Volume Data Correction for Single-Channel Advance Loop Detectors at Signalized Intersections

Yao-Jan Wu, Guohui Zhang, and Yinhai Wang

The single-channel detection scheme has been widely used in practice for traffic control. Differing from lane-by-lane detection, single-channel detection has all loop detectors across multiple lanes wired together and provides a single input to the controller. Although single-channel (wired-together) detectors are commonly used by traffic agencies, the traffic volume data collected from these single-channel loops are inaccurate because of high misdetection rates, compared with lane-by-lane detection. This study focuses on developing a probability-based approach for correcting the traffic volume data collected by the single-channel loop detectors at signalized intersections. The proposed probability-based nonlinear model (NM) explicitly describes a potential model compared with a multiple linear regression model. Both models were calibrated and validated using real-life data from seven two-lane intersection approaches and three three-lane intersection approaches under various traffic conditions. The results showed that the proposed NM requires much less effort to calibrate and can effectively reduce system errors by reducing absolute mean errors by 60% on average. The correction effect of the model positively scales with increases in the system error rate. The verification results indicate the proposed NMs are capable of correcting the misdetection resulting from single-channel loop detectors and accurately estimate hourly traffic volume. The NM demonstrated its transferability to different locations and can be a general application in practice.

Inductance loop detectors have become the most widely deployed type of sensor in traffic management systems since 1960 (*1, 2*). Nowadays, most cities in the United States primarily rely on loop detectors to support modern signal control systems, including actuated signal control and adaptive signal control (*3*). At these signalized intersections, two detection schemes are commonly adopted: single-channel detection and lane-by-lane detection. Single-channel detection means the loops in different lanes are physically wired together, whereas lane-by-lane detection can separate calls from the loops in different lanes. The single-channel (wired-together) detector has been commonly used in practice because of the lower costs of single-channel loop detector installation. The installation procedure is easier because the single conductor cable can be installed with fewer saw-cuts and sealant in the pavement. With fewer lead-in wires, the number of amplifier cards is reduced. When single-channel detectors are employed, multiple vehicles arriving concurrently in different lanes will be counted as one vehicle. As a result, the traffic volume data collected from the single-channel detector are inaccurate. In the past, traffic data could relatively easily be collected in real time by manual means. Real-time high-resolution volume data automatically collected by loop detectors were not necessary. With the development of the intelligent transportation system (ITS), collecting real-time data has become more and more important. Many arterial-related research projects (*4–6*) have relied heavily on data directly collected from the existing loop sensors because this is one of the most cost-effective ways to retrieve traffic data using existing data collection facilities without installing new sensors. Research is currently limited by the constraints imposed by the implementation of single-channel loops, leaving agencies to directly use the single-channel data without correction as a compromise.

Inductance loop-related issues have been investigated by many researchers. Most previous research focused on freeway loop detectors (*7–11*). For arterial loops, the impact of different detection schemes on traffic signal control was investigated (*12–15*). However, little research has been done for single-channel loop data correction. It may be because of the complexity of data correction or ground truth data collection that was not feasible in the past. To address these issues, this study aims to evaluate the data collection errors resulting from the single-channel detection scheme and develop mathematical models to correct the volume data collected by the single-channel detector. With such models single-channel loop outputs can be correctly tuned without altering the existing infrastructure, and data collection costs can be reduced.

This paper is organized as follows. In the next section, a problem investigation will provide a more detailed overview of the problem. Then, the paper will focus on the development of the probability-based model to further elaborate the issue mathematically. Two probability-based nonlinear models (NMs) will be constructed for both the two- and three-lane cases. Correspondingly, multiple linear regression models will be built for comparison purposes. Both nonlinear and linear models (LMs) will be calibrated based on real-life data, followed by model verification. A summary of the study and suggestions will be made at the end.

PROBLEM STATEMENT

As mentioned in the first section, the single-channel detector has multiple lanes wired together and cannot identify calls from each individual lane that is occupied by vehicles. Figure 1 illustrates three typical cases in which multiple vehicles crossing over the single-channel detector will be counted as one vehicle. Case *a* shows the

Y.-J. Wu and Y. Wang, Box 352700, Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195-2700. G. Zhang, Center for Transportation Research, Department of Civil, Architectural, and Environmental Engineering, University of Texas at Austin, 3208 Red River, Austin, TX 78705. Corresponding author: Y.-J. Wu, yaojan@u.washington.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2160, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 128–139.
DOI: 10.3141/2160-14

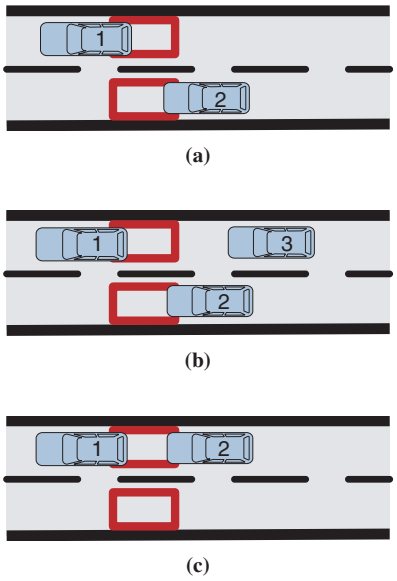


FIGURE 1 Data collection errors caused by single-channel loop: (*a*) Case *a* vehicles arriving consecutively and closely in different lanes; (*b*) Case *b* vehicles arriving closely in different lanes; and (*c*) Case *c* vehicles arriving closely in the same lane.

most common error in which two vehicles in different lanes are counted as one vehicle. Case *b* demonstrates how errors more easily occur as traffic becomes congested. Multiple vehicles in different lanes arrive consecutively with short headways. Then, multiple vehicles are counted as a vehicle. Case *c* happens when two vehicles move closely in the same lane and are counted as one vehicle. Longer loop detectors and highly congested conditions may increase this type of error. Under such a detection mechanism, vehicle volume observed by such detectors is distorted. This problem could become more serious with increasing numbers of lanes at one approach.

METHODOLOGY

Probability-Based Misdetection Model

To fully understand the data collection problems resulting from single-channel loop outputs, building a probability-based model can provide in-depth insight into the relationship between unsuccessful detection and traffic headway characteristics (*16*). As mentioned in the Problem Statement, the errors in Case *c* are relatively rare among three types of errors. Therefore, the errors in Cases *a* and *b* are the most common and were focused on in the modeling process. In this regard, the single-channel loop detector can count vehicles correctly when only one lane is occupied while the others are not. In other words, if more than one lane is occupied by vehicles, these will be mistakenly counted as a single vehicle. This case is regarded as "misdetection" in this paper. While this paper aims at modeling and correcting misdetection data for advance loop detectors [usually 100 ft (30.48 m) away from the stop bar] at signalized intersections, the methodology could be extended to address similar problems for various loops located at different locations at an intersection approach.

Let X be the number of vehicles crossing over a loop during a given time interval, T, and A_i represent the event that the advance loop detector on the ith lane is occupied by vehicles. $P(A_i)$ is the probability of at least one vehicle crossing over the detector in the ith lane. $P(A_i')$ is the probability of no vehicle crossing over the detector in the ith lane. It is assumed that the number of vehicle arrivals λ during time interval T follows a Poisson distribution. In general, λ may change over time and such a process is regarded as a nonhomogeneous Poisson process (*17, 18*). Thus, $P(A_i)$ and $P(A_i')$ can be calculated as

$$P(A_i') = P(X=0) = \frac{e^{-\lambda_i T}(\lambda_i T)^0}{0!} = e^{-\lambda_i T} \tag{1}$$

$$P(A_i) = P(X>0) = 1 - P(A_i = 0) = 1 - e^{-\lambda T} \tag{2}$$

and

$$T = \frac{L_{veh} + L_{loop}}{S_{veh}} \tag{3}$$

where

λ_i = average arrival rate for lane i,
$S_{veh,i}$ = vehicle speed, and
$L_{veh} + L_{loop}$ = average effective vehicle length of the traffic stream that passed the detector (*6*).

Since most vehicles on urban streets are passenger cars, it is assumed L_{veh} is 14 ft (4.27 m) and the effective electrical loop length L_{loop} is 6 ft (1.82 m). To simplify the model, S_{veh} and λ_i are assumed to be identical for both lanes.

For the two-lane scenario, misdetection happens when vehicles on each lane are crossing over the single-channel detector concurrently. Assuming the arrival on each lane is equal and mutually independent, the probability of misdetection can be defined as

$$P_{mis,2} = P(A_1 \cap A_2) = P(A_1)P(A_2) \tag{4}$$

After substituting Equation 2 into Equation 4, the probability of misdetection for the two-lane scenario can be expressed as

$$P_{mis,2} = (1 - e^{-\lambda T})^2 \tag{5}$$

For multilane scenarios, Equation 4 can be further extended by using set theory (*19*). Therefore, the probability of misdetection when "two or more than two lanes are occupied by a least one vehicle" in the three-lane scenario can be calculated using set operations and expressed as

$$\begin{aligned} P_{mis,3} &= P(A_1A_2) + P(A_2A_3) + P(A_1A_3) - 2P(A_1A_2A_3) \\ &= 3(1 - e^{-\lambda T})^2 - 2(1 - e^{-\lambda T})^3 \end{aligned} \tag{6}$$

Similarly, the probability of misdetection for the four-lane scenario is

$$\begin{aligned} P_{mis,4} &= P(A_1A_2) + P(A_2A_3) + P(A_3A_4) + P(A_1A_4) - P(A_1A_2A_3) \\ &\quad - P(A_1A_2A_4) - P(A_2A_3A_4) - P(A_1A_3A_4) + 3(A_1A_2A_3A_4) \\ &= 4(1 - e^{-\lambda T})^2 - 4(1 - e^{-\lambda T})^3 + 3(1 - e^{-\lambda T})^4 \end{aligned} \tag{7}$$

When more lanes are employed in one approach, the corresponding formula can be computed accordingly based on the mathematical induction fundamentals. Because more than four lanes are rarely

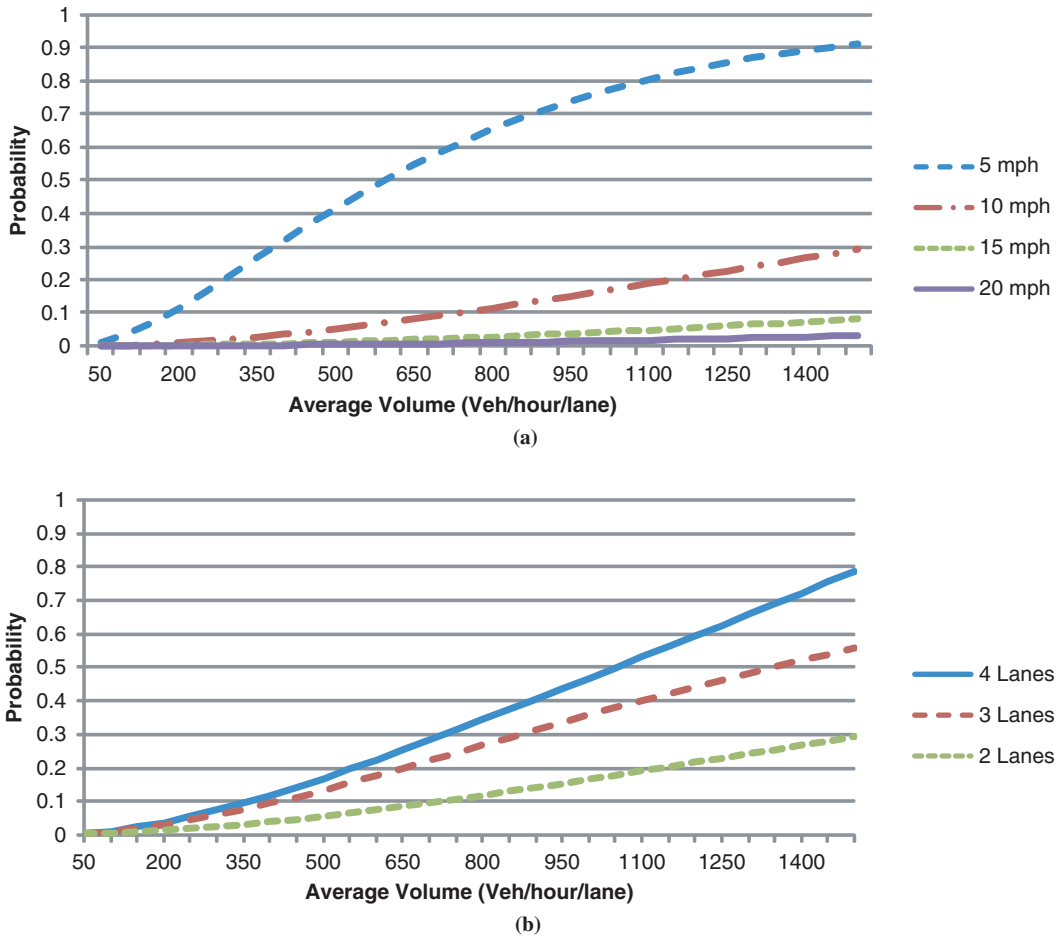


FIGURE 2 The relationship between arrival rate (average volume) and probability of misdetection (*a*) by different speeds (two lanes and effective vehicle length = 20 ft) and (*b*) by different number of lanes (speed = 10 mph).

observed in practice, their calculation will not be emphasized in this paper.

As shown in Equations 5 to 7, misdetection is determined by crossing time, closely associated with the vehicle speed, arrival rate, and the number of lanes on which the loop detectors are wired together. Figure 2*a* shows the relationship between different arrival rates (average volume per lane) and the probability of misdetection by different vehicle speeds in the two-lane scenario. As vehicle speed reduces, the probability of misdetection will rise significantly. This situation is common for other lane configurations. Figure 2*b* shows the relationships between different arrival rates (average volume per lane) and the probability of misdetection by different numbers of lanes assuming a constant vehicle speed of 10 mph (14.67 km/h). The more lanes are wired together, the higher the probability that a misdetection will occur. Note that the probability model does not fully consider the effect of error from Case c. In the event of high congestion, misdetections will happen more frequently than is shown.

Volume Data Correction Formulation

In this section, the volume correction models for the two- and three-lane scenarios are formulated, respectively. Two types of models are built for each scenario: the NM based on the probability-based misdetection model introduced earlier and traditional multiple linear regression models.

Probability-Based NM

Two-Lane Scenario

As mentioned, misdetection means the multiple vehicles in different lanes are counted as one vehicle. Theoretically, the volume data collected from a single-channel detector (hereafter, the term "single-channel volume" is used) should be lower than the ground truth volume entering the same detector because of the single-channel effect. If the portion of "misdetected" volume is deducted from the total ground truth volume, the result must be equal to the volume collected by the single-channel detector. The relationship can be formulated as

$$2q - qP_{\text{mis},2} = Q \tag{8}$$

where q = the ground truth volume per lane and Q = total volume collected from the single-channel detector. After substituting Equation 5 into Equation 8, the ground truth volume per lane can be calculated from

$$q = \frac{Q}{2 - P_{\text{mis},2}} = \frac{Q}{2 - \left(1 - e^{-\lambda T}\right)^2} \tag{9}$$

In Equation 9, T can be determined by the spot vehicle speed and the effective vehicle length as shown in Equation 3. As proposed by

Zhang (*6*), the spot speed for the vehicle can be derived from the relationship between volume and concentration.

$$S_{veh} = \frac{q}{d} \tag{10}$$

where q can be directly measured from the loop detector and concentration d (veh/mi) can be calculated from detector occupancy measurement Occ(%) and the average effective vehicle length ($L_{veh} + L_{loop}$) as shown in Equation 11

$$d = \frac{k \cdot \text{Occ}}{\left(L_{loop} + L_{veh}\right)} \tag{11}$$

where $k = 52.8$ is the unit conversion factor (from feet to miles). After substituting Equation 11 into Equation 10, the spot speed is formulated as

$$S_{veh} = \frac{q\left(L_{loop} + L_{veh}\right)}{k \cdot \text{Occ}} \tag{12}$$

Then, the travel time for each vehicle crossing a loop can be recalculated as

$$T = \frac{L_{veh} + L_{loop}}{S_{veh,i}} = \frac{k \cdot \text{Occ}}{q} \tag{13}$$

However, the single-channel (advance loop) detector cannot directly measure accurate Occ and q. Instead, the detector measures biased occupancy, Occ_s and total volume Q. Moreover, λ is the vehicle arrival rate entering the single-channel detector and may follow a mathematical relationship with Q [e.g., $Q = f(\lambda)$]. Therefore, the term k is replaced with a model parameter α. After calibration, this parameter should be able to capture the relationship between $\frac{\text{Occ}}{q}$ and $\frac{\text{Occ}_s}{Q}$ and the relationship between Q and λ, as well as other unexplainable factors. Therefore, the model shown in Equation 9 is redefined as

$$q = \frac{Q}{2 - \left(1 - e^{\alpha\frac{\text{Occ}_s}{Q}}\right)^2} \tag{14}$$

Three-Lane Scenario

To build the relationship between single-channel volume, Q, and lane-by-lane volume, q, the probability of misdetection in Equation 6 should be decomposed into two parts because the volume correction should be treated differently when different number lanes are occupied. Based on set theory, the probability of misdetection can be further separated into $P'_{mis,3}$ (exactly three lanes are occupied at the same time) and $P'_{mis,2}$ (exactly two lanes are occupied at the same time) as shown in Equation 15:

$$\begin{aligned} P_{mis,3} = P'_{mis,3} + P'_{mis,2} &= \left[P\left(A_1A_2A_3\right)\right] \\ &+ \left[P\left(A_1A_2\right) + P\left(A_2A_3\right) + P\left(A_1A_3\right) - 3P\left(A_1A_2A_3\right)\right] \end{aligned} \tag{15}$$

The single-channel volume, Q, should be equal to the total ground truth volume minus the volume collected when vehicles in three lanes are misdetected as one vehicle, $2qP'_{mis,3}$, and when vehicles in two lanes are misdetected as one vehicle, $qP'_{mis,2}$. This relationship can be formulated in Equation 16.

$$3q - 2qP'_{mis,3} - qP'_{mis,2} = Q \tag{16}$$

After Equations 2, 15, and 16 are combined, the volume correction model can be formulated as Equation 17.

$$q = \frac{Q}{3 - 2P'_{mis,3} - P'_{mis,2}} = \frac{Q}{3 + \left(1 - e^{-\lambda T}\right)^3 - 3\left(1 - e^{-\lambda T}\right)^2} \tag{17}$$

Similar to the derivation for the two-lane scenario in Equations 10 through 13, $-\lambda T$ can be replaced with Occ_s/Q multiplied by a parameter. In the three-lane scenario, two parameters, α and β, are used to capture more variations as shown in Equation 18.

$$q = \frac{Q}{3 + \left(1 - e^{\alpha\frac{\text{Occ}_s}{Q}}\right)^3 - 3\left(1 - e^{\beta\frac{\text{Occ}_s}{Q}}\right)^2} \tag{18}$$

DATA COLLECTION

To calibrate the proposed probability-based NM and multiple linear regression models, ground truth volumes are indispensable. However, it is difficult to collect ground truth volume using the same loop detectors since they are already wired together. To collect enough samples in this study, manual means may not be efficient and sample size would be small. Hence, collecting data from an upstream detector could be a possible solution assuming the upstream is not distant from the advance detector. Two types of detector data were collected from the City of Bellevue in Washington State: the advance detector and the system detector. The advance detector is located approximately 120 ft (36.6 m) upstream from the stop bar. As of 2009, there were more than 500 advance detectors, a majority of which are wired together and serve as primary data input to the traffic management center (TMC). The system detector is a lane-by-lane–based detector, located 250 ft (76.2m) to 300 ft (91.4m) upstream from the stop bar. Only a limited number of communication channels can send system detector data back to TMC. Thus, seven two-lane study sites (intersection approaches) and three three-lane study sites were selected based on various roadway characteristics and availability. The description of the study intersection approaches are listed in Table 1.

DATA CHARACTERISTICS

The real-time volume data were aggregated for each cycle. If the congestion is severe or cycle failure happens, the volume difference between advance loop data and system loop data will increase. Nevertheless, the effect of the volume difference is expected to be handled by the modeling process. Figure 3 shows two examples of the relationship between volume data collected from the advance detector and the system detector, representing single-channel data and (lane-by-lane) ground truth data, respectively. Because of the single-channel effect, most volume data were underestimated. The high volume samples circled by a dashed circle in Figure 3*b* are also with higher occupancy values (mostly greater than 50%). In other words, when traffic is congested, volume collected by the (single-channel) advance loop will likely be underestimated.

TABLE 1 Study Site Description

Intersection	Location	Direction	Attribute (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
041	NE 8th and 140th Ave. NE	EB	475	2,342	300	1	1	35	4	2
046	NE 8th and 143rd Ave. NE	WB	399	1,569	350	1	0	35	3	2
050	Main St. and 148th Ave. SE	NB	660	1,421	300	1	1	35	4	2
050	Main St. and 148th Ave. SE	SB	695	2,613	300	1	1	35	4	2
065	SE 8th and 148th Ave. SE	NB	590	1,496	300	1	0	35	3	2
065	SE 8th and 148th Ave. SE	SB	573	1,225	300	1	0	35	3	2
131	116th Ave. SE and SE 1st St.	SB	440	676	300	1	0	30	3	2
027	NE 8th and 110th Ave. NE	WB	711	716	300	1	0	30	4	3
047	Northup Way and 148th Ave. SE	SB	608	357	300	1	0	35	4	3
056	Landerholm Circle and 148th Ave. SE	SB	1,050	908	300	0	0	30	3	3

NOTE: (1) average daily traffic (7/7/2009 to 7/9/2009); (2) link length (feet); (3) system detector location (feet); (4) left-turn lane without a detector (yes: 1, no: 0); (5) additional right-turn lane without a detector (yes: 1, no: 0); (6) speed limit (mph); (7) total number of lanes at the approach; and (8) total number of lanes wired together.

MODEL CALIBRATION AND VERIFICATION

Model Calibration

Probability-Based NM

The nonlinear least squares technique (*20*) was applied to calibrate the proposed probability-based model. Details of the modeling techniques can be found in Ritz and Streibig (*20*). The model results are listed in Table 2 and the data set description is listed below:

NM1. The two-lane volume correction model (Equation 14) was calibrated using the data set collected at all of the seven test sites for 24 h on Wednesday, July 1, 2009.

NM2. The two-lane volume correction model (Equation 14) was calibrated using the data set collected at Intersections 41 eastbound (EB) and 46 westbound (WB) from 7:00 to 9:00 a.m. on Wednesday, July 1, 2009.

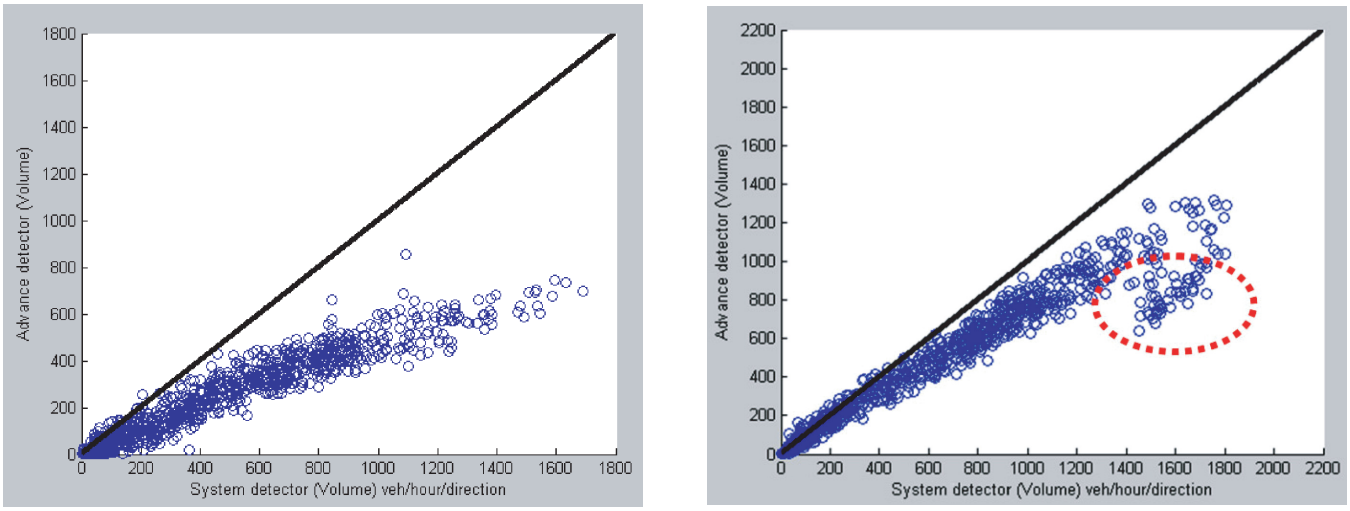
NM3. The three-lane volume correction model (Equation 18) was calibrated using the data set collected at all of the three-lane test sites for 24 h on Wednesday, July 1, 2009.

NM4. The three-lane volume correction model (Equation 18) was calibrated using the data set collected at Intersection 27 (WB) from 7:00 to 9:00 a.m. on Wednesday, July 1, 2009.

As shown in Table 2, all the calibrated parameters in these models were significant at the $p = .05$ level except NM 4. This could be because the sample size was relatively small or samples contained more random errors. Still, the model will be evaluated in the Model Verification section.

Multiple Linear Regression Model (LM)

Linear regression is one of the most common econometric techniques and suitable for modeling a wide variety of relationships between



(a) (b)

FIGURE 3 Comparisons between the volume collected from the system loop detector and advance detector (wired together) at (*a*) Intersection 41 (EB) (July 9, 2009) and (*b*) Intersection 65 (SB) (July 9, 2009).

TABLE 2 Nonlinear Model Parameters for Volume Correction

Parameter	Calibrated[a]	SE	*t*-Value	*p*-Value
NM1 (for two-lane scenario)				
Alpha (α)	−54.656	3.224	−16.95	.000
Sample size: 6,095				
NM2 (for two-lane scenario)				
Alpha (α)	−42.5845	0.2159	−197.20	.000
Sample size: 100				
NM3 (for three-lane scenario)				
Alpha (α)	−33.984	4.445	−7.65	.000
Beta (β)	−25.46	1.215	−20.96	.000
Sample size: 2,114				
NM4 (for three-lane scenario)				
Alpha (α)	−27.91	44.96	−0.62	.537
Beta (β)	−22.7	14.42	−1.57	.121
Sample size: 60				

[a]Calibrated-calibrated parameter.

variables (*21*). The objective of multiple linear regression is to model the relationships between the ground truth volume q and other independent variables. The model takes a very general form:

$$q = f(X_1, X_2, \ldots, X_k) + \epsilon \tag{19}$$

where $X_1, X_2, \ldots, X_k$ are the independent variables. The variables being used include single-channel volume and occupancy, cycle length, and the variables in Columns 2 through 8 of Table 1. In the modeling process, different forms of $f(X_1, X_2, \ldots, X_n)$ were investigated. The LMs being built considered all the first-order effects and variable interactions. Details of multiple regression techniques can be found in Washington et al. (*21*) and Faraway (*22*). Importantly, the model was forced to pass through (0.0). Compared with the same model with *y*-intercept, the study found that the models without *y*-intercept not only can provide better goodness of fit but also provide more accurate estimation. It is reasonable because the relationship between two types of volumes starts from origin (0.0).

Four LMs—LM1 through LM4—were built for this study. Each LM was built using the same data set used by the counterpart of the NM. For example, LM1 and NM1 used the same data set, so they can be comparable. Table 3 shows the final results for all the LMs. LM1 has the highest number of variables (seven) and all of them are significant at the $p = .05$ level. Note that the interaction effect between volume and occupancy (single-channel volume * single-channel occupancy) was captured in LM1. This shows volume and occupancy measured from a single-channel detector affect each other but the effect is minor (coefficient = 0.002).

All the LMs have fairly high goodness of fit (adjusted $R^2 > 0.9$) except for LM2. Since LM2 was modeled by the data collected during morning peak hours, the data collection sites were more congested than other study sites. Moreover, the sample size is fairly small (100 cycles). Therefore, it is not surprising that the nonlinear relationship between the dependent and other independent variables cannot be better fitted. Overall, the single-channel volume and single-channel occupancy variables are included in the four models. Moreover, the high *t*-values (low *p*-values) for both variables show their strong association with the LMs.

TABLE 3 Linear Models for Volume Correction

Independent Variable	Coefficients	SE	*t*-Value	*p*-Value
LM1 (for two-lane scenario)				
Single-channel volume (vehicles per hour)	0.493	0.003	149.51	.000
Single-channel occupancy (%)	2.539	0.088	28.94	.000
Link length (1,000 feet)	23.610	1.233	19.15	.000
System loop's distance to the intersection	−14.240	0.633	−22.49	.000
Right-turn lane (0: no, 1: yes)	16.950	1.376	12.32	.000
Cycle length (s)	0.450	0.013	33.89	.000
Single-channel volume * single-channel occupancy	0.002	0.000	16.89	.000
Multiple *R*-squared: .9914, adjusted *R*-squared: .9914				
LM2 (for two-lane scenario)				
Single-channel volume (vehicles per hour)	0.3429	0.0123	27.88	.000
Single-channel occupancy (%)	1.9211	0.3764	5.10	.000
Link length (1,000 feet)	64.0464	3.5101	18.25	.000
Multiple *R*-squared: .7376, adjusted *R*-squared: .732				
LM3 (for three-lane scenario)				
Single-channel volume (vehicles per hour)	0.353	0.003	122.60	.000
Single-channel occupancy (%)	2.661	0.058	45.83	.000
Link length (1,000 feet)	2.739	1.628	1.68	.093
Cycle length (second)	0.139	0.025	5.63	.000
Multiple *R*-squared: .9907, adjusted *R*-squared: .9907				
LM4 (for three-lane scenario)				
Single-channel volume (vehicles per hour)	0.440	0.027	16.41	.000
Single-channel occupancy (%)	1.484	0.469	3.16	.002
Multiple *R*-squared: .9928, adjusted *R*-squared: .9925				

Model Verification

To verify the robustness of the model, three measures of accuracy are used in this study: mean absolute error (MAE), mean absolute percentage error (MAPE), and root-mean-square error (RMSE) and are defined as follows (*21*).

$$\text{MAE} = \frac{\sum_{t=1}^{n} |G_t - F_t|}{n} \tag{20}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} (G_t - F_t)^2}{n}} \tag{21}$$

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{G_t - F_t}{G_t} \right| \tag{22}$$

where

G_t = ground truth volume at time interval t,
F_t = forecast (corrected) volume at time interval t, and
n = total number of samples for verification.

Estimation error is defined by the difference between G_t and F_t.

MAE provides an average view of all errors. The MAE difference between the volume collected by the single-channel detector and the volume corrected by the proposed model is regarded as improvement. RMSE also can measure the average magnitude of the error but it gives a relatively high weight to large errors. Hence, RMSE can provide better insight into the large errors. The MAE and RMSE can be used jointly to determine the variation of the errors. The greater the difference between these two measures, the greater the variance in the errors. MAPE can express the error as a percentage. Unlike RMSE, MAPE does not "exaggerate" the error.

To perform before-and-after comparisons, all F_ts in Equations 20 to 22 were replaced with the original single-channel volume collected by the advance detector and all the "before" measures were calculated accordingly. The verification data were collected from all the test sites during the entire period of Tuesday through Thursday, July 7 to 9, 2009.

Two-Lane Scenario

Table 4 and Figure 4 show the model verification results for the two-lane scenario. Figure 4 provides prompt visual comparisons between the single-channel and the corrected volume data by the proposed

TABLE 4 Results of Model Verification for Two-Lane Scenario

Location			MAE			RMSE		MAPE	
Intersection	DR[a]	Models	Before	After	Improvement (%)	Before	After	Before (%)	After (%)
041	EB	NM1	230.8	83.7	63.7	301.7	110.5	48.5	26.5
		NM2		76.4	66.9		101.6		25.5
		LM1		66.1	71.3		86.6		48.2
		LM2		163.6	29.1		195.4		231.9
046	WB	NM1	65.9	44.6	32.3	90.0	68.6	19.7	17.1
		NM2		54.0	18.0		84.4		19.7
		LM1		42.1	36.1		55.8		37.8
		LM2		100.2	−52.2		121.5		261.3
050	NB	NM1	214.7	65.7	69.4	285.3	90.1	30.4	14.6
		NM2		81.3	62.2		113.3		16.2
		LM1		50.8	76.3		68.9		14.6
		LM2		195.8	8.8		240.2		92.3
050	SB	NM1	234.5	81.5	65.2	304.6	107.2	34.5	17.2
		NM2		78.2	66.6		105.0		16.4
		LM1		62.4	73.4		79.0		28.8
		LM2		200.0	14.7		240.2		137.9
065	NB	NM1	145.2	55.4	61.8	208.6	79.0	24.9	16.4
		NM2		63.2	56.4		92.1		18.2
		LM1		44.1	69.6		60.5		21.4
		LM2		186.3	−28.3		221.1		168.0
065	SB	NM1	147.1	67.2	54.3	222.4	98.3	24.9	19.6
		NM2		86.8	41.0		125.2		23.5
		LM1		49.0	66.7		65.6		33.5
		LM2		177.3	−20.5		226.8		161.8
131	SB	NM1	130.9	61.3	53.2	176.6	85.3	27.9	18.3
		NM2		63.2	51.7		91.0		19.0
		LM1		55.8	57.3		71.6		30.7
		LM2		132.9	−1.6		177.4		74.9
Total		NM1	164.1	65.2	60.3	235.4	92.0	29.9	18.6
		NM2		69.4	57.7		102.5		20.0
		LM1		52.6	68.0		70.0		31.1
		LM2		164.1	0.0		205.9		164.5

NOTE: Total sample size: 19,335 cycles; unit: veh/h.
[a]DR = direction; NM = nonlinear model; LM = linear model.

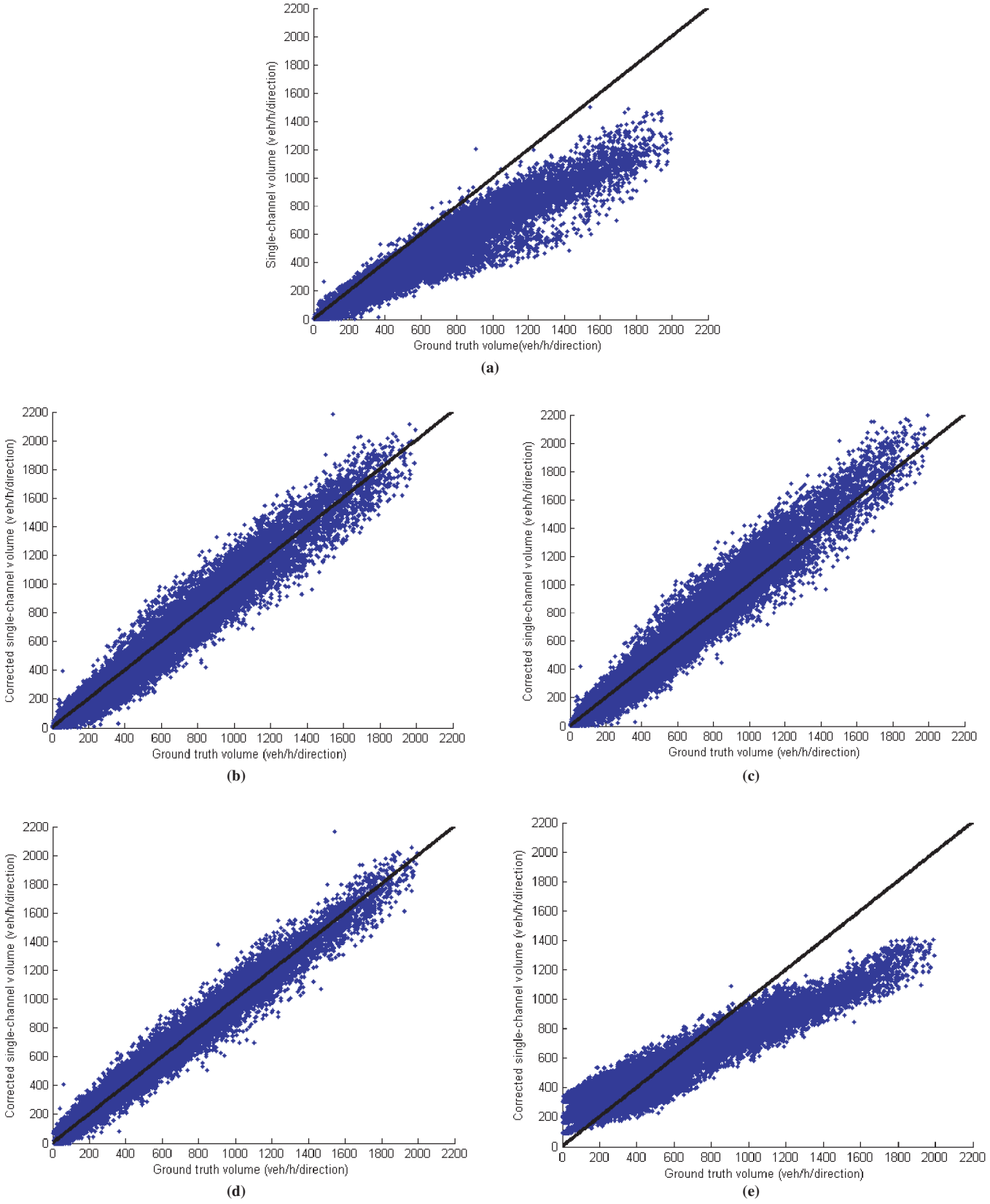


FIGURE 4 Results of volume correction for two-lane scenario (July 6 to 9): (*a*) before correction, (*b*) using NM1, (*c*) using NM2, (*d*) using LM1, and (*e*) using LM2.

models. Table 4 shows the details of the statistics of accuracy. Both NM1 and LM1 provide significant improvement on volume correction. Among all the test sites, NM1 improved the volume data quality by 60.3% (up to 69.4% at Intersection 50 northbound). Though NM1 introduced higher RMSE, it showed lower MAPE. This indicates the NM1 has a lower percentage error but was penalized by the erroneous estimation when the volume was high. On average, LM1 improved the volume data quality by 68.0% [up to 73.4% at Intersection 50 southbound (SB)] and shows relatively low error variance indicated by lower RMSE values. The RMSE difference between both models can be also observed by visually comparing the data "scatteredness" in Figures 4*b* and 4*d*. NM2 and LM2 were built using a very small data set with 100 samples (cycles). Obviously, NM2 is superior to LM2 but also provides similar correction capability as NM1 does. Importantly, this indicates the proposed NMs require minimal calibration effort and are suitable for practical application.

For both LMs and NMs being compared, the higher the original errors were, the more errors the models could correct. Importantly, NM1, NM2, and LM1 can effectively reduce the errors of single-channel volume to 70 veh/h. Overall, LM1 slightly outperformed NM1 by capturing more random effects and reducing the error variance effectively. However, NM2 significantly outperformed LM2 according to all measures of accuracy. In some cases, it even outperformed NM1. Thus, NM2 can be a suitable alternative for NM1. It should be noted that, in Table 4, both NM1 and NM2 have low MAPEs, 18.6% and 20.0%, respectively. This indicated that both NM1 and NM2 can fix the data evenly at all volume levels.

Three-Lane Scenario

Table 5 and Figure 5 show the model verification results for the three-lane scenario. Comparing Figure 5*d* with Figure 5*b*, LM3 fits the data better than NM3 because it effectively corrected the long tail circled in Table 5*a*. This "tail" was the data collected from Intersection 56 (SB). This study site has the longest link length [1,050 ft (320 m)] among all test sites and the parameter, Link length, in LM3 successfully captures this effect.

According to Table 5, all the results are similar to the ones in the two-lane scenario. Similar to the two-lane scenario, LM3 slightly outperforms NM3 by a 5.6% improvement in MAE. Compared with NM3 and NM4, LM3 has lower RMSEs but a higher MAPE. Both NM3 and NM4 have a relatively lower MAPE but higher RMSE than both LMs. It should be noted that NM4 only used a small data set to calibrate but provided similar improvements as NM3 and LM3 did. This is also true for NM2 in the two-lane scenario. This indicates that the NM is also appropriate for correcting single-channel volume data with little calibration effort.

Based on the results shown in Figures 4 and 5, the scatter plots of all NMs, LM1, and LM3 are nearly diagonally symmetric. This symmetry implies that system errors were mostly removed but the scatteredness of the plot also implies the random errors still exist. Though the LMs can capture more random errors, the random errors are fairly difficult to explain in the model. Theoretically, if the probability-based model is calibrated properly, these random errors should be canceled out with increased sample size. That is, the averaged hourly/daily volume estimated by the model should be very close to the ground truth volume.

Figure 6 demonstrated the examples of the application of the proposed models on July 7, 2009. The hourly volumes were estimated by averaging all estimated volumes within every hour. For the two-lane scenario, Intersection 46 westbound (WB) showed the least improvement in errors in Table 4, especially for NM1 (18.0%). However, NM2 only resulted in a slightly higher MAE (8.23 veh/h) compared with the best model's (LM1) MAE (7.43 veh/h). This shows the NM can effectively remove the system errors and random errors could be reduced by increasing sample size.

For the three-lane scenario, Intersection 56 (SB) showed the least improvement in error correction among all three-lane study sites. Compared with LM3's (best model) MAE (40.57 veh/h), NM4 resulted in a slightly lower MAE (35.51 veh/h) instead. It should be noted that NM2 and NM4 were built with 2 h of data collected from another two intersections on July 1, 2009 whereas the LM1 and LM3 were built based on the complete data set collected from all study sites on the same day.

TABLE 5 Results of Model Verification for Three-Lane Scenario

Location			MAE			RMSE		MAPE	
Intersection	DR[a]	Models	Before	After	Improvement (%)	Before	After	Before (%)	After (%)
027	WB	NM3	298.4	68.9	76.9	394.1	92.7	40.3	21.3
		NM4		68.1	77.2		92.1		20.8
		LM3		66.2	77.8		89.4		22.5
		LM4		91.5	69.3		120.9		25.0
047	SB	NM3	157.5	59.0	62.6	224.4	84.4	27.1	16.6
		NM4		55.7	64.6		77.2		16.0
		LM3		60.3	61.7		81.9		18.1
		LM4		86.1	45.3		110.3		23.1
056	SB	NM3	210.6	121.7	42.2	286.1	172.0	17.0	10.7
		NM4		131.4	37.6		184.0		11.3
		LM3		73.4	65.1		102.5		10.8
		LM4		139.7	33.7		171.7		18.7
Total		NM3	220.3	78.0	64.6	308.5	115.0	29.2	16.8
		NM4		78.9	64.2		117.4		16.5
		LM3		65.6	70.2		90.0		17.9
		LM4		101.3	54.0		131.7		22.7

NOTE: Total sample size: 6,247 cycles; unit: veh/h.
[a]DR = direction.

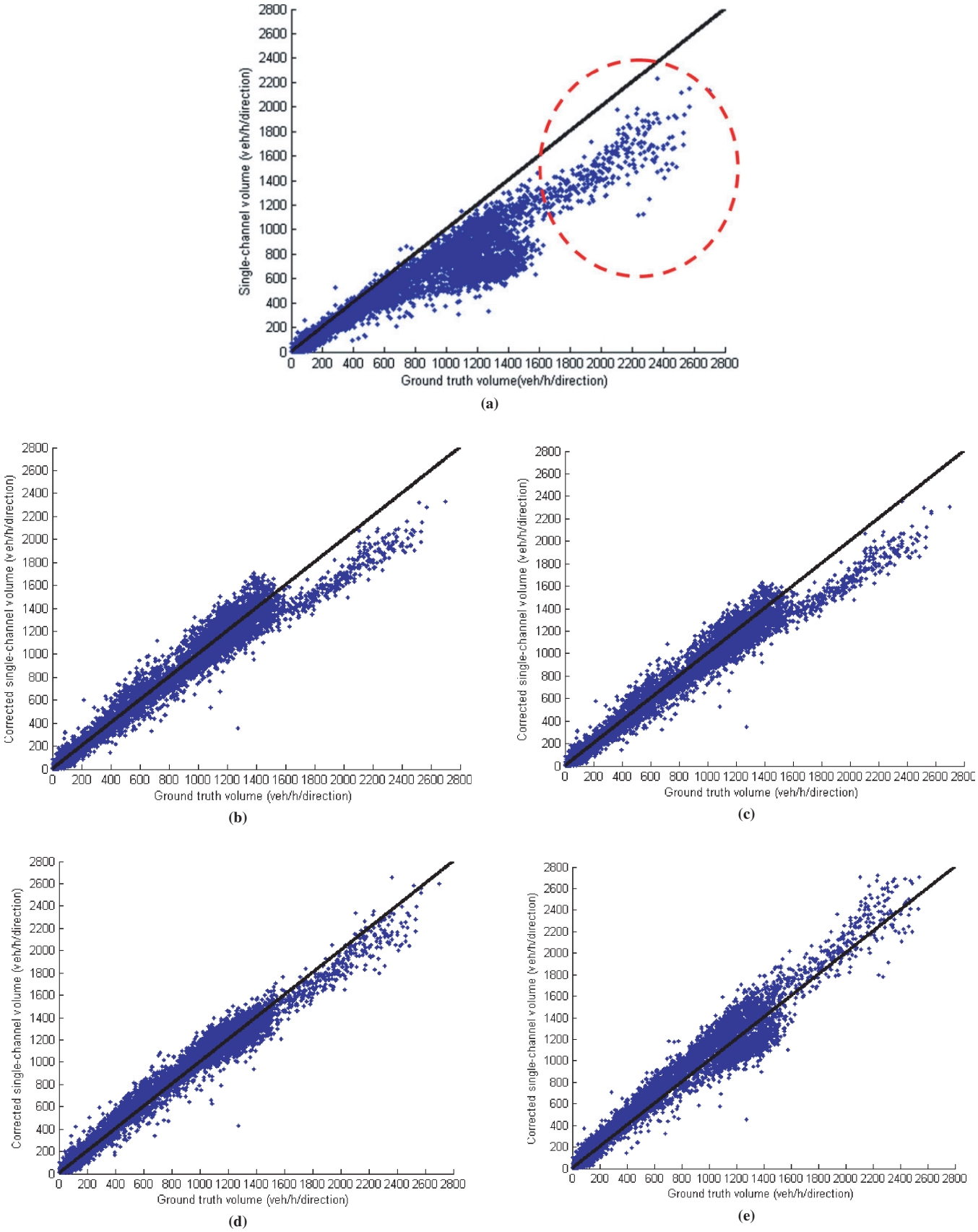


FIGURE 5 Results of volume correction for three-lane scenario (July 6 to 9): (*a*) before correction, (*b*) using NM3, (*c*) using NM4, (*d*) using LM3, and (*e*) using LM4.

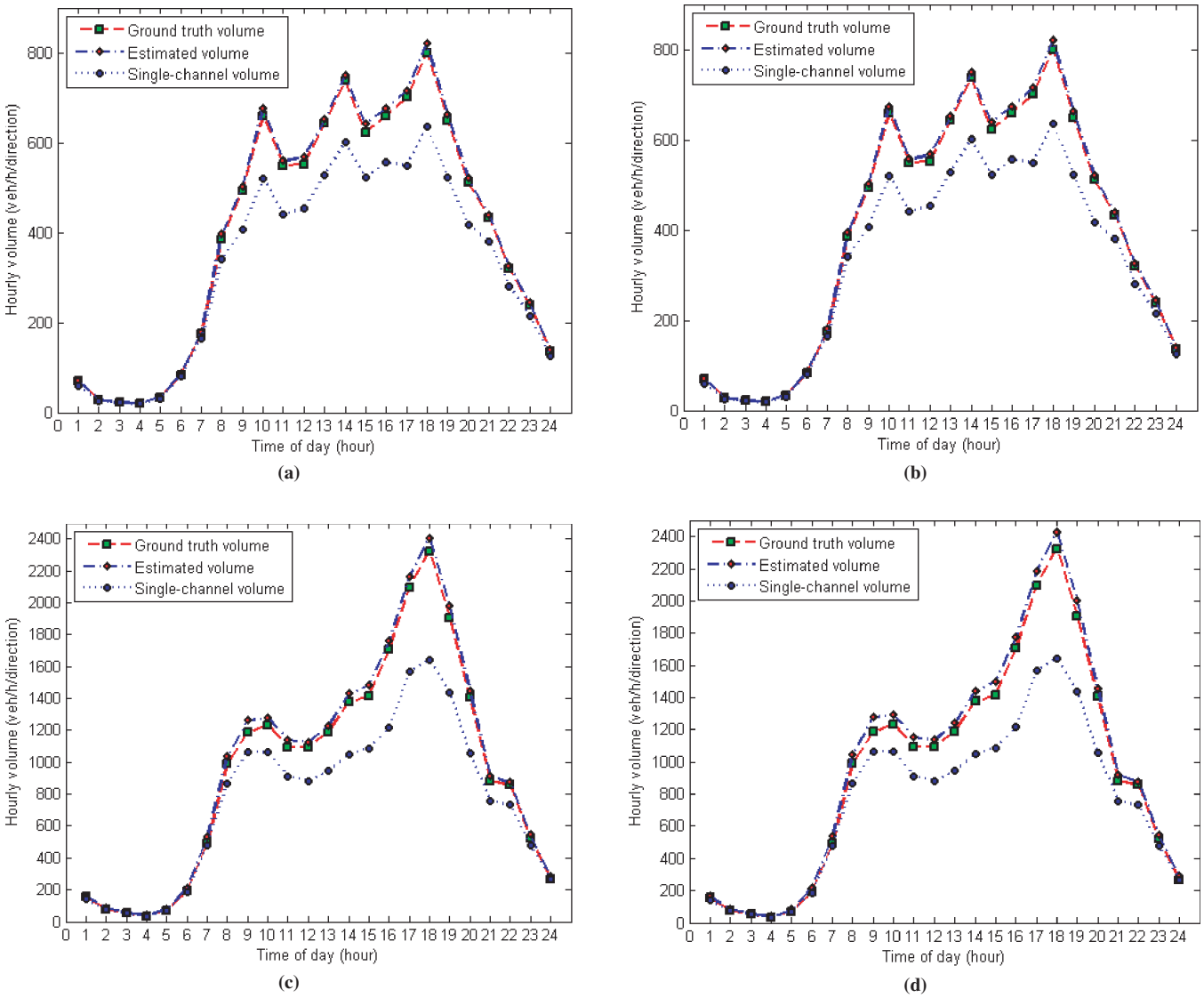


FIGURE 6 Application of models on July 7, 2009: (*a*) NM2, Intersection 46 (WB); (*b*) LM1, Intersection 46 (WB); (*c*) NM4, Intersection 56 (SB); and (*d*) LM3, Intersection 56 (SB).

Based on the model application results, NM2 and NM4, calibrated using a small data set, can estimate hourly volume as accurately as the other well-calibrated models in two- and three-lane scenarios, respectively. The results demonstrate the applicability, transferability (site-independent), and accuracy of the probability-based NMs that can effectively correct single-channel volume.

CONCLUSIONS AND RECOMMENDATIONS

This paper presented a probability-based approach for correcting data errors from single-channel loop detector outputs. Two probability-based NMs were proposed for two- and three-lane scenarios, respectively. Two corresponding LMs were proposed for comparisons. Each NM and LM were calibrated and verified by real-life data. The results showed that both NM and LM can effectively correct the single-channel volume data.

The LM can slightly outperform the NM by reducing more random errors only if all possible variables and a huge data set are included in the modeling process. Based on the model application results, the single-channel volume corrected by the NM was approximately equal to the ground truth volume with minor errors but the model is much simpler than LMs. Only one or two parameters need to be calibrated. Variables, such as cycle length and speed limit, are not required. Moreover, the NM requires much less effort in calibration. In other words, manual counts are sufficient for NM calibration. To provide better estimation, other types of ITS sensors can be used to increase sample size to calibrate the model more precisely. Importantly, the calibrated NMs are location-independent. If the locations have similar characteristics, the NM could be transferable and accuracy would still be maintained.

Though the proposed NMs demonstrated its strength in volume data correction for the single-channel loop detector, model applicability and further studies are recommended as follows. First, these models may not be applicable to the areas with diverse or varying traffic compositions. If the vehicle composition varies greatly in an area, the model parameter(s) may not effectively capture the randomness of traffic characteristics (e.g., varying effective vehicle length) and may lead to biased correction results. The study demonstrated

its applicability in low truck volume area (2% of trucks). However, impacts of high truck volume on the NMs should be further investigated, especially in the industrial areas where a large number of trucks are involved in the traffic. Second, this study was conducted using the data retrieved from the system loop, which may cause additional random errors. Calibrating the NMs using other data sources, especially manual count data, should be further evaluated in future studies. Lastly, the proposed NMs potentially can be extended to correct volume data collected by other configurations of single-channel loops. However, future studies will be needed to verify its applicability.

ACKNOWLEDGMENTS

The authors thank Fred Liang, Mike Whiteaker, and Brook Durant for providing arterial traffic data. The authors also thank Jonathan Corey for English consultation, and give special thanks to Chris Masek and Paul Coffelt for providing professional opinions on single-channel detectors.

REFERENCES

1. Klein, L. A. *Sensor Technologies and Data Requirements for ITS.* Artech House, Boston, Mass., 2001.
2. Klein, L. A., M. K. Mills, and D. R. P. Gibson. *Traffic Detector Handbook.* FHWA, U.S. Department of Transportation, Turner–Fairbank Highway Research Center, McLean, Va., 2006.
3. Mirchandani, P., and L. Head. A Real-Time Traffic Signal Control System: Architecture, Algorithms, and Analysis. *Transportation Research Part C,* Vol. 9, No. 6, 2001, pp. 415–432.
4. Liu, H. X., and W. Ma. A Virtual Vehicle Probe Model for Time-Dependent Travel Time Estimation on Signalized Arterials. *Transportation Research Part C,* Vol. 17, No. 1, 2009, pp. 11–26.
5. Tarko, A. P., K. Choocharukul, A. Bhargaa, and K. C. Sinha. Simple Method of Predicting: Travel Speed on Urban Arterial Streets for Planning Applications. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1988,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 48–55.
6. Zhang, H. M. Link–Journey–Speed Model for Arterial Traffic. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1676,* TRB, National Research Council, Washington, D.C., 1999, pp. 109–115.
7. Wall, Z. R., and D. J. Dailey. Algorithm for Detecting and Correcting Errors in Archived Traffic Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1855,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 183–190.
8. Nihan, N. L., Y. Wang, and P. Cheevarunothai. *Improving Dual-Loop Truck (and Speed) Data: Quick Detection of Malfunctioning Loops and Calculation of Required Adjustments.* Department of Civil and Environmental Engineering, University of Washington, Seattle, 2006.
9. Cheevarunothai, P., Y. Wang, and N. L. Nihan. Using Dual-Loop Event Data to Enhance Truck Data Accuracy. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1993,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 131–137.
10. Wang, Y., and N. L. Nihan. Can Single-Loop Detectors Do the Work of Dual-Loop Detectors? *ASCE Journal of Transportation Engineering,* Vol. 129, No. 2, 2003, pp. 169–176.
11. Yu, R., G. Zhang, and Y. Wang. Loop Detector Segmentation Error and Its Impacts on Traffic Speed Estimation. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2099,* Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 50–57.
12. Smaglik, E. J., D. M. Bullock, J. R. Sturdevant, and T. Urbanik II. Implementation of Lane-by-Lane Detection at Actuated Controlled Intersections. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2035,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 81–87.
13. Smaglik, E. J., D. M. Bullock, and T. Urbanik II. Evaluation of Lane-by-Lane Vehicle Detection for Actuated Controllers Serving Multilane Approaches. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1925,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 123–133.
14. Tian, Z. Z., and T. Urbanik II. Green Extension and Traffic Detection Schemes at Signalized Intersections. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1978,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 16–24.
15. Wang, X. Evaluation of Lane-by-Lane Detection at Signalized Intersections Using Simulation. *ITE Journal,* Vol. 78, No. 11, 2008, pp. 16–21.
16. Zhang, G., Y. Wang, H. Wei, and Y. Chen. Examining Headway Distribution Models with Urban Freeway Loop Event Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1999,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 141–149.
17. Mannering, F. L., S. S. Washburn, and W. P. Kilareski. *Principles of Highway Engineering and Traffic Analysis.* John Wiley and Sons, Inc., Hoboken, N.J., 2009.
18. May, A. D. *Traffic Flow Fundamentals.* Prentice Hall, Englewood Cliffs, N.J., 1990.
19. Leon-Garcia, A. *Probability and Random Processes for Electrical Engineering.* Addison Wesley, Reading, Mass., 1994.
20. Ritz, C., and J. C. Streibig. *Nonlinear Regression with R.* Springer, New York, 2008.
21. Washington, S., M. G. Karlaftis, and F. L. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis.* Chapman & Hall/CRC, Boca Raton, Fla., 2003.
22. Faraway, J. J. *Linear Models with R.* Chapman & Hall/CRC, Boca Raton, Fla., 2005.

The Highway Traffic Monitoring Committee peer-reviewed this paper.