Mixture Models for Fitting Freeway Travel Time Distributions and Measuring Travel Time Reliability

Shu Yang and Yao-Jan Wu

Travel time reliability has attracted increasing attention in recent years and is often listed as a major roadway performance and service quality measure for traffic engineers and travelers. Measuring travel time reliability is the first step toward improving it, ensuring on-time arrivals, and reducing travel costs. Most measures of travel time reliability derive from continuous probability distributions and apply to traffic data directly. However, little previous research shows a consensus for selection of a probability distribution family for travel time reliability. Different probability distribution families could yield different values for the same measure of travel time reliability (e.g., standard deviation). The authors believe that specific selection of probability distribution families has few effects on measuring travel time reliability. Therefore, they proposed two hypotheses for accurately measuring travel time reliability and designed an experiment to prove the two hypotheses. The first hypothesis was proved by (*a*) conducting the Kolmogorov–Smirnov test and (*b*) checking log likelihoods and the convergences of the corrected Akaike information criterion and of the Bayesian information criterion. The second hypothesis was proved by examining both moment- and percentile-based measures of travel time reliability. The results from testing the two hypotheses suggest that (*a*) underfitting may cause disagreement in distribution selection, (*b*) travel time can be precisely fitted by using mixture models with a higher value of *K* (regardless of distribution family), and (*c*) measures of travel time reliability are insensitive to the selection of the distribution family. These findings allow researchers and practitioners to avoid testing of various distributions, and travel time reliability can be more accurately measured by using mixture models because of the higher values of log likelihoods.

Travel time reliability has attracted increasing attention in recent years and is often listed as a major roadway performance and service quality measure for traffic engineers and travelers. The tendency for easily remembering a few bad traffic conditions rather than normal ones shows the importance of studying travel time reliability (*1*). In addition, many researchers have noted the value of measuring travel time reliability. For example, Small et al. translated both mean travel time and the standard deviation of travel time into monetary values ($2.60 to $8.00/h and $10 to $15/h, respectively) to indicate that unreliable trips are more costly for travelers than reliable trips (*2*).

Department of Civil Engineering and Engineering Mechanics, University of Arizona, Room 324F, 1209 East 2nd Street, Tucson, AZ 85721. Corresponding author: Y.-J. Wu, yaojan@email.arizona.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2594, Transportation Research Board, Washington, D.C., 2016, pp. 95–106.
DOI: 10.3141/2594-13

Additional examples illustrating the value of travel time reliability can be found in Carrion and Levinson's study (*3*) and a report by Sadabadi et al. (*4*).

Measuring travel time reliability is the first step toward improving it, ensuring on-time arrivals, and reducing expenses. Polus conducted the first study about measuring travel time reliability for arterials (*5*). In the past decade, an increasing number of studies about measuring travel time reliability have been published (*6–18*). Most of these studies measured travel time reliability in two steps: (*a*) build travel time distributions on the basis of given travel time data and (*b*) calculate statistical indicators from the distributions (*18*). Because travel time is a continuous variable, most of these studies fitted travel time data by using continuous single probability distributions (e.g., Gaussian, lognormal, and gamma distributions). For example, Polus applied the gamma distribution to fit 211 travel time samples on an arterial (*5*), and the Weibull distribution was used in a study by Emam and Al-Deek (*9*). The researchers usually tested multiple families of continuous probability distributions and selected the one best fitting the data. For example, Emam and Al-Deek tested four probability distribution families (Weibull, exponential, lognormal, and Gaussian) before settling on the lognormal after applying goodness-of-fit tests (*9*). The most commonly used continuous single probability distribution was the lognormal [e.g., Guo et al. (*12*), Guo et al. (*15*), and Pu (*17*)]. Yang et al. provides a more comprehensive examination of continuous probability usages (*18*). On the basis of the variety of distributions identified in these studies, few have demonstrated any agreement on the selection of a continuous single probability distribution for travel time reliability.

The reliability measures are more generally defined. Most measures of travel time reliability can be derived directly from continuous probability distributions, and they can also be applied to the data directly. For example, the 90th- or 95th-percentile travel time is considered the "simplest method to measure travel time reliability" (*1*). The standard deviation and the coefficient of variance derived from probability distributions can also be used to measure travel time reliability (*11*). Several measures of travel time reliability are further derived from these basic measures. For instance, the "buffer index" (BI) is the additional time that "travelers must add to their average travel time when planning trips to ensure on-time arrival" (*1*). Van Lint et al. defined BI as TT90 – *M*/*M*, where TT90 is the 90th percentile travel time and *M* is the mean value for travel time (*11*). Van Lint et al. also derived new measures on the basis of distribution skewness and kurtosis. The "planning time index" (PI), similar to BI, represents "how much total time a traveler should allow to ensure on-time arrival" (*1*). One definition of the PI used in a study by Pu is TT95/free–flow travel time, where TT95 is the 95th percentile travel time

(*17*). Further information about existing measures of travel time reliability was summarized in several studies (*11, 17, 18*).

In general, two categories of measures of travel time reliability can be broadly identified, including moment-based measures (e.g., mean, standard deviation, skewness, kurtosis, and coefficient of variance) and percentile-based measures (e.g., the 90th or 95th percentile, BI, and PI). These measures of travel time reliability are derived from probability distributions. However, few studies show agreement on the selection of probability distribution families, a lack indicating that different single probability distribution families could yield different values for the same measure of travel time reliability.

Intuitively, the same travel time data set should tell a consistent story about travel time reliability for a particular statistical indicator (e.g., moments and percentiles). The statistical indicators derived from the same travel time data set should not vary even after different probability distribution families are applied. The main reason for testing different probability distribution families before statistical indicators are derived is to check whether travel times statistically follow the tested distributions. If travel time data and the tested distributions have no statistical relationship, then measures of travel time reliability may be inaccurate and biased. According to their intuition, the authors believe that the specific selection of continuous probability distribution families has few effects on measuring travel time reliability. Two hypotheses are proposed in hope of accurately measuring travel time reliability: (*a*) travel times can be fitted with multiple probability distributions independent of distribution family, and (*b*) measures of travel time reliability are insensitive to the selection of distribution family. These hypotheses suggest that travel times can be fitted by using any probability family (e.g., Gaussian or lognormal) and that the measures of same travel time reliability (e.g., standard deviations or coefficient of variances) derived from the selected probability distributions should be insensitive to the selection. If these two hypotheses are valid, the benefits for conducting studies of travel time reliability could include that (*a*) the data fitting process can be simplified and the statistical tests on probability distribution family selection may not be required and, thus, that (*b*) travel time reliability could be more accurately measured regardless of probability distribution families and data sets.

The rest of this paper is organized as follows: first, the issue of underfitting and overfitting is discussed in relation to distributions of travel time reliability. Next, an experiment is designed to prove the hypotheses. This experiment is applied to a study corridor along I-270 in Saint Louis, Missouri. Finally, the conclusion addresses the results from the experiment and evaluates the validity of the hypotheses.

UNDERFITTING AND OVERFITTING ISSUES

Underfitting and overfitting are two major issues when mathematical models are used to fit a given data set (*19, 20*). Models may fail to capture the underlying trend of the data set when underfitting occurs, while they may capture excess noise from the data set when overfitting occurs. Essentially, "underfitting leads to biased estimation and overfitting leads to increased variances" (*20*). Using a linear model, say $y = \sum_{i=0}^{N} w_i x^i = w_n x^n + w_{n-1} x^{n-1} + \cdots + w_1 x + w_0$, to fit a given data set with N samples is a simple way to illustrate the problem of underfitting and overfitting. If $N = 1$, the linear model becomes $y = w_1 x + w_0$; a line with slope w_1 and intercept w_0 is equivalent to the model, and an optimal set of w_1 and w_0 can be estimated to best fit the given data set. However, $y = w_1 x + w_0$ may be too simple to capture the underlying data trend for which the data set clearly is not linear. In this case, underfitting occurs. A linear model with a higher value of N would be a better choice. In theory, a model to the $N - 1$ power would perfectly fit the given data with N samples (*21*), but fitting to the $N - 1$ power may cause overfitting, with no discernible data trend. To avoid both issues, the selection of N can be based on the bias–variance trade-off principle (*20*).

Similarly, a data set (e.g., a travel time data set) could be fitted with mixture probability models, denoted as $p(x|\Sigma) = \sum_{k=1}^{K} w_i p_i(x|\Sigma_i) = w_1 p_1(x|\Sigma_1) + \cdots + w_K p_K(x|\Sigma_K)$, where $p(x|\Sigma)$ is a continuous probability distribution, w_i is the weight of the distribution $p_i(x|\Sigma_i)$, K is the number of distributions, and Σ_i is a set of parameters in the ith distribution. Mixture probability models are simplified to single probability distributions if and only if $K = 1$, indicating that single probability distributions are special cases of mixture probability models. When single probability distributions are used, underfitting can occur and the goodness of fit may be low. Failure to capture the underlying probability trend may lead to an inaccurate measurement of travel time reliability.

The proposed hypotheses suggest that, if a given travel time data set could be better fitted by using mixture probability models, then the associated information and statistical indicators will be in agreement, regardless of which form of $p_i(x|\Sigma_i)$ is chosen (e.g., Gaussian, lognormal, or gamma). This agreement will be especially notable when underfitting is corrected.

The following section describes both an experiment to test the two hypotheses and a specific field study that uses the experiments.

METHODOLOGY

Mixture Probability Models

The general expression for mixture models was presented earlier. To implement mixture probability models, the form of $p(x|\Sigma)$ needs to be specified. Single probability distributions can generally be categorized as symmetric and skewed. To test the two hypotheses thoroughly, the most commonly used symmetric distribution, Gaussian, and two skewed distributions, lognormal and gamma, were selected to specify $p(x|\Sigma)$ in the field study described next. The mathematical expressions of these three distributions are listed in Table 1. The expressions for the corresponding mixture models are shown in Equations 1 through 3:

$$\text{Gaussian}(T \mid \mu, \sigma) = \sum_{k=1}^{K} w_i \text{Gaussian}_i(t|\mu_i, \sigma_i) \tag{1}$$

$$\text{lognormal}(T \mid \mu, \sigma) = \sum_{k=1}^{K} w_i \text{lognormal}_i(t|\mu_i, \sigma_i) \tag{2}$$

$$\text{gamma}(T \mid b, \theta) = \sum_{k=1}^{K} w_i \text{gamma}_i(t|b_i, \theta_i) \tag{3}$$

where

T = travel time data set;
w_i = weight of ith distribution;
μ_i and σ_i = mean and standard deviation of the ith Gaussian or lognormal distribution, respectively; and
b_i and θ_i = shape and scale of the ith gamma distribution, respectively.

TABLE 1 Statistics of Three Distributions

Distribution Family	Gaussian	Lognormal	Gamma
Probability density function	Gaussian (x; μ, σ) $= \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$	Lognormal (x; μ, σ) $= \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{(\ln x-\mu)^2}{2\sigma^2}}$	Gamma (x; b, θ) $= \frac{1}{\Gamma(b)*\theta^b} x^{b-1} e^{\frac{-x}{\theta}}$
First moment ($\mu^{(1)}$)	μ	$e^{\mu+\sigma^2}$	$b\theta$
Second moment ($\mu^{(2)}$)	σ^2	$\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}$	$b\theta^2$
Third moment ($\mu^{(3)}$)	0	$\left(e^{\sigma^2}+2\right)\sqrt{e^{\sigma^2}-1}$	$\frac{2}{\sqrt{b}}$

Several parameter values must be estimated for each mixture model, including the weights (w_i), the parameters for individual distributions in mixture models (either μ_i and σ_i or b_i and θ_i), and the number of individual distributions (K). The expectation–maximization algorithm is the most popular approach for estimating the weights and parameter values in individual distributions when K is given. Rogers and Girolami showed extensive detail about parameter estimation in the Gaussian mixture model by using the expectation–maximization algorithm (*21*). Following the procedure to estimate their parameter derivative, the parameters in the lognormal mixture model and the gamma mixture model can also be estimated by using the expectation–maximization algorithm. Unlike arbitrarily selected initial values for the weights and parameters in individual distributions, the K-means method was used to initialize these weights and parameters so that the optimal weights and parameters could be efficiently estimated (*22*). The three mixture models were implemented by using the R language (*23*), and the source code is available to the public at the authors' website (*24*).

Measures of Travel Time Reliability

Moment-Based Measures of Travel Time Reliability

The most popular moment-based travel time reliability measures include (*a*) the three basic moments [the first moment $\mu^{(1)}$ (the mean travel time), the second moment $\mu^{(2)}$ (the variance of travel time), and the third moment $\mu^{(3)}$ (the skewness of travel time)] and (*b*) two standardized moments (the coefficient of variance and the standardized skewness). Equations 4 through 7 show the expressions for the three basic moments by using mixture models. Equation 4 shows that the mean value of mixture models is the summation of weighted mean values of individual distributions. The mathematical expression for the ith moment of mixture model represents $E[(T-\mu)^i]$, and the expression can be rewritten as Equation 5 on the basis of a binomial expansion (*25*). Equations 6 and 7 show, respectively, the second and third moments of mixture models.

$$\mu^{(1)} = E[T] = \sum_{k=1}^{K} w_k \mu_k^{(1)} \tag{4}$$

$$\mu^{(i)} = E\left[(T-\mu)^i\right] = \sum_{k=1}^{K}\sum_{i=0}^{I}\binom{I}{i}\left(\mu_k-\mu^{(1)}\right)^{I-i} w_k E\left[(T_k-\mu_k)^i\right] \tag{5}$$

$$\mu^{(2)} = E\left[(T-\mu)^2\right] = \sum_{k=1}^{K} w_k\left[\left(\mu_k^{(1)}-\mu^{(1)}\right)^2+\mu_k^{(2)}\right] \tag{6}$$

$$\mu^{(3)} = E\left[(T-\mu)^3\right] = \sum_{k=1}^{K} w_k\left[\left(\mu_k^{(1)}-\mu^{(1)}\right)^3+3\left(\mu_k^{(1)}-\mu^{(1)}\right)+\mu_k^{(3)}\right] \tag{7}$$

where

$\mu^{(i)}$ = ith moment about mean value and $i \in \{1, 2, \ldots, I\}$,
$\mu_k^{(i)}$ = ith moment about mean value for kth distribution in mixture models,
$E[*]$ = expectation, and
w_k = weight for kth distribution in mixture models.

Expressions for the different choices on $p(x|\Sigma)$ are listed in Table 1.

From the expressions for the three basic moments, the two standardized moments are derived in Equations 8 and 9. Equation 8 incorporates Equations 4 and 6, while Equation 9 incorporates Equations 6 and 7.

$$\text{coefficient of variance} = \frac{\sqrt{\mu^{(2)}}}{\mu^{(1)}} = \frac{\sqrt{\sum_{k=1}^{K} w_k\left[\left(\mu_k^{(1)}-\mu^{(1)}\right)^2+\mu_k^{(2)}\right]}}{\sum_{k=1}^{K} w_k\mu_k^{(1)}} \tag{8}$$

$$\text{standardized skewness} = E\left[\left(\frac{T-\mu}{\sigma}\right)^3\right] = \frac{\mu^{(3)}}{\left(\mu^{(2)}\right)^{1.5}} = \frac{\sum_{k=1}^{K} w_k\left[\begin{array}{c}\left(\mu_k^{(1)}-\mu^{(1)}\right)^3\\ +3\left(\mu_k^{(1)}-\mu^{(1)}\right)+\mu_k^{(3)}\end{array}\right]}{\left(\sum_{k=1}^{K} w_k\left[\left(\mu_k^{(1)}-\mu^{(1)}\right)^2+\mu_k^{(2)}\right]\right)^{1.5}} \tag{9}$$

Percentile-Based Measures of Travel Time Reliability

The ith percentile of an arbitrary probability distribution is located at the value t that represents the area under the probability distribution curve ranging from negative infinity to t for which the area equals i/100. Figure 1 gives an example of calculating percentiles in an arbitrary distribution. The gray area is the given i, and t is the corresponding ith percentile. Because mixture models are the summations of weighted individual distributions, the expressions for calculating percentiles of mixture models are difficult to derive. The fundamental method of calculating mixture model percentiles is to estimate the area by discretizing the gray area to small portions and summing these portions.

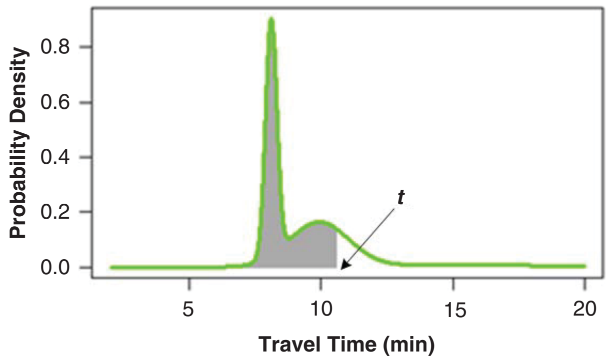


FIGURE 1 Example of calculating percentiles given arbitrary distribution.

Experimental Design

Figure 2 depicts an experimental design for testing the hypotheses. The three mixture models were estimated on the basis of a given travel time data set, and then the moment- and percentile-based measures of travel time reliability were calculated by using the estimated mixture models. The one-sample Kolmogorov–Smirnov (K-S) test was conducted to determine whether the given travel time data statistically follows the probability distributions generated from the mixture models. The log likelihoods, corrected Akaike information criterion (AIC), and Bayesian information criterion (BIC) of the mixture models were also calculated to determine convergence. If the one-sample K-S test suggests that the travel time data statistically follows the three mixture models and the log likelihoods of those mixture models converge, then the first hypothesis will be accepted; if the measures of travel time reliability estimated on the basis of mixture models remain unchanged, then the second hypothesis will be accepted.

The number of distributions (***K***) in the mixture models varied from one to six; this variation means that up to six individual distributions in the mixture models were used to fit the estimated travel times for the two cases. The first step was to apply each mixture model to fit the travel times with different K-values. Then, the one-sample K-S test was performed to examine whether the travel times statistically follow the three mixture models. The null hypothesis in the K-S test is defined as H_0: the travel times statistically follow the probability distributions generated from the three mixture models. The alternative hypothesis is H_1: the travel times do not follow the distributions generated from the mixture models. Two significance levels ($\alpha = 0.01$ and $\alpha = 0.05$) were selected as the thresholds for indicating whether the travel times follow those generated probability distributions. The resulting value of the one-sample K-S test is the maximum distance between an empirical cumulative probability function generated by the travel times and a cumulative probability function produced by one of the three mixture models.

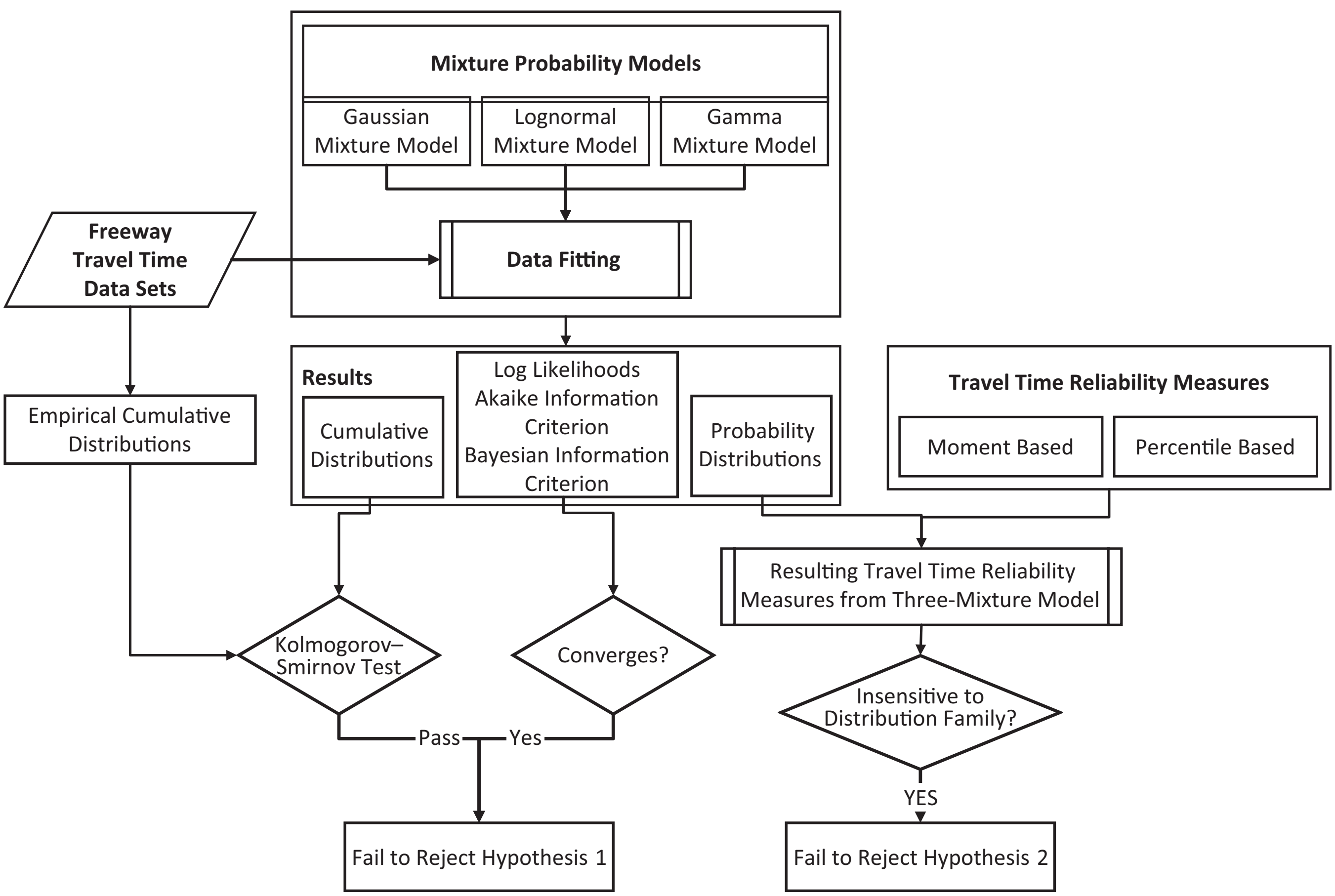


FIGURE 2 Work flowchart of testing hypotheses.

STUDY CORRIDOR AND DATA DESCRIPTION

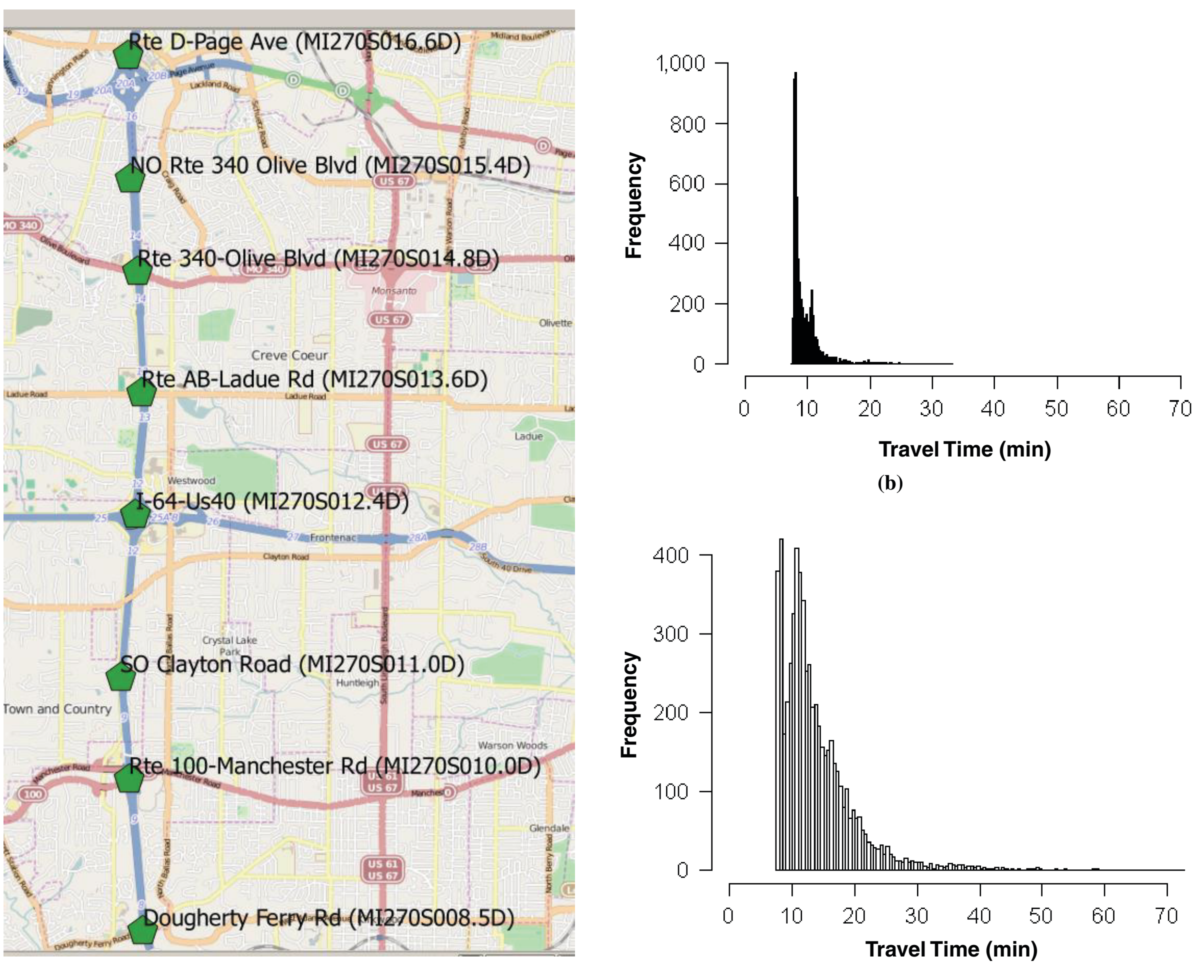
More than 700 intelligent transportation system sensors have been installed on major freeways and arterials in the Greater Saint Louis area. Traffic data collected from these sensors provide vehicle movement details every 30 s. An 8.1-mi segment of the I-270 southbound corridor (Figure 3*a*) was selected to be the study corridor for the experiment outlined earlier. Previous studies usually used a large amount of traffic data to measure travel time reliability by time of day and day of week [e.g., van Lint et al. (*11*) and Yang et al. (*18*)]. Similarly, this study included traffic data collected on the study corridor from October 2013 to March 2015 and grouped them by time of day and day of week to measure travel time reliability on the corridor. The time slice model was used to estimate travel time every 30 s, and the estimated travel times were aggregated every 5 min to avoid short-duration travel time fluctuation (*18*).

The study corridor usually suffers from recurring traffic congestion during the afternoon peak hours (4 to 7 p.m.) every weekday and experiences uncongested flow the rest of the day. To test the hypotheses, two cases (Case 1, 3 to 4 p.m. Tuesday, and Case 2, 6 to 7 p.m. Tuesday) were tested in this study. The two selected periods required additional attention because of the traffic complexity resulting from the congestion. Travel time histograms for the two cases are shown in Figure 3, *b* and *c*.

RESULTS

Test of First Hypothesis

The purpose of this section is to verify the first null hypothesis: travel times can be fitted with multiple probability distributions no matter which distribution family is chosen. Table 2 summarizes the maximum distances produced from the K-S test. Two critical values were determined from the two significance levels. If the distances were less than the critical values, then the null hypothesis could not be rejected; otherwise, the null hypothesis would be rejected; that is, the travel times did not statistically follow the distributions generated by the mixture models. In Table 2, a 1 indicates acceptance of the null hypothesis and a 0 indicates dismissal of the null hypothesis. The travel times did not statistically follow the single Gaussian distribution, the single lognormal distribution, or the single gamma distribution in both cases. This finding conflicted with many of the previous studies of travel time reliability. The null hypothesis in the



(a) (c)

FIGURE 3 Background information on study corridor on I-270 southbound: (*a*) map of study area (source of background image: https://www.openstreetmap.org/); (*b*) histogram of travel time, Case 1; and (*c*) histogram of travel time, Case 2.

TABLE 2 Testing Using One-Sample K-S Test

	Gaussian Mixture Model			Lognormal Mixture Model			Gamma Mixture Model		
K	Distance	$\alpha = 0.01$	$\alpha = 0.05$	Distance	$\alpha = 0.01$	$\alpha = 0.05$	Distance	$\alpha = 0.01$	$\alpha = 0.05$
Case 1									
1	0.21541	0	0	0.16713	0	0	0.16446	0	0
2	0.14159	0	0	0.07199	1	1	0.10861	1	1
3	0.02886	1	1	0.02672	1	1	0.07764	1	1
4	0.02936	1	1	0.01451	1	1	0.03403	1	1
5	0.01418	1	1	0.01368	1	1	0.06536	1	1
6	0.01385	1	1	0.00730	1	1	0.08614	1	1
Case 2									
1	0.14020	0	0	0.16556	0	0	0.18539	0	0
2	0.06273	1	1	0.04388	1	1	0.11997	1	1
3	0.04506	1	1	0.04487	1	1	0.05431	1	1
4	0.04761	1	1	0.00502	1	1	0.0827	1	1
5	0.00709	1	1	0.00526	1	1	0.05917	1	1
6	0.00661	1	1	0.00608	1	1	0.04492	1	1

NOTE: Critical values = 0.13309 [at significance level (α) 0.01] and 0.11104 [at significance level (α) 0.05].

K-S test was even rejected when the Gaussian mixture model with $K = 2$ was used in Case 1; but the travel times fit both the lognormal and gamma mixture models with $K = 2$. One major reason for the failure of the Gaussian mixture model with $K = 2$ could be related to the symmetric shape of the Gaussian distribution. The nonskewed Gaussian distribution failed to capture the long-tail shape of the travel times. However, the travel times can be accurately fitted by adding one more Gaussian distribution to capture the long-tail travel times ($K = 3$), and the null hypothesis could not be rejected. Because of their skewed shapes, the lognormal and gamma mixture models ($K = 2$, 3, 4, 5, and 6) could be used to fit the travel times and support the null hypothesis.

In addition to application of the K-S test, the log likelihoods of the probability distributions with variable K were also computed. Figure 4 clearly shows that the log likelihoods increased with an increase in K, a result indicating that the mixture models with a higher value of K can better fit the travel times. The log likelihoods converged when K was greater than four; that is, the addition of further individual distributions may not improve mixture model performance. The convergence curves in Figure 4 can help researchers to avoid both underfitting and overfitting and thereby to determine proper selection of K (*21*). In addition to the log likelihoods, the measures of statistical model quality AIC and BIC also appear in Figure 4. The values of log likelihoods, AIC, and BIC suggest that $K = 3$ was the appropriate level to mitigate both underfitting and overfitting issues in Case 1, while $K = 5$ was considered the appropriate level in Case 2. The low log likelihoods when $K = 1$ show that the model was underfitted by the travel times. No significant performance improvement can be found between the three mixture models with higher K-values. Two other interesting findings were these: (*a*) given the appropriate level of K, the Gaussian and the lognormal mixture models performed similarly and the differences in their log likelihood values were fairly small; and (*b*) both the Gaussian and lognormal mixture models performed slightly better than the gamma mixture model.

Figure 5 shows the generated probability distributions from the three mixture models when $K = 1$ and $K = 3$ for Case 1. The mixture models with $K = 3$ fit the travel times better than the three single distributions.

Overall, the test showed that the first hypothesis is valid. More specific findings from testing the first hypothesis are summarized below:

1. Single probability distributions are unsuitable for fitting the travel times in the two case studies because the travel times were underfitted. The potential reason is that travel times may be generated from different traffic flow states in the selected periods (*12, 15*).
2. Travel times can be precisely fitted by using the three mixture probability models with a proper number of individual distributions. Mixture models with symmetric probability distributions (e.g., the Gaussian distribution in this study) using a higher value of K can be used to fit travel times. In the two study scenarios, K had to be ≥ 3 to pass the one-sample K-S test. The other two mixture models with skewed shape distributions (i.e., the lognormal and gamma distributions) and $K \geq 2$ can be used to fit travel times.

Test of Second Hypothesis

This section examines the testing of the second hypothesis, that is, determining whether the measures of travel time reliability were insensitive to the form $p(t|\Sigma)$. More specific findings from testing the second hypothesis follow.

Moment-Based Measures of Travel Time Reliability

The first moment (mean travel time), the second moment (variance of travel time), the third moment (skewness of travel time), the coefficient of variance, and the standardized skewness were selected for measuring travel time reliability for the two cases. Figure 6 depicts

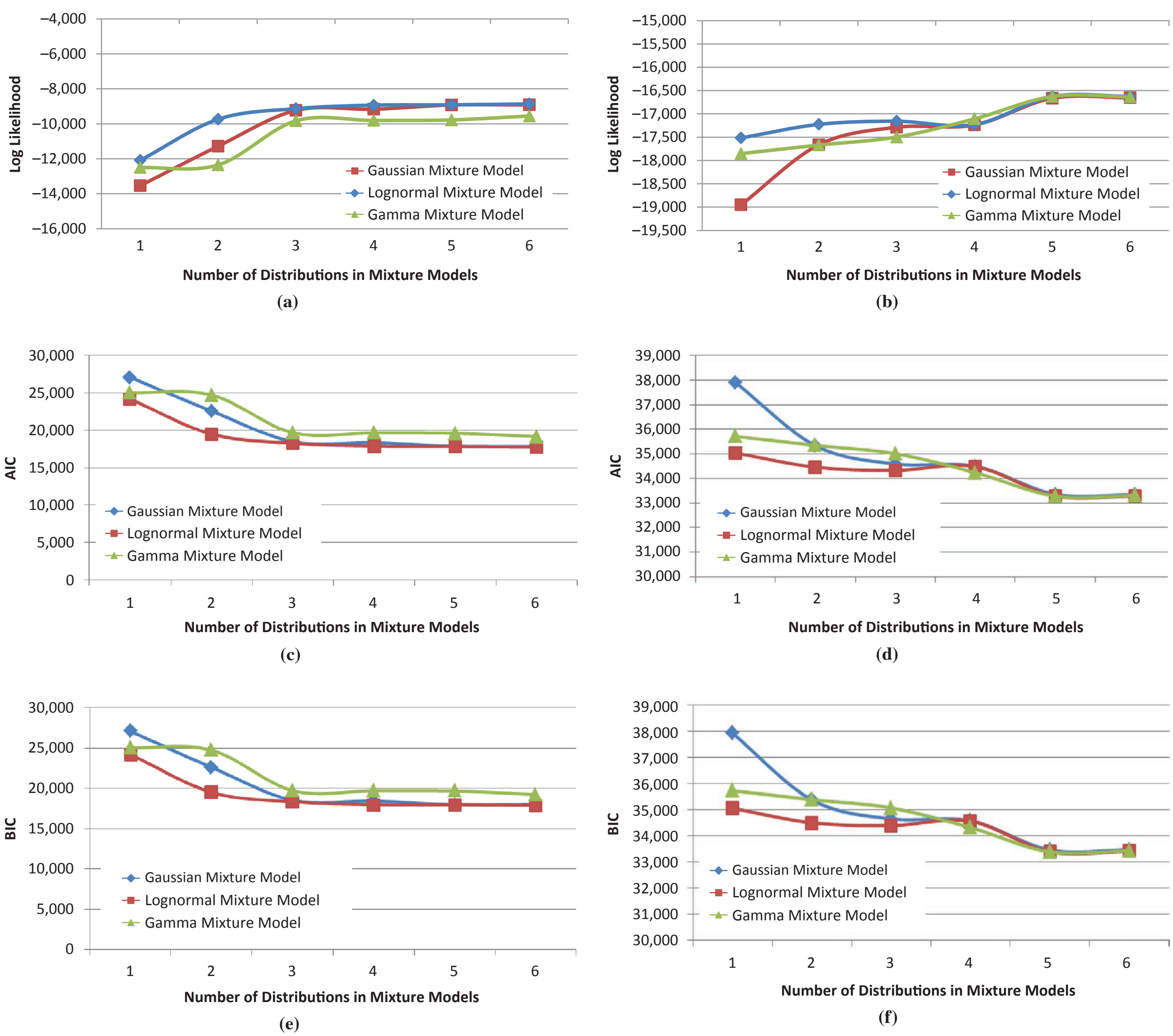


FIGURE 4 Log likelihoods of three mixture models with K lying in [1, 6]: (*a*) Case 1 with log likelihood model, (*b*) Case 2 with log likelihood model, (*c*) Case 1 with AIC model, (*d*) Case 2 with AIC model, (*e*) Case 1 with BIC model, and (*f*) Case 2 with BIC model.

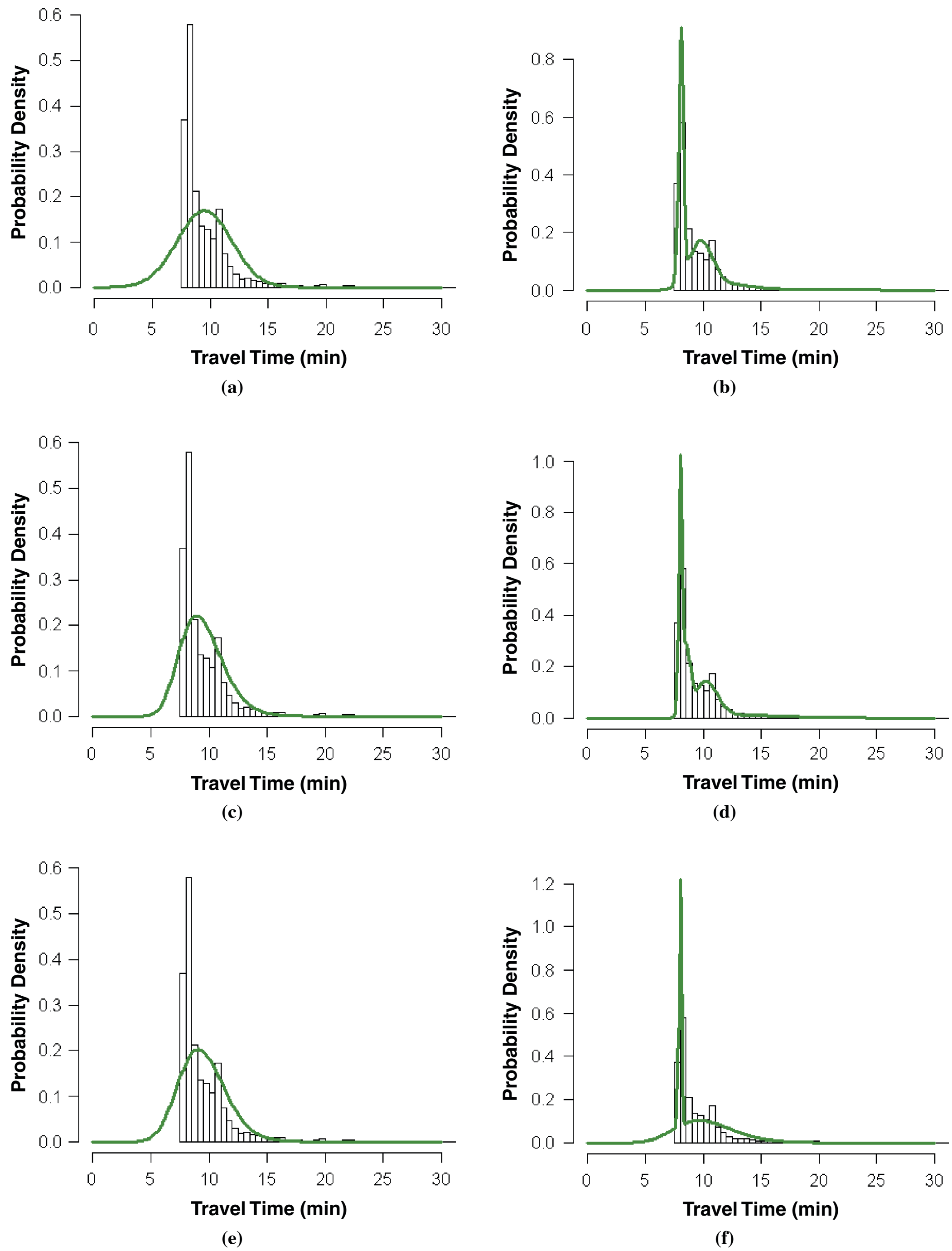


FIGURE 5 Mixture models versus histograms of travel times: (*a*) single Gaussian distribution, (*b*) Gaussian mixture model ($K = 3$), (*c*) single lognormal distribution, (*d*) lognormal mixture model ($K = 3$), (*e*) single gamma distribution, and (*f*) gamma mixture model ($K = 3$).

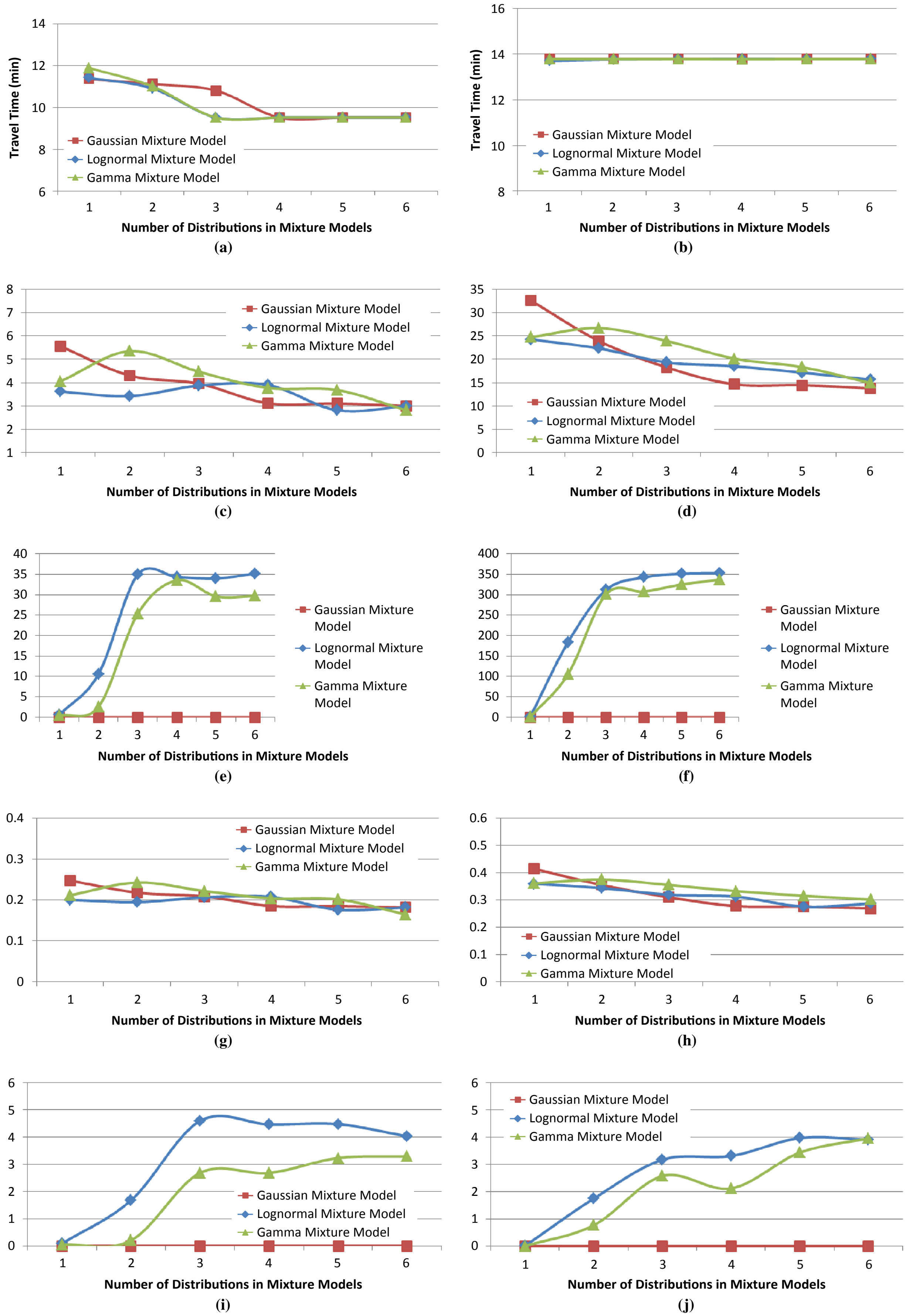


FIGURE 6 Moment-based measure of travel time reliability by using the three mixture models: (*a*) first moment (mean), Case 1; (*b*) first moment, Case 2; (*c*) second moment (variance), Case 1; (*d*) second moment, Case 2; (*e*) third moment (skewness), Case 1; (*f*) third moment, Case 2; (*g*) coefficient of variance, Case 1; (*h*) coefficient of variance, Case 2; (*i*) standardized skewness, Case 1; and (*j*) standardized skewness, Case 2.

the values of travel time reliability by means of the three mixture models with different K. Several findings are summarized below.

All the selected moment-based measures of travel time reliability trended closer together with increasing K. No major differences between these moment-based measures of travel time reliability could be observed when $K \geq 5$. These measures trending together reflected the log likelihood convergences and indicated that, if the travel times could be better fitted with the mixture models, the moment-based measures of travel time reliability would not change.

Because the Gaussian distribution is a symmetric probability distribution, the skewness and standardized skewness of the Gaussian distribution are zero, and thus the skewness and standardized skewness of the Gaussian mixture model are also zero by Equation 9.

Percentile-Based Measures of Travel Time Reliability

Figure 7 shows 10th, 50th, 90th, and 95th travel time percentiles, the BI, and the PI produced from the three mixture models with variable K. Similarly, the trend in Figure 6 can be observed in the percentile-based measures of travel time reliability measures; that is, no major differences between these percentile-based measures can be observed when $K \geq 5$. Insignificant differences exist when the three single probability distributions ($K = 1$) are used; those measures were not reliable because of underfitting. For example, the 90th and 95th percentile travel times based on the three single probability distributions were around 20 and 25 min, respectively. These two numbers may not convey the truth because they were calculated from

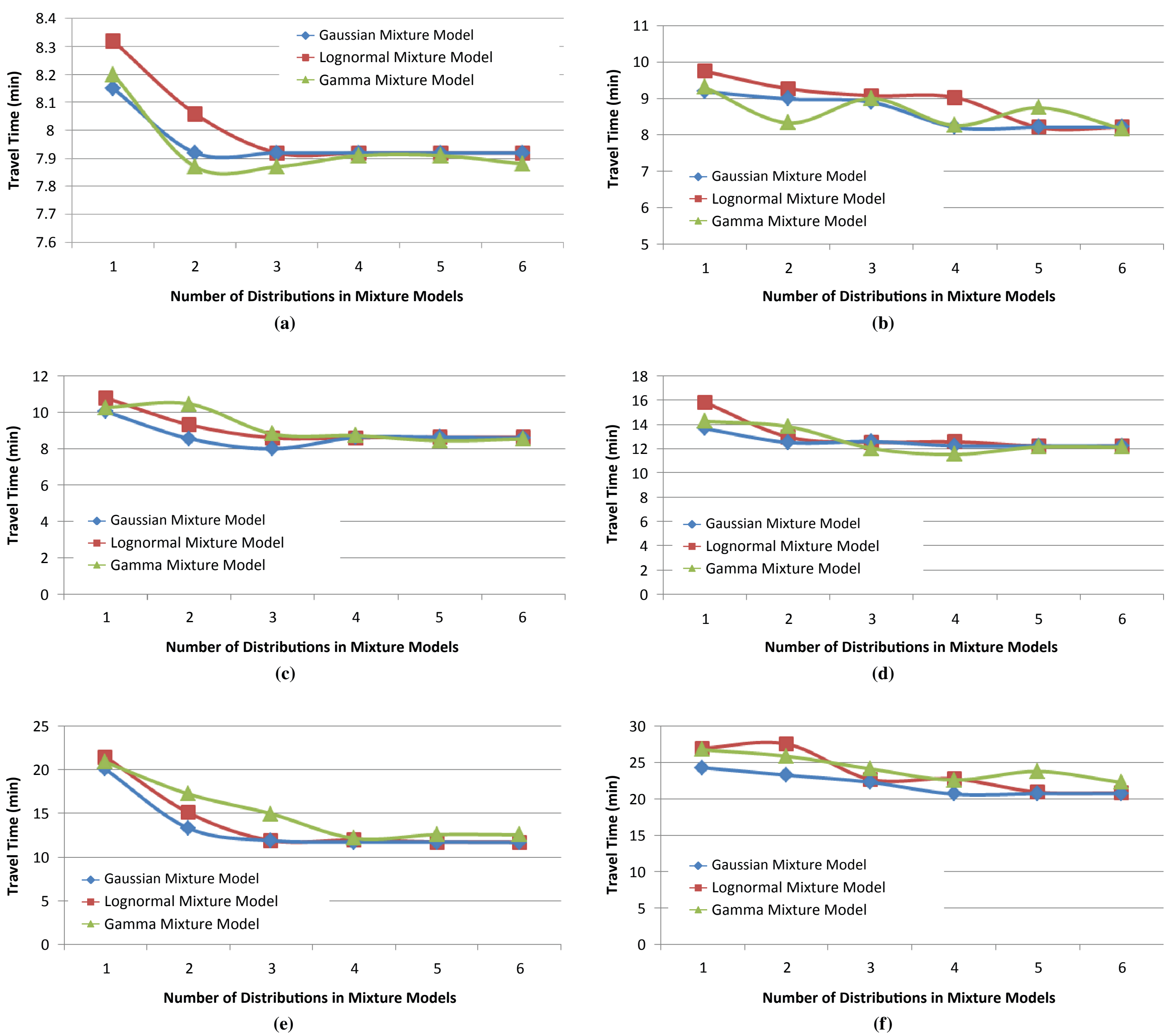


FIGURE 7 Percentile-based measure of travel time reliability by using the three mixture models: (*a*) 10th percentile travel time, Case 1; (*b*) 10th percentile travel time, Case 2; (*c*) 50th percentile travel time, Case 1; (*d*) 50th percentile travel time, Case 2; (*e*) 90th percentile travel time, Case 1; (*f*) 90th percentile travel time, Case 2.

(continued)

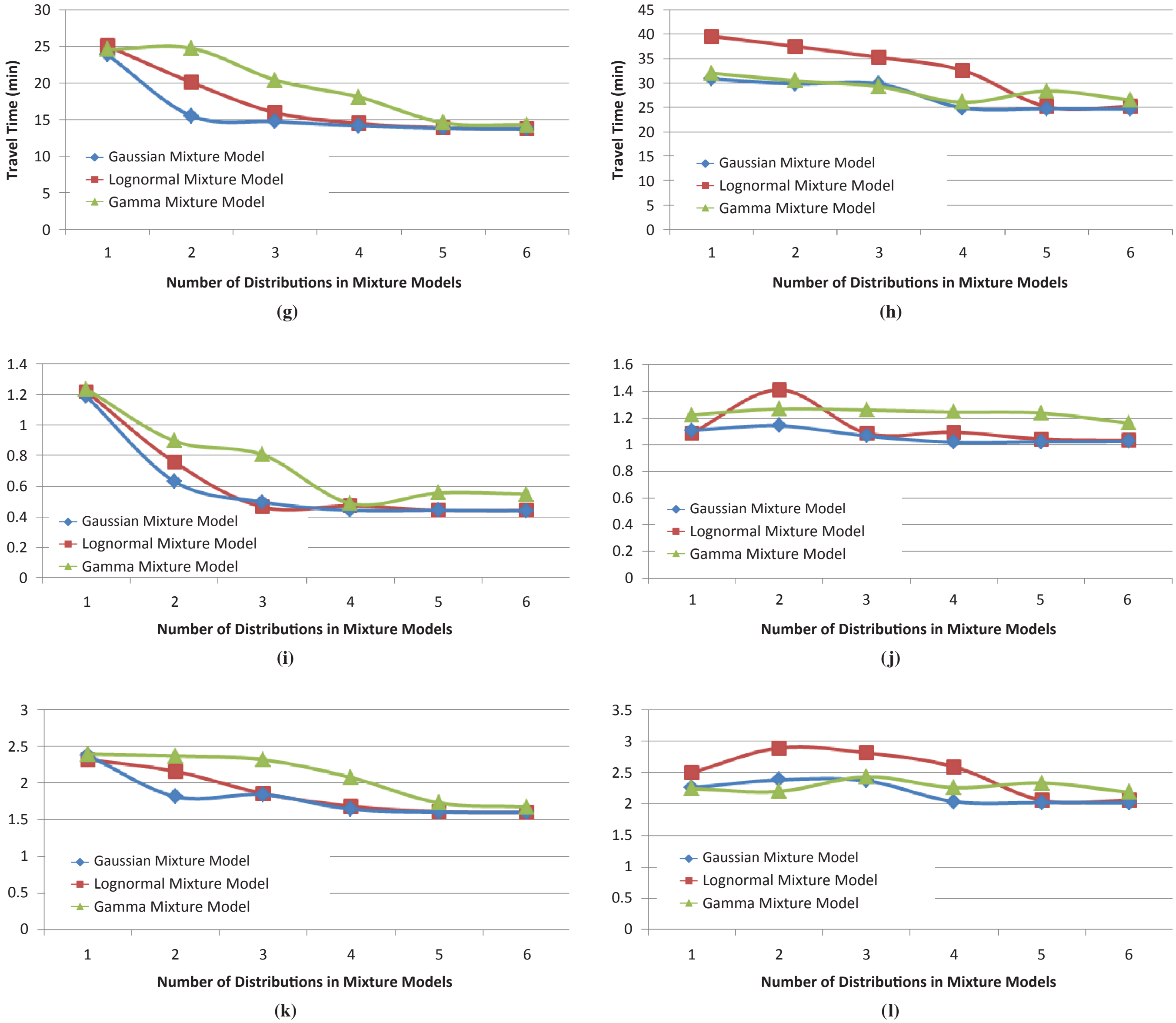


FIGURE 7 *(continued)* **Percentile-based measure of travel time reliability by using the three mixture models: (*g*) 95th percentile travel time, Case 1; (*h*) 95th percentile travel time, Case 2; (*i*) BI, Case 1; (*j*) BI, Case 2; (*k*) PI, Case 1; (*l*) PI, Case 2.**

the underfitted distributions. Instead, when the travel times were better fitted with the three mixture models, the 90th and 95th percentile travel times were, respectively, 11 and 14 min in Case 1.

CONCLUSIONS

Many previous studies also showed the importance of travel time reliability. To improve travel time reliability and ensure travelers' on-time arrivals, accurately measuring travel time reliability is the first step. Previous researchers contributed to the field by proposing new methods to measure reliability through two basic steps: (*a*) data fitting using continuous probability distributions and (*b*) measuring the reliability on the basis of the distributions. However, little agreement was reached on which distribution best fit all conditions, a situation suggesting that a new approach was necessary. Intuitively, measures of travel time reliability are supposed to be independent of the selection of the probability distribution family. Therefore, this study verified two hypotheses: (*a*) travel times can be fitted with multiple probability distributions and are independent of distribution family, and (*b*) measures of travel time reliability are insensitive to the selection of distribution family. A systematic experiment was designed and conducted to test the two hypotheses. Three mixture probability models were used to fit the travel times collected from a Saint Louis freeway corridor. The moment- and percentile-based measures of travel time reliability were derived from these mixture models. The K-S test was used to determine whether the travel times statistically follow the probability distributions generated from the mixture models. The results of the K-S test suggested that the travel times can be more accurately fitted by using the three mixture models with higher values of K to avoid the issue of underfitting; and the log likelihoods and AIC and BIC values suggested appropriate values of K ($K = 3$ and $K = 5$ were the most appropriate values of K in this study) to avoid overfitting. The successes in testing the first hypothesis suggest that (*a*) underfitting may cause disagreement in probability distribution selection and (*b*) travel time can be precisely fitted by using mixture models with higher values of K instead of using single probability distributions, regardless of the probability distribution family in the mixture models. The second hypothesis was tested through comprehensive investigation of moment- and percentile-based reliability measures. The results of these measures of travel time reliability suggested insensitivity to the selection of probability distribution family.

This research provided a detailed insight into the selection of probability distribution family for travel time reliability. The contribution is that researchers and practitioners can avoid the work of testing various distributions. In addition, travel time reliability can be more accurately measured by using mixture models because of the higher value of log likelihoods. Future research could focus on testing additional mixture models on multiple corridors with various traffic conditions. A mathematical proof on the two hypotheses would be conducted instead of designing experiments.

ACKNOWLEDGMENTS

The authors thank the Missouri Department of Transportation for funding and data support. Special thanks go to Payton E. Cooke for language editing of the draft manuscript.

REFERENCES

1. *Travel Time Reliability: Making It There On Time, All the Time.* FHWA, U.S. Department of Transportation, 2006. http://ops.fhwa.dot.gov/publications/tt_reliability/brochure/ttr_brochure.pdf.
2. Small, K.A., R. Noland, X. Chu, and D. Lewis. *NCHRP Report 431: Valuation of Travel-Time Savings and Predictability in Congested Conditions for Highway User-Cost Estimation.* TRB, National Research Council, Washington, D.C., 1999.
3. Carrion, C., and D. Levinson. Value of Travel Time Reliability: A Review of Current Evidence. *Transportation Research Part A,* Vol. 46, No. 4, 2012, pp. 720–741.
4. Sadabadi, K.F., T.H. Jacobs, S. Erdogan, F.W. Ducca, and L. Zhang. *Value of Travel Time Reliability in Transportation Decision Making: Proof of Concept—Maryland.* SHRP 2 Report S2-L35B-RW-1. Transportation Research Board of the National Academies, Washington, D.C., 2015.
5. Polus, A. A Study of Travel Time and Reliability on Arterial Routes. *Transportation,* Vol. 8, No. 2, 1979, pp. 141–151.
6. Lomax, T., D. Schrank, S. Turner, and R. Margiotta. *Selecting Travel Reliability Measures.* Texas A&M Transportation Institute, College Station, 2003.
7. van Lint, J.W.C., and H.J. van Zuylen. Monitoring and Predicting Freeway Travel Time Reliability: Using Width and Skew of Day-to-Day Travel Time Distribution. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1917,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 54–62.
8. Al-Deek, H., and E. Emam. New Methodology for Estimating Reliability in Transportation Networks with Degraded Link Capacities. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations,* Vol. 10, No. 3, 2006, pp. 117–129.
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. *The Florida Reliability Method.* Florida Department of Transportation, Tallahassee, 2007. http://www.dot.state.fl.us/planning/statistics/mobilitymeasures/reliability.pdf.
11. van Lint, J.W.C., H.J. van Zuylen, and H. Tu. Travel Time Unreliability on Freeways: Why Measures Based on Variance Tell Only Half the Story. *Transportation Research Part A,* Vol. 42, No. 1, 2008, pp. 258–277.
12. 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.
13. Kwon, J., T. Barkley, R. Hranac, K. Petty, and N. Compin. Decomposition of Travel Time Reliability into Various Sources: Incidents, Weather, Work Zones, Special Events, and Base Capacity. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2229,* Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 28–33.
14. Park, S., H. Rakha, and F. Guo. Calibration Issues for Multistate Model of 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. 74–84.
15. 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.
16. Lei, H., T. Xing, J.D. Taylor, and X. Zhou. Monitoring Travel Time Reliability from the Cloud: Cloud Computing–Based Architecture for Advanced Dissemination of Traffic Information. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2291,* Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 35–43.
17. Pu, W. Analytic Relationships Between Travel Time Reliability Measures. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2254,* Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 122–130.
18. Yang, S., A. Malik, and Y.-J. Wu. Travel Time Reliability Using the Hasofer–Lind–Rackwitz–Fiessler Algorithm and Kernel Density Estimation. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2442,* Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 85–95.
19. Rokach, L., and O. Maimon. *Data Mining with Decision Trees: Theory and Applications.* World Scientific Publishing Company, Singapore, 2008.
20. Yang, X., and X.G. Su. *Linear Regression Analysis: Theory and Computing.* World Scientific Publishing Company, Singapore, 2009.
21. Rogers, S., and M. Girolami. *A First Course in Machine Learning.* CRC Press, Boca Raton, Fla., 2011.
22. MacQueen, J.B. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability 1967,* Vol. 1, No. 233, 1967, pp. 281–297.
23. *Comprehensive R Archive Network.* R Foundation, 2015. https://cran.r-project.org/. Accessed June 26, 2015.
24. *Teach Traffic Smart.* On the Jeep Blog, 2015. https://onthejeep.wordpress.com/2015/07/12/three-mixture-models/. Accessed July 12, 2015.
25. Frühwirth-Schnatter, S. *Finite Mixture and Markov Switching Models.* Springer, New York, 2006.

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