Technical Report Documentation Page

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

4. Title and Subtitle	5. Report Date
INCIDENT EVALUATION PROCEDURES AND IMPLEMENTATION REQUIREMENTS	October 2005
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Cesar Quiroga, Khaled Hamad, and Eun Sug Park	Report 0-4745-2
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/0-4745-2.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 contains products 0-4745-P3 (which includes detailed incident evaluation procedures) and 0-4745-P4 (which addresses process definitions and implementation recommendations) that were developed during the second phase of research project 0-4745.

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	36	

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

INCIDENT EVALUATION PROCEDURES AND IMPLEMENTATION REQUIREMENTS

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 0-4745-2
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

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF ACRONYMS, ABBREVIATIONS, AND TERMS

ArcSDE	Arc Spatial Data Engine
CDT	Central daylight time
CST	Central standard time
DR	Detection rate
EAR	Effective alarm rate
ESRI	Environmental Systems Research Institute
FAF	False alarm frequency
FAR	False alarm rate
GIS	Geographic information systems
GUI	Graphical user interface
IDAT	Incident Detection Algorithm Tester
ITS	Intelligent transportation systems
LCU	Local control unit
ODBC	Open database connectivity
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.

This report contains products 0-4745-P3 (which includes detailed incident evaluation procedures) and 0-4745-P4 (which addresses process definitions and implementation recommendations) that were developed during the second phase of research project 0-4745. Report 0-4745-3 "Incident Detection Optimization and Data Quality Control" summarizes the activities conducted during the second phase of the research that led to the development of those products (*2*).

This report is organized as follows:

- Chapter 1 is this introductory chapter.
- Chapter 2 summarizes incident evaluation procedures.
- Chapter 3 discusses process definitions and recommendations for implementation.

CHAPTER 2. INCIDENT EVALUATION PROCEDURES

This chapter describes the incident detection evaluation procedures that constitute Product 0-4745-P3. The procedures include a process to match alarms and incidents using TransGuide data, an offline tool to evaluate incident detection algorithm performance, a list of quality control flags, and a methodology to assess lane data completeness.

PROCESS AND QUERIES TO MATCH ALARMS AND INCIDENTS

The analysis of TransGuide incident data relies on two datasets. The first dataset contains scenario database records that includes data about major accidents, minor accidents, stalled-vehicle incidents, and debris incidents. This database provides an adequate representation of the history of freeway incidents based on the results of an analysis completed during the first phase of the research 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*). The second dataset contains alarms triggered by the TransGuide incident detection algorithm in response to events on the road. Of specific interest are record types 5301 and 5303 from the event log files. Record type 5301 indicates the system has triggered a new alarm, whereas Record type 5303 indicates the operator has closed an alarm.

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 1 shows, there are three possible matching outcomes:

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

			Incident Detection Algorithm Triggered Alarm?	
			Yes	**No**
Scenario Deployed?	**Yes**	**Incident Occurred**	Incident Detected	False Negative
	No	**No Incident Occurred**	False Positive	

Figure 1. Possible Incident versus Alarm Dataset Matching Outcomes.

The lack of a common link between the two datasets (more specifically, an incident ID) led to the use of a "fuzzy" spatio-temporal query methodology whereby an incident is considered detected if the incident detection algorithm triggers an alarm within a pre-specified spatio-temporal window associated with an incident record (Figure 2). The reason behind this fuzzy range concept is 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 3 illustrates the query building process, which used the geodatabase structure described in report 0-4745-1 (*1*).

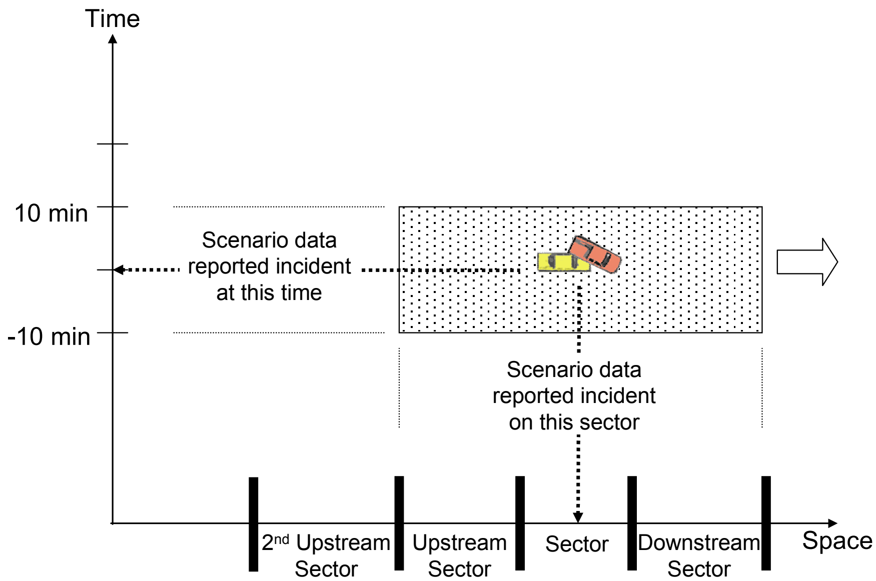


Figure 2. Spatio-Temporal Query Concept.

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. In practice, a sensitivity analysis would be necessary to measure the impact of the spatio-temporal window size on the number of matches between alarms and incidents. As an illustration, Figure 4 shows the results of the sensitivity analysis conducted during the research.

After the incident/alarm data matching, the next step would be to calculate performance measures such as incident detection rates and false alarm rates, which, in turn, would enable the production of summary reports and maps (Figure 5, Figure 6). For example, 3,828 out of 19,553 incident records from March 2002 – May 2004 had a matching alarm record. Likewise, 4,651 out of 202,690 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.

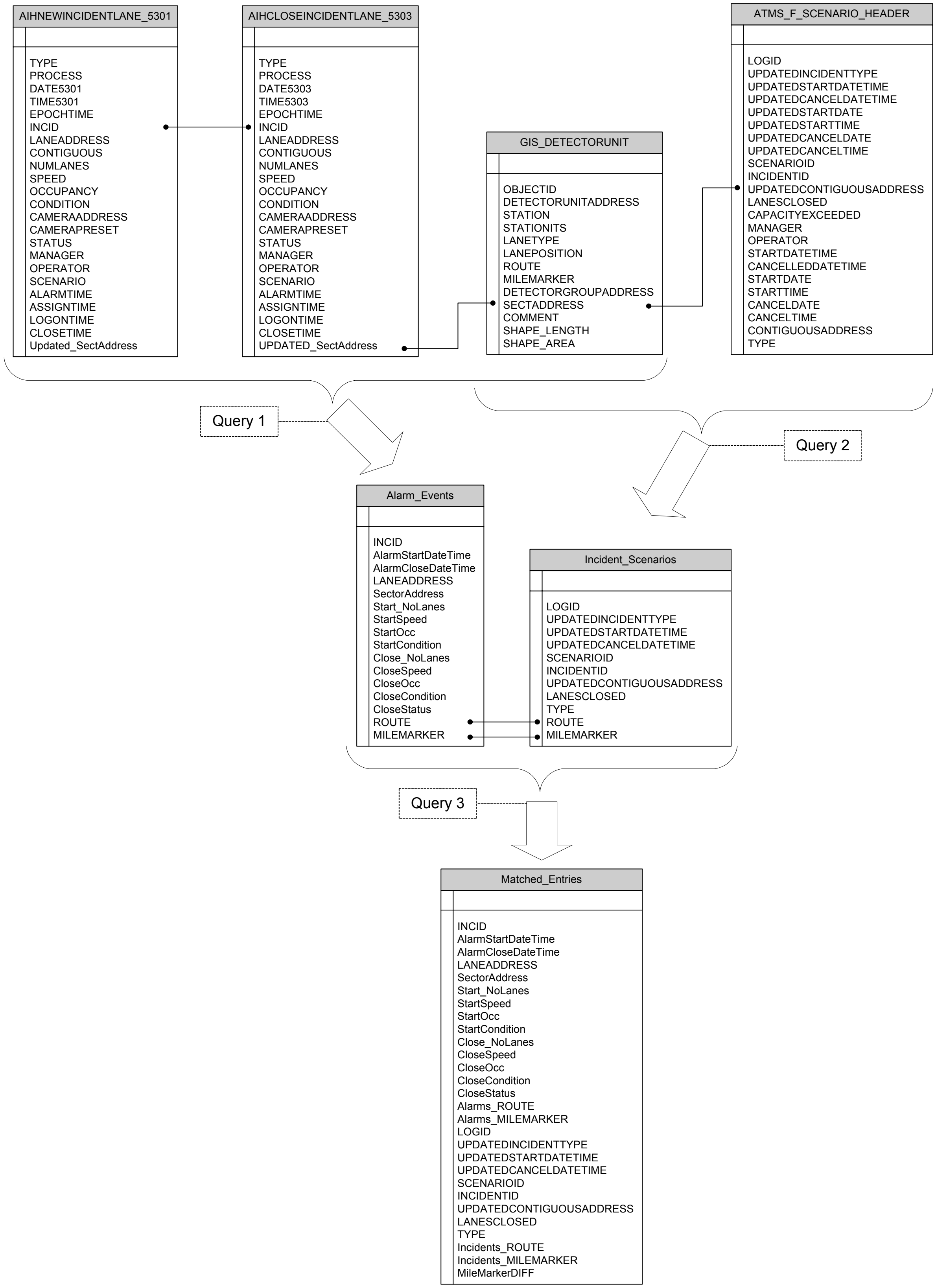


Figure 3. Query Building Process to Match Incidents to Alarms.

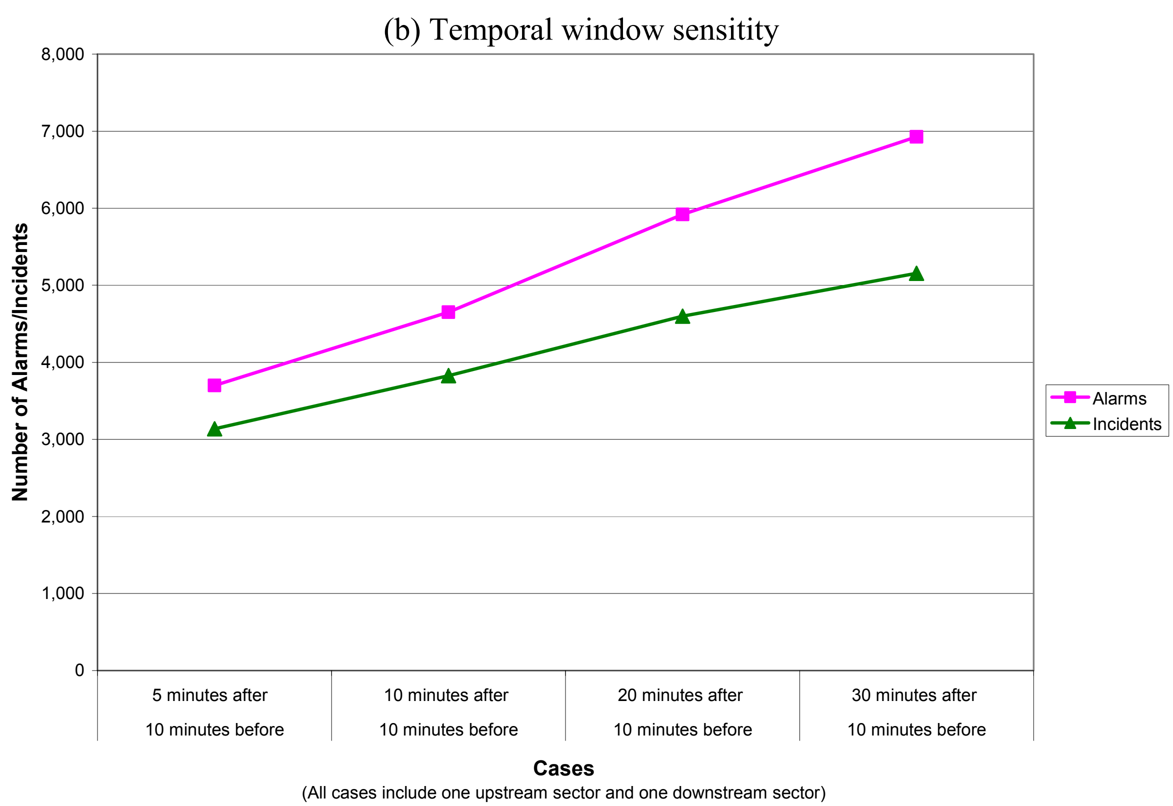


Figure 4. Sensitivity Results for Incident-Alarm Matching Query.

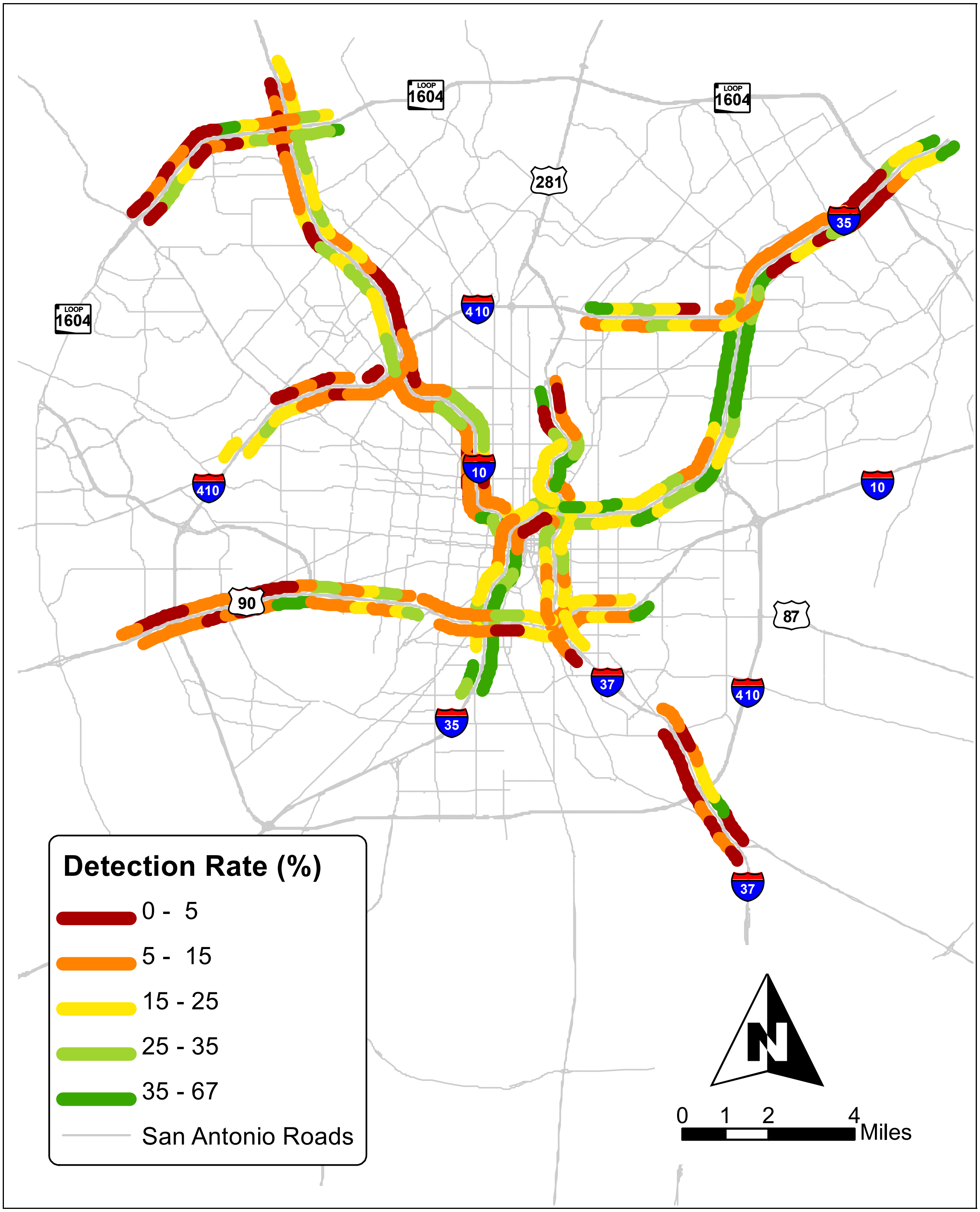


Figure 5. Distribution of Detection Rates by Sector.

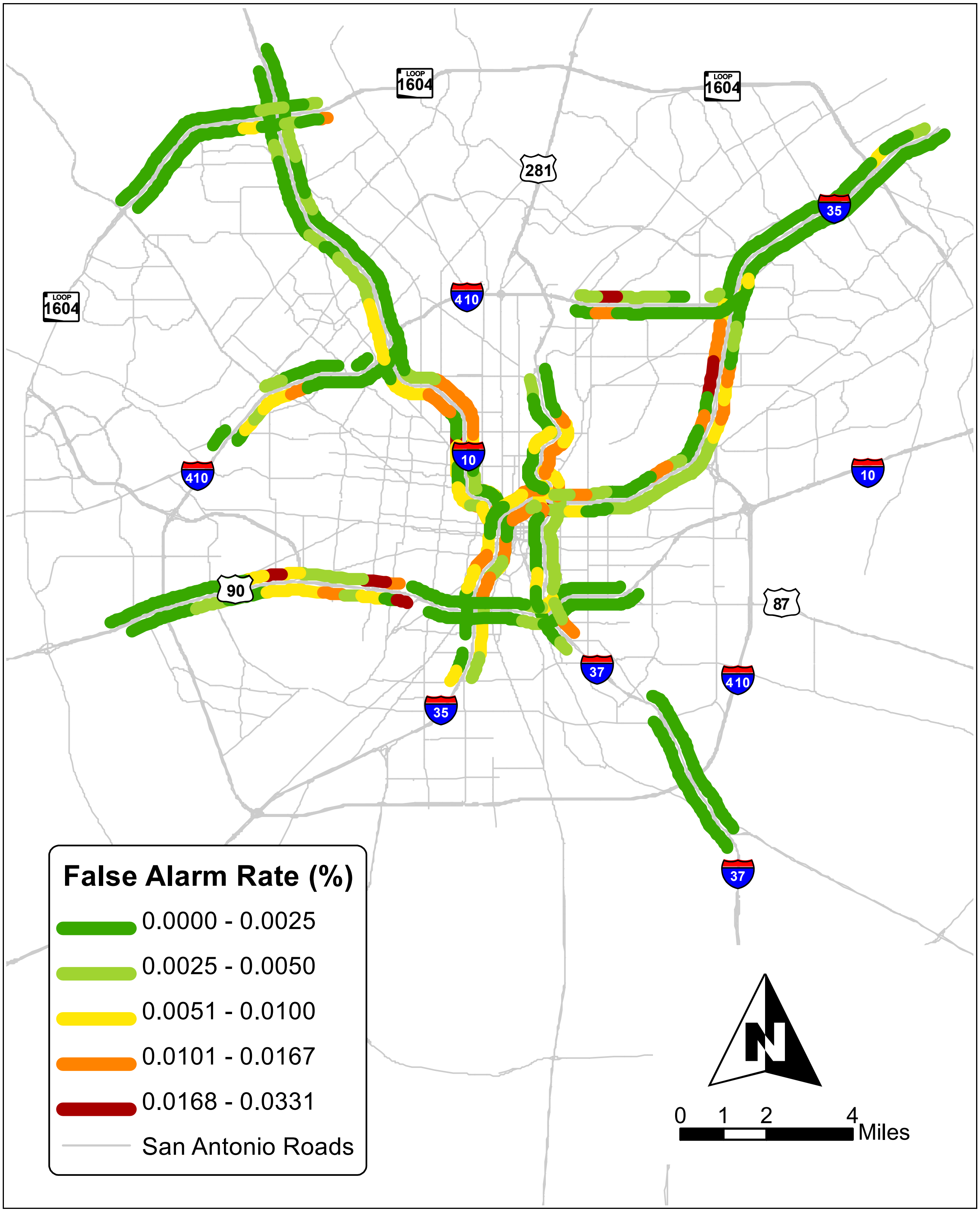


Figure 6. Distribution of False Alarm Rates by Sector.

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).

Although commonly used, the false alarm rate formulation described previously is not necessarily a good performance measure because it ignores the frequency of false alarms operators actually experience and the total number of alarms the algorithm triggers. A formulation for overall false alarm frequency is:

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

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.

INCIDENT DETECTION ALGORITHM PERFORMANCE EVALUATION TOOL

To test the feasibility of modifying alarm thresholds to optimize incident detection algorithm performance, 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 is to measure the impact of modifying speed alarm thresholds on the number and timing of alarms generated by the system. As Figure 7 shows, IDAT enables users to select one or more sectors of interest and a range of dates and time stamps. 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.

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, making it very difficult to fully replicate the archived alarm event database. To overcome this difficulty, IDAT uses an artificial "minimum recovery time" to enable an alarm to close automatically if the calculated moving average value is consistently larger than the minor alarm threshold (i.e., the moving average has "recovered") for the duration of that minimum recovery time. The minimum recovery time is subject to calibration. The current default is 15 minutes, which resulted from a calibration phase that involved varying the minimum recovery time from 5 minutes to 60 minutes in 5-minute increments until the number of generated alarms was closest to the number of alarms in the archive.

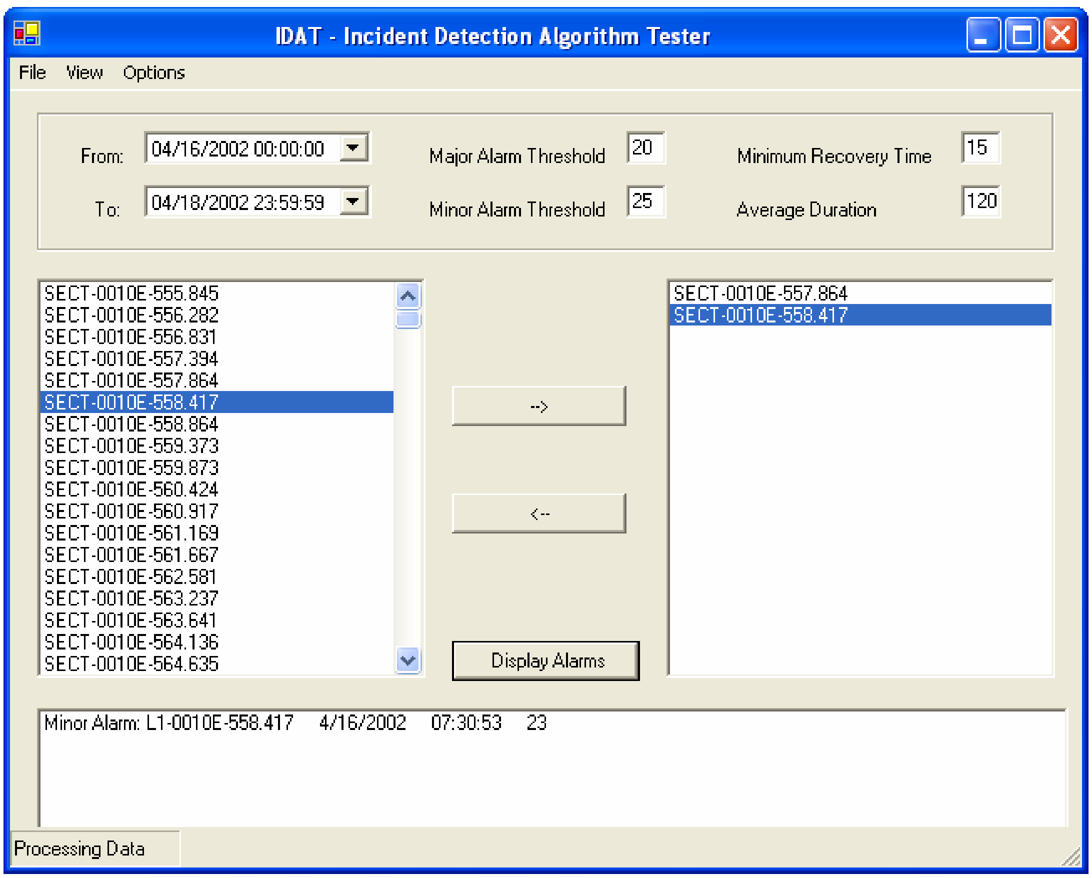


Figure 7. Incident Detection Algorithm Tester Interface.

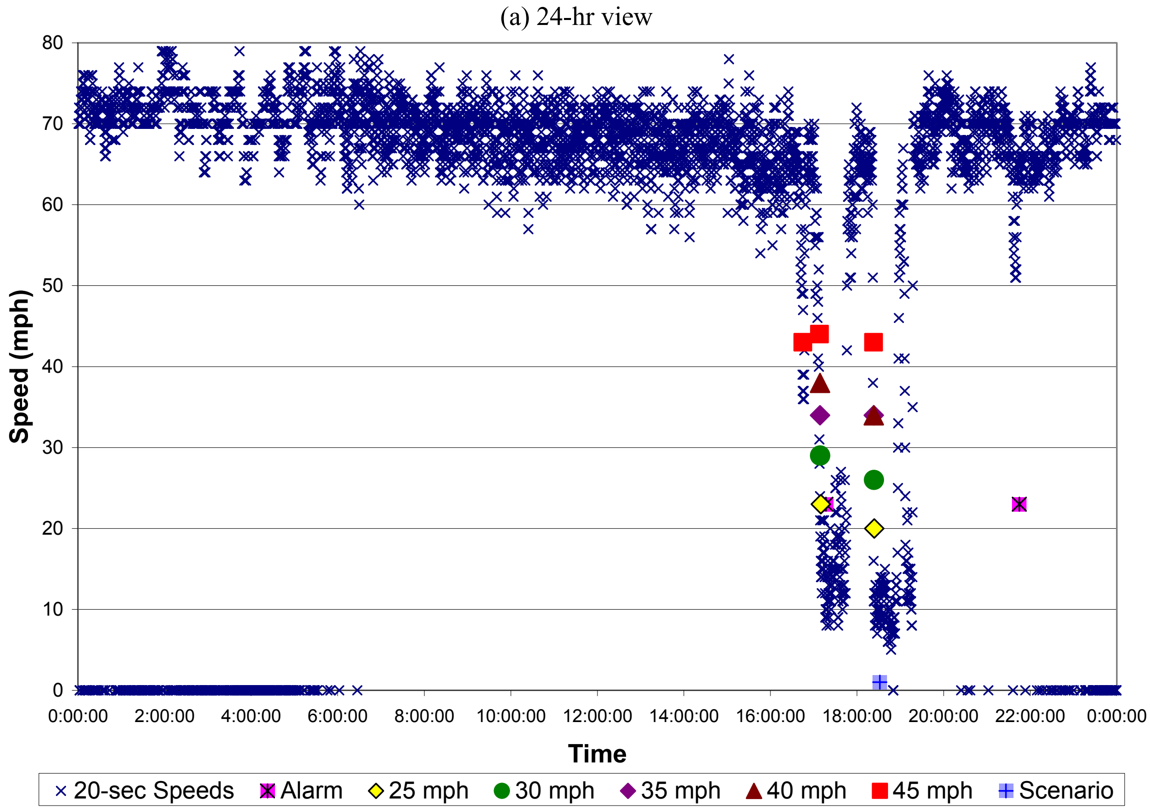
Although IDAT attempts to replicate the incident detection algorithm at TransGuide as closely as possible, its main objective is the measurement of changes in the number and timing of alarms in response to changes in alarm thresholds. IDAT outputs include listings of 2-minute moving average speeds and alarms, either in text file format or in Access format, which can serve as inputs for the generation of time series plots (using a third party application such as Excel) to document incidents and the impact of various alarm thresholds on the number and timing of alarms (Figure 8). In addition to lane data speed profiles, a typical plot could include all alarms generated at different alarm threshold levels, as well as all reported incidents for that particular sector and day. Using plots facilitates 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, it is possible to calculate measures such as average number of alarms and average incident detection time for all incidents within the same period (Figure 9).

DATA QUALITY CONTROL FLAGS

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 observation prompted the development of a set of quality control tests for detector data.

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 (*3*, *4*). 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 1 summarizes the quality control tests developed. In general, the tests 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 local control unit (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. As a reference, Figure 10 shows the location of detectors controlled by Naztec LCUs and the location of detectors controlled by TRF LCUs.

Because the inventory of lane detectors is in a GIS-based database, it is possible to examine spatial trends in the distribution of quality control flags. As an illustration, Figure 11 shows the spatial distributions for lane records with flags 2a – 2d. Assessment of temporal variations in the distribution of quality control flags is also possible. As an illustration, Figure 12 shows the temporal distribution of all quality control flags throughout the day.



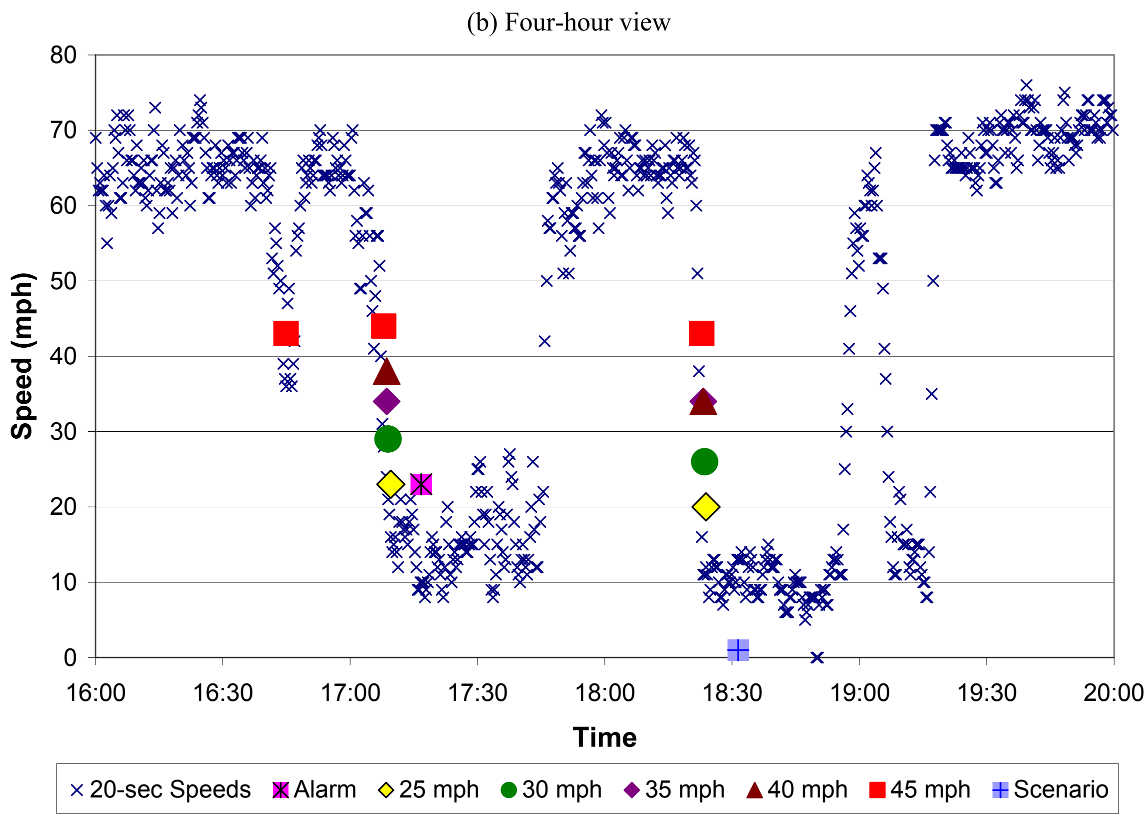


Figure 8. Sample 24-hour Speed Profile with Alarms for Different Threshold Levels.

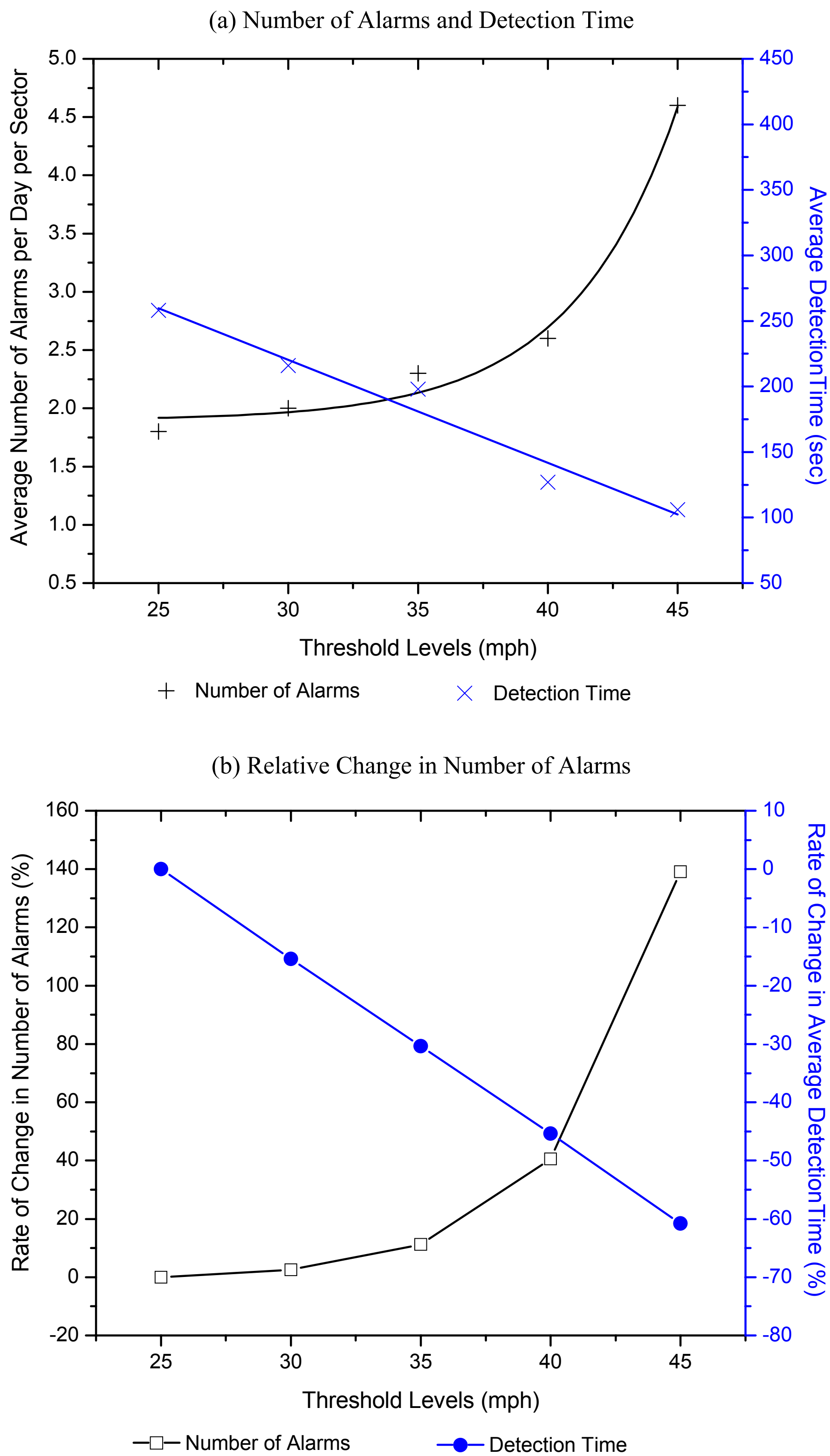


Figure 9. Impact of Alarm Thresholds on Number of Alarms and Detection Times.

Table 1. 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.

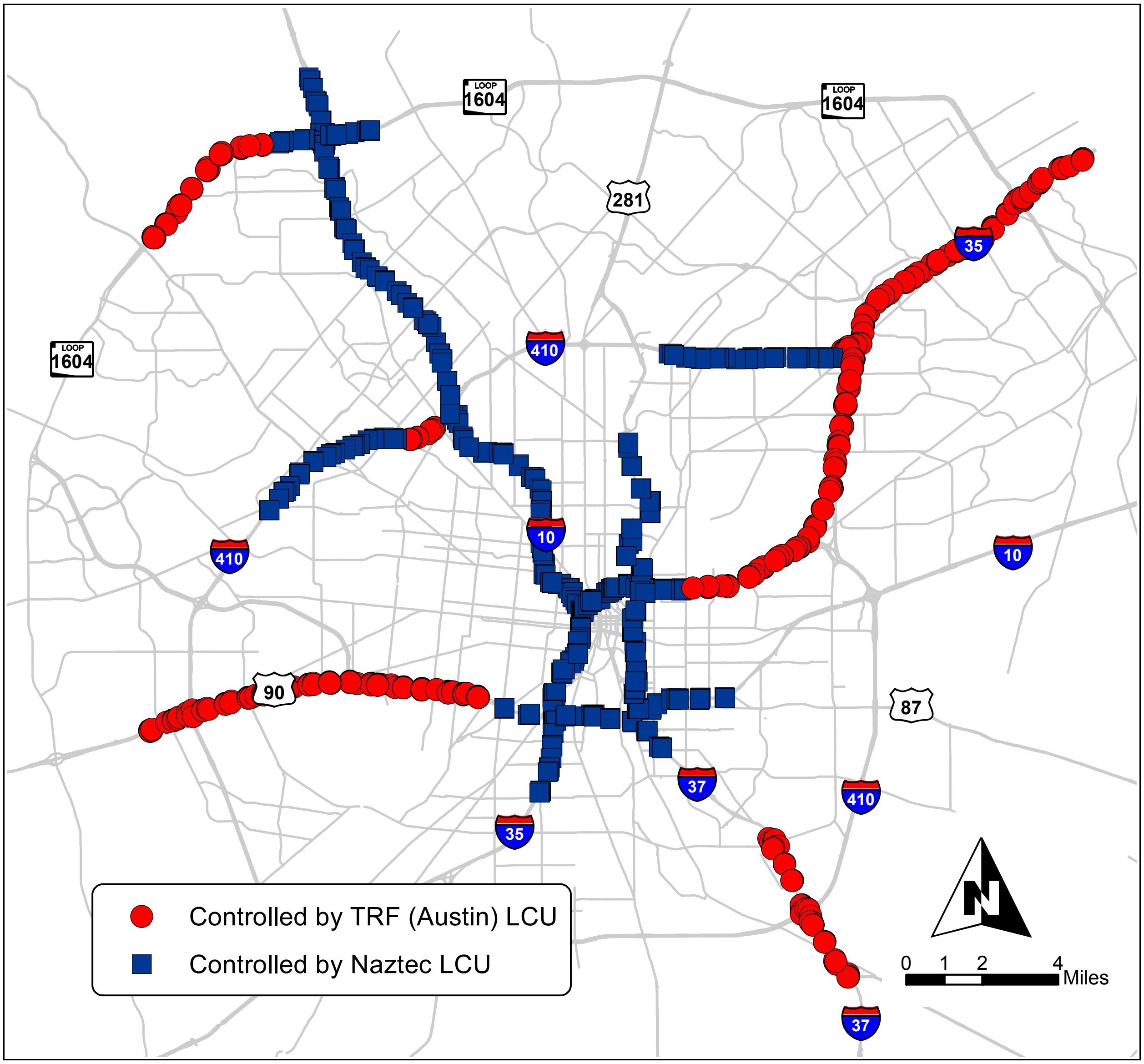
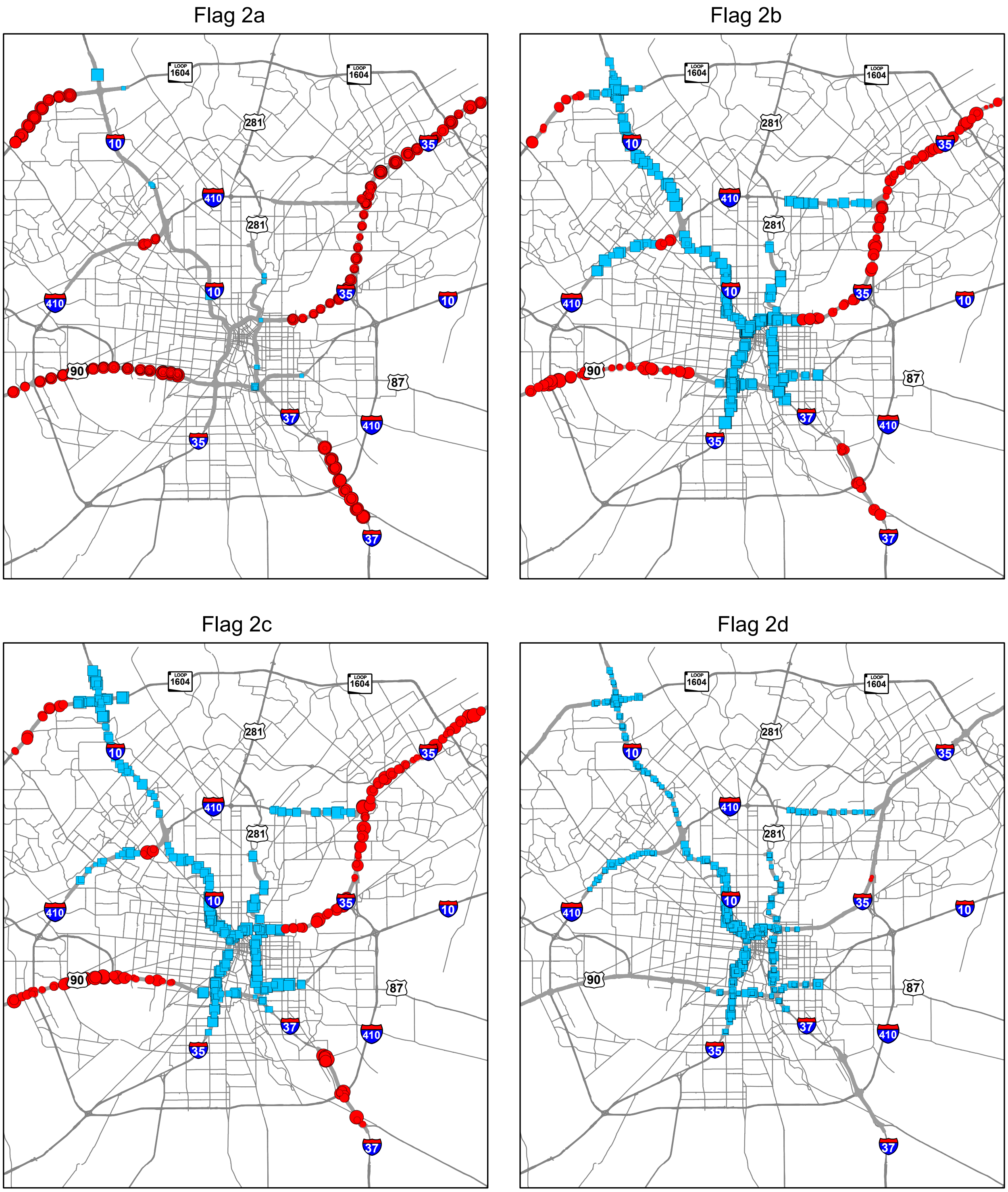


Figure 10. Detectors Controlled by Naztec LCUs and TRF LCUs.



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

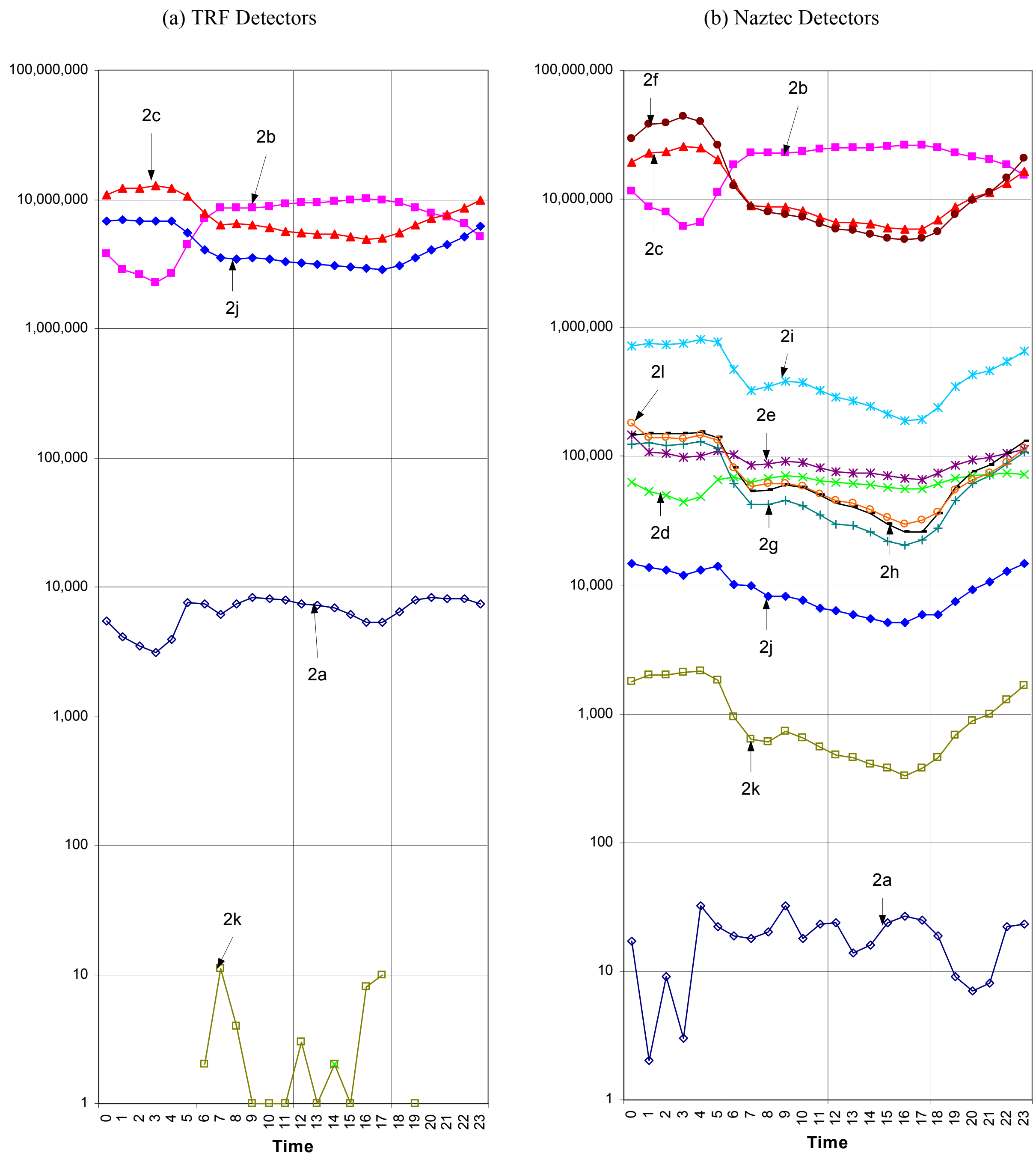


Figure 12. Temporal Distribution of Quality Control Records.

DATA COMPLETENESS ASSESSMENT

Table 1 includes quality control flag 2m to explicitly keep 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. Additional analyses may be necessary to characterize and manage missing data properly because missing data can account for a much larger percentage of flagged records than other quality control flags (20 percent versus 4 percent, respectively) (*2*).

The research focused on two types of data completeness analyses: an aggregate evaluation of completeness by LCU server and a detailed evaluation of completeness at the individual detector level. The purpose of the aggregate analysis at the LCU server level is to determine any potential completeness trends that could be attributed to system-wide causes rather than individual detectors. The analysis includes a determination of the total number of potential days with data, the effective number of days with data, and the calculation of an aggregated completeness rate (Table 2). It also includes a determination of relevant statistics such as average number of records per day, median, maximum, minimum, standard deviation, and coefficient of variation.

Table 2. 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%

Evaluating the history of records associated with individual lane detectors enables the production of aggregate summary charts and maps (Figure 13, Figure 14). For example, Figure 13 shows the percentage of detectors with completeness rates larger than a pre-specified completeness rate. According to the results, about 35 percent of detectors have a completeness rate of 95 percent or higher. At the same time, only about 10 percent of detectors have a completeness rate of 50 percent or lower. On average, the completeness rate for all detectors is 80 percent. Figure 13 also shows significant differences 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).

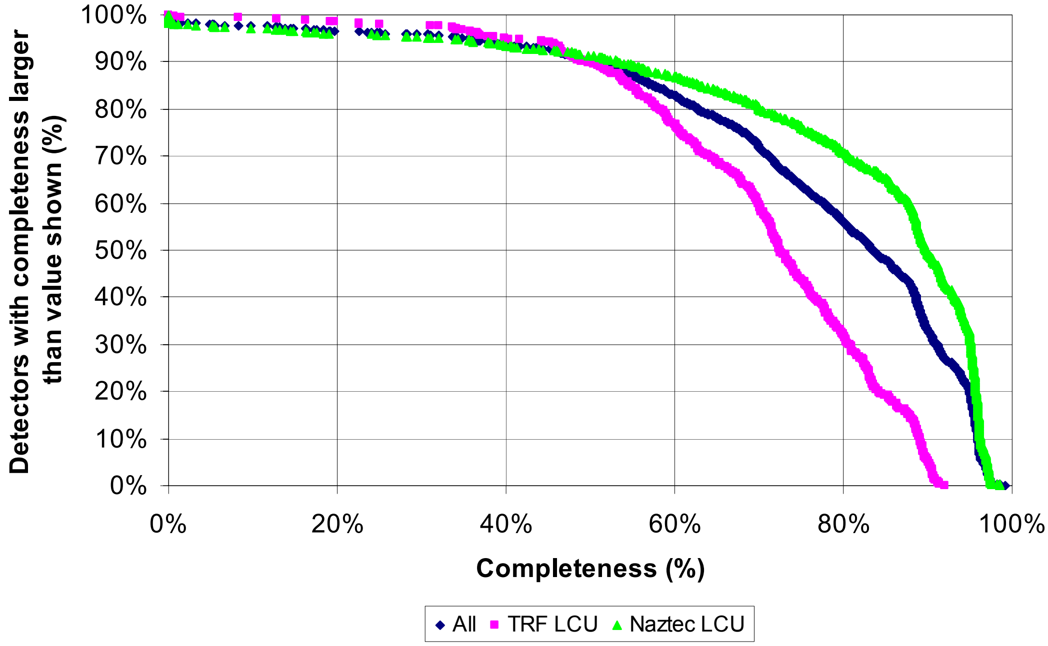


Figure 13. Detector Data Completeness Summary.

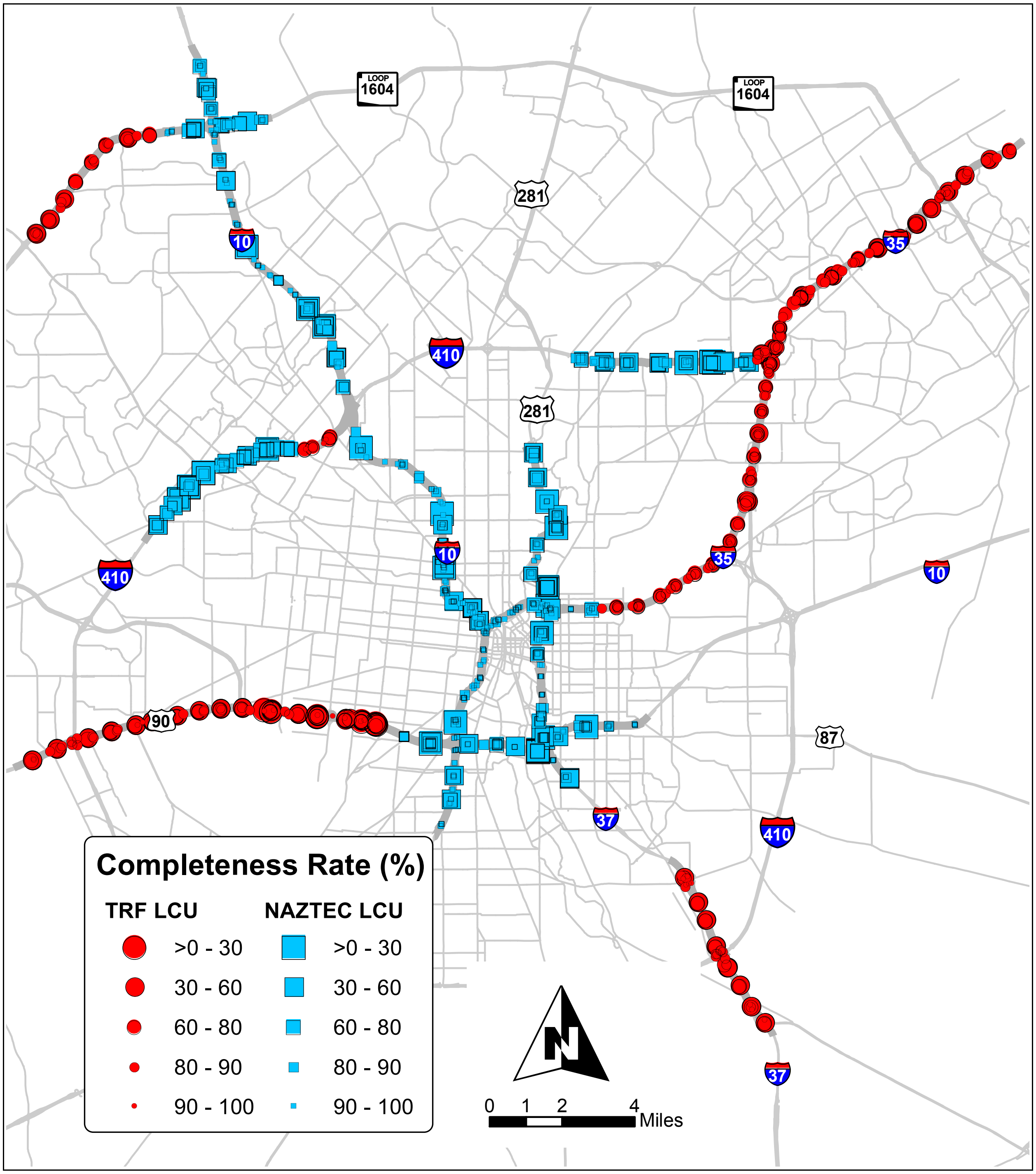


Figure 14. Spatial Distribution of Completeness Rates.

CHAPTER 3. PROCESS DEFINITIONS, SPECIFICATIONS, AND RECOMMENDATIONS FOR IMPLEMENTATION

This chapter describes Product 0-4745-P4. It includes an evaluation of process definitions and specifications for implementing the evaluation procedures discussed in Chapter 2 at TransGuide and recommendations for implementation at other TMCs around the state.

PROCESS DEFINITION AND IMPLEMENTATION SPECIFICATION ASSESSMENT

Chapter 2 described a number of procedures for evaluating incident detection practices and performance developed during the second phase of research project 0-4745. They 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. Rather than developing comprehensive process models describing how TMC personnel use incident-related data in general, which is beyond the scope of the research, the purpose of this chapter is to outline the changes that would need to take place at TransGuide in order to implement the research findings.

Process and Queries to Match Alarms and Incidents

Implementation of this procedure 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. Phase 1 of the research developed a prototype geodatabase and associated archived ITS data model, which the researchers used as a foundation for the development of the procedure to match alarm data and scenario data (*1*). Specific requirements for the implementation of the geodatabase (beyond the requirements already identified during the first phase of the research) and the alarm-scenario data matching procedure include the following:

- Incorporate the geodatabase tables into the TransGuide database design. The prototype geodatabase is in Environmental Systems Research Institute (ESRI) personal geodatabase format, which uses a Microsoft Access database engine. Exporting the geodatabase to other ESRI platforms such as Arc Spatial Data Engine (ArcSDE) or ArcInfo is straightforward. Exporting the detector unit attribute table to other database platforms such as Sybase is also straightforward.
- Develop a graphical user interface (GUI) to automate the query design shown in Figure 3. The graphical interface would provide users with the ability to select a range of dates and incident types, as well as printing and data exporting options. It is anticipated that the main users of the alarm-scenario matching query would be managers and analysts interested in evaluating incident detection trends.
- Modify the scenario header table population process to ensure the incident ID field is the same as the incident ID field in the alarm tables. As mentioned previously, the lack of a common incident ID link between the scenario tables and the event tables led to the development of a “fuzzy” spatio-temporal query methodology to match incident and

alarm data. While the "fuzzy" query approach produced meaningful results, a more accurate, effective approach would be to have the same incident ID on both datasets.

Incident Detection Algorithm Performance Evaluation Tool

Implementation of this procedure 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. The researchers had access to a copy of the incident detection algorithm source code provided by TransGuide officials. Unfortunately, the complexity of the code prevented its full replication under laboratory conditions, which led to the development of a separate prototype application. Nonetheless, the application developed mimicked the main components of the algorithm and successfully enabled the measurement of changes in the number and timing of alarms in response to changes in alarm thresholds.

Implementing a tool similar to IDAT within the TransGuide environment would involve the following requirements:

- Develop a relational database archive of 20-second speed, volume, and occupancy data. The research included the development of a prototype database in Oracle covering some 3.4 billion lane data records from March 2002 through April 2004. To accelerate data access and querying, the researchers divided the lane data archive into quarter lane data tables and associated index tables, each one containing some 500 million 20-second speed, volume, and occupancy records. Implementation of the lane data archive on the actual TransGuide Sybase production server would likely result in a different structure stemming from a careful analysis to balance data access speed requirements and hardware/software system characteristics and performance.
- Develop code and corresponding GUI to include the minimum recovery time concept implemented in IDAT (Figure 7). As mentioned previously, the purpose of the minimum recovery time is to mimic the alarm closing process by enabling an alarm to close automatically if the calculated moving average value is consistently larger than the minor alarm threshold for at least the duration of the minimum recovery time. By default, IDAT uses a minimum recovery time of 15 minutes. It is likely that actual implementation will require a calibration phase. 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, therefore contributing to optimize real-time operations.

While fully integrating a tool similar to IDAT within the TransGuide environment would be the most effective strategy, it would still be possible to use IDAT directly, provided a few minor changes are made, including the database source (the current version of IDAT automatically points to the prototype Oracle database using an open database connectivity (ODBC) protocol) and the lane table structure in Sybase resembles the lane table structure used in IDAT. For small experimental uses of IDAT, a potential additional feature would be to develop the capability to read flat files directly (and, if necessary, convert those files into temporary database tables).

Data Quality Control Flags

Implementation of this procedure would involve making changes to the way the LCU subsystem manages field data. Since there is relatively little control over the characteristics and functionality of the LCU software used in the field, the assumption here is that the implementation of the quality control tests and flags will take place at the LCU server level. Specific requirements for such implementation include the following:

- Create a lookup table in Sybase to list and describe the various quality control tests and flags used. That lookup table would be accessible to any users who need access to lane data, both within and outside TransGuide.
- Develop a module to conduct data quality control tests and assign flags to the affected records immediately after the LCU servers receive lane data from the field LCUs. The impact on overall system performance is expected to be minor because the quality control tests operate at the individual record level and do not require an examination of previous records (with the exception of test 1b, which can be executed right before appending the lane data records to the archive database).
- Add a unique date/time stamp to the lane data archive that does not depend on the seasonal changes between central standard time (CST) and central daylight time (CDT). The unique date/time stamp could be the Unix time function TransGuide uses throughout the rest of the system. As documented elsewhere, the use of the time function would effectively eliminate the duplicate record problem that occurs when time changes back one hour from CDT and CST (*2*). Because not too many users outside TransGuide use Unix, it would be advisable to use the time function in addition to (not instead of) the local date/time stamp used in the archive. For the same reason, it would also be advisable to flag all records from 1 – 2 AM right after the change from CDT to CST to alert users about the time change.

Data Completeness Assessment

Implementation of this procedure would involve developing the relational database containing 20-second speed, volume, and occupancy data mentioned previously. Additional requirements include the following:

- Automate the queries to derive aggregate data completeness measures, both at the LCU server level and at the individual detector level.
- Develop code and associated GUI to automate the query process needed to produce summary tables, charts, and maps such as those shown in Table 2, Figure 13, and Figure 14. The graphical interface would provide users with the ability to select a range of dates and LCU servers, as well as printing and data exporting options.

IMPLEMENTATION AT OTHER TMCS AROUND THE STATE

The procedures described in the previous sections are, by necessity, customized to suit TransGuide's characteristics and needs. However, some of the procedures, in particular those related to data quality control and completeness, are generic enough to enable partial or total implementation at other TMCs around the state. As mentioned previously, ITS data quality control and completeness has long been the subject of inquiry and research work (*3*, *4*). The quality control tests developed as part of this research built on those efforts by providing a formal spatial data modeling component that considerably enhances the ability to visualize and understand data quality trends. Some of the recommendations for implementation at other TMCs are similar to those already listed in the case of TransGuide. They are listed here once again for completeness:

- Develop formal data models and efficient data archives (preferably in relational database format) of raw speed, volume, and occupancy data. Conducting data quality analyses is considerably more difficult when using flat files, particularly if the structure of the data in those flat files does not facilitate efficient data accessing and indexing. Considering the increasing level of interest in archived ITS data by many users in the transportation community at large and the decreasing costs associated with the implementation of data archival systems, generating formal ITS data archives appears increasingly feasible.
- Add a lookup table to the database design to list and describe the various quality control tests and flags used. That lookup table would be accessible to any users who need access to lane data, both within and outside the TMC.
- Develop modules to conduct data quality control tests and assign flags to the affected records immediately after receiving lane data from the field. The impact on overall system performance is expected to be minor because most quality control tests operate at the individual record level and do not require an examination of previous records.
- Add unique date/time stamps to the lane data archive that do not depend on the seasonal changes between CST and CDT. Realistically, the unique date/time stamps would rely on a system time function or independent reference that would be added to (but not replace) the local date/time stamp used in the archive.

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 Detection Optimization and Data Quality Control*. Research Report 0-4745-3. Draft. Texas Department of Transportation, Austin, Texas, 2005.
3. Turner, S. *Guidelines for Developing ITS Data Archiving Systems*. Report 0-2127-3. Texas Department of Transportation, Austin, Texas, September 2001.
4. 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.