

# QUALITY CONTROL OF ARCHIVED INTELLIGENT TRANSPORTATION SYSTEMS DATA

**Khaled Hamad<sup>1</sup>**

<sup>1</sup> *Department of Civil & Environmental Engineering, University of Sharjah, Sharjah, United Arab Emirates*

*Received 29 December 2014; accepted 17 July 2015*

**Abstract:** Previous research has reported 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. This paper summarizes the work completed to address quality control and completeness issues associated with a very large archived ITS data set composed of some 3.4 billion 20-second lane detector data records from San Antonio's TransGuide. The paper includes a description of the quality control tests utilized, the results of the analysis conducted, and a discussion of ITS data completeness issues. 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 traffic was light. Consequently, there was a higher chance for time intervals without vehicles which produced abnormal detector readings. Finally, the researchers evaluated the data completeness both at the aggregate level (by server) and a more detailed individual detector level. While the analysis described in this paper uses data from one jurisdiction (San Antonio, Texas), the methodology is sufficiently generic to enable implementation at other traffic management centers.

**Keywords:** data quality control, data completeness, ITS archived data, TransGuide.

## 1. Introduction

There are many applications for archived ITS data such as incident detection and traffic condition forecasting. Implicit in these applications is the assumption of good quality of the detector data. Previous research has reported 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. This paper will touch

on a number of issues related to erroneous loop detector data (e.g., incorrect traffic records) and missing data (in relation to the need for data imputation).

Despite the importance of the ITS data quality assessment, few researchers have tackled this issue in the past; furthermore research in this area has not been very active in recent years.

Turochy and Smith (2000) used a screening procedure consisting of a set of five pass/

<sup>1</sup> Corresponding author: [khamad@sharjah.ac.ae](mailto:khamad@sharjah.ac.ae)

fail tests. A good loop detector data record is one which passes all five tests and an erroneous record is one that fails one or more of these tests. The set of tests consists of one threshold test for traffic volume, one for occupancy, one that screens for reasonable values of a derived average vehicle length, and two tests that filter out inconsistent combinations between the three major traffic data parameters (i.e. speed, volume, and occupancy).

Turner et al. (2000) proposed a similar approach with additional tests; the main ones of which are the consideration of spatial and historical consistency of a record. The spatial consistency test checks if the data reported by a particular loop detector are consistent with data reported by adjacent upstream and downstream detectors. The historical data consistency check verifies that the record is similar to historical mean values for that location, day of week, and time of day.

Ishak (2003) proposed a fuzzy clustering for measuring level of uncertainties in loop detector data. The developed approach clusters the input space of the three traffic parameters (speed, occupancy, and volume) into regions of highly-concentrated observations. A parameter, representing level of uncertainties, is derived from the membership grade and a decaying function of the normalized Euclidean distance. The parameter can be used to identify erroneous data.

Wall and Dailey (2003) used a time-series approach for detecting and correcting errors in archived traffic data. In their work, consistency of vehicle counts is used to evaluate the validity of the data: if vehicles counts are balanced, the data are valid; if

vehicle counts are not balanced, the data are not valid. A correction factor can also be determined.

Vanajakshi and Rilett (2004; 2006) used a check of conservation of vehicles over a series of detectors to identify and correct errors in traffic volume data. A constrained nonlinear optimization approach, namely generalized reduced gradient method, was adopted to select the objective function and constraints such that the result will follow the vehicle conservation principle with least deviation from original data. Weijermars and Van Berkum (2006) also used the principle of conservation of vehicles to detect invalid traffic data from single-loop detectors at signalized intersections.

Smith and Venkatanarayana (2007) developed a general methodology for tailoring data quality screening, and detector health monitoring. Similar to previous research, the proposed quality assessment methodology calls for checking the validity of traffic data values (i.e. reasonableness to represent a traffic state in the real world).

More recently, Widhalm et al. (2011) presented a methodology for identifying faulty traffic detectors. Floating-car data was used as an independent source of information for distinguishing actual changes in traffic conditions as opposed to detector faults. Instead of applying rule-of-thumb rules, fault detection was based on residuals of a nonlinear regression model fitted to detector and floating-car data.

Corey et al. (2011) proposed a method for identifying and correcting sensitivity errors of inductive loop data that impact the quality of data obtained from loop detectors. A Gaussian mixture model was used to identify

loop detectors with sensitivity problems, and to calculate a correction factor for correctible detectors.

Saavedra et al. (2011) developed an automated quality assurance methodology to identify unreliable archived transit data, namely Automatic Vehicle Location (AVL) and Automatic Passenger Counting (APC) data. Their method is based on observed/expected pattern limitations derived from archived data for bus travel time and passenger activity. Suspect data flagged using stop-level tests to identify irregular patterns.

Finally, Lin et al. (2012) developed a fuzzy logic approach to quantify the reliability of traffic detector data. The proposed method incorporates three separate assessments: fundamental consistency using thresholds checks, network consistency against neighboring detectors, and historical consistency against past data from same location.

This research utilizes quality control tests for loop detector data, which, for the most part, follow previous research efforts by Turochy and Smith (2000) and Turner et al. (2000). The set of quality control tests applied includes two levels: first-level quality control tests, which apply to raw data files, and second-level quality control tests, which could apply either to raw-data file records or database table records. The authors decided to utilize this threshold-based quality control assessment because it was simple to implement online prior to archiving the loop detector data.

The paper describes the quality control tests utilized, the results of the analysis conducted, and a discussion of ITS data completeness issues. Part of the effort also

includes an evaluation of temporal variations in the distribution of quality control flags for loop detector data.

## 2. Data Description

TransGuide, San Antonio's Traffic Management Center (TMC), uses a database environment to store data describing ITS equipment characteristics and data needed to support day-to-day operational activities at the center. TransGuide maintains a long-term data repository in compressed, flat file format, including: 20-second loop detector data, event data (triggered by the automatic incident detection algorithm), and a scenario log (archives of manually entered events and actions taken by operators). The 20-second loop-detector data archive includes speed, volume, and percent occupancy. Each record contains a date and time stamp, the detector address, and the corresponding average speed (in mph), volume (number of cars in 20-second period), and percent occupancy values.

The researchers evaluated two-year worth of data, totaled at some 3.4 billion 20-second individual lane loop-detector data records. Using this large sample size was beneficial because it enabled the observation of quality control trends. It may be worth noting that two types of Lane Control Unit (LCU) and associated software were operational at TransGuide: Naztec LCUs and TxDOT Traffic Operations Division (TRF) LCUs (also called Austin LCUs). It is therefore of interest to determine if different types of LCUs produce different quality control and data completeness results.

## 3. Data Quality Assessment Methodology

The quality control tests developed as part of this research were built on previous efforts;

although, by necessity, the quality control tests underwent modifications to suit the needs of the research. Table 1 shows the set of quality control tests the researchers applied. The set includes two levels of quality control tests. The first-level (namely tests 1a and 1b) tests for records with incorrect format and for duplicate records. Therefore, first-level tests normally do not result in individual record flags. The second-level quality control tests, which can result in quality control flags being assigned to individual records, were used to distinguish two types of records: “valid” records and “abnormal” records.

“Valid” records are records with valid volume and occupancy values but have invalid “by design” speed values (-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). Tests 2b, 2c, and 2f are the tests used to check for “Valid” records. “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. All tests except 2b, 2c, and 2f are used to check for “Abnormal” records.

Initially, the researchers considered adding a third quality control which takes into consideration basic functional relationships between flow rate, speed, and percent occupancy on the premise that such functional relationships could be used to identify potential outlier data that may be the result of equipment failure (Park et al., 2003). While theoretically sound, however, it was not possible to calibrate the parameters of the statistical model because the model also

depended on vehicle length, which currently is not part of the data collection protocol at TransGuide. Nonetheless, other TMCs that collect vehicle length data could potentially benefit from the application of that test.

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 to assess quality control flag values afterwards. 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 this threshold was exceeded. Because these values were too large for what would be typical volume for a 20 second period, researchers further examined these records. In all these cases, there were gaps in the data; and apparently volume data after the gaps were too large 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.

**Table 1**  
*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

## 4. Quality Control Analysis

### 4.1. First-Level Tests

Because of the structure of the flat file lane data archive, 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. 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. Similarly, 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.

### 4.2. Second-Level Tests

Table 2 summarizes the second-level quality control tests. 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.

- Four tests (2b, 2c, 2f, and 2j) accounted for about 99-percent of all flagged records. The distribution of these flags in both types of LCUs was almost similar with the exception of flag 2j which was experienced by TRF LCUs only.
- There were clear 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 to enable it to generate null speed (to avoid problems associated with the use of zero speeds), zero volume, and zero occupancy records when no vehicles were present during the 20-second recording interval.

Furthermore, the researchers examined temporal (by hour of the day) variations in the distribution of quality control flags. As Fig. 1 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. 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). TransGuide officials have indicated that during evening hours, TRF LCUs sometimes produce records with extremely high values, particularly speeds. The flag 2a trend in Fig. 1 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 per-cent 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). Fig. 2 summarizes the results of the analysis. Fig. 2(a) shows that most records were between 1 and 75 mph (typical of freeway driving conditions), with a rapidly de-creasing trend between 75 and 93 mph, and a few lingering records above 93 mph. Overall, Fig. 2(a) suggests a reasonable upper speed threshold of about 93 mph (which would translate to about 0.01 percent of records exceeding that threshold).

Fig. 2(b) 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, Fig. 2(b) suggests a reasonable upper volume threshold to be around 18 (which would translate to 0.1 percent of records exceeding that threshold). 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.

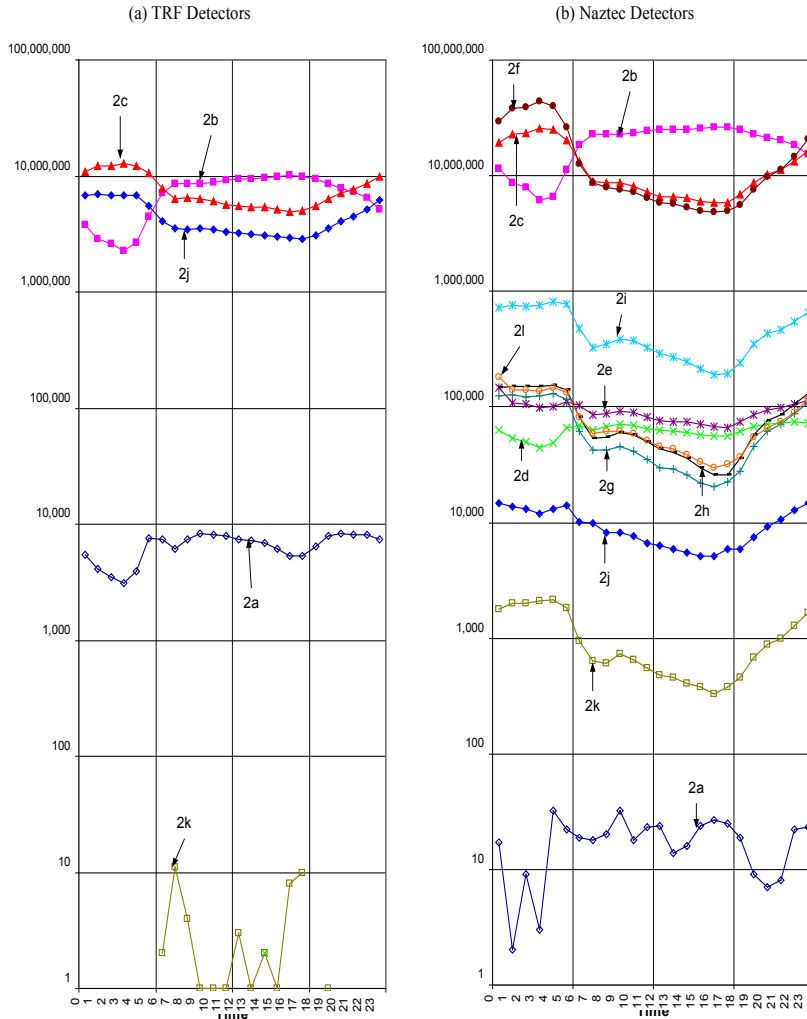
Fig. 2(c) shows that, with the exception of a handful of records, practically all records

had occupancy rates less than or equal to 100 percent. Overall, Fig. 2(c) suggests that a reasonable upper occupancy threshold is 99 percent (which would translate to only 0.006 percent of records exceeding that threshold).

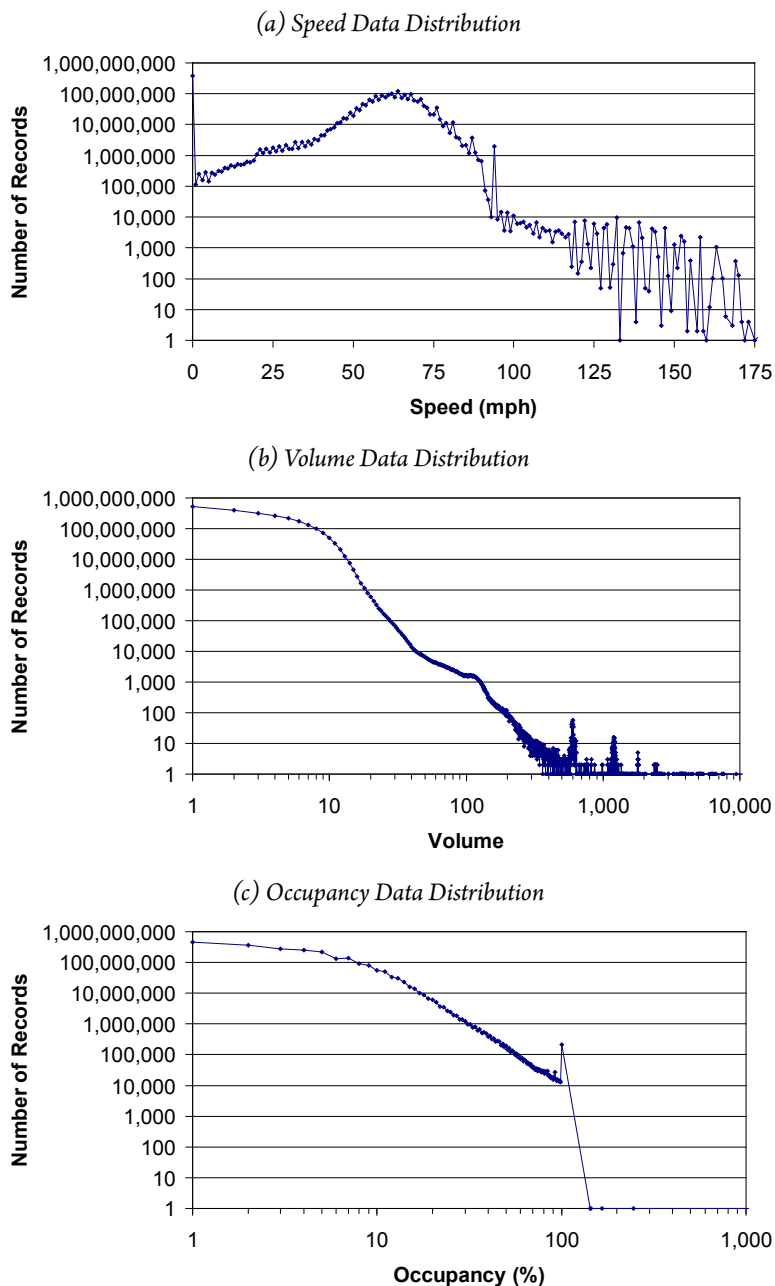
**Table 2**  
Summary of 20-Second Lane Records Flagged Using Level 2 Tests

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							





**Fig. 1.** Temporal Distribution of Quality Control Records (*x*-axis: hour of the day; *y*-axis: number of records)



**Fig. 2.**  
*Speed, Volume, and Occupancy Data Distributions*

## 5. 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 operates six LCU servers (Table 3). With the exception of Server 6, which started processing detector data at a later time compared to the others, the remaining servers were supposed to be operational and processing data during the 792-day analysis period. Table 3 shows there were several days during this 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.

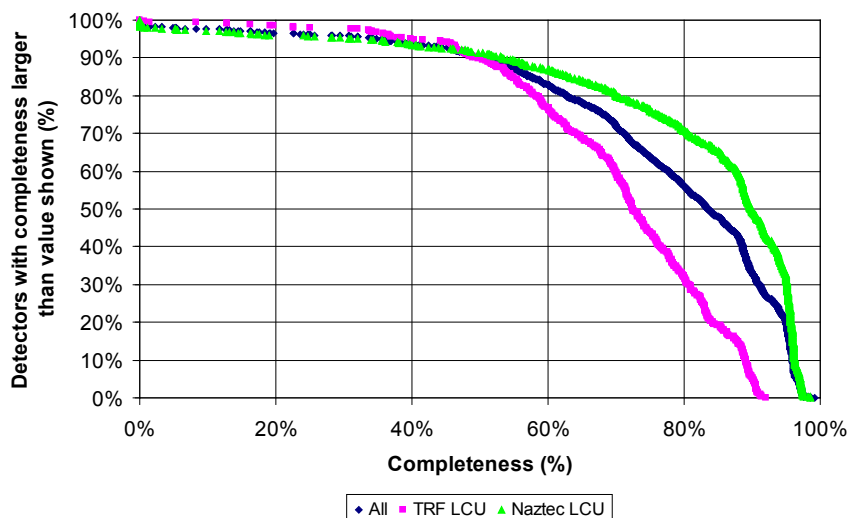
Table 3 also shows a wide range in the total number of records per day associated with

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 Fig. 3 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. 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.

**Table 3**  
*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%



**Fig. 3.**  
*Detector Data Completeness Summary*

## 6. Conclusion

This paper summarizes the work completed to address quality control and completeness issues associated with a very large archived ITS data set composed of some 3.4 billion 20-second lane detector data records from San Antonio's TransGuide. The paper includes a description of quality control tests utilized, the results of an analysis conducted, and a discussion of ITS data completeness issues. Approximately 1.5 billion flagged records were "valid" records and the remaining 126 million (about 3.7 percent) flagged records were "abnormal" records. 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. Finally, the researchers evaluated the data completeness both at the aggregate level (by server) and a more detailed individual detector level. At the LCU server level, the completeness rate 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.

The researchers recommend implementing the revised data quality control flags automatically. This 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.

The researchers also recommend continuing 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. Finally, though the analysis described in this paper used data from one jurisdiction (San Antonio, Texas), the methodology is sufficiently generic to enable implementation at other TMCs that also archive loop detector data.

## Acknowledgements

The work documented in this paper was funded by the Texas Department of Transportation. The authors appreciate the help and assistance provided by San Antonio's TransGuide. The views expressed by the authors do not necessarily reflect the views or policies of the Texas A&M Transportation Institute or the Texas Department of Transportation.

## References

- Corey, J.; Lao, Y.; Wu, Y.-J.; Wang, Y. 2011. Detection and Correction of Inductive Loop Detector Sensitivity Errors by Using Gaussian Mixture Models, *Transportation Research Record: Journal of the Transportation Research Board*. DOI: <http://dx.doi.org/10.3141/2256-15>, 2256: 120-129.
- Ishak, S. 2003. Fuzzy-Clustering Approach to Quantify Uncertainties of Freeway Detector Observations, *Transportation Research Record: Journal of the Transportation Research Board*. DOI: <http://dx.doi.org/10.3141/1856-02>, 1856: 6-15.
- Lin, D.-Y.; Boyles, S.; Valsaraj, V.; Waller, S.T. 2012. Reliability Assessment for Traffic Data, *Journal of the Chinese Institute of Engineers*. DOI: <http://dx.doi.org/10.1080/02533839.2012.655466>, 35(3), 285-297.
- Park, E.; Turner, S.; Spiegelman, C. 2003. Empirical Approaches to Outlier Detection in Intelligent Transportation Systems Data, *Transportation Research Record: Journal of the Transportation Research Board*. DOI: <http://dx.doi.org/10.3141/1840-03>, 1840: 21-30.
- Saavedra, M.; Hellinga, B.; Casello, J. 2011. Automated Quality Assurance Methodology for Archived Transit Data from Automatic Vehicle Location and Passenger Counting Systems, *Transportation Research Record: Journal of the Transportation Research Board*. DOI: <http://dx.doi.org/10.3141/2256-16>, 2256: 130-141.
- Smith, B.L.; Venkatanarayana, R. 2007. A New Methodology for Customizing Quality Assessment Techniques for Traffic Data Archives, *Transportation Research Record: Journal of the Transportation Research Board*. DOI: <http://dx.doi.org/10.3141/1993-22>, 1993: 165-174.
- Turner, S.; Albert, L.; Gajewski, B.; Eisele, W. 2000. Archived Intelligent Transportation System Data Quality: Preliminary Analyses of San Antonio Transguide Data, *Transportation Research Record: Journal of the Transportation Research Board*. DOI: <http://dx.doi.org/10.3141/1719-10>, 1719: 77-84.

Turochy, R.E.; Smith, B.L. 2000. New Procedure for Detector Data Screening in Traffic Management Systems, *Transportation Research Record: Journal of the Transportation Research Board*. DOI: <http://dx.doi.org/10.3141/1727-16>, 1727: 127-131.

Vanajakshi, L.; Rilett, L.R. 2004. Loop Detector Data Diagnostics Based on Conservation-of-Vehicles Principle, *Transportation Research Record: Journal of the Transportation Research Board*. DOI: <http://dx.doi.org/10.3141/1870-21>, 1870: 162-169.

Vanajakshi, L.; Rilett, L. 2006. System Wide Data Quality Control of Inductance Loop Data Using Nonlinear Optimization, *Journal of Computing in Civil Engineering*. DOI: [http://dx.doi.org/10.1061/\(ASCE\)0887-3801\(2006\)20:3\(187\)](http://dx.doi.org/10.1061/(ASCE)0887-3801(2006)20:3(187)), 20(3): 187-196.

Wall, Z.; Dailey, D. 2003. Algorithm for Detecting and Correcting Errors in Archived Traffic Data, *Transportation Research Record: Journal of the Transportation Research Board*. DOI: <http://dx.doi.org/10.3141/1855-23>, 1855: 183-190.

Weijermars, W.; Van Berkum, E. 2006. Detection of Invalid Loop Detector Data in Urban Areas, *Transportation Research Record: Journal of the Transportation Research Board*. DOI: <http://dx.doi.org/10.3141/1945-10>, 1945: 82-88.

Widhalm, P.; Koller, H.; Ponweiser, W. 2011. Identifying Faulty Traffic Detectors with Floating Car Data. In *Proceedings of 2011 IEEE Forum on Integrated and Sustainable Transportation System*, June 29-July 1 2011, Vienna, 103-108.