TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
2019, Vol. 2673(6) 165–175

DOI: 10.1177/0361198119841286
journals.sagepub.com/home/trr
SAGE

Designing a Comprehensive Procedure for Flagging Archived Traffic Data: A Case Study

Zhi Chen[1], Xiao Qin[1], Elizabeth Schneider[2], Yang Cheng[3], Steven Parker[3], and Md. Razaur Rahman Shaon[1]

Abstract

Archived data management systems (ADMS) are extensively used for storing historical traffic data (e.g., volume, speed, occupancy) collected from traffic sensors. Archived traffic data have important uses for engineering and planning applications such as ramp meter timing, work zone planning, and performance management. They are also an important data source for transportation research. Various flagging procedures have been implemented in ADMS to identify invalid or questionable archived traffic data, however, those flagging procedures may not be comprehensive enough to maintain adequate data quality. This study presents the findings of a literature search and a user survey to discuss the possible gap between the state-of-the-practice and the state-of-the-art validity tests, identifies complex yet effective validity tests which are favored by users, and recommends the procedure that prioritizes the implementation of validity tests in ADMS. To aid the implementation, different methods to establish quantitative rules and practical thresholds for candidate validity tests have been proposed. This study underscores the importance of keeping the basic validity tests required to maintain minimum data quality and adding more advanced tests to detect less obvious yet important data issues. The recommended tests along with the flagging procedure are demonstrated through a case study based on one detector station in Wisconsin. Results of the case study show that the guide is useful in the development of a comprehensive flagging procedure for better data quality.

Archived data management systems (ADMS) are used extensively to store historical traffic data collected from traffic sensors such as loop detectors and microwave detectors. Traffic data are essential for off-line analytics such as transportation planning, congestion monitoring, and performance measures. Quality control (QC) of archived data is critical for more efficient and practical use of this immense data source. Data quality is also a critical part of quality assurance (QA) to data customers. Therefore, data validity tests are a key component in any ADMS to ensure the provision of quality traffic data to support informed decision making in traffic operations, planning, and other traffic management activities.

Various ADMS have implemented data validity tests to flag obvious erroneous or potentially invalid traffic data. However, according to a thorough review of validity tests implemented in ADMS in eight different states (*1*), in the state of the practice, the validity tests implemented in various ADMS were less complex than those proposed in the literature. ADMS might fail to maintain sufficient data quality with currently available validity tests, especially if data quality issues are not easily discernable. A set of comprehensive yet tangible validity tests is needed.

The aim of this study is to identify and recommend useful and practical data validity tests and to guide the development of a flagging procedure for potential implementation. User preferences, programming complexity, and flagging performance were all considered during the design process. User preferences for potential validity tests can be collected from user surveys; the validity tests with high user preferences and low programming complexity are then evaluated. Based on the flagging performance of validity tests, a flagging procedure can be proposed to prioritize for implementation efficiency. The guide was demonstrated through a case study based on

[1]Department of Civil and Environmental Engineering, University of Wisconsin-Milwaukee, Milwaukee, WI
[2]Wisconsin Department of Transportation, Milwaukee, WI
[3]Traffic Operations and Safety (TOPS) Laboratory, Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI

Corresponding Author:
Address correspondence to Xiao Qin: qinx@uwm.edu

Wisconsin's ADMS: V-SPOC (i.e., Volume, Speed, Occupancy). A variety of methodologies for establishing rule-based or data-driven criteria for validity tests were developed, and the validity test criteria and flagging procedure are demonstrated in a case study.

Literature Review

Over the years, a number of validity tests have been proposed and applied to flag invalid traffic records and assess system-wide data quality for archived traffic data. Typical QA/QC procedures can be classified into two categories: (a) univariate and multivariate range checks, which validate a single traffic variable or a combination of traffic variables (e.g., traffic volume, speed, and occupancy) against predetermined thresholds (e.g., the minimum, maximum, and/or appropriate range of values); and (b) temporal consistency based on comparison between historical trends and patterns and spatial consistency checks which evaluate the consistency of traffic observations made from nearby detector locations (e.g., upstream or downstream locations or adjacent lanes if performed on lane-specific basis) (*1*). Additionally, diagnostics based on detailed sensor signal outputs like the sensor on/off time have been developed to identify detector errors such as the sensitivity issue, pulse-breakup errors, and chronic splashover errors (*2–6*). Detector diagnostics primarily focus on inductive loops, however, and cannot be carried out for other sensor types (*7*); therefore, they are not within the scope of this study.

Turner conducted a thorough review of validity tests implemented in the nine ADMS of different states (*1*): ADMS Virginia, California PeMS (Performance Measurement System), CATT (Center for Advanced Transportation Technology) Lab in Maryland, Central Florida Data Warehouse, FHWA Mobility Monitoring Program, Kentucky ADMS, Phoenix RADS (Regional Archived Data Server), PORTAL (Portland Oregon Regional Transportation Archive Listing), WisTransportal V-SPOC. One of the conclusions of his study is that the validity tests implemented across the nine ADMS are closely alike, and most are less complex than the contemporary methods in the research literature (*1*). A gap seems to exist between the state of the practice and the state of the art for validity tests, possibly because of the challenges in implementation. Common validity tests across the nine ADMS include univariate range checks (i.e., minimum and maximum values) for traffic variables and some zero consistency checks. Zero consistency checks flag invalid records which have a zero value in one variable but a non-zero value in another variable. Turner recommended the common validity tests for implementation in ADMS as basic validity tests because they were used in most ADMS and could guarantee the minimum data quality (*1*). The basic validity tests can flag obvious data errors, but they may not be able to discover data issues which are not easily discernable. For example, average effective vehicle length (AEVL) computed from traffic speed, flow, and density can effectively flag subtle data errors which may pass regular univariate and multivariate range checks. One study shows that the AEVL range test identified 77.2% of invalid data while basic checks flagged less than 30% of invalid data (*8*). Despite the proven effectiveness in research (*8*, *9*), AEVL has been implemented in only two ADMS: ADMS Virginia and Central Florida Data Warehouse. Therefore, the state-of-the-practice of QA/QC procedures in various ADMS may not be able to maintain sufficient data quality.

Other validity tests proposed in the literature to identify invalid or questionable traffic data are summarized in Table 1. These tests are comprised in three categories: multivariate range check, temporal consistency check, and spatial consistency check. The checks for multivariate range have been proposed to identify problematic trends which deviate from the fundamental relationship of any pair of two traffic variables (*10–13*). Notably, most of the proposed tests examine traffic volume or speed by occupancy because of its theoretically restricted range from 0% to 100%. Overall, the implementation of traffic flow theory based multivariate range checks in ADMS is limited.

Temporal consistency checks can be categorized into two groups: Type I and II. Type I considers data points in a single time interval. The typical applications include checking to see if there are too many zeros in a one-hour time period. Several ADMS implemented tests similar to those proposed in Hu et al. (*16*) and Schmoyer et al. (*17*). Type II compares the patterns in a single time interval with historical patterns. Typical applications include the similarity of the time-series volume profile between different days. Temporal consistency checks are lacking in the current ADMS, especially Type II checks. Among Type II checks, Turner's method was claimed to require a manual check on the plots (*1*), but this test can be automated with measures like the statistical correlation. The approach proposed by Lin et al. is based on a fuzzy classifier (*18*); however, the complex procedure and fuzzy pass/fail decision rules make the implementation less desirable.

Spatial consistency checks have not been implemented in any of the nine ADMS surveyed by Turner, possibly because spatial consistency checks are difficult to define and measure because of being affected by information other than traffic data (e.g., highway geometries, terrain, presence of ramps, lane position, and distance to the nearby detectors). Spatial consistency checks can be grouped by traffic pattern and the law of conservation. Tests of traffic patterns have been introduced in Lu et al. (*9*) and Kwon et al. (*19*). The principle of vehicle

Table 1. Alternative Validity Tests

Literature	Validity test	Rationale
Multivariate range check		
Nihan et al. (*10*) and Jacobson et al. (*11*)	Expected volume/occupancy ratio ranges based on four different occupancy ranges	Speed, derived from the volume/occupancy ratio, should be within the expected range depending on the occupancy region based on findings of Hall et al. (*14*, *15*)
Cleghorn et al. (*12*)	Maximum acceptable deviation from historical averages of traffic variable pairs (e.g., speed–density) based on established number of standard deviations (note: the data regime is divided into many small bins for practical purposes)	Traffic variable pairs should be within expected two-dimensional confidence interval based on historical data
Shi et al. (*13*)	1. If (VOL>166.67 or OCC>35%) and SPD>60, then invalid 2. If (SPD>60 or OCC<20%) and VOL>166.67, then invalid 3. If (SPD>60 or VOL <125) and OCC>35%, then invalid	1. Speed cannot be too high when the traffic is congested; 2. Flow volume cannot be too high in free-flow traffic; 3. Occupancy cannot be too high in free-flow traffic;
Temporal consistency check		
Type I		
Hu et al. (*16*) and Schmoyer et al. (*17*)	Maximum number of consecutive zeros and maximum number of consecutive repeating non-zeros in traffic volume counts	The probability of observing a specific traffic count during one time interval follows the Poisson distribution
Type II		
Turner (*1*)	Deviation of traffic profile of one day from traffic profiles of other similar days	Similar days should have similar overall traffic patterns
Lin et al. (*18*)	Temporal consistency using a fuzzy classifier based on data of one detector for the same day of the week, at the same time of day, or at the same time of day and day of the year	Traffic data in the present time interval should be similar to that in a similar past time interval
Spatial consistency check		
Traffic pattern		
Kwon et al. (*19*)[a]	The strong correlation between data from spatially close stations with no major disturbances such as freeway interchanges between them	Similar traffic pattern between spatially close stations
Lu et al. (*9*)	No significant difference should exist in the average AEVL between adjacent lanes or across neighboring stations	Similar traffic pattern between spatially close stations
Principle of vehicle conservation		
Nihan (*20*), Vanajakshi and Rilett (*21*)	The cumulative storage rate between the downstream and upstream detector cumulative flows cannot exceed the maximum number of vehicles that can be accommodated on the road between these two detectors at jam density	Principle of vehicle conservation
Nihan (*20*)	A steadily increasing or decreasing cumulative storage rate indicates that one location is consistently over- or under- counting vehicles	The trend of cumulative storage rate should fluctuate over time
Wall and Daily (*22*)	Cumulative differences between traffic counts from a valid reference station and those with a time lag from the target station should be around zero	Principle of vehicle conservation

[a]This study aims to identify mislabeled detectors, but the proposed method is claimed to have the potential to be extended to anomaly detection.

conservation is expressed as the storage rate in the validity checks in Nihan (*20*) and Vanajakshi and Rilett (*21*). Storage rate is the difference in vehicle volumes between the upstream mainline detector with on-ramp detectors and downstream mainline detector with off-ramp detectors in the same time interval. The cumulative storage rate over time should always be equal to the traffic on the segment between adjacent mainline detectors.

Table 2. Candidate Validity Tests

Validity test	Description
Basic check	
Missing data check[a]	If any traffic variable has a missing value, then invalid
Univariate range checks[a]	If the value of any traffic variable is out of the feasible range, then invalid
Zero consistency checks[a]	If not all three traffic variables are zero when there is no passing vehicle, then invalid
Multivariate range check	
High free-flow volume	If the volume is too high in the free-flow traffic, then invalid
Infeasible speed in congestion	If the speed is out of the feasible range in congestion, then invalid
Infeasible AEVL[b]	If AEVL is out of the feasible range, then invalid
Temporal consistency check	
Repeating zero volume	If there are too many repeated zero volumes, then invalid
Non-zero occupancy stuck	If there are too many repeated non-zero occupancies, then invalid
Abrupt change in volume	If the change in volume exceeds the maximum feasible volume change, then invalid
Abrupt change in speed	If the change in speed exceeds the maximum feasible speed change, then invalid
Anomalous daily traffic pattern	If the daily traffic profile deviates substantially from that in similar days based on a similarity measure (e.g., correlation coefficient), then further investigation is needed
Spatial consistency check	
Inconsistent traffic at adjacent upstream and downstream detector stations	If the traffic data two adjacent upstream and downstream detector stations are substantially different based on a similarity measure (e.g., correlation coefficient), then further investigation is needed
Maximum remaining vehicles in the segment	If remaining vehicles in the segment measured by the cumulative storage is larger than the maximum segment storage of vehicles (i.e., the product of the segment length and the jam density), then invalid
Infeasible trend of remaining vehicles in the segment	If remaining vehicles in the segment measured by the cumulative storage has a continuously increasing or decreasing trend based on a trend measure (e.g., linear trend coefficient), then invalid

[a]State of practice: implemented in all or most of the nine ADMS surveyed in Turner (*1*).
[b]State of practice: implemented in two ADMS: ADMS Virginia and Central Florida Data Warehouse.

Therefore, the value should not exceed the maximum allowable traffic volume or the product of the segment length and the jam density at any time. Moreover, the value should fluctuate over time rather than consistently increase or decrease. In general, spatial consistency checks usually need to use time-series data from different locations, and therefore require more effort in implementation than temporal consistency checks.

User Survey

ADMS users will determine how archived traffic data are applied, so their input is valuable for helping select validity tests and improve the design of the flagging procedure. A user survey can collect users' opinions of the flagging procedure. An online user survey was designed for this study to help guide crucial decisions regarding the development of a comprehensive flagging procedure. The survey included questions regarding how users apply traffic data, preferred traffic variables, data specifications like time frame and resolution, current methods of checking data quality, and user preferences regarding candidate validity tests which were proposed in the literature. The survey was sent to 118 V-SPOC users, and a 47% response rate was received. Among 54 responses, 37 had complete answers.

Survey responses showed that volume and speed are preferred traffic variables. Most respondents used data with low to moderate resolution (volume per 5–15 minutes or hourly). Very few used V-SPOC data for daily volume or intervals of one minute or less. The information from the survey suggests that new validity tests for volume and speed and those with a five-minute (5-min) resolution are preferred. Therefore, most of the proposed validity tests are for volume and speed and are evaluated using 5-min traffic data from the V-SPOC system.

Table 2 presents basic checks and eleven additional candidate validity tests based on users' suggestions and preferences. Note that basic checks are considered as required validity tests. Eleven additional candidate validity tests are ordered by user preference and programming complexity. Users generally prefer multivariate range checks and temporal consistency checks, and these two categories are also easier to program in ADMS than spatial consistency checks. In addition to low user preference and high programming complexity, the first candidate spatial consistency checks may perform inadequately when the distance between detectors is too large. Additionally, the latter two checks need additional information aside from traffic data. Moreover, it could be challenging to determine which detector carries invalid traffic data, as it is a possibility for all. Spatial consistency checks are therefore not recommended.

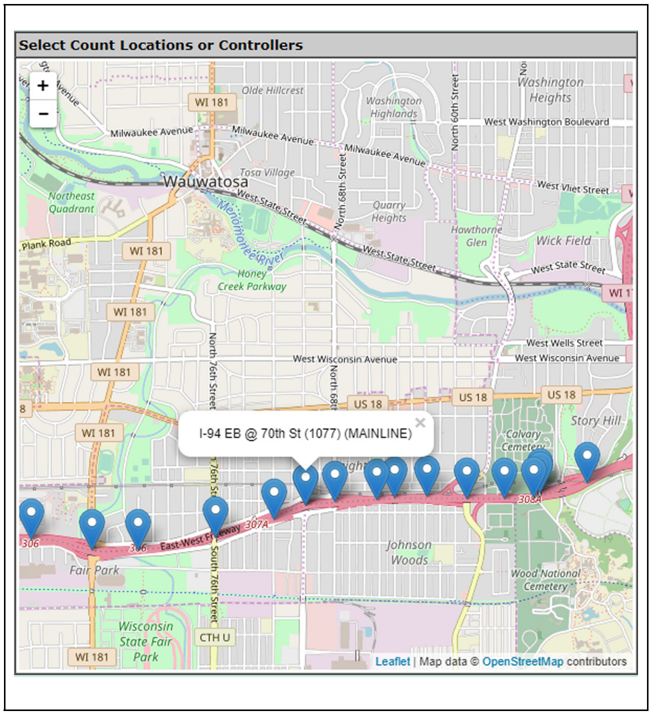

Figure 1. Location of the study station.

Validity Test Evaluation

Based on the literature review and user feedback, three types of validity tests were evaluated by their effectiveness in flagging invalid traffic data: basic checks, alternative multivariate range checks, and temporal consistency checks. Basic checks are fundamental univariate and multivariate range checks which can provide minimum data quality. Alternative multivariate range checks which examine one single traffic record, and temporal consistency checks which examine multiple traffic records, help flag additional invalid data. Corresponding rules of these checks are first specified using either rule-based or data-driven method. The rule-based method establishes the rule by specifying a reasonable threshold; the data-driven method is based on theory, and forms the expected threshold based on the sample traffic data. One mainline detector station (ID 1077) on I-94 EB in Wauwatosa, Wisconsin, was selected as the study station (Figure 1). The freeway section has three lanes with a posted speed limit of 55 mph.

Basic Checks

Turner (*1*) recommended some validity tests which are now widely accepted in various data archives to guarantee minimum data quality. These basic tests screen invalid records, and include: missing data check, univariate range check, and zero consistency check.

- *Missing data check*: This test flags missing records so that they will not be checked again by all the other validity tests.
- *Univariate range check*: Univariate range checks examine if traffic variables such as volume, speed, and occupancy exceed the possible range. Valid ranges for volume, speed, and occupancy are set as [0, 3,100 vphpl], [0, 100 mph], and [0%, 100%], respectively. Note that the 5-min volume is converted to the equivalent hourly volume.
- *Zero consistency check*: Zero consistency checks inspect whether all three traffic variables are recorded as zero when any one is zero. If there are no passing vehicles during the time interval when any of the three traffic variables is zero, all three variables should have zero values.

Alternative Multivariate Range Checks

Basic checks include univariate range checks and fundamental multivariate range checks. Alternative multivariate range checks complement basic checks by considering more complex relationships between multiple traffic variables. The three alternative multivariate range checks are: high free-flow volume, infeasible speed in congestion, and infeasible AEVL. The rules of these three checks were established using the data-driven method. The former two checks are based on the traffic occupancy to differentiate congestion from free-flow state. Based on the plots of volume against occupancy in Figure 2*a* with two red dashed lines representing 5% and 30% occupancy, respectively, it is safe to say that the traffic is in a free-flow state when the occupancy is below 5% and in congestion when the occupancy is above 30%. Figure 2*b* plots the speed against the occupancy. The speed spreads widely when the occupancy is below 5% and no clear pattern shows up. However, the speed seems to be within a certain range when the occupancy exceeds 30%. Therefore, the infeasible speed in congestion check would be constructed only for occupancies above 30%.

High Free-Flow Volume. A conservative maximum volume value is set as 1,200 vphpl when the occupancy is below 5%. This check is defined as: if VOL>1,200 vphpl and OCC < 5%, then invalid. Note that VOL and OCC stand for the equivalent hourly volume and occupancy, respectively.

Infeasible Speed in Congestion. The volume–occupancy plot in Figure 2*b* shows that a triangular fundamental diagram (TFD) may approximate the traffic records well. The right branch of a TFD depicts the linear relationship between the volume and the density in a state of

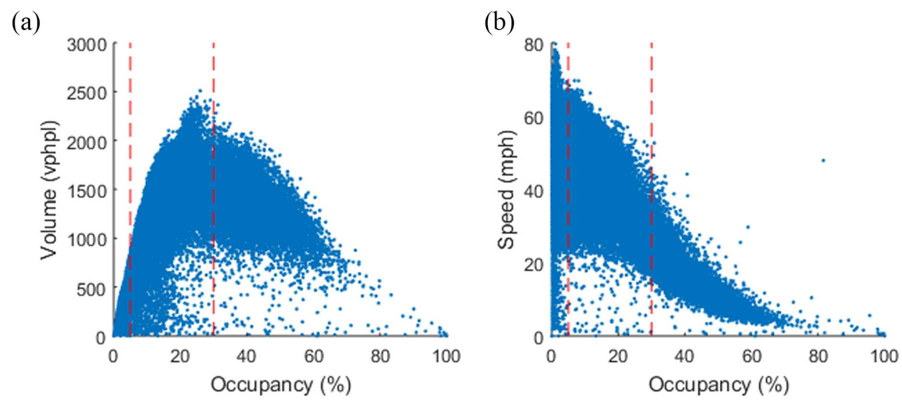

Figure 2. (*a*) Volume–occupancy plot and (*b*) Speed–occupancy plot.

congestion. The occupancy is proportional to the density given a g-factor, so the volume is defined by:

$$VOL = a{*}OCC + b \tag{1}$$

And the speed could be expressed as:

$$\begin{aligned} SPD &= \frac{VOL}{Density} = \frac{VOL}{g{*}OCC} = \frac{a{*}OCC + b}{g{*}OCC} \\ &= \frac{b}{g}{*}\frac{1}{OCC} + \frac{a}{g} = a'{*}\frac{1}{OCC} + b' \end{aligned} \tag{2}$$

where VOL, SPD, and OCC stand for volume, speed, and occupancy, respectively.

Figure 3 shows that the speed is linearly related with $\frac{1}{Occupancy}$. However, the variance of the speed varies as $\frac{1}{Occupancy}$ changes, which would violate the assumption of constant variance of error terms for the ordinary least squares (OLS) estimation method of linear regression. In this case, quantile regression was used to estimate the linear relationship (*23*). Two red dashed lines in Figure 3 present fitted linear lines based on 97.5th and 2.5th regression quantile estimates. The 95% confidence region is $798{*}\frac{1}{OCC} - 10 < SPD < 1658{*}\frac{1}{OCC} - 16$. Note the unit of OCC is %.

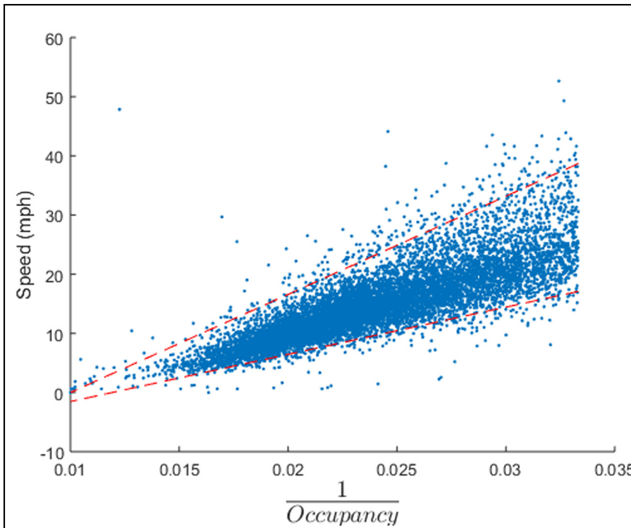

Figure 3. Plot of speed versus 1/occupancy.

Infeasible AEVL. This test checks the range of AEVL derived based on all three variables. The equation of AEVL is as follows:

$$\begin{aligned} AEVL\left(\frac{ft}{veh}\right) &= \frac{SPD\left(\frac{mi}{h}\right){*}\frac{1\,ft}{1mi}{*}OCC{*}1\%}{VOL\left(\frac{veh}{h}\right)} \\ &= \frac{SPD(mi/h){*}OCC}{VOL(veh/h)}{*}(5280/100) \end{aligned} \tag{3}$$

AEVL is a validity test which takes advantage of all three traffic variables and has shown its effectiveness in uncovering hidden errors that could pass some regular univariate or multivariate range checks, or both. Many researchers have used AEVL to identify data errors (*8*, *9*). The proposed range of AEVL values is [9 ft, 60 ft]. A traffic record is invalid if its AEVL value is beyond this range. Note that this test is not applicable when the volume is 0.

Temporal Consistency Checks

Five temporal consistency checks are proposed for consideration. Four of the proposed checks belong to Type I: repeating zero volume, non-zero occupancy stuck, abrupt changes in volume, and abrupt changes in speed. The rule-based method was used to determine the rules of the former two tests, and the data-driver method was applied to form the rules of the latter two tests. One of the proposed checks belongs to Type II, an anomalous daily traffic pattern, and its rule was decided using the data-driven method.

Repeating Zero Volume. Hu et al. (*16*) and Schmoyer et al. (*17*) introduced validity checks to test the maximum consecutive zeros in traffic volume counts based on the Poisson probability. However, according to Shi et al., stuck detectors or hardware failures may include cases in which the figures do not change over time, repeat every other record, or oscillate between two values (*13*). Therefore, a validity check for repeating zero volumes which do not have to be consecutive is more appropriate for the V-SPOC system. The rule-based method was applied to determine the rule of this test based on the probability of observing a certain number of repeated zero volumes assuming that the arrival rate of traffic flow follows the Poisson distribution.

According to the Poisson distribution, the possibility of observing zero vehicles in 5-min would be $P = e^{-\mu}$, where μ is the average 5-min volume. The possibility of observing k zeros in N 5-min intervals would be

$$p(k) = C_N^k \cdot P^k \cdot (1-P)^{N-k}, k = 0, 1, \ldots, N \quad (4)$$

and the possibility of observing more than J zeros in N 5-min intervals, which is also the probability of false flagging with J zeros as the threshold, would be

$$p(k>J) = \sum_{J+1}^{N} p(k) \quad (5)$$

Given the large volume of traffic data, a very small false flagging probability (e.g., 0.1%) should be adopted as the minimum value. N is set to be 8, and the criterion is determined as: if $\mathrm{VOL(t)} = 0$ & $\sum_{i=-4, i\neq 0}^{4} I(\mathrm{VOL(t)} = \mathrm{VOL(t+i)}) > J$ in 6:00 a.m. to 10:00 p.m./11:00 p.m. to 5:00 a.m. if the average historical 5-min volume is μ, then invalid. VOL(t) represents the 5-min volume in time interval t, and $I(\mathrm{VOL(t)} = \mathrm{VOL(t+i)})$ is the indicator function, which equals 1 if $\mathrm{VOL(t)} = \mathrm{VOL(t+i)}$, and 0 otherwise. For example, if the average historical 5-min volume in 6:00 a.m. to 10:00 p.m. is five vehicles, then J would be 2 for 6:00 a.m. to 10:00 p.m; if the average historical 5-min volume in 10:00 p.m. to 6:00 a.m. is one vehicle, J would be 7 for 11:00 p.m. to 5:00 a.m. Two time regimes, 6:00 a.m. to 10:00 p.m. and 11:00 p.m. to 5:00 a.m., are proposed since the average volume could be very different in these two regimes and may require different threshold values.

Non-Zero Occupancy Stuck. The non-zero occupancy stuck checks whether repeating non-zero traffic records are possible because of a device stuck issue. The occupancy, instead of volume or speed, is measured directly by the detector to check the stuck issue (*24*). This check has been adopted in California's data warehouse, California PeMS. The proposed criterion for this test was adjusted from a similar test in Shi et al. (*13*): if $1\% < \mathrm{OCC(t)} < 100\%$ and $\sum_{i=-6, i\neq 0}^{6} I(\mathrm{OCC(t)} = \mathrm{OCC(t+i)}) > 3$, then invalid. OCC(t) represents the 5-min occupancy in time interval t, and $I(\mathrm{OCC(t)} = \mathrm{OCC(t+i)})$ is the indicator function, which equals 1 if $\mathrm{OCC(t)} = \mathrm{OCC(t+i)}$, and 0 otherwise. The test checks to see if one non-zero 5-min occupancy repeats more than 15 minutes (i.e., three repeated 5-min occupancies) in a one-hour period when the occupancy is between 1% and 100%. The occupancy range is restricted because of the rounding issue when the occupancy is very low. The V-SPOC system rounds the occupancy to two decimal places, and it could make two very small but different occupancies appear the same. For example, 0.334% and 0.331% are both recorded as 0.33% in the V-SPOC system.

Abrupt Change in Volume. The abrupt change in volume test examines whether the change in volume is too abrupt. This and the *abrupt change in speed* test are inspired by Turner's theory that tests can identify invalid data given rapid fluctuations in values across successive time periods (*1*). This test compares the volume of one traffic record and the average volume of its adjacent records in time. The equation for this test is as follows:

$$\Delta V_t = V_t - \frac{V_{t-1} + V_{t+1}}{2} \quad (6)$$

where ΔV_t represents the change in volume at time t and V_{t-1}, V_t, and V_{t+1} are the volumes in time periods $t-1, t$ and $t+1$. Note all these volumes are in the same lane.

Figure 4 presents the time-series plot of ΔV_t. It shows the majority of ΔV_t s within the interval of [–500, 500] and no temporal variation is present. Since ΔV_t s are roughly symmetric, it would be more convenient to set a threshold for the absolute value of ΔV_t, that is, $|\Delta V_t|$. Based on the sensitivity analysis of flagged record counts with different threshold values of $|\Delta V_t|$ as shown in Figure 5, the maximum value of $|\Delta V_t|$ is set to be 600 vphpl beyond which the number of flagged records starts to decrease marginally. Hence, the criterion of this test is: if $|\Delta V_t| > 600$ vphpl, then invalid.

Abrupt Change in Speed. Similar to the previous test, this test examines whether the change in speed is too abrupt.

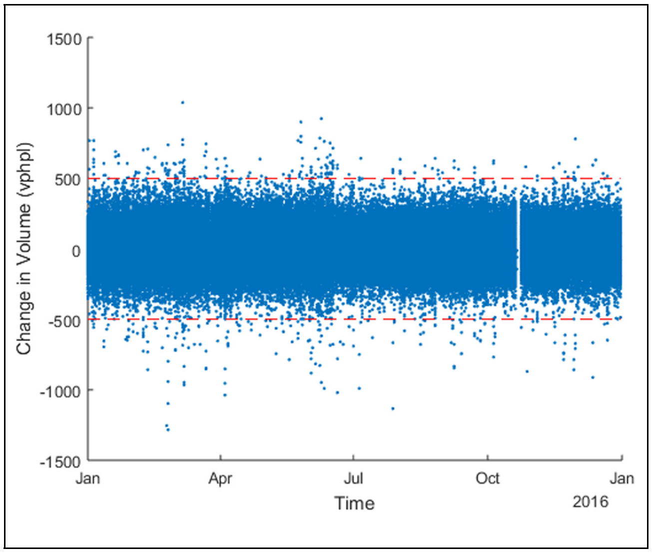

Figure 4. Plot of changes in volume.

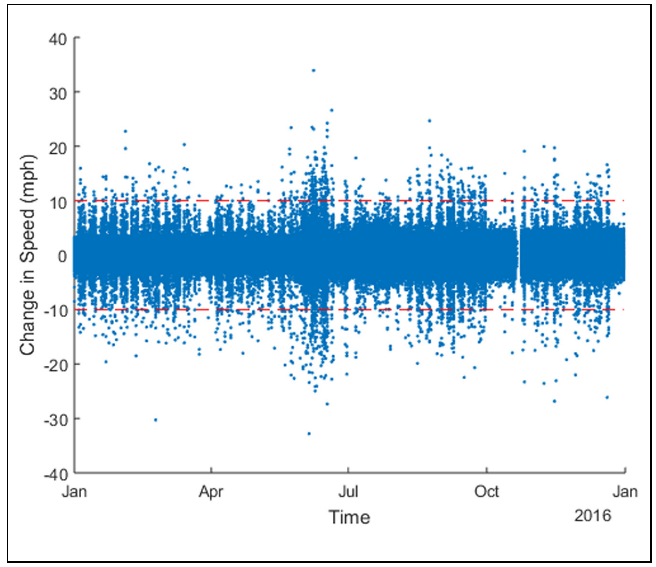

Figure 6. Plot of changes in speed.

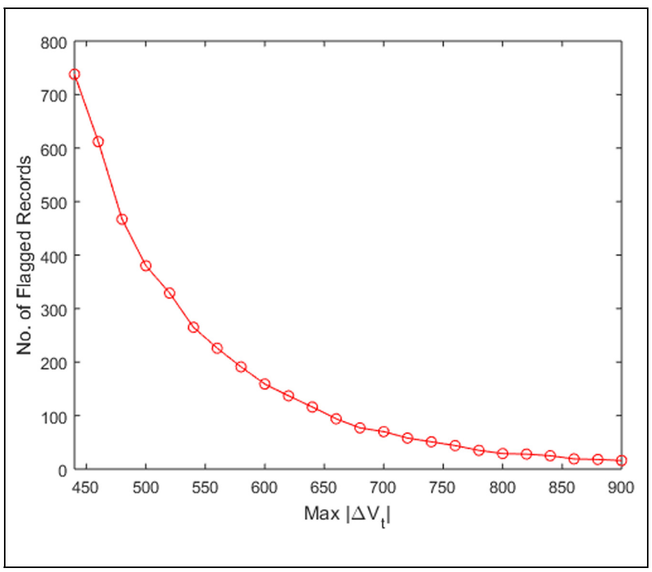

Figure 5. Sensitivity analysis on the maximum value of $|\Delta V_t|$.

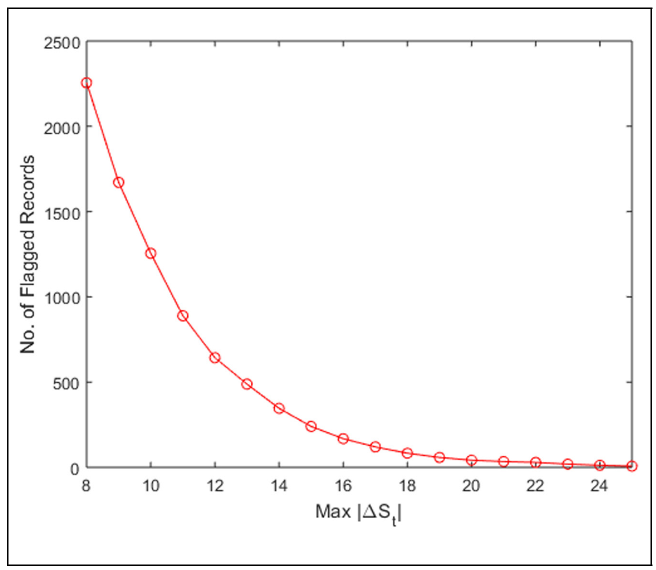

Figure 7. Sensitivity analysis on the maximum value of $|\Delta S_t|$.

It compares the speed of one traffic record and the average speed of its adjacent records in time. The equation for this test is as follows:

$$\Delta S_t = S_t - \frac{S_{t-1} + S_{t+1}}{2} \tag{7}$$

where ΔS_t represents the change in speed at time t and S_{t-1}, S_t, and S_{t+1} are the speeds in time periods, $t-1, t$ and $t+1$. Note that all of the speed values are from the same lane. The equation does not apply when any of S_{t-1}, S_t, and S_{t+1} is zero since the speed is not available when it is recorded as zero given no passing vehicles.

Figure 6 presents the time-series plot of ΔS_t. It shows that ΔS_t has no obvious temporal variation and the majority of ΔS_t s are within the interval of [–10, 10]. Similar to ΔV_t s, ΔS_t s are roughly symmetric, and it would be more convenient to set a threshold for the absolute value of ΔS_t, that is, $|\Delta S_t|$. Based on the sensitivity analysis of flagged record counts with different threshold values of $|\Delta S_t|$ as shown in Figure 7, the number of failed records starts to decrease marginally when max $|\Delta S_t|$ reaches 15 mph; thus, a favorable value of max $|\Delta S_t|$ is 15 mph. The criterion of this test is: if $|\Delta S_t| > 15$ mph, then invalid.

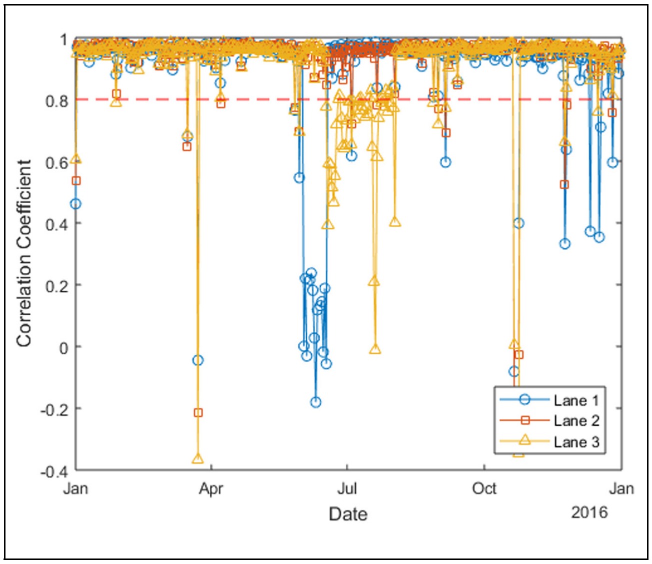

Figure 8. Plot of correlation coefficient by day.

Anomalous Daily Traffic Pattern. This test checks whether the traffic pattern of volume in one weekday/weekend is similar to that of similar days (i.e., other weekdays/weekends in the same month). Turner (*1*) proposed a similar temporal consistency check based on a daily speed profile of multiple days to identify plausible speed data. Instead of needing a manual check on the plots as claimed by Turner, this check can be programmed by measuring the correlation coefficient and establishing a threshold on the measure. The equation for correlation coefficient used to calculate the similarity is as follows:

$$r_i = \frac{\sum_{t=1}^{T}(V_{i,t} - \bar{V}_i)(V_t - \bar{V})}{\sqrt{\sum_{t=1}^{T}(V_{i,t} - \bar{V}_i)^2}\sqrt{\sum_{t=1}^{T}(V_t - \bar{V})^2}} \quad (8)$$

where r_i is the correlation coefficient of volume profile of Day i and average volume profile of all the same weekdays/weekends as Day i in the same month (for example, if Day i is a weekday in January 2016, then all weekdays in January 2016 would be used; if Day i is a weekend in January 2016, then all weekends in January 2016 would be used). $V_{i,t}$ is the volume at time t of Day i, and $\bar{V}_i$ is the average volume of Day i computed as $\bar{V}_i = \frac{\sum_{t=1}^{T} V_{i,t}}{T}$; V_t is the average volume at time t of all the weekdays/weekends same as Day i in the same month computed as $V_t = \frac{\sum_{j=1}^{J} V_{j,t}}{J}$, and $\bar{V}$ is the average of V_t s computed as $\bar{V} = \frac{\sum_{t=1}^{T} V_t}{T}$. The purpose of using the average volume profile is to smooth the noise and account for missing and invalid individual daily profiles.

Figure 8 plots the correlation coefficient by day and by lane (Lane 1, 2, and 3 are the median lane, middle lane, and shoulder lane, respectively), and shows that a threshold of 0.8 performs well in distinguishing valid daily profiles from those that are invalid. Therefore, the criterion of this test is: if the correlation coefficient calculated by Equation 7 is below 0.8, the traffic records of the corresponding day are flagged. Note that holidays should not be flagged, as they may have different patterns.

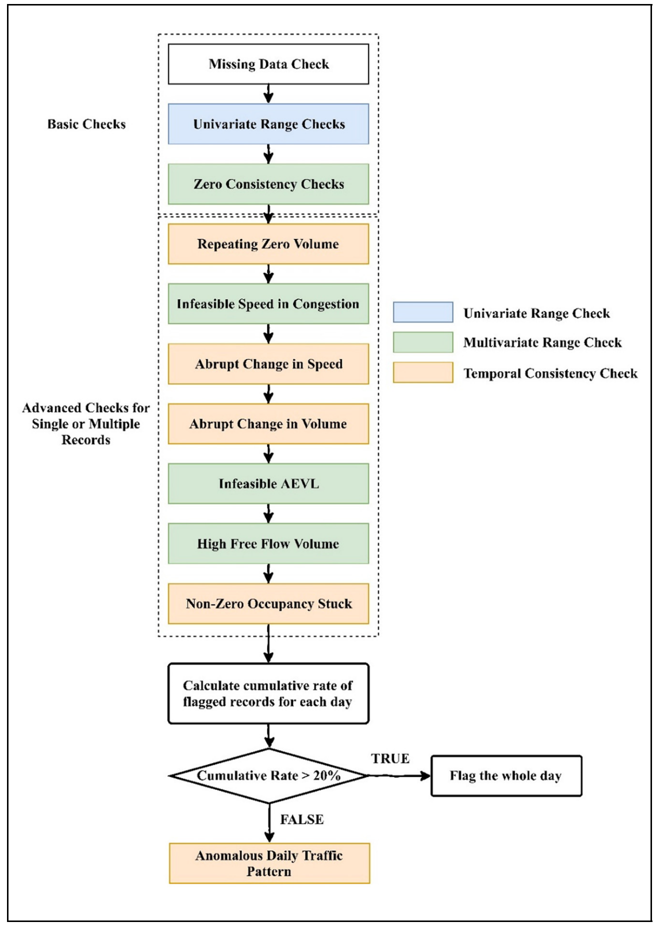

Figure 9. Flowchart of flagging procedure.

Flagging Procedure

The validity tests must be prioritized by their reliability and efficiency in error detection and error criticality since the proposed tests are more complex than what can actually be implemented. Multiple data issues may be identified in a traffic record given the varying and overlapping natures and theories of the validity tests; hence, the record will bear multiple flags, each representing a different problem. The intention is to avoid questionable (flagged) records rather than debugging the errors, meaning that whether or not a record is flagged is as meaningful as the number of flags, the type of flag, or both.

Figure 9 proposes the flowchart of a flagging procedure. This procedure works in sequence with no repeats,

Table 3. Performance of Validity Tests

Validity test	Flagged records	Cumulative flagged records
Basic		
Missing data check	41.4% (6,708)[a]	41.4% (6,708)
Univariate range checks	0% (0)	41.4% (6,708)
Zero consistency checks	0.1% (22)	41.5% (6,730)
Advanced		
Repeating zero volume	47.8% (7,745)	89.3% (14,475)
Infeasible AEVL	6.9% (1,116)	96.2% (15,591)
Infeasible speed in congestion	2.0% (321)	98.2% (15,912)
Abrupt change in speed	1.4% (238)	99.6% (16,150)
Abrupt change in volume	0.4% (54)	100% (16,204)
High free-flow volume	0% (0)	100% (16,204)
Non-zero occupancy stuck	0% (0)	100% (16,204)
	Flagged days	
Days with over 20% cumulative rates of flagged records	30 days	
Anomalous daily traffic pattern	13 days	
Total days	43 days	

[a]Percentage of flagged records with the count of flagged records in parentheses.

meaning a record that is flagged previously will not be flagged again.

For example, if one record is flagged by univariate range checks, it would not be checked by the other checks, regardless the types of data issues. The processing time can be saved in this way, as data that is already flagged does not need to be checked again. Traffic records run through basic checks first, which include missing data checks, univariate range checks, and zero consistency checks. Next, the records that are not flagged are examined by advanced checks which may involve a single record or consecutive records. Advanced checks are sorted by their efficiency in detecting questionable records. After records are flagged by basic checks, advanced checks, or both, the cumulative percentage of flagged records is calculated for the day. A cumulative rate of over 20% leads the corresponding day to be flagged, meaning it is considered not reliable for traffic analysis.

Table 3 presents the results of the flagging procedure. It shows that basic checks can flag only 41.5% of invalid data, which indicates that basic checks alone cannot provide sufficient data quality. Basic checks and the first two advanced checks (i.e., repeating zero volume, and infeasible AEVL) can flag more than 90% of invalid records, indicating that these validity tests are necessary for guaranteeing satisfactory performance of the flagging procedure. Anomalous daily traffic pattern can detect a considerable amount of questionable daily traffic patterns and is thus a necessary component of the procedure. Local agencies can pre-determine the desirable flagging percentage of invalid records and then decide which validity tests to implement. For example, in this case study, basic checks and the first four advanced checks are necessary for flagging more than 99% of invalid records (single or multiple).

Conclusions

This study has provided guidance on how to select validity tests and develop a flagging procedure for potential implementation in ADMS. Basic validity tests should be identified first to maintain minimum data quality. A user survey is desirable for incorporating user preferences in more complex validity tests found in the literature. Candidate validity tests should be identified based on both the user's preferences and the programming complexity. Rules of candidate validity tests can be established using the rule-based method and the data-driven method. Candidate validity tests should be evaluated regarding their effectiveness in flagging questionable data. Lastly, if all tests will not be implemented, a subset of candidate validity tests which collectively provide satisfactory data quality can be determined for implementation.

The proposed guide was demonstrated in a case study using traffic data from the Wisconsin ADMS, the V-SPOC system. Basic validity tests including missing data check, univariate range checks, and zero consistency checks were determined based on common validity tests which have been implemented in various ADMS. Basic checks may not be adequate for detecting less obvious yet important issues, and therefore more complex validity tests are needed. A literature search was run to collect potential validity tests for implementation. A user survey was conducted to collect users' preferences. Three alternative multivariate range checks and five temporal consistency checks which were favored by users and are easy to program were kept for evaluation. A flagging procedure consists of both basic checks and eight more complex checks. The flagging procedure works in a sequential manner to help prioritize validity tests for implementation so that sufficient data quality is provided and efficiency is improved.

Local agencies can apply the guide proposed in this study to customize the flagging procedure for their local ADMS. Methodologies proposed in this study for establishing rule-based or data-driven criteria for validity tests can be applied directly or can be adjusted using local data. Traffic data from one station was used in this study, testing more sites may result in different threshold values for individual validity tests, so data from more sites can be pooled to form rules that are more universally applicable or that could vary by site to provide superior data quality.

Acknowledgments

The authors thank the Wisconsin Department of Transportation for providing grant funding and guidance for this study.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: ZC, XQ, ES, YC, SP, MRRS; data collection: ZC; analysis and interpretation of results: ZC, XQ, ES, YC, SP; draft manuscript preparation: ZC, XQ. All authors reviewed the results and approved the final version of the manuscript.

References

1. Turner, S. *Quality Control Procedures for Archived Operations Traffic Data: Synthesis of Practice and Recommendations*. FHWA, U.S. Department of Transportation, 2007. http://www.fhwa.dot.gov/policy/ohpi/reports/qcproc/index.htm.
2. Lee, H., and B. Coifman. Identifying and Correcting Pulse-Breakup Errors from Freeway Loop Detectors. *Transportation Research Record: Journal of the Transportation Research Board*, 2011. 2256: 68–78.
3. Coifman, B. Using Dual Loop Speed Traps to Identify Detector Errors. *Transportation Research Record: Journal of the Transportation Research Board*, 1999. 1683: 47–58.
4. Lee, H., and B. Coifman. Quantifying Loop Detector Sensitivity and Correcting Detection Problems on Freeways. *Journal of Transportation Engineering*, Vol. 138, No. 7, 2011, pp. 871–881.
5. Chen, L., and A. D. May. Traffic Detector Errors and Diagnostics. *Transportation Research Record: Journal of the Transportation Research Board*, 1987. 1132: 82–93.
6. Coifman, B., and S. Dhoorjaty. Event Data-Based Traffic Detector Validation Tests. *Journal of Transportation Engineering*, Vol. 130, No. 3, 2004, pp. 313–321.
7. Peeta, S., and I. Anastassopoulos. Automatic Real-Time Detection and Correction of Erroneous Detector Data with Fourier Transforms for Online Traffic Control Architectures. *Transportation Research Record: Journal of the Transportation Research Board*, 2002. 1811: 1–11.
8. Turochy, R., and B. Smith. New Procedure for Detector Data Screening in Traffic Management Systems. *Transportation Research Record: Journal of the Transportation Research Board*, 2000. 1727: 127–131.
9. Lu, Y., X. Yang, and G. L. Chang. Algorithm for Detector-Error Screening on Basis of Temporal and Spatial Information. *Transportation Research Record: Journal of the Transportation Research Board*, 2014. 2443: 40–48.
10. Nihan, N. L. *Detector Data Validity*. Publication WA-RD 208.1. Washington State Department of Transportation, 1990.
11. Jacobson, L. N., N. L. Nihan, and J. D. Bender. Detecting Erroneous Loop Detector Data in a Freeway Traffic Management System. *Transportation Research Record: Journal of the Transportation Research Board*, 1990. 1287: 151–166.
12. Cleghorn, D., F. L. Hall, and D. Garbuio. Improved Data Screening Techniques for Freeway Traffic Management Systems. *Transportation Research Record: Journal of the Transportation Research Board*, 1991. 1320: 17–23.
13. Shi, X., S. T. Parker, Y. Cheng, J. Fang, B. Ran, and D. A. Noyce. A Comprehensive Traffic Data Quality Examination Tool-WisTransPortal V-SPOC QAQC Enhancement. Presented at 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.
14. Hall, F. L., and B. N. Persaud. Evaluation of Speed Estimates Made with Single-Detector Data from Freeway Traffic Management Systems. *Transportation Research Record: Journal of the Transportation Research Board*, 1989. 1232: 9–16.
15. Hall, F. L., and M. A. Gunter. Further Analysis of the Flow-Concentration Relationship. *Transportation Research Record: Journal of the Transportation Research Board*, 1986. 1091:1–9.
16. Hu, P. *Proof of Concept of ITS as an Alternative Data Resource: A Demonstration Project of Florida and New York Data*. ORNL Oak Ridge National Laboratory, Tennessee, 2001.
17. Schmoyer, R., P. Hu, and R. Goeltz. Statistical Data Filtering and Aggregation to Hour Totals of Intelligent Transportation System 30-s and 5-min Vehicle Counts. *Transportation Research Record: Journal of the Transportation Research Board*, 2001. 1769: 79–86.
18. Lin, D. Y., S. Boyles, V. Valsaraj, and S. T. Waller. Reliability Assessment for Traffic Data. *Journal of the Chinese Institute of Engineers*, Vol. 35, No. 3, 2012, pp. 285–297.
19. Kwon, J., C. Chen, and P. Varaiya. Statistical Methods for Detecting Spatial Configuration Errors in Traffic Surveillance Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, 2004. 1870: 124–132.
20. Nihan, N. L. Aid to Determining Freeway Metering Rates and Detecting Loop Errors. *Journal of Transportation Engineering*, Vol. 123, No. 6, 1997, pp. 454–458.
21. Vanajakshi, L., and L. Rilett. Loop Detector Data Diagnostics Based on Conservation-of-Vehicles Principle. *Transportation Research Record: Journal of the Transportation Research Board*, 2004. 1870: 162–169.
22. Wall, Z., and D. Dailey. Algorithm for Detecting and Correcting Errors in Archived Traffic Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2003. 1855: 183–190.
23. Qin, X., M. Ng, and P. E. Reyes. Identifying Crash-Prone Locations with Quantile Regression. *Accident Analysis & Prevention*, Vol. 42, No. 6, 2010, pp. 1531–1537.
24. Rajagopal, R., and P. Varaiya. *Health of California's Loop Detector System*. California PATH Program, Institute of Transportation Studies, University of California at Berkeley, 2007.

The Standing Committee on Highway Traffic Monitoring (ABJ35) peer-reviewed this paper (19-03528).