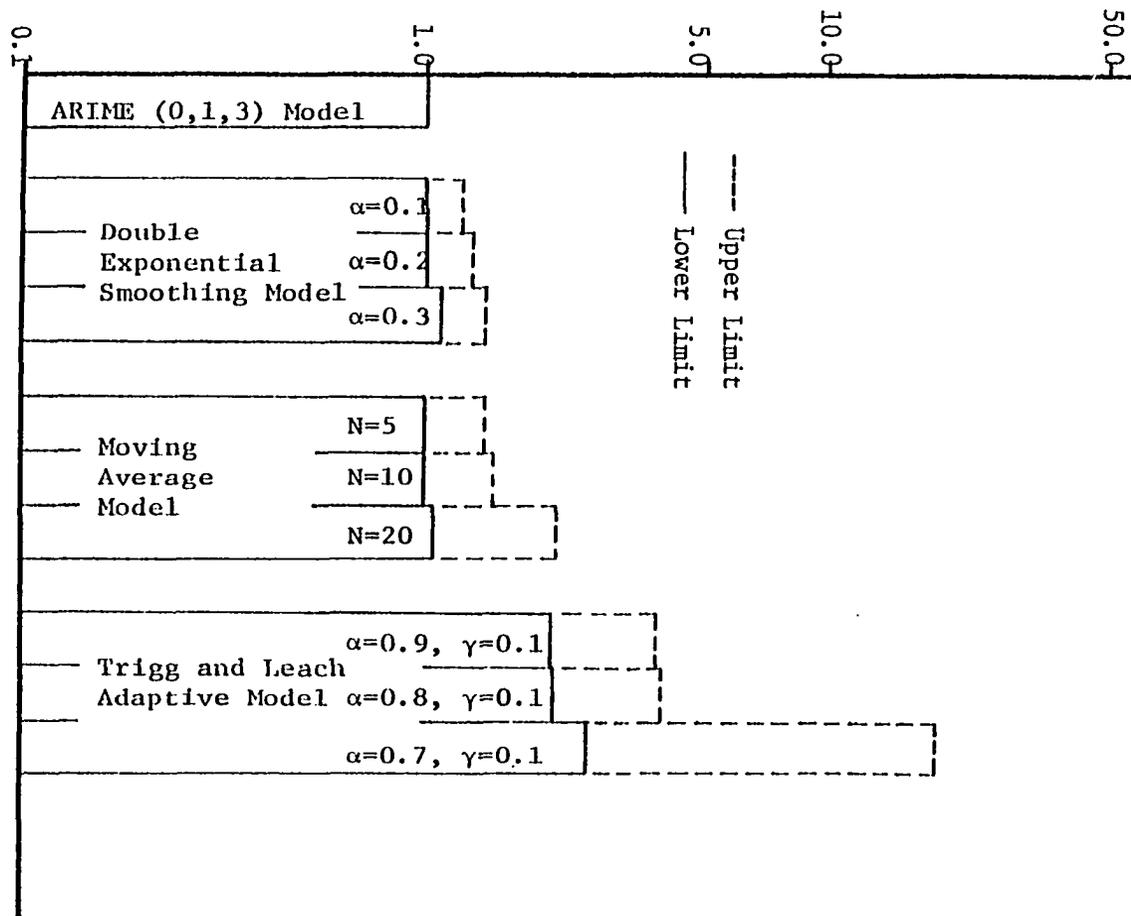


FIGURE 4.11

RATIO-TO-BOX-JENKINS FOR MEAN ABSOLUTE ERROR

RATIO-TO-BOX-JENKINS (log scale)



RATIO-TO-BOX-JENKINS FOR MEAN SQUARE ERROR

FIGURE 4.12

4.7 Summary

The class of autoregressive integrated moving average models, however complex as may seem, is no substitute for thinking about the dynamic behavior of many physical systems. These models have been successfully applied over the past few years in modeling economical, industrial and environmental systems. In this chapter, an attempt has been made to demonstrate the application of these models to freeway traffic data. An ARIMA (0,1,3) model did adequately represent traffic volume and occupancy time series from different freeway systems, thus providing evidence for the transferability of the form of the model. This model can be of potential utility in forecasting and simulating traffic systems. The next chapter investigates the application of this ARIMA model to freeway traffic management, particularly incident detection.

CHAPTER V

APPLICATION OF TIME SERIES ANALYSIS TO
URBAN FREEWAY INCIDENT DETECTION5.1 Introduction

Freeway capacity-reducing incidents are hazardous events occurring randomly at a rate which could be as high as one incident per directional mile per hour [70]. Moskowitz [55] reported that the most important problem in urban freeway operations is the determination of how to detect stopped vehicles and the necessary steps to remove the stoppage. The Traffic Control Systems Handbook [82] indicates that nonrecurrent freeway congestion caused by traffic incidents in large urban areas is responsible for as much motorist delay as the recurrent congestion due to geometric bottlenecks. In a metropolitan freeway network, at least one lane is expected to be blocked for 7.2 percent of the time because of accidents, 6.7 percent of the time as a result of stopped vehicles and for 27.8 percent of the time due to maintenance operations [82]. Another interesting finding by Goolsby [35] is that the magnitude of freeway capacity reduction due to an incident usually exceeds the magnitude of reduction in the physical width of the travelled way. For example, research on 6.5-mile section of the Gulf freeway in Houston indicated that the blockage of a single lane halved the capacity of the

three-lane freeway section, while the blockage of two lanes out of the three reduced the capacity by approximately 80 percent. In summary, incident detection is a very important function of any automated freeway surveillance and control system.

This chapter begins with a review of incident detection algorithms which can be used in real-time by the computer to determine the location, time of occurrence and magnitude of capacity reduction caused by freeway incidents. The review includes a general typology of detection algorithms, and the measures of effectiveness of these algorithms. Also, it attempts to survey the current structures of detection algorithms. There is then a description of the incident data base used in the analysis. Finally, there is a formulation and discussion of some detection algorithms based on the ARIMA (0,1,3) model, including an evaluation of one of these algorithms on a comparative basis with a number of previously developed algorithms.

5.2 Computer Incident Detection Algorithms

A computer incident detection algorithm is defined by Tignor [78] as "a specific logical and analytical procedure used along with data obtained from freeway surveillance traffic detectors to ascertain the presence or absence of a capacity reducing incident." As discussed earlier in section 2.6, when the capacity of a freeway section is reduced to a level below that of the prevailing demand as a result of traffic incident, certain flow-states will develop both upstream and downstream of the incident site. In general, downstream volume will decrease and the blocked lane or lanes will not be fully utilized for some distance by downstream traffic unless, of course, congestion have existed before the incident. Meanwhile, a

queue of vehicles will form upstream in the blocked lanes, and stop-and-go conditions will probably be generated in all lanes when the queued vehicles attempt to move to the unblocked lanes. Furthermore, occupancy will increase upstream and lower downstream relative to the conditions which have prevailed before the incident. Figure 5.1 illustrates representative volume and occupancy observations during a traffic incident which involved a disabled vehicle on the Lodge Freeway and lasted for approximately 14 minutes. Moreover, as long as the incident remains and the upstream demand exceeds the reduced capacity, shock waves of congestion and shock waves of expansion will continue to propagate upstream and downstream of the incident site, respectively.

In conclusion, incident detection algorithms are basically structured so as to register sudden changes in flow-states both upstream and downstream the incident site, and/or the passage of shock waves, particularly the wave of congestion travelling upstream. However, it is important to note that there exist some incident situations where the developed traffic patterns can not be exactly or immediately recognized by an incident detection algorithm. Examples of these situations are the case when free flowing traffic conditions prevail and the magnitude of capacity reduction is not severe, the case when congestion prevails before the incident both upstream and downstream of the incident location, and the case when heavy traffic conditions exist but traffic demand is less than the reduced capacity. On the other hand, there also exist some incident-free situations which tend to generate traffic patterns similar to those caused by incidents, and therefore produce false-alarms. Examples of these incident-free situations include the case of heavy traffic where the

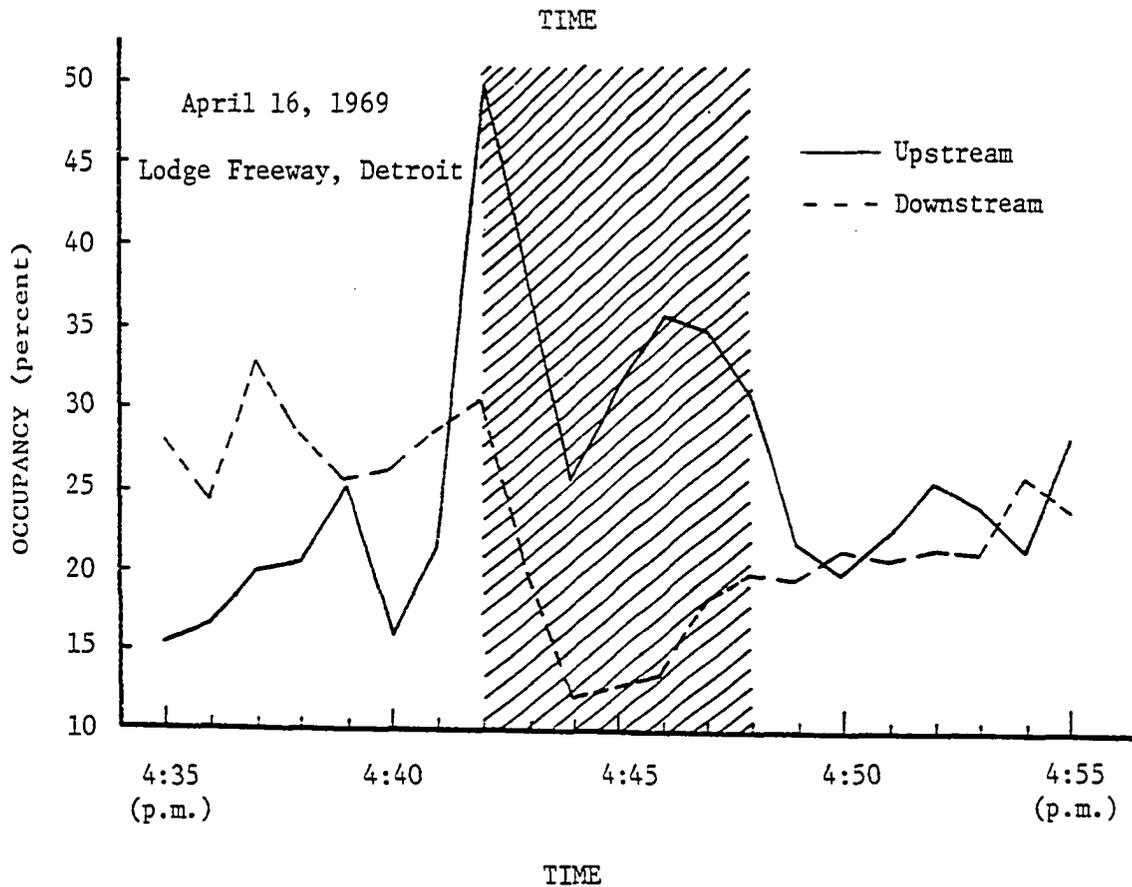
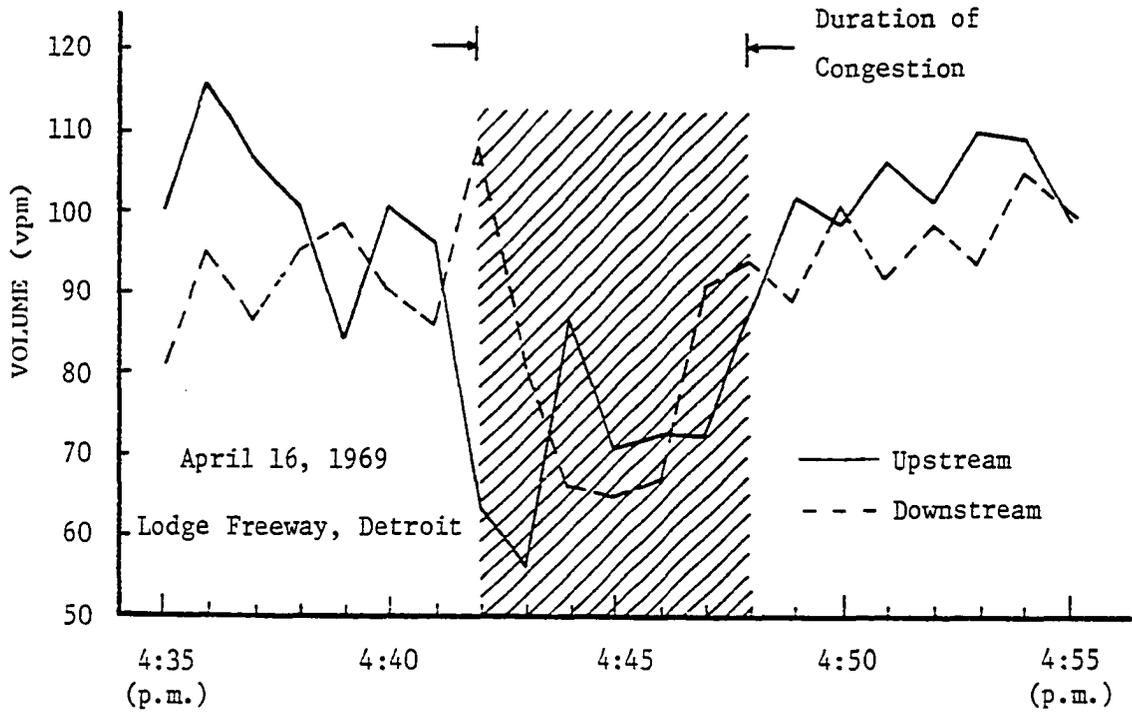


FIGURE 5.1

VOLUME AND OCCUPANCY CHANGES DURING
A CAPACITY-REDUCING INCIDENT

vehicles experience significant speed variations causing congestion waves to propagate opposite to the direction of flow, thus, creating significant spatial variations in traffic occupancy, and the case of abnormal geometrics at interchanges between freeways or at locations with a high volume of entrance ramp traffic.

5.2.1 Measures of Effectiveness for Detection Algorithms

To evaluate the performance of an incident detection algorithm, certain measures of effectiveness are usually used. These measures are highly pertinent to the applicability of an algorithm in automated freeway surveillance systems. Figure 5.2, adapted from Tignor [78], illustrates the outcome possibilities which arise when an incident detection algorithm is executed. The first and perhaps most obvious measure of effectiveness is the probability of detection, which is the percentage of incidents detected by the algorithm out of all incidents which occurred. Generally, the probability of detection can be increased, but only at the expense of an increase in the probability of false-alarms [65]. Two definitions of the probability of false-alarms exist in the literature, an on-line definition, and an off-line definition [49]. The on-line definition is the percentage of false incident messages out of total incident messages generated by the algorithm. Alternatively, the off-line definition is the percentage of incident messages out of all messages generated by the algorithm using representative incident-free data. This latter definition is particularly relevant in off-line studies as means of comparing different algorithms. Generally, the cost of false-alarms can be better understood if one recognizes that detection information is used to dispatch highway patrol units to the scene of the incident. On the

		ACTUAL CONDITION	
		Incident-Free	Incident Present
CONDITION INDICATED BY ALGORITHM	Incident-Free	Correct Detection	Correct Detection
	Incident Present	False Alarm	Correct Detection

FIGURE 5.2
OUTCOMES WHEN EXECUTING AN INCIDENT
DETECTION ALGORITHM

(Adapted from Reference 78)

other hand, the benefits derived from prompt detection include the receiving of the required type of assistance to the motorists in distress, and the reduction in delay to the other freeway users.

In addition to the probabilities of detection and false-alarms, the third measure of effectiveness is the average time lag to detection, which is the expected time between the actual occurrence of an incident and its detection for all detected incidents. As mentioned earlier in Chapter I, efforts to redistribute traffic demand during a freeway incident by diverting upstream drivers to other alternate routes become more effective in reducing total expected delay if the response starts in the early stages of congestion growth. An interesting rule of thumb which helps assessing the effect of faster response to incident situations is that motorist delay is proportional to the square of the total incident duration. Hence, if it is assumed that an incident duration of 25 minutes is reduced by 5 minutes, the proportionate reduction in total delay will be 36 percent. Also, faster response to incidents has the potential of improving safety due to reduction in time of exposure to hazards.

Other measures of effectiveness for detection algorithms include the probability of detecting the end of the incident, determining the exact incident location, assessing the magnitude of capacity reduction and determining the nature of the incident. Usually, however, one cannot expect that a single detection algorithm would satisfy all the performance requirements together. In a practical sense, the operating agency of an incident management system decides upon the acceptability of a particular algorithm based on its overall detection performance.

5.2.2 A Typology of Detection Algorithms

Detection algorithms can be categorized in a number of ways depending upon their structure and the data input required. Four dimensions can be proposed to describe the types of algorithms which might be constructed. These are: (1) station or subsystem algorithms, (2) pattern recognition or smoothing algorithms, (3) cross-sectional or lane-aggregated algorithms, and (4) empirical or simulation-based algorithms.

Station or Subsystem Algorithms. A station algorithm is one that identifies changes in macroscopic traffic stream measurements only at a detector station upstream of an incident site. Subsystem algorithms, on the other hand, attempt to identify discontinuities in traffic flow at detector stations on both sides of the incident. The freeway section between these stations is called a subsystem.

Pattern Recognition or Smoothing Algorithms. The basic purpose of a pattern recognition algorithm is to discriminate between incident and non-incident situations by means of one or more traffic features which significantly differ under both situations. A traffic feature can be a single variable like occupancy or a function of variables which is more sensitive to incident patterns. To a large extent, these algorithms deal with existing traffic patterns and pay no attention to past values of the features. Alternatively, smoothing algorithms utilize short-term forecasting techniques to identify sudden changes in traffic stream behavior which occur during incidents. This is similar to the problem of signal detection with noise present in communication systems. An important advantage of smoothing algorithms is that past trends of traffic features

may well forecast peak-period congestion, whereas incident congestion would be unexpected.

Cross-Sectional or Lane-Aggregated Algorithms. Cross-sectional algorithms are usually structured to compare traffic features of individual lanes at the same detector station. Basically, they attempt to identify whether the different freeway lanes are equally utilized by traffic or there is significant under-utilization of one or more lanes because of blockages. On the contrary, lane-aggregated algorithms deal with traffic features averaged across all lanes at a particular detector station.

Empirical or Simulation-Based Algorithms. Empirical algorithms can be built by utilizing mathematical and statistical techniques as well as a set of empirical data to serve as a basis for evaluating the relationships represented by the algorithm. Usually, however, such data is collected during peak-periods where the likelihood of occurrence of traffic incidents is relatively high. These situations are generally characterized by demand which exceeds the reduced capacity, and by extreme changes in traffic stream measurements. On the other hand, when demand is relatively low, as it does during off-peak periods, the sudden changes in traffic stream due to an incident become localized and unnoticeable in traffic data. Also, the frequency of incidents in light traffic conditions is very small, which makes it difficult to collect a reasonably adequate incident data base. In such cases, Monte Carlo simulation techniques can be utilized to generate traffic patterns similar to those occurring under incident conditions. Simulated traffic data may then be used in constructing and evaluating detection algorithms. The development of a simulation model to represent freeway incidents under light traffic

conditions is therefore an important aspect in building detection algorithms, and it usually involves several trial models before an adequate representation is obtained.

5.2.3 Review of the Structures of Current Detection Algorithms

Since 1961, several automated incident detection algorithms have been proposed by researchers in the field of traffic surveillance and control. Table 5.1 presents most of these algorithms classified according to the typological dimensions discussed in the previous section. To facilitate the discussion the following notation will be used throughout:

- $\xi(i,t)$ = occupancy in percent at detector station i and minute t , averaged across all lanes.
- $\xi(i,j,t)$ = occupancy in percent at detector station i , lane j , and minute t .
- $q(i,t)$ = volume (vehicles per minute) at detector station i and minute t , averaged across all lanes.
- $q(i,j,t)$ = volume (vehicles per minute) at detector station i , lane j , and minute t .
- $u(i,t)$ = $q(i,t)/\xi(i,t)$, surrogate for speed at detector station i and minute t , averaged across all lanes.
- $E(i,t)$ = $[q(i,t)]^2/\xi(i,t)$, surrogate for kinetic energy at detector station i and minute t , averaged across all lanes.
- $E(i,j,t)$ = $[q(i,j,t)]^2/\xi(i,j,t)$, surrogate for kinetic energy at detector station i , lane j , and minute t .
- $i, i+1$ = upstream and downstream detector stations, respectively.

Among the earliest detection systems is the one installed by the Port of New York Authority in the Lincoln Tunnel [28]. The algorithm used was based on the identification and tracking of individual vehicles within each of three sections of the tunnel by matching vehicle length

TABLE 5.1

INCIDENT DETECTION ALGORITHMS

Detection Algorithm	Typology Dimensions							
	Station Algorithm	Subsystem Algorithm	Pattern-Recognition Algorithm	Smoothing Algorithm	Cross-Sectional Algorithm	Lane-Aggregated Algorithm	Empirical Algorithm	Simulation-Based Algorithm
Port of New York Authority Algorithm		XX	XX			XX	XX	
Texas Transportation Inst. Algorithms	XX	XX	XX		XX	XX	XX	
California Algorithm	XX	XX	XX			XX	XX	
Technology Service Corp. Algorithms	XX	XX	XX			XX	XX	
Bayesian Algorithm		XX	XX			XX	XX	
Exponential Smoothing Algorithms	XX	XX		XX	XX	XX	XX	
Standard Normal Deviate Algorithm	XX			XX			XX	
ARIMA Algorithms	XX	XX		XX	XX	XX	XX	
Low-Volume Algorithms		XX	XX			XX		XX

patterns at each two consecutive sections. At the beginning of time period $(t+1)$, the number of vehicles present in a section can be expressed as

$$Y_{t+1} = Y_t + I_t - O_t \quad (5.1)$$

where:

Y_t = number of vehicles in the section at the beginning of time period t ,

I_t = traffic volume input to the section during time period t ,
and

O_t = traffic volume output from the section during time period t .

Traffic density at each section of the tunnel was estimated using equation 5.1, and incidents were predicted for unusually large values of density. This method was later improved by application of the Kalman filtering technique so as to account for occasional detector miscounts. The utility of this algorithm, however, is limited to tunnel operations where few lane changes take place.

In 1968, the Texas Transportation Institute, TTI, conducted incident detection research on the Lodge Freeway in Detroit as part of their study program. During their research, Courage and Levin [15] explored six detection algorithms in attempt to characterize traffic operations during a capacity-reducing incident. One of these algorithms is based on entrance ramp metering rates, while the other five algorithms employ different functions of one-minute volume and occupancy observations as detection features. These features are: station kinetic energy, station discontinuity, subsystem energy, subsystem shock wave, and subsystem discontinuity. A brief description of each of these features is given below:

Station Kinetic Feature, $E'(i,t) = E(i,t)/E_{\max}(i,t)$ (5.2)

where $E_{\max}(i,t)$ is the maximum kinetic energy determined by fitting representative peak-period data to a linear speed-occupancy model [15]. Unusually low values of this feature, those below a predetermined threshold level may indicate an incident.

$$\text{Station Discontinuity Feature} = \frac{(n-1) \{ \text{Min}[E'(i,j,t)]_{j=1}^n \}}{\sum_{j=1}^n E(i,j,t) - \text{Min}[E'(i,j,t)]_{j=1}^n} \quad (5.3)$$

where n is the number of lanes. Significantly low values of this feature actuate an incident signal.

$$\text{Subsystem Energy Feature} = E'(i,t) - E'(i+1,t) \quad (5.4)$$

Unusually large numerical differences in normalized kinetic energy values observed at adjacent detector stations may register an incident presence in that subsystem.

$$\text{Subsystem Shock Wave Feature} = q(i,t) - q(i+1,t) \quad (5.5)$$

Significantly large numerical differences in the volume recorded at adjacent detector stations may indicate the presence of an incident in that subsystem.

Subsystem Discontinuity Feature =

$$\left\{ \left[\frac{u(i,t)}{u_f(i,t)} - \frac{u(i+1,t)}{u_f(i+1,t)} \right]^2 + \left[\frac{\xi(i,t)}{\xi_j(i,t)} - \frac{\xi(i+1,t)}{\xi_j(i+1,t)} \right]^2 \right\}^{\frac{1}{2}} \quad (5.6)$$

where u_f is the free speed, and ξ_j is the jam occupancy. Significantly large values of this feature may indicate an incident presence in that subsystem.

For each of the above five features, a threshold of detection is determined such that its value would be exceeded only one percent of the time during non-incident operations. The process of estimating these thresholds is known as "calibration," and as one may readily expect the threshold values may vary by location and time of day. Courage and Levin concluded that their algorithms demonstrated some ability to detect incidents, and that despite the exhibited high false-alarm rates, the algorithms may merit further consideration.

In the late 1960's, the California Department of Transportation developed an incident detection algorithm for use on the Los Angeles freeway system. The California algorithm consists of three sequential tests based on spatial and temporal changes in occupancy features over short periods of time at the upstream and downstream detector stations of a freeway subsystem. An incident signal is actuated only when the threshold values for all three features are exceeded, indicating that the scenario of events associated with a typical capacity-reducing incident has occurred. Like the TTI algorithms, the California algorithm requires a calibration process to estimate the threshold values which may vary by location and time of day. The following is the structure of the California algorithm:

$$(1) \quad \xi(i,t) - \xi(i+1,t) = X_1 \geq T_1 \quad (5.7)$$

$$(2) \quad \frac{\xi(i,t) - \xi(i+1,t)}{\xi(i,t)} = X_2 \geq T_2 \quad (5.8)$$

$$(3) \quad \frac{\xi(i+1,t-B) - \xi(i+1,t)}{\xi(i+1,t-B)} = X_3 \geq T_3 \quad (5.9)$$

where:

B = five minutes in the original structure, and two minutes in the modified structure, and

T_1, T_2, T_3 = threshold levels for detection.

The first test examines the current difference in occupancy between two successive detector stations. If this difference is unusually large, it is likely that either a bottleneck or a capacity-reducing incident has occurred within the freeway subsystem. The second test is a normalization of the occupancy difference used in the first test to assure the existence of significant difference in the state of traffic operations at the two detector stations. Finally, the third test evaluates the relative temporal change in downstream occupancy over the past two minutes. Since a short-term decrease in occupancy is more characteristic of an incident than a bottleneck, the third test serves mainly as a discrimination tool between both situations.

More recently, the Technology Service Corporation in Santa Monica, California, conducted an empirical study on incident detection using data obtained from the Los Angeles and Minneapolis freeway surveillance systems. Payne and Tignor [78,61] defined several detection algorithms based upon extensions of the structure of the California algorithm. These algorithms have been referred to as "decision-tree algorithms." In general, the defined algorithms were intended to improve the performance of the California algorithm by including a persistence check and a compression wave presence check. To lower the false-alarm rate, the persistence check attempts to identify the short-lived disturbances in incident-free traffic by requiring discontinuity in traffic operations to continue for two or more time intervals. The compression wave presence check, on the other hand, utilizes the traffic feature used in the third test of the California algorithm, but in a more

elaborate way, to distinguish between traffic bottlenecks and capacity-reducing incidents in heavy traffic conditions.

Levin and Krause [49] proposed the application of the Bayesian concepts to the problem of incident detection. They used representative incident and incident-free data obtained from the Chicago freeway system to determine mathematical expressions for the distribution of the occupancy feature X_2 utilized in the second test of the California algorithm. Also, they estimated the prior probabilities of capacity-reducing incidents based on historical information from the highway patrol reports. The conditional probabilities of having an incident given that a sequence of incident signals has been generated by the computer were then calculated and used to calibrate the threshold level T_2 of equation 5.8. An incident was detected if the occupancy feature X_2 exceeded T_2 . The effectiveness of the algorithm was evaluated on a comparative basis with the California algorithm and one of the decision-tree algorithms using a sample of 17 incidents which occurred on the Kennedy Freeway in Chicago. Levin and Krause found that the Bayesian algorithm compared favorably with the other two algorithms included in their evaluation.

In addition to the pattern recognition algorithms discussed in the above sub-sections, detection algorithms based on smoothing techniques have been proposed and tested in some occasions. Whitson, et al. [86], suggested the use of a moving average of the most recent five minutes of volume data as a forecast of the next volume observation. When succeeding observations fell below the lower probability limits constructed two standard deviations away from the corresponding point

forecasts, an incident signal was actuated. This algorithm, however, was not fully evaluated with actual incident data.

Cook and Cleveland [13] conducted incident detection research on the Lodge Freeway in which they evaluated the California algorithm and five of the TTI algorithms. They also formulated 13 traffic features including the five features used with the TTI algorithms. All the 13 features were investigated with the double exponential smoothing model of equation 3.15 as a means for incident detection. A smoothing constant of 0.3 was used in calculating the forecasts. Based on the tracking signal of equation 3.17, an incident was detected when the successive values of this signal significantly deviated from zero. In computing the tracking signal, the mean absolute deviation was estimated by single exponential smoothing of the absolute values of forecast errors using a smoothing constant of 0.1. Similar to the TTI and the California algorithms, detection thresholds for the tracking signal require calibration. Among all the detection algorithms evaluated by Cook and Cleveland using a sample of 50 representative incidents, the most effective algorithms were those based on exponentially smoothed station occupancy, station volume, and station discontinuity features. Joint application of more than one feature resulted higher detection rate, but at the expense of increased probability of false-alarms.

As an extension to the approach of incident detection based on smoothing techniques, Dudek, et al. [23], proposed the application of a five-minute moving standard normal deviate to lane occupancy and lane energy features. The standard normal deviate, SND, as defined by equation 3.11 attempts to recognize rapid changes in traffic features over

short periods of time in relation to the expected changes caused by normal fluctuations in traffic flow. Values of SND corresponding to 90 percent detection rate and one percent false-alarms were used as thresholds of detection which, of course, need to be calibrated for different locations and times of day. The effectiveness of the SND algorithm was partially evaluated by Dudek, et al., using a sample of 35 incidents that took place on the Gulf Freeway in Houston during moderate and heavy traffic operations. Detection rate was found to be as high as 92 percent with one percent false-alarms.

Finally, research on incident detection under low-volume conditions has been pursued in limited studies based on simulation. As discussed earlier in this chapter, if the reduced capacity is not below the level of approaching traffic volume, it becomes difficult to distinguish between incident and non-incident traffic patterns. The reason for this difficulty is that flow discontinuities under low-volume conditions are less noticeable from those occurring under high-volume conditions, and stoppage waves are not likely to propagate [23]. The earliest suggestion is apparently that of Barker [3] who explored the discontinuity in traffic operations on the Connecticut turnpike using input-output analysis technique. Further work along these lines was done by Sakasita and May [69] based on incident data generated by a Microscopic Monte Carlo simulation model for a 1.5-mile freeway section. More recently, Dudek, et al. [20], applied time-scanning and event-scanning methods to individual vehicle counts, and produced a family of curves for determining detector-spacing using simulation. In general, the common denominator of these detection algorithms is that they use vehicle storage as a

feature. Despite their attractive nature, these algorithms are subject to some operational problems such as the presence of entrance or exit ramps between detector stations, the frequent vehicle stoppages on shoulders, and detector miscounts. In particular, these algorithms may have merit for tunnels, bridges, and freeway sections without ramps or shoulders, but their application elsewhere requires considerable refinement.

5.3 Incident Data Base

At the beginning of this research, incident data from the Los Angeles, Minneapolis, and Detroit freeway systems were available for subsequent analysis. The Los Angeles and Minneapolis data, described in [62], were obtained from the National Technical Information Service on three magnetic tapes. However, the documentation of the incident data contained on these tapes was incomplete, and much information needed for the analysis were missing. Therefore, it was decided to limit the study to the remaining incident data obtained from Detroit. In particular, this is the same data utilized by Cook and Cleveland [13] in their development and evaluation of the exponential smoothing algorithms.

The Detroit incident data base used in this research was collected from a 2-mile section of the National Proving Ground for Freeway Surveillance, Control, and Electronic Aids in the John C. Lodge Freeway. This section of the Lodge Freeway contained four detector stations, (Seward, Chicago, Calvert, and Glendale), located at variable spacing between 1460 and 4815 feet. The detection system consisted of fourteen

television cameras and a closed network of ultrasonic vehicle presence detectors connected to an IBM 1800 computer.

The data collection period covered a total of 13 months from December 1968 to December 1969. A total of 50 lane blockages consisting of 18 accidents, 28 stalls and breakdowns, 2 instances of debris, and 2 short maintenance operations were included in the incident data base. Most of the incidents took place during the afternoon peak-period, from 2:30 to 6:30, in the lane adjacent to the median with duration times ranging from 1 to 19 minutes. Observed volume levels which prevailed prior to the occurrence of incidents ranged from 1200 to 2000 vph per lane, while occupancies varied between 9 to 45 percent. During incidents, the prevailing flow levels were reduced by an average of 21 percent.

For each of the 50 incidents, historical records of one-minute average volume and occupancy data obtained from detectors were aggregated by the computer over all lanes, and recorded on punch cards for the nearest upstream and downstream detector stations. The historical records commenced about ten minutes preceding the television log time of occurrence of an incident, and continued until about ten minutes after the logged time of removal of the incident or until the dissipation of congestion. When two incidents occurred on the same day and in the same freeway subsystem, the data was continuously compiled throughout the interval between the two incidents. Thus, there was a net of 42 data records which contained the information on the 50 incidents. These 42 records represented a total of 1692 minutes of observations on traffic data associated with capacity-reducing incidents.

The time of occurrence of each of the 50 incidents was determined by inspecting the time the incident occurred as recorded in the television surveillance logs and the onset of incident-generated congestion as judged by the incident traffic data obtained from detectors [13]. It was noted that television surveillance detected the incidents from one to three minutes sooner than would the onset of congestion. For consistency, the incident time of occurrence was estimated from the onset of congestion as determined by traffic flow characteristics.

In conclusion, it is believed that the sample of the 50 on-freeway incidents used in this study is representative of the typical lane blockage incidents which take place during the peak periods on the freeways of most urban areas. However, the problem with applying the ARIMA (0,1,3) model to the individual time series of these 50 incidents is the limited number of observations contained in most of these series. Only three of the time series of incident data contained more than 100 observations. Two of these series were recorded at the Seward detector station, while the third series was recorded at the Calvert detector station.

5.4 ARIMA Model Application to Incident Detection

As indicated earlier, freeway incidents are usually associated with sudden changes in operating flow states which typically appear as pulsed input in the profiles of time series of traffic features [13]. The representative volume and occupancy plots depicted in Figure 5.1 tend to illustrate this concept. On the other hand, past trends of traffic features, those which can be determined by smoothing out the random variations, can predict the development of peak-period or bottleneck

congestion, whereas the unexpected impact of incidents cannot be reflected by these trends. Therefore, it is sensible to apply short-term forecasting techniques to detect incident generated irregularities in time series of traffic features.

Previously developed detection algorithms based on smoothing techniques, notably the exponential algorithms and the standard normal deviate algorithm, utilize ad hoc forecasting techniques to provide point forecasts of traffic features for one time interval in advance. Although these algorithms were found comparable if not superior to other pattern recognition algorithms, there is no solid theoretical reasoning for the particularly chosen forecasting techniques. In addition, a common weakness shared by all the detection algorithms described earlier is that they require calibration to determine the appropriate threshold levels for detection. Given the many factors contributing to variations in traffic flow, such as time of day, pavement and environmental conditions, several threshold levels need to be calibrated which require the development of frequency distributions for each set of factors at each freeway detector station. Furthermore, it may be difficult to account for all the factors involved. One way to overcome this problem is to use real-time estimates of the variability in traffic stream features as detection thresholds. Thus, false-alarms could possibly be reduced, and the thresholds would be responsive to the different factors. The approach used in this research is based upon constructing probability limits for the point forecasts rather than considering the point forecasts themselves. Importantly, this allows for the uncertainty associated with the forecasts to be included in the detection decisions, and

eliminates the need for threshold calibration. An incident is detected if the observed feature value lies outside the probability limits constructed two standard deviations away, for example, from the corresponding point forecasts.

A total of eight traffic features were chosen from those proposed by Cook and Cleveland [13], and by Courage and Levin [15], for application with the ARIMA (0,1,3) model. Using the same notation as in Section 5.2.3, these features are:

$$\begin{aligned}
 (1) \quad \text{Station Volume} &= q(i,t) \\
 (2) \quad \text{Station Occupancy} &= \xi(i,t) \\
 (3) \quad \text{Station Speed} &= u(i,t) \\
 (4) \quad \text{Station Energy} &= E(i,t) \\
 (5) \quad \text{Subsystem Volume} &= q(i,t) - q(i+1,t) \\
 (6) \quad \text{Subsystem Occupancy} &= \xi(i,t) - \xi(i+1,t) \\
 (7) \quad \text{Subsystem Speed} &= u(i,t) - u(i+1,t) \\
 (8) \quad \text{Subsystem Energy} &= E(i,t) - E(i+1,t)
 \end{aligned}
 \tag{5.10}$$

The time series of the station volume and station occupancy for the 50 incidents described in Section 5.3 were readily available in the form of 42 data records. Values of the other six features were computed minute-by-minute spanning the duration of the 50 incidents. Thus, there were 42 incident time series records for each of the above eight features. The ARIMA (0,1,3) model of equation 4.1 was then fitted to the different time series of these features, and diagnostic checks for the model adequacy were performed by inspecting the resulting residuals. It was found that the ARIMA (0,1,3) model did adequately represent the stochastic process generating all of the eight features.

Table 5.2 gives the ranges of values of the estimated moving average parameters for the 42 different time series of each of the features. As indicated in Section 5.3, the 42 time series of incident data were recorded during the afternoon peak periods of 42 different days at different detector stations. Hence, the variations in the estimated moving average parameters of Table 5.2 are due to location attributes and time. The variation in the parameters of the ARIMA(0,1,3) model by location was confirmed earlier by the results of Tables 4.2 and 4.3 using incident-free time series data which was recorded at the same time but at different detector stations. To have a closer look on the variation due to time, Table 5.3 gives the values of estimated moving average parameters of two incident records observed on 8/11/69 and 12/4/69, at the Seward detector station of the Lodge Freeway. The historical records of these incidents are the only available records which have reasonably large number of observations. It is evident from the results of Table 5.3 that the moving average parameters change over time. Although similar variation was noticed earlier in Figure 4.8, it was not possible to accurately conclude that the parameters vary over time because of the limited number of observations used in parameter estimation. However, with the confirmation noted by the results of Table 5.3, it is clear at this point that the parameters of the ARIMA(0,1,3) model vary from time period to another. As discussed earlier, parameter updating can be done occasionally, for example, at the beginning of peak and off-peak periods. Again, parameter updating was not explored in this research, in part because the available data consisted of afternoon peak-period time series only.

TABLE 5.2
MOVING AVERAGE PARAMETERS OF TRAFFIC FEATURES

Traffic Features	No. of Data Sets	Total No. of Observations	Moving Average Parameters		
			$\bar{\epsilon}_1$	$\bar{\epsilon}_2$	$\bar{\epsilon}_3$
Station Volume	42	1694	0.2414± 0.7576	0.3003± 0.6504	0.1409± 0.7997
Station Occupancy	42	1694	0.2077± 0.6243	0.2425± 0.6544	0.2043± 0.7666
Station Speed	42	1694	0.0772± 0.7351	0.0428± 0.7339	0.0097± 0.9489
Station Energy	42	1694	0.0793± 0.6767	0.7820± 0.1524	0.1304± 0.7935
Subsystem Volume	42	1694	0.3864± 0.5480	0.2493± 0.7283	0.0302± 0.8624
Subsystem Occupancy	42	1694	0.2215± 0.6561	0.0863± 0.8364	0.1435± 0.7980
Subsystem Speed	42	1694	0.0038± 0.9736	0.1057± 0.6853	0.0199± 0.9127
Subsystem Energy	42	1694	0.0470± 0.9222	0.0034± 0.9502	0.0935± 0.8368

TABLE 5.3
 VARIATION IN MOVING AVERAGE PARAMETERS
 OVER TIME AT SEWARD STATION

Traffic Feature	Observation Date	Number of Observations	Moving Average Parameters		
			$\hat{\epsilon}_1$	$\hat{\epsilon}_2$	$\hat{\epsilon}_3$
Station Volume	Aug. 11, 1969	124	0.6298	0.2171	-0.1494
	Dec. 4, 1969	154	0.6893	0.1747	0.0930
Station Occupancy	Aug. 11, 1969	124	0.5935	0.0991	-0.1263
	Dec. 4, 1969	154	0.3959	0.2515	-0.0990
Station Speed	Aug. 11, 1969	124	0.4726	0.1980	-0.0045
	Dec. 4, 1969	154	0.6465	0.0985	-0.1114
Station Energy	Aug. 11, 1969	124	0.4934	0.2140	-0.0172
	Dec. 4, 1969	154	0.6541	0.0911	-0.1000
Subsystem Volume	Aug. 11, 1969	124	0.6453	0.2437	0.0900
	Dec. 4, 1969	154	0.6669	0.1759	0.1436
Subsystem Occupancy	Aug. 11, 1969	124	0.7508	-0.1209	0.0365
	Dec. 4, 1969	154	0.2497	0.1981	0.0868
Subsystem Speed	Aug. 11, 1969	124	0.6635	0.0956	0.2176
	Dec. 4, 1969	154	0.5497	0.0653	0.0019
Subsystem Energy	Aug. 11, 1969	124	0.6226	0.1965	0.1619
	Dec. 4, 1969	154	0.6659	0.1149	0.0081

Having estimated the model parameters and checked the adequacy of the form of the ARIMA (0,1,3) model, equation 4.5 was then used to calculate the one-step-ahead forecasts. The moving average parameters estimated from a particular series were used in calculating the forecasts for the same series. Furthermore, approximately 95-percent confidence intervals were constructed utilizing equation 3.40. Sample plots for two station and two subsystem features for the incident of Figure 5.1 are depicted in Figure 5.3. The broken lines shown in the same figure represent the approximately 95-percent probability limits of the point forecasts, while the shaded area indicates the incident duration. For clarity, the point forecasts were not plotted. As expected, each of the four features is noted to respond for this particular incident as indicated by the deviation of the observed feature value from the corresponding probability limits at the moment when congestion started to develop.

The effectiveness of the ARIMA detection algorithms was partially evaluated on a comparative basis with some of the previously developed algorithms using the station occupancy feature as an example. In particular, this feature dominated all the other features explored by Cook and Cleveland [13] in their analysis of the exponential algorithms. Three measures of effectiveness were employed as evaluation criteria. These are detection rate, false-alarm rate, and mean time lag to detection. From an operational standpoint, these three measures are of particular importance to any effective computer-supervised incident detection capability.

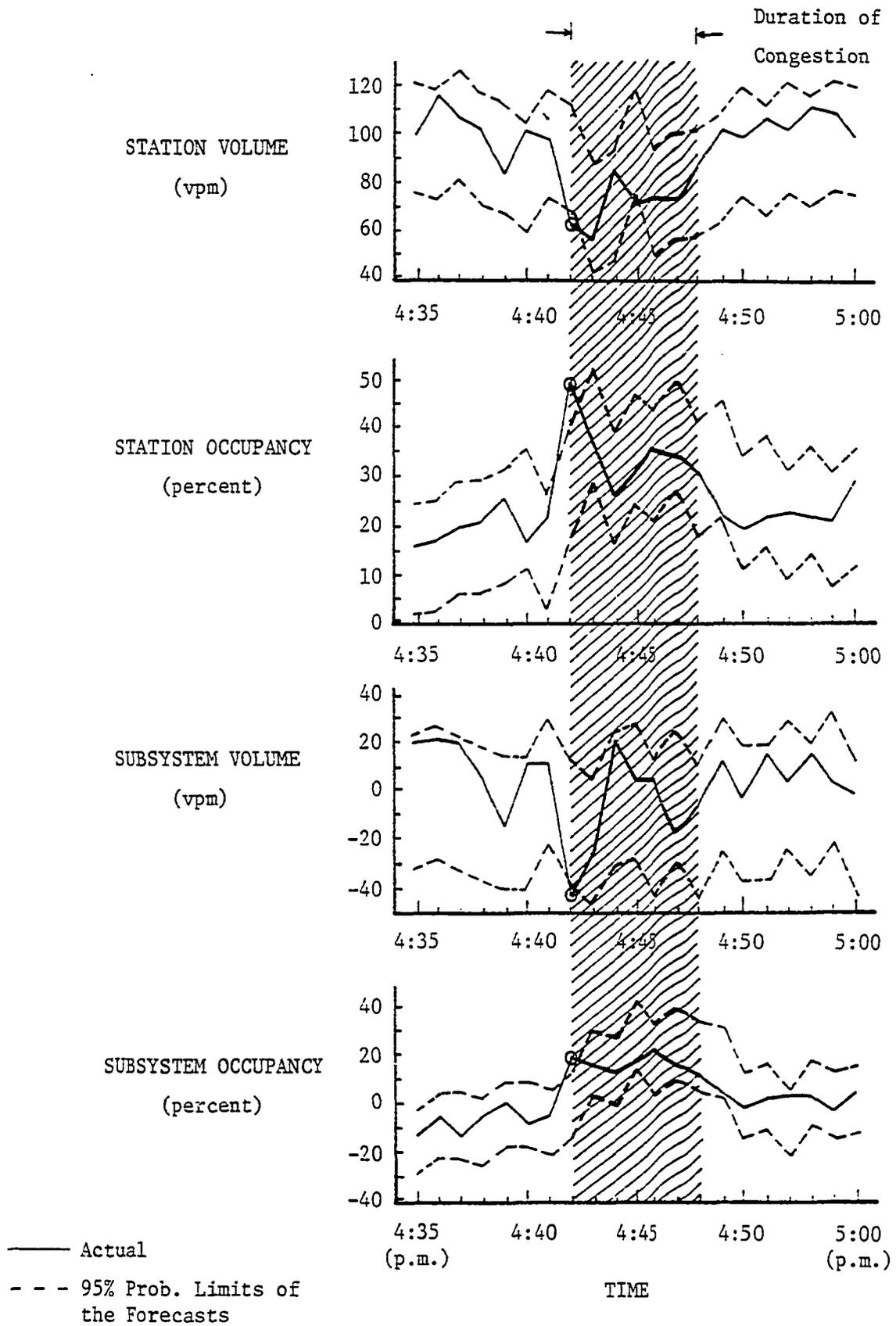


FIGURE 5.3

PERFORMANCE OF SOME TRAFFIC FEATURES
 UNDER AN INCIDENT

In determining the false-alarm rate, a total of 1554 minutes of non-incident observations preceding and following the 50 on-freeway incidents was used. Since these non-incident observations were closely associated with actual incidents, the estimated false-alarm rate using this sample may not be equivalent to the false-alarm rate estimated from representative incident-free data. Importantly, this same sample of 1554 minutes of non-incident observations was used by Cook and Cleveland [13] in evaluating the effectiveness of the exponential smoothing algorithms, the California algorithm, and five of the TTI algorithms. This provided the chance, without any bias in estimating the false-alarm rate, to compare the results of the ARIMA station occupancy algorithm with the results of those algorithms.

Figure 5.4 depicts the operating characteristic curve for the ARIMA station occupancy algorithm which was obtained by varying the width of the constructed confidence intervals from 3 to 5 standard deviations in increments of 1.0. This caused the false-alarm rate to range from nearly 2.5 percent to zero. The algorithm detected all of the 50 incidents with 1.4 percent false-alarms. For the purpose of comparison, results of the California, TTI, and exponential station occupancy algorithms were taken from the research findings of Cook and Cleveland [13] based on the same 50 incident data used in this study. These later results are shown in Figure 5.4. The operating points for the California and the five TTI algorithms were determined at threshold levels corresponding to one percent false-alarm rates, while the operating characteristic curve of the exponential station occupancy algorithm was obtained by varying the tracking signal thresholds from ± 1.50 to

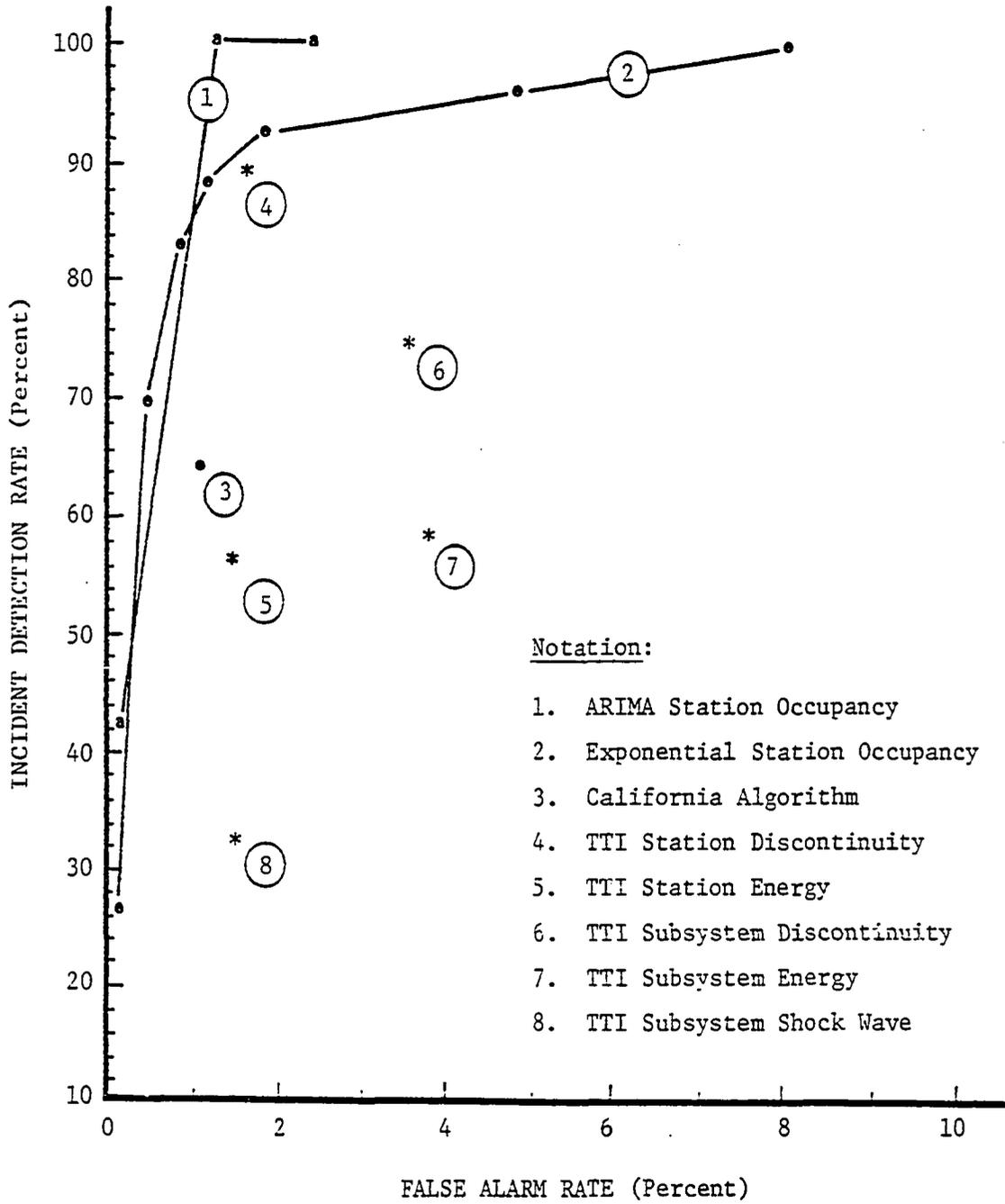


FIGURE 5.4

OPERATING CHARACTERISTIC CURVES FOR
INCIDENT DETECTION ALGORITHMS

±10.00 in 0.50 increments [13]. The performance of the ARIMA station occupancy algorithm is seen to be superior to all the other algorithms presented in Figure 5.4 by virtue of generating higher detection rates at the approximately zero, 1.4, and 2.5 percent false-alarm levels. Also, at higher levels of detection, between 90 and 100 percent, the ARIMA station occupancy algorithm has the lowest false-alarm rate.

Information on the mean and standard deviation of the time-lags to detection of the ARIMA station occupancy algorithm is presented in Table 5.4 at the low (near zero percent), medium (one to two percent), and high (about six percent) false alarm levels. As noted, there is no information on the time-lags at the high false-alarm level since the false-alarms did not exceed 2.5 percent. To help compare the relative effectiveness of the ARIMA station occupancy algorithm, the means and standard deviations of the time-lags for the exponential station occupancy and the TTI station discontinuity algorithms were taken from the results of Cook and Cleveland [13]. As one may readily expect, the mean time-lag decreased as the number of false-alarms increased. The ARIMA station occupancy algorithm which enjoyed the highest detection rate tended also to have the shortest mean time-lag, thus, eliminating the need for considering a trade-off between total detections and mean time-lag to detection.

5.5 Summary and Implications for Microcomputers

Although the evaluation of the ARIMA detection algorithms was limited in its extent, it did demonstrate the potential improvements which can be gained by applying the ARIMA (0,1,3) model to

TABLE 5.4

MEAN TIME LAGS TO DETECTION

Incident Detection Algorithm	False-Alarm Level		
	Low	Medium	High
ARIMA Station Occupancy	1.33 min. (1.58 min.)	0.39 min. (1.32 min.)	—
Exponential Station Occupancy*	1.46 min. (2.78 min.)	0.74 min. (1.40 min.)	0.35 min. (0.81 min.)
TTI Station Discontinuity*	3.13 min. (5.62 min.)	2.07 min. (4.05 min.)	0.83 min. (1.11 min.)

* Source: Reference 13

computer-supervised incident detection. These improvements include higher detection rate, lower false-alarms and shorter response time. In addition, the fact that the ARIMA detection algorithms do not require threshold calibration makes them more attractive algorithms from an operational standpoint. However, the only single difficulty with the adaptation of the ARIMA detection algorithms in automated incident detection systems is the computational requirements needed for estimating and updating the moving average parameters at each detector station. It is interesting in this regard to note that the parameters are estimated and updated using non-linear least-squares subroutines which are available in many computer libraries. In particular, program ESTIMATE which was used in this study needs 130K of computer core capacity for a time series length of 3000 observations. For incident detection purposes, a small portion of this computer core capacity will be actually utilized since a time series length of 100 observations is sufficient.

It is important to note that there are increasing trends toward the distribution of computational capabilities and the adaptation of microcomputers in large scale control systems. The parameters of the ARIMA (0,1,3) model can be estimated and occasionally updated using a central computer of medium size or a number of distributed microcomputers. Then, the estimated or updated parameters can be transmitted through cables or telephone lines to programmed microprocessors which compute the required forecasts and perform the logical tests of the ARIMA detection algorithms. One microprocessor can serve one or more detector stations depending on its computational capabilities. Thus, it is believed that the rapid advances in computer and

communication technologies can economically help overcoming any computational difficulties associated with the application of the ARIMA detection algorithms.

CHAPTER VI

CONCLUSIONS AND SUGGESTED RESEARCH

6.1 Conclusions

The research described by this dissertation has focused on investigating the dynamic nature of freeway traffic time series data using the analysis techniques described by Box and Jenkins. Throughout the investigation, two major developments have been achieved: first, the development of predictor model for traffic stream variables, and second, the development of computer algorithms for detecting accidents and other capacity-reducing incidents which are typical occurrences on urban freeways. The analysis was based on surveillance data obtained from the Los Angeles, Minneapolis, and Detroit freeway systems during the afternoon peak periods.

It has been indicated that most of the control strategies made by computer-supervised freeway surveillance and control systems should allow for the operational changes they are likely to cause in the immediate future, through the incorporation of real-time forecasts of traffic variables into the control logic. Other important applications which require short-term forecasting include incident detection, intersection control, automated transportation systems, and improved methods for handling traffic at maintenance and construction sites which involve

lane closures or during special events such as concerts and athletic events. Regardless of the type of application, to be effective, the development of efficient and informative forecasting techniques is extremely important.

During the review of current approaches to forecasting real-time traffic systems, it was apparent that these approaches are ad hoc in their nature, and that none of them has its solid theoretical justification. In contrast, the finite parameter autoregressive integrated moving average processes form a powerful and broad class of potential models for representing the dynamics of traffic time series. The primary virtue of this class of models is that the eventual form of the forecasting model is determined by the properties of the series in hand and, among other linear models, the resulting forecasts are optimal in terms of having minimum mean square error.

Application of the Box-Jenkins time series modeling framework to a total of 166 traffic volume and occupancy series, representing more than 27,000 minutes of observations, has revealed that the stochastic process generating traffic stream measurements can be best described by an ARIMA (0,1,3) model. Although the parameter estimates varied from one detector station to another, within and/or between the different freeway systems, there has been strong evidence for the transferability of the form of the model. The differences in parameter estimates arise from variations in driving situations, geometrics, and pavement and environmental conditions. Similar confirmation of the form of the model has been found with varying detector configurations and data aggregation intervals.

The theoretical discussion coupled with empirical analysis has demonstrated the operational utility of the ARIMA (0,1,3) model in providing real-time forecasts of freeway traffic stream variables at modest computer storage and computational requirements. Since the moving average parameters change over time, some improvements in the forecasts could be achieved by occasionally updating these parameters, for example, at the beginning of peak and off-peak periods. It is not necessarily warranted, however, to update the parameters in real-time for each new observation interval. Parameter updating requires the execution of non-linear least-squares routines which can be performed using microcomputers. With the increasing trend toward the distribution of computational capabilities and the adaptation of microcomputers in freeway surveillance and control systems, the parameters of the ARIMA (0,1,3) model can be estimated and occasionally updated using a central computer of medium size or a number of distributed microcomputers. Then, the estimated or updated parameters can be transmitted through cables or telephone lines to programmed microprocessors which compute the required forecasts. One microprocessor can serve one or more detector stations depending on its computational capabilities. In addition, it has been found that the ARIMA (0,1,3) model can be efficiently used with other computer programs to simulate the performance of complex traffic systems when new operational designs are implemented without the need to observe these systems for long periods of time.

Three of the widely used ad hoc forecasting models were evaluated on a comparative basis with the ARIMA (0,1,3) model. These are the moving average, the double exponential smoothing, and the Trigg

and Leach models. The mean absolute error, MAE, and mean square error, MSE, were used as evaluation criteria. Results of the moving average model indicated that both MAE and MSE increase as the number of past observations utilized in calculating the moving average gets larger. The best results of the double exponential smoothing model were associated with small values of the smoothing constant, while the Trigg and Leach adaptive scheme for updating the smoothing constant provided no improvement in the forecasts. In general, it has been concluded that the ARIMA (0,1,3) model is superior to the other models included in the evaluation by virtue of its more accurate representation of the stochastic process generating traffic data.

Using the ARIMA model developed earlier, it has been determined that freeway capacity-reducing incidents can be detected by the sudden and pulsed changes they generate in traffic stream time series data. In this context, eight traffic features sensitive to incident situations were compiled minute-by-minute spanning the duration of a sample of 50 on-freeway incidents. The form of the ARIMA (0,1,3) model was found representative of the different time series of these features. An incident was detected if the observed feature value laid outside the probability limits constructed two standard deviations away from the corresponding point forecasts. This approach eliminated the need for threshold calibration and allowed for the uncertainty associated with the forecasts to be included in the detection decisions.

The station occupancy feature was selected as an example to comparatively evaluate the performance of detection algorithms based on the ARIMA (0,1,3) model with some of the previously developed algorithms.

Three measures of effectiveness were used in the comparative evaluation: detection rate, false-alarm rate, and mean and variance of the time-lag to detection. In particular, the ARIMA station occupancy algorithm detected all of the 50 representative incidents at 1.4 percent false-alarm level. Based on the overall evaluation results, it has been concluded that the ARIMA station occupancy algorithm dominated the one developed for use by the Los Angeles Area Freeway Surveillance and Control Project, the five developed by Courage from the Texas Transportation Institute, and the exponential station occupancy algorithm which was found by Cook and Cleveland to be the best exponential detection algorithm. In addition to having the highest detection rate, the ARIMA station occupancy algorithm was found to have the shortest mean time-lag, thus, eliminating the need for considering a trade-off between total detections and mean time-lag to detection.

Despite the fact that this study has dealt primarily with freeway vehicular traffic, it is believed that the kind of approach presented has the potential for much broader applications in the field of transportation engineering. In many situations, discrete time series analysis could be successfully used to model the parameters of transportation systems which evolve over time. To a large extent, the limitations imposed on the range of applications are due to the lack of familiarity of many transportation analysts with the proper techniques. It is hoped that this dissertation will stimulate those who perform research in the transportation field to think along new lines utilizing the time series analysis approach to solving important problems in transportation research.

6.2 Prospects for Future Research

It was suggested in Chapter IV that some improvements in the forecasting performance of the developed ARIMA model may be achieved if the model parameters are occasionally updated over time. Research is required to evaluate the possibility of such improvements and, if any, to determine the optimal updating strategy which should be used. In a related concern, there is a need for developing transfer function models which could be efficiently applied to updating the parameter estimates. It is also important to investigate the magnitude of accuracy of the required forecasts relative to the proposed applications.

Further work on developing and evaluating incident detection algorithms based on the time series analysis approach presented in Chapter V is recommended. In addition, there is a need to explore the application of the developed ARIMA model in ramp signal as well as intersection control strategies.

Another area of interest is the application of microcomputers to freeway surveillance and control systems. Recent developments in microcomputer technology have resulted in increased computational capabilities at reduced cost. Over the past few years, some applications of microcomputers in ramp metering control and intersection control have been successfully made. Research is needed to expand these applications and explore the feasibility of using microcomputers in other aspects of traffic surveillance and control such as incident detection and variable message signs. It is of particular interest to program the developed ARIMA model on special purpose microprocessors for use in incident detection, ramp metering control, and variable message signs.

REFERENCES

1. Anderson, R. J., "A Revolution in Electronics." Traffic Engineering, Vol. 47, No. 4, 1977, pp. 17-18.
2. Athol, P., "Interdependence of Certain Operational Characteristics within a Moving Traffic Stream," Highway Research Record 72, 1965, pp. 58-87.
3. Barker, J. L., "Determination of Discontinuities in Traffic Flow as a Factor in Freeway Operational Control," Traffic Engineering, Vol. 31, No. 11, 1961, pp. 11-17.
4. Bartlett, M. S., "The Spectral Analysis of Point Processes," Journal of the Royal Statistical Society B., Vol. 25, 1963, pp. 264-296.
5. Bartlett, M. S., "Some Applications of Multivariate Point Processes," Stochastic Point Processes: Statistical Analysis, Theory and Applications, Papers presented at a conference held at the IBM Center, New York, on August 2-7, 1971, (Edited by: Lewis, P. A. W.), Wiley Interscience, Inc., New York, 1972, pp. 136-150.
6. Box, G. E. P., and Jenkins, G. M., "Some Recent Advances in Forecasting and Control," Applied Statistics, Vol. 1, No. 17, 1968, pp. 99-109.
7. Box, G. E. P., and Jenkins, G. M., Time Series Analysis: Forecasting and Control, Revised Edition, Holden-Day, Inc., San Francisco, 1976.
8. Box, G. E. P., and Pierce, D. A. "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models," Journal of the American Statistical Association, Vol. 65, No. 332, 1970, pp. 1509-1525.
9. Brown, R. G., Smoothing and Prediction of Discrete Time Series, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963.
10. Chandler, R. E., Herman, R., and Montrol, E. W., "Traffic Dynamics: Studies in Car Following," Operations Research, Vol. 6, No. 2, 1958, pp. 165-184.

11. Chow, W. M., "Adaptive Control of Exponential Smoothing Constant," Journal of Industrial Engineering, Vol. 16, No. 5, 1965, pp. 314-317.
12. Clayton, A. J., "Road Traffic Calculations," Journal of the Institution of Civil Engineering, No. 16, 1941, pp. 247-284.
13. Cook, A. R., and Cleveland, D. E., "Detection of Freeway Capacity-Reducing Incidents by Traffic Stream Measurements," Transportation Research Record 495, 1974, pp. 1-11.
14. Cook, A. R., "The Management of Urban Freeway Traffic," Canadian Journal of Civil Engineering, Vol. 1, No. 2, 1974, pp. 141-149.
15. Courage, K. G., and Levin, M., A Freeway Corridor Surveillance, Information, and Control System, Research Report 488-8, Texas Transportation Institute, Texas A&M University, College Station, Texas, December 1968, (cited by Reference 13).
16. Cox, D. R., and Lewis, P. A. W., The Statistical Analysis of Series of Events, John Wiley & Sons, Inc., 1966, pp. 186-209.
17. Darroch, J. N., and Rothery, R. W., "Car Following and Spectral Analysis," Proceedings of the Fifth International Symposium on the Theory of Traffic Flow and Transportation, (Edited by: Newell, G. F.), American Elsevier, Inc., New York, 1972, pp. 47-56.
18. Der, T. Y., Some Investigations of Forecasting Freeway Occupancies, M.Sc. Thesis, Illinois Institute of Technology, 1977.
19. Drew, D. R., Traffic Flow Theory and Control, McGraw-Hill, Inc., New York, 1968.
20. Dudek, C. L., Weaver, G. D., Ritch, G. P., and Messer, C. J., "Detecting Freeway Incidents under Low-Volume Conditions," Transportation Research Record 533, 1975, pp. 34-47.
21. Dudek, C. L., Huchingson, R. D., and Ritch, G., "Evaluation of a Prototype Warning System for Urban Freeways," Transportation Research Record 533, 1975, pp. 64-76.
22. Dudek, C. L., Messer, C. J., and Dutt, A. K., "Study of Detector Reliability for a Motorist Information System on the Gulf Freeway," Transportation Research Record 495, 1974, pp. 35-43.
23. Dudek, C. L., Messer, C. J., and Nuckles, N. B., "Incident Detection on Urban Freeways," Transportation Research Record 495, 1974, pp. 12-24.
24. Edie, L. C., Gazis, D. C., Helly, W., McNeil, D. R., and Weiss, G. H., Traffic Science, (Edited by: Gazis, D. C.), John Wiley & Sons, Inc., New York, 1974.

25. Edie, L. C., and Baverez, E., "Generation and Propagation of Stop-Start Traffic Waves," Vehicular Traffic Science, (Edited by: Edie, L. C., Rothery, R., and Herman R.), American Elsevier, Inc., New York, 1967, pp. 26-37.
26. Edie, L. C., "Discussion of Traffic Stream Measurements and Definitions," Proceedings of the Second International Symposium on the Theory of Traffic Flow, (Edited by: Almond, J.), The Organization for Economic Co-operation and Development, Paris, 1965, pp. 139-154.
27. Eldor, M., "Demand Predictors for Computerized Freeway Control Systems," Proceedings of the Seventh International Symposium on Transportation and Traffic Theory, (Edited by: Sasaki, T., and Yamaoka, T.), The Institute of Systems Science Research, Kyoto, Japan, 1977, pp. 341-370.
28. Foote, R. S., "Development of an Automatic Traffic Flow Monitor and Control System," Highway Research Bulletin 291, 1961.
29. Forbes, T. W., and Simpson, M. E., "Driver and Vehicle Response in Freeway Deceleration Waves," Transportation Science, Vol. 2, No. 1, 1968, pp. 77-104.
30. Franklin, R. E., "On the Flow-Concentration Relationship for Traffic," Proceedings of the Second International Symposium on the Theory of Traffic Flow, (Edited by: Almond, J.), The Organization for Economic Co-operation and Development, Paris, 1965, pp. 120-128.
31. Gazis, D. C., and Knapp, C. H., "On-Line Estimation of Traffic Densities from Time Series of Flow and Speed Data," Transportation Science, Vol. 5, No. 3, 1971, pp. 283-301.
32. Gazis, D. C., and Foote, R. S., "Surveillance and Control of Tunnel Traffic by an On-Line Digital Computer," Transportation Science, Vol. 3, No. 3, 1969, pp. 255-275.
33. Gazis, D. C., Herman, R., and Rothery, R. W., "Non-Linear Follow-the-Leader Models of Traffic Flow," Operations Research, Vol. 9, No. 4, 1961, pp. 545-567.
34. Gazis, D. C., Herman, R., and Potts, E. B., "Car-Following Theory of Steady-State Traffic Flow," Operations Research, Vol. 7, No. 4, 1959, pp. 499-505.
35. Goolsby, M. E., "Influence of Incidents on Freeway Quality of Service," Highway Research Record 349, 1971, pp. 41-46.
36. Greenberg, H., "An Analysis of Traffic Flow," Operations Research, Vol. 7, No. 1, 1959, pp. 79-85.
37. Haight, F. A., Mathematical Theories of Traffic Flow, Academic Press, Inc., New York, 1963.

38. Herman, R., and Rothery, R., "Frequency and Amplitude Dependence of Disturbances in a Traffic Stream," Proceedings of the Fourth International Symposium on the Theory of Traffic Flow, Karlsruhe, Germany, 1969, pp. 14-22.
39. Herman, R., and Rothery, R., "Propagation of Disturbances in Vehicular Platoons," Vehicular Science, (Edited by: Edie, L. C., Rothery, R., and Herman, R.), American Elsevier, Inc., New York, 1967, pp. 14-25.
40. Herman, R., Montrol. E., Potts, R., and Rothery, R., "Traffic Dynamics: Analysis of Stability in Car-Following," Operations Research, Vol. 7, No. 1, 1959, pp. 86-106.
41. Highway Research Board, "Highway Capacity Manual," Highway Research Special Report 87, 1965.
42. Hillegas, B. D., Houghton, D. G., and Athol, P. J., "Investigation of Flow-Density Discontinuity and Dual-Mode Traffic Behavior," Transportation Research Record 495, 1974, pp. 53-63.
43. Jenkins, G. M., and Watts, D. G., Spectral Analysis and its Applications, Holden-Day, Inc., San Francisco, 1968.
44. Kahn, H., and Wiener, A. J., The Year 2000: A Framework for Speculation on the Next Thirty-Three Years, Macmillan, Inc., New York, 1967.
45. Kolmogorov, A. N., Foundations of the Theory of Probability, Chelsea, Inc., New York, 1933.
46. Kreer, J. B., "A Comparison of Predictor Algorithms for Computerized Traffic Control Systems," Traffic Engineering, Vol. 45, No. 4, 1975, pp. 51-56.
47. Kreer, J. B., "Factors Affecting the Relative Performance of Traffic Responsive and Time-of-Day Traffic Signal Control," Transportation Research, Vol. 10, No. 2, 1976, pp. 75-81.
48. Lam, T., and Rothery, R., "The Spectral Analysis of Speed Fluctuations on a Freeway," Transportation Science, Vol. 4, No. 3, 1970, pp. 293-310.
49. Levin, M., and Krause, G. M., "Incident Detection-A Bayesian Approach," Paper Presented at the 57th Annual Meeting of Transportation Research Board, Washington, D.C., 1978.
50. Lighthill, M. J., and Whitham, G. B., "On Kinematic Waves II, A Theory of Traffic Flow on Long Crowded Roads," Proceedings of the Royal Society, London, Vol. 229A, 1955, pp. 317-345.

51. Marquardt, D. W., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," Journal of the Society for Industrial and Applied Mathematics, Vol. 2, 1963, pp. 431-441.
52. May, A. D., and Keller, H. E., "Non-Integer Car-Following Models," Highway Research Record 199, 1967, pp. 19-32.
53. Messer, C. J., Dudek, C. L., and Freibele, J. D., "Method for Predicting Travel Time and Other Operational Measures in Real-Time During Freeway Incident Conditions," Highway Research Record 461, 1973, pp. 1-16.
54. Mika, H. S., Kreer, J. B., and Yuan, L. S., "Dual Mode Behavior of Freeway Traffic," Highway Research Record 279, 1969, pp. 1-12.
55. Moskowitz, K., "Analysis and Projection of Research on Traffic Surveillance, Communication, and Control," NCHRP Report 84, 1970.
56. Nelson, C. R., "The First-Order Moving Average Processes: Identification, Estimation, and Prediction," Journal of Econometrics, Vol. 2, No. 2, 1974, pp. 121-141.
57. Nelson, R. C., Applied Time Series Analysis for Managerial Forecasting, Holden-Day, Inc., San Francisco, 1968.
58. Nerlove, M., and Wage, S., "On the Optimality of Adaptive Forecasting," Management Science, Vol. 10, No. 2, 1964, pp. 207-224.
59. Nicholson, H., and Swann, C. D., "The Prediction of Traffic Flow Volumes Based on Spectral Analysis," Transportation Research, Vol. 8, 1974, pp. 533-538.
60. Owen, D. B., Handbook of Statistical Tables, Addison-Wesley, Inc., Reading, Massachusetts, 1962.
61. Payne, H. J., and Tignor, S. C., "Freeway Incident Detection Algorithms Based on Decision Trees with States," Paper presented at the 57th Annual Meeting of the Transportation Research Board, Washington, D.C., 1978.
62. Payne, H. J., and Heifenbien, E. D., Freeway Surveillance Data, Report No. FHWA-RD-76-74, Federal Highway Administration, Washington, D.C., 1976.
63. Payne, H. J., Goodwin, D. N., and Teener, M. D., Evaluation of Existing Incident Detection Algorithms, Report No. FHWA-RD-75-39, Federal Highway Administration, Washington, D.C., 1975.
64. Polhemus, N. W., "Time Series Analysis of Local Fluctuations in Traffic Parameters," Transportation Research, Vol. 10, 1976, pp. 311-317.

65. Pollock, S. M., "Search Detection and Subsequent Action: Some Problems of the Interfaces," Operations Research, Vol. 19, No. 3, 1971, pp. 559-585.
66. Prigogine, I., and Herman, R., Kinetic Theory of Vehicular Traffic, American Elsevier, Inc., New York, 1971.
67. Richards, P. I., "Shock Waves on the Highway," Operations Research, Vol. 4, No. 1, 1956, pp. 42-51.
68. Roberts, S. D., and Reed, R., "The Development of a Self-Adaptive Forecasting Technique," AIIE Transactions, Vol. 1, No. 4, 1969, pp. 314-322.
69. Sakasita, M., and May, A. D., "Development and Evaluation of Incident-Detection Algorithms for Electronic Detector Systems on Freeways," Transportation Research Record 533, 1975, pp. 48-63.
70. Sakasita, M., and May, A. D., "Development and Evaluation of Incident-Detection Algorithms for Electronic Detector Systems on Freeways," Transportation Research Record 533, 1975, pp. 48-63, Citing: Kolenko, S. J., and Albegro, J. A., Disabled and Non-Disabled Vehicle Stoppage on the Congress Street Expressway, Chicago Area Expressway Surveillance project, Illinois Department of Transportation, 1962.
71. Sandell, N. R., Varaiya, P., Athans, M., and Safonov, M. G., "Survey of Decentralized Control Methods for Large Scale Systems," IEEE Transactions on Automatic Control, Vol. AC-23, No. 2, 1978, pp. 108-128.
72. Slutsky, E., "The Summation of Random Causes as the Source of Cyclic Processes," Econometrica, Vol. 5, 1937, pp. 105-146.
73. Smith, D. E., "Adaptive Response for Exponential Smoothing: Comparative System Analysis," Operations Research Quarterly, Vol. 25, No. 3, 1974, pp. 421-435.
74. Szeto, M. W., and Gazis, D. C., "Application of Kalman Filtering to the Surveillance and Control of Traffic Systems," Transportation Science, Vol. 6, 1972, pp. 419-439.
75. Tarnoff, P. J., "Data Errors in Urban Traffic Control Systems," Transportation Research Record 531, 1975, pp. 1-17.
76. Tarnoff, P. J., "Distribution of Data Processing Capabilities in an Urban Traffic Control System," Traffic Engineering, Vol. 44, No. 14, 1974, pp. 9-15.
77. Thomas, J. M., "Simple Ramp Metering Devices Reduce Rear-End Collisions," Traffic Engineering, Vol. 19, No. 7, 1969, pp. 22-25.

78. Tignor, S. C., and Payne, H. J., "Improved Freeway Incident Detection Algorithms," Public Roads, Vol. 41, No. 1, 1977, pp. 32-40.
79. Tignor, S. C., Operational Analyses and Improvements for Freeway Moving-Merge Systems, Ph.D. Dissertation, University of Michigan, Department of Civil Engineering, 1974.
80. Toong, H. D., "Microprocessors," Scientific American, Vol. 237, No. 3, 1977, pp. 146-160.
81. Trigg, D. W., and Leach, A. G., "Exponential Smoothing with an Adaptive Response Rate," Operations Research Quarterly, Vol. 18, No. 1, 1967, pp. 53-59.
82. U.S. Department of Transportation, Traffic Control Systems Handbook, Federal Highway Administration, Washington, D. C., 1976.
83. Wattleworth, J. A., "Accomplishments in Freeway Operations in the United States," Highway Research Record 368, 1971, pp. 26-32.
84. Welding, P. I., "Time Series Analysis as Applied to Traffic Flow," Proceedings of the Second International Symposium on the Theory of Road Traffic Flow, (Edited by: Almond, J.), The Organization for Economic Co-operation and Development, Paris, 1965, pp. 60-72.
85. Weinberg, M. I., "Traffic Surveillance and Means of Communicating with Drivers," Interim Report, NCHRP Report 9, 1964.
86. Whitson, R. H., Buhr, J. H., Drew, D. R., and McCasland, W. R., "Real-Time Evaluation of Freeway Quality of Traffic Service," Highway Research Record 289, 1969, pp. 38-50.
87. Winters, P. R., "Forecasting Sales by Exponentially Weighted Moving Averages," Management Science, Vol. 6, No. 3, 1960, pp. 324-342.
88. Wold, H., A Study in the Analysis of Stationary Time Series, Almqvist and Wiksell, Stockholm, 1938.
89. Wohl, M., and Martin, B. V., Traffic System Analysis for Engineers and Planners, McGraw-Hill, Inc., New York, 1967.
90. Wright, C. C., "Some Properties of the Fundamental Relations of Traffic Flow," Proceedings of the Fifth International Symposium on the Theory of Traffic Flow and Transportation, (Edited by: Newell, G. F.), American Elsevier, Inc., New York, 1972, pp. 47-56.
91. Yule, G. U., "on the Time-Correlation Problem, with Special Reference to the Variate-Difference Correlation Method," Journal of the Royal Statistical Society B, Vol. 84, 1921, pp. 497-526.

92. Yule, G. U., "On the Method of Investigating Periodicities in Distributed Series with Special Reference to Wolfer's Sunspot Numbers," Philosophical Transactions A, Vol. 226, 1927, pp. 267-298.