

Research Article

Estimation of Bimodal Urban Link Travel Time Distribution and Its Applications in Traffic Analysis

Yuxiong Ji,^{1,2} Shengchuan Jiang,¹ Yuchuan Du,¹ and H. Michael Zhang^{1,3}

¹*The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China*

²*Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Sipailou No. 2, Nanjing 210096, China*

³*Department of Civil and Environmental Engineering, University of California, Davis, CA 95616, USA*

Correspondence should be addressed to Yuchuan Du; ycdu@tongji.edu.cn

Received 30 December 2014; Accepted 20 February 2015

Academic Editor: Gisele Mophou

Copyright © 2015 Yuxiong Ji et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Vehicles travelling on urban streets are heavily influenced by traffic signal controls, pedestrian crossings, and conflicting traffic from cross streets, which would result in bimodal travel time distributions, with one mode corresponding to travels without delays and the other travels with delays. A hierarchical Bayesian bimodal travel time model is proposed to capture the interrupted nature of urban traffic flows. The travel time distributions obtained from the proposed model are then considered to analyze traffic operations and estimate travel time distribution in real time. The advantage of the proposed bimodal model is demonstrated using empirical data, and the results are encouraging.

1. Introduction

Travel time is an important piece of information for transportation planners, traffic operators, and road users. It has been widely used in the studies of route choice, origin-destination (OD) flow estimation, and transportation system reliability [1]. Travel time is also an essential input for the design and implementation of advanced route guidance and traveler information systems.

Loop detectors are the most common data source for travel time estimation, particularly on freeways. Since loop detectors provide traffic information, such as volume, speed, and occupancy, at fixed locations, additional assumptions need to be made to estimate vehicle travel times [2, 3]. Recently, probe vehicles utilizing automatic vehicle location (AVL) technologies such as those based on the global positioning system (GPS) have been used to collect vehicle travel times directly. Data from probe vehicles have been considered in various applications, such as congestion identification and incident detection [4–6].

Travel times are traditionally modeled with unimodal distributions, such as normal, lognormal, Gamma, and Burr

distributions [7–9]. Since the mean is a commonly used summary statistic for variables that are unimodally distributed, some studies specifically focused on the mean travel time (or speed) estimation [4, 10–12]. For example, Sen et al. [4] and Hellinga and Fu [10] analyzed probe link travel times and concluded that the sample mean does not approach asymptotically the population mean. Hellinga and Fu [11] developed a method based on stratified sampling techniques to reduce the bias in the mean travel time estimation. Pu et al. [12] developed Bayesian updating procedure to infer the mean speed of general vehicles from bus probe data.

However, recent studies have revealed that travel time distribution tends to be bimodal and even multimodal. Jintanakul et al. [13] argued that travel time distributions on freeways would not be unimodal due to the mixes of driving patterns and of vehicle types. They developed a bimodal model and illustrated it using simulation data. Guo et al. [14] proposed a multistate model to assess travel time distributions on freeways. They explained that travel time distributions would not be unimodal since multiple traffic states could exist in the same period.

Travel time distributions on urban streets are more complex than those on freeways. Vehicles travelling on urban streets are heavily influenced by traffic signal controls, pedestrian crossings, and conflicting traffic from cross streets. The interrupted nature of urban traffic flows would likely result in large fluctuations in observed travel times. For example, a vehicle passing a signal at the end of the green would experience quite different travel time than the vehicle following behind it that must make a stop for the red, although they traveled next to each other. Taylor and Somenahalli [1] have demonstrated empirically that urban link travel times are bimodally distributed. The bimodal property of urban link travel time distributions has also been studied analytically in [15, 16].

This paper develops methodologies to analyze urban link travel times from probe vehicles. Unlike previous travel time studies, this paper not only develops methodologies to capture important characteristics of the travel time distributions, but also explores potential applications of the travel time distributions resulting from the developed model. Specifically, the resulting travel time distributions are used to analyze traffic operations and estimate travel time distribution in real time. The advantage of the developed methodologies is demonstrated using empirical data.

The rest of the paper starts with the travel time model and the algorithms to estimate model parameters. Then we introduce the dataset used in the empirical study. Next we demonstrate how to use the travel time distributions to analyze traffic operations and estimate travel time distribution in real time. Finally, we conclude the paper with a summary of the key findings and possible future research.

2. The Methodologies

2.1. A Hierarchical Bayesian Bimodal Travel Time Model. Probe travel times collected on a given urban link in a given time-of-day period over multiple days are considered in this study. A hierarchical Bayesian bimodal model is developed to fit travel time distributions. Travel time distributions on urban streets are affected by various factors. Some of the factors, such as geometric characteristics (length, lane width, speed limits, etc.), are time-invariant, while others, such as traffic volume, traffic mix, and signal timing, could vary over days. Therefore, it is natural to model travel times hierarchically, with observed travel times modeled conditionally on certain parameters, which themselves are given a probabilistic specification in terms of further parameters, known as hyperparameters. The associations among travel times in different days are captured by using a joint probability distribution for model parameters in different days.

Given that travel times are positive and their distributions are positively skewed and that urban link travel time distributions tend to be bimodal, it is assumed that each component of the bimodal travel time distribution follows log normal distribution. That is, the model is fitted to the logarithms of the travel time observations. Specifically, the log travel times in a day j are modeled as a two-component mixture: one component is referred to as “fast vehicle” and the other is referred to as “slow vehicle.” A travel time observation is

considered to come from the “fast vehicle” distribution with probability $(1 - r)$ and from the “slow vehicle” distribution with probability r . In a day j , the “fast vehicle” distribution is assumed to be a normal distribution with mean θ_j and variance σ^2 , and the “slow vehicle” distribution is assumed to be a normal distribution with mean $\theta_j + \tau_j$ and variance ν^2 . The variation of the travel times between days is modeled by having the mean θ_j following a normal distribution with mean μ and variance ω^2 and τ_j following a normal distribution with mean β and variance δ^2 . Let y_{ij} represent the log travel time for an observation i in a day j ; the model can be written in the following hierarchical form:

$$\begin{aligned} y_{ij} &| \theta_j, \tau_j, \zeta_{ij}, \sigma^2, \nu^2 \\ &\sim N(\theta_j + \tau_j \zeta_{ij}, (1 - \zeta_{ij}) \sigma^2 + \zeta_{ij} \nu^2), \end{aligned} \quad (1)$$

$$\theta_j | \mu, \omega^2 \sim N(\mu, \omega^2), \quad (2)$$

$$\tau_j | \beta, \delta^2 \sim N(\beta, \delta^2), \quad (3)$$

$$\zeta_{ij} | r \sim \text{Bernoulli}(r), \quad (4)$$

where ζ_{ij} is an unobserved indicator variable that equals one if an observation i in a day j belongs to the “slow vehicle” group and zero if it belongs to the “fast vehicle” group. The parameter θ_j represents the expected log travel time in a day j if a vehicle is not delayed. The parameter $\theta_j + \tau_j$ represents the expected log travel time in a day j when a vehicle encounters a long delay. The mixture probability r represents the probability that a bus encounters a long delay. The expected travel times for vehicles in the “slow vehicle” and “fast vehicle” groups and the expected delay in a day j can be obtained by, respectively,

$$E_j^s = \exp\left(\theta_j + \tau_j + \frac{\nu^2}{2}\right), \quad (5)$$

$$E_j^f = \exp\left(\theta_j + \frac{\sigma^2}{2}\right), \quad (6)$$

$$E_j^d = E_j^s - E_j^f. \quad (7)$$

2.2. Posterior Distributions of the Model Parameters. Weak prior distributions (i.e., the prior densities are diffuse) are used in this paper such that the posterior distributions are dominated by the data. The prior distribution of the mixture probability r is taken to be uniform on $(0.01, 0.99)$ as values of zero or one would not correspond to mixture distributions. A Gamma distribution with the shape and rate parameters of 0.001 is used for the prior distributions of $1/\sigma^2$, $1/\nu^2$, $1/\omega^2$, and $1/\delta^2$. A normal distribution with mean 0 and variance 10000 is used for the prior distributions of the parameters μ and β . The parameter τ_j is restricted to be positive such that the model is identifiable [17].

Given the distributional assumptions of (1)–(4) and the assumptions of the prior distributions, the posterior distributions of the model parameters and the unobserved indicator variables are given by

$$\begin{aligned} p(\Psi, \zeta | y) &\propto \prod_j \prod_i f(y_{ij} | \theta_j, \tau_j, \zeta_{ij}, \sigma^2, \nu^2) \times f(\zeta_{ij} | r) \\ &\times \prod_j f(\theta_j | \mu, \omega^2) \times f(\tau_j | \beta, \delta^2) p(\mu) \\ &\times p(\beta) \times p(r) \times p\left(\frac{1}{\sigma^2}\right) \times p\left(\frac{1}{\nu^2}\right) \times p\left(\frac{1}{\omega^2}\right) \\ &\times p\left(\frac{1}{\delta^2}\right), \end{aligned} \quad (8)$$

where Ψ and ζ represent the sets of the model parameters and of the unobserved indicator variables, respectively. y represents the collection of the travel time observations across all days.

Even though the proposed model is developed based on standard distributions, the marginal posterior distributions of the model parameters are analytically intractable [17]. We consider the Gibbs sampler [17] to simulate the posterior distributions. The Gibbs sampler divides the parameter vector into several components and draws samples for a joint posterior distribution (e.g., (8)) from the posterior distribution of one component conditional on the values of the other components. This method is useful when it is difficult to sample from the marginal posterior distributions directly, while it is easy to sample from the conditional posterior distributions.

We adopt the “rjags” (stands for just another Gibbs sampler in R) package in the statistics software “R” to simulate the marginal posterior distributions of the model parameters [18]. The implementation is straightforward for the one who is familiar with “rjags” and, therefore, the algorithm is not discussed further in the paper.

2.3. Real-Time Estimation of Travel Time Distribution. Real-time estimation of travel time distribution is an important component in advanced traveler information system. Conditional on the real-time observed travel times, we develop a methodology to estimate the parameters of the travel time distribution for a given period of a given day. Historical data are considered in the methodology through the prior distribution of the model parameters.

The parameter set (Φ) of the travel time distribution in a given period of a given day consists of θ , τ , σ^2 , ν^2 , and r (see (1) and (4)). Note that the subscript j representing a day is omitted for simplicity in the following without the loss of generality. The prior information and the real-time available data are combined to produce the posterior distribution of

the parameter set (Φ) and the set of the unobserved indicator variables (ζ):

$$\begin{aligned} p(\Phi, \zeta | y) &\propto \prod_i f(y_i | \theta, \tau, \zeta_i, \sigma^2, \nu^2) \times f(\zeta_i | r) \\ &\times p(\theta | \mu, \omega^2) \times p(\tau | \beta, \delta^2) \times p(r) \\ &\times p\left(\frac{1}{\sigma^2}\right) \times p\left(\frac{1}{\nu^2}\right). \end{aligned} \quad (9)$$

The posterior distributions of the model parameters obtained based on historical data are used as the prior distributions in (9). Conjugate prior distributions are adopted to summarize the samples of the posterior distributions produced by the MCMC algorithm. The conjugacy property means that the posterior distribution follows the same parametric form as the prior distribution [17]. Specifically, the prior distributions of θ and τ are assumed to be normal distributions ((2) and (3), resp.). The prior distributions of $1/\sigma^2$ and $1/\nu^2$ are assumed to be Gamma(k_1, k_2) and Gamma(b_1, b_2), respectively. The prior distribution of r is assumed to be Dirichlet(α_1, α_0). The parameters of the prior distributions are estimated by matching their theoretical mean and variance to the corresponding values estimated from the MCMC samples.

The MCMC algorithm can be used to produce the posterior distributions of the model parameters. Nevertheless, the MCMC algorithm may not be suitable for real-time application due to its relatively high computational requirement. Therefore, we developed an expectation conditional maximization (ECM) algorithm [19] to produce the maximum likelihood estimates (MLE) of the model parameters (i.e., the parameter set Φ). The ECM algorithm, which includes an *E*- (expectation-) step and a *CM*- (conditional maximization-) step, is an iterative method for finding the marginal posterior mode from the joint posterior distribution [19]. In the context of the problem of interest, the ECM algorithm finds the mode of the marginal posterior distribution of the parameter set Φ from the joint posterior distribution of (9). The ECM algorithm, as applied to the problem at hand, consists of the following.

- (1) Start with some (likely crude) estimates of the model parameters $\theta(0)$, $\tau(0)$, $\sigma^2(0)$, $\nu^2(0)$, and $r(0)$.
- (2) For $h = 1, 2, \dots$, apply the following two steps iteratively:

(2.1) *E*-step: determine the expected value of ζ_i :

$$E(\zeta_i)$$

$$= \frac{N_2(y_i) \times r(h-1)}{N_1(y_i) \times (1 - r(h-1)) + N_2(y_i) \times r(h-1)}, \quad (10)$$

where $N_1(y_i)$ represents the density of an observation y_i , assuming that it follows the “fast vehicle” distribution $N(\theta(h-1), \sigma^2(h-1))$, and $N_2(y_i)$ represents the density of an observation y_i , assuming that it follows the “slow vehicle” distribution $N(\theta(h-1) + \tau(h-1), \nu^2(h-1))$.

(2.2) CM-steps:

(a) update the probability r by

$$r(h) = \frac{\sum_i E(\zeta_i) + \alpha_1 - 1}{n + \alpha_0 - 2}; \quad (11)$$

(b) update the expected parameter θ by

$$\begin{aligned} \theta(h) &= \left(\frac{\mu}{\omega^2} + \frac{\sum_i (y_i - \tau) \times E(\zeta_i)}{\nu^2 (h-1)} + \frac{\sum_i y_i \times (1 - E(\zeta_i))}{\sigma^2 (h-1)} \right) \\ &\cdot \left(\frac{1}{\omega^2} + \frac{\sum_i E(\zeta_i)}{\nu^2 (h-1)} + \frac{\sum_i (1 - E(\zeta_i))}{\sigma^2 (h-1)} \right)^{-1}; \end{aligned} \quad (12)$$

(c) update the expected parameter τ by

$$\begin{aligned} \tau(h) &= \left(\frac{\beta}{\delta^2} + \frac{\sum_i (y_i - \theta(h)) \times E(\zeta_i)}{\nu^2 (h-1)} \right) \\ &\cdot \left(\frac{1}{\delta^2} + \frac{\sum_i E(\zeta_i)}{\nu^2 (h-1)} \right)^{-1}; \end{aligned} \quad (13)$$

(d) update the variance σ^2 by

$$\begin{aligned} \sigma^2(h) &= \left(\frac{\sum_i (1 - E(\zeta_i)) \times (y - \theta(h))^2}{2} + k_2 \right) \\ &\cdot \left(\frac{\sum_i (1 - E(\zeta_i))}{2} + k_1 - 1 \right)^{-1}; \end{aligned} \quad (14)$$

(e) update the variance ν^2 by

$$\begin{aligned} \nu^2(h) &= \left(\frac{\sum_i E(\zeta_i) \times (y - \theta(h) - \tau(h))^2}{2} + b_2 \right) \\ &\cdot \left(\frac{\sum_i E(\zeta_i)}{2} + b_1 - 1 \right)^{-1}. \end{aligned} \quad (15)$$

Steps (2.1) and (2.2) are applied iteratively until some stopping criteria are satisfied.

3. Empirical Study

The methodologies proposed in Section 2 are general and can be applied to any type of probe vehicle data. This paper evaluates the proposed methodologies using probe bus data. Even though the driving patterns of buses are different from those of general vehicles, studies have confirmed that bus travel times are highly correlated with those of general vehicles and that bus travel times can be used to infer the travel times of general vehicles [12, 20, 21]. The empirical results presented below would provide insight for further methodology developments and potential applications of the travel time distributions obtained by the proposed model.

3.1. The Data. The probe bus data are provided by the Campus Area Bus Service (CABS) at the Ohio State University (OSU). The CABS serves approximately four million passengers annually on seven routes on and in the vicinity of the OSU Campus. GPS-based AVL systems have been used on all CABS buses since 2009. The AVL system records bus statuses (e.g., location and velocity) at a frequency of 1 Hz. This study considers AVL data collected by buses serving the Campus Loop South (CLS), Campus Loop North (CLN), and North Express (NE) bus routes. The advertised headways of the CLS, CLN, and NE bus routes are 9, 9, and 5 minutes, respectively.

Two links are considered in the empirical demonstration. The lengths of Links 1 and 2 are 248.1 and 216.8 meters, respectively. Link 1 contains a four-way signalized intersection and Link 2 contains a four-way signalized intersection and a pedestrian crossing. To capture the total time losses due to vehicle acceleration and deceleration caused by signal controls or pedestrians crossing the street in the travel times, links are defined such that the intersection and pedestrian crossing are located inside the links. In addition, although it is possible to eliminate the increase in travel time due to stopping at bus stops for passengers alighting and boarding [8], bus stops are excluded from the defined links to control the effect of the possible errors resulting from the travel time correction.

Travel times collected in a.m. (7:30 a.m.–7:45 a.m.) and p.m. (4:30 p.m.–4:45 p.m.) periods of 40 weekdays are considered for the empirical demonstration. Summary statistics of the travel time observations and the signal information for the corresponding travel direction are provided in Table 1. The three numbers in the third row represent the minimal, mean, and maximal numbers of travel time observations in the given period of a day.

As shown in Table 1, the sample size in a given period of a day is small. The mean travel times in both periods of Link 1 are close, while on Link 2, the mean travel time in a.m. period is 62.1% higher than that in p.m. period. In terms of the signal information, the green ratios (i.e., the ratio of the green phase to the cycle length) in a.m. period are 28.5% and 42.2% lower than those in p.m. period on Links 1 and 2, respectively.

Figure 1 presents the empirical distribution of the log travel time for each link and period combination. The distributions are obtained by pooling all observations collected in different days. On Link 1, the distributions in both periods have similar shape, while on Link 2, a.m. period has more “slow bus” observations than p.m. period. In addition, all distributions presented in Figure 1 indicate certain level of bimodality. The bimodality on Link 1 is more obvious than that on Link 2. The number inside the bracket of the title for each plot is the bimodality coefficient (BC) [22]. BC values greater than 0.556 (i.e., 5/9) indicate a bimodal or multimodal distribution. As can be seen, the BC values for three out of four link and period combinations are larger than 0.556 and the BC value for Link 2-a.m. period is close to 0.556. The empirical results reveal that unimodal distribution is not sufficient to describe urban link travel times.

3.2. Analysis of Traffic Operations. Two Markov chains are run for each link and period combination to evaluate the

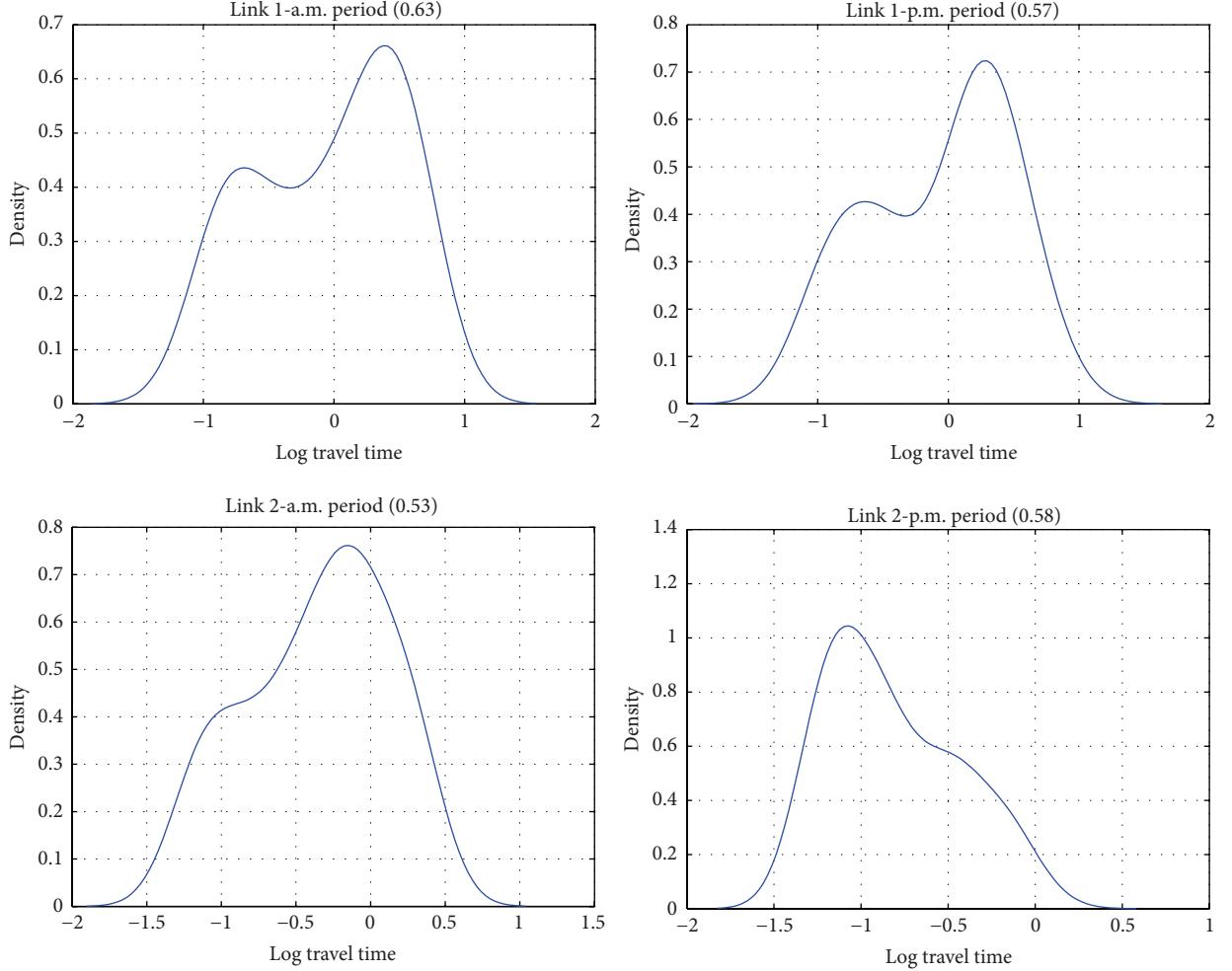


FIGURE 1: Empirical distributions of the log travel time for each link and period combination.

TABLE 1: Summary statistics for the two links of interest.

	Link 1		Link 2	
	a.m. period	p.m. period	a.m. period	p.m. period
Number of observations	(1, 4.2, 6)	(2, 3.3, 5)	(1, 4.2, 7)	(2, 3.2, 5)
Mean travel time (second)	66	64	47	29
Cycle length (second)	120	120	100	100
Green phase (second)	35	49	37	64

convergence of the MCMC algorithm. Each chain is run for 20,000 iterations and the first 10,000 iterations are discarded as burn-in. The rest of the sequence is thinned by keeping samples of every 10th iteration to reduce the correlations between samples, resulting in 2,000 samples (1,000 samples for one chain) to represent the posterior distributions of the model parameters. The MCMC samples of some model parameters are used in the following to analyze traffic operations.

Figure 2 presents the posterior distributions of the mixture probability. Figure 2 reveals that vehicles are more likely

to be delayed in the a.m. period than in the p.m. period since the mixture probability in a.m. period is greater than that in p.m. period on both links. Specifically, on Link 1, the posterior means of the mixture probability in a.m. and p.m. periods are 0.69 and 0.57, respectively. On Link 2, the posterior means of the mixture probability in a.m. and p.m. periods are 0.79 and 0.48, respectively.

Figure 3 presents the posterior distributions of the expected delay. On Link 1, the expected delays in a.m. period are slightly longer than those in p.m. period and the expected delay in p.m. period of the 24th day is abnormally long.

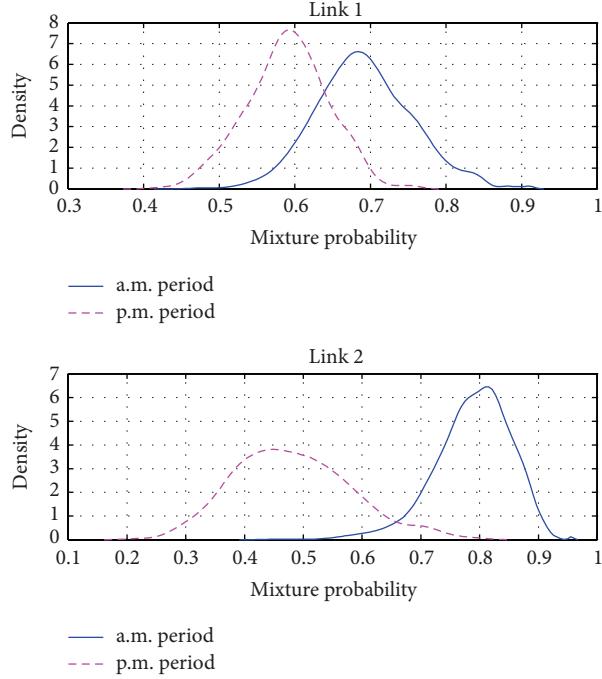


FIGURE 2: Posterior distributions of the mixture probability.

TABLE 2: Comparisons of signal timing and model parameters.

Link	Link 1		Link 2		
	Period	a.m.	p.m.	a.m.	p.m.
Red ratio	0.71	0.59	0.63	0.36	
Mixture probability	0.69	0.57	0.79	0.48	
Red phase (second)	85.0	71.0	63.0	36.0	
Expected delay (second)	56.2	52.6	31.8	18.3	

Specifically, the means of the expected delay in a.m. and p.m. periods are 56.2 and 52.6 seconds, respectively. The mean of expected delay in p.m. period of the 24th day is 79.9 seconds. On Link 2, the expected delays in a.m. period are much longer than those in p.m. period. The means of the expected delay in a.m. and p.m. periods are 31.8 and 18.3 seconds, respectively.

Table 2 compares the mixture probability and expected delay with the red ratio (i.e., the ratio of the red phase to the cycle length) and red phase, respectively. As can be seen in Table 2, the mixture probability is comparable with the red ratio. On Link 1, the red ratios are close to the mixture probabilities. On Link 2, the mixture probabilities are relatively higher than the red ratios, which is very likely due to the high pedestrian volume crossing the street. In addition, it can be seen that the expected delay increases approximately linearly with the red phase. The results suggest that the mixture probability and the expected delay could reflect the effect of signal timing and other factors that interrupt the flows of urban traffic.

Figure 4 presents the posterior distributions of the expected travel times for the “fast bus” and “slow bus” groups. The expected travel times for both groups are clearly distinguished from each other. On Link 1, the expected travel times for the “fast” group in both periods are close, and the expected travel times for the “slow” group in a.m. period are slightly higher than those in p.m. period. In addition, the expected travel time for the “slow” group in p.m. period of the 24th day is much higher than those in other days, indicating possible abnormal event. Specifically, the means of the expected travel time for the “fast bus” group in a.m. and p.m. periods are 33.1 and 32.9 seconds, respectively. The means of the expected travel time for the “slow bus” group in a.m. and p.m. periods are 89.3 and 85.5 seconds, respectively. The mean of the expected travel time for the “slow bus” group in p.m. period of the 24th day is 115.4 seconds.

On Link 2, the expected travel times for the “fast” group in a.m. period are slightly higher than those in p.m. period, and the expected travel times for the “slow” group in a.m. period are much higher than those in p.m. period. The higher expected travel times for the “slow” group in a.m. period are due to the greater value of the red ratio in a.m. period, as discussed when presenting Table 2. Specifically, the means of the expected travel time for the “fast bus” group in a.m. and p.m. periods are 21.6 and 20.2 seconds, respectively. The means of the expected travel time for the “slow bus” group in a.m. and p.m. periods are 53.3 and 38.5 seconds, respectively.

As demonstrated above, the proposed model provides useful information for analyzing traffic operations on urban streets. Such information cannot be obtained when travel

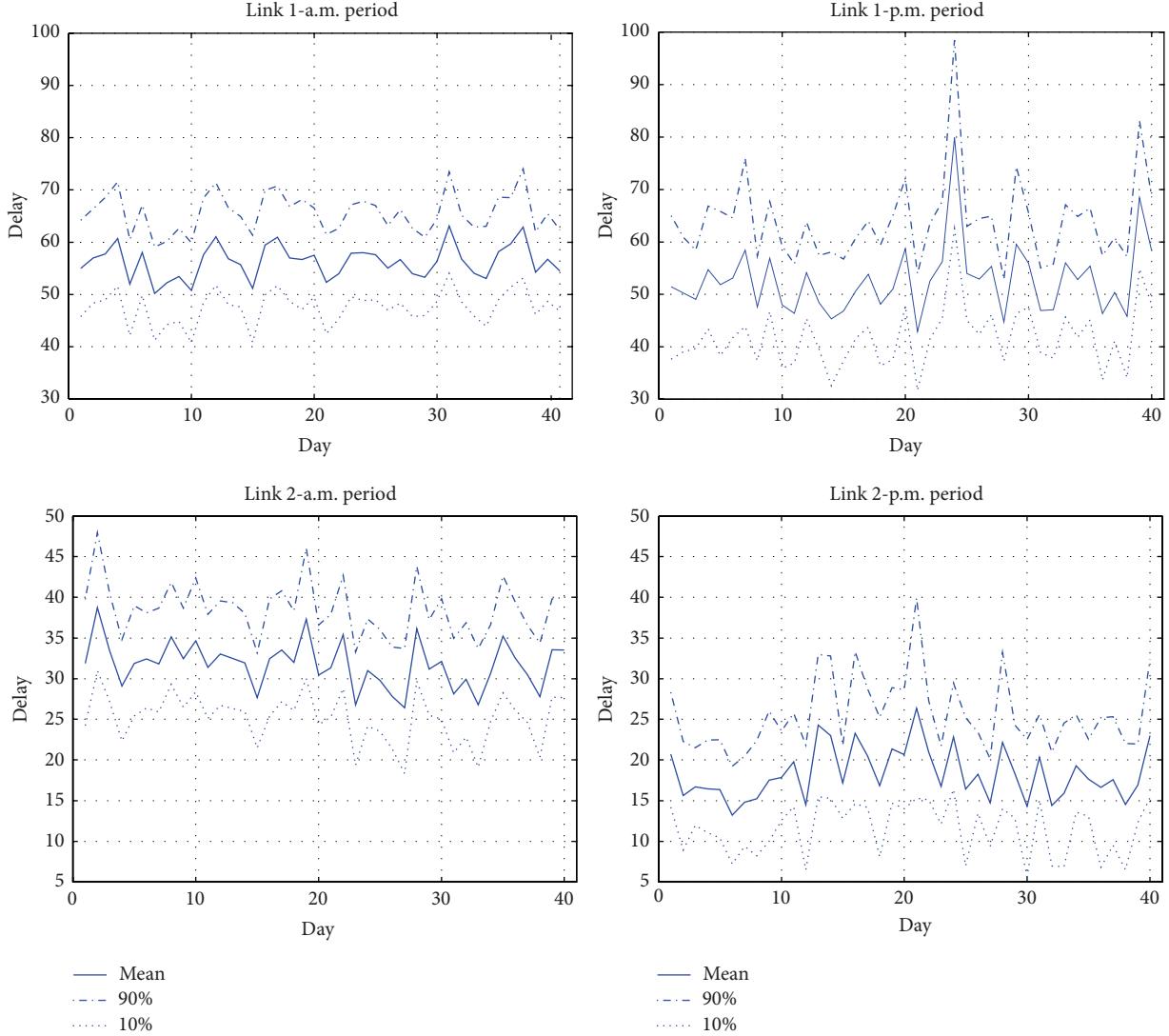


FIGURE 3: Posterior distributions of the expected delay (in seconds).

times are assumed to be unimodally distributed. The results presented in Figures 2–4 also provide new statistics to report travel times. For example, travel times on Link 1 in a.m. period could be summarized as follows: drivers would experience the travel time of 33.1 seconds with the probability of 31% and would experience the travel time of 89.1 seconds with the probability of 69%, or drivers would encounter the delay of 56.2 seconds with the probability of 69%.

3.3. Real-Time Estimation of Travel Time Distribution. To evaluate the performance of the proposed methodologies in estimating travel time distribution, travel times on Link 1 in a.m. period of a hypothetical day are simulated. The true model parameters for the hypothetical day are randomly selected from the MCMC samples. Based on the “assumed” true model parameters, various numbers of travel time observations are simulated using (1) and (4). The ECM algorithm is applied to produce the MLEs of the model parameters.

The model proposed in [12] (referred to as unimodal model) is also considered in the evaluation. The unimodal model ignores the interrupted nature of urban traffic flows and assumes that travel time distribution is unimodal. The unimodal model is a special case of the bimodal model. That is, it can be obtained by setting the mixture probability to be zero in the bimodal model.

The estimated travel time distributions produced by the two methods are presented in Figure 5. The “assumed” true distribution is also presented for comparison purposes. In addition, the simulated observations are presented as circles in the horizontal axis. The title of each subplot represents the number of travel time observations. Figure 5 suggests better performance of the proposed bimodal model. Regardless of the number of observations, the distribution produced by the bimodal model is closer to the true distribution. The modes of the distributions produced by the unimodal model are located near the saddle points of the

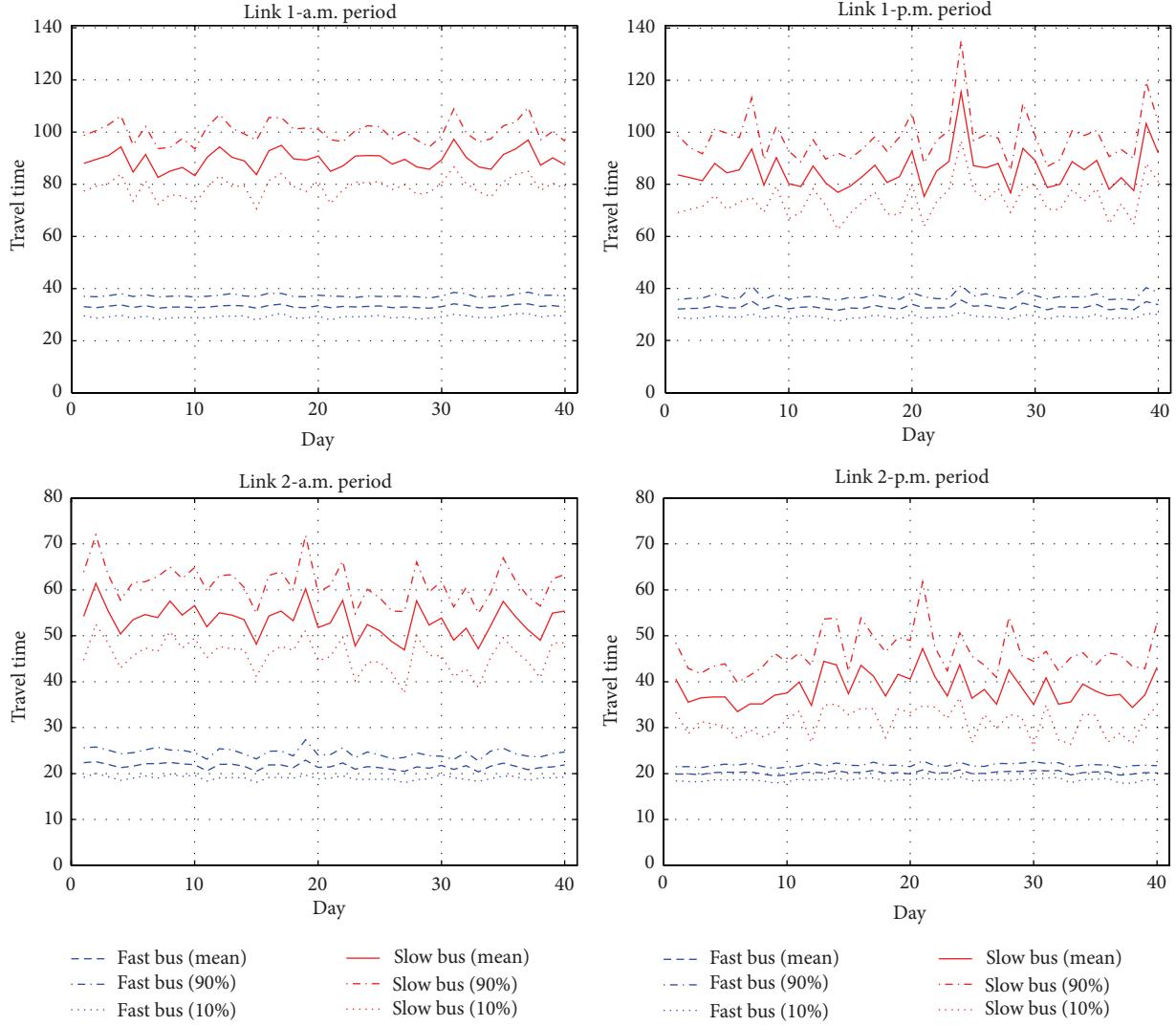


FIGURE 4: Posterior distributions of the expected travel time (in seconds) for the “fast bus” and “slow bus” groups.

true distributions, suggesting that the unimodal model is not sufficient for modelling vehicle travel times on urban streets.

Furthermore, we carry out the simulation for 100 times. In each time, the mean, lower quantile (i.e., 25% quantile), and the upper quantile (i.e., 75% quantile) of the travel time distributions produced by the bimodal and unimodal models are computed. The simulated distributions of the three summary statistics for 5 and 100 numbers of observations are presented in Figure 6. The “assumed” true summary statistics are presented as circles in the horizontal axis. Figure 6 further demonstrates the better performance of the bimodal model. The true values are close to the modes of the distributions of the summary statistics produced by the bimodal model. In terms of the distributions produced by the unimodal model, the true mean and upper quantile are located in the right sides of their corresponding distributions, while the true lower quantile is located in the left side of the

distribution. These results indicate that the unimodal model tends to underestimate the mean and the upper quantile and overestimate the lower quantile. The results on other link and period combinations also show similar patterns as in Figures 5 and 6.

4. Conclusions and Future Research

Empirical studies have revealed that the prevalent travel time distributions on urban links are bimodal, thanks to the presence of traffic signals. A hierarchical Bayesian bimodal model is developed to fit observed vehicle travel time data on urban streets. Since the bimodal travel time distributions capture the interrupted nature of traffic flows on urban streets more accurately, it offers a better model for a variety of applications. We demonstrate empirically that the bimodal model provides useful information for analyzing traffic operations and that

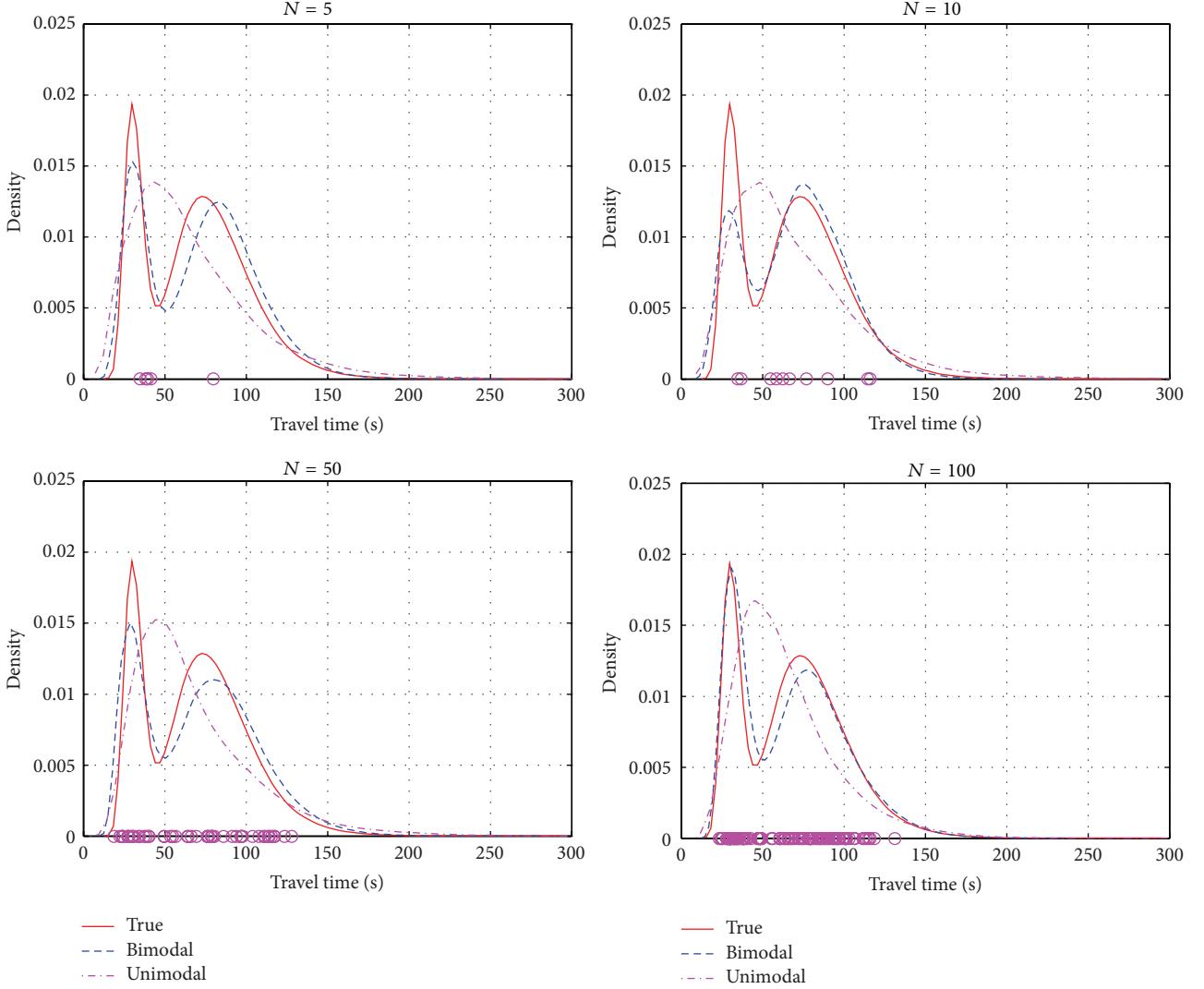


FIGURE 5: Estimated travel time distributions in a hypothetical day on Link 1 in a.m. period.

it provides higher estimation accuracy of the travel time distribution than does the unimodal model.

This study can be extended in several directions. For example, this study considers probe data alone to estimate travel time distributions. Considering traffic volume and traffic signal timing as covariate variables in the model would improve the accuracy of the travel time distribution estimation. The effect of some other observable factors, such as bad weather and accidents, could also be considered in an extension of the proposed model.

In addition, this study uses probe bus data to demonstrate the advantage of the proposed methodologies. It would be valuable to evaluate the proposed methodologies using other types of probe vehicle data, such as probe taxi data. Lastly, some other valuable applications based on the proposed model are worth pursuing in the future, such as congestion identification, accident detection, evaluation of coordinated signal control systems, and travel time reliability.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (51308410) and the Fundamental Research Funds for the Central Universities of China (1600219210). The work was also supported by Program for Changjiang Scholars and Innovative Research Team in University. The authors are grateful to Srah Blouch and Chris Kvitya of the OSU Campus Area Bus Service for providing the datasets used in this study, to Mr. Danie E. Moorhead and Mr. Gary J. Holt at the city of Columbus for providing the signal timings at the two intersections of interest, and to Mr. Xudong Hu for conducting a site survey. The authors also appreciate

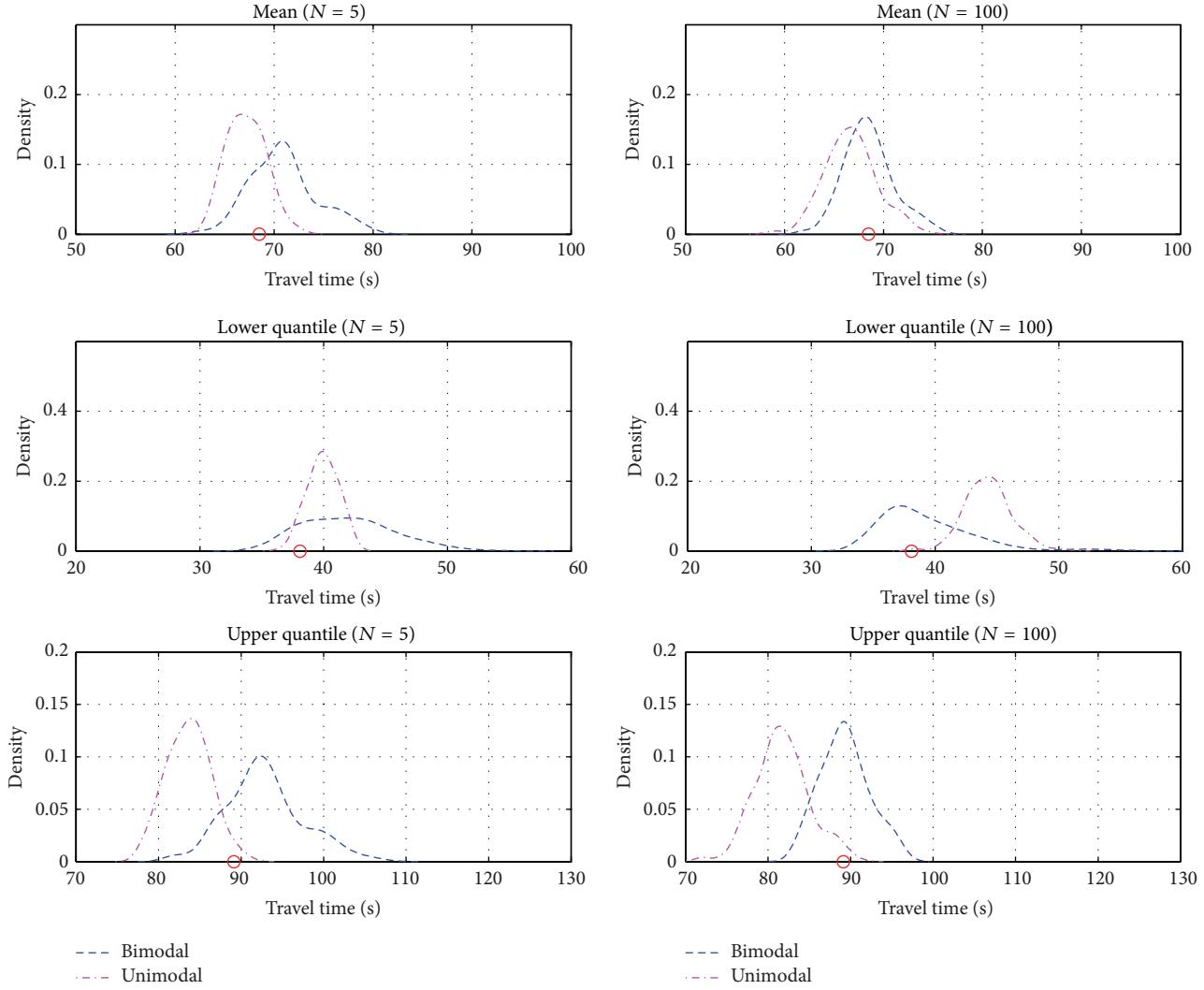


FIGURE 6: Simulated distributions of the mean, lower quantile, and upper quantile of the travel time distributions produced by the bimodal and unimodal models.

the technical support provided by the Campus Transit Lab (CTL) at the OSU.

References

- [1] M. A. P. Taylor and S. Somenahalli, "Travel time reliability and the bimodal travel time distribution for an arterial road," *Road & Transport Research*, vol. 19, no. 4, pp. 37–50, 2010.
- [2] C. Chen, K. Petty, A. Skabardonis, P. Varaiya, and Z. Jia, "Freeway performance measurement system: mining loop detector data," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1748, pp. 96–102, 2001.
- [3] B. Coifman, "Estimating travel times and vehicle trajectories on freeways using dual loop detectors," *Transportation Research Part A: Policy and Practice*, vol. 36, no. 4, pp. 351–364, 2002.
- [4] A. Sen, P. Thakuriah, X.-Q. Zhu, and A. Karr, "Frequency of probe reports and variance of travel time estimates," *Journal of Transportation Engineering*, vol. 123, no. 4, pp. 290–297, 1997.
- [5] D. E. Boyce, A. Kirson, and J. L. Schofer, "Design and implementation of advance: the Illinois dynamic navigation and route guidance demonstration program," in *Proceedings of the Vehicle Navigation & Information Systems Conference*, pp. 415–426, Dearborn, Mich, USA, October 1991.
- [6] C. A. Quiroga and D. Bullock, "Travel time studies with global positioning and geographic information systems: an integrated methodology," *Transportation Research Part C: Emerging Technologies*, vol. 6, no. 1-2, pp. 101–127, 1998.
- [7] "Reliability metrics for path travel time under log-normal distribution," in *Proceedings of the 3rd International Symposium on Transportation Network Reliability*, N.-E. E. Faouzi and M. Maurin, Eds., The Hague, The Netherlands, 2007.
- [8] N. Uno, F. Kurauchi, H. Tamura, and Y. Iida, "Using bus probe data for analysis of travel time variability," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 13, no. 1, pp. 2–15, 2009.
- [9] S. Susilawati, M. Taylor, and S. Somenahalli, Eds., *Modelling Urban Travel Time Variability with the Burr Regression Technique*, Australasian Transport Research Forum, Adelaide, Australia, 2011.

- [10] B. Hellinga and L. Fu, "Assessing expected accuracy of probe vehicle travel time reports," *Journal of Transportation Engineering*, vol. 125, no. 6, pp. 524–530, 1999.
- [11] B. R. Hellinga and L. Fu, "Reducing bias in probe-based arterial link travel time estimates," *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 4, pp. 257–273, 2002.
- [12] W. Pu, J. Lin, and L. Long, "Real-time estimation of Urban street segment travel time using buses as speed probes," *Transportation Research Record*, vol. 2129, pp. 81–89, 2009.
- [13] K. Jintanakul, L. Chu, and R. Jayakrishnan, "Bayesian mixture model for estimating freeway travel time distributions from small probe samples from multiple days," *Transportation Research Record*, no. 2136, pp. 37–44, 2009.
- [14] F. Guo, H. Rakha, and S. Park, "Multistate model for travel time reliability," *Transportation Research Record*, no. 2188, pp. 46–54, 2010.
- [15] F. Zheng and H. Van Zuylen, "Uncertainty and predictability of urban link travel time: delay distribution-based analysis," *Transportation Research Record*, no. 2192, pp. 136–146, 2010.
- [16] F. Zheng and H. van Zuylen, "Modeling variability of urban travel times by analyzing delay distribution for multiple signalized intersections," *Transportation Research Record*, vol. 2259, pp. 80–95, 2011.
- [17] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2nd edition, 2004.
- [18] "Jags: a program for analysis of Bayesian graphical models using gibbs sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, M. Plummer, Ed., Vienna, Austria, 2003.
- [19] X.-L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ECM algorithm: a general framework," *Biometrika*, vol. 80, no. 2, pp. 267–278, 1993.
- [20] F. W. Cathey and D. J. Dailey, "Estimating corridor travel time by using transit vehicles as probes," *Transportation Research Record*, vol. 1855, pp. 60–65, 2003.
- [21] B. Coifman and S. Kim, "Measuring freeway traffic conditions with transit vehicles," *Transportation Research Record*, no. 2121, pp. 90–101, 2009.
- [22] SAS Intitute Inc, 'Sas/Stat 12.1 User's Guide', SAS Intitute Inc., Cary, NC, USA, 2012.

