Application of Finite Mixture of Regression Model with Varying Mixing Probabilities to Estimation of Urban Arterial Travel Times

Peng Chen, Kai Yin, and Jian Sun

Travel time along an urban arterial is greatly affected by traffic signals. Most studies on urban travel time use statistical models to obtain the distribution directly without incorporating the effects of traffic signal timing. In this study, a finite mixture of regression model with varying mixing probabilities (weights) was proposed to gain a better understanding of urban travel time distribution through consideration of signal timing. Standard finite mixture models with constant mixing probabilities have a limited ability to adapt to underlying random structural changes for observed travel times. The model developed in this study can capture such dynamics by (*a*) modeling the mixing probabilities as a function of the explanatory variables associated with signal timing and (*b*) establishing a linear regression between the mean of each component and signal timing. The finite mixture of regression model was applied to the travel time data collected by the automatic vehicle identification system on one urban arterial with the Sydney coordinated adaptive traffic system (SCATS). The results demonstrate that the varying mixing probabilities can be used to classify the samples of travel time, and the mean values of components can capture the effects of signal timing. By comparing various types of mixture models, the proposed approach not only has a better statistical fitting performance but also provides useful information about travel time features.

Among the various measures used to evaluate the performance of urban arterials, travel time is probably one of the most important. Accurate and reliable travel time information has become a major concern for both travelers and traffic management agencies. However, because traffic flows on urban arterials can be interrupted, travel time estimation is more challenging on those roads than on freeways (*1–6*). How to accurately estimate arterial travel time is one of the most popular topics in transportation engineering.

Classical approaches for estimating travel times on arterials include relationships of speed versus volume-to-capacity ratio and procedures based on the *Highway Capacity Manual* (*7*). Average travel time is calculated as the sum of the running timing based on arterial link characteristics and the intersection delay by a deterministic point delay model. Such approaches are not suitable for real-time applications with variable traffic conditions.

Because of complex interactions between volatile traffic patterns and signal control strategies, as well as the interdependent relationships of vehicle queues at neighboring intersections under different levels of congestion, travel times on arterials often vary within a wide range and even have overdispersion characteristics. Analysis of travel time variability is as important as, if not more important than, the traditional analysis of average travel time. A quantitative evaluation of the spatial and temporal variations of travel time can facilitate better travel decision making and traffic management (*8*).

Most existing studies on travel time variability put significant effort into identifying the best statistical model for fitting travel times. Emam and Al-Deek compared the fit of single-mode distributions—for example, lognormal, gamma, Weibull, and exponential distributions—and suggested that lognormal be used to compute travel time reliability measures on freeway arterials (*9*). However, Guo et al. pointed out that a single-mode distribution may not sufficiently represent such variation, because mean travel times under free-flow conditions and under congested conditions can differ substantially (*10*). Instead, Guo et al. proposed a multistate model to fit a mixture of Gaussian distributions into travel time observations of one arterial. A direct connection between model parameters and the underlying travel time state was established, which resulted in better fitting. Meanwhile, it was suggested that the multistate of travel times was closely related to time of day and traffic conditions. In an extension of that work, Park et al. quantified the impact of traffic incidents on travel time distribution by using multistate (three states) models (*11*). It was demonstrated that incidents increased the travel time variability and made the state corresponding to the congested regime more dominant. Guo et al. further advanced the multistate model by using skewed component distributions, for example, the gamma and lognormal distributions, to accommodate nonsymmetrically distributed travel times, which were commonly observed in congested states (*12*).

By following a similar approach, Feng et al. used a mixture of normal distributions to estimate travel times for arterial routes with next generation simulation data. Four travel time states were defined for through vehicles (*2*). Then, a Bayesian approach was proposed for constructing and updating travel time distributions under various traffic conditions. Kazagli and Koutsopoulos modeled travel

P. Chen, Department of Transportation Science and Engineering, Beihang University, New Main Building H1101, 37 Xueyuan Road, Haidian District, Beijing 100191, China. K. Yin, Department of Civil Engineering, Dwight Look College of Engineering, Texas A&M University, College Station, TX 77843. J. Sun, Department of Traffic Engineering, Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, China. Corresponding author: K. Yin, yinkai1000@gmail.com.

Transportation Research Record: Journal of the Transportation Research Board, No. 2442, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 96–105.
DOI: 10.3141/2442-11

time distribution by using automatic vehicle identification (AVI) data (*4*). A finite mixture model of two lognormal distributions was proposed for processing noisy AVI data from urban routes. The model explicitly recognized that such AVI data were generated by two underlying populations: vehicles that travel through the network and those that may be delayed for various reasons. The results demonstrated that the model can give reliable estimates of the density of each subpopulation for the period of analysis.

Ji and Zhang developed a hierarchical Bayesian mixture travel time model to capture the interrupted nature of urban traffic flows (*5*). High-resolution bus probe data were used to estimate travel time on urban streets. Bimodal travel time distributions were found at the link level, with one mode corresponding to travels without delays and the other to travels with delays. Then the estimated distributions were used to analyze traffic operations and identify congestion. Barkley et al. presented a methodology for determining the optimal number of states for modeling travel time on freeways and linked results with nonrecurrent congestion (*13*).

Taken together, some recent applications of finite mixture models show high goodness of fit with the observed travel times. The multistate travel time model overcomes the limitations of the single-state model by allowing multiple travel time states simultaneously. In addition, the model parameters have been connected to the underlining traffic state. However, the standard finite mixture models with constant mixing probabilities have a low ability to adapt to changes in the underlying stochastic processes for the observed travel time. On urban arterials, the dynamic and stochastic vehicle behavior under signal control has a significant influence on travel time variability. Models that link the travel time states with signal timing allow agencies to predict travel time more accurately and to develop optimal strategies for improving travel time states (*14–16*). Therefore, this paper proposes a finite mixture of regression model with varying mixing probabilities, which incorporate the effects that have been ignored before, such as signal timing. The model will enable a better understanding of arterial travel time.

The remainder of the paper is organized as follows. The next section outlines the methodology with a brief introduction of finite mixture models and the adjustment of their parameters to incorporate the effect of signal timing. Then, the study site and data description are presented, and the results are discussed. Finally, the findings of the study are summarized, and issues for future research are provided.

METHODOLOGY

The finite mixture model exhibits a high degree of flexibility in that it allows a combination of representations for a heterogeneous population. It has been used in a wide array of applications. Conventionally, the mixed components are chosen as Gaussian distributions. Although such a model can deal successfully with problems involving stationary stochastic processes, it has limited ability in modeling nonstationary processes, which make up most of the cases in travel time estimation problems. A common approach that deals with nonstationary processes is to divide time into several intervals and then apply finite mixture models. This approach may involve a good understanding of the time when the underlying processes have structural changes, a difficult task in most applications. Motivated by recent progress in finite mixture regression models with varying parameters in statistical modeling (*17*), this study uses such a modeling technique to introduce a new aspect to traditional finite mixture models: allow the mixing probabilities to adapt to the changes of some quantity and relate the mean values of components to some explanatory variables.

Finite Mixture Models with Varying Parameters

The normal form of finite mixture (regression) models is as follows. Suppose that the sample $Y = (y_1, y_2, \ldots, y_n)$ is drawn independently from a random variable $\mathbf{Y}$ that follows a finite mixture distribution. For a variable y_i, the component k in the mixture distribution has a probability density function $f_k(y_i)$ with mixing probability (weight) π_k for $k = 1, 2, \ldots, K$, and the probability density function of y_i is given by

$$f(y_i) = \sum_{k=1}^{K} \pi_k f_k(y_i) \tag{1}$$

where $0 \le \pi_k \le 1$ is the mixing probability of component k satisfying $\sum_{k=1}^{K} \pi_k = 1$ for $k = 1, 2, \ldots, K$. In addition, $f_k(\cdot)$ is allowed to follow different probability distributions.

More generally, the probability function $f_k(y_i)$ for component k may depend on parameters θ_k and explanatory variables x_k, that is, $f_k(y_i) = f_k(y_i \mid \theta_k, x_k)$; θ_k and x_k are vectors. It is assumed that the variables x_k explain the changes to the parameters. Hence, the following linear regression can be applied:

$$\theta_k = \gamma_{0k} + \gamma_k^T \cdot x_k \tag{2}$$

where γ_{0k} and γ_k are constants and T is the transpose of the vector. Equation 2 does not explicitly associate θ_k with the index i since θ_k may depend on an embedded structure that generates the sample Y.

Furthermore, the finite mixture model can be extended by assuming that the mixing probabilities π_k, $k = 1, 2, \ldots, K$, for observation y_i, $i = 1, 2, \ldots, n$, are not fixed constants but dependent on explanatory variable $\mathbf{z}_0$. The mixing probabilities π_{ik} can be modeled with a multinomial logistic model,

$$\log\left(\frac{\pi_{ik}}{\pi_{i1}}\right) = \alpha_k^T z_{0i} \tag{3}$$

or, equivalently,

$$\pi_{ik} = \frac{\exp(\alpha_k^T z_{0i})}{\sum_{k=1}^{K} \exp(\alpha_k^T z_{0i})} \tag{4}$$

where α_k are parameters and $\alpha_1 = 0$, $k = 1, 2, \ldots, K$, $i = 1, 2, \ldots, n$. Therefore, Equation 1 has the following generalized form:

$$f(y_i) = \sum_{k=1}^{K} \pi_{ik} f_k(y_i \mid \theta_k, x_k) \tag{5}$$

where π_k and θ_k satisfy Equations 2 and 4.

This mixture regression model with varying mixing probabilities enables the capture of many specific properties of real data, such as multimodality, skewness, kurtosis, and unobserved heterogeneity (*17*). Particularly, if the explanatory variables $\mathbf{z}_0$ and x_k represent changes of an embedded time-varying structure for the underlying stochastic processes of Y, then Equation 5 can characterize the time-varying property for Y.

Parameter Estimation

The widely applied methods for estimating the parameters of mixture models include maximum likelihood estimation with the expectation–maximization (EM) algorithm and the Bayesian approach. The Bayesian approach applies Gibbs sampling techniques in Markov chain Monte Carlo procedures to find appropriate parameter values. But this approach is computationally demanding. The EM algorithm maximizes the complete data log likelihood function by treating the information about the component where a sample is drawn as missing (*18*). Then the algorithm alternates between the E step and the M step until convergence. In each iteration, the E step finds the conditional expectation of the complete log likelihood function given the parameter estimates obtained from the previous iteration, and then the M step maximizes the conditional expectation in the E step to get the current parameter estimates. When the mixing probabilities are modeled with a multinomial logistic model, the M step will involve additional procedures to fit the multinomial logistic model by the explanatory variables. Rigby and Stasinopoulos provide details of the procedures (*19*).

FIGURE 1 Study site: arterial area in Shanghai, China.

STUDY SITE AND DATA DESCRIPTION

A 767-m stretch of a major arterial, Changshou Road in Shanghai, China, was chosen as the study site for evaluating the proposed methodology. A map of the arterial area is shown in Figure 1. The area includes four intersections operated by the Sydney coordinated adaptive traffic system (SCATS; *20*). However, the intersection shown in gray in the figure was undergoing subway construction during the analysis period (August 25 to 31, 2008), and crossing traffic was not allowed to enter the arterial. The study section was simplified to a route with two links between three intersections, as illustrated in Figure 1. The lengths of the links are 194 m and 573 m, respectively.

The SCATS detectors installed at the stop lines of each approach provided traffic flow data and signal timing data on a cycle-by-cycle basis. Figure 2 illustrates the through traffic counts by SCATS detectors at Intersections 1 and 3 in 5-min intervals (August 25, 2008). The traffic flows at the two intersections were almost the same and had similar time-varying characteristics. In addition, afternoon peaks were not clear compared with morning peaks. Commercial activities extend the peak periods from morning to late evening. Figure 2 shows that only after 9 p.m. does traffic demand start decreasing.

Under the SCATS control system, a three-phase signal timing plan was applied at all three intersections; the cycle lengths and green splits in the coordination plan were time varying and dependent on the traffic demand. Because of a wide range of traffic conditions throughout the day—for example, low-volume off-peak conditions, peak conditions, and postpeak medium-flow conditions—the system cycle lengths ranged from 80 to more than 200 s over the course of a day. Figure 3*a* and Figure 3*b* show the cycle length

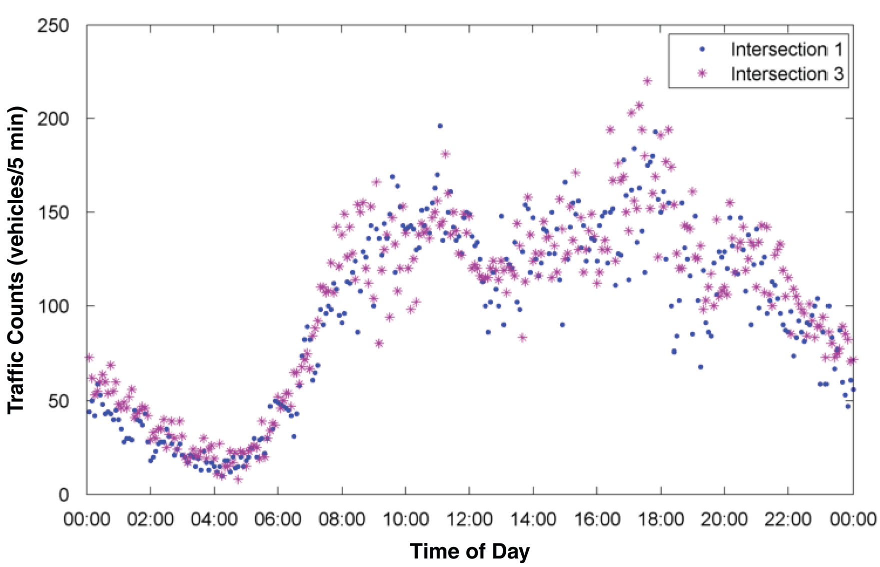


FIGURE 2 Through traffic counts made by SCATS detectors on August 25, 2008.

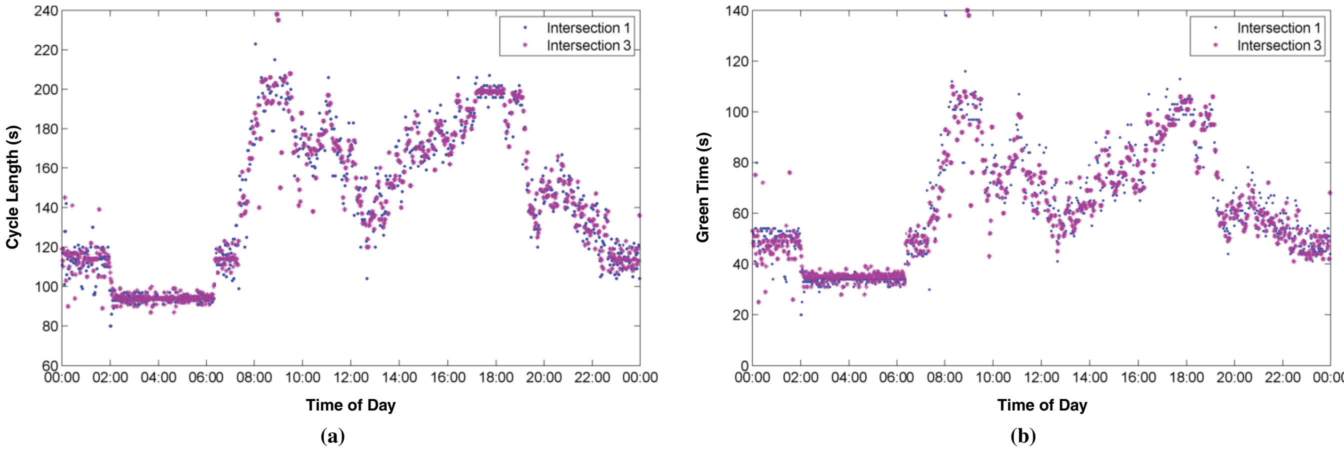


FIGURE 3 Intersections 1 and 3 on August 25, 2008: (*a*) cycle length and (*b*) through green time.

and through green time, respectively, at Intersections 1 and 3 on August 25, 2008. The two plots followed similar patterns for time-varying traffic volume.

AVI cameras installed at the downstream ends of Intersections 1 and 3 registered the license plates of passing vehicles and the corresponding time stamps. Actual travel times through the study route can be obtained by matching the license plates of the vehicles registered by both upstream and downstream cameras and comparing their time stamps. Figure 4 illustrates the traffic counts by SCATS detectors and the corresponding AVI observations at Intersections 1 and 3 in 1-h intervals. Even with less influence from secondary crossing traffic, it cannot be guaranteed that all passing vehicles will be recorded and then successfully matched, because AVI cameras are sensitive to overtaking vehicles with higher speeds and shorter time gaps. Across all travel times collected from midnight on August 25 to midnight on August 31, the average capture rate of the AVI cameras was around 66%, and the average matching rate (percentage of vehicles captured by both upstream and downstream AVI cameras) was about 24%. These ground truth travel times are used to validate the performance of the proposed methodology for multistate travel time distribution estimation.

EMPIRICAL ANALYSIS OF ROUTE TRAVEL TIME

This section illustrates an initial analysis of travel time data to gain insight into the observations. The previous section showed that the traffic has a peak during the morning and the afternoon on the studied arterial. Between these two peaks (Figures 2 and 4), traffic decreases at around noon and after 9 p.m. For this study, travel time data from

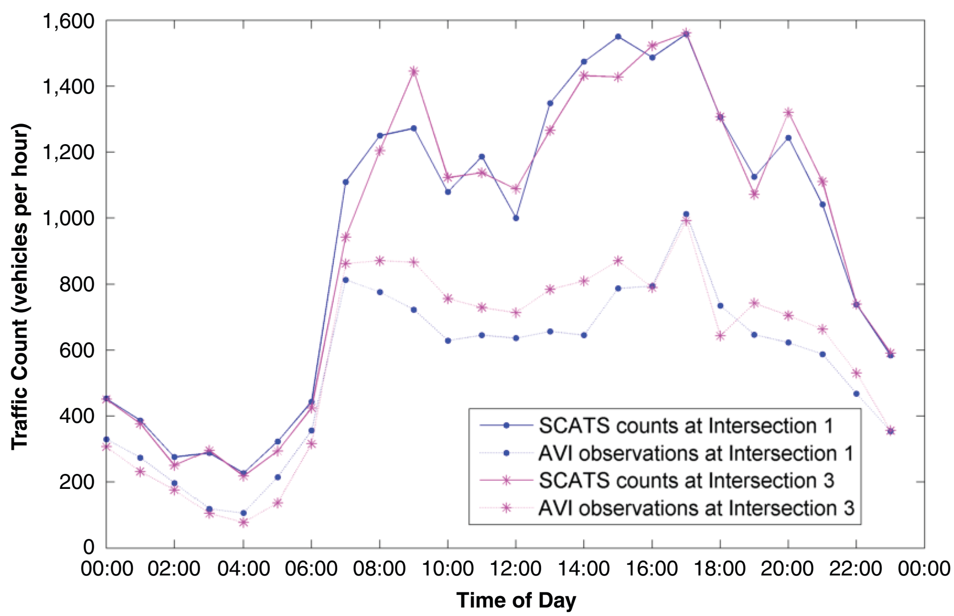


FIGURE 4 Traffic counts by SCATS detectors and corresponding AVI observations on August 25, 2008.

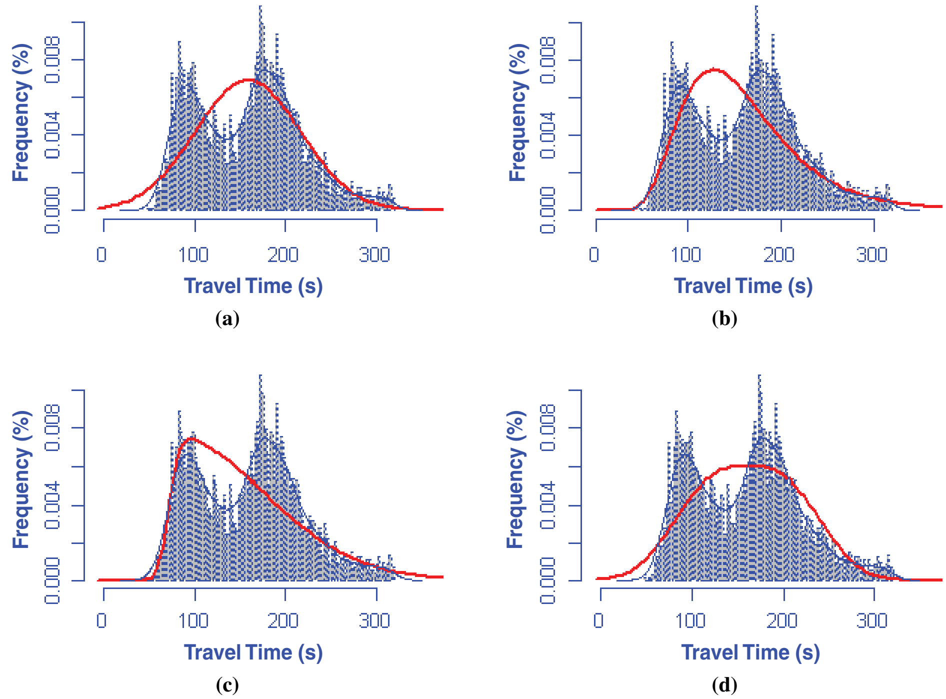


FIGURE 5 Single distribution density (red) versus frequency of travel time (smooth blue curves, obtained by nonparametric methods): (*a*) normal, (*b*) lognormal, (*c*) skew-*t* Type 1, and (*d*) generalized *t*.

noon to 10 p.m. during the week of August 25 to August 31, 2008, were used.

First, the frequency of a certain travel time and the empirical probability distribution were examined. The histogram in Figure 5 shows the normalized frequency of the travel time data. For illustration, a nonparametric smoothing method was used to obtain the empirical probability density (*21*), which is shown as the blue curve in Figure 5. The curve shows only the general trend of the travel time frequency and cannot be used as a probability density function. Furthermore, four nonmixture distributions—normal, lognormal, skew-*t* Type 1 (*22*), and generalized *t*—were calibrated to fit the travel time distribution to show the skewness and kurtosis of the entire distribution. The fitting results are shown by the red curves in Figure 5.

The figure shows that the travel time distribution has two main peaks, at around 100 and 200 s, and some small peaks at the right side. The difference between the two main peaks in terms of travel time is comparable to the red signal time. Neither nonmixture distribution can capture the general trend of the travel time distribution. However, the fitted four single distributions appear to exhibit some degree of similarity to one of the peaks. For example, the left trend of lognormal distribution appears similar to the first peak of travel time distribution, and the normal distribution appears similar to the second peak. In addition, the skew-*t* Type 1 and generalized *t* distributions may capture some local trends for the outliers. These facts indicate that if an attempt to establish finite mixture models may take some types of the distributions shown in Figure 5 into consideration as components.

A better understanding of the relationship between the green time for through traffic and the volume per green phase for the SCATS control system is also necessary (*23*). Figure 3 shows that the cycle length and the green time vary according to the time and show two peaks, during the morning and the afternoon. The two peaks clearly correspond to the traffic peaks during the day. A close correlation may be expected between them. If this is the case, both phase time and traffic count per cycle cannot be used simultaneously as the explanatory variables in modeling. Figure 6 shows green time versus volume per green phase between August 25 and 27. The regression analysis shows the corresponding R^2 value as .6529, indicating that there is a positive linear correlation. The red line in Figure 6 represents the results from linear regression.

Moreover, an attempt to apply finite mixture models to the travel time data requires determining the number of components. For the finite Gaussian mixture models with constant parameters and constant mixing probabilities, the number of components is varied and the Akaike information criterion (AIC) values are checked. AIC serves as a relative goodness-of-fit measure for the statistical models and provides a means for model selection. When the standard finite Gaussian mixture models are fitted, the resultant AIC values for the number of components ranging from two to seven are as follows:

- Two components, AIC = 39,221.49;
- Three components, AIC = 39,111.39;
- Four components, AIC = 39,096.34;
- Five components, AIC = 39,071.91;
- Six components, AIC = 39,108.52; and
- Seven components, AIC = 39,084.30.

When there are five components, the AIC value is its lowest. However, this result does not indicate that five is the best number for further modeling purposes. In the next part of this paper, the types of distributions for some components are changed and the mixing

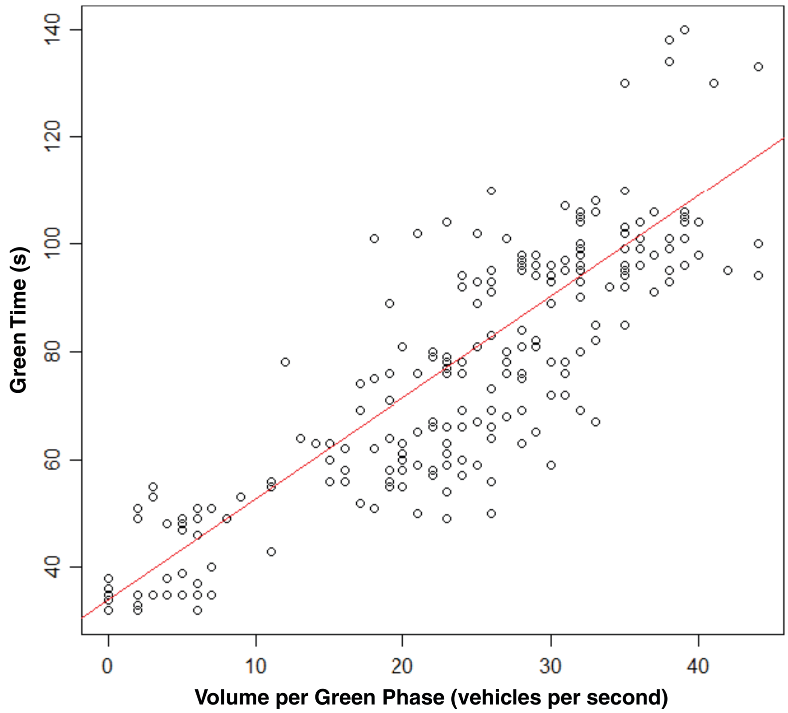


FIGURE 6 Positive correlation (R^2 = .6529) between each green time and through traffic volume in that phase.

probabilities varied. The increased complexity gives more flexibility to the model and increases the accuracy of prediction. Because the four-component model already has an AIC value close to that of the five-component model, the four-component model is expected to significantly decrease the AIC value with varying mixing probabilities. $K = 4$ was chosen in this study to achieve a good trade-off between the complexity of the models and the accuracy of prediction.

RESULTS

This section describes the modeling results of various finite mixture models for the travel time data. Then the best one, that is, the finite mixture regression model with varying mixing probabilities, was chosen for further analysis. In this study, the models were estimated with the GAMLSS package in R (*19*).

Various Finite Mixture Models

Several mixture models are compared and the best one for travel time data is obtained. In addition to the standard Gaussian distribution, lognormal distribution is used as the component for comparison. Other distributions, such as skew-*t* and generalized *t* distributions in Figure 5, are tested but they are heavy in computation effort and they do not have valuable analytical properties.

For modeling the mixing probabilities, it is assumed that the observed vehicles that pass the intersection may arrive during the current cycle or the red time in the previous cycle at the last signal. Because the timing of the second signal has a strong correlation with that of the third, only the timing of the third signal is used in the model. Moreover, it is possible for vehicles to arrive during the green time in the previous cycle and not depart until the current green phase. Here it is assumed that the arriving vehicles wait for at most two cycles. Because every cycle at the study site begins with through green, the variables that may affect the travel time for through vehicles are the red and green times during the previous cycle and the green time during the current cycle. Then Equation 3 has the following form in this study:

$$\log\left(\frac{\pi_{ik}}{\pi_{i1}}\right) = \alpha_{0,k} + \alpha_{1,k} z_{1,i} + \alpha_{2,k} z_{2,i} + \alpha_{3,k} z_{3,i} \tag{6}$$

where

i = ith sample,
$z_{1,i}$ = red time during previous cycle (s),
$z_{2,i}$ = green time during the previous cycle (s), and
$z_{3,i}$ = green time of the current cycle when observed vehicle departs (s).

Five four-component mixture models are compared:

Model 1 (standard). All components are Gaussian distributions with constant parameters and mixing probabilities.

Model 2 (varying mixing probabilities only). All components are Gaussian distributions with constant parameters. However, Equation 6 is used to model mixing probabilities.

Model 3 (mixture of Gaussian and lognormal). Three components are Gaussian and one is lognormal distribution. Equation 6 is applied as well.

Model 4 (mixture of Gaussian and lognormal). Two components are Gaussian and two are lognormal distribution. Other settings are the same as in Model 3.

Model 5 (mixture of regression model with varying mixing probabilities). All components are Gaussian distribution and Equation 6 is applied. Moreover, the mean in each Gaussian component is modeled as a regression given the explanatory variable, that is, the red time in the previous cycle. Specifically, the model has the following form:

$$f(y_i) = \sum_{k=1}^{4} \pi_{ik} f_k(y_i | \mu_k, \sigma_k, z_{1,i}) \tag{7}$$

where $f_k(\cdot | \mu_k, \sigma_k, z)$ is the form of Gaussian distribution N (μ_k, σ_k) and

$$\mu_k = \beta_{0k} + \beta_{1k} \cdot z_{1,i} \tag{8}$$

where β_{0k}, β_{1k} are constants.

Equation 8 can be regarded as the linear regression between the explanatory variable $z_{1,i}$ and the mean value of each component. The regression is motivated by the discussions in the previous section and the observations in Figure 5, where the travel time peaks have a difference approximately equal to the red time during the cycle. Because the red and cycle times fluctuate according to the traffic volume, Equation 8 reflects the dynamics of the changes in the mean of each component for the mixture regression model (Equation 7).

Goodness-of-Fit and Modeling Results

Table 1 presents the goodness-of-fit results of the five models in the previous section. The major concern is how to select the best model for the observed travel time. AIC values are applied in the comparison of models. The model with the smallest AIC value is of relatively good quality. In terms of both AIC and deviance values, Model 5 is preferred to the other models. This result appears to show that if one of the components changes to the lognormal distribution, the performance of models can be improved. However, if two of the components are lognormal and the other two are Gaussian, that is, Model 4, the performance is even worse than that of Model 3. Because the EM algorithm is used for parameter estimation, the mixture of lognormal and Gaussian distributions may lead to unstable results. That is, estimation procedures must be repeated many times to obtain the best estimates (in terms of AIC value) for the parameters of the model. If all the components are chosen as Gaussian, the computation results will be stable. Model 5 is significantly better than Model 2, which implies that the regression equation, Equation 8, does improve the performance of models. Also, Models 2 through 5 are all better than the standard Gaussian mixture model (Model 1).

On the basis of the preceding discussions, Model 5, the finite mixture of regression model with varying mixing probabilities, was chosen for further travel time study. For simplicity of presentation, the index of the component for the parameters in row 1 is dropped. The parameter estimates for Model 5 are shown in Tables 2 and 3. Parameters $\alpha_{0,1}$, $\alpha_{1,1}$, $\alpha_{2,1}$, and $\alpha_{3,1}$ must be 0 because π_1 is chosen as the reference in multinomial logistic models for mixing probabilities. The results show that the mean values of Components 1 and 4 have a negative correlation with the red time and the values of the other two components have a positive correlation. Moreover, as the β_1 values for Components 2 and 3 are close, it is anticipated that the mean values of these two components will be similar across time.

TABLE 2 Results of Component Parameter Estimates for Model 5

Parameter	β_0	β_1	σ
Component 1 estimate	105.33	−0.1787	15.29
Component 2 estimate	115.44	0.8597	19.63
Component 3 estimate	145.4	1.0	47.56
Component 4 estimate	201.95	−0.665	27.25

Results for Regression Mixture Model: Meaning of Classification

This section presents the results from which the meaning of mixed components is analyzed from the classification point of view. After the results for Model 5 are obtained, the mixing probabilities π_{ik} for each sample i can be computed according to Equation 6. The maximum value of π_{ik} for each i is selected, and this sample is considered as mainly drawn from the component k. In other words, the mixing probabilities serve as a measure for classifying the samples. According to this understanding, for the travel time data set, 30% of samples are from Component 1, 28% from Component 2, 34% from Component 3, and 8% from Component 4.

Figure 7 presents the classification results for 3 days (August 25, 26, and 27, 2008) of travel time data. In each subfigure the y-axis indicates values of travel time in seconds and the x-axis indicates the previous cycle length for each departure vehicle. Not all results are presented, as the explanations for other results are similar to what is presented here. The data points belonging to Component 4 are considerably less than those of other components, so Component 4 is not presented to avoid an inconvenient illustration. Equation 6 used in Model 5 involves information about previous cycle length for each sample. Hence, the combination of travel time and previous cycle length can differentiate the samples that belong to

TABLE 1 Goodness-of-Fit Statistics for Mixture Models

Model	Degrees of Freedom	AIC	Deviance
5	24	38,978.59	38,929.6
3	20	38,978.86	38,937.8
4	20	38,984.96	38,945.9
2	20	38,999.5	38,959.5
1	11	39,096.3	39,074.3

TABLE 3 Results of Mixing Probability Estimates for Model 5

Parameter	α_0	α_1	α_2	α_3
π_1 estimate	0	0	0	0
π_2 estimate	−1.6916	0.0118	−0.0172	0.0256
π_3 estimate	3.7136	−0.0891	0.0161	0.0213
π_4 estimate	−5.3679	0.0665	0.0088	0.0021

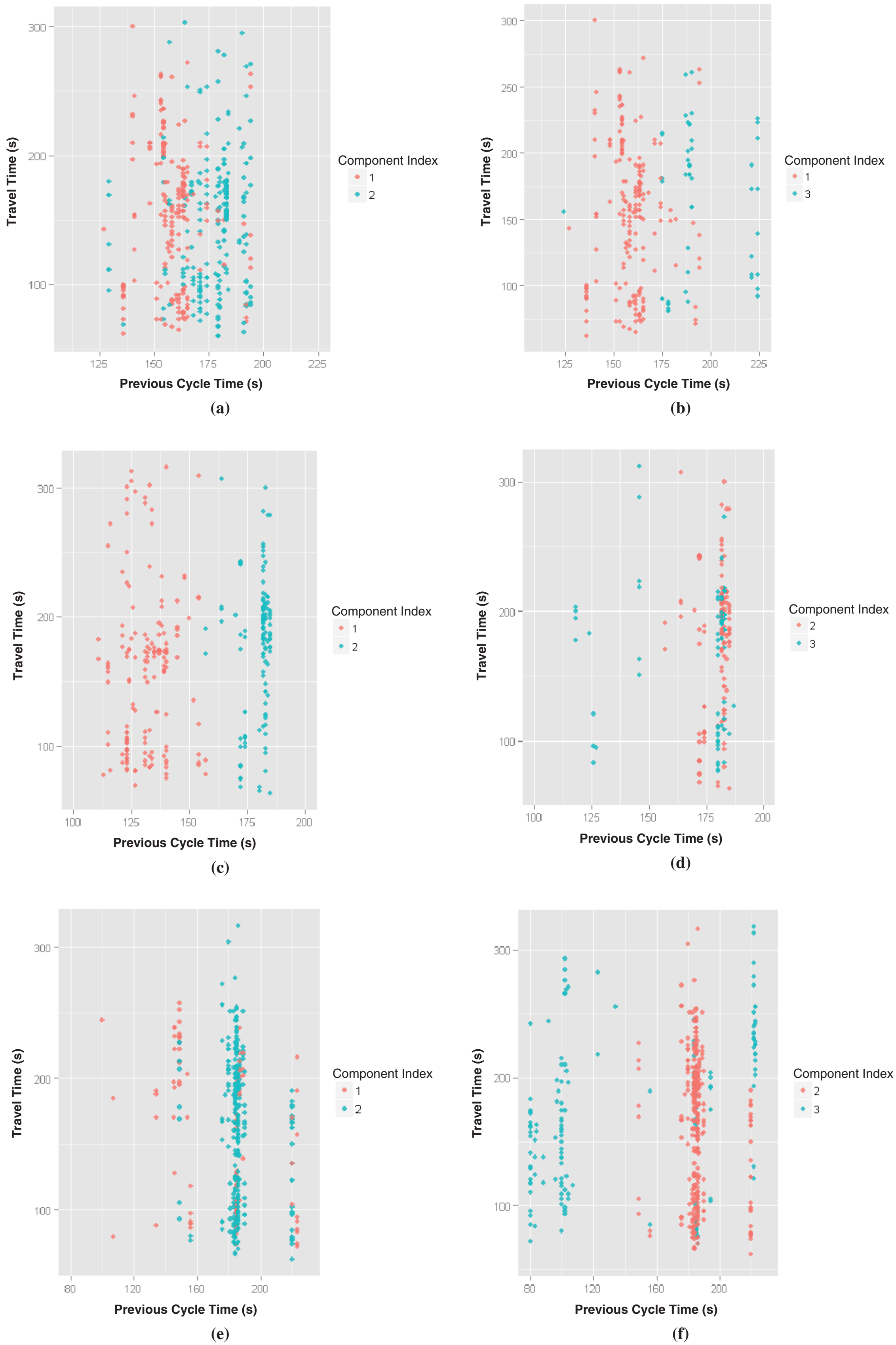


FIGURE 7 Sample classification by component: (*a* and *b*) August 25 data set, (*c* and *d*) August 26 data set, and (*e* and *f*) August 27 data set.

the various components in Model 5. Different components in the model have captured different features. In Figure 7, *a* and *b,* for travel time data on August 25, Components 2 and 3 characterize the samples with long previous cycle length, and Component 1 features the samples with low cycle length and moderate travel time. Figure 7*c* and Figure 7*e,* for the August 26 and 27 data, respectively, show that Components 1 and 2 differentiate travel time data according to cycle length. The travel times of the samples belonging to Component 2 tend to concentrate around 200 s, and the samples belonging to Component 1 are associated with low and medium travel times. From these results, it may be concluded that the choice of explanatory variables in Equation 6 is appropriate.

Mean of Each Component Across Time

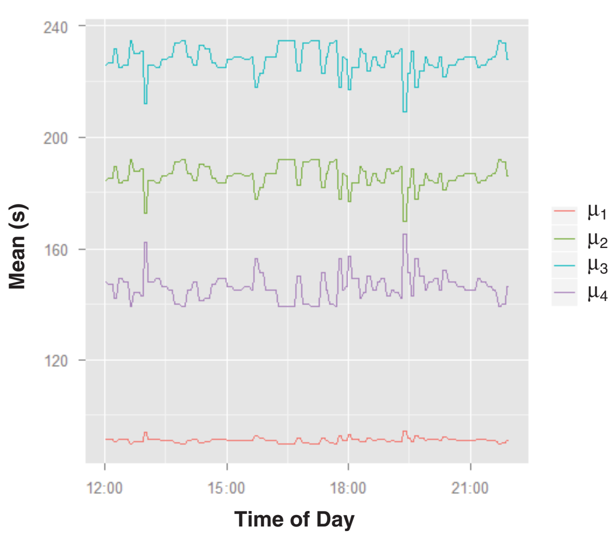
Equation 8 tracks the mean of each Gaussian component in Model 5 across time of day. Because the red time of the SCATS signals dynamically changes according to traffic conditions, Equation 8 implies that the mean values of components in mixture models also respond to traffic conditions. Moreover, the mean values can reflect the average of each classification group discussed in the previous section. Figure 8*a* and Figure 8*b* show the changes of mean for each component across time on August 25 and August 26, respectively. According to the results in Table 2, the mean values for Components 2 and 3 share a similar trend, and the trend for the mean for Component 4 is the opposite. It is clear from the figure that Component 1 always represents the lowest value of travel time, whereas Component 3 represents the highest value. The large fluctuations around 7 p.m. reflect traffic conditions during peak hours. The mean values of Components 2 and 3 exhibit a decreasing trend during peak hours. This does not indicate that the travel time actually decreases, since the mean of Component 4 dramatically increases during the same time. The mean travel time during peak hours should be computed by combining all components in the mixture model. In addition, the SCATS signal systems produce green times and cycle lengths that are longer in peak hours than in other periods. Hence, the trends in Figure 8 reflect that the chance of experiencing an extremely long travel time would be lower and the increase in green time decreases the travel time for a segment of through vehicles.

SUMMARY AND DISCUSSION

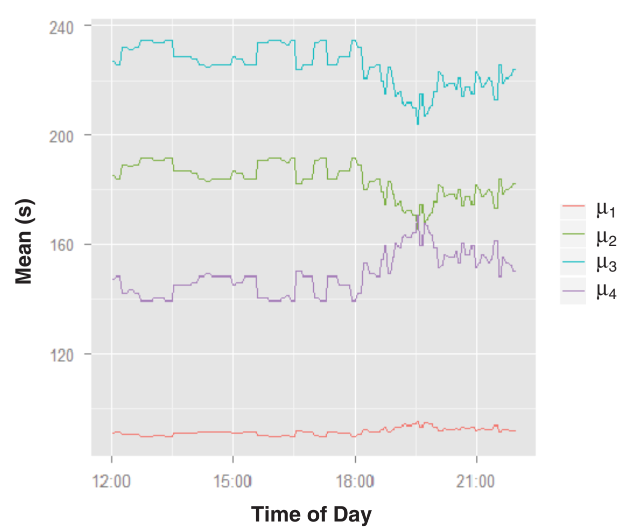
Some recent applications of finite mixture models show high goodness of fit to the multistate travel times on urban arterial rather than to single distribution. More significant, a direct connection between model parameters and the underlying travel time states has been achieved. However, a drawback of the standard finite mixture models is constant mixing probabilities. Little adaption can be made to dynamic changes in the observed travel time under traffic signal control. The study addressed this problem by developing a finite mixture of regression model with varying mixing probabilities (weights). Both mixing probabilities and mean values of components were modeled as a function of the explanatory variables associated with signal timing.

The finite mixture of regression models were then applied to the travel time data collected by the AVI system on one urban arterial with SCATS detectors. The results demonstrated that the varying mixing probabilities can be used to classify the samples of travel time, and the mean values of components can capture the dynamics of signal timing. In a comparison with various types of mixture models, the proposed approach not only has a better statistical fitting performance but also provides useful information about features of the travel time examined.

The results need further discussion. First, the performance of standard Gaussian mixture models can be improved either by replacing some components with other types of distribution that better fit the features of the data set or by introducing explanatory variables for modeling the mixing probabilities and the mean of each component. However, the trade-off among model complexity, computing efforts, and the accuracy of models must be considered. A model with increased complexity is likely to increase accuracy. However, whether a complex model is needed in practice should be determined to avoid the problem of overfitting. A mixture of Gaussian and other types of distribution may lead to unpleasant analytical properties



(a)



(b)

FIGURE 8 Average value of each component in Model 5 across time of day: (*a*) August 25 and (*b*) August 26.

when the likelihood functions are optimized. Hence, multiple runs must be performed and initial values of the parameters carefully chosen to overcome the local minimum problems for likelihood function optimization (*18*). Second, it is highly recommended that mixing probabilities and the mean in mixture models be modeled with appropriate explanatory variables when possible. If the explanatory variables represented critical features of the samples, the mixing probabilities then would provide proper results of classification. By examining the results of classification, one can also determine the appropriate explanatory variables and the best type of mixture model. Compared with popular artificial intelligent techniques like artificial neural networks (*24*), the finite mixture of regression model enjoys more convenient statistical properties. The varying mixing probabilities and the regression of the mean of each component can capture the dynamic trend of the studied quantity. Nevertheless, the results achieved need careful interpretation. By using advanced models like that proposed, traffic engineers can better design traffic signal timing plans for SCATS systems. Expanding the analysis with data covering more extensive sites and modeling mixing probabilities and the mean of each component in mixture models with alternative explanatory variables are promising future directions for this research. The developed methodology is also expected to be extended in future work to computing travel time distribution on large-scale road networks (*25*).

ACKNOWLEDGMENT

The authors thank Han Shu of Tongji University for providing advice and data analysis support.

REFERENCES

1. Feng, Y., G.A. Davis, and J. Hourdos. Arterial Travel Time Characterization and Real-Time Traffic Condition Identification Using GPS-Equipped Probe Vehicles. Presented at 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011.
2. Feng, Y., J. Hourdos, and G.A. Davis. Bayesian Model for Constructing Arterial Travel Time Distributions Using GPS Probe Vehicles. Presented at 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.
3. Susilawati S., M.A.P. Taylor, and S.V.C. Somenahalli. Distributions of Travel Time Variability on Urban Roads. *Journal of Advanced Transportation,* Vol. 47, No. 8, 2011, pp. 720–736.
4. Kazagli, E., and H.N. Koutsopoulos. Estimation of Arterial Travel Time from Automatic Number Plate Recognition Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2391,* Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 22–31.
5. Ji, Y., and H.M. Zhang. Travel Time Distributions on Urban Streets: Estimation with Hierarchical Bayesian Mixture Model and Application to Traffic Analysis Using High-Resolution Bus Probe Data. Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.
6. Ramezani, M., and N. Geroliminis. On the Estimation of Arterial Route Travel Time Distribution with Markov Chains. *Transportation Research Part B: Methodological,* Vol. 46, 2012, pp. 1576–1590.
7. *Highway Capacity Manual 2010.* Transportation Research Board of the National Academies, Washington, D.C., 2010.
8. *SHRP 2 Project L02: Establishing Monitoring Programs for Travel Time Reliability.* Final report. Transportation Research Board of the National Academies, Washington, D.C., 2013.
9. Emam, E.B., and H.M. Al-Deek. Using Real-Life Dual-Loop Detector Data to Develop New Methodology for Estimating Freeway Travel Time Reliability. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1959,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 140–150.
10. Guo, F., H. Rakha, and S. Park. Multistate Model for Travel Time Reliability. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2188,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 46–54.
11. Park, S., H. Rakha, and F. Guo. Multistate Travel Time Reliability Model: Impacts of Incidents on Travel Time Reliability. *Proc., 14th International IEEE Annual Conference on Intelligent Transportation Systems,* Washington, D.C., 2011.
12. Guo, F., Q. Li, and H. Rakha. Multistate Travel Time Reliability Models with Skewed Component Distributions. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2315,* Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 47–53.
13. Barkley, T., R. Hranac, and K. Petty. Relating Travel Time Reliability and Nonrecurrent Congestion with Multistate Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2278,* Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 13–20.
14. Liu, H., and W.T. Ma. A Virtual Vehicle Probe Model for Time-dependent Travel Time Estimation on Signalized Arterials. *Transportation Research Part C: Emerging Technologies,* Vol.17, No. 1, 2009, pp. 11–26.
15. Zheng, F., and H.J. van Zuylen. Uncertainty and Predictability of Urban Link Travel Time: Delay Distribution–Based Analysis. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2192,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 136–146.
16. Zheng, F., and H.J. van Zuylen. Modeling Variability of Urban Travel Times by Analyzing Delay Distribution for Multiple Signalized Intersections. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2259,* Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 80–95.
17. Rigby, R.A., and D.M. Stasinopoulos. Generalized Additive Models for Location, Scale and Shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* Vol. 54, No. 3, 2005, pp. 507–554.
18. Dempster, A.P., N.M. Laird, and D.B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B (Methodological),* Vol. 39, No. 1, 1977, pp. 1–38.
19. Rigby, R.A., and D.M. Stasinopoulos. A Flexible Regression Approach Using GAMLSS in R. 2009. http://www.gamlss.org/wp-content/uploads/2013/01/Lancaster-booklet.pdf.
20. *SCATS Traffic Reporter 6.4.2 User Manual.* Aldridge Traffic Controllers, Sydney, Australia. 2006.
21. Fan, J., and Q. Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer-Verlag, New York, 2002.
22. Zou, Y., and Y. Zhang. Use of Skew-Normal and Skew-*t* Distributions for Mixture Modeling of Freeway Speed Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2260,* Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 67–75.
23. Slavin, C., W. Feng, M. Figliozzi, and P. Koonce. Statistical Study of the Impact of Adaptive Traffic Signal Control on Traffic and Transit Performance. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2356,* Transportation Research Board of the National Academies,Washington, D.C., 2013, pp. 117–126.
24. Zeng, X., and Y. Zhang. Development of Recurrent Neural Network Considering Temporal-Spatial Input Dynamics for Freeway Travel Time Modeling. *Computer-Aided Civil and Infrastructure Engineering,* Vol. 28, 2013, pp. 359–371.
25. Hunter, T., A. Hofleitner, J. Reilly, W. Krichene, J. Thai, K. Anastasios, A. Pieter, and B. Alexandre. *Arriving on Time: Estimating Travel Time Distributions on Large-Scale Road Networks.* 2013. http://arXiv.org/abs/1302.6617.

The Urban Transportation Data and Information Systems Committee peer-reviewed this paper.