1-1-2017

Bayesian Nonparametric Model for Estimating Multistate Travel Time Distribution

Emmanuel Kidando
Cleveland State University, e.kidando@csuohio.edu

Ren Moses
FAMU-FSU College of Engineering

Eren E. Ozguven
FAMU-FSU College of Engineering

Thobias Sando
University of North Florida

Hindawi
Journal of Advanced Transportation
Volume 2017, Article ID 5069824, 9 pages
https://doi.org/10.1155/2017/5069824

Research Article

Bayesian Nonparametric Model for Estimating Multistate Travel Time Distribution

Emmanuel Kidando,[1] Ren Moses,[1] Eren E. Ozguven,[1] and Thobias Sando[2]

[1]*Department of Civil and Environmental Engineering, FAMU-FSU College of Engineering, Tallahassee, FL, USA*
[2]*School of Engineering, University of North Florida, Jacksonville, FL, USA*

Correspondence should be addressed to Emmanuel Kidando; ek15f@my.fsu.edu

Received 15 October 2016; Revised 18 December 2016; Accepted 28 December 2016; Published 20 February 2017

Academic Editor: Yuchuan Du

Multistate models, that is, models with more than two distributions, are preferred over single-state probability models in modeling the distribution of travel time. Literature review indicated that the finite multistate modeling of travel time using lognormal distribution is superior to other probability functions. In this study, we extend the finite multistate lognormal model of estimating the travel time distribution to unbounded lognormal distribution. In particular, a nonparametric Dirichlet Process Mixture Model (DPMM) with stick-breaking process representation was used. The strength of the DPMM is that it can choose the number of components dynamically as part of the algorithm during parameter estimation. To reduce computational complexity, the modeling process was limited to a maximum of six components. Then, the Markov Chain Monte Carlo (MCMC) sampling technique was employed to estimate the parameters' posterior distribution. Speed data from nine links of a freeway corridor, aggregated on a 5-minute basis, were used to calculate the corridor travel time. The results demonstrated that this model offers significant flexibility in modeling to account for complex mixture distributions of the travel time without specifying the number of components. The DPMM modeling further revealed that freeway travel time is characterized by multistate or single-state models depending on the inclusion of onset and offset of congestion periods.

1. Introduction

Modeling travel time distribution is essential for measuring the consistency of the traffic performance of a highway system. Moreover, the distribution of the travel time is useful in simulation and theoretical derivations regarding different traffic performance measures such as travel time reliability and variability. The accurate estimation and prediction of travel time are essential for traffic operators, planners, and traveler information systems [1].

This study develops a nonparametric Bayesian model to estimate the travel time distribution for freeways. The model is based on Dirichlet process distribution with an extension of a hierarchical structure to account for the mixture/multistate characteristics of a given dataset. During the modeling process, the proposed model is truncated with an upper bound of six mixture components to reduce computational cost. Unlike a parametric model, this model does not require specifying the true number of components; instead, the number of components grows with the dataset, which is automatically inferred using the Bayesian posterior inference framework. The posterior distributions of the model parameter are derived using the Metropolis-Hastings Markov Chain Monte Carlo (MCMC) sampler. For this study, an Interstate 295 freeway corridor located in Jacksonville, Florida, was studied using 2015 traffic data.

In the next section, review of relevant studies is undertaken, followed by the methodology framework used in this research. Then, the discussion of the dataset and a method used to estimate the travel time is presented. Next, the results and model evaluation using simulated data with known parameters is displayed, after which conclusions and recommendations for possible future research are made.

2. Literature Review

Literature review indicates that models of estimating the travel time distribution can be divided into two groups,

that is, single probability (unimodal) and multistate/mixture models. Unimodal distributions commonly used to estimate the travel time distribution are Gaussian, lognormal, gamma, Weibull, and Burr [2]. Findings from several comparative studies of unimodal distribution functions suggest that travel time distribution is skewed, which makes lognormal, gamma, Burr, and Weibull more accurate than the Gaussian distribution in modeling travel time distribution. For example, using hourly-based data, Kieu et al. [2] compared Gaussian, lognormal, gamma, Burr, and Weibull models and concluded that the lognormal function fits the travel time distribution better than the rest of the models. Similar findings are reported by Arroyo and Kornhauser [3], Rakha et al. [4], and Emam and Al-Deek [5]. On the other hand, Pu [6] reported that, during congested and free flow conditions, travel time distribution is close to symmetrical, suggesting the Gaussian distribution of travel time. However, at the onset and offset of the congestion, the distribution is skewed. The study by Pu [6] suggested that lognormal distribution fits these conditions well.

The multistate/mixture models refer to models comprising two or more distributions. In mixture modeling, the individual distribution forming the mixture is linearly added using a weighted sum of the individual distribution contributing to the model. The weights refer to the mixing probabilities of the model. Studies comparing the performance of mixture models to single models revealed that mixture models provide a superior fit of travel time distribution over single models [1, 7–9]. Using field data collected on the Interstate I-35 freeway in San Antonio, Texas, Guo et al. [7] compared different multistate models. The outcomes were that the lognormal multistate distribution outperforms the rest of the models in modeling travel time distribution. This finding is consistent with results by Yang and Wu [10]. As a result, our study also adopts lognormal distribution in the analysis. It should be understood that, with the same road geometric characteristics (e.g., lane width, pavement condition, posted speed limit, and the number of lanes), the multistate characteristic of travel time is attributed to different vehicle type, traffic conditions, incidents, and driving characteristics on freeways. In addition to the previously mentioned factors, arterial roads are influenced by signal light, conflicts with pedestrians, and other factors [9, 11, 12].

In multistate modeling, there are two commonly used methods for finding model parameters, that is, the maximum likelihood estimation-expectation maximization (MLE-EM) and the Bayesian approach (BA) [13]. The MLE-EM method treats components of the mixture as missing variables and iteratively alternates between the E-step and the M-step to find the parameters of the model [14]. In addition, the method uses random initial guess and, after sufficient iterations, parameters converge. Compared to the BA, the MLE-EM method is computationally less expensive. However, it is susceptible to local maxima trap problem, which could result in overfitting of the resulting model [14]. Unlike the MLE-EM estimation method, the BA treats the model parameters as distributions that can be updated after new data become available. The BA method also incorporates prior knowledge regarding travel time distribution [15], which can be obtained from previously observed characteristics of the data distribution. Moreover, studies indicate that, by using informative priors, the BA can estimate the posterior distributions with smaller number of sample sizes than the MLE-EM approach [15, 16].

Taken together, the probability distributions discussed above are parametric with either the single model or multistate characteristics, whereby the multistate model consists of a fixed number of mixture components. The number of mixture components is specified as input in the model. The information criterion, cross-validation, and Bayesian factor are procedures commonly used to select the best model among a set of candidates [13]. However, these procedures for selecting the best model sometimes result in the output model suffering from over- or underfitting problem, depending on the amount of data available and on the model bound complexity [17, 18].

However, there are two methods that can be used in modeling without causing overfitting or underfitting problems. The use of the infinite Dirichlet Process Mixture Model (DPMM) with a truncated number of mixture components overcomes the underfitting problem [17–20]. The overfitting problem can be overcome by the use of a BA to estimate the posterior distribution of the parameters [18]. In this study, both DPMM and BA were used in modeling the travel time distribution. As indicated above, the infinite DPMM was selected. The infinite number of mixture components is achieved through the application of the stick-breaking process in building mixing weight of the mixture. This property of the infinite set of mixture components makes a model to be considered as a typical nonparametric model [21, 22]. Although the model is taken as infinite, only a few nonempty components are drawn depending on the actual characteristics of the dataset given [23]. Generally, the nonempty components are less than the realized number of the sample sizes considered in the analysis.

The Bayesian nonparametric mixture models have been implemented in a wide range of applications, including topic modeling, image analysis, and lifetime distribution [21, 24–26]. The attractiveness of Bayesian nonparametric mixture models includes the ability to handle randomness of the mixing distribution of a noisy dataset. The randomness of the mixing component is estimated using infinite dimension priors, whereby during sampling, true mixture components are built automatically and the rest die out. This study constructed priors using the stick-breaking process [21]. This process represents an infinite discrete distribution with the probability of being repeated from the previous draws. This characteristic makes the stick-breaking process appropriate for clustering data with multistate characteristics. However, controlling infinite dimensional posterior distribution can be computationally expensive [27]. To reduce this problem, literature suggests the use of truncated dimension priors to reduce computational complexity [27].

3. Model Framework

The Dirichlet distribution is the generalization of a Beta distribution to account for higher order outcomes. The

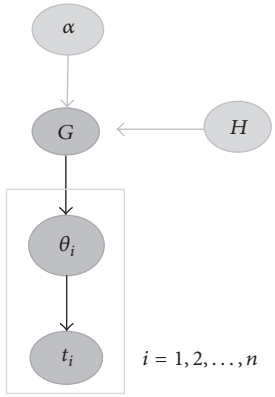


Figure 1: Graphical Dirichlet Mixture Model.

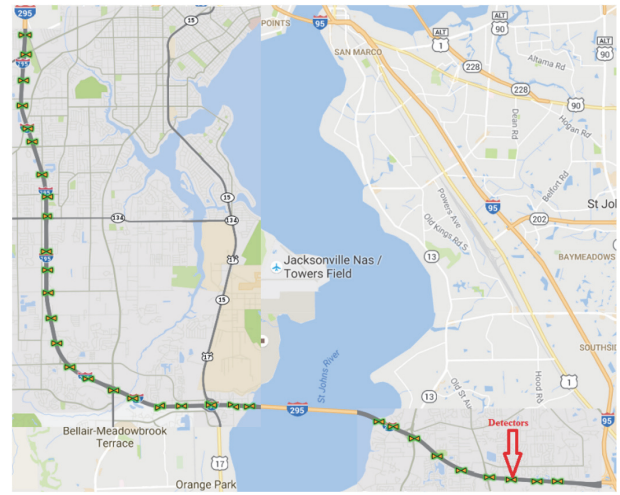


Figure 2: The study corridor.

distribution is parameterized by a concentration parameter α and k mixture components. Its probability density function is given by

$$f(x;\alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^{k} w_i^{\alpha_i - 1}, \quad \text{with } B(\alpha) = \frac{\prod_{i=1}^{k} \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}, \; w_1 + w_2 + \cdots + w_k = 1. \tag{1}$$

The definitions of the terms of (1) through (4) are given in the Abbreviations.

The Dirichlet process is described as a set of distributions over the infinite sample space or distributions [21]. A mixture model with a hierarchical structure can be constructed using the Dirichlet process, which is also referred to as the DPMM [21, 28]. Figure 1 shows a graphical representation of the hierarchal mixture model.

The model in Figure 1 can also be mathematically represented as follows:

$$\begin{aligned} t_i \mid \theta_i &\sim \text{LN}(t \mid \theta_i) \quad \text{for } 1 = 1, 2, 3, \ldots, n, \\ \theta_i \mid G &\sim G, \\ G \mid \alpha, H &\sim \text{DP}(\alpha, H), \end{aligned} \tag{2}$$

$$G = \sum_{k=1}^{\infty} \left(w_k^* \delta_{\theta_k^*}\right) \sim \text{DP}(\alpha, H), \quad \text{with } \sum_{k=1}^{\infty} \left(w_k^*\right) = 1. \tag{3}$$

In this study, the above model is implemented using the SBP, which involves breaking a unit length stick into infinite disjoint pieces repeatedly [20]. The initial break, $k = 1$, is determined randomly with a probability v_1, which is considered as the probability of the first mixture component. After the first break, the next break, $k = 2$, has the probability $(1 - v_1) * v_2$. The process of breaking continues until the infinite number of groups is created [22]. To reduce the computational complexity of the model, the breaking process can be truncated to $k = n$ groups. In this study, $k = 6$ was selected, which was checked later in the analysis to verify whether the truncation process did not bias the results of mixture component of our dataset. In particular, the highest mixture components used by the data were identified in the probability of the mixture component matrix. Recalling (3), the following conditions apply for the stick-breaking construction process:

$$\begin{aligned} w_k^* &= v_k \prod_{i=1}^{k-1} 1 - v_i, \\ \theta_k^* &\sim H, \\ v_k &\sim \text{Beta}(1, \alpha). \end{aligned} \tag{4}$$

Estimating the posterior distribution of the hierarchical Bayesian model is analytically difficult as it involves high dimensional integral in the marginal likelihood [1]. The common method used for approximating the posterior distribution of the model parameters is the MCMC simulation. In this study, we apply also the MCMC simulation to estimate the posterior distribution of the unknown parameters. In particular, we adopted Metropolis-Hastings sampling step through PyMC3, an open source package for approximating the posterior distribution of model parameters [29]. The Metropolis-Hastings sampling step uses the acceptance probability to draw a sample from the proposed posterior distribution [29]. The priors for distributions are taken as noninformative with Gamma(1, 1) for concentration parameter α, Normal(μ_1, σ_1) and HalfCauchy(0, 1) for mean μ, and sigma, σ, respectively. On the other hand, the hyperpriors for hyperparameters μ_1 and σ_1 are Normal(0, 0.001) and HalfCauchy(0, 1), respectively.

4. Study Data and Travel Time Estimation

For this study, data from the 20.4-mile corridor of the Interstate 295 freeway (Figure 2) in Jacksonville, Florida, were acquired. The corridor was divided into nine links running between interchanges. Each link had 65 miles per hour (mph) posted speed limit. The archived traffic data for analysis

Table 1: Summary of links.

Link ID	Location	Distance (miles)	Number of detectors
1	From I-95 to Old St. Augustine Rd	2.8	4
2	San Jose Blvd	1.6	4
3	Park Ave.	4.8	5
4	Blanding Blvd	2.0	3
5	Collins Rd	1.1	2
6	103rd St.	3.2	4
7	Wilson Blvd.	1.5	4
8	Normandy Blvd.	2	4
9	I-10	1.2	1

were provided by the Regional Integrated Transportation Information System (RITIS), which is comprised of speed data from microwave vehicle detectors (MVDs) aggregated at a 5-minute interval. The five-minute interval was selected to avoid fluctuation on short-duration travel time [10]. The data gathered were collected for the period of January 1, 2015, through December 31, 2015. Weekend and days in which incidents (crashes, work zones, etc.) happened were omitted from the dataset to reduce variability. If a link had more than one MVD in a lane, the average speed from the MVD was calculated and used to represent the link travel speed. Except for Link 9, other links have at least two MVDs in each lane (Table 1).

The travel time of each link was estimated using the average speed from the traffic speed reported by all MVDs in a segment. In addition, time of a day was an important parameter in the analysis. The travel time of a segment i at a time t is computed by the following equation:

$$\text{Travel time}\left(t_{i,t}\right) = \frac{n_i l_i}{\sum_{j=1}^{n} u_{i,j,t}}, \tag{5}$$

where n_i represents the number of the detectors on link i, l_i is the length of segment i, and $u_{i,j,t}$ is the speed reported by the MVD j on a segment i at time t.

We considered the same departure time in estimating the corridor travel time from individual link's travel time. By aggregating the travel time, the results showed that the morning peak hour for both directions (northbound and southbound) occurred between 7 a.m. and 8 a.m. while the evening peaking hour occurred between 5 p.m. and 6 p.m. Figure 3 shows the travel times plotted against time for the day for both the northbound and the southbound traffic. The data in Figure 3 reveal that southbound traffic frequently experiences longer travel times than northbound traffic, particularly during the morning peak hours.

5. Results and Discussion

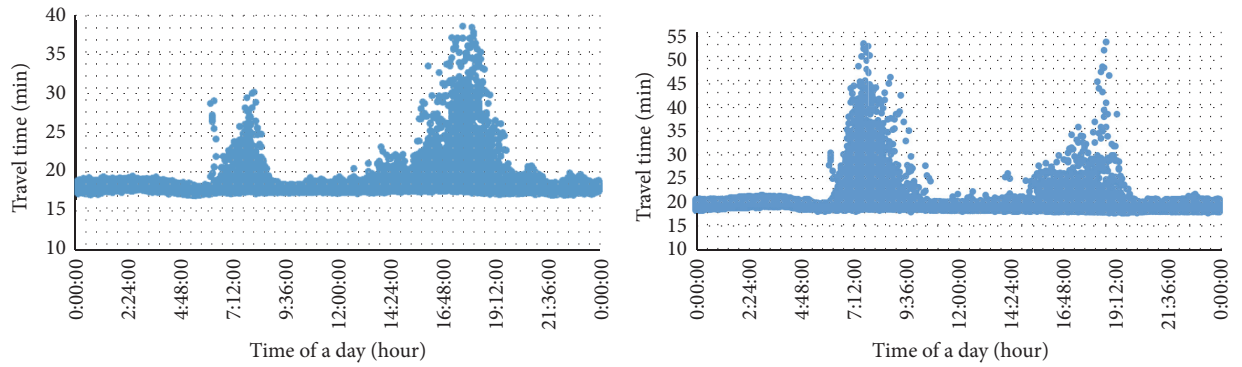
Two chains were drawn and the first 10,000 iterations were discarded as burn-in while the next 10,000 iterations were used for inference. To reduce correlations between drawn samples, the sequence of inference iterations were thinned by 10 iterations. Figure 4 presents predicted and actual data densities of some of the hours considered in the analysis. As shown in the figure, the proposed model provided a good fit such that actual and predicted probability densities are close. Furthermore, the quantitative test using the Kolmogorov-Smirnov (KS) goodness-of-fit was conducted testing the hypothesis whether the predicted and actual distributions are the same. The null hypothesis for the test is that the actual cumulative density of the travel time is equal to the predicted density. Results of analysis confirmed that the predicted cumulative travel time follows the empirical cumulative density (p value ≥ 0.05). Figure 5 compares cumulative predicted and empirical cumulative density. Table 2 depicts the number of mixture components, model parameters, and KS test of the predicted travel time distribution for some of the hours. The results from this table reveal that the truncation process of the mixture components using the maximum of six (6) did not bias the results of the dataset. The highest mixture component of the dataset was revealed at 3 mixture components.

As can be seen in Table 2, the travel time distributions in the northbound are predominantly two mixture components with two hours containing one component, while in the southbound, the distribution shows one, two, and three mixture components. However, the third component of the three components' distribution has a very low likelihood, less than 0.1.

5.1. Model Evaluation. To understand the performance of the DPMM in estimating the distribution of the travel time, four finite mixture models (i.e., single, two, and three mixture models) were simulated. The simulation was aimed at evaluating the accuracy of the models given the known parameters. The simulation was conducted with the known mean and variance, which were chosen randomly from link's average and variance of the travel time data. Subsequently, the true parameters were used to simulate various sample sizes including 100, 1,000, and 10,000 following the lognormal distribution with the predefined finite mixture. The reason for simulating different sample sizes was to evaluate the influence of sample size on the proposed model. The truncated DPMM with 6 numbers of components was applied to each sample data. Discarding first 10,000 iterations, the next 10,000 iterations were considered for inference of the model parameters. Table 3 illustrates the true and predicted parameters. According to this table, the number of mixture components, the mean, and the variance converged closer to the true parameters. Comparing the true to the predicted values, the results are promising as the number of components is predicted accurately while some of the data mean, the standard deviation, and mixing probability are somewhat deviating from the true parameters.

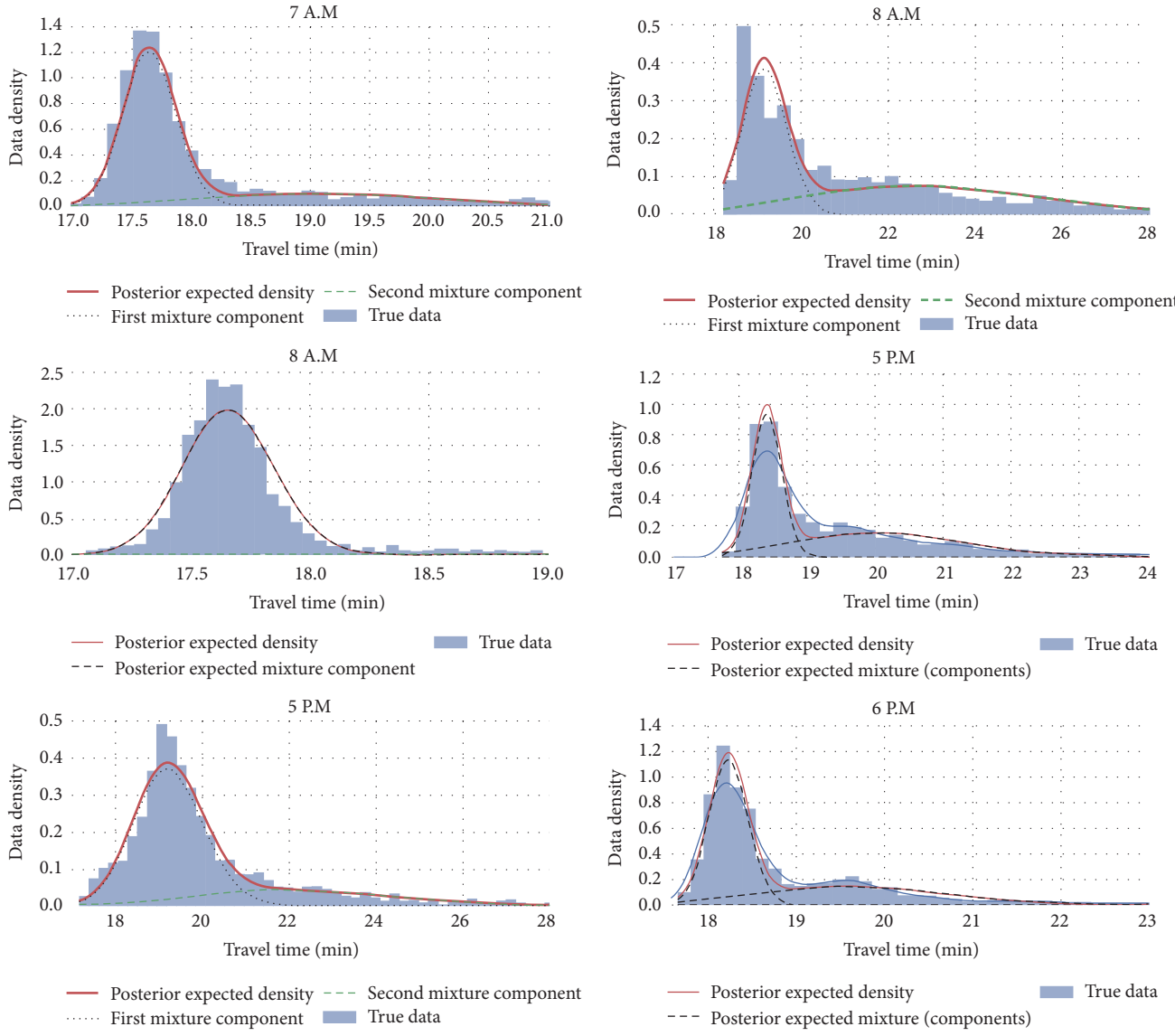
Regardless of the number of observations, the number of mixture components was predicted correctly. The true and the predicted distributions are plotted in Figure 6. The distributions predicted by the DPMM model are close to the true distributions, suggesting that this model can sufficiently approximate any unknown mixture component.

In addition, similar to travel time distribution goodness-of-fit test, the KS test was conducted to compare the actual



(a) Northbound traffic (b) Southbound traffic

Figure 3: Time of the day corridor travel times.



(a) Northbound traffic (b) Southbound traffic

Figure 4: Predicted distribution and actual data density.

Table 2: Corridor parameters and KS goodness-of-fit.

Time of day	Parameters			KS goodness-of-fit test	
	Number of components (mixture probability%)	Mean	Standard deviation	KS test stats	KS p value
Northbound (from I-95 to I-10)					
6:00 a.m.	2 (53, 47)	(2.89, 3.09)	(0.02, 0.14)	0.093	0.32
7:00 a.m.	2 (75, 25)	(2.87, 2.95)	(0.01, 0.05)	0.051	0.73
8:00 a.m.	2 (88, 12)	(2.87, 2.94)	(0.01, 3.19)	0.074	0.43
9:00 a.m.	1 (1)	(2.87)	(0.01)	0.079	0.99
10:00 a.m.	1 (1)	(2.88)	(0.01)	0.090	0.68
15:00 p.m.	2 (78, 22)	(2.88, 2.92)	(0.01, 0.04)	0.078	0.40
16:00 p.m.	2 (87, 13)	(2.93, 3.01)	(0.03, 0.07)	0.067	0.99
17:00 p.m.	2 (75, 25)	(2.96, 3.09)	(0.10, 0.13)	0.027	0.96
18:00 p.m.	2 (60, 40)	(2.89, 3)	(0.03, 0.08)	0.049	0.52
19:00 p.m.	2 (94, 6)	(2.87, 3)	(0.02, 0.08)	0.125	0.45
20:00 p.m.	2 (96, 4)	(2.87, 2.9)	(0.01, 0.06)	0.046	0.98
Southbound (from I-10 to I-95)					
6:00 a.m.	2 (70, 30)	(2.94, 3.09)	(0.02, 0.12)	0.070	0.07
7:00 a.m.	1 (1)	(3.25)	(0.17)	0.037	0.35
8:00 a.m.	2 (52, 48)	(2.96, 3.14)	(0.03, 0.12)	0.064	0.06
9:00 a.m.	2 (77, 23)	(2.94, 3.13)	(0.01, 0.83)	0.078	0.29
10:00 a.m.	1 (1)	(2.94)	(0.02)	0.108	0.17
15:00 p.m.	2 (68, 32)	(2.92, 2.96)	(0.01, 0.03)	0.083	0.31
16:00 p.m.	3 (49, 43, 8)	(2.95, 2.94, 1.38)	(0.04, 0.02, 0.09)	0.938	0.86
17:00 p.m.	2 (50, 50)	2.91, 3	(0.01, 0.07)	0.061	0.28
18:00 p.m.	2 (63, 37)	(2.9, 2.98)	(0.01, 0.05)	0.062	0.43
19:00 p.m.	3 (88, 2, 10)	(2.91, 2.72, 2.2)	(0.02, 0.86, 0.23)	0.081	0.27
20:00 p.m.	1 (1)	(2.93)	(0.2)	0.082	0.41

Table 3: Parameters of the study.

ID	True parameters	Predicted parameters
a	$0.51 * \mathrm{LN}(2.5, 0.09) + 0.49 * \mathrm{LN}(3.2, 0.2)$, $N = 1{,}000$	$0.52 * \mathrm{LN}(2.5, 0.09) + 0.48 * \mathrm{LN}(3.2, 0.19)$
b	$\mathrm{LN}(1.13, 0.19)$, $N = 200$	$\mathrm{LN}(1.14, 0.18)$
c	$\mathrm{LN}(1.13, 0.19)$, $N = 1{,}000$	$\mathrm{LN}(1.14, 0.19)$
d	$0.66 * \mathrm{LN}(1.13, 0.19) + 0.34 * \mathrm{LN}(0.89, 0.03)$, $N = 200$	$0.68 * \mathrm{LN}(1.13, 0.18) + 0.32 * \mathrm{LN}(0.89, 0.02)$
e	$0.66 * \mathrm{LN}(1.13, 0.19) + 0.34 * \mathrm{LN}(0.89, 0.03)$, $N = 1000$	$0.69 * \mathrm{LN}(1.13, 0.19) + 0.31 * \mathrm{LN}(0.89, 0.03)$
f	$0.42 * \mathrm{LN}(0.6, 0.04) + 0.45 * \mathrm{LN}(1, 0.01) + 0.13 * \mathrm{LN}(0.01, 0.07)$, $N = 100$	$0.44 * \mathrm{LN}(0.6, 0.04) + 0.40 * \mathrm{LN}(1, 0.01) + 0.16 * \mathrm{LN}(0.11, 0.05)$

cumulative distributions against those predicted by the DPMM. The results from analysis suggest that there is no evidence to reject the null hypothesis indicating that the predicted probability density follows the observed data.

As indicated in Table 4, the p value for each considered sample is greater than 0.05, suggesting that there is no significant difference between the distribution of the predicted and actual data.

6. Conclusions and Recommendations for Future Research

This study evaluated the application of a nonparametric Bayesian mixture model with the truncated DPMM through lognormal kernel density to estimate travel time distribution. The model developed here extends the commonly used mixture models to incorporate uncertainty about the number

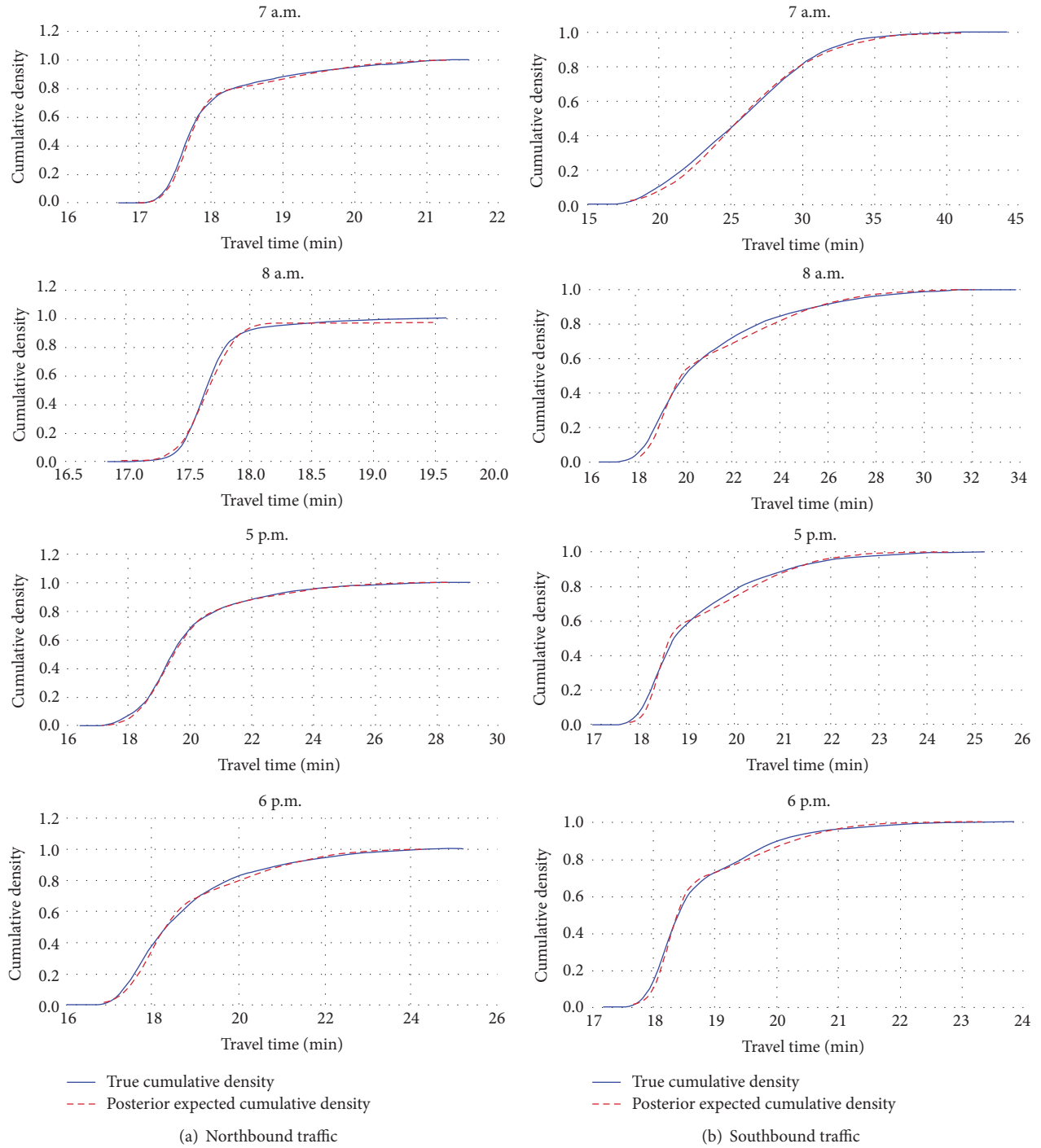


Figure 5: Cumulative predicted distribution and data cumulative density.

of mixture components of the model. In the DPMM, the number of components and the parameters of the travel time distribution were considered as random numbers. One-year spot speed data collected from a 20.4-mile corridor of the Interstate 295 freeway in Jacksonville, Florida, was used in the study. The peak and nonpeak hour travel times were aggregated at 5-minute intervals using data from MVD installed in various links in the corridor.

The findings have demonstrated that the developed model is capable of modeling the travel time distribution. Moreover, the results of the model support previous studies that travel time distribution is characterized by both multi-state and single-state model depending on the time window of the analysis. Furthermore, the results demonstrated that the proposed model can offer significant flexibility in modeling to account for complex mixture distributions of the travel

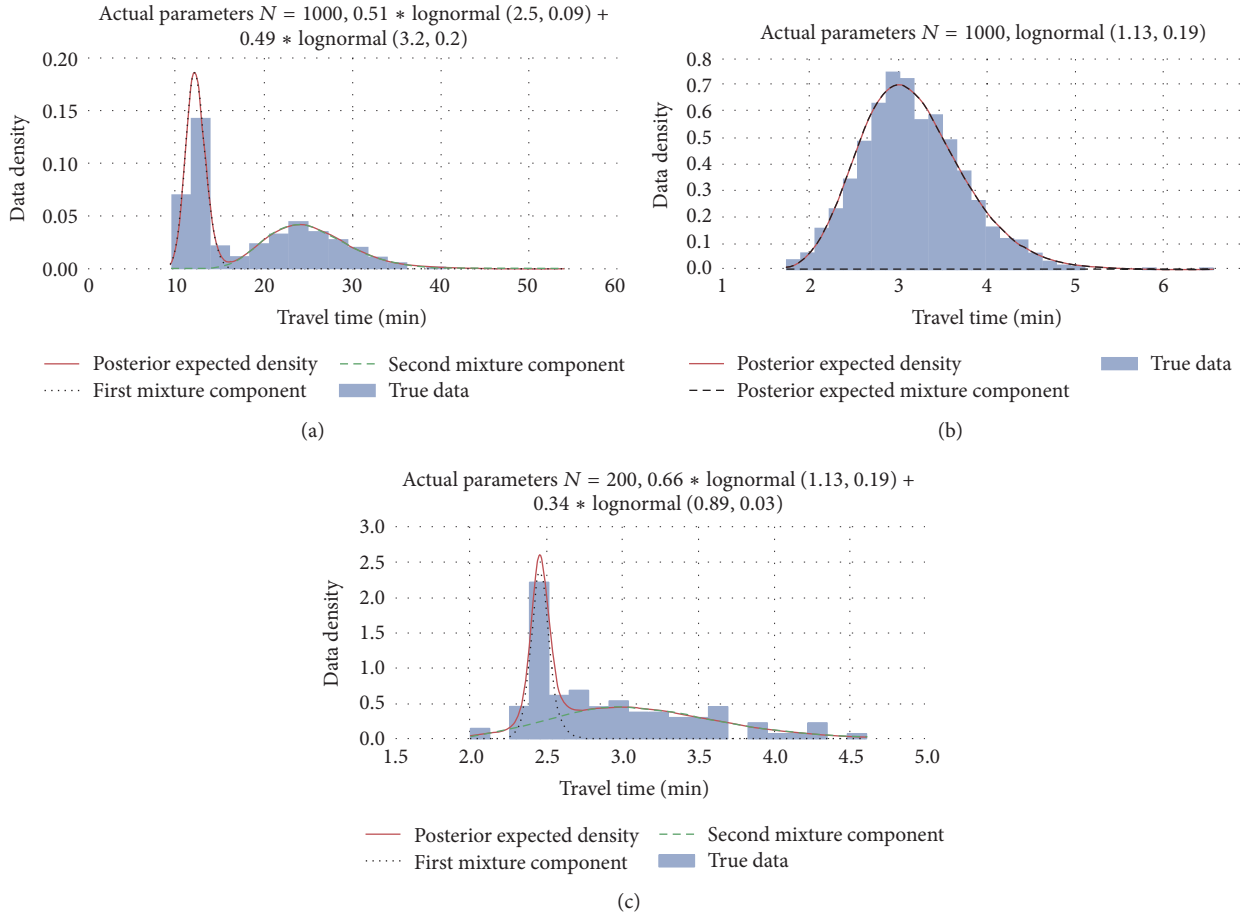


Figure 6: Estimated travel time distributions of the simulated data.

Table 4: The Kolmogorov-Smirnov goodness-of-fit test of the cumulative density.

ID	Test stats	p value
a	0.015	0.999
b	0.083	0.624
c	0.026	0.999
d	0.035	0.999
e	0.028	0.999
f	0.046	0.999

time without specifying the number of components. In the analysis, the uncertainties related to the number of mixture components were incorporated as well. The performance of the model based on the KS test on the actual and predicted cumulative probability density revealed promising results. Moreover, while testing the proposed model using simulated data, the number of true mixture components, mean, and the standard deviation value were correctly predicted.

It is important to note that in this study the travel time for the corridor was aggregated using the same departure time. This process may not represent the actual travel time of the corridor. Future studies may consider a vehicle trajectory-based method, dynamic time slice methods, or other methods to aggregate travel time across links. In addition, future studies could aim at analyzing and comparing the finite mixture and nonparametric mixture models using different sample sizes and other kernel functions such as gamma and normal distributions.

Abbreviations

DP(α, H): The random probability density function coming from the Dirichlet distribution with parameters α and H
H: The base measure
α: Concentration parameter
G: The random distribution drawn from the Dirichlet process DP(α, H)
θ_n: The parameter of G distribution which follows a stick-breaking process (SBP)
w_i: The nonnegative vector representing a probability mass function of length k
w_n^*: The mixing proportion
$\delta_{\theta_k^*}$: A Dirac delta function concentrated at θ_k
LN: The lognormal kernel distribution function with a parameter θ_i

k: The number of mixture components, usually equal to or less than a total number of realizations
t: Travel time
$\Gamma(x)$: The gamma function.

Competing Interests

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

References

[1] Y. Ji and H. M. Zhang, *Travel Time Distributions on Urban Streets: Their Estimation with a Hierarchical Bayesian Mixture Model and Application to Traffic Analysis Using High-Resolution Bus Probe Data*, Transportation Research Board, Washington, DC, USA, 2013.

[2] L.-M. Kieu, A. Bhaskar, and E. Chung, "Public transport travel-time variability definitions and monitoring," *Journal of Transportation Engineering*, vol. 141, no. 1, Article ID 04014068, 2015.

[3] S. Arroyo and A. L. Kornhauser, "Modeling travel time distributions on a road network," in *Proceedings of the 11th World Conference on Transport Research*, Berkeley, Calif, USA, 2007.

[4] H. Rakha, I. El-Shawarby, and M. Arafeh, "Trip travel-time reliability: issues and proposed solutions," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 14, no. 4, pp. 232–250, 2010.

[5] E. B. Emam and H. Al-Deek, "Using real-life dual-loop detector data to develop new methodology for estimating freeway travel time reliability," *Transportation Research Record*, no. 1959, pp. 140–150, 2006.

[6] W. Pu, *Analytic Relationships between Travel Time Reliability Measures*, Transportation Research Board, Washington, DC, USA, 2010.

[7] F. Guo, Q. Li, and H. Rakha, "Multistate travel time reliability models with skewed component distributions," *Transportation Research Record*, vol. 2315, pp. 47–53, 2012.

[8] Y. Ji, S. Jiang, Y. Du, and H. M. Zhang, "Estimation of bimodal urban link travel time distribution and its applications in traffic analysis," *Mathematical Problems in Engineering*, vol. 2015, Article ID 615468, 11 pages, 2015.

[9] P. Chen, K. Yin, and J. Sun, "Application of finite mixture of regression model with varying mixing probabilities to estimation of urban arterial travel times," *Transportation Research Record*, vol. 2442, pp. 96–105, 2014.

[10] S. Yang and Y. Wu, *Moving Ahead to Mixture Models for Fitting Freeway Travel Time Distributions and Measuring Travel Time Reliability*, Research Board, Washington, DC, USA, 2016.

[11] S. Susilawati, M. A. Taylor, and S. V. Somenahalli, "Travel time reliability and the bimodal travel time distribution for an arterial road," *Road and Transport Research*, vol. 19, no. 4, pp. 37–50, 2010.

[12] J. Klayut, L. Chu, and A. R. Jayakrishnan, "Bayesian mixture model for estimating freeway travel time distributions from small probe samples from multiple days," *Journal of the Transportation Research Board*, vol. 2136, no. 5, 2014.

[13] B.-J. Park, Y. Zhang, and D. Lord, "Bayesian mixture modeling approach to account for heterogeneity in speed data," *Transportation Research Part B: Methodological*, vol. 44, no. 5, pp. 662–673, 2010.

[14] N. Wan, G. Gomes, A. Vahidi, and R. Horowitz, *Prediction on Travel-Time Distribution for Freeways Using Online Expectation Maximization Algorithm*, Transportation Research Board, Washington, DC, USA, 2014.

[15] Y. Feng, J. Hourdos, and G. A. Davis, *Bayesian Model for Constructing Arterial Travel Time Distributions Using GPS Probe Vehicles*, Transportation Research Board, Washington, DC, USA, 2011.

[16] S. Yang and Y. Wu, "Mixture models for fitting freeway travel time distributions and measuring travel time reliability," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2594, pp. 95–106, 2016.

[17] M. Boullé, R. Guigourès, and F. Rossi, "Nonparametric hierarchical clustering of functional data," in *Advances in Knowledge Discovery and Management*, vol. 527 of *Studies in Computational Intelligence*, pp. 15–35, Springer, 2014.

[18] Y. W. Teh, "Dirichlet processes," in *Encyclopedia of Machine Learning*, Springer, 2010.

[19] N. Sharif-Razavian and A. Zollmann, "An overview of nonparametric bayesian models and applications to natural language processing," *Science*, pp. 71–93, 2008.

[20] N. B. Wentao Fan, "Variational learning of Dirichlet process mixtures of generalized Dirichlet distributions and its applications," in *Proceedings of the 8th International Conference on Advanced Data Mining and Applications (ADMA '12)*, Nanjing, China, 2012.

[21] A. Ahmed and E. Xing, "Dynamic non-parametric mixture models and the recurrent Chinese restaurant process: with applications to evolutionary clustering," in *Proceedings of the 8th SIAM International Conference on Data Mining Applied Mathematics 130*, pp. 219–230, SIAM, Atlanta, Ga, USA, April 2008.

[22] W. Fan and N. Bouguila, "Variational learning of a Dirichlet process of generalized Dirichlet distributions for simultaneous clustering and feature selection," *Pattern Recognition*, vol. 46, no. 10, pp. 2754–2769, 2013.

[23] M. Medvedovic and S. Sivaganesan, "Bayesian infinite mixture model based clustering of gene expression profiles," *Bioinformatics*, vol. 18, no. 9, pp. 1194–1206, 2002.

[24] P. Mokhtarian, M.-R. Namzi-Rad, T. K. Ho, and T. Suesse, "Bayesian nonparametric reliability analysis for a railway system at component level," in *Proceedings of the IEEE International Conference on Intelligent Rail Transportation (IEEE ICIRT '13)*, pp. 197–202, Beijing, China, September 2013.

[25] C.-H. Yang, T. Yuan, W. Kuo, and Y. Kuo, "Non-parametric Bayesian modeling of hazard rate with a change point for nanoelectronic devices," *IIE Transactions*, vol. 44, no. 7, pp. 496–506, 2012.

[26] V. Poynor and A. Kottas, "Nonparametric Bayesian inference for mean residual life functions in survival analysis," https://arxiv.org/abs/1411.7481.

[27] J. E. Griffn, "An adaptive truncation method for inference in Bayesian nonparametric models," *Statistics and Computing*, vol. 26, no. 1, pp. 423–441, 2016.

[28] L. Ren, D. B. Dunson, and L. Carin, "The dynamic hierarchical Dirichlet process," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 824–831, Helsinki, Finland, July 2008.

[29] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in Python using PyMC3," *PeerJ Computer Science*, vol. 2, article e55, 2016.