Travel Time Variability Analysis for Bluetooth Sensor Data in Highways

Mohammad Ghavidel, Navid Khademi, Ehsan Bahrami Samani and Sadaf Milani

October 12, 2020

Travel time variability analysis for Bluetooth sensor data in highways

Mohammad Ghavidel[1*], *Navid Khademi*[1], *Ehsan Bahrami*[2], Sadaf Milani[3]

[1]School of Civil Engineering, College of engineering, University of Tehran, Iran
[2]Department of Statistics, Shahid Beheshti University, Iran
[3]School of transportation and logistics, Dalian University of Technology, Dalian, China
[*]Corresponding author: mohamad.ghavidel@ut.ac.ir

Extended abstract submitted for presentation at the
8th International Conference on Transport Network Reliability (Stockholm, 24-26 June, 2020)

1. Introduction

Transportation systems have a complex, random, and dynamic nature. Decision-making behavior of travelers adds to this randomness and complexity. The route, mode, and departure time choices are basically the three main decision-making behavior of travelers which should be taken into account when characterizing the system dynamics (Saedi et al., 2020). Minimizing the perceived travel time is the core objective of travelers' choice-making act (Abkowitz and Engelstein, 1983).
Travel time variability and travel time reliability are the two sides of the same coin. Increasing travel time variability will lead to a decrease in the predictability of travel time, therefore, travel time reliability will also decline. Distribution of travel time data is characterized to mainly assess the variability of travel time. Variation in day-to-day travel time deteriorates the transportation system reliability (Al-Deek and Emam, 2006) by trespassing the user's expectations (Bates et al., 2001, Lam and Small, 2001, Sun et al., 2003). Variability of travel time is time-dependent and differently affects the departure time choice behavior of travelers at different times of the day. The most current researches on pricing strategies are based on the premise that the preferences of travelers depend on the time periods, such as the morning peak hour (e.g. see (Fosgerau and Engelson, 2011).

This research is aimed at providing a comprehensive model based on the division of time data with equal time intervals. This model is called the conditional model function with the random effect function. This model is built around the multiplication of the probability density function of each time interval with a random-effects. The causality of using random effect model is events of each time interval could be have impact on another time interval.

Current state-of-the-art travel time variability researches assume that travel times follow a unimodal distribution during off-peak hour and bimodal distribution such as lognormal and gamma mixture during peak-hour. Particularly, Burr, Weibull, or lognormal distributions are unimodal distributions employed to characterize the system reliability (Rakha et al., 2010, Rakha et al., 2006, Taylor and Susilawati, 2012). Dividing a day time period into constant time frames to pursue the travel time variability is the main technique applied in the previous studies. However, this ignores the interconnected nature of travel time among the different time periods of a day. To bridge this gap, the current study presents a conditional likelihood model which relates the different time windows using a random-effects model.

2. Data description

This study utilizes spatiotemporal travel data to determine the distribution of travel time. Bluetooth Sensor is one of the commonly used methods for recording travel time data. In this study, the data is collected via Bluetooth-based MAC address recognition technology. During the day, for any device whose Bluetooth is turned on (including a computer, tablet, mouse, cell phone, etc.), the identification and information of Bluetooth devices's Mac address is recorded (Aliari and Haghani, 2012). The study routes are illustrated in figure 1. They are the two major highways (Resalat and Modarres) in metropolitan area of Tehran, the capital of Iran. The city of Tehran is a very congested city with a total of 18.64 million trips per day, of which 41.8% are performed by private cars. The data collected in Resalat and Modarres highways are respectively employed to calibrate and validate the travel time estimation model. Data in Resalat highway was collected by 14 Bluetooth sensors over a length of 12.5 km. For Modarres highway, 7 Bluetooth sensors were exploited to collect the data for a length of 7 km. Data collection period was the two months of April and August 2015 which accounts for roughly 15% of the annual traffic volume. Cleansing the collected massive data by Bluetooth sensors is one of the contributions of this study. Travel time deducted from time, and spatial dependent data were recorded by Bluetooth sensor. Travel time is equal to the subtraction of arrival time to the first station from the departure time of the last station. Preparing dataset is generally time-consuming and needs to be checked repeatedly to ensure the accuracy of the recorded data. Only vehicles which pass entire path consider for this case study.

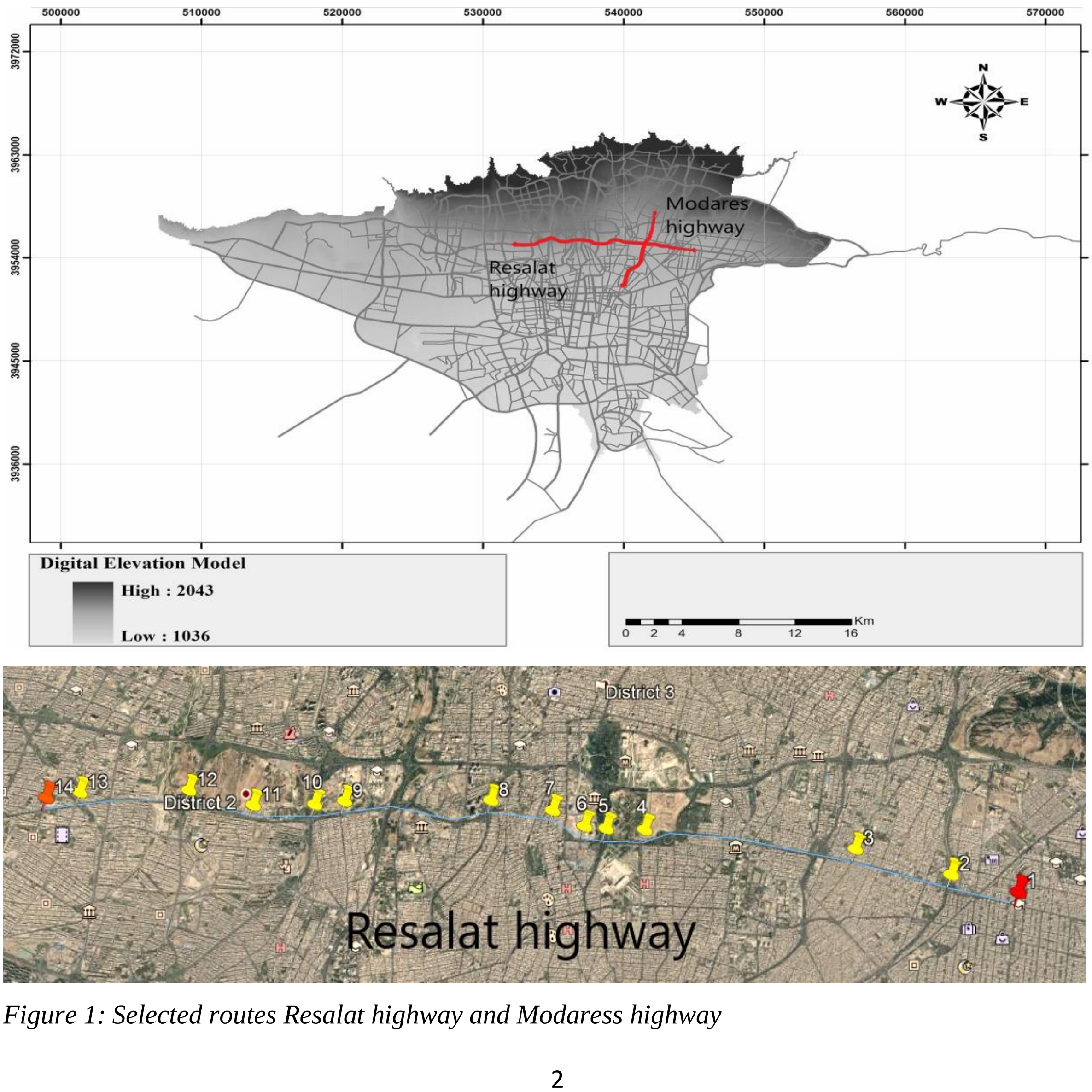

Figure 1: Selected routes Resalat highway and Modaress highway

Figure 2 illustrates the travel time variation for Resalat and Modarres highways. Aggregation in 15 minutes of travel time is performed to plot this graph. The morning and evening peak periods are observed in 6 AM to 10 AM and 4 PM to 8 PM, respectively. The blue graphs show the lower and upper bounds of the travel time collected in the two months of data gathering period. Travel time variation drastically increases in the transition intervals of peak to or from off-peak phases where the level of service of the highway is changed. During the peak hours, both highways operate at the level of service F and a stop-and-go traffic stream is observed. It implies that each car moves in lockstep with the car in front of it, requiring frequent slowing. At this point, a small interruption in downstream causes a large instability along with a significant fluctuating travel time.

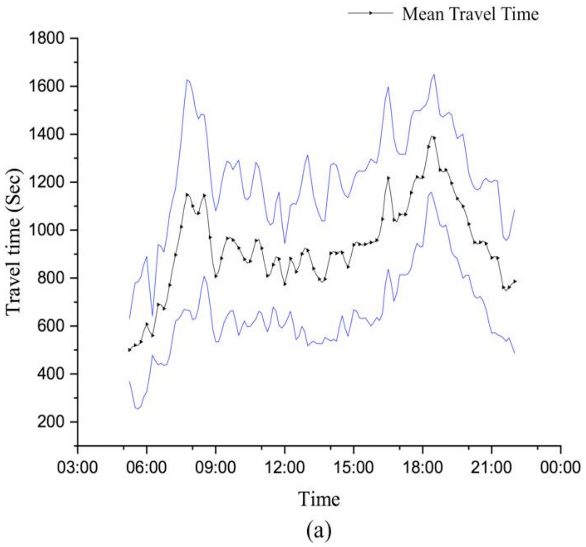

FIGURE 2: Travel time variation throughout a non-holiday day in (a) Resalat and (b) Modarres Highways

3. Travel Time Distribution

This section aims to identify the best density function of travel time distribution for one-hour time windows. The time period between 5 AM and 10 PM (which includes the morning and evening peak periods) are divided in on-hour time spans and multiple density function fits are statistically evaluated. Based on the Bayesian information criterion (BIC) (Schwarz, 1978) indicators at different time periods, the lognormal probability function is selected as the best data distribution descriptor.
First, the two Kolmogorov-Smirnov (KS) and chi-square tests (Stephens, 1974) are employed at the significance level of 95% to evaluate the candidate functions goodness of fit. For the functions which passed these tests, the BIC indicators are identified and the functions are ranked based on them. Table 2 shows the tests results and the best fitted functions for different time windows. Burr distribution passed the KS and chi-squared tests for 9 out of 17 time windows. This distribution shows a perfect fit for off-peak periods when the travel time is medium, positively skewed, and long-tailed. On the other hand, the Weibull distribution is a good fit for peak periods. The Normal distribution is selected only once as the best fitted distribution, while the Gamma and lognormal distributions show a good performance during the entire period of time. The main reason is that the travel time dataset has mild skew and kurtosis and the two distributions of Gamma and lognormal are the best descriptors of data with such a specification. Both distributions are proper tools to model the heavy-tailed and light-tailed data, despite the fact that higher-skewed and longer-tailed data are better represented by the lognormal distribution which are depicted in figure 3. Employing the method proposed by HCM 2010, the level of service for Resalat highway is identified. Table 1 shows the share of each level of services during a week for worst- and best-case scenarios (Chen et al., 2003).

TABLE 1: Tests results and the best fitted functions (B: Burr, W: Weibull, N: Normal, G: Gamma, L: Lognormal)

Time Window	Best LOS	Worst LOS	Sample Size	skewness	kurtosis	Is distribution passed the KS Test					Lowest BIC	Second Lowest BIC
						B	W	N	G	L		
05AM-06AM	A	B	202	8.7	2.9	✓	✓	✓	✓	✓	B	L
06AM-07AM	B	C	564	6.9	2.5	✓	✓	×	✓	✓	B	L
07AM-08AM	D	F	384	-1.1	0.6	×	✓	✓	✓	✓	W	N
08AM-09AM	D	F	384	-1.2	-0.3	×	✓	✓	✓	✓	W	L
09AM-10AM	D	E	334	0.5	-0.4	×	✓	✓	✓	✓	G	L
10AM-11AM	C	E	300	0.8	0.1	×	✓	✓	✓	✓	L	G
11AM-12AM	C	D	300	0.9	1.6	×	✓	✓	✓	✓	B	G
12AM-01PM	C	D	320	2.7	1.5	✓	✓	✓	✓	✓	L	B
01PM-02PM	C	D	334	1.1	1.4	✓	×	✓	✓	✓	B	L
02PM-03PM	D	E	350	0.6	1.0	×	✓	✓	✓	✓	L	B
03PM-04PM	E	E	418	0.3	0.7	×	✓	✓	✓	✓	G	L
04PM-05PM	E	F	274	-0.9	0.3	×	✓	✓	✓	✓	W	L
05PM-06PM	F	F	264	-0.1	0.2	✓	✓	×	✓	✓	G	L
06PM-07PM	F	F	270	-0.6	0.0	✓	✓	✓	✓	✓	W	N
07PM-08PM	F	F	256	-1.0	-0.2	✓	✓	✓	✓	✓	W	N
08PM-09PM	E	F	284	0.7	-0.1	✓	✓	✓	✓	✓	N	L
09PM-10PM	C	E	368	1.3	1.2	✓	✓	✓	✓	✓	L	G

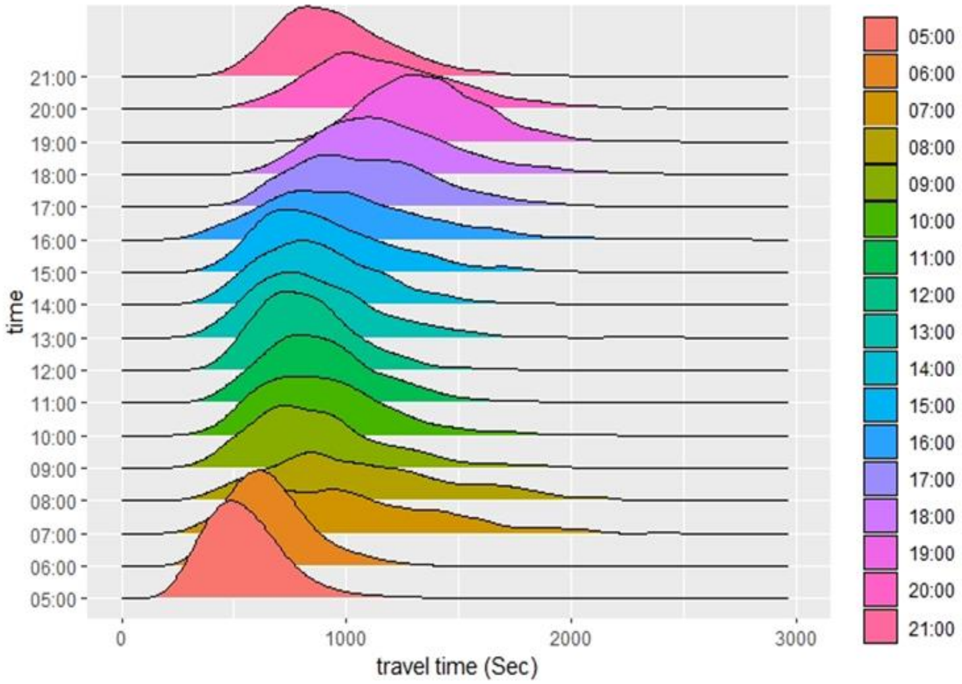

FIGURE 3: Fitted Lognormal distribution on travel time of one-hour windows

4. Methodology

In this section, a closed-form probability distribution function is fitted to the entire day travel time data. This model is classified as a random-effects distribution model (Fitzmaurice, *2008*). The random-effects model incorporates the correlation between the time intervals and establishes an interconnected travel time estimation tool. As a characteristic feature of such a model, it specifies the probability densities conditional on the random effects. Random-effects model fitting and inference are generally performed using the marginal model $f(TT)$; the latter is acquired by integrating over the random-effects distribution. Let $f(TT|b_w)$ denotes the density function of TT on window t conditional to random-effects b_w, and let $f(b_w)$ denotes the random effects distribution. The marginal density function is expressly provided as:

$$f(TT) = \int f(TT|b_w) \times f(b_w) db_w \tag{1}$$

Replacing the base probability density function of $f(TT|b_w)$ with the lognormal density function (which is identified as the best-fitted distribution function for hourly travel time distribution) Equation 1 transforms as follows; $L(TT)$ represents the likelihood function of $f(TT)$ which is presented in equation 3.

$$f(TT) = \int \frac{1}{TT \times \sqrt{2\pi(\sigma + \sigma_w)^2}} \exp\left(-\frac{\left(\ln(TT) - (\mu + \mu_w)\right)^2}{2(\sigma + \sigma_w)^2} \times f(b_w)\right) db_w \tag{2}$$

$$L(TT) = \prod_{t=1}^{n} \int \frac{1}{TT \times \sqrt{2\pi(\sigma + \sigma_w)^2}} \exp\left(-\frac{\left(\ln(TT) - (\mu + \mu_w)\right)^2}{2(\sigma + \lambda\sigma_w)^2} \times f(b_w)\right) db_w \tag{3}$$

$$b_w \sim N(\mu_w, \sigma_w)$$

Where μ and σ are the mean and standard deviation of the probability density function. σ_w and μ_w represent the effect of random variables on all elements affecting the travel time variations. b_w is random-effects which is assumed to follow the normal distribution with the mean of μ_w and the standard deviation of σ_w.

Using the Expectation–maximization (EM) as an optimization tool, the presented model is fitted to the travel time data collected in Resalat highway. μ_w is assumed equal to zero and σ_t is estimated as 0.991. The amount of log-likelihood is -2309.2. The estimated values of μ and σ for the different time windows are tabulated in Table 2. Five other models are also utilized to estimate the travel time distribution in an entire day – mixture of lognormal distribution, mixture of normal, mixture of gamma, mixture of Weibull distribution, and conditional model without random-effects; these models have been suggested by research as a best-fitted distribution (Aron et al., 2014, Kieu et al., 2014, Susilawati et al., 2013, Zheng et al., 2017).

TABLE 2: Estimated variables of the proposed model for different time windows

Time Window	Mean Travel Time (sec)	Mean of Lognormal Distribution (μ)	Standard Deviation of Lognormal Distribution (σ)
05AM-06AM	609	6.25	0.33
06AM-07AM	759	6.44	0.27
07AM-08AM	1148	6.85	0.44
08AM-09AM	1285	6.94	0.45
09AM-10AM	1084	6.75	0.34
10AM-11AM	1089	6.76	0.32
11AM-12AM	1041	6.75	0.30
12AM-01PM	1015	6.70	0.26
01PM-02PM	1004	6.70	0.33
02PM-03PM	1068	6.74	0.33
03PM-04PM	1131	6.78	0.33
04PM-05PM	1279	6.86	0.39
05PM-06PM	1352	6.96	0.30
06PM-07PM	1561	7.06	0.26
07PM-08PM	1404	7.20	0.18
08PM-09PM	1160	7.03	0.30
09PM-10PM	987	6.82	0.26

Table 3 presents the values of log-likelihood functions for all four models. Lower log-likelihood function values indicate better estimation results. According to Table 3, the conditional model with random-effects (Equation 2) shows the best-fit performance among all other models (log-likelihood for conditional model with random effect estimated from eq. 3). Also, in figure 4, mixture of the lognormal distribution, mixture of the normal, mixture of the gamma, and Weibull mixture distribution fitted to travel time for the entire day for comparison. From table 3 and figure 4, it can be deduced that mixture of the normal distribution is the best-fitted mixture distribution.

TABLE 3: Comparison of different models

Model	**Log-likelihood Function Value**
Mixture of lognormal distribution	-7317.2
Mixture of normal distribution	-7291.6
Mixture of Weibull distribution	-7310.1
Mixture of gamma distribution	-7309.3
Conditional model without Random-effects	-3265.2
Conditional model with Random-effects	-2309.2

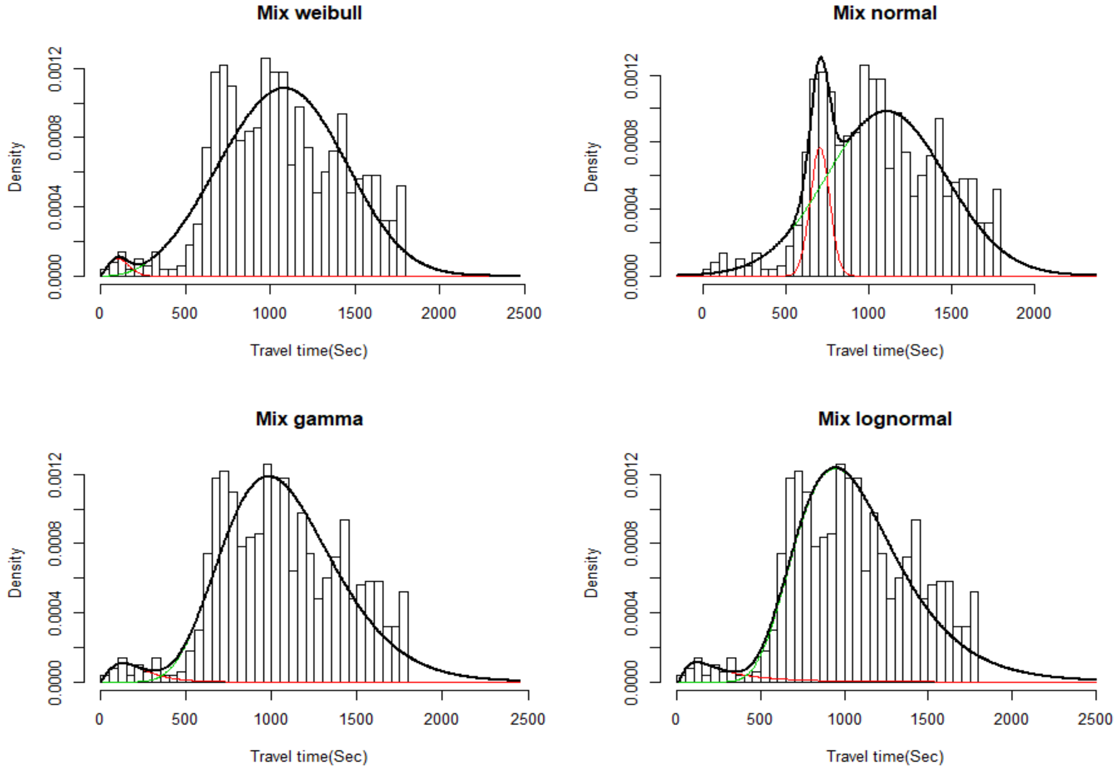

Figure 4: Fitted mixture distribution to travel time data for the entire day

Model Validation

The calibrated model (using the data collected in Resalat highway) in the previous section is utilized to estimate the travel time distribution collected in Modarres highway. The effectiveness of the validation is quantified using the average of mean absolute relative error (MARE) of the model (Swamidass, 2000). MARE is a widely used error metric that uses range normalization. It is given as:

$$\epsilon = \left|\frac{\hat{F}(TT) - F(TT)}{F(TT)}\right| \times 100 \tag{3}$$

Where, $\hat{F}(TT)$ and $F(TT)$ are the estimated (using the calibrated parameters based on the Resalat highway travel time data) and observed (optimized based on the Modarres travel time data) density function values. The error term is calculated as 4.74%. This shows a successful validation of data set using the travel time data of a different highway.

4. Results and conclusions

A mathematical model is presented to describe the travel time distribution over an entire day time period. First, different models are evaluated to identify the best-fitted probability density function for one-hour time windows. Lognormal distribution showed the best performance for hourly time spans. Then, a conditional model with random effects is employed to estimate the travel time variability over a complete day time period. This model can describe the connection between the different time windows. The presented model is the product of the probability density function of each time interval with the probability of normally distributed random effects. Employing the data collected from the two major highways of Tehran, the presented model is successfully calibrated and validated. Different models were also tested to compare the fitting performance with the proposed model. The conditional model with random-effects showed a better fit compared to all the other distributions.

Findings of this research can be utilized in travel time reliability assessment and departure time choice studies. Incorporating a proper travel time distribution model, which can provide a good estimation of travel time over the entire day, stipulates the design of travelers' navigation strategies and tools. Another possible direction for the future research is considering the other influential factors such as climate change, uncertainty of peak hour and departure time in travel time estimation.

References

Abkowitz, M. D. & Engelstein, I. (1983). Factors affecting running time on transit routes. *Transportation Research Part A: General,* 17**,** 107-113.

Al-Deek, H. & Emam, E. B. (2006). New Methodology for Estimating Reliability in Transportation Networks with Degraded Link Capacities. *Journal of Intelligent Transportation Systems,* 10**,** 117-129.

Aliari, Y. & Haghani, A. (2012). Bluetooth Sensor Data and Ground Truth Testing of Reported Travel Times. *Transportation Research Record,* 2308**,** 167-172.

Aron, M., Neila, B. & Guessous, Y. (2014). *Estimating Travel Time Distribution for Reliability Analysis*.

Bates, J., Polak, J., Jones, P. & Cook, A. (2001). *The valuation of reliability for personal travel*.

Chen, C., Zwet, E. & Variya, P. (2003). Travel-Time Reliability as a Measure of Service. *Transportation Research Record,* 1855**,** 74-79.

Fitzmaurice, G., Verbeke, G, (2008). *Parametric modeling of longitudinal data: Introduction and overview, Ch. 2 of Longitudinal data analysis,* Chapman and Hall/CRC

Fosgerau, M. & Engelson, L. (2011). The value of travel time variance. *Transportation Research Part B: Methodological,* 45**,** 1-8.

Kieu, L.-M., Bhaskar, A. & Chung, E. (2014). Public Transport Travel-Time Variability Definitions and Monitoring. *Journal of Transportation Engineering,* 141**,** 1-9.

Lam, T. C. & Small, K. A. (2001). The value of time and reliability: measurement from a value pricing experiment. *Transportation Research Part E: Logistics and Transportation Review,* 37**,** 231-251.

Rakha, H., El-Shawarby, I. & Arafeh, M. (2010). Trip travel-time reliability: Issues and proposed solutions. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations,* 14**,** 232-250.

Rakha, H., El-Shawarby, I., Arafeh, M. & Dion, F, (2006). Estimating path travel-time reliability. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2006. 236-241.

Saedi, R., Saeedmanesh, M., Zockaie, A., Saberi, M., Geroliminis, N. & Mahmassani, H. (2020). Estimating network travel time reliability with network partitioning. *Transportation Research Part C Emerging Technologies,* 112**,** 46-61.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Ann. Statist.,* 6**,** 461-464.

Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association,* 69**,** 730-737.

Sun, C., Arr, G. & Ramachandran, R. (2003). *Vehicle Reidentification as Method for Deriving Travel Time and Travel Time Distributions: Investigation*.

Susilawati, S., Taylor, M. A. P. & Somenahalli, S. V. C. (2013). Distributions of travel time variability on urban roads. *Journal of Advanced Transportation,* 47**,** 720-736.

Swamidass, P. M. (2000). MAPE (mean absolute percentage error) MEAN ABSOLUTE PERCENTAGE ERROR (MAPE). *Encyclopedia of Production and Manufacturing Management.* Boston, MA: Springer US.

Taylor, M. A. P. & Susilawati (2012). Modelling Travel Time Reliability with the Burr Distribution. *Procedia - Social and Behavioral Sciences,* 54**,** 75-83.

Zheng, F., Li, J., Van Zuylen, H., Liu, X. & Yang, H. (2017). *Urban travel time reliability at different traffic conditions*.