SCIENCE CHINA
Technological Sciences

• Article •

January 2019 Vol. 62 No. 1: 106–120
https://doi.org10.1007/s11431-018-9267-4

Analyzing distributions for travel time data collected using radio frequency identification technique in urban road networks

GUO JianHua[1,2], LI ChangGuang[1], QIN Xiao[3], HUANG Wei[1], WEI Yun[4] & CAO JinDe[2,5*]

[1] *Intelligent Transportation System Research Center, Southeast University, Nanjing 210018, China;*
[2] *Transportation Sensing and Cognition Research Center, Southeast University, Nanjing 210018, China;*
[3] *Department of Civil & Environmental Engineering, University of Wisconsin at Milwaukee, Wisconsin 53201, USA;*
[4] *National Engineering Laboratory for Green and Safe Construction Technology in Urban Rail Transit, Beijing 100037, China;*
[5] *School of Mathematics, Southeast University, Nanjing 210018, China*

Received January 10, 2018; accepted April 23, 2018; published online December 19, 2018

Travel time distribution studies are fundamental for supporting transportation system reliability studies, particularly for urban road networks. However, such studies are generally based on travel time data sets with limited sample sizes, which provide inconsistent findings. In this paper, a large amount of travel time data collected from the emerging radio frequency identification (RFID) technique are used to conduct empirical investigations and estimations of travel time distributions, and three major findings are determined. First, travel time data are shown to have a complex statistical structure: the travel time distribution is in general peaky, multi-modal, and skewed to the right, which cross validates findings shown in previous publications. Second, unimodal distribution models are shown to be unable to capture the complex statistical dynamics embedded in the travel time data; therefore, a multistate distribution model is more appropriate for modeling travel time distributions. In this respect, a three-component gaussian mixture model (GMM) is tested and results consistently outperform those of unimodal distribution models. Finally, the aggregation time interval is shown to have a trivial effect on the shape of travel time distributions: the travel time distribution is stable under different aggregation time intervals. Future work is recommended to investigate further travel time variabilities and travel time distribution estimations.

reliability, travel time, travel time distribution, time interval, RFID

Citation: Guo J H, Li C G, Qin X, et al. Analyzing distributions for travel time data collected using radio frequency identification technique in urban road networks. Sci China Tech Sci, 2019, 62: 106–120, https://doi.org/10.1007/s11431-018-9267-4

1 Introduction

Travel time distribution (TTD) is a stochastic characteristic of travel time that plays a fundamental role in evaluating the reliability of transportation systems; particularly for urban networks where traffic conditions are more complex because of a multitude of influencing factors. In this regard, many TTD studies have been proposed [1–9], and a variety of indices have been put forward to measure the reliability of transport systems using TTD models (such as plan time, buffer time, standard deviation, coefficient of variation, and skewness). However, such studies have, in general, focused on the reaction of traveler behavior to different TTD models [10–12], whereas scarce investigations have been conducted into the nature of the TTD model in support of TTD-based transportation system reliability studies [13].

TTD studies require data-intensive effort and demand the use of extensive travel time data. In this regard, TTD studies are heavily affected by the data source employed in the study. However, due to the limitation of commonly used traffic condition sensing techniques, it is not possible to directly collect large amounts of travel time data, and a limited sam-

*Corresponding author (email: jdcao@seu.edu.cn)

ple size is generally employed. In addition, travel time data are usually estimated from inductive loop detectors embedded underneath the pavement or derived using floating cars equipped with GPS (Global Positioning System), and bias can be introduced into the data if the vehicle used in the floating car approach is of particular type, such as a bus or taxi. Consequently, literature currently provides inconsistent results with respect to TTD studies, and multiple statistical TTD models have been identified, such as the normal distribution [14], skewed distribution [15], or Gaussian mixture distribution [16]. Nevertheless, due to the rapid development of advanced traffic sensing techniques in recent years, such as the mobile-phone based sensing technique, automatic license plate identification technique, or RFID (Radio Frequency Identification Technology) based sensing technique [17–21], travel time data are being increasingly accumulated, which thus facilitates TTD studies.

The aim of this paper is to empirically investigate TTDs using large amounts of travel time data collected via the Radio Frequency Identification (RFID) technique in real world urban transportation systems. A variety of commonly used statistical distribution models are tested to unravel the nature of TTDs. The aggregation time interval is an important factor determining the characteristics of traffic variables, and limited studies have researched the effect of the time interval on TTD. Therefore, in this paper, multiple time intervals are applied to travel time data to demonstrate the effect of the time interval on TTD. The remainder of the paper is organized as follows: Sect. 2 summarizes and analyzes related state-of-the-art of TTD studies; Sect. 3 provides commonly used statistical distributions in TTD studies; Sect. 4 presents the method used to collect and process RFID data to obtain travel time data in large quantities; Sect. 5 introduces the outline of experiments conducted in this paper; Sect. 6 presents results of empirical experiments; and conclusions and discussions are presented in Sect. 7.

2 Literature review

In this section, related travel time studies are reviewed, including those related to TTD and the effect of the time interval on TTD.

2.1 Travel time distribution

Knowledge of travel time is important for transportation systems, and many studies have been conducted to investigate the nature of TTDs. In general, investigations have been conducted by fitting travel time data with statistical distributions [15]. It was initially believed that the symmetric distribution characterized vehicle travel time well. For example, Taylor [3] analyzed bus running time data over 15 consecutive days and demonstrated that bus travel time followed a normal distribution. However, statistical analysis has identified that TTD is asymmetric and significantly skewed to the right [22], and Fosgerau and Karlström [23] found the distribution of normalized travel time in a single section was asymmetric and fat right-tailed, and that the normal distribution did not provide a good fit. Some studies have considered the lognormal distribution to be the most appropriate model because of its simplicity and good fit [2, 24]. Kieu et al. [25] analyzed the distribution of public transport day-to-day travel time using transit signal priority data, developed a comprehensive approach for estimating the bus route TTD, and recommended the lognormal distribution as the best descriptor of bus travel time on urban corridors. In addition, Uno et al. [26] used a lognormal distribution to fit bus running time on different routes. Polus [27] analyzed vehicle travel time data for Chicago, USA, and determined that travel time data on the city arterial approximately followed a gamma distribution under specific assumptions. Furthermore, Jordan and Turnquist [28] showed that the skewed distribution fitted bus travel time well in the morning peak period and recommended the gamma distribution as the best model. Al-Deek and Emam [29] modeled travel time variability with the Weibull distribution. In the insurance actuarial field, the Burr distribution has been widely applied to model the distribution of insurance claims. This distribution was developed by Burr [30] for fitting a cumulative distribution function (cdf) . Zimmer et al. [31] noted the advantage of the Burr distribution in modeling observed lifetime data. In addition, Susilawati et al. [8] tested various statistical distributions for travel time data, with the Burr distribution emerging as an appropriate model. Fosgerau and Fukuda [32] found that stable distributions describe the distribution of travel times well and that the family of stable distributions allows distributions with skewness and heavy tails, as observed in empirical data. In ref. [33], a multistate model was employed to fit a mixture of Gaussian distributions into travel time observations in an expressway corridor; each Gaussian distribution was associated with an underlying traffic state, thereby providing a quantitative uncertainty evaluation. These studies show that many distributions have been used to model travel time data, which has resulted in inconsistent results with respect to TTD models.

2.2 Effect of time interval on travel time distribution

The travel time aggregation time interval is an important factor determining the characteristics of TTDs. For example, Vlahogianm and Karlaftis [34] verified that the aggregation level of traffic data can alter its underlying stochastic characteristics, and Li et al. [5] demonstrated that the TTD tends to follow a normal distribution when the aggregation time interval is reduced. In addition, the TTD of passenger cars follows a lognormal distribution when the travel time aggregation level is large, but the distribution tends to be normal when the aggregation time interval is narrower. Mazloumi et al. [35] showed that the normal distribution can adequately fit the travel time data when a short aggregation time interval is used, but that the TTD becomes increasingly skewed for the off-peak period with an increase in the aggregation time interval. Therefore, it has been considered that the lognormal distribution provides a better fit than normal distributions, although distributions for peak periods are still normal. However, recent studies have shown that TTDs are still skewed, even in short aggregation time intervals (e.g., 5 min) [35]. With such inconsistent results in relation to TTDs, it is evident that the time interval has a significant effect on TTDs and that it requires further investigation.

2.3 Summary

In summary, although TTD is an important component supporting transportation system reliability study, investigations on TTD models have provided inconsistent results and multiple statistical distributions have been identified when fitting travel time data collected using various traffic condition sensing techniques. Therefore, the nature of TTDs requires further investigation. However, as conventional travel time data collection techniques are not capable of directly obtaining large amount of travel time data, related investigations into the nature of TTDs have been consequently hampered. However, in this paper, the recent and rapid development of traffic condition sensing techniques has enabled the collection of large quantities of travel time data using RFID traffic sensing technology, and a detailed investigation into TTDs is thus conducted and presented. In addition, the effect of time intervals on TTDs is also investigated in this paper.

3 Commonly used statistical distributions

This section presents an overview of commonly used statistical distributions that have been applied or adopted in literature to describe travel time data distributions. These commonly used statistical distribution models include unimodal models and multistate models.

3.1 Unimodal distribution

The most widely used unimodal models in TTD studies are the normal distribution, lognormal distribution, gamma distribution, Weibull distribution, and Burr distribution. The lognormal, gamma, and Burr distributions are asymmetrical and superior to the use of symmetrical distributions for fitting skewed data. The probability density function (PDF), cumulative density function (CDF), and parameter ranges of these unimodal models are given in Table 1.

3.2 Multistate model

The multistate model is a mixture of a finite number of component distributions, where each component distribution represents the distribution of collected data under different states. A mixture coefficient is associated with each component distribution, thereby representing the probability of each specific state occurring in the multistate model. A finite multistate model with K-component distributions has the following density function:

$$f(T|\lambda,\theta)=\sum_{k=1}^{K}\lambda_k f_k(T|\theta_k), \tag{1}$$

where T represents the collected data; $f(T|\lambda,\theta)$ is the density function of the distribution model of T; $\lambda=(\lambda_1,...,\lambda_K)$ is the vector of mixture coefficients, with $\sum_{k=1}^{K}\lambda_k = 1$; $\theta=(\theta_1,...,\theta_K)$ is the matrix of model parameters for each component distribution; $\theta_k=(\theta_{k1},...,\theta_{KI})$ is vector of model parameters for the kth component distribution; $f_k(T|\theta_k)$ is the density function for the kth component distribution.

Specifically, the density function of a three-component model based on normal distribution is defined as follows:

$$\begin{aligned} &f(T|\lambda_1,\lambda_2,\lambda_3,\mu_1,\mu_2,\mu_3,\sigma_1,\sigma_2,\sigma_3)\\ &=\lambda_1\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{\frac{(T-\mu_1)^2}{2\sigma_1^2}}+\lambda_2\frac{1}{\sqrt{2\pi}\sigma_2}\mathrm{e}^{\frac{(T-\mu_2)^2}{2\sigma_2^2}}\\ &\quad+\lambda_3\frac{1}{\sqrt{2\pi}\sigma_3}\mathrm{e}^{\frac{(T-\mu_3)^2}{2\sigma_3^2}}, \end{aligned} \tag{2}$$

where λ_1 is the mixture coefficient for the first component distribution with mean μ_1 and standard deviation σ_1; λ_2 is the mixture coefficient for the second component distribution with mean μ_2 and standard deviation σ_2; λ_3 is the mixture coefficient for the third component distribution with mean μ_3 and standard deviation σ_3.

4 Travel time calculation using RFID technique

As mentioned previously, large quantities of travel time data were collected using the RFID technique and are used in the TTD study presented in this paper. Therefore, in this section, travel time calculations and the general principle of the RFID technique, RFID data preprocessing, are described in detail to support an understanding of subsequent investigations.

Table 1 Unimodal models

Unimodal model	Probability density function	Cumulative distribution function	Parameter range
Normal	$f(t)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right]$	$F(t)=\frac{1}{\sigma\sqrt{2\pi}}\int_0^t\exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right]\mathrm{d}t$	$t>0, -\infty<\mu<+\infty, \sigma>0$
Lognormal	$f(t)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\ln t-\mu}{\sigma}\right)^2\right]$	$F(t)=\frac{1}{\sigma\sqrt{2\pi}}\int_0^t\exp\left[-\frac{1}{2}\left(\frac{\ln t-\mu}{\sigma}\right)^2\right]\mathrm{d}t$	$t>0, -\infty<\mu<+\infty, \sigma>0$
Gamma	$f(t)=\frac{1}{\lambda^{\alpha}\Gamma(\alpha)}t^{\alpha-1}\exp\left(-\frac{t}{\lambda}\right)$	$F(t)=\frac{1}{\lambda^{\alpha}\Gamma(\alpha)}\int_{-\infty}^t t^{\alpha-1}\exp\left(-\frac{t}{\lambda}\right)\mathrm{d}t$	$t>0, \alpha>0, \lambda>0$
Burr	$f(t)=\frac{\alpha\tau}{\lambda}\frac{x^{\tau-1}}{\left(1+\frac{x^{\tau}}{\lambda}\right)^{\alpha+1}}$	$F(t)=1-\left(1+\frac{x^{\tau}}{\lambda}\right)^{-\alpha}$	$t>0, \alpha>0, \lambda>0, \tau>0$
Weibull	$f(t)=\frac{\lambda}{\alpha}\left(\frac{t}{\lambda}\right)^{\alpha-1}\exp\left[-\left(-\frac{t}{\lambda}\right)^{\alpha}\right]$	$F(t)=1-\exp\left[-\left(-\frac{t}{\lambda}\right)^{\alpha}\right]$	$t>0, \alpha>0, \lambda>0$

4.1 General principle of RFID technique

Radio frequency identification (RFID) is an automatic identification technology that enables non-contact communication between a transmitter and a receiver so that an intended target can be automatically identified. Since its inception, the RFID technique has been applied in many fields, and one of its most recent applications is within the transportation field for detecting vehicles operating within the road network. To enable this, a fully deployed RFID-based traffic condition sensing system is deployed, which primarily includes an RFID base station, electronic tags, central processing system, and supporting systems. The RFID base station is installed next to the road, and it remotely reads the electronic tags installed in vehicles as they pass within a certain range. Note that identification information is preloaded into electronic tags; this primarily includes an electronic tag number, vehicle license plate number, and vehicle license plate color. When the electronic tag is read, a combination of the identification information, the current time stamp, and the base station information generate a so-called "vehicle passing record" (VPR), which is the fundamental output of the RFID-based traffic condition sensing system. Every vehicle that passes the RFID base station generates a passing record; therefore, a huge number of VPRs can be generated in heavy traffic; all of these are stored in a central system with the assistance of supporting systems, such as communication systems.

4.2 Vehicle passing record (VPR) screening

As with all traffic condition sensing systems, the VPR can sometimes contain anomalies, such as erroneous or redundant data, that need to be processed to increase data quality. Erroneous data are mainly related to vehicle license plate numbers that have been wrongly preloaded in the electronic tag, or which have been read mistakenly when the vehicle passes a station. Redundant data mainly relate to the vehicle detection process, for example where duplicate or similar VPRs are generated for the same vehicle passing the same station, and are usually identified as two or more VPRs with the same vehicle license plate number within a very short range of time. It is clear that vehicle license plate numbers are the major source or such anomalies; therefore, to remove abnormal VPRs, a screening procedure based on the vehicle license plate naming rule is developed in this study.

4.3 Travel time calculation

The VPR represents the fundamental data collected from the RFID-based traffic condition sensing system, and travel time data are then derived from processing the VPRs. In this respect, for the RFID-based traffic condition sensing system, a RFID station pair is defined as Origin Station, Destination Station. The difference between the time stamps for one vehicle passing the origin station and then the destination station, consecutively, is defined as the travel time of this vehicle for the RFID station pair. A procedure was developed in this study, whereby the VPRs for the origin station and the destination station were matched together using a combination of the license plate number and the license plate color as the key index. To enable use of the RFID-based traffic condition sensing system in this study, each vehicle requires a unique vehicle license plate number and vehicle plate license color combination, to conduct the matching process. The procedure used is shown in Figure 1.

Note that in Figure 1, a set of VPRs is obtained at the origin station for each VPR, and these then determine the set of VPRs for the destination station. These data include all the VPRs with timestamps falling within a predefined interval, starting with the timestamp of the selected VPR for the origin station. Clearly, the length of the predefined interval depends on the distance between the origin station and the destination station, and longer intervals are required for station pairs situated at greater distances. In this paper, the predefined interval is defined as 10 min, as the distance between station pairs

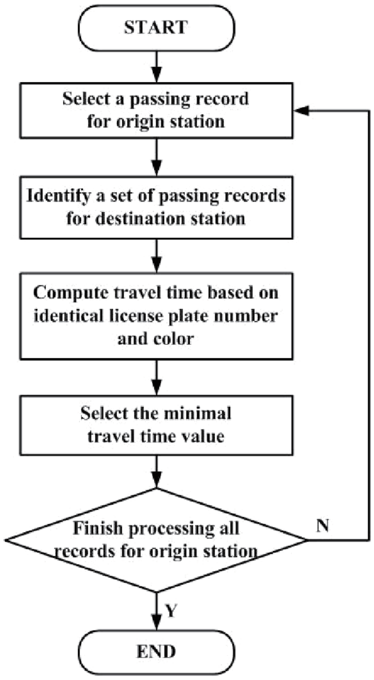

Figure 1 Calculation of travel time for arterial between RFID base station pair.

used in this study is in general around 1 km in the urban environment, which translates into a vehicle link travel speed of 6 km/h. This means that travel speeds of less than 6 km/h are regarded as suspicious, and in such instances the computed travel time is then be removed from investigation. In addition, due to potential possibility of a vehicle passing the destination station multiple times, the minimum travel time is selected when multiple vehicle passing records were found based on a single set of license plate number and color.

4.4 Travel time aggregation

In real world applications, travel time data are generally aggregated under different time intervals. In this regard, given a specific time interval, travel time data records can be divided into groups that are determined by a selected time interval according to the timestamp of the travel time record passing the origin RFID station. After dividing the travel time data records into different groups, the average of the travel time data records within each group can be used as the aggregated travel time for this group. For example, given a 5-minute time interval, a single day is divided into 288 time-groups, and travel time data records for each day are divided into these 288 time-groups, which generates 288 aggregated travel times for this day. Although it may appear to be simple and straightforward, travel time aggregation is an important step in making travel time data usable in real world applications.

5 Experimental design

In this paper, three investigations are conducted: empirical investigations into travel time variability, empirical investigations into TTDs, and estimations of TTDs, as described in the section below.

5.1 Empirical travel time investigation

The aim of this investigation is to show variability in collected travel time data. Three types of travel time variability have been previously defined in literature. The first type is vehicle-to-vehicle variability; this is the variability in travel time within a specific period of time that reflects the differences between travel times experienced by different vehicles traveling the same trip within the same period of time. Factors contributing to vehicle-to-vehicle variability relate to signal timing, driver behavior, or impedances from bikes or pedestrians. The second type is day-to-day variability, which is the variability in travel time for the same trip and same time-period across different days; this reflects day-to-day fluctuations in traffic demand, weather, driver behavior, and/or incidents. The third type is period-to-period variability, which is the variability in aggregated travel times across time; this reflects the continuous evolution of aggregated travel times with respect to different times of the day. Factors contributing to period-to-period variability include temporal variations in traffic demand, incidents, weather conditions, or the level of daylight.

A comparison between the definitions of these three types of variability shows that the first two types are helpful for gaining an understanding of the nature of travel time, whereas the final one is useful for developing transportation applications, as most transportation applications run in a continuous time fashion. Therefore, this study focuses on period-to-period variability, which can be demonstrated by showing the aggregated travel time for each time interval across time. Note that in this paper, the time intervals are selected as 5-min, 10-min, 30-min, and 60-min, respectively, to show the effect of time interval aggregation on the aggregated travel time.

5.2 Empirical travel time distribution investigation

The aim of this investigation is to show empirical TTDs. Three measures of central tendency are used in this paper: the mean, median, and mode of empirical travel time data. In addition, four measures of variability are selected: standard deviation (SD), skewness, kurtosis, and coefficient of variance (COV). Note that time intervals of 5-min, 10-min, 30-min, and 60-min are also selected to show the effect of aggregation of the time interval on TTDs. Of these measures, the mean, median, mode, and SD are widely used, therefore these are not defined here, but definitions of skewness, kurtosis, and COV are provided below.

Skewness is defined as the third-order central moment of a distribution, and it describes the degree of distribution symmetry; kurtosis is defined as the fourth-order central moment of a distribution, and it describes the degree of distribution flatness. Larger skewness and kurtosis indicate that the distribution is asymmetric: the peak of the distribution is shifted to one side and a long tail extends to the other side. Kurtosis can also be used to show the origin of the variance of data: a non-peaky distribution shows that the variance originates from throughout the whole value range, while a peaky distribution shows that the variance mainly originates from tails. Computationally, given a series sample of sample data, $x_t, t = 1, ...T$, skewness and kurtosis can be calculated as follows:

$$\text{Kurtosis} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t - \overline{x}}{\hat{\sigma}}\right)^4, \tag{3}$$

$$\text{Skewness} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t - \overline{x}}{\hat{\sigma}}\right)^3. \tag{4}$$

The coefficient of variation (COV) is defined as the ratio of the standard deviation to the mean. COV is usually used to show the distributions degree of dispersion. As COV is a dimensionless quantity, objective comparisons between different travel time data sets can be conducted. The COV is calculated using the following formula:

$$\text{COV} = \frac{\text{SD}_{\text{TT}}}{\text{E}_{\text{TT}}}, \tag{5}$$

where SD_{TT} is the SD of the TTD, and E_{TT} is the mean of the TTD.

5.3 Travel time distribution (TTD) estimation

The purpose of this investigation is to fit travel time data with the statistical models presented in Sect. 3, with the aim of identifying the most suitable statistical distribution model for modeling travel time data. Note that time intervals of 5-min, 10-min, 30-min, and 60-min are also applied in this investigation.

During this estimation or fitting process, three measures of goodness-of-fit are selected. The first is the Akaike information criterion (AIC), which measures the relative quality of a statistical model by trading off the model complexity (by considering the number of parameters) and the goodness-of-fit of the fitted model (by considering the maximized value of the log-likelihood). In general, complex models will fit the data better than simple models. However, the benefits in model fitting may not outweigh the extra model complexity. We assume a kth order model that specifies a probability density function $f(x|\boldsymbol{\theta_k})$ of n observations with a free parameter vector $\boldsymbol{\theta_k}$, and we then find the Maximum Likelihood Estimator $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}$ of $\boldsymbol{\theta_k}$ with $\boldsymbol{\theta_k} \in R^k$ (the K-dimensional Euclidean space) by maximizing the likelihood function with respect to $\boldsymbol{\theta_k}$. Hence, the likelihood is computed as

$$L(\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}|\hat{\boldsymbol{x}}) = L(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta_k}). \tag{6}$$

The AIC can then be computed as follows:

$$\text{AIC}(k) = -2\log f(x_i|\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}) + 2k = -2\log L(\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}) + 2k. \tag{7}$$

In addition to the AIC, the sum of squares for error (SSE) and the coefficient of determination (R^2) are also used for determining the effect of model fitting. The SSE reflects the difference between observed and fitted values, with SSE equals 0 indicating no difference between observed and fitted values. R^2 has a range between 0 and 1 and is used to explain the variability of dependent variables in fitting equations: if R^2 is close to 1, the variability of dependent variable is almost completely caused by the independent variable. These two measures are computed below as

$$\begin{aligned}
\text{SSE} &= \sum_{i=1}^{n}(y_i - \hat{y}_l)^2, \\
\text{SSR} &= \sum_{i=1}^{n}(\hat{y}_l - \bar{y})^2, \\
\text{SST} &= \sum_{i=1}^{n}(y_i - \bar{y}) = \sum_{i=1}^{n}(y_i - \hat{y}_l) + \sum_{i=1}^{n}(\hat{y}_l - \bar{y}) \\
&= \text{SSE} + \text{SSR}, \\
R^2 &= \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}},
\end{aligned} \tag{8}$$

where y_i is the observed data; $\hat{y}_l$ is the fitted or estimated data; $\bar{y}$ is the mean of the observed data; n is the total sample size of the data; SSR is the sum of squares for regression; and SST is the total sum of squares for regression.

6 Experimental results

6.1 Data collection

As an emerging traffic sensing technique, the RFID-based traffic condition detection system collects massive RFID data daily and has been implemented in the City of Nanjing, China. In this paper, a central region of Nanjing is selected as the study road network, as shown in Figure 2.

According to Figure 2, the study road network contains 43 RFID base stations (the numbers of RFID base stations and road names are shown in the figure). From the roads in the selected network, the major arterial roads of East Zhongshan Road, Middle Longpan Road, Zhujiang Road, Ruijin Road, Changfu Street, and Hubu Street are selected as study roads. There are 22 RFID base station pairs situated along these urban roads. As operational failures occurred at some of these RFID base stations, travel time data from 19 RFID base station pairs from April 1, 2014 to April 30, 2014 were selected for study (listed in Table 2), for time intervals of 5-min, 10-min, 30-min, and 60-min. It is clear from Table 2 that a large amount of travel time data were collected for these time intervals and for each station pair, thereby providing a good chance of unraveling the nature of travel time through investigating large quantities of travel time samples.

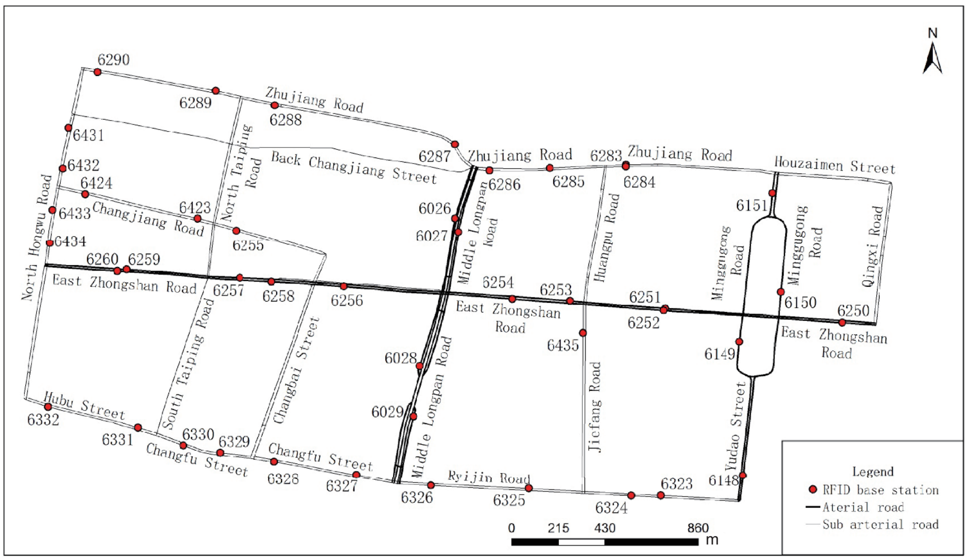

Figure 2 (Color online) Study road network.

Table 2 Travel time data overview

Road name	Direction	Base station pair	Sample size			
			5-min	10-min	30-min	60-min
Zhujiang Road	West bound	(6288, 6286)	8600	4318	1440	720
		(6286, 6284)	8588	4301	1435	718
	East bound	(6283, 6285)	8535	4274	1427	715
		(6285, 6287)	5561	2798	944	479
East Zhongshan Road	West bound	(6258, 6256)	8635	4319	1440	720
		(6256, 6254)	8627	4318	1440	720
		(6254, 6252)	8599	4304	1439	720
	East bound	(6251, 6253)	7877	3944	1316	658
		(6253, 6257)	7677	3891	1311	658
		(6257, 6259)	8610	4318	1440	720
Hubu Street-Ruijin Road	West bound	(6332, 6330)	8558	4318	1440	720
		(6330, 6328)	8623	4318	1440	720
		(6328, 6326)	8593	4319	1440	720
		(6326, 6324)	8637	4320	1440	720
	East bound	(6323, 6325)	8038	4029	1344	673
		(6325, 6327)	8009	4029	1345	673
		(6327, 6329)	8529	4313	1440	720
		(6329, 6331)	8590	4318	1440	720
Middle Longpan Road	North bound	(6026, 6028)	5287	2723	924	463

6.2 Empirical travel time description

As designed previously, this section shows the results of empirical travel time variability, namely, period-to-period variability. In doing so, four RFID station pairs were selected: base station pair (6026, 6028) on Middle Longpan Road, base station pair (6253, 6257) on East Zhongshan Road, base station pair (6283, 6285) on Zhujiang Road, and base station pair (6329, 6331) on Hubu Street-Ruijin Road. The travel times of these RFID base station pairs on April 12, 2014 were collected, and the period-to-period travel time variability is shown in Figure 3.

From an analysis of Figure 3, two observations can be made. First, all four RFID station pairs show clearly different travel time variability patterns. Specifically, pairs (6026,6028) and (6253,6257) show a clear peak period pattern during morning peak hours, indicating congested traffic at this time, whereas the other two pairs, pairs (6329,6331) and (6283,6285), do not show clear peak hour patterns. However, it is obvious that different types of roads will show different travel time patterns, and hence travel time variability will show different patterns across time. Second, regardless of the RFID station pairs, travel time is gradually smoothed out with an increase in the aggregation time interval. In other words, an increase in the time interval reduces the variability of travel time across time. This is helpful for determining an appropriate time interval when considering the time interval required to update traffic management and control systems. It is also important for use with transportation applications dealing with the variability of travel time under specific time intervals.

6.3 Empirical travel time distribution investigation

As described in Sect. 5, central tendency and variability measures are investigated for empirical travel time distributions, and measures of central tendency are shown in Table 3 for all the 19 RFID station pairs. From Table 3, it can be seen that the mean and median values of each selected travel time data set are generally very close to each other for different time intervals, which shows that the travel time aggregation process has a weak effect on the central location of the distributions. In addition, although the mode changes for different time intervals in several travel time data sets, the aggregation process does not generally significantly alter the modes for different time intervals. In summary, the location tendency of TTDs remains stable during the aggregation process.

Measurements of TTD variability are shown in Table 4 for all the 19 RFID station pairs. An observation of Table 4 firstly shows that all station pairs have different COVs and SDs, which indicates that the travel time dispersions are different for different links; this is normal as different links have different traveling patterns. It also shows that for each specific RFID station pair, there is a slight decrease in the SD and COV with an increase in the time interval, which indicates the smoothing effect of using long-time intervals when

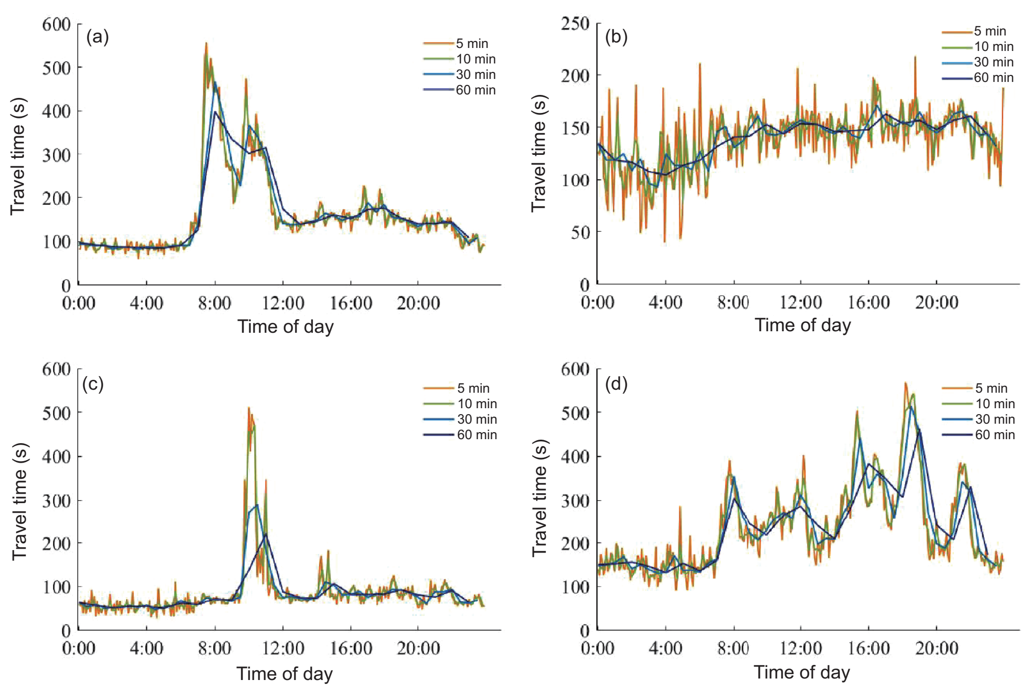

Figure 3 (Color online) Travel Time Variability for different base station pairs. (a) (6026, 6028); (b) (6329, 6331); (c) (6253, 6257); (d) (6283, 6285).

Table 3 Central tendency of travel time data measurements

Base station pair	5-min			10-min			30-min			60-min		
	Mean	Median	Mode	Mean	Median	Mode	Mean	Median	Mode	Mean	Median	Mode
(6288, 6286)	163.21	158	169	163.35	160	167	163.49	160	171	163.57	159	174
(6286, 6284)	87.89	85	86	87.99	86	92	88.08	85	71	88.18	85	84
(6283, 6285)	58.00	56	53	58.03	56	58	57.91	56	54	57.99	57	54
(6285, 6287)	123.87	109	107	123.84	107	72	123.64	109	93	123.65	109	94
(6258, 6256)	75.76	70	75	75.42	70	73	75.01	71	72	74.45	72	74
(6256, 6254)	141.40	126	135	141.83	127	126	141.44	130	132	141.05	132	81
(6254, 6252)	112.09	101	74	111.99	102	76	111.78	105	68	111.80	107	71
(6251, 6253)	119.85	82	61	119.83	81	52	118.62	80	63	117.91	80	60
(6253, 6257)	252.85	241	254	254.08	243	187	256.01	249	257	256.41	252	168
(6257, 6259)	116.70	97	77	116.86	95	77	116.65	93	78	116.38	94	76
(6332, 6330)	142.77	145	143	142.87	146	145	142.86	146	151	143.09	147	154
(6330, 6328)	75.36	71	71	75.52	72	74	75.59	73	77	75.69	74	75
(6328, 6326)	154.30	157	177	154.39	158	157	154.6	160	168	154.86	162	162
(6326, 6324)	142.11	118	113	142.28	119	109	142.60	119	106	142.73	120	108
(6323, 6325)	120.36	111	107	120.71	112	113	121.09	111	115	121.25	112	115
(6325, 6327)	136.99	120	102	136.97	121	98	137.41	120	94	137.33	119	101
(6327, 6329)	125.45	121	127	125.39	122	127	125.22	122	123	124.83	122	95
(6329, 6331)	103.01	88	64	102.97	88	67	103.24	89	71	103.42	88	71
(6026, 6028)	166.47	136	91	154.18	139	143	154.97	141	88	154.85	141	150

aggregating travel time. This result is in alignment with the findings of previous investigations. Furthermore, an observation of the skewness and kurtosis shows that all travel time distributions are asymmetric toward the right and that the variances of the distributions mainly originate from the tails of the distributions. This is important, as it shows that symmetric distributions, such as the normal distribution, cannot model real world travel time data, whereas asymmetric distribution models are more suitable. Finally, an analysis of the effect of time intervals on these measures shows that they do not significantly affect the shape of TTDs, which indicates that the travel time distribution is potentially stable under different aggregation time intervals.

In summary, this study shows that TTDs are peaked and asymmetrical, and that the aggregation time interval has a minimal effect on the shape of TTDs. This implies that TTDs are intrinsically stable across time and space.

6.4 Travel time distribution (TTD) estimation

As mentioned previously in Sect. 5, travel time data collected using the RFID sensing technique were used to fit selected distribution models. Note that the multistate model fitted here is a three-component gaussian mixture model (GMM), which is a combination of three normal distributions. In addition, three goodness-of-fit measures of AIC, SSE, and R^2 were used to gauge the effect of these distribution models on fitting travel time data. In the following, the average values of AIC, SSE, and R^2 in fitting TTD models are listed in Table 5, and Table 6 lists the estimated parameters of these fitted TTD models. It is also of note that all these outcomes are presented together with different time intervals.

An observation of Table 5 shows firstly that the GMM model consistently outperforms other models in terms of all the three goodness-of-fit measures and across all time intervals. This indicates that the statistical structure of travel time data is complex; therefore, single distribution models may not be able to capture all the dynamics embedded within the travel time data, and hence mixed models with multiple component distributions may be better choices for fitting travel time data. The GMM provides an established improvement over other models, and its degree of complexity is well rewarded by delivering a more realistic representation of travel time variations. Furthermore, the GMM model can be easily handled by standard statistical packages. It is thus considered to be flexible and robust enough to handle travel time data distributions. An observation of R^2 also shows a consistently decreasing trend with an increase in the time intervals (this measure has a high value of 0.9248 for the 5-min interval but decreases consistently to a moderate value of 0.6713 for the 60-min interval), which also indicates that aggregation of travel time data may cancel out the statistical dynamics of travel time data, thereby decreasing the model fitting effect.

Table 4 Variability in travel time data measurements

Base station pair	5-min				10-min			
	COV	Skewness	Kurtosis	SD	COV	Skewness	Kurtosis	SD
(6288, 6286)	36.94	1.84	6.58	60.30	34.99	2.03	7.56	57.16
(6286, 6284)	31.81	1.00	4.86	27.96	27.87	0.81	3.32	24.52
(6283, 6285)	36.92	2.41	18.85	21.41	2.26	2.39	18.38	18.72
(6285, 6287)	54.92	2.42	9.35	68.03	53.43	2.53	8.91	66.17
(6258, 6256)	49.93	4.73	36.28	37.83	44.70	4.57	32.96	33.71
(6256, 6254)	49.47	2.39	7.94	69.96	48.34	2.40	7.88	68.55
(6254, 6252)	46.64	1.80	5.92	52.28	44.25	1.71	5.17	49.56
(6251, 6253)	74.97	1.64	2.59	89.86	72.92	1.58	2.20	87.38
(6253, 6257)	37.84	0.99	0.95	95.67	37.43	1.03	0.97	95.10
(6257, 6259)	62.39	2.41	7.42	72.82	60.31	2.50	7.86	70.48
(6332, 6330)	24.90	1.45	13.16	35.55	21.60	1.77	18.99	30.86
(6330, 6328)	47.14	2.61	13.57	35.52	44.59	2.72	13.71	33.67
(6328, 6326)	29.40	0.97	4.62	45.36	27.27	1.02	4.97	42.11
(6326, 6324)	55.93	2.43	6.26	79.47	54.58	2.55	6.75	77.66
(6323, 6325)	43.18	2.50	9.64	51.97	40.64	2.79	11.44	49.06
(6325, 6327)	51.88	1.76	3.50	71.07	50.22	1.83	3.62	66.79
(6327, 6329)	38.74	3.58	21.45	48.60	36.16	3.77	23.19	45.33
(6329, 6331)	50.71	1.50	2.81	52.23	48.66	1.55	2.98	50.11
(6026, 6028)	58.98	1.75	3.52	98.19	54.83	2.38	6.05	84.54
Base station pair	30-min				60-min			
	COV	Skewness	Kurtosis	SD	COV	Skewness	Kurtosis	SD
(6288, 6286)	32.61	1.94	6.50	53.32	30.53	1.52	3.96	49.94
(6286, 6284)	24.24	0.83	3.93	21.35	23.07	0.59	1.13	20.34
(6283, 6285)	26.76	2.31	14.77	15.50	3.94	1.92	10.29	13.88
(6285, 6287)	49.03	2.26	6.84	60.63	46.84	2.10	6.01	57.91
(6258, 6256)	38.04	4.20	28.08	28.53	31.18	3.06	16.75	23.22
(6256, 6254)	44.05	2.10	6.45	62.30	40.38	1.71	4.63	56.95
(6254, 6252)	39.70	1.28	2.08	44.38	37.18	1.06	1.34	41.57
(6251, 6253)	67.65	1.37	1.15	80.25	64.48	1.27	0.8	76.03
(6253, 6257)	35.85	0.96	0.73	91.77	34.19	0.82	0.44	97.66
(6257, 6259)	56.04	2.34	6.51	65.38	52.08	2.05	4.7	60.61
(6332, 6330)	17.40	1.95	19.33	24.85	15.59	1.23	9.72	22.30
(6330, 6328)	39.92	2.47	10.44	30.18	36.85	2.15	8.53	27.89
(6328, 6326)	24.77	0.89	3.74	38.30	23.31	0.38	0.45	36.10
(6326, 6324)	51.89	2.62	7.04	73.99	48.76	2.58	6.94	69.59
(6323, 6325)	37.05	2.80	10.19	44.86	34.85	2.63	8.67	42.26
(6325, 6327)	47.24	1.80	3.13	64.91	45.42	1.70	2.71	62.38
(6327, 6329)	32.45	3.87	23.98	40.64	29.05	3.12	16.99	36.26
(6329, 6331)	45.18	1.62	3.30	46.64	42.46	1.45	2.44	49.92
(6026, 6028)	52.45	2.30	5.82	81.28	49.13	2.08	4.85	76.08

An observation of Table 6 shows the similarity of the fitted distributions for each RFID station pair across different time intervals. In other words, the TTD shapes remain stable for each RFID station pair with different aggregation time intervals. This confirms the findings in the previous investigation and supports the notion of plausible stability of travel time data under different time intervals.

To demonstrate intuitively the effect of fitting distribution models, Figure 4 to Figure 7 shows the fitting effect for four different RFID station pairs on different roads, including (6029, 6027), (6253, 6257), (6285, 6287) and (6329, 6331). In these figures, the histograms of travel time data and the fitted distribution models are shown collectively. An observation of Figure 4 to Figure 7 firstly shows that the histograms

Table 5 Distribution estimation effect of different time intervals

TTD model	5-min			10-min			30-min			60-min		
	SSE	R^2	AIC	SSE	R^2	AIC	SSE	R^2	AIC	SSE	R^2	AIC
Normnal	0.0311	0.4929	87244	0.0022	0.4084	43424	0.0037	0.2206	14237	0.0044	0.1704	6975
Lognormal	0.0136	0.8469	83875	0.0012	0.7672	41573	0.0024	0.5363	13645	0.0034	0.3789	6777
Gamma	0.0165	0.8078	84495	0.0013	0.7183	41953	0.0028	0.4932	13780	0.0036	0.3531	6924
Weibull	0.0302	0.5012	86206	0.0023	0.3566	41189	0.0039	0.1306	14160	0.0048	0.0778	7004
Burr	0.0089	0.9117	83395	0.0008	0.8634	41234	0.0018	0.7365	13508	0.0029	0.5575	6711
GMM	0.0004	0.9284	83128	0.0007	0.8804	40958	0.0016	0.7444	13443	0.0022	0.6713	6646

Table 6 Estimated parameters of travel time distribution models

Station Pair	Time Interval	GMM								Normal	
		λ_1	λ_2	μ_1	μ_2	μ_3	σ_1	σ_2	σ_3	μ	σ
(6253, 6257)	5-min	0.2533	0.4997	155.78	245.85	366.55	26.62	45.45	97.02	252.85	95.67
	10-min	0.2787	0.4684	159.18	249.34	367.41	23.15	41.42	95.42	254.08	95.10
	30-min	0.3351	0.2808	166.21	254.82	335.20	18.25	22.19	88.60	256.01	91.77
	60-min	0.3145	0.3488	164.37	256.87	341.88	14.84	30.67	78.74	256.41	87.66
(6283, 6285)	5-min	0.7062	0.2720	49.51	74.65	125.07	13.34	12.41	54.05	58	21.41
	10-min	0.9872	0.0030	56.95	123.13	145.89	15.34	3.83	49.41	58.03	18.72
	30-min	0.9834	0.0050	56.77	101.57	136.18	12.19	2.74	27.46	57.91	15.50
	60-min	0.9363	0.0456	56.43	66.15	118.23	10.95	1.69	17.57	57.99	13.88
(6329, 6331)	5-min	0.4798	0.3639	67.83	114.22	184.86	16.50	30.05	60.84	103.01	52.23
	10-min	0.4180	0.4240	67.98	108.07	181.87	13.85	29.69	58.76	102.97	50.11
	30-min	0.4106	0.4200	70.39	106.47	174.20	12.02	26.42	54.63	103.24	46.64
	60-min	0.4699	0.3813	71.68	114.46	175.41	10.39	25.71	45.65	103.42	43.92
(6026, 6028)	5-min	0.8529	0.0814	125.43	235.66	419.54	33.51	55.85	71.25	153.74	86.20
	10-min	0.8643	0.0775	126.95	252.98	427.43	32.75	55.65	62.17	154.18	84.54
	30-min	0.8581	0.0365	127.99	223.28	350.79	32.09	14.46	87.74	154.97	81.28
	60-min	0.8474	0.1299	128.78	273.48	448.21	32.32	58.67	26.09	154.85	76.08

Station Pair	Time Interval	Lognormal		Gamma		Weibull		Burr		
		μ	σ	α	λ	α	λ	α	λ	τ
(6253, 6257)	5-min	5.47	0.37	7.57	33.41	284.21	2.77	242.953	4.58	1.09
	10-min	5.47	0.36	7.85	32.38	285.48	2.79	233.69	4.88	0.96
	30-min	5.49	0.34	8.51	30.08	287.09	2.92	236.66	5	0.1293
	60-min	5.49	0.33	9.13	28.09	286.82	3.08	263.79	4.65	1.31
(6283, 6285)	5-min	3.99	0.35	8.39	6.91	64.75	2.63	63.99	4.49	1.61
	10-min	4.01	0.30	11.03	5.26	64.32	2.89	61.74	5.33	1.46
	30-min	4.03	0.25	16.19	3.58	63.59	3.33	57.73	7.18	1.13
	60-min	4.03	0.23	19.64	2.95	63.32	3.72	58.17	7.94	1.15
(6329, 6331)	5-min	4.52	0.47	4.64	22.20	116.83	2.10	77.65	4.45	0.6639
	10-min	4.53	0.44	5.14	20.05	116.72	2.18	73.69	5.29	0.53
	30-min	4.55	0.39	6.06	17.03	116.80	2.32	71.29	6.9	0.4031
	60-min	4.56	0.38	6.74	15.35	116.82	2.46	64.81	11.69	0.2056
(6026, 6028)	5-min	4.93	0.43	4.78	32.19	174.50	1.93	103.657	6.45	0.4425
	10-min	4.93	0.42	5.01	30.78	174.98	1.97	102.91	6.99	0.39
	30-min	4.95	0.41	5.33	29.05	175.81	2.04	100	7.93	0.3215
	60-min	4.95	0.39	5.75	26.92	175.51	2.15	103.42	7.51	0.3567

of selected travel time data clearly show that TTDs are complex and have multistate, peak, and right-skewness characteristics. These results cross-validate the findings of previous investigations. It is also clear that aggregation of travel time has no significant effect on the shape of TTDs, thereby indicating the stability of travel time under different time intervals. Furthermore, in terms of model fitting, the GMM model generally shows superiority of model fitting over the other distribution models. In particular, for the RFID station pair (6253,6257) with pronounced multistate phenomenon, as shown in Figure 4, the fitted GMM model clearly and closely tracks the two states shown in the data, while the other models are unable to capture the two states.

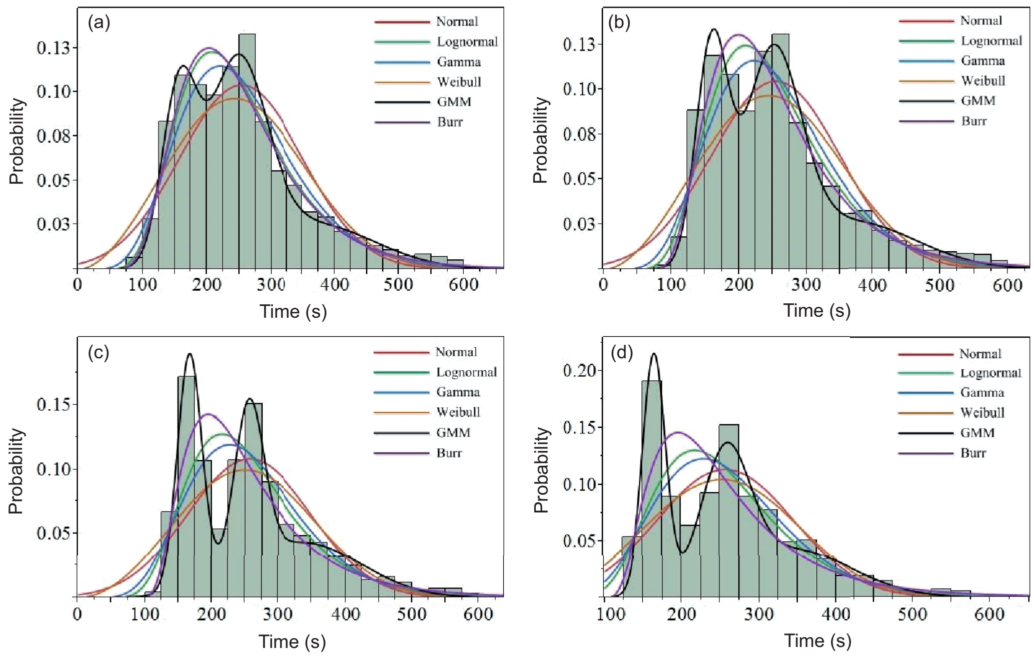

Figure 4 (Color online) Travel time distribution (TTD) for (6253, 6257) under different time intervals. (a) 5 min; (b) 10 min; (c) 30 min; (d) 60 min.

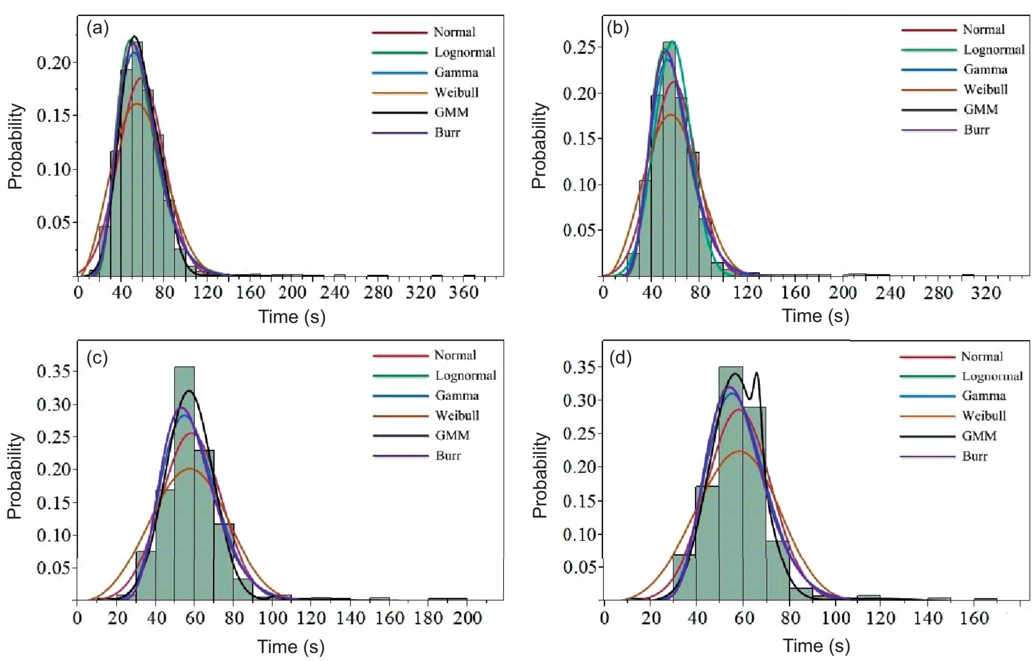

Figure 5 (Color online) Travel time distribution (TTD) for (6283, 6285) under different time intervals. (a) 5-min; (b) 10-min; (c) 30-min; (d) 60-min.

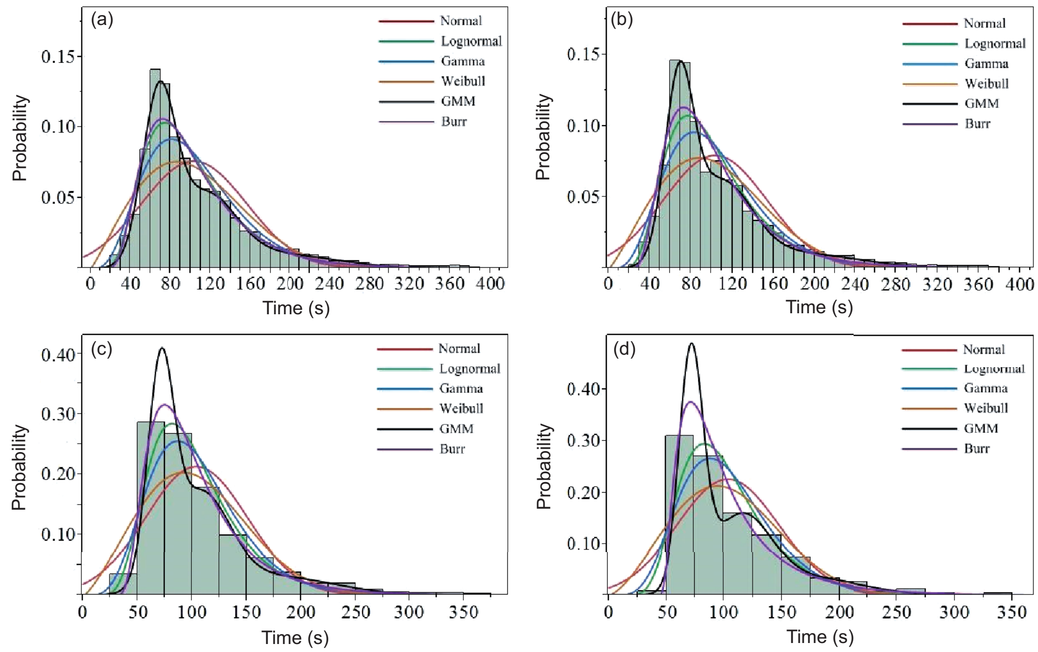

Figure 6 (Color online) Travel time distribution (TTD) for (6329, 6331) under different time intervals. (a) 5-min; (b) 10-min; (c) 30-min; (d) 60-min.

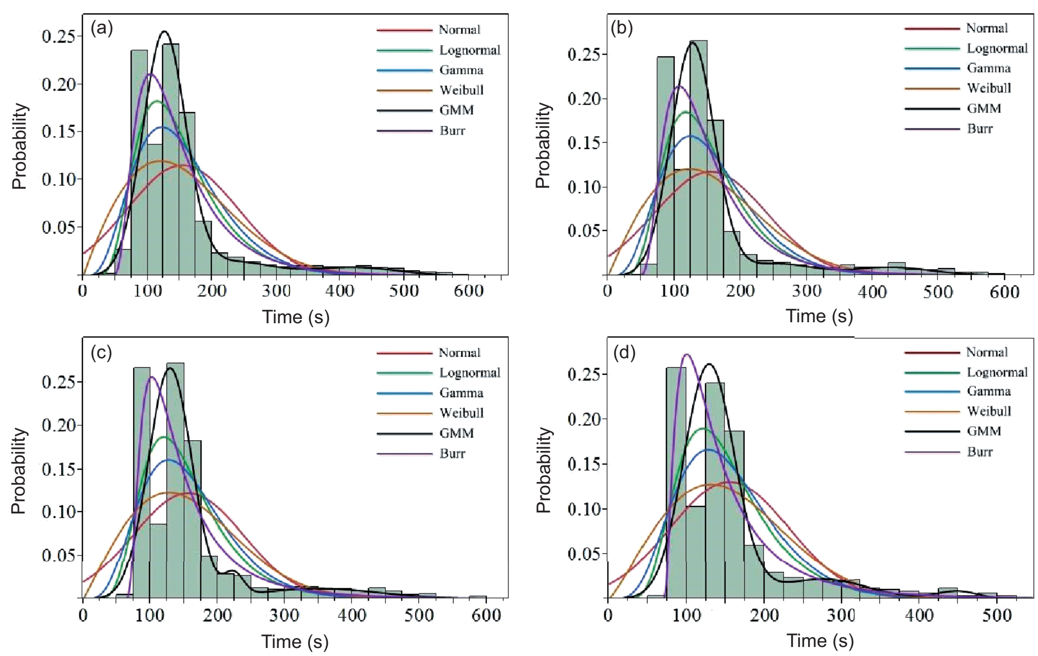

Figure 7 (Color online) Travel time distribution (TTD) for (6026, 6028) under different time intervals. (a) 5-min; (b) 10-min; (c) 30-min; (d) 60-min.

7 Conclusions and discussion

There has been a long-accepted consensus that TTD is fundamental for supporting transportation system reliability studies, particularly in relation to complex urban transportation systems, and many studies have been published with the aim of unraveling the nature of TTDs. However, due to the lack of adequate travel time data collection techniques, these studies have generally been based on a travel time data set with a limited sample size, resulting in inconsistent findings. Therefore, in this paper, large amounts of travel time data were collected using the emerging radio frequency identification

(RFID) technique, and extensive investigations into travel time data and its distributions were conducted to show the nature of travel time data. In addition, the aggregation time interval, an important determining factor for traffic variables, was also investigated together with the travel time data.

Three major findings can be drawn from the studies presented in this paper, and these are listed as follows. First, through empirical investigations into period-to-period travel time variability and TTDs, this paper shows that travel time data have a complex statistical structure: the TTD is in general peaky, multi-modal, and skewed to the right. This finding agrees with those of previous publications and is helpful in understanding the nature of travel time data. Second, based on previous findings, this paper shows that unimodal distribution models are unable to capture the complex statistical dynamics embedded in travel time data; therefore, use of a multistate distribution model is more appropriate for modeling TTDs. Specifically, in this paper, a three-component gaussian mixture model (GMM) is tested together with other typical unimodal distribution models, and results show that the GMM consistently outperforms the unimodal distribution models. Finally, results show that the aggregation time interval does not affect significantly the shape of TTDs: TTDs remains stable under different aggregation time intervals. Considering that time intervals are an important temporal factor for many transportation management and control applications, this finding is important as it supports the adoption of a single TTD model with various time intervals for transportation applications.

Future work is anticipated as follows. First, as travel time variability is an important topic and one of the primary focuses of transportation studies, as it is now possible to collect large amounts of travel time data, future work will be conducted to reveal patterns in travel time variability (in terms of vehicle-to-vehicle variability, data-to-day variability, and period-to-period variability). Second, TTD estimations could be further refined using mixture models with other types of component distributions, and the relationship between TTD and traffic conditions could also be investigated. This would be helpful for breaking travel time data into distributional groups with respect to different traffic condition states, and thus provide in-depth insights into travel time data. Finally, it is important to investigate adaptive procedures for estimating TTD models, as transportation management and control applications generally operate in an online fashion. In this sense, it is also important to generate a confidence level for travel time data using the TTD model, so that reliability of applications built upon travel time data can be measured in real time.

This work was supported by the National Natural Science Foundation of China (Grant No. 61573106), and the Jiangsu Provincial Key Laboratory of Networked Collective Intelligence (Grant No. BM2017002).

1 Bates J, Polak J, Jones P, et al. The valuation of reliability for personal travel. Transp Res Part E-Logistics Transp Rev, 2001, 37: 191–229
2 Hollander Y, Liu R. Estimation of the distribution of travel times by repeated simulation. Transp Res Part C-Emerging Technol, 2008, 16: 212–231
3 Taylor M A P. Travel time variability—The case of two public modes. Transp Sci, 1982, 16: 507–521
4 Wirasinghe S C, Liu G. Determination of the number and locations of time points in transit schedule design—Case of a single run. Ann Oper Res, 1995, 60: 161–191
5 Li R, Rose G, Sarvi M. Using automatic vehicle identification data to gain insight into travel time variability and its causes. Transp Res Record, 2006, 1945: 24–32
6 Dandy G C, McBean E A. Variability of individual travel time components. J Transp Eng, 1984, 110: 340–356
7 Kimpel T, Strathman J, Callas S. Improving scheduling through monitoring using AVL/APC data. In: Proceedings of the 9th International Conference on Computer-Aided Scheduling of Public Transport, CASPT. San Diego, 2004
8 Susilawati S, Taylor M A P, Somenahalli S V C. Distributions of travel time variability on urban roads. J Adv Transp, 2013, 47: 720–736
9 Van-Lint J, Van-Zuylen H. Monitoring and predicting freeway travel time reliability: Using width and skew of day-today travel time distribution. Transp Res Rec, 2005, 1917: 54–62
10 Pattanamekar P, Park D, Rilett L R, et al. Dynamic and stochastic shortest path in transportation networks with two components of travel time uncertainty. Transp Res Part C-Emerging Technol, 2003, 11: 331–354
11 Chang T S, Nozick L K, Turnquist M A. Multiobjective path finding in stochastic dynamic networks, with application to routing hazardous materials shipments. Transp Sci, 2005, 39: 383–399
12 Huang H, Gao S. Optimal paths in dynamic networks with dependent random link travel times. Transp Res Part B-Methodol, 2012, 46: 579–598
13 Sun L, Yang J, Mahmassani H. Travel time estimation based on piecewise truncated quadratic speed trajectory. Transp Res Part A-Policy Practice, 2008, 42: 173–186
14 May A, Bonsall P, Marler N. Travel time variability of a group of car commuters in north London. Working Paper, Institute of Transport Studies University of Leeds. Leeds, 1989
15 Clark S, Watling D. Modelling network travel time reliability under stochastic demand. Transp Res Part B-Methodol, 2005, 39: 119–140
16 Ma Z, Ferreira L, Mesbah M. Measuring Service Reliability Using Automatic Vehicle Location Data. Math Problems Eng, 2014, 2014: 1–12
17 Hellinga B, Izadpanah P, Takada H, et al. Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments. Transp Res Part C-Emerging Technol, 2008, 16: 768–782
18 Kazagli E, Koutsopoulos H N. Estimation of arterial travel time from automatic number plate recognition data. Transp Res Record, 2013, 2391: 22–31
19 Rahmani M, Koutsopoulos H N. Path inference from sparse floating car data for urban networks. Transp Res Part C-Emerging Technol, 2013, 30: 41–54
20 Zheng F, Van Zuylen H. Urban link travel time estimation based on sparse probe vehicle data. Transp Res Part C-Emerging Technol, 2013, 31: 145–157
21 Jenelius E, Koutsopoulos H N. Probe vehicle data sampled by time or space: Consistent travel time allocation and estimation. Transp Res Part B-Methodol, 2015, 71: 120–137
22 Richardson A, Taylor M. Travel time variability on commuter journeys. High Speed Ground Transp J, 1978, 12: 77–79
23 Fosgerau M, Karlström A. The value of reliability. Transp Res B, 2010,

43: 813–820
24 Sumalee A, Watling D, Nakayama S. Reliable network design problem: Case with uncertain demand and total travel time reliability. Transp Res Rec, 2006, 1964: 81–90
25 Kieu L M, Bhaskar A, Chung E. Public transport travel-time variability definitions and monitoring. J Transp Eng, 2015, 141: 04014068
26 Uno N, Kurauchi F, Tamura H, et al. Using bus probe data for analysis of travel time variability. J Intelligent Transp Syst, 2009, 13: 2–15
27 Polus A. A study of travel time and reliability on arterial routes. Transportation, 1979, 8: 141–151
28 Jordan W C, Turnquist M A. Zone scheduling of bus routes to improve service reliability. Transp Sci, 1979, 13: 242–268
29 Al-Deek H, Emam E B. New methodology for estimating reliability in transportation networks with degraded link capacities. J Intelligent Transp Syst, 2006, 10: 117–129
30 Burr I W. Cumulative frequency functions. Ann Math Statist, 1942, 13: 215–232
31 Zimmer W J, Keats J B, Wang F K. The burr XII distribution in reliability analysis. J Qual Tech, 1998, 30: 386–394
32 Fosgerau M, Fukuda D. Valuing travel time variability: Characteristics of the travel time distribution on an urban road. Transp Res Part C-Emerging Technol, 2012, 24: 83–101
33 Guo F, Rakha H, Park S. Multistate model for travel time reliability. Transp Res Record, 2010, 2188: 46–54
34 Vlahogianni E, Karlaftis M. Temporal aggregation in traffic data: implications for statistical characteristics and model choice. Transp Lett, 2011, 3: 37–49
35 Mazloumi E, Currie G, Rose G. Using GPS Data to Gain Insight into Public Transport Travel Time Variability. J Transp Eng, 2010, 136: 623–631