Travel Time Estimation with Correlation Analysis of Single Loop Detector Data

Huairui Guo and Jionghua (Judy) Jin

In this paper, the average travel time over a link is considered as a random variable following an identical probability distribution as the arrival process. A new estimation method of the average travel time uses a cross-correlation analysis of traffic flow measurement data. This method requires only traffic flow information, which is available from the measurements of single loop detectors upstream and downstream from one link. Different from the existing maximum cross-correlation analysis method, the proposed method considers average travel time as a random variable, with its mean value estimated from all significant cross-correlation coefficients rather than from only the maximum cross-correlation coefficient. Therefore, the inherent variability of average travel time among different vehicles can be considered. Moreover, different from the existing optimization method, the proposed method uses the statistical *t*-test of the significant cross-correlation coefficients to determine automatically and adaptively the fitting range of the probability density function of the average travel time. Thus, it avoids using the approximated car length factor and has no need to predetermine the range of the average travel time as required by the optimization method. Details of the average travel time estimation procedures are presented, and the effectiveness of the proposed method is demonstrated through both simulation study and a case study of real traffic data.

The accurate estimation of link travel time is essential to the development of an intelligent transportation system. Travel time estimation provides not only valuable information for traveler routing or transportation scheduling but also the potential capability for incident detection. Single loop detectors have been widely used in arterial roads and freeways for traffic condition monitoring and control. Generally, such detector data can be used to measure traffic flow directly but not vehicle speed and travel time.

Various models have been developed to estimate travel time. Among them, the most simple and fundamental model is to use the basic relationship among speed, volume, and occupancy (*1–3*). The basic idea of this model is briefly reviewed as follows.

For a given link, the relationship of average speed (v), flow (q), and occupancy (o) is expressed as

$$v = \frac{q}{o \cdot g} \tag{1}$$

H. Guo, Reliasoft Corporation, 115 South Sherwood Village Drive, Tucson, AZ 85710. J. Jin, Industrial and Operations Engineering Department, University of Michigan, 1205 Beal Avenue, Ann Arbor, MI 48109-2117.

Transportation Research Record: Journal of the Transportation Research Board, No. 1968, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 10–19.

where g is the car length factor that is a function of the lengths of vehicles and detectors. The average travel time (T) is estimated by

$$T = \frac{s}{v} \tag{2}$$

where s is the length of a given link.

Although the model of Equations 1 and 2 provides a simple way to estimate average travel time, its accuracy is sensitive to the car length factor g, which is further elaborated later. In addition, the model does not consider travel time variability among vehicles.

Recently, some other techniques were developed to estimate travel time. For example, Coifman used instant velocity measurement by a dual loop detector (*4*). This method is effective for links that have been equipped with dual loop detectors. However, it is not suitable for links that are equipped with only single loop detectors, because a single loop detector cannot measure speed directly.

Another technique is to use a cumulative flow plot to calculate total travel delay and estimate average travel time (*5, 6*). This technique requires the priori knowledge about the number of vehicles existing in the link initially, which usually is not available in practice. In addition, there is no systematic way to choose the start time to plot the departure curves. Thus, implementation of this technique is limited. Nam and Drew report use of a cumulative flow plot to estimate the travel time in a detailed analysis and present the new development of the flow–density–speed relationship of their models (*7*). However, this method has the same problem as other methods that use a cumulative flow plot (*5, 6*).

Some different methods were developed to estimate travel time by applying advanced signal processing and intelligent classification techniques such as state space modeling, neural network classification, and car identification and signature matching. Park et al. use neural networks to forecast real-time travel time for 5 min ahead (*8*). Hoogendoorn et al. propose a method that describes state space dynamics of traffic with recurrent neural networks (*9*). Both methods require the appropriate selection of training data sets to conduct extensive training of the neural network models. Coifman and Cassidy developed an algorithm to estimate travel time through vehicle reidentification, in which vehicle length is used as a "signature" to match the vehicle in the downstream to the vehicle in the upstream (*5*). Thus, the vehicle length must be measured or estimated accurately to ensure the vehicles are correctly reidentified.

To consider the stochastic nature of traffic, some stochastic models of traffic flow have been developed in the past decade. Dailey uses cross-correlation analysis to estimate travel time, assuming that the average travel time is a deterministic variable (*10*). Because it uses only the maximum cross-correlation coefficient, it is effective when different vehicles have close travel speeds. More recently, Dailey modifies the model of Equation 1 by considering the measurement errors of the car length and occupancy (*11*). The modified model considers the occupancy, speed, and vehicle length as random variables following certain probability distributions. Then, a Kalman filter is used to estimate the actual speed of each vehicle. This method requires a prior understanding of the probability distribution function of various random variables involved in the model. A significant estimation error—even instability or nonconvergence—may occur if the mean or variance of the random variables were not properly estimated.

In addition, by considering the random characteristics of average travel times, Petty et al. developed a probabilistic regression model with the assumption that average travel time is a random variable with the same probability distribution function over a given link for a given estimation window of travel time (*12*). Although this method is robust to congestion conditions, its effectiveness depends highly on prior knowledge of the fit range of the probability density function of the average travel time. Petty et al. adaptively estimate the center of the fit range on the basis of Equation 1 with a fixed value of *g*. The change of the car length factor due to the instant change of occupancy is not considered; this could lead to an estimation error on the fit center of the probability density function. Moreover, the selection of a fit range also affects estimation performance. The influence of these two factors on the estimation of average travel time are further illustrated later.

In summary, the existing travel time estimation methods limited by the use of traffic flow data from single loop detectors can be classified in three categories. The first method is rooted in the model of Equation 1 for travel speed estimation, which is sensitive to the car length factor estimation error. The second method considers the stochastic nature of the traffic flow and uses the maximum cross-correlation coefficient between the upstream and downstream flow data to estimate average travel time, which ignores travel time variability among different vehicles. The third method develops a probabilistic regression model to estimate the probability density function of the average travel time, which requires prior determination of the fit range of the estimated probability density function of the average travel time. Therefore, the aim of this paper is to improve existing methods of travel time estimation on the basis of flow measurement data from upstream and downstream single loop detectors.

In this paper, a new method is developed to estimate travel time by integrating the cross-correlation analysis technique with the probabilistic model of the random average travel time. Resembling the method of Petty et al. (*12*), a probabilistic regression model is used to consider the randomness of the average travel time. The advantage of the proposed analysis method is to improve estimation by automatically determining the fit range of the probability density function of the average travel time adaptively. Moreover, the proposed estimation method does not require the estimation of a car length factor, which may lead to significant estimation error. The only information needed is the downstream and upstream flow measurement.

In the following sections, the probabilistic regression model is briefly reviewed, and the impact of errors in car length factor estimation on the existing travel time estimation method is discussed. Then, a new cross-correlation analysis method is proposed to consider the travel time variability among different vehicles. Statistical hypothesis testing is used to automatically determine the fit range of the probability density function of the average travel time. Next, a simulation study using VISSIM software and a case study of real traffic data are conducted to demonstrate the analysis procedures and the effectiveness of the proposed method. Finally, findings are summarized and conclusions presented.

REVIEW: ESTIMATING TRAVEL TIME WITH FLOW MEASUREMENTS

Probabilistic Regression Model

The proposed travel flow model is based on the probabilistic regression model proposed by Petty et al. (*12*). It assumes that the arrivals measured at the upstream point within a given time interval (called the estimation window) have the same probability density function as the average travel time over the link. If at sampling time t the numbers of upstream and downstream arrivals during the sampling interval Δ are $x(t)$ and $y(t)$, respectively, then the relationship between these upstream and downstream arrivals can be modeled as

$$y(t) = \sum_{i=a_1}^{a_2} \left[x(t-i)\, f_i \right] \tag{3}$$

where f_i is the probability density function of the average travel time over the link, and a_1 and a_2 correspond to the shortest and longest average travel times, respectively, over all upstream vehicles; (a_1, a_2) is called the fit range of the probability density function. Petty et al. estimate f_i to minimize the sum of squares of the regression residual errors by solving the following constrained optimization problem:

$$\min_{f_i} \left\{ \sum_{t=b_1}^{b_2} \left[y(t) - \sum_{i=a_1}^{a_2} x(t-i)\, f_i \right]^2 \right\} \tag{4}$$

subject to

$$\left\{ \mathbf{f}: f_i \geq 0 \qquad \sum_{i=a_1}^{a_2} f_i = 1 \right\}$$

where b_1 and b_2 are the start time index and end time index of the estimation window (b_1, b_2) (*12*). Thus, the mean estimate of the average travel time is

$$\hat{T} = \Delta \sum_{i=a_1}^{a_2} f_i \cdot i \tag{5}$$

From Equation 4, it can be seen that if the predefined fit range of (a_1, a_2) is too wide, there will be too many parameters f_i to be estimated; this may lead to an unstable solution in the optimization solver. Meantime, if the fit range is too narrow, the resolution and accuracy of the estimated probability density function may be too poor to get an appropriate probability distribution function. Therefore, it is critical to define an appropriate fit range of (a_1, a_2). As reported by Petty et al. (*12*), the fit range is fixed as 20 s, independent of the occupancy change, and the center of the mean travel time is adaptively estimated on the basis of Equation 2 with a fixed car length factor of g in Equation 1. The following analysis illustrates how this assumption of fixed car length affects the fit center of (a_1, a_2) used in the travel time estimation of Equation 4.

Impact Analysis of Car Length Factor Estimation Error

Hall and Persaud thoroughly discuss the relationship between speed, occupancy, and car length factor g (*2*). Their results show a strong correlation between occupancy and average car length. With increasing occupancy, average car length also increases. So, car length should be adaptively estimated for different occupancies. Now, how the estimation error of using a fixed car length factor affects the accuracy of travel time estimation is further illustrated.

From Equations 1 and 2, the fit center, $\hat{T}_c$, of (a_1, a_2) used in the Petty et al. estimation method (*12*) is obtained as

$$\hat{T}_c = \frac{s \cdot o}{q \cdot l} \tag{6}$$

where l (which represents the average effective car length, or the sum of the average car length and the width of the loop detector) is equal to $1/g$. When dual loop detectors are used, traffic flow (q) and traffic occupancy (o) are directly obtained and adaptively used to estimate $\hat{T}_c$, which is used as the center of the fit range of (a_1, a_2) in every estimation window of (b_1, b_2). In this case, a fixed car length is used in Equation 6. The resulting deviation error Δl due to incorrectly using the fixed average car length can cause the estimation error ΔT_c as

$$\Delta T_c = -\frac{s \cdot o}{l^2 \cdot q}\Delta l \tag{7}$$

The impact of Δl on ΔT_c is further investigated through a simulation study with VISSIM simulation software. The conditions used in the simulation study are set as follows: link distance $s = 300$ m, average traffic flow $q = 1{,}000$ vehicles/h (0.28 vehicles/s). The estimation window is 1 min. For each estimation window, the occupancy and the corresponding average car length are calculated from simulation data. The relationship between occupancy and average car length is plotted in Figure 1. The overall average estimated car length is 6.52 m, and the corresponding occupancy is 0.12. Dividing Equation 7 by Equation 6 yields

$$\frac{\Delta T}{\hat{T}_c} = \frac{-\Delta l}{l} \tag{8}$$

From Equation 8 it can be seen that the estimation error of the average travel time is proportional to the estimation error of the average car length. If the overall average car length of 6.52 m is used for the value of l in Equation 6, the range of $\Delta l/l$ will be (–0.29, 0.36) in the above simulation. The effect of the estimation error of the average car length on the estimated average travel time $\hat{T}_c$ is obvious. So, if this inaccurate $\hat{T}_c$ value is further used as the center of the fit range (a_1, a_2), an estimate of average travel time could be biased because the predefined fit width $\Delta = a_2 - a_1$ may not be sufficiently cover all possible travel time values under different occupancy conditions. The influence of using a fixed car length on the travel time estimation is further illustrated through a simulation analysis later, in the section on the comparison of different travel time estimation methods.

In the following sections, a new method is proposed to automatically determine the fit range of (a_1, a_2) on the basis of statistical correlation analysis and t-test.

PROPOSED CORRELATION ANALYSIS METHOD

Upstream Flow and Downstream Flow

Two data series $x(t)$ and $y(t)$ denote travel flow upstream and downstream, respectively. The cross-covariance $\gamma_{yx}(k)$ of $x(t)$ and $y(t)$ and the autocovariance $\gamma_x(k)$ of $x(t)$ with lag k are defined as

$$\gamma_{yx}(k) = E\{[y(t) - u_y][x(t-k) - u_x]\} \tag{9}$$

$$\gamma_x(k) = E\{[x(t) - u_x][x(t-k) - u_x]\} \tag{10}$$

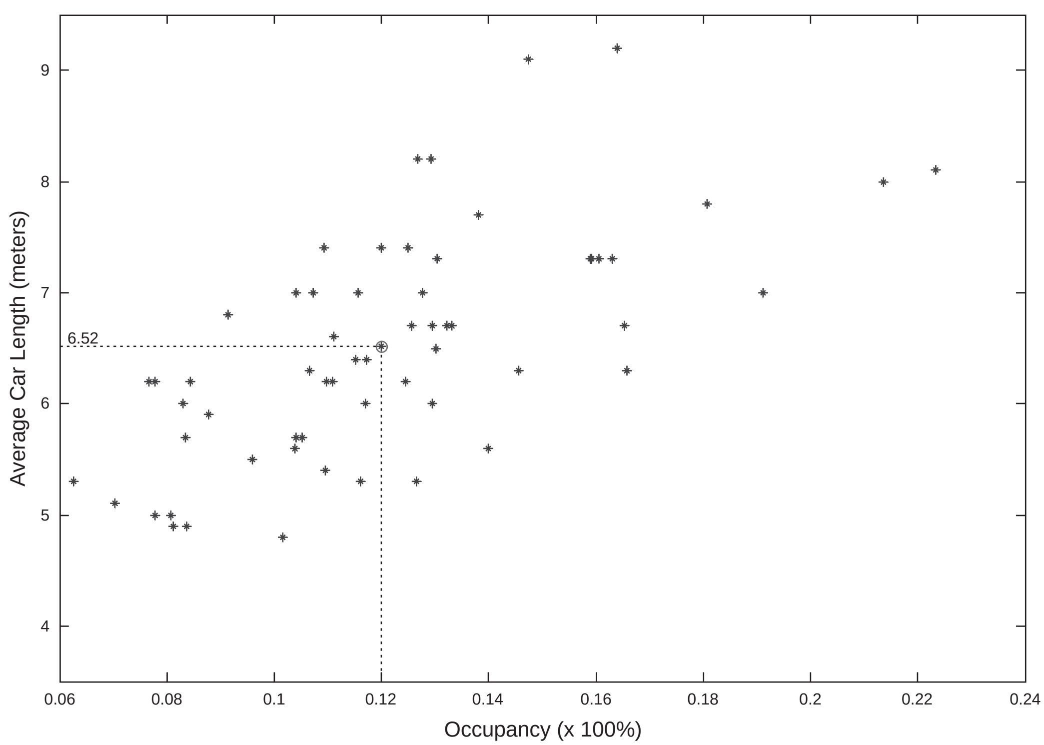

FIGURE 1 Correlation between average car length and occupancy.

Thus, the cross-correlation $\rho_{yx}(k)$ and autocorrelation coefficients $\rho_x(k)$ are

$$\rho_{yx}(k) = \frac{\gamma_{yx}(k)}{\sigma_x \sigma_y} \tag{11}$$

$$\rho_x(k) = \frac{\gamma_x(k)}{\sigma_x^2} \tag{12}$$

where σ_x and σ_y are the standard deviation of $x(t)$ and $y(t)$, respectively, and $x(t)$ and $y(t)$ are assumed to be stationary series within the estimation window of (b_1, b_2). From Equation 3,

$$u_y = \sum_{i=a_1}^{a_2} f_i u_x \tag{13}$$

where $u_y = E(y_t)$ and $u_x = E(x_t)$. On the basis of Equations 3 and 9–13,

$$\begin{aligned} \rho_{yx}(k) &= \frac{1}{\sigma_x \sigma_y} \cdot E\left\{\left[\sum_{i=a_1}^{a_2} f_i x(t-i) - \sum_{i=a_1}^{a_2} f_i u_x\right]\left[x(t-k) - u_x\right]\right\} \\ &= \frac{1}{\sigma_x \sigma_y} \cdot \sum_{i=a_1}^{a_2} f_i \gamma_x(i-k) \\ &= \frac{\sigma_x}{\sigma_y} \cdot \sum_{i=a_1}^{a_2} f_i \rho_x(i-k) \end{aligned} \tag{14}$$

The unknown parameters f_i $[i \in (a_1, a_2)]$ can be estimated through the correlation analysis of upstream and downstream flow.

Although the proposed approach uses the correlation analysis to estimate the travel time, there is a major difference from Dailey's maximum cross-correlation method (*10*). Dailey assumes that the average travel time is one fixed value and uses only the maximum cross-correlation coefficient for average travel time estimation and thus does not consider the variability of average travel time due to different travelers. In the proposed method, the average travel time is assumed to be a random variable over different travelers, and Equation 14 is used to estimate a set of f_i as the discretized probability density function by the sampling interval Δ, in which multiple significant correlation coefficients of $\rho_{yx}(k)$ will be used as follows.

Under free-flow conditions, upstream $x(t)$ are assumed to be independent samples, thus

$$\rho_x(i-k) = \begin{cases} 1 & i = k \\ 0 & \text{others} \end{cases} \tag{15}$$

On the basis of Equations 14 and 15,

$$f_k = \frac{\sigma_y}{\sigma_x} \rho_{yx}(k) \tag{16}$$

From Equation 16, f_k can be determined solely by the cross-correlation coefficients of $\rho_{yx}(k)$. The task of estimating f_k is transferred to find the nonzero cross-correlation coefficients, which is simpler and straightforward. The proposed method does not need to predetermine the car length factor and the fit ranges of (a_1, a_2) for f_i. Moreover, the proposed method using Equation 16 can avoid the unstable solution problem that occurs in the optimization approach.

Hypothesis Testing of Cross-Correlation Coefficients

The aforementioned analysis indicates that the discretized average travel time distribution can be estimated directly from the nonzero cross-correlation coefficients. In practice, the true value of $\rho_{yx}(k)$ is unknown. Thus, the sample cross-correlation function $\hat{\rho}_{yx}(k)$ is used, which is calculated from the flow data within the estimation window (b_1, b_2):

$$\hat{\rho}_{yx}(k) = \frac{\sum_{t=b_1+k}^{b_2} \left[x(t-k) - \bar{x}\right]\left[y(t) - \bar{y}\right]}{\sqrt{\sum_{t=b_1}^{b_2-k} \left\{\left[x(t) - \bar{x}\right]^2\right\} \left\{\sum_{t=b_1+k}^{b_2} \left[y(t) - \bar{y}\right]^2\right\}}} \tag{17}$$

where

$$\bar{x} = \frac{\left[\sum_{t=b_1}^{b_2} x(t)\right]}{(b_2 - b_1 + 1)}$$

and

$$\bar{y} = \frac{\left[\sum_{t=b_1}^{b_2} y(t)\right]}{(b_2 - b_1 + 1)}$$

To determine the nonzero cross-correlation coefficients, the following statistical hypothesis test is used:

H_0:
$\rho_{yx}(k) = 0$

H_1:
$$\rho_{yx}(k) \neq 0 \tag{18}$$

When the estimation window is appropriately selected to include enough number of data, $\hat{\rho}_{yx}(k)$ under H_0 approximately follows a normal distribution with a mean of 0, and the standard deviation $\sigma_{\hat{\rho}}$ is

$$\sigma_{\hat{\rho}} = \frac{1}{\sqrt{n}} \tag{19}$$

where n [the number of data points used to estimate $\hat{\rho}_{yx}(k)$] is equal to $b_2 - b_1 - k + 1$ (*13*). Thus, a t-test can be used to check whether $\rho_{xy}(k) = 0$ under a given α value. The statistic $t_{\rho,k}$ is defined as

$$t_{\rho,k} = \frac{\hat{\rho}_{yx}(k)}{\sigma_{\hat{\rho}}} \tag{20}$$

Under the given α error, if the condition of $t_{\rho,k} > t_{1-\alpha/2,n}$ holds, H_0 is rejected. Thus, the cross-correlation at lag k is significant, and f_k calculated from Equation 16 is the estimated probability corresponding to the average travel time $k \cdot \Delta$.

Under severe congestion conditions during the estimation window, no significant cross-correlation coefficients will be found. In fact, under congestion, it is more important to monitor traffic condition than to estimate travel time. The lack of significant cross-correlation coefficients is an effective method of detecting congestion. One example of how to automatically determine significant cross-correlation coefficients is presented in the section about the simulation and case studies.

The proposed model is based on cross-correlation values of upstream and downstream volume, which require real-time flow information from each sampling interval. Traditional flow-density models do not have this information and thus cannot be used to validate the proposed model analytically. Therefore, VISSIM is used to conduct a series of simulations and study how the traffic volume will affect the cross-correlation values. A 300-m-long link with a capacity of 1,900 vehicles/h is used. The estimation window is 10 min. According to the above analysis, $t_{0.95,600} = 1.96$. Therefore, the threshold value of rejecting H_0 is 0.08. The traffic flow is set initially at 200 vehicles/h and ends at 1,900 vehicles/h, with step increases of 100 vehicles/h. For each setting, the simulation lasted 1 h. Averages of all the maximum cross-correlation values of each estimation window are calculated and plotted in Figure 2.

Under the free-flow condition, increases in traffic volume can lead to decreases in the maximum cross-correlation coefficients. However, for all the volumes that are less than the road capacity, the cross-correlation values are >0.08, which indicates a strong cross-correlation between upstream and downstream traffic flows. Therefore, the proposed method can be used effectively to estimate travel time under the free-flow condition.

Mean Estimate of Average Travel Time

The estimation of f_i using Equation 16 considers only those nonzero probability values of f_i that correspond to a certain percentage of vehicles that have the same average travel time over the link. The summation of all estimated f_i represents the percentage of upstream vehicles within the estimation window of (b_1, b_2) used for estimating average travel time, that is,

$$\sum_{i \in c_1}^{c_m} f_i = P$$

Therefore, among the upstream vehicles used for the travel time estimation, $100\, f_i/P\%$ of vehicles have the travel time $c_i \cdot \Delta$, $i = 1, \ldots, m$.

So, the mean estimate of the average travel time based on these $100P\%$ upstream vehicles is

$$\hat{\mu}_T = \Delta \cdot \sum_{i=c_1}^{c_m} \left[\left(f_i / P \right) \cdot i \right] \tag{21}$$

The mean estimate using Equation 21 ignores the effect of the small probability of the vehicles having either extremely long travel time (or never arriving at the downstream detector) or extremely short travel time; this provides a more realistic value of the mean estimate of the average travel time. With the sample cross-correlation function of Equation 17 substituted into Equation 21, the mean estimate of the average travel time can be obtained:

$$\hat{\mu}_T = \Delta \cdot \frac{\sum_{i=c_1}^{c_m} \left\{ \left[\hat{\rho}_{yx}(i) \right] \cdot i \right\}}{\sum_{i=c_1}^{c_m} \left[\hat{\rho}_{yx}(i) \right]} \tag{22}$$

SIMULATION STUDY AND CASE STUDY

Description of Simulation Conditions

The four simulation studies conducted are discussed in the following subsections. VISSIM simulation software was used to generate the travel flow data and obtain references of the true travel time values for the simulation studies. The total simulation time was 5 hours, but only data generated in the past 4 hours were used for the analysis. The common conditions used for the four simulation studies are

- traffic flow $q = 1{,}000$ vehicles/h,
- speed $v = 45$ to 60 km/h (12.5 to 16.67 m/s),
- link distance $s = 300$ m,

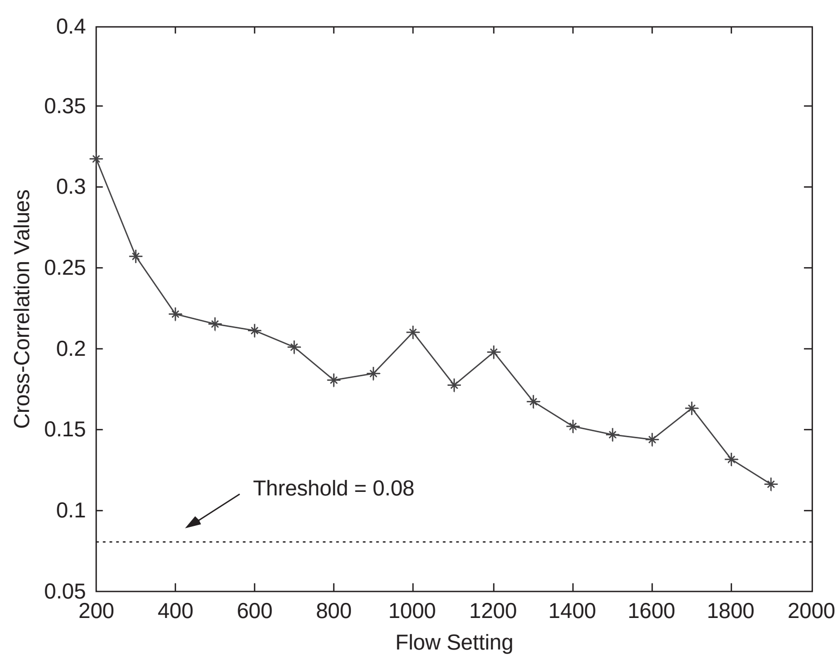

FIGURE 2 Cross-correlation between *x* and *y* under different flow conditions.

- data sampling interval of flow measurement $\Delta = 1$ s, and
- width of the estimation window $b_2 - b_1 = 10$ min.

Cross-Correlation Analysis and *t*-Test

This simulation study illustrates the procedures of the cross-correlation analysis and the *t*-test discussed in the section on hypothesis testing of cross-correlation coefficients. As an example, upstream and downstream data are selected from an arbitrary estimation window. The sample cross-correlation function $\hat{\rho}_{yx}(k)$ $(0 < k \leq 35)$ is calculated on the basis of Equation 17, as shown in Figure 3. Because $n = b_2 - b_1 - k + 1$ is large, the value of $t_{1-\alpha/2, b_2-b_1-k+1}$ will not depend on the lag k. If α error is selected as 5%, then $t_{1-\alpha/2, b_2-b_1-k+1} = t_{0.95,600-k+1} = 1.96$. So, on the basis of Equation 20, the threshold η for determining the significant cross-correlation coefficients is obtained as

$$\eta = \frac{t_{0.95,600-k+1}}{\sqrt{n}} = \frac{1.96}{\sqrt{600-k+1}} \approx \frac{1.96}{\sqrt{600}} = 0.08$$

From Figure 3, it is easy to see that the cross-correlation coefficients from lag 20 to lag 25 are significant because they are larger than the threshold, 0.08. So, on the basis of Equation 2, the mean estimate of the average travel time over the link is obtained as

$$\hat{\mu}_T = \frac{\Delta \cdot \sum_{i=20}^{25} \{[\hat{\rho}_{xy}(i)] \cdot i\}}{\sum_{i=20}^{25} [\hat{\rho}_{xy}(i)]} = 22.35$$

This analysis indicates that, with the *t*-test, the proposed estimation method has no requirement on the preselection of the fit range of (a_1, a_2) and the optimization solver is not needed.

Comparison of Different Travel Time Estimation Methods

This simulation is used to compare the performance of different methods of travel time estimation. As is discussed earlier, two parameters need to be set using the optimization method of Petty et al. (*12*). One is the width of the fit range, equal to $a_2 - a_1 = 20$ s. The other is the fixed average car length (used in Equation 6 for the estimation of the center of the fit range). For comparison of the optimal performance of the Petty et al. method (ignoring the bias in the estimation of average car length), the true average car length of 6.5 m, obtained from VISSIM software, is used in the simulation. The travel time estimation results—obtained by using the optimization method, a single (maximum) cross-correlation coefficient, and the proposed method—are all compared with true travel time in Figure 4.

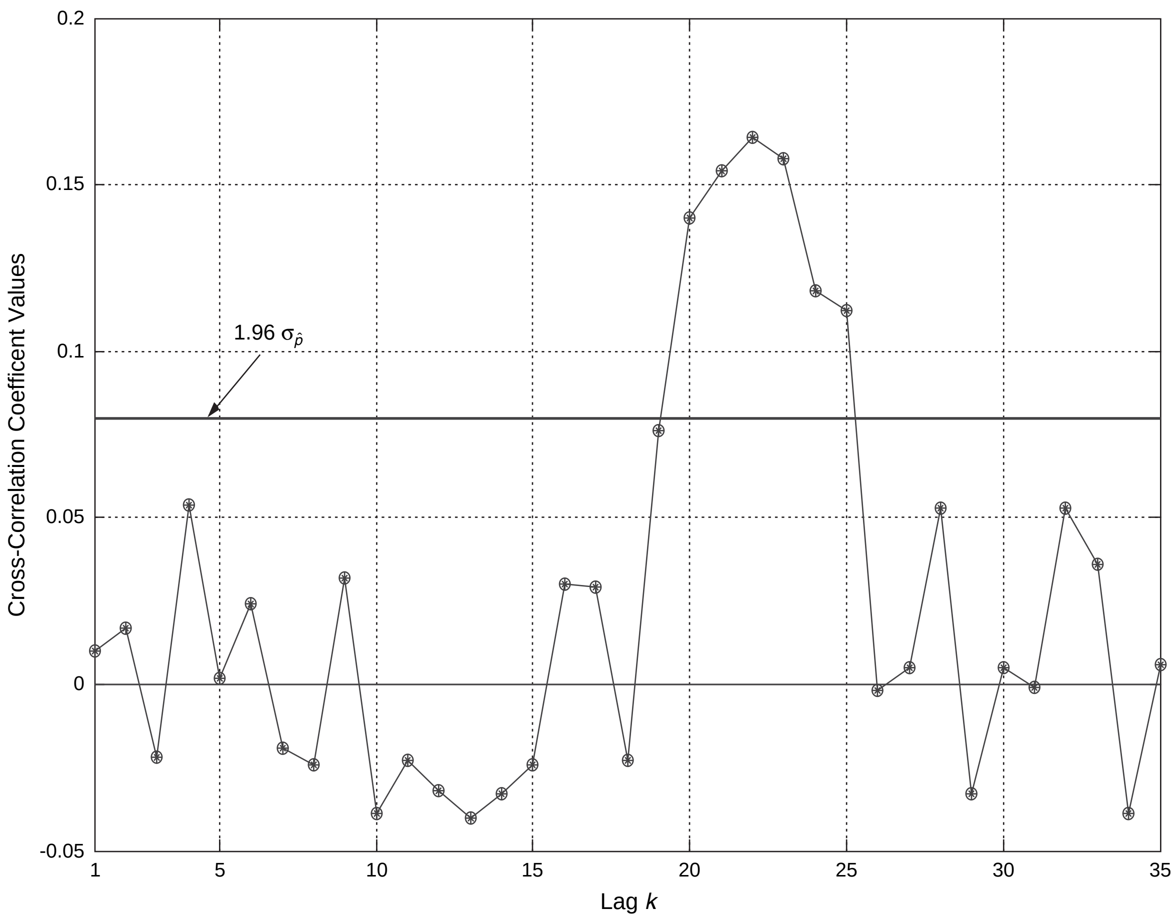

FIGURE 3 Cross-correlation analyses between upstream and downstream.

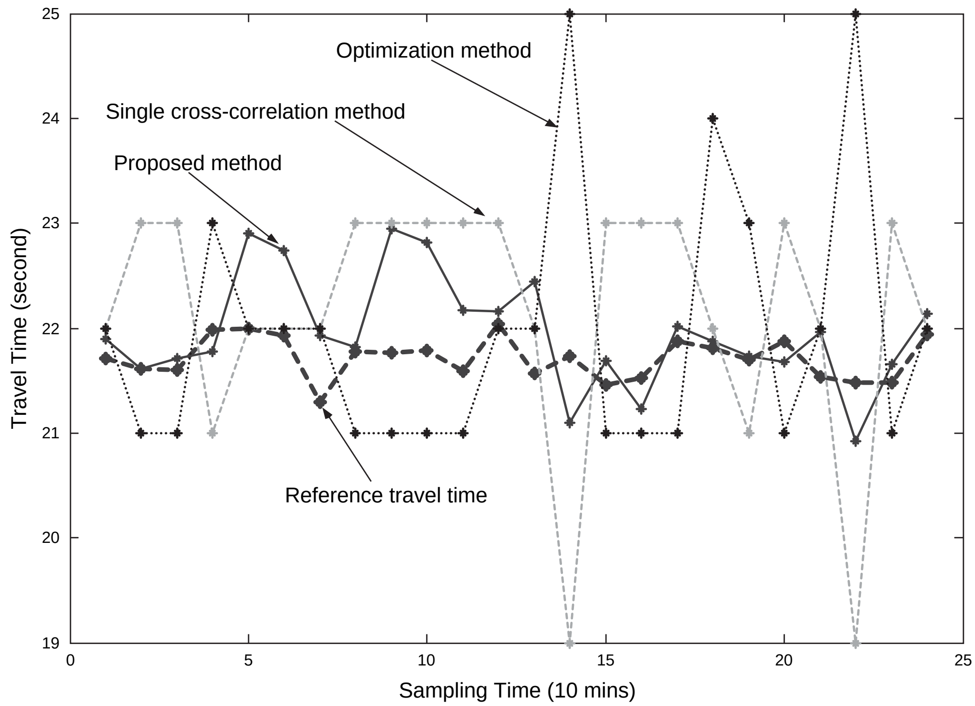

FIGURE 4 Comparison of different travel time estimation methods.

For further comparison, the estimation errors of each method ($i = 1, 2, 3$) are calculated by the deviation e_i of the estimated travel time $\hat{\mu}_{T_i}$ from the true travel time $\hat{\mu}_{T_0}$, that is, $e_i = \hat{\mu}_{T_i} - \hat{\mu}_{T_0}$. The average and standard deviation of the estimation errors of each method are compared in Table 1, which indicates that the performance of the proposed method is better than that of the other two methods.

Effect of Window Length Estimation on Travel Time Estimation

This simulation is used to study how the length of the estimation window affects the performance of each estimation method. Five estimated window lengths ($b_2 - b_1$) are used in the simulation: $b_2 - b_1 = 90$, 150, 300, 450, and 600 s. The other simulation parameters are kept the same as in the section on the effect of window length estimation on different travel time estimation methods. The three methods of travel time estimation are compared in Figures 5*a* and 5*b*, which correspond to the average of the estimation errors and the standard deviation of the estimation errors, respectively.

The plots in Figure 5 clearly indicate that for all estimation methods, the standard deviations of the estimation errors are significantly reduced when the window length estimation is ≥300 s. However, if too large an estimation window is selected, the dynamic behavior of the change in traffic condition will be lost. So, selection of an appropriate window length estimation is important. Figure 5 shows that under the given simulation conditions, 600 s is a relatively good estimated window length for all three estimation methods.

When the maximum cross-correlation method is compared with the proposed method and the estimated window length is <300 s, estimation performance is similar for both methods (Figure 5). However, as the estimated window length increases, the performance of the proposed method is better than that of the maximum cross-correlation method, because more upstream arrivals with different travel times are included in longer estimated window lengths. The maximum cross-correlation method, which relies on only a single maximum cross-correlation coefficient, cannot consider traveler variability and is more sensitive to noise influence in determining the maximum cross-correlation coefficient. The proposed method, which uses all significant cross-correlation coefficients for travel time estimation, can consider traveler variability with different travel times. Moreover, the use of the multiple significant cross-correlation coefficients can smooth the estimation error of the cross-correlation coefficients. That is why the proposed estimation method is better than the maximum cross-correlation method with longer estimated window lengths.

Case Study

A case study illustrates how the proposed method estimates travel time under real traffic conditions. The flow data used in this study

TABLE 1 Comparison of Estimation Errors

Value	Optimization Method	Single Cross-Correlation Method	Proposed Multi-Cross-Correlation Method
Mean of error	0.2464	0.4547	0.2427
Standard deviation of error	1.238	1.1759	0.4814

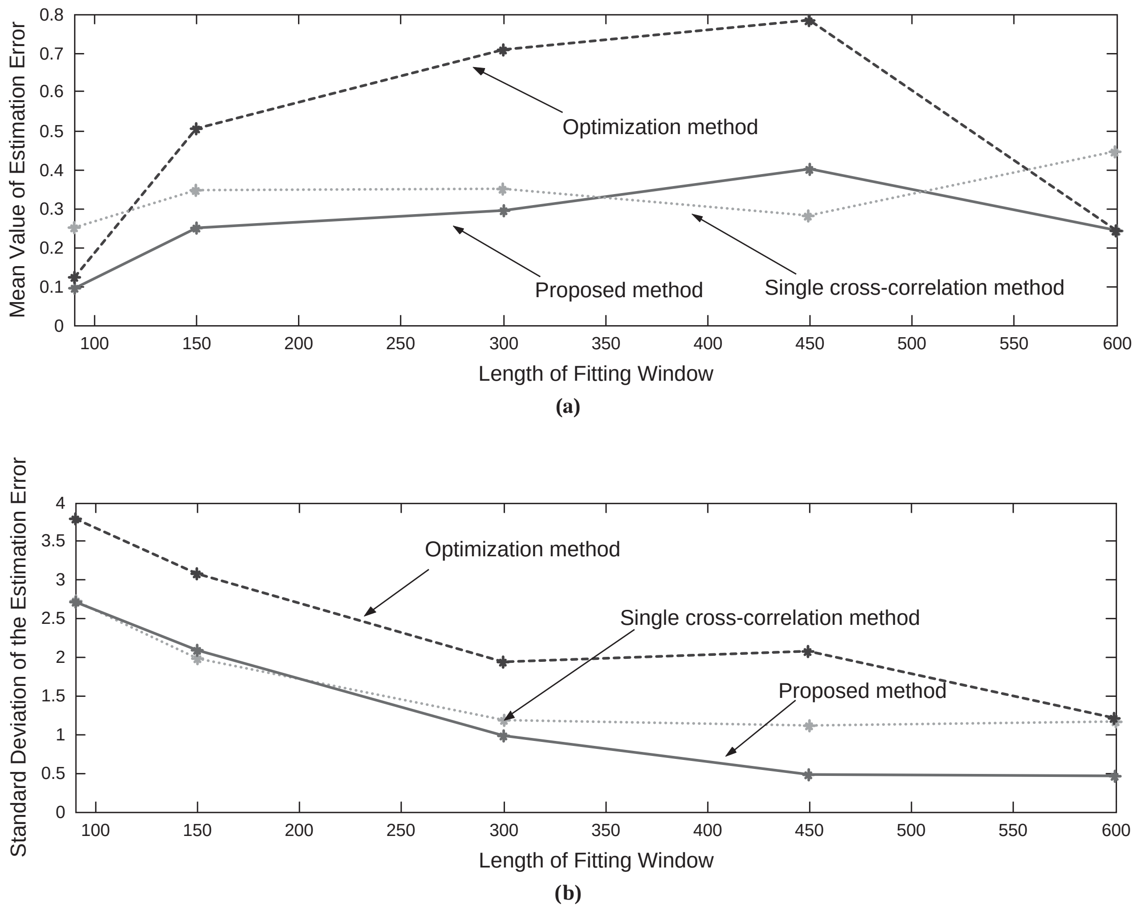

FIGURE 5 Comparison of estimation errors under different fitting window lengths.

were collected at the intersection of Speedway Blvd. and Tucson Blvd. in Tucson, Ariz., from 11:30 a.m. to 12:30 p.m. on December 13, 2002, by single loop detectors used to measure traffic flow in each direction.

The detector locations for the case study are illustrated in Figure 6. Speedway Blvd. has three lanes of traffic. Traffic data for the center lane from Tucson Blvd. to Country Club Blvd. were collected by Detector 2 and used to estimate travel time; the link distance is 2,689 ft (817 m), and the speed limit is 35 mph (15.6 m/s). Travel time estimation was updated every 10 min.

Figure 7 shows the cross-correlation analysis results in each estimation window, and travel time estimates are plotted in Figure 8. A reasonable trend of increased estimated travel time is observed approaching the noontime lunch break.

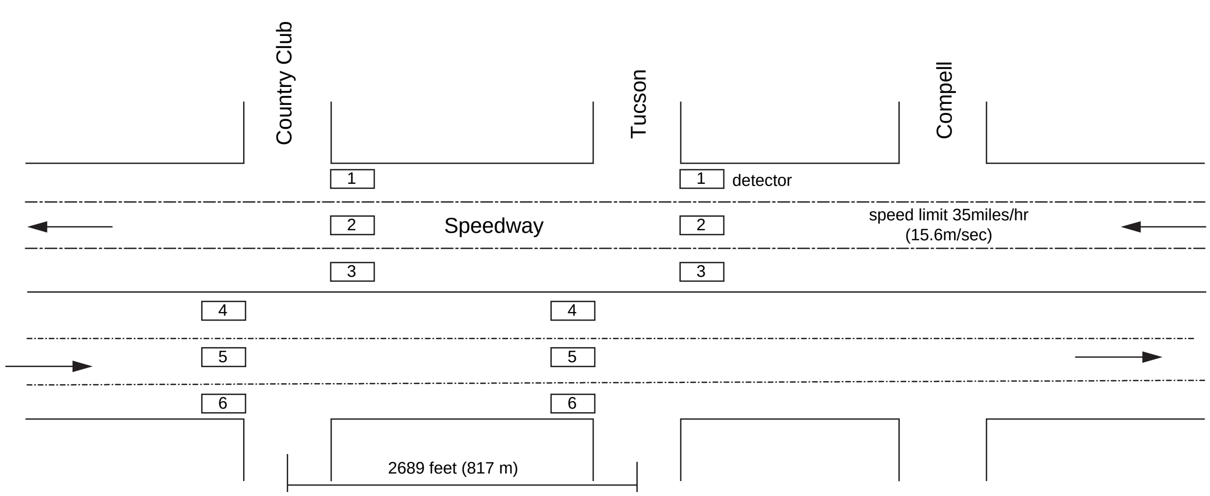

FIGURE 6 Single loop detector locations at Speedway Blvd. in Tucson, Ariz.

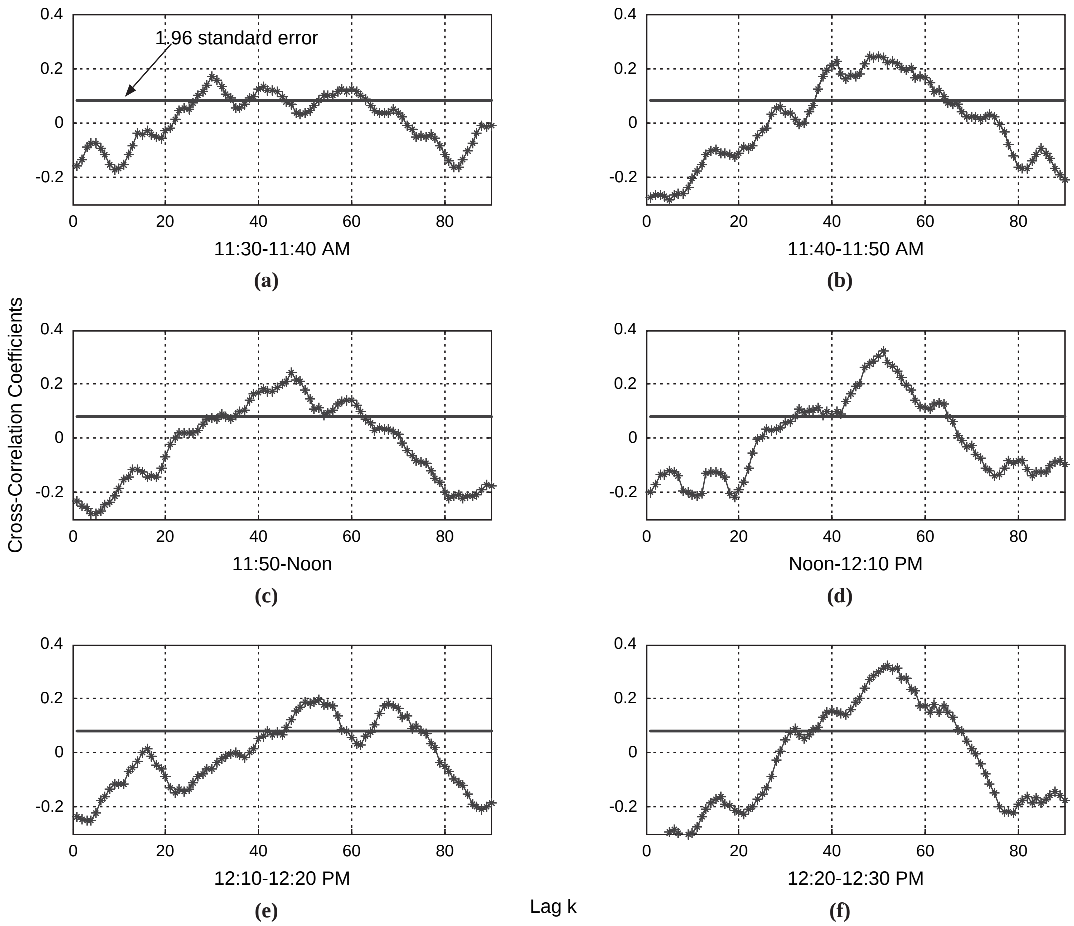

FIGURE 7 Cross-correlation analyses at different time periods.

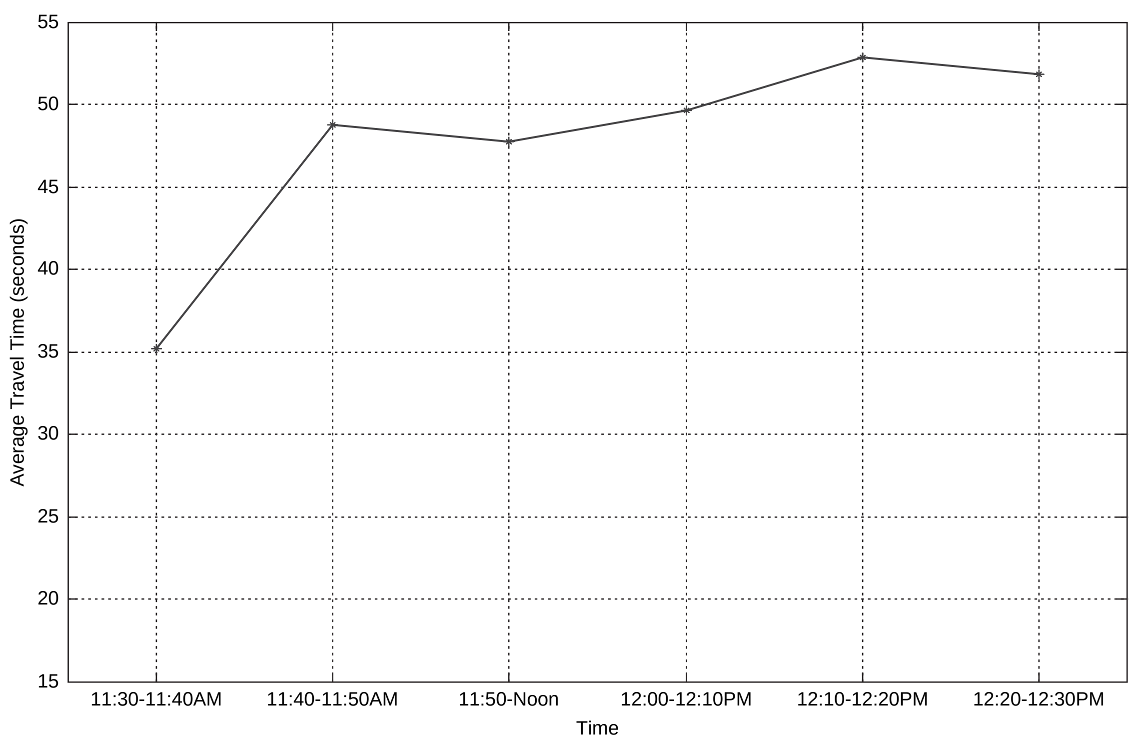

FIGURE 8 Estimated average travel time using the proposed method.

CONCLUSION

This paper proposed a new correlation analysis method for travel time estimation. With this method, only travel flow information measured by single loop detectors is required. Therefore, estimation errors caused in the measurement of occupancy and vehicle length are avoided. By using the *t*-test, the proposed method can automatically determine the center and width of the estimation window of the probability distribution function of average travel time.

Several comparison analyses between the proposed method and two existing methods were conducted as simulation studies, and a case study shows the effectiveness of the proposed method for the data analysis of real traffic.

ACKNOWLEDGMENTS

The authors thank the reviewers for insightful comments and suggestions that significantly improved the quality and readability of the paper. The authors gratefully acknowledge financial support from the U.S. Department of Transportation and Bureau of Transportation Statistics; the Federal Highway Administration and Arizona Department of Transportation; and the National Science Foundation.

REFERENCES

1. Athol, P. Interdependence of Certain Operational Characteristics Within a Moving Traffic Stream. In *Highway Research Record 72,* HRB, National Research Council, Washington, D.C., 1965, pp. 58–87.
2. Hall, F. L., and B. N. Persaud. Evaluation of Speed Estimates Made with Single-Detector Data from Freeway Traffic Management Systems. In *Transportation Research Record 1232,* TRB, National Research Council, Washington, D.C., 1989, pp. 9–16.
3. Jacobson, L. N., N. L. Nihan, and J. D. Bender. Detecting Erroneous Loop Detector Data in a Freeway Traffic Management System. In *Transportation Research Record 1287,* TRB, National Research Council, Washington, D.C., 1990, pp. 151–166.
4. Coifman, B. Estimating Travel Times and Vehicle Trajectories on Freeways Using Dual Loop Detectors. *Transportation Research, Part A,* Vol. 36, No. 4, 2002, pp. 351–364.
5. Coifman, B., and M. Cassidy. Vehicle Reidentification and Travel Time Measurement on Congested Freeway. *Transportation Research, Part A,* Vol. 36, No. 10, 2002, pp. 899–917.
6. Nam, D. H., and D. R. Drew. Traffic Dynamics: Method for Estimating Freeway Travel Times in Real Time from Flow Measurements. *Journal of Transportation Engineering,* Vol. 122, No. 3, 1996, pp. 185–191.
7. Nam, D. H., and D. R. Drew. Automatic Measurement of Traffic Variables for Intelligent Transportation System Applications. *Transportation Research, Part B,* Vol. 33, No. 6, 1999, pp. 437–457.
8. Park, D., L. R. Rilett, and G. Han. Spectral Basis Neural Networks for Real-Time Travel Time Forecasting. *Journal of Transportation Engineering,* Vol. 125, No. 6, 1999, pp. 515–523.
9. Hoogendoorn, S. P., J. W. C. van Lint, and H. J. van Zuylen. Freeway Travel Time Prediction with State-Space Neural Networks: Modeling State-Space Dynamics with Recurrent Neural Networks. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1811,* Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 30–39.
10. Dailey, D. J. Travel Time Estimation Using Cross-Correlation Techniques. *Transportation Research, Part B,* Vol. 27, No. 2, 1993, pp. 97–107.
11. Dailey, D. J. A Statistical Algorithm for Estimating Speed from Single Loop Volume and Occupancy Measurement. *Transportation Research, Part B,* Vol. 33, No. 5, 1999, pp. 313–322.
12. Petty, K., P. Bickel, M. Ostland, J. Rice, F. Schoenberg, J. Jiang, and Y. Ritov. Accurate Estimation of Travel Times from Single-Loop Detectors. *Transportation Research, Part A,* Vol. 32, No. 1, 1998, pp. 1–17.
13. Shumway, R. H., and D. S. Stoffer. *Time Series Analysis and Its Applications.* Springer-Verlag, New York, 1999.

The Artificial Intelligence and Advanced Computing Applications Committee sponsored publication of this paper.