TRAFFIC CONGESTION FORECASTING MODEL FOR THE INFORM SYSTEM

FINAL REPORT

Department of Advanced Technology
Brookhaven National Laboratory
Upton, New York 11973

May 1995

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency, contractor or subcontractor thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency, contractor or subcontractor thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

**TRAFFIC CONGESTION FORECASTING MODEL
FOR THE INFORM SYSTEM**

FINAL REPORT

Principal Authors:

A. Azarm
S. Mughabghab
D. Stock

Contributors:

H. Thode
W. McShane**
A. Tzes**

Project Manager: R. Hall

Department of Advanced Technology
Brookhaven National Laboratory
Upton, New York 11973

May 1995

*This work was sponsored by the U.S. Department of Transportation in a facility supported by the U.S. Department of Energy.

**Polytechnic University

MASTER
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

for

ABSTRACT

This report describes a computerized traffic forecasting model, developed by Brookhaven National Laboratory (BNL) for a portion of the Long Island INFORM Traffic Corridor. The model has gone through a testing phase, and currently is able to make accurate traffic predictions up to one hour forward in time. The model will eventually take on-line traffic data from the INFORM system roadway sensors and make projections as to future traffic patterns, thus allowing operators at the New York State Department of Transportation (D.O.T.) INFORM Traffic Management Center to more optimally manage traffic. It can also form the basis of a travel information system.

The BNL computer model developed for this project is called ATOP for Advanced Traffic Occupancy Prediction. The various modules of the ATOP computer code are currently written in Fortran and run on PC computers (pentium machine) faster than real time for the section of the INFORM corridor under study. The following summarizes the various routines currently contained in the ATOP code.

- Statistical forecasting of traffic flow and occupancy using historical data for similar days and time (long term knowledge), and the recent information from the past hour (short term knowledge).
- Estimation of the empirical relationships between traffic flow and occupancy using long and short term information.
- Mechanistic interpolation using macroscopic traffic models and based on the traffic flow and occupancy forecasted (item-1), and the empirical relationships (item-2) for the specific highway configuration at the time of simulation (construction, lane closure, etc.).
- Statistical routine for detection and classification of anomalies and their impact on the highway capacity which are fed back to previous items.
- Adaptive correction and updating to control the prediction errors within a pre-specified limit.

The BNL program for developing the ATOP computer code was divided into three major tasks. These tasks are:

- a) Historical Traffic Data Base (HTDB): In this task a computerized database containing loop detector data, construction data, weather data (precipitation, snow, and ice accumulation), and traffic data during major accidents was developed. Various computerized routines, with the help of SPSS package, were written to translate the collected data into a format usable by the ATOP code. The initial effort in developing the HTDB was to develop the needed computerized system and testing the routines for a period of about one year (1993).
- b) Kernel: The mechanistic interpolation routine was developed in this task. A macroscopic view of the continuity of traffic flow was described through a first order, non-linear, partial differential equation, using the empirical relationship of speed vs. traffic density. The Kernel is capable of simulating the traffic flow at any segment of the selected INFORM corridor, for any specific configuration, using the traffic flow and density forecasted at the boundary conditions by the other routines.

- c) Shell: All other routines discussed above is developed by the Shell task. Furthermore, all validation, verification, and integration of the various tasks was performed in this task.

At the current stage of development, the individual components of the model (namely the Shell, Kernel, and HTDB) have been all developed and have gone through initial testing, and validation. The Shell and Kernel are currently integrated in ATOP and the communication between them and HTDB is being performed interactively through manual input file modifications.

Traffic Occupancy (%) Profile for L.I.E.

THE LAX METHOD

Data : May 24, 1993, 17:00 — 18:30

CONTENTS

	<u>Page</u>
ABSTRACT.	iii
LIST OF FIGURES	vii
LIST OF TABLES	xi
EXECUTIVE SUMMARY	xiii
ACKNOWLEDGEMENTS	xv
ACRONYMS	xvi
 1. INTRODUCTION	 1-1
1.1 Background	1-1
1.2 Project Objective	1-2
 2. INFORM SYSTEM, INFORM CORRIDOR, AND SELECTED CORRIDOR FOR PHASE-1	 2-1
2.1 Overview of INFORM System	2-1
2.2 Corridor Section Selected for Pilot Model	2-2
 3. INFORM SYSTEM SIMULATION: MODEL OVERVIEW	 3-1
 4. THE EMPIRICAL KNOWLEDGE DATABASE	 4-1
 5. THE SHELL	 5-1
5.1 Overview	5-1
5.2 Determination of Speed from Single Loops	5-3
5.3 Speed-Density Relations	5-5
5.4 Short-Term Updating	5-7
5.5 Adaptive Routine	5-13
 6. THE KERNEL	 6-1
6.1 Introduction	6-1
6.2 Brief Descriptions of the Various Macroscopic Traffic Models	6-1
6.2.1 The Forward Differencing Method	6-2
6.2.2 The Lax Method	6-3
6.2.3 The First Upwind Differencing Method	6-4
6.2.4 The Second Upwind Differencing Method.	6-5
6.3 Model Parameters Determinations.	6-5
6.3.1 Average Vehicle Length.	6-6
6.3.2 Free-Flow Speeds and Critical Densities	6-14
6.4 Model Applications to the Long Island Expressway	6-16
6.4.1 L.I.E. Configuration	6-16
6.4.2 Initial and Boundary Conditions	6-17
6.5 Tests and Validations of the Models	6-17
6.6 Comparisons Between the Lax and the First Upwind Differencing Methods	6-32
6.7 Future Refinements	6-33
6.7.1 Kernel Improvements	6-33
6.7.2 Algorithm Refinements	6-36

	<u>Page</u>
7. PERFORMANCE OF INTEGRATED MODEL	7-1
7.1 Integrated ATOP Performance Excluding the Adaptive Routine	7-1
7.2 Integrated ATOP Performance Including the Adaptive Routine	7-5
8. ATOP PROGRAM DESCRIPTION AND USER GUIDE	8-1
8.1 Program Structure	8-1
8.2 User Guide for the ATOP Code	8-3
8.3 SENSOR Program	8-7
8.4 Historical Databases	8-8
8.5 Future Programming Improvement	8-8
9. INSIGHTS AND AREAS FOR FUTURE RESEARCH	9-1
10. REFERENCES	10-1
APPENDIX A. DATA MISSING FROM INFORM'S ARCHIVAL RECORDS	A-1
APPENDIX B. SINGLE LOOP SPEED ESTIMATION TECHNIQUE USED BY THE INFORM SYSTEM"	B-1
APPENDIX C. ADDITIONAL KERNEL PREDICTIONS OF OCCUPANCY AND FLOW	C-1

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
2.1 Map of the INFORM Corridor	2-3
3.1 A flow chart of the routines embedded in current ATOP computer code	3-1
3.2 INFORM system simulation: model overview	3-3
5.1 Functional diagram of the Shell	5-2
5.2 Regression of speed on the ratio of volume to occupancy. Data is shown with regression line and a 95% prediction interval	5-4
5.3 Speed-density relationship fit to data from one section of highway during the month of June 1993	5-6
5.4 Absolute percent error 15 minutes ahead as predicted by time/series retained	5-10
5.5 Absolute percent error 30 minutes ahead as predicted by time/series retained	5-10
5.6 Absolute percent error 45 minutes ahead as predicted by time/series retained	5-11
5.7 Absolute percent error 60 minutes ahead as predicted by time/series retained	5-11
5.8 Actual occupancy recorded on July 19, 1993 for zone 100, and the historical norm for that zone	5-12
5.9 Differences after the historical norms have been subtracted from the actual data	5-12
5.10 Differences between the actual data and predictions based upon the historical norms and an ARIMA (0,1,1) model	5-13
6.1 A configuration of the L.I.E. as modeled in the current study. In the analysis, the nodes are positioned at the locations of the loop detectors	6-7
6.2 A comparison between flows calculated from occupancies, speeds and measured INFORM flows for Sunday, March 20, 1994. A least squares fit to the data yielded an average vehicle length of 13.66 feet for zone 94 on the basis of 3 open lanes.	6-8
6.3 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Sunday, March 20, 1994. A least squares fit to the data gave an average vehicle length of 13.87 feet for zone 95 on the basis of 3 open lanes.	6-8
6.4 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Sunday, March 20, 1994. A least squares fit analysis yielded an average vehicle length of 11.18 feet for zone 100 based on 3 open lanes.	6-9

<u>Figure</u>	<u>Page</u>
6.5 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Sunday, March 20, 1994. A least squares fit to the data gave an average vehicle length of 10.52 feet for zone 100.	6-9
6.6 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Sunday, March 20, 1994. From a least squares fit analysis, on an average vehicle length of 10.18 feet is determined for zone 105.. . . .	6-10
6.7 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Monday, March 21, 1994. A least squares fit analysis of the data gave an average vehicle length of 17.11 feet for zone 94. Note the absence of data between 9:30 a.m. and 4:30 a.m..	6-10
6.8 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Monday, March 21, 1994. A least squares fit analysis to the data resulted in an average vehicle length of 17.5 feet for zone 95 based on 3 open lanes.. . . .	6-11
6.9 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Monday, March 21, 1994. A least squares fit to the data yielded an average vehicle length of 16.38 feet for zone 100 based on 3 open lanes.. . . .	6-11
6.10 A comparison between flows calculated from occupancies, speeds and measured INFORM flows for Monday, March 21, 1994. A least squares fit analysis of the data yielded an average vehicle length of 15.69 feet for zone 102 based on 3 open lanes.. . . .	6-12
6.11 A comparison between flows calculated from occupancies, speeds and measured INFORM flows for Monday, March 21, 1994. A least squares fit analysis of the data gave an average vehicle length of 18.45 feet for zone 105 based on 3 open lanes.	6-12
6.12 Variation of the average vehicle length with zone number derived on the basis of the INFORM, as well as the BNL speed algorithms for Sunday, May 20, 1994.	6-13
6.13 Variation of the average vehicle length with zone number derived on the basis of the INFORM, as well as the BNL speed algorithms for Monday, May 21, 1994. An average value of 15.5 feet is deduced for the average vehicle length.	6-13
6.14 A Guassian form for the speed-density relation is derived for the L.I.E. from a non-linear squares fit analysis for Mondays through Fridays, March 5-19, 1994. For zone 105, the free flow speed and the optimum occupancy are 70.46 miles/hr and 23.5% respectively.	6-14
6.15 The speed density relation can alternatively be derived from an analysis of flow as a function of occupancy for Mondays through Fridays of March 5-19, 1994. Note that at optimum occupancy of 22.6%, the flow attains its maximum value. Below optimum density, the slope is positive, while above it, the slope is negative.	6-15

<u>Figure</u>	<u>Page</u>
6.16 Occupancy profile for the L.I.E. as a function of segment number (zone) and time generated by the Kernel on the basis of the Lax method for May 24, 1993, 15:30-17:00. The direction of traffic is from right to left. Segments 1 and 14 correspond to zones 94 and 107 respectively and are considered as the boundary conditions for the problem. The time is expressed in units of minutes.	6-18
6.17 Occupancy profile for the L.I.E. as generated by the Kernel on the basis of the Lax method for May 24, 1993 from 17:15 to 19:00. See Figure 6.16 for more details.	6-19
6.18 A comparison between predicted and measured INFORM occupancies as a function of zone number at 15:45 on May 24, 1993. The calculations are carried out by the Kernel on the basis of the Lax method.	6-29
6.19 A comparison between predicted and measured INFORM occupancies as a function of zone number at 16:00 on May 24, 1993. The calculations generated by the Kernel on the basis of the Lax method.	6-29
6.20 A comparison between predicted and measured INFORM occupancies as a function of zone number at 16:15 on May 24, 1993. The calculations are carried out by the Kernel on the basis of the Lax method.	6-30
6.21 A comparison between predicted and measured INFORM occupancies as a function of zone number at 16:30 on May 24, 1993. The calculations are generated by the Kernel on the basis of the Lax method.	6-30
6.22 A comparison between predicted and measured INFORM occupancies as a function of zone number at 16:45 on May 24, 1993. The predictions are based on the Lax method.	6-31
6.23 A comparison between predicted and measured INFORM occupancies versus zone number at 17:00 on May 24, 1993. The predictions are based on the Lax method.	6-31
6.24 A comparison between predictions based on the Lax and the upwind differencing methods and measured INFORM occupancies as a function of zone number at 16:15 on May 24, 1993	6-33
6.25 A comparison between predictions based on the Lax and the upwind differencing methods and measured INFORM occupancies as a function of zone number at 16:30 on May 24, 1993	6-34
6.26 A comparison between predictions based on the Lax and the upwind differencing methods and measured INFORM occupancies as a function of zone number at 17:00 on May 24, 1993.	6-34
6.27 A comparison between predictions based on the Lax and the upwind differencing methods and measured INFORM occupancies as a function of zone number at 17:15 on May 24, 1993.	6-35

<u>Figure</u>	<u>Page</u>
6.28 A comparison between predictions based on the Lax and the upwind differencing methods and measured INFORM occupancies as a function of zone number at 17:45 on May 24, 1993	6-35
6.29 A comparison between predictions based on the Lax and the upwind differencing methods and measured INFORM occupancies as a function of zone number at 18:15 on May 24, 1993	6-36
7.1 Monday May 10, 1993: Z-94 starting at 08:00	7-2
7.2 Monday May 10, 1993: Z-103 starting at 08:00	7-2
7.3 CDF for mean prediction error of traffic density	7-3
7.4 CDF for the percent mean error in flow prediction	7-4
7.5 Scatter plot of mean error of flow vs. density	7-4
7.6 The performance of adaptive routine 6:00 P.M. (June 14, 1993)	7-6
7.7 The performance of adaptive routine 6:00 P.M. (October 12, 1993) severe rain	7-7
7.8 The performance of adaptive routine 5:00 P.M. (May 6, 1993) lanes closed	7-7

LIST OF TABLES

<u>Table</u>	<u>Page</u>
5.1 Speed-Density Parameters Obtained on the L.I.E. During June-July of 1993	5-6
6.1 Speed-Density and Optimum Density Parameters Obtained for the L.I.E. During March 1994	6-15
6.2 Speed-Density and Critical Density Parameters for Exits and Entrances Obtained for the L.I.E. During March 1994	6-16
6.3 Comparison Between Model Predictions (Lax Method) Carried Out in the Present Study with INFORM for May 24, 1993 from 15:30 to 17:00	6-20
6.4 Comparison Between Model Predictions (Lax Method) Carried Out in the Present Study with INFORM for May 24, 1993 from 17:00 to 18:30	6-24
8.1 List of Files and File Sizes Contained in the Tape	8-4

EXECUTIVE SUMMARY

The Intelligent Vehicle Highway System (IVHS) Program is envisioned to improve the efficiency and effectiveness of surface transportation in the United States. IVHS focuses on developing, deploying, and evaluating advanced and emerging technologies. One of several program areas of IVHS is Advanced Traffic Management Systems or ATMS. Based on the National Program and New York State IVHS needs, Brookhaven National Laboratory (BNL) has entered an interagency agreement with Federal Highway Administration (with the concurrence of the N.Y.S. D.O.T.) to work in the ATMS area of IVHS for New York State. Specifically, BNL will develop a Congestion Forecasting Model for Long Island's INFORM system.

Information for Motorists (INFORM) is one of the first federal-state programs to develop, test, and implement an Advanced Traffic Management System (ATMS), combined with an advanced traveler's information system. INFORM was designed to aid drivers in the busy central corridor of Long Island, which includes the major east/west highways and connectors (Long Island Expressway, Northern State Parkway, Jericho Turnpike and Sagtikos/Sunken Meadow Parkways). INFORM is an advanced integrated computerized system designed to aid in the improvement of travel time within the INFORM corridor. The system consists of roadway sensors, merge lights, signal lights, and variable message signs, all linked by a communications network and a computer system at the INFORM Traffic Information Center in Hauppauge.

The objective of the current BNL project is to assist in the advancement of the INFORM to a state-of-the-art ATMS. The project has developed and demonstrated an advanced traffic model that takes data from many sources, including the INFORM system roadway traffic sensors and then rapidly predict impending traffic patterns for up to one hour in advance. This will allow the INFORM system operators to manage traffic much more effectively and efficiently, through the use of public announcements and variable roadway message signs.

The BNL computer model developed for this project is called ATOP for Advanced Traffic Occupancy Prediction. The various modules of the ATOP computer code are currently written in Fortran and run on PC computers (pentium machine) faster than real time for the section of the INFORM corridor under study. The following summarizes the various routines currently contained in the ATOP code.

- Statistical forecasting of traffic flow and occupancy using historical data for similar days and time (long term knowledge), and the recent information from the past hour (short term knowledge).
- Estimation of the empirical relationships between traffic flow and occupancy using long and short term information.
- Mechanistic interpolation using macroscopic traffic models and based on the traffic flow and occupancy forecasted, and the empirical relationships for the specific highway configuration at the time of simulation (construction, lane closure, etc.).
- Statistical routine for detection and classification of anomalies and their impact on the highway capacity which are fed back to previous items.
- Adaptive correction and updating to control the prediction errors within a pre-specified limit.

The BNL program for developing the ATOP computer code was divided into three major tasks. These tasks are:

- a) Historical Traffic Data Base (HTDB): In this task a computerized database containing loop detector data, construction data, weather data (precipitation, snow, and ice accumulation), and traffic data during major accidents was developed. Various computerized routines, with the help of SPSS package, were written to translate the collected data into a format usable by the ATOP code. The initial effort in developing the HTDB was to develop the needed computerized system and testing the routines for a period of about one year (1993). The HTDB is discussed in the second volume of this report.
- b) Kernel: The mechanistic interpolation routine was developed in this task. A macroscopic view of the continuity of traffic flow was described through a first order, non-linear, partial differential equation, using the empirical relationship of speed vs. traffic density. The Kernel is capable of simulating the traffic flow at any segment of the selected INFORM corridor, for any specific configuration, using the traffic flow and density forecasted at the boundary conditions by the other routines.
- c) Shell: All other routines discussed above are developed by the Shell task. Furthermore, all validation, verification, and integration of the various tasks was performed in this task.

At the current stage of development, the individual components of the model (namely the Shell, Kernel, and HTDB) have been all developed and have gone through initial testing, and validation. The Shell and Kernel is currently integrated in ATOP and the communication between them and HTDB is being performed interactively through manual input file modifications.

Beyond the development of the traffic simulation and prediction model, discussed above, several other spin-off products have resulted from the work. These interim results vary in type, level of sophistication, and maturity. They include Traffic Modeling and Analysis Tools such as: a flow split algorithm accounting for both inlet and outlet density, a speed estimator to be used with a single loop detector, a refined speed estimator accounting for rate of variation of traffic density as a function of distance, a PC-based code for data analysis and estimation, and the traffic flow perturbation as a result of closing a lane for a long distance compare to closing a lane for relatively very short distance. Highway insights were also identified such as the identification of specific roadway sector congestion sensitivities and empirical values for maximum roadway capacity at varying locations and weather conditions. Potential future products include HOV lane analysis techniques, variable message sign effectiveness measures, and day-ahead traffic forecasting techniques.

ACKNOWLEDGEMENTS

The authors would like to thank several people at both the federal and the New York State Department of Transportation for their support and assistance in this project, Matthew Miller, Thomas Werner, and Susan Lauffer. We also particularly want to thank Robert Rosendahl, the Director of the INFORM Center, as well as his staff for their cooperation and valuable insights. This work would not have been possible without their enthusiastic cooperation and assistance. Fruitful technical discussions with H. Nourbakhsh are gratefully acknowledged.

We also want to thank Melissa Collichio and Jeanne Madaia for coordinating the interim version of this report and Pamela Ciufo for her excellent work in preparing the current version of this report.

ACRONYMS

α_l	-	Diffusion coefficient in the Lax method
α_u	-	Diffusion coefficient in the upwind method
ATMS	-	Advanced Transportation Management Systems
BNL	-	Brookhaven National Laboratory
C	-	Courant limit (dimensionless)
DOT	-	Department of Transportation
FHWA	-	Federal Highway Administration
HTDB	-	Historical Traffic Data Base
HOV	-	High Occupancy Vehicle
INFORM	-	Information for Motorists
ISTEA	-	Intermodal Surface Transportation Efficiency Act
IVHS	-	Intelligent Vehicle Highway System
K	-	Traffic density (vehicles/mile)
K'	-	Traffic occupancy (dimensionless)
K_m	-	Optimum (critical) traffic density (vehicles/mile)
K'_m	-	Optimum (critical) traffic occupancy (dimensionless)
L_d	-	The separation distance between two loop detectors (ft)
L_v	-	Average vehicle length (ft)
O-D	-	Origin Destination
Q	-	Traffic flow (vehicles/hour)
U_f	-	Free-flow speed (miles/hour)
U_w	-	Wave speed (miles/hour)
VTIP	-	Visual Traffic Information Project

1. INTRODUCTION

1.1 Background

The Intelligent Vehicle Highway System (IVHS) is designed to improve the efficiency and effectiveness of surface transportation in the United States.¹ The IVHS is specifically designed to accomplish its mission in such ways as to minimize the need to increase current highway capacity. These will include not only the vehicle and car driver, but trains, commercial trucks and public transportation. The effective application of advanced technologies can produce major societal benefits including improved use of infrastructure, reduced need for energy resources, improved transportation system safety (e.g., roadway, vehicles, and drivers), and reduction of environmental pollution associated with vehicular emissions. In addition, as improvements in these areas accrue collectively, they can contribute to improved societal productivity and global competitiveness for the United States.

The major user services to be provided by IVHS are: Travel and Traffic Management, Public Transportation Management, Electronic Payment Services, Commercial Vehicle Operations, Emergency Management, and Advanced Vehicle Safety Systems. IVHS program plans are divided into a number of key areas including: research and development, operational tests, national architecture, standards and guidelines, and supporting IVHS deployment.

Based on the National Program and N.Y.S. IVHS needs, Brookhaven National Laboratory (BNL) has entered an interagency agreement with FHWA with the concurrence of the N.Y.S. D.O.T. to develop a Congestion Forecasting Model for Long Island's INFORM system. The Brookhaven National Laboratory project consists of three phases. The Phase I study was awarded, the work is already completed, and documented in this report. This phase is designed to show the feasibility of forecasting traffic congestion an hour ahead, and demonstrate this technology for a portion of the corridor being managed by Information for Motorists (INFORM) System. Phase II will expand the project to the full INFORM system, and demonstrate the technology in an integrated on-line fashion. Finally, efforts will be directed towards new corridors including the Southern State Parkway for Phase III.

The goal of the project, as stated in the June 14, 1993 letter² to the BNL Director from the N.Y.S. D.O.T. Commissioner is that, "Advanced Traffic Management Systems (ATMS) should prevent traffic congestion by developing on-line traffic control measures based on anticipated degradation in the throughput of networks using current real time traffic volumes and origin-destination (OD) information. Rather than detecting, analyzing, and reacting to traffic congestion problems (break-downs), new ATMS systems must forecast and implement mitigation strategies prior to degradation."

An incident on a major highway can affect areas which are quite removed from the incident site (5-10 miles). If preventive measures could be taken, such as altering signal timing or posting appropriate messages on variable message signs, the impact would be reduced. Initiating these preventive measures will be further facilitated with a large scale predictive model which uses real time traffic information, as well as historical data on traffic system performance (e.g., congestion and clearance information). This new technology could also be used to assist in formulating construction and maintenance detours.

1.2 Project Objective

The objective of the current BNL project is the development and demonstration of a large scale predictive model using adaptive simulation techniques.

Prior to the concept of Advanced Traffic Management Systems (ATMS), the focus of a number of traffic simulation models has been "off-line", e.g., performed in an environment separated from ongoing traffic conditions. Using such simulation capabilities, the traffic engineer could then plan highway systems, intersections, traffic signal timing, and other engineering tasks.

The ATMS application is different from the aforementioned traditional traffic simulation, because conditions in the traffic network are largely known and traffic engineers need to be able to forecast developments over the next hour or two. Preferably, the simulation would be rapid, thereby allowing the engineer to explore the results of contemplated actions before they are implemented.

In order to comply with the project objectives in the most efficient manner, the existing modeling techniques^{3,4} in traffic engineering were reviewed and an attempt was made to tailor these techniques for the development of the ATOP code. We consider the success of the ATOP code to be indebted to and built upon the successes and lessons learned from this previous work.

2. INFORM SYSTEM, INFORM CORRIDOR, AND THE SELECTED CORRIDOR FOR PHASE-I

2.1 Overview of INFORM System

Information for Motorists (INFORM) is one of the first federal-state programs to develop, test, and implement a computerized traffic monitoring system, combined with a traveler's information system. INFORM was designed to aid drivers in the busy central corridor of Long Island, which includes the major east/west highways (Long Island Expressway, Northern State Parkway, Jericho Turnpike and Sagtikos/Sunken Meadow). INFORM was initiated in the late 1970's and was implemented by the New York State Department of Transportation (N.Y.S.D.O.T.) in 1987.⁵ The INFORM system covers the area from Hauppauge to Kew Gardens on the Long Island Expressway, the stretch from Hauppauge to Shea Stadium on the Northern Parkway (which becomes Grand Central Parkway at the Queens border) and from Veterans Memorial Highway to the Meadowbrook Parkway on Jericho Turnpike. The system also covers major north/south crossovers and connecting routes. This area is called the INFORM corridor. A corridor, as defined,⁵ is a parallel roadway or transportation facility that generally serves major metropolitan areas. The INFORM corridor is host to one of the largest and most advanced traffic information systems in the nation⁶ (Figure 2.1).

INFORM is an advanced integrated computerized system designed to aid in the improvement of travel time within the INFORM corridor. The system consists of roadway sensors, merge lights, signal lights, and variable message signs, all linked by a communications network to a central computer system at the INFORM Traffic Information Center in Hauppauge. The INFORM system is staffed and operated by the N.Y.S. D.O.T. 24 hours a day, 365 days a year.

Information on current traffic patterns is received by the over 2,400 electronic sensors (loop detectors), placed in these roadways. The loop detectors are placed at about 1/2 mile intervals throughout the corridor. Vehicle movement over a loop detector causes changes in the loop inductance, and the electric current is sampled four times per second and transmitted to central computers in the INFORM Traffic Information Center. The information is then processed to estimate the current traffic occupancies and flow averaged over one minute, three minutes, and fifteen minutes. The processed information is fed to various pieces of software for further analysis and display. One of these pieces of software, Visual Traffic Information Project (VTIP) displays a color coded illustration of the INFORM corridor to the operators. The color coding of VTIP is based on average vehicle speed over a fifteen minute time window. VTIP therefore provides a concise illustration of the monitored highway traffic in an easily understandable graph for the INFORM staff. The VTIP display is examined by Traffic Coordinators, and it is used to identify potential problem areas which need to be further examined by video pictures transmitted by the cameras. There are many other pieces of software and routines embedded in the INFORM platform, however the limited scope of this report does not allow detailed discussions of these routines. Some of these routines are: a rule-based routine for detection of incidents using the loop detector data, a routine for speed prediction, a routine for signal validation, and a routine to estimate a queue length, etc.

West of Hauppauge, the drivers entering these highways pass through an automated light linkage (a metering device) on the entrance ramps and certain intersections. The lights are computer adjusted to allow for easier flow of vehicular traffic through a particular area in response to traffic patterns.

The INFORM system is also linked with traffic lights at over 100 intersections on major connector roads and parallel routes. Traffic lights at these intersections are monitored all day; however they are fully automated and automatically adjusted to maintain efficient flow.

INFORM alerts drivers to the conditions ahead by flashing computerized variable message signs on the highway. These signs alert drivers to dangerous or congested driving conditions due to such factors as weather, vehicular crashes or construction. They can be used to also suggest alternate routing of traffic in case of one of these incidents affects one of the routes although this feature is rarely used.

Whether information is sent to the INFORM center by a computerized signal line from one of the three major airports on Long Island, LaGuardia, Kennedy, or Islip MacArthur, information on construction is radioed in ahead of time by field supervisors to the INFORM center. Motor vehicle crashes, disabled vehicles, and other incidents occurring on the system are identified by video cameras, police and emergency calls, and passer-by phone calls.

At the INFORM Center, Traffic Information Coordinators use data from computer monitors and a 25ft wide digital map that monitors the 35 mile INFORM corridor, to determine volume and speed of traffic. With this information they can then transmit motorist advisories on the variable message signs on conditions present in the area.⁵ The Traffic Information Coordinators monitor the road for emergency situations, and radio in problems to the police and fire authorities at the nearest exit. The information is also delivered to the news media.

INFORM has been deemed Long Island's "test bed" corridor for IVHS technology. This is seen in INFORM'S "natural outgrowth" to Visual Traffic Information Project (K),⁷ which is a computer based system that delivers information from the Center to television and radio, which is then available to motorists. INFORM is currently being upgraded with new computers and software and it is planned to be expended to the Southern State Parkway. Recently, as part of the INFORM Corridor Project, the HOV lanes were opened in May 1994.

2.2 Corridor Section Selected for Pilot Model

The study region for Phase I of the current project (as described in this report), includes a well-defined area within Suffolk County, namely eastbound Long Island Expressway from Deer Park Road to Wicks Road; eastbound Northern State Parkway from Deer Park Road to Sagtikos State Parkway; and Sagtikos State Parkway (North and Southbound). This study region (network subsystem) was selected because construction within the area was largely completed by May 1994, and no additional construction has been planned. Therefore, this network subsystem is anticipated to remain stable throughout the BNL program. It also allows a before/after comparison of factors affecting congestion, including incident identification, response, and management; weather conditions; and construction. However, due to the absence of ramp data on all except the L.I.E., the study was subsequently limited to the L.I.E.

Figure 2.1 Map of the INFORM corridor

3. INFORM SYSTEM SIMULATION: MODEL OVERVIEW

As discussed in Section 1.2, the objective of the first phase of this project is to develop a traffic model for the selected portion of the INFORM corridor which can quickly and accurately predict traffic congestion about one hour ahead based on inputs from the INFORM system detectors. This section provides a brief overview of the various models and routines which are described in considerably more detail in Sections 5, 6, and 7.

The structure of the various routines developed for this project, and currently contained in the ATOP computer code, is depicted in Figure 3.1. The various routines and databases developed under the three major tasks of this project, the Kernel, the Shell, and the Historical Traffic Data Base (HTDB) are also identified in this figure. The basic concept which was repeatedly used in developing the ATOP computer code is that of constructing knowledge from historical information (referred to as static information) and dynamically updating this information based on the most recent applicable information (referred to as short term updating). An extensive amount of effort was allocated for the proper construction of historical (static) knowledge in the form of parametric mathematical models and databases. Many standard packages such as SPSS, SAS, IMSL, Statgraphics, and in-house routines were used for this purpose, but are not depicted in Figure 3.1. These additional computer packages were used for construction and estimation of the parameters for static speed-density relationship, single loop speed predictor, estimation of the scaling parameters for flow, simulation of one minute averaged data from fifteen minutes average data, etc. Some discussion of these routines and the results are detailed in Sections 5 and 6.

Figure 3.1 A flow chart of the routines embedded in current ATOP computer code

The processed data on recent traffic flow (presently 15 minutes averages) are translated to the format required by the ATOP code by the translation routine (top of Figure 3.1). Historical information and static relationships are fed to a short term updating and forecasting routine to arrive at the dynamic estimation and forecasting of traffic flow, density, and parameters of speed-density relationship at every segment of the highway including the forecast of flow and density at the boundaries of the corridor and every exit and entrances. State-of-the-art techniques, such as Bayesian dynamic updating, and ARIMA forecasting with seasonal differencing, have been implemented in the ATOP code. The information generated is then fed to a Kernel routine which predicts the traffic flow movement by integrating all the input and forecast information generated and then satisfying the flow continuity equation, while also accounting for any potential road construction or lane closure. A state-of-the-art numerical routine is used to solve the first order, partial differential equation describing the dynamics of the flow continuity. The predictions from the Kernel are compared to the earlier forecasts generated by ARIMA method, and any additional information received from the loop detectors while the computations were being performed, through the use of a comparator routine. The current version of ATOP uses a control chart approach, based on tolerance bounds, to identify significant deviations. Once significant deviations are identified, a mechanistic traffic model is used to decide whether or not the deviation is due to an incident or is the result of other traffic conditions such as; larger than expected inflow/outflow from entrance/exits, severe weather conditions, etc. If an accident/incident is identified, then a flow/density comparison is used to classify the accident in terms of the segments affected, and the most probable number of lanes that are expected to be closed. If the adaptive routine is activated, a feedback loop initiates the short term updating and forecasting routines to correct for any potential changes in traffic conditions.

A detailed description of each of these modules, along with identification of potential areas for improvements, is provided in Sections 5 and 6. As an example, the current ATOP version uses confidence bounds to identify anomalies, and uses mechanistic rules to classify them. This process, even though satisfactory at the present time, can be made more efficient. A faster response time can be obtained if the classification process is made by intelligent comparison of the current anomalous signature with the historical anomalous signatures developed under HTDB databases. Neural network techniques and potentially AI routines, are prime candidates for such future refinement.

Figure 3.2 shows a conceptual expanded version of the future ATOP program, if implemented. The future ATOP works as an electronic manager, analyst, and historian, continually tracking conditions on the highway network, and frequently updating the empirical relationships as required. ATOP then performs traffic simulations, thereby projecting traffic conditions ahead for up to one hour. The accuracy of the simulations to a large degree depends on accurate empirical relations with respect to the traffic density-flow relationships, the flow forecast for exits/entrances and interchanges, and projected traffic conditions at the system boundaries (as a function of time). Once the first set of computations (simulation) is completed, the ATOP adaptive routines will assess the trends for apparent anomalies. If there are such anomalies apparent, they will be classified to a set of categories to allow proper correction and updating to be performed. If an accident has occurred and is identified, the incident signature will be categorized based on previous anomalous signatures, and an early warning will be provided to the INFORM System Operator. After the incident is verified, and more specific information is obtained by the INFORM operator, then information can be interactively entered to the ATOP code for future simulation.

The system can also be used off-line to arrive at an optimum emergency procedure by postulating an accident or severe congestion and evaluating the effectiveness of the re-routing procedures in terms of clearing time or travel time or a combination thereof. This system is also useful for construction planning, to assure the least possible impact on highway capacity and traffic flow. In case of predicted congestion by the ATOP code, or the occurrence of unexpected congestion, the ATOP code is designed to perform its calculations several times faster than real time. This allows the INFORM staff to simulate the effect of any remedial actions that they are about to take.

In summary, the INFORM system simulation codes, ATOP, provides a catalog of historical traffic data, powerful estimating ability, and the insights available through several seemingly unrelated data bases. It is not intended to replace the traffic management system operator, but it should make the operators much smarter, more efficient, and more effective in performing their jobs. Future applications of ATOP Code may include traveler information systems through variable message signs, and other communication media. The authors envision that the information could be posted on a computer system available to both TV and radio traffic reporters and any Internet users. Cellular phone companies could even make recordings of up-to-the-minute traffic predictions for callers approaching main roads.

Figure 3.2 INFORM system simulation: model overview

4. THE EMPIRICAL KNOWLEDGE DATABASE

For the development, testing, and validation of the simulation model of this report, necessary input data were collected and compiled. The empirical knowledge database is comprised of inductive loop detector data, construction data, weather data (precipitation, snow, and ice accumulation), and police data on major accidents on the highway. A brief description of the sources for the empirical knowledge database is presented here. Detailed information can be found elsewhere.⁸

- *INFORM traffic data*

The computerized INFORM traffic file contains 15-minute averaged data of volume (flow) and occupancy for the main highways, as well as on-and off-ramps for each zone. The status of the data is shown in Appendix A. The INFORM data played a major role in model testing, as well as in the derivation of the parameters for the speed density relationship (Equation 5.4) and the speed algorithm (Equation 5.3).

- *Fatal Accident Reporting System (FARS) of the federal database*

FARS contains information on all motor vehicle crashes which results in one or more fatalities. FARS data were available for years prior to 1992. In addition, data were available for a portion of 1993 at the time of data retrieval by this project. The data include: 1) date and time of crash, 2) location of occurrence, 3) type of collision, and 4) related conditions such as weather and road conditions.

- *New York State Department of Motor Vehicles police crash reports*

Police crash reports involving accidents occurring within the current INFORM corridor were obtained from the New York State Department of Motor Vehicles. These accident reports include information on: 1) data and hour of accident, 2) location of the accident i.e., county, municipality and road, 3) collision type, and 4) weather and road conditions.

- *NOAA weather data*

The NOAA reports contain daily summaries of weather type, temperature, precipitation, barometric pressure, and wind. In addition, three hour summaries of temperature, visibility, and wind are provided. However, precipitation, which has an impact on traffic flow and the equation of state (Equation 5.4), is not available on a three-hour basis.

- *Brookhaven National Laboratory weather data*

Weather data which is comprised of one hour, as well as daily, summaries of temperature, precipitation, barometric pressure, and wind speeds are available from the Brookhaven National Laboratory Meteorology Division

- *INFORM incident reports*

Computerized reports of incidents occurring within the INFORM corridor were obtained from the INFORM Center for the period of January through October 1993. Incident reports include

information on anomalous traffic buildup, phoned in reports from the police, radio reports, visual identifications obtained from closed circuit video cameras on the highway. The available data contained in these incident reports, are 1) date and time of occurrence, 2) date and time of end of the incident, 3) duration of incidents, 4) location and incident description, and 5) number of highway lanes closed. Knowledge of the latter information is important in the modeling analysis.

- *Highway history construction data*

A highway history database which contains microfilmed plans for construction projects is available from New York State. However, construction data from this source were not included at this time.

The initial effort was the development of the needed computerized systems and testing the routines, utilizing data collected for a period of about a year (1993). With the exception of FARS, an independent computerized database readable by the SPSS software package was developed and can be consulted for information.

Of primary importance in the validation of the ATOP simulation model is the INFORM traffic occupancy and volume (flow) data which were processed by the computer for conversion into a format readable by the Shell (Section 5) and then integrated with the ATOP model. This portion of the integrated empirical database is labeled and referred to in the report as the Historical Traffic Database (HTDB). The integrated data structure of the HTDB is discussed in some detail in Section 8. However, the other information, which was collected as part of the empirical database but not incorporated in the ATOP model because of its secondary importance, will be evaluated for potential use in other possible aspects in the future phase of this program.

5. THE SHELL

The Shell acts as the central manager, analyst, predictor, and decision maker in the BNL model. The Shell is responsible for the prediction of traffic flow, establishing the required empirical relationships, maintaining current and updated traffic parameters, and the detection of anomalies. To perform these tasks, the Shell uses time series techniques, Bayesian statistics, and extensive database of historical traffic patterns, and adaptive routines. The remainder of this section will discuss these functions in detail.

5.1 Overview

A functional diagram of the Shell, and its relationships with the loop detectors and the Kernel, is shown in Figure 5.1. The top half of this diagram, above the box labeled "Kernel," outlines the data interpretation functions of the Shell. The bottom half of the diagram shows the sections of the Shell which monitor the accuracy of the Kernel's predictions and perform corrective actions.

The Shell starts with processed loop detector data, which is shown at the top of Figure 5.1. If, in the future, the Kernel and Shell are integrated into the INFORM system, the processed loop detector data will consist of short cycle averages. At the present time, however, BNL has been using archival loop detector data from the INFORM system. There are two problems regarding the use of this data. First, only 15 minute averages are currently available from INFORM archives. Second, the archives contain no data for any of the entrances and exits on two major highways in the system: the Northern State Parkway, and the Sagtikos. At present, BNL is handling the first problem by using the patterns present in the 15 minute data to simulate short cycle data. The second problem, lack of data on two of the major highways, has forced BNL to limit the scope of this project to the Long Island Expressway. A description of the data missing from the INFORM archives, can be found in Appendix A.

This processed data is interpreted in the light of historical information, shown on the third line of Figure 5.1. The box labeled "Historical Information" represents a file containing both physical and historical information for each zone in the highway system. The physical information consists of information on the existence and type of ramps, the number of highway lanes and ramp lanes, and the identities of the loop detectors. The historical information consists of the parameters that govern the speed-density relationship, profiles of the typical traffic density as a function of time and type of day, and profiles of the traffic flow at entrances and exits. The speed-density parameters are estimated from historical data. The traffic density and ramp flow profiles consists of median values calculated from historical data, and stored for each 15 minute time interval. Profiles have been stored for different types of days.

The processed loop detector data, and the historical information, are manipulated by the short term updating functions. These functions consist of routines which perform Bayesian⁹ updating of the speed-density parameters, and time series routines which predict boundary conditions and ramp flows. Historically, the actual data coming from the loop detectors has been noisy and sometimes non-existent. Since this situation is unlikely to change, the interpretation process has been developed to obtain the best possible estimates of the conditions on the highway. The end product of this process is a set of parameter estimates which reflect both the current state of the highway and historical information. The parameters passed to the Kernel are information about the relations

Figure 5.1 Functional diagram of the Shell*

between traffic speed and traffic density, the traffic densities present in each section of the highway at the start of the simulation, and predictions of the traffic densities which are expected to be present at the boundaries of the system.

Predictions made by the Shell for the boundary conditions and ramp flows are fed to the Kernel for extrapolation to other segments of the highway. The results from the Kernel are then fed to a comparator routine in the Shell which determines whether or not the difference between the extrapolations and the actual data is acceptably small, e.g., within the normally observed range of such differences. If the difference is within the normal range, then the predictions from the Kernel are accepted as they are. If the predictions made by the Kernel deviate significantly from the incoming data, then the Shell attempts to determine the cause of the deviation, and adapt to that cause.

The sections that follow discuss the data interpretation aspects of the Shell in more detail. Section 5.2 discusses how the Shell is able to determine traffic speed from single-loop detectors. Section 5.3 describes the speed-density relationship being used by the Kernel. Section 5.4 describes

*This figure is the same as Figure 3.1 and is duplicated here to facilitate the discussion in this section.

the Bayesian routine that is being used by the Shell to update the speed-density relationship, as well as the time series methods that are used to forecast traffic density and ramp flow. Finally, Section 5.5 discusses the adaptive routines.

5.2 Determination of Speed from Single Loops

Information about traffic speed in each segment of highway is required to update the speed-density relationships required by the Kernel. Speed can easily be determined from a pair of loop detectors located in close proximity, called "speed traps." However, most of the INFORM corridor is instrumented with isolated, single-loop detectors, which report only occupancy and volume. Since speed is an important parameter, and is not directly measured, an exploration was made into various algorithms for determining speed using the data from these isolated detectors.

Consideration of the functional characteristics of loop detectors leads to the conclusion that traffic speed should be proportional to the volume divided by the occupancy. As measured by loop detectors, occupancy is the amount of time that a loop registers a car above it, divided by the total amount of time measured. Given a six foot loop detector, and a vehicle of length L_d , traveling at speed U , the total time for the vehicle to completely pass over the loop detector will be $(L_d + 6)/U$.

If the traffic volume is Q , then the total occupancy, is the sum of the occupancies due to each car, and hence:

$$K' = Qd(L_d + 6)/U . \quad (5.1)$$

where d is a correction factor for any potential delay in the triggering of the loop detectors.

Solving for the speed, U , yields:

$$U = d(L_d + 6) (Q/K') . \quad (5.2)$$

If L_d and d are not known exactly, it can only be said that speed is proportional to the volume divided by the occupancy.

The INFORM system determines speed in each zone by measuring car length at a neighboring speed trap, and then using car length to calculate speed directly. The algorithm used by INFORM is presented in Appendix B. For the purposes of this project, Equation 5.2 was calibrated using empirical regression techniques.

Calibration of the relationship in Equation 5.2 was attempted using data from the speed trap located at zone 100 on the Long Island Expressway, which is located just to the east of Straight Path. However, it was found that the model of strict proportionality exhibited systematic deviations from

the data. The shape of these deviations suggested that the addition of a quadratic term would improve the fit, and this proved to be the case. Thus, the final model selected for predicting speed from the ratio of volume to occupancy was:

$$U = b_1(Q/K') + b_2(Q/K')^2 \quad (5.3)$$

where b_1 and b_2 are regression coefficients, Q is vehicles per hour, and K' is occupancy, measured on a scale from 0 to 100. The coefficient b_1 was estimated to be 0.197 miles, with a standard error of 0.0018 ($t=109.3$, $p<.0001$), and the parameter b_2 was estimated to be $-1.15E-04$ (hours.mile), with a standard error of $3.8E-06$ ($t=-29.9$, $p<.0001$). The mean square error for this regression was 3.87 mph.

The actual data, and the model fit from Equation 5.3 are shown in Figure 5.2. Each data point is plotted with actual speed on the ordinate, and the ratio of volume to occupancy on the abscissa. Notice that the lack of an intercept term in Equation 5.3 forces the regression line through the origin. The fit is generally good, but is dominated by the large cluster of points in the upper right hand corner. There is also a slight tendency for the equation to over predict the lower speeds. The overall fit of the model can be improved slightly by adding an intercept term to the model. However, when the model was fit with an intercept, the estimate of the intercept was negative. Since the negative intercept results in the prediction of negative speeds, the term was dropped.

Figure 5.2 Regression of speed on the ratio of volume to occupancy. Data is shown with regression line and a 95 % prediction interval

5.3 Speed-Density Relations

The Kernel contains a set of difference equations (6.6, 6.11, 6.14, and 6.19) which are driven by traffic density. Hence, the Kernel requires a relationship between density and speed from which it can derive traffic speed and traffic flow. A number of models have been proposed to describe the speed-density relationship, and all of them have advantages and disadvantages.¹⁰ After empirically testing several of these models, the single regime model proposed by Northwestern University was adopted.

The form of the Northwestern speed-density model is:

$$U = U_f \exp \left\{ -\frac{1}{2} \left(\frac{K}{K_m} \right)^2 \right\} \quad (5.4)$$

Here, U is the traffic speed, U_f is the free-flow speed, K is the traffic density, and K_m is the density at which optimum flow occurs. Both U_f and K_m are constants which are estimated from the data.

Equation 5.4 has been fit to historical data from 34 sets of loop detectors present in the mainline lanes of the highway system being modeled. However, since the INFORM loop detectors supply occupancy data, occupancy has been used in place of density. Scaling from occupancy to density can be done by assuming an average vehicle length. An example of a typical fit to Equation 5.4 is shown in Figure 5.3. This data was collected from a set of loop detectors on the Long Island Expressway, located just to the east of Straight Path (zone 100). The data are shown as "box and whisker" plots, where the boxes represent the interquartile range, and the ends of the whiskers represent the most extreme value. The function itself is drawn as the smooth line passing through the data.

The parameters of the function shown in Figure 5.3 are fairly typical for the Long Island Expressway. The parameter U_f (free-flow speed) was estimated as 70.6 mph, and parameter K'_m was estimated at an occupancy of 0.22. Assuming a combined vehicle and loop length of 24 feet, these parameter values imply that a maximum flow of 6183 vehicles per hour will occur at a density of 48.4 vehicles/mile/lane, and a speed of 42.6 mph. Table 5.1 shows the parameter estimates that were obtained, using data from June and July of 1993, for the parts of the Long Island Expressway that are being modeled. The table shows a wide range of parameter estimates, as well as large variations in capacity, from one zone to the next. For example, zones 96 and 97 show relatively slow free flow speeds. Examination of those zones one month later, in August of the same year, shows that the parameters values came back in line with the rest of the zones. In August, zone 96 had U_f estimated as 65.9, and K'_m estimated as 0.29, while zone 97 had U_f estimated at 64.1 and K'_m estimated at 0.28.

This data analysis indicates a large variation in capacity along sections of the L.I.E. The speed-density relationships have been used to predict the theoretical maximum flow for each zone, and are shown in the third column of the Table 5.1. Shown in the fourth column of this table is the maximum flow actually observed in each zone, which provides a check on the theoretical calculations. The observed and predicted flows agree well, and are highly correlated ($r=.88$). However, for zones 96, 97, and 102, the difference between the predicted maximum flow and observed maximum flow

Figure 5.3 Speed-density relationship fit to data from one section of highway during the month of June 1993.

Table 5.1 Speed-Density Parameters Obtained on the L.I.E During June-July 1993
(flow calculations are based on an assumed combined vehicle
and loop length of 24 feet and a three-lane highway)

Zone	U_r (miles/hr)	K_m (occupancy scaled from 0 to 1.0)	Predicted Maximum Flow (vehicles/hour)	Observed Maximum Flow (vehicles/hour)
94	66.6	0.26	6933	6151
95*				-----
96	54.6	0.28	6117	7343
97	53.0	0.35	7415	6973
98*				-----
99	68.3	0.11	3006	3975
100	71.2	0.21	5988	6017
101	73.4	0.19	5580	5917
102	75.7	0.14	4241	5258
103	72.8	0.15	4366	4534
104	72.0	0.15	4326	4154
105	72.3	0.19	5493	5641
106	79.4	0.17	5408	5362
107	79.1	0.20	6336	6495

* parameters not estimable from the present data

is about 1000 vehicles/hour, or about 20% underestimation. The absolute differences exist even though there is quite an agreement in the relative values of predicted vs. observed. There are two important notes that the reader shall be aware of:

- 1) The predicted values are the expected sample averages whereas, the observed is a sample representing maximum flow observed (not necessarily at $K' = K'_m$). Comparing the sample average with a maximum sample from the population, may not be appropriate and can result in a large difference.
- 2) Stochastic variations of the parameters of speed-occupancy relationship on a day by day basis is accounted for through Bayesian updating of the parameters of the empirical fit. This will minimize the discrepancies between the predicted vs. observed flow, as will be discussed later in this section.

However, the factors behind this variation in capacity, both between zones and in a given zone over time, are not yet fully understood. Another important feature of the table is that the parameters for zones 95 and 98 could not be estimated. During the months of June and July, very little loop detector data was obtained from these two zones. However, a determination of the parameters for zones 95 and 98 was possible from data analysis of traffic data in March 1994, which was received in the latter part of this project (Table 5.1).

Two problems are clear from the static estimation presented here. First, there is substantial variation in the speed-density parameters. Second, these parameters may vary from month to month due to construction, or from hour to hour due to weather conditions, segment density, or unanticipated incidents (e.g., accidents). This was demonstrated by zones 95 and 98, which failed to produce adequate data to allow estimation. Solutions to both of these problems will be addressed in the next section.

5.4 Short-Term Updating

The Kernel requires continually updated information on speed-density relationships, and predictions of traffic densities and traffic flow. Two different techniques are used to process this information. Bayesian updating is used to process the loop detector and historical data in order to estimate the parameters governing the speed-density relationship. Time-series methods, which use data from the past to predict future data, are used to project the traffic densities at the boundaries of the highway system, as well as entrance and exit behavior, into the future. These two techniques will now be described in detail.

The parameters governing the speed-density relationship, U_f and K_m , will change due to such things as inclement weather, narrowing of road width by events on the shoulders, and lane closures. These events will be reflected back to the Shell through the loop detector data. The Shell must report these changes to the Kernel if predictions are to be accurate. Since loop detector data is both noisy and unreliable, it is important that this data be filtered through the use of historical information. Bayesian statistical methods provide a quantitative method to accomplish this filtering. In cases where the loops have failed, and no current information is available, it will be necessary to rely completely on the historical data.

Bayesian methods combine two sources of information and produce a result that is a weighted average of both of the sources. The weights attached to each source are inversely proportional to the uncertainty associated with that source. The first source of information represents prior knowledge, and is called the "prior distribution," or simply the "prior." The prior is a function that represents the probability of the parameter of interest taking on different values. For instance, one might know from previous measurements that free flow speeds on a particular stretch of highway are normally distributed, with a mean of 70 mph and a variance of 5 mph. Here, the variance is equivalent to the uncertainty in the prior distribution. A large variance implies a large uncertainty, and therefore a weak state of prior knowledge. A small variance implies strong prior knowledge. The second source of knowledge represents current information, and is referred to as the likelihood. This function represents the likelihood of each possible parameter value as a function of observed data. A large amount of similar data represents a strong likelihood, with small uncertainty. A small amount of data, or very dissimilar data, represents a weak likelihood, with large amounts of uncertainty.

For example, if a winter storm were to cover the highway with ice, and the loop detectors consistently indicated low-density traffic moving at 15 mph, this would constitute strong evidence that the speed-density relationship had changed. In this case, the short-term updating routine would radically move the speed-density parameters away from their historical averages. On the other hand, if a transient event were to cause a large, but short-lived, perturbation in the traffic flow, the short-term updating routines would maintain the speed-density parameters near their historical average.

The Bayesian updating routine in the Shell assumes that the parameters U_f and K_m have prior distributions which are independent and normally distributed. The means of these distributions are determined from historical information for the relevant section of highway, and the variances are determined by the observed range of variation in the means. Now, let $\{(k_1, q_1), (k_2, q_2), \dots, (k_n, q_n)\}$ be the observed values of density and flow. It is assumed that each k_i has been measured without error. Thus, we condition upon k_i , and base the estimate on the probability of the observed flow. The estimation procedure starts with a random draw from the prior distribution, which results in values that will be denoted by U_f^* and K_m^* . Given k_i , U_f^* , and K_m^* , the speed-density relation can be used to determine the expected value of the flow, $E(q_i)$. Assuming that the flow is normally distributed, with mean and variance equal to $E(q_i)$, the probability of each observed q_i , $P(q_i | k_i, U_f^*, K_m^*)$, can be determined. Multiplication yields the likelihood:

$$L(q_i | k_i, K_m^*, U_f^*) = \prod_{i=1}^n P(q_i | k_i, K_m^*, U_f^*) \quad (5.5)$$

The likelihood is then combined with the prior to obtain the posterior distribution and an estimate of U_f and K_m . Sampling from the prior distribution, and subsequent evaluation of the posterior is continued in order to obtain a large sample from the posterior distribution. The posterior sample is then evaluated to obtain the final estimate of U_f and K_m .

Forecasting the traffic density at the boundaries of the highway system, as well as the ramp flows, is handled in the Shell by time-series techniques. The first step in creating a time series model for this data is to subtract out the regular variation. Seasonal differencing is a standard technique for removing regular variation from a data series. If the variation has period p , then the observation at time t is differenced by subtracting the observation at time, $t - p$. In the case of traffic data, the period might be considered to be seven days. Thus, occupancy data from Monday at 3:00 p.m. would be differenced by subtracting the occupancy observed exactly one week prior.

Strict seasonal differencing has a number of disadvantages for this application. First, it would require storing large amounts of information for each section of highway. Second, due to the unreliability of loop detectors, this data may not be available. These irregular gaps in the data, which may be as long as a couple of months, would be much more than any automated statistical analysis could handle. Thus, some other method of removing the regular variation is needed.

In order to overcome the problems of seasonal differencing, BNL employed a technique that we have labeled as "historical differencing." Specifically, for each section of highway, the data is examined and the median occupancy is calculated for each 15 minute time period. This series of historical occupancies is stored and is then subtracted out of data coming from the loops. This method overcomes the problems with seasonal differencing. First, the amount of data that needs to be stored is greatly reduced. BNL has empirically determined that the traffic pattern for Mondays, Tuesdays, Wednesdays, and Thursdays are quite similar. Thus, the historical profile consists of three series per section of highway: one series for Friday, one series for Saturday and Sunday, and one series for the remainder of the week. Naturally, additional series will be needed for holidays, and for different times of the year. Historical differencing also solves the missing data problem. A large amount of data is used to create each series, and thus the possibility of having no data for a particular time and location is minimized.

Overall, the time-series approach adopted here works well. Under most conditions the error in prediction runs about 10%, even when predictions are made an hour into the future. Figures 5.4 through 5.7 show the absolute value of the percent error in flow as a function of time of day. These plots show the median value to the absolute percent error for predictions of occupancy made 15, 30, 45, and 60 minutes into the future. Only during the early morning hours is the error in prediction large. During these times, the occupancies on the road are small, typically about 1%. Thus, small errors of prediction result in large percent errors at these times.

For most locations and time periods, the historical series is very close to the actual data series. However, there are occasions where the two series diverge. Figure 5.8 illustrates this point. Shown here is the actual data collected at zone 100, on Thursday, July 19, 1993. Also shown in the figure is an historical series which was developed using data from the month of June, 1993. The two series run close together until they diverge at 17:00 hours. The difference between these two series is shown in Figure 5.9. From 17:00 to 18:15, the occupancy on this section of highway was significantly less than the historical norm. However, the rest of the series is quite close to the historical norm and presents little problem.

Box-Jenkins¹¹ analysis, applied to the series of difference shown in Figure 5.9, will work well, except for the deviant period between 17:00 and 18:15 hours. To illustrate this point, this set of data was modeled as an ARIMA (0,1,1) process. After the model was fit, the predictions from the model were summed with the historical data to produce a time series prediction for this zone. These predictions were subtracted from the actual data. The differences between the actual and predicted series are shown in Figure 5.10. As can be seen in the figure, the predictions do well, except in the deviant period. The type of deviations that occurred from 17:00 to 18:15 hours, are seen frequently in the INFORM data. Generally, these deviant periods last from one-half to two hours. Modeling these deviations has proved to be difficult. Since the deviations are caused by unobserved, transient variables from outside the system, transfer functions (e.g., regression equations) cannot be used. Since the causes of the deviation are indeed external to the series, standard Box-Jenkins type analysis cannot capture the phenomenon.

Figure 5.4 Absolute percent error 15 minutes ahead as predicted by time/series retained

Figure 5.5 Absolute percent error 30 minutes ahead as predicted by time/series retained

Figure 5.6 Absolute percent error 45 minutes ahead as predicted by time/series retained

Figure 5.7 Absolute percent error 60 minutes ahead as predicted by time/series retained

Figure 5.8 Actual occupancy recorded on July 19, 1993 for Zone 100, and the historical norm for that zone

Figure 5.9 Differences after the historical norms have been subtracted from the actual data

Figure 5.10 Differences between the actual data and predictions based upon the historical norms and an ARIMA (0,1,1) model

5.5 Adaptive Routine

The adaptive routines in the Shell watch for discrepancies between the predictions from the Kernel and the data coming from the highway. When discrepancies are found, these routines attempt to determine the cause of the problem and execute corrective actions. There are two classes of problems that these routines can handle: lane changes and changes in the volume of traffic entering the highway. These problems are handled by reducing the number of lanes used in the Kernel's calculations or by re-forecasting the entrance/exit flows and boundary conditions.

A series of conditional tests allow the adaptive routines to determine if a problem exists, and what that problem might be. First, each zone is tested to see if the observed occupancy deviates by more than seven units of occupancy. This roughly corresponds to the fifth and ninety-fifth percentiles historical distribution of the differences between actual and projected occupancies. If a discrepancy, greater than seven units (7% occupancy) is found, and an entrance or exit exists in the zone, the routine checks the difference between the projected and actual ramp flows. If these two differ by more than 30%, the ramp flows are re-forecasted and the Kernel is run again. If no problems are detected at the ramps, then the routines check for lane closures and report any suspected lane closures to the Kernel. If neither problems on the ramps or lane closures are detected, then the boundary conditions are re-forecasted.

Lane closures are detected by comparing the observed flow to the flow predicted by using the observed occupancy and the parameters of the speed density relationship. For example, assume the observed occupancy is 30%, $U_f = 65$ mph, and $K'_m = 20\%$. If the vehicle plus loop length is equal to 24 feet, then one would expect a flow of 4178 car/hour on a three lane road. If only two thirds of this volume is observed, then it is likely that only two of the three lanes are open. If only one third of this volume is observed, then it is likely that only one of three lanes is open. This hypothesis has been empirically tested against several cases of closure of one lane out of three which indicated a reduction of flow from 30 to 40 percents. It is due to this limited empirical evidence that we have used the above guideline.

6. THE KERNEL

6.1 Introduction

Traffic flow models are generally divided into two categories; microscopic and macroscopic. In the microscopic models,¹² each individual vehicle's motion is modeled according to some algorithm while in the macroscopic models average flow variables such as traffic densities and flows are determined by solving the partial differential equation for the conservation of vehicles. In the present investigation, we followed the latter approach.

Conventionally, the partial differential equation can be solved either by the finite element or the finite difference method. Since the latter method is computationally more time efficient than the former, the finite difference method is employed in the present analysis.

6.2 Brief Descriptions of the Various Macroscopic Traffic Models

The partial differential equation for the conservation of vehicles in one dimension is described by:

$$\frac{\partial}{\partial t}K(x,t) + \frac{\partial}{\partial x}Q(x,t) = G(x,t) - S(x,t) \quad (6.1)$$

where: $K(x,t)$ = vehicle density (vehicles/mile)

$Q(x,t)$ = traffic flow (vehicles/hour)

$G(x,t)$ = vehicle generation term from on-ramps (vehicles/hour.mile)

$S(x,t)$ = vehicle disappearance term from off-ramps (vehicles/hour.mile)

The above physical quantities are averaged over the number of lanes. From the definitions of aggregate flow and density, it follows that:

$$Q(x,t) = U(x,t) \cdot K(x,t) \quad (6.2)$$

where: U = vehicle speed (miles/hr)

In order to solve Equation 6.1, another relationship, called the speed-density relation, is required. Various forms for this relation were proposed by different investigators.¹³ Subsequent studies¹⁴ showed that these different speed-density relations can be derived from the car-following equation developed at Michigan by General Motors Corporation.¹⁴⁻¹⁶ Extensive and detailed analysis of data for the Long Island Expressway (L.I.E.) obtained from the INFORM Center, and carried out for this study, revealed that the best representation of the speed-density relation is given by the Gaussian form (Figure 5.3 and 6.14):

$$U(K) = U_f \exp \left\{ -0.5 \left[\frac{K}{K_m} \right]^2 \right\} \quad (6.3)$$

where: U_f = free flow speed
 K = vehicle density
 K_m = optimum density

The latter parameter, K_m is alternatively known as the critical density and represents that density at which maximum flow is reached. For the L.I.E., the critical density parameters were determined previously and range in values from 0.11 - 0.35 while the free-flow speeds are in the range from 53 to 79 miles/hr, depending on the zone number (Table 5.1)

On substituting Equation 6.2 and 6.3 into 6.1, one obtains the following partial differential equation for traffic density,

$$\frac{\partial K}{\partial t} = V_w(K) \frac{\partial K}{\partial x} \quad (6.4)$$

where:

$$V_w(K) = U(K) \left\{ 1 - \left[\frac{K}{K_m} \right]^2 \right\} \quad (6.5)$$

which is called the wave velocity and represents the propagation of traffic density. For traffic densities below the optimum value, K_m , the wave velocity is positive, whereas for $K > K_m$, the wave velocity is negative. As will be described shortly, the wave velocity plays a major role in the upwind differencing methods.

Various methods for solving the conservation equation in the discrete form have been developed in the literature.¹⁷ These techniques have been examined, modeled, programmed, and tested in the present study. A selection of the method is made on the basis of stability of solution, and good representation of the traffic data at high congestion. A brief review of each method follows.

6.2.1 The Forward Differencing Method

In the discrete form, the solution of Equation 6.1 is given by:

$$K(j,n+1) = K(j,n) - \frac{\Delta t}{\Delta x} [(Q(j,n) - Q(j-1,n)) + \Delta t [G(j,n) - S(j,n)] \quad (6.6)$$

where: j,n = integers represent space and time respectively.
 Δx = segment increment
 Δt = time increment

The time and distance variables, when defined relative to a prescribed time, t_0 , and distance, x_0 , are then given by:

$$\begin{aligned} x(j) &= x_0 + j \cdot \Delta x \\ t(n) &= t_0 + n \cdot \Delta t \end{aligned} \quad (6.7)$$

The stability requirement for the various solutions of Equation 6.1 was shown¹⁷ to lead to the following condition:

$$C = U_f \frac{\Delta t}{\Delta x} \leq 1 \quad (6.8)$$

which is known as the Courant-Friedrichs-Lewy stability criterion, generally cited in the literature as the Courant condition for brevity. This method was applied by Papageorgiou et al.¹⁸ to model 6 km of a freeway in Paris and by Cremer and May²⁰ to model a Santa Monica freeway.

6.2.2 The Lax Method

The above method (Equation 6.6) sometimes leads to instabilities. Lax and Wendroff¹⁹ showed that stability is restored if an average density at node j is substituted for the density from two adjacent nodes $j+1$ and $j-1$ i.e.,:

$$K(j,n) = 0.5 [K(j-1,n) + K(j+1,n)] \quad (6.9)$$

and the partial derivative of flow with respect to distance is represented by

$$\frac{\Delta Q}{\Delta x} = \frac{Q(j-1,n) - Q(j+1,n)}{2\Delta x} \quad (6.10)$$

Then the solution of Equation 6.1 becomes:

$$\begin{aligned} K(j,n+1) &= 0.5 [K(j-1,n) + K(j+1,n)] - 0.5 \frac{\Delta t}{\Delta x} [Q(j+1,n) + Q(j-1,n)] \\ &+ \Delta t [G(j,n) - S(j,n)] \end{aligned} \quad (6.11)$$

It can be easily shown by a Taylor series expansion that the Lax representation given by Equation 6.11 leads to the following equation for the case with no sources or sinks.

$$\frac{\partial K}{\partial t} = -V_w \frac{\partial K}{\partial x} + \frac{\Delta x^2}{\Delta t} (1-C^2) \frac{\partial^2 K}{\partial x^2} \quad (6.12)$$

As is clearly shown in Equation 6.12, an artificial diffusivity is introduced into the problem. Stability dictates that the artificial diffusion coefficient

$$\alpha_1 = \frac{\Delta x^2}{\Delta t} (1 - C^2) \quad (6.13)$$

is positive, which leads to the usual Courant condition, $C \leq 1$.

One can view the second term on the right hand side of Equation 6.12 as representing an error in the approximations. Therefore, in the application of the Lax method to traffic modeling, it is necessary to stipulate that the selected time mesh Δt be such that C is close to unity (Equation 6.13) in order to minimize the effect of the diffusion term. In the present modeling work, the time mesh satisfying this criterion is found to be 15 seconds.

In previous traffic modeling, the Lax method was extensively investigated by Michalopoulos et al.²¹⁻²² at the University of Minnesota by modeling both interrupted and uninterrupted traffic flow. The results of those detailed studies validated the Lax approach and demonstrated the stability of the solutions.

6.2.3 The First Upwind Differencing Method

This method is utilized in transport calculations. The solution is determined by the sign of the wave velocity, $V_w(j,n)$, and is given by:

$$\begin{aligned} K(j,n+1) &= K(j,n) - \frac{\Delta t}{\Delta x} [Q(j,n) - Q(j-1,n)] + \Delta t [G(j,n) - S(j,n)] & \text{for } V_w(j,n) > 0 \\ K(j,n+1) &= K(j,n) - \frac{\Delta t}{\Delta x} [Q(j+1,n) - Q(j,n)] + \Delta t [G(j,n) - S(j,n)] & \text{for } V_w(j,n) < 0 \end{aligned} \quad (6.14)$$

In this method, the conditions at any node are either affected by up-stream or down-stream conditions based on the direction of the wave velocity ($V_w(j,n)$) at that node. The sign of V_w is determined from the slope of the flow-density relation:²³

$$V_w = \frac{\partial}{\partial K} Q(K) \quad (6.15)$$

For a speed-density relation of a form given by Equation 6.3, the wave velocity at node j and time n is represented by:

$$V_w(j,n) = U(j,n) \left\{ 1 - \left(\frac{K(j,n)}{K_m(j)} \right)^2 \right\} \quad (6.16)$$

Thus, for $K(j,n) > K_m(j)$, then $V_w(j,n) < 0$ and the wave propagation is in the backward direction to the traffic direction, while for $K(j,n) < K_m(j)$, $V_w(j,n) > 0$, and the wave propagation is in the forward direction to the traffic direction.

As in the Lax method, a Taylor series expansion of Equation 6.14 leads to the result

$$\frac{\partial K}{\partial t} = V_w \frac{\partial K}{\partial x} + \frac{U \Delta x}{2} (1-C) \frac{\partial^2 K}{\partial x^2} \quad (6.17)$$

Thus, an artificial diffusivity is introduced with a diffusion coefficient of

$$\alpha_U = \frac{U \Delta x}{2} (1-C) \quad (6.18)$$

6.2.4 The Second Upwind Differencing Method

In this differencing technique, employed in thermal hydrodynamical calculations,^{17,24} average values for the interface velocities are considered. The conservation equation is then written in the form:

$$\Delta K_j = - \frac{\Delta t}{\Delta x} (V_R K_R - V_L K_L) + \Delta t [G(x,t) - S(x,t)] \quad (6.19)$$

where:

$$\begin{aligned} K_R &= K_j & \text{for } V_R > 0 & & K_R &= K_{j+1} & \text{for } V_R < 0 \\ K_L &= K_{j-1} & \text{for } V_L > 0 & & K_L &= K_j & \text{for } V_L < 0 \end{aligned} \quad (6.20)$$

and some average values (such as those in Equation 6.21) are considered for the interface velocities.

$$\begin{aligned} V_R &= 0.5 (V_j + V_{j+1}) \\ V_L &= 0.5 (V_j + V_{j-1}) \end{aligned} \quad (6.21)$$

Other averaging schemes can also be considered.¹⁷

6.3 Model Parameter Determinations

To subject the model to tests and validations, it is necessary to determine, as accurately as possible from the inductive loop detector data, the parameters which enter into the solution of the conservation equation of traffic as represented by Equation 6.6, 6.11, 6.14, and 6.19. Two of the parameters, free-flow speed, U_f , and critical density, K_m , have been determined previously from an analysis of the loop detector data collected during June-July 1993 (Table 5.1). An additional important parameter which is required in the modeling analysis is the average vehicle length.

Furthermore, in order to carry out sensitivity analysis, additional factors which were taken into considerations are: (1) the range and variability of these parameters as a function of time and zone, and (2) the quality of the data. This can be accomplished by carrying out a non-linear least squares fits (LSF) to the data and observing consistencies, as well as any inconsistencies which may result in this procedure.

6.3.1 Average Vehicle Length

Since the traffic density variable in Equation 6.1 is difficult to measure directly, the occupancy variable, defined previously and which is determined by loop detectors, is generally employed in road traffic and control. The relationship between these two variables is described by:¹⁰

$$K = \frac{52.80 K'}{(L_v + L_d)} \quad (6.22)$$

where: K' = traffic occupancy expressed in percentage units
 L_v = average vehicle length in a zone expressed in feet
 L_d = loop detector length expressed in feet

The loop detector length employed in the present analysis is 6.0 feet.

The INFORM loop detector data gives occupancies per lane and total flows, both averaged over 15 minute intervals, as a function of time, day, and zone number. From the measured occupancies, traffic flows are calculated by the relation:

$$Q(K', N) = \frac{N \cdot 52.80 \cdot K' \cdot U(K')}{(L_v + 6.0)} \quad (6.23)$$

where: N = the effective number of lanes (1-3)
 U = average vehicle speed in a zone

The average vehicle speeds employed in the present analysis can be either those values deduced by the INFORM Center or those obtained with the aid of Equation 5.3. A non-linear least squares fit to Equation 6.23 can determine the average vehicle length, as well as the number of open lanes for a zone. A representation of the section of the L.I.E. of interest for the modeling is represented in Figure 6.1. The zone number designations, locations of the loop detectors, and the distances between loop detectors are indicated in Figure 6.1.

It is noteworthy that in contrast to the 1993 data, March 5th-19th, 1994 INFORM loop detector data, which was received lately, contains the INFORM deduced speeds. As a result, it is of particular interest to carry out a LSF analysis in terms of the BNL, as well as the INFORM speeds. The results of the model fits for a fixed number of lanes, 3 are described in Figures 6.2 through 6.16 for L.I.E. zones 94, 95, 100, 102, 105 for Sunday, May 20, 1994, as well as Monday, May 21, 1994. Note that for zone 94 (Figure 6.7) no data is available between about 9:30 a.m. and 16:30 p.m. Similar situations of lack of data are frequent occurrences for zones 95, 98, and 99. It is of interest and importance to note that the BNL speed predictor performs slightly better than the INFORM speed predictor as evidenced by the goodness of fit. A summary of the deduced average vehicle lengths as a function of zone number (94-107) are depicted in Figures 6.12 and 6.13. Four interesting features which emerge from the present analysis and are graphically described Figures 6.12 and 6.13 are:

- 1) The vehicle length values, when derived from a linear-speed predictor (INFORM algorithm which is similar to Equation 6.23), seem to be decreasing with zone number.

- 2) The above observed trend is smoothed out to a large extent when the BNL, instead of the INFORM, speed predictor is employed.
- 3) The average vehicle length on Sunday is less than that on Monday.
- 4) The occupancy values are generally consistent with the traffic flows. However, there are cases where large deviations between measured and calculated flows according to Equation 6.23 do exist (Figure 6.11 during the same time interval 16:00-18:00).

Figure 6.1 A configuration of the L.I.E. as modeled in the current study. In the analysis, the nodes are positioned at the locations of the loop detectors.

Figure 6.2 A comparison between flows calculated from occupancies, speeds and measured INFORM flows for Sunday, March 20, 1994. A least squares fit to the data yielded an average vehicle length of 13.66 feet for zone 94 on the basis of 3 open lanes.

Figure 6.3 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Sunday, March 20, 1994. A least squares fit to the data gave an average vehicle length of 13.87 feet for zone 95 on the basis of 3 open lanes.

... Measured Flow ◇ ◇ ◇ Calculated Flow-INFORM Speed
 Figure 6.4 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Sunday, March 20, 1994. A least squares fit analysis yielded an average vehicle length of 11.18 feet for zone 100 based on 3 open lanes.

... Measured Flow ◇ ◇ ◇ Calculated Flow
 Figure 6.5 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Sunday, March 20, 1994. A least squares fit to the data gave an average vehicle length of 10.52 feet for zone 102.

... Measured Flow ◇ ◇ ◇ Calculated Flow

Figure 6.6 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Sunday, March 20, 1994. From a least squares fit analysis, on an average vehicle length of 10.18 feet is determined for zone 105.

... Measured Flow ◇ ◇ ◇ Calculated Flow—BNL Speed

Figure 6.7 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Monday, March 21, 1994. A least squares fit analysis of the data gave an average vehicle length of 17.11 feet for zone 94. Note the absence of data between 9:30 a.m. and 4:30 a.m.

... LIE flow ◇ ◇ ◇ LIE calculated flow-BNL speed

Figure 6.8 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Monday, March 21, 1994. A least squares fit analysis, to the data resulted in an average vehicle length of 17.5 feet for zone 95 based on 3 open lanes.

... Measured Flow ◇ ◇ ◇ Calculated Flow-BNL Speed

Figure 6.9 A comparison between flows computed from occupancies, speeds and measured INFORM flows for Monday, March 21, 1994. A least squares fit to the data yielded an average vehicle length of 16.38 feet for zone 100 based on 3 open lanes.

... Measured Flow ◇ ◇ ◇ Calculated Flow-BNL Speed

Figure 6.10 A comparison between flows calculated from occupancies, speeds and measured INFORM flows for Monday, March 21, 1994. A least squares fit analysis of the data yielded an average vehicle length of 15.69 feet for zone 102 based on 3 open lanes.

... Measured Flow ◇ ◇ ◇ Calculated Flow-BNL Speed

Figure 6.11 A comparison between flows calculated from occupancies, speeds and measured INFORM flows for Monday, March 21, 1994. A least squares fit analysis of the data gave an average vehicle length of 18.45 feet for zone 105 based on 3 open lanes.

Figure 6.12 Variation of the average vehicle length with zone number derived on the basis of the INFORM, as well as the BNL speed algorithms for Sunday, May 20, 1994.

Figure 6.13 Variation of the average vehicle length with zone number derived on the basis of the INFORM, as well as the BNL speed algorithms for Monday, May 21, 1994. An average value of 15.5 feet is deduced for the average vehicle length.

Observations 1 and 2 can be possibly interpreted in terms of the existence of an additional term in the speed, U , as in the BNL predictor, which describes the change of speed with traffic density, as well as the change of density with distance. Note that when $\delta K/\delta x = 0$, this proposed additional term disappears. Trend 3 can be explained in terms of a lack or absence of commercial traffic on Sundays. Observation 4 can be possibly attributed to a change in the traffic flow pattern i.e., an increase in commercial traffic during that time.

6.3.2 Free-Flow Speeds and Critical Densities

In order to gain insight into the variability of the free-flow speed, U_f , and the critical density, K_m , a detailed analysis of the March 1994 loop detector data was carried out. Non-linear least squares fit analysis was carried out on two dependent variables, the speed or the traffic flow, in terms of the independent parameter, traffic occupancy, K' . The results of the analysis, representing these two approaches, are depicted in Figures 6.14 and 6.15. The values of the determined parameters are summarized in Table 6.1 for a section of the main highway which is pertinent to the model and in Table 6.2 for the corresponding ramps. In the fourth column of Table 6.1 are shown the critical occupancies for zones 102-107 when extracted from data analysis of flows.

Figure 6.14 A Gaussian form for the speed-density relation is derived for the L.I.E. from a non-linear squares fit analysis for Mondays through Fridays, March 5-19, 1994. For zone 105, the free flow speed and the optimum occupancy are 70.46 miles/hr and 23.5% respectively.

18 feet average vehicle length is assumed
least squares fit for occupancies > 0.08

Figure 6.15 The speed density relation can alternatively be derived from an analysis of flow as a function of occupancy for Mondays through Fridays, March 5-19, 1994. Note that at optimum occupancy of 22.6%, the flow attains its maximum value. Below optimum density, the slope is positive, while above it, the slope is negative.

Table 6.1 Speed-Density and Optimum Density Parameters Obtained for the L.I.E. During March 1994

ZONE	U_f (miles/hr)	$K'_m/100$ (%)	$K'_m/100$ (%)
94	68.7 ± 1.0	0.218 ± 0.014	
95	68.4 ± 0.7	0.236 ± 0.014	
96	62.0 ± 1.3	0.122 ± 0.023	
97	80.9 ± 0.7	0.175 ± 0.009	
98	77.8 ± 1.0	0.114 ± 0.004	
99	79.5 ± 1.1	0.145 ± 0.006	
100	72.5 ± 0.8	0.230 ± 0.018	
101	76.3 ± 1.0	0.171 ± 0.014	
102	75.5 ± 1.8	0.131 ± 0.012	0.116
103	74.9 ± 0.3	0.167 ± 0.012	0.172
104	78.3 ± 0.7	0.127 ± 0.006	0.129
105	70.5 ± 0.6	0.235 ± 0.006	0.226
106	75.5 ± 0.6	0.189 ± 0.005	0.184
107	81.3 ± 0.4	0.193 ± 0.004	0.165

A detailed comparison between Tables 6.1 and 6.2 clearly shows good agreements in the results for zones 94 and 100-107 while for the remainder of the zones large disagreements exist. To a large extent this situation can be interpreted in terms of the quality of the data or possibly lane closure. For zones 94, 100-107, the loop detector data are in good shape in contrast to the other zones.

Table 6.2 Speed-Density and Critical Density Parameters for Exits and Entrances
Obtained for the L.I.E. During March 1994

Zone	Name	Ramp Type	Ramp Lanes (miles/hr)	U_r (miles/hr)	$K'_m/100$ (%)
96	Deer Park	exit	1		
98	Deer Park	entrance	1		
101	Commack Rd.	exit	1		
102	C/D Rd. Eb	exit	2	63.9 ± 1.4	0.090 ± 0.010
105	Sagtikos	entrance	1	24.6 ± 0.5	0.165 ± 0.020
107	C/D Rd. Eb	entrance	2	69.2 ± 1.0	0.074 ± 0.004

6.4 Model Applications to the Long Island Expressway

6.4.1 L.I.E. Configuration

The original plan for traffic modeling of the network included a region within Suffolk County defined by:

- Eastbound Long Island expressway from Deer Park Road to Wicks Road
- Eastbound Northern State Parkway from Deer Park Road to Sagtikos State Parkway.
- Sagtikos State Parkway (north and southbound).

Soon it became evident that 100% of the ramp data for the Northern State Parkway and the Sagtikos Parkway are missing (Appendix A). In order to test and validate the models, the study was restricted to the eastbound Long Island Expressway for which traffic data are to a large extent complete.

The specific section of the Long Island Expressway pertinent to our modeling effort is bounded between zones 94 and 107, and is 6.14 miles long (Figure 6.1). The segment sections or nodes for the solutions are adopted at the location of the loop detectors. The traffic direction is from left to right (Figure 6.1). The distances between loop detectors (nodes) are extracted from the INFORM document (INFORM L.I.E. Record Plans, Suffolk County). Fourteen nodes of roughly equal meshes resulted. In the L.I.E. section of the main highway heading towards eastern Long Island, there are three exits located close to and just before zones 96, 101, and 102 and three entrances before and close to the detectors designated by zones 98, 105, 107 (Table 6.2).

6.4.2 Initial and Boundary Conditions

In order to determine solutions to the continuity equation, as described by the partial differential equation (Equation 6.1), initial conditions at time, t_0 , and boundary conditions, which are a function of time, need to be specified. In addition, the generation and dissipation terms for exits and entrances have to be known as a function of time. To satisfy the Courant condition (Equation 6.8) for the present configuration, and minimize the error incurred due to the artificial diffusivity, a 15.0 second time mesh satisfied these criteria and was considered in the modeling analysis. Since the INFORM loop detector data are averaged over 15 minute intervals, linear interpolation routines are used in order to generate 15 second values for the boundary conditions, as well as the generation and dissipation terms.

The loop detector data for zones 94 and 107, which are in very good condition, are considered as the boundary conditions of the problem. One of the basic problems facing us is the frequent lack of traffic occupancy data for certain zones, examples of which are zones 95, 98 and 99. In the present modeling analysis, if such an occurrence happens, occupancy data from adjoining zones are assumed. Such a procedure did not lead to any inconsistencies with the outcome of the results.

Initially, the historic parameters, U_f and K'_m , for the speed-occupancy Equation 6.3 for the various zones were obtained from the Shell and are adopted in the analysis.

6.5 Tests and Validations of the Lax Model

The Lax method was followed in the present work because of the results of previous macroscopic traffic modeling.^{21,22} However, the results of the first and second upwind differencing methods (Equations 6.14 and 6.19), were explored in order to gain additional insight into the methodology. To validate and test the Lax method for the L.I.E. configuration, the Kernel predictions were compared with the actual data of the INFORM Center. Additional supporting basis for the Lax technique is derived on the basis of comparative studies with the predictions of the two upwind differencing methods. The results of such comparisons are described briefly in the following section and a few illustrative examples are presented. However, the details of these studies, which were carried out after the conclusion of the present task, will be reported elsewhere.²⁵

To carry out the computations, the Kernel requires the following information:

- 1) Initial and boundary conditions
- 2) On-ramp and off-ramp flows
- 3) The parameters of the speed-density relation for each segment
- 4) The average vehicle length during the particular day under study

The initial conditions are comprised of the occupancy data for zones 94-107 at time t_0 . The boundary conditions are the time dependent occupancy data for zones 94 and 107, bounding the present configurations. The generation and dissipation terms, $G(j,n)$ and $S(j,n)$ in Equation 6.11, are the traffic flows per unit distance for the on-ramps at zones 98, 105, and 107; and off-ramps for zones 96, 101, and 102. The historical parameters K'_m and U_f described in previous sections are adopted. The average vehicle lengths, 18.0 and 15.5 feet, deduced in previous analysis are used.

For the purpose of testing and validation, it is important to note that, in contrast to the integrated Shell-Kernel model, the actual time dependent INFORM information, is utilized initially as an input to the Kernel. Note that in the integrated model, the Shell considers the loop detector data (occupancies for zones 94 and 107 and pertinent ramp flows) for one hour. Subsequently, the Shell carries out future projections of the boundary conditions and the ramp flows for the next hour by application of time-series analysis. It is those projected values which are utilized by the Kernel in the computations in the integrated model.

Since the main objective of this project is the prediction of congestion, the various models were subjected to tests under congested conditions e.g., for those cases of the time of the day when occupancies close to, or larger than, the optimum value ($\sim 20\%$) are reached. On this basis, INFORM data at congested times 15:30 -19:00 were closely scrutinised to determine whether the flows and occupancies correspond to normal traffic conditions. Abnormal traffic conditions consist of lane closures, traffic accidents, or bad weather conditions. Occupancies, as well as speeds and flows, were computed by the Kernel for each segment for a prescribed total time interval of 90 minutes. Extensive and detailed computations were carried out under a variety of congested conditions. In all cases considered, the Lax method proved to be quite stable, while the forward differencing and the first upwind differencing methods exhibited unstable behavior when dealing with large congested conditions. The temporal profiles of the occupancy as a function of distance (in segment units) and time (in minutes) are generated by the Kernel on the basis of the Lax method and are shown in Figures 6.16 and 6.17 for May 24, 1993.

Figure 6.16 Occupancy profile for the L.I.E. as a function of segment number (zone) and time generated by the Kernel on the basis of the Lax method for May 24, 1993, 15:30-17:00. The direction of traffic is from right to left. Segments 1 and 14 correspond to zones 94 and 107 respectively and are considered as the boundary conditions for the problem. The time is expressed in units of minutes.

Figure 6.17 Occupancy profile for the L.I.E as generated by the Kernel on the basis of the Lax method for May 24, 1993 from 17:15 to 19:00. See Figure 6.16 for more details.

Since the predicted occupancies and flows are computed every 15 seconds, it is necessary to collapse the calculated values into 15-minute intervals for comparison with the actual measurements. Unweighted average values for each 900 seconds are computed by the Kernel for each segment. The 15-minute results are summarized in Tables 6.3 and 6.4. Values which are enclosed in parenthesis are assumed in the analysis for the initial and boundary conditions. The dashed lines indicate lack of INFORM data at those entries.

Shown in Columns 1 and 2 are the time and the zone number designation, respectively. The predicted and measured flows and occupancies are given in Columns 3 and 4, respectively while the predicted and measured occupancies are summarized in Columns 5 and 6, respectively.

Graphical representations of the model predictions and comparison with the measurements are shown in Figures 6.18 through 6.23. The predicted and measured occupancy variables are plotted versus zone number for May 24, 1993 and for the time of day 15:45 -17:00 in 15-minute intervals. In these calculations, an average vehicle length, L_v , of 17.6 feet was assumed. The sensitivity of the Kernel predictions was explored by changing L_v by $\pm 20\%$. The solid, dashed, and chaindashed curves represent the predictions for average vehicle lengths of 14.1, 17.6, and 21.1 feet, respectively. The solid squares are the INFORM data.

Another set of calculations for May 24, 1993 which covers the time interval from 17:00 - 18:30, was carried out. A graphical display of the results for occupancies (Figures C-1 through C-6) and flows (Figures C-7 through C-18), are included in Appendix C.

Table 6.3 Comparison Between Model Predictions (Lax Method) Carried Out in the Present Study with INFORM for May 24, 1993 from 15:30 to 17:00

Time	Zone	Predicted Flow (vehicles/hr)	Measured Flow (vehicles/hr)	Predicted Occupancy (%)	Measured Occupancy (%)
15:30	94	5433	5007	(12)	12
15:30	95	5544	---	(12)	--
15:30	96	4690	4028	(11)	11
15:30	97	4320	4084	(10)	10
15:30	98	2920	3090	(6)	6
15:30	99	3060	---	(6)	--
15:30	100	4800	4650	(10)	10
15:30	101	4010	4123	(8)	8
15:30	102	2140	3000	(4)	4
15:30	103	3010	3246	(6)	6
15:30	104	4370	2995	(10)	10
15:30	105	5090	4125	(11)	11
15:30	106	5250	4461	(21)	21
15:30	107	5860	5194	(23)	23
15:45	94	5430	5420	(12)	12
15:45	95	5470	---	11.8	--
15:45	96	4560	4458	10.7	12
15:45	97	4490	4525	10.4	11
15:45	98	4590	---	10.3	--
15:45	99	4680	---	9.7	--
15:45	100	4420	4947	9.1	10
15:45	101	3930	4512	7.8	9
15:45	102	2860	3302	5.5	4
15:45	103	2950	3496	5.9	7

Table 6.3 Comparison Between Model Predictions (Lax Method) Carried Out in the Present Study with INFORM for May 24, 1993 from 15:30 to 17:00 (Cont'd)

Time	Zone	Predicted Flow (vehicles/hr)	Measured Flow (vehicles/hr)	Predicted Occupancy (%)	Measured Occupancy (%)
15:45	104	3370	3374	7.0	7
15:45	105	4840	4761	10.3	10
15:45	106	5000	4780	13.8	9
15:45	107	5440	5814	(14)	14
16:00	94	5764	5711	(13)	13
16:00	95	5620	---	12.2	0
16:00	96	4630	4713	10.9	12
16:00	97	4570	4948	10.6	12
16:00	98	4650	---	10.5	--
16:00	99	4770	---	9.9	--
16:00	100	4510	5522	9.3	11
16:00	101	3970	4987	7.9	10
16:00	102	2880	3909	5.5	5
16:00	103	2930	4125	5.8	8
16:00	104	3220	3946	6.6	8
16:00	105	4480	5497	9.3	13
16:00	106	4660	5434	11.7	12
16:00	107	5920	6103	(18)	18
16:15	94	5770	5592	(13)	13
16:15	95	5810	---	12.7	--
16:15	96	4840	4672	11.4	12
16:15	97	4780	4780	11.2	12

Table 6.3 Comparison Between Model Predictions (Lax Method) Carried Out in the Present Study with INFORM for May 24, 1993 from 15:30 to 17:00 (Cont'd)

Time	Zone	Predicted Flow (vehicles/hr)	Measured Flow (vehicles/hr)	Predicted Occupancy (%)	Measured Occupancy (%)
16:15	98	4820	----	11.0	--
16:15	99	4960	----	10.4	--
16:15	100	4680	5284	9.7	11
16:15	101	4090	4667	8.2	9
16:15	102	3030	3310	5.8	4
16:15	103	3190	3678	6.4	7
16:15	104	3760	3416	8.1	7
16:15	105	5510	4964	12.5	12
16:15	106	5290	4759	16.8	16
16:15	107	5530	5405	(26)	26
					0
					12
16:30	94	6840	5878	(17)	17
16:30	95	6390	----	14.4	--
16:30	96	5360	4793	12.8	16
16:30	97	5300	4916	12.6	12
16:30	98	5180	----	12.3	--
16:30	99	5350	----	11.5	--
16:30	100	5040	5289	10.6	11
16:30	101	4410	4760	8.9	9
16:30	102	3290	3303	6.4	4
16:30	103	3670	3667	7.6	7
16:30	104	4490	3385	10.6	7

Table 6.3 Comparison Between Model Predictions (Lax Method) Carried Out in the Present Study with INFORM for May 24, 1993 from 15:30 to 17:00 (Cont'd)

Time	Zone	Predicted Flow (vehicles/hr)	Measured Flow (vehicles/hr)	Predicted Occupancy (%)	Measured Occupancy (%)
16:30	105	6340	4836	16.8	10
16:30	106	5180	4559	21.7	16
16:30	107	5650	5476	(25)	(25)
16:45	94	5790	5590	(13.0)	13
16:45	95	6480	---	14.7	--
16:45	96	5510	4691	3.3	12
16:45	97	5560	4763	3.3	13
16:45	98	5390	---	3.2	0
16:45	99	5670	---	2.5	0
16:45	100	5420	5428	1.7	13
16:45	101	4860	4817	0.1	10
16:45	102	3840	3240	7.7	4
16:45	103	4500	3475	9.9	6
16:45	104	5032	3325	3.8	6
16:45	105	6530	4674	0.0	10
16:45	106	4920	4292	3.7	19
16:45	107	5360	4860	(27)	27
17:00	94	6080	5845	(14)	14
17:00	95	5950	---	13.1	--
17:00	96	4900	4890	11.6	12
17:00	97	4850	5051	11.4	13
17:00	98	4860	---	11.1	0

Table 6.3 Comparison Between Model Predictions (Lax Method) Carried Out in the Present Study with INFORM for May 24, 1993 from 15:30 to 17:00 (Cont'd)

Time	Zone	Predicted Flow (vehicles/hr)	Measured Flow (vehicles/hr)	Predicted Occupancy (%)	Measured Occupancy (%)
17:00	99	5000	—	10.5	0
17:00	100	4690	5599	9.7	12
17:00	101	4000	4896	8.0	10
17:00	102	2530	3381	4.8	4
17:00	103	3032	3522	6.1	7
17:00	104	4300	3419	9.8	7
17:00	105	6420	4040	17.5	25
17:00	106	4842	3523	24.1	30
17:00	107	4620	4436	(31)	31

Table 6.4 Comparisons Between Model Predictions (Lax Method) of the Present Study with the INFORM Data for May 24, 1993 from 17:00 to 18:30

Time	Zone	Predicted Flow (vehicles/hr)	Measured Flow (vehicles/hr)	Predicted Occupancy (%)	Measured Occupancy (%)
17:00	94	6080	5590	(14)	14
17:00	95	5900	—	(13)	—
17:00	96	5060	4691	(12)	12
17:00	97	5500	4763	(13)	13
17:00	98	5360	—	(13)	—
17:00	99	5520	—	(12)	—
17:00	100	5520	5428	(12)	12
17:00	101	4820	4817	(10)	10
17:00	102	2140	3240	(4)	4
17:00	103	3440	3475	(7)	7

Table 6.4 Comparisons Between Model Predictions (Lax Method) of the Present Study with the INFORM Data for May 24, 1993 from 17:00 to 18:30 (Cont'd)

Time	Zone	Predicted Flow (vehicles/hr)	Measured Flow (vehicles/hr)	Predicted Occupancy (%)	Measured Occupancy (%)
17:00	104	3380	3325	(7)	7
17:00	105	6160	4674	(25)	25
17:00	106	3690	4292	(30)	30
17:00	107	4600	4860	(31)	31
17:15	94	6080	5806	(14)	14
17:15	95	6120	—	13.6	—
17:15	96	5080	4910	12.0	12
17:15	97	5040	4831	11.9	12
17:15	98	5000	—	11.6	—
17:15	99	5180	—	11.0	—
17:15	100	4880	5364	10.2	12
17:15	101	4190	4768	8.4	9
17:15	102	2740	3099	5.2	4
17:15	103	3400	3265	6.9	6
17:15	104	4630	3094	11.4	7
17:15	105	6470	4167	19.2	26
17:15	106	4590	4140	25.5	26
17:15	107	5180	5195	(28)	28
17:30	94	6080	5867	(14)	14
17:30	95	6120	—	13.6	—
17:30	96	5070	5190	12.0	13
17:30	97	5030	5243	11.9	13
17:30	98	5000	—	11.6	—

Table 6.4 Comparisons Between Model Predictions (Lax Method) of the Present Study with the INFORM Data for May 24, 1993 from 17:00 to 18:30 (Cont'd)

Time	Zone	Predicted Flow (vehicles/hr)	Measured Flow (vehicles/hr)	Predicted Occupancy (%)	Measured Occupancy (%)
17:30	99	5150	---	10.9	--
17:30	100	4820	5872	10.0	13
17:30	101	4100	5263	8.2	11
17:30	102	2350	3146	4.4	4
17:30	103	2770	3364	5.5	6
17:30	104	4050	3216	8.9	7
17:30	105	6320	3808	16.4	30
17:30	106	4930	3546	23.6	32
17:30	107	4810	4649	(30)	30
17:45	94	5780	5742	(13)	13
17:45	95	5960	---	13.1	--
17:45	96	4890	4837	11.5	12
17:45	97	4850	4851	11.4	11
17:45	98	4880	---	11.2	--
17:45	99	5000	---	10.5	--
17:45	100	4670	5532	9.7	11
17:45	101	3910	4897	7.7	10
17:45	102	1920	2941	3.6	4
17:45	103	2240	3190	4.4	6
17:45	104	3500	3079	7.3	12
17:45	105	6020	3600	14.	37
17:45	106	5060	3752	22.	29
17:45	107	5180	4828	(28)	28

Table 6.4 Comparisons Between Model Predictions (Lax Method) of the Present Study with the INFORM Data for May 24, 1993 from 17:00 to 18:30 (Cont'd)

Time	Zone	Predicted Flow (vehicles/hr)	Measured Flow (vehicles/hr)	Predicted Occupancy (%)	Measured Occupancy (%)
18:00	94	6080	5962	(14)	14
18:00	95	5950	---	13.2	--
18:00	96	4930	4918	11.6	12
18:00	97	4870	4916	11.4	12
18:00	98	4870	---	11.2	--
18:00	99	5000	---	10.5	--
18:00	100	4660	5511	9.6	11
18:00	101	3910	4913	7.6	10
18:00	102	2010	3086	3.8	4
18:00	103	2180	3289	4.2	6
18:00	104	3030	3152	6.1	7
18:00	105	5340	4027	11.8	15
18:00	106	5350	4106	19.4	18
18:00	107	5510	5022	(26)	26
18:15	94	5780	5679	(13)	13
18:15	95	5980	---	13.2	--
18:15	96	4990	4925	11.8	12
18:15	97	4950	4740	11.7	11
18:15	98	4950	---	11.4	--
18:15	99	5100	---	10.8	--
18:15	100	4780	5384	10.0	12
18:15	101	4080	4812	8.2	10
18:15	102	2380	3163	4.5	4
18:15	103	2540	3357	5.0	6

Table 6.4 Comparisons Between Model Predictions (Lax Method) of the Present Study with the INFORM Data for May 24, 1993 from 17:00 to 18:30 (Cont'd)

Time	Zone	Predicted Flow (vehicles/hr)	Measured Flow (vehicles/hr)	Predicted Occupancy (%)	Measured Occupancy (%)
18:15	104	3260	3189	7.0	6
18:15	105	5320	4523	11.7	17
18:15	106	5390	4521	18.2	16
18:15	107	5760	5300	(24)	24
18:30	94	6080	5883	(14)	14
18:30	95	5960	---	13.1	--
18:30	96	4970	4691	11.7	12
18:30	97	4900	4763	11.5	12
18:30	98	4890	----	11.3	--
18:30	99	5030	----	10.6	--
18:30	100	4710	5428	9.8	12
18:30	101	4030	4817	8.1	10
18:30	102	2530	3240	4.8	4
18:30	103	2590	3475	5.1	7
18:30	104	3000	3325	6.1	7
18:30	105	4430	4674	9.2	10
18:30	106	4750	4292	12.7	8
18:30	107	4730	4860	(11)	11

Figure 6.18 A comparison between predicted and measured INFORM occupancies as a function of zone number at 15:45 on May 24, 1993. The calculations are carried out by the Kernel on the basis of the Lax method.

Figure 6.19 A comparison between predicted and measured INFORM occupancies as a function of zone number at 16:00 on May 24, 1993. The calculations generated by the Kernel on the basis of the Lax method.

Figure 6.20 A comparison between predicted and measured INFORM occupancies as a function of zone number at 16:15 on May 24, 1993. The calculations are carried out by the Kernel on the basis of the Lax method.

Figure 6.21 A comparison between predicted and measured INFORM occupancies as a function of zone number time 16:30 on May 24, 1993. The calculations are generated by the Kernel on the basis of the Lax method.

Figure 6.22 A comparison between predicted and measured INFORM occupancies as a function of zone number at 16:45 on May 24, 1993. The predictions are based on the Lax method.

Figure 6.23 A comparison between predicted and measured INFORM occupancies versus zone number at 17:00 on May 24, 1993. The predictions are based on the Lax method.

On the basis of Tables 6.3 and 6.4, Figures 6.18 through 6.23, and Figures C-1 through C-6, the following observations can be made:

- 1) At zero time, corresponding to time 15:30 and 17:00 in Tables 6.3 and 6.4, the predicted and measured flows serve as a check on the consistency of the historic parameters of the equation of state with the present data for May 24, 1993, as well as the scaling factor which takes into account the conversion of occupancy to density and vice versa (Equation 6.23).
- 2) Qualitatively, the model predictions on the basis of the Lax method reproduce very well the general trend of the data. Quantitatively, agreements between Kernel model predictions and measurements are generally within 10%-20%, particularly for zones 95-104.
- 3) In some cases, the worst cases of disagreement generally occur for zones 105 - 106 (Figures 6.21 through 6.23). On the other hand, good agreement between measurements and predictions are obtained for these zones (Figure 6.20).
- 4) Occurrence of large deviations between predictions and measurements can be interpreted as signatures of anomalies which can be corrected by the adaptive routines (Section 5.5).

6.6 Comparisons Between the Lax and the First Upwind Differencing Methods

To carry out an assessment of the performance of the Lax method, its predictions were compared not only with the actual data, but also with the predictions generated by the first upwind differencing method. The mathematical formulations of the first upwind differencing method was described previously and is represented by Equation 6.14. At this point, it is interesting to note the major differences between the first and second upwind differencing methods. In the first upwind differencing method, the decision of whether to carry out forward or backward differencing is solely determined by the condition as to whether the traffic flow is below or above saturation at a single node j . In contrast, in the second upwind differencing method, the condition is based on two adjacent nodes. In addition, an averaging of the interface speeds is carried out at the adjacent nodes. The latter feature has some similarity to the Lax method. The second upwind differencing method is second-order accurate in the calculations of the spatial derivatives. This method is presently under investigation and the results will be reported elsewhere.²⁵

Two sets of calculations were carried out for May 24, 1993:

- 1) one covering the time interval from 15:30 to 17:00 where the occupancy data at 15:00 were considered as the initial conditions, and
- 2) the other covering the time interval from 17:00 to 18:30, for which the occupancy data at 17:00 were considered as the initial conditions for the problem.

The results of the computations are graphically displayed in Figures 6.24 through 6.29. A detailed examination of these computations reveal the following significant features:

- Very good descriptions of the low occupancy data for zones 95 - 104 are achieved by the predictions of the Lax and the first upwind differencing methods.

- The predictions of the Lax method are generally in better agreement with the data for the time interval 15:45-17:00, as represented in Figures 6.24-6.26, than the other method.
- The first upwind differencing was able to capture a description of the congested conditions at zones 105 and 106 better than the Lax methods between 17:15 and 17:45 (Figures 6.27-6.28).
- However, during the time interval 18:00-18:30, the first upwind differencing generated predictions which deviated significantly from the data for zones 105 and 106 while the Lax method produced good agreement with the measurements.

In summary, the predictions of the Lax and the first upwind differencing methods are in very good agreement with the data for the low occupancy zones. However, the first upwind differencing method seems to predict congestions for a limited time close to the initial conditions, but fails to stabilize subsequently.

6.7 Future Refinements

Future improvements in the Kernel model and the speed algorithm, will be briefly discussed in this section.

6.7.1 Kernel Improvements

On the basis of the conclusions described in Section 6.6, studies will be carried out to develop an additional model which combines the best features of the Lax and upwind differencing methods.

Figure 6.24 A comparison between predicted and measured INFORM occupancies as a function of zone number at 16:15 on May 24, 1993. The predictions are based on the Lax and the first upwind differencing methods.

Figure 6.25 A comparison between predicted and measured INFORM occupancies as a function of zone number at 16:30 on May 24, 1993. The predictions are based on the Lax and the first upwind differencing methods.

Figure 6.26 A comparison between predicted and measured INFORM occupancies as a function of zone number at 17:00 on May 24, 1993. The predictions are based on the Lax and the first upwind differencing methods.

Figure 6.27 A comparison between predicted and measured INFORM occupancies as a function of zone number at 17:15 on May 24, 1993. The predictions are based on the Lax and the first upwind differencing methods.

Figure 6.28 A comparison between predicted and measured INFORM occupancies as a function of zone number at 17:45 on May 24, 1993. The predictions are based on the Lax and the first upwind differencing methods.

Figure 6.29 A comparison between predicted and measured INFORM occupancies as a function of zone number at 18:15 on May 24, 1993. The predictions are based on the Lax and the first upwind differencing methods.

6.7.2 Algorithm Refinements

As pointed out in the derivation of the average vehicle length described Section 6.3.1, the results indicated that another term in the speed algorithm exists. This additional term depends on the variation of speed with density, as well as variation of density with distance. It is of great interest to carry out detailed analysis, similar to the ones described in Section 6.3.1, for a determination of these parameters and for inclusion in the new model. In addition, if the average vehicle length turns out to be zone dependent due largely to the variation of commercial traffic on the main highway, such a variation should be included into the model.

7. PERFORMANCE OF INTEGRATED MODEL

In addition to testing and validation of each individual routine, the performance of the integrated ATOP package has also been extensively evaluated. This evaluation has been performed in two stages. In the first stage, the ATOP code was evaluated without using any adaptive or anomalous signature correction. In the second stage, the integrated performance, including the adaptive and incident correction routine, was assessed. The results of these evaluations are provided in the following sections.

7.1 Integrated ATOP Performance Excluding the Adaptive Routine

A large number of the test runs were performed using the ATOP code, however the adaptive and anomalous signature correction routines were disabled. The purpose of this evaluation was two fold; (1) to assess the specific parameters to be used for the adaptive routine, and (2) to evaluate the integrated performance of the ATOP code in terms of the expected prediction error.

A total of sixty case runs were performed. The graphical results of these runs illustrating the predicted flow and density vs. the measured flow and density for five selected zones within the section of the highway currently modeled were generated. A sample case run, showing the graphical results for Monday (5/10/93) from 8:00 A.M. to 9:00 A.M. for two zones (Z-94, and Z-103) are shown in Figures 7.1 and 7.2. The x-axis of these graphs shows the time of simulation in minutes, where 60 indicates the beginning of the prediction (in this example 8:00). At the start of simulation, (60 on x-axis) the predicted density is basically the initial condition and is the same as the observed density. However, the predicted flow is estimated based on the speed-density relationship (Equation of State) and may not exactly correspond to the observed flow. The good agreement between the predicted vs. the observed flow, at the start of prediction, indicates that the dynamic Bayesian routine is performing as expected and adjusts for the day and time specific variations of speed-density relationship (even the potential severe weather conditions). The satisfactory performance of the Bayesian routine is further substantiated by these graphs since at any time when the prediction of density is close to observed density, the prediction of flow would closely agree with the observed flow.

Another important observation that can be made based on these graphs relates to the agreement between prediction and observed flow and density in zone 94. Zone 94 is the inlet boundary for the section of the highway currently modeled in the ATOP. As discussed previously, the traffic density has to be forecasted at the boundaries using ARIMA and a differencing method described in Section 4. The close agreement observed in Figure 7.1, indicates the excellent performance manifested by this routine.

This is by no means an indication that forecasting always produces excellent agreements. As discussed previously, forecasting faces fundamental problems in capturing traffic waves, or sudden changes in highway configurations (e.g., reduction of lanes). It is to this end that the forecasted results are fed to the Kernel, and the forecasting is adaptively adjusted based on the deviation of observed vs. predicted traffic conditions within the network.

Finally, the integrated performance of the ATOP code, excluding its adaptive routine, is illustrated in Figures 7.3 through 7.5. Figure 7.3 shows the cumulative probability distribution

Figure 7.1 Monday May 10, 1993: Z-94 Starting at 08:00

Figure 7.2 Monday May 10, 1993: Z-103 Starting at 08:00

function for the mean errors of predicted occupancy. The mean error of predicted occupancy is obtained through the difference between the predicted occupancy and the actual occupancy for each zone averaged over 12 samples, each of five minutes average (one hour prediction period). The fitted distribution is normal with a mean of less than 1 and 95 percentile confidence bound of (-7) to (+7). In summary, this graph shows that the integrated ATOP code predictions are unbiased and in most cases the maximum error in the predicted occupancy is less than seven (e.g., if the ATOP predicts an occupancy of 9, the highest occupancy that may be actually in the highway is less than 16).

Figure 7.4 shows a similar graph to Figure 7.3 for the mean error in the predicted flow. However, it is shown in terms of percent of the mean of the actual flow. The 95 percentile for the percent error in flow is between -40% to +40%. Again, the ATOP estimates are approximately unbiased since the mean of the fitted normal distribution is about -5%. The large errors in flow prediction is expected since maximum flow takes place at occupancies around 20%. Therefore, an error of 7% in occupancy can produce an error of about 40% in flow. It is important to note that the ATOP flow prediction error should not be confused with flow prediction which were made using actual measured occupancies through the empirical relationship of speed and density (Equation of State). Based on the discussion provided in Section 4, the errors associated with the predicted flow using the equation of state is always less than 10%. All the adaptive routines discussed in Section 4, which are based on significant deviation of flow, use the flow prediction from the equation of state using measured occupancies. The ATOP flow predictions are not used in the adaptive routines, due to the potentially large errors that may be produced.

Finally, Figure 7.5 shows the scatter plot of mean-percent errors in flow vs. mean absolute errors in occupancy. As expected, the scatter plot shows strong correlation; that is, the higher the error in occupancy the higher the error in flow predicted by the ATOP.

Figure 7.3 CDF for mean prediction error of traffic density

Figure 7.4 CDF for the percent mean error in flow prediction

Figure 7.5 Scatter plot of mean error of flow vs. density

In conclusion the integrated performance of the ATOP code, excluding the adaptive routine, indicates that:

- 1) The ATOP code, excluding the adaptive routine, provides an unbiased estimate of flow and occupancy.
- 2) The absolute error of prediction for the occupancies made by the ATOP code is less than 7% in occupancy, when there is no anomalous condition or lane closure.
- 3) The percent error of prediction for the flows predicted by the ATOP code is less than 40% when there is no anomalous condition or lane closure. The prediction of flow based on the equation of state and the measured occupancies is less than 10%. Therefore, a large portion of the percent errors of the ATOP predicted flow is caused by inaccuracies in the prediction of occupancies.

In cases where an accident has taken place or certain lanes have been closed, the performance of the ATOP code without the adaptive routine is quite poor and this is not reflected in the earlier conclusions discussed. The performance of the ATOP code with the adaptive routine enabled is discussed below.

7.2 Integrated ATOP Performance Including the Adaptive Routine

A full evaluation of the systematic integrated performance of the ATOP code (including adaptive routine) is left for completion in Phase II of the proposed program. Several case studies have been performed which illustrate the power of the adaptive routine in both controlling the errors of prediction and detecting incidents and potential lane closures. Several areas of refinement have already been identified for future implementations. Here we will be discussing the results of three case studies to show the improvements gained from the currently implemented adaptive routine.

Figure 7.6 shows a traffic density graph for zone 105 on June 14, 1993 at 6:00 P.M. In this case study, the entrance and exit flows to the main highway were significantly higher than the historical data and the short term trend of the previous hour. The reason for this anomaly is unknown but it is suspected to have been caused by an anomalous condition outside the section of the highway modeled. The adaptive routine associated with the ramp adjustment was activated exactly 15 minutes after the start of simulation. It should be noted that the 15 minutes for adaptive correction currently programmed within ATOP code is due to the fact that the data is being received from INFORM averaged over 15 minutes, in every 15 minute interval. Therefore, adaptive adjustments can be made at the time that the INFORM data had arrived (every 15 minutes). This restriction will be removed in the future when INFORM will be capable of supplying data every minute. Figure 7.6 shows that with only three adaptive corrections much closer agreement and better predictive performance was manifested by the ATOP code.

Figure 7.6 The performance of adaptive routine 6:00 P.M. (June 14, 1993)

Figure 7.7 shows a traffic density graph for zone 106 on October 12, 1993 at 6:00 P.M. In this case, the database developed under HTDB shows severe weather (1.8" of precipitation) on that evening. This would affect both the parameters of the equation of state and the exit/entrance flow forecasts. Again, the adaptive routine was enabled four times to adjust for the anomalies being experienced. As a result of the adaptive routine, the performance of the ATOP code appears to have been improved.

Figure 7.8 shows the traffic density graph for zone 103 at 5:00 P.M. on May 6, 1993. Contrary of the other two case studies, in which the adaptive routine did not close any lanes, in this case at different adaptive corrections, three segments of highway lost one lane. Even though the construction and the incident database of the HTDB recorded an anomaly on that day, this information was not fed to the ATOP code. The information was withheld in order to assess whether the adaptive routine is capable of detection of these anomalies, adjusting the number of lanes per segment and other traffic parameters. More importantly, the information was withheld to test if the ATOP will be reset when two anomalies are cleared. Figure 7.8 shows two curves with adaptive corrections. The first curve, which has only two adaptive corrections, has identified the anomalous condition and has closed the proper lanes for the remainder of prediction without knowing that these lanes are opened about one hour after simulation. The second curve with five adaptive corrections had not only detected which lanes has to be closed but sometimes later (in about 135 minutes after the simulation), the adaptive routine recognizes that the closed lanes have to be opened up and the predictions need to be re-adjusted. The behavior of the adaptive routine in this case explicitly shows the increased performance and more accurate prediction.

Figure 7.7 The performance of adaptive routine 6:00 P.M. (October 12, 1993) severe rain

Figure 7.8 The performance of adaptive routine 5:00 P.M. (May 6, 1993) lanes closed

In summary, the first version of the adaptive routine currently incorporated in the ATOP code has performed satisfactorily in:

- a) controlling the errors in the ATOP predictions,
- b) providing intelligent detection of various kinds of anomalies, and
- c) correcting and adjusting for the anomalies with a maximum response time of one half hour.

The authors believe that the adaptive routine can be refined significantly. In addition, its response time will be improved in the future specifically when the one minute data would be available from INFORM. More systematic testing and implementation of the refinements is to be a major research task in Phase II of this program.

As discussed in the previous sections, the ATOP code has demonstrated the capability of predicting traffic conditions an hour ahead (within 5 minutes computation time) under normal traffic conditions.

In anomalous conditions, the adaptive routine will detect and adjust the prediction as necessary. The response time of the adaptive routine is heavily dependent on the time interval for receiving loop detector data. INFORM, at the time of this study, provided 15 minutes averaged data in 15 minute intervals. Under this constraint, the response time of the adaptive routine is currently about 30 minutes (about 2 data points). However, this limitation will be reduced substantially once INFORM begins reporting one minute data.

8. ATOP PROGRAM DESCRIPTION AND USER GUIDE

8.1 Program Structure

ATOP is a large program intended for demonstrating the feasibility of traffic congestion forecasting and detection of anomalies in terms of the equivalent number of lanes closed. This program is provided on an "as is" basis, without any warranties expressed or implied.

The key element of this program is a custom data structure called simply "Zones." Each section of highway is called a zone, and has many pieces of information attached to it. All such pieces of information for a single zone is an element of this data structure. The exact form of this data structure is shown in the included file "DECLARE.INC." Custom data structures are an extension specific to Microsoft FORTRAN, and a feature of most civilized computing languages. If you don't have Microsoft Power Station FORTRAN, you probably will not be able to use this program. It is intended that this program will later be rewritten in C. Therefore, the program was written to take advantage of features that will be available in that language.

The program logic is as follows:

- 1) Prompt user for date and time to start simulation
- 2) Initialize the data structure
 - get historical equation of state parameters
 - get historical traffic pattern (volume & occupancy)
- 3) Simulate loop detector data using archival information from INFORM
- 4) Update the equation of state parameters
- 5) Produce time series projections at the boundaries and ramps
- 6) Run the KERNEL
- 7) End Program - produce all reports and exit

The program terminates after four adaptive corrections, or after the three hours of data have been used up. The number of adaptive corrections can be easily changed by the user.

This program is quite large and has a number of files associated with it. The filetype conventions used for these files are as follows:

- FOR FORTRAN
- INC Include file
- HIS Historical data file
- CON Configuration file
- LOP Loop detector data file
- PRT Files written by the Shell to be printed

.FOR Files

- SHELL.FOR This is the main program file

.INC Files

DECLARE.INC	Contains the main data structure: zones
GET_HIST.INC	Retrieves historical data for the zones
BAYES_V3.INC	Bayesian updating of the equation of state
INTERPOL.INC	Linear interpolation
GET_LOOP.INC	Retrieves archival loop detector data
KERNLAX.INC	Solves the traffic flow and density conservation
PRINT.INC	Prints the contents of zones
DATEFUNC.INC	Day/date functions

.HIS Files

RAMP.HIS	EOS parameters for the ramps	UNIT = 20
MAIN.HIS	EOS parameters for the mainline	UNIT = 21
VOL_OCC.HIS	Historical volume and occupancies	UNIT = 22

.CON Files

DETAIL.CON	Detailed information on each zone	UNIT = 11
INIT.CON	Initialization information	UNIT = 10

.LOP Files

MAY.LOP	Loop detector data from May, 1993	UNIT = 30
JUN.LOP	Loop detector data from June, 1993	UNIT = 30
JUL.LOP	Loop detector data from July, 1993	UNIT = 30
AUG.LOP	Loop detector data from August, 1993	UNIT = 30
SEP.LOP	Loop detector data from September, 1993	UNIT = 30
OCT.LOP	Loop detector data from October, 1993	UNIT = 30
NOV.LOP	Loop detector data from November, 1993	UNIT = 30
DEC.LOP	Loop detector data from December, 1993	UNIT = 30

.PRT Files

ZONEDATA.PRT	Contents of the data structure	UNIT = 40
--------------	--------------------------------	-----------

For this version of the program, the Shell will always extract 3 hours worth of data from the data resource files. The first hours worth of data will be used to develop the Bayesian updates and ARIMA forecasts. The second two hours of data will be used to test the Kernel's predictions. Since the data are only available in 15 minute chunks from INFORM, the data resource files contain data in 15 minute intervals. After being retrieved, the data are interpolated into one minute intervals. This choice of time structure will be recoded to reflect the INFORM system as it exists on-line.

Historical and loop data have occupancies on a scale from zero to 100. The parameter K_m for the equation of state ranges from 0 to 1. This factor of 1100 must be accounted for in any calculations. The problem will be fixed in the next release.

Missing Data Conventions:

- 9 = always indicates data that is truly missing, and
- 1 = always indicates data you might expect to see, but should be missing. For example, ramp detector data in a zone without ramps.

8.2 User Guide for the ATOP Code

A copy of executable programs and the necessary files are provided in a tape submitted with this report. A directory listing of the files and the associated file sizes are provided in Table 8.1. The user may copy all the files from the tape to a directory of his/her desire on the hard disk. We also have included the executable file for a graphic interface user program called "SENSOR", which is written in C language and has a capability to read the output from the ATOP program and to graphically present the results in a color coded format on a highway network. This program was developed as a part of this contract for the future use and automated interface with the ATOP code. Integration and automated interface has not been completed yet. A brief description of the SENSOR program is provided in Section 8.3.

Table 8.1 List of Files and File Sizes Contained in the Tape

FILE NAME	FILE SIZE
MAY.LOP	9,668,751
NOV.LOP	9,360,000
DEC.LOP	9,672,000
OCT.LOP	9,672,000
JUN.LOP	9,360,001
JUL.LOP	9,672,001
AUG.LOP	9,672,000
SEP.LOP	9,360,000
SENSOR.EXE	227,335
SHELL2.EXE	324,608
RAMP.HIS	5,330
MAIN.HIS	5,330
VOL_OCC.HIS	1,216,801
INIT.CON	1,782
DETAIL.CON	5,330
OCC_VOL.DAT	153,852
SENSOR.CPP	15,978
SENSOR.IDE	23,622
SVGA16.BGI	8,910
SVGA16.H	815
LI-CN.IDF	4
LI-CN.LDF	12,000
LI-CN.SDF	26,400
LI-NC.IDF	4
LI-NC.LDF	12,000
LI-NC.SDF	26,400
LIC.IDF	4
LIC.LDF	12,000
LIC.SDF	26,400
LITT.CHR	5,131

Prior to running the ATOP code, the user shall assure that the following statements are in the autoexec.bat file:

```

PATH = C:\F32\BIN
Set LIB = C:\IMSL\FORTRAN\LIB;C:\F32\LIB
Set include = C:\IMSL\FORTRAN\INCLUDE;C:\F32\Include

```

Both the IMSL package and FORTRAN Power Station compiler is needed for this version of the ATOP.

To start running the program, the user shall type "SHELL2" at the desired directory prompt. The program will respond by:

This is a temporary user interface. Any mistakes made in this interface may cause the program to crash or produce bad data. Answer all the following prompts with integers.

The next line shown of the number monitor would be:

Input month (5-12)

At this time, the user shall select a month, May through December; i.e., the year 1993). As an example, the user may enter 6 for the month of June and then press enter.

The next input required is the day of the month. An example of input requested and the response is given below:

Input day (1-30/31)28

The last input required by the program would be when the simulation starts. There are two important notes that a user should know:

- 1) The time has to be entered in military format; e.g., 22:30, which is 8:30 pm.
- 2) The actual simulation will start an hour after the time specified since the first hour data is used for initiation. It is due to this reason that the time entered shall be less than 22:00 (or 10:00 pm).

From this point on, the program will run through various routines and keeps the user informed of what routine is being run through a message line displayed on the monitor. As an example:

- Initializing data structure
- Loading loop data and historical profiles
- Bayesian updating in progress

After completion of the first hour of forecasting, and no adaptive correction, the program displays the following message:

A DETAILED REPORT IS IN THE FILE OCC_VOL.PRN.
OCC_VOL.PRN CAN BE IMPORTED INTO QUATRO PRO

The program then provides the user with information about missing loop detectors and their treatment using historical data.

The program then repeats the same calculation up to four times to perform adaptive corrections. After the run is complete, the user shall examine the output file contained in OCC_VOL.PRN.

If the user uses the "edit" command to see the content of OCC_VOL.PRN, he/she may notice the following:

"Month = "6
 "Day = "28
 "Starting at = "1900
 " 9 = missing data"
 " 1 = data that does not, and should not, exist."
 "Actual occupancies & Kernel predictions every 5 minutes."
 "Type of adaptive correction = none"
 "Number of adaptive corrections = 0"
 "Time of adaptive correction = "

Effective number of lanes in each zone

"Zone"	Number of Lanes"
94	3
95	3
.	.
.	.
.	.
.	.
.	.
107	3

The file then follows by providing a tabular form of zone numbers, actual measured occupancy (denoted by zone number; e.g., Z94) and forecasted occupancy (denoted by K) as a function of time in minutes as shown below:

"Time"	"Z94"	"K"	"Z95"	"K".....
60	11.0	11.0	0	11.0
65	10.7	11.1	0	9.1
70	10.3	11.2	0	8.9
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
180	4.0	7.4	0	6.0

A similar table is also provided for traffic flow right below the table above. The remainder of the file contains the duplication of the above information for various numbers of adaptive corrections.

The information in this file can be imported into a spreadsheet package for a graphical presentation, or used as input to SENSOR programs for displaying color coded highway networks on the monitor.

8.3 SENSOR Program

The "sensor.cpp" program serves as the front-end module for ATOP. It is a graphical interface (GUI) which allows ATOP to display the outcome of its simulation. It has been written in C++, and is modular and expandable. It runs under DOS on a standard VGA-monitor, and a mouse is recommended for convenient data entry.

The program displays a geographical map of the network. The loop-detector sensors are displayed as squares with a small description next to them related to the zone-identification. Any sensor can be selected as the reference point. As the mouse moves, the distance (from the reference sensor) of any point within the geographical boundaries is displayed on the screen. The scale (ft/mile-motion-inch) has been selected for the INFORM corridor accordingly. Straight lines are used to indicate a flow connection between different loop detectors. These lines are colored according to the recorded/estimated vehicle speed. The displayed color is adopted according to the speed calibration diagram which appears to the right part of the display. The lack of a colored link between two sensors indicate either that the loop detector is malfunctioning (not reporting), or that there is indeed no flow between these sensors. The user can point to the particular zone, and request information about its structure. The zone structure (C++ object class) contains: a) the loop detectors in this zone, and b) the zones(s) to which its flow will propagate. The user can disable/enable specific loop any item within the zone structure. Furthermore, the user can define new zones, or eliminate existing ones. In this way, fictitious simulation test-runs can be implemented in the form "what if...a loop detector fails/...reroute the traffic".

The GUI-program has been currently interfaced with ATOP to read (manually the speed-prediction data files. Subsequently, the retrieved predicted speeds are displayed. The user can alter the duration of the update interval. The default value is one second, while the actual predicted speed update interval is 15 minutes. The program currently can handle up to 65535 zones, and can be interfaced to directly record the loop detector positions (x-y coordinates) from commercial mapping packages. For the INFORM corridor, the zone locations were recorded through a manual interface with the MAPinfo software package.

A demonstration version of this program is also provided in the computer tape submitted with this report. The user can start running the code by entering "SENSOR" at the desired directory prompt. A color coded vehicle speed will be shown on the right side of the monitor with several menu items at the top of the display. A beginner may type "L" for load. In response to the file name, the user may enter "LI-CN". A display of all the highway junctions on the L.I.E will be shown on the monitor. To start the simulation and connect all of the junctions, the user shall click on the right button of the mouse. In response, all the junctions will be connected and the color of the connected segments will change reflecting the change of speed in the vehicles. The simulation can be stopped by the user at any time by entering any key. The program will go back to the menu mode. A more experienced user may construct a new highway network by using the menu with the "connect" heading and continue to follow the instructions.

8.4 Historical Databases

A large amount of data was retrieved from historical databases and used for the purpose of developing, testing, and validating the ATOP code. Some of the data dealing with traffic and occupancy flow collected under the empirical knowledge database task, was directly transformed and used as input to the ATOP code (files with the extension "LOP" in Section 8.1). Other data files were incorporated into a SPSS database which contains not only traffic flow and occupancy, but also weather condition (e.g., precipitation, snow, etc.), average temperature, and construction. FARS data was also collected and it is presently available in hard copy format.

The SPSS data files are provided on tapes labeled as tapes 1 and 2. Each tape contains several SPSS files with the extension "SAV", and the file name corresponding to the name of the month to which the file relates, e.g., February.SAV. This is historical data which provides weather and road conditions, traffic flow, traffic occupancy, average temperature, specifies the type of day (working day, specific holiday) etc. These files cover all of the months in 1993 and provide information averaged over 15 minutes intervals.

The files can be displayed using the SPSS package. Also included in tapes 1 and 2 is the SPSS routine that was used to generate these data files.

These databases, in addition to the files with a "LOP" extension in the ATOP input, were the foundation of all the empirical knowledge used in this research. The empirical knowledge database supported the development, testing, and validation of the ATOP code as discussed earlier.

8.5 Future Programming Improvement

There are several areas of improvements envisioned for enhancement of the ATOP code. These include:

- use of partially missing data; partially missing data is currently treated as completely missing data,
- addition of a signal validation routine to detect incorrect data reported by loop detectors,
- enhancement of adaptive initiation and anomalous detection routine as discussed previously,
- graphical display routine to substitute current manual interface to the Quatro Pro software,
- automated interface routine between ATOP and SENSOR program,
- development of a user interface routine for confirmation of the anomalies detected or eliminated,
- development of a user interface routine for input of external information, such as construction, severe weather conditions, etc., and
- improvement of the forecasting routine which accounts for loop detector signals from adjacent segments.

9. INSIGHTS AND AREAS FOR FUTURE RESEARCH

The main product of Phase I of this program is the traffic simulations and predictions model. Additionally, several other spin-off products have resulted from the work. These interim results vary in type, level of sophistication, and maturity. They are discussed here in summary fashion.

1. **Historical Traffic Data Base:** The HTDB brings together many databases to allow the generation of insights that were not previously available. The scope of HTDB can be extended beyond the needs of the ATOP code; however, other areas where HTDB can be utilized should be explored.
2. **Identification of roadway congestion sensitivity:** The analyses performed to date have shown that certain sections of the small part of the corridor, chosen for this phase of the study, are more susceptible than others to traffic congestion on a routine basis. The reasons for this need to be explored. Also it seems clear that, as the study scope is broadened, there will be other such areas identified.
3. **Maximum roadway capacity:** We now have quantitative data as to the typical maximum speeds on various sections of the L.I.E., and the maximum flow rate of cars both absolutely and with respect to speed. This information is valuable to highway designers, traffic engineers, and traffic managers.
4. **The entrance and exit flow** are presently estimated through statistical methods. Mechanistic methods which account for inlet (origination) and outlet (destination) densities can improve the performance of the statistical forecast.
5. **The single loop equation for speed predictor** was discussed in Section 5. Surprisingly, this equation was a non-linear function of the ratio of flow over density (linearity was previously expected). Mechanistic models have been explored to determine the cause of this non-linear behavior and improve the performance of the predictor.
6. **The more efficient numerical algorithms to solve the mechanistic continuity equation (Kernel),** have been explored in this report. The incorporation of these novel techniques, in the second phase of this project, will further improve the performance of the ATOP code.
7. **The use of historical anomalous signatures generated by HTDB,** through an intelligent classification scheme, will substantially improve the performance of incident detection and correction.

In addition to the above areas, which have already been explored and planned for incorporation into the next version of the ATOP code, several future applications have also been identified.

1. **HOV lane analysis techniques:** The data analysis techniques currently in use for the project are easily adaptable for measuring the effectiveness of the new HOV lanes. They can compare the traffic flow prior to the opening of the HOV lanes to the flow now with the lanes in operation, under a variety of times and conditions.

2. Variable message sign effectiveness measures: We have some initial ideas about methods to quantitatively measure and analyze the effectiveness of the variable message signs in use on the INFORM corridor. These analyses will allow the INFORM center to fine tune their strategies for traffic management and to improve the overall traffic flow through the use of optimized message sign use.
3. Automated interface for HTDB: Through an integrated automation of HTDB and ATOP, best estimate traffic forecast for a variety of weather conditions, constructions, and incidents would be possible. This will allow optimization of construction planning, emergency re-routing procedures, and response to anticipated congestion patterns.

10. REFERENCES

1. The National Program Plan For Intelligent Vehicle-Highway Systems (IVHS) Volume 1, Draft May 1994.
2. Letter to Nick Samios, the Director of Brookhaven National Laboratories (BNL) from the NYS DOT Commissioner, John C. Egan.
3. Santiago, A.J., Chen, H., "Traffic Modeling Software for IVHS Applications," Proceedings of the 1990 Winter Simulation Conference, O. Balci, R.P. Sadowski, R.E. Nance, editors, 1990.
4. New York State Department of Transportation. "New York Moves. Through Intelligent Vehicle Highway Systems." Draft, no date.
5. IVHS AMERICA "Federal IVHS Program Recommendations for Fiscal Years 1994 and 1995. IVHS AMERICA, Washington D.C., October 15, 1992.
6. New York State Department of Transportation. "INFORM - The Traffic Information System For Long Island Drivers." Pamphlet, no date.
7. New York State Department of Transportation. "Long Island INFORM System. K (Visual Traffic Information Project)." Draft, no date.
8. Barancik, J., Kramer, C., and Thode, H., Private Communication.
9. Box, G.E.P., and Tiao, G.C., "Bayesian Inference in Statistical Analysis," John Wiley and Sons, New York, NY, 1973.
10. May, Adolf D., "Traffic Flow Fundamentals," Prentice Hall, Englewood Cliffs, N.J. 1990
11. Pankratz, A., "Forecasting with Dynamic Regression Models," John Wiley and Sons, New York, NY, 1991.
12. KLD Associates, Inc., "Development and Testing of INTRAS (Integrated Traffic Simulations), A Microscopic Freeway Simulation Model, FHWA-RD-80-106, 1980.
13. Papageorgiou, Markos, and References therein "Concise Encyclopedia of Traffic and Transportation Systems," Pergamon Press, Oxford, 1991.
14. Gazis, D.C., Herman, R., and Potts, R.B., "Car-Following Theory of Steady State Flow," Operations Research, Vol. 7, No. 4, 1959, Pages 499-505.
15. 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, pages 545-567.
16. Herman, R., Montroll, W., Potts, R., and Rothery, R.W., "Traffic Dynamics: Analysis of Stability in Car-Following," Operations Research, Vol 1., No. 7, 1959, pages 86-106.

17. Roach, P.J., and References therein. "Computational Fluid Dynamics," Hermosa Publishers, Albuquerque, N.M., 1982.
18. Papageorgiou, M., Blosseville, J., Hadji-Salem, H., "Modeling and Real-Time Control of Traffic Flow on the Southern Part of Boulevard Peripheral in Paris," Transp. Res. A, Vol. 24A, pp.345-359 (1990).
19. Lax, P., and Wendroff, B., "Systems of Conservation Laws," Communications on Pure and Applied Mathematics 13, 1960, Pages 217-237.
20. Cremer, M., and May, A.D., "An Extended Traffic Flow Model for Inner Urban Freeways," IFAC Control in Transport Systems," p. 239 (1986), Vienna, Austria.
21. Michalopoulos, P., Lin, J. and Beskos, D., "Integrated modeling and Numerical Treatment of Freeway Flow," Applied Math. Modeling, Vol. 11, pp. 447-457 (1987).
22. Michalopoulos, P., Beskos, D., and Lin, J., "Analysis of Integrated Traffic Flow by Finite Difference Methods," Transp. Res. B. Vol. 18B, pp.409-421 (1984).
23. Lighthill, M.J., and Whitham, G.B., "On Kinematic Waves II: A Theory of Traffic Flow on Long Crowded Roads," Proc. Roy., Soc., A229, 1955, pages 317-245.
24. Centry, R.A., Martin, R.E., and Daly, J.B., "An Eulerian Difference Method for Understanding Compressible Flow Problems," Journal of Computerized Physics, Vol. 1, pp. 87-118 (1966).
25. Mughabghab, S., Azarm, M.A., and Stock, D., "Macroscopic Traffic Modeling with the Finite Difference Method," Brookhaven National Laboratory Report, BNL-61408, 1995.

APPENDIX A: DATA MISSING FROM INFORM'S ARCHIVAL RECORDS

Large amounts of data are missing from INFORM's archival records of traffic flow and occupancy. In the region studied for this project, the missing data was concentrated on the ramps of the Northern State Parkway (NSP), the Sagtikos Parkway (SAG), and the Connecting/Distributing (C/D) roads. Table A.1 shows the percentage of missing data from the roads under study. As shown in this table, absolutely all of the data was missing from the ramps of the NSP, SAG, and C/D roads. Fifty percent of the data was missing from the mainline detectors of the SAG, and two thirds of the data was missing from the C/D roads.

**Table A.1 Data Missing From the INFORM Archives
as a Function of Roadway**

Roadway	Percent Missing	
	Main Line	Ramp
Long Island Expressway	24.2%	14.9%
Northern State Parkway	19.4%	100%
Sagtikos Parkway	48.5%	100%
Connecting/Distributing Roads	66.6%	100%

APPENDIX B: SINGLE LOOP SPEED ESTIMATION TECHNIQUE USED BY THE INFORM SYSTEM

The two memos contained in this section contain the computer code which is used by INFORM to determine traffic speed and vehicle length. These memos were provided to BNL by Mr. Ray Schemel of the New York State INFORM System.

MEMO TO: MR MUGABAB
BROOKHAVEN NATIONAL LABORATORY
BLDG 130
FROM: RAY SCHIEMEL
N.Y.S. INFORM
DATE: 06/10/94
SUBJECT: SPEED CALCULATION REQUEST

THE FOLLOWING FORTRAN CODE SEEMS TO CALCULATE SPEED FOR THE ENTIRE SYSTEM.
IT'S RESIDENT IN THE MINUTE TASK WHICH EXECUTES ONCE A MINUTE.
SPEED TRAPS ARE LOCATED APX EVERY 10 ZONES.

```
TEMPR3= ((LVBAR(LZONE)+LLBAR(IZON)*CNTMZ(IZON)*3600.)/(OCCMZ(IZON)*88.)  
IF(SPEED(IZON).LT.TEMPR3+.05.AND.SPEED(IZON).GT.TEMPR3-.05) THEN  
    TEMPR3=SPEED(IZON)  
ENDIF  
SPEED(IZON)=SPEED(IZON)+FILTER(7)*(TEMPR3-SPEED(IZON))
```

TEMPR3: CALCULATED SPEED THIS MINUTE
! LVBAR: AVERAGE VEHICLE LENGTH FOR THIS ZONE
IZON: ZONE NUMBER (COMPUTER INTERNAL)
LZONE: ZONE NUMBER WHERE SPEED TRAP IS LOCATED FOR THIS ZONE
LLBAR: LOOP LENGTH
CNTMZ: VOLUME COUNT
OCCMZ: OCCUPANCY (UNITS=1/60th SEC)
SPEED: SMOOTHED SPEED
FILTER(7): CONSTANT 0.25

MEMO TO: MR MUGABAB
BROOKHAVEN NATIONAL LABORATORY.
BLDG 130
FROM: RAY SCHIEMEL
N.Y.S. INFORM
DATE: 06/16/94
SUBJECT: VEHICLE LENGTH CALCULATION REQUEST

THE FOLLOWING FORTRAN CODE CALCULATES VEHICLE LENGTH.
IT EXECUTES ONCE A MINUTE.

```
TEMPR = (LS*CNITZ(IZON))/(PASSTZ(IZON)/60.)  
TEMPR1 = TEMPR*(OCCMZ(IZON)/(CNITZ(IZON)*60))-LLBAR(IZON)  
  
IF(TEMPR1 .LT. 13) TEMPR1 = 13  
IF(TEMPR1 .GT. 50) TEMPR1 = 50  
  
IF(LVBAR(IZON) .LT. TEMPR1+.05 .AND. LVBAR(IZON) .GT. TEMPR1-.05) THEN  
    TEMPR1 = LVBAR(IZON)  
ENDIF  
  
LVBAR(IZON) = LVBAR(IZON) + .005*(TEMPR1-LVBAR(IZON))
```

LS: LOOP SEPARATION
CNITZ: ACCUMULATED COUNT FOR TRAP
IZON: ZONE NUMBER (COMPUTER INTERNAL)
PASSTZ: ACCUMULATED TRAVEL TIME FOR TRAP
OCCMZ: ACCUMULATED PRESENCE (OCCUPANCY) FOR THE MAINLINE
LLBAR: LOOP LENGTH
LVBAR: VEHICLE LENGTH

APPENDIX C: ADDITIONAL KERNEL PREDICTIONS OF OCCUPANCY AND FLOW

Additional tests of the Lax method are included in this appendix. The partial differential equation determines traffic densities or occupancies, K' , by Equation 5.11 which are converted to flows by Equation 5.2 and 5.23. The sensitivity of the predictions on the assumed value of the vehicle length, L_v , is explored.

Figure C.1 A comparison between predicted and measured INFORM occupancies at 17:15 on May 24, 1993. The predictions are based on the Lax method.

Figure C.2 A comparison between predicted and measured INFORM occupancies at 17:30 on May 24, 1993. The predictions are based on the Lax method.

Figure C.3 A comparison between predicted and measured INFORM occupancies at 17:45 on May 24, 1993. The predictions are based on the Lax method.

Figure C.4 A comparison between predicted and measured INFORM occupancies at 18:00 on May 24, 1993. The predictions are based on the Lax method.

Figure C.5 A comparison between predicted and measured INFORM occupancies at 18:15 on May 24, 1993. The predictions are based on the Lax method.

Figure C.6 A comparison between predicted and measured INFORM occupancies at 18:30 on May 24, 1993. The predictions are based on the Lax method.

Figure C.7 A comparison between predicted and measured INFORM flows at 15:45 on May 24, 1993. The predictions are based on the Lax method.

Figure C.8 A comparison between predicted and measured INFORM flows at 16:00 on May 24, 1993. The predictions are based on the Lax method.

Figure C.9 A comparison between predicted and measured INFORM flows at 16:15 on May 24, 1993. The predictions are based on the Lax method.

Figure C.10 A comparison between predicted and measured INFORM flows at 16:30 on May 24, 1993. The predictions are based on the Lax method.

Figure C.11 A comparison between predicted and measured INFORM flows at 16:45 on May 24, 1993. The predictions are based on the Lax method.

Figure C.12 A comparison between predicted and measured INFORM flows at 17:00 on May 24, 1993. The predictions are based on the Lax method.

Figure C.13 A comparison between predicted and measured INFORM flows at 17:15 on May 24, 1993. The predictions are based on the Lax method.

Figure C.14 A comparison between predicted and measured INFORM flows at 17:30 on May 24, 1993. The predictions are based on the Lax method.

Figure C.15 A comparison between predicted and measured INFORM flows at 17:45 on May 24, 1993. The predictions are based on the Lax method.

Figure C.16 A comparison between predicted and measured INFORM flows at 18:00 on May 24, 1993. The predictions are based on the Lax method.

Figure C.17 A comparison between predicted and measured INFORM flows at 18:15 on May 24, 1993. The predictions are based on the Lax method.

Figure C.18 A comparison between predicted and measured INFORM flows at 18:30 on May 24, 1993. The predictions are based on the Lax method.