

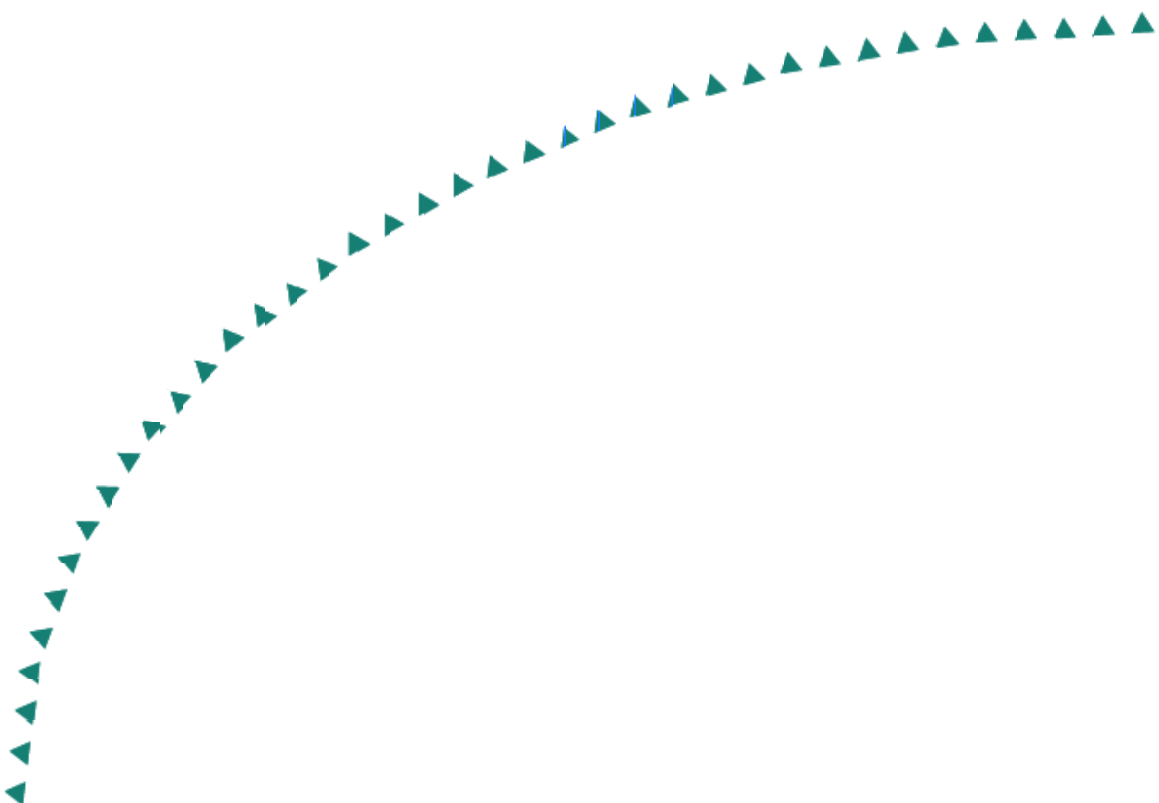
2003-06

Final Report

Employment of the Traffic
Management Laboratory (TRAMLAB) for
Evaluating Ramp Control Strategies
in the Twin Cities



Research



Technical Report Documentation Page

1. Report No. MN/RC 2003-06	2.	3. Recipients Accession No.	
4. Title and Subtitle EMPLOYMENT OF THE TRAFFIC MANAGEMENT LABORATORY (TRAMLAB) FOR EVALUATING RAMP CONTROL STRATEGIES IN THE TWIN CITIES		5. Report Date June 2002	
		6.	
7. Author(s) P. Michalopoulos, J. Hourdakis, K. Muralidhar, A. Sekhar, and V. K. Subramaniam		8. Performing Organization Report No.	
9. Performing Organization Name and Address University of Minnesota Civil Engineering Department 500 Pillsbury Dr SE, Minneapolis MN 55455		10. Project/Task/Work Unit No.	
		11. Contract (C) or Grant (G) No. (c) 74708 (wo) 164	
12. Sponsoring Organization Name and Address Minnesota Department of Transportation Office of Research Services 395 John Ireland Boulevard Mail Stop 330 St. Paul, Minnesota 55155		13. Type of Report and Period Covered Final Report 2002	
		14. Sponsoring Agency Code	
15. Supplementary Notes There is also a 24-page summary report which you may obtain by visiting the Mn/DOT web site at www.dot.state.mn.us , click on "Search Products," and type in 2003-06S.			
16. Abstract (Limit: 200 words) As freeway traffic congestion spreads ramp metering is implemented to address the problem. However, recently there is increasing opposition to freeway ramp control caused by excessive ramp delays. The objective of this research is to employ a recently developed tool called Traffic Management Laboratory (TRAMLAB) for assessing the effectiveness of Mn/DOT's control strategy in three Twin Cities freeway sections totaling approximately 65 miles. The feasibility of a corridor simulation will be followed by the selection and preliminary model development of the combination of an arterial and a freeway in the Twin Cities. As a result of this testing, TRAMLAB will evolve into an effective tool for developing control strategies that could reduce ramp delays without excessively increasing freeway congestion. Finally, a new traffic management concept for early detection of incident prone traffic conditions will be developed and integrated to traffic management through Ramp Metering and Variable Message Signs in order to smooth flow and prevent (to the extend possible) incident occurrence, thereby further reducing delays and improving safety. Even though this proposal focuses on evaluating ramp metering and implementing a concept recently developed in a current project, we also address the more general issue of research continuity and suggest a strategic partnership with MnDOT.			
17. Document Analysis/Descriptors Traffic Simulation Traffic Control Corridor Control Ramp Metering		18. Availability Statement No restrictions. Document available from: National Technical Information Services, Springfield, Virginia 22161	
19. Security Class (this report) Unclassified		20. Security Class (this page) Unclassified	
		21. No. of Pages 319	22. Price

**Employment of the Traffic Management Laboratory
(TRAMLAB) for Evaluating Ramp Control Strategies in the
Twin Cities**

Final Report

Panos Michalopoulos
John Hourdakis
Koka Muralidhar
Adarsh Sekhar
Vijay Konduru Subramaniam
Department of Civil Engineering
University of Minnesota

June 2002

Published by
Minnesota Department of Transportation
Office of Research Services
Mail Stop 330
395 John Ireland Boulevard
St. Paul, Minnesota 55155-1899

This report represents the results of research conducted by the authors and does not necessarily represent the view or policy of the Minnesota Department of Transportation and/or the Center for Transportation Studies. This report does not contain a standard or specified technique.

Acknowledgments

The University team wishes to thank the Mn/DOT Metro division and specifically Mr. Rich Lau, Mr. James Aswegan, Mr. Frank Lilja, Mr. David Berg and Mrs. Linda Taylor for their valuable help throughout the project. We wish to thank the people at the Twin Cities Metro Council and especially Mr. Mark Fillipi for their effort and assistance. A lot more people from Hennepin county, City of Bloomington, and other departments of the university have contributed into this project and we wish to thank them all.

1	INTRODUCTION AND PROBLEM STATEMENT	1
2	BACKGROUND	5
2.1	RAMP METERING IN THE UNITED STATES	5
2.1.1	Control Algorithms used in Ramp Metering Strategies	6
2.1.2	Minnesota Ramp Metering Strategy: An integrated adaptive ramp control algorithm	10
2.1.3	Local access ramp metering case study	14
2.1.4	Freeway-to-freeway ramp metering case studies	14
2.1.5	TH169 case study	14
2.2	SIMULATOR REVIEW	17
2.2.1	Models Currently in Use.....	22
2.2.2	Modeling in AIMSUN2	23
2.3	IMPROVING THE AIMSUN RAMP EXITING MODEL	34
3	SIMULATION ENHANCEMENTS	38
4	DESIGN OF EXPERIMENT	44
4.1	INTRODUCTION.....	46
4.2	SITE SELECTION.....	47
4.2.1	Site Selection Criteria	48
4.3	DATA COLLECTION.....	49
4.3.1	Test Sites Selected	50
4.3.2	Data Collection	58
5	CALIBRATION & VALIDATION.....	82
5.1	INTRODUCTION.....	82
5.2	STATISTICAL MEASURES USED IN CALIBRATION	83
5.3	CALIBRATION FOR NON-PEAK PERIOD FLOWS	90
5.4	PEAK-PERIOD CALIBRATION WITHOUT METERING	91
5.4.1	Volume Calibration.....	92
5.4.2	Speed Calibration.....	92
5.4.3	Occupancy Calibration	93
5.5	CALIBRATION WITH MN/DOT RAMP METERING.....	94
5.6	PROBLEMS ENCOUNTERED WITH THE TEST SITE ON TH 169	96
5.7	PROBLEMS ENCOUNTERED WITH THE TEST SITE ON I-94.....	98
6	RAMP METERING EVALUATION.....	104

6.1	INTRODUCTION.....	104
6.2	GENERAL OBSERVATIONS: FREEWAY AND DEMAND CHARACTERISTICS.....	106
6.3	OBSERVATIONS FROM THE EXPERIMENTS.....	107
6.4	EXTRACTION OF MOES FROM AIMSUN2 OUTPUT.....	111
6.4.1	Quarter Mile Statistics.....	115
6.4.2	Individual Ramp MOEs.....	116
6.5	RESULTS INTERPRETATION.....	117
6.5.1	Entire simulation period.....	117
6.6	METERING PERIOD ONLY.....	125
6.7	INDIVIDUAL RAMP RESULTS.....	132
6.8	CONCLUSIONS.....	139
7	FEASIBILITY OF CORRIDOR WIDE RAMP METERING EVALUATION.....	141
7.1	NEED FOR THE STUDY.....	141
7.2	STUDY OBJECTIVES.....	141
7.3	PROCESS OVERVIEW.....	142
8	CORRIDOR SITE SELECTION.....	143
8.1	SITE SELECTION CRITERIA.....	143
8.2	TEST SITE DESCRIPTION.....	144
8.2.1	ITS Deployment in the ICTM Corridor.....	144
8.2.2	Corridor Boundaries:.....	145
8.2.3	Freeways in the ICTM Corridor.....	146
8.2.4	Principal Arterials in the ICTM Corridor.....	146
8.2.5	Geometric Properties of the Freeways in the ICTM Corridor.....	148
9	CORRIDOR DATA COLLECTION.....	150
9.1	DATA NEEDS.....	150
9.1.1	Geometric Data.....	150
9.1.2	Traffic Data.....	152
9.1.3	Vehicle and Driver Characteristics.....	153
9.1.4	Traffic Control Data.....	153
9.2	DATA AVAILABLE.....	154
9.2.1	Geometric Data Available.....	155
9.2.2	Traffic Data Available.....	159
9.2.3	Traffic Control Data Available.....	163
9.3	DATA AVAILABILITY ISSUES.....	163
9.3.1	Summary of Data Needs & Availability.....	163

9.3.2	Effects of data unavailability	165
9.4	CONCLUSIONS	165
10	CALIBRATION ISSUES IN CORRIDOR SIMULATION.....	166
10.1	CALIBRATION METHODOLOGY FOR CORRIDOR SIMULATION	166
10.1.1	Calibration Methodology for the Turning Percentages Mode.....	169
10.1.2	Calibration Issues and Parameters in the Route Based Mode	169
10.2	IMPLEMENTATION OF CALIBRATION.....	175
10.2.1	Volume-based Calibration in the Freeway network.....	179
10.2.2	Speed-based Calibration in the ICTM Corridor Freeway network	180
10.2.3	Occupancy adjustment for the ICTM Corridor Freeway system	182
10.3	CONCLUSIONS	183
11	ESTIMATION OF O/D MATRIXES FOR CORRIDOR SIMULATION	184
11.1	DEFINITION OF TERMS:.....	184
11.2	TRAVERSAL O/D ESTIMATION METHODOLOGY:.....	187
11.3	ISSUES ENCOUNTERED IN IMPLEMENTATION OF TRAVERSAL OD ESTIMATION METHODOLOGY FOR THE ICTM CORRIDOR	190
11.4	ALTERNATIVE OD ESTIMATION METHODOLOGIES	190
11.4.1	ICTM Corridor Zones	191
11.4.2	Metro-council based method.....	191
11.4.3	Problems encountered in the Implementation of the Metro-council based Methodology.....	193
11.4.4	AADT Volumes method	194
11.5	EVALUATION EXPERIMENT	197
11.6	CONCLUSIONS	197
12	DETECTION OF INCIDENT PRONE CONDITIONS.....	ERROR! BOOKMARK NOT DEFINED.
12.1	PROBLEM STATEMENT	199
12.2	RESEARCH OBJECTIVES.....	201
13	DATA COLLECTION FOR DETECTION OF IPC.....	202
13.1	SITE SELECTION CRITERIA	202
13.2	TEST SITE DESCRIPTION.....	203
13.3	DATA COLLECTION	207
14	METHODOLOGY.....	212
14.1	BRIEF DESCRIPTION OF METHODOLOGY	212
14.2	TRAFFIC MEASUREMENTS CONSIDERED.....	215
14.2.1	Time Headway	216

14.2.2	Space Headway	217
14.2.3	Quality of Flow	221
14.2.4	Speed Variance	225
14.2.5	Acceleration Noise	226
14.2.6	Traffic Pressure	228
14.3	ALGORITHM EFFECTIVENESS:	230
15	UTC DETECTION ALGORITHM DEVELOPMENT.....	232
15.1	SELECTION OF TRAFFIC MEASUREMENTS FOR ALGORITHM DEVELOPMENT	232
15.2	DEVELOPMENT OF ALGORITHMS.....	237
15.2.1	Space Headway Algorithm	237
15.2.2	Traffic Pressure Algorithm	244
15.2.3	Quality of Flow Index Algorithm	249
15.2.4	Heuristic Integrated Algorithm	255
15.2.5	Neural Network Integrated Algorithm	261
16	UTC DETECTION ALGORITHM IMPLEMENTATION.....	282
17	UTC DETECTION IMPLEMENTATION RESULTS	287
17.1	HEADWAY ALGORITHM TEST RESULTS.....	287
17.2	TRAFFIC PRESSURE TEST RESULTS	289
17.3	QUALITY OF FLOW INDEX ALGORITHM TEST RESULTS.....	292
17.4	COMPARISON OF INTEGRATED ALGORITHM TEST RESULTS	295
18	INCIDENT PRONE CONDITIONS: CONCLUSIONS.....	ERROR! BOOKMARK NOT DEFINED.
19	ASSISTANCE TO MN/DOT IN EXPANDING SIMULATION USE	300

TABLE OF TABLES

TABLE 2.1: MAJOR RAMP METER INSTALLATIONS (MORE THAN 50 RAMP METERS) IN THE U.S.6

TABLE 2.2: MINOR RAMP METER INSTALLATIONS (LESS THAN 50 METERS) IN THE U.S.7

TABLE 2.3: MODEL COMPARISONS BASED ON QUANTITATIVE AND QUALITATIVE TESTS.....17

TABLE 4.1. GEOMETRIC PROPERTIES OF TEST SITES53

TABLE 4.2: VEHICLE CLASSIFICATION63

TABLE 4.3. DATA COLLECTION DATES AND DISCARDED DATA69

TABLE 4.4. CAR CHARACTERISTICS74

TABLE 4.5. TRUCK CHARACTERISTICS74

TABLE 4.6. SEMI-TRAILER CHARACTERISTICS75

TABLE 4.7. CAR FOLLOWING AND OVERTAKING MANEUVER MODEL PARAMETERS76

TABLE 4.8. FUEL CONSUMPTION FORMULAE FOR DIFFERENT VEHICLE STATES77

TABLE 4.9. FUEL CONSUMPTION PARAMETERS USED79

TABLE 4.10. POLLUTION EMISSION RATES FOR CARS.....80

TABLE 4.11. POLLUTION EMISSION RATES FOR SEMI TRAILERS81

TABLE 4.12. POLLUTION EMISSION RATES FOR TRUCKS81

TABLE 6.1: SECTION SIMULATION OUTPUT MEASUREMENTS109

TABLE 6.2. SYSTEM MOES112

TABLE 6.3. TH169 NB, ENTIRE CYCLE RESULTS (VS. METERED PERIOD ONLY).....118

TABLE 6.4. TH169 NB, ENTIRE CYCLE RESULTS (VS. METERED PERIOD ONLY).....119

TABLE 6.5. TH169 NB, ENTIRE CYCLE RESULTS (VS. METERED PERIOD ONLY).....120

TABLE 6.6. I-94EB, ENTIRE CYCLE RESULTS (VS. METERED PERIOD ONLY).....121

TABLE 6.7. I-94 EB, ENTIRE CYCLE RESULTS (VS. METERED PERIOD ONLY).....122

TABLE 6.8. I-94 EB, ENTIRE CYCLE RESULTS (VS. METERED PERIOD ONLY).....123

TABLE 6.9. TH-169 NB, METERED PERIOD RESULTS126

TABLE 6.10. TH-169 NB, METERED PERIOD RESULTS127

TABLE 6.11. TH-169 NB, METERED PERIOD RESULTS128

TABLE 6.12. I-94 EB, METERED PERIOD RESULTS129

TABLE 6.13. I-94 EB, METERED PERIOD RESULTS130

TABLE 6.14. I-94 EB, METERED PERIOD RESULTS131

TABLE 6.15. TH-169 NB, METERED PERIOD ONLY, METERED RAMPS.....133

TABLE 6.16. TH-169 NB, METERED PERIOD ONLY, METERED RAMPS.....134

TABLE 6.17. TH-169 NB, METERED PERIOD ONLY, METERED RAMPS.....135

TABLE 6.18. I-94 EB, METERED PERIOD ONLY, METERED RAMPS136

TABLE 6.19. I-94 EB, METERED PERIOD ONLY, METERED RAMPS137

TABLE 6.20. I-94 EB, METERED PERIOD ONLY, METERED RAMPS138

TABLE 8.1. INTERSTATE 494 (I-494):148

TABLE 8.2. TRUNK HIGHWAY 77 (TH77):	148
TABLE 8.3. INTERSTATE 35W (I-35W):	149
TABLE 8.4. TRUNK HIGHWAY 100 (TH100):	149
TABLE 5.13. AVERAGE STATISTICAL MEASURES FOR MAINLINE STATION VOLUMES	96
TABLE 9.1 PRELIMINARY VEHICLE CHARACTERISTICS FOR CARS	161
TABLE 9.2 PRELIMINARY VEHICLE CHARACTERISTICS FOR TRUCKS	161
TABLE 9.3 PRELIMINARY VEHICLE CHARACTERISTICS FOR SEMI-TRAILERS	162
TABLE 9.4: PRELIMINARY CAR FOLLOWING MODEL PARAMETERS	162
TABLE 9.5: PRELIMINARY LANE CHANGING MODEL PARAMETERS	162
TABLE 9.6 GEOMETRIC DATA:	163
TABLE 9.7 TRAFFIC DATA:	164
TABLE 9.8 TRAFFIC CONTROL DATA:	164
TABLE 10.1 MALFUNCTIONING FREEWAY DETECTOR STATIONS	176
TABLE 10.2 CALIBRATED VEHICLE CHARACTERISTICS FOR CARS.....	177
TABLE 10.3 CALIBRATED VEHICLE CHARACTERISTICS FOR TRUCKS.....	177
TABLE 10.4 CALIBRATED VEHICLE CHARACTERISTICS FOR SEMI-TRAILERS	178
TABLE 10.5 FINAL CALIBRATED VALUES FOR THE CAR-FOLLOWING MODEL PARAMETERS	178
TABLE 10.6 FINAL CALIBRATED VALUES FOR THE LANE-CHANGING MODEL PARAMETERS	178
TABLE 10.7 AVERAGE STATISTICAL MEASURES FOR MAINLINE STATION VOLUMES	179
TABLE 10.8 AVERAGE STATISTICAL MEASURES FOR MAINLINE STATION OCCUPANCIES	183
TABLE 13.1: SPEEDS AND VOLUMES AT MVS LOCATION NEAR PENN AVENUE.....	207
TABLE 13.2: SUMMARY OF ALGORITHM DEVELOPMENT DATA COLLECTED	210
TABLE 13.3: SUMMARY OF ALGORITHM TESTING DATA COLLECTED.....	211
TABLE 14.1: CIRCUMSTANCES UNDER WHICH REAR END COLLISIONS MOST COMMONLY OCCUR	214
TABLE 15.1: HEADWAY ALGORITHM RESULTS	243
TABLE 15.2: TRAFFIC PRESSURE ALGORITHM RESULTS	248
TABLE 15.3: QUALITY OF FLOW INDEX ALGORITHM RESULTS	254
TABLE 15.4: VALUES OF PARAMETERS USED IN HEURISTIC INTEGRATED ALGORITHM	258
TABLE 15.5: HEURISTIC INTEGRATED ALGORITHM.....	259
TABLE 15.5 CONTINUED: HEURISTIC INTEGRATED ALGORITHM.....	260
TABLE 15.6: CHARACTERISTICS OF DIFFERENT NEURAL NETWORKS (REPRODUCED FROM [8]).....	266
TABLE 15.8: NEURAL NETWORKS TRAINING AND TESTING DETAILS AND RESULTS	272
TABLE 15.8 : WEIGHTS FROM INPUT SIGNALS TO PROCESSING ELEMENTS IN 1ST LAYER.....	274
TABLE 15.9: BIASES ADDED TO WEIGHTED SUMS OF INPUT SIGNALS (PROCESSING ELEMENTS IN 1ST LAYER).....	274
TABLE 15.10: EIGHTS FROM PROCESSING ELEMENTS IN 1ST LAYER TO PROCESSING ELEMENTS IN 2ND LAYER	274
TABLE 15.11: BIASES ADDED TO WEIGHTED SUM OF OUTPUTS OF 1ST LAYER (DONE IN 2ND LAYER)	274
TABLE 15.12: WEIGHTS FROM INPUT SIGNALS TO PROCESSING ELEMENTS IN 1ST LAYER.....	277

TABLE 15.13: BIASES ADDED TO WEIGHTED SUMS OF INPUT SIGNALS (PROCESSING ELEMENTS IN 1ST LAYER).....	277
TABLE 15.14: EIGHTS FROM PROCESSING ELEMENTS IN 1ST LAYER TO PROCESSING ELEMENTS IN 2ND LAYER	277
TABLE 15.15: BIASES ADDED TO WEIGHTED SUM OF OUTPUTS OF 1ST LAYER (DONE IN 2ND LAYER)	277
TABLE 15.16: WEIGHTS FROM INPUT SIGNALS TO PROCESSING ELEMENTS IN 1ST LAYER.....	280
TABLE 15.17 : BIASES ADDED TO WEIGHTED SUMS OF INPUT SIGNALS (PROCESSING ELEMENTS IN 1 ST LAYER)	280
TABLE 15.18: EIGHTS FROM PROCESSING ELEMENTS IN 1ST LAYER TO PROCESSING ELEMENTS IN 2ND LAYER	280
TABLE 15.19: BIASES ADDED TO WEIGHTED SUM OF OUTPUTS OF 1ST LAYER (DONE IN 2ND LAYER)	280
TABLE 17.1: HEADWAY ALGORITHM RESULTS OBTAINED WITH SECOND DATA SET	288
TABLE 17.2: TRAFFIC PRESSURE ALGORITHM RESULTS OBTAINED WITH SECOND DATA SET.....	291
TABLE 17.3: QUALITY OF FLOW INDEX ALGORITHM RESULTS OBTAINED WITH SECOND DATA SET.....	293
TABLE 17.4: COMPARISON OF INTEGRATED ALGORITHM RESULTS (SECOND DATA SET)	296

TABLE OF FIGURES

FIGURE 2.1. CONCEPTUAL STRUCTURE OF GETRAM/AIMSUN2 18

FIGURE 2.2. AIMSUN2 SIMULATION PROCESS. 21

FIGURE 2.3. MEASURED AND SIMULATED SPEEDS 27

FIGURE 2.4. MEASURED AND SIMULATED RELATIVE DISTANCES LEADER-FOLLOWER 27

FIGURE 2.5. EMPIRICAL VERSUS SIMULATED FLOW-DENSITY CURVES..... 28

FIGURE 2.6. LANE CHANGING ZONES..... 30

FIGURE 2.7. VARIABILITY OF LANE CHANGING ZONES 31

FIGURE 2.8. EXAMPLE OF CONTINUOUS LINES 33

FIGURE 2.9. LANE CHANGING DECISION TREE 35

FIGURE 2.10. EXTENDING BACK ZONES 2 AND 3 37

FIGURE 3.1. GLOBAL VIEW OF THE AIMSUN2, CPI, RAMP CONTROL LOGIC AND G2..... 40

FIGURE 3.2. DETAILED INTERACTION OF CPI WITH THE SIMULATOR AND THE RAMP CONTROL ALGORITHM 42

FIGURE 4.1. TWIN CITIES FREEWAY NETWORK WITH THE THREE TEST SITES 52

FIGURE 4.2. TH-169 TEST SITE..... 54

FIGURE 4.3. I-94 TEST SITE..... 56

FIGURE 4.5. TYPICAL FREEWAY ENTRANCE RAMP AND MAINLINE SECTION 59

FIGURE 4.6. TYPICAL FREEWAY EXIT RAMP 60

FIGURE 4.7. TYPICAL FREEWAY LANE ADD SECTION 60

FIGURE 4.8. TYPICAL FREEWAY LANE DROP SECTION..... 61

FIGURE 4.9. TYPICAL FREEWAY WEAVING SECTION 61

FIGURE 4.10. SCREEN SHOT CAPTURE OF THE VIDEO TAKEN BY STATE PATROL 62

FIGURE 4.11. TYPICAL MAINLINE, ENTRANCE RAMP, AND RAMP DETECTORS 64

FIGURE 4.12. VISUAL INSPECTION OF DETECTOR DATA OF 3134, 3135 AND 3136..... 66

FIGURE 4.13. MAINLINE STATION DETECTORS AND EXIT RAMP DETECTOR..... 68

FIGURE 4.14. CORRELATION COEFFICIENT VARIATION WITH DIFFERENT CAR CATEGORIES AND TRUCKCAT3 AND SEMICAT3..... 73

FIGURE 4.15. CORRELATION COEFFICIENT VARIATION WITH DIFFERENT TRUCK CATEGORIES AND CARCAT2 AND SEMICAT3..... 73

FIGURE 4.16. CORRELATION COEFFICIENT VARIATION WITH DIFFERENT SEMI-TRAILERS AND CARCAT2 AND TRUCKCAT3 74

FIGURE 5.16. GRAPH SHOWING COMPARISON OF SIMULATED AND ACTUAL STATION VOLUMES 89

FIGURE 5.17. GRAPH SHOWING THE COMPARISON OF THE MERGE DETECTOR COUNTS 95

FIGURE 5.18. QUEUE DETECTOR AT I-394 COUNTING HOV VEHICLES ALSO 97

FIGURE 5.19: DETECTORS ON THE EXCELSIOR ENTRANCE RAMP..... 98

FIGURE 8.1. MAP OF THE ICTM CORRIDOR 146

FIGURE 9.1. MAP OF ICTM CORRIDOR 155

FIGURE 9.2. SCREENSHOT OF DIGITAL AERIAL PHOTOGRAPHS USED AS A BACKGROUND IN “TEDI”	159
FIGURE 9.2. ZONES IN THE ICTM CORRIDOR	160
FIGURE 10.1. FLOW CHART DESCRIBING THE CALIBRATION METHODOLOGY FOR CORRIDOR SIMULATION	168
FIGURE 10.3 ACTUAL SPEEDS ON EASTBOUND I-494	180
FIGURE 10.4 SIMULATED SPEEDS ON EASTBOUND I-494	181
FIGURE 10.5 ACTUAL SPEEDS ON WESTBOUND I-494	181
FIGURE 10.6 SIMULATED SPEEDS ON WESTBOUND I-494	182
FIGURE 11.2. CONCEPT OF TRAVERSAL MATRIX	186
FIGURE 11.3. FLOW-CHART DESCRIBING TRAVERSAL O/D ESTIMATION METHODOLOGY	189
FIGURE 11.4. ICTM CORRIDOR ZONES	191
FIGURE 11. 5. FLOW CHART DESCRIBING THE METRO-COUNCIL BASED METHODOLOGY	193
FIGURE 11.6. FLOW CHART DESCRIBING THE AADT VOLUMES METHODOLOGY	196
FIGURE 13.2: FIGURE SHOWING THE FIELD OF VIEW OF THE MACHINE VISION SENSOR	205
FIGURE 13.3: FIGURE SHOWING THE PLACEMENT OF THE MVS CAMERA	206
FIGURE 14.1A LEADING VEHICLE JUST STARTS TO BRAKE	219
FIGURE 14.1B DISTANCES NEEDED BY BOTH VEHICLES TO STOP	219
FIGURE 14.2: FLOW CHART SHOWING THE METHODOLOGY USED TO CALCULATE POINT QOF	224
FIGURE 15.1: EXAMPLE OF PATTERN OBSERVED BETWEEN SPACE HEADWAYS AND OCCURRENCES OF MANUAL ALARMS	234
FIGURE 15.2: EXAMPLE OF PATTERN OBSERVED BETWEEN HIGH VALUES OF TRAFFIC PRESSURE AND OCCURRENCE OF MANUAL ALARMS	235
FIGURE 15.3: EXAMPLE OF PATTERN EXISTING BETWEEN SHARP DECREASE IN VALUES OF QUALITY OF FLOW INDEX AND MANUAL ALARMS	237
FIGURE 15.4: FLOW CHART DESCRIBING THE HEADWAY ALGORITHM	239
FIGURE 15.5: SCHEMATIC REPRESENTATION OF “UNSAFE GROUP EXTENSION CHECK”	242
FIGURE 15.6: PERFORMANCE CURVES OBTAINED FOR HEADWAY ALGORITHM	244
FIGURE 15.7: FLOW CHART DESCRIBING THE TRAFFIC PRESSURE ALGORITHM	246
FIGURE 15.8: PERFORMANCE CURVES OBTAINED FOR TRAFFIC PRESSURE ALGORITHM	249
FIGURE 15.9: FLOW CHART DESCRIBING THE QUALITY OF FLOW INDEX ALGORITHM	250
FIGURE 15.10: MODIFICATION MADE TO QUALITY OF FLOW INDEX ALGORITHM FOR RIGHT AND MIDDLE LANES	252
FIGURE 15.11: PERFORMANCE CURVES OBTAINED FOR QUALITY OF FLOW INDEX ALGORITHM	253
FIGURE 15.12: FLOW CHART DESCRIBING THE HEURISTIC INTEGRATED ALGORITHM	256
FIGURE 15.13: PERFORMANCE CURVES OBTAINED FOR QUALITY OF FLOW INDEX ALGORITHM	261
FIGURE 15.14: REPRESENTATION OF A NEURAL NETWORK (PARTIALLY SHOWN)	263
FIGURE 15.15: TOPOLOGY OF A SIMPLE MULTI-LAYER FEED-FORWARD NETWORK	268
FIGURE 15.16: SCHEMATIC REPRESENTATION OF NNRL	275
FIGURE 15.17: TRAINING CURVE OBTAINED FOR NNRL	275

FIGURE 15.18: SCHEMATIC REPRESENTATION OF NNML	278
FIGURE 15.19: TRAINING CURVE OBTAINED FOR NNML	278
FIGURE 15.20: SCHEMATIC REPRESENTATION OF NNLL	281
FIGURE 15.21: TRAINING CURVE OBTAINED FOR NNLL	281
FIGURE 16.1: FLOW CHART DEPICTING THE FUNCTIONING OF THE PROGRAM	283
FIGURE 16.2: RAW MVS DATA	FIGURE 16.3: PROCESSED LEFT LANE DATA.....
	284
FIGURE 17.1: PERFORMANCE CURVES OBTAINED FOR HEADWAY ALG WITH THE SECOND DATA SET	289
FIGURE 17.2: PERF CURVES OBTAINED FOR TRAFFIC PRESSURE ALG WITH THE SECOND DATA SET	292
FIGURE 17.3: PERFORMANCE CURVES OBTAINED FOR QUALITY OF FLOW ALG WITH THE SECOND DATA SET	294

1 INTRODUCTION AND PROBLEM STATEMENT

Peak period travel demand has exceeded un-managed road capacity on most of the Twin Cities metropolitan area freeways for more than two decades. Freeway delays are expected to almost triple in the ten-year period between 1995 and 2005 [1].

Since the early 1970's, the Minnesota Department of Transportation (Mn/DOT) has developed, implemented, and continues to expand its Freeway Traffic Management System (FTMS). The overall goal of the FTMS is to optimize traffic flow in metro area freeway corridors by making efficient use of the available transportation facilities. Specifically, the objectives include maximizing the traffic volumes that freeway bottleneck locations can accommodate, minimizing the magnitude and duration of congestion, and minimizing accident rates among others [2].

A metered freeway with an automated, on-line traffic detection and response system can be expected to carry as much as 2,400 vphpl. A few on-line freeway segments in the Twin Cities area have consistently carried 2,700 vphpl, whereas un-metered freeways typically have a capacity of 1,800 to 1,900 vphpl. Similarly, it is common for congested freeways to be characterized by peak-hour speeds of 25 to 35 mph prior to activation of ramp metering systems. Fully automated ramp metering systems typically operate at 45 to 50 mph, the speed at which optimum capacity is achieved. Based on a study by Mn/DOT [2], the number of accidents on Twin Cities area metered freeways reduced from an average of 9,200 crashes per year in 1988 and 1989 to about 7,900 per year from 1991 to 1996. There are also other benefits like lower fuel consumption and lower emissions.

In spite of the tangible benefits that can accrue from ramp control, some of the public questions the effectiveness of ramp metering, through either press or their elected representatives. This is mainly because of the excessive ramp delays they experience during peak hours. According to a perception tracking study done in June 1999 for Mn/DOT by MarketLine Research, Inc. [3], the

opinion rating on a 10-point scale (1 being “Not at all effective” to 10 being “Highly effective”) by the public for ramp metering was 5.32, the lowest among twelve traffic management tools. More importantly, the rating was down significantly from 1998 and the number of people who said their opinion of ramp metering has become negative over the past year was up from the same period in 1998 (27% compared to 21%). This can partially be attributed to the fact that in its effort to maximize freeway throughput, FTMS as implemented in the Twin Cities did not consider ramps as part of the system as well as the effects of ramp metering on the entire corridor.

Ramp metering has not found much favor recently with the Legislature either. On March 20, 2000 the Minnesota State Senate passed Senator Dick Day’s amendment “to shut off all ramp meters in the metro on access ramps to freeways for a month in 2000, to monitor the effects of traffic flow without the artificial restrictions.” [4]

Hence, there is a need to develop a scientifically sound procedure to determine the effectiveness and validity of Mn/DOT’s current and future ramp control strategies (which includes quantifying the benefits of ramp metering) and prepare a platform to test any changes to the algorithm. The control algorithm, like many others developed in the United States, is empirical and fine-tuned over a period of time but its deviation from optimality is not known. Moreover, currently any changes to the algorithm have to be tested in “real life” before an effective decision can be made to implement them. This trial and error process is time consuming and risky because some of the changes are bound to be ineffective and thus might reduce the level of acceptance the public has of ramp metering.

Simulation is the most commonly accepted tool to shorten and minimize the risk of the process but this rarely occurs in practice. Some of the problems are that entering geometric and traffic data into the simulator is tedious, the output data is cumbersome and difficult to analyze, and there are some doubts about how well real conditions can be replicated in simulation. Furthermore, simulators are, at best, designed to implement only a particular control strategy. Thus, a flexible and uniform practical tool for selecting the best control strategy and optimizing it or developing and testing new concepts is currently lacking. As part of this research, one of the

objectives is to make simulation a viable tool and simplify its use. This implies automating as many tasks as possible so that the users can test a variety of scenarios.

This study is attempting to address these issues by:

- Developing an automated simulation tool for testing, and comparing (new) ramp control strategies. This required selection of a high performance simulator and for enabling emulation of any user specified control strategy among others.
- Emulating Mn/DOT's ramp control strategy by integrating it with the selected simulator.
- Employing the resulting simulation package to test Mn/DOT's control strategy at two test sites on Twin Cities freeways representing different functional designs, various traffic congestion levels, etc. Compare the results with and without control and to quantify the advantages and disadvantages of ramp control.
- Explore the feasibility of expanding the aforementioned analysis on a corridor (freeway and arterials).

This report is organized into eleven chapters. Chapter 2 starts with a general overview of ramp metering practices in the US following by details of Mn/DOT's ramp control algorithm and finally the microscopic simulator used in this research is reviewed pointing out its salient features and modeling technologies. In Chapter 3, the simulator enhancements that were carried out to make it a viable tool for this study are presented. Chapter 4 describes the overall experimental design that includes, the selection criteria used for choosing test sites, the detailed procedure used in collecting geometric, traffic and other data needed for simulation, and the various assumptions that were made in this study. Chapter 5 gives a detailed description of the calibration/validation methodology and results. Chapter 6 describes in detail the procedure used by the simulator to collect traffic measurements and the Measures Of Effectiveness (MOEs) used in this study. Chapter 7 presents the results of the study for each site and chapter 8 contains conclusions and recommendations.

Chapter 9 describes the results from the feasibility analysis of a corridor-wide evaluation of ramp metering through simulation. Chapter 10 contains the material and conclusions from the project results Technology Transfer that includes the Simulator training material and the feedback from

the training course contacted. Finally Chapter 11 contains the first phase on the attempt to detect Incident prone conditions on freeways.

2 BACKGROUND

Ramp metering is being currently used on over 2,000 miles of freeways in the United States with about 2,200 ramp meters deployed [5]. To be sure ramp control has been applied since the 1960s in Chicago, San Diego and Los Angeles areas. Success of these early applications led to the extension of ramp metering systems to 23 metropolitan areas by 1995, according to a 1995 update on ramp metering status in North America [6]. In this chapter a brief overview of ramp metering in the U.S. is presented. Specifically, the various types of ramp control strategies being used are described at a high level without going into the details. Many ramp-metering systems evaluation studies have been done and some of these are presented in this chapter. As the objective of this research is to assess the effectiveness of Mn/DOT's ramp control strategy, it is described in more detail.

2.1 Ramp Metering in the United States

The first ramp meter was installed in Chicago on the Eisenhower Expressway in 1963. This initial Chicago study employed a police officer, stationed on the entrance ramp, which stopped traffic and released vehicles at a rate determined from a pilot detection program. Since then more and more cities have been embracing ramp metering as a traffic management strategy. From 1989 to 1995, the number of operating meters in North America has increased from 1600 to over 2300 and they were being used in 23 metropolitan regions. At this time it was estimated that by the year 2000, 33 cities in the United States and Canada would have functional ramp meters.

Table 2.1 lists most of the metropolitan areas in the U.S. with extensive installations (more than 50 ramp meters) [5, 6, 7] whereas Table 2.2 presents some of the cities with smaller installations (less than 50 ramp meters).

Table 2.1: Major ramp meter installations (more than 50 ramp meters) in the U.S.

Metropolitan Area	City	First Installation	Number of Ramp Meters
Arizona	Phoenix	Early 1970s	65
California	Los Angeles	1968	1200
	Orange County	N/A	278
	San Diego	1968	250
	San Francisco	N/A	96
Illinois	Chicago	1963	113
Michigan	Detroit	1981	49
Minnesota	Minneapolis	1970	367
New York	Long Island	1989	75
Oregon	Portland	1981	58
Washington	Seattle	1993	126

2.1.1 Control Algorithms used in Ramp Metering Strategies

Although the ramp metering strategies employed by various DOTs in many of the above metropolitan areas are fine-tuned over a period of time and customized to meet their unique needs, they can be broadly classified into three different types:

Fixed Time Control: The ramp metering rates are pre-determined and vary according to the time of the day. These rates are calculated offline based on average traffic conditions expected at different times of the day. Fixed time metering is the simplest form of ramp metering to implement and provides benefits associated with accident reductions as only one vehicle is allowed to merge with the mainline traffic at a time, but is not very effective in regulating traffic volume as it doesn't consider real-time mainline traffic conditions

Table 2.2: Minor ramp meter installations (less than 50 meters) in the U.S.

Metropolitan Area	City	First Installation	Number of Meters
California	Fresno	1993	15
	Riverside	N/A	43
	Sacramento	1983	19
	San Bernardino	N/A	35
Colorado	Denver	1981	30
Georgia	Atlanta	1996	5
Ohio	Columbus	1973	6
Virginia	Northern Virginia	1985	26
Wisconsin	Milwaukee	1969	43

Local Traffic Responsive Control: This type of ramp metering strategies need some of traffic detection on the mainline and (maybe) on the ramps, and use this information to calculate new rates. The new rates can be either chosen from a set of pre-determined rates corresponding to different parameters thresholds or could be calculated in real-time. These strategies can respond to local disturbances in traffic and hence are more effective than fixed time control.

System-wide Control: The metering rates are computed based on traffic conditions system-wide rather than at a particular location. Similar to the local traffic-responsive control, the rates could be either pre-determined or calculated in real-time. These strategies are the most difficult to implement because they need sensors for traffic detection and transmission of this data to a central computing facility so that new rates can be calculated based on system-wide conditions but they are also more effective compared to the local traffic-responsive strategies as case studies in Denver [8] and Minneapolis [9] have shown.

According to Kwon [7], 16 of the 24 cities in North America employing ramp-metering strategies implement fixed time control. From the remaining, three (Chicago, Toronto and San Diego) use local traffic responsive strategies whereas Denver, Seattle, Los Angeles, Northern Virginia and Minneapolis implement traffic responsive coordinated control.

The Colorado DOT operates 30 ramp meters in Denver using a combination of local control, group coordination and queue override. There are queue, presence and merge detectors on the on-ramps. Each meter has predetermined turn-on and turn-off transition intervals, volume and occupancy thresholds and six metering rates [7]. Washington State DOT (WSDOT) is metering about 126 ramps in the Seattle metro area. Recently, WSDOT switched to a "Fuzzy" metering algorithm [10]. Until then, an integrated, traffic-responsive metering algorithm (bottleneck algorithm) was used [11].

A system-wide adaptive ramp-metering algorithm is being installed in the Los Angeles area freeway network to meter 1200 ramps. The algorithm employs empirically determined bottlenecks and associated saturation densities [7]. The Virginia DOT operates 26 centralized computer-controlled ramp meters near Washington D. C. Virginia has adopted a link-based, zone control algorithm in which a given freeway is divided into zones. Meters are turned on/off when the measured link volume exceeds/drops a predetermined threshold. Each ramp has a predetermined minimum and maximum rate [7]. Minneapolis uses a zone based ramp-metering algorithm [12] and is described in detail in Section 2.3.

Ramp Metering Benefits

Success and benefits of ramp metering have been documented in a number of reports [8, 9, 13, 14, 15, 16]. Benefits attributed to ramp metering in the literature include lowered travel times, increased freeway capacity and throughput, decreased accident rates, fuel consumption and vehicle emissions. Piotrowicz and Robinson surveyed eight ramp-metering systems in the U.S. [7] for the Federal Highway Administration. Most of the ramps metering evaluations were carried out in the eighties. Average highway speed increases of, up-to, 60% were reported after installation of ramp meters. Ramp metering reduces congestion by managing traffic demand, improving the efficiency of merging, and reducing accidents. Efficient merging results in fewer accidents, which typically cause major traffic congestion. Accidents were reduced 27%-50% as a result of ramp metering [7]. Improved traffic flow, particularly the reduction in stop-and-go traffic, decreases fuel consumption and vehicle emissions. The ramp metering benefits of some of the sites surveyed by Piotrowicz and Robinson for ramp metering benefits are presented below except those for the Twin Cities, which are presented separately.

Portland, Oregon

Average speed on I-5 NB increased from 26 to 44 km/h during the PM peak hour after metering

Average speed on I-5 SB increased from 64 to 69 km/h only, as the pre-metered conditions were less severe

43% reduction in peak period accidents with reduction in rear-end and side-sweep accidents

Fuel consumption, including additional consumption caused by ramp delay, was reduced by 2040 liters of gasoline per weekday

Seattle, Washington

Evaluation of 22 ramps (17 and 5 on I-5 SB and NB respectively) between 1981 and 1987 showed that mainline volumes during the peak periods increased 86% in NB and 62% in SB direction

Accident rate decreased by 39%

Travel time on a specific 11 km course reduced from 22 to 11.5 minutes

Denver, Colorado

Metering started in 1981 on a 4.7 km section of I-25. An 18-month study showed that during the peak period average speed increased by 57%, average travel time and accidents decreased by 37% and 5% respectively.

Another evaluation in 1988 showed that the measured speeds in 1988 were less than in the above evaluation (80 vs. 80 km/hr) but more than before meters were installed (69 km/hr)

A study of local traffic responsive metering vs. centralized control showed that for speeds below 90 km/hr, the latter is very effective with speed increases of 35% and vehicle travel hours reduction of 13%. Hence, this study stresses the importance of implementing system-wide traffic responsive ramp-metering strategies.

Detroit, Michigan

Speeds on I-94 (25 ramps were metered) increased 8% whereas the typical peak hour volume on three eastbound lanes increased from 5600 to 6400 vph

Accidents reduced by nearly 50%

Long Island, New York

INFORM (Information For Motorists) project covers a 64 km long by 8 km wide corridor which includes the Long Island Expressway, and the Northern State/Grand Central Parkway

A 1991 study showed that while throughput increased by only 2%, the average mainline speeds improved by 9%

At two separate bottleneck locations, the speeds improved from 53 to 84 and 53 to 89 km/hr

Accident frequency was reduced by 15%

Even though there are a number of ramp control strategies being used by various state transportation agencies, there are no uniform or standard evaluation criteria and the measures of effectiveness vary with the system objectives. Most U.S. evaluations are almost a decade or more old. Nevertheless most of the implemented ramp metering strategies achieved substantial system wide benefits. While it is reasonable to assume that numerous hurdles had to be overcome (especially in educating and convincing the public) and significant costs had to be incurred to install and operate ramp meters (building the infrastructure), they were not highlighted in the evaluations. It has been argued that many evaluations fail to fully analyze negative effects, such as the impacts of diversion onto surface networks, and queuing on ramps. Continuous traffic growth suggests that more rigorous evaluations are needed to conclusively assess ramp meter performance. Hopefully, the research being conducted in this study would be a positive step in that direction.

2.1.2 Minnesota Ramp Metering Strategy: An integrated adaptive ramp control algorithm

The Minnesota control strategy [11], active until October 2000, begins by dividing the freeway into zones. A zone is defined as a freeway section, traveling in one direction, and is typically three to six miles in length. The beginning or upstream end of a zone is usually a free-flow area not subject to high incident rates. The downstream end of a zone is a bottleneck, where the demand to capacity ratio is highest on that freeway section. Lane drop locations, high volume entrance ramps, and high volume weaving sections are typical bottleneck locations.

The zone control algorithm is built on the basic concept of balancing the volume of traffic entering the zone with that leaving the zone. All volumes of entering and exiting traffic are measured in real time every 30 seconds. When these total volumes are balanced, the density of traffic in the zone should remain within a narrow range. When the density of traffic in the zone is low, there is space available within the zone for additional entering traffic.

The metering zone conservation equation can be expressed as:

$$A + U + M + F = X + B + S$$

where,

A = Upstream mainline volume (veh/5 min), a measured variable

U = Σ (Unmetered entrance ramp volumes) (veh/5 min), a measured variable

M = Σ (Metered local access ramp volumes) (veh/5 min), a controlled variable

F = Σ (Metered freeway to freeway access ramp volumes) (veh/5 min), a controlled variable

X = Σ (Exit ramp volumes) (veh/5 min), a measured variable

B = Downstream bottleneck capacity volume (veh/5 min), a constant

S = Space available within the zone (veh), a computed volume based on occupancy of mainline detectors

Stated as the sum of metered ramp volumes, the equation becomes:

$$M + F = X + B + S - A - U$$

Any measured variation in $X + B + S - A - U$ is equaled by a controlled variation in $M + F$.

Each individual variable in the above equation has a target value (noted by t). The zone conservation equation written in the target volume form is: $M_t + F_t = X_t + B_t + S_t - A_t - U_t$. All variables, except S_t , are assigned a one-hour volume derived from the detector data. This volume is the median value from fifteen days, of the highest floating sixty-minute flow rate. The volumes for each variable are expressed as five-minute flow rates. The value of S_t is set to zero, indicating the target condition has no space available in the zone. When these target volumes are placed in the equation, an exact balance may not appear. For this reason, a minor adjustment to the incoming volume (A_t) is made to balance the equation. The values in this balanced equation are the target volumes.

The controlled variables $M + F$ are expressed as the products of the target volumes and a series of factors. The metering factors for local access ramps range from 0.5 to 1.5, and for freeway-to-freeway ramps, they range from 0.75 to 1.25. For the most restrictive rate,

$$M = 0.5 \times M_t \text{ and}$$

$$F = 0.75 \times F_t.$$

The narrower range for freeway-to-freeway ramps reflects the fact that these ramps carry much higher volumes and a smaller percentage change provides the desired numerical volume change. A factor of 0.75 is used as the most restrictive rate on freeways to also lessen the likelihood of queuing problems on the upstream freeway.

The variables in the equation are assigned real-time values using data collected from detectors placed on the mainline, entrance and exit ramps. Detector data is recorded every 30 seconds and the latest two to ten values (selectable by zone) are then expanded to a five-minute flow rate. Detector data for variables U (un-metered ramp volumes) and X (exit ramp volumes) are recorded once every five minutes and then used for the next five minutes. The five-minute value for B (usually 185 vehicles per lane) is normally a constant equal to $1/12$ of the highest recorded one-hour flow.

Values for the variable S are calculated using the difference between the capacity of the zone at an average density of 32 vehicles/lane-mile (selectable by zone) and the actual number of vehicles presently in the zone. The actual number of vehicles in the zone is calculated by first determining the average occupancy of each mainline detector in the zone. These values are added together and then multiplied by 1.1. The factor 1.1 is used because detectors spaced at $\frac{1}{2}$ mile with a "full zone" occupancy of 15 percent multiplied by 1.1 approximately equals target density of 32 vehicles/lane-mile. The actual number of vehicles in the zone is then subtracted from the full capacity of the zone. The remaining capacity, if any, is the space available (S) in the zone.

Each metered ramp is assigned six metering rates. On local access ramps, these rates over a five-minute time period would correspond to 1.5, 1.3, 1.1, 0.9, 0.7, and 0.5 times the target volume. On freeway-to-freeway ramps, rates over a five-minute period are 1.25, 1.15, 1.05, 0.95, 0.85,

and 0.75 times the target volume. The selection of which rate to use is then determined by a comparison of the measured variables ($X + B + S - A - U$) to a series of thresholds.

While volume control by zone is quite effective during a wide range of flow conditions, it relies on roadway capacities throughout the zone remaining undisturbed by incidents in the zone. To minimize the effects of an incident, which produces a temporary bottleneck at any location on a freeway, Mn/DOT uses occupancy control. Using detectors, occupancy is measured every 30 seconds across all lanes on the freeway mainline in one direction. These stations are located at intervals of 800 meters. The occupancy recorded at each station is compared to a series of thresholds (typically 17%, 18%, 23%, and 40%). These thresholds translate approximately to traffic densities of 26, 28, 35 and 62 vehicles per km assuming 5% trucks. The 17% would call for a rate of 3, 18% to rate 4, 23% to rate 5, and 40% to rate 6.

Each metered ramp is assigned stations up to 3 miles downstream to be used for occupancy control. Some stations may be outside the zone (beyond the zone bottleneck). The occupancy rate selection is calculated using the highest occupancy measured for one minute at any one station. If the latter is more restrictive than the volume control rate selected, then it is used for the next 30 seconds.

The AM and PM peak periods, usually from 6:00 A.M. to 9:00 A.M. and 2:30 P.M. to 7:00 P.M. respectively, used by Mn/DOT are defined to bracket the heavy work trip times on weekdays. Each zone is designated an AM zone, a PM zone, or both. Zones are metered when the expected flow rate at the bottleneck is equal to capacity for one or more hours during the peak period, but there are no specific ramp-metering start and stop times. The peak periods have a turn-on period (starting from 6:00 A.M. for the morning peak and 2:30 P.M. for the afternoon peak) during which ramp meters may be turned on individually or in groups by operators when the algorithm calls for a restrictive rate 5 or rate 6. Turn-on is not initiated when less restrictive rates are called for because typical flow early in the peak hours is characterized by dramatic swings in volume under open flow conditions. Once ramp metering has begun, the rate used will be variable between rate 1 and rate 6, which can be overridden by the operators, if necessary. The operators turn off ramps meters when the arrival rate of vehicles falls and the ramp empties which can be

observed in the form of short queues. Just before turn off, the ramp meters are put in the dump mode (rate 1), which allows faster queue dissipation (if any).

Effects of Mn/DOT's Ramp Metering

As described earlier, Mn/DOT's control strategy meters both local access ramps and freeway-to-freeway ramps. The results of some of earlier studies performed [6, 14, 16] to determine the effects of Mn/DOT's ramp-metering strategy are summarized below:

2.1.3 Local access ramp metering case study

An evaluation of an 8-km section of I-35E going through St. Paul downtown, in 1985, reported average peak hour speed increases of 16%, peak period volume increases of 25% and peak period accident rate reduction of 38% [6].

An evaluation of a 27-km segment of I-35W SB, in 1985, showed that average peak period mainline speeds increased from 55 to 74 km/hr, peak period volumes increased 32%, accident rate reduced 38% and peak period pollutants emissions reduced by just under 2 million kilograms per year [14].

2.1.4 Freeway-to-freeway ramp metering case studies

Mn/DOT meters 74 freeway-to-freeway ramps, the largest in the U.S.

One study of the ramp joining I-94 EB to SR65 SB showed that mainline volumes increased 17%, peak hour speeds increased by 29% and accidents reduced 21% though the wait time can reach 8 minutes with queues extending one-half kilometer [6].

2.1.5 TH169 case study

Changing the isolated ramp metering system on TH 169 SB to a dynamic metering system increased mainline speeds 20%, volumes at bottleneck locations increased by 13% though the average wait time increased from less than a minute to about 6 minutes during both AM and PM peak period [9].

Project Background

This project is part of a bigger program which aims to develop TRAMLAB as part of the ITS Laboratory at the University of Minnesota. Such an environment will contain state-of-the-art traffic simulation programs, and allow the development of viable, intelligent, and automated

traffic-flow simulation systems which can function as both operational and research tools. In the course of this program, six well-known freeway simulators were evaluated in 1996 [17]. These simulators were FREFLO [18], FRESIM [19], INTEGRATION [20,21], FREQ11 [22], KRONOS [23,24,25], and AIMSUN2 [26,27,28,29]. The six simulators were initially tested in order to get a better feel of the state-of-the-art in traffic simulation, benchmark their performance, and determine the feasibility of allowing easy selection of any simulator through a common user-interface. The simulators were divided into three groups based on their modeling approach, i.e., microscopic (2 simulators), macroscopic (3 simulators) and mesoscopic¹ (1 simulator). The first microscopic model tested was FRESIM. This is an improved version of the Federal Highway Administration's (FHWA) INTRAS model. The second microscopic model AIMSUN2 was developed at the University of Catalunya, Spain. The first macroscopic model, FREFLO, is part of the FHWA's CORFLO package while the second, FREQ11, is one of the earliest simulation models available. Mn/DOT is currently using the macroscopic model, KRONOS, developed at the University of Minnesota in 1983. The only mesoscopic model selected, INTEGRATION, was developed in Canada and at the time of the study it was widely used in ITS projects.

A systematic procedure was developed for testing the models both quantitatively and qualitatively [17]. Three quantitative tests corresponding to real life situations were performed for model accuracy in estimating the MOEs, and traffic volumes for which actual measurements were available. Six qualitative tests were performed to check the conservation and spill-back characteristics of each model using the same data sets for consistent evaluation. Input data requirements, capabilities and output relevance were compared during the evaluation, while difficulties associated with data entry, interpretation of the output, and calibration procedures were determined. A general overview of the test results is presented in Table 2.3.

During the tests, the suspected reasons preventing widespread use of simulation were confirmed. Specifically, all of the simulation models tested suffered from a lack of automation for

¹ A mesoscopic model considers platoons of vehicles as independent entities and traffic is emulated by following rules describing the interaction between platoons.

minimizing the data entry effort. Even for the simple test sites selected, entering the geometry, traffic volumes, and control information took anywhere from a few hours to a week to complete. In addition, none of the simulators had, at the time of testing, all of the features needed for emulating state of the art ATMS applications. Such features include user specified adaptive control schemes, origin-destination tracking, map-based geometry interfaces, automated demand data access, and capability to model all common geometric layouts.

After considerable evaluation of the options available, it was determined that, although desirable, cost and lack of uniformity made a generic user interface for all programs impractical if not impossible. Two simulators were selected for the project covered by this report, one macroscopic (KRONOS) and one microscopic (AIMSUN2), to be joined under the same input interface and database to demonstrate the project concepts and complete a working laboratory [30]. The programs performed very well comparatively in the initial tests and were selected because full access to the code was available as well as support from its developers. In addition, the programs were field-proven in many real-life projects and have been continuously enhanced and tested over the last 12 years.

Following model selection, the AIMSUN2 simulator was enhanced to communicate with external user-defined ATMS applications and, as a demonstration, the current control strategy developed by Mn/DOT over the last 28 years was interfaced with the simulator. In previous work, the KRONOS simulator has been enhanced to include the same control strategy [31, 32]; however, in KRONOS the strategy cannot be changed without major revisions to the simulator code. In short, AIMSUN2 was the most amenable to integration with user-specified ATMS applications.

Table 2.3: Model Comparisons based on Quantitative and Qualitative Tests

SUBJECT	FRESIM	AIMSUN2	FREFLO	FREQ11	KRONOS	INTEGRATION
I-494 WB volume est. (1)	Good	Good	Very good	Very good	Very good	Good
I-494 WB volume est. (2)	Good	Average	Good	Good	Good	Average
I-94 EB volume est.	Good	Average	Good	Good	Good	Good
Congestion spill-back	Average	Average	Good	Fair	Average	Good
Incident spill-back	Good	Good	Average	Fair	Average	Good
Dissipation times	Good	Fair	Average	Average	Average	Good
Vehicle conservation	Fair	Fair	Good	Very good	Very good	Average
Geometric data entry	Average	Good	Average	Good	Very good	Average
Geometric abilities	Good	Good	Fair	Average	Good	Good
Flow data entry	Good	Fair	Average	Good	Good	Fair
Data entry time	Average	Fair	Average	Good	Good	Poor
Calibration process	Fair	Fair	Average	Average	Good	Average
Calibration time	Average	Average	Good	Good	Good	Fair
Simulation time	Average	Average	Good	Very good	Good	Fair
Output files	Fair	Good	Fair	Good	Fair	Average
Output graphics	None	Good	None	Fair	Average	Average

2.2 Simulator Review

AIMSUN2 (Advanced Interactive Microscopic Simulator for Urban and Non-Urban Networks) is a simulation package that can model complex geometry (such as interchanges, collectors-distributors, roundabouts, etc.) of realistic, large-scale urban networks consisting of both freeways and surface streets. In this chapter, an overview of the simulator is given in order to get a better insight and appreciate the functionality and capabilities of the simulator. A detailed review of the various models used by AIMSUN2 is also presented.

AIMSUN2 follows a microscopic simulation approach. This means that the behavior of each vehicle in the network is continuously modeled throughout the simulation period using several driver behavior models (car following, lane changing, gap acceptance, etc.). AIMSUN2 is a combined discrete-continuous simulator; this implies that there are some elements of the system (i.e., vehicles, detectors) whose states change constantly every simulation time step throughout the simulation time, while there are other elements (i.e., traffic lights, entrance points) whose states change discretely at specific times. AIMSUN2 distinguishes between different classes of vehicles and drivers and can simulate incidents, conflicting movements and other complexities encountered in reality such as weaving sections, HOVs, etc.

AIMSUN2 is integrated in GETRAM (Generic Environment for TRaffic Analysis and Modeling) [26], a simulation environment comprising a traffic network graphical editor called TEDI, a network database, and an API (Applications Programming Interface). GETRAM also includes a module for storing and presenting results as shown in Figure 2.1.

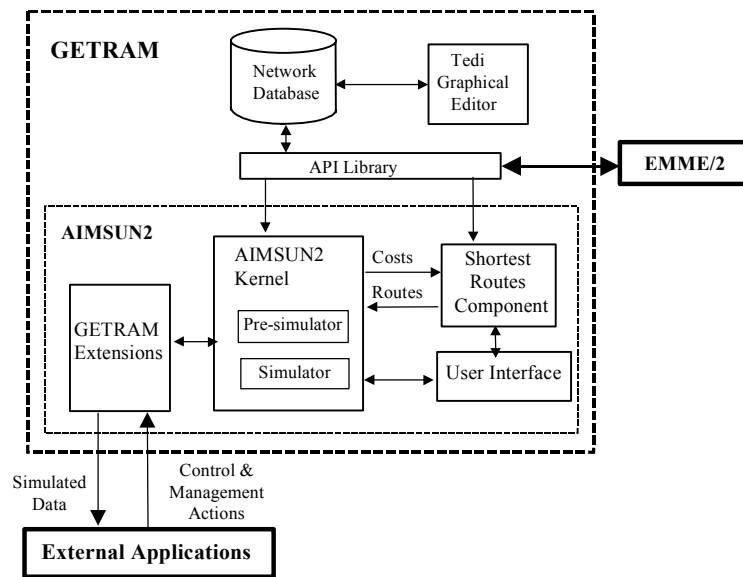


Figure 2.1. Conceptual structure of GETRAM/AIMSUN2

AIMSUN2 has the capabilities to model all combinations of roadway geometries. This includes urban networks, freeways, complex interchanges and intersections. Different types of traffic control such as pre-timed or actuated signals, unsignalized intersections and ramp control can

also be modeled. The vehicle behavior models are functions of several parameters that allow modeling of different types of vehicles, i.e., cars, buses, trucks, etc. They can be grouped in classes and can move on exclusive lanes. Due to the detailed modeling of each vehicle in the network, AIMSUN2 can simulate all types of traffic measurements that traditional or advanced detectors can provide, i.e., vehicle counts, occupancy, presence, speed, queue length, and classification.

The simulator provides detailed statistical outputs for all MOEs and traffic variables such as flows, speeds, travel times, delays, etc. These can be global (for all of the simulation period) or at user-defined periodic intervals. These measurements can be aggregated for the entire system or disaggregated for sections and turnings. The output is quantitative as well as graphical, including a drawing of the network and an animated representation of the vehicles.

TEDI is a graphical editor for the creation of road geometry and demand patterns that are needed by simulators like AIMSUN2. To facilitate this task, TEDI accepts, as a background, a graphical description of the network area, so that sections and nodes can be built subsequently into the foreground. The editor supports both urban and interurban roads, which means that the level of detail goes down to elements such as turning or acceleration/deceleration lanes, entrance and exit ramps, intersections, traffic signals, and ramp meters. The geometry of the links is specified at the microscopic level, but the ease of use of the editor makes it as fast as specifying one-dimensional links in some macroscopic systems, and nodes are created automatically.

The input required by AIMSUN2 is composed of three types of data: network description, traffic control plans and traffic conditions. The network description contains information about the geometry of the network, turning movements, layout of links (or sections), junctions and location of detectors along the network. The traffic control plans are composed of the description of phases and their duration for signal-controlled junctions, the priority definition for unsignalized junctions, and any required ramp-control information. AIMSUN2 accepts two classes of inputs for the simulator depending on how traffic is to be simulated. The combination of traffic volumes at the input sections, the turning proportions at junctions and the initial state of the network, is the classical input to most of the microscopic traffic simulation models moving cars

probabilistically through a network. The other class of input data consists of Origin-Destination (O-D) matrices [29]. In this case, explicit routes are computed for vehicles in the network according to their origins and destinations, and vehicles are allocated to the routes following specific route choice models. Routes between every origin, or between every section and every destination can be determined using various models such as time-dependent shortest paths. A route choice model uses these routes for the allocation of vehicles to routes at the next time interval. The shortest route component takes into account turning penalties as different turning movements at the end of a section have, in general, unequal travel times.

The simulation process, when simulating in the Route Based mode, includes an initial computation of routes going from every section to every destination, a shortest route component which calculates periodically the new shortest routes according to the new travel times provided by the simulator, and a route selection model which allocates vehicles to the routes according to the corresponding decision rules.

At each time interval, the simulation cycle updates the unconditional events scheduling list, e.g., events such as traffic signal changes. After this updating process, a set of nested loops updates the states of the entities and vehicles in the model. Once the last entity has been updated, the simulator performs the remaining operations. This includes generating new vehicles according to the specified or estimated time sliced O/D matrix, allocating vehicles to routes according to the route choice model, calculating the new routes if it is the case, collecting data, and other record keeping. If traffic diversion is not included in the network, the use of the O/D matrix is not needed. In this case, the user can enter a percentage of exiting or turning flows at freeway exits and intersections, respectively. The system's open architecture implies that all vital models such as traffic control, vehicle behavior, route calculation, route choice and others are independent of the simulation logic and, therefore, could be replaced by alternative ones. The logic of the simulation process is illustrated in Figure 2.2.

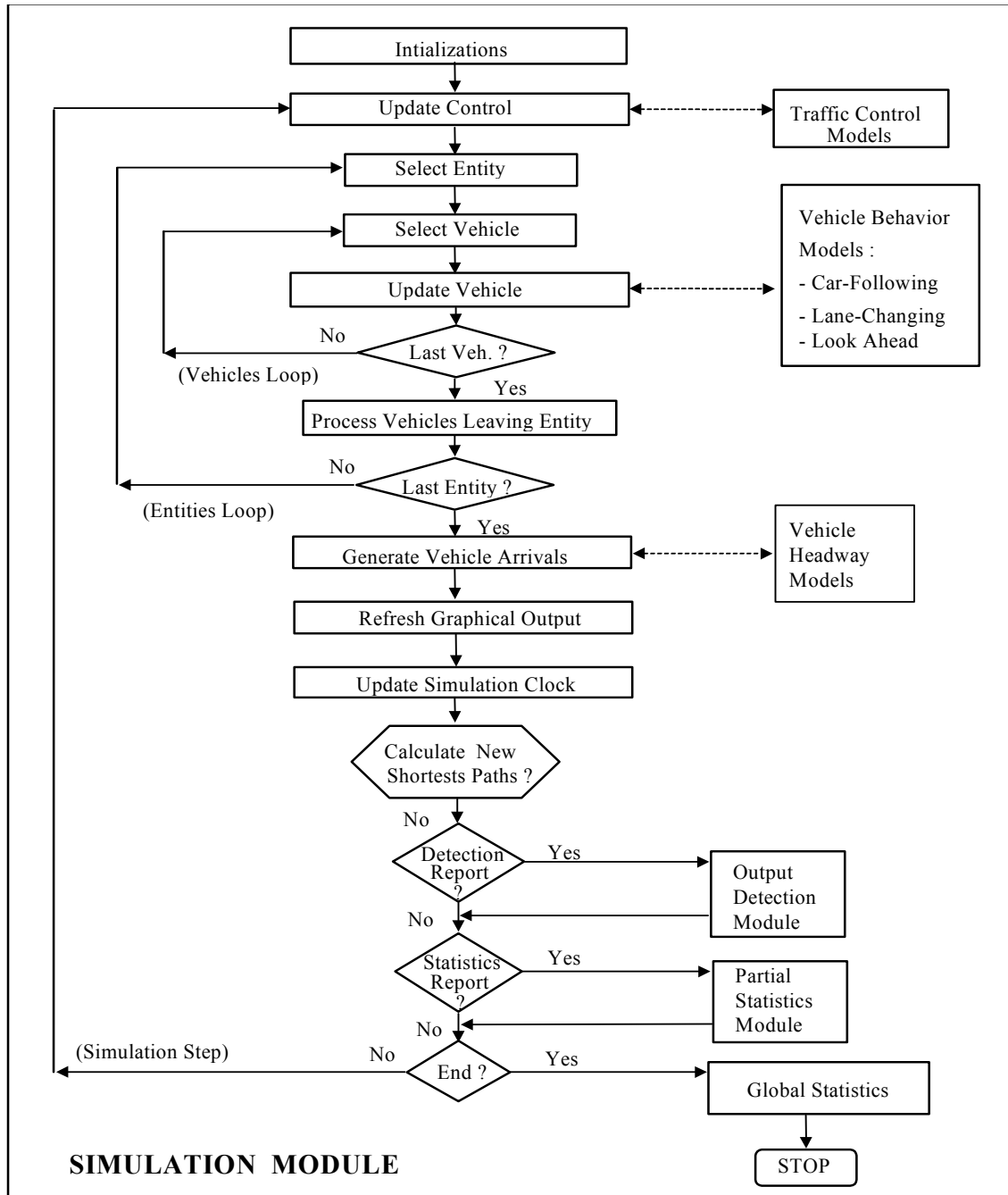


Figure 2.2. AIMSUN2 Simulation Process.

2.2.1 Models Currently in Use

Most of the currently existing microscopic traffic simulators are based on the family of car-following, lane changing and gap acceptance models to model a vehicle's behavior. Among such models, one can find Helly's model [34] implemented in SITRA-B+ [35], Herman's model [36] or its improved version by Wicks [37] implemented in MITSIM [38], the psychophysical model of Wiedemann used in VISSIM [39], or Gipps model [40] used in AIMSUN2 [28, 41]. Other microscopic simulators such as INTEGRATION [20] and PARAMICS [42] employ heuristics or modeling not disclosed.

Helly's model uses the following expression for the acceleration of the follower $\ddot{x}_{n+1}(t+T)$:

$$\ddot{x}_{n+1}(t+T) = c_1[\dot{x}_n(t) - \dot{x}_{n+1}(t)] + c_2[x_n(t) - x_{n+1}(t) - D]$$

where T is the reaction time for the vehicle-driver system, c1 and c2 are the relative velocity and headway control parameters and D the desired headway, typically expressed as: $D = l_n + \tau_{n+1}\dot{x}_{n+1}(t)$ with l_n being the length of the leader vehicle, n, and τ_{n+1} the time headway for the follower, n+1.

Herman's model assumes an acceleration rate given by

$$\ddot{x}_{n+1}(t) = \alpha^\pm \frac{\dot{x}_{n+1}^{\beta^\pm}(t)}{g_{n+1}^{\gamma^\pm}(t)} (\dot{x}_n(t) - \dot{x}_{n+1}(t))$$

where α^\pm , β^\pm and γ^\pm are model parameters, α^+ , β^+ , γ^+ are used for acceleration ($\dot{x}_{n+1}(t) \leq \dot{x}_n(t)$), and α^- , β^- , γ^- for deceleration ($\dot{x}_{n+1}(t) > \dot{x}_n(t)$), and $g_{n+1} = x_{n+1} - x_n - l_n$ represents the gap distance from the leading vehicle.

The Gipps model consists of two components: acceleration and deceleration. The first represents the intention of a vehicle to achieve certain desired speed, while the second reproduces the limitations imposed by the preceding vehicle when attempting to drive at the desired speed. This model states that the maximum speed at which a vehicle (n) can accelerate during a time period (t, t+T) is given by:

$$V_a(n, t + T) = V(n, t) + 2.5a(n)T \left(1 - \frac{V(n, t)}{V^*(n)}\right) \sqrt{0.025 + \frac{V(n, t)}{V^*(n)}} \quad \dots (1)$$

where: $V(n, t)$ is the speed of vehicle n at time t ; $V^*(n)$ is the desired speed of the vehicle (n); $a(n)$ is the maximum acceleration for vehicle n ; T is the reaction time.

The maximum speed that the same vehicle (n) can reach during $(t, t+T)$, according to its own characteristics and the limitations imposed by the presence of the leader vehicle is:

$$V_b(n, t + T) = d(n)T + \sqrt{d(n)^2 T^2 - d(n) \left[2\{x(n-1, t) - s(n-1) - x(n, t)\} - V(n, t)T - \frac{V(n-1, t)^2}{d'(n-1)} \right]}$$

where: $d(n)$ (< 0) is the maximum deceleration desired by vehicle n ; $x(n, t)$ is position of vehicle n at time t ; $x(n-1, t)$ is position of preceding vehicle ($n-1$) at time t ; $s(n-1)$ is the effective length of vehicle ($n-1$); $d'(n-1)$ is an estimation of vehicle ($n-1$) desired deceleration. The final speed for vehicle n during time interval $(t, t+T)$ is the minimum of those previously defined speeds:

$$V(n, t + T) = \min\{V_a(n, t + T), V_b(n, t + T)\}$$

The position of vehicle n in the current lane is updated by taking the speed into the movement equation:

$$x(n, t + T) = x(n, t) + V(n, t + T)T$$

A common drawback of most of these models is that the model parameters are global, i.e., constant for the entire network whereas it is well known that a driver's behavior is affected by geometry and traffic conditions. Long-term experimentation revealed that a more realistic car-following modeling for microscopic simulation accounting for local behavior and driving conditions significantly improves performance.

2.2.2 Modeling in AIMSUN2

The AIMSUN2 car following model evolved after the seminal Gipps model, which was improved to meet the local requirements, described earlier. Three main aspects of the model have been enhanced based on the empirical evidence gathered calibrating the model for observed data:

- The way in which the vehicle speed $V^*(n)$ used in the Gipps model is calculated

- How vehicles in adjacent lanes influence a vehicle's behavior
- Accounting for grade effects in car-following
-

Speed calculation

The first improvement is related to the vehicle speed $V^*(n)$ used in equation (1). In AIMSUN2 implementation, $V^*(n)$ is the desired speed of vehicle n for the current section. In car-following model a leading vehicle, attempts to drive to its maximum desired speed. Three parameters are used to calculate the maximum speed of leading vehicle n while driving on a particular section or turning:

Maximum desired speed of n : $v_{\max}(n)$. This is a vehicle parameter.

Speed acceptance of n : $\theta(n)$. This is a vehicle parameter measuring the driver's degree of compliance of the speed limits on the section. $\theta(n) = 1$, represents the perfect compliance. $\theta(n) < 1$, a driver driving below the speed limits, and $\theta(n) > 1$, faster than the speed limits. $\theta(n)$ in AIMSUN2 is sampled from a probability distribution, when such information is available, implicitly modeling the aggressiveness of the drivers.

Speed limit in section or turning s : $S_{\lim_{nt}}(s)$. This is a section parameter.

The actual speed limit for a vehicle n on a section or turning s , $S_{\lim_{nt}}(n, s)$, is given from:

$$S_{\lim_{nt}}(n, s) = S_{\lim_{nt}}(s) \cdot \theta(n)$$

The maximum desired speed of vehicle n on a section or turning s , $v_{\max}(n, s)$ is:

$$v_{\max}(n, s) = \text{MIN}[S_{\lim_{nt}}(n, s), v_{\max}(n)]$$

Thus the local maximum desired speed $v_{\max}(n, s)$ equals the desired speed $V^*(n)$ in Eq.(1).

Influence of adjacent lanes

When the leading vehicle is driving along a section, the AIMSUN2 car-following model takes into account the potential influence of certain number of vehicles ("Nvehicles") driving slower in the adjacent right-side lane or left-side lane, when driving on the left. The model calculates first the mean speed for Nvehicles driving downstream of the vehicle in the adjacent slower lane ("MeanSpeedVehiclesDown"). Only vehicles within a certain distance ("MaximumDistance")

from the current vehicle are taken into account. Two cases are distinguished: 1) the adjacent lane is an on-ramp, or acceleration lane, and 2) the adjacent lane is any other type of lane. Apart from *Nvehicles* and *MaximumDistance* parameters, the user can define two additional parameters, “*MaximumSpeedDifference*” and “*MaximumSpeedDifferenceOnRamp*.” Then, the final desired speed of a vehicle on a section is given from the following logic:

```

if (the adjacent slower lane is an on-ramp) {
    MaximumSpeed = MeanSpeedVehiclesDown + MaximumSpeedDifferenceOnRamp
}
else {
    MaximumSpeed = MeanSpeedVehiclesDown + MaximumSpeedDifference
}
DesiredSpeed = Minimum ( $v_{\max}(n, s)$ ,  $\theta(n) * \text{MaximumSpeed}$ )

```

This procedure ensures that the differences of speeds between two adjacent lanes will always be lower than *MaximumSpeedDifference* or *MaximumSpeedDifferenceOnRamp*, depending on the case.

Effects of Grades

The maximum desired acceleration for a vehicle (*veh_acc*) is a vehicle parameter defined all vehicles belonging to the same class. The influence of the section grade on the vehicle movement is taken into account by increasing or reducing the acceleration and deceleration rate. The maximum acceleration for a vehicle on a section is a function of the grade and the maximum desired acceleration for the vehicle i.e.:

$$\text{accel} = \text{Maximum} (\text{veh_acc} - \text{grade} * 9.81 / 100.0, \text{veh_acc} * 0.1)$$

In order to avoid zero or negative acceleration values, a minimum value of 10% of the maximum desired acceleration for the vehicle is used.

Model Calibration and Testing

In addition to numerous tests that were performed, the car-following model in AIMSUN2 was tested and calibrated in various real life projects; due to space limitations only the benchmark test performed based on the data and the methodology supplied and proposed by a research group from Robert Bosch GmbH [443,44] is presented here. This test employed a set of field data and most of the microsimulator developers in Europe and North America were invited to participate. The test for the car-following model was defined as follows:

“The primary task of a car-following model is to reproduce realistic car-following behavior. The reality has been measured with a radar sensor equipped vehicle recording distance and relative speed to the front car (additionally to the own car's speed) in a 100 ms cycle, (see [44] for further details). One specific sequence of 5 min length has been chosen to perform the comparison. This sequence has been recorded under stop and go conditions during an afternoon peak on a one-lane-per-direction fairly straight road in Stuttgart, Germany. Stop and go is most challenging to the models because the free flow behavior is relatively easily reproducible by any model. During the five min sequence several decelerations and accelerations of the front car have been observed. At one moment after 144 sec the front car turned off resulting in a distance step of about 40 m. Because such a maneuver can always happen in real traffic the models have to be able to deal with it. Note that it can't be the target of a simulation model to reproduce exactly the measured behavior of this specific driver in the specific test vehicle. Driver and vehicle variations have to be respected. Hence, the main focus lies on qualitative differences. But the fairly good reproduction of the behavior indicates a model's realism. To give an impression of similarity to the measured behavior a quantitative error metric on the distance seems to be reasonable. To avoid overrating discrepancies for large distances a relative metric was chosen weighted by the logarithm and squared. Only the values after each second have been considered.”

The error metric used to measure the accuracy of the fitting between measured and simulated

values was:
$$Em = \sqrt{\sum \left[\log \left(\frac{d_{sim}}{d_{meas}} \right) \right]^2}$$
 where d_{sim} is the distance of the simulated vehicle,

d_{meas} is the distance measured with the test vehicle, and \log denotes the logarithm base 10.

Figures 2.3 and 2.4 depict the observed versus simulated relative speed and distance between

leader and follower. The error metric for the AIMSUN2 model was 3.4726. These results demonstrate that the AIMSUN2 car-following model is capable of a fairly good reproduction of the observed values. According to the Bosch tests [43] the accuracy of the AIMSUN2 model was the second best of all tested, resulting in the second lowest Em.

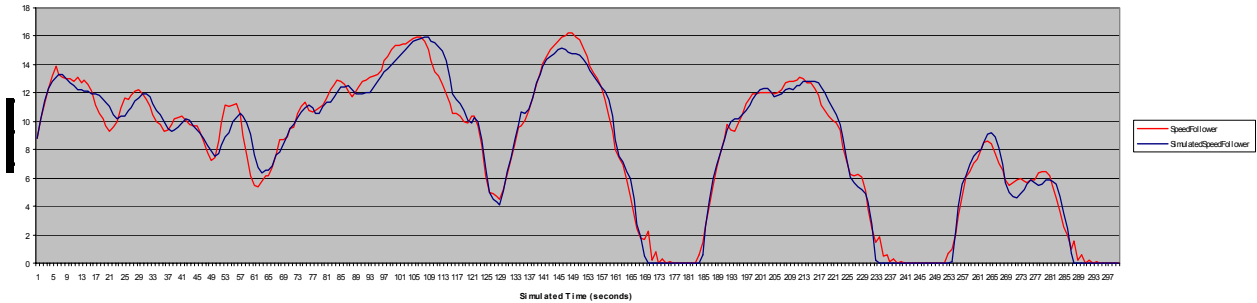


Figure 2.3. Measured and Simulated Speeds

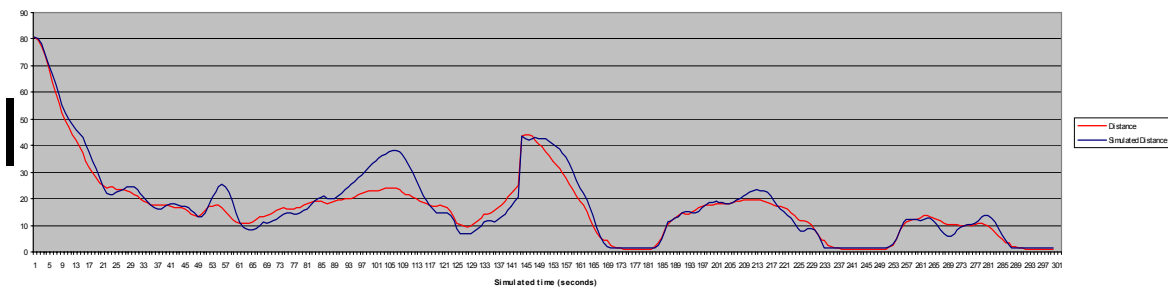


Figure 2.4. Measured and Simulated Relative Distances Leader-Follower

An additional test to analyze the quality of the microscopic simulator is to check the ability to reproduce macroscopic behavior. Bleile et al. [44] performed a related test to compare various microscopic simulators, described as follows: “The macroscopic behavior of a microscopic model can be most easily tested by simulating the traffic on cyclic, one-lane roads. This excludes any effects of lane changes and node passing and concentrates on the car-following task. For this study, a cyclic road of 1000 m length was used. A fixed number of vehicles had been initially set with speed value 0 km/h at randomly determined positions. All vehicles had the same length of 4.5 m and the drivers had the same free flow speed of 54 km/h. Starting with this initial situation a 10 minute time period was simulated without any measurements to reach traffic conditions which are achievable by the model's behavior itself. After the starting phase the traffic behavior (exact passing time and speed value of each vehicle) was recorded at one local measurement

point during a simulation time of 2 hours. The fixed number of vehicles for the simulation run was varied in discrete steps to realize different traffic densities. To visualize the results, the traffic flow has been drawn versus the density (given as the number of initially set vehicles on the 1 km ring). The maximal mean traffic flow value of about 1800 veh/h is known as a quite realistic value for longer periods of measurement time. Under urban traffic conditions this maximal flow is typically reached at higher density values than for freeway traffic.”

The results of AIMSUN2 for the simulated flow density curve versus the one derived from data for the independent test are displayed in Figure 2.5; as the figure suggests the agreement between the two is satisfactory. This graphical conclusion was confirmed by the values of the same error metric for finding the fitness between the measured and simulated values. The error in this case was $E_m=0.063381$.

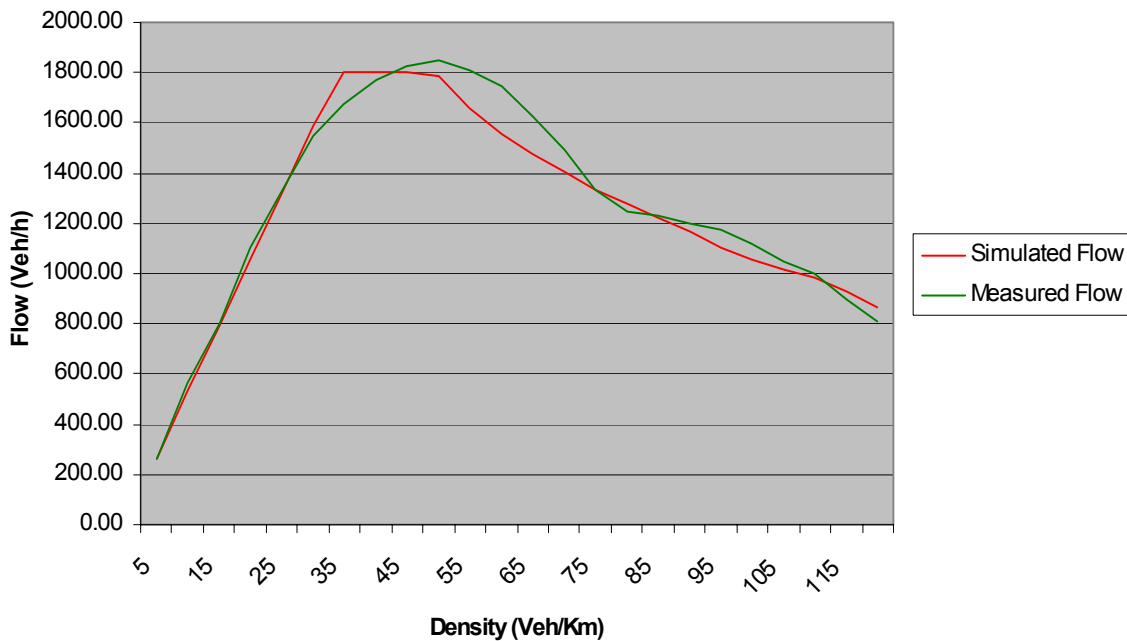


Figure 2.5. Empirical versus simulated flow-density curves

2.2.2.1 Lane Changing Model

The lane changing process in AIMSUN2 can also be considered as a further evolution of the Gipps lane change model [45]. Lane changing is modeled as a decision process considering the

necessity of the lane change (as in the case of turning maneuvers determined by the route), the desirability of the lane change (as for example to overpass a slow moving vehicle), and the feasibility conditions for the lane change. The lane-changing model is a decision model that emulates the driver's behavior according to the following logic:

Each time a vehicle has to be updated, the model asks the question: Is it necessary to change lanes? The decision depends on several factors: the turning feasibility at current lane, the distance to next turning and the traffic conditions in the current lane. The traffic conditions are measured in terms of speed and queue lengths. When a vehicle is moving slower than the driver wishes, he tries to overtake the preceding vehicle. On the other hand, when he is traveling fast enough, he tends to go back to the slower lane.

If the answer to the previous question is affirmative, to proceed with lane changing two additional questions need to be asked?

Is it desirable to change lanes?

This requires checking if there will be any improvement in the traffic conditions for the driver as a result of the lane changing. This improvement is measured in terms of speed and distance. If the speed in the future lane is faster (i.e., a user specified threshold is exceeded) than the current lane or if the queue is shorter than a threshold, then it is desirable to change lanes.

Is it possible to change lanes?

This requires verifying if there is sufficient gap to make the lane change safely. For this purpose, both the braking imposed by the next downstream vehicle to the changing vehicle and the braking applied by the changing vehicle to the future upstream vehicle is calculated. If both braking ratios are acceptable, then lane changing is possible.

In order to achieve a more accurate representation of the driver's behavior in the lane changing decision process, three different zones inside a section are considered, each one corresponding to a different lane changing motivation. The distance up to the end of the section (the next turning

point) characterizes these zones. Figure 2.6 depicts the structure of these zones that are defined as follows:

Zone 1: This is the farthest from the next turning point. The lane changing decisions are mainly governed by the traffic conditions of the lanes involved; the feasibility of the next desired turning movement is not yet taken into account. To measure the improvement that the driver will get on changing lanes, several parameters need to be considered: the desired speed of the driver, speed and distance of the current preceding vehicle and speed and distance of the future preceding vehicle.

Zone 2: This is the intermediate zone. Mainly it is the desired turning lane that affects the lane changing decision. Vehicles not in a valid lane (i.e., a lane where the desired turning movement can be made) tend to get closer to the correct side of the road where the turn is allowed. In this zone, drivers look for a gap and may try to accept it without affecting the behavior of vehicles in the adjacent lanes.

Zone 3: This is the nearest to the next turning lane-changing point. Vehicles are forced to reach their desired turning lanes, reducing the speed if necessary and even coming to a complete stop in order to make the lane change possible. Also, drivers in the adjacent lane can modify their behavior in order to allow a gap big enough for the lane-changing vehicle.

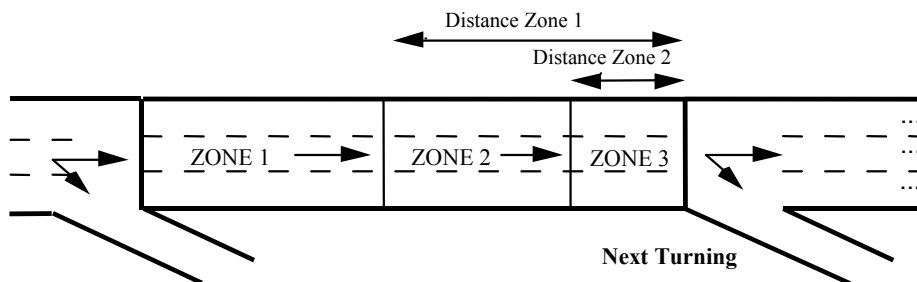


Figure 2.6. Lane Changing Zones

Lane changing zones are defined by two parameters, “Distance to Zone 1” and “Distance to Zone 2.” These parameters are defined in time (seconds) and they are converted into distance whenever it is required for each vehicle at each section using the “Vehicle Desired Speed” at a section. This means that these distances are then local parameters, their value depending on the current traffic conditions in the section. AIMSUN2 provides default values based on the user’s experience; however, it should be noted that finding the right values for these parameters is part of the model calibration exercise that the user has to perform.

Turning Maneuvers

When a vehicle reaches the end of a lane in a section and enters into a junction it may have the possibility of choosing among different connections, i.e., movements from origin lane (the current one) to the destination lanes (the lanes in the next section).

For instance, in Figure 2.7, vehicle a is going to turn right from the rightmost lane of section 17 to section 11. However it can choose among the different destination lanes in section 11. As this vehicle’s ahead turn is again to the right it will enter section 11 through the rightmost lane instead of using the central one.

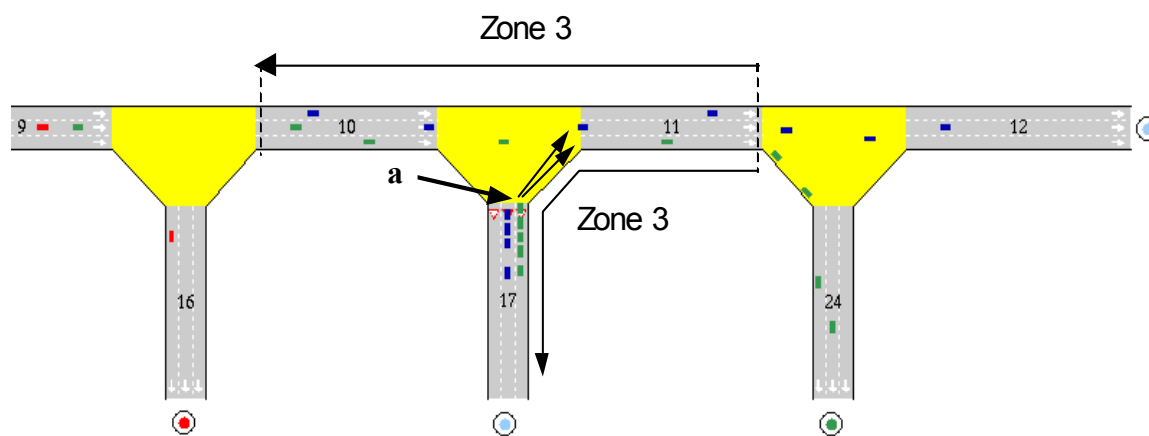


Figure 2.7. Variability of Lane Changing Zones

When a vehicle crosses from zone 1 to zone 2 there is a change in the vehicle’s behavior, as the next turn becomes relevant now. Also, the crossing from zone 2 to zone 3 produces a change in

the behavioral rules of the vehicles, as now reaching the turning lane becomes urgent. In order to distribute these changes of behavior along a longer distance a greater variability is given to the lane changing zones. These zones are calculated particularly for each vehicle according to the following equation:

Distance Zone n for vehicle v in section s (in meters) =

Distance Zone n (in seconds) * Speed Limit of Section s * Vehicle v Coefficient

Vehicle v Coefficient = Speed Limit of Section s / Desired speed of Vehicle v in section s

This algorithm ensures that for vehicles whose desired speed is slower than the speed limit the lane changing zones will be longer than for vehicles whose desired speed is greater than the speed limit. It means, for instance, that a heavy truck will try to reach the appropriate turning lane earlier than a speed car.

Lane Changing Modeling at On-Ramps

A special lane changing modeling is applied at entrance ramps. An additional zone parameter may be defined, “Time Distance On Ramp”. This is the distance (in seconds, converted into distance as before) from those lateral lanes considered to be on-ramp lanes, in order to distinguish between a common lateral lane, that is a long lane used for overtaking which drops down, from the proper on-ramp lanes, which are never used for overtaking. Vehicles driving on a lateral lane that are farther than the above time distance on ramp from the end behave as if they were in Zone 1 of a normal lane. When they are closer than the time distance on ramp to the end of the lane, they behave as having to merge from an on-ramp. Merge from the on-ramp model takes into account whether a vehicle is stopped or not, and if it is stopped, whether it is at the beginning of the on-ramp queue or not and how long has been waiting. There is another vehicle parameter, “Maximum Waiting Time”, that determines how long a driver is willing to wait before getting impatient.

Lane Changing Prohibitions

Prohibitions to changing lanes are modeled through the definition of continuous lines between lanes. Lane changing is not allowed wherever there is a continuous line between two lanes. Through this feature, the user may avoid undesired lane changing in some conflicting points such as a nearby on-ramp. For each pair of adjacent lanes, the user may define a continuous line at the right, at the left or on both sides. A continuous line at the right means that no lane change is allowed from the right lane towards the left lane. A continuous line at the left means that no lane change is allowed from the left lane towards the right lane. A continuous line at both sides means that no lane change at all is allowed between both lanes. Continuous lines can be defined for the whole lane length or only part of it. Figure 2.8 shows an example where there is a short continuous line between the on-ramp lane (lane number 1) and the rightmost main lane (lane number 2), so vehicles can not merge from the on-ramp during the first 20 meters, as is the case of vehicle 2. There is also a continuous line between the left lane (lane number 3) and the right lane with the purpose of creating more gaps for the nearby on-ramp. This is the case for vehicle 4, which is not allowed to change to the right lane, which would disturb the merging of vehicles 1 and 2.

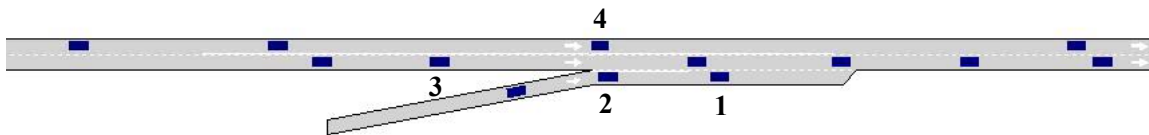


Figure 2.8. Example of continuous lines

Overtaking Maneuver

An overtaking maneuver takes place mainly in zone 1, although it can also take place in zones 2 and 3 when the vehicle is in the appropriate turning lane. In order to promote or discourage the overtaking, there are two parameters that the user can define: “PercentOvertake” and “PercentRecover.” PercentOvertake is the percentage of the desired speed of a vehicle below which the driver may decide to overtake. This implies that whenever the lead vehicle is moving slower than PercentOvertake of the follower’s desired speed, the follower vehicle will try to overtake. The default value is 0.90. PercentRecover is the percentage of the desired speed of a

vehicle above which a driver may decide return to the slower lane. It means that whenever the lead vehicle is driving faster than PercentRecover of the follower's desired speed, the follower driver will try to recover the right (or left) lane. The default value is 0.95.

The whole lane change process is modeled formally as a decision tree model whose logic structure is depicted in Figure 2.9. The process identifies the type of entity (central lane, off-ramp lane, junction, on-ramp etc.) in which the maneuver is going to take place and then determines how the zone modeling should be applied. The current traffic conditions are analyzed and the level at which the lane change can be performed is determined, and then the corresponding model is applied.

2.3 Improving the AIMSUN Ramp Exiting Model

When traffic conditions are very congested it may happen that some vehicles cannot reach the appropriate turning lane and consequently miss their designated turn. This situation could appear either in urban networks where there are short sections or in freeways where weaving sections may be relatively short. It gets worse as the sections get more congested.

Tuning some modeling parameters such as lane changing zone distances, simulation step, acceleration rates, etc., could improve the behavior in order to minimize the number of vehicles which cannot make their designated turns. Also, when feasible, using polysections instead of sections in modeling geometry such weaving areas might help to improve the situation, but it may not be enough.

To override these drawbacks a major improvement has been done in the lane change model to enhance turning behavior of vehicles. The objective is to provide vehicles with the knowledge of various next turning movements, so that they will be able to make decisions not only based on the immediate next turning movement, but on a set of next turning movements.

The improved lane changing exit model consists of four steps:

At any time, each vehicle knows the next two turning movements, so the lane changing decisions are influenced by two consecutive turns.

The lane changing zones 2 and 3 of any section is extended back beyond the limits of the section, therefore affecting the upstream sections.

The next turning movement also influences the turning maneuvers so the selection of destination lane is done based also on the next turn.

A greater variability is given to the lane changing zones in order to distribute the lane changing maneuvers along a longer distance.

These steps are described in more detail.

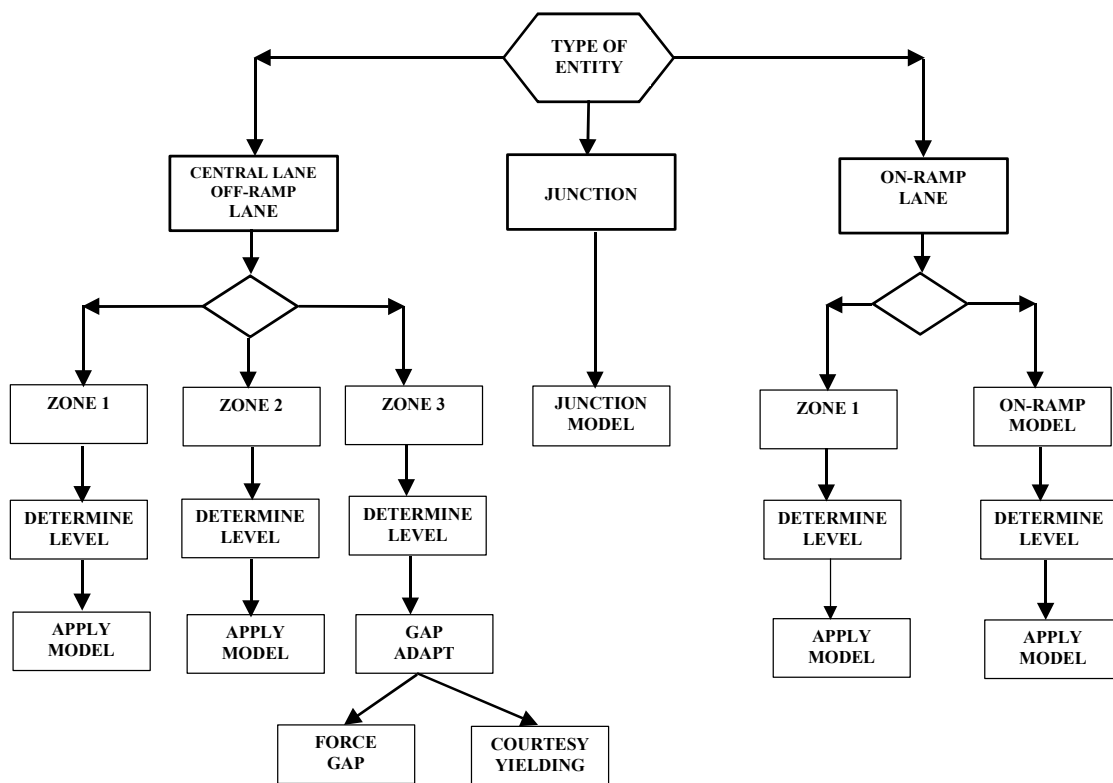


Figure 2.9. Lane changing decision tree

Influence of next two turning movements in lane changing

The lane change model has been recently improved. Under the directions of the University of Minnesota team, the software developers implemented a more accurate lane changing model. Although the simulator improvements made for this project are described in detail in the next chapter, this improvement was not only implemented in the simulator version used in this study but also joined the production version as a standard feature. The reason for this improvement is based on the inability of the previous version of the simulator to realistically describe traffic conditions in short weaving sections.

When a vehicle is generated and introduced into the network, two turning movements are calculated: the next turn and the ahead turn. This means that the vehicle has knowledge of the first three consecutive sections (or polysections) that it has to follow: the current input section, the second defined by the next turn, and a third section that is defined by the second section and the ahead turn. Then, each time a vehicle enters a new section (or polysection), the ahead turn becomes the next turn and a new ahead turn is calculated.

The behavior of a vehicle driving in zones 2 or 3 of a section is mainly governed by the next turn, therefore it will first try to reach a lane where the next turn is feasible. Once it is driving in an appropriate lane with respect to the next turn, it will take into account the ahead turn. It consists in checking whether it is already driving inside the extended zone 2 or 3 of the next section. In that case it will try to find the best lane in the current section that, still allowing next turn, will drive the vehicle either to a lane in the next section where the ahead turn is feasible or to the lane closest to a feasible one.

This is illustrated in the example of Figure 2.10. Vehicles that are dark (example, vehicle c) take the left exit whereas vehicles that are light in color (such as vehicle b) take the right exit. Vehicle a in section 3 is already located in the left lane in order to be as close as possible to the left when entering section 2. In this case the next turn is from 3 to 2 while the ahead turn will be from 2 to 4. Therefore, although the vehicle a being in the right lane would meet the next turn requirements, it wouldn't be in an ideal lane to make the ahead turn. Similar behavior for vehicle b can be observed: it has changed to the rightmost lane in section 1 in order to reach the ahead turn to the right.

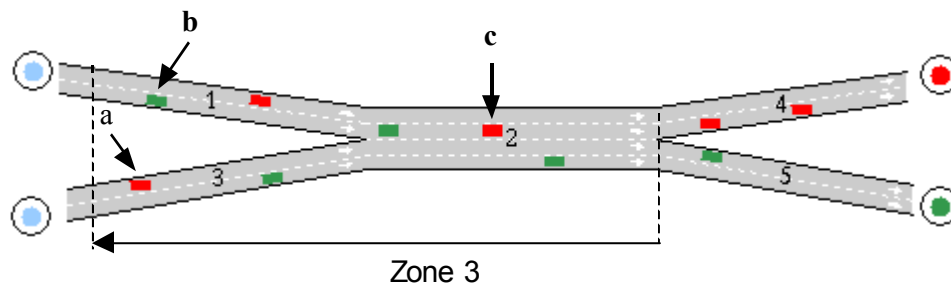


Figure 2.10. Extending back zones 2 and 3

Extending back zones 2 and 3

In case if zone 2 or zone 3 of a section is longer than the length of that section, the zone is extended beyond the limits of the section, therefore extending into the upstream section. In order to take full advantages of the lane changing exit model, it is now very important to define the lane changing zones appropriately. The improvements to the lane changing model will not be applied whenever the lane changing zones 2 or 3 are shorter than a section. In the case of a polysection, the user must take care to define the zones appropriately for each section of a polysection. The zone extension is also illustrated in Figure 2.10.

3 SIMULATION ENHANCEMENTS

As mentioned in the introduction, there is a need to develop and test alternate ramp control strategies before implementing them. In general, most of the simulators have been designed to include one or more known ramp control schemes. This allows the user to compare the available schemes within a single simulator, but it doesn't allow comparison with other non-supported schemes. To compare non-supported schemes using a single simulator would require modifying a major part of the simulator's source code to incorporate the non-supported strategy. Otherwise different simulators need to be used thus casting a doubt on the results because they might use different modeling techniques and different criteria in evaluating measures of effectiveness such as delay, queue-lengths, throughput, etc. However, any simulator can emulate only a predetermined control strategy specified at the time of development and AIMSUN2 was no exception. Clearly, for comparing alternate strategies we need the capability to simulate a wide variety of them. It was also mentioned in the introduction that one of the problems in the widespread use of simulation is the tedious job of entering initial and boundary conditions. For these purposes, three major enhancements had to be made to the simulator. The first was the development of the Control Plan Interface (CPI), which facilitates the simulation of alternate ramp control strategies, by connecting the simulator with an external user defined ramp control strategy. The second enhancement was the development of Traffic Demands Creation Application in which the process of creation of boundary conditions was introduced and automated. The needs for the development of the above two enhancements, their design and implementation are described in detail in this chapter. The third enhancement was to emulate Mn/DOT's ramp control strategy described in Section 2.3 in order to evaluate the effectiveness of the test sites described in chapter 4. Finally, as described in section 2.5.9.1, the simulator was improved with the implementation of a new lane changing model. The implementation was contacted by the software developers under the directions of the university team.

Design and Development of the CPI

The design of the CPI can be better understood by knowing how the traffic control systems operate in real life. From a conceptual viewpoint, the operation of advanced traffic control

systems can be described as follows: The road network is equipped with traffic detectors with a specific layout corresponding to the requirements of the control strategy. The detectors supply the necessary real-time traffic data (flows, speeds, occupancies, etc.) to the traffic control or management system, which, after suitable processing, makes control decisions, e.g., extend the green phase, change to the red phase, apply some traffic calming strategies, etc. These decisions must then be communicated to the traffic control devices such as traffic signals, ramp meters etc. for implementation. To evaluate and test any of these systems, a simulator must be capable of incorporating in the model the corresponding traffic devices and emulate their functions, e.g., provide the specific traffic measurements at the required time intervals, increase the phase timing in a given amount of time, implement a traffic calming strategy such as reduce the speed limit, recommend an alternative route, etc. AIMSUN2 models these traffic devices and emulates their functions. The CPI is an interface that integrates AIMSUN2 with external user-defined ramp control logic.

It has been widely established that one of the most critical features of any traffic control scheme involves some form of operator interaction with the system. This interaction can be in the form of incident management, overriding ramp meters in response to long queues, displaying real-time traffic information or alternative routes through variable message signs etc. Ways in which CPI provides this functionality can be illustrated with the help of Figure 3.1, which provides a high level view of the all the components involved. The human interaction is implemented with the help of G2, an expert system and the communication is through the CPI.

The interaction of the various components can be explained as follows: The simulator provides the necessary measurements, which the CPI transfers to the external ramp control logic. In its turn, the control logic calculates the new rates and transfers them to the simulator through the CPI. The communication between AIMSUN2 and CPI occurs at every simulation step (0.5 to 1.0 sec) whereas the communication between the CPI and the ramp control logic is at a user defined time period. The communication between G2 and CPI is always open and any decisions made by the operator using G2 are implemented in the next simulation step. One more reason behind designing the system as shown above is to isolate the ramp control logic as much as possible from the implementation of the CPI and G2. The intention is to allow the user to replace one

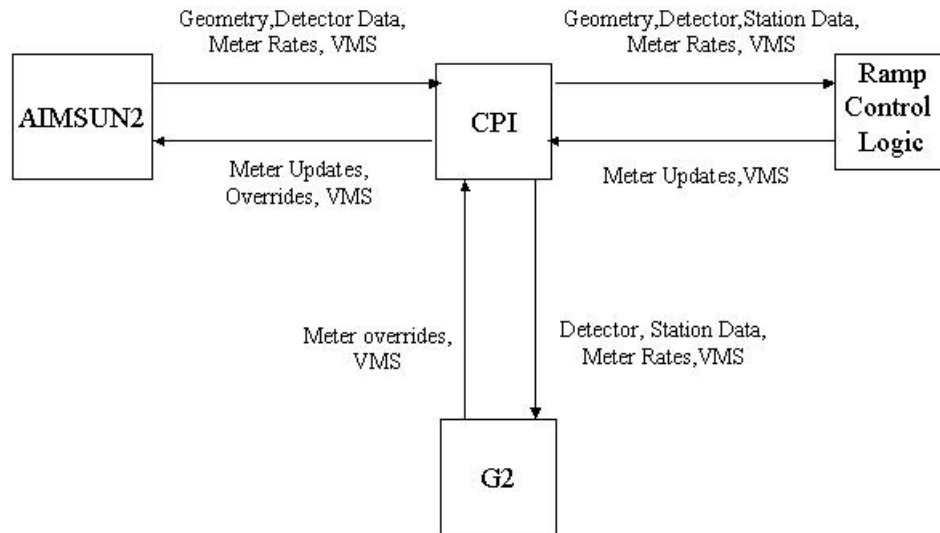


Figure 3.1. Global View of the AIMSUN2, CPI, Ramp Control Logic and G2

ramp control strategy with another without resorting to major changes to the code. Hence, using the CPI, the user can simulate different ramp control schemes (and potentially new ones) on the same network using AIMSUN2. The user can also optimize a particular strategy by simulating various network scenarios with different values of the parameters that are to be optimized.

AIMSUN2 is capable of communicating with an external application. The CPI enhances this ability by allowing operator interaction with the simulator and grouping the necessary information specifically needed for ramp control schemes. In addition, new functions have been added to allow easier access to the simulator. This makes the job of the end users easier because they now have more tools available to integrate their ramp control logic with the simulator. Specifically, in the CPI, the notion of detector stations was added, as most of the current ramp control strategies require measurements over all lanes of the mainline instead of lane by lane. Additionally, the user is now able to define the collection rate of measurements, which can now be different from the one specified in the simulator. For example, the simulator may collect lane-by-lane detector data every 30 seconds but the user's ramp control logic could request and receive 5-minute detector station data.

With the original external interface, the user's logic had to be designed specifically for the network under consideration. With the CPI, the user may access information at run-time about the road geometry, traffic detection, and control devices as well as their mode of operation. This allows the user to design logic in a more generic way and be able to use it on a variety of different networks. Finally, the implementation allows for customized output to be saved including information specific to the operation, effectiveness, and general performance of the control logic. For example the users can get the ramp rates implemented by the simulator and compare them with the actual rates to determine how close the system is to reality. The detailed design and the flow of control between the CPI and other components are shown in Figure 3.2.

The (green) circles with numbers 1 to 16 in the figure 3.2 indicate the flow of control between AIMSUN2 and the CPI and between the CPI and the external ramp control algorithm. To simplify the figure, the communication from G2 was not shown. The prefix "circle" is omitted while describing the communication; hence 1 refers to circle with number 1 inside the figure. The first communication call to be used is 1, which invokes the function `GetExtInit`. In this function the input data such as the frequency at which traffic data needs to be obtained from the simulator by the CPI and the frequency at which control needs to be sent to the external ramp control logic are read from an input text file. The data structures needed by the CPI such as the detector maps, station maps and meter maps are created. These data structures are used to collect the traffic data from the simulator so that it can be passed on to the ramp control algorithm or G2 at user defined time intervals. Also, the communication between CPI and G2 (if necessary) is set up in this function and it remains open through out the simulation period. The next communication call i.e., 2, invokes the function `USER_INITIALIZE`, which in turn may create its own data structures needed by the ramp control logic.

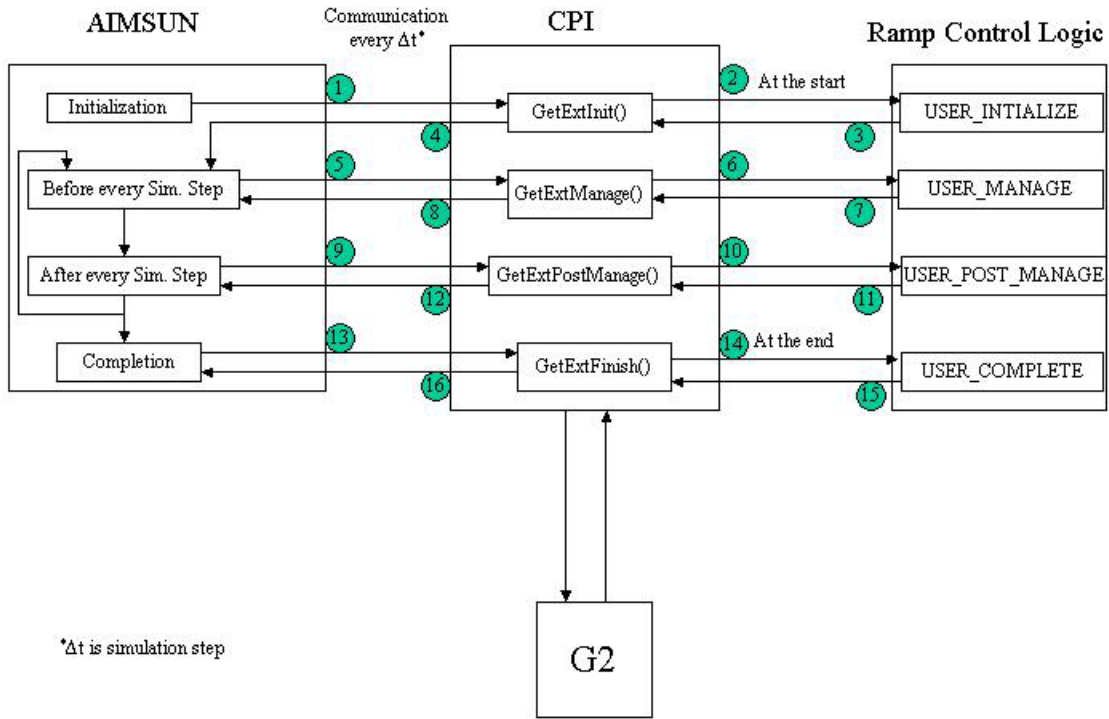


Figure 3.2. Detailed interaction of CPI with the simulator and the Ramp Control Algorithm

The USER_INITIALIZE function returns the default ramp metering rates with which the ramp meters should operate at the start of simulation which are then passed on to the simulator by the CPI (through the communication call 4).

Before every simulation step, control is transferred to the function GetExtManage. If it is time to update the CPI data structures, they are updated by accessing the required data from the simulator. If it is time to pass control to the ramp control algorithm the function USER_MANAGE is invoked which in turn may obtain some traffic data from the CPI. The ramp control logic may calculate new ramp metering rates that are to be implemented and passes them on to the CPI (through the communication call 7), which are then sent to the simulator. Data is also sent to G2 based on the requests made by G2. Before the control goes back to the simulator, any communication from G2 is processed. For example, if the operator decides to override a particular ramp, a flag is set indicating that the ramp is in the override mode and that the

algorithm can no longer control the ramp. Only the operator can remove the ramp from the override mode so that the control of that ramp returns back to the algorithm.

After the completion of a simulation step, the control is passed to the function `GetExtPostManage` to allow the CPI and the users to complete any tasks that are needed. No functionality has been added in this step, except that control is passed to the function `USER_POST_MANGE` at user specified time intervals. At the end of the simulation, the function `GetExtFinish` is invoked which helps the CPI to destroy (de-allocate memory) the created data structures. The function `USER_COMPLETE` is invoked which helps the user in destroying the data structures that might have been created for the ramp-metering algorithm. The communication with G2 is then terminated and control is returned to the simulator.

The CPI has been developed on Visual C++, using Microsoft Foundation Classes (MFCs) and it runs on Windows 95/NT machines. It has the form of a Dynamic Link Library (DLL) that the user attaches to the simulator.

Traffic Demands Creation Application (TDCA)

As mentioned in Chapter 2, the Mn/DOT algorithm measures real time volume and occupancy every 30 seconds and uses them to calculate new ramp metering rates that are to be implemented for the next 30 seconds. Hence there is a need to create input for the simulator with the flexibility of changing it every 30 seconds. The primary objective was to develop a program that minimizes the effort required to create the input files needed by the simulator. The traffic demands creation application described in this section substantially reduces the time and effort needed to create the traffic demands. For instance, for a 50 km (30 mile) section of I-35W in Minneapolis it now takes only a few minutes to create the input demands as opposed to several days.

In the Twin Cities freeway network more than 3000 loop detectors are currently installed measuring lane volume and occupancy. This data is collected every 5 minutes by the Mn/DOT's TMC and saved in binary format. The detector data is extracted and stored in a relational database. The result is historical five-minute lane volume and occupancy data for more than 800 locations in the freeway network. The detector location and identity information is also stored in

the models and is used by the TDCA to create the input demands. The user provides the application with the name of the network, date for which demands are to be created and the starting and ending time of the simulation. The user defined time period is divided into five-minute intervals, and the program automatically scans the network and identifies all boundary locations; this includes all the entrance sections and sections from which there is an exit(s). For each such section the applications reads the identities of the detectors that influence the flow in that section. For every time interval, the system retrieves from the sensor database the volumes of all detectors identified and creates the input flows or the turning percentages or both for all the boundary locations for all vehicle types defined in the model.

In the initial versions of the TDCA only fixed vehicle composition was used i.e., the ratios of the vehicles entering and exiting from the network was assumed to be the same across all boundary locations. But during data collection it was observed that the vehicle compositions (especially the percentage of trucks) vary significantly from one entrance location to another. Hence the program has been enhanced so that vehicle compositions can now be customized for every entrance section. Another important reason was that during the initial testing of the simulator, it was found that the simulation results were sensitive with respect to the percentage of trucks. Hence it was necessary for the input demands to be as close to reality as possible.

The TDCA has been of immense help in this study. During the course of the study it was found that there were some errors in the detector data used for simulation. If not for the TDCA it would have been next to impossible to create the input files again. The importance of this application also has to be viewed in the context that if the data were to have been entered manually, it would have lead to inevitable errors and, hence, more time would have been needed to determine whether the data entered is correct.

Emulation of Mn/DOT's ramp control strategy

Having developed the CPI, the next step in meeting this study's objectives was to implement the Mn/DOT ramp control strategy described in Section 2.3. For this, Mn/DOT's engineers were contacted to resolve some finer issues such about the working of the algorithm in the field such as the calculation of red times for different ramp metering rates, etc. The algorithm was

implemented in the simulator by developing the necessary code in C++ using MFCs. Zone control was implemented first, followed by occupancy control. To test the algorithm's implementation, a five-minute simulated data was taken at random and ramp-metering rates were calculated manually based on Mn/DOT's logic and were compared with the rates produced by the algorithm implemented in the simulator. Care was also taken to check that the algorithm produced not only the correct final ramp metering rates but also correct output at each and every stage of the rates calculation.

“Look Ahead” Model

During the course of an earlier study, which aimed in evaluating the feasibility of freeway simulation with the selected microscopic simulator, a serious problem with the lane-changing model was detected. Specifically, the previous lane changing model allowed for very short areas where vehicles were performing route choice related lane changes. This problem was especially evident on short weaving sections, a geometry frequently encountered in the Twin Cities. Since the new improved lane changing model, called “look ahead” is substantially better than the original the software developers incorporated it in the current production version of the software and guarantee its maintenance through subsequent version as long as it remains the best solution. The details of this improvement are described in detail in section 2.5.9.1 since it is now an integral part of the simulator.

4 DESIGN OF EXPERIMENT

4.1 Introduction

The need driving this study was to develop a scientifically sound procedure to determine the effectiveness and validity of Mn/DOT's current and future ramp control strategies (which includes quantifying the benefits of ramp metering) and prepare a platform to test any changes to the algorithm.

As described in earlier chapters, this study succeeded in developing an automated simulation tool for testing, and comparing (new) ramp control strategies. This tool has been integrated with Mn/DOT's ramp control strategy. In this chapter the employment of the resulting simulation package to test Mn/DOT's control strategy is described. The following steps towards this are as follows:

- Define the requirements of the test sites to be selected in order to provide enough information for an accurate evaluation of the ramp metering strategy.
- Select the locations (freeways) that offer a balance between the aforementioned need for accuracy and the time and resources available.
- Define the necessary geometry and traffic information needed in order to build realistic simulation models of the selected freeways.

Collect the aforementioned data and solve any problems due to erroneous or insufficient information. Such information includes:

- Geometry of the selected sites.
- Traffic measurements on selected locations of the freeway network.
- Initial values of the simulation parameters (vehicle/driver characteristics, etc.)

4.2 Site Selection

The Twin Cities metropolitan area has 500 miles of freeways of which 210 miles are metered. There are about 430 ramp meters of which 350 are metered using the Mn/DOT's zone control algorithm, the rest being operated on a pre-timed basis. As mentioned in the introduction and the literature review, earlier studies by Mn/DOT suggest that ramp metering is beneficial and results in increased speeds on the mainline and reduced accident rates, apart from additional benefits due to lower vehicle emissions [6, 9, 14]. Also, when the freeway is near capacity, ramp metering discourages short freeway trips by diverting freeway bound traffic to the arterials.

Determining the effectiveness of the Mn/DOT ramp metering would require collecting extensive data (depending upon the MOEs) such as vehicle speed, distance traveled, wait time at signals or meters, fuel consumed, vehicle emissions, number of stops and others. The data needs to be collected with and without ramp metering for the complete set of freeways, arterials and surface streets in the region being metered. In practice, this process is simplified through macroscopic data collection i.e., by taking average counts or sampling. Some MOEs such as number of stops, fuel consumption, emissions and others require detailed marking of all vehicles entering and leaving the system. Clearly, collecting such enormous data is expensive and impractical. This is the reason why simulation is the most accepted scientific alternative as the movement of all vehicles can be tracked in the network, and, hence adequate data can be obtained for analysis. There are two problems with this approach, however. First, modeling the complete Twin Cities freeway network along with the surface streets would be extremely difficult given the constraint of limited resources such as time and human-personnel. Secondly, detailed volume, geometric and other traffic data, especially at all intersections and surface streets, needed by the simulator, are missing. Currently, Mn/DOT has extensive traffic volume and occupancy data for only freeways and ramps. Hence, this study had to be curtailed to freeways and ramps only. It should be noted that the simulator could model both freeways and surface streets with equal ease. Plans are underway to do simulation studies of the I-494 corridor.

A number of test sites were considered and three of them were selected for testing the effectiveness of ramp metering. This chapter gives a detailed description of the selection criteria

used. The salient features of the test sites are described and finally, diagrams of the test sites and a table summarizing their properties are presented.

4.2.1 Site Selection Criteria

Through discussions with the Mn/DOT engineers at the TMC, three test sites were selected having geometric properties and traffic characteristics that are representative of the Twin Cities freeway network to the extent possible. The following criteria were used in choosing the sites:

Minimum length (number of zones)

As mentioned in the description of the Mn/DOT control strategy, the algorithm assumes that the freeway is divided into zones for volume and occupancy control. To implement volume control, a test site is needed that includes at least one zone. To implement occupancy control, for each ramp, the Mn/DOT algorithm uses occupancies from all stations located within two to three miles downstream of the ramp. This generally requires the use of detector data beyond a zone boundary. Hence, at least two zones are usually needed to implement occupancy control. Also, having two zones allows the study of inter-zonal effects (if any). To be on the safe side, sites that have a minimum of three zones have been chosen.

Mix of geometries representing typical Twin Cities freeways

The Twin Cities freeway network includes typical geometric configurations such as weaving sections, lane drop locations, high volume entrance ramps, high volume exit ramps, etc. An effort was made to select sites, which included most of these features but also avoided those that had too many weaving sections and complex geometric sections within a short span. Also, the freeways can be classified into one of the following general categories: radial, circumferential, central business district connector i.e. connecting the Minneapolis and St. Paul downtown districts. Sites representing each of these categories were selected.

Different levels of traffic congestion

One of the objectives of this research was to determine how the Mn/DOT algorithm performs under different traffic congestion levels. Hence, test sites were selected representing low traffic, moderately heavy traffic and very heavy traffic.

Test site(s) with alternate routes

One of the goals of this research is to have the flexibility to expand a test site(s) to include arterials associated with that freeway for future research. This would allow a study of the effects of diverted freeway-bound traffic to adjoining arterials due to ramp metering and evaluation of the impacts of ramp control on the corridor as a whole. Through information about the traffic diversion due to metering is not currently available, one site with alternate routes was selected so that corridor simulation can be done when data becomes available.

Upstream and Downstream boundary conditions

For simulation purposes and for the application of the Mn/DOT algorithm, the upstream and downstream boundaries should have free flow conditions. The test sites selected were such that as far as possible this requirement is met.

4.3 Data Collection

One of the factors that had to be considered was the feasibility of collecting reliable data. The intent was to choose sites, which had most of the data collection equipment in working condition so that required repairs will be minimized. Particularly, all the detectors on the entrance ramps, exit ramps and designated upstream stations of zones needed to be in working condition as the Mn/DOT algorithm relies heavily on these detector measurements for periodically calculating new ramp metering rates. Also, as will be explained in the next chapter on data collection, personnel had to be allocated to perform manual data counts for estimating the real demand at the entrance ramps. This placed a restriction on the maximum length of the sites that could be chosen for modeling.

Common test site with other ramp metering research project

One of the goals of this project was to compare the effects of Mn/DOT metering using two different modeling approaches: microscopic and macroscopic. One of the test sites was chosen is common with other project.

4.3.1 Test Sites Selected

Based on the above selection criteria and in consultation with Mn/DOT engineers, a number of test sites were considered for testing by considering the Twin Cities freeway network. An example of a test site which was considered but discarded is the corridor of I-94 Westbound from Marion St. To TH 55 as the downstream end at the interchange with TH 55 is congested. Following such considerations, three test sites were selected that meet the aforementioned criteria shown in Figure 4.1. A brief overview of each site is given below. The test site's general geometric characteristics are presented in Table 4.1.

TH 169 Northbound

The first test site selected is a 20 km (12 miles) long section of Trunk Highway 169 in the northbound direction starting from the interchange with I-494 and ending at 63rd Avenue North (Figure 4.2). Most of the test site consists of two lanes with 10 weaving sections. It has 24 entrance ramps of which one is un-metered. The metered ramps include 4 HOV bypasses and two freeway-to-freeway ramps from TH 62 and I-394. The test site contains 25 exit ramps. The upstream and downstream boundaries are uncongested. Among the three test sites, TH 169 is the least complex in terms of the geometry and generally carries lower traffic volumes compared to others. It is a circumferential freeway.

I-94 Eastbound

The second test site is I-94 Eastbound beginning at I-394 and ending at the off ramp at 9th St. It falls under the category of central business districts connector because it connects the Minneapolis and St. Paul downtown districts. It is about 18 km (11 miles) long, carries a high volume of traffic and is often known to be congested during peak hours (Figure 4.3). In terms of complexity, it is more complex than TH 169 but lesser complex than the I-35W test site. The upstream boundary of the test section is an un-congested area just ahead of a tunnel near I-394

and the downstream boundary is relatively un-congested. It includes 19 entrance ramps and 14 exit ramps of which 4 are un-metered. It contains 6 weaving sections and has about 3 lane-drop sections. One of the unique features of this site is the section where I-94 merges with I-35E near downtown St. Paul. It was one of the most difficult sections to model but it provided an opportunity to study the interaction between the two freeways.

I-35W Southbound

A 24 km (15 mile) long section of I-35W Southbound was selected as the third test site. It begins at downtown Minneapolis and ends at the interchange with TH 13 (Figure 4.4). It includes 22 entrance and 20 exit ramps that are controlled during peak hours. Four entrances are freeway-to-freeway ramps, carrying very high volumes in the range of 1,200 veh/hr with long spillback queues. The geometry includes six weaving sections and also has three lane-drop sections. The test site is divided into three zones and has three bottleneck locations. It also has a single HOV lane from the I-494 interchange to TH 13 that is about 10 km (6 miles) long.

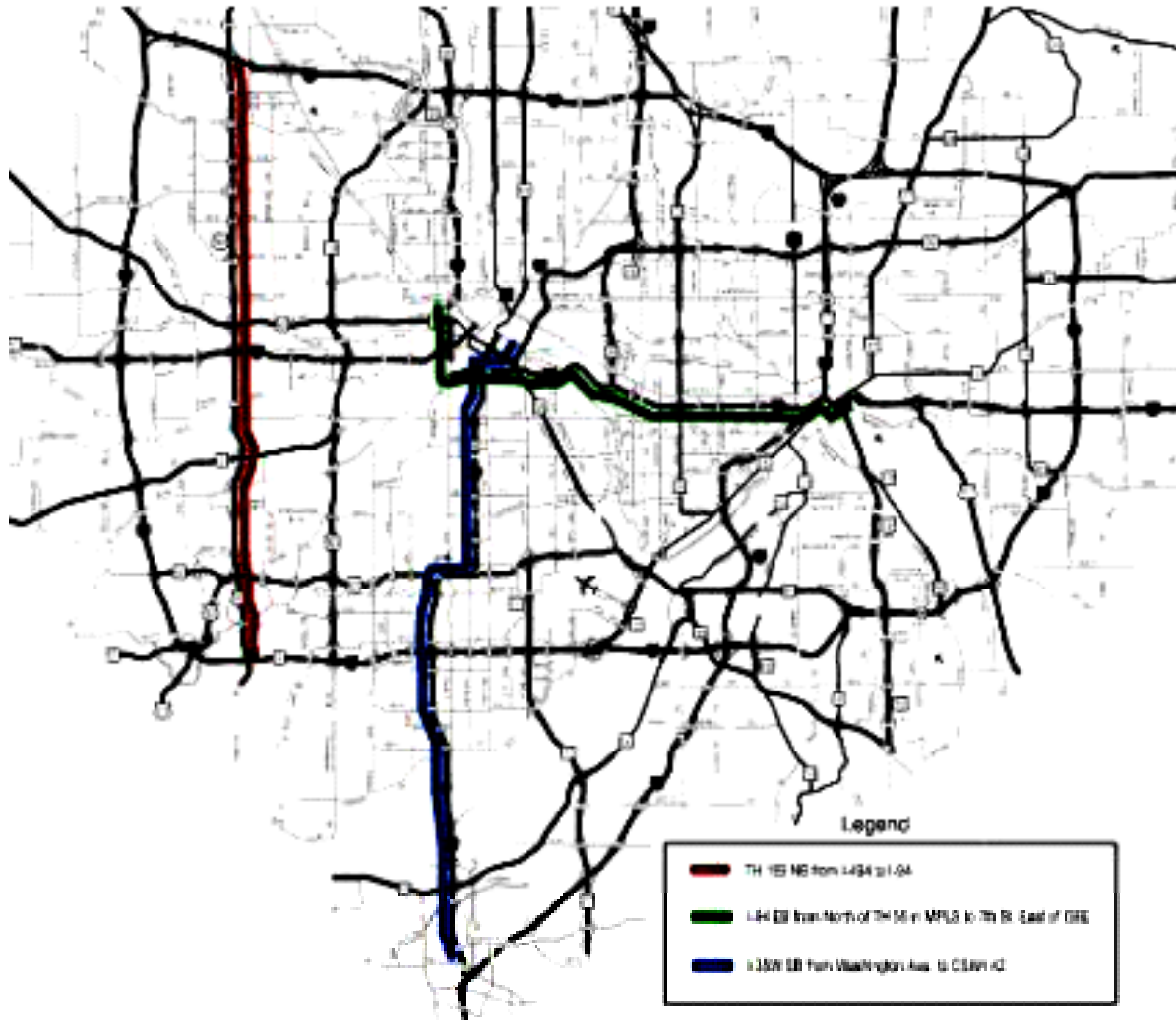


Figure 4.1. Twin Cities freeway network with the three test sites

Table 4.1. Geometric Properties of Test Sites

Characteristic	TH 169	I-94	I-35W
Direction	Northbound	Eastbound	Southbound
Length (km)	20	18	24
Freeway Category	Circumferential	CBD Connector	Radial
Upstream Boundary	I-494 interchange	I-394 interchange	Downtown Mpls.
Downstream Boundary	63rd Avenue North	9th St. Off ramp	TH 13 Interchange
Metering Period	AM and PM	AM and PM	AM and PM
Zones (names)	3 (3A, 3B, 3C)	3 (8E, 8F, 8G)	3 (6M, 6N, 6P)
Bottleneck Locations	3A: 7th Street	8E: Portland Ave.	6M: TH 62 Exit
	3B: 23rd Street	8F: Franklin Ave.	6N: I-494
	3C: Medicine Lake Rd.	8G: Marion St.	6P: 113th St.
Metered on-ramps	23	15	17
Unmetered on-ramps	1	4	4
Off ramps	25	14	20
Weaving sections	10	6	6
Lane drop locations	1	3	2
Geometric Complexity	Simple	Complex	Very Complex
Traffic Levels	Low	High	Very High
Alternative Routes (Non Freeway)	No	Yes	No
Sensors	Good	Good	Bad

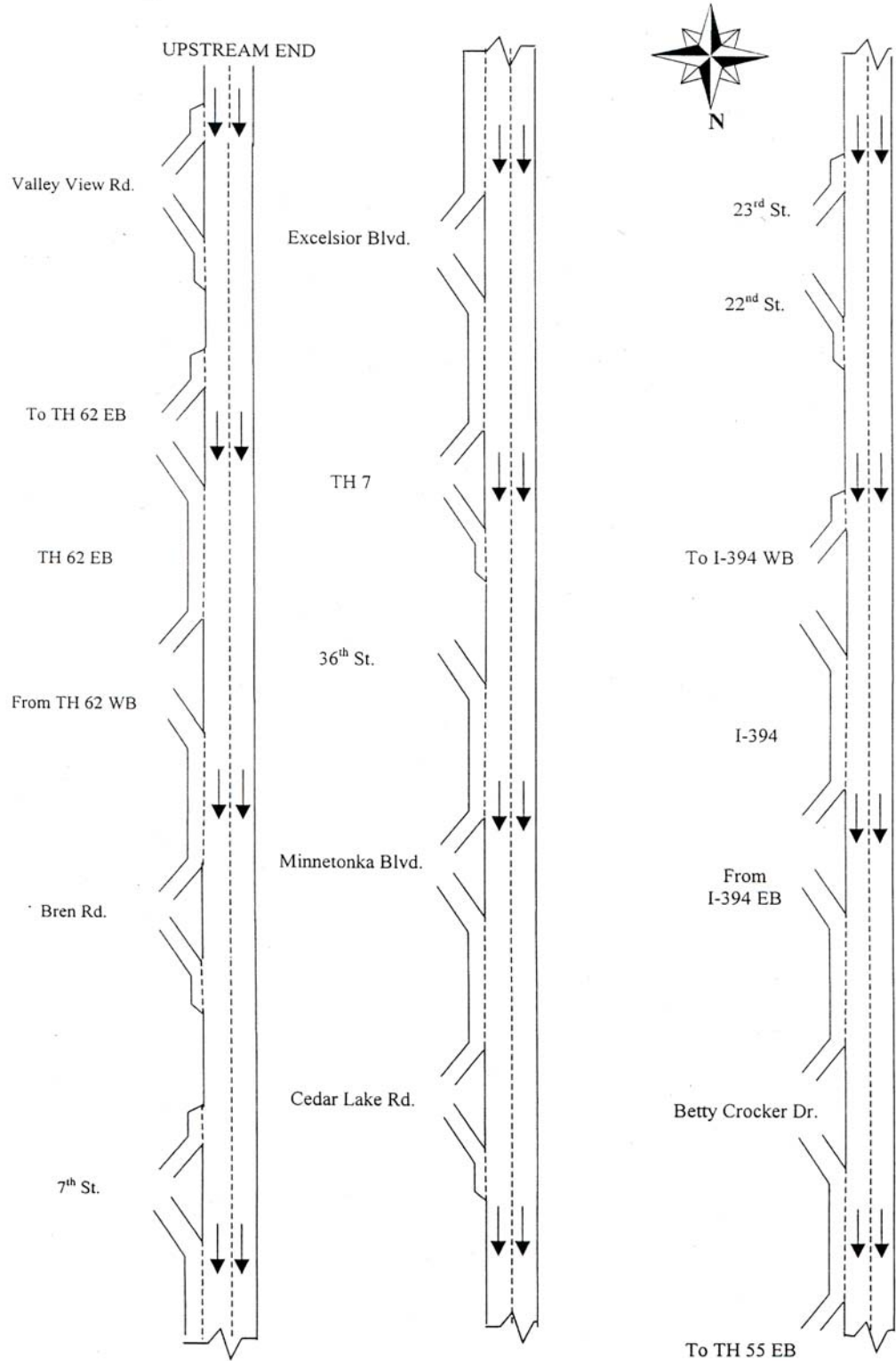


Figure 4.2. TH-169 Test Site

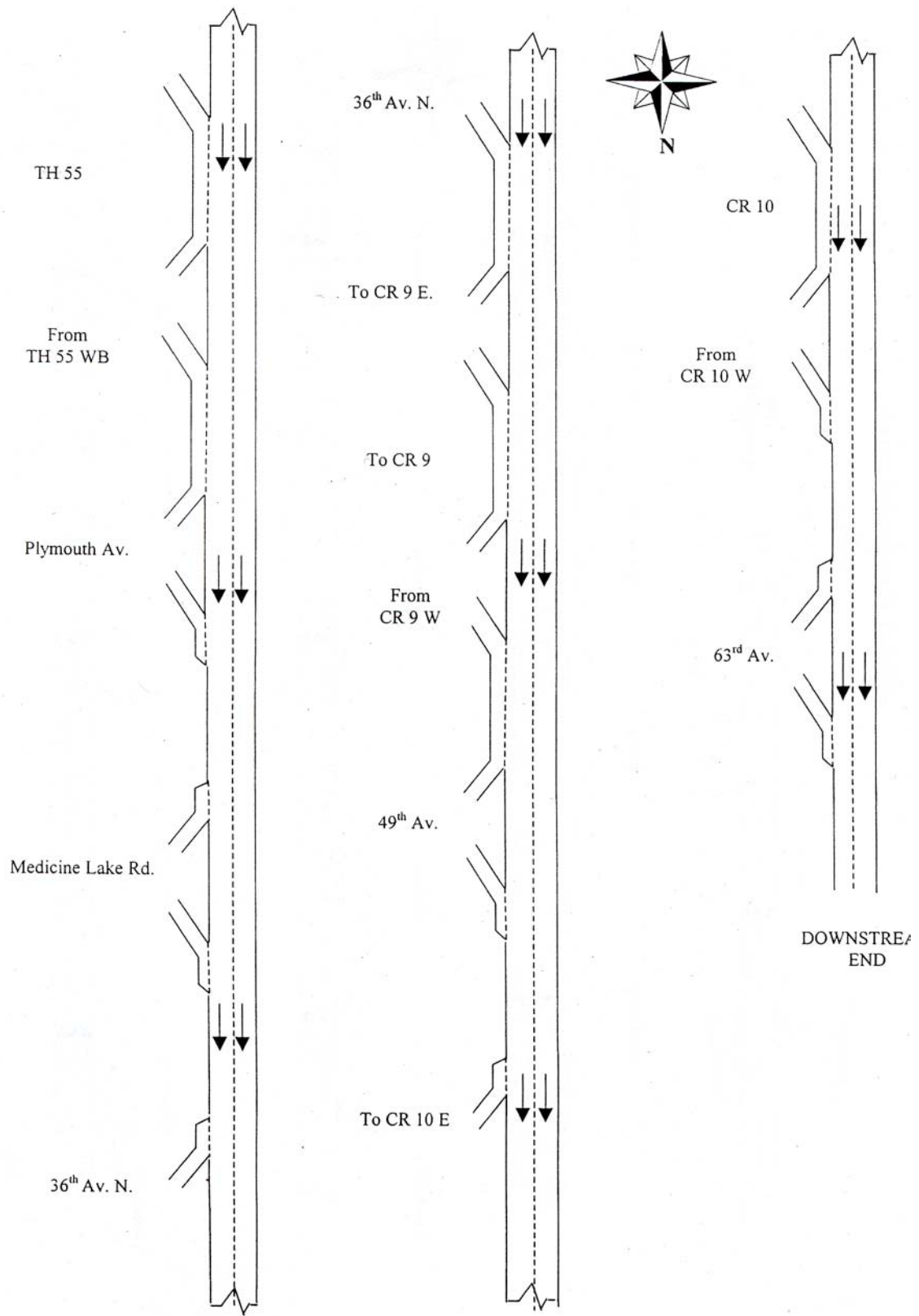


Figure 4.2. TH-169 Test Site (cont.)

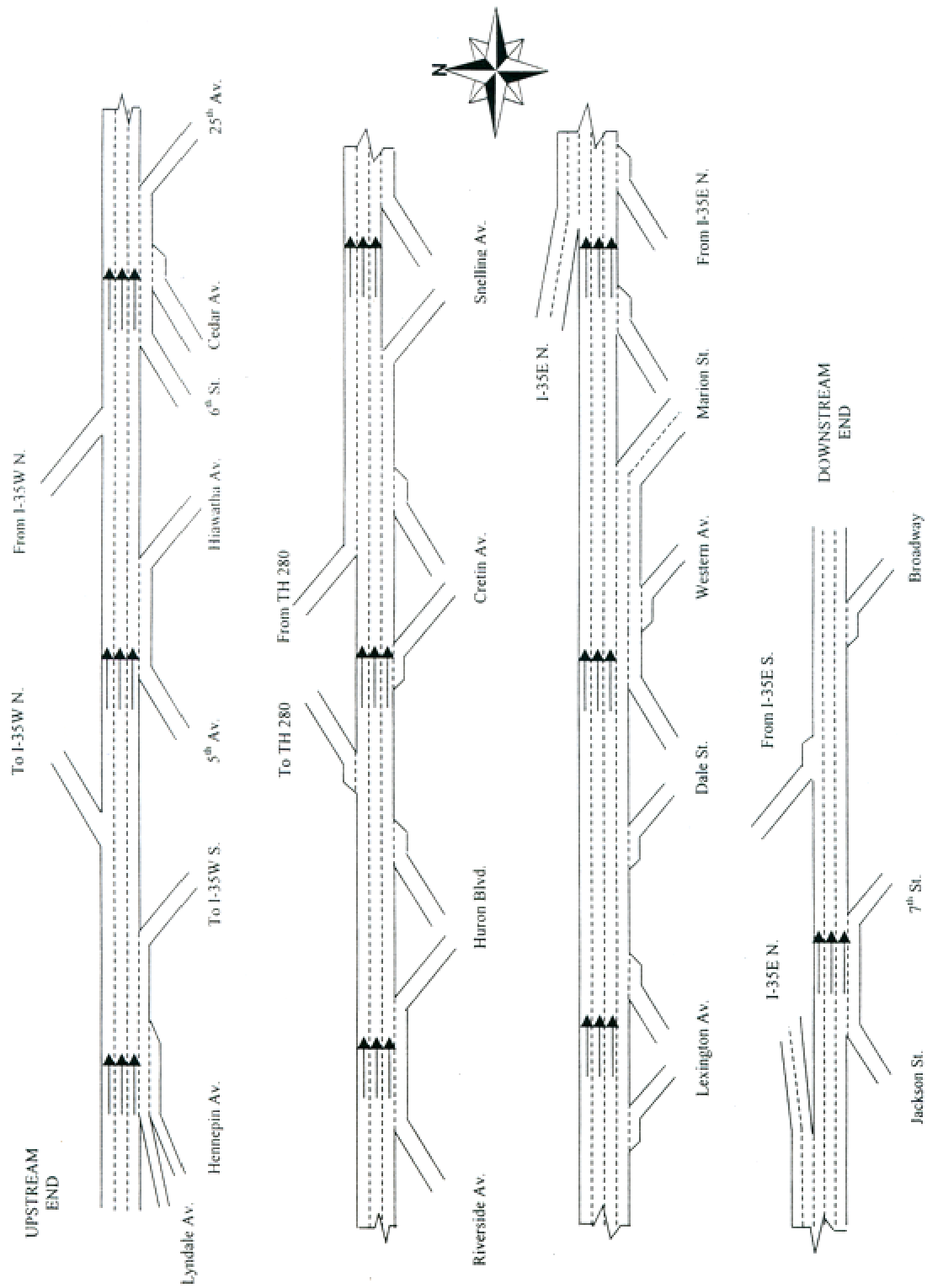


Figure 4.3. I-94 Test Site

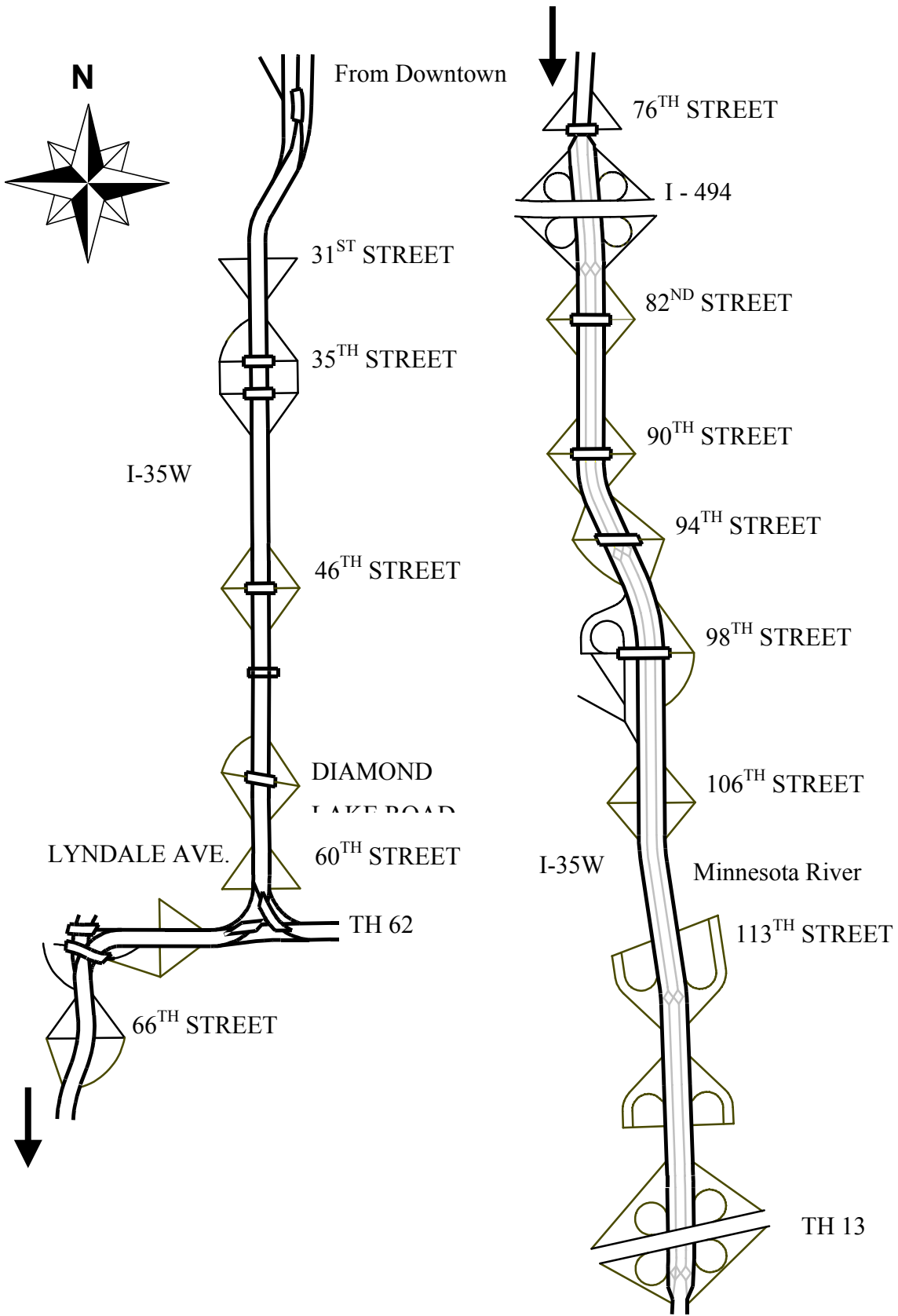


Figure 4.4. I-35W test site

4.3.2 Data Collection

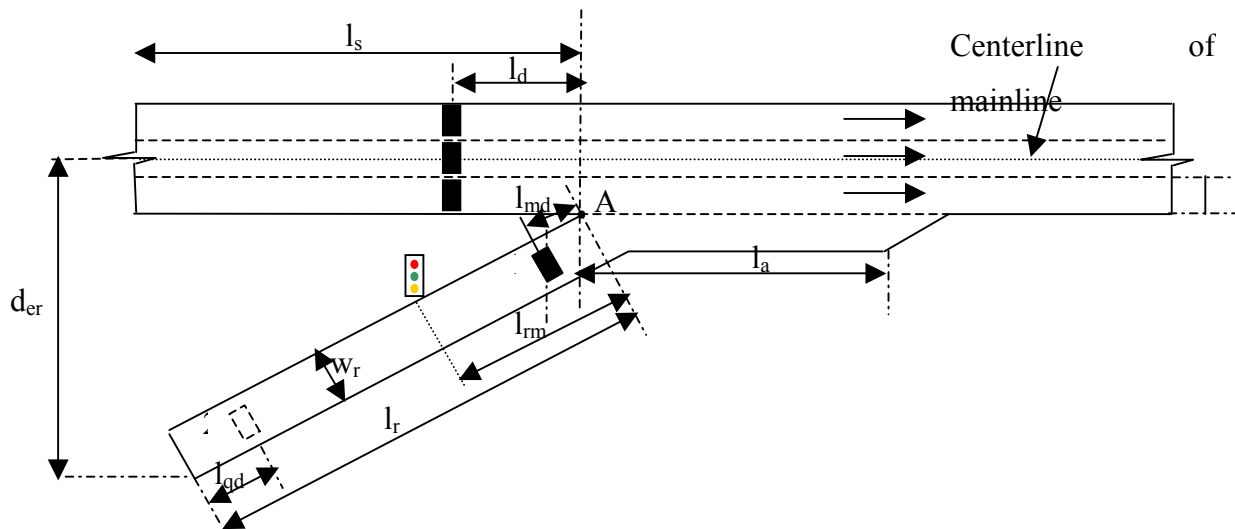
To simulate traffic flow in a network, two types of information are needed: geometric data to model the roadway and traffic data to simulate vehicles. The geometric data consists of the physical properties of the freeway such as the number of lanes, width of the lanes, length of the mainline between ramps, length of the entrance ramps, location of the detectors and ramp meters, etc., which is described in detail in this chapter. It was extracted from sources such as AutoCAD files, construction maps and aerial video of the test sites. The traffic data needed by the simulator includes traffic composition, entrance demands, exit percentages, vehicle and driver characteristics, among others. Unfortunately Mn/DOT is not collecting all of the required data routinely except volume counts and occupancies on the mainline detectors and merge detectors (detectors just downstream of a ramp meter). A description of the data usually collected by Mn/DOT is presented in section 4.3.2.1. As the simulator needs the entrance demands and traffic compositions at the entrance ramps which are not collected by Mn/DOT, they had to be collected manually or indirectly extracted as described in section 4.3.2.2 Vehicle and driver characteristics such as the acceleration and deceleration rates, fuel consumption and emission rates, reaction time, etc., were obtained from various sources which include the Highway Capacity Manual, Traffic Engineering Handbook, among others. These characteristics are presented in detail in section 4.3.2.4.

4.3.2.1 Geometric Data

In general, though the freeway geometry may be composed of different types of segments, they can be broadly classified into five types: entrance ramps, exit ramps, lane-add sections, lane drop sections, and weaving sections. It is assumed that the orientation of the freeway is known in the form of a centerline passing through the middle of the mainline as shown in Figure 4.5. In this study, the centerline was obtained from an AutoCAD file containing information about the freeway geometry. The geometric data needed for each of the different segments is presented in this section.

Entrance Ramp and Mainline section

For a typical freeway entrance ramp and mainline section shown in Figure 4.5, point A's location is calculated based on its distance l_s from a known upstream reference point. A typical upstream reference point may be the location of an exit ramp on the mainline. All distances are calculated with respect to point A. The information needed is as follows:



- 1: merge detector (present on every ramp)
- 2: queue detector (present on some of the ramps)

Figure 4.5. Typical freeway entrance ramp and mainline section

Number of lanes on the mainline (3 in the Figure 4.5)

Width of the lanes (w_l)

Distance of the mainline detectors from point A (l_d)

Length of the acceleration lane (l_a)

Number of lanes in the entrance ramp (1 in Figure 4.5)

Width of the entrance ramp (w_r)

Length of the entrance ramp (l_r)

Distance between the ramp meter and point A (l_{rm})

Location of the merge detector from point A (l_{md})

Distance of the end of the ramp from the centerline of the mainline (d_{er})

Location of the queue detector from the end of the ramp (l_{qd})

Exit Ramp

For a usual exit ramp shown in Figure 4.6, assuming that the location of point A (where the exit ramp actually starts) is known (as discussed in section 4.3), the data needed is:

Number of exit lanes (1 in Figure 4.6)

Length of the deceleration lane (l_{dec})

Length of the exit ramp (l_e)

Distance of the end of the exit ramp from the centerline of the mainline (d_{er})

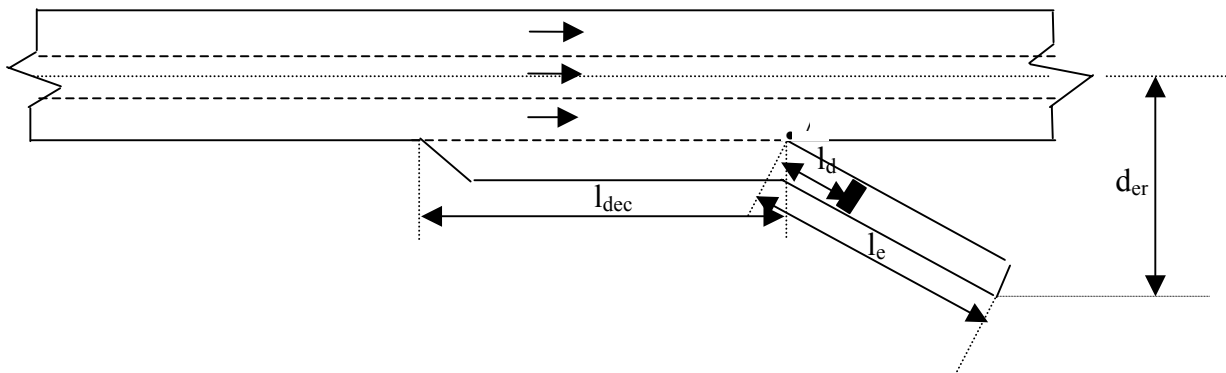


Figure 4.6. Typical freeway exit ramp

Lane-add Sections

For a typical lane-add section (for example shown in Figure 4.7), the data needed is the location of point A, where the additional lane begins, from a known upstream point.

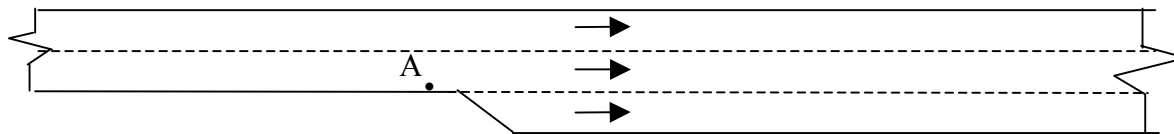


Figure 4.7. Typical freeway lane add section

Lane-drop Sections

For a usual lane drop section shown in Figure 4.8, the information needed is the location of point A, where the lane drop is complete, with respect to a known upstream point.

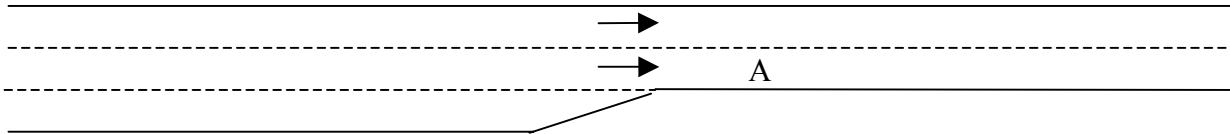


Figure 4.8. Typical freeway lane drop section

Weaving Sections

For a typical weaving section shown in Figure 4.9, other than the regular data that may be needed for an entrance and exit ramps, the extra information needed is the length of the weaving section (l_w).

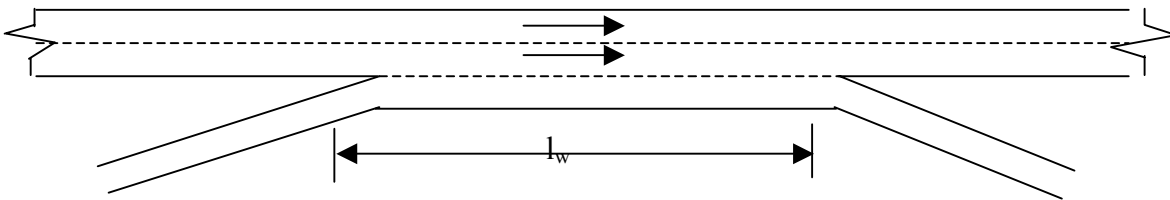


Figure 4.9. Typical freeway weaving section

As a starting point for modeling the test sites, AutoCAD diagrams of the freeway background, obtained from Mn/DOT, were used. The diagrams contain information about the orientation of the roadway geometry, lane markings, and the location of the traffic control devices such as detectors and ramp meters. Using this background as a reference, the data needed was extracted and the model was built in the simulator. The speed limits were obtained from Mn/DOT and added to the model. At that point, rough models of the test sites were available.

To determine whether the modeled network was accurate, some members of the research team from the University and Mn/DOT drove along the test sites. It was found that though most of the information is accurate in the AutoCAD files, it was not up-to-date at some locations, especially where there have been modifications to the original freeway geometry such as addition of lanes and HOV bypasses. There were also some uncertainties about the length of the acceleration and deceleration lanes. Construction plans were used to extract more details about the freeway geometry. To obtain the current freeway geometry, the State Patrol was requested to capture the test sites on video with one of their specially equipped helicopters. This video was very helpful in defining the test sites as close as possible to the real geometry. A screen shot of the video is shown in Figure 4.10.



Figure 4.10. Screen shot capture of the video taken by State Patrol

4.3.2.2 Traffic Data

For each test site, the traffic data needed are the demands and the traffic composition at each of the entrance ramps as well as the upstream end of the network. The turning percentages of the mainline volumes and the traffic composition are also needed at the exit ramps. In this study, following meetings with Mn/DOT engineers, it was decided that the vehicles would be classified into three types: cars, trucks and semi-trailers based on the criteria shown in Table 4.2.

Table 4.2: Vehicle Classification

Vehicle Type	Criteria
Passenger Cars	Less than 6 tires
Trucks	Single body and ≥ 6 tires
Semi Trailers	Double body units

For each test site, it was decided that the traffic data would be collected simultaneously at all the ramps and the mainline on three different days. Collecting data simultaneously at all entrance and exit ramps for a test site ensures that the input to the simulator is identical to that in the field. Also, if data were collected on three different days, the probability of the actual collected data deviating from the average arrival patterns for the test sites would be lesser rather than collecting on a single day.

An additional constraint on the data to be collected was that it should be incident free. This restriction was placed because at the time of data collection, Mn/DOT had an active mechanism to deal with incidents that includes broadcasting about the incidents on traffic radio and displaying the same on variable message signs. It can be safely assumed that this would have an effect on the drivers with some of them deciding to use alternate routes. As the number of drivers choosing alternate routes was not available, it was felt that it would be appropriate to do the study with incident-free data. Before going into how the data is collected for this study, a brief description of the data collected by Mn/DOT is given in the next section.

Traffic Data Collected by Mn/DOT

The traffic detection system currently used by Mn/DOT consists of two types: loop detectors and closed circuit television (CCTV) cameras. The detectors are 6' x 6' inductive loops that are embedded in each travel lane of the freeway and spaced at approximately one-half mile increments on the mainline. The detectors at each location are grouped into a station. Loop detectors are also placed on all entrance and exit ramps to measure the traffic entering and leaving the freeway. The detectors on the entrance ramp are located downstream of the ramp meter and are called merge detectors. Some of the entrance ramps also have queue detectors.

Each loop detector collects two types of data: volume and occupancy. Volume is a measurement of the number of vehicles that have passed over the detector. Each detector measures a lane volume and the sum of all detector volumes in a station measures the total traffic volume crossing that location. Occupancy is a measure of the percentage of time that a vehicle is located over the detector. Again, each detector provides the lane occupancy of the roadway but the station occupancy is an average of the detector occupancies at the station.

The second type of traffic detection is through visual surveillance via CCTV cameras. The cameras are usually mounted on 50-foot poles at one mile mainline spacing and on some entrance ramps. Operators at the TMC monitor the cameras to determine if the central computer is regulating the system appropriately. They also use CCTV cameras to verify and detect freeway incidents, assist in the timely arrival of motorist aid, and help minimize traffic congestion.

An example of a mainline station (442) having two detectors (1942 and 1943) and detectors on an entrance (1944) and exit ramp (1945) is shown in Figure 4.11.

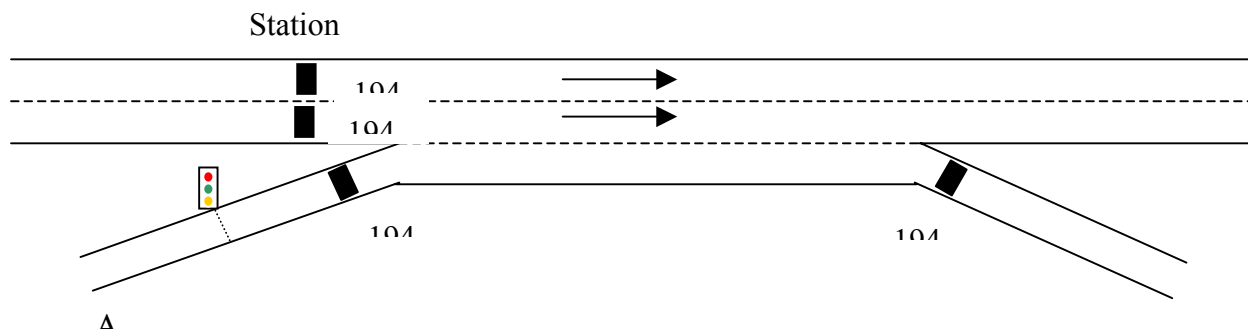


Figure 4.11. Typical mainline, entrance ramp, and ramp detectors

The detector station data is aggregated every 30 seconds and transmitted to the TMC whereas the individual detector data (including those on the entrance and exit ramps) is aggregated every five minutes and transmitted to the TMC. It should be noted that Mn/DOT's detectors couldn't determine the traffic composition.

Extra Traffic Data Collected

As mentioned earlier, AIMSUN2 needs the entrance demands and the traffic composition at the ramps as well as the upstream end of the network, which are not routinely collected by Mn/DOT. This data was extracted by looking at the video recordings of CCTV cameras and by stationing human personnel in the field. As the data collection involves stationing human personnel in the field, it was important to ensure that all the detectors were functioning properly before data is collected. For this, both detector and station data (previously collected during March 2000) were analyzed to identify any malfunctioning detectors. If only the station data is analyzed and if one of the detectors in a station is malfunctioning (a detector not counting any vehicles), it may not be observed in the station data. Incident logs obtained from Mn/DOT were used to discard those days of data, during which there were incidents on any of the three test sites. The remaining days were analyzed to check if there were any extreme values such as too high or too low volume counts or occupancy measures. Specifically, the data was checked for conditions where the volume count is zero but the occupancy value is non-zero or vice-versa. Also, a visual comparison of the volume and occupancy values was carried out for different days to see if the flow patterns were similar. A screen shot comparing detector data on two different days is shown in Figure 4.12.

Figure 4.12 shows volume counts on March 22nd and March 23rd from 10:00 A.M. to 8:00 P.M. for station 554 (near I-94 and 25th Ave.) comprising detectors 3134, 3135, and 3136 for. It is evident that the volume patterns on those two days were similar with the maximum counts recorded between 4:00 P.M. and 6:00 P.M. Hence, it can be inferred that the detectors were functioning well on those days.

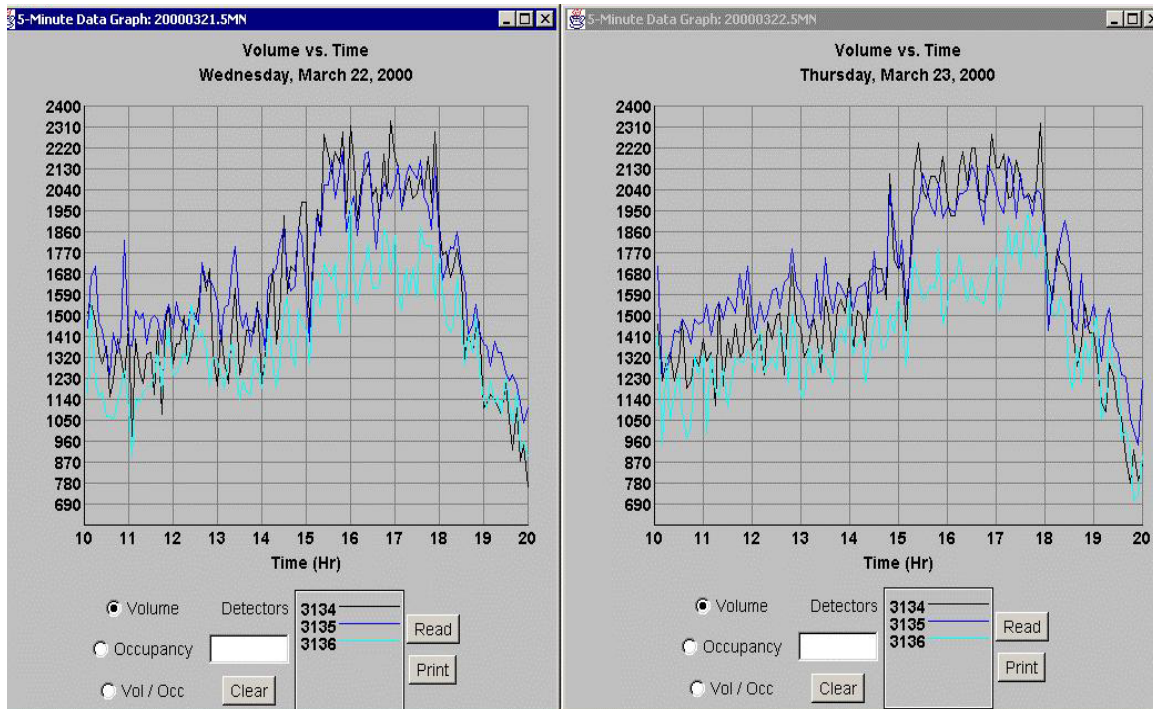


Figure 4.12. Visual Inspection of detector data of 3134, 3135 and 3136

Analysis of all the station data resulted in the identification of six malfunctioning detectors on the I-94 test site: four on the mainline, one on an exit ramp and one on an entrance ramp. The list of malfunctioning detectors was given to Mn/DOT so that corrective measures could be taken.

Entrance Demands and Traffic Composition

As can be seen from Figure 4.11, the detector on the entrance ramp is located downstream of the ramp meter and before the ramp joins the freeway. The merge detectors count the number of vehicles that were allowed by the ramp meters and hence those that have actually entered the network from entrance ramps. The data collected by the merge detectors is therefore called “metered demand”. The simulator needs real demand, i.e., the vehicle arrival rate at the upstream end of the ramp (location A in Figure 4.11) to calculate the delay at the ramps.

On some of the ramps, Mn/DOT uses queue detectors located upstream of the ramp meters to measure the length of the queues. In such cases, the queue detector counts give the vehicle arrival patterns at the ramps. But it should be noted that most of the ramps do not have usable

queue detectors and hence Mn/DOT cannot measure the arrival rates at those ramps with the current equipment.

On those entrance ramps that do not have any queue detectors, the following approach is adopted to estimate the arrival rates. Assuming that the number of the cars waiting in the queue at the ramp meter can be found every five minutes, the arrival rate is calculated based on the following formula:

For time period t_n ,

$$\text{arriving volumetn} = \text{merge-detector-voltn} - (\text{cars-in-queueetn} - \text{cars-in-queueetn-1})$$

The number of cars in the queue was calculated manually every five minutes. If the ramp could be observed through one of the surveillance cameras, then the number of cars in the queue was recorded every five minutes (counting the cars in the queue was simplified by noting the number of cars it takes for the queue to reach certain landmarks on the ramp). Otherwise, personnel were sent to the ramp to count the queue lengths every five minutes. With the last two methods, the vehicles types were also noted simultaneously so that they could be used to calculate the traffic composition. At the end of each day, consistency checks were performed by comparing the arrival volumes with the merge detector volumes (they should be equal). Only the vehicles present on the ramps were counted as part of the queues. If the queues spilled on to the surface streets (this scenario was encountered at the 6th Street ramp on the I-94 test site), they were discarded because it was difficult to estimate the number of vehicles in the queue that are trying to enter the freeway in those situations.

Exit Turning Percentages

AIMSUN2 requires the turning percentages of the mainline volumes at exit ramps and the traffic composition. The loop detectors placed on the exit ramps can only record the number of vehicles that have left the freeway and not the type of vehicle. Using this information, the percentage of the mainline vehicles that take the exit can be calculated. For example, for the exit ramp shown in Figure 4.13.

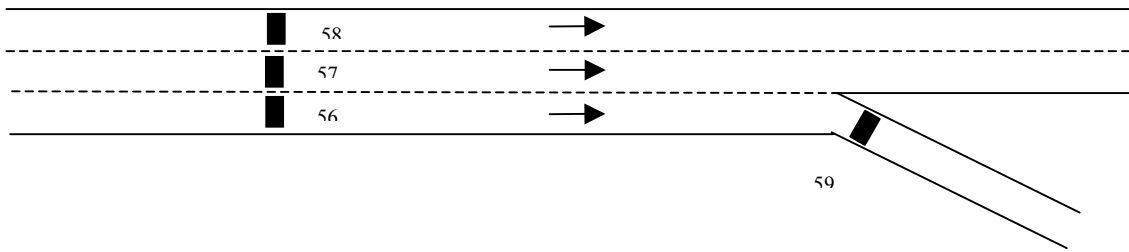


Figure 4.13. Mainline station detectors and exit ramp detector

Exit turning percentage during a five-minute interval = output-volume/input-volume where,

Output-volume = volume recorded by the exit ramp detector during the five-minute interval (counts recorded by detector 59)

Input-volume = volume recorded on the mainline station just upstream of the exit ramp during the five-minute interval (sum of the volume counts recorded by detectors 56, 57, and 58)

For some exit ramps where the mainline upstream detector station is much further away than the mainline downstream station, the latter station was used to calculate the input flow rather than the former. In such situations, the input flow is the sum of flows on the mainline downstream station and the exit ramp detector.

Ideally, the traffic compositions at the exit ramps are required by AIMSUN2. Collecting this data would have required human personnel to be physically present at the exit ramps. Alternatively CCTV cameras could be employed to collect the data. Due to the constraints on the number of personnel that could be assigned for data collection and the limitations on the CCTV cameras that can observe the exit ramps, this data was not collected and it was assumed that the turning percentages remain constant for all vehicle types at an exit. For example, if the percentage of the mainline traffic taking an exit is 10%, then it assumed that 10% of cars, 10% of trucks and 10% of semi trailers from the mainline take the exit ramp.

At the time of choosing test sites it was thought that data could be collected on all the test sites. But after a close estimation of the human personnel needed for manual data collection, it was decided not to collect data on the I-35W test site. Plans are underway to use the I-35W test site

for alternate experiments that do not need manual data for e.g., the data gathered during the ramp holiday can be used to test the effectiveness of the ramp metering. The dates for which the data was collected for each test site and the dates for which the data was discarded are shown in Table 4.3. Due to time and manpower constraints, data was not collected on the I-35W test site.

Table 4.3. Data collection dates and discarded data

Test Site	Date	Incidents	Result (Used/Discarded)
TH-169	March 21, 2000	No	Used
	March 22, 2000	No	Used
	March 23, 2000	No	Used
I-94	March 29, 2000	No	Used
	March 30, 2000	No	Used
	April 3, 2000	No	Used
	April 4, 2000	Yes	Discarded

Once the actual data was collected for the days shown in Table 4.3, it was checked to make sure there were no unknown malfunctioning detectors on those days.

Traffic Signal Timings: Platoon Simulation

One of the benefits of ramp metering is that vehicle-platoons released by signals upstream of the ramp are broken up by the ramp meters, thereby allowing only one vehicle at a time to merge with the mainline traffic. Without ramp metering, a platoon of vehicles trying to merge with freeway traffic, which is nearing its capacity, may cause a breakdown of traffic flow. Hence, platoons had to be simulated to find their effects on mainline traffic. This was done by introducing traffic signals into the model that required measurement of the traffic signal timings at the intersections upstream of the ramps and incorporating them into the model. It should be

noted that simulating platoons does not extend to freeway-to-freeway ramps, as intersections do not exist at these locations.

Vehicle and Driver Characteristics

AIMSUN2 needs a number of parameters such as vehicle and driver characteristics, car-following model parameters, etc., for model calibration. A literature survey was done to estimate the values of these parameters used by other simulators/models so that they can be refined further during the calibration. A brief description of these parameters, the sources from which they were obtained and the values used in this study are presented in this section.

4.3.2.3 Vehicle Characteristics

Every vehicle type modeled in AIMSUN2 needs the following set of parameters:

Length: The length of the vehicle in meters or feet. Length has a direct influence on traffic modeling, as vehicle length is taken into account in all vehicle behavior models.

Width: The width of the vehicle in meters or feet. Width is used only for graphical purposes. It does not have any impact on modeling of traffic.

Maximum Desired Speed: The maximum speed that a particular type of vehicle can travel anywhere in the network.

Maximum Acceleration: The maximum acceleration that a vehicle of a particular type can develop under any circumstances at any location in the network.

Normal Deceleration: The maximum deceleration that a vehicle can develop under normal circumstances.

Maximum Deceleration: The maximum deceleration that the vehicle can develop under special circumstances such as sudden braking etc.

Give-Way Time: The maximum time a vehicle waits prior to accepting shorter gaps to change lanes.

Speed Acceptance: Determines whether the vehicles moving on a particular section of the network try to adopt a speed higher or lower than the speed limit for that section.

Minimum Distance between Vehicles: The distance a vehicle maintains between itself and the preceding vehicle when stopping.

Guidance Acceptance: The probability that a driver will obey a recommendation such as one displayed by a variable message sign or a vehicle guidance system.

For each vehicle characteristic, AIMSUN2 needs the mean, deviation, minimum and maximum values. Data for the parameters was collected from the following sources:

AASHTO Policy on Geometric Design

Highway Capacity Manual

Traffic Engineering Handbook

CORSIM's user guide

Institute of Traffic Engineers electronic library.

Default values used by AIMSUN

Values used by VISSIM for American roads

From the aforementioned sources, the maximum and minimum values, where available, were selected for length, desired speed, maximum acceleration, maximum deceleration and normal deceleration. The objective was to find the range of values for each vehicle type that best suited the study's needs. For each vehicle type, the search was further divided into five categories:

Cat1 : minimum value

Cat2 : minimum + 25%(maximum – minimum)

Cat3 : minimum + 50%(maximum – minimum)

Cat4 : minimum + 75%(maximum – minimum)

Cat5 : maximum value

Hence, each vehicle type resulted in five different categories:

Car: CarCat1, CarCat2, CarCat3, CarCat4 and CarCat5

Truck: TruckCat1, TruckCat2, TruckCat3, TruckCat4 and TruckCat5

Semi-Trailers: SemiCat1, SemiCat2, SemiCat3, SemiCat4 and SemiCat5

As the majority of the vehicles on the freeways are cars, their characteristics have a higher impact on the model than those of the trucks and semi-trailers. Therefore, as a first try, it was

decided to keep the trucks and semi-trailers category constant, while all the five car types were tested. It was also agreed that the mean values would be better guesses than the extreme values. Hence, the first configuration tested was <CarCatX, TruckCat3 and SemiCat3>, where X takes values from 1 to 5. The other parameters, namely, speed acceptance, minimum distance between vehicles, give way time and guidance acceptance were kept constant between the simulation runs as they were not expected to change much with the physical properties of the vehicles.

Approximately ten simulation runs were performed for each of the above configurations and the correlation coefficient (ρ) values were calculated for the simulated mainline station volumes vs. actual volumes recorded in the field. It was found that the configuration with CarCat2 had the maximum correlation coefficient. The results are shown in the Figure 4.14. After the car category was fixed (CarCat2), simulations were carried out for different categories of trucks. The results are shown in Figure 4.15. At the third stage, the semi-trailer categories were varied and the results are presented in Figure 4.16. It was found that the configuration <Cartype2, Trucktype3, Semitype3> gave the best correlation coefficient value.

It should be noted that the vehicle characteristics selected for the above configuration were not fixed. They were just a starting point for the calibration and some of the values have been modified (as will be explained in chapter 8 on calibration). The vehicle characteristics obtained at the end of calibration are presented in Tables 4.4, 4.5 and 4.6.

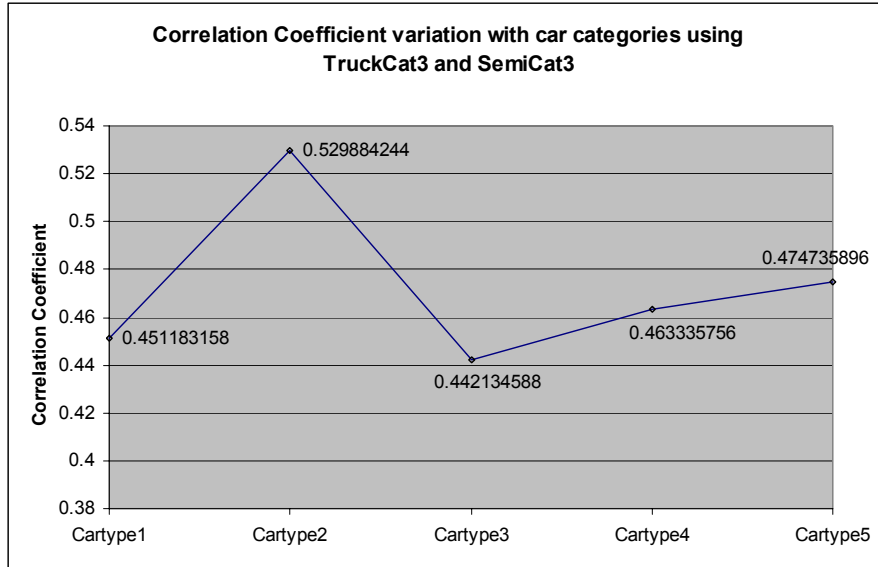


Figure 4.14. Correlation Coefficient variation with different car categories and TruckCat3 and SemiCat3

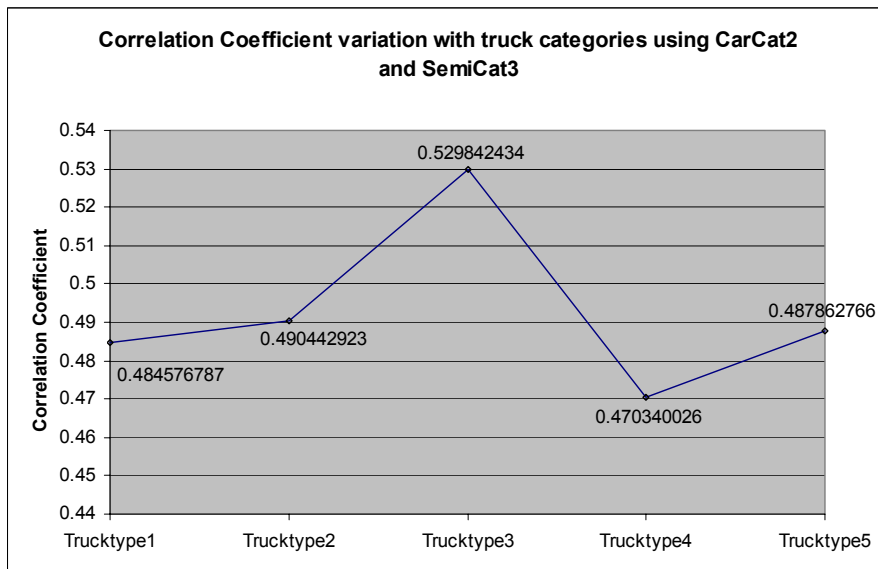


Figure 4.15. Correlation Coefficient variation with different truck categories and CarCat2 and SemiCat3

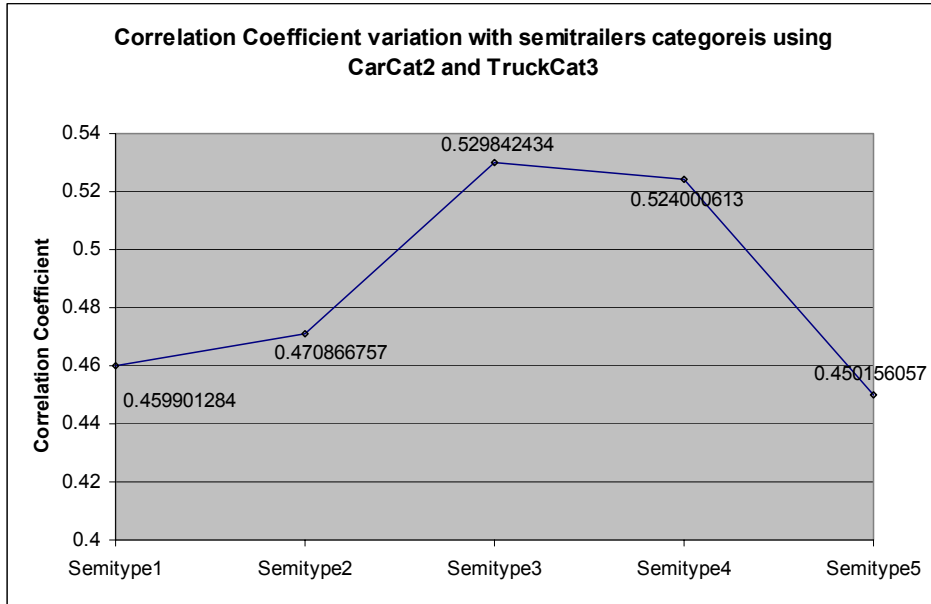


Figure 4.16. Correlation Coefficient variation with different semi-trailers and CarCat2 and TruckCat3

Table 4.4. Car Characteristics

Parameter	Mean	Deviation	Min.	Max.	Units
Length	4.47	0.31	3.81	5.03	m
Width	2.13	0.00	2.13	2.13	m
Maximum Desired Speed	110.00	30.00	95.00	140.00	km/h
Maximum Acceleration	3.00	0.30	2.70	3.30	m/s ²
Normal deceleration	4.00	0.20	2.70	4.30	m/s ²
Maximum Deceleration	7.00	0.20	6.40	7.60	m/s ²
Speed Acceptance	1.10	0.20	1.00	1.30	
Minimum Distance between vehicles	0.60	0.05	0.50	1.30	m
Give Way Time	5.00	1.00	3.00	7.00	sec
Guidance Acceptance	0.00	0.00	0.00	0.00	

Table 4.5. Truck Characteristics

Parameter	Mean	Deviation	Min.	Max.	Units
Length	9.18	0.89	8.08	10.21	m
Width	2.74	0.00	2.74	2.74	m
Maximum Desired Speed	90.00	30.00	88.00	110.00	km/h
Maximum Acceleration	2.10	0.10	1.80	2.40	m/s ²
Normal deceleration	3.00	0.10	1.80	2.40	m/s ²
Maximum Deceleration	5.00	0.10	2.70	3.30	m/s ²
Speed Acceptance	1.00	0.00	1.00	1.00	
Minimum Distance between vehicles	0.50	0.00	0.50	0.50	m
Give Way Time	5.00	0.00	5.00	5.00	sec
Guidance Acceptance	0.00	0.00	0.00	0.00	

Table 4.6. Semi-trailer Characteristics

Parameter	Mean	Deviation	Min.	Max.	Units
Length	18.94	3.51	15.24	22.44	m
Width	3.05	0.00	3.05	3.05	m
Maximum Desired Speed	80.00	30.00	70.00	100.00	km/h
Maximum Acceleration	1.60	0.05	1.50	1.70	m/s ²
Normal deceleration	2.40	0.10	2.10	2.70	m/s ²
Maximum Deceleration	3.80	0.10	3.50	4.10	m/s ²
Speed Acceptance	1.00	0.00	1.00	1.00	
Minimum Distance between vehicles	0.50	0.00	0.50	0.50	m
Give Way Time	5.00	0.00	5.00	5.00	sec
Guidance Acceptance	0.00	0.00	0.00	0.00	

Speed Acceptance and Guidance Acceptance

During discussions with Mn/DOT, it was noted that it is not uncommon for the drivers to drive at speeds in excess of five to ten miles per hour above the posted speed limit. It was also pointed out that this was mostly observed only for cars. But the simulator uses an absolute value as the speed limit on a section. One of the input parameters to the simulator is the degree of acceptance of the speed limit by the drivers. For cars, this speed limit has been increased by an average of 10% to model the aggressive drivers, subject to further refinement during calibration. For trucks and semi-trailers, it was assumed that the drivers do not cross the speed limit. As there was no vehicle guidance system in use, all the guidance acceptance values were set to zero.

Car Following Model Parameters

AIMSUN2 uses the 2-lane variation of the car-following model and needs the following parameters:

Number of Vehicles: Maximum number of vehicles to be considered for modeling the influence of adjacent lanes

Maximum Distance: The distance downstream of a vehicle to be considered in the model.

Maximum Speed Difference: The difference in speed between one lane and the adjacent lane in the model.

Maximum Speed Difference on On-Ramp: The difference in speed between the mainline and the on-ramp lane in the model.

The values used in this study were based on a literature survey of the sources mentioned earlier and are presented in Table 4.7.

Table 4.7. Car Following and Overtaking Maneuver Model Parameters

Parameter	Value
Number of Vehicles	8
Maximum Distance	100 m
Maximum Speed Difference	40 km/h
Maximum Speed Difference on on-ramp	50 km/h
Percent Overtake	0.95

Percent Recover	1.00
-----------------	------

Overtaking Maneuver Parameters

AIMSUN2 uses the “percent overtake,” and “percent recover,” to model the overtaking maneuvers as described in chapter 2 on simulator review. The parameters are presented here again (their values are shown in Table 4.7):

Percent Overtake: The percentage of the speed at which a driver decides to overtake. This value has to be greater than zero but less than one.

Percent Recover: The percentage of the speed at which a driver decides to return to the slower lane after overtaking. This value, also, has to be greater than zero but less than or equal to one.

Fuel Consumption and Pollution Emission Model Parameters

The fuel consumption model used in AIMSUN2 assumes that a vehicle is either idling, cruising at a constant speed, accelerating or decelerating. The state of each vehicle is determined every simulation step and an appropriate formula from Table 4.8 is used to calculate the fuel consumed for that state.

Table 4.8. Fuel Consumption formulae for different vehicle states

Vehicle State	Fuel Consumed during Δt
Idling	$F_i \Delta t$
Accelerating with acceleration a (m/s ²) and speed v (m/s)	$(c_1 + c_2 a v) \Delta t$
Decelerating	$F_d \Delta t$
Cruising at Speed v (m/s)	$(k_1 (1 + (\frac{v}{v_m})^3) + k_2 v) \Delta t$

where,

F_i = fuel consumption rate for idling vehicles in ml/s

c_1 and c_2 : constants in ml/s used for calculating fuel consumed

F_d : the fuel consumption rate for decelerating vehicles in ml/s

v_m : speed at which the fuel consumption rate, in ml/s, is at a minimum for a vehicle cruising at a constant speed

k_1 and k_2 are constants dependent on the variables F_1 and F_2 where,

F_1 : fuel consumption rate, in liters per 100 km, for vehicles traveling at a constant speed of 90 km/hr

F_2 : fuel consumption rate, in liters per 100 km, for vehicles traveling at a constant speed of 120 km/hr

Ackelic's [46] equation has been used to calculate the fuel consumption equation for a cruising vehicle moving at speed v . If F_1 and F_2 are the fuel consumption rates in liters per 100 km for a vehicle traveling at a constant speed of v_1 and v_2 respectively, then:

$$k_1 = \frac{(F_1 - F_2)v_1v_2v_m^3}{180(2v_2v_m^3 - 2v_1v_m^3 + v_2v_1^3 - v_1v_2^3)}$$
$$k_2 = \frac{2F_2v_2v_m^3 - 2F_1v_1v_m^3 + F_2v_2v_1^3 - F_1v_1v_2^3}{360(2v_2v_m^3 - 2v_1v_m^3 + v_2v_1^3 - v_1v_2^3)}$$

The values for fuel consumption in this study were taken from the results of a test conducted on a 1992 Oldsmobile at Orlando, Florida. This test conducted over a period of five months (to consider both summer and winter driving conditions) was part of the research "Trav-Tek Evaluation Modeling Study" conducted by the FHWA in 1992-1993, the results of which were published in March 1996. The vehicle was fully equipped with the Trav-Tek route guidance system. The values shown in Table 4.9 were used for the various parameters in this study.

Table 4.9. Fuel Consumption parameters used

Parameter	Value
Fi	0.97 ml/s
C1	1.20 ml/s
C2	0.60 ml/s
F1 @ 90 kmph	8.90 liters per 100 km
F2 @ 120 kmph	9.20 liters per 100 km
Fd	1.50 ml/s
Vm	60 kmph

Pollution Emissions

AIMSUN2 has the capability to model different types of emissions for different vehicle types. As in the fuel consumption model, the vehicle state (idling, cruising, accelerating or decelerating), vehicle speed and acceleration are used to evaluate the emission from each vehicle for each simulation step. In this study the pollutants modeled were carbon monoxide (CO), Nitrogen Oxides (NO_x), and unburned Hydrocarbons (HC). The inputs required (for each vehicle class and for each pollutant) for the pollution emission model are:

- 1) Emission rate for accelerating vehicles in g/s (Ea)
- 2) Emission rate for decelerating vehicles in g/s (Ed)
- 3) Emission rate for idling vehicles in g/s (Ei)
- 4) Emission rates for vehicles cruising at speeds from 10km/h to 70km/h in g/s

In this study, the values shown in Tables 4.10 to 4.12 were used in the pollution emission model.

Table 4.10. Pollution Emission rates for Cars

Emission rates	CO	NOx	HC
Idling emission rate (g/s)	0.060	0.001	0.007
Accelerating emission rate (g/s)	0.377	0.010	0.020
Decelerating emission rate (g/s)	0.072	0.010	0.007
Cruising emission rate (g/s)			
10 kmph	0.060	0.001	0.006
20 kmph	0.091	0.001	0.008
30 kmph	0.130	0.002	0.008
40 kmph	0.129	0.002	0.013
50 kmph	0.090	0.004	0.010
60 kmph	0.110	0.005	0.012
70 kmph	0.177	0.006	0.014

Table 4.11. Pollution Emission rates for Semi trailers

Emission rates	CO	NOx	HC
Idling emission rate (g/s)	0.050	0.005	0.038
Accelerating emission rate (g/s)	0.377	0.010	0.020
Decelerating emission rate (g/s)	0.072	0.001	0.007
Cruising emission rate (g/s)			
10 kmph	0.097	0.002	0.078
20 kmph	0.056	0.002	0.044
30 kmph	0.050	0.002	0.042
40 kmph	0.069	0.004	0.056
50 kmph	0.056	0.007	0.078
60 kmph	0.042	0.008	0.067
70 kmph	0.060	0.013	0.067

Table 4.12. Pollution Emission rates for Trucks

Emission rates	CO	NOx	HC
Idling emission rate (g/s)	0.050	0.001	0.007
Accelerating emission rate (g/s)	0.377	0.010	0.020
Decelerating emission rate (g/s)	0.072	0.001	0.007
Cruising emission rate (g/s)			
10 kmph	0.060	0.001	0.006
20 kmph	0.091	0.001	0.008
30 kmph	0.130	0.002	0.008
40 kmph	0.129	0.002	0.013
50 kmph	0.090	0.004	0.010
60 kmph	0.110	0.005	0.012
70 kmph	0.177	0.006	0.014

5 CALIBRATION & VALIDATION

5.1 Introduction

From a methodological point of view it is widely accepted that simulation is a useful technique to provide an experimental test bed to compare alternate system designs, replacing the experiments on the physical system by experiments on its representation in a computer in terms of a simulation model. The outcomes of the experiments on the simulation model can provide a basis for quantitative support to decision makers.

The reliability of this decision making process depends on the ability to produce a simulation model representing the physical system behavior close enough so that the model can be thought of as a substitute of the actual system. This is true for any simulation analysis in general and traffic simulation is no exception. The process of determining whether the simulation model is close enough to the actual system is usually achieved through the validation of the model, an iterative process involving the calibration of the model parameters and comparing the model's behavior with that of the actual system and using the discrepancies between the two, and the insight gained, to improve the model until the accuracy is judged to be acceptable.

Calibration/validation of the models was one of the most difficult phases of this research. The models were calibrated in three different steps: calibration for non-peak period flows, calibration during the peak period with metered demands, and finally calibration during the peak period with estimated real demands along with the implementation of the Mn/DOT ramp control algorithm. This chapter describes all the above three steps in detail and the reasons why calibration was done as a sequence of the above three steps rather than as one single step. Also, a brief overview of the statistical measures used in calibration is presented.

5.2 Statistical Measures Used in Calibration

In the case of traffic systems the behavior of the actual system is usually defined in terms of traffic variables such as flows, speeds, occupancies, queue lengths, etc., which can be measured by detectors at specific locations on the road network. To validate the simulation model the simulator should be able to emulate the traffic detection process and produce a series of simulated observations which when compared with the actual measurements can be used to determine whether the desired accuracy in reproducing the system behavior is achieved. Statistical techniques can be used for such determination.

For a microscopic simulation model in general, and AIMSUN2 in particular, the model behavior depends on a number of parameters, most of which have already been described in the previous sections. Summarizing, if one considers the model to be composed of entities such as vehicles, sections, junctions, intersections, etc., each of them described by a set of attributes, i.e. parameters of the car-following model, the lane changing model, gap acceptance model, speed limits and speed acceptance on sections, etc.; the numerical values of these parameters determines the model behavior. The calibration process objective is to find these parameters' values that would result in model as close to reality as possible within the study's constraints.

Some examples will help to illustrate the dependency between parameter values and the model behavior. Vehicle length has a clear influence on flow: as the vehicle length increases flow decreases and queue length increases. When discussing the AIMSUN2 car-following model it was shown that the section speed limit, the vehicle's target speed, and speed acceptance, among others, define the desired speed for each vehicle on each section. Higher target speed would mean higher desired speed for any given section, resulting in an increase in flow according to the speed-flow relationships. Also, the vehicle acceleration and breaking parameters influence the capacity of the sections, specifically of weaving sections. Similarly, the global percent recover and percent overtake parameters and a section's Zone 1 and Zone 2 lengths, have an influence on the lane-changing model, traffic distribution in various lanes, etc.

The statistical methods and techniques for validating simulation models are described in many textbooks and specialized papers [47,48]. What follows is an adaptation of the general process to

the problem of validating a microscopic simulation model. The measured data in the actual system should be split into two data sets: the data set that will be used to develop and calibrate the model, and a separate data set that will be used for the validation test.

At each step in the iterative validation process a simulation experiment is conducted. Each of these simulation experiments is defined by the data input to the simulation model and the set of values of the model parameters that identify the experiment. The output of the simulation experiment is a set of simulated values of the variables of interest, which in this study are the volume and occupancy measured at each traffic detector station every five-minutes. For example, assuming that the model statistics are gathered every five minutes, and that the sampling variable is the simulated flow w , the output of the simulation model will be characterized by the set of values w_{ij} , of the simulated flow at station i at time j , where index i identifies the detector station ($i=1,2,\dots,n$; where n is the number of stations), and index j the sampling interval ($j=1,2,\dots,m$; where m is the number of sampling intervals in the simulation time period T). If v_{ij} are the corresponding actual measures for station i at sampling interval j , a typical statistical technique to validate the model would be to compare both series of observations to determine if they are close enough. For a station i , the comparison could be based on testing whether the difference

$$d_i = w_{ij} - v_{ij}, j=1,\dots,m$$

has a mean \bar{d}_i significantly different from zero or not. This can be determined using the t-statistics:

$$\bar{t}_{m-1} = \frac{\bar{d}_i - \delta_i}{\bar{s}_d / \sqrt{m}}$$

where δ_i is the expected value of \bar{d}_i and \bar{s}_d the standard deviation of \bar{d}_i , for testing the null hypothesis:

$$H_0 : \delta_i = 0 \quad \left(|\bar{t}_{m-1}| > t_{m-1;\alpha/2} \right)$$

If for $\delta_i = 0$ the calculated value \bar{t}_{m-1} of the Student's t distribution is significant to the specified significance level α then it can be concluded that the model is not reproducing close enough to the system behavior and the model has to be rejected.

If $\delta_i = 0$ gives a non-significant \bar{t}_{m-1} then it can be concluded that the simulated and the real means are “practically” the same so the simulation is “valid enough”.

This process is repeated for each of the n detector stations. The model is accepted when all stations (or a specific subset of stations, depending on the model purposes and taking into account that the simulation is only a model, and therefore an approximation, so δ_i will never be exactly zero) pass the test.

For the statistical tests however some special considerations need to be taken into account [48], specifically in the case of the traffic simulation analysis.

The statistical procedure assumes identically and independently distributed (i.i.d) observations whereas the actual system measures and the corresponding simulated measures are for a time series. Therefore it would be desirable that at least the m paired (correlated) differences $d_i = w_{ij} - v_{ij}$, $j=1, \dots, m$ are i.i.d. This can be achieved when the w_{ij} and the v_{ij} are average values of independently replicated experiments.

The bigger the sample is, the smaller the critical value $\bar{t}_{m-1; \alpha/2}$ is, and this implies that a simulation model has a higher chance of being rejected, as the sample grows bigger. Therefore the t statistics may be significant and yet unimportant if the sample is very large, and the simulation model can be good enough for practical purposes.

These considerations imply that the validation of the model should not be relied on only one type of statistical test. An alternative test is to check whether w and v are positively correlated, that is test the significance of the null hypothesis:

$$H_0 : \rho > 0 \text{ (}\rho \text{ linear correlation coefficient)}$$

This represents a less stringent validation test acceptance because simulated and real responses do not necessarily need to have the same mean; what is significant is that whether or not they are

positively correlated. The test can be implemented using the ordinary least squares technique to estimate the regression model:

$$E(v|w) = \beta_0 + \beta_1 w + \varepsilon \quad (\varepsilon \text{ random error term})$$

This test deals with the one-sided hypothesis $H_0 : \beta_1 \leq 0$. The null hypothesis is rejected and the simulation model accepted if there is strong evidence that the simulated and the real responses are positively correlated. The variance analysis of the regression model is the usual way of implementing this test. This test may be strengthened, making it equivalent to the first test if this hypothesis is replaced by the composite hypothesis $H_0 : \beta_0 = 0$ and $\beta_1 = 1$, implying that the means of the actual measurements and the simulated responses are identical and when an actual measurement exceeds its mean then the simulated observation exceeds its mean too.

A third family of statistical tests for the validation of simulation models is rooted in the former observation that the actual measured and the simulated series, v_{ij} and w_{ij} respectively, are time series. In this case the actual measured series could be interpreted as the original one and the simulated series the “prediction” of the observed series. In that case the quality of the simulation model could be established in terms of the quality of the prediction, and that would mean to resort to time series forecasting techniques for that purpose. If it can be considered that what is observed as output of the system as well as output of the simulation model is dependent on two type of components: the functional relationships governing the system (the pattern) and the randomness (the error), and that the measured as well as the observed data are related to these components by the relationship:

$$\text{Data} = \text{pattern} + \text{error}$$

Then the critical task in forecasting can be interpreted in terms of separating the pattern from the error component so that the former can be used for forecasting. The general procedure for estimating the pattern of a relationship is through fitting some functional form in such a way as to minimize the error component. A way of achieving that could be through the regression analysis as in the former test.

If for station i the error of the j -th “prediction” is $d_i = w_{ij} - v_{ij}$, $j=1, \dots, m$, then a typical way of estimating the error of the predictions for the station i is “Root Mean Square Error” , rms_i defined by:

$$rms_i = \sqrt{\frac{1}{m} \sum_{j=1}^m (w_{ij} - v_{ij})^2}$$

This error estimate has been perhaps the most widely used in traffic simulation, and although a smaller rms_i implies a better model it has an important drawback as it emphasizes large errors by squaring the error. Therefore it would be helpful to have a measure that considers both, the disproportionate weight of large errors and provides a basis for comparison with other methods.

Theil’s U-Statistic measure achieves both these objectives. In general, if X_j is the observed and Y_j the forecasted series, $j = 1, \dots, m$, then Theil’s U-Statistic is defined as:

$$U = \sqrt{\frac{\sum_{j=1}^{m-1} (FRC_{j+1} - ARC_{j+1})^2 / (m-1)}{\sum_{j=1}^{m-1} (ARC_{j+1})^2 / (m-1)}}$$

where:

$$FRC_{j+1} = \frac{Y_{j+1} - X_j}{X_j} \quad \text{is the forecasted relative change, and}$$

$$ARC_{j+1} = \frac{X_{j+1} - X_j}{X_j} \quad \text{is the actual relative change}$$

An immediate interpretation of Theil’s U-Statistic, is the following:

$U = 0 \Leftrightarrow FRC_{j+1} = ARC_{j+1}$, then the forecast is perfect

$U = 1 \Leftrightarrow FRC_{j+1} = 0$, then the forecast is as bad as possible

Therefore the closer to zero the Theil's U-Statistics is, the better the forecasted series is or, in other words, the better the simulation model. When Theil's U-statistic is close to 1 the forecasted series, and therefore the simulation model, should be rejected.

When the forecast efficiency is based on the regression model $E(v|w) = \beta_0 + \beta_1 w + \varepsilon$ (ε random error term) the most efficient forecast would correspond to $\beta_0 = 0$ and $\beta_1 = 1$, that can be tested by the application of variance analysis to the regression model as indicated earlier. But taking into account that the average squared forecast error:

$$D_m^2 = \frac{1}{m} \sum_{j=1}^m (Y_j - X_j)^2$$

can be decomposed (Theil) in the following way:

$$D_m^2 = \frac{1}{m} \sum_{j=1}^m (Y_j - X_j)^2 = (\bar{Y} - \bar{X})^2 + (S_Y - S_X)^2 + 2(1 - \rho)S_Y S_X$$

where \bar{Y} and \bar{X} are the sample means of the forecasted and the observed series respectively, S_Y and S_X are the sample standard deviations and ρ is the sample correlation coefficient between the two series, the following indices can be defined:

$$\left. \begin{aligned} U_M &= \frac{(\bar{Y} - \bar{X})^2}{D_m^2} \\ U_S &= \frac{(S_Y - S_X)^2}{D_m^2} \\ U_C &= \frac{2(1 - \rho)S_Y S_X}{D_m^2} \end{aligned} \right\} \Rightarrow U_M + U_S + U_C = 1$$

UM is the "Bias Proportion," index and can be interpreted in terms of a measure of systematic error, US is the "Variance Proportion," index and provides an indication of the forecasted series ability to replicate the degree of variability of the original series or, in other words, the simulation model's ability to replicate the variable of interest of the actual system. Finally UC or "Covariance Proportion," index is a measure of the unsystematic error. The best forecasts, and hence the best simulation models, are those for which UM and US do not differ significantly from zero and UC is close to unity. It can be shown that this happens when β_0 and β_1 in the regression do not differ significantly from zero and one respectively.

For efficiency and speed all of the above calculations were implemented in a computer program that queries both the simulation output database and the database containing the real (actual) values and produces the desired statistical measures (correlation coefficient, various Theil's coefficients, etc.). It also displays the actual and simulated volumes and occupancies along with the statistical measures for each mainline station used in the calibration. Thus the program reduces substantially the effort necessary for iterating until the best parameter values are obtained. An easy to use interactive graphical user interface was developed for this purpose as shown in Figure 5.16.

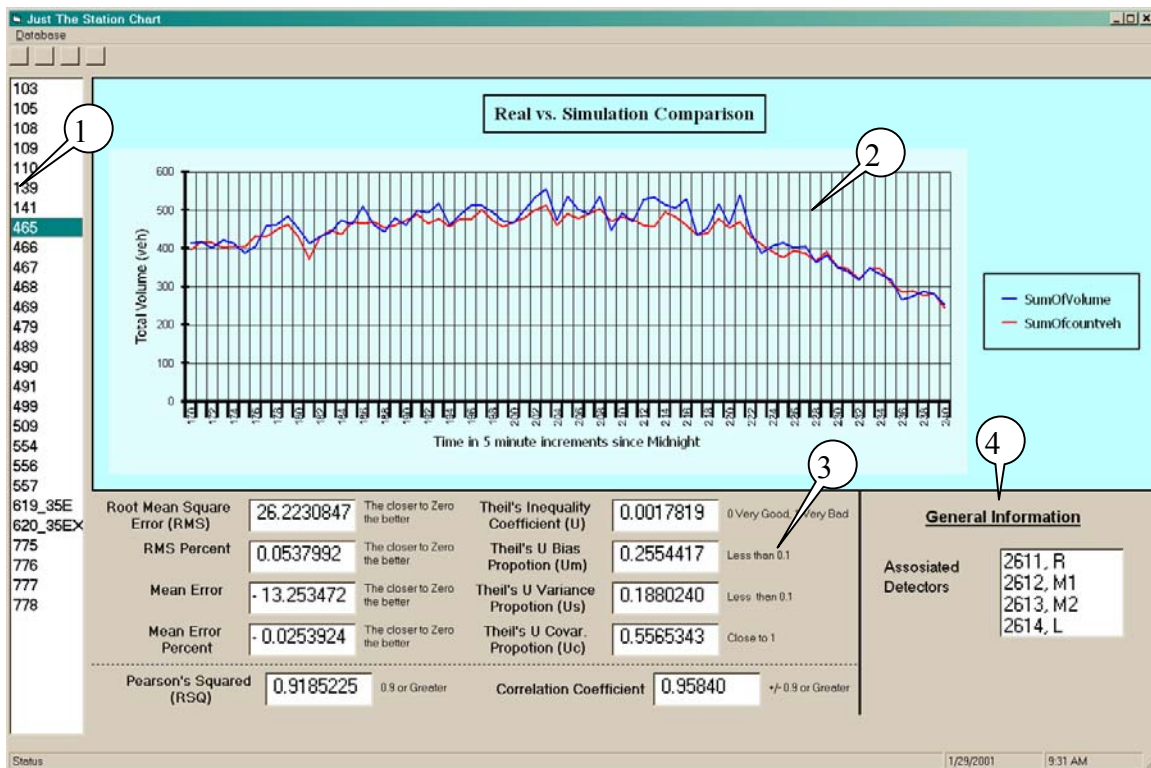


Figure 5.16. Graph showing comparison of simulated and actual station volumes

Area 1 of Figure 5.16 shows a list of all the detector stations in the test site (TH-169 in this case). The user can select each one of these stations with the mouse and observe the corresponding real and simulated measures plotted against time as shown in area 2 of the figure. The blue curve represents the real measurements from the field whereas the red curve represents the simulation output for the same location. The program can extract and plot data for volumes (which are

displayed in Figure 5.16), occupancies, speeds and densities. As speed and density are not measured in the field, only their simulated values can be displayed. In area 3 the various statistical measures calculated for the data shown in area 2 corresponding to the chosen detection station is displayed with some tips on how to interpret them. In area 4 the individual detectors that comprise the selected detector station are displayed for information purposes.

5.3 Calibration for Non-peak Period Flows

By the beginning of this step, models of the test sites already had been prepared with traffic demand volumes for the non-peak period hours i.e., 11:00 A.M. to 1:00 P.M. The correlation coefficient of the simulation results obtained was only 0.48, which indicated that the model was not accurate. A decision had to be made on how to progress further, i.e., whether the calibration of the model should be done in conjunction with Mn/DOT ramp metering control for the peak-period or try to improve the model before ramp metering was tested.

It should be noted that if the model is not accurate and ramp metering is invoked, the decisions that would have been made by the Mn/DOT algorithm in the model would differ from those made in the field. These deviations would be small at first and then increase rapidly as the simulation proceeds in time because the model would be going further away from reality. This is because the Mn/DOT algorithm uses the mainline station data for calculating the ramp metering rates and the data in the model could be different from that in the field, depending upon the quality of the model. Calibrating the model using the Mn/DOT algorithm would mean trying to calibrate not only the mainline station volume and occupancy but also the ramp metering decisions made by the Mn/DOT algorithm that would make it very complex. This approach would have required considerably more runs to calibrate the model.

Hence, a different approach was adopted. It was felt that it would be easier to calibrate the model for peak period traffic conditions if the model was working well for non-peak period traffic conditions. Data from 11:00 A.M. to 1:00 P.M. on March 21st was used for calibrating the test site on TH 169, and April 3rd data for the same time period was used to calibrate the test site on I-94. The entrance flows and the exit turning percentages were calculated based on the loop

detector data. For the remaining data such as vehicle characteristics and car-following parameters, best guesses were used. For example, the vehicle characteristics used was the best set of vehicle types data chosen as explained in Section 4.3.2.4. So is the case with the speed limits on the sections.

One of the inputs to the simulator was the initial state of the network. This initial state consists of specifying the number of vehicles in the model, including their location and speed. Instead of calculating the initial number of vehicles along with their speeds in the model, warm up period of fifteen minutes was used in the study. Hence, when the simulator starts collecting traffic data from 11:00 A.M., the initial state of the network would not be empty, as it would have already simulated traffic for fifteen minutes. Simulation runs were done with varying warm-up time periods, during which the simulator is using the first demand time slice as input, and it was found that the simulation results were more or less stable for warm-up time of fifteen minutes and above. Hence, a fifteen-minute warm-up time period was used in the study.

For calibration, only the simulated station volume counts were compared with those in the field. The simulated occupancy values were not compared with their real counterparts because prior to that detectors with correct lengths need to be modeled and at that stage those lengths were not available. The vehicle parameters, speed limits, car-following parameters were calibrated so that the simulated station volumes were as close as possible to those in the field. A correlation coefficient of 0.85 was achieved at the end of this step. It should be noted that a correlation coefficient higher than 0.85 was not achieved, no matter what set of input parameters were used for the study. This behavior can be attributed to the fact that at very low flows, which are the typical flows expected between 11:00 A.M. and 1:00 P.M., the driver behavior is too random to exactly replicate in the model.

5.4 Peak-period Calibration Without Metering

As mentioned earlier in Section 4.3.2.2, Mn/DOT collects “metered demand”, i.e., the number of vehicles that have actually been allowed through the ramp meters every five minutes, as part of its normal data collection efforts. Though, this was presented as a limitation in section 4.3.2.2, as

AIMSUN2 needs traffic arrival patterns upstream of the ramp (i.e., “real demand”) to calculate the delay on the ramps, for calibration purposes it may be considered a blessing in disguise.

Simulating the metered demand without any ramp meters in the model is equivalent to simulating the Mn/DOT ramp-metering algorithm in which the decisions made in the model are identical to those in the field. This is because the Mn/DOT algorithm computes the ramp metering rates to achieve maximum throughput at bottleneck locations based on the mainline conditions without taking into account the number of vehicles waiting on the ramps. Hence, this approach was adopted to calibrate the model during the peak period conditions. Also, as the ramp meters release the vehicles at a constant headway (for a given ramp metering rate), the vehicle arrivals in the model were assumed to have a constant headway between them.

5.4.1 Volume Calibration

The models were initially calibrated for mainline volumes. The test site on TH 169 was calibrated for March 21st data, and March 22nd data was used to refine it further. For the test site on I-94, March 29th and April 3rd data were used for similar purposes. Starting from the upstream end of the networks, the five-minute mainline station volumes in the model were visually compared against the volumes in the field. The desired speed, acceleration, normal deceleration and maximum deceleration of the individual vehicle types were calibrated so that the simulated volumes matched their field volumes as much as possible. For each modification of the above parameters, five to six runs were made so that the average behavior could be captured. The volume calibration needed about 500 runs resulting in an average correlation coefficient of 0.90 for mainline station volumes.

5.4.2 Speed Calibration

As mentioned in Section 2.3, the Mn/DOT algorithm performs volume and occupancy control. Therefore, before testing the model with the Mn/DOT ramp-metering algorithm, it is necessary to calibrate the model for both volume and occupancy. But it is known that the flow could be on either side of the Q-K curve for the same volumes, i.e., it could be congested or uncongested. Hence, there was a need to calibrate the model more before testing it with ramp metering. In this

step, speeds calculated from real data were used in calibration. The effective lengths of the detectors (which ranged from 15 to 32 feet) and vehicles were obtained from Mn/DOT and speeds were calculated every 5 minutes using the volume and occupancy as follows:

$$V = \frac{Q * (L_v + L_d)}{300 * O_{cc}}$$

where,

V = speed in ft/sec,

Q = detector vehicle count in the five-minute interval,

L_v = effective length of the vehicle in ft,

L_d = effective length of the detector in ft,

Occ= Occupancy recorded by the detector

The simulator measures the speeds of all the vehicles that pass over the detectors. These speeds were averaged for all the detectors at a location to get the simulated station speeds. The simulated station speeds were visually compared with the station speeds calculated from the field data. The speed limits and grades on the sections in the model were modified to get as close a match as possible. Hence, at the end of this step, the models were calibrated for mainline volumes and speeds.

5.4.3 Occupancy Calibration

It should be noted that calibrating the model for occupancy involves calibrating the length of the detectors in the model. It should also be noted that changing the length of the modeled detectors doesn't have any effect on the volume or the speed calibration done earlier. In this step, the modeled detector lengths were modified for occupancy calibration. It was found that, generally, there is a one-to-one correspondence between the effective length of the detectors in the field and the modeled detectors except at two locations.

While performing the speed and occupancy calibrations, each modification is checked by doing five to ten simulation runs. Also, the random seeds used by the simulator were changed frequently to capture the overall behavior. These two phases needed about 1,000 simulation runs,

with the models having an average correlation coefficient value of 0.97 and 0.90 for mainline volume and occupancy respectively.

5.5 Calibration With Mn/DOT Ramp Metering

The final step in the models' calibration was simulating real demands with the Mn/DOT ramp-metering algorithm. The real demands were computed as described in Section 4.3.2. As real demands are the traffic arrival patterns upstream of the ramp, the vehicles in the model were generated with exponential headway, i.e., the time intervals between two consecutive vehicle arrivals at entrance sections were sampled from an exponential distribution.

As mentioned in Section 2.3 in the Mn/DOT ramp control algorithm, there are no specific start and stop times for ramp metering. The operators switch on ramp meters manually and they also have the capability to override the metering rates calculated by the Mn/DOT algorithm if they think it is necessary. The ramp meters are usually overridden during incidents. The effect of this is that individual ramp meters are switched on at different times which may range anywhere between 2:00 to 3:45 P.M. and they may also be overridden during the P.M. peak period. Also, Mn/DOT operators invoke a "dump mode," at the end of metering, in which before a ramp meter is turned off; the waiting vehicles (if any) are released at rate 1 till the queue dissipates.

Hence, to simulate the ramp metering, metering logs containing ramp metering rates implemented in the field during the data collection days were obtained from Mn/DOT and the ramp metering start and stop times for each ramp meter were determined. The logs were also analyzed to determine if the ramp meters were overridden during the peak period. The override files were created for each test day using this information. The models were then simulated with the Mn/DOT ramp control algorithm. In additions to the mainline stations volume and occupancy comparisons, the actual merge detector counts, simulated merge detector counts and the actual arrival rates are plotted against time and compared visually as shown in Figure 5.17.

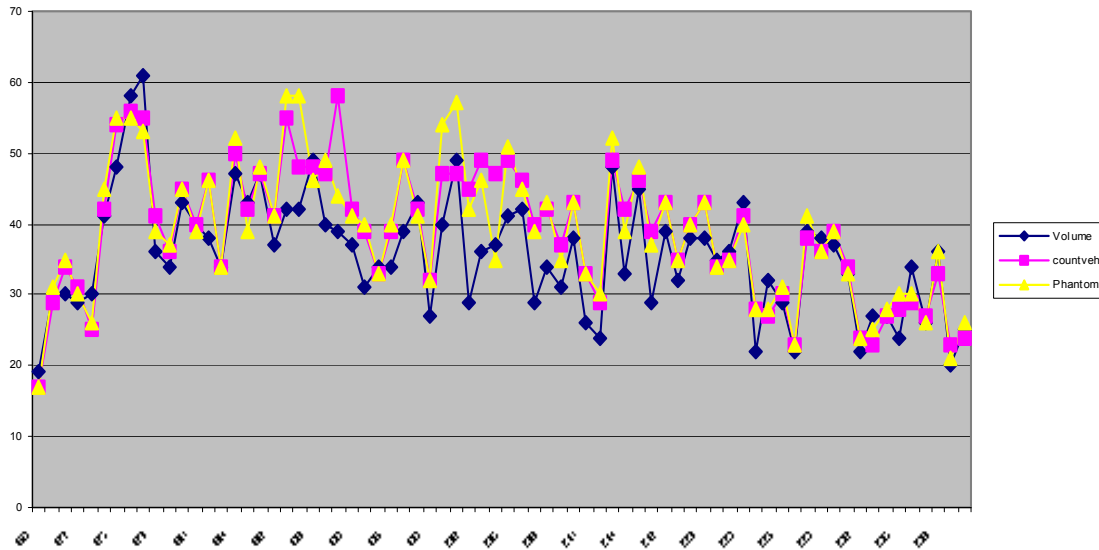


Figure 5.17. Graph showing the comparison of the merge detector counts (simulated and real) and the arrival rates

The line with the triangle markers indicates the vehicle arrival at the upstream end of the ramp. The line having rhombuses as point markers indicates the actual merge detector counts and the line with square markers represents the simulated merge detector counts. From the graph, it can be seen that the actual merge detector counts and simulated merge detector counts coincide with the arrival patterns before the start of ramp metering which is expected as there would not be any waiting at the ramp meters. This is an indication that the model performed well until the ramp meters were invoked. The simulated and actual merge detector counts do not match very closely during the ramp-metering period, as the decisions made in the model do not coincide exactly with those in the field. It can be observed that the simulated and actual merge detector counts being more or less the same once ramp-meters are turned off. In the first few attempts to calibrate the model, it was found that merge detector counts in the model were diverging from their real values. At this point, the speed limits and grades on sections were modified further to improve the model. The reaction times also were modified to achieve the simulated algorithm compliance with reality. It should be noted that modifying a specific section properties would have effects not only on a single ramp but also on a number of ramps, some of which may be in the same zone in which the section is present and others may be in the upstream zones. Hence, every time a parameter is modified, its effect on the whole model should be observed to determine if the modification of that parameter takes the model closer to reality. Also, the

cumulative rates of the simulated ramp meters were compared for each ramp with those in the field to check if there was any bias in the rates being implemented in the model. Performing simulation runs with the March 22nd and March 23rd data validated the TH-169 model. The I-94 model was validated by performing simulation runs with the March 30th and April 3rd, 2000 data. The average statistical measures calculated for mainline station volume for the two test sites are shown in Table 5.13.

Table 5.13. Average Statistical Measures for mainline station volumes

	TH-169			I-94		
	Mar 21st	Mar 22nd	Mar 23rd	Mar 29th	Mar 30th	Apr 3rd
Root Mean Square Error %	10.62	6.42	7.39	5.02	6.14	5.57
Correlation coefficient	0.98	0.97	0.96	0.97	0.95	0.97
Theil's Inequality Coefficient (0 is good)	0.00426	0.00154	0.00238	0.00122	0.00178	0.00139
Theil's Bias Proportion (0 is good)	0.68070	0.12352	0.08826	0.11589	0.12152	0.13044
Theil's Variance Proportion (0 is good)	0.01052	0.05365	0.03098	0.09125	0.03374	0.07182
Theil's Covariance Proportion (1 is good)	0.30877	0.82281	0.88075	0.79285	0.84473	0.79773

5.6 Problems Encountered with the Test Site on TH 169

During the calibration, it was found that before invoking the “dump mode” (i.e., turning off the ramp meters), there were long queues on the entrance ramps from I-394 and from Excelsior Blvd., which wasn't the case in the field. When the dump mode was invoked, the freeway segments upstream of the ramp were blocked due to high traffic volume that entered the freeway. But the metering rates computed in the model were more or less similar to those implemented in the field, indicating that the Mn/DOT algorithm in the model is making similar decisions. Reaction time, detectors lengths, speed limits, etc., were modified in the hope that the problem

might be rectified, but with no result. After exhausting all possible scenarios as to why long queues were being formed at these ramps, the input ramp data was cross checked with the data given by Mn/DOT. As described in Section 4.3.2.2, the queue detector measurements can be used as real demand. The layout of the detectors is shown in Figure 5.18. It was thought that the queue detector was only counting the flow that was being metered but in reality it is counting both the HOV flow as well as the metered flow.

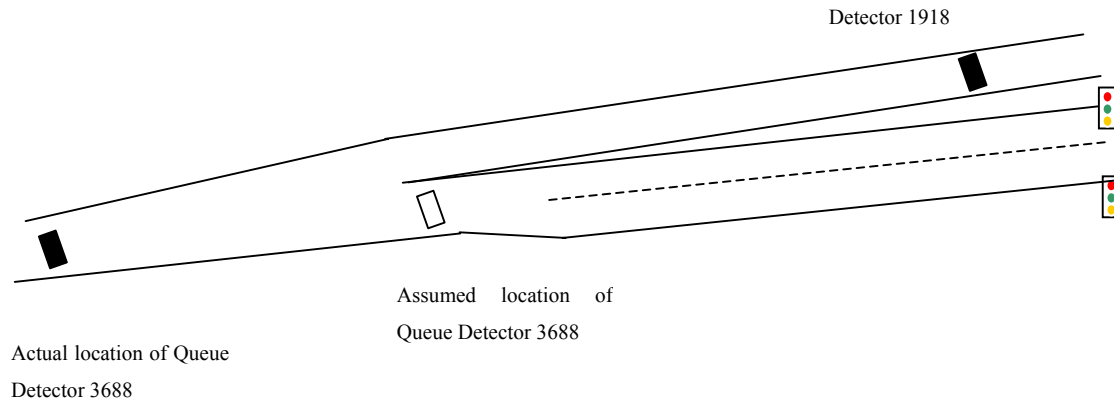


Figure 5.18. Queue detector at I-394 counting HOV vehicles also

Hence, the real demand was being overestimated by a number equal to the number of HOV vehicles. For the whole ramp metering time period, this constituted a significant number of vehicles, which resulted in long queues. The real demand was then corrected, which rectified the problem.

The ramp from Excelsior Blvd. also was the site of another problem exhibiting a similar discrepancy. The data were checked for errors by comparing the maximum number of vehicles that can be allowed by the ramp meter with the merge detector volume counts reported every five minutes. It was found that the merge detector indicated counts, which were higher than the maximum allowed by the ramp. But this increase in the number of vehicles was less than the number of vehicles counted by the HOV detector. Hence, additional counting of the HOV vehicles by the merge detector was ruled out. Violations were investigated as another possible reason for this discrepancy. But the difference between the merge detector counts and the maximum allowed by the ramp meters was as high as 40% and it was felt that such high violation rates were not feasible. Another reason that was considered was the possibility that the

red times reported for the ramp metering rates were different from those that actually occurred in the field. Hence, the red times were manually measured in the field on a weekday and were compared with those that were recorded by Mn/DOT; but there was no difference.

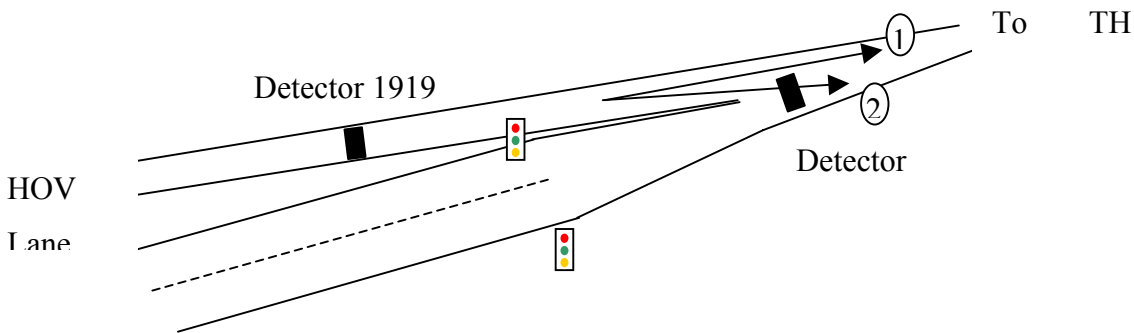


Figure 5.19: Detectors on the Excelsior entrance ramp

After further inspection, it was determined that the HOV vehicles were the likely reason for the discrepancy. Although not drawn to scale, Figure 5.19 shows the approximate detectors layout on the Excelsior Ramp. Detector 1919 is the merge detector for the un-metered HOV flow and detector 1908 is the merge detector for the metered flow. The HOV vehicles that take the path labeled by arrow 2 were actually going over the metered-flow merge detector and hence were being counted twice which should not be the case. As described in Section 4.3.2.2, the arrival demand patterns were estimated based on the detector 1908-volume counts and the queue lengths at the ramps, which resulted in generating more vehicles than those that have actually crossed the ramp. It was felt that this was the only suitable explanation that would describe the discrepancy. The real demand for that ramp was rectified by estimating the number of vehicles that would have used the ramp in a five-minute interval as equal to the number of greens displayed by the ramp meter during that five-minute period.

5.7 Problems Encountered with the Test Site on I-94

When the model was run with the Mn/DOT algorithm and real demands, it was found that the simulation volumes and occupancies were very different from their real values. Reaction time, section speed limits, detector lengths, etc. had been changed in the hope of getting the model closer, but that did not have much effect. Then the algorithm parameters were crosschecked with

Mn/DOT and it was realized that the wrong parameters were being used in the model. When the correct parameters were used, the model performed along expected lines. To simulate the ramp metering with correct parameters, it was necessary to go only one step up, i.e., the model was calibrated for the peak period without ramp metering rather than starting all over again from the beginning. This also indicates the advantages of the methodology adopted to do the calibration in small steps rather than as one single large step.

Even after using the correct algorithm values, it was found that the entrance ramp from 6th Street South had queues that were much longer than actually observed in the field. At that stage the I-94 model calibration was still in its early stage. Even the Mn/DOT engineers felt that similar to the problems faced with the test site on TH 169 at Excelsior Blvd., the merge detector on the entrance ramp from 6th Street South was counting the HOVs also. Hence, the HOV counts were subtracted from the merge detector counts. The result was that there were much smaller queues at the ramp. When the model was calibrated further, it was found that the queues on the ramps were smaller than observed actually. Hence, the data was checked once again and it was observed that the merge detector counts did not include the HOV counts. Adding the subtracted HOV counts back to the entrance demand and creating the input demand again rectified this error.

It should be noted that the amount of time spent in calibrating the test site on I-94 was much less than the time needed for the test site on TH 169. Though, the I-94 test site was more complex in terms of the geometry and traffic that was simulated, the experience gained in calibrating TH 169 was put to good use. Hence, the authors feel that it would be a good approach to model a less complex network before trying to attempt more complex networks.

REFERENCES

- “Transit Services Needs Assessment”, Regional Transit Board, Minneapolis, Final Report, 1986.
- Mn/DOT Report (1996). “Traffic Management Program Overview, Twin Cities Metro Area”, Report # TMC 070403-0196, Mn/DOT, February 1996.
- MarketLine Research, Inc. “4th Annual Perception Tracking Study”, prepared for Mn/DOT, 1999.
- Contents of, ”Transportation Bill”, passed by the Minnesota State Senate on March 20, 2000.
- “Development of advanced traffic management systems in the US by state agencies”, ITSA Fact Sheets (1995), 1995.
- Piotrowicz, G. and Robinsins, J. “ Ramp metering status in North America: 1995 Update”. DOT-T-95-17, FHWA. 1995.
- Eil Kwon, Sreeman Nanduri. “Traffic Responsive Metering Systems in North America”, Interim Report, 2000.
- Lipp, Louis E., Corcora, Lawrence J. and Hickman, Gordon A. “Freeway Ramp Metering Effects in Denver”, TRB Annual Meeting, 1991.
- “TH 169: Dynamic Ramp Metering Evaluation”, Mn/DOT, Report 1998-14, March 1998.
- Cynthia Taylor, Deirdre Meldurm. “Evaluation of a Fuzzy Logic Ramp Metering Algorithm: A comparative study among three ramp metering algorithms used in the greater Seattle Area”, Draft Research Report, Research Project Agreement No. T9903, Task 84.
- Jacobson, L., Henry, K., and Mehyar, O. “Real-Time Metering Algorithm for Centralized Control”, Transportation Research Record 1232, pp 17-26.
- Lau Rich P. “MnDOT Ramp Metering Algorithm”. Internal Report of Minnesota Department of Transportation, Minneapolis, 1996.
- “I-5 North Freeway Ramp Metering, Portland, Oregon: Project Development-Operation”, Oregon Department of Transportation-Metropolitan Branch, June 1982.
- “Traffic Management Program Overview: Twin Cities Metro Area”, Minnesota Department of Transportation, Report TMC 07043-0196, 1996.
- Henry Kim C., Mehyar, Omar. “Six-Year FLOW Evaluation”, Washington Department of Transportation, District 1, 1989.

“Before and After Study, Ramp Metering, Eastbound Long Island Expressway, Suffolk County”, Technical Memo, James H. Kell & Associates, 1989.

Kevin Sommers, “Comparative Evaluation of Freeway Simulation Models”, M.S. Thesis, University of Minnesota, Minneapolis, 1996.

Payne, H. J. “FREFLO: A Macroscopic Simulation Model of Freeway Traffic”, Transportation Research Record 722, 1979, pp.68-77.

“FRESIM User Guide, Version 4.5”, FHWA, 1994.

Van Aerde, M. “INTEGRATION: A Model for Simulating IVHS in Integrated Traffic Networks, User’s Guide for Model Version 1.5e”, Transportation Systems Research Group, Queens University, Kingston, Ontario, Canada, 1994.

Van Aerde, M. and Baker, M. “Weaving Capacity Sensitivity Analysis Using the INTEGRATION Model”, Queens University, Kingston, Ontario, Canada, 1996.

Tsutomu I. and May A. A., “FREQ8PE: A Freeway Corridor Simulation and Ramp Metering Optimization Model”, Institute of Transportation Studies, University of California, Berkeley, 1985.

Kwon E., Kota R., Coyle M., Michalopoulos P. G., Song S. “Enhancement of the KRONOS Simulation Package and Database for Geometric Design, Planning, Operations and Traffic Management in Freeway Networks/Corridors (Phase III)”. Project Report, Mn/DOT, , Minnesota, 1997.

Michalopoulos P. G., Kwon E., and Kang J.C. “Enhancement and field testing of a Dynamic Simulation Program”. Transportation Research Review #1320, pp. 203-215. 1991.

Michalopoulos P. G. “A dynamic Freeway Simulation Program for Personal Computers”. Transportation Research Record 971, pp. 68-79, 1984.

Barceló, J., Ferrer J.L. and Grau R. “AIMSUN2 and the GETRAM Simulation Environment”. Internal Report. Departamento de Estadística e Investigación Operativa. Facultad de Informática. Universitat Politècnica de Catalunya. 1994.

Barceló, J., Ferrer J.L. and Grau R. “Microscopic Traffic Simulation for ATT Systems Analysis”. Research Report, Departament de Estadística e Investigación Operativa, Universitat Politecnica de Catalunya. 1996.

Barceló, J., Ferrer J.L., Grau R., Florian M. and Le Saux E. "A Route Based Variant of the AIMSUN2 Microsimulation Model". 2nd. World Congress on Intelligent Transport Systems, Yokohama, 1995.

Ferrer, J.L. and Barceló J. "AIMSUN2: Advanced Interactive Microscopic Simulator for Urban and non-urban Networks". Internal Report. Departamento de Estadística e Investigación Operativa. Facultad de Informática. Universitat Politècnica de Catalunya. 1993.

Reynhout K., and Michalopoulos P. G. "Development of Traffic simulation Laboratory For Design, Planning, and Traffic operations (Phase I)". Project Report, Minnesota DOT, Office of Research Administration, Minnesota, 1996.

Kwon Eil. "Development of a New Simulation Module for Traffic Responsive, Automatic Rate-Selection Ramp Metering Strategies". Project Report Mn/DOT, Office of Research Administration, Minnesota. 1997.

Ramesh Kota. "Macroscopic Modeling of Traffic Flow in HOV Lanes", MS Thesis, University of Minnesota, Minneapolis, Minnesota. 1999.

Muralidhar Koka, John Hourdakakis and Panos G. Michalopoulos. "Computer Aided Testing and Evaluation of Adaptive Ramp Control Strategies" presented at 79th Annual Meeting, Transportation Research Board, 2000.

W. Helly. "Simulation of Bottlenecks in Single-Lane Traffic Flow.", in R.C. Herman (Ed.), Theory of Traffic Flow, Proc. Symp. Theory of Traffic Flow, Elsevier, Amsterdam, pp. 207-238, (1961).

J.F. Gabbard. "Car-Following Models", Concise Encyclopedia of Traffic and Transportation Systems, Pergamon Press, Oxford, 1991.

R. Herman, E. W. Montroll, R. Potts and R.W. Rothery. "Traffic Dynamics: Analysis of Stability in Car-following". Operations Research, 1(7), pp. 86-106, 1959.

D.A. Wicks. "INTRAS-a microscopic freeway corridor simulation model. Overview of simulation in highway transportation," 1, pp.95-107, 1977.

Qi Yang and H.N. Koutsopoulos. "A Microscopic Traffic Simulator for Evaluation of Dynamic Traffic Management Systems," Transportation Res. C. Vol.4, No. 3, pp. 113-129, 1996.

M. Fellendorf, VISSIM. "A microscopic simulation tool to evaluate actuated signal control including bus priority," Technical paper, 64th ITE Annual Meeting, Dallas, 1994.

- P.G. Gipps, A behavioral car-following model for computer simulation. *Transp. Res. B*, Vol. 15, pp 105-111, (1981).
- J. Barceló, J.L. Ferrer, D. García, M. Florian and E. Le Saux. "Parallelization of Microscopic Traffic simulation for ATT Systems Analysis", *Equilibrium and Advanced Transportation Modeling*, Kluwer Academic Publishers, 1998.
- G. Duncan. "Simulation at the Microscopic Level", *Traffic Technology International*, pp. 62-66, 1996.
- D. Manstetten, W. Krautter and T. Schwab. "Traffic Simulation Supporting Urban Control System Development," Robert Bosch GmbH, Stuttgart, Germany, 1998.
- T. Bleile, W. Krautter, D. Manstetten and T. Schwab. "Traffic Simulation at Robert Bosch GmbH," *Proc. Euromotor Seminar Telematic / Vehicle and Environment*, Aachen, Germany, Nov. 11-12, 1996.
- P.G. Gipps. "A Model for the Structure of Lane-Changing Decisions," *Transpn. Res.* 20B, pp.403-414 1986.
- Akcelic. "Progress in Fuel Consumption Modeling for Urban Traffic Management", *Australian Road Research Board Research Report*, ARR No. 124. 1982.
- O. Balci. "Verification, Validation and Testing," in "Handbook of Simulation: Principles, Methodology, Advances, Applications and Practice," Ed. by J. Banks, John Wiley, 1998.
- J.P.C: Kleijnen. "Theory and Methodology: Verification and Validation of Simulation Models," *European Journal of Operational Research*, Vol. 82, pp. 145-162, 1995.

6 RAMP METERING EVALUATION

6.1 Introduction

This is the preliminary report containing the results and progress from the completion of task 2 of the “Employment of the Traffic Management Laboratory (TRAMLAB) for Evaluating Ramp Control Strategies in the Twin Cities” project conducted at the University of Minnesota. This report is just a part of the final report and contains only information relevant to the final results.

The results presented in this report are products of traffic simulations conducted for two freeway sections in the Twin Cities, Minnesota. The two sections are:

TH-169 Northbound from the intersection with I-494 up to 36th St.

I-94 Eastbound from the Intersection with I-394 up to 7th St. in St. Paul

The experiments were based on data collected during March 2000 specifically the following days:

On TH-169: Tuesday March 21, Wednesday March 22, and Thursday March 23.

On I-94: Wednesday March 29, Thursday March 30, and Monday April 3.

In this preliminary report the results from the comprehensive runs of the simulation are presented. The MOEs defined in Task 1 were extracted from each test site for each of the three days and are presented in two sets of tables and two sets of graphs. The first set of tables contains MOEs for the entire simulation period whereas the second set contains the same MOEs for the ramp-metering period only. The MOEs collected are the following:

Volume: The total number of vehicles that have entered the system extracted for all the metered ramps and for the entire system.

Total Travel Time: The total travel time in veh-hrs spent by all the vehicles at three different levels: mainline, all metered ramps, and the entire system.

Total Travel: The total number of veh-miles traveled by all the vehicles in the system extracted for all the metered ramps and the entire system.

Average Speed: The average speed in mph on the mainline.

Total Delay: The delay in veh-hrs accumulated by all the vehicles calculated at three levels: mainline, all metered ramps and the entire system.

Average Delay per vehicle: The delay in minutes faced on an average by a vehicle while traveling on the ramps, mainline and the entire system.

Total Number of Stops: The total number of stops experienced by all the vehicles while using the mainline.

Average Number of Stops per vehicle: The average number of stops experienced by a vehicle while traveling on the mainline.

Fuel Consumption: The total fuel consumed in gallons by all the vehicles while traveling in the entire system.

Pollutants: The total emissions in kgs calculated separately for CO, HC, NO_x made by all the vehicles while they were in the system.

The third set of tables for each test site presents detailed results for every individual metered ramp of the system. This information is extracted only for the metered period since these MOEs are generally zero for the un-metered period. The MOEs extracted are as follows:

Total Ramp Delay: The total delay in veh-hrs, experienced by all the vehicles at the ramp.

Average Delay per vehicle: The average delay in minutes experienced by a vehicle at the ramp.

Maximum Delay: The maximum delay, in minutes, of all the five-minute average delays experienced at the ramp. Hence, the actual maximum delay experienced by an individual vehicle could be slightly higher than the maximum delay extracted here.

Average Queue: The average queue (in number of vehicles) observed at the ramp.

Maximum Queue: The maximum queue (in number of vehicles) observed at the ramp.

For the ramps that have been identified as problematic i.e., ramps which may have excessive delays, queues, etc., the graphs presented for each day for each test site show the variation of

average delay experienced at those ramps with respect to time. In addition to the above graphs, three 3-D graphs are presented for each test site displaying the following:

Volume vs. Time vs. Length. The average flow aggregated every quarter of a mile every five minutes.

Speed vs. Time vs. Length. The average speed of vehicles aggregated every quarter of a mile every five minutes.

Density vs. Time vs. Length. The average density of vehicles aggregated every quarter of a mile every five minutes.

6.2 General Observations: Freeway and Demand Characteristics

Before understanding the results of the simulation a few observations that were made during the course of this study have to be noted. As expected, the two freeways used in this study have some noticeable differences that may affect the control strategy operation. TH-169 is a relatively straight freeway with mild grade changes. It has very small visibility problems or any other obstructions that might affect the speed on certain sections of the freeway.

On the other hand on I-94 there are quite a few turns and also some noticeable grade changes that have been observed to affect traffic. In addition, although I-94 has more lanes than TH-169 it carries a higher number of vehicles that can easily reach unstable flow conditions. On TH-169, entrance and exit ramps are textbook cases; they are on the right hand side of the mainline and in most cases have adequate acceleration/deceleration lanes. I-94 has four entrance ramps joining the mainline from the left side introducing slow traffic into the fastest lane of the mainline and three exit ramps located on the left hand side of the mainline; these entrance and exit ramps result in increased weaving and slowing down of the fastest lane. For example, the exit ramp to I-35W NB in conjunction with the large volumes on the upstream entrance ramps generates a high number of lane changes that create one of the reported bottlenecks.

The days on which data was collected were supposedly free from capacity reducing incidents. This means that no incidents were reported to the TMC on those days but it does not guarantee

that traffic was typical or disturbance free. For some applications in traffic engineering a large truck moving down the freeway with speed below the speed limit might be considered an incident but events like this are rarely reported or logged. The TMC engineers were of the opinion that during the last week of March 2000 they observed less than normal traffic volumes on I-94, mainly due to the fact that the University of Minnesota was closed for spring break during that period. The data collected on these days had to be used for the project because immediately after data collection the 6th Street entrance ramp was closed for reconstruction. For the three test days on which data was collected for TH-169, the TMC engineers were also of the opinion that though there was nothing unusual about the observed traffic, the station counts were slightly on the lower side. Under a closer inspection of the demand data, March 22nd, 2000 produced a different traffic pattern than the other two days. Specifically, the entrance volumes of the ramps are higher while the upstream end entrance flow is slightly lower.

6.3 Observations from the Experiments

As the simulation models were calibrated with data collected when the Mn/DOT ramp control algorithm was in operation, the scenarios with ramp control had more or less similar traffic patterns observed in the field. Specifically, since the Mn/DOT ramp control algorithm is designed to maximize the zone bottlenecks' throughput, all the bottlenecks on both test sites had normal traffic flows that were close to their corresponding capacities, but never exceeding them. For the no-control simulation scenarios, as expected, the bottlenecks were the first locations to display flow breakdown. On these scenarios it was observed that TH-169 broke down at the intersection with I-394. I-94 on March 29th broke down on the first bottleneck encountered (5th Street) resulting in normal conditions on the subsequent bottlenecks. For the remaining two days, although the no control scenarios revealed higher congestion, the demands were not high enough to produce a severe breakdown. This resulted in lower speeds and delays but not stop and go traffic.

As in reality, forced flow (stop and go conditions) occurred suddenly and then spread very rapidly; further more, the existence and onset of forced flow conditions as well as their duration have a significant impact on the simulation results. Often, even for the same demand patterns

forced flow conditions are generated by random events such as slow moving vehicles, etc. The latter was observed at the upstream bottleneck, mentioned earlier (I-94). In order to get statistically correct results, 30 iterations of each scenario were done and the output for each section was averaged as described in the next Section to extract the MOEs presented here.

MOE Calculation

AIMSUN2 has the capability of collecting measurements at different levels of aggregation in space and time. It can gather them for the whole system, for each section, for each turning movement and for every stream (set of consecutive sections) defined by the user. It can also gather them at two time scopes: global, for the whole duration of simulation, and at regular user-defined collection intervals. The following sections describe the output that is measured by the simulator and the methodology involved. In this study, MOEs were extracted from AIMSUN2 and summarized for the freeway mainline only, for all the ramps in aggregate, for the entire system, and for each individual ramp. The individual ramp MOEs help in identifying problem ramps i.e., ramps which may have excessive vehicle delays, queues, etc.

AIMSUN2 Output

At the section level, vehicle data such as entry and exit times, number of stops and total stopped time is collected when a vehicle enters and leaves a section. From this information, the section traffic measurements shown in Table 6.1 are collected. The process adopted by AIMSUN2 to compute these measurements is described in this section.

The information gathered from an individual vehicle at every section, s , is:

Entrance time of the i th vehicle in the section (seconds), $TENS_i$

Exit time of the i th vehicle from the section (seconds), $TEXS_i$

Total stop time accumulated in the section by the i th vehicle (seconds), $TSTS_i$

Total number of stops accumulated in the section by the i th vehicle, TNS_i

Using this information, when a vehicle enters into a new section the following variables are calculated:

The average section travel time of the i th vehicle (seconds), TT_i

$$TT_i = TEXS_i - TENS_i$$

Table 6.1: Section simulation output measurements

Traffic Measurement	Flow	Units	Definition
Mean Flow		veh/hr	Average number of vehicles per hour that have crossed the section
Density		veh/km	Average number of vehicles per kilometer in the section per lane
Mean Speed		km/hr	Average speed for all the vehicles that have traversed the section
Travel Time		s/veh	Average time a vehicle needs to traverse the section
Delay Time		s/veh	Average delay time per vehicle in the section
Stop Time		s/veh	Average time at standstill per vehicle while traveling in the section
Number of Stops		per veh	Average number of stops per vehicle while traveling in the section
Mean Queue Length	Queue	veh/lane	Average length of the queue in the section, expressed as the number of vehicles per lane
Maximum Queue Length	Queue	veh/lane	Maximum length of the queue in the section, expressed as the number of vehicles per lane
Total Travel		Km	Total number of kilometers traveled by all the vehicles in the section
Fuel Consumed		Liters	Total liters of fuel consumed inside the section, by all the vehicles that have traversed it
Pollutants Emitted		Kg	For each pollutant, total kilograms of pollution emitted inside the section by all the vehicles that have traversed the section

The average section delay time of the i th vehicle (seconds), DT_i , is computed from:

$$DT_i = TT_i - \left[\frac{L_s}{\text{Min}(S_{Max_i}, S_s * \theta_i)} + \frac{L_t}{\text{Min}(S_{Max_i}, S_t * \theta_i)} \right]$$

where S_s = speed limit on section s (m/s).

S_t = speed limit of turning t (m/s).

θ_i = speed acceptance of i th vehicle

S_{Maxi} = maximum desired speed of vehicle i (m/s).

L_s = length of section s (meters).

L_t = length of turning t (meters).

The average section speed of the i th vehicle (m/s), S_i , is calculated as:

$$S_i = \frac{L_s + L_t}{TT_i}$$

The inverse of S_i (s/m), used for calculating the harmonic mean speed, HS_i , is obtained as:

$$HS_i = \frac{1}{\frac{L_s + L_t}{TT_i}}$$

The stop time in the section of the i th vehicle (seconds), ST_i , is measured directly. Hence,

$$ST_i = TSTs_i$$

Number of stops of the i th vehicle, NS_i , is also measured directly. Hence,

$$NS_i = TNSs_i$$

For every user defined time slice, I (in seconds), for which the simulation output measurements are requested as well as for the entire simulation period, the following measurements are calculated by section:

The mean flow in a section, F_{sec} is calculated as:

$$F_{sec} = \frac{N_{sec}}{I} * 3600$$

where,

N_{sec} = Number of vehicles that exit a section

The average travel time per vehicle in a section, TT_{sec} is computed as:

$$TT_{sec} = \frac{\sum_{i=1}^{N_{sec}} TT_i}{N_{sec}}$$

The average speed per vehicle in a section (km/h), S_{sec} , is calculated as:

$$S_{sec} = \frac{\sum_{i=1}^{N_{sec}} S_i}{N_{sec}} * 3.6$$

The harmonic mean speed per vehicle in a section (km/h), HS_{sec} is computed as:

$$HS_{sec} = \frac{N_{sec}}{\sum_{i=1}^{N_{sec}} S_i} * 3.6$$

The average delay time per vehicle in a section, DT_{sec} is obtained as:

$$DT_{sec} = \frac{\sum_{i=1}^{N_{sec}} DT_i}{N_{sec}}$$

The average stop time per vehicle of a section, ST_{sec}, is calculated as:

$$ST_{sec} = \frac{\sum_{i=1}^{N_{sec}} ST_i}{N_{sec}}$$

The average number of stops per vehicle of a section, NS_{sec}, is computed as:

$$NS_{sec} = \frac{\sum_{i=1}^{N_{sec}} NS_i}{N_{sec}}$$

The density of a section, DEN_{sec} is calculated as:

$$DEN_{sec} = \frac{\sum_{l \in sec} \left(\sum_{t_i \in T_l} \left[NVeh_{l,t(i-1)} * (t_{l,i} - t_{l,(i-1)}) \right] / I \right)}{\sum_{l \in sec} L_l} * 1000$$

where,

L_l = Length of lane l (m)

NVeh_{l,t} = Number of vehicles in the lane l at time t

T_l = (0, t_{l,1}, ..., t_{l,m}, I) : instants when the number of vehicles in lane l changes

6.4 Extraction of MOEs from AIMSUN2 Output

The original section output from the simulator was the set of measurements defined in Table 6.1.

In this study, the time slice was defined to be five minutes for every section. To be statistically

correct, data from 30 simulation runs was used in the study. The output data per section was then averaged for all the runs. The averaged output was then used to compute the system MOEs for different sections or levels i.e., the freeway mainline, all the entrance ramps and for the entire system. Table 6.2 presents the MOEs and units used for the first three levels. Individual entrance ramp MOEs were also computed in this fashion.

Table 6.2. System MOEs

System MOE	Units	Aggregation Level
Volume	Vehicles	Ramps
		Entire system
Total Travel Time	Vehicle-hours	Mainline
		Ramps
		Entire system
Total Travel	Vehicle-miles	Ramps
		Entire system
Average Speed	Miles per hour	Mainline
Total Delay	Vehicle-hours	Mainline
		Ramps
		Entire system
Average Delay per vehicle	Minutes per vehicle	Mainline
		Ramps
		Entire system
Total number of stops		Mainline
Number of stops per veh	Per vehicle	Mainline
Fuel Consumption (gallons)	Gallons	Entire system
Pollutants (CO, HC, NO _x)	Kilograms	Entire system

The system MOEs were computed for the control vs. the no-control scenarios not only for the whole simulation time-period but also for the ramp metering time-period so that the performance

of the system can be studied during that time. The methodology for extracting the system MOEs from the AIMSUN2 output measurements is described in this section.

The mainline was modeled in the simulator as a set of consecutive sections. Hence, data from all the sections that comprise the mainline needs to be used to calculate the mainline MOEs. Assuming that the mainline consists of sections 1 to s , the total travel time (TTTi), the total delay time (TDTi), the total travel (TTi), and the total number of stops (TSi) of all the vehicles that have entered the mainline during the i th five-minute interval is computed as:

$$\begin{aligned} \text{TTTi} &= \sum_{j=1}^s \frac{T_{j,i} * Q_{j,i}}{12} \\ \text{TDTi} &= \sum_{j=1}^s \frac{D_{j,i} * Q_{j,i}}{12} \\ \text{TTi} &= \sum_{j=1}^s TT_{j,i} \\ \text{TSi} &= \sum_{j=1}^s \frac{AS_{j,i} * Q_{j,i}}{12} \end{aligned}$$

where,

$T_{j,i}$ = average travel time per vehicle (seconds), in section j during the i th five-minute interval

$Q_{j,i}$ = average flow in veh/hr through section j during the i th five-minute interval

$D_{j,i}$ = average delay time per vehicle (seconds), in section j during the i th five-minute interval

$TT_{j,i}$ = total travel in kilometers in section j during the i th five-minute interval

$AS_{j,i}$ = average number of stops per vehicle in section j during the i th five-minute interval

As the average flow is in veh/hr, it is converted into an equivalent flow for a five-minute interval by dividing it by 12. The summation of the total travel time, over all the five-minute intervals, gives the total travel time (TTT) on the mainline for the entire simulation period. Similarly, the total delay time (TDT), total travel (TT), and the total number of stops (TS) on the mainline for the entire simulation period were extracted. The equations used are:

$$TTT \text{ (veh-hrs)} = \frac{\sum_{i=1}^N TTT_i}{3600}$$

$$TDT \text{ (veh-hrs)} = \frac{\sum_{i=1}^N TDT_i}{3600}$$

$$TT \text{ (veh-miles)} = \frac{\sum_{i=1}^N TT_{r,i}}{1.607}$$

$$TS = \sum_{i=1}^N TS_i$$

where,

N = total number of five-minute time intervals for the entire simulation period

The average speed (Q) on the mainline is computed as:

$$Q \text{ (mph)} = \frac{\sum_{i=1}^N \sum_{j=1}^s (V_{j,i} * Q_{j,i})}{1.607 * \sum_{i=1}^N \sum_{j=1}^s Q_{j,i}}$$

where,

V_{j,i} = mean speed in km/hr through section j during the ith five-minute interval

The average delay per vehicle and the average number of stops per vehicle on the mainline is computed as follows:

$$\text{Average delay per vehicle (minutes)} = \frac{TDT}{TN}$$

$$\text{Average number of stops per vehicle} = \frac{TS}{TN}$$

where,

TN = total number of vehicles that entered the mainline, i.e., the summation of the total number of vehicles that entered from the upstream end of the network and from all the entrance ramps over the entire simulation period.

A similar procedure was used to calculate the MOEs for the ramps. The MOEs of the ramps and the mainline were used to obtain the corresponding MOEs for the entire system. For example, the total travel time for the entire system is the sum of the total travel time on the mainline and the total travel time on the ramps. So, is the case with total travel, total delay, and total number of stops for the entire system. Average speed for the entire system is a weighted average of the speeds of all the vehicles that have entered the system. It was calculated similar to the average speed on the mainline, except that the summation is over all the sections in the entire system. For the entire system, the average delay is obtained by dividing the sum of the total delay on the mainline and the total delay on the ramps with the total number of vehicles that have entered the system.

6.4.1 Quarter Mile Statistics

To study the variation of flow, speed, and travel time along the mainline in space and time, average speeds, flows, and travel times were calculated for the mainline every quarter mile, for every five-minute interval. To create the quarter mile statistics, the whole network was split into quarter mile segments with the first segment starting at the upstream end of the network. All the sections on the mainline being between 0 and 0.25 miles from the upstream end of the network are assigned to the first segment. This process is repeated for all the mainline sections in the network. It should be noted that if the location of a section is such that it straddles the boundary between two quarter-mile segments, it is placed in only one of them, depending on which quarter mile segment includes most of the section.

The measurements of all the sections lying in the quarter-mile segments are aggregated. For example, the average speed for a quarter-mile segment for every five minutes was calculated as the average of the section speeds that comprise that segment. The total travel time in the quarter-mile segment is calculated as the sum of the travel times on the individual sections. The quarter-mile segment flow is calculated as the average of the all the section flows. These results are presented in the form of graphs in later sections.

6.4.2 Individual Ramp MOEs

The following MOEs were computed for an individual ramp during the metering period: total ramp delay, average ramp delay, maximum ramp delay, average queue length and maximum queue length. Total delay time on the ramp is the sum of the delay of all the vehicles that have entered the ramp. For a ramp consisting of sections 1 to s , the total delay time (TDT) on the ramp is calculated as:

$$\text{TDT (veh-hrs)} = \frac{\sum_{i=1}^N \sum_{j=1}^s \frac{D_{j,i} * Q_{j,i}}{12}}{3600}$$

where,

$Q_{j,i}$ = average flow in veh/hr through section j during the i th five-minute interval

$D_{j,i}$ = average delay time per vehicle (seconds), in section j during the i th five-minute interval

Average delay per vehicle on the ramp is obtained by dividing the total delay time on the ramp by the total number of vehicles that have entered the ramp. It should be noted that the delay time shown in Table 6.1. is the average delay per vehicle that has crossed the section during a five-minute interval. Information about the maximum delay of all the vehicles that have crossed a section during a five-minute interval is not available. To calculate the maximum delay and the maximum queue length experienced at a ramp the following reasoning is used. Vehicles at the upstream end of the ramp would be delayed/queued only when there is congestion/queue at the downstream end of the ramp. Hence, the summation of the average delay during a five-minute interval for all the sections in a ramp gives the average delay at the ramp during that five-minute interval. The maximum of these average delays for the entire simulation period is defined as the maximum delay at the ramp, which is one of the MOEs. It should be noted that the individual maximum delays of the vehicles that entered the ramp could be expected to be higher than this. The mean queue length per section obtained from AIMSUN2 (Table 6.1) is aggregated for all the sections in the ramp which gives the mean queue length for that five-minute interval for that ramp. This mean queue length is then averaged for the entire simulation period to obtain the average queue length MOE. The summation of the maximum queue lengths observed in individual sections during a five-minute interval gives the maximum queue length observed on

that ramp during that five-minute interval. The maximum value of the maximum queue length observed every five-minutes over the entire simulation period gives the maximum queue length ever observed at a ramp.

6.5 Results Interpretation

6.5.1 Entire simulation period

The following pages containing Tables 6.3 to 6.8 present the MOEs extracted for the entire simulation period for the two test sites.

**Table 6.3. TH169 NB, Entire Cycle Results (vs. metered period only)
2:00 to 8:00 P.M., March 21st, 2000.**

MOE	Aggregation Level	With Control ♦	Without Control	Percentage Improvement	
Total Travel Time (veh-hours)	Mainline	3804	4603	+21%	
	Ramps	599	81	-86%	
	Entire Site	4403	4685	+6%	
Total Travel (veh-miles)	Entire Site	219715	219754	Negligible*	
	Ramps	5792	5792	Negligible*	
Speed (mph)	Mainline	56.2	46.5	+18%	
Total Delay (veh-hours)	Mainline	303	1099	+263%	
	Ramps	503	0	NA**	
	Entire Site	806	1099	+36%	
Average Delay/veh (min)	Mainline	0.365	1.322	+262%	
	Ramps	0.772	0	NA**	
	Entire Site	0.970	1.322	+36%	
Total no. of stops ²	Mainline	7256	153177	+2011%	
No. of stops per veh	Mainline	0.145	3.071	+2017%	
Volume (veh entered)	Entire Site	49884	49884	0	
	Ramps	39102	39102	0	
Fuel Consumed (gal.)	Entire Site	13736	18371	+34%	
Pollutants:	CO (kg)	Entire Site	3493	4128	+18%
	HC (kg)		231	262	+14%
	NOx (kg)		67	84	+25%

Negligible since results presented are for entire congestion cycle

** No ramp congestion occurs without control throughout this freeway

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

² A vehicle is considered stopped if it's speed drops below 2.2 mph

**Table 6.4. TH169 NB, Entire Cycle Results (vs. metered period only)
2:00 to 8:00 P.M., March 22nd, 2000.**

MOE	Aggregation Level	With Control ♦	Without Control	Percentage Improvement
Total Travel Time (veh-hours)	Mainline	3863	5213	+35%
	Ramps	689	83	-88%
	Entire Site	4552	5295	+16%
Total Travel (veh-miles)	Entire Site	221936	221882	Negligible*
	Ramps	5869	5870	Negligible*
Speed (mph)	Mainline	55.9	41.4	+27%
Total Delay (veh-hours)	Mainline	325	1668	+413%
	Ramps	592	0	NA**
	Entire Site	917	1668	+82%
Average Delay/veh (min)	Mainline	0.385	1.976	+413%
	Ramps	0.896	0	NA**
	Entire Site	1.086	1.976	+82%
Total no. of stops ³	Mainline	10232	276005	+2597%
No. of Stops per veh	Mainline	0.202	5.450	+2597%
Volume (veh entered)	Entire Site	50642	50642	0
	Ramps	39625	39625	0
Fuel Consumed (gal.)	Entire Site	14101	21840	+55%
CO (kg)	Entire Site	3562	4709	+32%
Pollutants: HC (kg)		236	297	+26%
NOx (kg)		68	99	+46%

Negligible since results presented are for entire congestion cycle

** No ramp congestion occurs without control throughout this freeway

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

³ A vehicle is considered stopped if it's speed drops below 2.2 mph

**Table 6.5. TH169 NB, Entire Cycle Results (vs. metered period only)
2:00 to 8:00 P.M., March 23rd, 2000.**

MOE	Aggregation Level	With Control ♦	Without Control	Percentage Improvement
Total Travel Time (veh-hours)	Mainline	3827	4675	+22%
	Ramps	599	80	-87%
	Entire Site	4427	4755	+7%
Total Travel (veh-miles)	Entire Site	221516	221572	Negligible*
	Ramps	5712	5712	Negligible*
Speed (mph)	Mainline	56.4	46.2	+18%
Total Delay (veh-hours)	Mainline	296	1140	+285%
	Ramps	505	0	NA**
	Entire Site	802	1140	+42%
Average Delay/veh (min)	Mainline	0.355	1.370	+285%
	Ramps	0.784	0	NA**
	Entire Site	0.963	1.370	+42%
Total no. of stops ⁴	Mainline	7058	156920	+2123%
No. of Stops per veh	Mainline	0.141	3.143	+2123%
Volume (veh entered)	Entire Site	49923	49923	0
	Ramps	38654	38654	0
Fuel Consumed (gal.)	Entire Site	13694	18329	+34%
CO (kg)	Entire Site	3496	4182	+20%
Pollutants: HC (kg)		231	266	+15%
NOx (kg)		67	85	+28%

Negligible since results presented are for entire congestion cycle

** No ramp congestion occurs without control throughout this freeway

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

⁴ A vehicle is considered stopped if it's speed drops below 2.2 mph

**Table 6.6. I-94EB, Entire Cycle Results (vs. metered period only)
2:00 to 8:00 P.M., March 29th, 2000.**

MOE	Aggregation Level	With Control ♦	Without Control	Percentage Improvement
Total Travel Time (veh-hours)	Mainline	6943	8726	+26%
	Ramps	1004	516	-49%
	Entire Site	7948	9242	+16%
Total Travel (veh-miles)	Entire Site	399205	399289	Negligible*
	Ramps	19350	19356	Negligible*
Speed (mph)	Mainline	54.7	43.5	+20%
Total Delay (veh-hours)	Mainline	371	2068	+458%
	Ramps	413	60	-85%
	Entire Site	784	2128	+172%
Average Delay/veh (min)	Mainline	0.219	1.220	+457%
	Ramps	0.279	0.049	-85%
	Entire Site	0.463	1.256	+171%
Total no. of stops ⁵	Mainline	15615	306849	+1865%
No. of Stops per veh	Mainline	0.154	3.017	+1865%
Volume (veh entered)	Entire Site	101652	101692	0
	Ramps	88740	88749	0
Fuel Consumed (gal.)	Entire Site	23344	34326	+47%
CO (kg) Pollutants: HC (kg) NOx (kg)	Entire Site	5152	6549	+27%
		342	431	+26%
		97	137	+41%

Negligible since results presented are for entire congestion cycle

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

⁵ A vehicle is considered stopped if it's speed drops below 2.2 mph

**Table 6.7. I-94 EB, Entire Cycle Results (vs. metered period only)
2:00 to 8:00 P.M., March 30th, 2000.**

MOE	Aggregation Level	With Control ♦	Without Control	Percentage Improvement
Total Travel Time (veh-hours)	Mainline	6221	7438	+20%
	Ramps	732	457	-38%
	Entire Site	6953	7895	+14%
Total Travel (veh-miles)	Entire Site	403888	404301	Negligible*
	Ramps	19681	19677	Negligible*
Speed (mph)	Mainline	61.7	51.7	+16%
Total Delay (veh-hours)	Mainline	308	759	+146%
	Ramps	249	13	-95%
	Entire Site	557	772	+39%
Average Delay/veh (min)	Mainline	0.179	0.441	+147%
	Ramps	0.165	0.008	-95%
	Entire Site	0.323	0.448	+39%
Total no. of stops ⁶	Mainline	39943	58557	+47%
No. of Stops per veh	Mainline	0.386	0.567	+47%
Volume (veh entered)	Entire Site	103424	103344	0
	Ramps	90383	90314	0
Fuel Consumed (gal.)	Entire Site	23813	26315	+11%
CO (kg) Pollutants: HC (kg) NOx (kg)	Entire Site	4824	5346	+11%
		320	349	+9%
		92	103	+11%

Negligible since results presented are for entire congestion cycle

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

⁶ A vehicle is considered stopped if it's speed drops below 2.2 mph

**Table 6.8. I-94 EB, Entire Cycle Results (vs. metered period only)
2:00 to 8:00 P.M., April 3rd, 2000.**

MOE	Aggregation Level	With Control ♦	Without Control	Percentage Improvement
Total Travel Time (veh-hours)	Mainline	5691	6542	+15%
	Ramps	470	429	-9%
	Entire Site	6161	6971	+13%
Total Travel (veh-miles)	Entire Site	377993	378700	Negligible*
	Ramps	18593	18599	Negligible*
Speed (mph)	Mainline	63.1	55.0	+13%
Total Delay (veh-hours)	Mainline	203	340	+67%
	Ramps	72	12	-83%
	Entire Site	275	351	+28%
Average Delay/veh (min)	Mainline	0.126	0.210	+67%
	Ramps	0.051	0.008	-83%
	Entire Site	0.170	0.217	+28%
Total no. of stops ⁷	Mainline	17048	21191	+24%
No. of Stops per veh	Mainline	0.175	0.218	+24%
Volume (veh entered)	Entire Site	97127	97152	0
	Ramps	84895	84860	0
Fuel Consumed (gal.)	Entire Site	20191	20591	+2%
CO (kg) Pollutants: HC (kg) NOx (kg)	Entire Site	4304	4622	+7%
		287	305	+6%
		81	85	+6%

Negligible since results presented are for entire congestion cycle

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

⁷ A vehicle is considered stopped if it's speed drops below 2.2 mph

Total System Travel Time (TTT) at both sites has reduced with ramp control strategy, the reduction ranged from 6% to 16% (Table 6.3 and Table 6.6). For TH-169, the 6% reduction in TTT was observed on March 21st, whereas the 16% decrease in TTT was observed on TH-169 on March 22nd and on I-94 on March 29th. As seen from Tables 6.3 to 6.8, the Total Travel (TT) has not changed with control as expected since it refers to the entire simulation period (starting uncongested, and ending uncongested). Average speed for the ramp control scenario increased by 13% to 26% on the freeway mainline. The increases in speed on TH-169 ranged from 17% to 26%; on I-94 the increase in speed on the mainline ranged from 13% to 20%.

Total Delay (TD) was the most sensitive of all MOEs, with decreases ranging from 28% to 172%. Specifically, as Tables 6.3 to 6.5 indicate the system decrease of TD on TH-169 ranged between 36% to 82% while on I-94 ramp metering resulted in improvements from 28% to 172%. At first glance, the numbers on TH-169 appear to be more focused when compared to I-94 but the reader must keep in mind that I-94 is carrying almost twice as much volume as TH-169 with the result that congestion affects a much larger number of vehicles resulting to greater unpredictability. All tables suggest that delays on the ramps increase substantially with control as expected.

On both freeways the total number of stops with ramp control has been relatively low. However, without control the total number of stops increased tenfold. Specifically, with control, the average number of stops varied from 0.14 to 0.38 per vehicle on TH-169 and I-94 respectively (Table 6.3 and Table 6.7), while without control the same average number of stops varied from 0.22 to 5.45 per vehicle. The smallest increase was observed during two days on I-94 where congestion increased but the flow did not break down to stop-and-go conditions.

Fuel consumption and pollutant emissions are greatly affected by the smoothness of flow (fewer acceleration-deceleration cycles) so it should be expected to follow the trends observed in the number of stops per vehicle. In fact the simulation confirmed this as seen from Tables 6.3 to 6.8, wherein while on TH-169 fuel consumption increased without control from 34% to 55%, while on I-94 the latter ranged from 2% to 47%. Pollutant emissions followed the same trend as above.

Finally, the figures at the end of this section present graphs of speed, volume and density vs. time vs. space for the test site on TH-169 for the control vs. no control scenarios on all three days. Similar graphs for the test site on I-94 are shown. As can be seen most visibly from the speed graphs, ramp-metering smoothes out the flow considerably resulting in substantial reduction of stops and pollution levels.

6.6 Metering Period Only

The following pages contain the second set of tables presenting the same MOEs as before but are extracted only when the ramp meters were (or would have been in the no-control case) in operation. During the metering period TTT displayed similar to slightly higher improvements compared to the entire 2 - 6 P.M. cycle. With the ramp control strategy, the reduction ranged from 8% to 16%. From Table 6.10 the least reduction of TTT on TH-169 with control can be observed on March 21st, while the largest decrease in TTT was realized on I-94 on March 29th. Speed in the control scenarios increased on the freeway mainline by 14% to 33%. The increase in speed on TH-169 ranged from 25% to 33% while on I-94 the increase of speed on the mainline ranged from 14% to 25%.

Total Delay (TD), the most sensitive of all MOE's, ranged widely from 23% to 133%. On TH-169 the decrease of TD on the whole system varied from 34% to 62% without control, while on I-94 it was from 23% to 133%. On the ramps, as mentioned before, delays increase substantially with control as expected.

In all cases while the control strategy was in effect the number of stops was relatively low. When the control was turned off the number of stops had increased tenfold. Specifically, the average number of stops with control during the metering period was in the range of 0.23 to 0.68 per vehicle. In the no-control cases this increased and ranged from 0.37 to 8.68 per vehicle. As in the entire 2 - 8 P.M. simulation period the fuel consumption and pollutants are greatly reduced with control as seen in Tables 6.9 to 6.14. Specifically, on TH-169 scenarios fuel consumption increased, during the metering period, without control from 44% to 61%, whereas on I-94 from 3% to 47%. Pollutant emissions followed a similar trend.

**Table 6.9. TH-169 NB, Metered Period Results
2:30 to 6:30 P.M., March 21st, 2000.**

MOE	Aggregation Level	With Control ♦	Without Control	Percentage Improvement
Total Travel Time (veh-hours)	Mainline	2308	3039	+32%
	Ramps	542	42	-92%
	Entire Site	2850	3081	+8%
Total Travel (veh-miles)	Entire Site	129258	128460	+1%
	Ramps	3050	3052	0
Speed (mph)	Mainline	54.7	41.3	+25%
Total Delay (veh-hours)	Mainline	238	976	+311%
	Ramps	489	0	NA
	Entire Site	727	976	+34%
Average Delay/veh (min)	Mainline	0.531	2.184	+311%
	Ramps	1.424	0	NA
	Entire Site	1.627	2.184	+34%
Total no. of stops ⁸	Mainline	6591	144647	+2095%
No. of Stops per veh	Mainline	0.246	5.393	+2094%
Volume (veh entered)	Entire Site	26809	26819	0
	Ramps	20619	20630	0
Fuel Consumed (gal.)	Entire Site	9435	13795	+46%
CO (kg)	Entire Site	2346	2789	+19%
Pollutants: HC (kg)		154	173	+12%
NOx (kg)		46	59	+30%

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

⁸ A vehicle is considered stopped if it's speed drops below 2.2 mph

**Table 6.10. TH-169 NB, Metered Period Results
2:30 to 6:30 P.M., March 22nd, 2000.**

MOE	Aggregation Level	With Control ♦	Without Control	Percentage Improvement
Total Travel Time (veh-hours)	Mainline	2328	3395	+46%
	Ramps	627	42	-93%
	Entire Site	2956	3437	+16%
Total Travel (veh-miles)	Entire Site	129316	127224	+2%
	Ramps	3064	3068	0
Speed (mph)	Mainline	54.2	36.6	+33%
Total Delay (veh-hours)	Mainline	256	1347	+426%
	Ramps	575	0	NA
	Entire Site	831	1348	+62%
Average Delay/veh (min)	Mainline	0.569	2.992	+426%
	Ramps	1.664	0	NA
	Entire Site	1.845	2.992	+62%
Total no. of stops ⁹	Mainline	9502	234725	+2370%
No. of Stops per veh	Mainline	0.352	8.685	+2370%
Volume (veh entered)	Entire Site	27018	27025	0
	Ramps	20724	20732	0
Fuel Consumed (gal.)	Entire Site	9655	15555	+61%
CO (kg)	Entire Site	2237	3118	+39%
Pollutants: HC (kg)		148	193	+30%
NOx (kg)		44	68	+55%

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

⁹ A vehicle is considered stopped if it's speed drops below 2.2 mph

**Table 6.11. TH-169 NB, Metered Period Results
2:30 to 6:30 P.M., March 23rd, 2000.**

MOE	Aggregation Level	With Control ♦	Without Control	Percentage Improvement
Total Travel Time (veh-hours)	Mainline	2289	3037	+33%
	Ramps	545	41	-92%
	Entire Site	2833	3078	+9%
Total Travel (veh-miles)	Entire Site	128708	127664	+1%
	Ramps	2993	2994	0
Speed (mph)	Mainline	54.9	41.0	+25%
Total Delay (veh-hours)	Mainline	227	987	+334%
	Ramps	494	0	NA
	Entire Site	721	988	+37%
Average Delay/veh (min)	Mainline	0.511	2.219	+334%
	Ramps	1.457	0	NA
	Entire Site	1.621	2.219	+37%
Total no. of stops ¹⁰	Mainline	6329	144187	+2178%
No. of Stops per veh	Mainline	0.237	5.401	+2177%
Volume (veh entered)	Entire Site	26691	26698	0
	Ramps	20338	20345	0
Fuel Consumed (gal.)	Entire Site	9234	13324	+44%
CO (kg)	Entire Site	2170	2772	+28%
Pollutants: HC (kg)		143	173	+21%
NOx (kg)		42	59	+40%

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

¹⁰ A vehicle is considered stopped if it's speed drops below 2.2 mph

**Table 6.12. I-94 EB, Metered Period Results
2:30 to 6:30 P.M., March 29th, 2000.**

MOE	Aggregation Level	With Control ♦	Without Control	Percentage Improvement
Total Travel Time (veh-hours)	Mainline	4043	5293	+31%
	Ramps	785	293	-63%
	Entire Site	4829	5587	+16%
Total Travel (veh-miles)	Entire Site	224426	219202	-2%
	Ramps	10599	10540	-1%
Speed (mph)	Mainline	52.8	39.4	+25%
Total Delay (veh-hours)	Mainline	300	1589	+430%
	Ramps	403	45	-89%
	Entire Site	703	1634	+133%
Average Delay/veh (min)	Mainline	0.330	1.765	+435%
	Ramps	0.505	0.057	-89%
	Entire Site	0.774	1.815	+135%
Total no. of stops ¹¹	Mainline	13627	250149	+1736%
No. of Stops per veh	Mainline	0.250	4.630	+1752%
Volume (veh entered)	Entire Site	54507	54025	+1%
	Ramps	47853	47343	+1%
Fuel Consumed (gal.)	Entire Site	15409	22584	+47%
CO (kg)	Entire Site	3138	3240	+3%
Pollutants: HC (kg)		210	209	-1%
NOx (kg)		62	70	+12%

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

¹¹ A vehicle is considered stopped if it's speed drops below 2.2 mph

**Table 6.13. I-94 EB, Metered Period Results
2:30 to 6:30 P.M., March 30th, 2000.**

MOE	Aggregation Level	With Control ♦	Without Control	Percentage Improvement
Total Travel Time (veh-hours)	Mainline	3647	4513	+24%
	Ramps	544	248	-54%
	Entire Site	4192	4761	+14%
Total Travel (veh-miles)	Entire Site	228049	227882	0
	Ramps	10803	10799	0
Speed (mph)	Mainline	59.6	48.1	+19%
Total Delay (veh-hours)	Mainline	268	686	+156%
	Ramps	243	8	-97%
	Entire Site	511	694	+36%
Average Delay/veh (min)	Mainline	0.289	0.741	+156%
	Ramps	0.299	0.009	-97%
	Entire Site	0.552	0.749	+36%
Total no. of stops ¹²	Mainline	38005	56324	+48%
No. of Stops per veh	Mainline	0.684	1.013	+48%
Volume (veh entered)	Entire Site	55602	55577	0
	Ramps	48871	48837	0
Fuel Consumed (gal.)	Entire Site	15848	18303	+15%
CO (kg)	Entire Site	2934	3333	+14%
Pollutants: HC (kg)		194	218	+12%
NOx (kg)		59	68	+15%

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

¹² A vehicle is considered stopped if it's speed drops below 2.2 mph

**Table 6.14. I-94 EB, Metered Period Results
2:30 to 6:30 P.M., April 3rd, 2000.**

MOE	Aggregation Level	With Control ♦	Without Control	Percentage Improvement
Total Travel Time (veh-hours)	Mainline	3370	3914	+16%
	Ramps	303	243	-20%
	Entire Site	3673	4158	+13%
Total Travel (veh-miles)	Entire Site	218018	218418	0
	Ramps	10548	10561	0
Speed (mph)	Mainline	61.6	53.1	+14%
Total Delay (veh-hours)	Mainline	175	290	+66%
	Ramps	68	8	-88%
	Entire Site	243	299	+23%
Average Delay/veh (min)	Mainline	0.196	0.324	+65%
	Ramps	0.087	0.011	-88%
	Entire Site	0.272	0.334	+23%
Total no. of stops ¹³	Mainline	15640	19698	+26%
No. of Stops per veh	Mainline	0.291	0.366	+26%
Volume (veh entered)	Entire Site	53757	53798	0
	Ramps	47388	47419	0
Fuel Consumed (gal.)	Entire Site	13526	13963	+3%
CO (kg)	Entire Site	2594	2806	+8%
Pollutants: HC (kg)		172	185	+8%
NOx (kg)		51	55	+7%

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

¹³ A vehicle is considered stopped if it's speed drops below 2.2 mph

6.7 Individual Ramp Results

Detailed results for individual metered ramps on each test day for both the sites are presented in Tables 6.15 to 6.20. As mentioned earlier, the MOEs extracted are the total ramp delay, average ramp delay per vehicle, maximum ramp delay, average queue and the maximum queue. This information is presented only for the ramp-metering period since these MOEs were usually zero during the non-metering period. As can be seen from the tables, the MOEs extracted provide a systematic way of identifying problematic ramps i.e., ramps at which queues or delays were excessive. These ramps are marked in each table assuming a five-minute average delay threshold (with a single asterisk) or a ten-minute maximum delay threshold (with two asterisks) or a maximum queue more than the ramp capacity i.e., the number of vehicles the ramp can hold before the queue spills onto the arterials (with three asterisks). For example, vehicles on the entrance ramp from TH-55 WB on TH-169 experienced average delays more than five minutes on March 22nd, 2000. The last threshold can only be considered as an example because the way the ramp demands were measured did not account for possible spill backs that happened on those days. Since it wasn't measured in the field, it could not happen during the simulation either.

For each of the ramps that were identified as problematic, the variation of average delay(/average queue) with respect to time is plotted. These graphs aid in finding out if the delay on the ramps is uniform with respect to time or if the delay exhibits a specific pattern such as regular crests and troughs. For example, as shown in the first Ramp Delay figure, the average delay exceeded the five-minute threshold limit during most of the ramp metering period. Such problems strongly suggest that even though the control strategy is beneficial, the strategy should be modified to reduce driver frustration or spill backs onto arterials. The methodology developed in this study can be used to test the effectiveness of the modified strategy and to refine it further.

Table 6.15. TH-169 NB, Metered Period only, Metered Ramps ♦

2:30 to 6:30 P.M., March 21st, 2000

Ramps	Average Ramp Delay (minutes)	Max Ramp Delay (minutes)	Total Ramp Delay (veh-hr)	Average Queue (veh)	Maximum Queue (veh)
Valley View Rd	0.19	1.27	6.3	2	34
T.H.62 EB	0.43	2.32	16.6	6	48
T.H.62 WB	0.27	1.94	9.2	5	39
Bren Rd	1.01	3.55	22.5	8	43
Lincoln Dr	3.69	8.45	39.0	13	30
Excelsior Blvd	4.20	10.03**	76.0	27	68***
T.H.7	1.17	3.31	22.9	8	27
36th St	2.62	11.40**	30.3	10	47
Minnetonka Blvd	2.66	6.06	26.0	10	33
Cedar Lake Rd	1.20	7.90	9.7	4	33
I-394 EB	0.04	0.06	1.0	5	9
I-394 WB	1.51	5.12	45.8	16	63
Betty Crocker Dr	1.97	7.36	18.3	8	35
T.H.55 EB	2.49	8.27	23.5	13	46
T.H.55 WB	3.97	10.40**	41.9	14	41
Plymouth Ave	2.96	5.76	38.8	14	29
Medicine Lake Rd	4.44	10.02**	61.7	21	49

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

Table 6.16. TH-169 NB, Metered Period only, Metered Ramps ♦

2:30 to 6:30 P.M., March 22nd, 2000

Ramps	Average Ramp Delay (minutes)	Max Ramp Delay (minutes)	Total Ramp Delay (veh-hr)	Average Queue (veh)	Maximum Queue (veh)
Valley View Rd	0.23	1.47	7.9	3	42
T.H.62 EB	0.26	1.68	9.6	3	36
T.H.62 WB	0.42	1.55	14.4	6	32
Bren Rd	0.82	3.17	19.4	7	39
Lincoln Dr	4.56	10.77**	48.4	17	41
Excelsior Blvd	2.52	9.95**	46.0	16	68***
T.H.7	1.46	6.87	29.9	10	50
36th St	0.94	3.98	10.5	3	20
Minnetonka Blvd	2.94	8.44	30.4	14	42
Cedar Lake Rd	0.63	2.73	5.3	4	16
I-394 EB	0.04	0.05	1.0	5	9
I-394 WB	2.49	4.84	76.0	27	58
Betty Crocker Dr	3.08	6.3	28.9	12	32
T.H.55 EB	3.85	11.12**	37.6	18	57
T.H.55 WB	11.61*	21.54**	126.5	44	82***
Plymouth Ave	1.94	4.54	24.6	9	25
Medicine Lake Rd	4.36	9.63	58.3	20	46

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

Table 6.17. TH-169 NB, Metered Period only, Metered Ramps ♦

2:30 to 6:30 P.M., March 23rd, 2000

Ramps	Average Ramp Delay (minutes)	Max Ramp Delay (minutes)	Total Ramp Delay (veh-hr)	Average Queue (veh)	Maximum Queue (veh)
Valley View Rd	0.18	0.70	6.0	2	17
T.H.62 EB	0.22	1.40	8.1	3	31
T.H.62 WB	0.09	0.53	2.8	2	14
Bren Rd	1.38	6.09	29.8	10	65***
Lincoln Dr	5.42*	13.37**	54.5	19	48
Excelsior Blvd	1.96	7.92	33.5	12	44
T.H.7	2.25	10.03**	44.1	15	64
36th St	0.95	3.70	10.6	4	20
Minnetonka Blvd	2.08	5.84	20.8	9	31
Cedar Lake Rd	0.63	1.79	5.6	4	13
I-394 EB	0.04	0.05	1.0	4	8
I-394 WB	2.44	6.88	74.4	26	86***
Betty Crocker Dr	2.87	7.39	27.1	12	36
T.H.55 EB	2.33	6.96	22.8	12	39
T.H.55 WB	3.87	8.57	40.7	14	35
Plymouth Ave	3.09	7.97	39.7	14	43
Medicine Lake Rd	5.22*	12.88**	72.5	25	57***

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

**Table 6.18. I-94 EB, Metered Period only, Metered Ramps ♦
2:30 to 6:30 P.M., March 29th, 2000**

Ramps	Average Ramp Delay (minutes)	Max Ramp Delay (minutes)	Total Ramp Delay (veh-hr)	Average Queue (veh)	Max Queue (veh)
Lyndale Ave.	0.23	1.62	4.0	2	22
Hennepin Ave.	0.47	1.82	24.1	13	58
5th Ave.	0.71	2.08	12.2	7	24
6th Street	3.77	8.62	142.7	70	130***
Cedar Ave.	0.64	3.33	14.0	6	32
Riverside Ave.	0.77	3.61	20.5	10	42
Huron Blvd.	1.24	4.98	25.3	12	43
Cretin Ave.	0.81	3.64	25.1	12	53
Snelling Ave.	0.64	3.11	24.3	13	52
Lexington Ave.	1.33	4.10	51.3	24	64
Dale Street	0.70	2.09	17.2	8	28
Marion Street	1.18	5.66	38.7	18	76

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

**Table 6.19. I-94 EB, Metered Period only, Metered Ramps ♦
2:30 to 6:30 P.M., March 30th, 2000**

Ramps	Average Ramp Delay (minutes)	Max Ramp Delay (minutes)	Total Ramp Delay (veh-hr)	Average Queue (veh)	Maximum Queue (veh)
Lyndale Ave.	0.10	0.28	1.8	1	7
Hennepin Ave.	0.08	0.24	3.8	2	19
5th Ave.	0.57	1.50	9.1	5	18
6th Street	2.41	5.99	92.1	46	101***
Cedar Ave.	0.48	1.68	10.3	5	20
Riverside Ave.	2.01	6.42	54.7	26	65
Huron Blvd.	0.28	0.88	5.9	3	13
Cretin Ave.	0.30	1.19	9.4	4	22
Snelling Ave.	0.27	1.76	10.6	6	33
Lexington Ave.	0.61	2.54	24.6	11	45
Dale Street	0.26	1.50	6.3	3	17
Marion Street	0.42	3.11	14.1	6	44

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

**Table 6.20. I-94 EB, Metered Period only, Metered Ramps ♦
2:30 to 6:30 P.M., April 3rd, 2000**

Ramps	Average Ramp Delay (minutes)	Max Ramp Delay (minutes)	Total Ramp Delay (veh-hr)	Average Queue (veh)	Maximum Queue (veh)
Lyndale Ave.	0.14	1.21	2.2	1	19
Hennepin Ave.	0.05	0.29	2.5	1	18
5th Ave.	0.46	2.67	8.2	5	27
6th Street	0.30	1.59	11.7	7	36
Cedar Ave.	0.23	1.58	4.7	2	20
Riverside Ave.	0.34	2.69	9.2	4	35
Huron Blvd.	0.27	1.91	6.1	3	23
Cretin Ave.	0.19	1.40	5.6	2	27
Snelling Ave.	0.12	0.89	4.7	3	23
Lexington Ave.	0.13	0.94	4.8	2	22
Dale Street	0.14	1.11	3.2	1	16
Marion Street	0.13	0.43	4.1	2	18

♦ Original MN Algorithm as used throughout metering system from March 1997 to October 2000

6.8 CONCLUSIONS

In this research a practical methodology was developed for selecting, testing and calibrating the most suitable integrated ramp control strategy for any freeway. As part of this methodology several enhancements were done on a widely used microscopic simulator to simulate any external user-defined ramp control algorithm, allow operator interaction with the system (such as overriding the ramp meters), and simplify the creation of boundary conditions by automating a previously manual process. The methodology was successfully tested by implementing the Mn/DOT ramp control algorithm on three selected test sites having geometric properties and traffic characteristics that were considered representative of the Twin Cities freeway network. The test sites are: TH-169 NB from I-494 to I-94 interchange, I-94 EB from 7th St. in Minneapolis to 7th St. in St. Paul, and I-35W SB from downtown Minneapolis to TH-13. Extensive detailed traffic data was collected both manually and automatically and at all sites which along with detailed geometric roadway representation and they were modeled in the simulator. However only a preliminary testing was performed on the I-35W test site due to reconstruction. The calibration of the test sites was completed successfully with accuracy exceeding expectations through successive and elaborate iterations.

The results obtained by emulating the Mn/DOT algorithm on the two remaining test sites lead to some general conclusions that appear to be reasonable and even expected. Such findings include a substantial variation in the effects of ramp metering by location and the particular demand patterns on the day the experiments were performed. In short, ramp metering effectiveness depends heavily on the freeway geometry and demand patterns. Even at the same site, ramp-metering decisions vary from day to day depending on traffic patterns. Hence, it was felt that standards such as maximum allowable ramp delays, maximum queue lengths, and others need to be established to determine the effectiveness of ramp metering on a particular site. Another expected result was that ramp delays increase substantially with ramp control while freeway delays decrease but overall system delay decreases.

Beyond such obvious general conclusions, however, the specific results obtained suggest that at least in engineering terms, ramp metering is effective in terms of improving measurable MOEs

such as Total Travel Time, Total Delay and Average Speed for the entire system (mainline and ramps). The results also indicate that the improvements are more substantial on I-94 where the demand patterns and freeway volumes are much higher resulting in severe congestion. In addition, it can be inferred that due to the considerable reduction in the number of stops and increased smoothness of flow, ramp metering should be expected to result in lower accident rates, even though no factual data is available as of now to support this. The specific results also suggest that drivers on some of the ramps experience excessive delays; this can be improved by adjusting the control strategy and retesting through the methodology developed here prior to implementation.

Another unexpected benefit of the developed methodology was the discovery of certain operational problems related to the surveillance and control system deployed. For instance, the simulation revealed malfunctioning detecting equipment and errors regarding its location and operation. Even though these problems are easy to fix they were not noticed earlier; such problems should be expected in real life due to the complexity and the size of the system. The developed methodology is a useful tool for identifying and fixing such problems.

As the developed methodology allows easier simulation of different scenarios and quantification of the results, ramp-metering strategies can now be tested more rigorously prior to deployment. It can also be used for improving Mn/DOT's ramp control logic or testing new ones which alleviate driver frustration by distributing delays more evenly between the freeway and the ramps. This subject along with the expansion of the methodology to freeway corridors is left for future research. The latter can also include dynamic assignment thereby incorporating diversion.

7 FEASIBILITY OF CORRIDOR WIDE RAMP METERING EVALUATION

7.1 Need for the Study

The research and evaluation of the Minnesota Ramp Control strategy as described up to this point was confined to the freeway-ramp system only and therefore did not consider diversion effects [2]. Diversion of traffic onto the arterials could be considerable during the peak periods when there is heavy congestion on the freeways or high traffic demand on the on-ramps and therefore have to be accounted for in the simulation [1]. Due to this limitation, this research team was asked by Mn/DOT to proceed with the next step, i.e., investigate the feasibility of corridor simulation. As a result, a well-instrumented freeway corridor was selected and the feasibility of simulating it was investigated. (In terms of data requirements, data availability and effort involved)

7.2 Study Objectives

The study objectives were to:

1. Develop a systematic methodology for corridor simulation
2. Determine a methodology for estimating OD matrices for corridor simulation from the data available
3. Determine an effective calibration methodology for corridor simulation
4. Attempt to implement the methodology to the selected corridor
5. Identify problems related to implementing the methodology and devise ways to solve them

7.3 Process Overview

The Integrated Corridor Traffic Management (ICTM) corridor [3] was the ITS deployment site for the SCATS [6] control system and as a result is the most heavily instrumented corridor in the Twin Cities. Although in the project proposal two more freeways were named as candidate sites the availability of the advanced instrumentation on the ICTM corridor immediately prompted the research team in choosing that as test site for implementation of corridor simulation. The data needs for corridor simulation were determined and the data availability for the selected corridor was investigated. The most critical step in the entire process was the estimation of traffic demands for the corridor in the form of O/D matrices. This step involves the estimation of Traversal (local) O/D matrices for selected corridor from area-wide O/D information that are available from planning. Estimation of Traversal O/D matrices for the selected corridor turned out to be infeasible with the available data. Consequently two alternative methods were developed to compensate for the lack of data. The feasibility of the implementation of these methods on the selected corridor was explored. The results of the O/D estimation methods are presented here along with the issues encountered in their test implementation. The calibration of the entire corridor simulation model could not be performed due lack of link-volume data on the surface streets. However this data was available for the freeways and consequently the freeways in the ICTM corridor were calibrated. The results of the freeway calibration are also presented.

8 CORRIDOR SITE SELECTION

8.1 Site Selection Criteria

The Twin Cities freeway network is extensive with 500 miles of freeways of which 210 miles are metered [5]. Several of these freeway sections have prominent arterials running parallel to them thereby presenting alternate routes to the motorists using the freeway. Thus several possible choices of freeway corridors were available for the current study. Microscopic simulation is highly data intensive and it requires detailed geometric and traffic demand & traffic control data. This data is readily available for the freeways as a result of the automation in the data collection process, especially with respect to traffic demand data. Similar data for the arterials and other surface streets is usually difficult to obtain or unavailable in several cases and hence the arterials in the selected corridor should contain automated data collection devices along with the freeways.

Four major criteria for site selection were identified. They are as follows:

Existence of non-freeway alternate routes

The corridor should be such that there exists a significant number of non-freeway alternative routes for trips that traverse the corridor. This would be ensured through the existence of parallel arterials (on either side of the freeway) throughout the corridor.

Size of Corridor

The corridor has to be sufficiently large to include short trips and long trips as the majority of diversion is caused by short trips seeking non-freeway routes when the freeway is congested.

The ease of data collection

Data would readily available when the data collection process is automated and the data is stored electronically. Such automatically stored data can be retrieved in order to avoid costly and time-consuming manual data collection. Thus a freeway corridor where the

surface streets are instrumented (with detectors etc.) would clearly satisfy the above criterion.

Reliability of data collected

Manually collected data is generally less reliable than data that is collected through instrumentation unless there is a malfunction in the instrumented system. Thus a freeway corridor that is instrumented is preferable to a non-instrumented freeway corridor

Thus all the above criteria point towards the selection of a well-instrumented freeway corridor for the purpose of this study.

8.2 Test Site Description

Based on the selection criteria specified above the Interstate-494 (I-494) transportation corridor was chosen as the test site. The I-494 transportation corridor satisfied the site selection criterion specified earlier and clearly stood out as the most instrumented corridor in the Twin Cities. The I-494 corridor was part of the Integrated Corridor Traffic Management (ICTM) project undertaken by Mn/DOT in association with Hennepin County and the cities of Bloomington, Richfield and Edina[Ref: ICTM evaluation report]. The objective of this project was to optimize corridor capacity, traffic operations and safety by application of advanced technologies such as adaptive ramp metering, adaptive traffic signals, and motorist information and surveillance systems.

8.2.1 ITS Deployment in the ICTM Corridor

8.2.1.1 Traffic Control System:

The ICTM corridor employs the SCATS [6] adaptive Traffic control system to control and manage 75 signalized intersections and 27 freeway metered ramps. The objective of using the SCATS system was to develop a real time traffic control system that could automatically adjust signal timing plans at arterial traffic lights and freeway ramp meters in response to changing traffic conditions and incidents. The SCATS system also collects periodic traffic data (volumes

and occupancies) from the SCATS detectors installed in the signalized intersection approaches in the corridor.

8.2.1.2 Motorist Information & Surveillance Systems:

An extensive motorist information system was established in the ICTM corridor by freeway and arterial street surveillance and the installation of traveler information field devices. These devices were placed strategically in the corridor and included 2 freeway Variable Message Signs, 9 arterial dynamic message signs, 11 Arterial Closed Circuit TV Cameras and 81 Trailblazing Signs [3]. The freeway variable message signs provided alternate non-freeway routes in the event of an incident or congestion. The arterial dynamic message signs provided complementary route guidance information and the trailblazing signs on local streets aided in directing motorists around problem areas.

The above description of ITS deployment indicates a very high level of instrumentation in the selected test site. The ICTM corridor in fact is the most highly instrumented corridor in Minnesota and is an ideal choice for the present study.

8.2.2 Corridor Boundaries:

The test site is an 8-mile stretch of I-494 transportation corridor (ICTM Corridor) south of the Twin Cities encompassing I-494, sections of I-35 W, TH77 & TH100, four parallel arterial streets and ten perpendicular arterial streets crossing five jurisdictions. The boundaries of the ICTM Corridor are 76th St. on the north side, 82nd St. on the south and from East Bush Lake Road to 34th Avenue East .

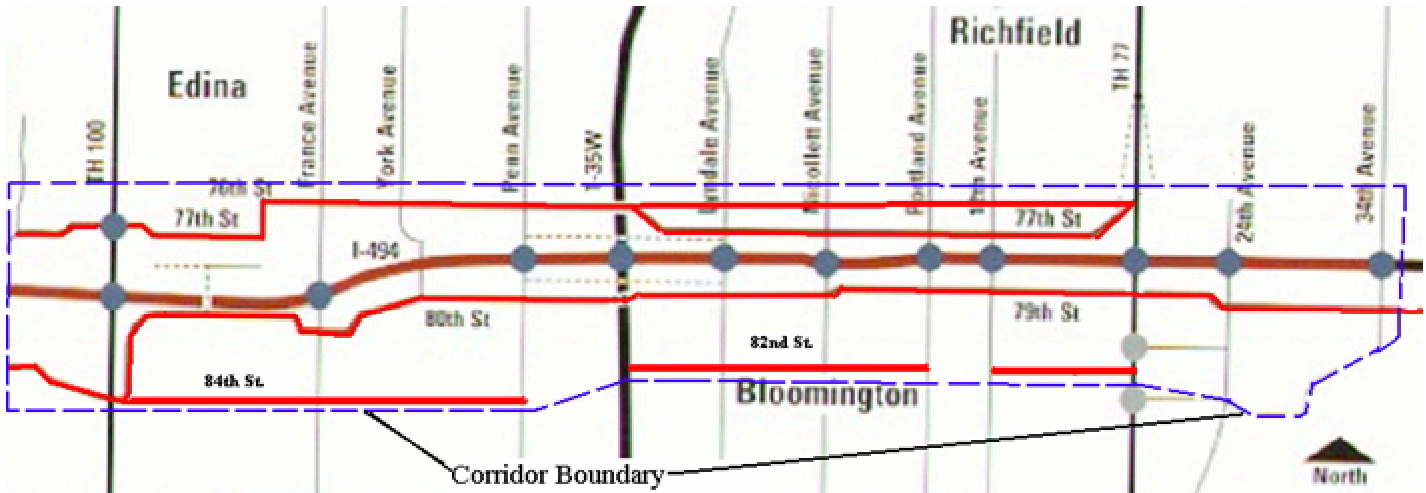


Figure 8.1. Map of the ICTM Corridor

The corridor consists of 75 signalized intersections and 250 un-signalized intersections. Fig 8.1 shows the layout of the ICTM corridor including the names of all the freeways and principal arterials in the corridor.

8.2.3 Freeways in the ICTM Corridor

The ICTM corridor is defined by an 8 mile stretch of I-494 in both directions and includes 3 perpendicular to 494 freeways, I-35W, TH77, and TH100 for approximately 1 mile each in both directions.

8.2.4 Principal Arterials in the ICTM Corridor

The principal arterials that run parallel to Interstate 494, the primary freeway in the corridor are:

1. 76th Street: Between TH77 & France Avenue
2. 77th Street: Between TH77 & I-35W
3. 79th Street: Between TH77 & I-35W
4. 80th Street: Between I-35W & TH100

The principal arterials that run perpendicular to Interstate 494 are as follows:

1. 34th Avenue: Between E 78th Street & E 80th Street
2. 24th Avenue: Between E 78th Street & Killebrew Drive
3. 12th Avenue: Between E 76th Street & 82nd Street
4. Portland Avenue: Between E 76th Street & E 80th Street
5. Nicollet Avenue: Between W 76th Street & W 82nd Street
6. Lyndale Avenue: Between W 76th Street & W 82nd Street
7. Penn Avenue: Between W 76th Street & W 82nd Street
8. York Avenue: Between W 76th Street & W 80th Street
9. France Avenue: Between W 76th Street & W 80th Street
10. East Bush Lake Road: Between W 78th Street & W 84th Street

8.2.5 Geometric Properties of the Freeways in the ICTM Corridor

Table 8.1. Interstate 494 (I-494):

Property	I-494 Eastbound	I-494 Westbound
Length (km)	13	13
Freeway Category	Metropolitan Area Beltway	Metropolitan Area Beltway
Upstream Boundary	E Bush Lake Road	34th Ave
Downstream Boundary	34th Ave	E Bush Lake Road
Metering Period	AM & PM	AM & PM
Metering Zones Contained	4D, 4E, 4F	4G, 4H
Bottleneck Locations	4D: Xerxes Avenue 4E: Nicollet Avenue 4F: TH 5	4G: Xerxes Avenue 4H: East Bush Lake Road
# of On-ramps	14	14
# of Off-ramps	13	13
# of Weaving Sections	3	3
# of Lane Drop Locations	0	1
Alternative Routes (Non-Freeway)	Present	Present

Table 8.2. Trunk Highway 77 (TH77):

Property	TH77 Northbound	TH77 Southbound
Length (km)	1.7	1.7
Freeway Category	Circumferential	Circumferential
Upstream Boundary	Killebrew Dr.	76th Street
Downstream Boundary	76th Street	Killebrew Dr.
Metering Period	AM & PM	AM & PM
Metering Zones Contained	None	None
Bottleneck Locations	N/a	N/a
# of On-ramps	2	3
# of Off-ramps	2	4
# of Weaving Sections	1	1
# of Lane Drop Locations	0	0

Table 8.3. Interstate 35W (I-35W):

Property	I-35W Northbound	I-35W Southbound
Length (km)	1.6	1.6
Freeway Category	Radial	Radial
Upstream Boundary	82nd Street	76th Street
Downstream Boundary	76th Street	82nd Street
Metering Period	AM & PM	AM & PM
Metering Zones Contained	None	None
Bottleneck Locations	N/a	N/a
# of On-ramps	3	3
# of Off-ramps	2	3
# of Weaving Sections	1	1
# of Lane Drop Locations	1	0

Table 8.4. Trunk Highway 100 (TH100):

Property	TH100 Northbound	TH100 Southbound
Length (km)	1.7	1.7
Freeway Category	Circumferential	Circumferential
Upstream Boundary	I-494	77th Street
Downstream Boundary	77th Street	I-494
Metering Period	AM & PM	AM & PM
Metering Zones Contained	None	None
Bottleneck Locations	N/a	N/a
# of On-ramps	3	3
# of Off-ramps	3	3
# of Weaving Sections	1	1
# of Lane Drop Locations	0	0

9 CORRIDOR DATA COLLECTION

9.1 Data Needs

Microscopic traffic simulation is highly data intensive, i.e., to create a simulation model of a real traffic network a large amount of data is required. The data required for creating the simulation model can be broadly classified into three categories. They are: geometric data to model the roadway, traffic data to simulate the flow of vehicles and traffic control data to control the traffic flows in the model. Not all the data components specified above are readily available and need to be extracted indirectly. Data Collection is one of the most crucial steps in the simulation process as the accuracy of the data collected ultimately determines the accuracy of the simulation model.

9.1.1 Geometric Data

The geometric data required by the simulator consists of the physical properties of the road networks such as their orientation, dimensions and layout. The corridor geometry has two distinct components namely the freeway geometry and the surface street geometry and the data required for these components are also distinct. Hence they are explained separately in the subsequent sections.

9.1.1.1 Geometric Data for Freeway

The geometric data required to model the freeway are:

The number of lanes on the freeway mainline

Width of lanes in the freeway sections

Length of acceleration and deceleration lanes

Length of entrance and exit ramps

Length of weaving sections

Location of mainline detectors and merge detectors on the ramps

Location of queue detectors on the entrance ramps

Location of lane-add and lane drop sections on the freeway

Location of ramp meters

Location of special purpose (Eg: HOV) lanes on the freeway
Orientation of the centerline of the freeway mainline

9.1.1.2 Geometric Data for the Surface Streets

The geometric data required to model the surface streets are as follows:

Number of lanes on the arterials and local streets
Width of lanes on the surface streets
Layout of intersections
Length of left turn bays at intersections
Distance between adjacent intersections
Permitted turning movements at intersections
Placement of stop signs at unsignalized intersections
Phasing at signalized intersections

9.1.1.3 Sources of Geometric Data

Digital aerial photographs of the road network can be used as a background for entering the corridor geometry. The digital aerial photographs are preferred to AutoCAD drawings as they include more information and are easier to interpret.

Sketches of the freeway containing the precise location of the detectors on the freeway mainline and the ramps.

Sketches of the freeway (drawn to scale) indicating the location of special purpose lanes on the freeway such as high occupancy vehicle (HOV) lanes.

Detailed layouts of the surface street intersections indicating the placement of detectors in the intersection approaches.

Detailed sketches indicating the feasible turning movements at all signalized intersections in the network and the placement of stop signs at unsignalized intersections

9.1.2 Traffic Data

The traffic data primarily consists of the traffic demands in the network and traffic composition in the network. However there are other data components such as vehicle attributes and driver behavior model parameters, which affect the simulation accuracy. Consequently reasonable initial values need to be determined for these parameters.

9.1.2.1 Traffic Demand Data

The traffic demand data for the simulator can be classified into two categories. They are as follows:

Boundary Volumes and turning percentages

Origin Destination Matrices

9.1.2.2 Boundary Volumes and Turning Percentages

The Turning percentages mode of the simulator requires the volumes at the boundaries of the network and the turning percentages at every decision point in the network (such as an off ramp or intersection) throughout the simulation period. This information can be obtained from detector volumes at the boundaries of the network. The turning proportions are obtained from detector volumes at the surface street intersections and off ramps. The objective of the current study is to model diversion at on-ramps and hence this form of Traffic Demand is unsuitable. This is due to the fact that by using Boundary volumes and fixed turning proportions at decision points the upstream demands at on-ramps are fixed regardless of traffic conditions and control decisions. Though this data cannot be used directly it is useful in performing stage 1 of the calibration, which is described in Section 4.2 of Chapter 4 “Calibration Issues in Corridor Simulation”.

9.1.2.3 Origin Destination Matrices

For the current study traffic demands in the form of origin destination matrices are necessary. The O/D flows need to be assigned to the network through a dynamic traffic assignment in the simulator to model diversion. This is ensured by the dynamic traffic assignment model of the simulator, which assigns trips from their origin to destination through non-freeway routes when the freeway becomes congested.

The O/D data required are as follows:

Origin Destination Zones for the network: These are usually obtained from planning. The Origin Destination Zones are areas that have homogeneous trip-making characteristics and all the trips that originate and terminate in the zone are assumed to do so at a specific point in the zone called the “centroid”.

Time-sliced origin destination matrices: These indicate the number of vehicles traveling from every origin centroid to every destination centroid during that time slice. (These matrices are required for the entire period of simulation)

9.1.3 Vehicle and Driver Characteristics

These are parameters required by the simulator to emulate the behavior of drivers in the simulation model and it is important that reasonable values for these parameters are determined. These parameters are modified as a part of the model calibration process but a set of good initial values is required to carry out simulations.

9.1.4 Traffic Control Data

The traffic control data required by the simulator can be classified into two types. They are as follows:

Traffic Control Data for Freeways

The traffic control data for the freeways consist of parameters used by the external application [5] that emulates Mn/DOT’s adaptive ramp control strategy. This application was developed in

prior research at the ITSLAB in the University of Minnesota. Examples of input parameters required by the ramp control program are:

Ramp metering periods for the freeways

Definition of Mn/DOT's metering zones

Bottleneck locations and capacities for each zone

Red times for every metered ramp

Occupancy control stations

Traffic Control Data for Surface Streets

The traffic control data required for the surface streets are:

A program that emulates the control logic of the adaptive traffic control system functioning in the real network

Type of control at every intersection in the network (uncontrolled, pre-timed, actuated or adaptive)

Phase splits and offsets for all (non-adaptive) signalized intersections

9.2 Data Available

The test site chosen for the current study was the Interstate-494 transportation corridor also known as the Integrated Corridor Traffic Management (ICTM) corridor in the Twin Cities. The corridor is relatively large with 4 interacting freeways, 15 major arterials 75 signalized intersections and 250 un-signalized intersections. The data required for simulating a network of this size was considerable. Fig 9.1. represents the map of the corridor. The data available for the ICTM corridor are described in detail in the subsequent sections.

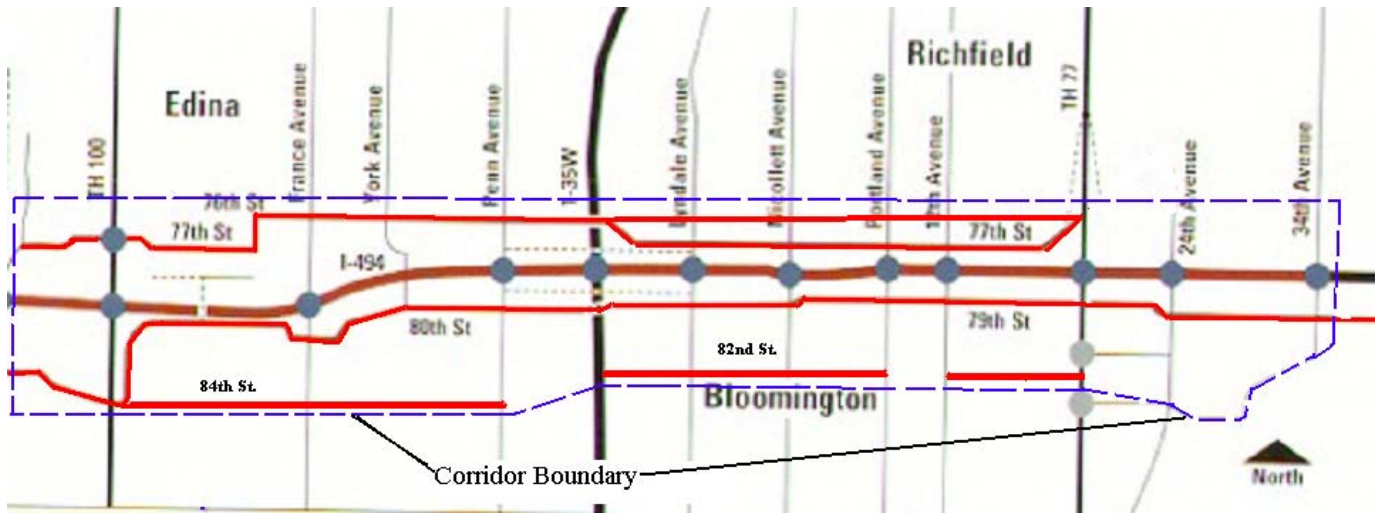


Figure 9.1. Map of ICTM Corridor

9.2.1 Geometric Data Available

High resolution, digital aerial photographs of the ICTM corridor were obtained from the Metropolitan council [7]. These photographs had a resolution of 1pixel = 0.6m and the lane markings on the roads could be clearly identified. They were just sufficient for the purposes of building the geometry in the simulator. Any resolution lower than this would be unsuitable for use as a background for geometry creation.

Sketches of the freeways in the ICTM Corridor indicating the detector placement were obtained from Mn/DOT. These sketches were drawn to scale and in most cases clearly represented the locations of the detectors on the freeways and the ramps as well as the layout of the Mn/DOT metering zones.

The layouts of the signalized intersections in the ICTM corridor were obtained from Hennepin County and the City of Bloomington. These were instrumental in accurately creating the intersections in the simulation model, especially in places where the intersection layout was not clearly represented in the aerial photographs.

9.2.1.1 Advantages of using Digital Aerial Photographs over AutoCAD files

Digital Aerial Photographs are a superior source of geometric information than AutoCAD files for the following reasons:

AutoCAD drawings of freeway segments are not readily available and are normally obtained from construction plans and joined together, which is a tedious process. Putting together a background for a freeway corridor, which contains several arterials and local streets in addition to the freeways, would be practically infeasible using AutoCAD drawings. Digital Aerial Photographs are ideal for geometry entry for freeways and other traffic networks as they can directly be used as backgrounds in simulator.

AutoCAD drawings usually contain parts with geometric information that are missing or difficult to interpret and it may be necessary to refer to alternative sources of geometric information, e.g. the drawings do not describe the lane markings which in some cases differ a lot from the hard boundaries of the street. Such problems are not encountered when Aerial Photographs are used. Thus digital aerial photographs are a superior source of geometric information and their use significantly simplifies the geometry entry in the simulator.

9.2.1.2 Process of Geometry Entry

The first step in the geometry entry process was the creation of a detailed and comprehensive map of the corridor, drawn to scale. This map was developed using AutoCAD 2000 and it contained the following information:

The names of all the freeways, arterials and other local streets in the ICTM corridor

The location of all the signalized intersections in the corridor.

The distances between adjacent intersections in the corridor

The distances along all the freeways in the corridor

The lengths of all the entrance and exit ramps at all interchanges

This map is a highly useful as a reference and simplifies the data collection process. It is also an important aid in the entry of corridor geometry in TEDI.

9.2.1.3 Obtaining Backgrounds for Geometry Entry

Putting together AutoCAD drawings for freeway & surface street segments is a tedious and time-consuming task. The only alternative is to obtain digital aerial photographs or Satellite Imagery of the corridor to serve as backgrounds for entering geometry. The Aerial Photographs need to be of sufficiently high resolution so as to distinguish lane markings on the freeways and arterials. Tracking down a source of such data involved considerable time and effort, as most of the sources of Satellite Imagery & Photogrammetry data of such high resolution were inaccessible due to government regulations. Finally, digital aerial photographs (of sufficient resolution) of the corridor were obtained from the Metropolitan Council of the Twin Cities. These photographs had a resolution of 1pixel = 0.6 m on the ground and were adequate for the purposes of the current study.

9.2.1.4 Preparation of Backgrounds for Geometry Entry

The ICTM corridor was not contained in a single digital aerial photo-quadrangle. Hence portions of the ICTM corridor had to be cut out from about 4 digital aerial photo-quadrangles. These portions had to be cut into smaller portions to meet the memory constraints of the computer. Thus the corridor consisted of 9 background images (bitmaps). These background pieces had to be arranged in the form of a mosaic to create the background for the ICTM corridor. Prior to arranging these background files, the scale to which these photographs had to be set to in “TEDI” had to be determined. This was necessary to ensure that a 1m section of road created in “TEDI” would represent 1m of the road in the background or a 3 lane road section created in TEDI would be in alignment with the 3 lane section on the background. The scale was determined by a trial and error process wherein the background was set to different scale values starting from the default value of 1:500. For each scale value that was set subsequently, a section of road would be created in TEDI and the alignment (lane width) of the section would be compared to its background. The scale was finally determined as 1:3064 corresponding to a dots per inch (DPI) value of 126. The backgrounds were then set to the appropriate scale and were aligned in a mosaic consisting of 9 backgrounds that covered the entire corridor.

9.2.1.5 Entering Geometry in TEDI

After the backgrounds were arranged in a mosaic and set to the appropriate scale, the geometry of the corridor was entered in the Traffic Editor “TEDI” by tracing the backgrounds.

The geometry was entered in two phases. They are as follows:

- Freeway Geometry

- Surface Street Geometry

The geometry entry for the ICTM corridor was accomplished in 3 months. The same process would take considerably longer if AutoCAD drawings or other alternatives were used as backgrounds. Thus the geometry entry process was simplified by the use of aerial photographs

Fig 9.2. is a screenshot of the ICTM corridor in TEDI, which represents the use of digital aerial photographs as a background for development of corridor geometry

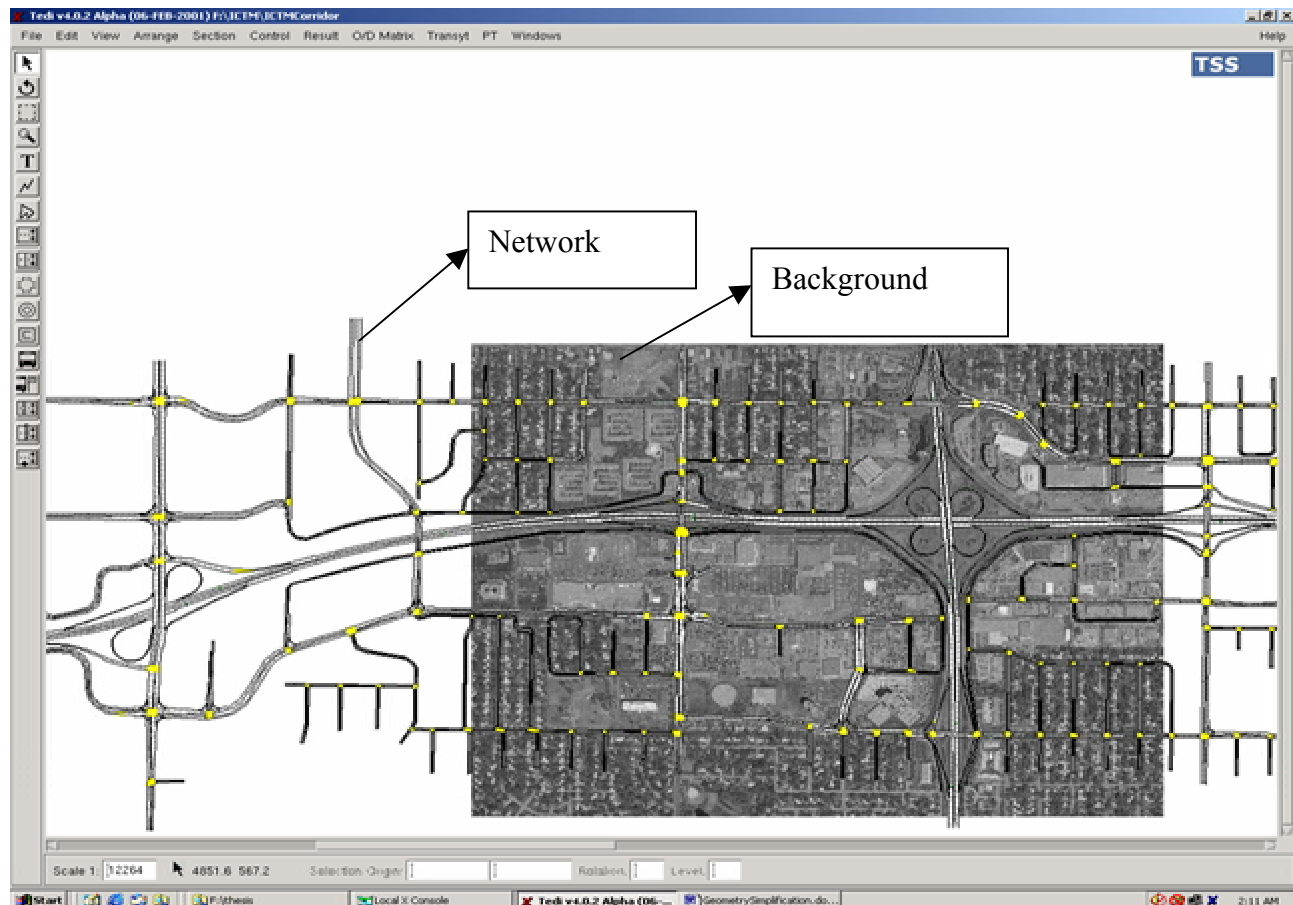


Figure 9.2. Screenshot of digital aerial photographs used as a background in “TEDI”

9.2.2 Traffic Data Available

Five-minute detector volumes and occupancies were obtained for all the freeway detectors in the ICTM corridor. These were extracted for October 11th 2000, just prior to the ramp meter shutoff. The individual detector data is aggregated every five minutes and stored in a database. It is important to note that these detectors cannot determine traffic composition.

Three primary vehicle types were considered: cars, trucks and semi-trailers. Traffic composition data for the ICTM Corridor from the Metropolitan Council. The proportion of trucks and semi-trailers using I-494 were obtained from the Metropolitan Council. It was found to be 7% of the total traffic.

Traffic counts were obtained only for 20 unsignalized intersections in the corridor. There are 250 unsignalized intersections in the ICTM corridor and traffic demands could be obtained only for a small fraction of them.

Annual Average Daily Traffic Data for the year 2000, for the ICTM corridor links was used to estimate hourly volumes on the corridor links as the actual link volumes were not available.

Temporal Distribution of traffic in the ICTM corridor was determined from the AADT values and the hourly detector volumes on the freeways.

The traffic zoning system for the ICTM corridor was obtained from the Metropolitan Council. The Transportation Analysis Zones (TAZ's) defined by the Metropolitan Council [8] were used to represent the zones in the corridor. However certain additional zones were added at the freeway boundaries to account for external trips (these are described in detail in Chapter 5). In all 30 zones were defined for the ICTM Corridor. Fig 9.2. represents the zones in the ICTM corridor.

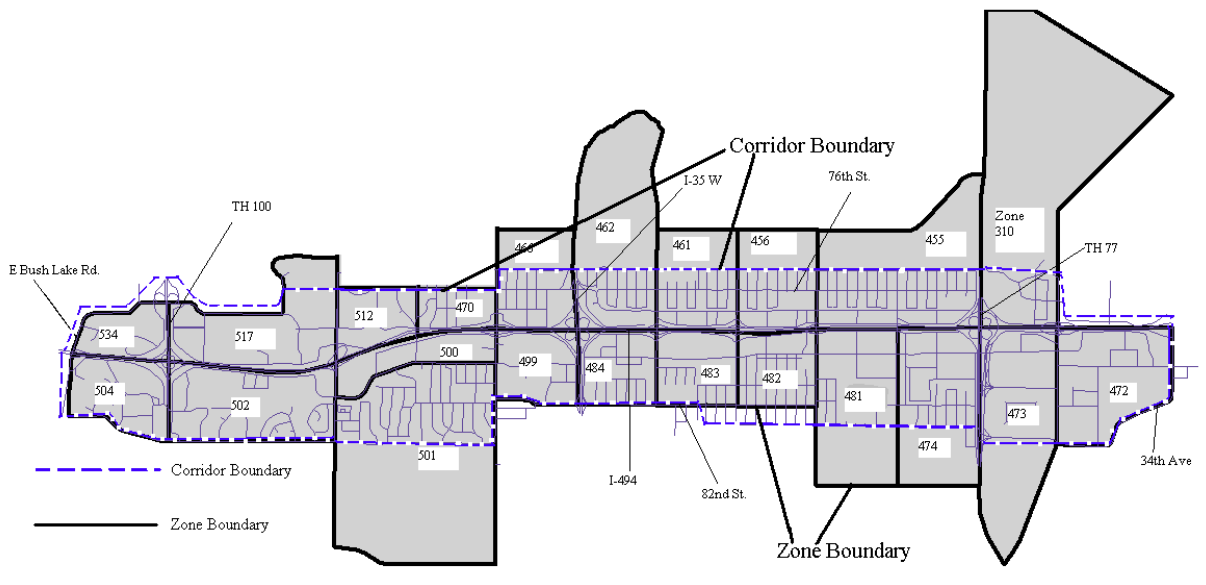


Figure 9.2. Zones in the ICTM Corridor

Preliminary Origin Destination information for the ICTM corridor was obtained from the Metropolitan Council Regional Planning Model. This information included socio-economic data for each TAZ, based on which the trip productions and attractions were determined for those zones and a preliminary OD matrix for the ICTM corridor. These are not the final O/D matrices for the corridor. The final O/D matrices for the corridor are extracted from the preliminary O/D matrix and link volumes.

A reasonable set of initial vehicle parameters was obtained from prior research at the University of Minnesota. These values are given in Tables 9.1, 9.2 and 9.3.

The initial values of other global model parameters such as the car following and lane changing model parameters are given in Table 9.4 and 9.5.

Table 9.1 Preliminary vehicle characteristics for cars

Parameter	Mean	Deviation	Min.	Max.	Units
Length	4.47	0.31	3.81	5.03	m
Width	2.13	0.00	2.13	2.13	m
Maximum Desired Speed	110.00	30.00	95.00	140.00	km/h
Maximum Acceleration	3.00	0.30	2.70	3.30	m/s ²
Normal deceleration	4.00	0.20	2.70	4.30	m/s ²
Maximum Deceleration	7.00	0.20	6.40	7.60	m/s ²
Speed Acceptance	1.10	0.20	1.00	1.30	
Minimum Distance between vehicles	0.60	0.05	0.50	1.30	m
Give Way Time	5.00	1.00	3.00	7.00	sec
Guidance Acceptance	1.00	0.00	1.00	1.00	

Table 9.2 Preliminary vehicle characteristics for trucks

Parameter	Mean	Deviation	Min.	Max.	Units
Length	9.18	0.89	8.08	10.21	m
Width	2.74	0.00	2.74	2.74	m
Maximum Desired Speed	90.00	30.00	88.00	110.00	km/h
Maximum Acceleration	2.10	0.10	1.80	2.40	m/s ²
Normal deceleration	3.00	0.10	1.80	2.40	m/s ²
Maximum Deceleration	5.00	0.10	2.70	3.30	m/s ²
Speed Acceptance	1.00	0.00	1.00	1.00	
Minimum Distance between vehicles	0.50	0.00	0.50	0.50	m
Give Way Time	5.00	0.00	5.00	5.00	sec
Guidance Acceptance	1.00	0.00	1.00	1.00	

Table 9.3 Preliminary vehicle characteristics for semi-trailers

Parameter	Mean	Deviation	Min.	Max.	Units
Length	18.94	3.51	15.24	22.44	m
Width	3.05	0.00	3.05	3.05	m
Maximum Desired Speed	80.00	30.00	70.00	100.00	km/h
Maximum Acceleration	1.60	0.05	1.50	1.70	m/s ²
Normal deceleration	2.40	0.10	2.10	2.70	m/s ²
Maximum Deceleration	3.80	0.10	3.50	4.10	m/s ²
Speed Acceptance	1.00	0.00	1.00	1.00	
Minimum Distance between vehicles	0.50	0.00	0.50	0.50	m
Give Way Time	5.00	0.00	5.00	5.00	sec
Guidance Acceptance	1.00	0.00	1.00	1.00	

Table 9.4: Preliminary Car Following model parameters

Parameter Name	Value	Units
Number of Vehicles	4	n/a
Maximum Distance	100	m
Maximum Speed Difference	50	Kmph
Maximum Speed Difference On-Ramp	70	Kmph

Table 9.5: Preliminary Lane Changing model parameters

Parameter Name	Value	Units
Percent Overtake	0.90	n/a
Percent Recover	0.95	n/a

9.2.3 Traffic Control Data Available

Ramp Control parameters corresponding to the Mn/DOT control algorithm was obtained for the freeways in the ICTM corridor from the Traffic Management Center (TMC) in Minneapolis.

Timing plans were obtained for the signalized intersections in the corridor from Hennepin County and the City of Bloomington. This data was very difficult to obtain (almost 1 year)

9.3 Data Availability Issues

It is obvious at this stage that the data available did not match the data needs of the simulator. In fact a substantial amount of data that is necessary for corridor simulation could not be obtained. The consequences of the data unavailability are discussed in detail in the following sections. The data needs for corridor simulation and the data availability for the ICTM corridor are summarized in Tables 9.6 through 9.8.

9.3.1 Summary of Data Needs & Availability

Table 9.6 Geometric Data:

Data Required	Available?(Yes/No)	Source of Data (if available)
General Freeway & Surface Street Geometric Data	Yes	Digital Aerial Photographs
Freeway Detector Placement	Yes	Mn/DOT Detector Map
Surface Street Detector Placement	Yes	
Ramp Meter Placement	Yes	Mn/DOT
Stop Sign Placement	Yes	Hennepin County
Signalized Intersection Information	Yes	Hennepin County, City of Bloomington

Table 9.7 Traffic Data:

Data Required	Available?(Yes/No)	Source of Data (if available)
Vehicle Characteristics	Yes	ITS Lab, University of Minnesota
Traffic Composition	Yes	Metropolitan Council Heavy Vehicle Data
Traffic Demands: Freeway boundary volumes and turning proportions	Yes	Mn/DOT freeway detector database
Traffic Demands: Surface-street boundary volumes and turning proportions	No	
Traffic Demands: Time-sliced origin destination matrices for the network	(not readily available but are estimated)	Metropolitan Council-Regional Planning Model
Volumes on all the network links (required for calibration and OD estimation)	(available only for the freeway mainline links and ramps)	Mn/DOT freeway detector database

Table 9.8 Traffic Control Data:

Data Required	Available?(Yes/No)	Source of Data (if available)
Minnesota Ramp Control Algorithm Parameters	Yes	Mn/DOT
Signal phasing & timing plans for intersections	Yes	Hennepin County, City of Bloomington
External program emulating Minnesota ramp control logic	Yes	ITSLAB, University of Minnesota
Signal timing plans for intersections	Yes	Hennepin County, City of Bloomington
External program emulating SCATS intersection control logic ¹⁴	No	

¹⁴ The SCATS intersection control logic is proprietary.

9.3.2 Effects of data unavailability

9.3.2.1 Unavailability of traffic volumes on surface streets

This was a serious limitation as the traffic volumes for the surface streets are essential during the calibration process (explained in Chapter 4). They are also essential in validating the Origin Destination matrices (explained in Chapter 5) to be used in the simulator.

Thus the unavailability of traffic volumes for the surface streets caused the following:

Calibration of the entire corridor was not possible. Only the freeways in the corridor could be calibrated

The O/D matrix estimation methodology had to be modified to compensate for the lack of link volume data.

Thus two very important steps in the overall process namely O/D Estimation and Calibration were affected due to unavailability of link volume data for the surface streets.

9.3.2.2 Unavailability of the SCATS control program

Most of the signals in the ICTM corridor are controlled by SCATS [6] an adaptive traffic control system. The SCATS simulator port that emulates SCATS control logic in the simulator could not be obtained and thus the adaptive control had to be approximated by pre-timed signals based on SCATS timing plans.

9.4 Conclusions

Thus at this stage it was established that the data required for performing a rigorous and accurate corridor simulation is not available. However there was sufficient data available to perform a preliminary corridor simulation by making some approximations and assumptions with regard to the data and the methodology. Thus one of the objectives of the study, namely determining the feasibility of corridor simulation was attained. The feasibility of a rigorous corridor simulation is constrained not by the ability of the simulator but by the extent of data availability. The subsequent chapters describe in detail, the problems caused due to the data unavailability in the corridor simulation process and ways to get around these problems.

10 CALIBRATION ISSUES IN CORRIDOR SIMULATION

Validation of the model is a crucial step in any simulation process as it determines its credibility. Validation of a model involves the calibration of the model parameters so as to improve its performance and make it representative of reality. The same rule applies to microscopic traffic simulation wherein the parameters of the network models are modified in an iterative fashion to improve the model accuracy and make them representative of the real traffic networks they are trying to emulate.

Corridor simulation requires extensive model calibration. Not surprisingly, calibration is one of the most difficult phases in the process of corridor simulation. Since calibration has already been discussed earlier in this document in the following sections only the additional considerations for corridor model calibration are to be discussed.

10.1 Calibration Methodology for Corridor Simulation

The calibration methodology for microscopic corridor simulation in AIMSUN2 is described in this section. This methodology consists of two stages corresponding to the two simulation modes in AIMSUN2.

Stage 1:

Calibration in the Turning Percentages Mode of AIMSUN2

The first stage involves the calibration of the simulation model in the turning percentages mode of the simulator. The Turning percentages mode requires the volumes at the boundaries of the network and the turning percentages at every decision point in the network (such as an off ramp or intersection) throughout the simulation period [Ref]. This information can be obtained from loop detectors at the boundaries of the network. The turning proportions are obtained from detector volumes at the surface street intersections and off ramps. The simulated traffic measurements (such as link volumes and speeds) are compared to the actual values after every simulation run and the model parameters are changed based on the fit between the actual and simulated values. The actual traffic measurements are required for carrying out this process. This process is performed iteratively until the simulated volumes and speeds on the links closely

match the actual values. The objective of this stage in the calibration process is to ensure that the actual traffic conditions observed on the particular day of interest are replicated in the simulation model.

Stage 2:

Calibration in the Route Based Mode of AIMSUN2

The second stage is the calibration of the corridor model in the route-based mode of the simulator. The route-based mode requires traffic demands in the form of time-sliced origin destination matrices [4]. Stage 1 is necessary, as the O/D matrices are not estimated on a daily basis but reflect the generalized travel pattern in the corridor. The network model is calibrated first in the turning percentages mode of the simulator and the exact traffic conditions of the day of interest are incorporated into the simulation model. This model is used as the starting point for Stage 2 of the calibration process. Calibration in the Route Based mode of AIMSUN2 involves modifying only the Dynamic traffic assignment model parameters to ensure that the paths taken by the vehicles are such that the O/D flows replicate the actual traffic volumes observed on the corridor links. All the other simulation parameters such as vehicle attributes, driver behavior model parameters etc. have already been calibrated in Stage 1 and are not modified in Stage 2. After Stage 2, is completed , the simulation model is close to reality and can be used for testing various Traffic Management Schemes.

Fig 10.1 describes the ideal calibration methodology for corridor simulation.

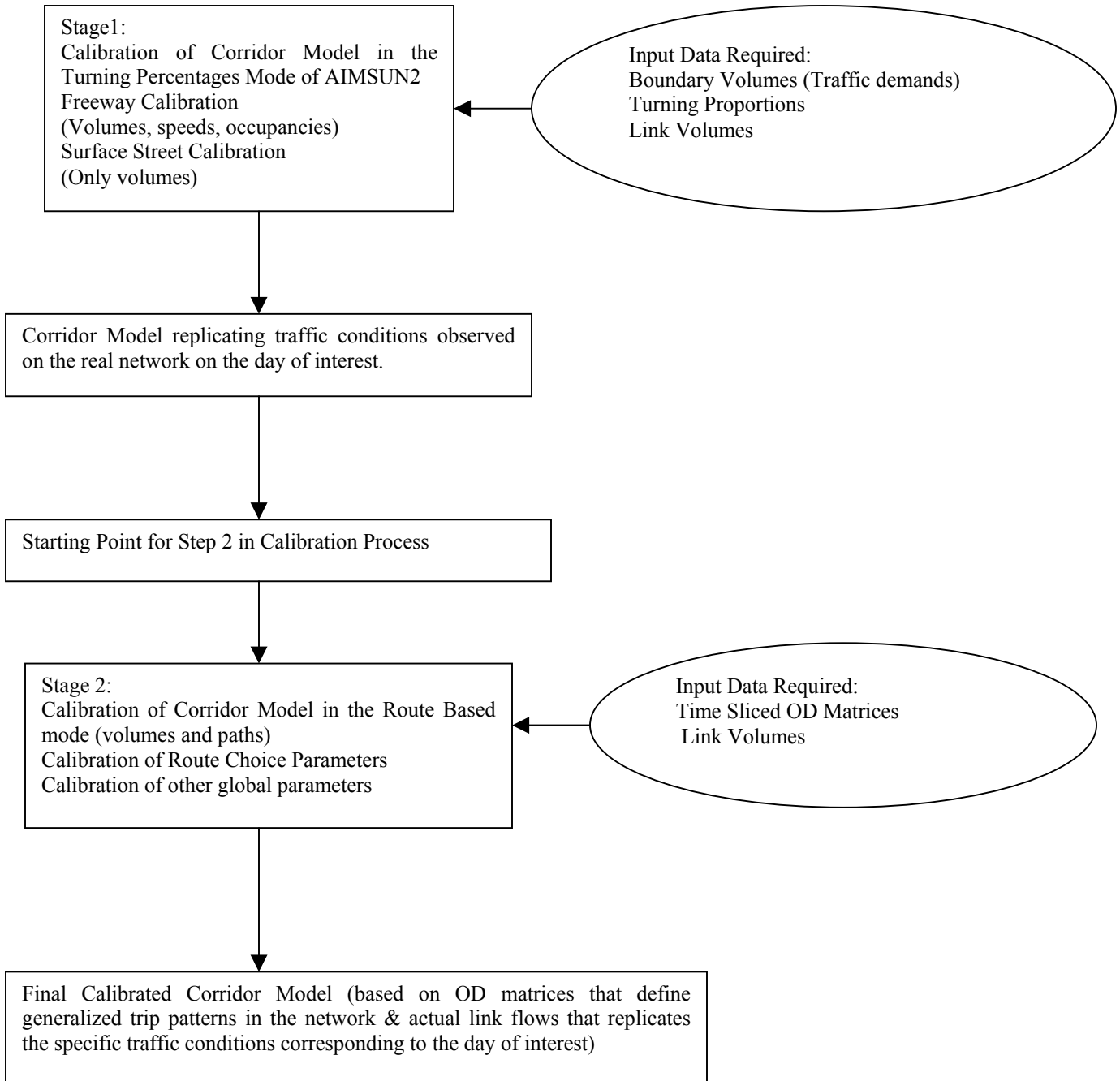


Figure 10.1. Flow chart describing the Calibration Methodology for Corridor Simulation

10.1.1 Calibration Methodology for the Turning Percentages Mode

This calibration methodology consists of three important steps, which need to be performed sequentially in the order specified. They are:

Volume-based Calibration

Speed-based Calibration

Occupancy-based Calibration

Volume-based Calibration is necessary step for both the freeways as well as the surface streets whereas speed-based and occupancy-based calibration is essential only for the freeways. The methodology of the aforementioned steps has already been discussed in earlier chapters.

10.1.2 Calibration Issues and Parameters in the Route Based Mode

The primary objective of calibrating a network in the Route Based mode of AIMSUN2 is to ensure that the origin destination flows are assigned to the network as they are in reality. In other words the paths chosen by the vehicles between any Origin-Destination pair in the simulated network is representative of the paths chosen in the real network. Another aspect of calibration in the route-based mode is to ensure that the actual traffic volumes observed on the links closely match the volumes resulting from the assignment of the O/D flows to the network.

10.1.2.1 Origin-Destination Matrices in AIMSUN2

In the Route Based Mode of AIMSUN2 traffic demand data is input in the form of time-sliced O/D matrices. An origin destination matrix contains a record of the number of trips from every origin centroid to every destination centroid for each time slice and for each vehicle type. Vehicles are generated at each origin centroid and input into the network by means of “To-connectors”. The generated vehicles pass through the network and reach their destination centroid. The vehicles are consumed from the network by means of “From-connectors” associated with the destination centroids.

When a vehicle is generated, the assignment of the vehicle to the objects connected to the centroid (i.e. sections and nodes) can be made in two ways. They are as follows:

Probabilistic Approach: The vehicles generated at the origin centroids are assigned to the various possible input sections and nodes based on the pre-assigned probabilities assigned to those objects by the user.

Destination Dependent Approach: The vehicles generated at an origin centroid are assigned to the input sections and nodes by calculating the shortest path to the actual destination of each generated vehicle and assigning it to that input section or node that lies on the shortest path to its destination.

The probabilistic approach is rarely used in practice as the probabilities associated with the input sections and nodes are usually unavailable.

10.1.2.2 Arc Cost Functions

In the representation of the network used in the route-based mode of AIMSUN2, an “arc” corresponds to a particular turning movement of a section. A cost function is assigned to each arc, which is a function of the travel time in the section including a penalty for the turning movement and the capacity of that arc. Thus the routes between every OD pair comprise a series of arcs and the route cost function is obtained by summing up the cost functions of all its constituent arcs.

These cost functions have weights for the capacity term of the cost function, which can be varied. There is also a provision in AIMSUN2 to specify user defined cost functions for the arcs. At the beginning of the simulation, shortest paths are computed from every section in the network to every destination centroid based on the initial arc cost functions (based on free flow travel time). The shortest paths are calculated based on a Route Choice model. The commonly used route choice models are the Binomial, Multinomial Logit and C-Logit models. Each of these models has a set of parameters that need to be calibrated.

During the simulation, new routes are calculated every time-slice (say 5minutes) that is defined prior to starting the simulation.

10.1.2.3 Dynamic Route Choice or Traffic Assignment

When an OD matrix is loaded into the simulator, it specifies the number of vehicles that travel between a particular origin and destination every time slice (say 15 minutes). Vehicles are generated from the origin centroid based on a pre-specified probability distribution (Uniform, Constant etc.). The vehicles generated thus are assigned to the traffic network based on the initial shortest paths that were computed at the instant the vehicle was generated. However the path assigned to the vehicle may no longer be the shortest path as it moves through the network (as the simulation progresses). Thus new routes are calculated every time interval (specified by the user-say every 5 minutes) and the vehicles can change their route during the trip. Thus information about new shortest routes are available to each vehicle every pre-specified time interval based on which it changes its path during the trip. Thus dynamic route choice can be used to minimize the travel times of the individual vehicles.

10.1.2.4 Route Choice Model Calibration Parameters

The route choice models used widely in AIMSUN2 are as follows:

Binomial Model

Multinomial Logit Model

C-Logit Model

Binomial Model:

A Binomial model employs a binomial distribution $(k-1,p)$ to find the probability of selecting each path. The parameter “k” represents the number of available paths and “p” is the probability of “success” (probability that the specified path is taken). This model does not consider the travel costs in the decision process, but only the time at which the path was calculated. Selecting a small value of p would mean that oldest paths are more likely to be used whereas selecting a high value of p would ensure that new paths are taken more frequently.[4]

Multinomial Logit Model:

In a multinomial logit model, the choice probability P_k of a given path k between a particular OD pair is a function of the measured utilities of that path and all other alternative paths between that OD pair.

$$P_k = \frac{1}{1 + \sum_{l \neq k} \exp (v_l - v_k) \theta}$$

where $V_i(t)$ is the perceived utility for alternative path i at time t . In AIMSUN2 the utility is assumed to be the negative of the current travel time for path ‘ i ’ at time ‘ t ’: $T_i(t)$, measured in hours computed by summing up the cost functions of all the arcs that constitute path i , as explained earlier.

$$V_i(t) = -T_i(t)$$

The parameter “ θ ” in the probability function is called the shape or scale parameter. The scale parameter plays a two-fold role, making the decisions (based on differences in utilities) independent of measurement units, and influencing the standard error of the distribution of expected travel times.

Variance of expected travel times for all paths between an OD pair = $\frac{\sigma^2}{\theta^2}$

Thus when $\theta < 1$, there is a high perception of variance, in other words a trend towards utilizing several alternative routes is observed

When $\theta > 1$, alternative choices are concentrated in very few routes [4]

C-Logit Model:

The multinomial logit model exhibits a tendency towards route oscillations with the corresponding instability creating a flip flop process. The primary reason for this behavior is the inability of the logit function to distinguish between two alternative routes when there is a high degree of overlapping. Thus when alternate routes with a high degree of overlap exist, they are considered as independent choices and the part that is common to the routes is overloaded.

The above instability can be improved substantially by changing the shape factor θ and re-computing the routes very frequently. However this is not an acceptable solution for large networks with substantial overlapping between alternative routes.

To address this limitation, the C-Logit model was introduced. It is a variation of the Multinomial Logit model and the choice probability P_k , of each alternative path k between the OD pair “r-s”, is expressed as:

$$P_k = \frac{\exp[\theta (v_k - CF_k)]}{\sum_{l \in I_{rs}} \exp[\theta (v_l - CF_l)]}$$

where I_{rs} is the set of all available paths between origin “r” and destination “s”

$V_i(t)$ is the perceived utility for alternative path i at time t . In AIMSUN2 the utility is assumed to be the negative of the current travel time for path i at time t : $T_i(t)$, measured in hours computed by summing up the cost functions of all the arcs that constitute path i , as explained earlier.

$$V_i(t) = -T_i(t)$$

The term CF_k denoted as “commonality factor” of path k , is directly proportional to the degree of overlapping of path k with other alternative paths. Thus highly overlapped paths have a larger CF factor and therefore smaller utility factor with respect to similar paths.

CF_k is calculated as follows:

$$CF_k = \beta \ln \left\{ \frac{L_k}{L_l^{0.5} L_k^{0.5}} \right\}^\gamma$$

$$l \in I_{rs}$$

where L_{lk} is the length of arcs common to paths l and k , while L_l and L_k are the lengths of paths l and k respectively. Depending on the values of the parameters β and γ , a greater or lesser weighting is given to the “commonality factor”. Larger values of β imply that the overlapping factor has greater importance with respect to the utility V_i . It is recommended that β is taken in the range $[t_{min}, t_{max}]$ where t_{min} is the minimum travel time on path “i” and t_{max} is the maximum travel time on path “i”. Parameter γ is a positive parameter usually taken in the range $[0,2]$ whose influence is smaller but opposite to that of parameter β [4].

10.1.2.5 Calibration Issues

Calibration of the traffic network in the Route-Based mode involves the modification of the parameters of the cost functions and the Route Choice model to replicate the link volumes observed in reality. Certain issues arise when a network is to be calibrated in the Route-Based mode of AIMSUN2. They are as follows:

The definition of connectors is a very important issue, as it determines the number of alternative paths available between each OD pair. Ideally the connectors have to be defined such that all the feasible paths between each OD pair are available. In practice however this may result in absurd results (such as freeway routes being under-utilized) if the route-choice model parameters are incorrectly defined.

The route choice model used affects the traffic assignment significantly. The simulations are highly sensitive to the parameters of the route-choice model. One of the most critical parameters in the Route-Choice models is the “Cycle” which is the time interval at which new routes are calculated. Caution should be exercised in fixing a value for this parameter. If the value is assigned to the cycle is very low, say 1 minute, the vehicles have too much information and will take the best available paths to their destinations and consequently the network congestion would be reduced significantly. If too high a value is assigned to the cycle, the vehicles do not have enough information and stick to their original (non-optimal) routes thereby causing unwarranted congestion. Thus fixing a reasonable value for the cycle is extremely important and is one of the critical aspects of the calibration of the network in the Route-Based mode.

Certain route choice models have limitations, for instance, the multinomial logit model exhibits a tendency towards route oscillations, with the corresponding instability creating a flip-flop process. The primary reason for this behavior is the inability of the model to distinguish between alternative routes when there is a high degree of over-lapping. Thus such limitations have to be addressed by changing the model parameters to improve the performance of the model.

Only guided vehicles have the ability to change their paths en-route to their destinations, in other words the dynamic route choice models influence only such vehicles. Thus the proportion of

guided vehicles among the total population of vehicles (which is a simulation parameter) is critical in determining the effectiveness of the dynamic traffic assignment.

The issues mentioned above are very general and are applicable to any multi-path traffic network. Several new issues may arise which are network specific when calibration is performed. Since the data required for calibration (volumes on the arterials) was unavailable for the ICTM corridor, these issues could not be identified.

10.2 Implementation of Calibration

All the data required for calibrating the ICTM corridor network was not available. The traffic volumes on the arterials could not be obtained. Traffic counts for most of the local streets were also not available. The only data available were the traffic demands on the freeways in the ICTM corridor. Thus only freeway calibration was possible. Consequently the freeways in the ICTM corridor were calibrated in the turning percentages mode of the simulator (Stage 1 of the calibration process). The methodology specified earlier (in Section 4.1.1) was adopted and the calibration was performed in three stages that are described in detail in the subsequent sections.

The ICTM corridor freeway network consisted of the following freeway sections:

- A 7.5mile (12km) stretch of Interstate 494 Eastbound
- A 7.5mile (12km) stretch of Interstate 494 Westbound
- A 1mile stretch of I-35W Northbound
- A 1mile stretch of I-35W Southbound
- A 1mile stretch of TH77 Northbound
- A 1mile stretch of TH77 Southbound
- A 1mile stretch of TH100 Northbound
- A 1mile stretch of TH100 Southbound

The ICTM freeway network has a total of 45 mainline detector stations of which 6 were malfunctioning. The following table gives the list of malfunctioning detector stations and their location.

Table 10.1 Malfunctioning freeway detector stations

Station ID	Detector ID's	Location
116	1250,1251,1252	Eastbound I-494
493	1537,1538,1539,1540,1541	Eastbound I-494
115	1246,1247,1248,1249	Westbound I-494
191	801,802	Westbound I-494
22	238,239,351	Southbound I-35W
45	285,286	Northbound I-35W

After the initial checks were performed on the input to the simulation model, the volume-based calibration was commenced. The traffic demands were those observed on October 11th 2000, which immediately preceded the ramp meter shutoff in the Twin Cities. The network was calibrated for the PM period (2PM-8PM), as the traffic in the afternoon peak was higher than that in the morning peak for this network.

This calibration methodology was tested earlier for uni-directional freeway segments and was found to be effective but was not tested for a system of interacting freeways. This was the first attempt to calibrate a network of interacting freeways in AIMSUN2. It was noted that the difficulty and complexity of calibration is significantly higher for a network of freeways when compared to a uni-directional freeway segment.

The calibration process was started with a set of good initial parameters obtained from prior research at the University of Minnesota. The initial vehicle attributes & driver behavior model parameters are provided in Tables 9.1 through 9.5 of the Data Collection chapter.

These parameters are global parameters, which affect the entire model. These were calibrated first to ensure that the global effects on the model were optimized before the local effects were considered. The global vehicle parameters that affect the simulated volumes such as maximum

desired speed, maximum acceleration, normal deceleration and maximum deceleration of individual vehicle types had the most significant effect on the model.

The modified vehicle parameters obtained after the calibration were as follows:

Table 10.2 Calibrated vehicle characteristics for cars

Parameter Name	Mean	Deviation	Minimum	Maximum	Units
Length	4.47	0.31	3.81	5.03	m
Width	2.13	0	2.13	2.13	m
Maximum Desired Speed	110.0	30.0	88.0	140.0	km/h
Maximum Acceleration	4.00	0.30	2.70	4.30	m/s ²
Normal deceleration	4.70	0.10	2.70	5.30	m/s ²
Maximum Deceleration	5.70	0.20	5.40	7.60	m/s ²
Speed Acceptance	1.10	0.20	0.70	1.30	
Minimum Distance between veh	1.00	0.50	0.50	1.80	m
Give Way Time	2.00	1.00	1.00	3.00	sec
Guidance Acceptance	0	0	0	0	

Table 10.3 Calibrated vehicle characteristics for trucks

Parameter Name	Mean	Deviation	Minimum	Maximum	Units
Length	9.18	0.89	8.08	10.21	m
Width	2.74	0	2.74	2.74	m
Maximum Desired Speed	90.0	30.0	88.0	100.0	km/h
Maximum Acceleration	2.0	0.10	1.80	2.40	m/s ²
Normal deceleration	3.00	0.10	2.70	3.30	m/s ²
Maximum Deceleration	4.9	0.10	4.70	5.30	m/s ²
Speed Acceptance	1.00	0	1.00	1.00	
Minimum Distance between veh	0.50	0	0.5	0.5	M
Give Way Time	5.00	0	5.0	5.0	Sec
Guidance Acceptance	0	0	0	0	

Table 10.4 Calibrated vehicle characteristics for semi-trailers

Parameter Name	Mean	Deviation	Minimum	Maximum	Units
Length	18.94	3.51	15.24	22.44	M
Width	3.05	0	3.05	3.05	M
Maximum Desired Speed	85.0	30.0	70.0	100.0	km/h
Maximum Acceleration	1.60	0.05	1.50	1.70	m/s ²
Normal deceleration	2.50	0.10	2.10	2.70	m/s ²
Maximum Deceleration	3.70	0.10	3.50	4.10	m/s ²
Speed Acceptance	1.0	0	1.0	1.0	
Minimum Distance between vehicles	0.5	0.0	0.5	0.5	M
Give Way Time	5.00	0	5.0	5.0	Sec
Guidance Acceptance	0	0	0	0	

Table 10.5 Final calibrated values for the Car-Following Model parameters

Parameter Name	Value	Units
Number of Vehicles	5	na
Maximum Distance	100	m
Maximum Speed Difference	45	km/h
Maximum Speed Difference On-Ramp	60	km/h

Table 10.6 Final calibrated values for the Lane-Changing Model parameters

Parameter Name	Value	Units
Percent Overtake	0.92	na
Percent Recover	0.95	na

Local parameters such as section speed limits, distance zone1 and zone2 were then modified in order to bring the simulated traffic measurements as close as possible to the actual values.

10.2.1 Volume-based Calibration in the Freeway network

Volume-based calibration was performed in accordance with the methodology specified earlier (Section 4.1.1). This was an iterative process and around 100 iterations were made to obtain a satisfactory fit between the simulated and observed detector station volumes for the network. The short freeway sections namely I-35W, TH77 and TH100, which have 3 detector stations in each direction, were easily validated and the convergence of the statistical measures was very rapid compared to the longer freeway sections of Interstate 494. The results of the volume-based calibration are given below. These include the average values of correlation coefficients between the actual and simulated volumes, mean percentage error and other goodness of fit coefficients:

Table 10.7 Average Statistical Measures for Mainline Station Volumes

Statistical Measure	Average Value
Root Mean Square Error %	12
R-squared Coefficient	0.87
Correlation Coefficient	0.93
Theil's Inequality Coefficient (U) (zero is good)	0.003
Theil's Bias Proportion (Um) (zero is good)	0.072
Theil's Variance Proportion (Us) (zero is good)	0.045
Theil's Covariance Proportion (Uc) (1 is good)	0.88

The results presented above clearly reflect that the network model has been validated with respect to traffic volumes, i.e., a satisfactory fit has been obtained between the actual and simulated volumes for the network.

10.2.2 Speed-based Calibration in the ICTM Corridor Freeway network

Speed-based Calibration was performed primarily for Interstate 494, as it is the major freeway in the corridor. The other freeway sections are short (about 1 mile in length), therefore elaborate speed calibration was not required for these. An inspection of the percentage errors between the actual and simulated speeds revealed an average error of around 5%, which is satisfactory. About 100 iterations were performed and parameters such as section speed limits, vehicle parameters, car following and lane changing model parameters were modified. 3-D graphs and contour plots were generated for the actual and simulated detector station speeds on the freeways. The 3-D graphs have the Station number on the X-axis, Time on the Y-axis and Speeds on the Z-axis. The 3-D speed distribution graphs for the actual and simulated speeds observed on Eastbound I-494 and Westbound I-494 respectively are as follows:

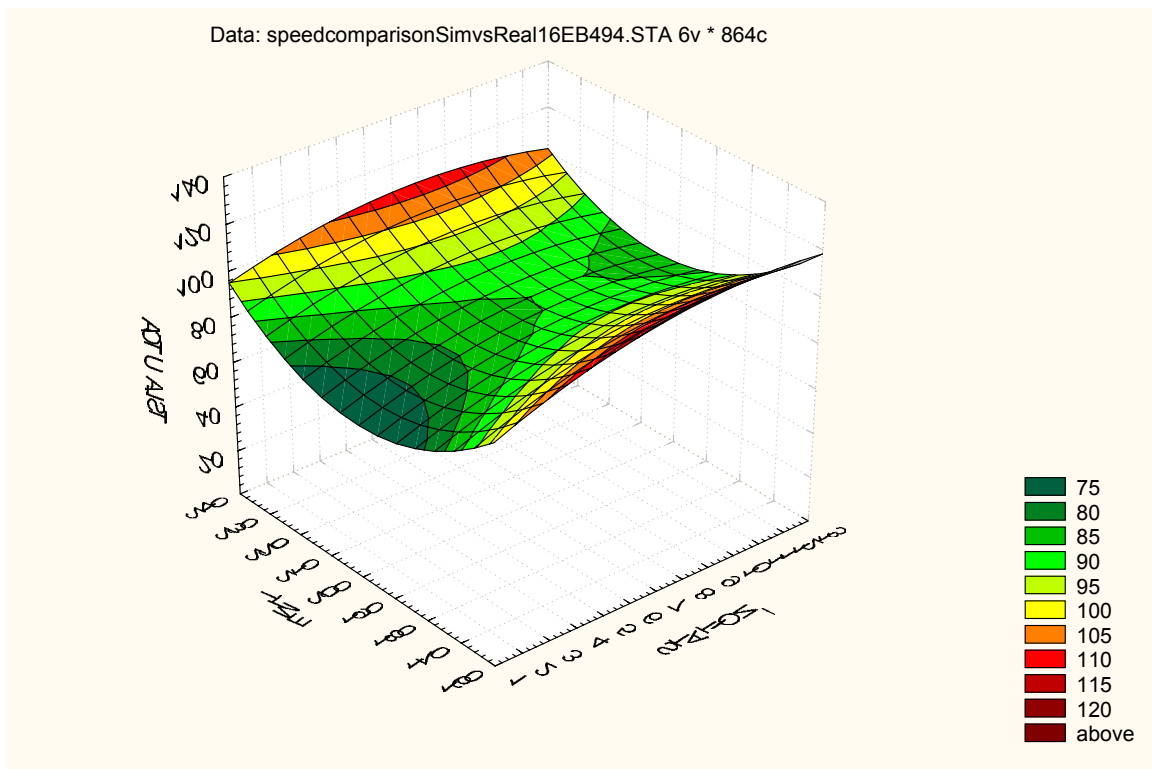


Figure 10.3 Actual Speeds in Km/hr on Eastbound I-494

Data: speedcomparisonSimvsReal19EB494.STA 6v * 864c

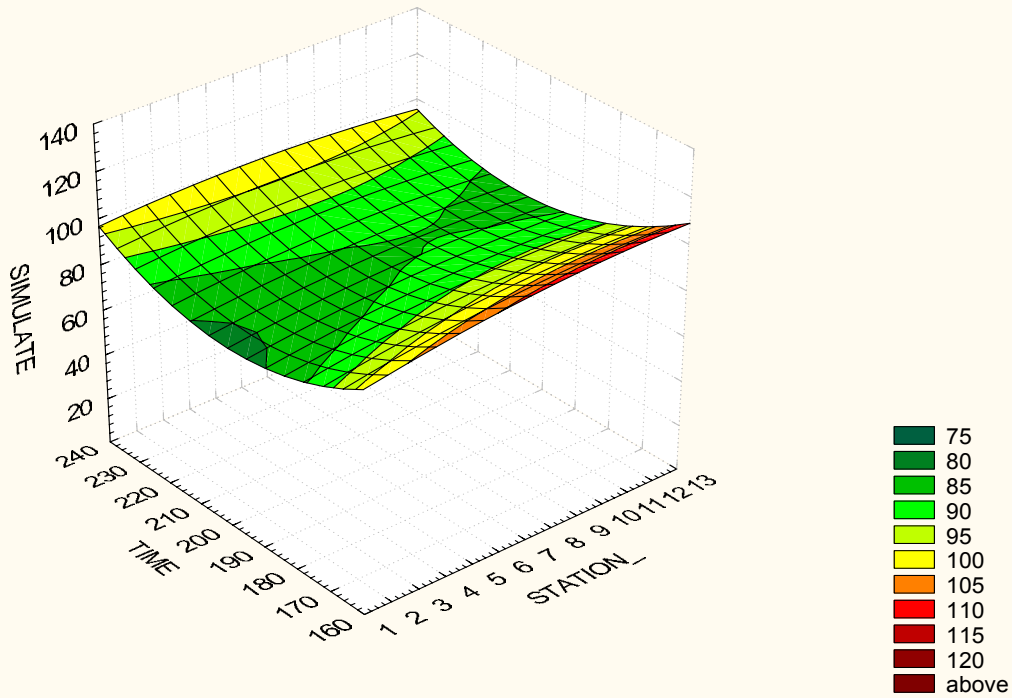


Figure 10.4 Simulated Speeds in Km/hr on Eastbound I-494

Data: speedcomparisonSimvsReal16WB494.STA 6v * 936c

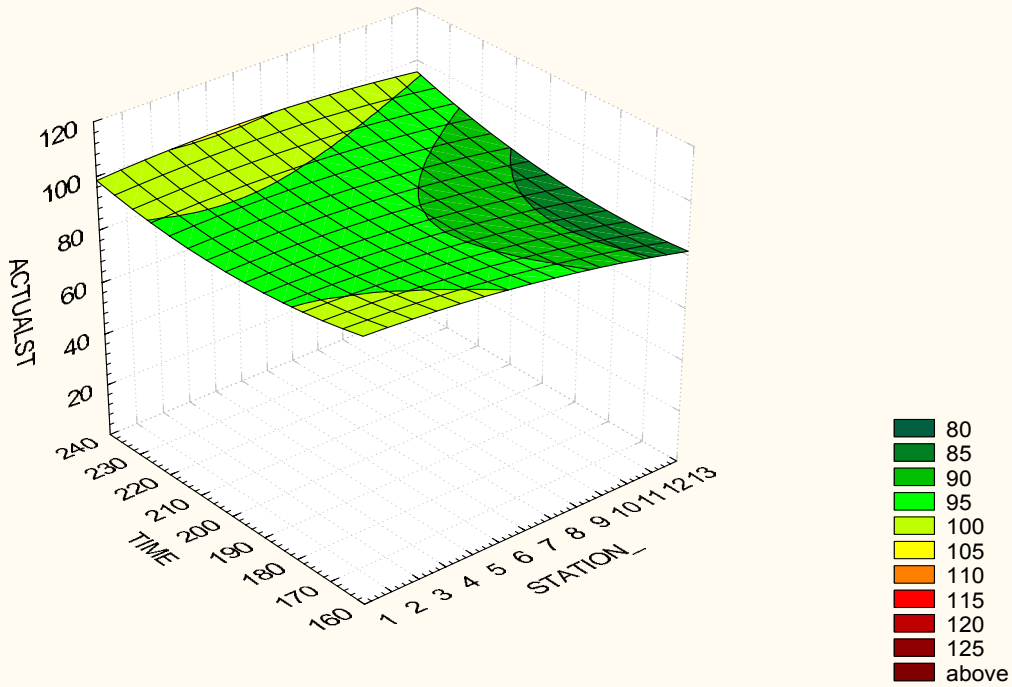


Figure 10.5 Actual Speeds in Km/hr on Westbound I-494

Data: speedcomparisonSimvsReal19WB494.STA 6v * 936c

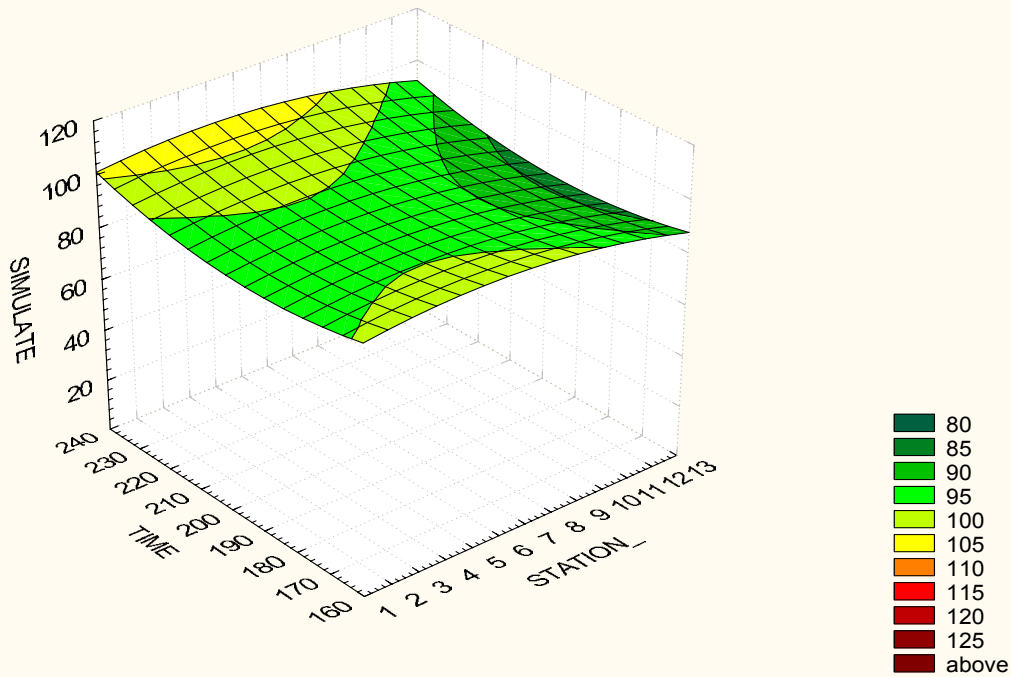


Figure 10.6 Simulated Speeds in Km/hr on Westbound I-494

From Fig 10.3 through 4.6 it can clearly be inferred that the locations of the bottlenecks in simulation are consistent with those observed in reality over both space and time. Also the speed profiles observed in simulation are similar to the actual speed profiles. The speed calibration was better for westbound Interstate 494 than the eastbound direction. This was due to the fact that there was congestion close to the upstream end of the eastbound 494. Congestion near the upstream end of the freeway is very difficult to replicate in the simulation model. Merely reducing the speed limit of those sections would not only be insufficient but also inconsistent with reality. Thus the global simulation parameters were modified iteratively until the three dimensional speed profiles from the simulations closely resembled the actual speed profile observed for that freeway. Thus speed-based calibration was successfully carried out for the freeways in the ICTM corridor.

10.2.3 Occupancy adjustment for the ICTM Corridor Freeway system

Occupancy adjustment was performed after the completion of volume and speed-based calibrations. This procedure is relatively easy when compared to volume and speed-based

calibrations. This is due to the fact that the only parameter that needs to be modified is the detector length. Usually a one to one correspondence is observed between the detector length and the occupancy of that detector. A rule of thumb that was followed during occupancy calibration was: to increase the occupancy of a detector by 10% the detector length needs to be increased by 1m. Around 50 iterations were made and a satisfactory fit was obtained between the actual and simulated occupancies. The average statistical measures of the occupancy calibration for the freeways in the ICTM corridor are as follows:

Table 10.8 Average Statistical measures for Mainline Station Occupancies

Statistical Measure	Average Value
Root Mean Square Error %	18
Correlation Coefficient	0.85
Theil's Inequality Coefficient (U) (zero is good)	0.036

10.3 Conclusions

The data available for the ICTM Corridor was insufficient to implement the calibration methodology in its entirety. The data available was just sufficient to perform a rigorous calibration of the freeways in the ICTM Corridor. The freeway network however was extensive and included four freeways that interacted with each other. Earlier freeway calibration efforts were restricted to uni-directional freeway sections, which did not interact with other freeways. Thus calibrating the freeways in the ICTM corridor was a new challenge as the size and complexity of the network was substantially higher than that of uni-directional freeway sections. It was successfully demonstrated in this study that complex freeway networks could be successfully calibrated. However it should be noted that calibration is a time-consuming and challenging task, in the present case the calibration effort took three months and involved close to 250 simulation runs. The effectiveness of the calibration methodology specified here can be tested at the corridor level once traffic demand data are available for the arterials and surface streets.

11 ESTIMATIONS OF O/D MATRIXES FOR CORRIDOR SIMULATION

This chapter describes several methodologies for the estimation of traffic demands for a freeway corridor in the form of time-sliced Origin-Destination matrices. The freeway corridor is a “multi-path” network unlike a freeway, which is a “single-path” network. This means that a number of alternate routes exist between each Origin-Destination pair which, adds another level of complexity to the O/D estimation problem.

The primary reason for seeking a method to derive origin destination by analytical or empirical methods is economic in-viability of O/D data collection. Ideally, accurate origin destination data can be obtained only through comprehensive surveys or elaborate area surveillance techniques. Both these approaches are very expensive when compared to estimation methods that use traffic counts. Traffic counts are collected using detectors or sensors, which are widely used today. The objective in this case is to obtain O/D matrices for a freeway corridor, which is a relatively small study area. In most cases, O/D survey data exist at the metropolitan area scale collected as a part of planning activities [9]. In a majority of the cases such primary origin destination data are too sparse to be used for a small study area. Thus primarily due to the lack of actual origin destination data one needs to employ analytical or heuristic techniques that reasonably estimate O/D matrices for the study area based on routinely collected data such as traffic counts.

11.1 Definition of Terms:

Centroids and Connectors:

For simplicity in modeling, the origin-destination zones are represented by centroids. A centroid can be considered as a traffic source, a traffic sink or both. All trips originating/terminating in a zone do so at the centroid. The centroid is connected to the links in the network by means of connectors. The connectors are classified into two types: “To-Connectors” which load trips from the centroid (source) to the network and “From-Connectors” that absorb trips from the network to the centroid (sink). Fig 11.1 represents the concept of centroids and connectors.

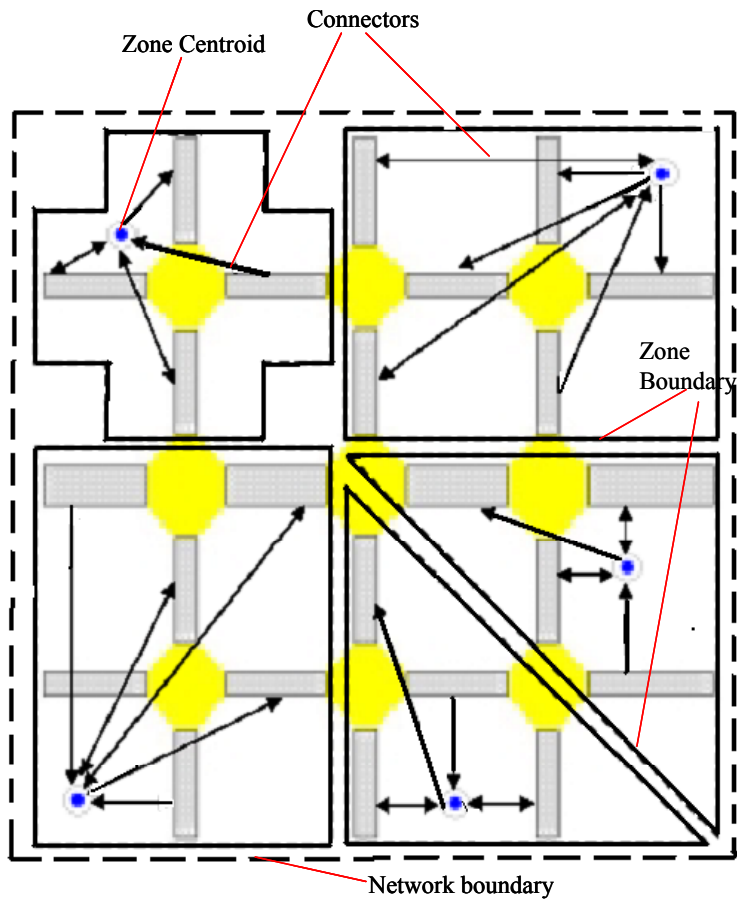


Figure 11.1. Zones, centroids and connectors

Trip Classification:

Internal Trips: These are trips that originate as well as end in the study area

Internal to External Trips: These are trips that originate within the study area but end outside

External to Internal Trips: These are trips that originate outside the study area but end within the study area

External Through Trips: These are trips that neither originate nor end in the study area but merely pass through it. They are also called “Through” trips.

Traversal OD Matrix:

The O/D matrix of a sub-network (corridor) which is extracted from the O/D matrix of the larger network (of which the sub-network is a part) by performing a traffic assignment for the larger network and translating the O/D of all the external trips that use the sub-network (corridor) to the input/output gates at the boundaries of the corridor is called a Traversal O/D matrix. Fig 11.2. represents the concept of a Traversal matrix

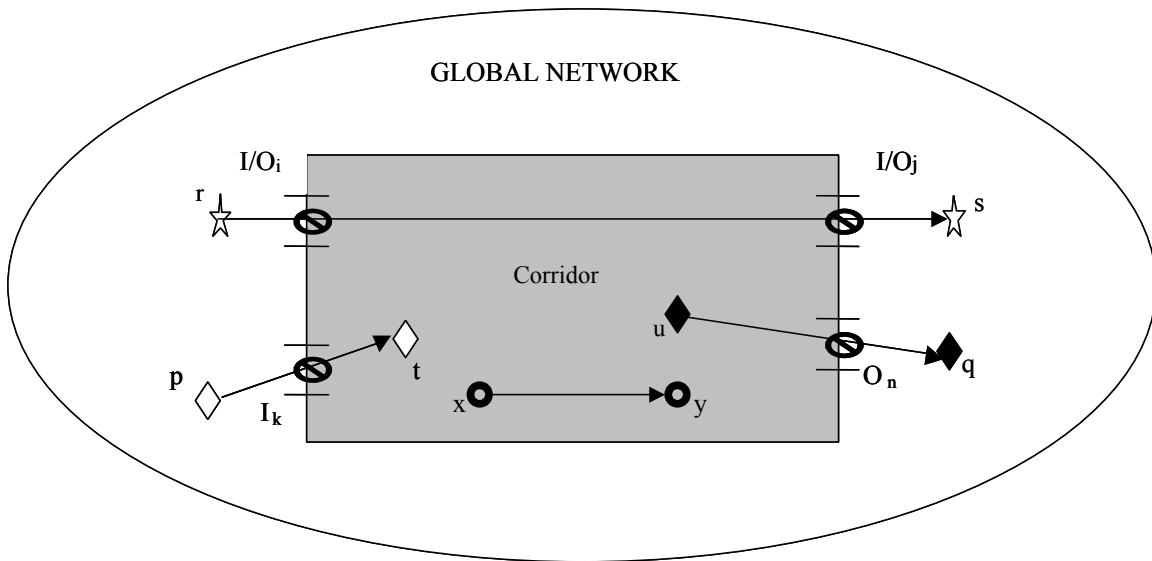


Figure 11.2. Concept of Traversal Matrix

The traversal matrix is the local OD matrix for the shaded area inside the rectangle, which represents the corridor. The traversal matrix is composed of the original Origin and Destination Zones (centroids) in the area (sub-network) along with other dummy origins and destinations, which represent the input and output gates of flows into, from and through the area. In the figure I/O_i and I/O_j correspond to the i -th and j -th input/output gates, i.e., new dummy centroids corresponding to the flows from centroid 'r' to centroid 's' which pass through the sub-network. That is the Traversal matrix for the sub-network would include the flows between external zones "r" and "s" passing through the gates "i" and "j" as an OD flow between centroid "i" and

centroid “j” of the sub-network. I_k is the k-th input gate for the flows with origin at centroid “p”, outside the sub-network, ending the trips inside the sub-network. The flow between “p” and the internal zone “t” would appear in the Traversal matrix as an OD flow between centroids “k” and “t” of the sub-network. O_n is the n-th output gate, for flows generated at a centroid “u” inside the area, leaving the area through this output gate and ending the trip in centroid “q” outside the area. The flow between the internal zone “u” and “q” would appear in the Traversal matrix as an OD flow between centroids “u” and “n” of the sub-network. The flows between internal zones “x” and “y” would remain the same as in the internal trip O/D matrix, i.e., as a flow between zone-centroids “x” and “y”.

11.2 Traversal O/D Estimation Methodology:

There are no sound analytical procedures known to date for estimating time sliced O-D matrices for complex traffic networks. However from a practical viewpoint, there is a heuristic approximate procedure based on empirical data that can provide useful estimates. The heuristic methodology is based on the following assumptions:

The network for which the time sliced OD matrix is to be estimated is a sub-network of a larger network for which an approximate time sliced OD matrix is known.

Traffic counts are available on a significant number of links of the study network.

The procedure consists of the following steps:

1. Starting from a global OD matrix for the larger network for a time horizon (such as the whole day, morning peak etc.), information pertaining to the time distribution of trips is used to generate time sliced (such as hourly) OD matrices for the larger network.
2. The network model is created in a planning tool (such as EMME2).
3. The sub-network (corridor), which is of interest is identified and its boundaries are marked in the network model.

4. Given an OD matrix for the whole area, traffic assignment is performed on the main network.
5. The traversal OD flows are estimated by identifying the traffic flows occurring between the gates defined by the boundary of the sub-network and assigning those traffic flows as OD flows between the zone/dummy centroids in the sub-network, i.e., the sub-matrix corresponding to the selected sub-network is extracted from the global OD matrix and the flows resulting from the traffic assignment of the global OD matrix.
6. The traversal matrix is extracted from a global OD matrix whose information corresponds to an average long-term representation of trip patterns, therefore it could have significant deviations with respect to the actual trip patterns for the time interval under consideration. If information about current traffic flows on a significant number of the links in the sub-network are available, this information can be used to adjust the local OD matrix and get a better representation of the trip patterns.
7. The method used for adjusting the local OD matrices is an ad-hoc version of the Speiss bi-level optimization method. [10]. This method involves the calculation of a sequence of OD matrices that consecutively reduce the mean square error between the traffic flows estimated by the traffic assignment and the observed traffic flows.
8. The resulting adjusted time-sliced OD matrices are the final OD estimates from this method and can be input to the simulator.

The flowchart in Fig 11.3.elucidates the above methodology.

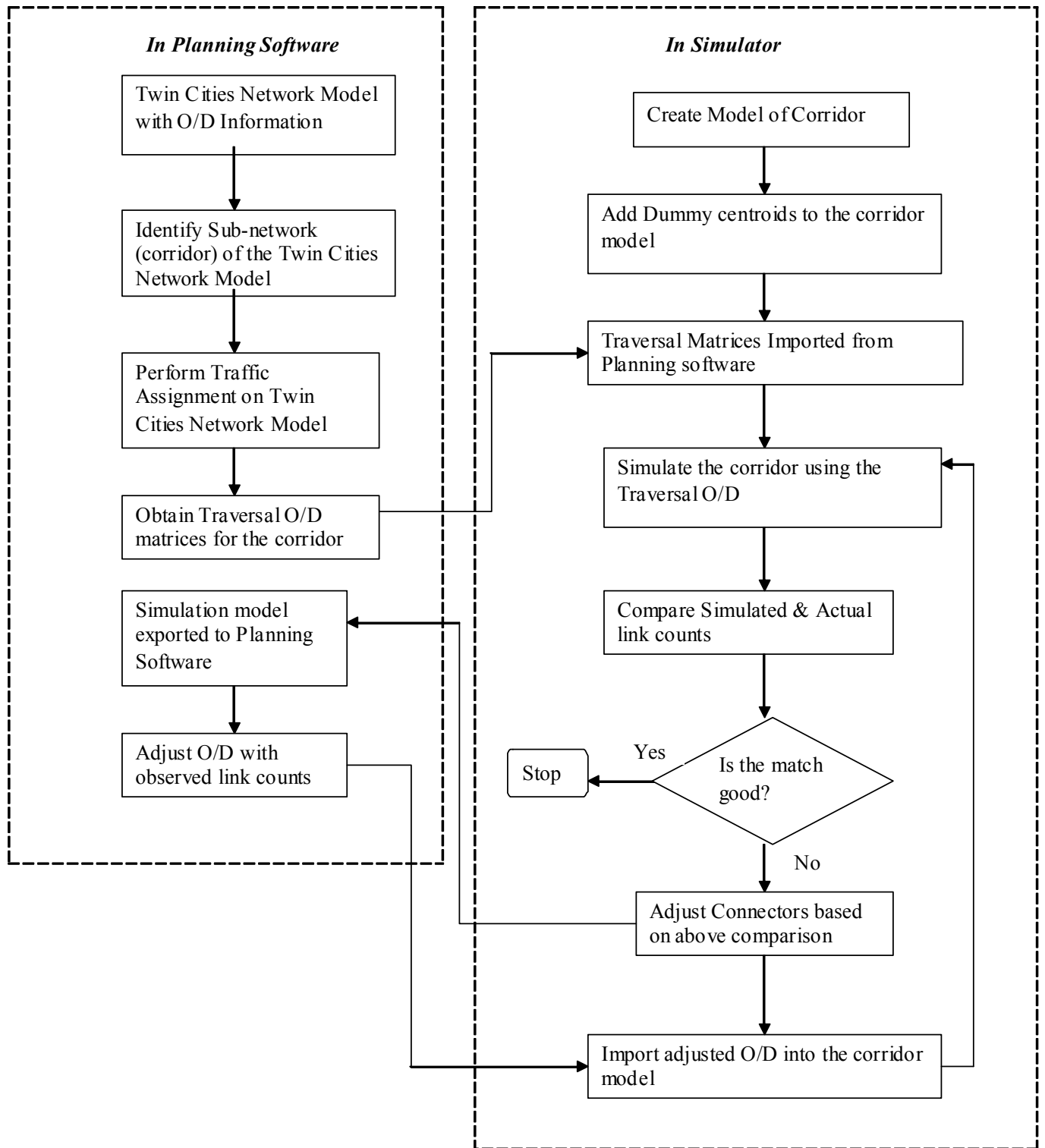


Figure 11.3. Flow-chart describing Traversal O/D Estimation Methodology

11.3 Issues Encountered in Implementation of Traversal OD Estimation Methodology for the ICTM Corridor

The estimation of a traversal matrix requires information about the routes used by trips contained in the global OD matrix and the trip proportions using each route between all origin destination pairs in the global network. Such a large amount of information is very difficult to handle and store in traffic databases. Thus a computer application is necessary to estimate traversal matrices. Currently EMME2 is the only Transportation Planning package, which has this feature.

Thus to implement the above methodology for the ICTM corridor, the entire Twin Cities network (global network) needs to be modeled in EMME2. This is a time-consuming and cumbersome task and could not be performed under the scope of this study. The Metropolitan Council has coded the Twin Cities road-network model in TRANPLAN, another Transportation Planning tool. Unlike EMME2, TRANPLAN does not have the capability of estimating traversal matrices. The network in TRANPLAN cannot be translated into EMME2. Thus the Twin Cities network could not be coded in EMME2 and consequently the methodology specified earlier could not be used for the estimation of time-sliced OD matrices for the ICTM Corridor.

Even if the Twin Cities network could be coded in EMME2, link counts on many of the links in the ICTM corridor could not be obtained. Thus even if an initial traversal matrix were estimated, sufficient link volume information was not available to adjust the traversal matrices based on the bi-level optimization algorithm.

Thus the methodology could not be adopted in its original form to estimate time sliced OD matrices for the ICTM corridor. Alternative methodologies are proposed for the OD matrix estimation for the corridor, which are described in detail in the subsequent sections.

11.4 Alternative OD Estimation Methodologies

Two alternative methods were devised based on available data. They are as follows:

Metro-council based method

AADT Volumes method

These methods are explained in detail in the subsequent sections.

11.4.1 ICTM Corridor Zones

The ICTM Corridor is divided into zones based on the Metropolitan Council's Transportation Analysis Zones (TAZ's). The TAZ's are defined in such a manner that they have homogeneous socio-economic characteristics, such as a shopping zone, or households etc. This is to ensure that the trips making characteristics in each zone is uniform. The TAZ's do not coincide with the boundaries of the corridor and the corridor zones were modified accordingly. Fig 11.4. represents the ICTM corridor zones. 22 zones were defined in the ICTM Corridor.

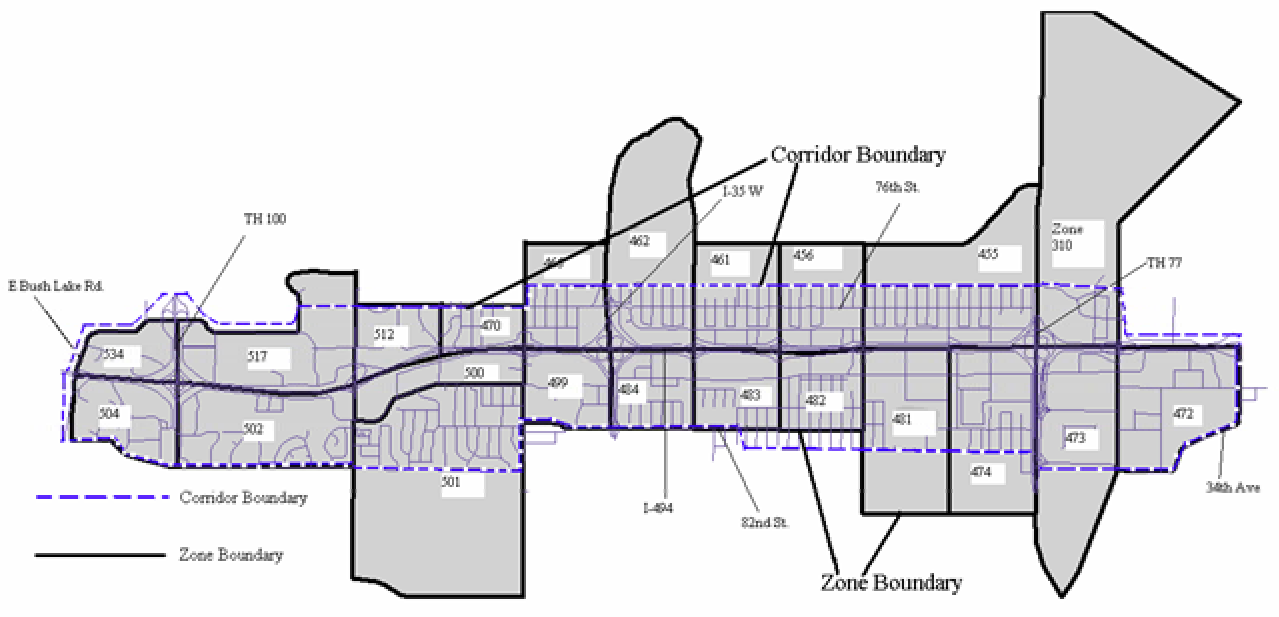


Figure 11.4. ICTM Corridor Zones

11.4.2 Metro-council based method

This method uses the 4-Step Planning Process applied to the Twin Cities Regional Planning Model developed by the Metropolitan Council [8,11] to estimate Origin Destination matrices for the Twin Cities network, from which, O/D data for the specified corridor are extracted.

The following steps explain the methodology:

1. Time-sliced (hourly) O-D matrices are obtained for the entire Twin Cities network for the simulation period. Hourly O-D matrices for the corridor are obtained representing all the internal trips in the corridor (trips originating and terminating within the corridor)
2. Since actual traffic volume data for the links in the corridor are not available, an equilibrium traffic assignment is performed for the entire Twin Cities network in a planning software and the resulting traffic volumes for the corridor links are assumed to represent the actual link volumes.
3. Another traffic assignment is performed in a planning software for only the internal trips in the corridor and the resulting link volumes compared with the corridor link volumes observed as a result of the traffic assignment for the entire Twin Cities network.
4. Based on the above comparison of link volumes in the two scenarios, an external trips percentage is obtained and the internal trips O/D matrix for the corridor is adjusted to account for external trips. Thus the initial O/D matrix for the corridor is estimated.
5. A Traffic assignment is then performed in a planning software for the initial O/D matrix (obtained in step 4) and the resulting corridor link volumes compared to the corridor link volumes from the equilibrium assignment for the Twin Cities network (obtained in Step 2).
6. Based on the comparison an error measure such as root mean square error is evaluated. A cutoff value of 15 % is specified for this error measure to determine a satisfactory fit between the estimated and actual volumes.
7. If the error measure is higher than the cutoff value, the O/D matrix is adjusted through an optimization process and a traffic assignment performed in a planning software with the adjusted O/D matrix. The error measure is computed again and compared to the cutoff value.

The above procedure is performed iteratively until the error measure has a lower value than the specified cutoff value. The O/D matrix obtained at this stage is a good approximation of the actual O/D matrix for the corridor. The following flow chart in Fig 11.5. elucidates the above methodology:

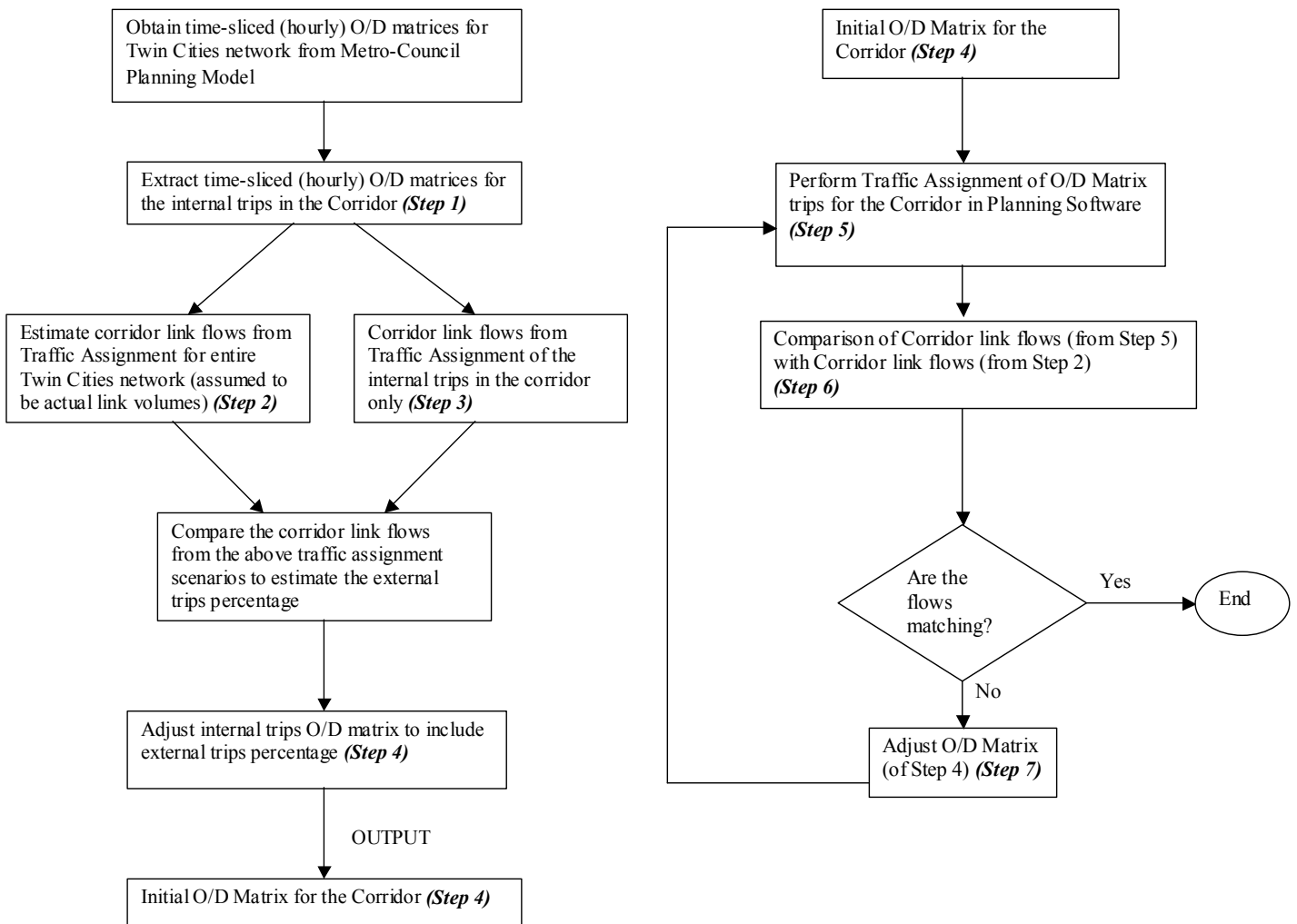


Figure 11. 5. Flow Chart describing the Metro-council based methodology

11.4.3 Problems encountered in the Implementation of the Metro-council based Methodology

The link volumes observed in the arterials of the corridor after the Equilibrium Traffic Assignment for the entire Twin Cities Network in TRANPLAN were very low (of the order of 100 vehicles an hour). This is due to the fact that the Traffic Assignment model used by Metro-Council is validated primarily for the freeways. Thus the corridor link volumes obtained from the Twin Cities equilibrium Traffic assignment could not be assumed representative of the real link volumes for the corridor.

Thus the above methodology could not be implemented to obtain a reasonable estimate of the OD matrix for the corridor.

11.4.4 AADT Volumes method

This methodology is a variant of the previous methodology. It differs primarily from the Metro-Council based methodology in that it estimates the actual volumes on corridor links from Annual Average Daily Traffic (AADT) data.

The steps in the methodology are as follows:

1. As before (in the Metro-council based method), the internal trip O/D matrix is obtained directly from the Metro-council Planning model.
2. The external trip OD matrix is determined by including additional dummy centroids at the boundaries of the corridor to account for external trips. The Metro-council gravity model is used to distribute trips from the dummy centroids (freeway boundaries) among the zones of the corridor. The resulting OD matrix is the external trip OD matrix for the corridor
3. The internal trip O/D matrix and the external trip O/D matrix are combined together to determine the initial OD matrix for the corridor. (This is performed for every hourly time-slice)
4. Reasonable estimates of the actual traffic volumes for the corridor links need to be obtained. The best available approach is to use Annual Average Daily Traffic (AADT) data for the corridor links and convert them into hourly link flows using a suitable temporal distribution of traffic.
5. An equilibrium traffic assignment for the corridor is performed in a planning software based on the Initial time-sliced OD matrix estimate (obtained in Step 3)

6. The arterial link volumes resulting from the traffic assignment (obtained in Step 5) were compared with the corresponding estimates from the AADT values (obtained in Step 4).
7. Based on the comparison an error measure such as root mean square error is evaluated. A cutoff value of 15 % is specified for this error measure to determine a satisfactory fit between the estimated and actual volumes. If the error measure is higher than the cutoff value, the O/D matrix is adjusted through an optimization process and an equilibrium assignment performed in the planning software with the adjusted O/D matrix. The error measure is computed again and compared to the cutoff value.

The above procedure is performed iteratively until the error measure has a lower value than the specified cutoff value. The O/D matrix obtained at this stage is a good approximation of the actual O/D matrix for the corridor.

The following flow-chart in Fig 11.6 describes the methodology

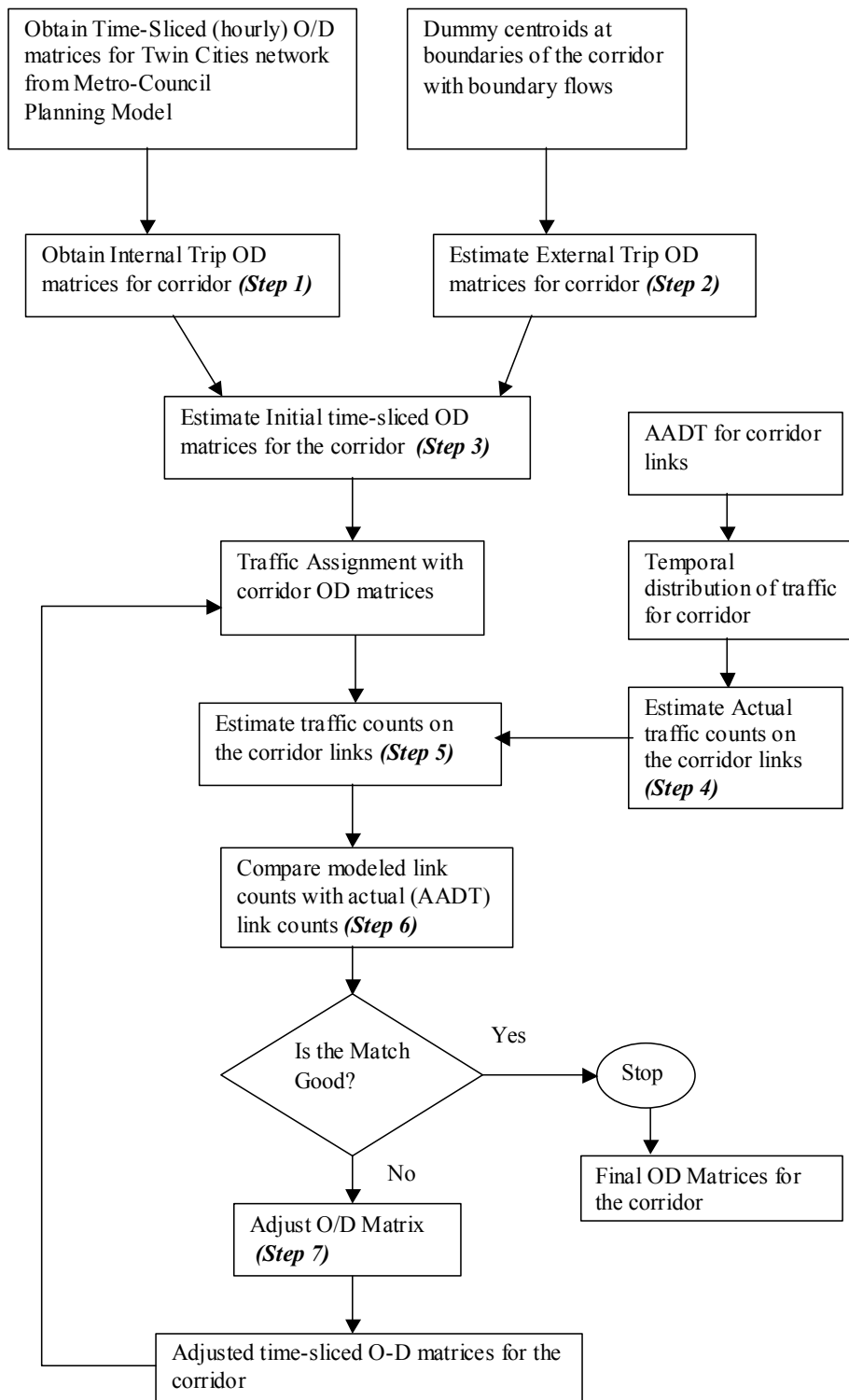


Figure 11.6. Flow Chart describing the AADT Volumes methodology

11.5 Evaluation Experiment

After the Traversal O/D matrices are obtained for the corridor, they need to be loaded into the simulator. Prior to this, the zone centroids and the connectors in the corridor need to be defined. The time-sliced O/D matrices are to be loaded into the simulator. The simulations are to be performed using the Dynamic Traffic Assignment model of the simulator. Pre-timed signal control is to be implemented in the corridor based on the signal timing plans obtained for the 75 intersections. Two different scenarios are to be simulated: ramp-control & no ramp-control. The freeway entrance ramps demands would be compared in the two scenarios to estimate the diversion caused by ramp metering. Observations will also be made to determine conditions (such as time of day, specific arterial traffic conditions etc.), which significantly affect diversion. Ramp metering effectiveness will be evaluated by simulating and comparing the MOEs for the ramp-control and no ramp-control scenarios.

11.6 CONCLUSIONS

In this task the need for corridor wide evaluation of ramp metering has been established. The objective was to define a methodology for such an evaluation and explore it's feasibility. Based upon information in the literature and after extensive meetings with experts in the field a methodology that promises to yield the most accurate results has been defined. This methodology depends on the availability of an accurate planning model for the entire urban area and the integration of such a model with planning software capable of extracting the traversal O/D matrix for the selected corridor. In the case of the Twin Cities the aforementioned methodology is not feasible, mainly because of the unavailability of such a software tool. To compensate for this alternative methodologies were explored.

Two alternative methodologies are proposed in this task. Both of these methodologies seek to estimate OD matrixes from more general information like a traffic assignment for the whole urban area or based on the Annual Average Daily Traffic. The first of these alternative methodologies failed to produce satisfactory results due to the nature of the available planning model. Due to time and budget constraints the second methodology was not tested. Nevertheless, it is clear from the research conducted that without having measured link volumes on a sufficient

number of urban links there is no way to determine the accuracy of any of these methods since all of them depend on O/D information that are usually many years old. Therefore, for this research to be continued in the future a way of obtaining recent link counts must be found. With this information available, following the methodologies defined in this study will yield the most accurate results.

REFERENCES:

1. “Twin Cities Ramp Meter Evaluation”, Final report prepared for Mn/DOT by Cambridge Systematics, Inc. Oakland ,CA, (2001)
2. John Hourdakakis, Panos G. Michalopoulos, “Evaluation of Ramp Control Effectiveness in Two Twin Cities Freeways”, Transportation Research Board 02-2675, (2002)
3. “Integrated Corridor Traffic Management Final Evaluation Report”, prepared for Mn/DOT by Booz-Allen & Hamilton, (2000)
4. AIMSUN2, Version 4.0., User Manual, Draft Version, Transport Simulation Systems,(2001)
5. Muralidhar Koka, “Assessing the Effectiveness of Ramp Control on Twin Cities
a. Freeways” ,Thesis report, University of Minnesota, (2001)
6. <http://www.mrwa.wa.gov.au/traffic/its/scats.htm>
7. http://gis.metc.state.mn.us/order_info_doq.asp
8. “1990 Travel Behavior Inventory Summary Report-Twin Cities Metropolitan Area”, Metropolitan Council, Minnesota,(1994)
9. Anthony F. Han, Richard G. Dowling , Edward C. Sullivan and Adolf D. May, “Deriving Origin Destination Information from Routinely Collected Traffic Counts Volume II”, Research Report 81-9, Institute of Transportation Studies, University of California – Berkeley,(1981)
10. H.Speiss, “ A Gradient Approach for the O/D Matrix Adjustment Problem,”, Publication # 693, Centre de Recherche sur les Transports, Universite de Montreal, (1990)
11. “ 1990 TBI based Trip Production Model Calibration” ,Technical Memo prepared for Metropolitan Council by Barton- Aschman Associates, Inc., Minneapolis, MN, (1993)

12 DETECTION OF INCIDENT PRONE CONDITIONS

12.1 Problem Statement

The population of urban areas has continued to grow unabated even as we make the transition from the 20th century to the 21st. This increase in population coupled with ever-increasing automobile usage has had a debilitating effect on transportation infrastructure all over the world. A typical household in the United States of America owns an average of 1.85 automobiles. Fifty five percent of peak period travel is carried out under congested conditions. Since 1970, the number of cars on the road has increased by 147%, the number of automobiles per household by 53%, the number of daily vehicle trips by 102%, the number of daily miles traveled by 110% and the number of peak period incidents by 72% [W1, W2]. However, highway capacity has only increased by 6%. Traffic congestion costs the roadway users in the United States an estimated \$100 billion in lost productivity, wastes approximately 2 billion hours and consumes 1 billion more gallons of gasoline annually [W3]. These numbers help in further ratifying the ill effects of increasing levels of congestion. The most immediate upshots of increasing volumes on the roadways are increased travel times, increased delays, productivity losses, increased levels of pollution, increased fuel consumption and most importantly, increased number of associated accidents and fatalities.

It is a well-documented fact that more accidents occur in congested traffic conditions than in free flow traffic conditions [1]. Congestion leads to accidents and accidents lead to further congestion. Thus, it is possible to visualize the relationship between accidents and congestion as a vicious, never ending cycle.

While recurrent congestion can almost always be expected to occur at specific locations during rush hour, non-recurrent congestion, is characterized by the disturbances caused to traffic flow as a result of any random event or incident (collisions, vehicle breakdowns, road repair and other

such unexpected events). Non-recurrent congestion accounts for 50% to 75% of the total urban congestion [1].

Accidents are the major cause of non-recurrent congestion. Around 65% to 80% of non-recurrent congestion occurs due to collisions and accidents and the remaining occurs due to such events as stalled vehicles, material spills etc [1]. If traffic conditions likely to result in accidents could be detected, then the occurrence of accidents should be reduced thereby resulting in reduced congestion and its side effects.

The need for research to detect those conditions that cause accidents becomes even more evident when the number of fatalities and the cost of accidents are considered. Motor crash related deaths are the highest unnatural occurring deaths in the United States. On the average, 42,000 crash-related deaths have occurred every year in the U.S over the past decade [W4]. According to 1998 NHTSA accident statistics [W5], there were 6,335,000 accidents of which 2,029,000 resulted in injuries, 37,081 resulted in fatalities and the rest in property damage in the United States in 1998. In Minnesota, 65 fatal crashes occurred in the year on Interstate freeways in which 111 vehicles were involved and 179 persons were killed in the year 2000 [W4]. The total average cost per non-fatal accident is \$9,716 (this sum is only reflective of costs borne by the person involved in the accident not including other property damage or administrative costs) [W6]. According to the same reference, the total economic cost of motor vehicle crashes in 1994 was \$150.5 billion. This sum includes property damage costs of \$52.1 billion, lost market productivity costs of \$42.4 billion and medical expenses equaling \$17 billion; the loss of human life is not included in these costs. Crashes where the driver was exceeding the speed limit cost \$27.7 billion in 1994 while each taxpayer was effectively taxed an additional \$144 to pay for crashes.

Having considered the effects of congestion, the number of accidents taking place and the heavy price that is extorted from everyone, the importance of a research undertaking that detects accident prone conditions becomes highly magnified. Hence, by preventing accidents, not only can life and property be saved, but, delays following the accident will be prevented along with secondary accidents that often accompany the first accident, resulting in substantial economic benefits.

Advanced Traffic Management Systems (ATMS) use advanced technologies to assist the operators at Traffic Management Centers (TMCs) in better managing the freeway networks. Incident detection, verification and clearance form integral parts of an ATMS. Much research has been devoted to incident detection rather than detection of accident prone conditions because of the claim that since accidents are random events, detection of these conditions will not be particularly easy or even possible. Although, incident detection does help an operator in identifying locations where incidents have taken place and thus ensuring quick clearing of debris and lane blockage, it cannot prevent the occurrence of incidents (accidents) in the first place.

12.2 Research Objectives

The objective of the task presented in this chapter is to explore the feasibility of identifying those traffic conditions that are likely to cause accidents and to begin to develop a methodology to detect these conditions. The research relies on individual vehicle (microscopic) data (vehicle speeds, lengths etc) and visual observation of traffic (live video surveillance) from a Machine Vision Sensor placed on Interstate 394 East Bound in the Twin Cities just near the exit to Penn Avenue.

Specifically, the scope of the task is to detect visually undesirable traffic flow conditions. Experienced operators at the Traffic Management Center of the Minnesota Department of Transportation were asked to describe visual or other signs of traffic flow that are likely to cause accidents. Such undesirable conditions were then observed on video recordings while simultaneously extracting traffic measurements. This information was then used to develop algorithms for detecting undesirable flow conditions with reasonable success. Such a development can be viewed as the first step towards the identification and detection of accident prone conditions.

A basic consideration in detecting undesirable traffic flow conditions is that the detection task must only rely on common measurable traffic data from field sensors such as volume, occupancy, speed and time headways. Towards this objective, additional traffic flow measurements that can be derived from these data were defined and used in developing undesirable flow conditions algorithms.

13 Data Collection for detection of IPC

13.1 Site Selection Criteria

The methodology proposed for the detection of undesirable traffic flow conditions requires the use of microscopic (individual vehicle) traffic data measurements (speeds and lengths) and video surveillance of the site. Therefore, a major consideration in selecting the site was that that it should have access to both video surveillance and microscopic traffic measurements. Another important factor governing site selection was the occurrence of a number of accidents so that sufficient accident prone conditions could be recorded. The site should also be subject to high traffic volumes and experience many shock waves and undesirable traffic flow conditions. Because of the requirement for microscopic traffic data (individual detections, on-off times, vehicle lengths and speeds) and video surveillance, it was determined that the best way to meet this requirement would be by video detection; individual detections, on-off times, vehicle lengths and speeds could not be accessed or were unavailable from Mn/DOT loop detectors which only provide aggregate measurements of volumes and occupancy. In addition, no Mn/DOT surveillance camera could be dedicated for video detection of the site as they were needed by TMC operators and daily operations at TMC could not be hampered. However, because of the lack of such devices in the Twin Cities' freeways, it was necessary to confine the scope of the research to the only existing installation available on Interstate-394 just east of the Penn Avenue Interchange in Minneapolis. Another factor that played a significant role in selecting this site was the presence of an overpass bridge that could be used for overhead mounting of the camera.

The Machine Vision Sensor (MVS) at this location is the Autoscope Solo, an advanced traffic surveillance system that offers the advantage of wide-area vehicle detection and accurate traffic parameter extraction without the high costs of installation and maintenance. The MVS allowed measurement of detailed individual vehicle measurements such as speed, time headways, vehicle type as well macroscopic measurements such as volume and occupancy over a wide area. It is also capable of detecting and verifying incidents as well as stopped vehicles within the field of the camera's view. Last, but not the least, it allows quick and easy no-cost relocation of the

sensors for optimal detection placement as well as the placement of multiple sensors within the field of view of the cameras.

The user of the Machine Vision Sensor can easily set detection zones by making use of an interactive graphics browser. This was another reason for selecting a site that had access to data from a Machine Vision Sensor because it was initially proposed that two detection zones be set on each lane at some distance from each other so that the behavior of a vehicle as it traveled from one point to another on a lane could be studied and an attempt to correlate this behavior with the occurrence of accident prone conditions could be made. In order to allow for a wide area of detection that will permit the placement of a second row of detectors, the camera had to be tilted too close to the horizon therefore reducing the accuracy of the volume measurements due to occlusion from high vehicles (trucks) and the glare from the setting sun. A positioning of the camera on the other side of the bridge might have solved the latter problem but the effort too great to justify it.

13.2 Test Site Description

Interstate-394 has 3 basic lanes 2 reversible high occupancy vehicle lanes. In this study, the video output from two camera positions on the I-394 Eastbound were used. The schematic representation of the experimental setup used is shown in Figure 13.1.

The camera from which the MVS extracts the traffic measurements is mounted on Penn Avenue Bridge and looks down westwards on eastbound freeway traffic. As shown in Figure 13.2, the field of view of the camera extends to around 92 feet of the eastbound traffic on the I-394 freeway. The rectangular boxes over the lanes in figure 13.2 represent the locations of the detection zones. As shown in Figure 13.3, the Machine Vision Sensor camera is placed at a height of 65 feet over the I-394 freeway on the Penn Avenue over pass. Though it was planned to place more than one detection zone on each lane, owing to the absence of a wide-area lens in the Machine Vision Sensor camera, this plan could not be carried out i.e. the view of the camera was quite restrictive. For this reason, the detection methodology was restricted to just one detection zone on a lane. The second camera at the site is only for surveillance and is mounted Wirth

Parkway looks down in a northerly direction on the eastbound traffic. The camera at Wirth Parkway is merely used for additional observation purposes and not for any data collection.

The site carries high traffic volumes throughout the day with recurrent congestion during the morning and afternoon peak periods and relatively low traffic volumes during evenings and weekends. The right lane at the site experiences excessive stop and go traffic during the afternoon peak period. The middle and left lane, though, not as slow as the right lane, do slow down considerably during the peak periods. One reason for the right lane being considerably slower than the other two lanes at the present site is because of the high entering volumes from Penn Avenue (located a little east of the site) and the high volumes exiting from Interstate-394 to Interstate-94 East (located further east from the site). A number of shock waves can be noticed at the site during morning and afternoon peak periods. As a result of these shockwaves this site has an above normal occurrence of accidents therefore covering the last of the requirements defined in the previous section. Unfortunately, the instrumentation available did not produce any video recordings of such accidents.

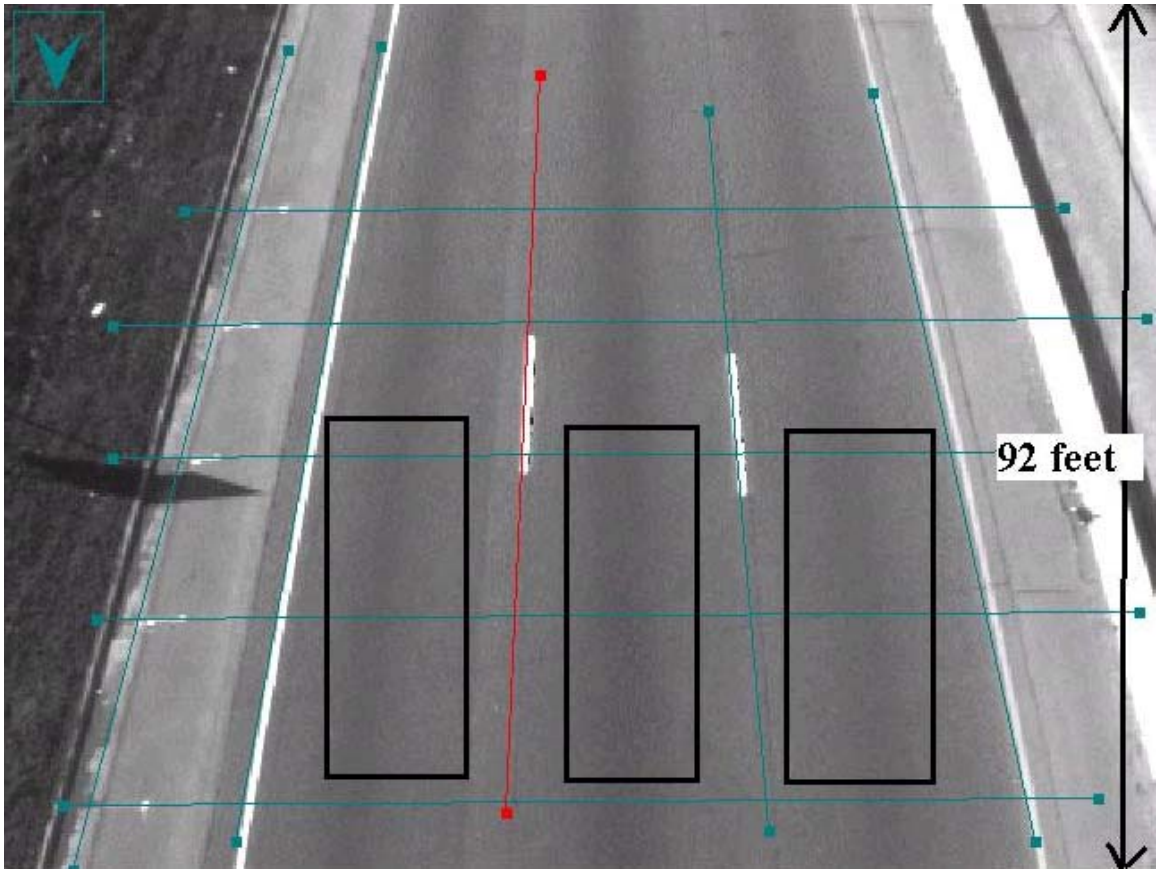


Figure 13.2: Figure showing the field of view of the Machine Vision Sensor and approximate location of detection zones and calibration lines

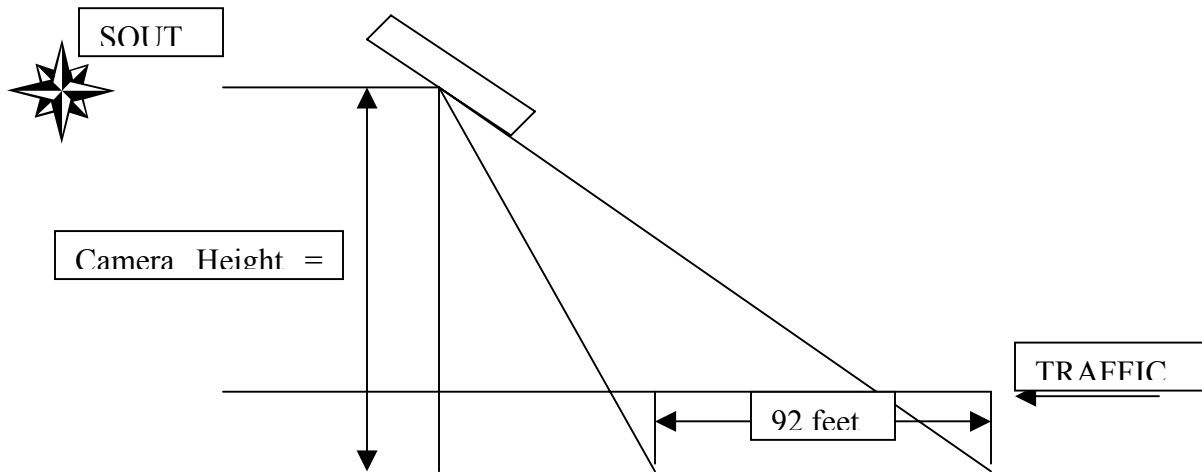


Figure 13.3: Figure showing the placement of the MVS Camera on the Penn Avenue Over-bridge (drawing not to scale)

Table 13.1 gives a fair indication of the traffic conditions on the three lanes at the MVS location near Penn Avenue than has a posted speed limit of 55mph. The table shows representative values of average volumes and average speeds on all the three lanes during only one morning and one afternoon peak period. As can be seen, the right lane is the slowest and most congested lane at the site and the left lane is the fastest and most congested. The morning peak period does not experience as much congestion as the afternoon peak period. Speeds of vehicles on the right lane sometimes drop to as low as 5 mph during the afternoon peak period. While speeds are relatively higher during the morning peak, the same pattern is observed with respect to speed distribution among lanes, albeit to a lesser level. Speeds start dropping at the site at around 07:45 in the morning and around 15:30 in the afternoon. The afternoon peak period almost always experiences more shock waves than the morning peak period in all the three lanes. The right lane also experiences more sudden fluctuations in speed than the other two lanes. Sometimes, during the afternoon peak, congestion on the right lane extends to as far back as Wirth Parkway from the MVS location at Penn Avenue

Table 13.1: Speeds and Volumes at MVS Location near Penn Avenue

Period of Observation	AM Peak Period (05/25/2000, 6:00 – 10:00)			PM Peak Period (05/24/2000, 15:00 – 19:00)				
		Right lane	Middle lane	Left lane		Right lane	Middle lane	Left lane
Traffic Characteristics	Avg. Speed	52.33 mph	58.44 mph	58.57 mph	Avg. Speed	28.54 mph	42.82 mph	48.62 mph
	Avg. Volume	1186 vph	992 vph	1265 vph	Avg. Volume	1249 vph	1481 vph	1889 vph

Even though the freeway experiences periods of stop and go traffic and large fluctuations in speed and is a high accident risk location, it has not had an accident take place within the field of view of the camera in any of the observation periods. However, this was the only location at which data collection was feasible and therefore as explained earlier, the scope of the study was changed to detect undesirable flow conditions only.

13.3 Data Collection

The data collection process involved the video taping of traffic at the test site while simultaneously obtaining real time microscopic (individual) vehicle data measurements from the MVS. The microscopic measurements collected were speeds and lengths of vehicles and detector on (occupied by a vehicle), off times. The macroscopic measurements, volume and occupancy though collected were not used in this study.

The MVS processes the video images from the camera and stores the pertinent information continuously in a text file. The data as stored by the MVS is unformatted. So a program was written in Visual C++ to format the data and derive only the required variables (on-off times, vehicle speeds and lengths). The MVS outputs three types of data, microscopic individual vehicle data, and data aggregated over user specified intervals, which in this case were 10 seconds and 30 seconds.

As mentioned earlier, since there were no occurrences of accidents at the site, undesirable traffic flow conditions that are likely to cause accidents were visually defined. Thus, an Accident Prone Condition was defined as any undesirable traffic condition that was judged to have the potential to lead to an accident. Such a condition is not only dangerous, it is also highly discomforting for freeway users.

After an extensive review of literature and discussion with TMC operators, a list of visible undesirable flow conditions was narrowed down to the following:

1. Sudden drops in speed on a lane.
2. Excessive weaving and lane changing.
3. Short headways at high speeds.
4. Excessive speed differentials between lanes.
5. Excessive fluctuation of speeds among vehicles in a lane.
6. Adverse weather, pavement and visibility conditions.

Since these conditions are qualitative, they are subjective in terms of visual identification. Therefore, the algorithms developed here should be recalibrated by having more than one person observe the data. The underlying basis for identifying accident prone conditions while visually observing traffic was that “Visual Observation of Traffic gives the viewer a fair indication of the dangerousness of driving conditions”. This premise was arrived at based on the fact that operators at Traffic Management Centers make most decisions based on the video images that they view. Experienced traffic operators know when and why accidents are likely to occur just by visual observation of traffic [2]. The algorithm developed in this thesis attempts to capture and formalize this ability of traffic operators so that a real time pro-active approach to detect undesirable flow conditions will be developed.

The list of undesirable flow conditions arrived at is in accordance with the causes of crashes, in particular, rear end collisions [W7]. According to an article in the Volpe Journal, the causes of rear end collisions are almost always poor road conditions, excessive speed, poor road alignment, driver inattention, following too closely, external distraction, and poor judgment.

Because accidents due to failed human judgments are difficult to measure due to the sheer unpredictability of human decision-making and corresponding actions, the list of Accident Prone Conditions being used in the current research only includes measurable traffic conditions that may cause accidents and no other factors.

While data was being collected for analysis, traffic was also visually observed on the video monitors. The objective of this exercise was to try and identify the times at which undesirable traffic flow conditions occurred by manually setting alarms. Based on the six criteria set, traffic was observed and visual (manual) alarms corresponding to undesirable traffic conditions were raised.

Around 56 hours of traffic were visually observed. MVS data was also stored for these 56 hours so that they could be analyzed. Around 20 hours of data were used for development of the algorithm and around 36 hours of data were used for testing purposes. Table 13.2 shows the dates on which traffic was visually observed and data collected for algorithm development purposes (approximately 20 hours of data) and table 13.3 shows the summary of the data that was collected for algorithm testing purposes. Since weather extremes are experienced at the test site (snow, sleet, rain, fog, a wide range of temperatures etc), data was collected and analyzed for dry, wet and snowy conditions.

As can be seen from the table, an average of around 5 manual alarms is raised per lane per hour. More alarms are raised on the left lane because vehicles travel at higher speeds and closer to each other more times than on the other two lanes. Most of the alarms raised on the right lane are those corresponding to sudden drops in speed.

The algorithm developed to define thresholds of certain traffic variables (space headway, traffic pressure and quality of flow index) based on the correlation between the values of these variables and the logs containing the alarms as raised by the visual observations are described in subsequent chapters.

Table 13.2: Summary of Algorithm Development Data Collected

DATES	DURATION OF VISUAL OBSERVATION	CHARACTERISTICS	LANE	NUMBER OF MANUAL ALARMS
			Right	18
05/23/2000	6:00 – 9:40	AM Peak, Dry	Middle	17
	(220 minutes)		Left	17
			Right	15
05/24/2000	15:00 – 19:00	PM Peak, Dry	Middle	27
	(240 minutes)		Left	29
			Right	21
05/25/2000	6:00 – 9:40	AM Peak, Dry	Middle	16
	(220 minutes)		Left	24
			Right	17
05/25/2000	15:00 – 19:00	PM Peak, Dry	Middle	23
	(240 minutes)		Left	26
			Right	17
05/31/2000	15:00 – 19:00	PM Peak, Dry	Middle	17
	(240 minutes)		Left	22

Table 13.3: Summary of Algorithm Testing Data Collected

DATES	DURATION OF VISUAL OBSERVATION	CHARACTERISTICS	LANE	NUMBER OF MANUAL ALARMS
			Right	13
02/28/2001	14:30 – 18:30 (240 minutes)	PM Peak, Snow	Middle	14
			Left	14
			Right	18
03/28/2001	17:40 – 19:15 (235 minutes)	PM Peak, Dry	Middle	16
			Left	23
			Right	18
04/16/2001	16:00 – 18:30 (150 minutes)	PM Peak, Dry	Middle	19
			Left	14
			Right	10
04/19/2001	16:00 – 18:15 (135 minutes)	PM Peak, Dry	Middle	16
			Left	26
			Right	12
04/23/2001	11:00 – 14:00 (180 minutes))	Non Peak Dry	Middle	14
			Left	15
			Right	19
04/23/2001	14:00 – 17:15 (195 minutes)	PM Peak, Dry	Middle	20
			Left	23
			Right	16
04/26/2001	14:25 – 18:50 (265 minutes)	PM Peak, Dry	Middle	35
			Left	31
			Right	17
05/01/2001	13:15 – 17:40 (265 minutes)	PM Peak, Dry	Middle	19
			Left	27
			Right	16
05/10/2001	15:00 – 19:15 (255 minutes)	PM Peak, Rain	Middle	19
			Left	18
			Right	19
05/23/2001	8:45 – 12:30 (225 minutes)	Non Peak, Dry	Middle	20
			Left	21

14 METHODOLOGY

14.1 Brief Description of Methodology

As mentioned earlier, the main objective of the current research was initially to explore the feasibility of identifying accident prone conditions and to develop a methodology that will detect these conditions whenever they occur. An accident prone condition is any undesirable traffic condition that has a high potential to lead to an accident. Successful detection of such conditions should go a long way towards reducing accidents on freeways by warning drivers and alerting TMC operators.

It was initially planned to collect and analyze microscopic traffic data and video footage pertaining to traffic conditions leading up to accidents in order to identify patterns of traffic conditions that caused those accidents. As indicated in the previous chapter, access to microscopic traffic measurements and video surveillance is possible only at those locations that have Machine Vision Sensors installed. However, due to the lack of sites on the Twin Cities freeways having such instrumentation, it was necessary to confine the data collection to the only existing installation of a Machine Vision Sensor available. This installation is located on Interstate-394 just east of the Penn Avenue Interchange in Minneapolis.

Unfortunately, though the afore-mentioned site is prone to the occurrence of a number of shock waves every peak period, it was not subjected to the occurrence of any accidents during the periods of traffic video observation and data collection. Hence, the methodology of the current research was modified to detect a pre-defined set of conditions that were deemed to lead to accidents based on earlier literature and discussions with several TMC operators and engineers. These conditions were visually detected by observing live video and algorithms were developed from actual data (volume, occupancy, speed, headways and other derivative measurements) that were extracted simultaneously. The reasoning is that if detecting such conditions is indeed feasible, then detection of actual accident prone conditions should also be feasible.

After an extensive review of literature, communication with operators working at Traffic Management Centers and visual observation of traffic, a list of undesirable traffic flow conditions was arrived at. As mentioned in the previous chapter, the visible accident prone or undesirable conditions considered are:

1. Sudden drops in speed on a lane.
2. Excessive weaving and lane changing.
3. Short headways at high speeds.
4. Excessive speed differentials between lanes.
5. Excessive fluctuation of speeds among vehicles in a lane.
6. Adverse weather, pavement and visibility conditions.

The main types of accidents that may occur due to the presence of any of the above conditions are:

1. Rear end collisions
2. Collisions that occur when vehicles are involved in lane changing or merging maneuvers (for example, side swipes)
3. Accidents in which only one vehicle is involved. In these accidents, the vehicle slides off the roadway without colliding with any other vehicle

The above three types of accidents alone constitute approximately 48% of all automobile collisions and 36% of all automobile related crashes on all types of roadways [W7]. However, traffic conditions that may cause rear end collisions have been given the most importance in the present research because they are the accidents that occur most often due to prevailing traffic conditions rather than human error and judgment and also because, given the present experimental set-up, these conditions are the easiest to visually observe. Rear end collisions occur most often in stop and go traffic or due to a sudden reduction of speeds or vehicles traveling in a platoon of closely following vehicles. The statistics presented in table 13.1 [W7] give an indication about the conditions in which rear end collisions occur:

Table 14.1: Circumstances under which rear end collisions most commonly occur

Pre-collision Circumstance	% Of Rear end collisions
Both vehicles traveling at the same speed with the lead vehicle decelerating	37
Following vehicle encounters a stopped vehicle	33
Leading vehicle traveling much slower when compared to following vehicle	14
Both vehicles decelerating with the leading vehicle decelerating at a higher rate	5

Thus, it can be seen that most rear end collisions occur when vehicles are traveling at almost the same speed, but the driver of the following vehicle fails to detect that the leading vehicle is decelerating.

Having defined an accident prone condition and having arrived at the list of undesirable flow conditions, video recordings and microscopic traffic measurements from the Machine Vision Sensor were initially collected for approximately twenty hours. Based on this data, the traffic parameters associated with undesirable traffic flow conditions were identified and detection algorithms were developed subsequently. Thirty additional hours of data and videotapes, which included rainy and snowy conditions, were collected later in order to further test and improve the detection algorithms. These videotapes were then analyzed and the times at which conditions corresponding to any of the undesirable flow conditions occurred were noted along with the conditions observed. In this manner, a log of undesirable flow condition occurrences was constructed so that it could be used for ground truth while developing the detection algorithms.

The next step in the methodology involved identifying the traffic flow measurements that are potentially contributing to accidents and then re-defining them so that they could be derived from raw sensor data such as vehicle speeds, lengths and on-off detection times. Thus, the detection algorithms could also use data from loop detectors. The measurements considered were raw data of the machine vision sensor and included time headway, space headway, speed variance, density, traffic pressure, quality of flow index, acceleration noise, momentum and kinetic energy.

However, some of these measurements were not in the final undesirable conditions detection algorithms. Only those measurements that exhibited a fair degree of correlation or some semblance of a pattern with the logs of visual observations were selected. The process of selecting the most relevant measurements is described in the next chapter.

Having selected the traffic measurements that could best detect undesirable conditions, individual algorithms (based only on each of the above mentioned traffic measurements) were developed so that the contribution of each measurement could be determined and threshold values established. Having determined the relevance of a measurement, optimum thresholds were selected by determining which threshold produced the best combination of detection rate and false alarm rate. Thus, the initial algorithms are “stand-alone” in nature and can be used independently of any other traffic measurement to detect undesirable traffic flow conditions. Algorithm development is explained in detail in the next chapter.

After developing the “stand-alone” or individual algorithms, an algorithm that included all the relevant traffic measurements was developed. This algorithm was heuristic in nature and assigned scores to each traffic measurements. Thresholds of the cumulative scores were used as criteria for raising alarms. Here again, the optimum score was that which produced the best combination of detection rate and false alarm rate.

A final approach was also explored in an attempt to integrate all the traffic measurements into one algorithm. Neural networks were employed to determine optimum coefficients for each of the traffic measurement. Thus, whenever values of the traffic measurements are multiplied by their corresponding coefficients and processed by the selected squashing functions, alarms are raised based on the value of the output obtained.

14.2 Traffic Measurements Considered

This section defines all the traffic measurements that were initially considered to contribute in detecting accident prone or undesirable traffic flow conditions. The measurements considered

include time headway, space headway, speed variance, density, traffic pressure, quality of flow index, acceleration noise and kinetic energy. Because some of these measurements require spatial monitoring of vehicles, adjustments were made so that they could be estimated from the raw point detection data. Kinetic energy was not explored separately because it uses the same formula as acceleration noise, albeit with a multiplying correction factor. What is important to note is that the basic detector measurements i.e. volume, speed, time headway derived from on-off detection times were used for detecting undesirable conditions because of the requirement to only rely on measurements from equipment already in the field.

14.2.1 Time Headway

Time headway is a microscopic traffic measurement that measures the temporal spacing between an individual pair of vehicles. Roess et al. [3] define time headway as “the time between successive vehicles as they pass a point along the lane, also measured between common reference points on the vehicles”. Time headways are one of the most important judgments that the driver of a vehicle should make correctly. According to a report submitted to the Federal Highway Administration (FHWA) on the “Development of Countermeasures for Driver Maneuver Errors” [W8], a driver who has selected an inappropriate following headway may not always find himself involved in a collision. However, unpredictable circumstances may cause the lead vehicle to slow down suddenly and this may result in a collision. Hence, the report states that, in such a case, when a vehicle follows another with “below-safe” time headway, the potential for a collision or accident is extremely high and always present. This was one of the main reasons for having included time headway to the list of traffic measurements that could be used in detecting accident prone or undesirable traffic flow conditions. Time headway between two vehicles has been calculated from,

$$T_{\text{current}} = T_{i+1} - T_i \quad (3.1)$$

Where,

T_{existing} is the time headway that exists between leading vehicle V_i and following vehicle V_{i+1}

T_i is the time at which the sensor detects the leading vehicle

T_{i+1} is the time at which the sensor detects the following vehicle

In determining the relevance of the time headways for detecting undesirable flow conditions, each current headway was compared with the minimum safe time headway. (T_{safe}). A safe headway is that amount of time that gives the driver of the following vehicle sufficient time to react to any sudden action on the part of the drivers ahead of him under prevailing pavement conditions (dry, wet and icy). Extensive review of literature was done to determine the values that should be used for safe time headway. Most literature mentions that every driver should follow the “thumb rule of 2”. This means that the minimum headway that should exist between every pair of vehicles is 2 seconds. But this rule is applicable only in dry conditions and for vehicles that are in fairly good operating condition. Drivers always have to consider the road conditions, brake condition, tire condition, level roads as opposed to going upgrade or downgrade, the weight of vehicle, and even wind resistance. In the present scenario, it has been assumed that all vehicles are of good operating conditions and that there is negligible wind resistance. However, a minimum headway of 3 seconds for rainy conditions and a minimum headway of 4 seconds for snowy conditions have been adopted. These values correspond to the average safe time headways that drivers seek to maintain [W7, 3].

2 seconds, dry conditions	}
$T_{safe} = 3$ seconds, wet conditions	
4 seconds, snowy/icy conditions	

14.2.2 Space Headway

Roess et al. [3] define space headway or spacing as the distance between successive vehicles in a traffic lane, measured from some common reference point on the vehicles. Space headway is thus a microscopic measurement that measures the spatial displacement between an individual pair of vehicles. Space headway was also selected as one of the traffic measurements that could be used to assist in the detection of undesirable or accident prone conditions for the same reason, as was time headway. It is imperative that drivers of vehicles maintain safe space headways and

adjust them accordingly to the prevailing traffic, weather and pavement conditions and operating condition of the vehicles. Whenever vehicles travel too close to each other with headways that are inappropriate taking into consideration their speeds and other conditions, the probability that a collision may occur is high. Space headway complements time headway as it describes the space between vehicles in a different manner. Space headway is estimated from speed and time headway.

The existing space between two vehicles was estimated from:

$$H_{\text{current}} = S_{i+1} * (T_{i+1} - T_i) \quad (3.2)$$

Where,

H_{current} is the space headway that exists between leading vehicle V_i and following vehicle V_{i+1}

T_i is the time at which the sensor detects the leading vehicle

T_{i+1} is the time at which the sensor detects the following vehicle

Just as in the case of time headways, it isn't sufficient if just the space headways between vehicles are measured. It is therefore reasonable to compare the spacing between successive vehicles with a safe headway that should exist between every pair of vehicles (H_{safe}). Drew [4] defines safe headway as "that headway between two vehicles in a lane that will allow the following vehicle to stop safely even if the vehicle in front makes an emergency stop". A rule of thumb that is generally suggested for use is that the following vehicle maintains one car length of distance between itself and the leading vehicle for every ten miles per hour of its speed. The choice of a safe space headway to be used for vehicle following so that no rear-end collision occurs when the leading vehicle applies its brakes during emergencies depends on many factors, including the braking capabilities of the vehicles involved, the perception-reaction time of the driver of the following vehicle, the friction coefficient etc. All these factors are taken into account while calculating the safe space headway.

Figure 14.1a shows the position of the vehicles just when the following vehicle is at the sensor. Figure 14.1b shows the positions of the vehicles when they ultimately come to a stop. From figures a and b, it can be seen that, the minimum safe headway that should exist between

vehicles V_i and V_{i+1} , 'Hsafe' should be equal to the difference between the distance traveled by the following vehicle and the distance traveled by the leading vehicle.

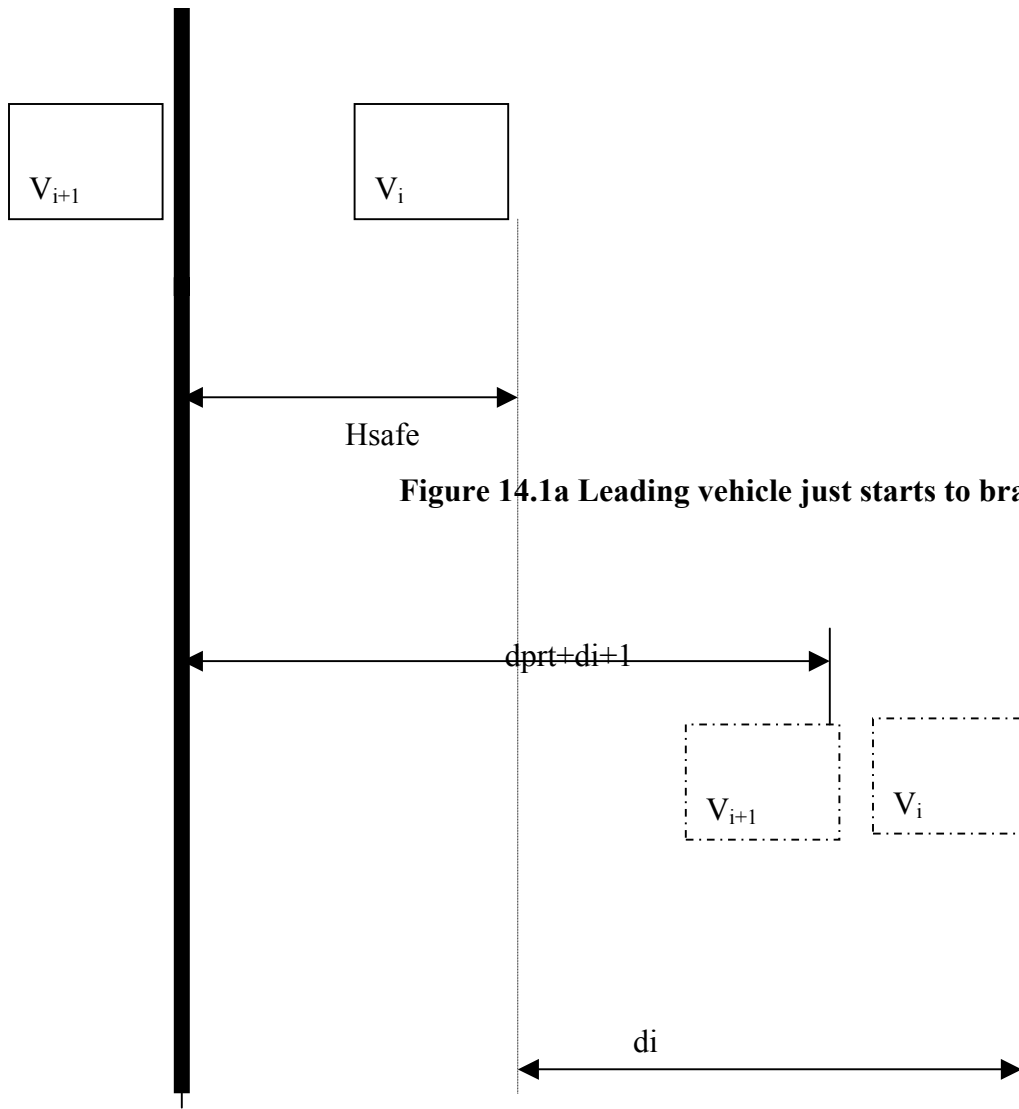


Figure 14.1b Distances needed by both vehicles to stop

Hence,

$$H_{safe} = 1.47 * S_i * t + d_{i+1} - d_i \quad (3.3)$$

where, d_{i+1} and d_i are respective stopping distances and d_{prt} is the distance traveled by V_{i+1} during the response time

H_{safe} is the safe space headway that should exist, in feet

t is the Response time of the following driver in seconds

S_i and S_{i+1} are the speeds of the leading and the following vehicle respectively in mph

Equation 3.3 can be rewritten as

$$H_{safe} = 1.47 * S_i * t + S_{i2} / (30(f+g)) - S_{i-12} / (30(f+g))$$

Where,

f is the co-efficient of friction

g is the roadway grade expressed as a decimal

The above equation can be rewritten as,

$$H_{safe} = 1.47 * S_i * t + (S_{i2} - S_{i-12}) / (30(f+g)) \quad (3.4)$$

Response time is the total time it takes for a vehicle driver to perceive, evaluate, decide and react to a situation on the roadway, such as a vehicle in front of him coming to a sudden stop. Response time is a sum of Perception-Decision Time and Reaction Time.

$$\text{Response time} = \text{Perception-Decision Time} + \text{Reaction Time}$$

The perception-decision time is the time it takes to view a hazard such as a suddenly stopping vehicle and figure out what to do about it. The reaction time is the time it takes to perform a particular function such as braking once a decision has been made. Just like other human characteristics, response time is a highly variable quantity. For the same situation on the road, the response time of drivers varies markedly as a function of age, time of day, weather conditions, chemical ingestion, and fatigue. For a given driver, response times vary depending on type of situation that he has to react to. [4]

According to a 1983 study, Highway Design and Operations Standards Affected by Driver Characteristics [W9, 5], the 85th percentile of the response time was 3.19 seconds. However, the 85th percentile is often chosen as the upper boundary for design analyses. The design standards of the American Association of State Highway and Transportation Officials (AASHTO) allow 1.5 seconds for perception time and 1.0 second for reaction time [W10]. It is worth noting that design standards are conservative and set to accommodate the large majority of drivers, not just the average driver. A study conducted by the Transportation Research Board in 1998 found that most people can perceive and react to an unexpected roadway condition in 2 seconds or less, concluding that AASHTO's total average perception-reaction time of 2.5 seconds "encompasses

most of the driving population and is an appropriate value for highway design”. Therefore, in the present research, a response time of 2.0 seconds has been adopted for algorithm development purposes, but this is a variable that can be calibrated for each site.

The co-efficient of friction, f , between the tires and pavement is not constant during deceleration. Actual measurements show that f is not the same at all speeds but varies inversely with speeds. According to an article in “The Source” [W10], the value most commonly used for co-efficient of friction of dry pavements was 0.72. The value of co-efficient of friction for wet pavements was 0.40 and for that of icy pavements was 0.24. These values have been adopted for the calculation of safe headway in the current research.

The response time, co-efficient of friction and current vehicle speeds obtained from the sensors in real time are used to calculate the safe space headway in the algorithm development described in the next chapter.

14.2.3 Quality of Flow

Early accident studies attempted to relate accidents with exogenous variables such as speed and density. These studies were conducted to analyze the association between the occurrence of the accident and the roadway level of usage. Greenshields [6, 4] attempted to correlate the frequency of accidents and the quality of flow by developing a quality traffic flow index based on “drivers’ frustration factors”. Greenshields contended that the amount of speed changes and the frequency of speed changes are undesirable factors that irritate or frustrate road users, increase the freeway operation costs and thus were inversely proportional to the quality of flow that existed on a segment of the roadway. As the first suggestion to the study of several variables within the same model, Greenshields concluded that there should exist “devices” to measure the ability of drivers to deal with unexpected situations on the roadway and that the features of a roadway that influence the driver’s ability consist of its geometric design, its surface condition and the appearance of the road and its immediate surroundings.

The quality of flow index as formulated by Greenshields [4] is

$$QOF = K * S * (f)^{0.5} \quad (3.5)$$

Σs

where,

QOF is the Quality of Flow Index

S is the average speed of a vehicle in miles per hour

f is the frequency of speed changes of the vehicle per mile that is incremented at different intervals within 1 mile

Σs is the absolute sum of speed changes per mile calculated by adding the absolute difference between the speeds of the vehicles at successive intervals

K is a constant in the order of 1000

However, the index developed by Greenshields requires tracking speeds of individual vehicles as they travel across a roadway. This can be measured by following a vehicle using an aerial speed-monitoring device or by using an instrumented vehicle that reports speed at periodic intervals of time. This, however, was not possible with the sensors available because even the machine vision sensors have a limited range of operation and do not track vehicles for distances long enough to be meaningful.

Hence, a methodology was devised for estimating the quality of flow (QOF) from point measurements of speeds and headways, as shown in figure 14.2. From these, we estimate the trajectories and speeds of vehicles as they cross the sensor, by guessing how many vehicles are in a platoon within approximately a mile ahead of the current vehicle and assuming that the current vehicle's speed will change in space as the speeds of the preceding vehicles in the platoon changed in time. Whenever a vehicle, $V_{current}$ passes over the sensor, the speeds of all the vehicles prior to this vehicle that lie within an estimated distance of approximately equal to a mile are accounted for by the proposed methodology while calculating the QOF. This space is calculated by projecting the positions of successive vehicles ahead of the current vehicle until we find a vehicle that lies about a mile ahead. The first vehicle in the platoon is the one which is 1 mile away from the present vehicle and it is identified by proceeding through an iterative process by looking at successively preceding vehicles and estimating their positions from the currently

detected vehicle assuming constant speeds and travel times equal to the difference between their times of arrival at the sensor and the current time. The first vehicle that is about 1 mile ahead of the sensor is the leading vehicle, V_{leader} . In this manner, the number of vehicles in the platoon is easily calculated as it includes the leading vehicle and all preceding vehicles up to the current vehicle.

Since vehicle, $V_{current}$ cannot be followed as it travels across 1 mile on the roadway, the methodology proposed to calculate QOF assumes that $V_{current}$ will emulate the behavior (speeds) of the vehicles in front of it. Thus, it is assumed that when $V_{current}$ reaches the position where vehicle $V_{current-1}$ is at present, it will have the same speed as that of $V_{current-1}$. In this manner, according to the proposed methodology, at the end of 1 mile, $V_{current}$ will have a speed equal to that of V_{leader} .

After the speed profile of vehicle $V_{current}$, as it travels a distance of 1 mile from the sensor, has been acquired in the above manner, the average speed, the amount of speed changes and the frequency of speed changes are calculated as follows. If the speed of the current vehicle changes as it moves from point 'a' to point 'b', then the amount of speed changes, Σs is incremented by the absolute difference of speeds of the current vehicle at points a and b and the frequency of speed changes, f is incremented by 1. In this manner, Σs and f are obtained as the current vehicle travels for a distance of 1 mile from the sensor along the roadway. These values of Σs , f and average speed, S are used in equation 3.5 to obtain the value of quality of flow.

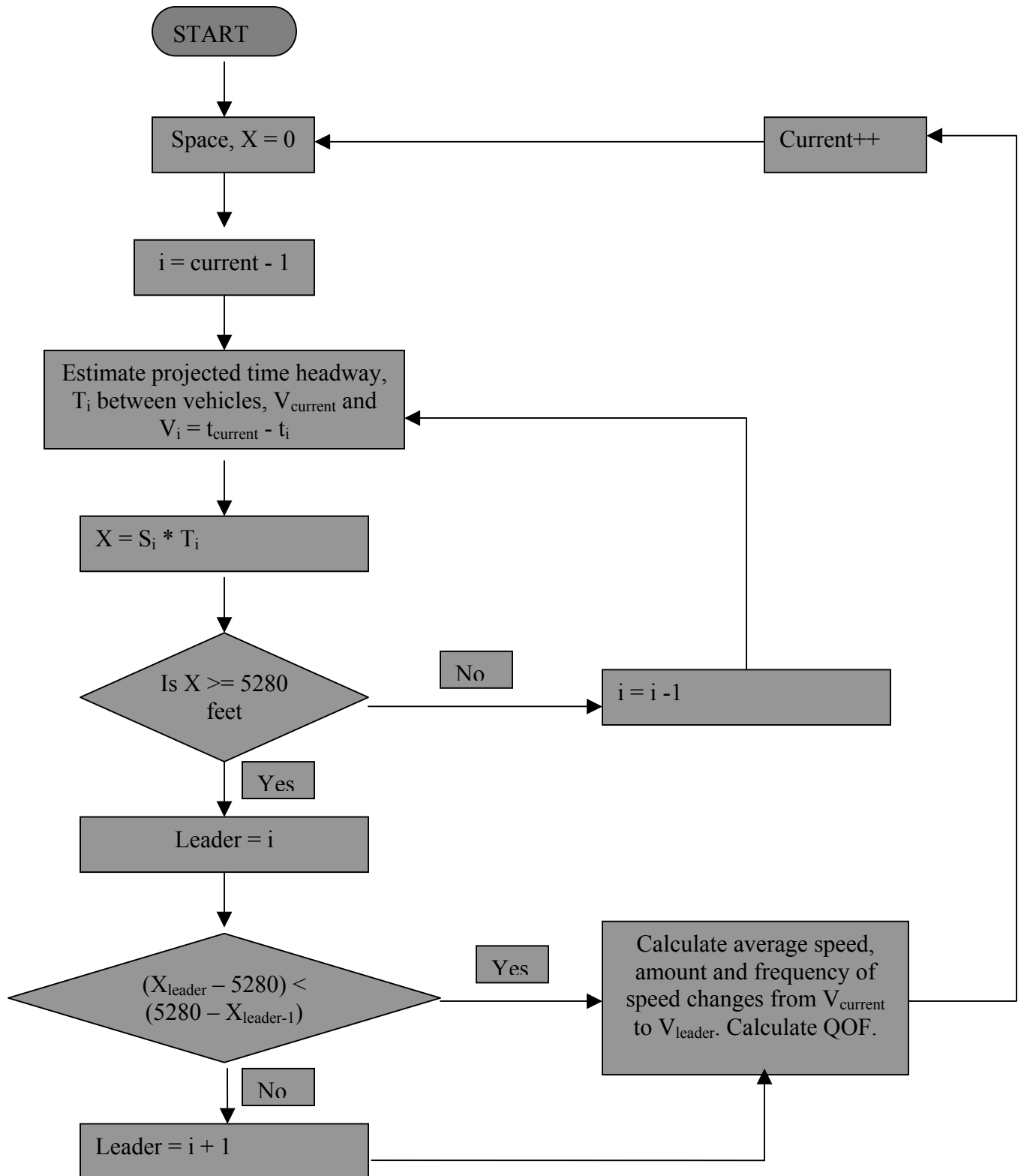


Figure 14.2: Flow chart showing the methodology used to calculate Point QOF

14.2.4 Speed Variance

The most common refrain that can be heard about speed variance is that “it kills”. While this statement may be a bit too extreme, it is well established that speed variance is a major causal factor in the likelihood of traffic crashes and that actions that reduce speed variance should also reduce crash rates. Speed variance measures the spread of distribution of speeds. It is a statistical measure of variation or dispersion or scatter of set of values of speeds from their mean value. This was the reason why speed variance was selected as one of the measurements that could be used to help in detecting undesirable traffic flow conditions. The greater an individual vehicle’s deviation is from the mean speed of its traffic stream, the higher its probability is of being involved in an accident. This is because, the greater the speed variance, or the distribution of speeds, the greater the number of interactions among vehicles. Thus, passing maneuvers and opportunities for collision increase. Hence, speed variance and not necessarily high speeds is associated with an increase in the frequency of crashes.

The following formula is used for calculating the speed variance at any instant on a roadway,

$$SV = \frac{1}{N} \sum (s_i - s_{mean})^2 \quad (3.6)$$

N

Where,

SV is the speed variance (point measurement)

N is the number of vehicles that have passed over a point in the last 60 seconds

s_i is the speed of vehicle V_i

s_{mean} is the average of the speeds of the ‘N’ vehicles

i is the vehicle id which ranges from 1 to N.

Just as in the estimation of the Point Quality of Flow Index, the methodology developed for calculating the speed variance also adopts a moving window. However, the unit of size of the moving window in this case is time (seconds) instead of length (feet). In this case, whenever the sensor detects a vehicle $V_{current}$, the speeds of all the vehicles that have been detected in a time

period, t of duration 60 seconds before the detection of V_{current} are considered for calculating the speed variance. This time period, t is calculated by adding the time headways between successively preceding vehicles until we find a vehicle that was detected by the sensor approximately sixty seconds before the current time. The first vehicle that was detected approximately a minute before the detection of the current vehicle is the leading vehicle, V_{leader} . In this manner, the number of which arrived within approximately one minute from the current vehicle is easily calculated.

After the vehicles that were detected by the sensor for a period of approximately 1 minute before the detection of the current vehicle have been obtained, the speed variance is calculated from 3.6.

14.2.5 Acceleration Noise

Acceleration noise is the standard deviation of accelerations of a vehicle and can be considered as the disturbance of the vehicle's speed from a mean speed and can be identified as an indicator of the smoothness of flow [4]. Higher disturbances of the speed of a vehicle from the uniform speed result in higher values of acceleration noise. Acceleration noise varies with the amount and frequency of acceleration and deceleration of a vehicle. The more violent and frequent these accelerating and decelerating maneuvers are, the higher is the acceleration noise. Violent decelerations affect the acceleration noise to a great deal and hence the value of acceleration noise also indicates the dangerousness of prevalent traffic conditions. Results of studies show that acceleration noise increased on highways with increasing levels of congestion [4]. Results of the same study also showed that higher accident rates were observed when acceleration noise values were of higher levels. Hence, this justifies the selection of acceleration noise as one of the measurements to help in the detection of undesirable traffic flow conditions. Acceleration noise is calculated from [4],

$$\sigma = \left(\frac{\sum 1t(a_i - a_{\text{mean}})^2}{t} \right)^{1/2} \quad (3.7)$$

Where,

σ is the acceleration noise

i refers to the current location of the vehicle

a_i is the acceleration of vehicle V as it travels across the interval ' i '

a_{mean} is the average acceleration calculated by taking the mean of the accelerations vehicle V as it travels across a roadway of length 1 mile

t is the travel time of the vehicle over a distance of 1 mile

As in the Quality of Flow, the determination of acceleration noise also requires the speeds and accelerations and decelerations of each vehicle at different intervals as they travel across a roadway. The accelerations of the vehicle can be measured directly by an accelerometer or approximate values can be obtained from the speed-time graph of the vehicle's trip. However, the available sensors only facilitate the measurement of speeds of vehicles at a point. Hence, just as in quality of flow, a methodology was developed to calculate the value of the acceleration noise by using approximated values of speeds of the vehicle, $V_{current}$. From the point measurements of speeds of vehicles, we estimate the trajectories and speeds of vehicles as they cross the sensor, by guessing how many vehicles are in a platoon within approximately a mile ahead of the current vehicle and assuming that the current vehicle's speed will change in space as the speeds of the preceding vehicles in the platoon changed in time. Whenever a vehicle, $V_{current}$ passes over the sensor, the speeds of all the vehicles prior to this vehicle that lie within an estimated distance of approximately equal to a mile are accounted for by the proposed methodology while calculating the acceleration noise. This space is calculated by projecting the positions of successive vehicles ahead of the current vehicle until we find a vehicle that lies about a mile ahead. The leading vehicle in the platoon is the one which is about 1 mile away from the present vehicle and it is identified by proceeding through an iterative process by looking at successively preceding vehicles and estimating their positions from the currently detected vehicle assuming constant speeds and travel times equal to the difference between their times of arrival at the sensor and the current time. The first vehicle that is about 1 mile ahead of the sensor is the leading vehicle, V_{leader} . In this manner, the number of vehicles in the platoon is easily calculated as they include the leading vehicle and all preceding vehicles up to the current vehicle.

Since vehicle, $V_{current}$ cannot be followed as it travels across 1 mile on the roadway, the methodology proposed to calculate acceleration noise assumes that $V_{current}$ will emulate the

behavior (speeds) of the vehicles in front of it. Thus, it is assumed that when $V_{current}$ reaches the position where vehicle $V_{current-1}$ is at present, it will have the same speed as that of $V_{current-1}$. In this manner, according to the proposed methodology, at the end of 1 mile, $V_{current}$ will have a speed equal to that of V_{leader} .

After the speed profile of vehicle $V_{current}$, as it travels a distance of 1 mile from the sensor, has been acquired in the above manner, the mean acceleration and the deviation of the vehicle's accelerations from this mean can be estimated from the above assumption. Specifically, the acceleration or deceleration of the current vehicle is its rate of change of speed or the difference of the speeds of the vehicle at points between its current position and its projected position divided by the travel time it would have taken for the vehicle to reach its projected position from its current position. In this manner, the accelerations or decelerations of the current vehicle as it travels for a distance of 1 mile from the sensor along the roadway are obtained and the acceleration profile is constructed. These values are then averaged to obtain the mean acceleration and then substituted in 3.7 to obtain the value of acceleration noise. Acceleration noise is updated every time a new vehicle is detected.

14.2.6 Traffic Pressure

Traffic pressure is one of the best indicators of the degree to which drivers on a lane must interact with each other given the existing traffic flow conditions. The word pressure is used because its principle is similar to that of gas pressure in the kinetic theory of gases [7]. Traffic pressure is defined as the product of density and speed variance at the point of measurement. If density on a lane is low, then drivers on that lane will have very little interaction with each other. However, if the variance is small, then, even at high densities, the interactions between drivers on a lane are minimal. Traffic pressure was chosen as one of the traffic measurements to help in detecting undesirable traffic flow conditions because it combines the properties of both density and speed variance in one equation. High values of both speed variance and density are undesirable and the probability of an accident occurring at these values is always high.

The formula used to calculate traffic pressure is

$$P = K * SV \quad (3.8)$$

Where,

P is the traffic pressure

K is the estimated density in the vicinity of the sensor obtained from 3.9.

SV is the speed variance.

Since speed variance is generally calculated for periods of a minute or 60 seconds, it was decided that traffic pressure too would be calculated over periods of 60 seconds. The methodology developed for calculating the speed variance also adopts a moving window. The unit of size of the moving window in this case is time (seconds) instead of length (feet). In this case, whenever the sensor detects a vehicle $V_{current}$, the speeds of all the vehicles that have been detected in a time period, t of duration 60 seconds before the detection of $V_{current}$ are considered for calculating the speed variance. This time period, t is calculated by adding the time headways between successively preceding vehicles until we find a vehicle that was detected by the sensor approximately sixty seconds before the current time. The first vehicle that was detected approximately a minute before the detection of the current vehicle is the leading vehicle, V_{leader} . In this manner, the number of which arrived within approximately one minute from the current vehicle is easily calculated.

After the vehicles that were detected by the sensor for a period of approximately 1 minute before the detection of the current vehicle have been obtained, the speed variance is calculated from 3.6.

In this methodology, the total space that exists between vehicles that traveled over the sensor over the past one minute, X is also determined by adding up all the space headways that exist between them. The density is calculated using the formula shown below

$$K = \frac{N * 5280}{X} \quad (3.9)$$

Where,

N is the number of vehicles that pass over the sensor in 1 minute

X is the total space between these N vehicles

The traffic pressure is then obtained by multiplying the speed variance by the density from 3.8.

14.3 Algorithm Effectiveness:

The major indicators of the performance of undesirable flow conditions detection algorithms are the detection rate and the false alarm rate.

The detection rate is defined as the ratio of accident prone or undesirable traffic flow conditions that are detected by the algorithm(s) to all the accident prone or undesirable traffic flow conditions that occur during a given time period. Detection rate is generally expressed as a percentage. Undesirable conditions that occur without the algorithm(s) detecting them are classified as missed detections. Thus,

$$\text{Detection Rate, DR} = \frac{\text{Number of Undesirable Conditions Detected} \times 100}{\text{Total Number of Undesirable Conditions}} \quad (3.10)$$

False Alarm rate is defined as the ratio of number of incorrect detections (alarms) to the total number of detections of undesirable traffic flow conditions generated by the algorithm(s). A detection is classified as a false alarm when the algorithm(s) detects an undesirable condition when there actually is none. The false alarm rate is generally expressed as a percentage but may also be expressed as the number of false alarms per time period.

$$\text{False Alarm Rate, FAR} = \frac{\text{Number of Incorrect Undesirable Conditions Detected} \times 100}{\text{Total Number of Detections (correct and incorrect) generated}} \quad (3.11)$$

These measures of effectiveness are dependent on each other. There exists a trade-off between them that is similar to statistical hypotheses tests (alpha and beta statistics). Normally, a higher detection rate is also accompanied with a higher false alarm rate because algorithms calibrated to detect a large proportion of undesirable flow conditions must be highly sensitive and thus, they also tend to set off a large number of false alarms. Similarly, less sensitive algorithms produce fewer false alarms and this is offset by detection of fewer undesirable flow conditions. Because

of this trade-off, a range of detection rate and false alarm rate combinations describes an algorithm's performance. There is no specific rule to determine the best combination of detection rate and false alarm rate and hence, there is not necessarily one optimum choice. Rather, the accident prone condition detection logic should balance the detection rate and false alarm rate based on the fact that the costs and consequences of a missed detection are much more severe from that of a false alarm. Satisfactory values of detection rate and false alarm rate for undesirable flow conditions are yet too early to establish and are therefore dependent on judgment.

15 UTC DETECTION ALGORITHM DEVELOPMENT

This chapter discusses all the algorithms that were developed to detect undesirable traffic flow conditions. The process by which, the most relevant traffic measurements were selected for the development of algorithms is discussed initially. The individual algorithms that were developed with each of the selected measurements are subsequently presented. These algorithms determine optimum threshold values for the selected measurements, which if violated result in alarms. Finally, the algorithms combining the selected traffic measurements together are developed. The initial data that was used for algorithm development is summarized in table 13.2. As mentioned earlier, the initial data set comprised of Machine Vision Sensor data and logs containing the times at which accident prone conditions were detected visually by simultaneous visual observations.

15.1 Selection of Traffic Measurements for Algorithm Development

As mentioned in the previous chapter, the traffic measurements that were initially considered to be used as indicators of accident prone conditions include time headway, space headway, speed variance, density, traffic pressure, quality of flow index, acceleration noise and kinetic energy. However, all the above measurements were not employed in the final algorithm development, instead; only the ones that passed an initial screening by exhibiting a fair degree of correlation or some semblance of a pattern with the occurrence of the manual alarms were selected.

Two periods of observation were analyzed for selecting the best traffic measurements. These periods comprised of a typical morning peak (05/25/2000, 6:00 – 9:40) and a typical afternoon peak period (05/25/2000, 15:00 – 19:00) observations. Having developed methodologies to calculate all the traffic measurements, time-series plots of all these measurements were developed for the two considered periods of observation. These plots were then superimposed with the times at which the manual alarms were raised. Patterns that existed between these traffic measurements and the manual logs were then identified.

The following criteria were used for the initial screening process in the time series plots of the traffic measurements in the initial screening for detecting their relevance in determining undesirable flow conditions:

Whether manual alarms occurred in the vicinity of the time at which a particularly high or low value of the traffic measurement.

Whether manual alarms occurred in the vicinity of the time at which the value of traffic measurements increased or decreased sharply with respect to the previous value.

All the traffic measurements except space and time headway had to be screened in the above manner because there is no pre-set formula or literature establishing their relevance for detecting undesirable traffic flow conditions. The minimum safe space headway that should exist between any pair of vehicles can be calculated by using equation 3.3 developed in the previous chapter. The existing values of space and time headways between a pair of vehicles can be compared with the values of the minimum headways that should exist between them. Hence, time series plots were drawn with both the existing and safe headway and the times at which manual alarms were raised were then marked.

Having drawn the time series plots and analyzed the data visually for patterns as mentioned above, space headway, traffic pressure and quality of flow index were selected as the most relevant traffic measurements that could help in detecting undesirable flow conditions.

Time headway was not selected as an individual measurement even though some manual alarms were observed at times when the value of the current time headway was less than that of the minimum safe time headway because the time headway thresholds do not take into account the vehicle speeds. Hence, if the algorithm were set up to raise an alarm for violation of the safe time headway, there would be many false alarms because manual alarms would not have been raised for vehicles traveling close to each other (with respect to time) at very low speeds. Time headway, however, is used as a raw measurement to derive current space headway. Hence, even though time headway is not considered individually, it is still an important measurement.

The equations developed to calculate current and minimum space headways on the other hand, incorporate time headways between vehicles, their speeds and the friction coefficient. The friction factor inclusion accounts for the weather and pavement conditions. Many manual alarms were noticed at instances when groups of vehicles traveled really close to each other with average to high speeds leading to the selection of space headway. One such example is shown in figure 15.1, where it can be seen that a manual alarm was raised close to the time at which the minimum safe headways were successively violated (5 consecutive points below the minimum safe headway line before the large raise).

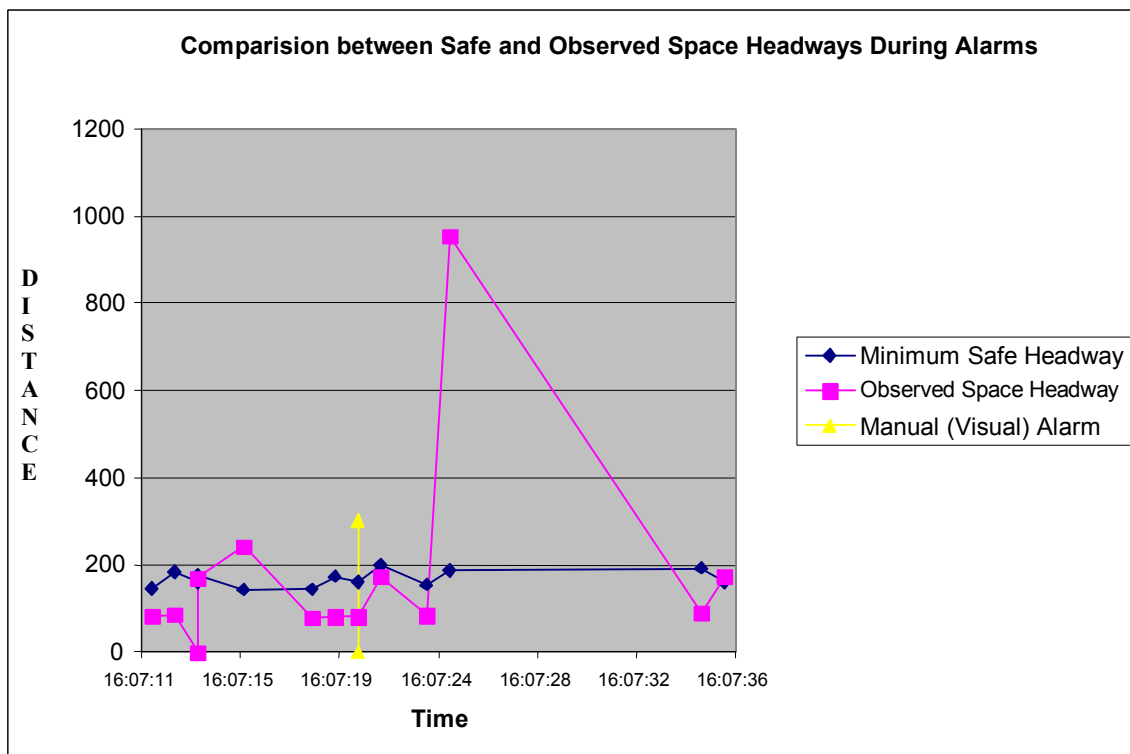


Figure 15.1: Example of pattern observed between space headways and occurrences of manual alarms

Though some manual alarms were observed at high values of both speed variance (measured over a moving period of 60 seconds) and density, they were not selected individually as major contributing measurements. This is because, if the algorithm were to be set up such that alarms were to be raised whenever high speed variances are observed, then many false alarms would be raised by the algorithm at times when the vehicles are very far apart but traveling at considerably differing speeds. Similar considerations were made for density as well i.e. when vehicles were

traveling really close to each other, albeit with almost identical speeds, it was not judged as an undesirable traffic flow condition. Hence, the alarms raised by a density only algorithm at such instances would be false.

Figure 15.2 shows the occurrence of a manual alarm in the vicinity of the time when high values of traffic pressure were reported.

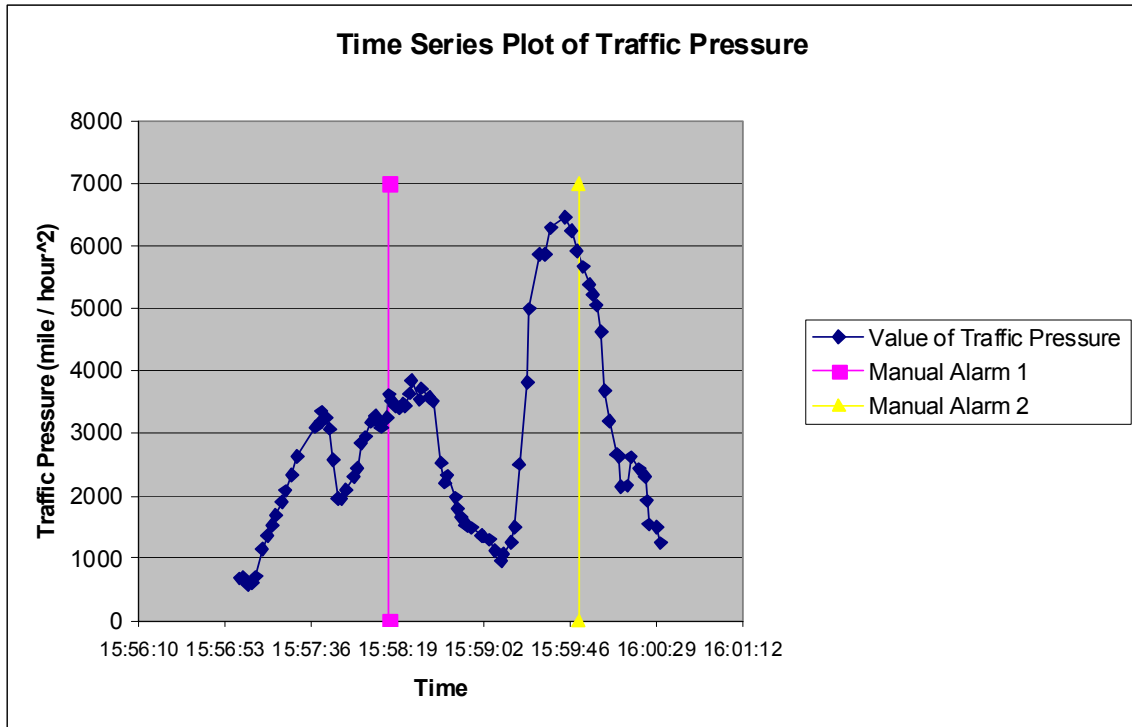


Figure 15.2: Example of pattern observed between high values of traffic pressure and occurrence of manual alarms

Traffic pressure combines the qualities of both speed variance and density since it is the product of speed variance by the density. Many manual alarms were noticed at instances when high values of traffic pressure were observed.

Some false alarms were be raised whenever only the value of either speed variance or density were extremely high resulting in high pressure values. However, the number of these false alarms is much lower than the number that would have been obtained if the algorithms were developed individually for speed variance and density.

From the plots of time series curves of the quality of flow index, it was observed that this measurement was well correlated with the manual alarms at low speeds. This is because, the formula to calculate quality of flow index employs both average speed and absolute sum of speed changes. Thus, the lower the average speed and the higher value of absolute sum of speed changes, the more unsafe the traffic condition. This situation corresponds to a lower quality of flow index values. Hence, many manual alarms corresponding to the times at which the speeds of vehicles dropped sharply within a short period were noticed at instances when the quality of flow index value decreased sharply with respect to the previous quality of flow index value. One such example is shown in figure 15.3, where a manual alarm can be observed close to the time at which the quality of flow index decreased by a large margin when compared to its previous value.

The time series plots constructed for acceleration noise and kinetic energy showed that it would not be worthwhile in pursuing the effectiveness of these variables in detecting accident prone conditions. This was because; only a few manual alarms were noticed near the vicinity of high values of acceleration noise and kinetic energy. Since acceleration noise and kinetic energy were extrapolated from point measurements, they may not have been accurate estimations. This could be the reason why no patterns existed between acceleration noise and kinetic energy and the occurrence of manual alarms.

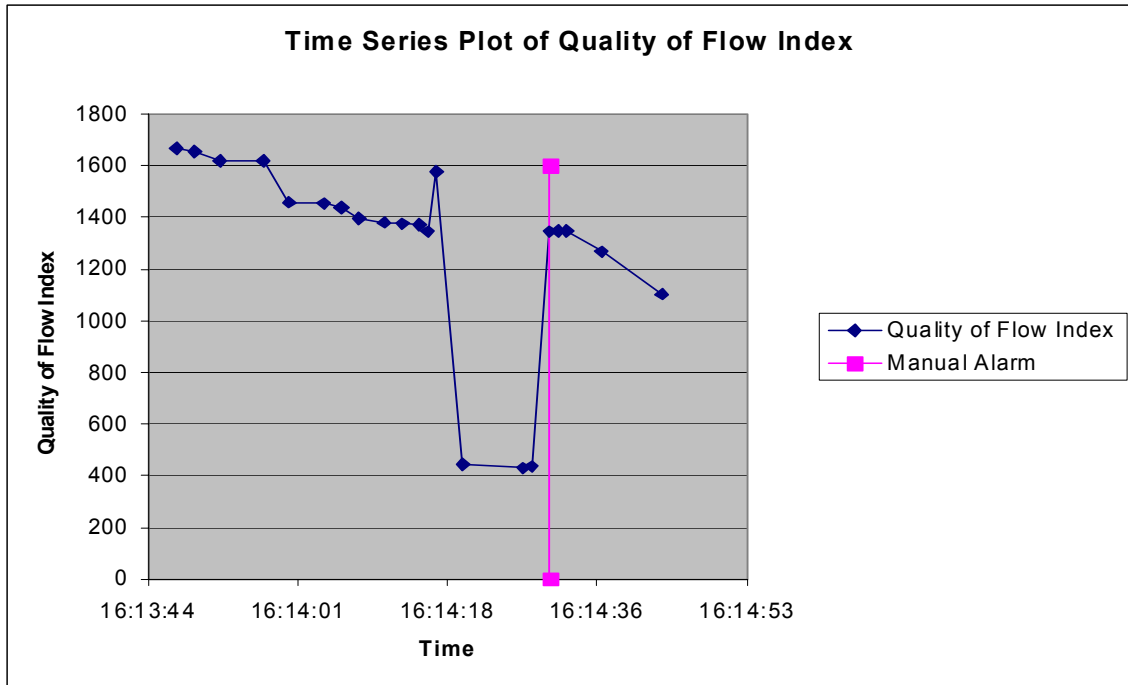


Figure 15.3: Example of pattern existing between sharp decrease in values of quality of flow index and manual alarms

15.2 Development of Algorithms

After the completion of the preliminary screening and identification of the most relevant measurements, contributing to the detection of undesirable flow conditions, we proceeded with the development of algorithms involving the selected measurements and combinations of them as described next.

15.2.1 Space Headway Algorithm

As mentioned earlier, many manual alarms were found to have been raised in the vicinity of the times at which vehicles traveled very close to each other at fairly high speeds. When this occurs, the actual space headway is lower than the minimum safe headway, it follows that when the leading vehicle stops abruptly then there will not be enough space for the following vehicle to come to a stop without colliding with the leading vehicle. Of all the manual alarms raised, around 60% comprised of those alarms that were raised when groups of vehicles traveling too

fast and too close to each other were observed. Hence, the Headway Algorithm was designed to detect more than one successive violation of the minimum safe space headway, as depicted in the flow chart in figure 15.4. A set of vehicles traveling such that each pair of vehicles in the group violate the safe headway is referred to as a “Group” and the number of vehicles in such a Group is referred to by the term “Group-Size”.

The main objective of the Headway Algorithm was to find the optimal group-size that would produce an acceptable combination of high detection rate and low false alarm rate. The process by which alarms are raised by the Headway Algorithm consists of the following steps:

Step 1: The current headway between a pair of vehicles is first calculated from,

$$H_{current} = S_{i+1} * (T_{i+1} - T_i) \quad (4.1)$$

Where,

$H_{current}$ is the space headway that exists between leading vehicle V_i and following vehicle V_{i+1}

T_i is the time at which the sensor detects the leading vehicle

T_{i+1} is the time at which the sensor detects the following vehicle

Step 2: The minimum safe headway that should exist between V_i and V_{i+1} for the following vehicle to be able to come to a stop before colliding with the leading vehicle in a sudden braking situation is then calculated from,

$$H_{safe} = 1.47 * S_i * t + (S_i^2 - S_{i+1}^2) / (30(f+g)) \quad (4.2)$$

Where,

H_{safe} (in feet) is the minimum headway that should exist between leading vehicle V_i and following vehicle V_{i+1}

t is the following driver’s response time

S_i and S_{i+1} are the speeds of the leading vehicle and the following vehicle respectively

F is the co-efficient of friction

g is the grade

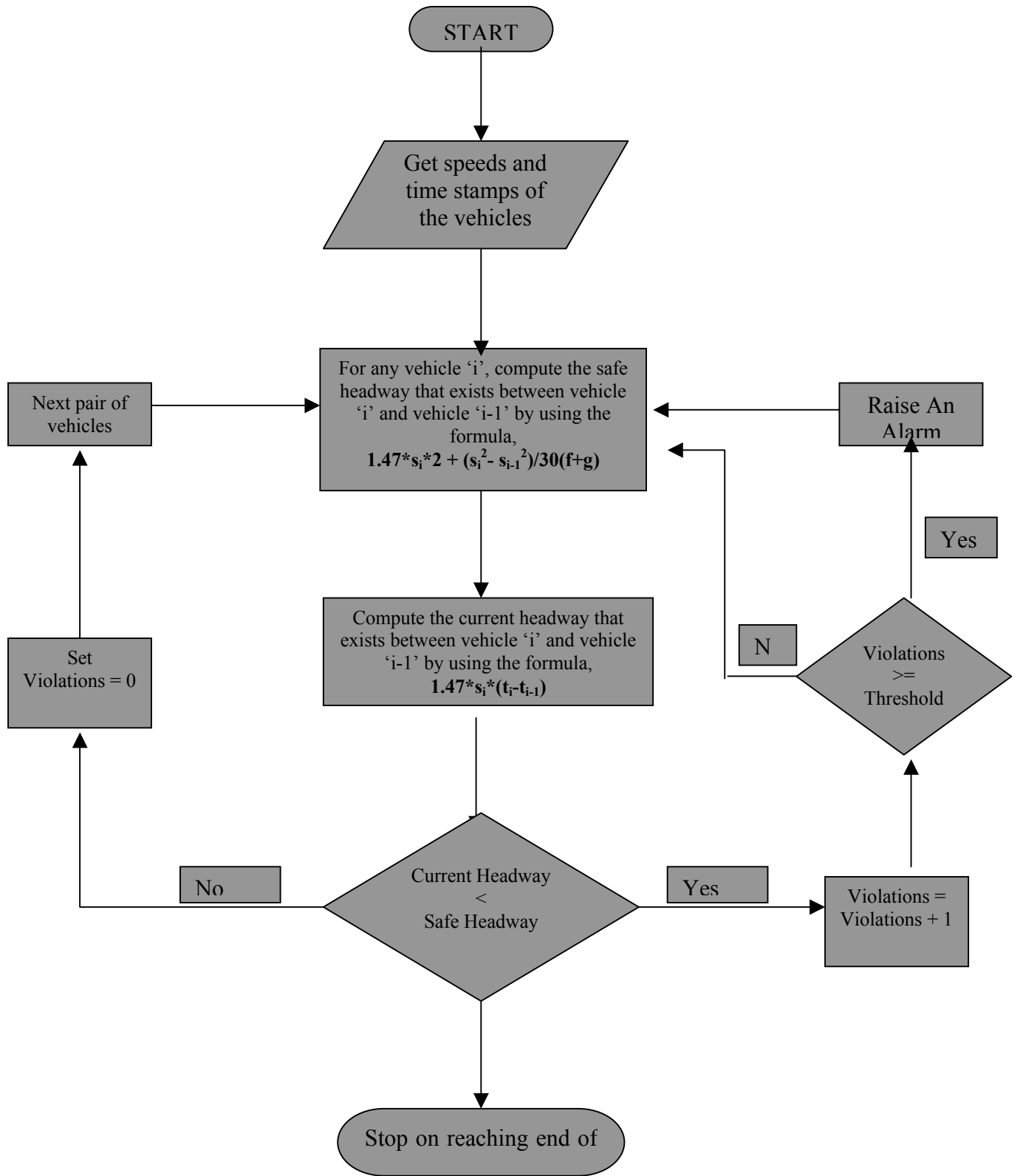


Figure 15.4: Flow Chart Describing the Headway Algorithm

In the present research, the value of grade has been considered to be 0.0 because the vertical alignment of the freeway at the test site is fairly horizontal. The value of the co-efficient of friction was taken as 0.72 if the pavement was dry, 0.40 if the pavement was wet and 0.24 if the pavement had traces of snow or ice (refer [section 3.2.2](#)).

Step 3: The current headway is then compared with the minimum headway. If the current headway is higher than the minimum headway, then there is no cause for alarm and the algorithm proceeds to the next pair of vehicles. However, if the minimum headway is greater than the current headway, then this is cause to start the alarm process if such additional and successive violations of the minimum safe headway are observed.

Step 4: In this step, the Headway Algorithm determines the number of such successive violations of the minimum safe headway that occur (Group-Size - 1).

Step 5: When the Group-Size exceeds a preset threshold, then the Headway Algorithm raises an alarm.

The Headway Algorithm was implemented on the data (separately for each lane) for a number of thresholds of the Group-Size i.e. from 2 to 15. Alarms were raised when each of these thresholds were violated (the Group-Size was greater than or equal to the threshold). The resulting Algorithm alarms were then compared with the manual alarms to determine the number of correct, missed and false alarms. The detection rate and false alarm rate were then calculated and the performance curves (Detection Rate v/s False Alarm Rate Plots) were drawn.

Analysis of the detection rate and false alarm rate of each group-size threshold suggested that a group-size of 5 vehicles (4 successive violations of the safe headway) was optimum for all the three lanes.

This group-size produced an average detection rate of only 41.5% and an average false alarm rate of 21% (values of detection rate and false alarm averaged across all the lanes for all the data). The value of the detection rate was considered low. It was therefore decided that some sort of improvement should be made to the algorithm.

It was noticed that the average detection rate corresponding to a group-size of 4 vehicles (64.5%) was much higher than the average detection rate corresponding to a group-size of vehicles 5 (41.5%). In spite of this, group-size of 5 vehicles was selected as the best group-size, because it produced the best combination of detection rate and false alarm rate. Group-size of 4 vehicles was not selected even though it had a higher detection rate because it produced a fairly high false alarm rate (67.5%).

Further analyses of algorithm alarms and manual alarms revealed that the Algorithm missed some manual alarms because there was one vehicle in between two successive violation groups that just managed to meet the requirement of the safe headway. Because of this, many double alarms were observed, especially when the threshold of the group-size was less than or equal to five vehicles.

Therefore, to reduce the false alarm rate produced by group-size thresholds of lower values (≤ 4) and to increase the detection rate produced by group-size thresholds of higher values (> 4), the algorithm was modified by adding what was called a “Unsafe Group Extension Check (UGEC)” to identify and disregard or overlook the presence of one safe headway within a string of unsafe ones. In this process, the current headway between that one “safe” pair of vehicles is reduced by 5%.

The following steps are carried out:

- I. If a “safe” current headway is detected immediately after a string of at least 3 successive violations of the minimum safe space headway and immediately before another violation of the safe headway, then, this “safe” current headway is reduced by 5%. This is the UGEC. The reason for implementing the UGEC after at least 3 successive violations is mentioned later.
- II. If this modified current headway violates the minimum safe space headway, then it is counted as a standard violation, if not the number of violations is reset to 0 and the algorithm proceeds to the next pair of vehicles.

The algorithm was implemented with the UGEC being performed after 1 violation and 2 successive violations, but very high false alarm rates were observed. Hence, it was decided that the UGEC would only be performed after more than 2 successive violations have been observed.

An example of the functioning of the FAC is shown in figure 15.5.

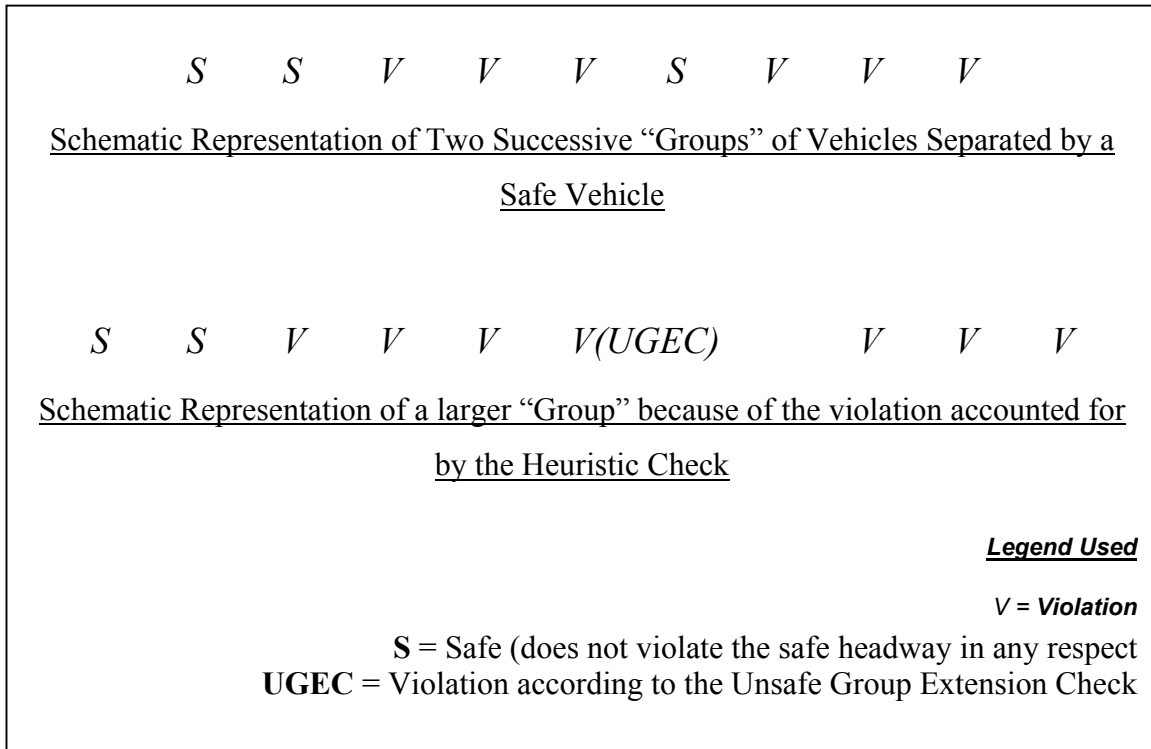


Figure 15.5: Schematic Representation of “Unsafe Group Extension Check”

The top part of the figure shows the schematic representation of the violations determined by the original headway algorithm. Thus, the algorithm would have raised two alarms corresponding to a group-size of 4 vehicles within the space of a few seconds. However, by applying the UGEC, the one safe vehicle in between the two violating groups violated the heuristic condition (95% of current headway < minimum safe headway), so the algorithm incremented the group-size.

Table 15.1 shows the average detection rate and false alarm rate per lane for all time periods that were obtained with the Headway Algorithm. Figure 15.6, derived from table 15.1, depict the detection rate v/s false alarm rate curves (performance curves) for each lane. As can be seen, group-size of 5 vehicles produces the best combination of detection rate and false alarm rate for the right and the middle lane. For the left lane, a group-size of 6 vehicles produces the best combination of detection rate and false alarm rate.

Table 15.1: Headway Algorithm Results

Headway Algorithm Results					
Right Lane Results					
Manual Alarms = 4.55 / hour					
Group-Size Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
11	0.05	0.05	0.00	1.33	0.00
10	0.16	0.16	0.00	3.84	0.00
9	0.21	0.21	0.00	5.02	0.00
8	0.41	0.41	0.00	9.75	0.00
7	0.72	0.67	0.05	15.34	4.00
6	1.81	1.50	0.31	34.06	13.44
5	3.47	2.38	1.09	52.77	28.08
4	7.55	2.90	4.66	64.49	60.14
3	16.81	3.93	12.83	87.04	75.40
2	37.24	4.50	32.74	98.89	87.83
Middle Lane Results					
Manual Alarms = 5.17 / hour					
Group-Size Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
7	0.26	0.26	0.00	4.78	0.00
6	1.19	1.03	0.16	20.15	7.30
5	2.84	2.33	0.52	44.53	14.85
4	7.71	2.95	4.76	56.05	58.76
3	18.21	4.14	14.07	79.33	76.27
2	31.34	5.02	26.33	97.27	83.55
Left Lane Results					
Manual Alarms = 5.69 / hour					
Group-Size Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
8	0.51	0.51	0.00	8.96	0.00
7	1.60	1.46	0.14	25.74	8.57
6	4.43	3.18	1.25	55.93	28.16
5	7.05	3.75	3.30	65.93	46.77
4	11.75	4.52	7.24	79.39	61.56
3	20.89	5.19	15.70	91.14	75.17
2	34.73	5.59	29.14	98.28	83.90

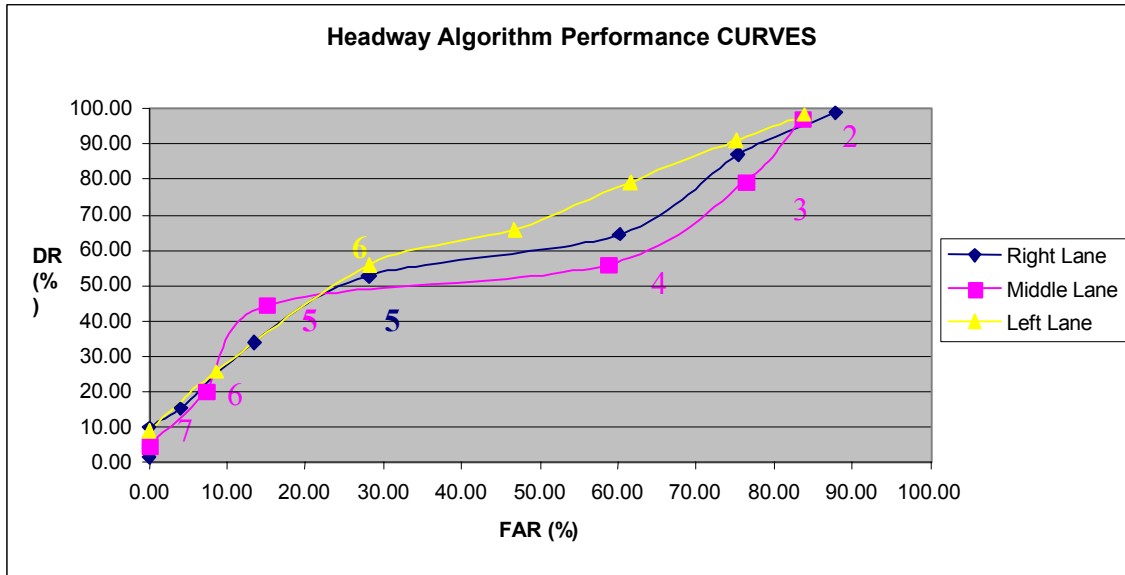


Figure 15.6: Performance Curves Obtained For Headway Algorithm

15.2.2 Traffic Pressure Algorithm

As described earlier in the traffic measurement selection section, some manual alarms were found to have been raised in the vicinity of the times at which values of traffic pressure were particularly high. High values of traffic pressure correspond to those traffic flow conditions at which vehicles travel very close to each other with highly fluctuating speeds from vehicle to vehicle. The value of traffic pressure gives a good indication of the degree to which drivers on a lane interact with each other. Interactions between drivers on a lane are minimal either at low densities or at high densities when the speed variance is low. Traffic conditions corresponding to high values of traffic pressure are characterized by excessive accelerations and decelerations. Since, traffic pressure is the product of speed variance and density, high values of both density and speed variance are undesirable and the likelihood that an accident will occur when these both values are high is large. The traffic pressure algorithm was designed to raise alarms when high values of traffic pressure are detected as shown in figure 15.7. The process by which the traffic pressure algorithm raises alarms is described below:

Step 1: Whenever the presence of a vehicle is reported and its speed and length obtained, the traffic pressure is calculated by using the methodology described in the flowchart in figure 15.5.

Step 2: The methodology developed for calculating the traffic pressure adopts a moving window. The unit of size of the moving window in this case is seconds. This moving window is similar to a stack (a stack is an abstract data type used in computer science which adopts first in, last out basis). In this case, the vehicle last detected is at the bottom of the stack. Thus, whenever the Sensor detects a vehicle V_i , the speeds of all the vehicles that have been detected in a time period, t of duration 60 seconds before the detection of V_i are considered for calculating the speed variance and density. The duration of this time period, t is calculated by adding up the existing time headways between all the vehicles until 't' is closest (greater or lesser) to 1 minute or 60 seconds. All the speeds and lengths of the vehicles within this moving window are used to calculate speed variance and density.

Step 3: The speed variance is calculated from,

$$SV = \frac{\sum 1n(s_i - s_{mean})^2}{N} \quad (4.3)$$

Where,

SV is the speed variance

s_i is the speed of vehicle V_i

Step 4: The density is then calculated from,

$$K = \frac{N * 5280}{X} \quad (4.4)$$

Where,

K is the density

N is the number of vehicles that pass over the sensor in 1 minute

T is the total enclosed space between these N vehicles

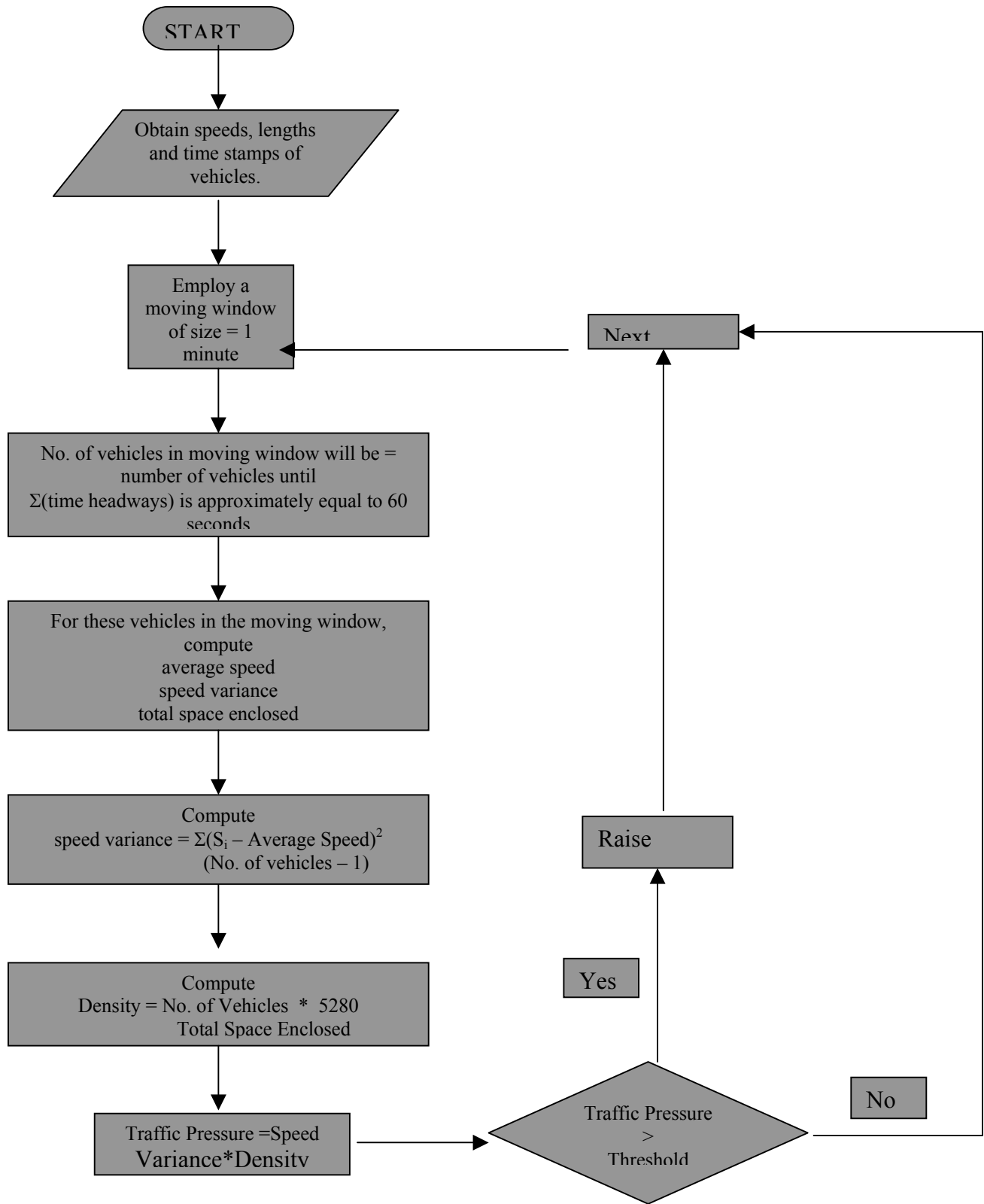


Figure 15.7: Flow Chart Describing the Traffic Pressure Algorithm

Step 5: The traffic pressure is calculated by multiplying the speed variance by the density from the equation shown below,

$$P = K * SV \quad (4.5)$$

Where,

P is the traffic pressure in veh-mile/hour²

Step 6: If the value of the traffic pressure is greater than a preset threshold, then the algorithm raises an alarm. In this manner, the Traffic Pressure Algorithm was implemented on the data (separately for each lane) for a number of values of the threshold of Traffic Pressure from 500 to 6500 veh-mile/hour² in increments of 500. Alarms were raised when each of these thresholds were violated (the traffic pressure value obtained was greater than or equal to the threshold). These Algorithm alarms were then compared with the manual alarms to determine the number of correct, missed and false alarms. The detection rate and false alarm rate were then calculated and the performance curves (Detection Rate v/s False Alarm Rate Plots) were drawn.

Table 15.2 shows the average detection rate and false alarm rate per lane for all time periods that were obtained with the Headway Algorithm. Figure 15.8, derived from table 15.2 depicts the detection rate v/s false alarm rate curves (performance curves) for each lane. As can be seen, a threshold of 3500 produces the best combination of detection rate and false alarm rate for the right lane. For the middle lane, a threshold of 2000 produces the best combination of detection rate and false alarm rate and for the left lane, a threshold of 1500 was found to be optimum.

Table 15.2: Traffic Pressure Algorithm Results

Traffic Pressure Algorithm Results							
Right Lane Results		Manual Alarms = 4.55 / hour					
Traffic Pressure Value Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Rate	Alarm	
6000	0.58	0.58	0.00	12.68	0.00		
5500	0.75	0.72	0.03	15.92	4.00		
5000	1.02	0.98	0.04	21.44	4.00		
4500	1.48	1.30	0.18	28.59	11.90		
4000	2.22	1.70	0.52	37.36	23.49		
3500	2.93	2.02	0.91	44.44	31.05		
3000	4.35	2.28	2.07	50.19	47.49		
2500	5.93	2.66	3.26	58.52	55.07		
2000	8.21	2.98	5.23	65.60	63.65		
1500	11.21	3.30	7.91	72.53	70.57		
1000	16.19	3.80	12.39	83.42	76.56		
500	23.23	4.29	18.94	94.32	81.53		
Middle Lane Results		Manual Alarms = 5.17 / hour					
Traffic Pressure Value Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Rate	Alarm	
6000	0.17	0.17	0.00	3.37	0.00		
5500	0.32	0.32	0.00	6.16	0.00		
5000	0.36	0.36	0.00	7.03	0.00		
4500	0.51	0.51	0.00	9.81	0.00		
4000	0.95	0.89	0.06	17.25	6.19		
3500	1.37	1.24	0.13	24.07	9.45		
3000	1.68	1.45	0.23	28.03	13.71		
2500	2.40	1.79	0.61	34.55	25.50		
2000	3.25	2.12	1.13	40.99	34.85		
1500	4.65	2.70	1.95	52.32	41.87		
1000	9.34	3.26	6.08	63.03	65.11		
500	18.28	4.33	13.95	83.67	76.33		
Left Lane Results		Manual Alarms = 5.69 / hour					
Traffic Pressure Value Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Rate	Alarm	
6000	0.14	0.14	0.00	2.50	0.00		
5500	0.14	0.14	0.00	2.50	0.00		
5000	0.14	0.14	0.00	2.50	0.00		
4500	0.19	0.19	0.00	3.27	0.00		
4000	0.30	0.30	0.00	5.29	0.00		
3500	0.52	0.50	0.02	8.77	4.00		
3000	0.81	0.75	0.06	13.20	7.33		
2500	1.31	1.12	0.19	19.64	14.72		
2000	1.97	1.63	0.34	28.62	17.39		
1500	2.95	2.14	0.82	37.53	27.69		
1000	6.01	2.73	3.28	48.02	54.54		
500	15.10	4.58	10.53	80.45	69.69		

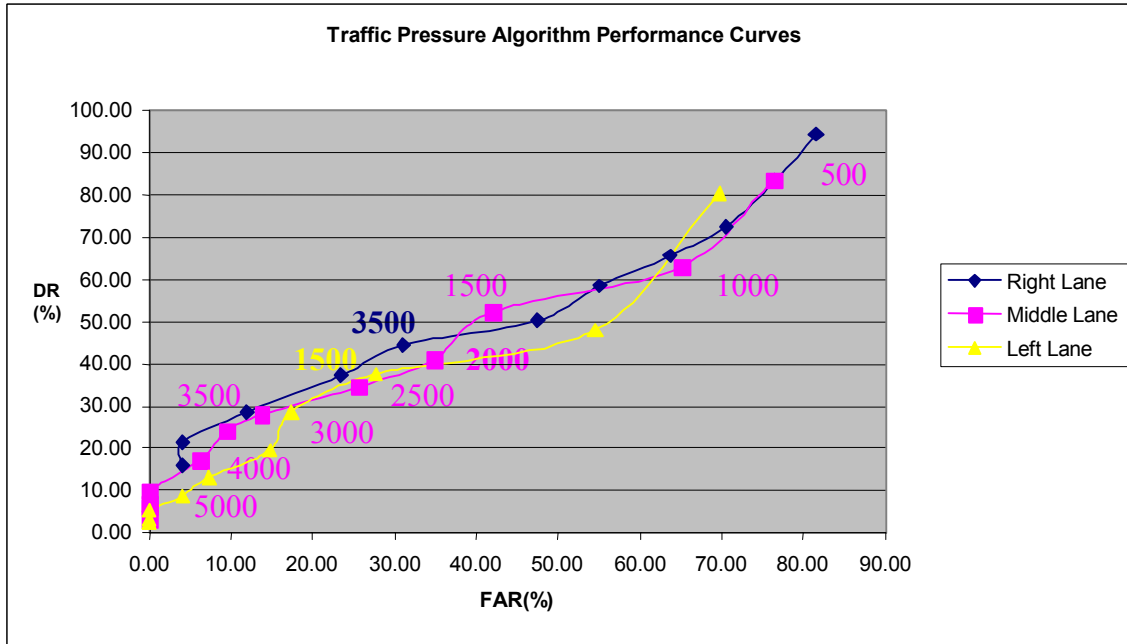


Figure 15.8: Performance Curves Obtained For Traffic Pressure Algorithm

15.2.3 Quality of Flow Index Algorithm

Low values of quality of flow index (QOF) are indicators of extremely low operating speeds on the roadway. A sudden decrease in the value of quality of flow index is an indicator of a sudden decrease in speeds of the vehicles on the roadway. An increase in quality of flow signifies improved traffic conditions. The quality of flow is directly proportional to average speed and the frequency of speed changes and inversely proportional to the absolute sum of speed changes. Hence, whenever low average speeds and high values of absolute sums of speed changes are observed, the resulting value of quality of flow is also low. These values also correspond to a more dangerous or undesirable traffic flow condition since low values of speed and high values of absolute sum of speed changes are undesirable and the likelihood that an accident will occur when speeds suddenly decrease is extremely high. Many manual alarms that were raised when the speeds of vehicles dropped sharply were observed in the vicinity of the times at which the value of the quality of flow decreased sharply with respect to the previous value of quality of flow. Hence, the quality of flow index algorithm was designed to raise alarms whenever sharp decreases in its value are observed as shown in figure 15.9.

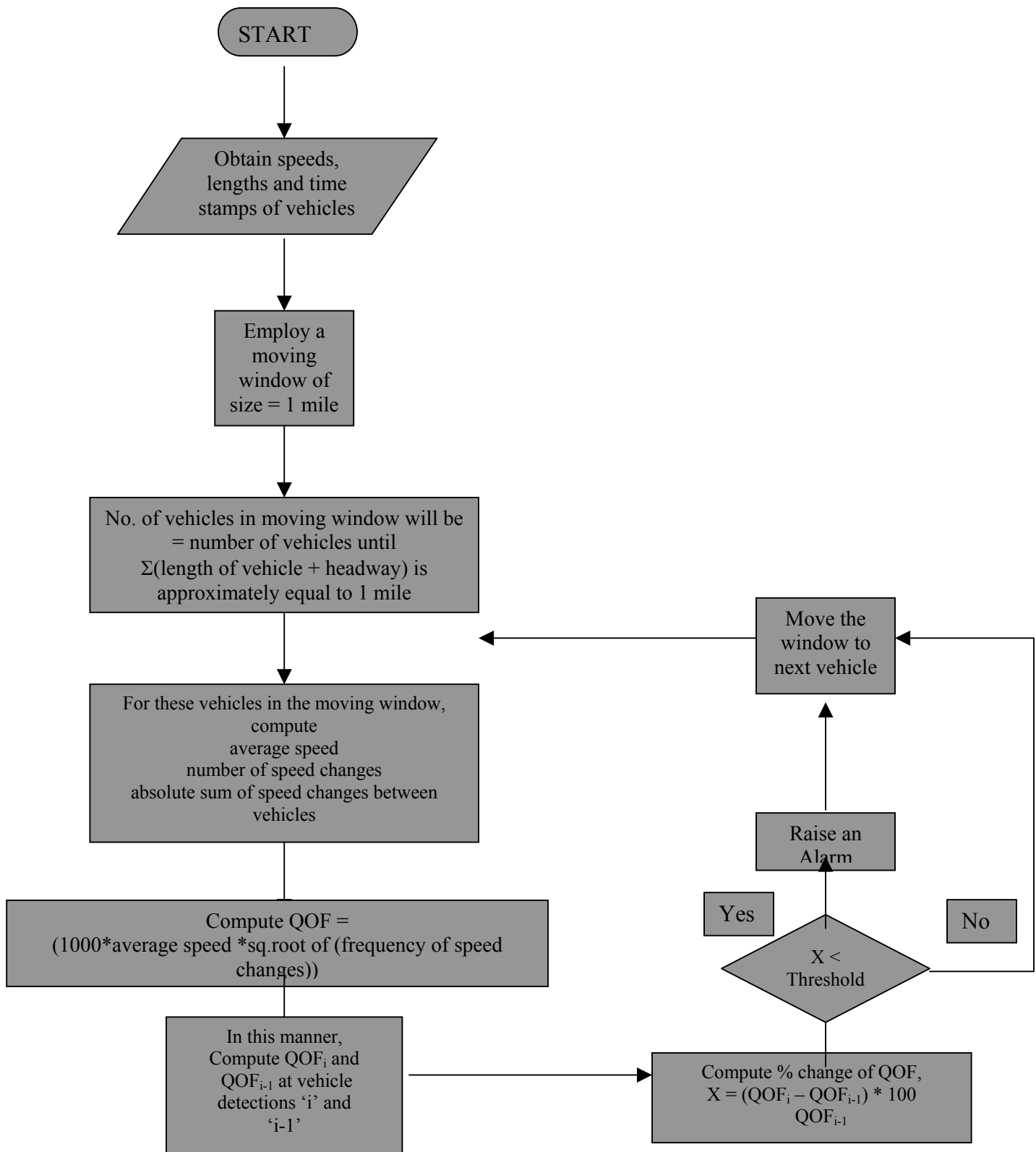


Figure 15.9: Flow Chart Describing the Quality of Flow Index Algorithm

Step 1: Whenever the presence of a vehicle is reported and its speed and length obtained, the traffic pressure is calculated by using the methodology described in the flowchart in figure 15.2.

Step 2: The methodology developed for calculating the quality of flow adopts a moving window. The unit of size of the moving window in this case is miles.

Thus, whenever the Machine Vision Sensor detects a vehicle V_i , the speeds of all the vehicles that have been detected within the moving window of size 1 mile before the detection of V_i are considered for calculating the quality of flow index. The length of the total enclosed space, X is calculated by adding up the lengths of the vehicles and the space headways existing between them, until ' X ' is closest (greater or lesser) to 1 mile or 5280 feet. All the speeds of the vehicles within this moving window are used to calculate average speed, frequency of speed changes and the absolute sum of speed changes.

Step 3: The quality of flow index is then calculated from,

$$QOF = \frac{K * S * (f)^{0.5}}{\Sigma s} \quad (4.6)$$

where,

QOF is the Quality of Flow Index

S is the average speed of a vehicle in miles per hour

f is the frequency of speed changes of the vehicle per mile that is incremented at different intervals within 1 mile

Σs is the absolute sum of speed changes per mile calculated by adding the absolute difference between the speeds of the vehicles at successive intervals

K is a constant in the order of 1000

Step 4: The % change of quality of flow, X is then calculated from,

$$X = \frac{(QOF_i - QOF_{i-1}) * 100}{QOF_{i-1}} \quad (4.7)$$

Where,

QOF_i is the quality of flow that is estimated when vehicle V_i is reported by the sensor

QOF_{i-1} is the quality of flow that is estimated when vehicle V_{i-1} is reported by the sensor

Step 5: If the value of the % change of quality of flow is lesser than a preset threshold, then the algorithm raises an alarm. In this manner, the Quality of Flow Algorithm was implemented on the data (separately for each lane) for a number of values of the threshold of rate of change of Quality of Flow, -5%, -10%, -15%,, -65%. Alarms were raised when each of these thresholds were violated (the rate of change of quality of flow was lesser than or equal to the threshold). These Algorithm alarms were then compared with the manual alarms to determine the number of correct, missed and false alarms. The detection rate and false alarm rate were then calculated and the performance curves (Detection Rate v/s False Alarm Rate Plots) were drawn.

On analyzing the performance curves and the corresponding results, it was found that the thresholds that were optimum for the right, middle and left lanes were -40%, -35% and -20%. However, for the right and middle lanes, the detection rates obtained for the optimum thresholds were extremely low (28.75% and 31.24% respectively). It was noticed that, for thresholds lower than the above optimum thresholds for the right and the middle lane, while the detection rates were much higher, the false alarm rates were also correspondingly more. Hence, the methodology had to be changed in such a way that some of these detections could be retained while eliminating some false alarms at the same time. On careful analyses of the time series plots that were drawn with quality of flow index for the right and middle lanes, it was observed that most manual alarms were raised in the vicinity of the times at which the quality of flow reduced sharply to values less than 2500. Hence, this additional check was also included for determining whether an alarm should be raised or not. Thus the algorithm was modified to raise alarms when the threshold of rate of change of quality of flow was violated and if the quality of flow was lesser than 2500 as shown in figure 15.10.

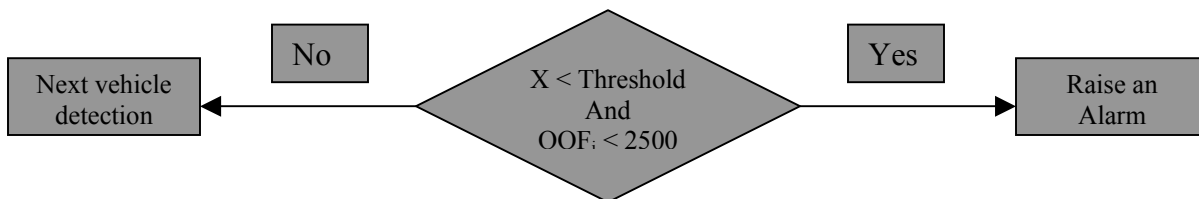


Figure 15.10: Modification made to Quality of Flow Index Algorithm for right and middle lanes

Table 15.3 shows the average detection rate and false alarm rate per lane for all time periods that were obtained with the Headway Algorithm. Figure 15.11, derived from table 15.3 depicts the detection rate v/s false alarm rate curves (performance curves) for each lane.. As can be seen, a threshold of 20% when the quality of flow is lesser than 2500 produces the best combination of detection rate and false alarm rate for the right lane. For the left lane, a threshold of 20% again produces the optimum combination of detection rate and false alarm rate but without the additional check that was employed for the right and middle lanes.

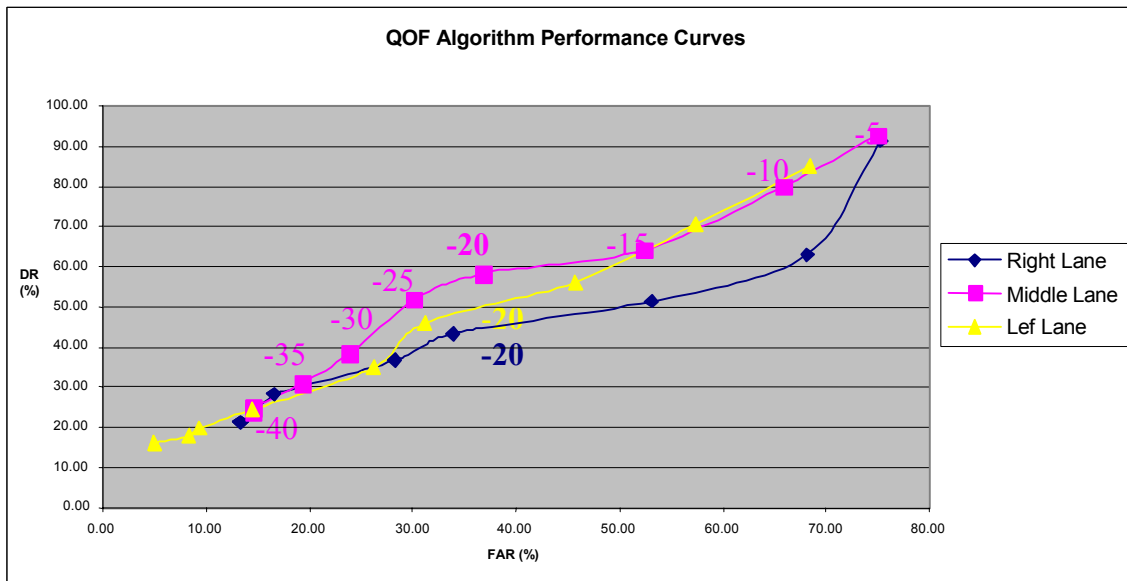


Figure 15.11: Performance Curves Obtained For Quality of Flow Index Algorithm

Table 15.3: Quality of Flow Index Algorithm Results

Quality of Flow Index Algorithm Results				
Right Lane Results		Manual Alarms = 4.55 / hour		
"% Decrease of QOF" Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Ra
50%	1.11	0.96	0.15	21.16
45%	1.11	0.96	0.15	21.16
40%	1.11	0.96	0.15	21.16
35%	1.54	1.29	0.25	28.26
30%	1.54	1.29	0.25	28.26
25%	2.32	1.67	0.66	36.64
20%	2.98	1.97	1.01	43.35
15%	4.99	2.34	2.65	51.45
10%	8.99	2.86	6.13	62.96
5%	16.83	4.16	12.67	91.51
Middle Lane Results		Manual Alarms = 5.17 / hour		
"% Decrease of QOF" Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Ra
50%	1.45	1.24	0.21	24.01
45%	1.51	1.29	0.22	24.94
40%	1.51	1.29	0.22	24.94
35%	1.99	1.61	0.38	31.07
30%	2.61	1.99	0.62	38.45
25%	3.83	2.68	1.15	51.77
20%	4.77	3.01	1.75	58.27
15%	6.97	3.32	3.65	64.19
10%	12.09	4.14	7.95	80.00
5%	19.12	4.79	14.32	92.72
Left Lane Results		Manual Alarms = 5.69 / hour		
"% Decrease of QOF" Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Ra
50%	0.97	0.92	0.05	16.20
45%	0.97	0.92	0.05	16.20
40%	1.10	1.01	0.09	17.80
35%	1.24	1.13	0.12	19.82
30%	1.62	1.39	0.24	24.41
25%	2.69	1.99	0.71	34.91
20%	3.79	2.61	1.18	45.88
15%	5.89	3.20	2.69	56.24
10%	9.41	4.01	5.40	70.56
5%	15.34	4.84	10.50	85.09

15.2.4 Heuristic Integrated Algorithm

It has already been proved from the detection rate and false alarm rate values obtained by the stand-alone algorithms that there are some typical patterns that exist whenever an undesirable flow condition occurs. Undesirable traffic flow conditions are normally detected when a threshold number of successive headway violations is exceeded or a certain percentage reduction in quality of flow index value is observed or when a certain threshold value of traffic pressure is exceeded. However, the three stand-alone algorithms have to be integrated into one algorithm or methodology so that better detection rates can be obtained. Hence, a heuristic technique for amalgamating the three stand-alone algorithms was developed. The algorithm developed assigns scores to each of the outputs produced by the individual (stand-alone) algorithms, then sums them up together and compares the sum with a preset threshold. The algorithm raises alarms whenever the sum of the scores exceeds the thresholds as shown in the flowchart in figure 15.12.

Step 1: Whenever the presence of a vehicle is reported and the relevant traffic measurements obtained, the three individual algorithms, i.e. the headway algorithm, the quality of flow algorithm and the traffic pressure algorithm are executed and the number of successive violations, the percentage change of quality of flow, the value of quality of flow and the value of traffic pressure are determined as shown in the flow charts shown in figures 15.4, 15.7 and 15.9.

Step 2: These values are then transformed into scores as shown in the flow chart in figure 15.10. Each score transformation is associated with a multiplication factor that is unique for each lane. This multiplication factor is obtained for each lane and for each individual algorithm by determining the integer value of the ratio of detection rate to false alarm rate of the most optimum threshold obtained by that particular algorithm. Thus,

$$\text{Multiplication Factor} = \text{INT}(\text{Detection Rate} / \text{False Alarm Rate}) \quad (4.8)$$

The values of multiplication factors that were used in the integrated algorithms are shown in table 15.4.

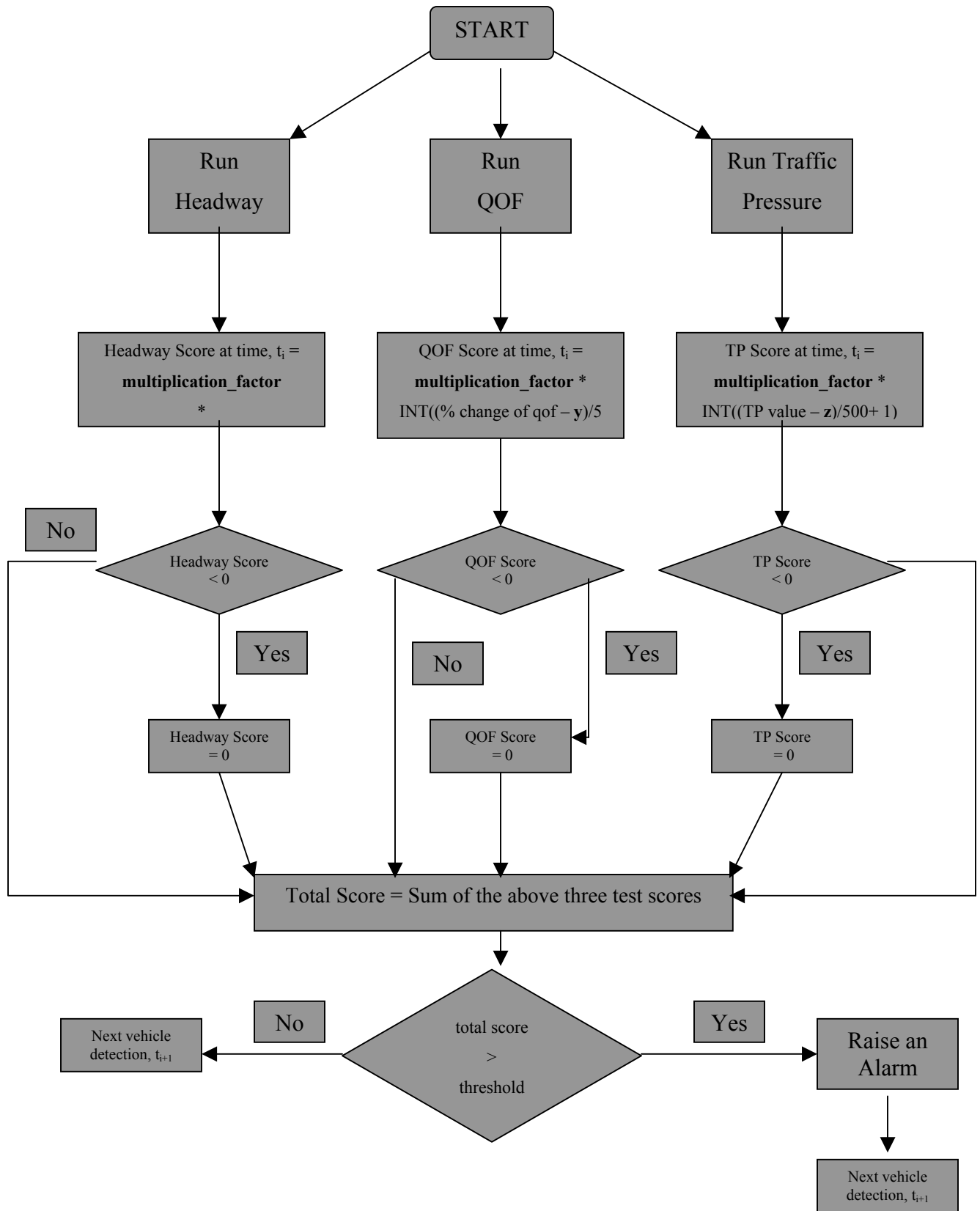


Figure 15.12: Flow Chart Describing the Heuristic Integrated Algorithm

Thus each score transformation consists of determining the displacements of the number of violations or the percentage change of quality of flow or the value of traffic pressure from the corresponding optimum values and then multiplying it by a multiplication factor.

Thus,

$$\text{Headway Score} = \text{multiplication factor} * (\text{violations} + 1 - x) \quad (4.9)$$

Where,

x is the optimum number of headway violations as determined by executing the headway algorithm on the data

$$\text{QOF Score} = \text{multiplication_factor} * \text{INT}((\% \text{ change of qof} - y)/5 + 1) \quad (4.10)$$

Where,

y is the optimum value of percentage change of quality of flow as determined by executing the quality of flow algorithm on the data

Thus, for QOF score, the distance of the five percent interval in which the current value of percent change of quality of flow lies from the optimum value is being determined.

$$\text{TP Score} = \text{multiplication_factor} * \text{INT}((\text{TP value} - z)/500 + 1) \quad (4.11)$$

Where,

z is the optimum value of traffic pressure as determined by executing the traffic pressure algorithm on the data

Thus, for TP score, the distance of the interval (each interval = 500) in which the current value of traffic pressure lies from the optimum value is being determined.

In this manner, the headway score, the quality of flow (QOF) score and the traffic pressure (TP) score is calculated each time a vehicle is detected by the sensor.

Step 3: Having obtained the headway score, the quality of flow score and the traffic pressure, the total score is determined by summing up the three above scores. Thus,

$$\text{Total Score} = \text{Headway Score} + \text{QOF Score} + \text{TP Score} \quad (4.12)$$

If the value of the total score is greater than a preset threshold, then the algorithm raises an alarm. In this manner, the Heuristic Integrated Algorithm was implemented on the data (separately for each lane) for a number of values of the threshold of total score, 1, 2, 3... 15. Alarms were raised when each of these thresholds were violated (the total score was greater than

or equal to the threshold). These Algorithm alarms were then compared with the manual alarms to determine the number of correct, missed and false alarms. The detection rate and false alarm rate were then calculated and the performance curves (Detection Rate v/s False Alarm Rate Plots) were drawn.

Table 15.4: Values of Parameters used in Heuristic Integrated Algorithm

Lane	Right	Middle	Left
Parameter			
Multiplication Factor (Headway Score)	2	1	1
Value of 'x'	4	4	5
Multiplication Factor (QOF Score)	3	2	1
Value of 'y'	20	20	20
Multiplication Factor (TP Score)	2	1	1
Value of 'z'	3500	2000	1500

On analyzing the performance curves and the corresponding results (table 15.5 and figure 15.13), it was found that the thresholds that were optimum for the right, middle and left lanes were total scores of 4, 6 and 4 respectively. These produced detection and false alarm rates of 66.55% and 36.79% for the right lane, 71.36% and 43.05% for the middle lane and 61.34% and 41.65% for the left lane respectively. The detection rates are substantially more improved than those produced by the individual algorithms. However, the false alarm rates are also more than those produced by the individual algorithms.

Table 15.5: Heuristic Integrated Algorithm

Heuristic Integrated Algorithm Results					
Right Lane Results	Manual Alarms = 4.55 / hour				
"Scores" Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
11	0.91	0.89	0.02	19.63	2.17
10	0.97	0.89	0.08	19.63	8.28
9	1.26	1.05	0.21	23.09	16.54
8	1.48	1.20	0.28	26.33	18.83
7	1.74	1.26	0.48	27.66	27.59
6	3.19	2.23	0.95	49.10	29.91
5	3.47	2.34	1.13	51.39	32.64
4	4.79	3.03	1.76	66.55	36.79
3	6.32	3.74	2.58	82.24	40.82
2	6.87	3.80	3.08	83.42	44.79
1	8.70	4.10	4.60	90.21	52.83
Middle Lane Results	Manual Alarms = 5.17 / hour				
"Scores" Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
15	1.87	1.53	0.35	29.50	18.45
14	1.90	1.53	0.38	29.50	19.79
13	2.10	1.53	0.57	29.50	27.29
12	2.84	2.02	0.82	39.03	28.94
11	2.90	2.02	0.88	39.03	30.37
10	3.26	2.18	1.08	42.13	33.12
9	3.82	2.47	1.35	47.85	35.26
8	4.16	2.58	1.58	49.86	37.98
7	4.25	2.58	1.67	49.86	39.36
6	5.85	3.45	2.40	66.79	40.96
5	5.99	3.50	2.48	67.72	41.51
4	6.48	3.69	2.79	71.36	43.05
3	7.50	4.10	3.41	79.26	45.39
2	8.14	4.22	3.92	81.66	48.12
1	10.09	4.71	5.39	91.02	53.38

Table 15.5 Continued: Heuristic Integrated Algorithm

Left Lane Results	Manual Alarms = 5.69 / hour				
"Scores" Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
15	0.35	0.32	0.03	5.67	8.21
14	0.35	0.32	0.03	5.67	8.21
13	0.35	0.32	0.03	5.67	8.21
12	0.67	0.59	0.07	10.43	11.14
11	0.69	0.59	0.10	10.43	14.32
10	0.73	0.59	0.14	10.43	18.56
9	1.36	1.03	0.33	18.11	24.52
8	1.46	1.03	0.43	18.11	29.23
7	1.55	1.03	0.52	18.11	33.61
6	3.67	2.36	1.32	41.43	35.86
5	4.03	2.47	1.56	43.37	38.71
4	5.98	3.49	2.49	61.34	41.65
3	7.02	3.77	3.25	66.20	46.32
2	7.87	4.02	3.85	70.58	48.96
1	9.85	4.73	5.12	83.11	51.97

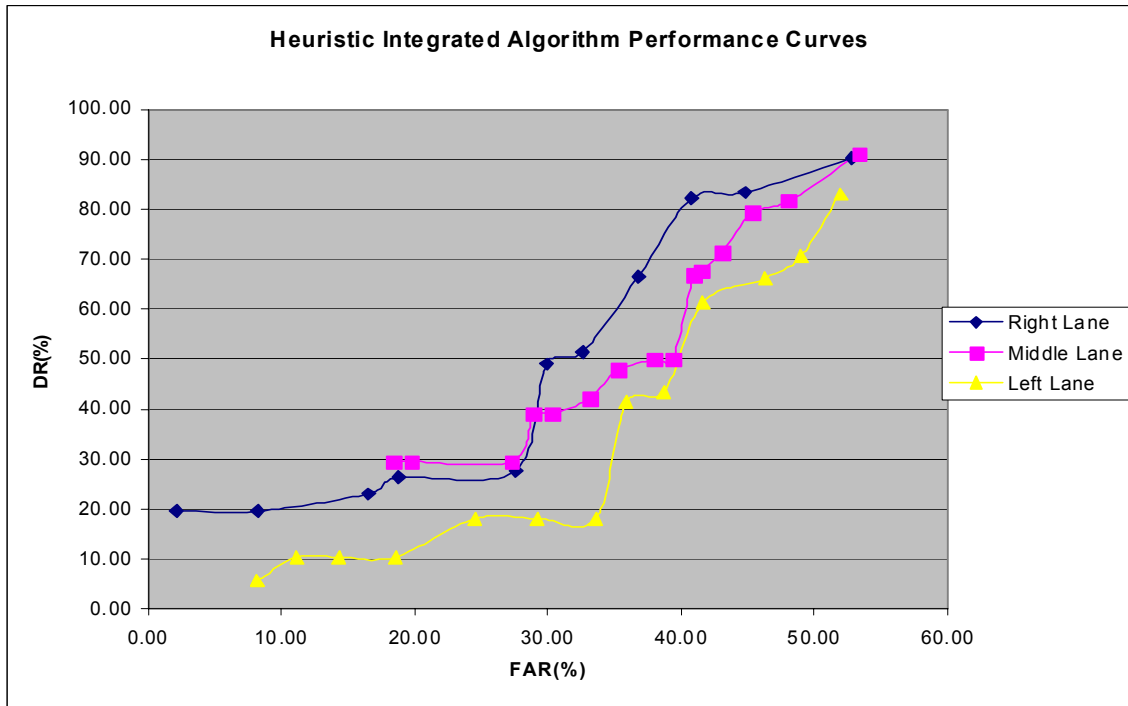


Figure 15.13: Performance Curves Obtained For Quality of Flow Index Algorithm

15.2.5 Neural Network Integrated Algorithm

Though, the heuristic integrated algorithm proved that better detection rates can be obtained by integrating the three individual algorithms, it also substantially increased the false alarm rate by much. Hence, a scientific method to integrate the three individual algorithms was sought and this was the main reason for selecting the neural networks approach to integrate the three algorithms.

An Artificial Neural Network is an information-processing paradigm that implements a simplified model of a biological neural network. An Artificial Neural System can be compared to a nonlinear directed graph with a weighted edge that is able to store patterns by changing the edge weights. According to Hecht-Nielsen, “a neural network is a parallel, distributed information processing structure consisting of processing elements interconnected together with unidirectional signal channels called connections”. ANNs, like animal brains, learn by example. An ANN is configured to serve a particular application through a learning process. Learning involves adjustment of the synaptic connections that exist between neurons. Neural networks,

have in the past, been extensively used for pattern recognition such as speech processing, character recognition and image classification. In transportation engineering also neural networks have begun to be used. Traffic assignment, traffic flow prediction, traffic signal setting and incident detection are some of the many problems to which the neural networks' approach has been used.

Neural networks were selected to integrate the variables used in the different individual "stand-alone" algorithms because of all the advantages that using neural networks brings to a problem. Since detection and quantification of patterns existing between the variables and the manual (visual) alarms was required and lots of examples representing the sought behavioral patterns were available from the data, it was decided that neural networks would be ideally suited to generate solutions. This decision was reached at taking into consideration the fact that neural networks have a remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are normally too complex to be noticed by either humans or other computer techniques. A trained neural network is extremely proficient in the category of information it has been given to analyze. This proficiency can then be used to provide forecasts or solutions given new situations. The main advantages of neural networks are fast processing speed, parallel and distributed representation and processing of complex information, ability to be trained to perform non-linear mapping of patterns and the ability to produce good results even with imperfect data. Other advantages include:

Adaptive learning: Neural networks can learn how to perform certain tasks based on the data given as training.

Self-Organization: Neural networks can create their representation of the information that they receive during training.

Real Time Operation: Neural network computations may be carried out in parallel, thus giving users the advantage of real time results.

All these advantages can be harnessed towards detection of accident prone or undesirable flow conditions. Hence, it was decided that using neural networks would be the best approach to meet the objectives of the present research.

15.2.5.1 Characteristics of Neural Networks

Processing elements and their connections form the basic components of neural networks. Processing Elements (PEs) also referred to as nodes, short term memory, neurons or threshold logic units are the artificial neural network system components where almost all, if not all, of the computing is done. Each processing element (PE) or neuron has one output connection which fans out into as many collateral connections as desired with each collateral connection carrying the same PE output signal. The PE output signal can be of any mathematical type desired. All of the processing that goes on within each PE is entirely local. Thus, in a neural network, each processing element receives input signals either from an external source or other processing elements, processes these signals and transmits the resulting signal to other processing elements. In this manner, complex information is represented and processed in a parallel and distributed method. The processing elements in an artificial neural network are organized into layers as shown in figure 15.14.

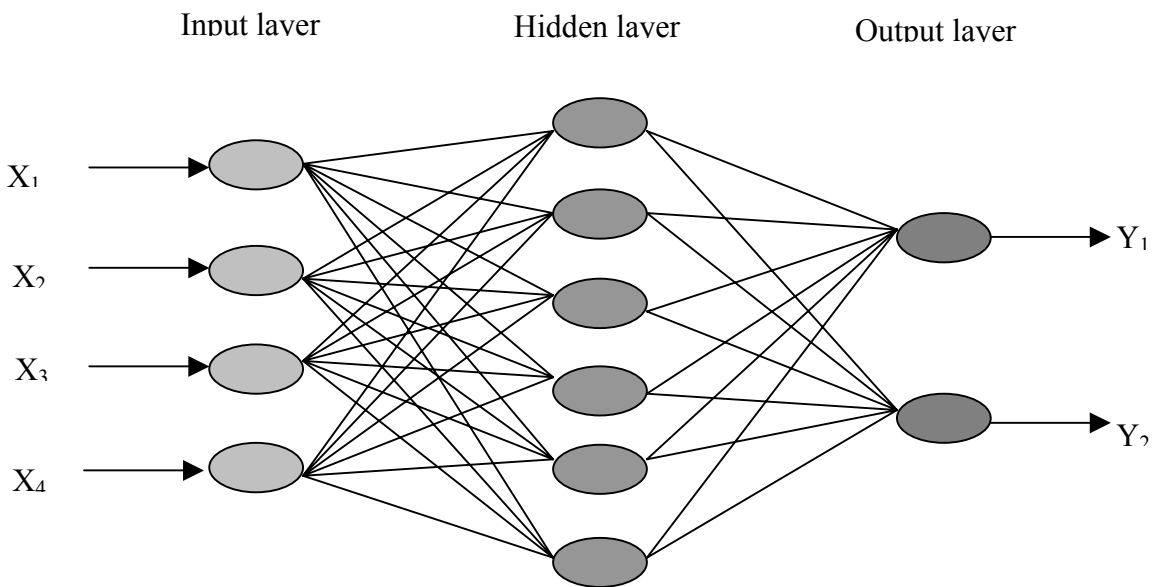


Figure 15.14: Representation of a Neural Network (partially shown)

The layer of processing elements that receives input signals from external sources is called the input layer. The layer of processing elements that transmits signals as output to external sources is called the output layer. Layers of processing elements in between the input layer and the

output layers are called as hidden layers. The set of signals that are supplied to the processing elements in the input layer (X_1, X_2, \dots, X_M) is called the input vector. The corresponding output signals (Y_1, Y_2, \dots, Y_M) is called the output vector. Each processing element is associated with an activation function. This activation function transforms the weighted sum of the input signals into the processing elements' output.

Mathematically, the operations within a neural network can be summarized as follows:

Input Neurons: $X_1, X_2, X_3, \dots, X_N$

Output Neurons: $Y_1, Y_2, Y_3, \dots, Y_M$

Kth Hidden Layer: $H_{k1}, H_{k2}, \dots, H_{kp}$

$$I_j = \sum_{i=0}^n W_{ji} X_i \tag{4.13}$$

$$Y_j = f(I_j) \tag{4.14}$$

Training forms an integral part of the entire process of developing an artificial neural network. Training is necessary so that the neural network can adapt the connection weights and parameters of the activation function so that desired mapping between the input and output patterns can be performed. Training is performed by successively applying the input vectors and adjusting the corresponding connection weights according to a preset procedure. Ultimately the weights normally converge to certain values in such a way that the network detects and learns the patterns occurring between the input and the output. There are two types of training algorithms, supervised and unsupervised. Both the input and the output vectors are presented to the neural network in supervised training. The outputs produced by the neural network are compared with the desired output and the weights are changed so as to reduce the output error. In unsupervised training, only the input vectors are presented to the neural network. The network adjusts the weights by some preset procedure so that similar input vectors will produce similar output patterns. This type of training detects different patterns and groups together those input vectors causing these patterns.

15.2.5.2 Model Selection

Different neural networks have been designed to perform specific tasks appropriate for their problem sets. Each network is designed to receive certain types of inputs and produce certain outputs. Simpson [8] listed out 26 different types of neural networks and tabulated their properties. Table 15.6 has been reproduced from Simpson's book and it lists out the important properties of all the neural networks. The properties listed out are:

- I. Number of processing element layers
- II. Forms of the input and output signals
- III. Mode of operation
- IV. Off-line or on-line learning capability
- V. Supervised or unsupervised learning or training

The input and output signals presented to and produced by the neural networks are of two forms, analog and digital. Analog signals are continuous and binary signals are discrete with values of either 0 or 1. The operation of neural networks can be either discrete or continuous. A discrete neural network will produce an output vector every time an input vector is presented to it. A continuous neural network can receive continuous input signals that vary with time and simultaneously produce output signals. Some neural networks can learn with online data while some others can only train with offline data. Some neural networks are required to self-adapt to the operating conditions while some neural networks are required to train with prior data.

For application in the present study, the following factors were considered while selecting the type of neural network:

1. Number of Processing Element Layers: Mapping of a large amount of information, such as number of headway violations, value of quality of flow index, percentage reduction of quality of flow and value of traffic pressure which form numerous combinations had to be done. Therefore pattern associator neural networks and auto associator neural networks were considered to be not suitable. Also, a network that could memorize and store hetero-associative patterns between input and output vectors was needed. Hence, the network had to have at least two layers of processing elements. Also, a network that had the ability to perform non-linear mapping was sought.

Application of these constraints excluded BAM, ABAM, TAM, FAM, AG, SG, DA, CH, BSB, FCM, LAM, OLAM and ADALINE/MADALINE types of neural networks.

network name	layer of PEs	type of signals*		mode of operation**	method of training	
		input	output		on/offline	un/supervised
Additive Grossberg (AG)	1	A	B	C	online	unsupervised
Shunting Grossberg (SG)	1	A	B	C	online	unsupervised
Adaptive Resonance Theory 1 (ART1)	2	B	B	D & C	online	unsupervised
Adaptive Resonance Theory 2 (ART2)	2	A	B	D & C	online	unsupervised
Discrete Autocorrelator (DA)	1	B	B	D	offline	unsupervised
Continuous Hopfield (CH)	1	A	A	D	offline	unsupervised
Bidirectional Associative Memory (BAM)	2	B	B	D	online	unsupervised
Adaptive Bidirectional Associative Memory (ABAM)	2	A	A	C	offline	unsupervised
Temporal Associate Memory (TAM)	2	B	B	D	offline	unsupervised
Learning Matrix (LM)	2	A	A	C	offline	unsupervised
Drive Reinforcement	2	A	A	D	offline	unsupervised
Sparse Distributed Memory (SDM)	3	B	A	D	offline	unsupervised
Linear Associate Memory (LAM)	2	A	A	D	offline	unsupervised
Optimal Linear Associative Memory (OLAM)	2	A	A	D	offline	unsupervised
Fuzzy Associative Memory (FAM)	2	A	A	D	offline	unsupervised
Learning Vector Quantization (LVQ)	2	A	A	D	offline	unsupervised
Counterpropagation (CPN)	3	A	A	D	offline	unsupervised
Self-Organizing Feature Map (SOFM)	2	A	B	D	offline	unsupervised
Brain-State-in-a-Box (BSB)	1	A	A	D	offline	supervised
Fuzzy Cognitive Map (FCM)	1	A	A	D	offline	supervised
Multi-Layer Feed-Forward Net (MLF)	23	A	A	D	offline	supervised
Adaline/Madaline	>1	A	B	D	offline	supervised
Boltzman Machine (BM)	3	B	B	D	offline	supervised
Cauchy Machine (CM)	3	B	B	D	offline	supervised
Adaptive Heuristic Critic (AHC)	2 or 3	A	B	D	offline	supervised
Associate Reward-Penalty (ARP)	2	A	B	D	offline	supervised
Avalanche Matched Filter (AMF)	2	A	A	D	offline	supervised

* type of signals: A=analog, B=binary

** mode of operation: C=continuous, D=discrete

Table 15.6: Characteristics of Different Neural Networks (Reproduced from [8])

2. Input Data: The data that was to be presented to the neural network consisted of continuous values for the number of headway violations, value of quality of flow index, percentage reduction of quality of flow and value of traffic pressure variables. Hence a network was sought that could receive continuous (analog) values. The application of this constraint excluded ART1, DA, BAM, TAM, SDM, BM and CM types of neural networks. A point to be remembered is that networks designed to receive analog inputs can also receive values of 0 or 1.
3. Type of Training: The desired output vector or target values of output had to be presented to the neural network so that it could detect and memorize the patterns that existed between the input and output vectors. The application of this constraint

excluded AG, SG, ART1, ART2, DA, CH, BAM, ABAM, TAM, LM, Drive Reinforcement, SDM, LAM, OLAM, FAM, LVQ, CPN and SOFM types of neural networks.

4. Training Capability: Training patterns should be selected carefully so that the network parameters trained to detect the existing typical patterns can be stored. In the current study, online training is not possible because any occurrence of an accident prone or undesirable traffic flow condition had to be verified by looking at videotapes or live feed of the traffic at the Machine Vision Sensor location and this process would involve considerable delay. Training therefore had to be done offline. The application of this constraint excluded AG, SG, ART1, ART2 and BAM types of neural networks.
5. Mode of Operation: An output vector or value was required every time the Machine Vision Sensor detected a vehicle. In other words, the neural network had to output a value every time an input vector was presented to it. This corresponds to discrete mode of operation. The application of this constraint excluded AG and SG types of neural networks.

Having applied the above constraints, the remaining neural networks from the table that satisfy the conditions are MLF, AHC and AMF types of neural networks. Among these networks, the Multi-Layer Feed-Forward Neural Network is the most frequently applied network. Hence, it was selected for this study.

15.2.5.3 The Multi-Layer Feed-Forward Neural Network

The multi-layer feed-forward neural network is also called as the back propagation network or the multi-layer perceptron. Rosenblatt [8] first studied the concept of the perceptron. Extensive demonstrations and applications of the multi-layer feed-forward neural networks have been reported by many researchers, the first few being D.B Parker, P.J Werbos, G.E Hinton, R.J Williams, and D.E Rumelhart [8]. This sub-section describes the operation of the MLF.

Feed-forward networks usually consist of three to four layers in which the processing elements are logically arranged. The first and last layers are the input and output layers respectively, and there are usually one or more hidden layers in between the outer two layers. Many studies have

concluded that a minimum of three layers (input and output layers with one hidden layer) is required to solve complex problems. The term Feed-Forward is used because the information is only allowed to "travel" in one direction within the neural network. This means that the output of one layer becomes the input of the next layer, and so on and so forth. In order for this to occur, each layer is fully connected to next layer (each processing element is connected by a weight to a processing element in the next layer). The input layer does have a part in computation, rather than just receiving the inputs. The raw data is computed, and activation or threshold functions are applied in all layers. This process occurs until the data reaches the processing elements in the outer layer.

When a set of input values (X_1, X_2, \dots, X_M) is presented to the network, each of the processing elements in the hidden layer receives the weighted sum of all the input values (the interconnected processing elements have connection weights of W_{ij}, V_{jk} , as shown in the figure). This is passed through a transfer or activation function and produces an output value Y_k .

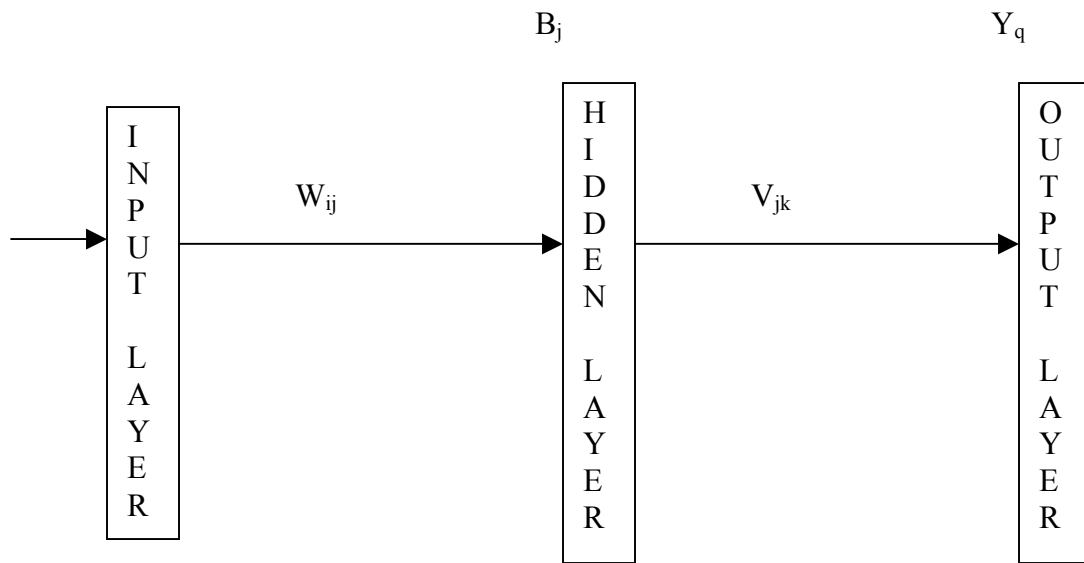


Figure 15.15: Topology of A Simple Multi-Layer Feed-Forward Network

Mathematically,

$$B_j = f(\sum_n (w_{ij} X_{ij} + \Theta_j)) \quad \forall j = 1, \dots, p \quad (4.15)$$

where the threshold value of the j th processing element in the hidden layer, Θ_j , adjusts the magnitude of the output, and $f(\)$ is the activation function. There are three main types of

activation functions, hard-limit, tan-sigmoid and log-sigmoid functions. The activation function has to be a non-linear function, that when applied to the weighted sum of inputs, including the bias (the net input), will determine the output of the network. A bias is an assumed input that is always equal to 1.

The same process is repeated for processing elements in the output layer, where,

$$Y_j = f(\sum_n (V_{ij} B_{ij} + \Phi_j)) \quad \forall k = 1, \dots, q \quad (4.16)$$

where the threshold value of the kth processing element in the output layer, Φ_k . The vector $(Y_1, Y_2, \dots, Y_q)_n$ is the output vector.

The most common and widely used training algorithm for multi-layer feed-forward neural networks is the back-propagation algorithm. It is based on the Delta Rule that basically states that if the difference (delta) between the user's desired output and the network's actual output is to be minimized, the weights must be continually modified. The result of the transfer function changes the delta error in the output layer. The error in the output layer is adjusted or fixed, and therefore it can be used to change the input connection weights so that the desired output may be achieved.

The training process starts by converting all input weights to small non-zero values. A subset of training samples is presented to the network. Each pattern is fed into the network separately and the obtained outputs are compared to the desired outputs and the size of the error is measured. The input connection weights are adjusted in such a way that the error will be minimized. This process is repeated (many epochs) several times until satisfactory results are obtained. Training can stop when the error obtained is less a certain threshold or limit. A mean square of errors value of 0.001 is considered good. (W11). The mean square of errors is computed by summing the squared differences between what the value of an output variable should be versus what it actually is, then dividing by the number of components that went into the sum.

The simplest implementation of back propagation learning updates the network weights in the direction in which the performance function, usually the mean square of errors, decreases most rapidly- the negative of the gradient. Mathematically the iteration can be expressed as:

$$X_{k+1} = X_k - \alpha k g_k \quad (4.17)$$

Where,

X_k is a vector of current weights,

g_k is the current gradient

and α_k is the learning rate (a constant used in neural networks that affects the speed at which neural networks detect and memorize the typical patterns in the data set)

15.2.5.4 Training and Testing Results

Neural networks were designed, trained and tested on the data corresponding to all three lanes (one for each lane: NNRL, NNML, NNLL). 80% of the data was used for training the neural networks and 20% of the data was used for testing or simulation. This means that all the data corresponding to a lane was arranged together and 80% of it was used to be presented input and desired output vectors while training the neural networks and 20% of the data was used for obtaining neural network outputs and testing these with the desired output. This was done for data pertaining to all three lanes.

As described previously, positive correlations were observed between successive headway violations, the actual value and the percentage reduction of the quality of index and the traffic pressure value and the manual (visual) alarms for the right and the middle lane. For the left lane, positive correlations were observed between all the above-mentioned variables except the actual value of the quality of flow index and the manual alarms.

Hence, the input vector presented to the neural networks for the right and middle lanes were of the form:

{number of successive headway violations, percentage reduction of quality of flow index, value of quality of flow index, value of traffic pressure}

The input vector presented to the neural network for the left lane was of the form:

{number of successive headway violations, percentage reduction of quality of flow index, value of traffic pressure}

The target vector (desired output values) presented to the neural networks for training comprised of values of 1 at times when an accident-prone condition was detected and values of 0 otherwise.

Since, multi-layered feed-forward networks normally use tan-sigmoid or log-sigmoid functions for activation or transfer, these functions were persisted with in the present study. The log sigmoid transfer function generates outputs whose values range from 0 to 1 while the neuron's input ranges from negative to positive infinity. The tan sigmoid function generates outputs whose values are squashed with the range of -1 to +1.

Mathematically,

$$\text{Log-sig}(x) = 1 / (1 + e^{-x}) \quad (4.18)$$

$$\text{Tan-sig}(x) = (e^x - e^{-x}) / (e^x + e^{-x}) \quad (4.19)$$

MATLAB software was used to train and test the neural networks because the software includes a Neural Network Toolbox that has most of the common features of neural networks. One major advantage of using this toolbox is that it has a user-friendly interface and it has different types of numerical optimization techniques, which reduce the memory requirements and the convergence time.

The training and testing results are shown in table 15.8. Batch-training is performed on all the three neural networks using the steepest descent with momentum algorithm (traingdm). This algorithm was selected because it provided the fastest convergence. Momentum discounts the presence of shallow local minimums by responding to the recent trends in the error surface along with the local gradient. Another training algorithm, Levenberg-Marquardt (trainlm), which is a standard numerical optimization technique, was used in conjunction with the above batch-training algorithm to provide for faster convergence.

Table 15.8: Neural Networks Training and Testing Details and Results

Details	Lane	Right Lane	Middle Lane	Left Lane
	Network	NN_{RL}	NN_{ML}	NN_L
<i>Neural Adopted</i>		Multi-layer Feed-forward Network	Multi-layer Feed-forward Network	Multi-layer Feed-forward Network
<i>Training method</i>		Backpropagation	Backpropagation	Backpropagation
<i>Number of Layers</i>		2	2	2
<i>Mode of Operation</i>		Discrete	Discrete	Discrete
<i>Training Capability</i>		Offline	Offline	Offline
<i>Input Vector</i>		Every vehicle detection	Every vehicle detection	Every vehicle detection
<i>Output Vector</i>		Every vehicle detection	Every vehicle detection	Every vehicle detection
<i>Performance Function</i>		Mean square of errors	Mean square of errors	Mean square of errors
<i>Amount of Data Used for Training</i>		80% of right lane data	80% of middle lane data	80% of left lane data
<i>Amount of Data Used for Testing</i>		20% of right lane data	20% of middle lane data	20% of left lane data
<i>Input Data</i>		<ol style="list-style-type: none"> 1. number of headway violations 2. % reduction of quality of flow index value 3. value of quality of flow index 4. value of traffic pressure 	<ol style="list-style-type: none"> 1. number of headway violations 2. % reduction of quality of flow index value 3. value of quality of flow index 4. value of traffic pressure 	<ol style="list-style-type: none"> 1. number of headway violations 2. % reduction of quality of flow index value 3. value of traffic pressure
<i>Batch Algorithm</i>		TRAINGDM	TRAINGDM	TRAINGDM
<i>Optimization Technique</i>		Levenberg-Marquardt method	Levenberg-Marquardt method	Levenberg-Marquardt method
<i>Transfer function from layer 1 to hidden layer</i>		Log-Sigmoid	Tan-Sigmoid	Log-Sigmoid
<i>Transfer function from layer 1 to output layer</i>		Log-Sigmoid	Log-Sigmoid	Log-Sigmoid
<i>Number of neurons in layer1</i>		4	4	6
<i>Number of neurons in output layer</i>		1	1	1
<i>Value of Performance Function</i>		0.000971927	0.000967234	0.00902552
<i>Manual Alarms</i>		22	19	26
<i>Correct Alarms</i>		18	17	21
<i>False Alarms</i>		5	3	8
<i>Detection Rate</i>		81.81%	89.47%	80.76%
<i>False Alarm Rate</i>		21.74%	30%	27.59%

Neural Network Developed for Right Lane (NNRL)

The neural network designed for the right lane comprised of two layers. The form of the input vector has already been described above. The log-sigmoid function was used as the transfer function for the output produced by the weighted sum of the original inputs and the corresponding biases (first layer). The log-sigmoid function was again used as the transfer or activation function for the weighted sum of the first layer outputs and the corresponding bias (second layer). Four neurons were used in the first layer and 1 neuron was used in the second layer. The goal set for the performance function, the mean square of errors was 0.001 or lesser. Training was stopped as soon as this goal was met.

Each processing element in a layer is connected to every processing element in the next layer.

The schematic representation of the neural network that was trained with input vectors and output vectors corresponding to data from the right lane, NNRL is shown in figure 15.16. As mentioned before, there are 4 input signals each time. These input signals are multiplied by the corresponding weights shown in Table 15.8 and added to the corresponding biases shown in Table 15.9 and deposited in the corresponding processing elements in the 1st layer. Each of these sums (4, because there are four processing elements in the 1st layer) is processed by the log-sigmoid function. The outputs are then multiplied by the corresponding weights shown in table 15.10 and added to the bias value shown in table 15.11 and deposited in the solitary processing element in the 2nd layer. This sum is then processed by the log-sigmoid function again and the output value is the final output signal.

The final values of the weights and biases obtained at the end of training of NNRL are

Weights from Input Signals to Processing Elements in 1st Layer

Table 15.8 : Weights from Input Signals to Processing Elements in 1st Layer

From \ To	Input 1	Input 2	Input 3	Input 4
PE 1 in 1	1.5224	-3.6266	0.060627	0.038513
PE 2 in 1	0.13378	-0.29211	0.00055159	0.023359
PE 3 in 1	-0.61824	-0.0039624	1.5055	-0.09664
PE 4 in 1	-1462.7108	164.0777	0.10274	0.36481

Biases Added to Weighted Sums of Input Signals (Processing Elements in 1st Layer)

Table 15.9: Biases Added to Weighted Sums of Input Signals (Processing Elements in 1st Layer)

To \ From	BIAS
PE 1 in 1	0.27755
PE 2 in 1	-78.0363
PE 3 in 1	-12.7867
PE 4 in 1	4454.79381

Weights from Processing Elements in 1st Layer to Processing Elements in 2nd Layer

Table 15.10: eights from Processing Elements in 1st Layer to Processing Elements in 2nd Layer

From \ To	PE 1 in 1	PE 2 in 1	PE 3 in 1	PE 4 in 1
PE in 2	1.5224	-3.6266	0.060627	0.038513

Biases Added to Weighted Sum of Outputs of 1st Layer (Done in 2nd Layer)

Table 15.11: Biases Added to Weighted Sum of Outputs of 1st Layer (Done in 2nd Layer)

To \ From	BIAS
PE in 2	-8.927

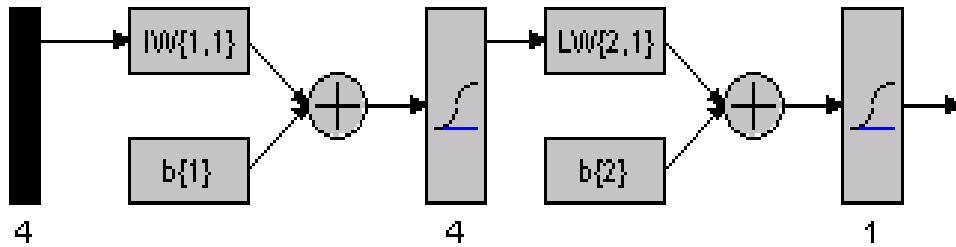


Figure 15.16: Schematic Representation of NNRL

The behavior of the training curve is shown in fig 15.17. As shown, the training was stopped after 43 epochs because the goal was met.

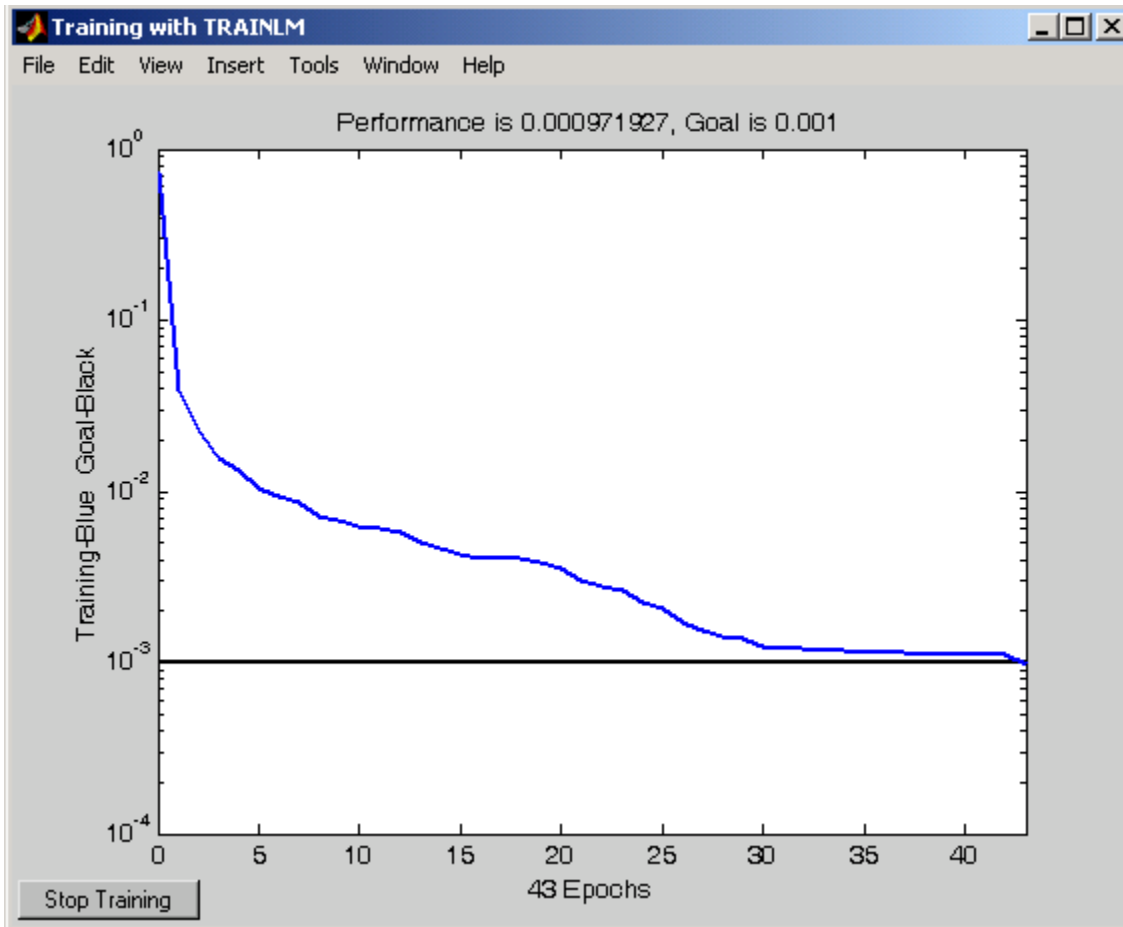


Figure 15.17: Training Curve Obtained for NNRL

The neural network was then simulated or tested on the remaining 20% of right lane data and the outputs as produced by it were obtained. When these output values were compared with the actual output (desired output), the neural network was found to have a detection rate of 81.81% and a false alarm of 21.74%.

Neural Network Developed for Middle Lane (NNML)

The neural network designed for the middle lane also comprised of two layers. The tan-sigmoid function was used as the transfer function for the output produced by the weighted sum of the original inputs and the corresponding biases (first layer). Here again, the log-sigmoid function was used as the transfer or activation function for the weighted sum of the first layer outputs and the corresponding bias (second layer). Four neurons were used in the first layer and 1 neuron was used in the second layer. The goal set for the performance function, the mean square of errors, was a value of 0.001 or lesser.

The schematic representation of the neural network that was trained with input vectors and output vectors corresponding to data from the middle lane, NNML is shown in figure 15.18. As mentioned before, there are 4 input signals each time. These input signals are multiplied by the corresponding weights shown in Table 15.12 and added to the corresponding biases shown in Table 15.13 and deposited in the corresponding processing elements in the 1st layer. Each of these sums (4, because there are four processing elements in the 1st layer) is processed by the tan-sigmoid function. The outputs are then multiplied by the corresponding weights shown in table 15.14 and added to the bias value shown in table 15.15 and deposited in the solitary processing element in the 2nd layer. This sum is then processed by the log-sigmoid function again and the output value is the final output signal.

Weights from Input Signals to Processing Elements in 1st Layer

Table 15.12: Weights from Input Signals to Processing Elements in 1st Layer

From \ To	Input 1	Input 2	Input 3	Input 4
PE 1 in 1	[-0.56261	-0.18204	0.0013379	-0.048964
PE 2 in 1	0.11736	-0.056611	0.058916	0.11211;
PE 3 in 1	0.095746	-0.60694	-0.0009784	0.00029439
PE 4 in 1	3.5249	-1.5871	-4.4515	3.8467

Biases Added to Weighted Sums of Input Signals (Processing Elements in 1st Layer)

Table 15.13: Biases Added to Weighted Sums of Input Signals (Processing Elements in 1st Layer)

To \ BIAS	BIAS
PE 1 in 1	121.9383
PE 2 in 1	-7.462
PE 3 in 1	-10.6543
PE 4 in 1	8.6103

Weights from Processing Elements in 1st Layer to Processing Elements in 2nd Layer

Table 15.14: eights from Processing Elements in 1st Layer to Processing Elements in 2nd Layer

From \ To	PE 1 in 1	PE 2 in 1	PE 3 in 1	PE 4 in 1
PE in 2	-32.4886	12.0539	13.4009	-4.6666

Biases Added to Weighted Sum of Outputs of 1st Layer (Done in 2nd Layer)

Table 15.15: Biases Added to Weighted Sum of Outputs of 1st Layer (Done in 2nd Layer)

To \ BIAS	BIAS
PE in 2	13.9409

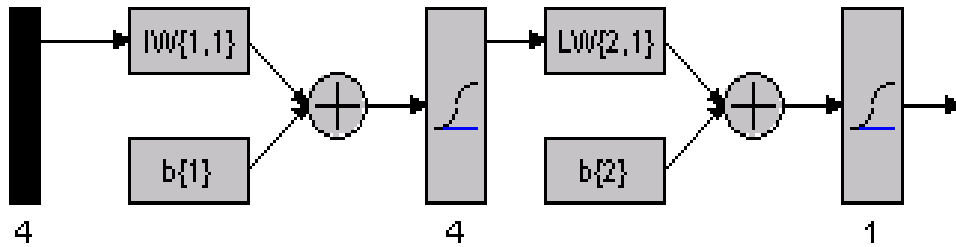


Figure 15.18: Schematic Representation of NNML

The behavior of the training curve is shown in fig 15.19. As shown, the training was stopped after 39 epochs because the goal was met.

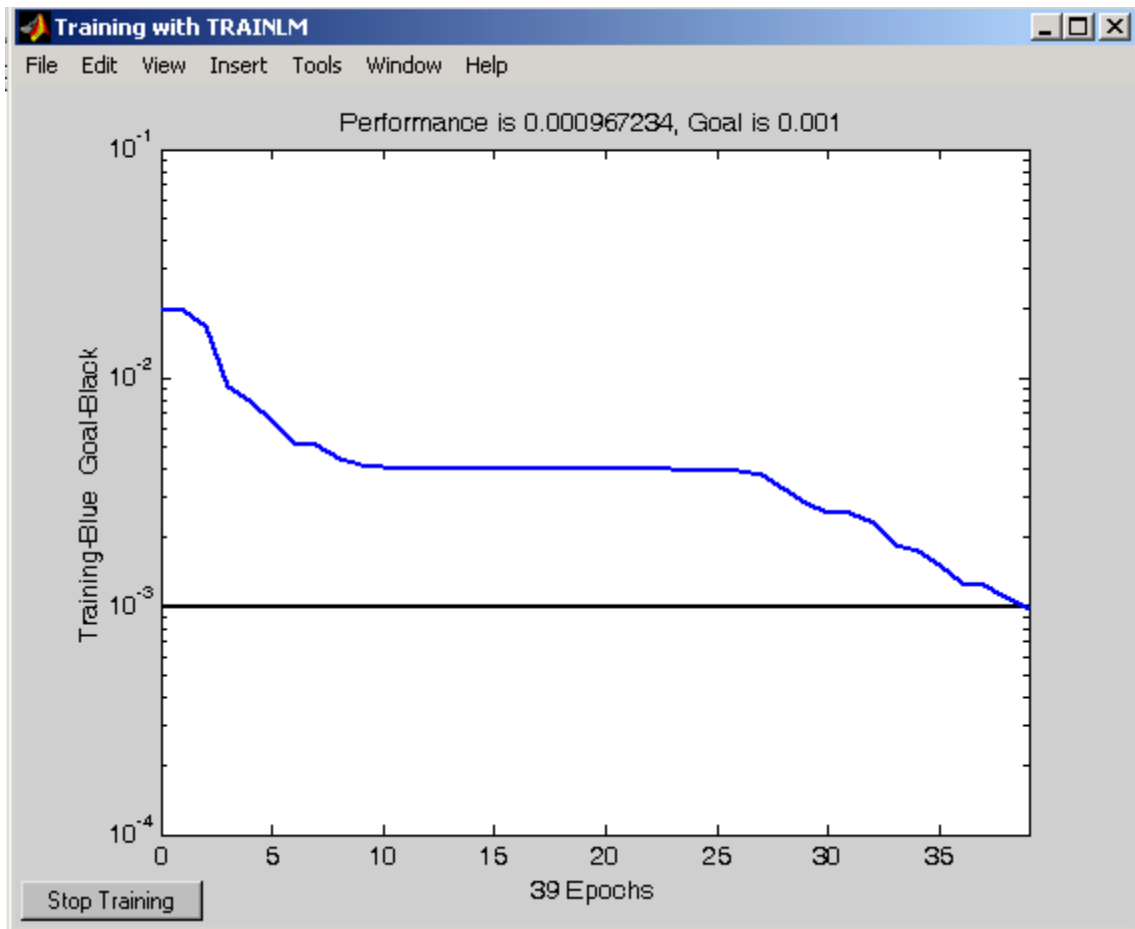


Figure 15.19: Training Curve Obtained for NNML

The neural network was then simulated or tested on the remaining 20% hours of middle lane data and the outputs as produced by it were obtained. When these output values were compared with the actual output (desired output), the neural network was found to have a detection rate of 89.47% and a false alarm of 30.0%.

Neural Network Developed for Left Lane (NNLL)

The neural network designed for the left lane also comprised of two layers. The log-sigmoid function was used as the transfer function for the output produced by the weighted sum of the original inputs and the corresponding biases (first layer). As in the previous two cases, the log-sigmoid function was used as the transfer or activation function for the weighted sum of the first layer outputs and the corresponding bias (second layer). However, for NNLL, 6 neurons were used in the first layer and 1 neuron was used in the second layer. The goal set for the performance function, the mean square of errors, was a value of 0.01 or lesser. Though this goal is larger than what was set for the previous two cases, this value still signifies a good correlation between the desired outputs and the neural network outputs.

The schematic representation of the neural network that was trained with input vectors and output vectors corresponding to data from the left lane, NNLL is shown in figure 15.20. As mentioned before, there are 3 input signals each time. These input signals are multiplied by the corresponding weights shown in Table 15.16 and added to the corresponding biases shown in Table 15.17 and deposited in the corresponding processing elements in the 1st layer. Each of these sums (6, because there are six processing elements in the 1st layer) is processed by the log-sigmoid function. The outputs are then multiplied by the corresponding weights shown in table 15.18 and added to the bias value shown in table 15.19 and deposited in the solitary processing element in the 2nd layer. This sum is then processed by the log-sigmoid function again and the output value is the final output signal.

Weights from Input Signals to Processing Elements in 1st Layer

Table 15.16: Weights from Input Signals to Processing Elements in 1st Layer

From \ To	Input 1	Input 2	Input 3
PE 1 in 1	-7.8741	-18.8819	-0.11789
PE 2 in 1	26.8112	12.4685	-1.2779
PE 3 in 1	0.02118	0.013351	0.0024606
PE 4 in 1	3.8223	-7.5393	-0.060864
PE 5 in 1	-3.5688	10.5729	-0.011964
PE 6 in 1	-5.8169	6.7698	0.016129

Biases Added to Weighted Sums of Input Signals (Processing Elements in 1st Layer)

Table 15.17 : Biases Added to Weighted Sums of Input Signals (Processing Elements in 1st Layer)

To \ From	BIAS
PE 1 in 1	-5.0257
PE 2 in 1	-31.086
PE 3 in 1	-10.0075
PE 4 in 1	-46.4342
PE 5 in 1	; -4.9675
PE 6 in 1	3.6487

Weights from Processing Elements in 1st Layer to Processing Elements in 2nd Layer

Table 15.18: eights from Processing Elements in 1st Layer to Processing Elements in 2nd Layer

From \ To	PE 1 in 1	PE 2 in 1	PE 3 in 1	PE 4 in 1	PE 5 in 1	PE 6 in 1
PE in 2	3.7683	-51.5047	92.167	-30.1797	0.31391	74.2733

Biases Added to Weighted Sum of Outputs of 1st Layer (Done in 2nd Layer)

Table 15.19: Biases Added to Weighted Sum of Outputs of 1st Layer (Done in 2nd Layer)

To \ From	BIAS
PE in 2	-77.8085

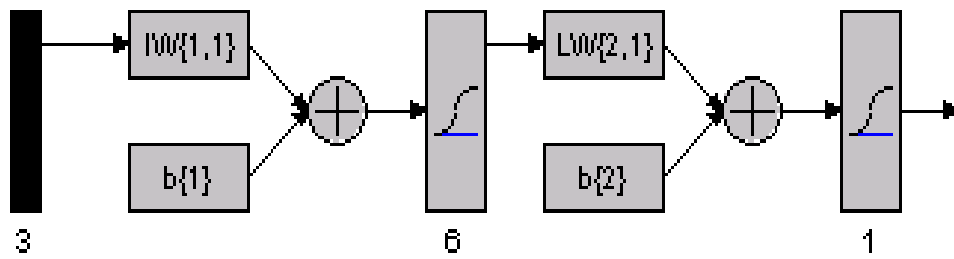


Figure 15.20: Schematic Representation of NNLL

The behavior of the training curve is shown in fig 15.21. As shown, the training was stopped after 17 epochs because the goal was met.

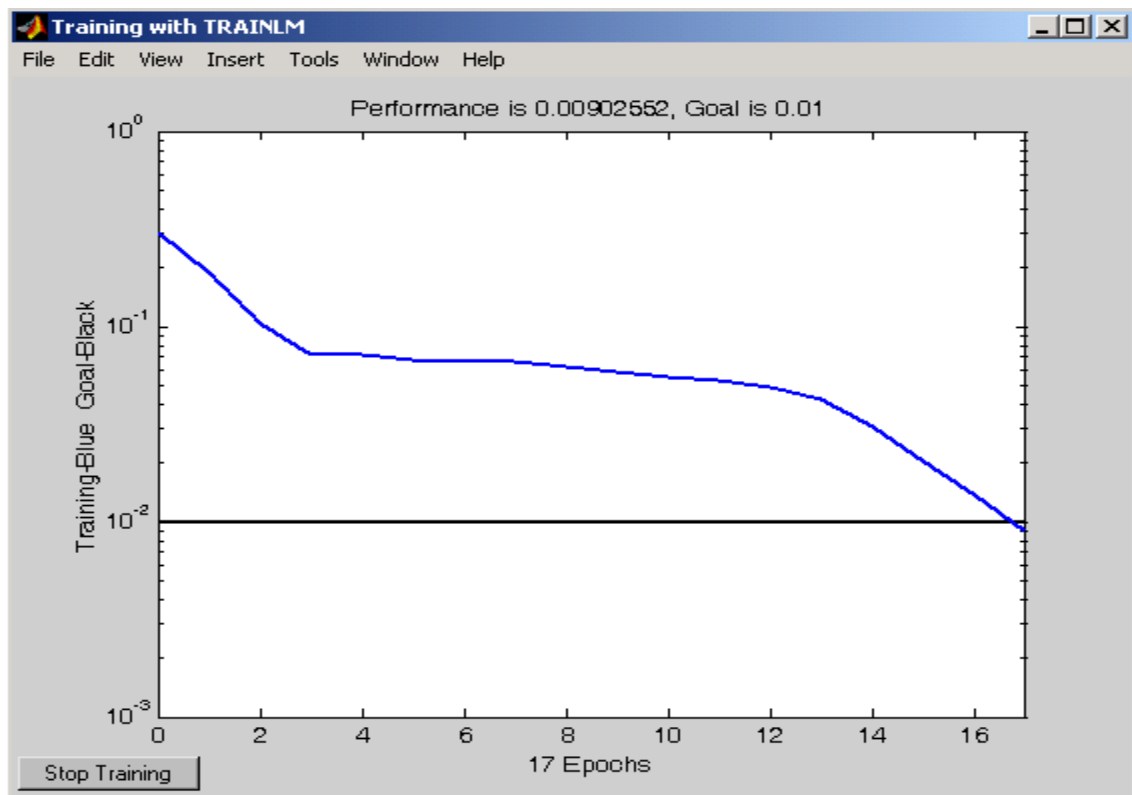


Figure 15.21: Training Curve Obtained for NNLL

The neural network was then simulated or tested on the remaining 20% hours of middle lane data and the outputs as produced by it were obtained. When these output values were compared with the actual output (desired output), the neural network was found to have a detection rate of 80.76% and a false alarm of 27.59%.

16 UTC DETECTION ALGORITHM IMPLEMENTATION

Code was developed in Visual C++ to facilitate easy processing of the raw Machine Vision Sensor data and implementation of the stand-alone algorithms and the heuristic integrated algorithm described in chapter 4. The program helped in expediting the entire process of getting the required text fields from the raw data and also allowed for changes in logic in the algorithms without taking up too much time or effort. Figure 16.1 shows a flow chart which depicts the manner in which the program functions.

The program first prompts the user to input the name of the text file containing the raw Machine Vision Sensor Data. The user is then presented with a set of choices representing the weather conditions that were prevalent at the time of data collection. The choices that are available to the user are:

1. Rainy conditions / Slick pavement
2. Snowy conditions / Icy pavement
3. Dry conditions

If the user inputs a value of either 1 or 2 that corresponds to rainy or snowy weather conditions, he is asked to input the times between which the weather conditions could be described as snowy or rainy.

The program then opens the text file that contains the raw Machine Vision Sensor data. This text file into which the MVS outputs all the data during the period of observation contains traffic measurements from detector station across all three lanes. These measurements are outputted to the text file every time a vehicle passes over the sensors.

The program processes the data so that all the extraneous text is discarded and also disaggregates the data in such a way that data pertinent to each lane is categorized together.

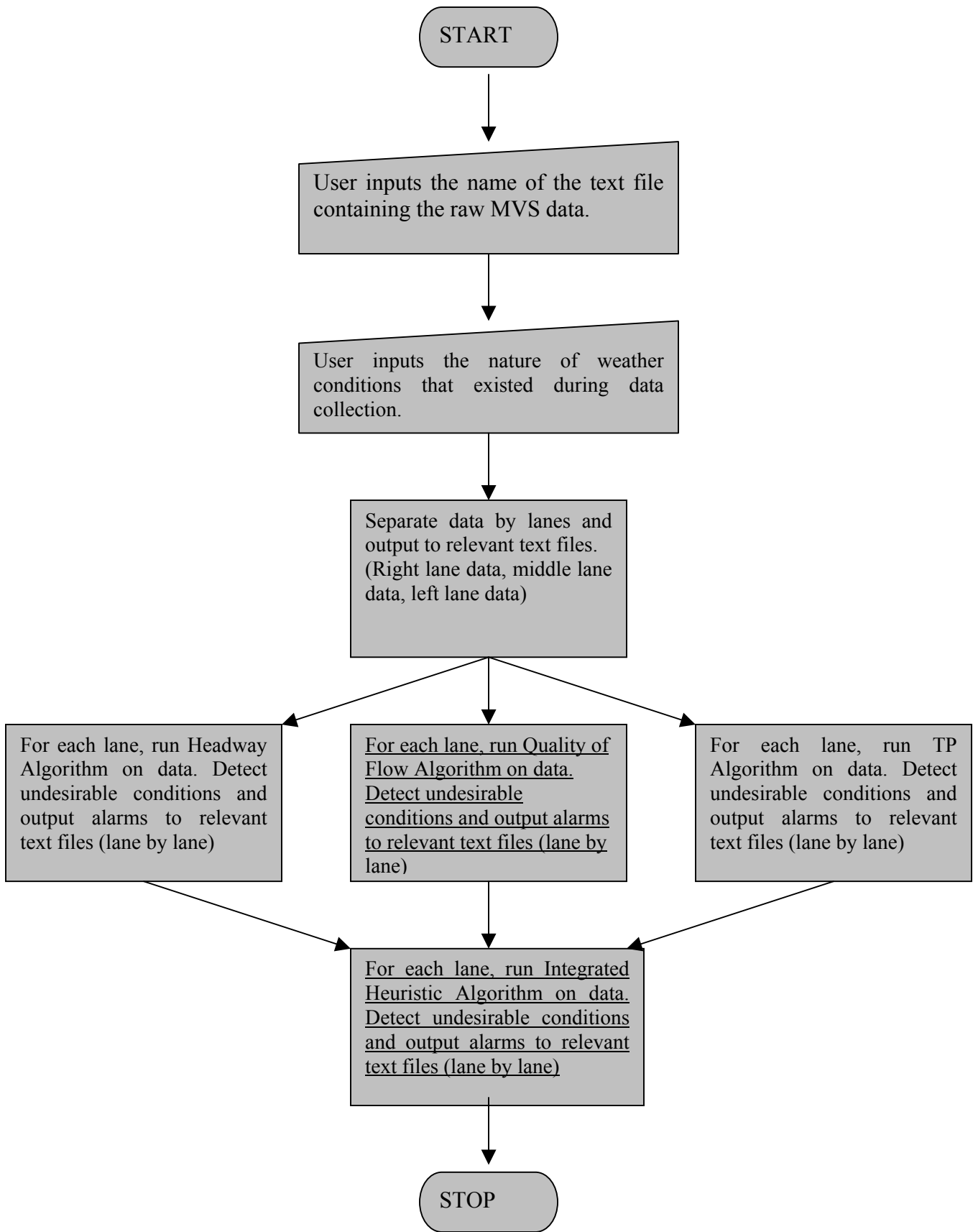


Figure 16.1: Flow Chart depicting the functioning of the Program

Only the detector data, timestamps (the time at which the vehicle arrives at the sensor), speeds and lengths of the vehicles are conserved from the raw data. All other text is shed. The program opens three text files (one for each lane's detector station) and outputs the relevant data to the applicable text file. Figure 16.2 shows some unformatted raw MVS data. Figure 16.3 shows some text from the file corresponding to the left lane after the program has output relevant data to corresponding files.

MVP Description;Det ID;Date & Time;Time Interval;Interval	Hour	Minute	Sec	Speed	Length
Type;State;Speed;Length;Class;Average Flow Rate;Total Volume Count;Arithmetic Mean Speed;Auto Volume;Single-Unit Volume;Tractor-Trailer Volume;Average Time Headway;Average Time Occupancy;Level of Service;Space Mean Speed;Space Occupancy;Density;	16	00	00	55	16
Solo I-394 at Penn Ave.;107;20000531;16:00:00;60;5;,,,,;1680;14;57.141;10;2;2.141;18.070;B;56.973;14.836;29.484;	16	00	02	59	15
Solo I-394 at Penn Ave.;108;20000531;16:00:00;60;5;,,,,;1560;13;58.613;13;0;0;2.305;13.051;B;58.332;11.715;26.742;	16	00	03	58	15
Solo I-394 at Penn Ave.;109;20000531;16:00:00;60;5;,,,,;1320;11;60.453;11;0;0;2.727;10.715;?;60.398;9.688;21.852;	16	00	06	57	17
Solo I-394 at Penn Ave.;101;20000531;16:00:00;;0;1;55;0;7;,,,,,,,,,,,,;	16	00	08	54	16
Solo I-394 at Penn Ave.;103;20000531;16:00:00;;0;1;64;0;7;,,,,,,,,,,,,;	16	00	09	54	14
Solo I-394 at Penn Ave.;103;20000531;	16	00	10	51	16
	16	00	12	52	16
	16	00	13	50	15
	16	00	14	49	12
	16	00	15	56	16
	16	00	16	54	16
	16	00	18	58	15
	16	00	19	60	16
	16	00	22	60	17

Figure 16.2: Raw MVS Data

Figure 16.3: Processed Left Lane Data

After the program has output data from each lane to corresponding text files, it invokes the function containing the headway algorithm on the data in the three files. The headway algorithm detects undesirable traffic flow conditions and outputs the times at which these conditions are detected and the number of successive headway violations at these times to different files, with each file corresponding to headway algorithm detections for one lane.

In the same manner, the program calls upon the quality of flow algorithm and the traffic pressure algorithm on data corresponding to the three lanes. These two algorithms also detect undesirable traffic flow conditions and output the times at which these conditions are detected to text files relevant to each lane and to each algorithm. The files containing the quality of flow algorithm alarms receive data corresponding to detection times, reduction in value of quality of flow and value of quality of flow. The files containing the traffic pressure algorithm alarms receive data corresponding to detection times and values of traffic pressure.

The function implementing the integrated heuristic algorithm recalls the number of successive headway violations as determined by the headway algorithm, the values of quality of flow and the percentage reduction in its value as determined by the quality of flow algorithm and the values of traffic pressure as determined by the traffic pressure algorithm. Based on the logic described in the previous chapter, the program raises alarms corresponding to the detection of undesirable traffic flow conditions and outputs them to relevant text files pertaining to each lane. The columns in these text files contain the times at which undesirable conditions are detected and the scores that the algorithm assigns to the undesirability of the condition.

In summary, the program receives as input:

- On-off times
- Individual vehicle Speeds
- Individual vehicle lengths

And outputs:

- The times at which undesirable flow conditions are detected by the headway algorithm and the number of successive headway violations at these times
- The times at which the quality of flow algorithm detects undesirable flow conditions and the value of quality of flow and the percentage reduction of quality of flow at these times
- The times at which undesirable flow conditions are detected by the traffic pressure algorithm and the value of traffic pressures at these times
- The times at which the heuristic integrated algorithm detects undesirable flow conditions and the scores that the algorithm assigns to the undesirability of the condition at these times.

Assumptions had to be made in the program in some rare instances when the Machine Vision Sensor failed to report correct values for speeds or lengths of the vehicles. The assumptions made are listed below:

1. Undetected speeds: There were some instances where the Machine Vision Sensor failed to detect the speed of a vehicle. In these cases, the speed of the vehicle was reported by the MVS as 255 miles per hour. To fix this problem, it was assumed that the speed of the vehicle was the same as the speed of the previous vehicle.
2. Undetected lengths: In instances, where the Machine Vision Sensor reported the length of a vehicle as 0, the length of that vehicle was assumed to be 18 feet (a value obtained by averaging the lengths of vehicles over a period of one hour).
3. 0-second time headways: The least count of time stamps as reported by the Machine Vision Sensors in the data was 1 second. Hence, when vehicles passed so close to each other that their headway was less than 1 second, the program calculated the time-headway as 0 second. In these cases, the time headway between vehicles was assumed to be 0.5 seconds.
4. Values of Change of Speed while calculating quality of flow and traffic pressure: If the absolute difference of speeds between present vehicle and previous vehicle was less than 5% of the speed of the present vehicle, the speed change was considered to be 0. This is because the speed outputs by the Machine Vision Sensor have an accuracy of up to 95%.

SQL queries were also developed using Microsoft Access so that the tedious process of counting the number of correct, missed and false alarms would be accelerated. The queries, when presented with the list of manual alarms and the list of algorithm alarms, determine which program alarms are correct alarms, which manual alarms have been missed and which program alarms are false alarms.

In order to make input data compatible for the neural network tool box in Matlab, some minor programs were developed using C++. These programs essentially obtain data from all the stand-alone algorithm alarm text files and then format them so that they maybe used as input in Matlab.

17 UTC DETECTION IMPLEMENTATION RESULTS

This chapter presents and discusses the test results that were obtained by implementing the algorithms developed in chapter 5 on the second data set of table 17.3. The results that were obtained by testing the individual algorithms are first discussed. The results produced by the integrated algorithms are then compared and discussed.

The algorithms that were developed in chapter 4 identified optimum thresholds that produced the best combinations of detection rate and false alarm rate (these thresholds correspond to number of headway violations, % change of quality of flow index, value of traffic pressure). Additional testing of the algorithms was performed to check whether the thresholds identified in chapter 4 were satisfactory or required recalibration..

17.1 Headway Algorithm Test Results

The headway algorithm developed in chapter 4 detected the number of successive violations of the minimum safe headway and raised alarms whenever this number exceeded a certain preset threshold. As chapter 4 suggests, a group-size of 5 vehicles (4 successive violations of the safe headway) produced the best combination of detection rate and false alarm rate for the right and middle lanes and a group-size of 6 vehicles (5 successive violations of the safe headway) produced the optimum combination of detection rate and false alarm rate for the left lane. The Headway Algorithm was implemented on the second data set (separately for each lane) for a number of values of the threshold of Group-Size, 2, 3, 4,, 15. Alarms were raised when each of these thresholds were violated (i.e. when the Group-Size was greater than or equal to the group-size threshold). These Algorithm alarms were then compared with the manual alarms to determine the number of correct, missed and false alarms. The detection rate and false alarm rate were then calculated and the performance curves (Detection Rate v/s False Alarm Rate Plots) were drawn.

Table 17.1 shows the average detection rate and false alarm rate per lane for all time periods that were obtained with the Headway Algorithm.

Table 17.1: Headway Algorithm Results Obtained with Second data set

Headway Algorithm Results-Second data set					
Right Lane Results	Manual Alarms = 4.41 / hour				
Group-Size Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
11	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00
8	0.59	0.56	0.02	12.74	4.00
7	1.12	1.07	0.04	24.33	4.00
6	2.03	1.86	0.17	42.07	8.52
5	3.19	2.29	0.91	51.85	28.37
4	7.58	2.85	4.73	64.64	62.41
3	16.47	3.91	12.56	88.65	76.26
2	36.22	4.41	31.81	99.92	87.83
Middle Lane Results	Manual Alarms = 5.37 / hour				
Group-Size Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
8	0.34	0.34	0.00	6.24	0.00
7	0.34	0.34	0.00	6.24	0.00
6	1.11	1.01	0.10	18.75	9.36
5	2.89	2.49	0.40	46.33	13.92
4	7.27	2.96	4.31	55.21	59.24
3	23.35	4.41	18.95	82.06	81.13
2	33.61	5.19	28.42	96.64	84.56
Left Lane Results	Manual Alarms = 5.93 / hour				
Group-Size Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
9	0.00	0.00	0.00	0.00	0.00
8	0.32	0.32	0.00	5.33	0.00
7	1.14	1.08	0.06	18.27	5.20
6	4.67	3.24	1.43	54.68	30.56
5	7.11	3.68	3.43	62.03	48.24
4	15.43	4.81	10.62	81.14	68.82
3	20.94	5.20	15.74	87.65	75.17
2	49.36	5.81	43.55	97.98	88.23

Figure 17.1, derived from table 17.1, depicts the detection rate v/s false alarm rate curves (performance curves) for each lane. From table 17.1 and figure 17.1, it can be seen that the optimum thresholds obtained for the second data set were the same as those obtained for the first data set i.e. 5, 5 and 6 vehicles for the right, middle and left lanes respectively. As can be observed in table 17.1, the optimum threshold of 5 vehicles produces a detection rate and false alarm rate of 51.85% and 28.37% respectively for the right lane. For the middle lane, the optimum threshold of 5 vehicles produces a detection rate and false alarm rate of 46.33% and

54.68% respectively. A threshold of 6 vehicles produces the best combination of detection rate and false alarm rate of 54.68% and 30.56% for the left lane. However, these results were not considered satisfactory as they suggest relatively low detection rates and high false alarm rates. The results suggested that the headway algorithm was promising for further consideration in a more complex algorithm that was subsequently developed. Before concluding the findings of the headway algorithm, it is worth noting that, when excluding rainy and snowy days from the above results and analyzing them separately, the same optimum thresholds were found, albeit the algorithm performance was slightly worse; an average detection rate of 48.6% and average false alarm rate of 31.42% was observed for the snowy day and an average detection rate of 50.14% and an average false alarm rate of 26.36% was obtained for the rainy day.

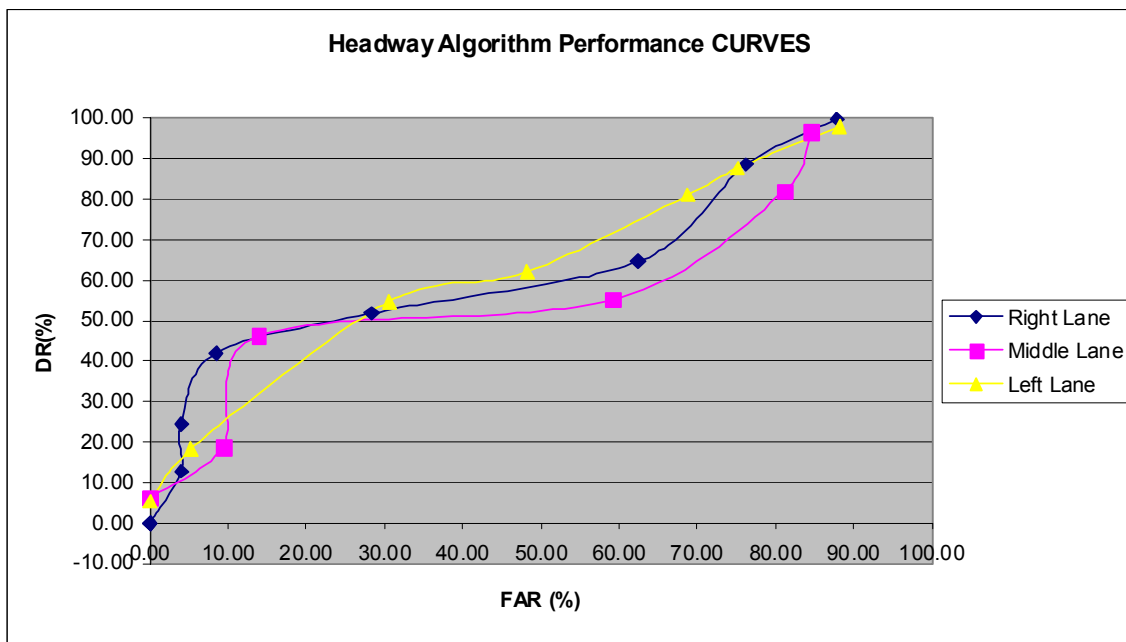


Figure 17.1: Performance Curves Obtained for Headway Alg with the Second data set

This similarity of results shows that the Headway Algorithm can be applied successfully on different data sets since uniformity of performance can be observed between all data. This implies that the optimum group-size threshold does not have to be recalibrated for the same site.

17.2 Traffic Pressure Test Results

As mentioned in chapter 5, the traffic pressure algorithm was designed to raise alarms when high values of traffic pressure are detected. On implementing the traffic pressure algorithm on the first

data set, it was determined that the thresholds that produced the best combinations of detection rate and false alarm rate for the right, middle and left lanes were traffic pressure values of 3500, 2000 and 1500 vehicle-mile /hour² respectively. Following the same procedure, the Traffic Pressure Algorithm was implemented on the second data set (separately for each lane) for a number of values of the threshold of traffic pressure. Table 17.2 shows the average detection rate and false alarm per lane for all time periods that were obtained with the Traffic Pressure Algorithm. The optimum thresholds were again found to be traffic pressure values of 3500, 2000 and 1500 vehicle -mile/hour² for the right, middle and left lanes respectively. Even the values of the detection rate and false alarm rates obtained for each of the thresholds are almost identical to the values obtained with the first data set.

This similarity of results shows that the Traffic Pressure Algorithm can also be applied successfully on to different data sets since uniformity of performance can be observed between all data. As can be observed in table 17.2 and figure 17.2, the optimum traffic pressure value of 3500 vehicle-mile/hour² produces a detection rate and false alarm rate of 45.92% and 32.33% respectively for the right lane. For the middle lane, the optimum traffic pressure value of 2000 vehicl-milee/hour² produces a detection rate and false alarm rate of 43.14% and 36.04% respectively. A threshold of 1500 vehicle-mile/hour² produces the best combination of detection rate and false alarm rate of 39.11% and 26.28% for the left lane. The traffic pressure algorithm, though not very successful individually, was also promising for consideration in a combined algorithm.

Table 17.2: Traffic Pressure Algorithm Results Obtained with Second data set

Traffic Pressure Algorithm Results- Second data set					
Right Lane Results	Manual Alarms = 4.41 / hour				
Traffic Pressure Value Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
6000	0.00	0.00	0.00	0.00	0.00
5500	0.39	0.39	0.00	8.50	0.00
5000	0.96	0.85	0.11	18.65	11.92
4500	1.30	1.09	0.20	24.00	15.80
4000	2.43	1.79	0.65	39.24	26.58
3500	3.09	2.09	1.00	45.92	32.33
3000	4.27	2.21	2.06	48.62	48.19
2500	5.87	2.72	3.15	59.76	53.66
2000	7.27	2.95	4.31	64.90	59.36
1500	12.34	3.39	8.94	74.58	72.49
1000	18.18	3.78	14.40	83.07	79.21
500	25.63	4.19	21.44	92.15	83.64
Middle Lane Results	Manual Alarms = 5.37 / hour				
Traffic Pressure Value Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
6000	0.00	0.00	0.00	0.00	0.00
5500	0.16	0.16	0.00	3.02	0.00
5000	0.34	0.34	0.00	6.25	0.00
4500	0.52	0.52	0.00	9.67	0.00
4000	0.77	0.77	0.00	14.37	0.00
3500	1.31	1.17	0.14	21.81	10.81
3000	1.81	1.53	0.28	28.54	15.55
2500	2.38	1.75	0.63	32.62	26.48
2000	3.62	2.32	1.30	43.14	36.03
1500	6.57	3.05	3.52	56.85	53.55
1000	13.61	3.94	9.67	73.33	71.07
500	26.83	4.77	22.07	88.75	82.24
Left Lane Results	Manual Alarms = 5.93/ hour				
Traffic Pressure Value Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
6000	0.00	0.00	0.00	0.00	0.00
5500	0.19	0.19	0.00	3.12	0.00
5000	0.19	0.19	0.00	3.12	0.00
4500	0.35	0.35	0.00	5.87	0.00
4000	0.48	0.48	0.00	8.14	0.00
3500	0.80	0.73	0.07	12.38	8.33
3000	1.31	1.14	0.16	19.27	12.50
2500	1.81	1.47	0.34	24.83	18.75
2000	2.39	1.88	0.51	31.77	21.33
1500	3.15	2.32	0.83	39.11	26.28
1000	6.94	3.03	3.91	51.16	56.29
500	22.82	4.29	18.52	72.42	81.18

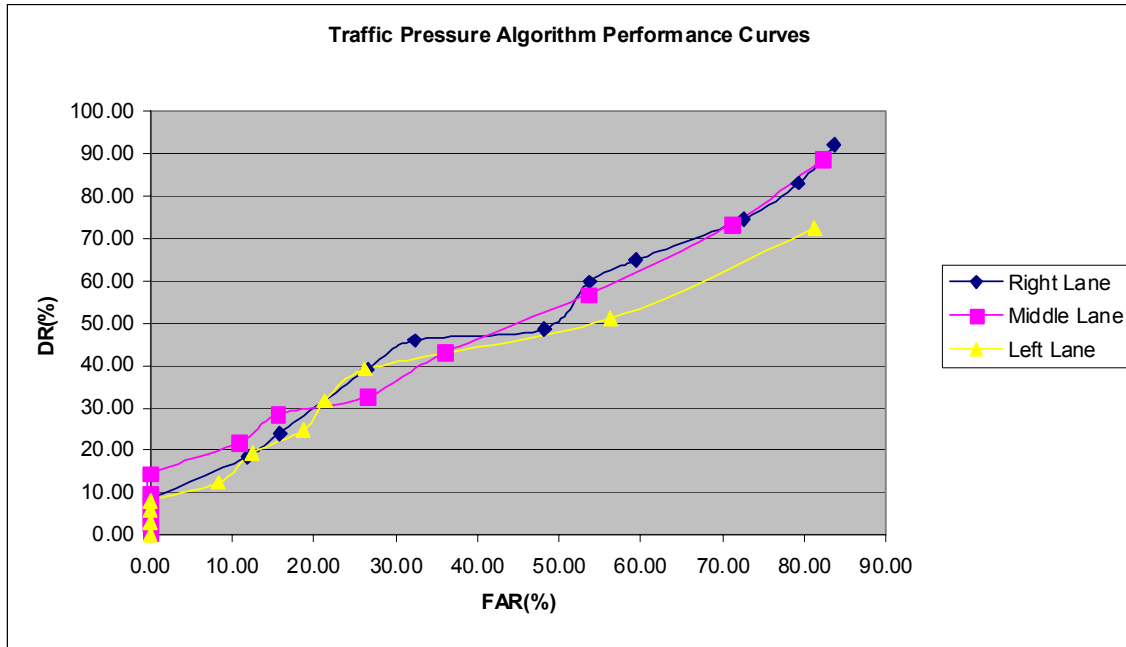


Figure 17.2: Perf Curves Obtained for Traffic Pressure Alg with the Second data set

17.3 Quality of Flow Index Algorithm Test Results

As described in chapter 4, the quality of flow index algorithm was developed to raise alarms whenever sharp decreases in the value of quality of flow are observed. If the value of the % change of quality of flow is lower than a preset threshold, then the algorithm raises an alarm. For the right and the middle lanes, the algorithm was modified to raise alarms when the threshold of % change of quality of flow was violated and if the quality of flow was lesser than 2500. On implementing the quality of flow algorithm on the first data set, it was determined that the thresholds that produced the best combinations of detection rate and false alarm rate for the right, middle and left lanes were when the quality of flow decreased by 20%. The Quality of Flow algorithm was implemented on the second data set (separately for each lane) by following the same procedure as was done with the first data set. Table 17.3 shows the average detection rate and false alarm per lane for all time periods that were obtained with the quality of flow index algorithm.

Table 17.3: Quality of Flow Index Algorithm Results Obtained with Second data set

Quality of Flow Index Algorithm Results-Second data set					
Right Lane Results	Manual Alarms = 4.41 / hour				
"% Decrease of QOF" Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
50%	0.77	0.71	0.06	16.18	7.85
45%	1.73	1.36	0.37	30.82	21.24
40%	1.73	1.36	0.37	30.82	21.24
35%	1.73	1.36	0.37	30.82	21.24
30%	1.73	1.36	0.37	30.82	21.24
25%	2.34	1.68	0.66	38.05	28.15
20%	2.89	1.88	1.02	42.52	35.16
15%	5.58	2.42	3.16	54.90	56.64
10%	11.01	3.06	7.94	69.45	72.17
5%	22.56	4.16	18.40	94.33	81.56
Middle Lane Results	Manual Alarms = 5.37 / hour				
"% Decrease of QOF" Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
50%	0.45	0.45	0.00	8.42	0.00
45%	0.99	0.93	0.06	17.27	6.24
40%	0.99	0.93	0.06	17.27	6.24
35%	2.27	1.91	0.36	35.60	15.94
30%	2.27	1.91	0.36	35.60	15.94
25%	3.32	2.50	0.82	46.58	24.59
20%	4.34	2.94	1.40	54.78	32.18
15%	8.04	3.29	4.75	61.33	59.06
10%	13.29	4.21	9.08	78.31	68.35
5%	18.65	4.63	14.02	86.27	75.16
Left Lane Results	Manual Alarms = 5.93 / hour				
"% Decrease of QOF" Threshold	Program Alarms / hr	Correct Alarms / hr	False Alarms / hr	Detection Rate	False Alarm Rate
50%	0.48	0.48	0.00	8.15	0.00
45%	0.48	0.48	0.00	8.15	0.00
40%	1.04	0.96	0.08	16.20	7.45
35%	1.04	0.96	0.08	16.20	7.45
30%	2.26	1.87	0.40	31.48	17.56
25%	2.83	2.15	0.68	36.27	24.00
20%	4.35	2.93	1.43	49.33	32.82
15%	10.49	3.98	6.51	67.18	62.04
10%	19.53	5.00	14.53	84.26	74.41
5%	30.72	5.49	25.23	92.58	82.13

Figure 17.3, derived from table 17.3, depicts the performance curves for each lane. From table 17.3 and figure 17.3, it can be seen that the optimum thresholds were again found to be a decrease of 20% for the right, middle and left lanes respectively with almost identical values of detection rate and false alarm rate as were observed with the first data set. As can be observed in table 17.3, the optimum threshold value of quality of flow decrease (20%) produces a detection rate and false alarm rate of 42.52% and 35.16% respectively for the right lane. For the middle lane, it produces a detection rate and false alarm rate of 54.78% and 32.18% respectively. The 20% decrease threshold produces the best combination of detection rate and false alarm rate of 49.33% and 32.82% for the left lane.

This consistency of results obtained shows that the Quality of Flow Algorithm can also be transferred successfully on to different data sets since uniformity of performance can be observed between all data.

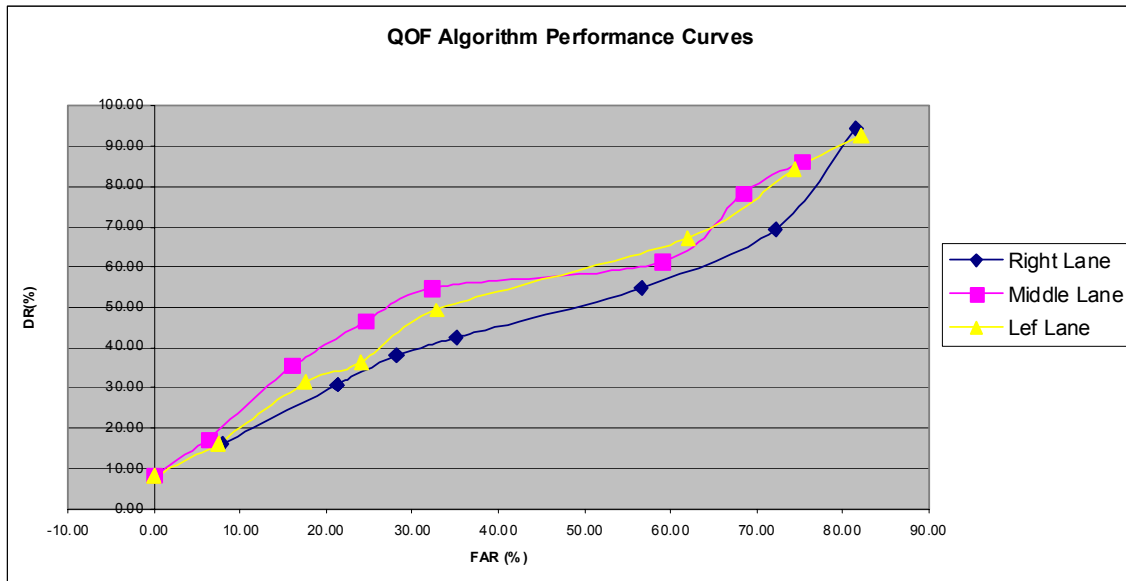


Figure 17.3: Performance Curves Obtained for Quality of Flow Alg with the Second data set

17.4 Comparison of Integrated Algorithm Test Results

As mentioned in chapter 4, the three individual algorithms were integrated into one algorithm or methodology using two approaches to obtain detection rates. Initially, a heuristic technique for amalgamating the three stand-alone algorithms was developed. However, since this did not increase the detection rates by much, a more scientific approach was sought and therefore the Neural Network Integrated Algorithm was developed.

For each vehicle detection, the individual algorithms calculate the number of successive headway violations, the % decrease of quality of flow and the value of traffic pressure and raise alarms when the optimum thresholds are violated. The heuristic algorithm developed produces a weighted score based on the individual algorithm calculations and compares the sum with a preset threshold, which if exceeded, produces an alarm. On implementing the heuristic integrated algorithm on the first data set, it was determined that the combined-score thresholds that produced the best combinations of detection rate and false alarm rate for the right, middle and left lanes were scores of 4, 6 and 4 respectively. When the Heuristic Integrated algorithm was implemented on the second data set, following the same procedure, the optimum thresholds were again found to be a score of 4 for the right lane, a score of 6 for the middle lane and a score of 4 for the left lane respectively. Even the values of the detection rate and false alarm rates obtained for each of the thresholds are almost identical to the values obtained with the first data set. This similarity of results shows that the Heuristic Integrated Algorithm can also be applied successfully on to different data sets since consistent performance can be observed between all data. Table 17.4 shows the best performance of the Heuristic Integrated Algorithm when implemented on the second data set. Though, the results suggest considerable improvement in performance when compared to the individual algorithms in terms of detection rate, the heuristic integrated algorithm did not succeed in reducing the false alarm rates by much. Hence, a scientific method to combine the three individual algorithms was sought and this was the main reason for employing the neural networks approach.

Neural networks were designed, trained and tested on the data corresponding to all three lanes (one for each lane: NNRL, NNML, NNLL). 80% of the first data set was used for training the neural networks and 20% of the data was used for testing or simulation. This was done for data pertaining to all three lanes. Successive headway violations, the actual value and the percentage reduction of the quality of index and the traffic pressure value were used as variables in the input vector for the right and the middle lane. For the left lane, all the above variables except the actual value of the quality of flow index were submitted as the input vectors. Table 17.4 shows the results obtained by implementing the relevant neural networks on data corresponding to relevant lanes and compares these results with those obtained by the Heuristic Integrated Algorithm.

Heuristic Integrated Algorithm			Neural Network Algorithm		
Threshold Score (lane)	Detection Rate	False Alarm Rate	Neural Network (lane)	Detection Rate	False Alarm Rate
4 (right)	67.33%	38.12%	NNRL(right)	79.68%	23.54%
6 (middle)	68.92%	38.46%	NNML(middle)	84.08%	32.17%
4 (left)	60.28%	40.42%	NNLL(left)	78.51%	28.63%

Table 17.4: Comparison of Integrated Algorithm Results (second data set)

As can be seen from the table, the neural network algorithm outperforms the heuristic integrated algorithm in terms of both detection rate and false alarm rate. The neural network algorithm increases the detection rate and reduces the false alarm rate considerably. Even though these results are promising and much better than all the other algorithms, they still suggest high false alarm rate values. This implies that additional research is needed to improve the detection algorithms or even design new ones by analyzing and collecting more data from additional sites which may have more than one detection zone in each lane within the field of view of the camera.

18 INCIDENT PRONE CONDITIONS:CONCLUSIONS

The objective of this task was to explore the feasibility of identifying traffic conditions that are likely to cause accidents and to develop a methodology to detect these conditions. Since no accidents occurred at the test site during the periods of observation, the scope was modified to detect undesirable traffic flow conditions that could cause accidents. After an extensive literature review and discussion with TMC operators, a list of visual undesirable flow conditions was determined. A library of video recordings and raw traffic data was created containing events matching the ones in the previously created list of undesirable conditions. This information was then used to develop algorithms for detecting undesirable flow conditions. This development can be viewed as the first step towards the identification and detection of accident-prone conditions which is the subject of newly funded research.

The data collected was split into two sets – the first data set was used to develop and calibrate the algorithms and the second data set was used to test them. From all the measurements considered, minimum space headway estimated from time headway and speeds, quality of flow index and traffic pressure were found to significantly contribute in detecting undesirable flow conditions. Individual algorithms were developed to determine optimum threshold values for these measurements, which if violated causes the algorithms to raise alarms. Algorithms were then developed combining the above-mentioned traffic measurements. These algorithms were subsequently tested on the second data set to determine whether they produced consistent results.

Analysis of the results obtained by implementing the algorithms on the second data set shows that the algorithms can be implemented on different data sets without any recalibration. The results also show that combining the individual measurements in a single algorithm produce better results i.e. better detection rates with only marginal increase in false alarm rates. Based on this, a neural network algorithm was developed to combine the individual measurements. The performance of the neural network was found to be much better than that of the individual algorithms and the heuristic integrated algorithm.

The neural network produced 80.75% and a false alarm rate of 28.11%. This suggests that the algorithms developed are successful, albeit, they should be further improved for practical implementation with regards to detecting the undesirable traffic flow conditions specified.

From the results achieved, it is hoped that actual accident-prone conditions can also be detected by extending the methodologies developed. Detection of actual accident-prone conditions is in fact being currently pursued in another project in which several Machine Vision Sensors are being installed in high accident risk locations. The cameras are being placed higher so that wide area detection of traffic can be possible. This will ensure improved estimation of measurements such as quality of flow. The methodology developed in the current project will be a starting point for developing algorithms to detect actual accident-prone conditions.

Successful detection of accident-prone conditions can enable a traffic operator to issue warnings and take pro-active measures to avert potentially dangerous driving conditions before they become severe enough to cause an accident. There are various ideas on how to transmit these warnings to the drivers and it is as much as a human factors problem as it is a traffic engineering one. Variable Message signs although can be easily used are not present in locations that allow such an application. Technologies currently available in Europe like the RDMS radio announcement system and the BMW incident warning devices can be used as a start.

Because the occurrence of accidents can be reduced or prevented, the non-recurrent congestion that accompanies accidents should also be averted. Hence, detection of accident prone conditions can not only help in preventing the occurrence of accidents, it can also help in easing congestion on freeways. As an example of the potential benefits of this, consider the statistics according to which the property-damage only accidents cost a total of approximately \$127, 500, 000. Although it is hard to predict the number of accidents that can actually be prevented by practical implementation of the methodologies presented here, even if a percentage of accidents can be prevented, the benefits would be significant. For instance, assuming that 50% of the property-damage only accidents can be prevented in Minnesota, this translates to \$63, 000, 000 savings per year to the users.

Reference:

1. Lindley, Jeffrey A., "Urban Freeway Congestion: Quantification of the Problem and Effectiveness of Potential Solutions," *Journal of the Institute of Transportation Engineers*, 57, no. 1 (January), pp 27-32. 1987
2. Stavropoulos, Antonios, "Development of Incident Prediction Algorithms for Freeway Traffic Management", University of Texas, Austin, 1997
3. Roess, R.P., McShane, W. and Prassas, E., "Traffic Engineering", Prentice-Hall, 1999
4. Drew, Donald. R, "Traffic Flow Control and Theory", McGraw Hill, New York, 1968
5. "Highway Design and Operations Standards Affected by Driver Characteristics", Report No. FHWA-RD-83-015, Bellorro-McGee Inc., Vienna, Virginia, 1983
6. Greenshields, B.D, "Traffic Accidents and the Quality of Traffic Flow", *Traffic Accident Studies*, 1958, Highway Research Board, Bulletin 208, pp 1-15, Washington DC, 1958
7. Simpson, P, "Artificial Neural Systems: Foundations, Paradigms, Applications, and Implementations", Pergamn Press, New York, 1990
8. May, D., "Traffic Flow Fundamentals", Prentice Hall, 1990

Web References

1. <http://www.lamaroutdoor.com/main/whylamar/industryTrends.cfm>
2. <http://www.highways.org/> (American Highway Users Alliance)
3. http://www.co.fulton.ga.us/publicworks/its_atms.htm
4. <http://www-fars.nhtsa.dot.gov/>
5. <http://www.nhtsa.dot.gov/people/ncsa/StateData/app-b.htm>
6. www.driversalert.com
7. <http://www.volpe.dot.gov>
8. <http://www.fhwa.dot.gov/tfhrc/safety/pubs/00022/intro.htm>
9. <http://www.claimsmag.com/Issues/archives/perception.asp>
10. <http://www.usroads.com/journals/rej/9707/re970701.htm>
11. <http://sunsite.univie.ac.at>

19 ASSISTANCE TO Mn/DOT IN EXPANDING SIMULATION USE

The final task of this project aims in assisting Mn/DOT in incorporating the tools, methods and findings described in this report into its regular operations. Towards that end a training course was organized and offered to all the interested Mn/DOT engineers as well as other people involved with Mn/DOT traffic operations from FHWA and the university.

The course offered lasted for three working days and covered almost all subjects related to microscopic traffic simulation. The agenda of the course was as follows:

Day 1 April 22, 2002 at the CTS Lab

- 8:00 Arrivals, Welcome and Introductions
 - Expectations overview
 - Introduction to Simulation and Traffic Simulation
- 9:30 Coffee break
- 9:45 Q & A
- 10:00 Introduction to Tedi Editor and GETRAM modelling components
 - Networks objects: sections, junctions, etc
 - Traffic devices: detectors, traffic lights, ramp metering, VMS, etc
 - Vehicles: types and classes, attributes
- 11:00 Hands-on exercises
 - Editing a network with Tedi
- 12:30 Lunch
- 13:00 GETRAM modelling components (continued)
 - Traffic demand data: input flows and turning proportions or O/D matrices.
 - U of M state generation application based on Loop Detector data.
 - Traffic control: signal groups, phases, control plans, ramp metering.
 - Public Transport Plan (optional)
- 14:15 Coffee break
- 14:30 Hands-on exercises
 - Editing traffic demand and control plan with Tedi
- 16:30 End of session

Day 2 April 23, 2002 at the CTS Lab

- 8:00 Expectations overview
- 8:15 AIMSUN2 Modelling
 - Network modelling
 - Traffic modelling
 - Vehicle generation models
 - Vehicle behaviour models: Car following, lane changing, etc
- 9:30 Coffee break
- 9:45 Q & A
- 10:00 AIMSUN2 Modelling (continued)
 - Modelling Control and Management
 - Route based simulation (optional, time permitted)
- 11:00 Hands-on exercises
 - Preparing AIMSUN simulation experiments
 - Running the simulation
- 12:30 Lunch
- 13:00 Simulation Outputs
 - AIMSUN Graphical User Interface
- 14:15 Coffee break
- 14:30 Hands-on exercises
 - Use of AIMSUN GUI
- 16:00 Discussion
- 16:30 End of Session

Day 3 April 24, 2002 at the CTS Lab

- 8:00 Calibration and validation of AIMSUN2 models
 - Data Collection
 - Design of simulation experiments
 - Calibration methodology

- Influence of Modelling Parameters
- 9:30 Coffee break
- 9:45 Q & A
- 10:00 Calibration and validation of AIMSUN2 models
Use of U of M Calibration & Validation tools.
- 11:00 Hands-on exercises
Check the influence of modelling parameters
- 12:00 Lunch
- 13:00 Results Analysis
Detailed Statistical Output and Data Base Definition
Use of ODBC
- 14:15 Coffee break
- 14:30 Advanced topics on AIMSUN Modelling
Vehicle behaviour models
Variable Message Signs
GETRAM Extensions
- 15:30 Discussion
- 16:30 End of Course

The hands-on exercises used in the training were based on an active project by Mn/DOT involving the I-494 freeway in Minneapolis. At the end of the course a CD containing all material used in the training as well as a large number of documents, manuals, and tools was distributed on all participants and accompanies this report.

An evaluation form was also distributed on all participants. Based on returned the course was rated as follows:

Rated 1-5 with 1 = Strongly disagree and 5= Strongly agree

The Instructor was knowledgeable in the subject.	5
The Instructor presented the content effectively.	4.75
The facility/room was conducive to learning.	4
The participant materials enhanced the course content.	4.5
The visual aids (video, projected presentations, overhead transparencies, etc.) contributed to my learning.	4.5
I effectively participated in the course.	4.25
Received adequate reference material.	4.75
The course content will make me more effective in my work.	4.25
The class met the objectives described in the announcement.	4.5

Questions

1. Include any explanations for the above ratings, especially those marked less than “neutral” or “3.”
 - “Having hands on participation is important. Unfortunately I couldn’t stay for the whole thing.”
 - “I have been to a couple other computer program classes and this was the first that provided reference material that are actually useful outside of the classroom. John has a good knowledge of the workings of the AIMSUN package and was able to answer any questions the class had.”
 - “I thought it was a very good class and I look forward to using AIMSUN.”
2. How could this learning experience be improved?
 - “The room was a little crowded but I think we all got enough time to get some hands on experience.”
 - “A smaller roadway network that could be built and populated might be better suited for the classroom exercise. This way the students could get some extra time entering the volumes and examining the output files.”

- “Have a follow-up class (advanced class) after having used AIMSUN on a project.”
 - “Increase the lab time to allow more hands-on time with the software. You may want to consider adding a lab to cover the U of M tools developed to process the TMC data and to format the output.”
3. What were the most valuable aspects of the course?
- “The materials and the PowerPoint Presentations were very useful in teaching me how to use the program. Getting hands on experience is also very valuable.”
 - “Learning about the new developments within the simulation package. It was good exposure to another microscopic modeling package for those new to the modeling experience.”
 - “The use of I-494 as a project area – I am familiar with the area and had modeled with a different program.”
 - “The model calibration process was the most valuable aspect of the course. John did an excellent job of walking us through the process, explaining the statistical tools used and passing on some of the lessons they had learned.”

Specifically the project Technical Liaison made the following comments:

- Course was well organized and structured.
- Course description and outline accurately reflected the materials covered.
- Course objectives were met.
- Course documentation (Getting Started, TEDI and AIMSUN Users' manual) are excellent resources.
- Course handouts well organized and address the topics very well. John did an excellent job explaining how the model worked, using TEDI to build the network geometries, pointing out common mistakes, and populating the model. He also walked us through the calibration and validation process and shared the lessons he had learned. The data processing tool developed by

the U of M streamlines the data entry process. The post processing tools help siphon through the mounds of output data and restructure it into a useable format.

- Basing the labs exercises on a real world example definitely allowed the attendees to apply what they had learned.

Overall, John did an excellent job on the training.