Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/TX-06/0-4745-3		

4. Title and Subtitle	5. Report Date
INCIDENT DETECTION OPTIMIZATION AND DATA QUALITY CONTROL	October 2005
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Cesar Quiroga, Khaled Hamad, and Eun Sug Park	Report 0-4745-3
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Texas Transportation Institute The Texas A&M University System College Station, Texas 77843-3135	11. Contract or Grant No. Project 0-4745
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Texas Department of Transportation Research and Technology Implementation Office P. O. Box 5080 Austin, Texas 78763-5080	Technical Report: September 2004 – August 2005
	14. Sponsoring Agency Code

15. Supplementary Notes

Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.
Project Title: Using Archived ITS Data and Spatial Statistics for Optimizing Incident Response at Transportation Management Centers
URL: http://tti.tamu.edu/documents/0-4745-3.pdf

16. Abstract

Project 0-4745 developed a geographic information system (GIS)-based approach for the determination of patterns in the spatial and temporal distribution of incidents along freeway corridors. The research documented incident detection and data archival at several transportation management centers (TMCs) in Texas, a process to develop a data model and geodatabase of intelligent transportation system (ITS) equipment and archived ITS data, and a process to determine patterns in the spatial and temporal distribution of freeway incidents.

This report describes the procedures and activities completed during the second phase of the project. The analysis included an evaluation of incident detection procedures at a sample TMC (TransGuide) and an assessment of the feasibility to modify/calibrate alarm threshold values to help optimize incident detection practices at that TMC. The research involved the use of two performance measures (detection rate and false alarm rate) and the development of a prototype offline tool to evaluate automatic incident detection algorithm performance. The analysis also included an evaluation of archived loop detector data completeness and quality control.

17. Key Words	18. Distribution Statement
Intelligent Transportation Systems, Incidents, GIS, Transportation Management Centers, Archived ITS Data, Automatic Incident Detection Algorithms	No restrictions. This document is available to the public through NTIS: National Technical Information Service Springfield, Virginia 22161 http://www.ntis.gov

19. Security Classif.(of this report)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	84	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

INCIDENT DETECTION OPTIMIZATION AND DATA QUALITY CONTROL

by

Cesar Quiroga, P.E.
Associate Research Engineer
Texas Transportation Institute

Khaled Hamad
Assistant Research Scientist
Texas Transportation Institute

and

Eun Sug Park
Assistant Research Scientist
Texas Transportation Institute

Report 0-4745-3
Project 0-4745
Project Title: Using Archived ITS Data and Spatial Statistics for Optimizing Incident Response at Transportation Management Centers

Performed in cooperation with the
Texas Department of Transportation
and the
Federal Highway Administration

October 2005

TEXAS TRANSPORTATION INSTITUTE
The Texas A&M University System
College Station, Texas 77843-3135

DISCLAIMER

The contents of this document reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Federal Highway Administration (FHWA) or the Texas Department of Transportation (TxDOT). This document does not constitute a standard, specification, or regulation, nor is it intended for construction, bidding, or permit purposes. The engineer in charge of the project was Cesar Quiroga, P.E. (Texas Registration #84274).

The United States Government and the State of Texas do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

ACKNOWLEDGMENTS

This project was conducted in cooperation with TxDOT and FHWA. The researchers would like to gratefully acknowledge the assistance provided by TxDOT officials, in particular the following:

- Bill Jurczyn – San Antonio District (project director),
- Brian Burk – Austin District,
- Steve Connell – Fort Worth District,
- Guillermo Dougherty – Laredo District,
- Brian Fariello – San Antonio District,
- David Fink – Houston District,
- Ron Holtz – San Antonio District,
- Tai Nguyen – Fort Worth District, and
- David Rodrigues – San Antonio District.

The researchers would also like to acknowledge Robert Pina, who wrote the incident detection tester application, and the students, in particular James Williams, who produced the traffic figures for all the cases studied, and Mia-Andrea Veliz, who assisted in writing code for the incident detection tester application.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF ACRONYMS, ABBREVIATIONS, AND TERMS

ADM	Administrative
AIH	Alarm incident handler
ArcIMS	Arc Internet Map Server
ARIMA	Autoregressive integrated moving averages
ATIS	Advanced traveler information system
ATMS	Advanced traffic management system
AWARD	Advanced warning to avoid railroad delays
CCTV	Closed-circuit television
CMS	Changeable message sign
CDT	Central daylight time
CST	Central standard time
DDD	Dynamic data distribution
DMS	Dynamic message sign
DR	Detection rate
EAR	Effective alarm rate
ETT	Estimated travel time
FAF	False alarm frequency
FAR	False alarm rate
FHWA	Federal Highway Administration
GIS	Geographic information systems
GUI	Graphical user interface
IDAT	Incident Detection Algorithm Tester

ITS	Intelligent transportation systems
LCS	Lane control signal
LCU	Local control unit
NTCIP	National Transportation Communications for ITS Protocol
PLAN	Personalized assistance and notification
SCM	Scenario management
TMC	Transportation management center
TRF	Traffic Operations Division
TxDOT	Texas Department of Transportation
TxDPS	Texas Department of Public Safety

CHAPTER 1. INTRODUCTION

Report 0-4745-1 "Incident Characteristics and Impact on Freeway Traffic" summarized the activities conducted during the first phase of research project 0-4745 (*1*). It described a process to determine patterns in the spatial and temporal distribution of incidents along freeway corridors using geographic information system (GIS), traffic engineering, and statistical analysis techniques. It also illustrated incident detection and data archival practices at several Texas Department of Transportation (TxDOT) transportation management centers (TMCs), a process to develop a geodatabase of intelligent transportation system (ITS) equipment and archived ITS data using a variety of data sources at the San Antonio TMC (TransGuide), a process to determine patterns in the spatial and temporal distribution of freeway incidents in San Antonio, and a process to calculate the impact of incidents on traffic delay. The report contained products 0-4745-P1 (which described incident evaluation procedures) and 0-4745-P2 (which described steps for incident evaluation procedure implementation).

This report summarizes the procedures and activities completed during the second phase of the research. Those activities resulted in two products (0-4745-P3—detailed incident evaluation procedures—and 0-4745-P4—process definitions and implementation recommendations), which are included in Report 0-4745-2 (*2*). During the second phase, the research team extended the first phase analysis to evaluate in greater detail incident detection procedures at a sample TMC (TransGuide) and assess the feasibility to modify/calibrate alarm threshold values to help optimize incident detection practices at that TMC. The research involved the use of two performance measures (detection rate and false alarm rate) and the development of a prototype offline tool to evaluate automatic incident detection algorithm performance. It also extended an analysis from the first phase to evaluate the completeness and quality control of archived loop detector data.

This report is organized as follows:

- Chapter 1 is this introductory chapter.
- Chapter 2 describes the incident detection process at TransGuide, with a focus on the automatic incident detection algorithm, and the process to extract meaningful incident data from archived ITS data sources.
- Chapter 3 summarizes the analysis conducted to evaluate the feasibility to modify alarm threshold values at TransGuide.
- Chapter 4 describes the analysis to evaluate the completeness and quality control of archived loop detector data.
- Chapter 5 summarizes conclusions and recommendations for implementation.

CHAPTER 2. INCIDENT DETECTION AT TRANSGUIDE

This chapter describes the incident detection process at TransGuide, with a focus on the automatic incident detection algorithm, and the process to extract meaningful incident data from archived ITS data sources. It starts with a brief introduction to automatic incident detection concepts, followed by a summarized description of the incident detection process at TransGuide, the process to extract incident data from archived ITS data sources, and an overall assessment of effectiveness of the automatic incident detection process at TransGuide.

INCIDENT DETECTION ALGORITHMS

TMCs use a variety of techniques to detect roadway incidents. Examples include detector-based alarms, 911-based alarms, closed-circuit television (CCTV) camera scanning, police radio scanning, courtesy patrols, motorist assistant dispatch, and commercial traffic services. TMCs are increasingly relying on non-detector-based procedures to detect incidents, which raises questions about the feasibility to continue making considerable investments on road-based detectors and associated hardware and software infrastructure. Nevertheless, in jurisdictions where road detectors are already in place, detector-based incident detection remains an important incident management tool.

Detector-based incident detection algorithms typically follow one of the following approaches:

- **Comparative Approach**. Algorithms that follow this approach compare measured traffic conditions against predetermined thresholds and trigger an alarm if the field measures cross the thresholds. Examples of this type of algorithm are the California algorithm series, which use absolute and relative differences in occupancy values (*3*) and the Texas algorithm, which uses moving average occupancy values (*4*). The TransGuide algorithm falls within this category, except that it uses speed data from speed-trap detectors and percent occupancy data from non-speed-trap loop detectors. Comparative algorithms are simpler than other algorithms. Many implementations rely on static thresholds, making them relatively inefficient for handling fluctuating traffic demands (*5*). Some implementations enable managers to vary thresholds using pre-specified criteria, e.g., by time of day, but populating threshold lookup tables frequently remains an incomplete task.
- **Statistical Approach**. Algorithms that follow this approach use statistical procedures to detect significant deviations in traffic patterns over time as compared to predictable patterns. Examples of this type of model include the standard normal deviate model (*6*), which uses the mean and standard deviation of occupancy values, time-series models (*7*), which use autoregressive integrated moving average (ARIMA) predictions of occupancy values, and the Minnesota algorithm (*8*), which uses a low-pass filter to remove high-frequency components in observed data. Statistical models require data to follow pre-specified statistical theory models, thus limiting their wide applicability.
- **Traffic Modeling Approach**. Algorithms that follow this approach use complex traffic-flow theoretical models to predict deviations from normal conditions using current traffic measurements as well as historical trends. An example of this type of algorithm is the

McMaster algorithm, which relies on the volume-occupancy relationship to determine when conditions change at individual detection stations (*9*).

- **Artificial Intelligence Approach**. Algorithms that follow this approach use artificial intelligence techniques such as neural networks (*10*) and fuzzy set theory (*11*). Although these techniques do not pre-assume theoretical traffic models, they nonetheless require extensive calibration. They are also among the most recent examples of algorithm development work and for the most part remain untested under real-world operating conditions.

TxDOT has funded a number of research studies related to the implementation and effectiveness of incident detection algorithms in Texas. For example, in 1993 Project 0-1232 evaluated several algorithms based on performance measures reported in the literature and site visits to operating TMCs in the United States and Canada (*9*). The project recommended implementation of the California 7, California 8, and McMaster algorithms. In 1996, Project 0-1795 tested the California 8, Minnesota, and Texas algorithms using TransGuide data, and developed a holistic data fusion model that combined detector data and indicators from different algorithms (*5*). Project 0-2917 compared the performance of the TransGuide speed-based algorithm against the California 8 algorithm and a fuzzy logic algorithm (*12*). Using data from seven incidents in 1996, the study concluded that the TransGuide's algorithm performed well compared to the other algorithms. More recently, Project 0-4156 and Project 0-4957 explored the integration of loop detector data and automated vehicle identification (AVI) data for incident detection at TransGuide (*4*, *13*). These projects concluded AVI-based algorithms were feasible, but did not perform as well as loop detector-based algorithms. As a side note, it may be worth noting that in 2003 TransGuide abandoned the AVI data collection program.

TRANSGUIDE INCIDENT DETECTION PROCESS

This section describes the general system configuration and incident detection and response process at TransGuide. For convenience, it summarizes relevant material from Chapter 2 of Report 0-4745-1, but expands on the description of the automated incident detection algorithm and alarm incident handling process (*1*).

Configuration

TransGuide's ITS deployment covers some 87 miles of freeway. It includes 1,463 loop detector units (both speed-trap and non-speed-trap) and sonic detectors organized in 325 sensor locations located roughly every half a mile, 140 CCTV cameras located roughly every mile, 80 main lane dynamic message signs (DMSs) located roughly every 3 miles, 121 frontage road DMSs, and 236 lane control signals (LCSs) located roughly every mile. No longer operational is an AVI subsystem that TransGuide used to collect travel time and speed data on corridors that did not have loop detector coverage. Currently, TransGuide is deploying Autoscope cameras to collect speed, volume, and occupancy data on several periphery corridors.

TransGuide's transportation management software operates as a client/server-based system that runs on Sun workstations in a Unix Solaris environment (*14*, *15*). The system includes several subsystems (Table 1), each with a number of components, including menu bars, processes, services, and servers. Table 2 describes subsystem components that are most relevant for understanding the ITS data archival process.

Table 1. TransGuide Subsystems.

Subsystem	Description
Administrative (ADM) Subsystem	It accomplishes basic administrative tasks and contains the main user interface—called the Advanced Traffic Management System (ATMS) Menu Bar—that sends requests to all other graphical user interface (GUI) servers in the system.
Alarm Incident Handler (AIH) Subsystem	It handles incident alarms using data from four subsystems: Local Control Unit (LCU), Advanced Warning to Avoid Railroad Delays (AWARD), AIH 911, and Pump Station. It also executes incident responses.
Advanced Traveler Information System (ATIS) Subsystem	It distributes travel information managed by the ATIS data server process.
AVI Subsystem	No longer operational, it handled real-time speed and travel time data using field data collected from vehicle AVI tags.
AWARD Subsystem	It provides railroad crossing information to motorist and emergency response vehicles. Using loop detector sensors, it calculates and predicts the arrival and duration of closures along the Union Pacific Kerrville Line.
CCTV Subsystem	It controls the operation of the CCTV cameras in the TransGuide ATMS.
Changeable Message Sign (CMS) or DMS Subsystem	It manages and controls DMSs through interaction with the Map Application and Scenario Management (SCM) subsystems.
Data Server Subsystem	It is the main centerpoint of access for all data in the TransGuide ATMS. It collects, stores, and distributes data to the TransGuide ATMS.
Dynamic Data Distribution (DDD) Subsystem	It distributes real-time ATMS data to the appropriate collection point. It interacts with all ATMS master processes and collects and sends equipment and incident data every 20 seconds to the Data Server Subsystem.
Estimated Travel Time (ETT) Subsystem	It provides current traffic and estimated travel time data to drivers through field equipment such as the DMSs.
Lane Closure GUI Subsystem	It allows operators to manually edit information about lane closures in a database table.
LCS Subsystem	It manages and controls LCS units. This subsystem interacts with the SCM and map display subsystems.
LCU Subsystem	It manages and controls LCUs in the TransGuide ATMS.
Map Application Subsystem	It is a set of map application tools (Real-Time Map Display, Real-Time Map Generation Tool, and World Wide Web Real-Time Map Display) that display TransGuide ATMS data using a map interface.
SCM Subsystem	It manages scenarios in the TransGuide ATMS.
Paging Subsystem	It sends alphanumeric pages from ATMS operators.
Personalized Assistance and Notification (PLAN) Subsystem	No longer operational, it enabled users to select routes for which they wanted to receive incident information from the TransGuide ATMS via e-mail.
Pump Station Subsystem	It handles alarms from the drainage pumps.

Table 2. Sample of Relevant ITS Data TransGuide Subsystem Components.

Component	Function
AIH Background Process	It handles alarm requests from all subsystems that produce incident events.
AIH Management Process	It manages all ATMS alarms and incidents.
AIH 911 Process	It reads current San Antonio Police Department incidents and updates the AIH background process.
AIH GUI Server	It displays ATMS incident alarms and messages on the screen.
CCTV Master	It manages all connection, disconnection, and control command requests.
CCTV GUI	It makes requests for connection and control commands to cameras.
CMS Master	It manages the interactions of all DMSs in the TransGuide ATMS.
CMS GUI Server	It manages the screens for the CMS interface.
TransGuide CMS Poll Server	It establishes connections to DMSs and polls DMSs once per polling cycle.
National Transportation Communications for ITS Protocol (NTCIP) CMS Poll Server	It is similar to the TransGuide CMS Poll Server, except it uses the statewide driver client library that supports the NTCIP to communicate with DMSs.
LCS Master	It manages lane control signal interactions.
LCS GUI Server	It manages the screens for the LCS interface.
TransGuide LCS Poll Server	It establishes connections to LCSs and polls LCSs once per polling cycle.
LCU Master	It manages LCU interactions.
LCU GUI Server	It manages screens for the LCU interface.
LCU Driver	It pushes commands from the LCU Master to the LCU poll servers. It also sends alarm packages to the AIH Subsystem.
Austin LCU Poll Server	It establishes connections to Austin LCUs and polls those LCUs once per polling cycle.
NazTech LCU Poll Server	It establishes connections to Naztech LCUs and polls those LCUs once per polling cycle.
Map Display Application	It provides access to real-time speed data, status data about road segments and traffic equipment, incident data, and lane closure data.
Map Generation Application	It creates a geographic representation of roadway segments and ITS equipment.
Scenario Master	It manages scenarios along with interactions with the field equipment.
Scenario GUI Server	It manages screens for the scenario interface.

Incident Detection and Response

Incident detection relies on a combination of detector-based alarms and 911-based alarms, CCTV camera scanning, police radio scanning, and courtesy patrols. The AIH subsystem handles detector-based alarms and 911-based alarms. For 911-based alarms, the AIH subsystem manages these alarms only if they are on or near TransGuide LCU-instrumented roadways. Detector-based alarms rely on speed for speed-trap detectors (installed on main lanes and some ramps) and percent occupancy for non-speed-trap detectors (mostly installed on entrance and exit ramps). LCUs continuously poll data from the detectors and relay 20-second aggregated data to the LCU driver. For speed-trap detectors, if a moving 2-minute average speed drops below 25 mph, the LCU driver automatically triggers a minor (yellow) alarm. If the moving 2-minute average speed drops below 20 mph, the alarm becomes a major (red) alarm. For non-speed-trap detectors, the default minor and major alarm thresholds are 25 percent occupancy and 35 percent occupancy, respectively.

It may be worth noting that these thresholds are default values and that the AIH subsystem allows users to set up different thresholds by time of day, day of week, or day of the year (*16*).

The system also allows users to vary LCU polling intervals from 10 to 60 seconds (the default is 20 seconds) and moving average lengths from 1 to 10 minutes (the default is 2 minutes). TransGuide officials rarely modify the default settings, partly because of the lack of a formalized procedure to access and analyze archived ITS data trends that could suggest that modifying default values could result in a more effective incident detection and alarm handling process.

The manager on duty receives all alarms, decides what further action is necessary, and assigns alarms to operators (Figure 1). After the manager assigns an alarm to an operator, the alarm becomes an incident. In practice, operators are responsible for specific corridors and tend to handle most incidents that happen on those corridors. However, if a corridor is experiencing too many alarms, the system manager can forward alarms to other operators to distribute the work load. At the operator's desk, all incidents on the network appear both on the system map and in the form of icons that identify the process that gave origin to the alarm (e.g., "LA" for lane alarm, "PD" for police department alarm, "RR" for railroad alarm, and "PS" for pumping station alarm) and a color code to indicate the alarm condition (green, yellow, or red).

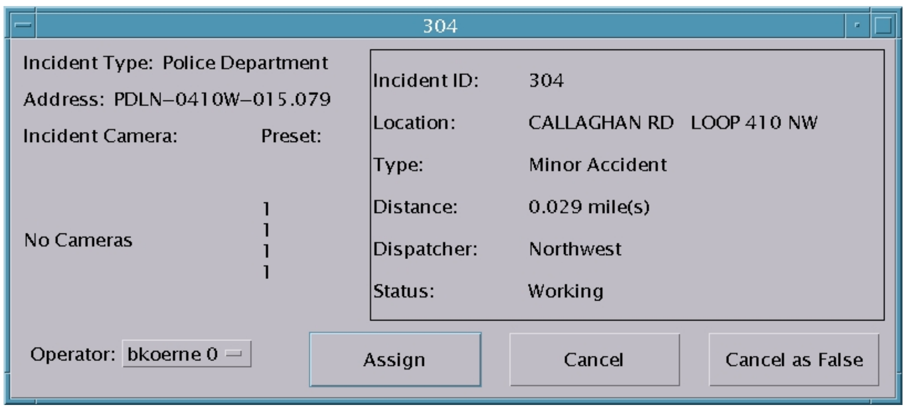


Figure 1. TransGuide Incident Assignment Screen.

After an operator acknowledges an incident, the system displays a modified version of the incident assignment screen and the CCTV subsystem attempts to display the primary incident camera listed on the incident screen (Figure 2). After verifying the incident with the CCTV camera, the operator has the option to execute a scenario (Figure 3). A scenario is a pre-defined set of messages that operators can apply to a pre-selected set of DMSs and/or LCSs, depending on incident type, extent, and location. In practice, operators also have the option to create new scenarios or modify existing scenarios to fit the needs of the specific incidents the operators are managing. The system displays all active scenarios using "S" icons (Figure 2).

The original TransGuide ATMS design allowed operators to load scenarios only if an incident record already existed in the system. After a system design change several years ago, operators were able to load scenarios even if an incident record did not previously exist. This resulted in added flexibility because operators could display DMS and LCS messages to manage incidents detected by processes such as CCTV camera scanning and courtesy patrols, i.e., incidents not handled by the AIH subsystem. In practice, the system design change did not include an alternate procedure to generate an incident record for those incidents, leaving the scenario record as the only data repository for incidents not handled by the AIH subsystem.

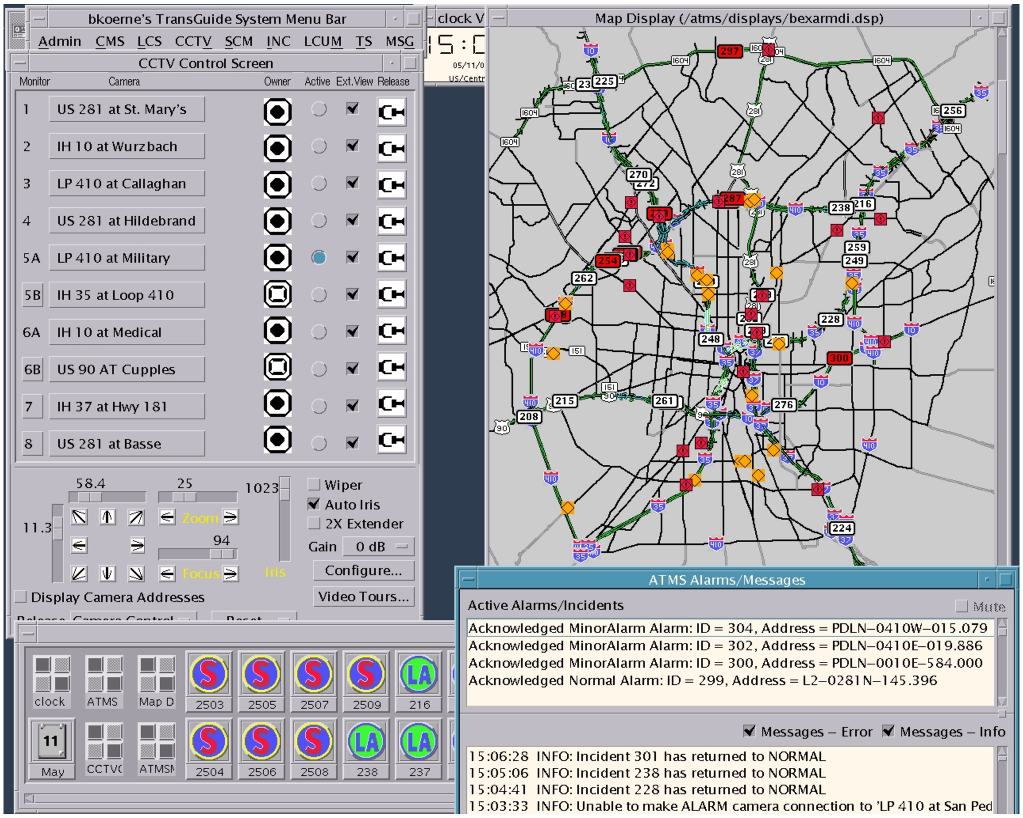


Figure 2. TransGuide Operator Console Screen.

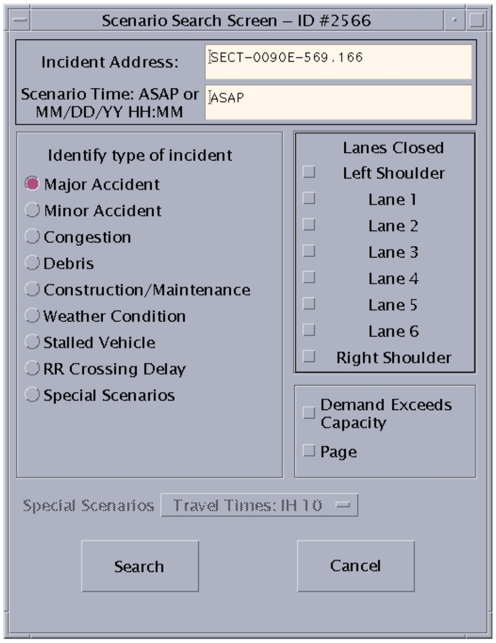


Figure 3. TransGuide Scenario Search Screen.

Because of the current structure of the detector-based incident detection algorithm, which relies on speed for main lanes and percent occupancy for ramps, many alarms are actually the result of recurrent roadway congestion. TransGuide has a policy of displaying congestion-related DMS messages to alert motorists about congested traffic conditions. Experienced operators are aware of the locations where the system usually triggers congestion-related alarms and prepare scenarios accordingly ahead of time. Typically, operators watch camera feeds for specific corridors to monitor congestion buildup. When the system begins to generate congestion-related alarms, or at the discretion of the operator, the operator may execute the scenario prepared in advance. In theory, operators could cancel congestion-related alarms at any time. In practice, they typically "iconize" congestion-related alarms and wait until speeds increase again before closing the alarms to prevent new alarm triggers at the same locations within a short period of time.

In general, the system design is such that, as long as there is an active alarm for a specific highway segment (whether the alarm is congestion-related or in response to an actual incident), the system does not trigger any new alarms for that segment. For the system to generate new alarms for the segment in question, the operator first has to close any previous active alarm associated with that segment. Because the operators' response is not automatic and can vary substantially from case to case, it becomes very difficult to replicate or predict exactly when operators close alarms, which, in turn, makes it difficult to fully characterize the incident detection and alarm incident handling processes using archived incident data. This also makes it difficult to identify and test strategies to optimize the incident detection process.

Traffic and Incident Data

TransGuide maintains a long-term data repository in compressed file format, which includes 20-second detector data (since July 1997) and event data (since January 1998) (*17*). TransGuide also maintains a scenario log in Sybase, which includes a scenario header table and a scenario execution table (since February 2002). The current 20-second detector data archive includes speed, volume, and percent occupancy. As Figure 4 shows, each record contains a date and time stamp, the detector address, and the corresponding average speed (in mph), volume, and percent occupancy values. The detector address has three components separated by a dash: detector location and designation (where "L" represents main lane, "EN" represents entrance ramp, "EX" represents exit lane, and the number represents the lane number beginning with the lane closest to the median), freeway number and direction, and mile marker. The system reports speeds on non-speed-trap detectors as -1.

```
02/04/2003 00:30:36 EN1-0035S-166.340   Speed=-1 Vol=001 Occ=001
02/04/2003 00:30:36 EX1-0035S-166.239   Speed=-1 Vol=000 Occ=000
02/04/2003 00:30:36 EX2-0035S-166.239   Speed=-1 Vol=001 Occ=001
02/04/2003 00:30:36 L1-0035N-166.450    Speed=61 Vol=001 Occ=001
02/04/2003 00:30:36 L2-0035N-166.450    Speed=54 Vol=001 Occ=001
02/04/2003 00:30:36 L2-0035S-166.450    Speed=60 Vol=004 Occ=005
02/04/2003 00:30:36 L3-0035N-166.450    Speed=54 Vol=004 Occ=005
02/04/2003 00:30:36 L3-0035S-166.450    Speed=57 Vol=004 Occ=005
```

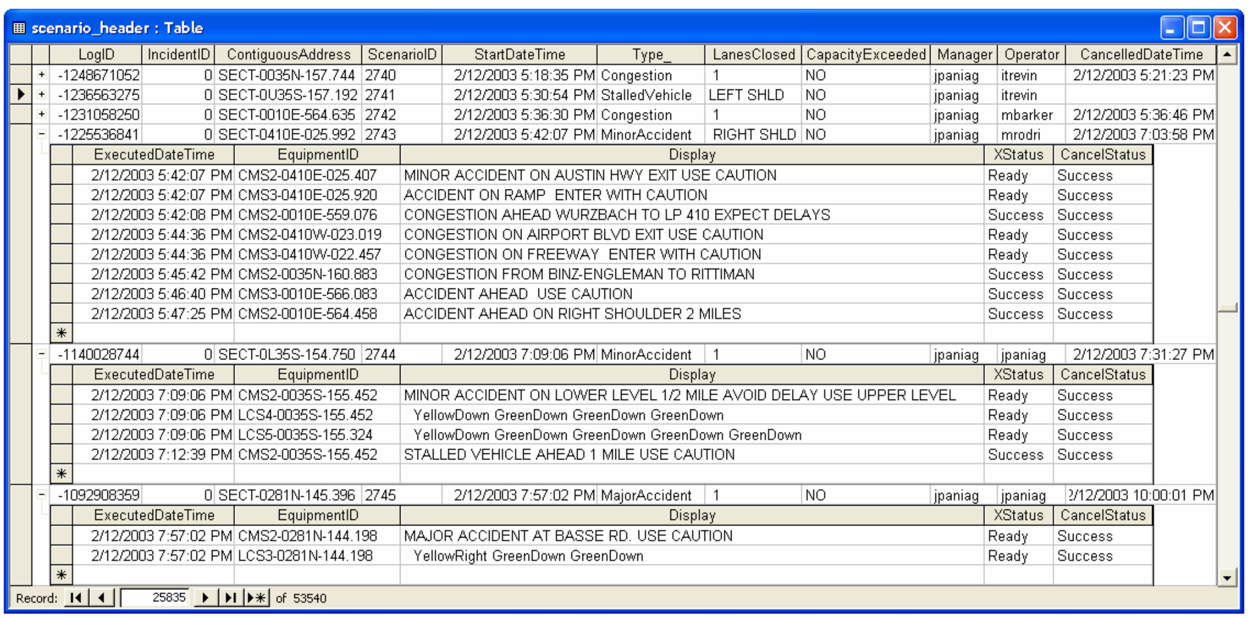
Figure 4. Sample 20-Second Detector Data at TransGuide.

The current event data archive includes 30 different major record types (such as 2301, 2303, and 8354), with several record types including more than one record subtype (Figure 5). The original intent of the event data archive was to serve as a debugging tool for ATMS, but over time, the archive has grown to become a very extensive data repository. Of particular interest in this research are record types 5301, 5302, and 5303, which contain incident data records.

```
8337 lcu_driver5 2003/05/20 16:00:10 1053464410 L2-1604W-032.121 1 61 5 15 25 144
2301 cms_master 2003/05/20 15:57:00 1053464220 CMS CMS2-0410W-025.558 Display Return: msgID=2643
    text='TRAVEL TIME TO|US281 4-6 MINS|IH10 9-11 MINS|||'
2301 cms_master 2003/05/20 15:57:12 1053464232 CMS CMS2-0090W-568.933 Display Return: msgID=2646
    text='TRAVEL TIME TO|LP410 5-7 MINS|HUNT LN 6-8 MINS|||'
5304 aih_back 2003/05/20 16:00:33 1053464433 258 'L2-0035S-164.412' 'SECT-0035S-164.412' 3 28 23
    'Normal' 'CCTV-0035N-163.955'.1 'ToBeAssigned' 'blopez' '' '' 1053461876 0 0 0
5302 aih_back 2003/05/20 16:00:33 1053464433 258 'L2-0035S-164.412' 'SECT-0035S-164.412' 3 23 25
    'MinorAlarm' 'CCTV-0035N-163.955'.1 'ToBeAssigned' 'blopez' '' '' 1053461876 0 0 0
5301 aih_back 2003/05/20 16:00:33 1053464433 274 'L3-0035N-164.412' 'SECT-0035N-164.412' 3 20 16
    'MinorAlarm' 'CCTV-0035N-164.835'.1 'ToBeAssigned' 'blopez' '' '' 1053464433 0 0 0
8341 aih_mgmt 2003/05/20 16:00:33 1053464433 274 SECT-0035N-164.412 blopez 3 0
2301 cms_master 2003/05/20 15:57:21 1053464241 CMS CMS3-0035N-164.308 Display Return: msgID=2805
    text='CONGESTION|ON FREEWAY||ENTER WITH|CAUTION|'
2301 cms_master 2003/05/20 15:57:21 1053464241 CMS CMS2-0035N-168.672 Display Return: msgID=2645
    text='TRAVEL TIME TO|LOOP 1604|UNDER 5 MINS|||'
5302 aih_back 2003/05/20 16:00:49 1053464449 267 'EN2-0035S-153.608' 'SECT-0035S-153.608' 2 -1 35
    'MinorAlarm' 'CCTV-0035N-153.619'.1 'ToBeAssigned' 'blopez' '' '' 1053463712 0 0 0
2301 cms_master 2003/05/20 15:57:42 1053464262 CMS CMS2-0035S-168.645 Display Return: msgID=2645
    text='TRAVEL TIME TO|LP410 UNDER 5 MINS|US281 12-14 MINS|||'
```

Figure 5. Sample Event Data at TransGuide.

Figure 6 shows sample records from the archived scenario database. Each record includes a header that summarizes basic data from the scenario loaded by the operator and a linked table that contains actual DMS and LCS messages displayed in the field.



scenario_header : Table

LogID	IncidentID	ContiguousAddress	ScenarioID	StartDateTime	Type_	LanesClosed	CapacityExceeded	Manager	Operator	CancelledDateTime
-1248671052	0	SECT-0035N-157.744	2740	2/12/2003 5:18:35 PM	Congestion	1	NO	jpaniag	itrevin	2/12/2003 5:21:23 PM
-1236563275	0	SECT-0U35S-157.192	2741	2/12/2003 5:30:54 PM	StalledVehicle	LEFT SHLD	NO	jpaniag	itrevin	
-1231058250	0	SECT-0010E-564.635	2742	2/12/2003 5:36:30 PM	Congestion	1	NO	jpaniag	mbarker	2/12/2003 5:36:46 PM
-1225536841	0	SECT-0410E-025.992	2743	2/12/2003 5:42:07 PM	MinorAccident	RIGHT SHLD	NO	jpaniag	mrodri	2/12/2003 7:03:58 PM

ExecutedDateTime	EquipmentID	Display	XStatus	CancelStatus
2/12/2003 5:42:07 PM	CMS2-0410E-025.407	MINOR ACCIDENT ON AUSTIN HWY EXIT USE CAUTION	Ready	Success
2/12/2003 5:42:07 PM	CMS3-0410E-025.920	ACCIDENT ON RAMP ENTER WITH CAUTION	Ready	Success
2/12/2003 5:42:08 PM	CMS2-0010E-559.076	CONGESTION AHEAD WURZBACH TO LP 410 EXPECT DELAYS	Success	Success
2/12/2003 5:44:36 PM	CMS2-0410W-023.019	CONGESTION ON AIRPORT BLVD EXIT USE CAUTION	Ready	Success
2/12/2003 5:44:36 PM	CMS3-0410W-022.457	CONGESTION ON FREEWAY ENTER WITH CAUTION	Ready	Success
2/12/2003 5:45:42 PM	CMS2-0035N-160.883	CONGESTION FROM BINZ-ENGLEMAN TO RITTIMAN	Success	Success
2/12/2003 5:46:40 PM	CMS3-0010E-566.083	ACCIDENT AHEAD USE CAUTION	Success	Success
2/12/2003 5:47:25 PM	CMS2-0010E-564.458	ACCIDENT AHEAD ON RIGHT SHOULDER 2 MILES	Success	Success

LogID	IncidentID	ContiguousAddress	ScenarioID	StartDateTime	Type_	LanesClosed	CapacityExceeded	Manager	Operator	CancelledDateTime
-1140028744	0	SECT-0L35S-154.750	2744	2/12/2003 7:09:06 PM	MinorAccident	1	NO	jpaniag	jpaniag	2/12/2003 7:31:27 PM

ExecutedDateTime	EquipmentID	Display	XStatus	CancelStatus
2/12/2003 7:09:06 PM	CMS2-0035S-155.452	MINOR ACCIDENT ON LOWER LEVEL 1/2 MILE AVOID DELAY USE UPPER LEVEL	Ready	Success
2/12/2003 7:09:06 PM	LCS4-0035S-155.452	YellowDown GreenDown GreenDown GreenDown	Ready	Success
2/12/2003 7:09:06 PM	LCS5-0035S-155.324	YellowDown GreenDown GreenDown GreenDown GreenDown	Ready	Success
2/12/2003 7:12:39 PM	CMS2-0035S-155.452	STALLED VEHICLE AHEAD 1 MILE USE CAUTION	Success	Success

LogID	IncidentID	ContiguousAddress	ScenarioID	StartDateTime	Type_	LanesClosed	CapacityExceeded	Manager	Operator	CancelledDateTime
-1092908359	0	SECT-0281N-145.396	2745	2/12/2003 7:57:02 PM	MajorAccident	1	NO	jpaniag	jpaniag	2/12/2003 10:00:01 PM

ExecutedDateTime	EquipmentID	Display	XStatus	CancelStatus
2/12/2003 7:57:02 PM	CMS2-0281N-144.198	MAJOR ACCIDENT AT BASSE RD. USE CAUTION	Ready	Success
2/12/2003 7:57:02 PM	LCS3-0281N-144.198	YellowRight GreenDown GreenDown	Ready	Success

Record: 25835 of 53540

Figure 6. Sample Scenario Data at TransGuide.

INCIDENT DETECTION ALGORITHM ASSESSMENT

Three commonly used measures to conceptualize and/or assess the performance of incident detection algorithms are:

- Detection Rate (DR): It is the ratio of the number of detected incidents to the total number of recorded incidents.
- False Alarm Rate (FAR): It is the ratio of incorrect decisions (false positives) to the total number of algorithm decisions made.
- Detection Time (DT): It is the time interval between the moment the incident occurred and the time the incident was detected.

As Figure 7 shows, detection rate is directly proportional to the detection time. Likewise, the false alarm rate is inversely proportional to the detection time. It follows that by increasing the time it takes for the algorithm to detect incidents (which would result, e.g., from using a more sophisticated algorithm), it is possible to increase the detection rate while, at the same time, reducing false alarm rates. Unfortunately, a longer detection time would also result in a longer incident response time, which is normally undesirable. Likewise, a too short detection time (which would result, e.g., from using a relatively simple algorithm), while desirable, would also result in low detection rates and high false alarm rates. Consequently, it becomes necessary to calibrate the incident detection algorithm to achieve an acceptable balance between detection rates, false alarm rates, and detection times.

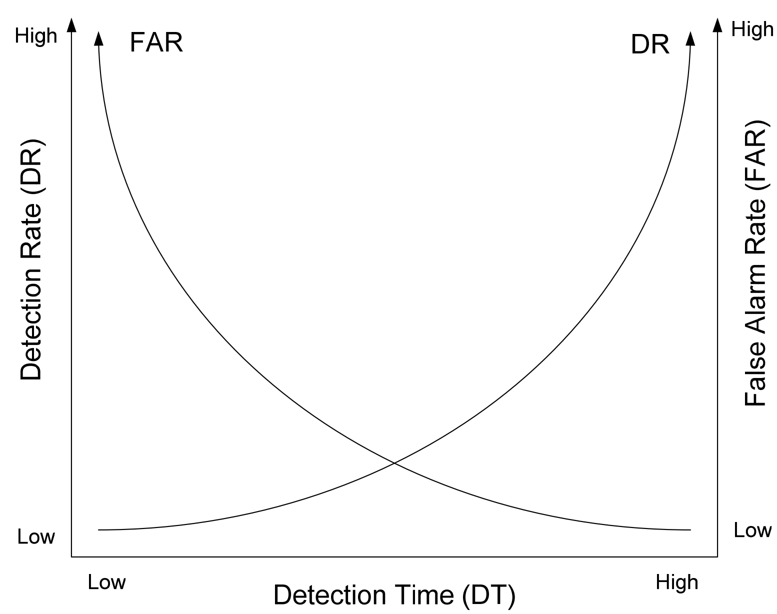


Figure 7. Incident Detection Algorithm Performance Measure Relationships (*9*).

It may be worth noting that one of the original design objectives at TransGuide called for fast incident detection—which resulted in a simple one-parameter incident detection algorithm—even at the expense of the false alarm rate (*16*). The reasoning behind this decision was that

TransGuide operators would be able to quickly confirm individual alarms using the video subsystem, rendering the requirement for a more sophisticated incident detection algorithm (that would most likely result in longer incident detection times) unnecessary.

Matching Alarms to Incidents

The researchers prepared two datasets for the analysis. The first dataset contained 19,553 scenario database records from March 2002 – May 2004 that included data for four types of incidents: major accidents, minor accidents, stalled-vehicle incidents, and debris incidents. The researchers assumed this database provided an adequate representation of the history of freeway incidents based on the results of an analysis completed during the first phase that found similarities between incidents (major and minor accidents) from the TransGuide scenario database and crash data from the Texas Department of Public Safety (TxDPS) (*1*). Although the analysis showed differences by time of day (the number of TransGuide-reported incidents during daytime hours was larger, but at night, the number of TxDPS-reported crashes was larger), the impact on the total number of incidents as a result of potential nighttime underreporting at TransGuide should be relatively minor because the vast majority of incidents happen during daytime hours.

The second dataset contained alarms triggered by the TransGuide incident detection algorithm in response to events on the road. For the analysis, the researchers focused on record types 5301 and 5303 from the event log files. In general, the AIH subsystem creates a new 5301 record every time it receives a new alarm package from the LCU subsystem. It also creates a new 5303 record every time an operator closes an alarm. The system also generates other record types while an alarm is active, although those records are of no particular interest for this analysis. In total, for the March 2002 – May 2004 period, the dataset contained records for 202,690 alarms.

In an ideal situation, the number of records in the two datasets would be the same, with a record in the incident dataset having a corresponding matching record in the alarm dataset. In practice, because of false alarms, potentially erroneous scenario records, and other factors, there is not a perfect match between incident records and alarm records. In general, as Figure 8 shows, there are three possible matching outcomes:

- Incident Detected. This occurs if an incident actually happened (a scenario was deployed) and the alarm incident handler *triggered* an alarm.
- False Negative. This occurs if an incident actually happened (a scenario was deployed) and the alarm incident handler *did not trigger* an alarm.
- False Positive. This occurs if an incident did not happen (a scenario was not deployed) but the alarm incident handler triggered an alarm.

			LCU Subsystem Triggered Alarm?	
			Yes	**No**
Scenario Deployed?	**Yes**	**Incident Occurred**	Incident Detected	False Negative
	No	**No Incident Occurred**	False Positive	

Figure 8. Possible Incident versus Alarm Dataset Matching Outcomes.

To find the number of detected incidents, the researchers attempted to match incidents reported in the scenario database to alarms recorded in the event database. Because of the lack of a common link between these two datasets (more specifically, an incident ID), the researchers had to develop a "fuzzy" spatio-temporal query methodology whereby an incident would be considered detected if the LCU subsystem triggered an alarm within a pre-specified spatio-temporal window associated with an incident record (Figure 9). The reason behind this fuzzy range concept was to account for situations such as an alarm being triggered before or after operators deployed a scenario (which almost always happens because the two datasets are not synchronous), an alarm being triggered on a sector other than where the incident actually happened, and scenarios being reported on the wrong sector. Figure 10 illustrates the query building process, which used the geodatabase structure described in report 0-4745-1 (*1*).

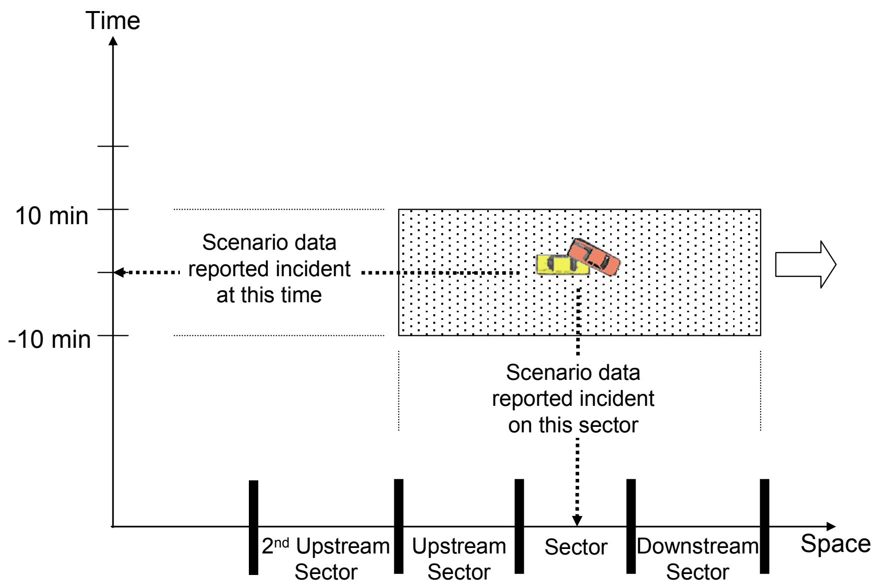


Figure 9. Spatio-Temporal Query Concept.

A preliminary analysis suggested using a spatio-temporal window composed of three highway sectors (including the sector of interest as well as the adjacent upstream and downstream sectors) and a 10-minute range before and after the scenario execution time. To test this hypothesis, the researchers conducted a sensitivity analysis (Figure 11). As Figure 11a shows, the number of matched incidents and alarms increased with the number of sectors considered. However, the rate of increase in the number of matches flattened after including more than three sectors in the query (the sector of interest as well as the adjacent upstream and downstream sectors), clearly suggesting that the chances of sector mismatch decreased considerably outside the three sector window. Figure 11b shows that the number of matched incidents and alarms increased as the time window size increased. In this case, the number of matches did not flatten, suggesting the possibility of an increasing number of alarm records incorrectly matching incident records and that using time window size was not necessarily a strong query parameter. Nonetheless, since it was necessary to use a time window factor for the query building process anyway, the researchers decided to maintain the 10-minute range before and after the scenario execution time.

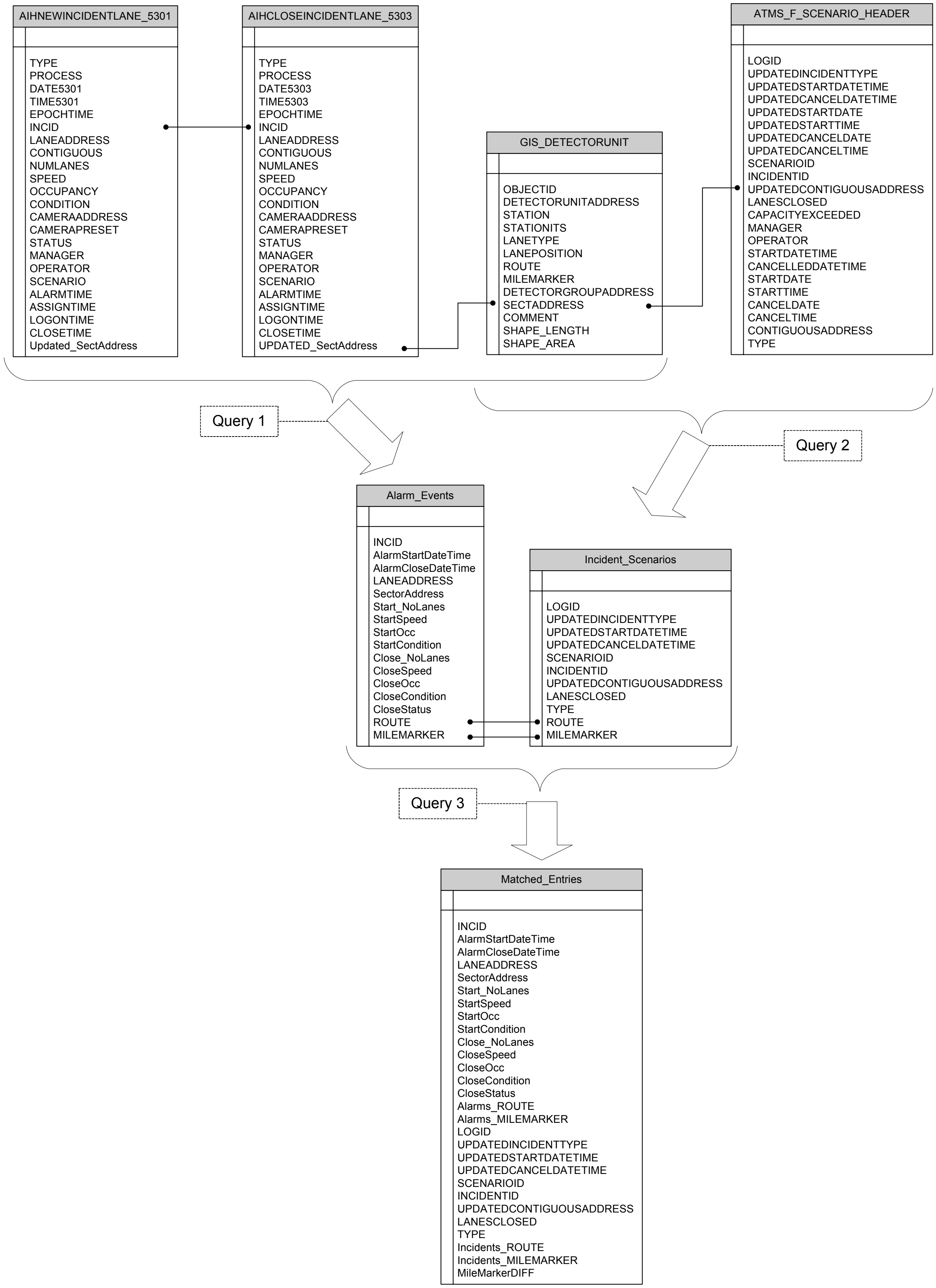


Figure 10. Query Building Process to Match Incidents to Alarms.

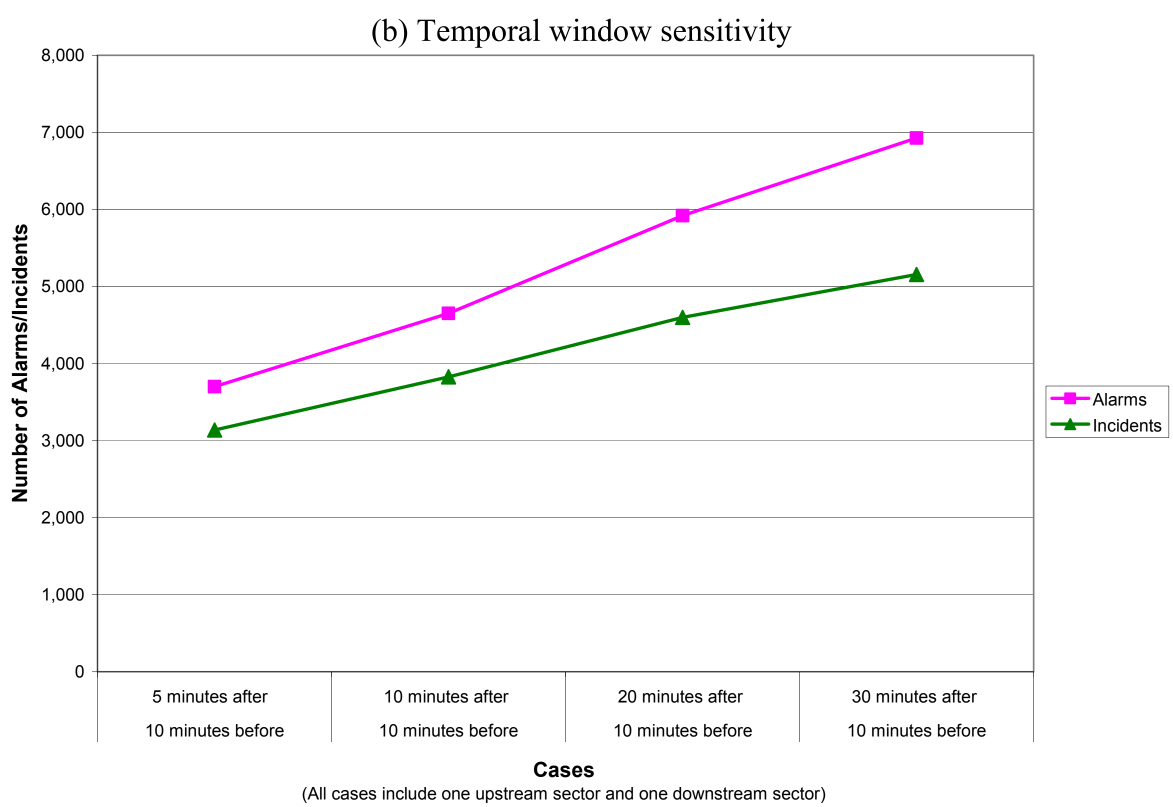


Figure 11. Sensitivity Results for Incident-Alarm Matching Query.

Analysis

Figure 12 summarizes the results of the matching operation. From Figure 11, out of 19,553 incidents during the March 2002 – May 2004 analysis period, 3,828 incident records had a matching alarm record. Likewise, 4,651 alarm records had a matching incident record. Therefore,

$$\text{Detection Rate (DR)} = \frac{\text{No. of detected incidents}}{\text{No. of recorded incidents}} = \frac{3{,}828}{19{,}553} 100\% = 19.6\%$$

$$\text{False Alarm Rate (FAR)} = \frac{\text{No. of false positives}}{\text{No. of algorithm decisions}} = \frac{198{,}039}{4{,}320 \times 1{,}463 \times 792} 100\% = 0.0039\%$$

This calculation assumed for simplicity that the algorithm made 4,320 decisions per detector per day (once every 20 seconds) and that all 1,463 detectors in the geodatabase were operational all the time during the 792-day analysis period from March 2002 to May 2004.

It was not possible to calculate the third performance measure (detection time) because the archived incident data did not provide a measure for when incidents actually happened in relation to the time the system detected the incidents.

			LCU Subsystem Triggered Alarm?	
			Yes	**No**
Scenario Deployed?	**Yes**	**Incident Occurred**	Incident Detected	False Negative
	No	**No Incident Occurred**	False Positive	

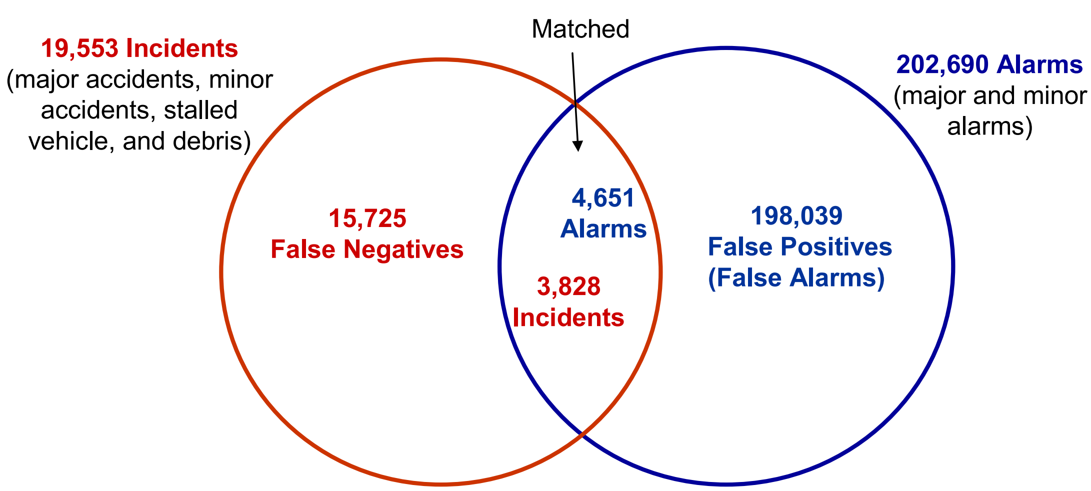


Figure 12. Summary of Matching Results.

The 19.6 percent incident detection rate included major and minor accidents, stalled vehicles, and debris. After excluding debris incidents from the analysis, the incident detection rate would grow to 20.0 percent (3,695 detected incidents relative to 18,427 recorded incidents). Likewise, excluding debris and stalled vehicle incidents from the analysis would result in an incident detection rate of 24.8 percent (2,755 detected incidents relative to 11,083 recorded incidents). Excluding debris, stalled vehicles, and minor accidents would result in an incident detection rate of 27.2 percent (1,789 detected incidents relative to 6,571 recorded incidents). In general, these percentages indicate the incident detection algorithm is responsible for the detection of 20 – 27 percent of incidents detected at TransGuide. The literature reports detection rates that are typically much higher—between 60 and 100 percent (*4*, *9*), but it also includes references to detection rates in the 30 – 50 percent range (*18*). Readers should be aware that many high detection rate reports in the literature use very small sample sizes and/or pre-set thresholds calibrated under the assumption of "normal flow" conditions (*19*). Actual performance on the ground tends to be lower (*18*).

A false alarm rate of 0.0039 percent is relatively low compared to rates typically found in the literature—between 0.0018 and 1.9 percent (*9*). However, a low false alarm rate, although desirable, is not necessarily a good performance measure because it ignores the frequency of false alarms operators actually experience (*19*). Another disadvantage is that it ignores the total number of alarms the algorithm triggers. The overall false alarm frequency is:

$$\text{False Alarm Frequency (FAF)} = \frac{\text{No. of false positives}}{\text{No. of hours}} = \frac{198{,}039}{792 \times 24} = 10 \text{ false alarms/hour}$$

which indicates that, on average, operators need to respond to false alarms every 6 minutes. Because most alarms happen during daytime hours, the false alarm frequency during daytime hours (17 alarms per hour or a false alarm roughly every 4 minutes) is much higher than the overall average.

A formulation that measures the effectiveness of the incident detection algorithm in terms of the number of incident-confirmed alarms relative to the total number of alarms actually triggered is:

$$\text{Effective Alarm Rate (EAR)} = \frac{\text{No. of confirmed alarms}}{\text{No. of alarms}} = \frac{4{,}651}{202{,}690} 100\% = 2.3\%$$

which would correspond to an "alternative" false alarm rate—that relates the number of false alarms to the number of alarms actually triggered—of 97.7 percent. This result indicates that between two and three alarms for every 100 alarms correspond to detected incidents.

Both detection rate and false rate measures are average values. As Figure 13 and Figure 14 show, detection rates and false alarm rates vary considerably by location. Interestingly, many sectors with low detection rates are located on corridor sections with relatively low volumes, such as SL 1604 west of IH-10, US 90 west of SH 151, and IH-37 south of SE Military Drive. One possible reason for the low detection rates on those corridors is the minor alarm threshold of 25 mph currently in place, which might not be high enough to enable an effective detection of incidents. However, a few other more heavily traveled corridors also experienced low detection rates. Likewise, Figure 14 shows several sectors with unusually high false alarm rates. The research could not identify reasons for such apparent anomalies. Further analyses would be necessary to clarify the issue.

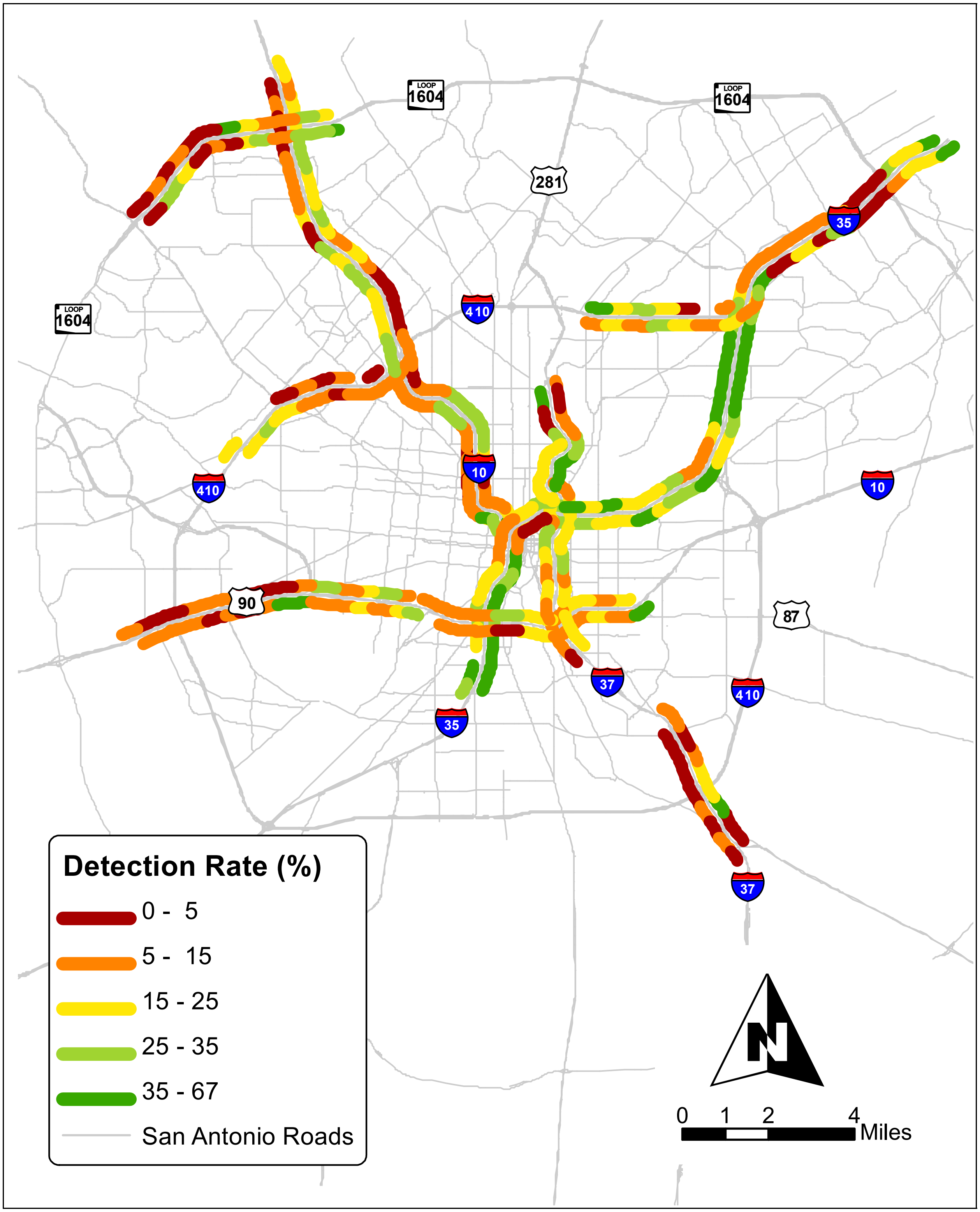


Figure 13. Distribution of Detection Rates by Sector.

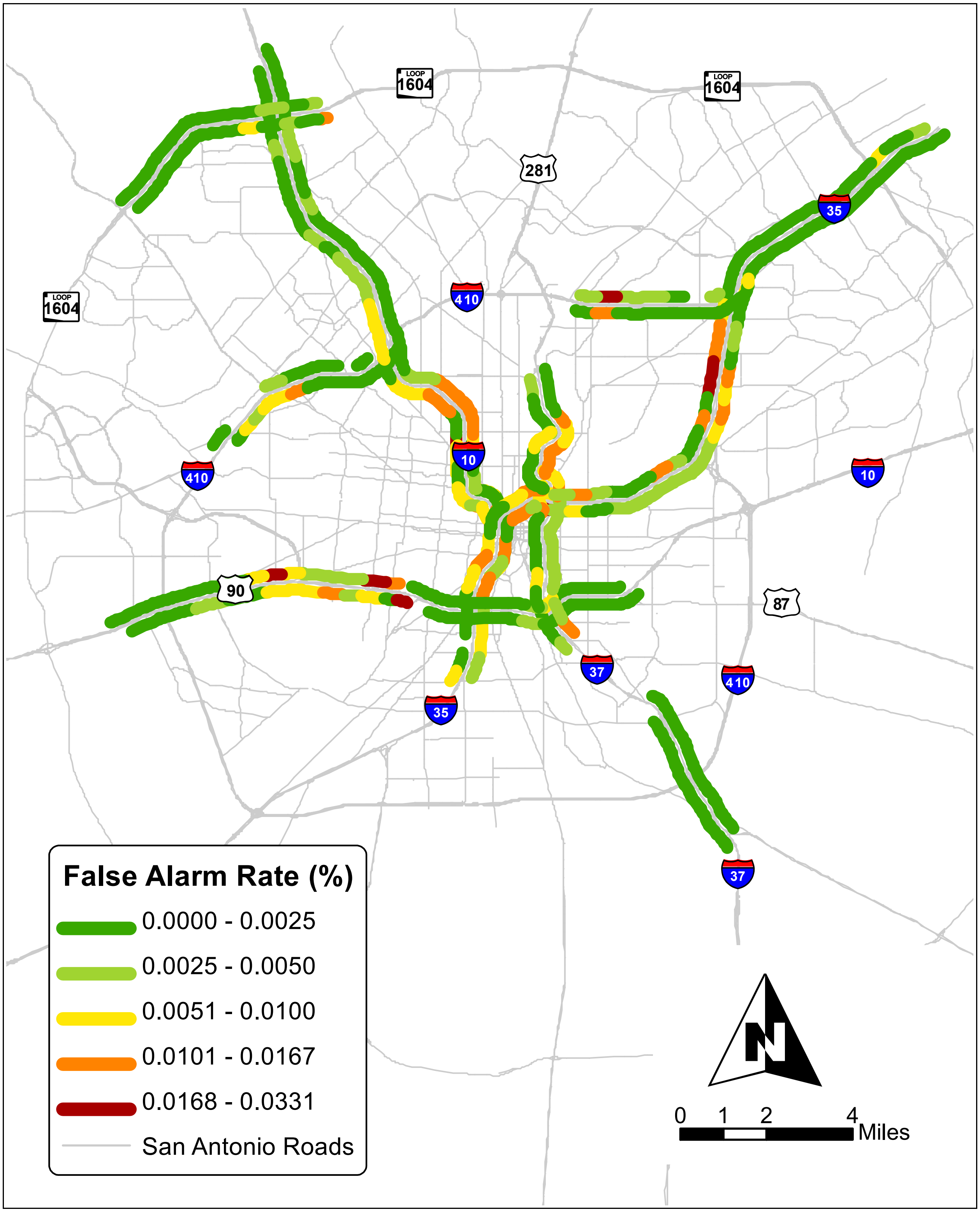


Figure 14. Distribution of False Alarm Rates by Sector.

CHAPTER 3. INCIDENT DETECTION ALARM THRESHOLD OPTIMIZATION

This chapter describes the work completed to assess the feasibility of modifying current incident detection alarm thresholds to help optimize TMC incident detection practices. It describes the analytical methodology followed, a prototype offline tool to evaluate incident detection algorithm performance, and ways to increase detection rate while minimizing the impact on false alarm rates.

ANALYTICAL APPROACHES

A number of approaches may be possible to optimize TMC incident detection algorithm performance. In the specific case of Texas TMCs, one approach would be to increase from single-parameter incident detection algorithms (e.g., TransGuide uses speed for speed-trap detectors and occupancy rate for non-speed-trap detectors) to multiple-parameter incident detection algorithms. There are several ways to increase the number of parameters, e.g., by using speed, volume, and occupancy data from a single sector or speed, volume, and/or occupancy data from multiple sectors (e.g., the sector in question, the upstream sector, and the downstream sector). The rationale behind this approach is that traffic flow relationships tend to behave differently during abnormal traffic conditions as opposed to normal conditions and that tracking those relationships in real-time would enhance/optimize incident detection (Figure 15).

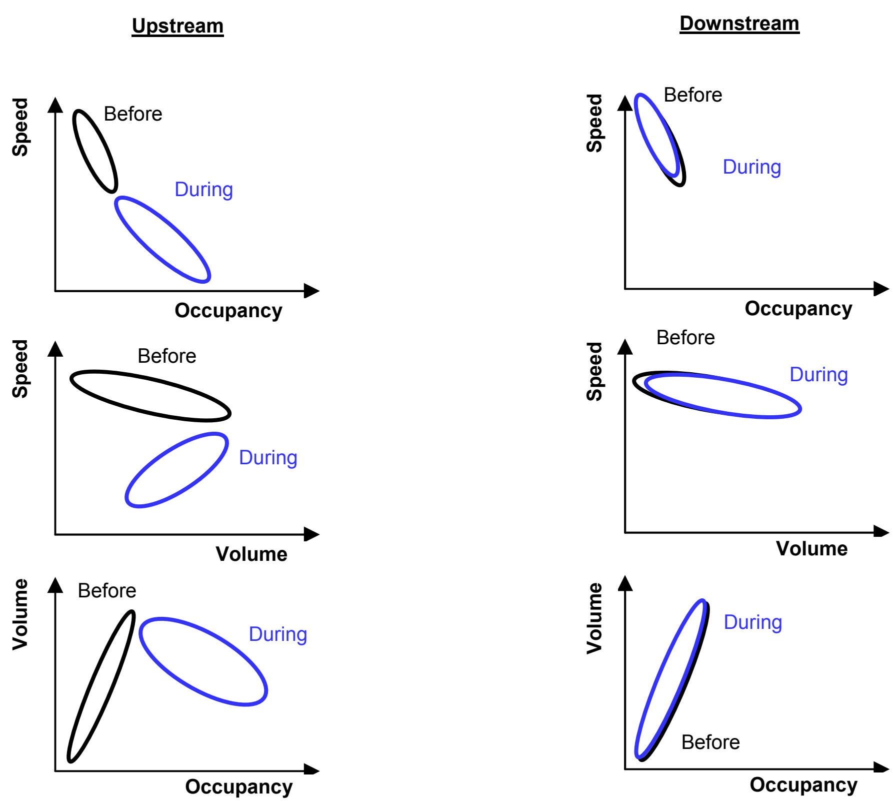


Figure 15. Typical Traffic Flow Relationships Before and During Incidents.

Another approach would be to vary alarm thresholds according to some pre-specified criteria (e.g., by time of day or based on traffic flow conditions) to enable a more effective incident detection algorithm response. Figure 16 illustrates this concept. An incident happened at about 5:23 PM (TMC operators displayed DMS and LCS messages to warn motorists about the incident at 5:31:48 PM). During the incident, speeds decreased from about 50 mph to 26 mph. However, the incident detection algorithm did not trigger an alarm because the speed alarm threshold was set at 25 mph. By raising the threshold to 30 mph, the algorithm could have triggered an alarm on lane 2 at 5:33:22 PM (still later than the TMC response, though). Raising the alarm threshold to 35 mph could have enabled the incident detection algorithm to trigger an alarm on lane 1 at 5:25:21 PM and potentially result in an earlier TMC response to the incident. Further, raising the alarm threshold to 40 mph could have enabled the algorithm to trigger an alarm at 5:24:33 PM. Raising the alarm threshold even further, however, would not have produced any additional benefit. For example, at 45 mph, it would have produced at least one false alarm (at 5:07:44 PM).

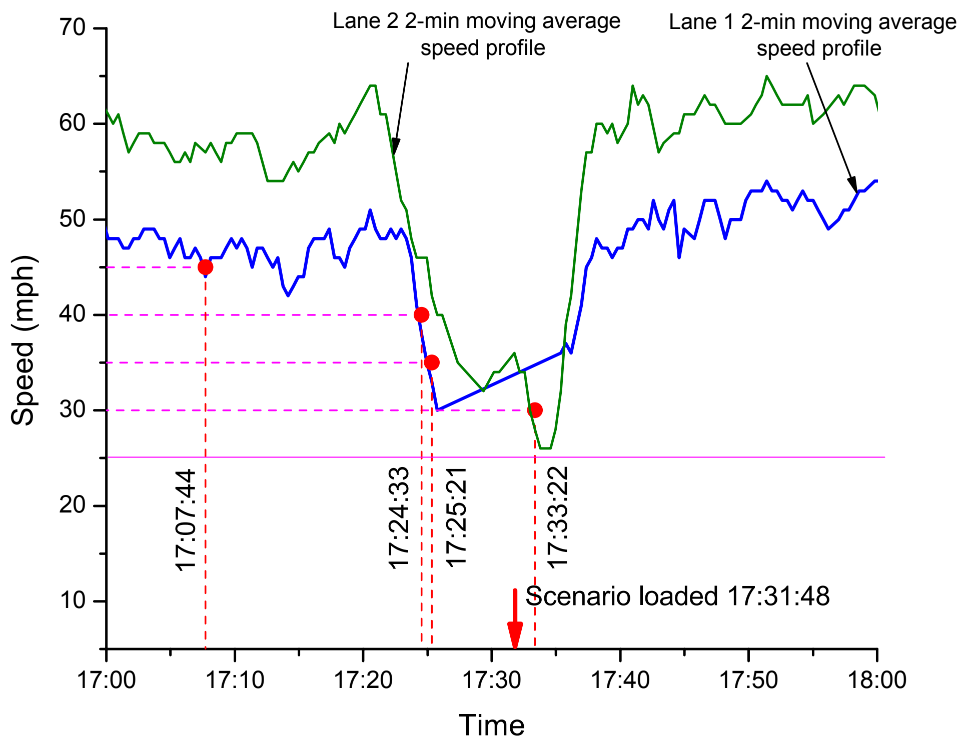


Figure 16. Illustration of Dynamic Threshold Levels Approach for AID Optimization.

SAMPLE CASES

For the analysis, the researchers used incident data—and corresponding archived 20-second lane data—from the list of 19,553 incidents identified in Chapter 2. Following the structure in Figure 12, the researchers selected cases that fell under the Detected Incident, False Negative, and False Positive (i.e., False Alarm) categories. Evaluating all 19,553 incident cases would have been ideal but unfeasible during the course of the research. First, to analyze incident detection algorithm behavior and response properly, it was necessary to analyze in great detail all available data associated with each incident (speed, volume, occupancy, incident response messages, event data, and so on, both for the sector in question as well as adjacent sectors). Basically, this meant analyzing incident data on a case by case basis. Second, in the process to identify suitable sample sites, the researchers found many cases of gaps in the lane data that effectively prevented the use of the data to properly characterize incidents or to understand incident detection algorithm response. Chapter 4 will discuss the issue of data completeness in greater detail. Third, rather than focusing on the development of automated procedures to optimize/enhance incident detection algorithm performance in a batch mode, it was of a higher priority to adequately understand incident cases in a controlled laboratory setting with the hope that the lessons learned from that analysis could be exported later to the rest of the cases. For these reasons, the decision was to select a sample that could still be considered representative of the entire population. Figure 17 shows the location of the sample cases selected. The 75 sample cases covered a wide range of situations, including detected incidents (26), false negatives (35), and false positives (14). For completeness, the dataset included data from both Traffic Operations Division (TRF) LCUs (30 cases) and Naztec LCUs (45 cases).

INCIDENT CHARACTERIZATION

For each case selected, the researchers prepared plots depicting speed, volume, and occupancy, relationships for the main sector, as well as the upstream and downstream sectors (Figure 18 through Figure 21). To better understand differences and similarities among cases, the researchers categorized each case according to the criteria in Table 3 (for both detected incidents and false negatives) and Table 4 (for false positives—false alarms). These classification criteria follow previous work documented in the literature (*3*, *9*, *22*). Table 5 summarizes the results of the categorization effort.

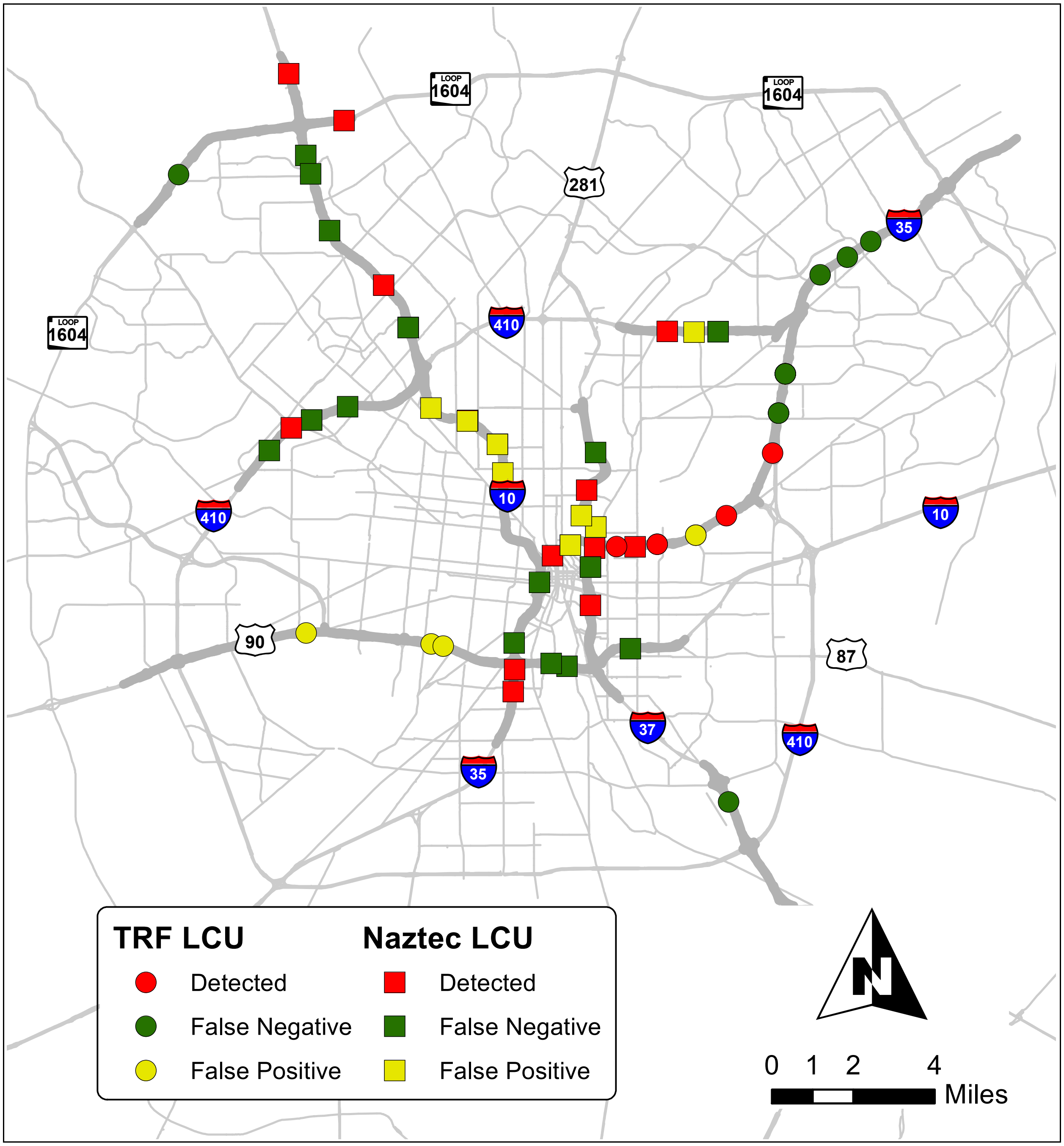
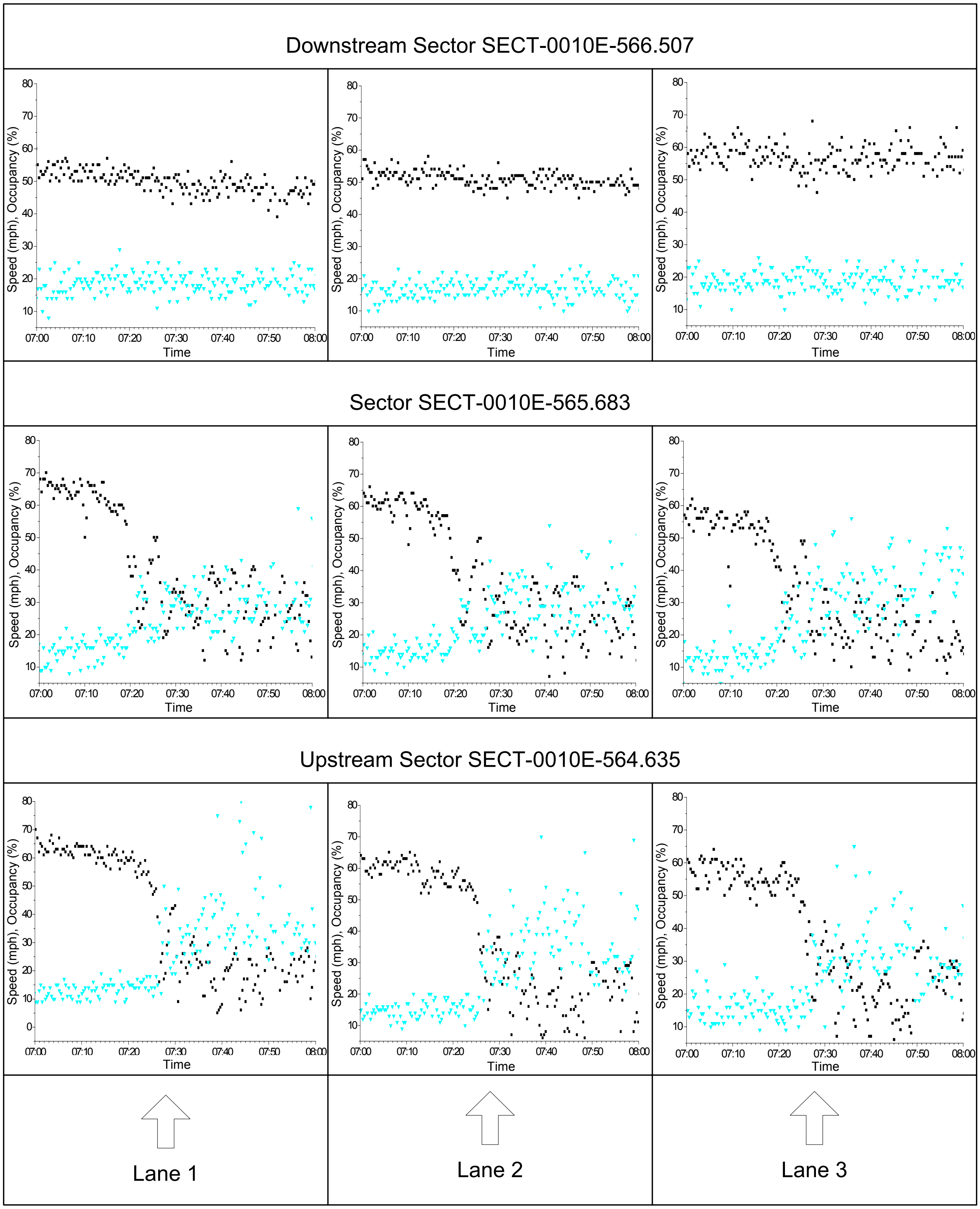


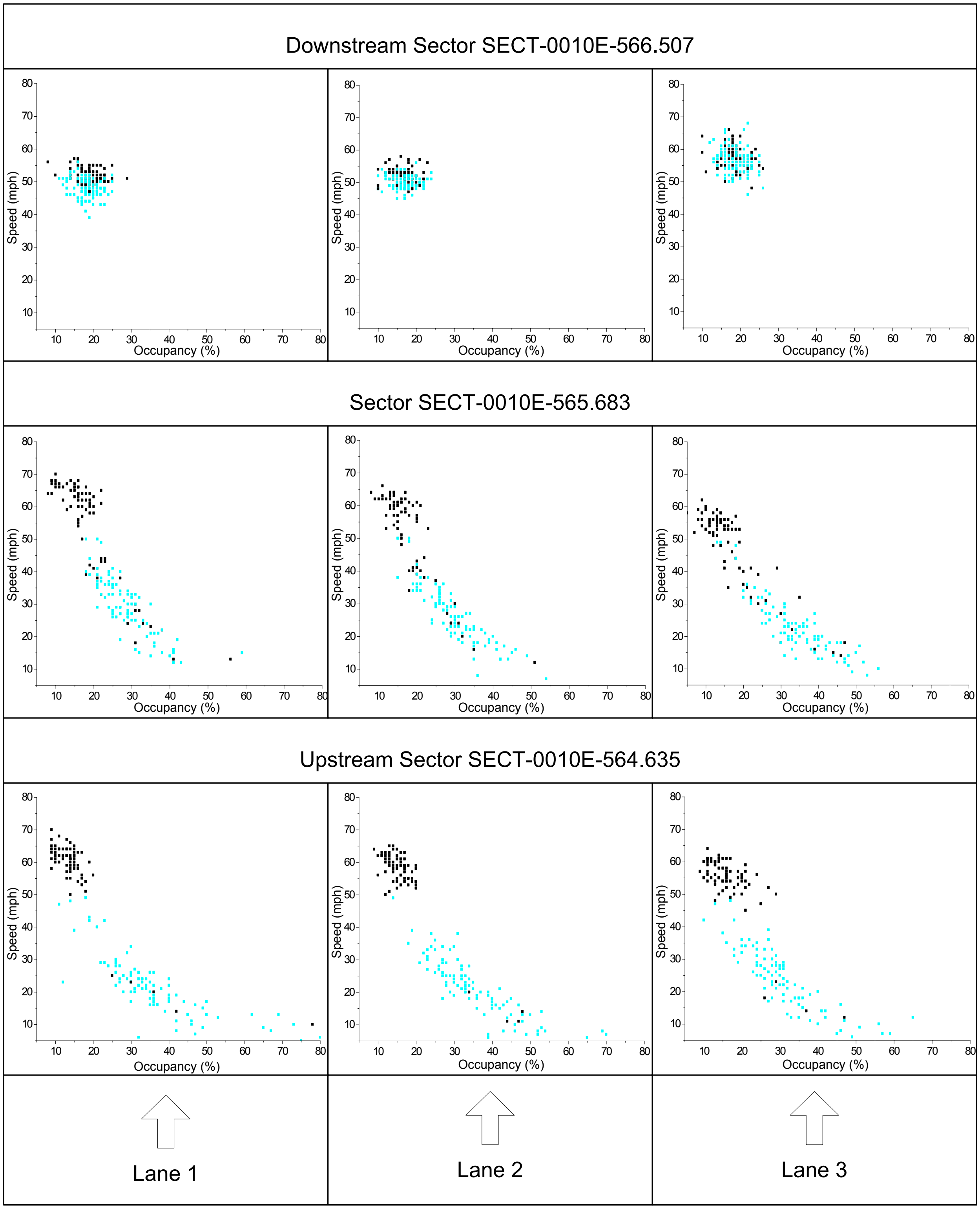
Figure 17. Spatial Distribution of Sample Cases.



Black squares represent speed and blue triangles represent occupancy.

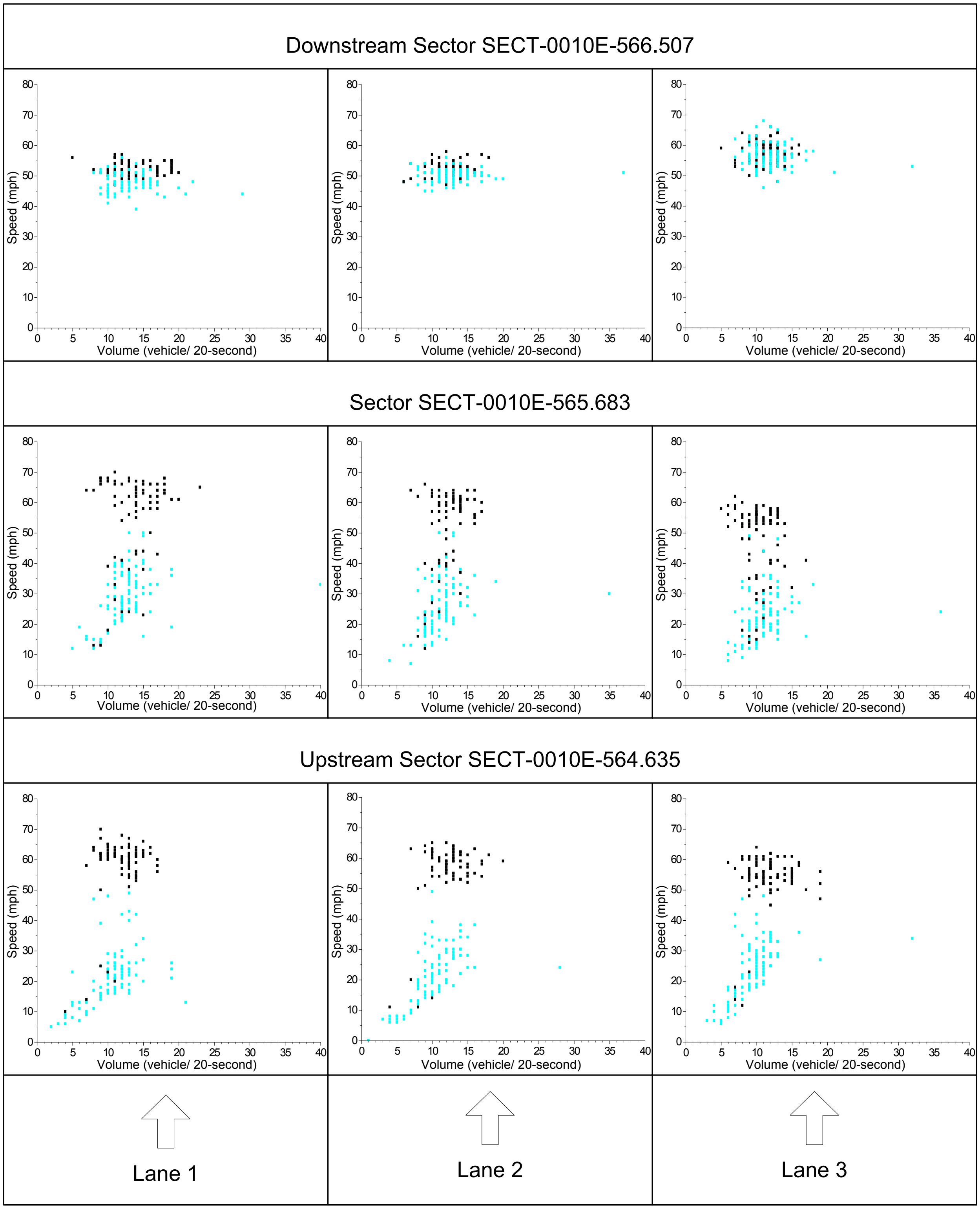
An alarm event was triggered at 7:29 AM on lane 2 of SECT-0010E-565.683.

Figure 18. Speed and Occupancy Profiles for Sample Case No. 45.



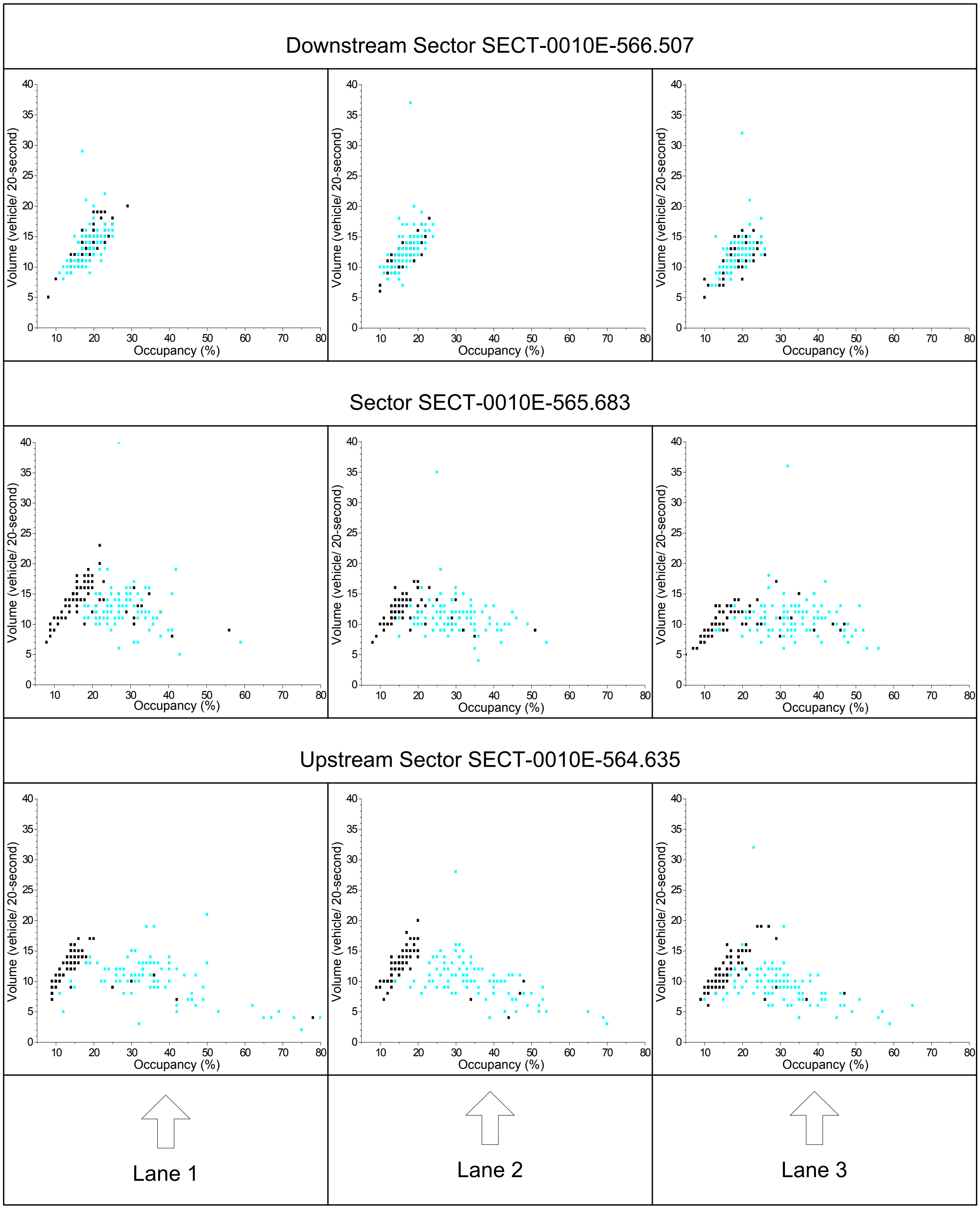
Blue squares represent conditions after an alarm was triggered at 7:29 AM on lane 2 of SECT-0010E-565.683.

Figure 19. Speed versus Occupancy Plots for Sample Case No. 45.



Blue squares represent conditions after an alarm was triggered at 7:29 AM on lane 2 of SECT-0010E-565.683.

Figure 20. Speed versusVolume Plots for Sample Case No. 45.



Blue squares represent conditions after an alarm was triggered at 7:29 AM on lane 2 of SECT-0010E-565.683.

Figure 21. Volume versus Occupancy Plots for Sample Case No. 45.

Table 3. Categories for Incident Pattern Characterization.

Type	Conditions before incident	Conditions during incident	Comment
Type I-1	Uncongested traffic	Roadway capacity at the site of incident is less than the volume of oncoming traffic. A queue develops upstream of the incident site while at the same time a region of light traffic develops downstream.	This is the easiest pattern to detect.
Type I-2	Uncongested traffic	Incident partially and shortly blocks a lane or two in a multi-lane sector. Roadway capacity at the incident site exceeds incoming traffic. Only minor, or possibly no queues, form in the immediate vicinity of the incident.	Impact of incident is less severe than Type I-1. It is more difficult to detect.
Type I-3	Light or free-flowing traffic	Incident has no noticeable impact on traffic. It can also happen in situations with moderate traffic conditions with a stalled vehicle or debris on the shoulder.	This type is extremely difficult for any algorithm to detect.
Type I-4	Congested or heavy traffic	Roadway capacity at the incident site is less than the volume (and capacity) of the traffic downstream. Since the incident meters traffic entering downstream sector, demand in downstream decreases and thus congestion starts to slowly clear downstream while congestion persists upstream of incident.	Some algorithms may detect this type of incident but only after considerable delay and/or processing.
Type I-5	Congested or heavy traffic	Roadway capacity (and volume) at the incident site is greater than the volume of the traffic downstream. This situation usually happens when incident occurs in the midst of queues caused by earlier incident or recurrent congestion. Since traffic is already congested, incident impact will be hardly distinguishable.	Similar to Type I-3, this type is extremely difficult for any algorithm to detect.

Table 4. Categories for False Alarm Pattern Characterization.

Type	Conditions causing alarm generation	Comment
Type A-1	Abnormal traffic data due to malfunctioning detector(s) reporting high occupancy or very-low speed values.	Bad data produces bad decisions. Quality control tests should be used to validate data before using algorithm.
Type A-2	Heavy traffic (stop-and-go) condition due to recurrent congestion causing significant speed variations similar to those experienced during incidents.	Most significant contributor to false alarms.
Type A-3	Abnormal roadway geometrics, such as sharp horizontal curves, severe vertical grades, or intermediate ramps and interchanges can cause speed to decrease.	TMC operators can identify these locations relatively easily. It requires localized customization of incident detection parameters.
Type A-4	Bottlenecks caused by excessive entrance ramp traffic can cause significant variations in occupancy and speed values.	Excessive entrance ramp traffic can occur during peak hours or special events. TMC operators can frequently identify these situations.
Type A-5	Slow-moving vehicle(s), such as trucks or rubberneckers, can cause isolated variations in speed prompting an alarm.	Improper calibration of incident detection algorithm parameters could trigger this type of false alarm.

Table 5. Categorization of Sample Cases.

Type	Incident Cases						False Positive (False Alarm) Cases	
	Total		Detected Incident		False Negative			
Type I-1	25	42%	22	85%	3	9%		
Type I-2	7	12%	1	4%	6	18%		
Type I-3	11	18%	0	0%	11	32%		
Type I-4	6	10%	2	8%	4	12%		
Type I-5	4	7%	1	4%	3	9%		
Type A-1							3	20%
Type A-2							7	47%
Type A-3							3	20%
Type A-4							2	13%
Type A-5							0	0%
Missing Data[1]	7	11%		0%	7	21%		0%
Total	**60**	**100%**	**26**	**100%**	**34**	**100%**	**15**	**100%**

[1] Either data were missing completely or critical data were missing for incident detection purposes.

An analysis of the data in Table 5 yields the following results:

- Overall, 25 (or 42 percent) of incidents were of Type I-1. Of this total, 22 (or 85 percent of detected incidents) were of Type I-1. Only 3 cases were false negatives. It is likely that incident detection algorithm optimization could lower this number even more considering that Type I-1 cases are normally the easiest to detect. Readers should be aware that 42 percent does not represent a true "incident detection rate," but rather an indication of the relative proportion of Type I-1 cases in the sample. Such a percentage turned out to be higher that the incident detection rates documented in Chapter 2 and may provide an indication that the selected sample was probably biased. Nonetheless, the result in Table 5 is still useful because it provides at least some approximation to the relative frequency of cases that might be encountered in the actual population.
- There were 15 Type I-3 or Type I-5 incidents (or 25 percent). Of this total, 14 were false negatives (i.e., there was an incident but the incident detection algorithm did not detect them). Considering that these two incident types are usually the most difficult to detect, it is highly unlikely that incident detection algorithm optimization alone could lower the number of false negatives significantly under this category.
- There were 13 Type I-2 or Type I-4 incidents (or 22 percent). Of this total, 10 were false negatives. It is likely that incident detection optimization could reduce this number, although it is not clear at this point by how much.
- There were 7 (or 11 percent) of incidents with critical lane data missing, which prevented the incident detection algorithm from generating alarms and also prevented an assessment of the feasibility to optimize the algorithm to increase the chances of detection. Notice that 89 percent does not represent a true measure of data completeness, since some of the other cases also had missing data, but in those cases the missing data were not critical for incident detection. Overall, in about 27 percent of all cases studied, there was at least one lane in the immediate vicinity of the incident with missing data. Further, only 42 percent of all cases had complete data from the main sector as well as its downstream and upstream sectors. Chapter 4 will explore the issue of data completeness in greater detail.

- There were 7 Type A-2 false positive (false alarm) cases (or 47 percent), which were associated with typical recurrent congestion conditions. It is likely that optimizing the incident detection algorithm could lower the number of false alarms under this category, particularly during peak hours when most of the Type A-2 alarms tend to occur.
- There were 3 Type A-1 false positive cases (or 20 percent), which were likely the result of abnormal lane detector data. The most effective way to address this type of false alarm cases would be by introducing quality control tests to the lane data prior to their use by the incident detection algorithm.
- There were 5 Type A-3 or Type A-4 false positive cases (or 33 percent), which were likely influenced by localized road geometric characteristics or perhaps by heavy entrance ramp traffic that could have caused sudden changes in speed and/or occupancy values. In most cases, localized customization of incident detection parameters would be necessary to reduce the number of false alarms under this category.

ALARM THRESHOLD MODIFICATION TOOL

The analysis from the previous section led to the conclusion that adding multiple parameters to the incident detection algorithm at TransGuide would not necessarily result in more effective incident detection, particularly during congested periods, when the need is highest. An additional challenge was that lane data time stamps from contiguous sectors were frequently asynchronous, making the matching process between corresponding records more difficult, therefore less feasible. These reasons prompted an assessment of the feasibility of the second approach, i.e., varying alarm thresholds according to some pre-specified criteria (e.g., by time of day or based on traffic flow conditions). To test this approach, the researchers developed an offline tool called Incident Detection Algorithm Tester (IDAT) that simulates the alarm generation process at TransGuide. The purpose of the tool was to measure the impact of modifying speed alarm thresholds on the number and timing of alarms generated by the system.

As Figure 22 shows, IDAT enables users to select one or more sectors of interest and a range of dates. With this information, the tool reads archived 20-second lane data from the archived lane data database, calculates 2-minute moving average speeds, and “triggers” minor and major alarms if the moving averages fall below the pre-specified thresholds. IDAT also enables users to export the minor and alarm data to comma-delimited text files.

Conceptually, the process to generate alarms using moving average speed values based on pre-specified thresholds is straightforward. In practice, simulating archived alarm events can be quite challenging because, in reality, as long as TMC operators are managing active alarms, the system ignores (and therefore does not archive) any new alarms from any of the lane detectors within the affected sectors. Floor personnel are supposed to close the alarms when they no longer need to manage the incidents, but the exact time when this happens varies considerably from case to case. Frequently during recurrent congestion conditions, operators “iconize” alarm windows to prevent the system from generating new alarms for that sector, sometimes through the rest of the peak period or when the congestion ends. Because of the uncertainty associated with the time an alarm effectively closes, it is not always possible to determine if the event archive contains all the alarms the system could have generated, therefore making it very difficult to fully replicate the archived alarm event database.

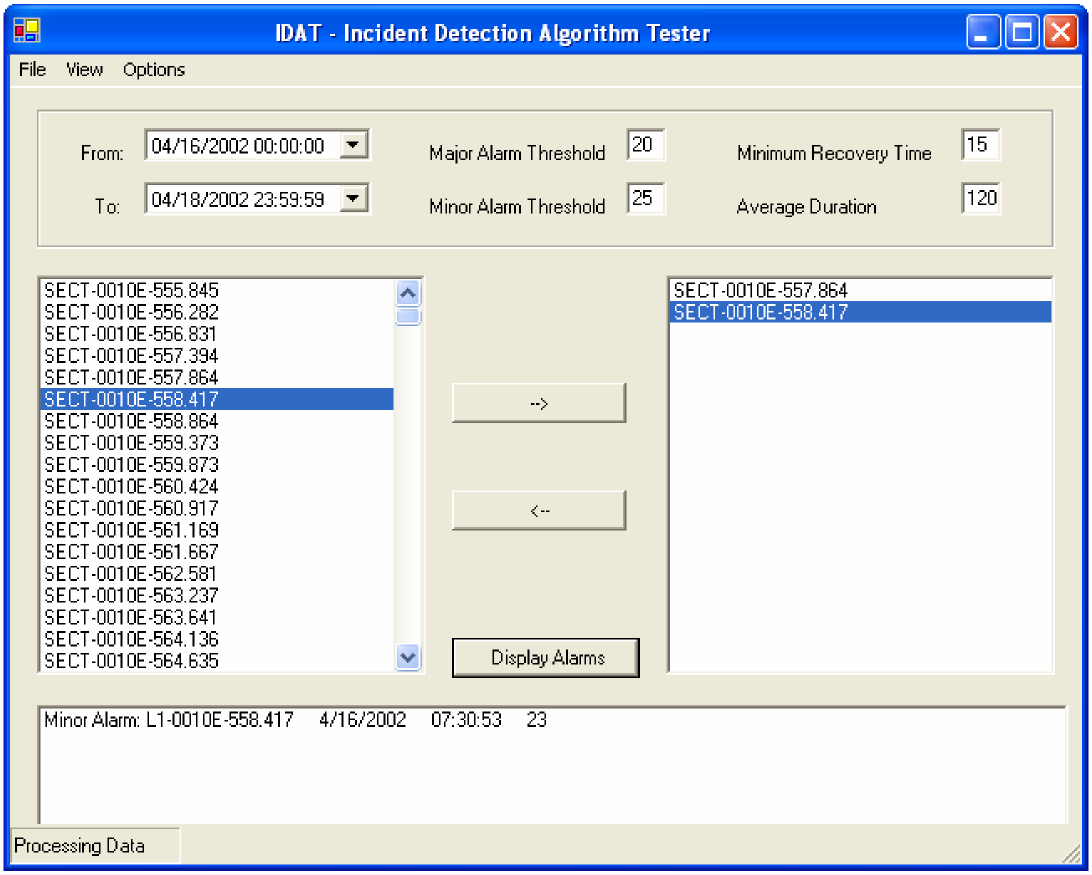


Figure 22. Incident Detection Algorithm Tester Interface.

To overcome this difficulty, the researchers introduced an artificial "minimum recovery time" to enable an alarm to close automatically if the calculated moving average value was consistently larger than the minor alarm threshold (i.e., the moving average "recovered") for the duration of that minimum recovery time. After a calibration phase that involved varying the minimum recovery time from 5 minutes to 60 minutes in 5-minute increments, the researchers selected a default value of 15 minutes. The 15-minute recovery time produced a number of generated alarms that was closest to the number of alarms in the archive.

The actual incident detection algorithm at TransGuide is extremely complex because it needs to deal with many special situations, e.g., records containing zero speeds or zero occupancy values, zero speeds and non-zero occupancy values, data from two different types of LCU software, and data gaps of various durations. The researchers attempted to replicate the algorithm as closely as possible, but at some point it became evident that rather than achieving a 100 percent matching rate between simulated alarms and archived alarms, it was more important to develop a tool that would enable the measurement of *changes* in the number and timing of alarms in response to changes in alarm thresholds. The simulation actually focused on detecting minor alarm thresholds because the incident detection algorithm at TransGuide simply modifies the status of already existing alarms when the moving averages drop below the major alarm threshold.

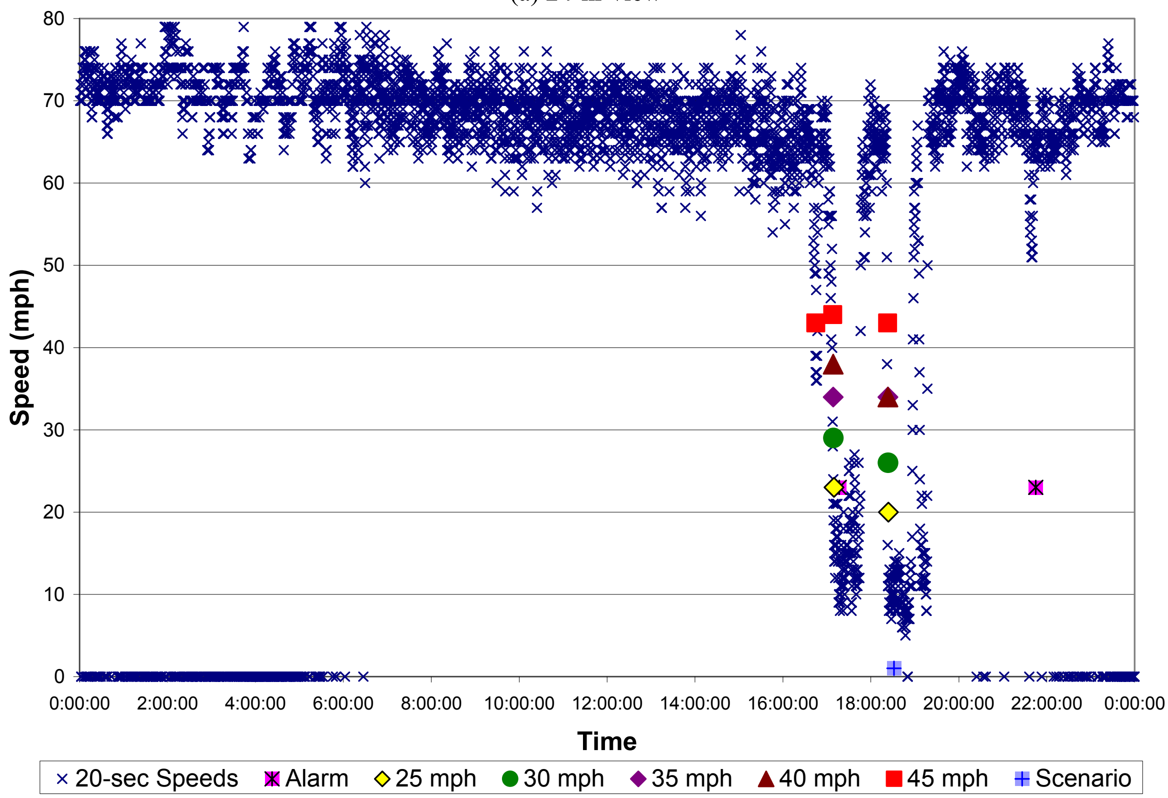
ANALYSIS AND RESULTS

To conduct the analysis, the researchers selected 27 cases from the 75 cases identified previously. The 27 cases were the only cases that involved incidents that the incident detection algorithm either detected or could have detected if the alarm thresholds were modified. For each case, the researchers ran IDAT for a 24-hour period and recorded the number of alarms generated in addition to the corresponding time stamps. The simulation involved using five different alarm threshold values: 25 mph (current minor alarm threshold), 30 mph, 35 mph, 40 mph, and 45 mph. To better visualize traffic conditions surrounding each of the 27 sample cases studied, the researchers plotted speed profiles covering the 24-hour period of analysis. In addition to the speed profile, each plot included all alarms generated at different alarm threshold levels, as well as all reported incidents for that particular sector and day.

As an illustration, Figure 23 shows the plot for one of the sample cases considered. According to the speed profile, there was a decrease in speed at approximately 6:23 PM that turned out to be associated with an incident (as documented by a scenario database record executed at 6:31 PM). During the incident, speeds dropped from about 65 mph to 10 mph. Figure 23 shows the speed data points IDAT selected for various alarm threshold values from 25 mph to 45 mph in 5-mph increments. Interestingly, there was no record for an alarm in the event database. However, the speed profile also shows that between 5 and 6 PM there was another drop in speed, which turned out to be associated with recurrent congestion (as documented by an alarm at about 5:15 PM and a congestion-related scenario deployed by TMC operators). Figure 23 shows the corresponding speed data points IDAT selected. There was no record for when TMC operators closed the congestion-related alarm, which raises the possibility that the alarm remained active long after the congestion ended at about 7:20 PM. If the alarm was still active when the actual incident at 6:23 PM happened, this would explain why there was no alarm record in the database.

Using the plots facilitated the determination of incident detection times as well as an assessment of a number of factors (e.g., moving average speed structure, congestion levels, and data gaps). After plotting and analyzing all cases, the researchers calculated average number of alarms per 24-hour period and average incident detection time for all incidents within the same period. Figure 24 shows the relationship between average number of alarms per 24-hour period, average incident detection time, and alarm threshold level. For completeness, Figure 25 shows the effect of congestion levels on the average number of alarms per 24-hour period, and Figure 26 shows the effect of congestion levels on average incident detection time. For completeness, all figures show both average values and relative changes with respect to the current 25-mph alarm threshold.

(a) 24-hr view



(b) Four-hour view

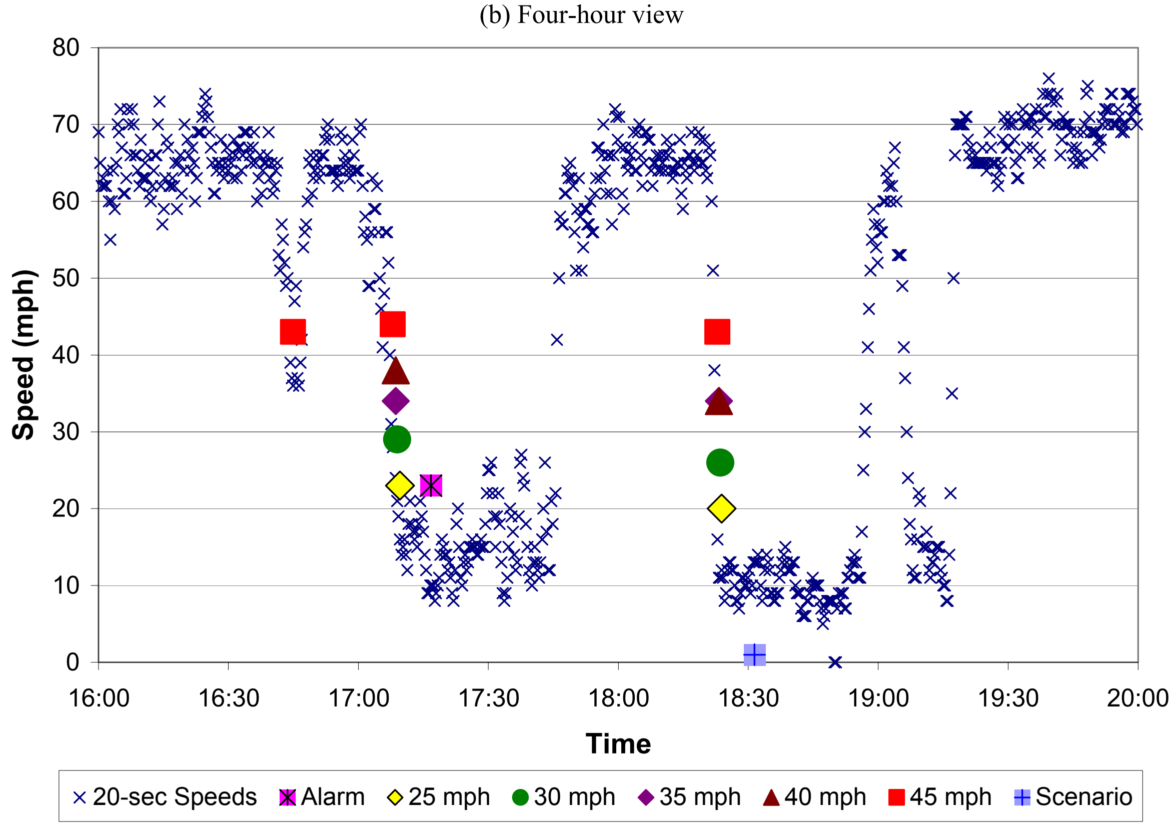


Figure 23. Sample 24-hour Speed Profile with Alarms for Different Threshold Levels.

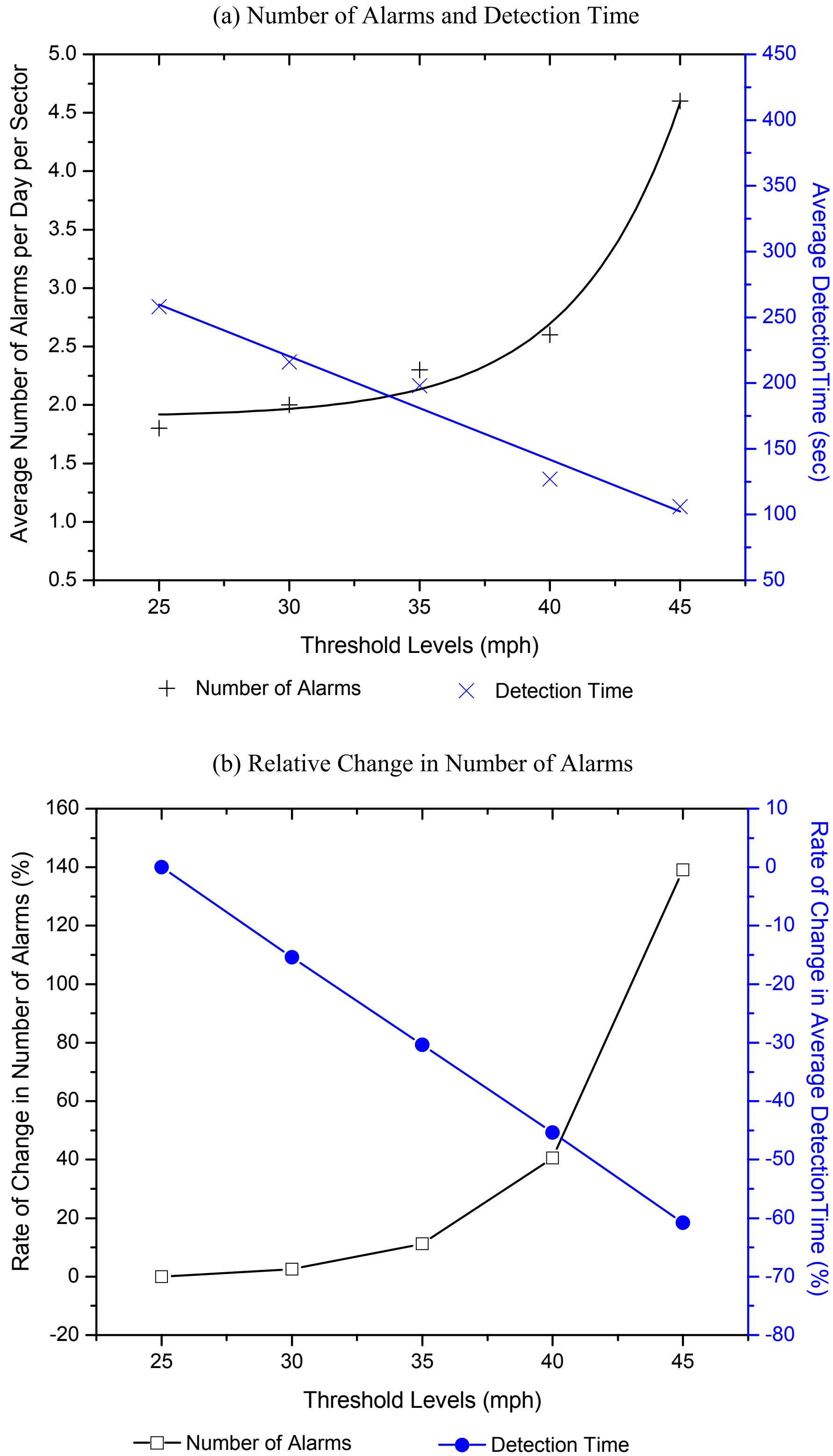


Figure 24. Impact of Alarm Thresholds on Number of Alarms and Detection Times.

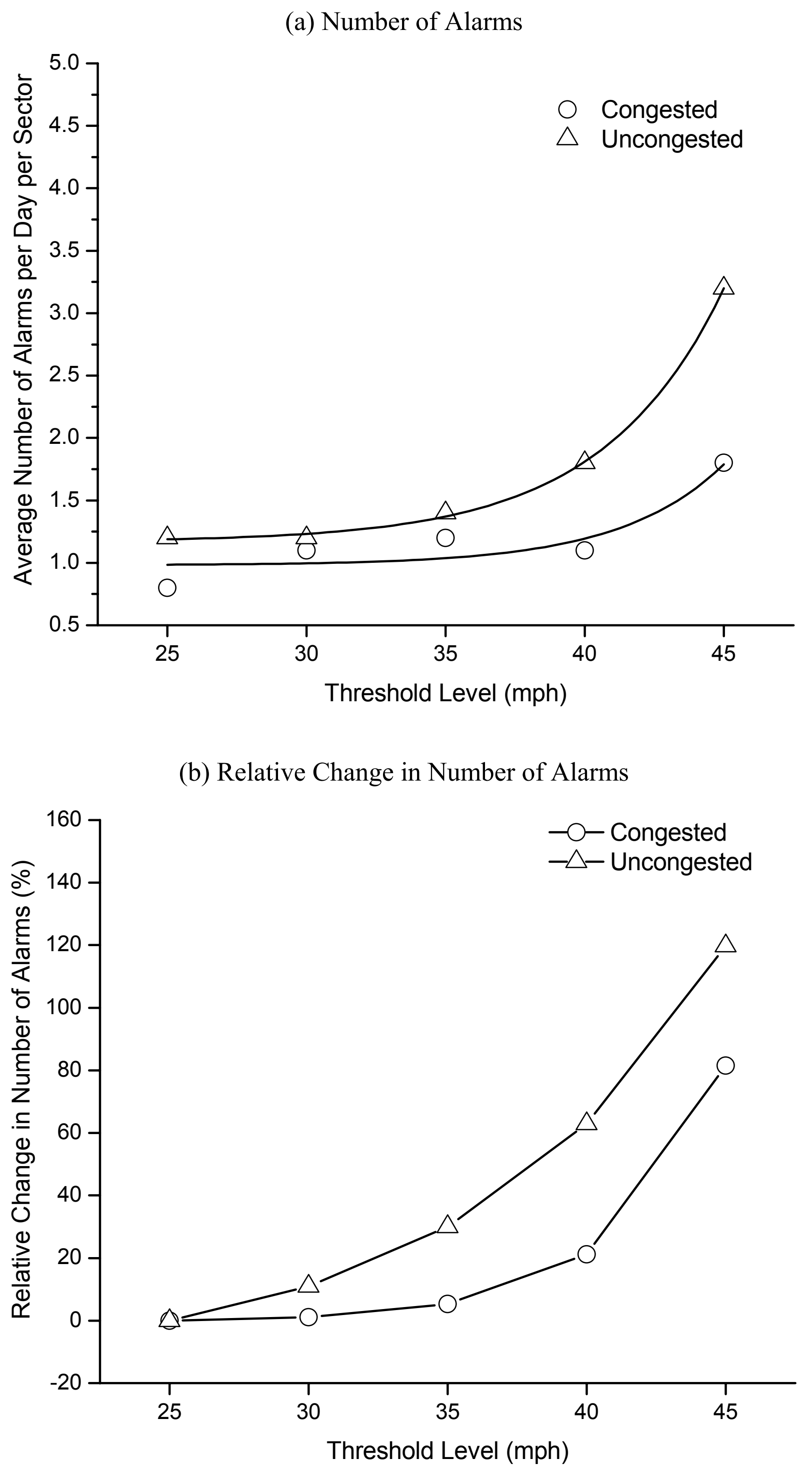
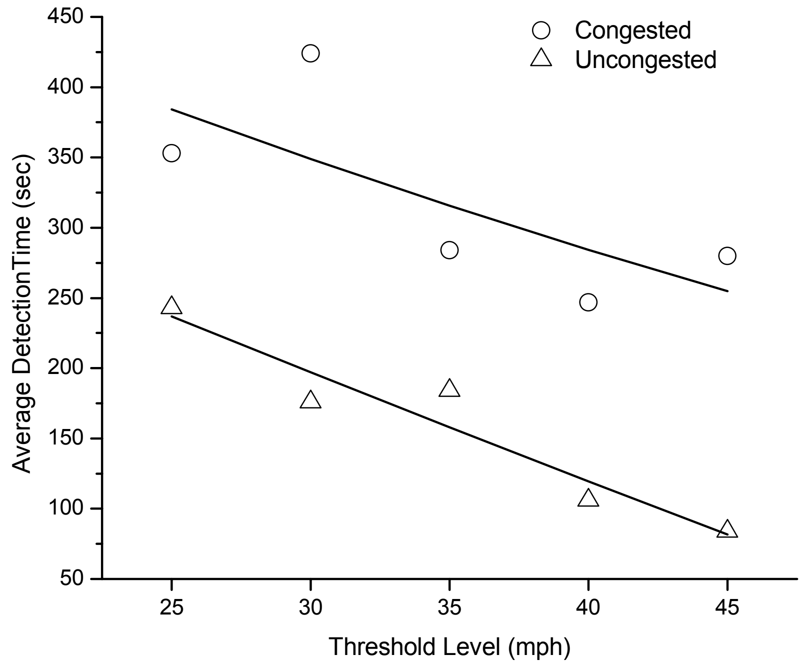


Figure 25. Impact of Congestion Levels on the Number of Alarms per Day.

(a) Detection Time



(b) Relative Change in Detection Time

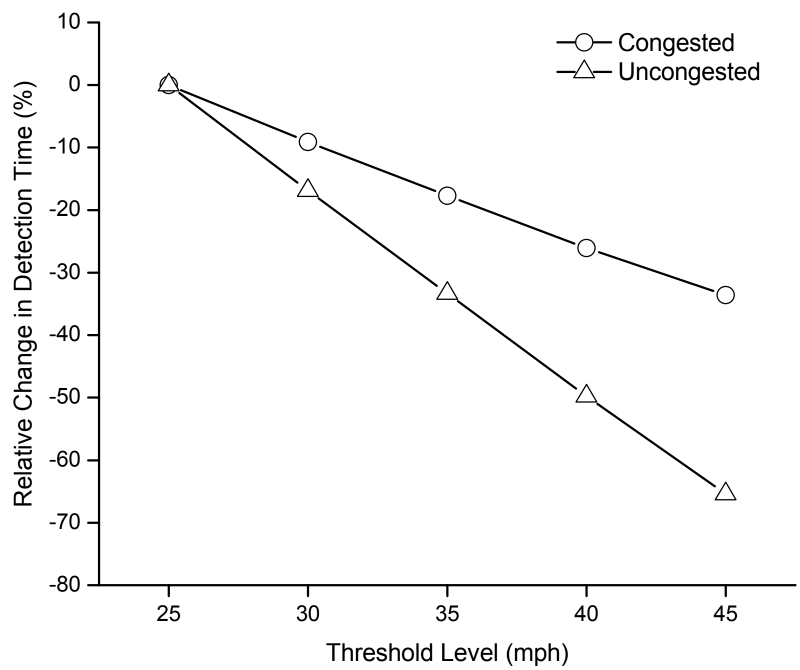


Figure 26. Impact of Congestion Levels on Average Detection Time.

An analysis of the data yields the following results:

- As Figure 24 shows, as the alarm threshold level increased, the average number of alarms increased and the average incident detection time decreased. The best-fit curves indicated a growth trend for number of alarms that was approximately exponential with relatively low growth rates for threshold levels between 25 mph and 35 mph and accelerated growth rates for threshold levels between 40 and 45 mph. The best-fit curves also indicated a trend for average incident detection time that was more linear than the trend for number of alarms. As a result, raising the alarm threshold from 25 mph to 45 mph resulted in a 140 percent increase in the average number of alarms while average incident detection times decreased by about 60 percent.
- The number of alarms in Figure 24 included both true alarms, i.e., alarms that corresponded to actual incidents on the ground, and false alarms, i.e., cases where there was an alarm but an incident did not actually happen. For simplicity, the researchers decided to keep both types of alarms in the analysis because of the realization that TMC operators need to react to both true and false alarms anyway. The percentage of false alarms increased as the alarm threshold increased. For example, for a 25 mph alarm threshold, 42 percent of the alarms were false alarms. For 35 mph, 47 percent of the alarms were false alarms. For 40 mph, 54 percent of the alarms were false alarms, and for 45 mph, 74 percent of the alarms were false alarms. This trend is an indication that one of the main effects of increasing the alarm threshold, particularly at the 40 mph or 45 mph level, would be to increase the number of false alarms.
- Figure 25 shows an impact of congestion levels on the number of alarms generated by the algorithm. In general, there were more alarms under uncongested traffic conditions than under congested conditions. However, readers should be aware that part of the reason was that uncongested traffic conditions spanned over a much longer period of time than congested traffic conditions. Overall, the ratio of number of alarms during uncongested periods to number of alarms during congested periods was similar to the corresponding ratio of incidents recorded during the first phase of the research, which was based on some 20,000 incident data points (*1*).
- Figure 26 shows there was a correlation between congestion levels and average incident detection times. In general, incident detection took considerably longer (between 60 and 100 percent longer) under congested traffic conditions than under uncongested traffic conditions. Increasing the alarm threshold level resulted in a decrease in average incident detection times that was much more noticeable under uncongested traffic conditions than under congested traffic conditions. As opposed to the number of alarms, there was not a decrease in incident detection time performance as the alarm threshold increased from 25 mph to 45 mph, suggesting that performance in terms of incident detection times would generally improve by increasing the alarm threshold level. Increase in performance would be more noticeable under uncongested conditions than under congested conditions.
- Average incident detection time was about 4.3 minutes (260 seconds) at the 25-mph alarm threshold level, with average values ranging from about 4 minutes for uncongested traffic to 6 minutes for congested traffic. While based on simulated runs, these numbers are not unreasonable considering the lag effect caused by the use of 2-minute moving average speeds and the time it takes for recorded speeds to physically drop below the alarm threshold. As Figure 27 shows, moving average speeds introduce a lag to the

speed profile that is typically between 1 and 1.5 minutes, but that can be as large as 2 minutes. How fast speeds drop below the threshold depends on traffic conditions and the nature of the incident. In the case of the speed profile shown in Figure 27, it took about 40 seconds for the speed to drop from 40 mph to less than 25 mph.

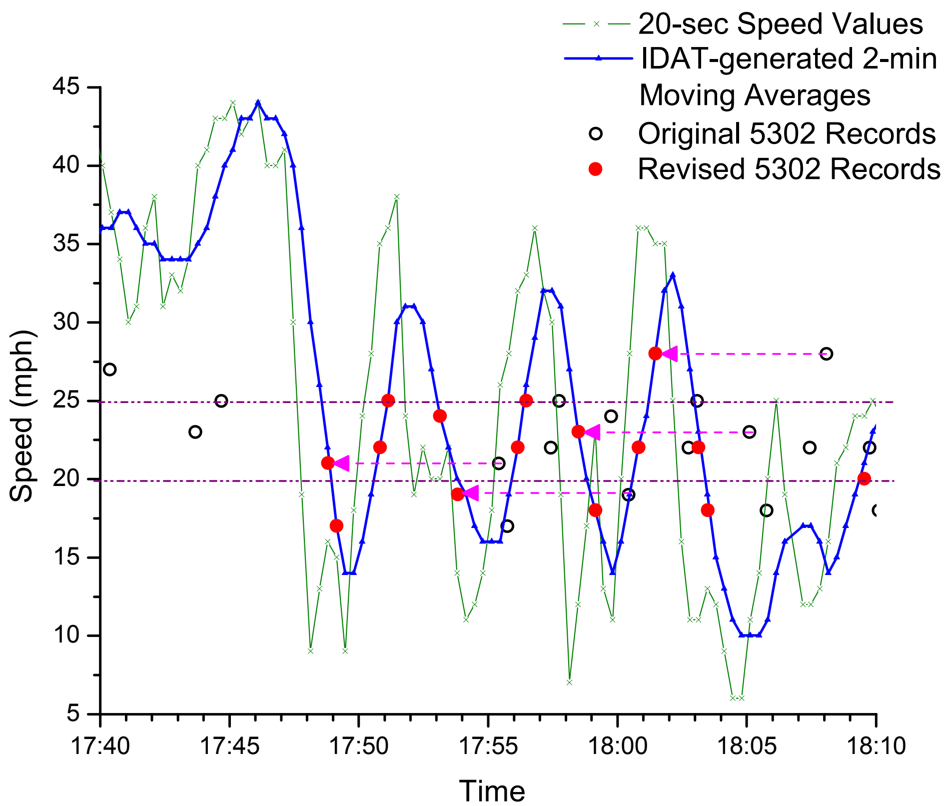


Figure 27. Comparison between 20-Second Speed Data and 2-Minute Moving Averages.

- Figure 27 also shows points associated with record type 5302, which the archived event logs use to keep track of instances when the moving average speeds cross the 25-mph and 20-mph thresholds while an alarm is active (i.e., any time after record type 5301 and before record type 5303). Notice the significant offset associated with the original points (approximately 6.5 minutes), which forced the recalculation of their time stamps to properly align the data with the 2-minute average profile. This observation is important because record types 5301 and 5303 (which provided the foundation for the matched event-incident dataset) are also affected by similar offset issues. Unfortunately, a review of several incident and event cases revealed that offsets were not consistent (they varied anywhere from -1 to 4 minutes, or, as in the case of Figure 27, 6.5 minutes), limiting the applicability of a generic offset correction factor.
- To obtain an estimate of actual average detection times, the researchers used 24 cases that included detected incidents from the 75 sample cases discussed in Chapter 3. The original list included 26 cases, but the researchers eliminated 2 cases that did not have incident times clearly identified. Since there is no record for actual incident time stamps, the researchers used archived lane data to identify approximate incident times and then

calculated incident detection times as the difference between event data alarm times (from the 5301 type records) and incident times. The resulting average incident detection time was 7.6 minutes. After taking into account the offset associated with the 5301 type records for each incident, the overall average detection time was 5.8 minutes.

IMPLEMENTATION STRATEGIES

From the previous section, 35 mph is the maximum minor alarm threshold value that could reasonably be expected without increasing the number of false alarms to an unacceptable level. Because increasing the minor alarm threshold can increase the number of alarms and decrease incident detection times, but the relative impact depends on traffic conditions, a number of implementation strategies might be possible:

1. **Do nothing.** In this scenario, TransGuide would maintain current alarm threshold values, i.e., 25 mph for minor alarms and 20 mph for major alarms, regardless of congestion level (or time of day). Obviously, this scenario would not result in any improvements in incident detection capabilities.
2. **Increase alarm thresholds during peak hours only.** In this scenario, TransGuide would increase the minor alarm threshold to 35 mph during peak hours while maintaining the current 25-mph minor alarm threshold for the remaining hours. This scenario would result in a modest increase in the number of alarms during peak hours and a significant (possibly around 20 percent) reduction in incident detection times. It may be worth noting that not necessarily all traffic during peak hours is congested. It is possible, therefore, that the net benefit for the entire network would be different—potentially higher as discussed below.
3. **Increase alarm thresholds during off-peak hours only.** In this scenario, TransGuide would increase the minor alarm threshold to 35 mph during off-peak hours (i.e., when most uncongested traffic occurs) while maintaining the current 25-mph minor alarm threshold during peak hours. This scenario would result in a significant increase in the number of alarms during off-peak hours (possibly around 30 percent) and a significant reduction in incident detection times (possibly around 35 percent). Since at 35 mph the number of false alarms would not be much higher than at the current 25-mph level, it is reasonable to expect that most of the increase in the number of alarms would be in the form of true alarms. Overall, Scenario 3 would yield a higher benefit than either Scenario 1 or 2.
4. **Increase alarm thresholds for the entire day.** In this scenario, TransGuide would increase the minor alarm threshold to 35 mph throughout the day, regardless of congestion level (or time of day period). This scenario would result in a significant increase in the number alarms during off-peak hours (possibly around 30 percent), a modest increase in the number of alarms during peak hours, a significant reduction in incident detection times during off-peak hours (possibly around 35 percent), and a significant reduction in incident detection times during peak hours (possibly around 20 percent). The overall impact would be a 10 percent increase in the number of alarms (at least half of which would be in the form of true alarms) and a 30 percent decrease in incident detection times. Overall, Scenario 4 is the most favorable.

A question that might surface at this point is whether it would be feasible (or even necessary) to use archived lane data to modify alarm thresholds for individual detectors. Presumably, one of the expected benefits of using archive lane data would be to develop the ability to customize alarm thresholds for different time periods based on the actual history of speed data (in the case of speed-trap detectors) or occupancy data (in the case of non speed-trap detectors). Since significant variations from free flow conditions typically occur during peak hours, the ability to customize alarm thresholds would serve its purpose primarily during those hours of the day, not during off-peak hours when traffic usually travels unimpeded. However, as the analysis above points out, modifying alarm thresholds would be more beneficial during off-peak hours than during peak hours. During off-peak hours, using archive lane data to customize alarm thresholds does not appear to be particularly promising because speeds during those time periods are for the most part free flow speeds and, therefore, predictable. From this perspective, there would not be considerable value in using archived data to customize alarm thresholds.

An argument sometimes presented to justify using archived data to customize alarm thresholds is that customized lower alarm threshold levels during peak hours can reduce the number of false alarms. It is true that using a lower alarm threshold during peak hours could result in a lower number of false alarms (according to the analysis above, some 5 percent when decreasing the alarm threshold from 35 to 25 mph or 32 percent when decreasing the alarm threshold from 45 to 25 mph). However, the trade-off would be considerably longer incident detection times (around 50 percent when decreasing the alarm threshold from 35 to 25 mph or 130 percent when decreasing the alarm threshold from 45 to 25 mph) and potentially a reduction in the number of true alarms. Overall, it appears that the benefit in terms of lower number of false alarms would not compensate the trade-off in terms of longer incident detection times. These reasons lead to the conclusion that not even during congested periods it would be necessary to either use archived lane data or use a minor alarm threshold lower than 35 mph.

CHAPTER 4. ARCHIVED ITS DATA QUALITY AND COMPLETENESS

This chapter summarizes the work completed to address quality control and completeness issues associated with archived ITS data at TransGuide. It includes a description of quality control tests, the results of an analysis conducted on some 3.4 billion 20-second lane detector data records from the TransGuide TMC, and a discussion of ITS data completeness issues.

DATA QUALITY ASSESSMENT

Methodology

While examining archived ITS data for the analysis during the first phase of the research, the researchers encountered situations such as erroneous data (e.g., incorrect scenario type characterization), missing data (in relation to the need to do data imputation), and comparability of ITS data to similar data sources (in relation to the normalization of the number of incidents using traffic volume data) (*1*). This prompted a data quality control analysis and the development of a preliminary set of quality control tests for detector data. During the second phase of the research, the researchers extended the first phase data quality control analysis and evaluated some 3.4 billion 20-second lane detector data records from March 2002 to April 2004. Using this large sample size was beneficial because it enabled the observation of quality control trends as TransGuide was installing new detectors on the ground.

Previous research has reported extensively on the need to implement quality control programs for ITS data to address critical issues such as suspicious or erroneous data, nature and extent of missing data, and accuracy and comparability of ITS data to similar data sources (*23*, *24*). The quality control tests developed as part of this research built on those efforts, although, by necessity, the quality control tests underwent modifications to suit the needs of the research. Table 6 shows a preliminary list of quality control tests the researchers evaluated. In general, the tests in Table 6 apply to two types of records: "valid" records and "abnormal" records. "Valid" records are records with valid volume and occupancy values but invalid "by design" speed values, e.g., -1 in the case of non-speed-trap detectors located on entrance and exit ramps, or zero in the case of main lane detectors when no vehicle has passed the detection zone during the detection time period. "Abnormal" records are records with "abnormal" combinations of speed, volume, and percent occupancy values (e.g., zero speed, zero volume, but larger than zero occupancy) that might result from causes such as faulty detectors or faulty LCU software logic. It may be worth noting that two types of LCU and associated software are currently operational at TransGuide: Naztec LCUs and TxDOT Traffic Operations Division (TRF) LCUs (also called Austin LCUs). It was therefore of interest to determine if different types of LCU produced different quality control test results. As a reference, Figure 28 shows the location of detectors controlled by Naztec LCUs and the location of detectors controlled by TRF LCUs.

Table 6. Preliminary Speed, Volume, and Occupancy Quality Control Tests.

Quality Control Name and Description		Test	Action
	First-Level Tests		
1a	**Record format error** Record is in incorrect format	Record is in incorrect format	Move record to dump file
1b	**Duplicate records**	Detector ID and date/time stamp are identical	Move duplicate record to dump file
	Second-Level Tests		
2a	**Extreme values** Unknown cause	Speed < -1 or Speed > 100 Volume < 0 or Volume > 3000 Occupancy < 0 or Occupancy > 100	Flag record
2b	**Entrance or exit ramp: Valid record**	Speed = -1 0 < Volume ≤ 3000 0 < Occupancy ≤ 100	Flag record Set Speed = <null>
2c	**Entrance or exit ramp: No vehicle present** No vehicle passed the detection zone during the detection time period	Speed = -1 Volume = 0 Occupancy = 0	Flag record Set Speed = <null>
2d	**Entrance or exit ramp: Volume is zero when occupancy is not zero**	Speed = -1 Volume = 0 0 < Occupancy ≤ 100	Flag record Set Speed = <null>
2e	**Entrance or exit ramp: Occupancy is zero when volume is not zero**	Speed = -1 0 < Volume ≤ 3000 Occupancy = 0	Flag record Set Speed = <null>
2f	**Main lane: No vehicle present** No vehicle passed the detection zone during the detection time period	Speed = 0 Volume = 0 Occupancy = 0	Flag record
2g	**Main lane: Speed and volume are zero when occupancy is not zero**	Speed = 0 Volume = 0 0 < Occupancy ≤ 100	Flag record
2h	**Main lane: Speed and occupancy are zero when volume is not zero**	Speed = 0 0 < Volume ≤ 3000 Occupancy = 0	Flag record
2i	**Main lane: Speed trap not functioning properly**	Speed = 0 0 < Volume ≤ 3000 0 < Occupancy ≤ 100	Flag record
2j	**Main lane: Volume and occupancy are zero when speed is not zero**	0 < Speed ≤ 100 Volume = 0 Occupancy = 0	Flag record
2k	**Main lane: Volume is zero when speed and occupancy are not zero**	0 < Speed ≤ 100 Volume = 0 0 < Occupancy ≤ 100	Flag record
2l	**Main lane: Occupancy is zero when speed and volume are not zero**	0 < Speed ≤ 100 0 < Volume ≤ 3000 Occupancy = 0	Flag record

Given the extremely large number of lane records in the database, the researchers found it computationally more efficient to add quality control flag values to an indexed quality control field as the script was populating the lane data table rather than running queries after the fact using volume, speed, and occupancy data to assess quality control flag values. Using an indexed quality control field accelerated the query building process considerably, but the downside was that the researchers had to make preliminary assumptions with respect to certain thresholds. For example, in the case of the 3,000 volume threshold, the researchers examined sample lane detector data files and found a few cases where there were gaps in the data and volume data after the gaps that were "too" large for what would be typical of a 20 second period (suggesting that the LCU apparently had not reset the volume counter). Since there was no way of knowing ahead of time if the volume data would be necessarily invalid, the decision was to use a large enough threshold and examine the results once all the quality control flag data were in the database. Similarly, in the case of the 100 mph speed threshold, since there was no way of knowing ahead of time if a speed value between 90 and 100 mph was an anomaly or simply the result of extremely aggressive driving, the decision was to use a large enough speed threshold and examine the results after all the data were in the database. The following section describes the results of these analyses.

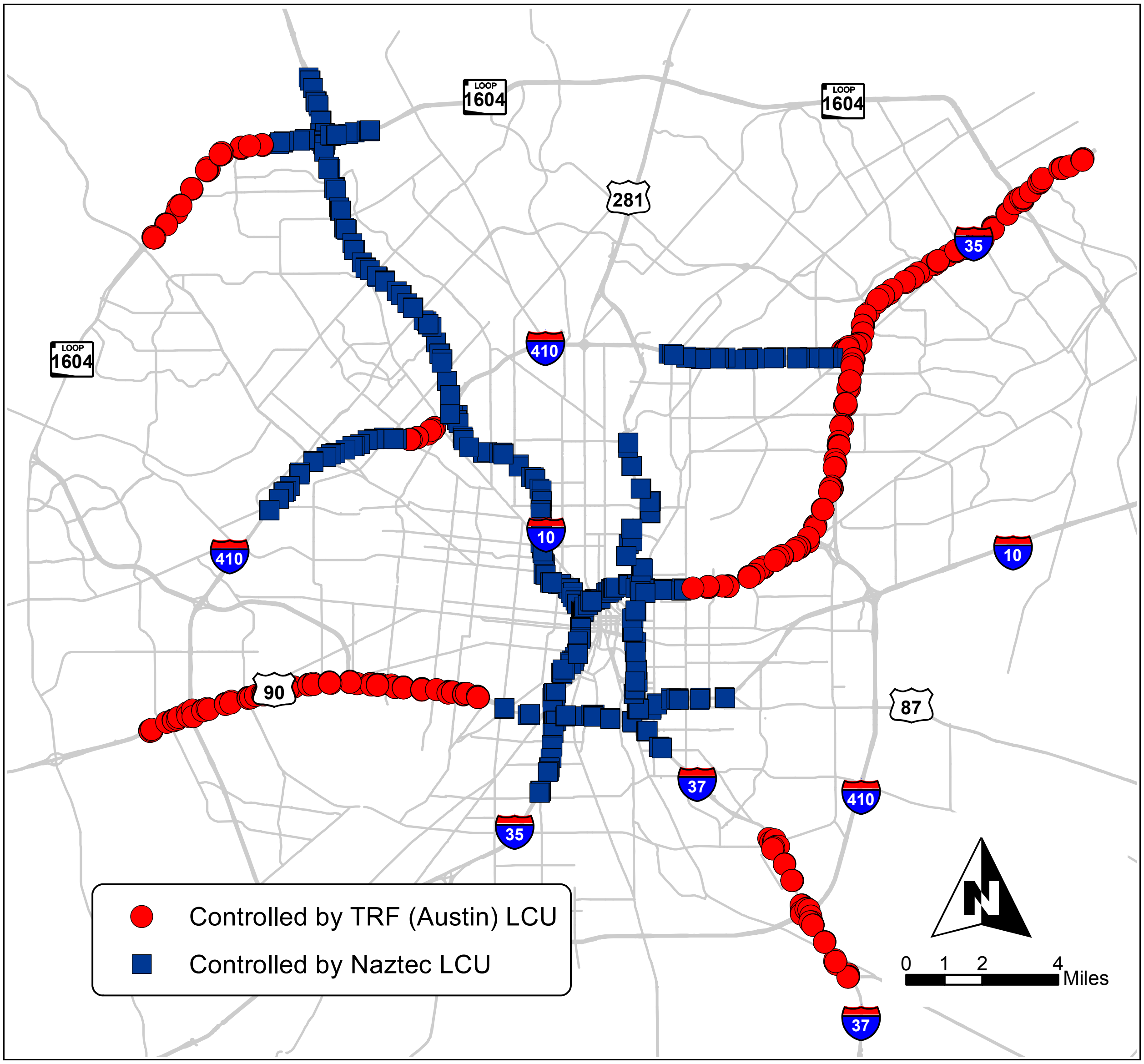


Figure 28. Detectors Controlled by Naztec LCUs and TRF LCUs.

Quality Control Analysis

First-Level Tests

Because of the structure of the flat file lane data archive (Figure 1), there were no records that failed quality control test 1a (record format errors). In the case of quality control test 1b (duplicate records), the analysis detected a few instances of duplicate detector ID and date/time stamp records whenever the time changed from central daylight time (CDT) to central standard time (CST) in October. Internally, TransGuide uses the Unix time function to assign unique time stamps to events. However, the lane data archive does not use the Unix time function, relying instead on local date/time stamps (Figure 4). As a result, when time changes back one hour at 2:00 AM the last Sunday in October, the time stamps of the records following that change begin at 1:00 AM. Since the archive already contains lane records with time stamps beginning at 1:00 AM (from the previous hour), there is a very good chance that some, if not all, of the new records will contain duplicate detector ID/time stamp information. In practice, this phenomenon does not always happen because time stamps sometimes fluctuate by a second or two (i.e., the time interval is not always exactly 20 seconds). If this happens when the time changes one hour, the database will not exactly contain duplicate records—although there is still a problem because it can be very difficult to sort the records and, consequently, recreate the time series reliably.

Interestingly, another implication of the yearly time change is that in April, when time changes forward one hour from CST to CDT, there is a one-hour gap in the lane data archive.

Second-Level Tests

Table 7 summarizes the second-level quality control tests on some 3.4 billion 20-second speed, volume, and occupancy data from March 1, 2002, through April 30, 2004. An analysis of the data yields the following results:

- Some 1.6 billion speed, volume, and occupancy records had a quality control flag, accounting for nearly 48 percent of the 3.4 billion lane data record set. Approximately 1.5 billion flagged records were "valid" records and the remaining 126 million flagged records were "abnormal" records. The "valid" flagged records had a speed value of -1 or 0, but the volume and occupancy values were most likely valid. A total of 126 million "abnormal" records translate to an overall "abnormal" record rate of about 3.7 percent.
- Of the 126 million "abnormal" records, the vast majority (106 million or 84 percent) had flag 2j (speed > 0, volume = 0, and occupancy = 0). The remaining 11 flags accounted for 16-percent of the "abnormal" records.
- There were significant differences between TRF LCU records and Naztec LCU records. For example, even though 32 percent of LCUs were TRF LCUs, the percent of "abnormal" records associated with detectors controlled by TRF LCUs was 84 percent. The vast majority of these records had flag 2j (speed > 0, volume = 0, and occupancy = 0), with practically no records under the other flag categories (except flag 2a). In contrast, Naztec LCU records, even though they were the minority, had representation in every single flag category. Some 54 percent of Naztec LCU records had flag 2i (speed = 0, volume > 0, occupancy > 0). Interestingly, while only 433 Naztec LCU records had

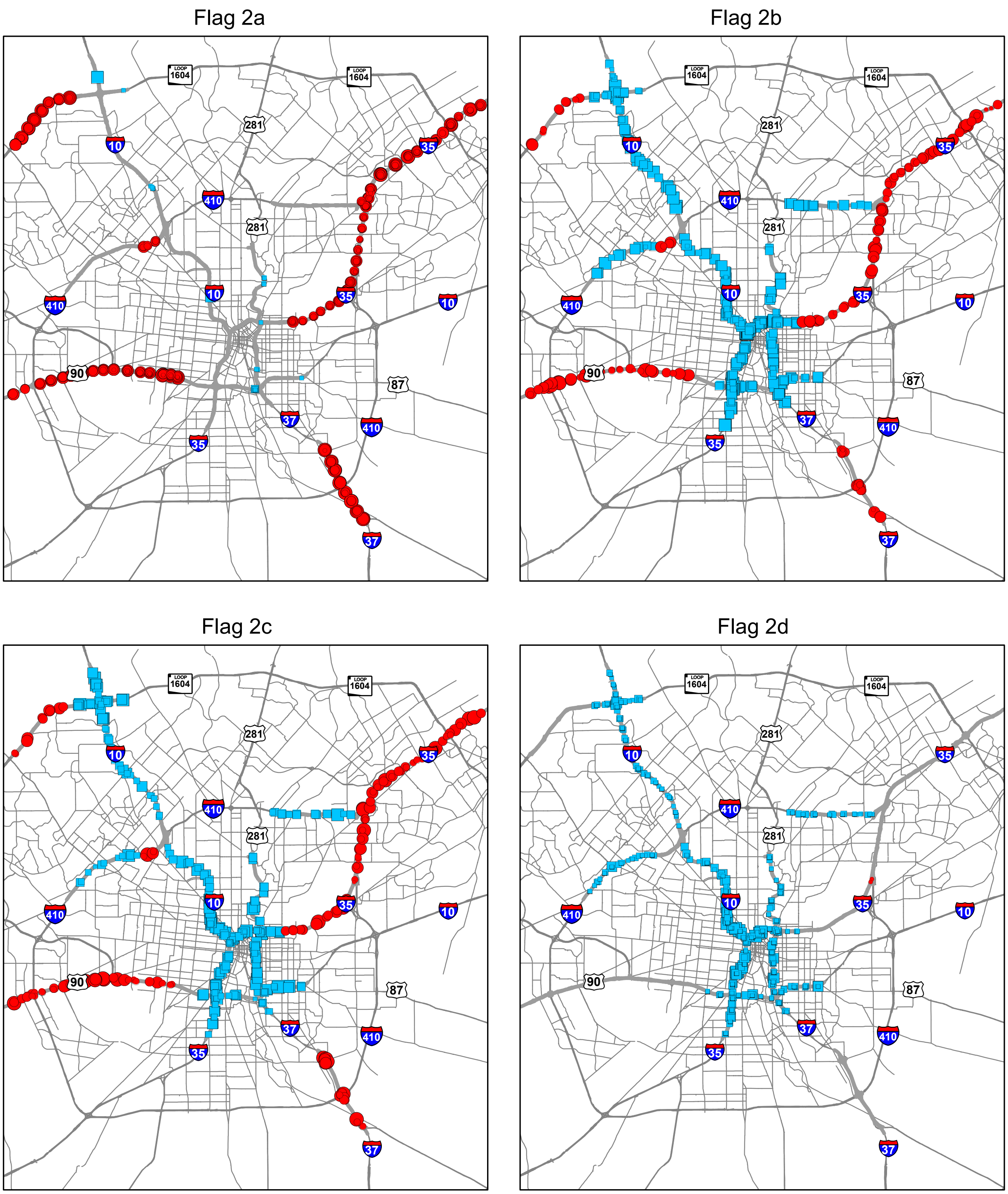
flag 2a (extreme values: speed > 100, volume > 3,000, occupancy > 100), more than 157,000 TRF LCU records had that flag.

- The differences between TRF LCU records and Naztec LCU records point to other important differences between the two types of LCU software. For example, there were 369 million Naztec LCU records with flag 2f (speed = 0, volume = 0, and occupancy = 0). In contrast, there was not a single TRF LCU record with that flag. The reason is that the TRF LCU software does not generate lane detector records if no vehicles cross the associated detectors during the 20-second recording interval. While the result is a more compact lane data repository, it makes it practically impossible to recreate what actually happened in the field, since a missing record does not automatically mean that no vehicles crossed the detectors. To address this limitation, it would be advisable to modify the TRF LCU software so that it can generate null speed (to avoid problems associated with the use of zero speeds), zero volume, and zero occupancy records when no vehicles are present during the 20-second recording interval.

Table 7. Summary of 20-Second Lane Records Flagged from March 2002 to April 2004.

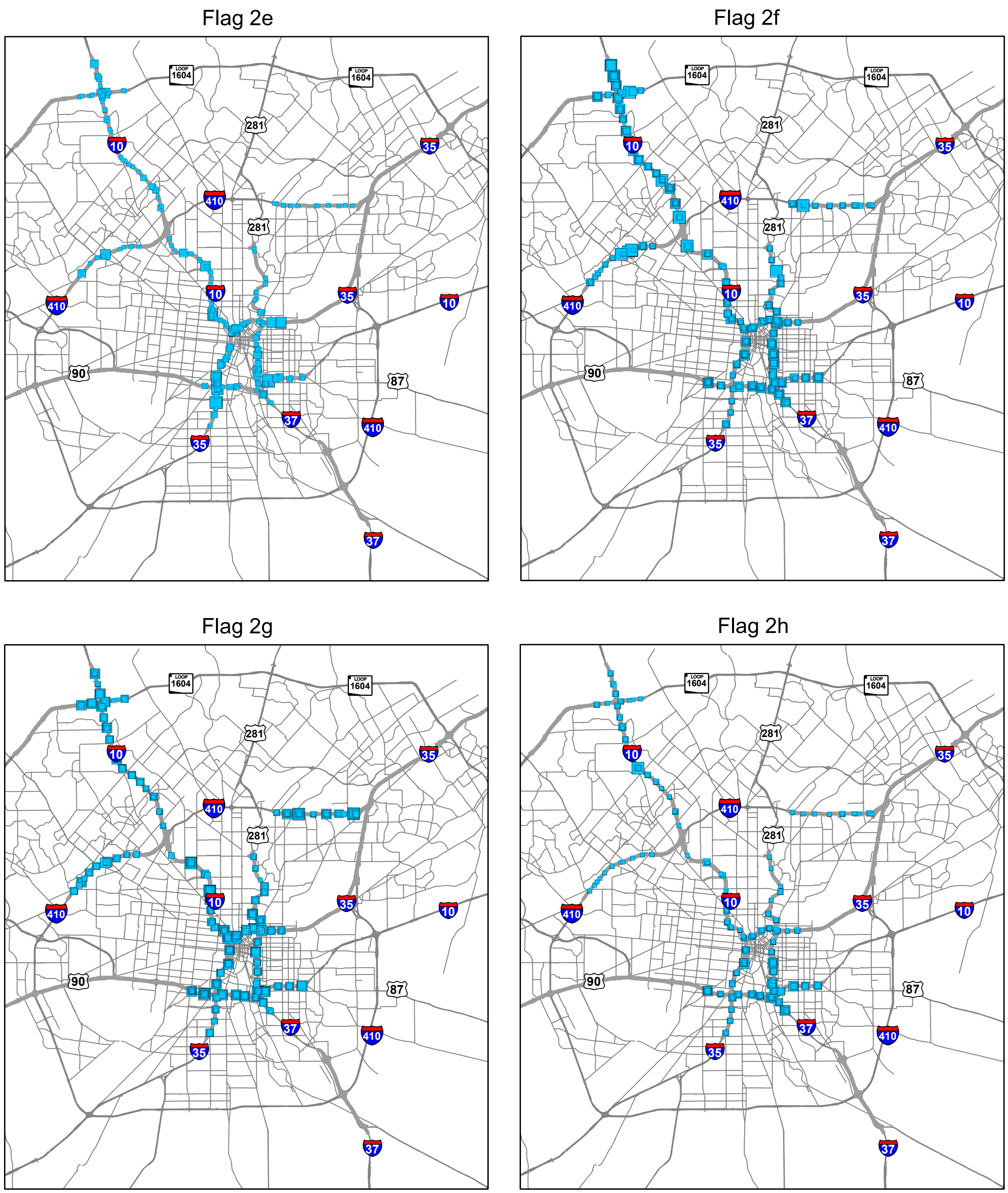
Quality Control Flag	TRF LCU				Naztec LCU			
	"Valid" Records		"Abnormal" Records		"Valid" Records		"Abnormal" Records	
2a			157,470	<1%			433	<1%
2b	172,315,686	17%			464,394,214	20%		
2c	186,139,423	18%			295,773,277	13%		
2d			2	<1%			1,510,386	<1%
2e							2,211,955	<1%
2f					368,902,192	16%		
2g							1,563,112	<1%
2h							1,946,840	<1%
2i							10,935,197	<1%
2j			105,533,470	10%			226,056	<1%
2k			46	<1%			24,369	<1%
2l							1,917,923	<1%
Subtotal	**358,455,109**	**34%**	**105,690,988**	**10%**	**1,129,069,683**	**48%**	**20,336,271**	**1%**
Total	**464,146,097**	**45%**			**1,149,405,954**	**49%**		
Total Flags	**1,613,552,051**	**48%**						
Lane Records	**1,042,089,780**				**2,351,336,786**			
Total Lane Records	**3,393,426,566**							

The researchers also examined spatial trends in the distribution of quality control flags. As Figure 29, Figure 30, and Figure 31 show, in several cases the spatial distribution was roughly uniform, although there were some significant exceptions. For example, in the case of flags 2d, 2e, 2g, 2h, 2i, and 2l, there was a higher concentration of flagged records in the central part of town than on or outside Loop 410. Likewise, in the case of flag 2j, there was a higher concentration of flagged records on US 90 west of downtown than in other parts of town.



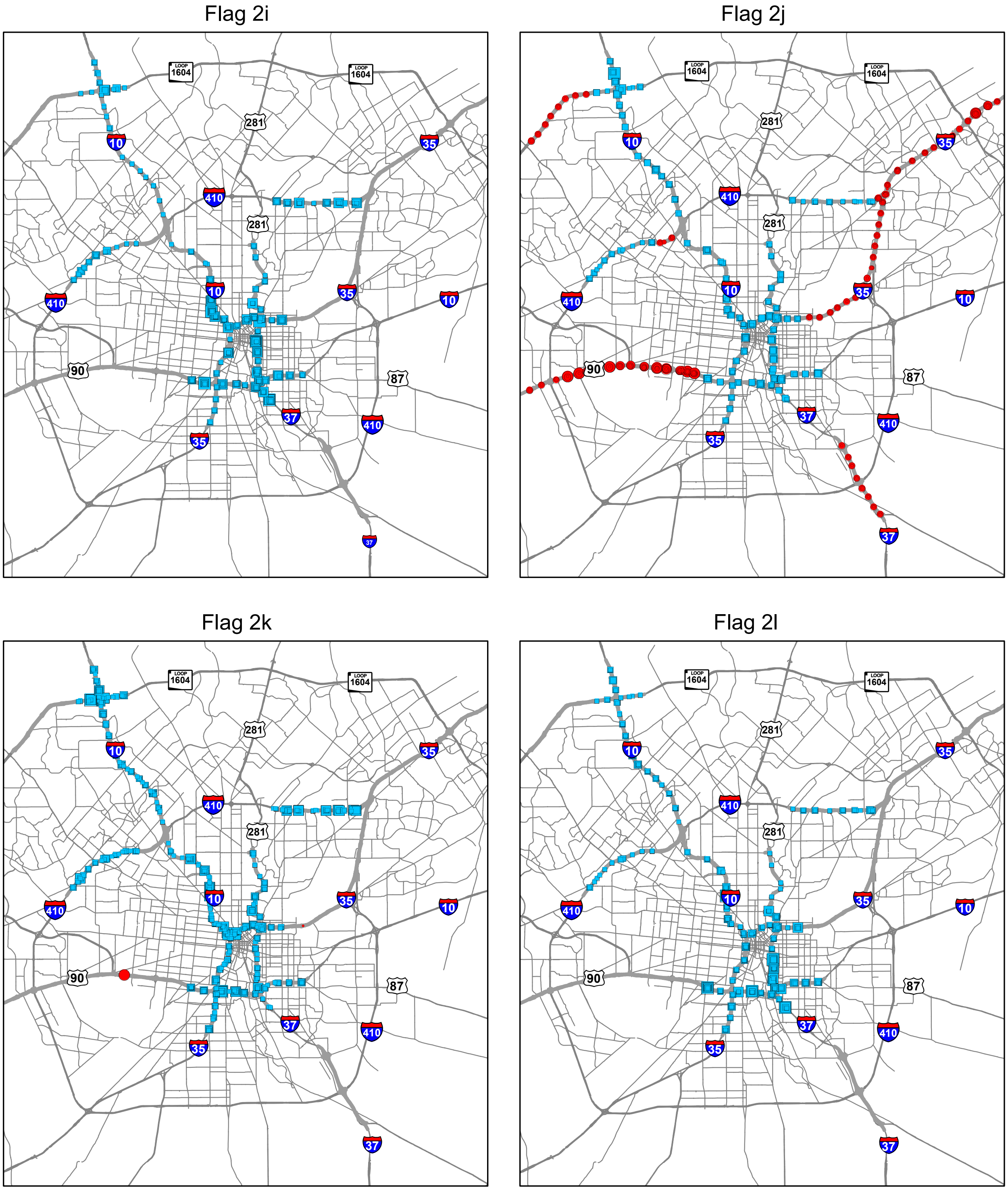
Blue squares represent Naztec LCU detectors. Red circles represent TRF LCU detectors.
Symbols represent ratio of number of flagged records to number of potential records per detector. Symbol sizes are relative to each map and are not necessarily comparable across maps.

Figure 29. Spatial Distribution of Quality Control Records (Flags 2a – 2d).



Blue squares represent Naztec LCU detectors. Red circles represent TRF LCU detectors.
Symbols represent ratio of number of flagged records to number of potential records per detector. Symbol sizes are relative to each map and are not necessarily comparable across maps.

Figure 30. Spatial Distribution of Quality Control Records (Flags 2e – 2h).



Blue squares represent Naztec LCU detectors. Red circles represent TRF LCU detectors.
Symbols represent ratio of number of flagged records to number of potential records per detector. Symbol sizes are relative to each map and are not necessarily comparable across maps.

Figure 31. Spatial Distribution of Quality Control Records (Flags 2i – 2l).

In addition to spatial variations, the researchers examined temporal variations in the distribution of quality control flags. As Figure 32 shows, the distribution of quality control flags varied widely throughout the day. In most cases, the highest concentration of flagged records occurred at night, when there was relatively little traffic and, consequently, there was a higher chance either for time intervals with no vehicles crossing the detectors (e.g., 2f) or for isolated detector readings producing abnormal speed, volume, and occupancy combinations (e.g., flags 2g, 2h, 2i, 2j, 2k, and 2l). Not surprisingly, most records associated with flag 2b (which were valid records, except the speed was recorded as -1) happened during the day, when most of the traffic took place.

TRF LCU flag 2a records yielded an interesting pattern characterized by a peak at about 6 AM, a minor dip at about 8 AM, a second peak at 10 AM, a steady decline until about 5 PM, a third peak at 9 PM, and a decline until 4 AM. Naztec LCU flag 2a records produced a completely different pattern, but the sample size was very small, making the observed trend unreliable. The trend for TRF LCU flag 2a records was interesting because flag 2a involved extreme value records (speed > 100 mph, volume > 3,000, occupancy > 100 percent). Anecdotally, TransGuide officials have indicated that during evening hours, TRF LCUs sometimes produce records with extremely high values, particularly speeds. The flag 2a trend in Figure 32 confirms that observation, although it also points to other times of the day when extreme value records are also relatively high (e.g., from 6 – 10 AM).

To assess the feasibility of the quality control thresholds (100 mph for speed, 3,000 for volume, and 100 percent for occupancy rate), the researchers analyzed the speed, volume, and occupancy data distributions of the 3.4 billion 20-second lane records in the database (actually 2.2 billion in the case of speed data records because not all lane records were from speed-trap detectors). Figure 33 summarizes the results of the analysis. Figure 33a shows that most records were between 1 and 75 mph (typical of freeway driving conditions), with a rapidly decreasing trend between 75 and 93 mph, and a few lingering records above 93 mph. Three data points stood out: 0, 87, and 94 mph. At first sight, the trend between 1 and 5 mph would suggest the number of zero speed records to be around 100,000 (presumably, actual number of records where vehicles physically stopped). However, it is more likely that the vast majority of the nearly 400 million zero speed records—all of them associated with Naztec LCUs—resulted from cases where no vehicles crossed the detectors. In the case of the 87- and 94-mph records, the trends clearly showed an anomaly, but it was unclear from the analysis what could have caused that anomaly. Overall, Figure 33a suggests a reasonable upper speed threshold of about 93 mph (which would translate to about 0.01 percent of records exceeding that threshold). By comparison, the equivalent percentages for 90 and 100 mph thresholds are 0.1 and 0.007 percent, respectively.

Figure 33b shows that most records had volumes lower than 10 vehicles, with a rapidly decreasing trend between 10 and 300, and a few lingering records above 300. By and large, Figure 33b suggests a reasonable upper volume threshold to be around 18 (which would translate to 0.1 percent of records exceeding that threshold). By comparison, the equivalent percentage for 37, 100, and 3,000 would be 0.01, 0.002, and 0.000003 percent, respectively. It may be worth noting that 18 vehicles over a 20-second period are equivalent to a flow rate of 3,240 vehicles per hour, which is higher than the maximum hourly flow rate of 2,400 passenger cars per hour per lane normally associated with freeway traffic at capacity.

Figure 33c shows that, with the exception of a handful of records, practically all records had occupancy rates less than or equal to 100 percent. Interestingly, there was a jump between 99 and 100 percent, which the analysis could not explain. Overall, Figure 33c suggests that a reasonable upper occupancy threshold is 99 percent (which would translate to 0.006 percent of records exceeding that threshold). By comparison, the equivalent percentage for 100 percent occupancy rate is 0.0000004 percent.

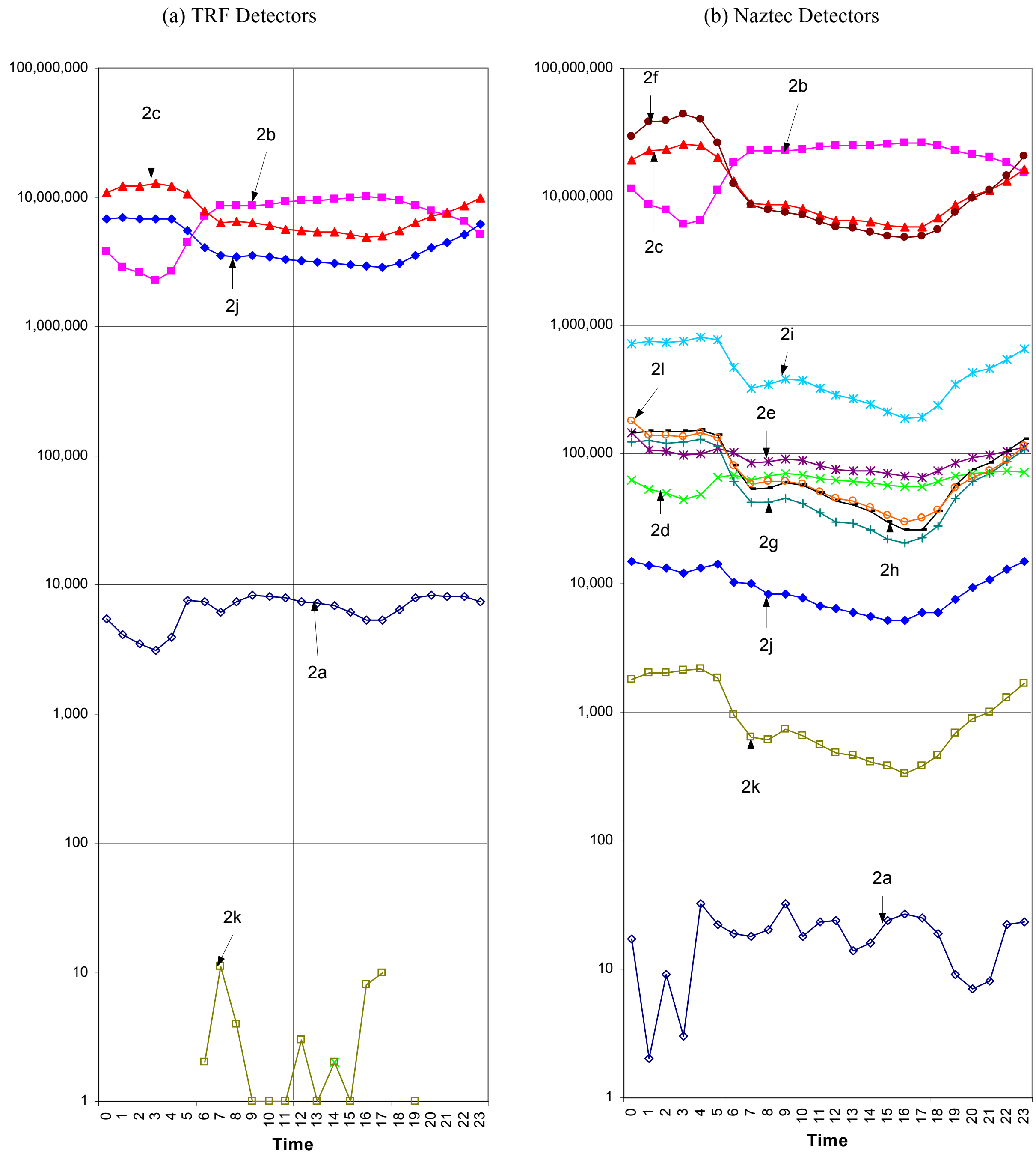
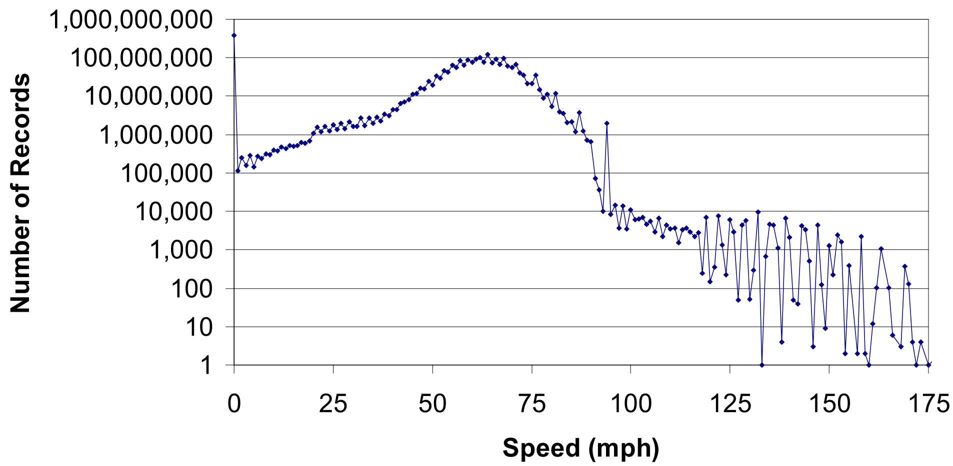
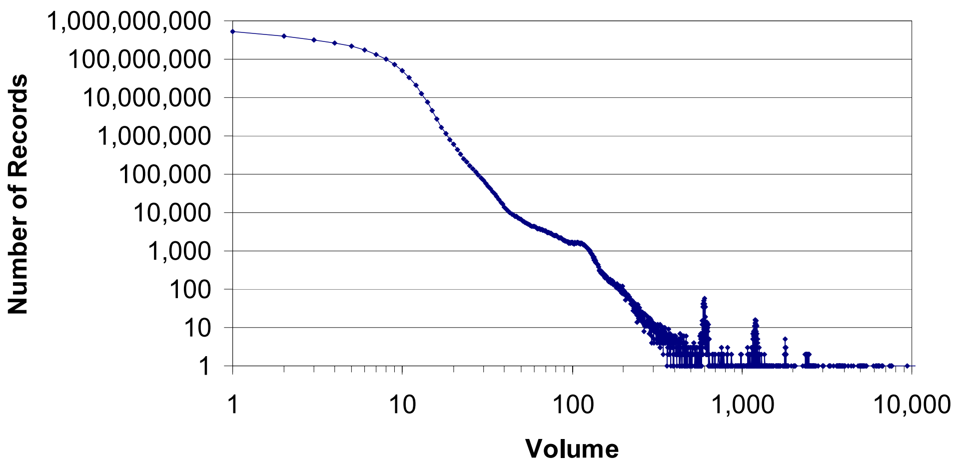


Figure 32. Temporal Distribution of Quality Control Records.

(a) Speed Data Distribution



(b) Volume Data Distribution



(c) Occupancy Data Distribution

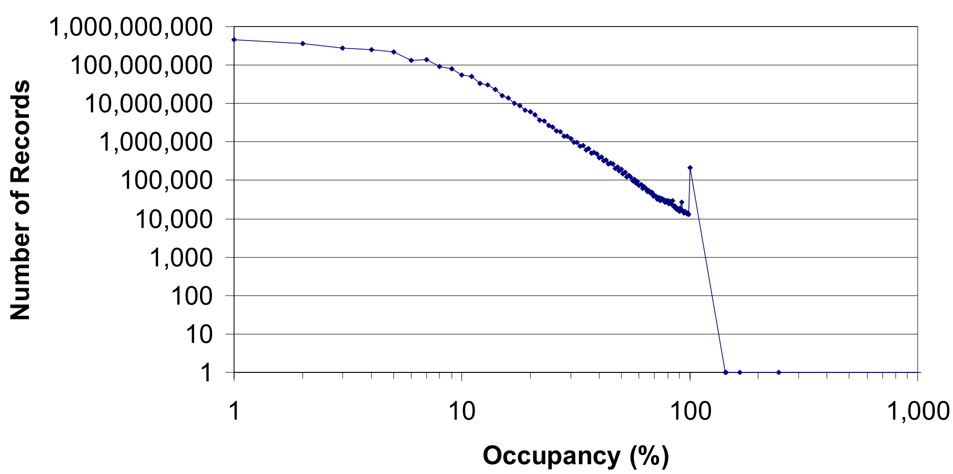


Figure 33. Speed, Volume, and Occupancy Data Distributions.

DATA COMPLETENESS ASSESSMENT

The researchers conducted a data completeness analysis to complement the quality control analysis presented in the previous section. The completeness analysis included an aggregate evaluation of completeness by LCU server as well as a detailed evaluation of completeness at the individual detector level.

The purpose of the aggregate completeness analysis at the LCU server level was to determine any trends that could be attributed to system-wide causes rather than individual detectors. TransGuide currently operates six LCU servers (Table 8). With the exception of Server 6, which started processing detector data in November 2003, the remaining servers were supposed to be operational and process data during the 792-day analysis period from March 1, 2002, through April 30, 2004. Table 8 shows there were several days during this analysis period when the archive did not include any data. Overall, the completeness rate—measured as number of days with data to total number of potential days with data—varied from 95 to 100 percent.

Table 8. Summary Data Completeness Results by LCU Server.

Statistic	Server 0	Server 1	Server 2	Server 3	Server 4	Server 5	Server 6
Count	792	790	752	781	781	781	160
Max No. of Days	792	792	792	792	792	792	168
Days with No Data	0	2	40	11	11	11	8
Completeness Rate	100%	99.7%	94.9%	98.6%	98.6%	98.6%	95.2%
No. of Records	744,776,647	995,158,370	614,461,919	393,337,241	294,779,400	292,259,242	62,881,334
Daily Median	938,913	1,269,771	826,466	533,595	402,643	362,357	399,601
Daily Average	940,375	1,259,694	817,104	503,633	377,438	374,212	393,008
Daily Maximum	1,131,003	1,378,409	1,181,130	638,521	477,290	469,178	423,382
Daily Minimum	504,519	103,095	53,609	199,558	118,722	139,183	146,807
Standard Deviation	104,864	78,724	135,003	92,929	82,284	54,524	30,771
Coefficient of Variation	11.2%	6.2%	16.5%	18.4%	21.8%	14.6%	7.8%

Table 8 also shows a wide range in the total number of records per day associated with each server, suggesting the possibility of large gaps in the data. To measure this effect, the researchers looked at the history of records associated with individual lane detectors over the 792-day analysis period. For each detector, the researchers determined the earliest date/time stamp with data and the latest date/time stamp with data to calculate the maximum number of potential records that could be associated with that detector. The researchers also counted the effective number of records for each detector and then calculated a completeness rate.

As Figure 34 shows, very few detectors had high completeness rates. For example, only about 35 percent of detectors had a completeness rate of 95 percent or higher. Likewise, very few detectors had low completeness rates. For example, only about 10 percent of detectors had a completeness rate of 50 percent or lower. On average, the completeness rate for all detectors was 80 percent. There was a significant difference between TRF LCU detectors and Naztec LCU detectors. The overall completeness rate for Naztec LCU detectors was higher than the overall completeness rate for TRF LCUs (84 percent versus 71 percent, respectively). This difference is reasonable considering that TRF LCUs do not generate records when vehicles do not cross the detectors during the 20-second polling period.

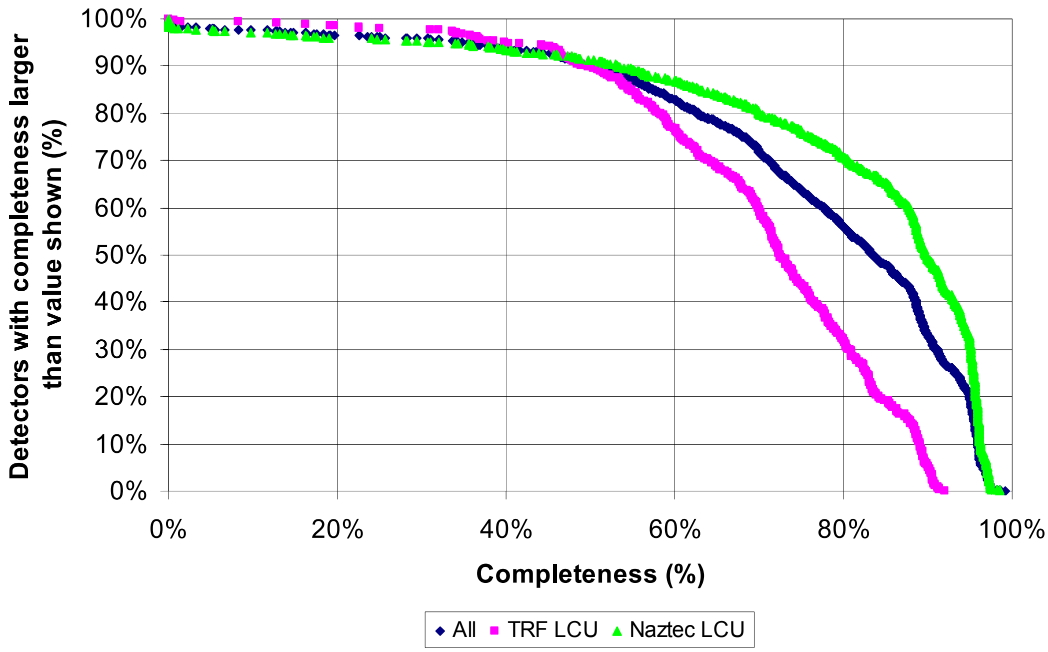


Figure 34. Detector Data Completeness Summary.

As Figure 35 shows, the spatial distribution of completeness rates by detector was not uniform. In general, there was more variability in the case of Naztec LCUs than in the case of TRF LCUs, with a larger number of Naztect LCU detectors having very low completeness rates. This result is consistent with the trend shown in Figure 34 for detectors having completeness rates less than 50 percent. Even though the spatial distribution of completeness rates was more uniform for TRF LCUs, there were a few detectors, particularly on US 90 west of downtown, that had very low completeness rates.

Data Completeness during Incidents

Following the characterization of sample cases in Chapter 3, the researchers noticed what appeared to be more gaps in the lane data during incidents than during normal traffic conditions. This observation prompted an analysis to test the data gap hypothesis. For the analysis, the researchers calculated completeness rates for incident and non-incident periods using 53 cases that included detected incidents and false negatives from the 75 sample cases discussed in Chapter 3. The original list included 60 cases, but the researchers eliminated seven cases that had missing data before the incidents happened. To calculate completeness rates during incidents, the researchers included all the lanes associated with the sector where the incident occurred as well as the lanes associated with the upstream and downstream sectors. To calculate completeness rates during non-incident periods, the researchers used the time windows that characterized the selected incidents (essentially incident start time and incident end time) and queried the archived database to gather several days' worth of non-incident lane data before and after the dates when the incidents occurred.

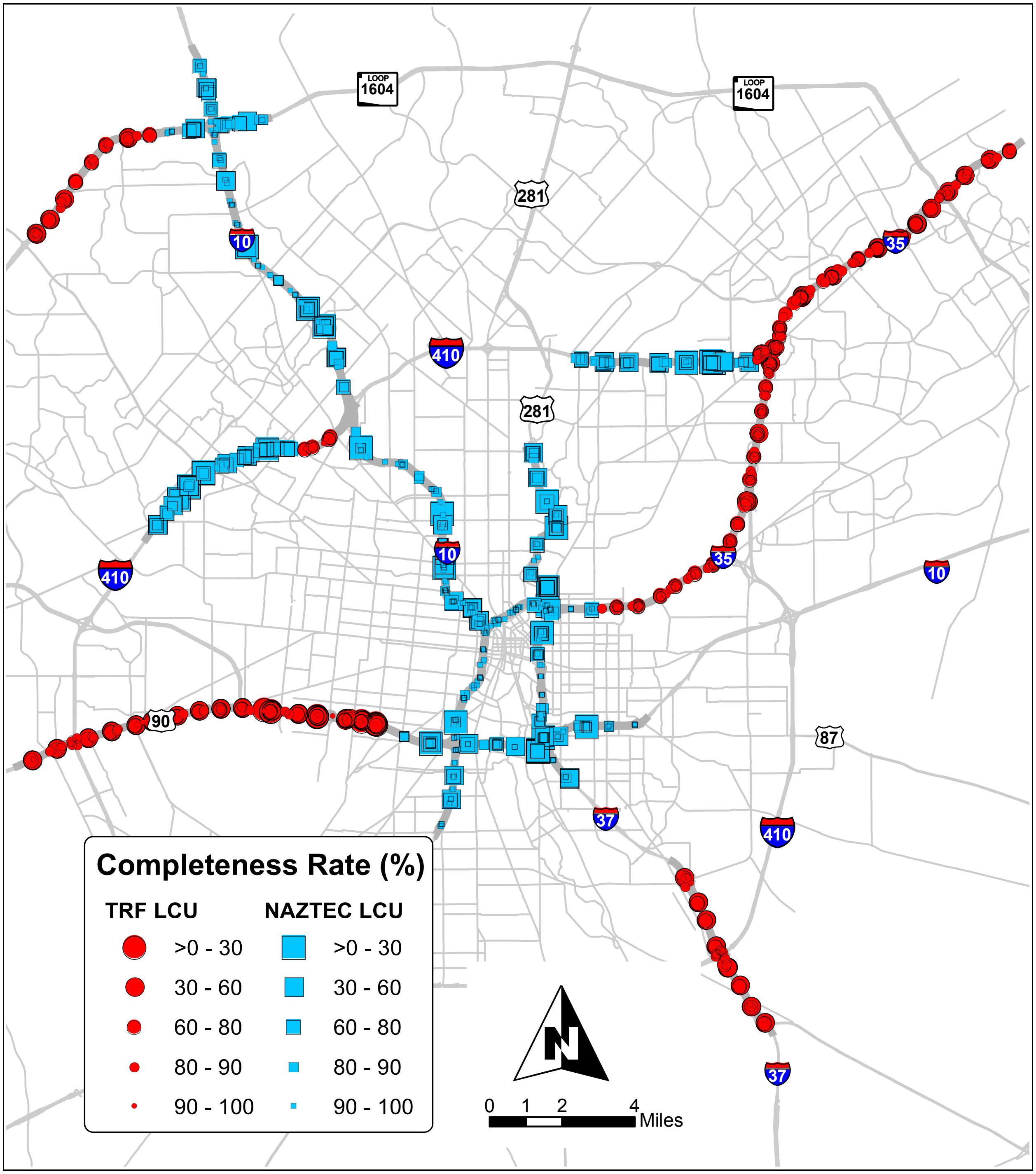


Figure 35. Spatial Distribution of Completeness Rates.

Table 9 summarizes the results. Because TRF LCUs and Naztec LCUs treat zero-speed, zero-occupancy data differently, the results include two tabulations: one assuming that zero-speed, zero-occupancy data were *valid*, and another one assuming that zero-speed, zero-occupancy data were *invalid*. For 27 cases (or about 51 percent), the completeness rate during incidents was significantly lower than during non-incident periods. This result assumed zero-speed, zero-occupancy data to be valid. When zero-speed, zero-occupancy data were treated as invalid data, and therefore excluded from the analysis, it turned out that 32 out of 53 cases (or about 60 percent) had a completeness rate during incidents that was significantly lower than during non-incident periods. Notice that TRF LCUs resulted in a higher proportion of cases with completeness rates during incidents lower than during non-incident periods (e.g., 65 percent versus 40 percent, assuming zero-speed, zero-occupancy data to be valid).

Table 9. Data Completeness Results for Incident vs. Non-Incident Periods.

(a) For Naztec LCUs, zero-speed, zero occupancy observations are counted as *valid* observations

LCU Type	Total No. of Cases	No. of Cases with Completeness Rate during Incidents lower than Non-Incident Periods	Ratio
TRF LCU	23	15	65%
Naztec LCU	30	12	40%
Both LCU Types	53	27	51%

(b) For Naztec LCUs, zero-speed, zero occupancy observations are counted as *invalid* observations

LCU Type	Total No. of Cases	No. of Cases with Completeness Rate during Incidents lower than Non-Incident Periods	Ratio
TRF LCU	23	15	65%
Naztec LCU	30	17	57%
Both LCU Types	53	32	60%

The researchers also calculated overall completeness rates by merging all the lane data for the 53 sample cases into two separate groups: one for incidents and the other one for non-incident periods. The overall completeness rate for incident periods was 64 percent as opposed to 78 percent for non-incident periods (which was very close to the 80 percent overall completeness rate for all detectors, as the previous section documented). After excluding zero-speed, zero-occupancy from the analysis, the difference was even greater: 52 percent for incident periods versus 70 percent for non-incident periods. Unfortunately, during the course of the research it was not possible to explore in detail potential reasons that could explain why data gaps seemed to be more prevalent during incidents. It is possible that traffic flow during incidents can be erratic in ways that the LCU software cannot process properly, causing it to simply reject more records than normal.

Influence of Lane Closures on ITS Data Completeness

TransGuide maintains a database to keep track of lane closure events, which includes data elements such as route name and direction, beginning and ending crossing streets, beginning and ending dates and times, nature of the work being performed, number of lanes affected, associated ramps affected, and detour information. The lane closure database also includes information in the form of latitude-longitude pairs associated with individual lane closure events. Figure 36 shows a sample of records from that database.

The researchers attempted to link the lane closure database to the prototype geodatabase of ITS features developed in this project in an effort to establish a correlation between lane closure data and potential gaps in the lane data repository. Manual matching of individual lane closure events and lane data records is certainly possible because, with the route information and beginning and ending crossing streets from the lane closure database, it would be possible to manually identify the TransGuide sectors of interest. With the beginning and ending dates and time stamps, it would then be possible to run a query on the lane data archive to gather all the corresponding lane data records and determine completeness rates. For example, for the lane closure event on US 90 between SH 151 and Acme Road (Figure 36), the researchers identified five lane detectors on sector SECT-0090W-568.156 that could be affected by the closure. For the lane closure duration (from August 5, 2002, at 10:00 PM to August 4, 2002 at 5:00 AM), they gathered all the corresponding lane detector speed, volume, and occupancy records and calculated the corresponding completeness rates: 25, 43, 42, 43, and 25 percent for lanes 1 (left most), 2, 3, 4, and 5 (right most), respectively. The average completeness rate for the sector was 36 percent. Notice that, according to the lane closure database, the lane closure affected the two right lanes. However, the lane data archive showed a completeness rate for those lanes that was similar to the other three lanes.

lane_closures : Table

Latitude	Longitude	Highway	Direction	Start_date	End_date	Time_from	Time_to	Nature	Lane_cl
2945708	-9841528	LP 410 S CONNECTOR RAMP TO IH 35 SOUT	SOUTH	8/2/2002	8/2/2002	1230	1330	REPLACE CCTV (BINZ-ENGLEMAN)	1 RIGHT
2943208	-9850083	IH 10E CONNECTOR RAMP TO IH 35N (FINES	EAST	8/2/2002	8/2/2002	0930	1530	REPAIR SIGNAL CABLE	ONE RIGHT LAN
2945611	-9840180	LP 410S FROM IH 35/FM 78 TO DIETRICH RD	SOUTH	8/3/2002	8/4/2002	2100	1600	EXTEND ENTRANCE RAMP, MILL AND F	ONE RIGHT LAN
2951583	-9852805	LP 410 EB VANCE JACKSON TO BLANCO	EASTBOUND M	8/3/2002	8/3/2002	0200	0900	SET BRIDGE BEAMS	4
2951958	-9839944	IH 35S CONNECTOR RAMP TO LP 410W	SOUTH	8/4/2002	8/4/2002	1130	1300	LCS REPAIR (LCS3-0035S-166.745)	ONE LEFT LANE
2939555	-9852264	US 90 EB CUPPLES TO IH 35	EASTBOUND	8/4/2002	8/5/2002	2330	0530	SIGN REPLACEMENT	3 LEFT LANES
2951736	-9840014	LP 410 E FROM PERRIN BEITEL TO IH 35	EAST	8/4/2002	8/4/2002	1030	1130	LCS REPAIR (LCS4-0410E-026.656)	TWO RIGHT LAN
2950027	-9855375	IH 10E FROM MEDICAL DR TO CALLAGHAN F	EAST	8/4/2002	8/4/2002	0800	1100	OVERHEAD SIGN WORK	TWO LEFT LANE
2946111	-9851319	IH 10 W @ FULTON AVE	WEST	8/4/2002	8/4/2002	0700	1300	ATTENUATOR REPAIR	1 RIGHT (UPPEF
2949611	-9854903	IH 10E EXIT RAMP TO TRANSGUIDE (OLD LP	EAST	8/4/2002	8/4/2002	1200	1600	OVERHEAD SIGN WORK	ONE
2940708	-9847944	IH 37N FROM IH 10 TO DURANGO BLVD	NORTH	8/4/2002	8/4/2002	0800	0930	VMS REPAIR AT CAROLINA (CMS2-0037	2 LEFT LANES
2943722	-9847680	IH 37N CONNECTOR RAMP TO IH 35S	NORTH	8/4/2002	8/4/2002	0930	1030	CCTV REPAIR (CCTV-0037N-142.825)	ONE
2960902	-9846888	US 281 FROM DONELLA TO LP 1604	NORTH	8/4/2002	8/4/2002	0800	1400	REPLACE ATTENUATOR	BOTH
2949652	-9854930	IH 10 E/B FROM LP-410 CONNECTOR RAMP	EAST	8/4/2002	8/4/2002	0700	1000	LONG LINE STRIPING	ALTERNATING
2949041	-9856708	BABCOCK FRNTG RD	BOTH	8/5/2002	8/9/2002	0900	1600	BRIDGE OVERHANG	NB 1 RIGHT LAN
2940569	-9857416	US 90 E FROM 36TH ST. TO GEN MCMULLEN	EASTBOUND	8/5/2002	8/6/2002	2200	0500	SIGN INSTALLATION	LEFT 3 LANES
2944805	-9864555	SH 151E FRNTG RD FROM SLICK RANCH CF	EAST	8/5/2002	8/9/2002	0900	1600	INSTALL WALL PANELS / HEADER BANI	ONE
2940764	-9858111	US 90 W FROM SH 151 TO ACME RD.	WESTBOUND	8/5/2002	8/6/2002	2200	0500	SIGN INSTALLATION	RIGHT 2 LANES
2948972	-9855763	LP 410E @ FREDERICKSBURG RD	EAST	8/5/2002	8/9/2002	0800	1700	POURING CONCRETE	1 RIGHT LANE
2944638	-9854041	BANDERA RD (SP 421) @ HILLCREST	NORTH	8/5/2002	8/7/2002	0830	1700	ASPHALT OVERLAY	ALTERNATING
2939555	-9851624	US 90 E 1000' BEFORE IH 35 TO IH 35 SOUTI	EASTBOUND	8/5/2002	8/6/2002	2200	0500	SIGN INSTALLATION	RIGHT LANE
2954569	-9836569	IH 35 N FRONTAGE RD @ OCONNOR RD	NORTH	8/6/2002	8/6/2002	0800	1300	CUT CURB APPROCH	1 RIGHT
2958319	-9830569	IH-35 N/B JUST NORTH OF LP 337 TO FM 104	N/B	8/6/2002	8/6/2002	1315	1400	MOVE BARRIERS TO RELOCATE EXIT R	1 (RIGHT LANE)
2940375	-9851083	IH 35 SB MAINLANES @ NOGALITOS	SOUTHBOUND	8/6/2002	8/7/2002	2200	0500	OVERHEAD SIGN REPLACEMENT	2 (RIGHT)
2951916	-9848180	LP 410E FRNTG RD FROM JONES MALTSBEI	EAST	8/6/2002	8/6/2002	0330	0530	MOVE FORM WORK	ONE LEFT LANE
2940027	-9851249	IH 35 S FRNTG BEFORE NOGALITOS	SOUTH	8/6/2002	8/7/2002	2200	0500	OVERHEAD SIGN REPLACEMENT	ALTERNATING
2948847	-9856583	LP 410E FROM BABCOCK TO IH 10E	EAST	8/6/2002	8/7/2002	2200	0530	POUR BRIDGE DECK	3 (COMPLETE C
2954583	-9836903	IH 35 N AT THE O'CONNOR OVERPASS	NORTHBOUND	8/6/2002	8/9/2002	2000	0600	ROAD REPAIR AND BRIDGE WORK	OVERPASS
2942597	-9871083	LP 1604/POTRANCO RD. INTERSECTION*	BOTH	8/7/2002	8/9/2002	0830	1700	ASPHALT OVERLAY	ALTERNATING
2943611	-9849708	IH 35 S ON UPPER LEVEL FROM ST MARYS T	SOUTH	8/7/2002	8/7/2002	1900	2359	REPAIR BARRELS	1 LEFT
2950791	-9855347	IH 10E FRNTG. RD. FROM CALLAGHAN TO LF	EAST	8/7/2002	8/8/2002	0900	1530	REMOVE CONCRETE TRAFFIC BARRIEI	ALTERNATING
2937416	-9853278	US 90 ZARZAMORA @ NOGALITOS INTERSE	WEST	8/7/2002	8/7/2002	0800	1600	TIE-IN TYPE A ASPHALT ON NOGALITOS	1
2951666	-9845152	LP 410 WB MAINLANES @ NACOGDOCHES F	WEST	8/7/2002	8/7/2002	2000	2330	REPAIR ACOUSTIC SENSORS	2 RIGHT *SEE N
2951638	-9846458	LP 410 EB MAINLANES @ BROADWAY	EAST	8/8/2002	8/8/2002	2000	2330	REPAIR ACOUSTIC SENSORS	2 RIGHT *SEE N
2952597	-9839125	IH 35 N/B FROM WALZEM TO WEIDNER RD.	NORTHBOUND	8/8/2002	8/9/2002	2100	0530	PAVEMENT REPAIR	LEFT 2 LANES, A
2951666	-9863486	BANDERA RD FROM REINDEER TRAIL TO LE	NORTH	8/8/2002	8/10/2002	0830	1600	ASPHALT OVERLAY	ALTERNATING
2936624	-9844985	IH 37 MAINLANES FROM FAIR AVE TO HOT W	BOTH	8/8/2002	8/8/2002	0900	1400	STRIPING	ALTERNATING

Record: 2059 of 3146

Figure 36. Sample Lane Closure Database Records.

Interestingly, the overall completeness rates for the same lanes from March 2002 – April 2004 were 46, 52, 63, 71, and 60 percent (or 58 percent overall for the sector). At first sight, this result could suggest that the lane closure event was responsible for the lower completeness rates. However, all the detectors in this part of town are controlled by TRF LCUs, which do not report data if no vehicles cross the detectors during the corresponding 20-second recording intervals. Since the lane closure took place at night, when there was less traffic and consequently fewer lane data records in the database, identifying the impact of the lane closure on the completeness rate became much less clear. As a reference, the overall completeness rates for the same lanes using data from two different random nights when there were no incidents and no lane closures were 21, 0.2, 12, 61, and 50 percent (or 29 percent overall for the sector).

In general, while the process to obtain completeness rates using the lane closure database as a data resource is conceptually simple, it is not trivial to execute, particularly if the objective is to be able to match lane closure data and archived lane data automatically. As Figure 36 shows, the field containing the highway name in the lane closure database does not follow consistent naming conventions, which would make it very difficult to use automated scripts to extract road name information reliably. As a result, there are many instances of records that refer to the same highway, but use very different spelling, e.g., I 10, I 10 (i.e., two spaces between "I" and "10"), IH 10, IH-10, IH 10E, IH 10W, IH 10E/W, IH 10EAST, IH 10EB, IH 10WEST, and IH 10WB. It is not always clear whether the first street named corresponds to the corridor along which the lane closure takes place. Further, the road name field frequently uses local street names instead of state highway names and does not always make a clear distinction between main lanes and frontage roads. A similar difficulty arises from the street names representing the beginning and ending locations, which means that an automated procedure would have very low chances of success not just matching corridors, but also matching beginning and ending locations.

Because the lane closure database included latitude-longitude data, the researchers generated a layer in the GIS to represent the location associated with every lane closure in the sample. Figure 37 shows the resulting map. At first sight, it appeared that the mapped locations would be adequate to match lane closure locations to sectors. However, a zoomed-in view quickly revealed significant discrepancies between lane closure locations and ITS infrastructure locations that made it extremely difficult to derive meaningful information consistently. For example, because lane closure locations only provided single data points representing entire lane closure events, it was not clear from the map whether the points represented the beginning, the end, or any point somewhere in the middle of the lane closure events. As an illustration, all the highlighted road closure events in Figure 37 included US 90 and 36th Street in the description. With only that information provided, however, it would be very difficult to (a) determine the correct relationship between lane closure points, lane closure event extents, and 36th Street; and (b) map the lane closure points to the correct ITS detectors and sectors.

The zoomed-in view in Figure 37 highlighted another challenge. In general, neighborhood operations facilitate the mapping between features in a GIS environment. However, the success in using those procedures depends, among other factors, on the positional accuracy of the various features involved. Highly inaccurate features dramatically increase the probability that features will be incorrectly mapped. As an illustration, several of the points in Figure 37 could be mapped to the wrong sector (e.g., a point representing a lane closure on the eastbound direction would be mapped to a westbound sector), because their physical location on the map is not consistent with the corresponding database description.

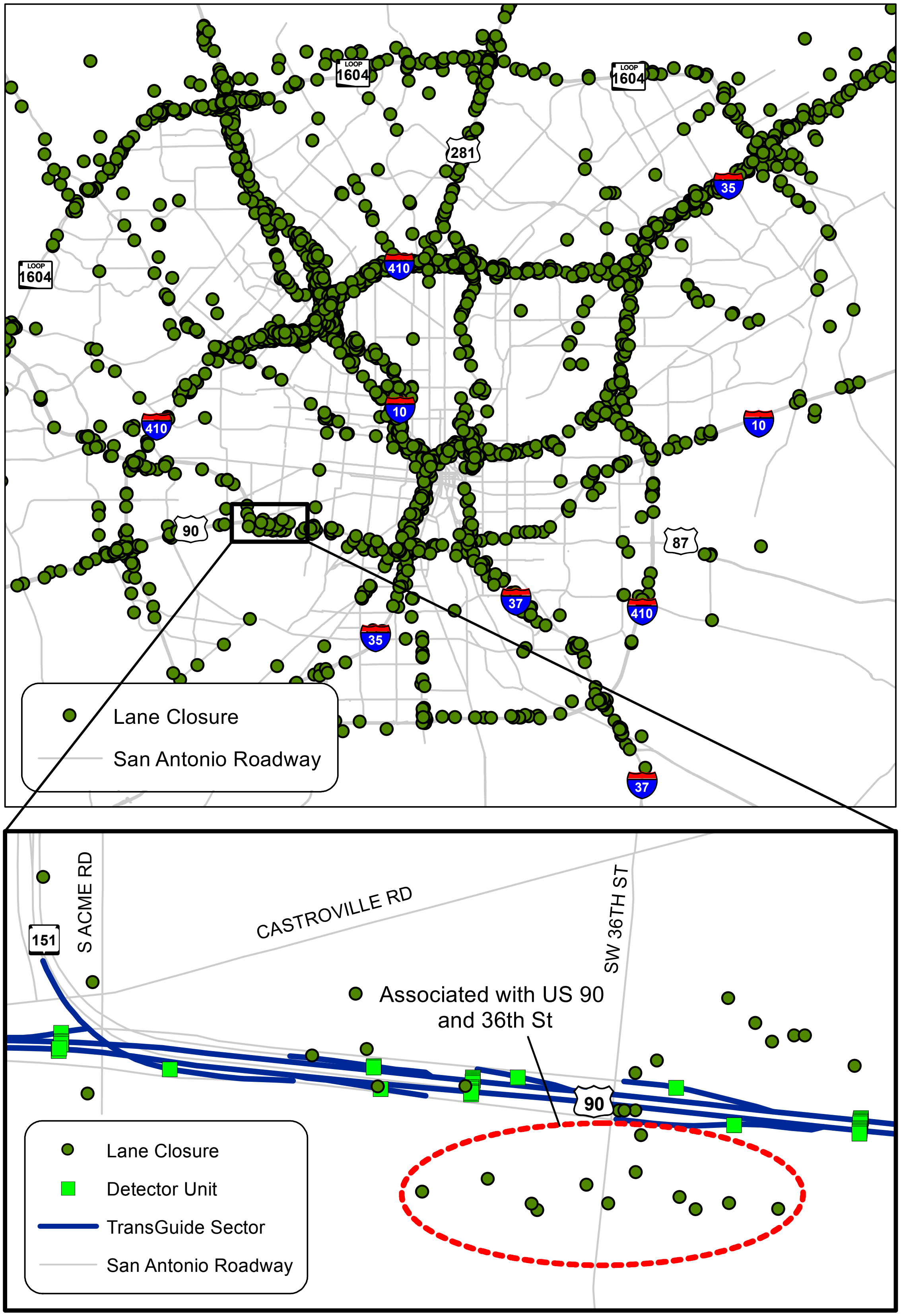


Figure 37. Lane Closure Locations.

In its present form, the TransGuide lane closure database is not very useful to help assess ITS data completeness. To make the lane closure database useful for this purpose, it would be necessary to modify its database structure and data entry procedures. It may be worth noting that the Highway Condition Reporting System (HCRS) already enables districts to enter roadway closure data that become part of the official TxDOT highway condition data repository (*26*). Because there are two separate data entry processes entering very similar information (one for the TransGuide lane closure database and the other one for HCRS), it would be advisable to develop a single data entry interface that could address the needs of both systems. If this is not feasible or practical, at the very least it would be advisable to modify the TransGuide lane closure database. The following is a list of potential changes that would be necessary:

- Replace the highway and limit fields with at least three fields, where the first field represents the route where the lane closure will take place, the second field represents the beginning point, and the third field represents the ending point.
- In the data entry form, replace text boxes with drop-down lists to ensure compliance with pre-established roadway naming conventions. Effectively, this strategy would eliminate the problem resulting from using a multiplicity of names to represent the same corridor. To facilitate data entry into HCRS, the drop-down lists should use official TxDOT route ID designations, e.g., "IH0010" instead of "IH 10" or "IH-10," or "US0090" instead of "US 90."
- Develop a different procedure to represent lane closure locations. Single point coordinate data pairs to represent lane closure locations are inadequate. Since there is a one-to-many relationship between lane closure events and highway segments or sectors, the required change would involve using either multiple coordinate data pairs or the ability to associate several linear features with lane closure events. Using online mapping techniques (e.g., using Arc Internet Map Server (ArcIMS), which is part of TxDOT's core GIS architecture) would enable the complete representation of road closure events by interactively clicking on the affected highway segments or sectors. This approach would also enable explicit modeling of lane closures affecting both directions of travel.

To implement the changes to the lane closure database it might be necessary to take into account requirements included in the TxDOT GIS architecture (*27*), in particular those that pertain to the integration of absolute location measures and relative location measures and temporal and spatial querying.

REVISED DATA QUALITY CONTROL TESTS

The results from the previous sections confirmed the need to introduce some changes to the preliminary list of quality control tests shown in Table 6. Table 10 shows the revised list. A summarized description of the new table structure follows:

- Speed threshold. The updated threshold is 93 mph. Based on the results of the analysis, recorded speeds exceed this threshold 0.01 percent of the time.
- Volume threshold. The updated threshold is 18 vehicles. On average, recorded volumes exceed this threshold 0.1 percent of the time.

Table 10. Revised Speed, Volume, and Occupancy Quality Control Tests.

	Quality Control Name and Description	Test (LCU Subsystem Level)	Action before Database Archival	Further Action before Future Use (Query Level after Archival)
		First-Level Tests		
1a	Record format error	Record is in incorrect format	Move record to dump file	
1b	Duplicate records	Detector ID and date/time stamp are identical	Flag record Add system time function date/time stamp	
		Second-Level Tests		
2a	Extreme values	Speed < -1 or Speed > 93 Or (Volume < 0 or Volume > 18) Or (Occupancy < 0 or Occupancy > 99)	Flag record	Set Speed = <null> Set Volume = <null> Set Occupancy = <null> Impute missing values [1]
2b	Entrance or exit ramp (valid record)	Speed = -1 0 < Volume ≤ 18 0 < Occupancy ≤ 99	Flag record Set Speed = <null>	
2c	Entrance or exit ramp: No vehicle present (valid record)	Speed = -1 Volume = 0 Occupancy = 0	Flag record Set Speed = <null>	
2d	Entrance or exit ramp: Volume is zero when occupancy is not zero	Speed = -1 Volume = 0 0 < Occupancy ≤ 99	Flag record Set Speed = <null>	Set Volume = <null> Set Occupancy = <null> Impute missing values [1]
2e	Entrance or exit ramp: Occupancy is zero when volume is not zero	Speed = -1 0 < Volume ≤ 18 Occupancy = 0	Flag record Set Speed = <null>	Set Volume = <null> Set Occupancy = <null> Impute missing values [1]
2f	Main lane: No vehicle present (valid record)	Speed = 0 or Speed = <null> Volume = 0 Occupancy = 0	Flag record Set Speed = <null>	
2g	Main lane: Speed and volume are zero when occupancy is not zero	Speed = 0 Volume = 0 0 < Occupancy ≤ 99	Flag record	Set Speed = <null> Set Volume = <null> Set Occupancy = <null> Impute missing values [1]
2h	Main lane: Speed and occupancy are zero when volume is not zero	Speed = 0 0 < Volume ≤ 18 Occupancy = 0	Flag record	Set Speed = <null> Set Volume = <null> Set Occupancy = <null> Impute missing values [1]
2i	Main lane: Speed is zero when volume and occupancy are not zero	Speed = 0 0 < Volume ≤ 18 0 < Occupancy ≤ 99	Flag record	Set Speed = <null> Set Volume = <null> Set Occupancy = <null> Impute missing values [1]
2j	Main lane: Volume and occupancy are zero when speed is not zero	0 < Speed ≤ 93 Volume = 0 Occupancy = 0	Flag record	Set Speed = <null> Set Volume = <null> Set Occupancy = <null> Impute missing values [1]
2k	Main lane: Volume is zero when speed and occupancy are not zero	0 < Speed ≤ 93 Volume = 0 0 < Occupancy ≤ 99	Flag record	Set Speed = <null> Set Volume = <null> Set Occupancy = <null> Impute missing values [1]
2l	Main lane: Occupancy is zero when speed and volume are not zero	0 < Speed ≤ 93 0 < Volume ≤ 18 Occupancy = 0	Flag record	Set Speed = <null> Set Volume = <null> Set Occupancy = <null> Impute missing values [1]
2m	Missing records: either field or LCU server cause	Record is missing	Insert record Set Speed = <null> Set Volume = <null> Set Occupancy = <null>	Impute missing values [1]

[1] If needed for the analysis.

- Percent occupancy threshold. The updated threshold is 99 percent. On average, recorded occupancy values exceed this threshold 0.0000004 percent of the time.
- Actions. For clarity, Table 10 shows two types of actions: actions before database archival and additional actions before using the archived data in the future. Since there is relatively little control over the characteristics and functionality of the LCU software used in the field, the assumption here is that implementation of any action before database archival will likely take place at the LCU server level. With the exceptions of tests 2b, 2c, 2d, 2e, and 2f (which change the speed value from -1 or 0 to null), actions before database archival should not result in any changes to the raw data. Additional actions before future use are suggested and include changing speed, volume, and occupancy values to null for "abnormal" records and imputing missing values as needed, depending on the needs and purposes of the analysis.
- Test 1b. The only cases where the analysis found duplicate records (meaning the detector ID and the date/time stamp were duplicate) were when the time changed back one hour from CDT to CST. Because the affected records were still valid, it would have been inappropriate to move the duplicate records to a dump file. To address this issue, the revised version of Test 1b simply flags those records. In addition, it adds a unique date/time stamp field using the Unix time function, which TransGuide already uses throughout the rest of the system.
- Test 2m. This test explicitly keeps track of missing records that may be caused by reasons other than the system not being able to physically append records to the database. A typical example would be if there is a malfunctioning detector and/or LCU that prevents the LCU driver from receiving data from the field. Adding records with null speed, volume, and occupancy values can increase the physical size of the database. However, adding those records to the database (with the corresponding flag) can provide a useful, positive confirmation that the missing record was due to problems in the field, not at the TMC. In general, flag 2m addresses cases where there is traffic crossing the detection zone, but the system in effect cannot generate a record. In contrast, flag 2f addresses cases where the detector is functioning properly but there is no traffic. Currently, Naztec LCUs generate this type of record, but TRF LCUs do not. To provide a positive confirmation that any missing TRF LCU record is indeed due to problems in the field and not simply due to lack of traffic, it would be advisable to modify the TRF LCU software to enable the generation of records having zero (or null) speed, zero volume, and zero occupancy.

CHAPTER 5. CONCLUSIONS AND RECOMMENDATIONS

Previous chapters described incident detection practices at TransGuide, a process to enhance incident detection algorithm performance, and ITS data quality control and completeness tests. This chapter summarizes the research findings and outlines recommendations for implementation and further work.

SUMMARY OF FINDINGS

Incident Detection Assessment

Chapter 2 discussed the incident detection process at TransGuide, with a focus on the incident detection algorithm, and the process to extract meaningful incident data from archived ITS data sources. It described TransGuide's general system configuration and the incident response process at TransGuide. For incident detection, TransGuide relies on a combination of detector-based alarms and 911-based alarms, CCTV camera scanning, police radio scanning, and courtesy patrols. Detector-based alarms rely on speed for speed-trap detectors on main lanes and percent occupancy for non-speed-trap detectors on entrance and exit ramps. The current thresholds for speed-trap detectors are 25 mph (minor alarms) and 20 mph (major alarms). For non-speed-trap detectors, the current thresholds are 25 percent occupancy (minor alarms) and 35 percent occupancy (major alarms).

To assess the incident detection algorithm effectiveness, the researchers prepared two datasets. The first dataset contained data from the scenario database, under the assumption that this database provided an adequate representation of the history of incidents along the freeway network covered by TransGuide. The second dataset contained alarms triggered by the incident detection algorithm in response to events on the road. The lack of a common link between the two datasets led to the use of a "fuzzy" spatio-temporal query methodology that considered an incident to be detected if the incident detection algorithm triggered an alarm within a pre-specified spatio-temporal window associated with an incident record.

Matching alarm and scenario data enabled the determination of performance measures such as incident detection rates, false alarm rates, false alarm frequeny, and effective alarm rate. To complete the analysis, the researchers also prepared maps showing the spatial distribution of incident detection rates and false alarm rates on a sector by sector basis. The incident detection rate, which included major and minor accidents, stalled vehicles, and debris, was 20 percent. After excluding debris, stalled vehicles, and minor accidents, the incident detection rate increased to 27 percent. The literature reports detection rates that are typically much higher (60 – 100 percent), although there are also references to detection rates in the 30 – 50 percent range. Readers should be aware that many high detection rates in the literature are based on very small sample sizes and/or pre-set thresholds calibrated under the assumption of "normal flow" conditions, and that actual performance on the ground tends to be lower. The false alarm rate was 0.0039 percent, which was low compared to rates typically found in the literature (0.002 – 1.9 percent). The false alarm frequency was 10 false alarms per hour (or a false alarm roughly every 6 minutes). The effective alarm rate, or number of incident-confirmed alarms relative to the total number of alarms actually triggered, was 2.3 percent.

These results raise questions about the effectiveness of using road sensors for incident detection, not just at TransGuide but also at many other TMCs that rely on similar technologies. These questions are relevant given the increasing use of non-sensor-based incident detection procedures (such as drivers calling on their cell phones to report incidents). Unfortunately, the data available for the research only provided answers relative to the effectiveness of the incident detection algorithm, but not whether or when incidents would be detected if the road sensors were not in place. As a result, it was not possible to conclusively determine whether eliminating the use of sensors for incident detection would be advisable. It may be worth noting that TMCs use detectors not just for incident detection but also for the production of performance measures such as travel times and delays, which TMCs are increasingly disseminating to the public through traveler information system implementations. Public perception is that travel time information is useful and timely, therefore providing justification to the continuous investment on detector technology and related infrastructure. In this regard, TMC officials are facing a number of issues related to the spatial and temporal resolution of the data collection (i.e., detector spacing and data collection interval) because the data needs for travel time calculations are not necessarily the same as those for incident detection.

Alarm Threshold Optimization

Chapter 3 discussed the feasibility of modifying current incident detection alarm thresholds to help optimize TMC incident detection practices. For the analysis, the researchers used incident data, alarm event data, and archived 20-second lane data. The researchers analyzed 75 sample cases that covered a wide range of situations, including detected incidents (scenario loaded and alarm triggered), false negatives (scenario loaded, but no alarm was triggered), and false positives (no scenario was loaded but an alarm was triggered anyway). For completeness, the dataset included data from both TRF LCUs and Naztec LCUs. To better understand differences and similarities among cases, the researchers categorized cases based on traffic conditions before and during incidents (for detected incidents and false negatives) and based on conditions causing alarm generation (for false alarms).

The researchers used a prototype offline tool to evaluate incident detection algorithm performance by measuring the impact of modifying speed alarm thresholds on the number and timing of alarms generated by the system. The tool, called Incident Detection Algorithm Tester (IDAT), enables users to select one or more sectors of interest and a range of dates. With this information, the tool reads 20-second data from the archived lane data database, calculates 2-minute moving average speeds, and "triggers" minor and major alarms if the moving averages fall below the pre-specified thresholds. IDAT also enables users to export the minor and alarm data to comma-delimited text files.

For each case analyzed, the researchers ran IDAT for a 24-hour period and recorded the number of alarms generated in addition to the corresponding time stamps. The simulation involved using five different alarm threshold values: 25 mph (current minor alarm threshold), 30 mph, 35 mph, 40 mph, and 45 mph. Analysis of data showed that as the alarm threshold level increased, the average number of alarms increased exponentially and the average incident detection time decreased linearly. There was a correlation between congestion levels and the number of alarms generated by the algorithm, as well as a correlation between congestion levels and average incident detection times. In general, incident detection took considerably longer (between 60 and 100 percent longer) under congested traffic conditions than under uncongested traffic conditions. Increasing the alarm threshold level resulted in a decrease in average incident detection times that were much more noticeable under uncongested traffic conditions than under congested traffic conditions.

The analysis indicates that increasing the minor alarm threshold from 25 mph to 35 mph would result in tangible benefits in terms of shorter incident detection times without increasing the number of false alarms to an unacceptable level. The analysis also indicates that the best strategy would be to increase the minor alarm threshold to 35 mph throughout the day, regardless of congestion level or time of day period. The overall impact of increasing the minor alarm threshold to 35 mph would be a 10 percent increase in the number of alarms (at least half of which would be in the form of true alarms) and a 30 percent decrease in incident detection times.

The analysis resulted in some additional observations. For example, average incident detection time was about 4.3 minutes (260 seconds) at the 25-mph alarm threshold level, with average values ranging from 4 minutes for uncongested traffic to 6 minutes for congested traffic. In general, using 2-minute moving average speeds in the incident detection algorithm (as opposed to the original 20-second speeds) resulted in 1 – 2 minutes of delay in the incident detection time.

Data Quality and Completeness

Chapter 4 discussed ITS data quality control and completeness issues. For the analysis, the researchers evaluated approximately 3.4 billion 20-second lane detector data records from March 2002 to April 2004. In general, there were two types of tests: tests for "valid" records and tests for "abnormal" records. "Valid" records were records with valid volume and occupancy values but invalid "by design" speed values, e.g., -1 in the case of non-speed-trap detectors located on entrance and exit ramps, or zero in the case of main lane detectors when no vehicle passed the detection zone during the detection time period. "Abnormal" records were records with "abnormal" combinations of speed, volume, and percent occupancy values (e.g., zero speed, zero volume, but larger than zero occupancy) that could have resulted from causes such as faulty detectors or faulty LCU software logic. Some 1.6 billion speed, volume, and occupancy records had a quality control flag, accounting for nearly 48 percent of the 3.4 billion lane data record set. Approximately 1.5 billion flagged records were "valid" records and the remaining 126 million flagged records were "abnormal" records. The 126 million "abnormal" records resulted in an overall "abnormal" record rate of about 3.7 percent.

There were significant differences between TRF LCU records and Naztec LCU records. For example, even though 32 percent of LCUs were TRF LCUs, the percent of "abnormal" records associated with detectors controlled by TRF LCUs was 84 percent. The vast majority of these records had flag 2j (speed > 0, volume = 0, and occupancy = 0), with practically no records under the other flag categories (except flag 2a). In contrast, Naztec LCU records, even though they were the minority, had representation in every single flag category. Some 54 percent of Naztec LCU records had flag 2i (speed = 0, volume > 0, occupancy > 0). There were 369 million Naztec LCU records with flag 2f (speed = 0, volume = 0, and occupancy = 0). In contrast, there was not a single TRF LCU record with that flag. The reason is that the TRF LCU software does not generate lane detector records if no vehicles have crossed the associated detectors during the 20-second recording interval.

An evaluation of spatial trends in the distribution of quality control flags showed cases where the spatial distribution was roughly uniform, but also cases where there were significant exceptions. An evaluation of temporal variations in the distribution of quality control flags showed that, in most cases, the highest concentration of flagged records occurred at night, when there was relatively little traffic and, consequently, there was a higher chance for time intervals without vehicles or for isolated detector readings producing abnormal data.

An analysis of the speed data series found 93 mph to be an adequate upper threshold for quality control purposes. Similar analyses of the volume and percent occupancy data found 18 vehicles and 99 percent, respectively, to be adequate upper thresholds for quality control purposes. Three speed values stood out in the analysis because of their abnormal frequencies: 0, 87, and 94 mph. Most zero-speed records—all of them associated with Naztec LCUs—resulted from cases where no vehicles crossed the detectors. In the case of the 87- and 94-mph records, the trends clearly showed an anomaly, but it was unclear from the analysis what could have caused that anomaly. In the case of the percent occupancy data, there was a jump between 99 and 100 percent, which the analysis could not explain.

The data completeness analysis included an aggregate evaluation of completeness by LCU server and a detailed evaluation of completeness at the individual detector level. At the LCU server level, the completeness rate—measured as the ratio of number of days with data to total number of potential days with data—varied from 95 to 100 percent. At the individual detector level, the analysis showed that, on average, the completeness rate for all detectors was about 80 percent. Very few detectors had high completeness rates. For example, only about 35 percent of detectors had a completeness rate of 95 percent or higher. At the same time, very few detectors had very low completeness rates. For example, only about 10 percent of detectors had a completeness rate of 50 percent or lower. The overall completeness rate for Naztec LCU detectors was higher than for TRF LCUs (84 percent versus 71 percent, respectively). Interestingly, the analysis showed a higher than average frequency of gaps in the lane data during incidents than during normal traffic conditions.

RECOMMENDATIONS FOR IMPLEMENTATION

Report 0-4745-2 contains products 0-4745-P3 (which includes detailed incident evaluation procedures) and 0-4745-P4 (which addresses process definitions and implementation recommendations) (*2*). That report already describes recommendations for implementation in detail. This section, therefore, only summarizes some of the most relevant aspects.

The research developed a number of procedures for evaluating incident detection practices and performance. The procedures cover a wide range of activities such as extracting meaningful incident data for analysis, evaluating incident detection algorithm performance, and assessing data quality control and completeness. Implementation of the research findings would likely involve changes in the way managers and operators interact with, manage, and interpret incident-related data. For example, implementation of the process and queries to match alarms and incidents requires the use of queries to match alarm data and scenario data, which, in turn, requires the use of relational database structures to handle event data, scenario header and execution data, and ITS infrastructure data. Specific recommendations include incorporating those database structures into the TransGuide database design, developing a GUI to automate the query design process, and modifying the scenario header table population process to ensure that the incident ID field is the same as the incident ID field in the alarm tables.

Implementation of the incident detection algorithm performance evaluation tool would require the development and installation of an offline tool similar to the IDAT tool the researchers developed to simulate the alarm generation process at TransGuide. Specific recommendations include developing a relational database archive of 20-second speed, volume, and occupancy data and developing code and corresponding GUI to include the minimum recovery time concept implemented in IDAT. It may be worth noting that the minimum recovery time concept has potential beyond the offline incident evaluation environment evaluated during the research. Incorporating a minimum recovery time into the real-time incident management process at TxDOT would enable the system to automatically close alarms after moving average speeds have “recovered” after a reasonable period of time; thus, reducing further interference for operators.

As mentioned previously, the analysis showed that increasing the minor alarm threshold would result in tangible benefits in terms of shorter incident detection times without increasing the number of false alarms to an unacceptable level. The recommendation is to increase the minor alarm threshold to 35 mph throughout the day, regardless of congestion level or time of day period. The expected impact of increasing the minor alarm threshold to 35 mph would be a 10 percent increase in the number of alarms (at least half of which would be in the form of true alarms) and a 30 percent decrease in incident detection times.

Implementation of the data quality control flags (Table 10) would involve making changes to the way the LCU subsystem manages field data. Specific recommendations include creating a lookup table in the archive database to list and describe the various quality control tests and flags used, developing a module to conduct data quality control tests and assign flags to the affected records immediately after receiving lane data from the field, adding a unique date/time stamp to the lane data archive that does not depend on the seasonal changes between CST and CDT, and developing code and GUIs to automate the query building process.

RECOMMENDATIONS FOR FURTHER RESEARCH WORK

This report has outlined a number of areas that need further work. A summary of research needs follows:

- Continue the development of ITS data quality control and completeness testing procedures. This research described tests that, for the most part, involve individual lane records and, therefore, ignored trends that would require analyses of consecutive lane detector records. Examples include tests to verify the validity of volume data over longer periods of time, such as 15 minutes, one hour, or 24 hours; as well as tests to verify the validity of the relationship between speed, volume, and occupancy in cases where none of the values is zero (which this research addressed). The importance of developing more comprehensive quality control and completeness tests becomes apparent as TMCs see their roles evolving towards the management and distribution of both real-time and archived data packages to interested stakeholders.

- Develop a prototype lane closure database and associated data entry and management procedures to address both district needs and TxDOT highway condition reporting needs. This report outlined a few recommendations concerning changes that would be necessary to make the lane closure database at TransGuide useful as a data resource for ITS data completeness assessments. In the larger picture, however, it appears that both HCRS and the local lane closure database would need enhancements to avoid duplication of data entry efforts and to ensure the resulting database design addresses both local district and division needs. It would be advisable to develop a prototype that takes into account modern web-based mapping and data management tools to facilitate the data entry, query, and reporting processes.
- Investigate the correlation between missing ITS data and incidents. This research project found unusually high gaps in archived lane data during incidents. It is possible that traffic flow during incidents can be erratic in ways that the LCU software does not how to process properly, causing it to simply reject more records than normal. Unfortunately, during the course of the research it was not possible to explore in detail any potential reasons that could explain why data gaps seemed to be more prevalent during incidents. Further research would be needed to explain this finding.

REFERENCES

1. Quiroga, C., Kraus, E., Pina, R., Hamad, K., and Park, E. *Incident Characteristics and Impact on Freeway Traffic*. Research Report 0-4745-1. Texas Department of Transportation, Austin, Texas, 2004.
2. Quiroga, C., Hamad, K., and Park, E. *Incident Evaluation Procedures and Implementation Requirements*. Research Report 0-4745-2. Draft. Texas Department of Transportation, Austin, Texas, 2005.
3. Payne, H., Helfenbein, E., and Knobel, H. *Development and Testing of Incident Detection Algorithms, Volume 2: Research Methodology and Detailed Results*. Report No. FHWA-RD-76-20. Federal Highway Administration, U.S. Department of Transportation, Washington D.C., April 1976.
4. Haas, C., Mahmassani, H., Haynes, M. Khoury, T., and Logman, H. *Evaluation of Automatic Vehicle Identification for San Antonio's TransGuide for Incident Detection and Advanced Traveler Information Systems.* Research Report 0-4957-1. Texas Department of Transportation, Austin, Texas, 2001.
5. Mahmassani, H., Haas, C., Zhou, S., and Peterman, J. *Evaluation of Incident Detection Methodologies.* Research Report 0-1795-1. Texas Department of Transportation, Austin, Texas, 1999.
6. Dudek, C., Messer, J., and Nuckles, N. *Incident Detection on Urban Freeways*. Transportation Research Record 495, TRB, Washington, D.C., 1974.
7. Ahmed, S. A., and Cook, A. R. *Application of Time Series Analysis Techniques to Freeway Incident Detection*. Transportation Research Record 682, TRB, Washington, D.C., 1980, pp. 19-21.
8. Stephanedes, Y. J., and Chassiakos, A. P. Application of Filtering Techniques for Incident Detection. *Journal of Transportation Engineering*, ASCE, Vol. 119, No.1, 1993, pp. 13-26.
9. Balke, K. *An Evaluation of Existing Incident Detection Algorithms.* Research Report 0-1232-20. Texas Department of Transportation, Austin, Texas, 1993.
10. Cheu, R., and Ritchie, S. Automated Detection of Lane-blocking Freeway Incidents Using Artificial Neural Networks. *Transportation Research Part C*, Vol. 3, No. 6, 1995, pp. 371-388.
11. Chang, E., and Wang, S. *Improved Freeway Incident Detection Using Fuzzy Set Theory*. Transportation Research Record 1453, TRB, Washington, D.C., 1994, pp. 75-82.
12. Chang, E., Kolb, S., Park, B., Tam, K., and Urbanik, T. *San Antonio Incident Detection Algorithm Study*. Research Report 2917-1F. Texas Department of Transportation, Austin, Texas, 1996.
13. Mahmassani, H., Haas, C., Logman, H., and Shin, H. *Integration of Point-based and Linked-based Incident Detection and Traffic Estimation.* Research Report 0-4156-1. Texas Department of Transportation, Austin, Texas, 2004.
14. *TransGuide Intelligent Transportation System. Software Design Document. Version 2.0*. Report TGITS-SDD-2.0. Southwest Research Institute, Texas Department of Transportation, San Antonio, Texas, May 2003.
15. *TransGuide Intelligent Transportation System. Software Users Manual. Version 2.0*. Report TGITS-SUM-2.0. Southwest Research Institute, Texas Department of Transportation, San Antonio, Texas, May 2003.

16. *TransGuide ITS Design Report*. SwRI Project No. 10-6315, Southwest Research Institute, 1995.
17. Traffic Statistics. Texas Department of Transportation, San Antonio, Texas. http://www.transguide.dot.state.tx.us/docs/statistics.html. Accessed August 26, 2005.
18. Al-Deek, H. *Implementation of Incident Detection Algorithms*. Summary of Final Report, WPI 0510672. Florida Department of Transportation, 1999.
19. Martin, P., Perrin, J., and Hansen, B. *Incident Detection Algorithm Evaluation*. Utah Department of Transportation, 2001. http://www.ndsu.nodak.edu/ndsu/ugpti/MPC_Pubs/html/MPC01-122/index.html. Accessed August 30, 2005.
20. Black, J., and Sreedevi, I. Detection Algorithms. *ITS Decision Web site*. http://www.calccit.org/itsdecision/serv_and_tech/Incident_management/Incident_detection/detection_algorithms/detection_algorithms_report.html. Accessed August 2005.
21. Karim, A., and Adeli, H. Incident Detection Algorithm Using Wavelet Energy Representation of Traffic Patterns. *Journal of Transportation Engineering*, May/June 2002, pp. 232-242.
22. Weil, R., Wootton, J., and Gracia-Ortiz, A. Traffic Incident Detection: Sensors and Algorithms. *International Journal of Mathematical and Computer Modelling*, Vol. 27, No. 9-11, 1998, pp. 257-291.
23. Turner, S. *Guidelines for Developing ITS Data Archiving Systems*. Report 0-2127-3. Texas Department of Transportation, Austin, Texas, September 2001.
24. Turner, S., Eisele, W., Gajewski, B., Albert, L., and Benz, R. *ITS Data Archiving: Case Study Analyses of San Antonio TransGuide Data*. Report No. FHWA-PL-99-024. Federal Highway Administration, Washington, D.C., August 1999.
25. Lane Closures. Texas Department of Transportation, San Antonio, Texas. http://www.transguide.dot.state.tx.us/Traffic/lclist.php. Accessed August 26, 2005.
26. TxDOT Roadway Information Portal. Texas Department of Transportation, Austin, Texas. http://www.dot.state.tx.us/hcr/main.htm. Accessed August 26, 2005.
27. Wyman, M., and Wyman R. GAIP: A Technical Discussion. *GIS Architecture and Infrastructure Project (GAIP)*, version 1.0. Texas Department of Transportation, Austin, Texas, 2002.