Application of Trajectory based Data for analyzing Travel Time Variability

Fahmida Rahman
Graduate Research Assistant, Department of Civil Engineering
University of Kentucky, Lexington, KY- 40508
Email:fra228@uky.edu

4th May 2018

Word count: 4,041 words text + 10*250 (7 FIGUREs and 3 TABLEs) = 6,541 words

ABSTRACT

With the advancement of Intelligence Transportation System (ITS), the identification of the location of moving objects using the Global Positioning System (GPS) has been an important part of traffic and transport analysis. Detailed traffic analysis can be carried out using these GPS based data. This article summarizes various functionality of the GPS based data using probe vehicle. These data can be processed by map matching for a specific segment of a facility type. Then, processed trajectory data is used to evaluate travel time variability and the level of service (LOS) of a selected urban street. For this purpose, a step by step process is discussed to evaluate the street from the viewpoint of travel time stability and reliability. Observing the variability in the travel time in that segment, statistical analysis is done to confirm the distribution of the observed travel time data. Finally, the selected distribution is exploited to assess the LOS of that street segment based on the concept of travel time reliability.

Keywords: Global Positioning System (GPS), level of service (LOS), Travel Time Variability.

INTRODUCTION

Real-time traffic information is important to observe the performance of the roadways. This information is available on freeways and major arterials. However, real-time traffic monitoring in limited on urban arterials and streets. Besides, one of the widely used source for real-time traffic monitoring data is Global Positioning System (GPS) coordinated data. Performance of the freeways and arterials are being assessed using this type of data (*1*). Wireless technology already made it possible to have real-time data using GPS tracking.

Most new mobile phones and smart phones have embedded GPS and are capable to access the Internet. In this system, the positions mobile phones within a specified geographical area are tracked anonymously over time. This process is called location referencing. The location referencing process is usually carried out by the wireless carrier and the resulting data consists of a randomly assigned probe vehicle identification number, time stamp, and position. This position indicates the latitude and longitude for corresponding time stamp and the tracking process gives the trajectory of a trip traversing through a roadway.

Availability of accurate and frequent positions of GPS probes over time for a large spatial region provides new opportunities to acquire traffic condition that is not readily available through traditional traffic surveillance systems. One of these opportunities is travel time reliability. Travel time reliability is defined as a measure of trip uniformity during a specific time in a specific location. It considers more than daily congestion and is attributed to route inconsistencies due to unexpected delay (*2*). Commuter face traffic congestion on a daily basis; they design their trip based on their experience of the network. However, unexpected congestion may cause delay and lead to them arriving late at their destination. Therefore, travel time reliability has been recognized as the most important factor that affects ridership satisfaction (*3*). Provision of accurate, robust and reliable travel time information helps road users to make more informed decisions and assists road managers in the management and operation of the network. This accurate travel time information can be utilized to quantify the existing situation by estimating travel time measures, whereas ideas on qualitative measures (e.g. level of service) is determined.

An important starting input of measuring travel time reliability is the probability distribution of travel time Most travel time reliability measures can be directly derived from continuous probability distributions once the probability distribution is known. Different statistical characteristics of probability distributions would produce various travel time reliability measures. Thus, choosing an appropriate probability distribution is essential.

In this study, a method is applied to report travel time reliability including Travel Time Index and Buffer Index. This method is implemented using Global Positioning System (GPS) based trajectory data in Chengdu, China. during four time periods. The remainder of the paper is organized as follows: Existing researches on travel time reliability measures for different facility types and methods for observing travel time distribution is covered in next section. Step by step method for evaluating an urban street segment from the viewpoint of travel time stability and reliability using highly precise probe data. Based on the selected travel time distributions of the urban corridor covered by the probe data, level of service (LOS) is then assessed following the concept of travel time reliability. Finally, the findings of the paper are summarized along with some concluding remarks and recommendations for future research in the last section.

RELATED WORKS

Limited attention has been paid off to assess LOS of an urban street based on travel time reliability concept. No established method is found to estimate travel time reliability of urban street that has millions of data per day recorded by probe vehicle. In (*4*), the author proposed a generic real-time estimation framework and examine the real-time sensitivity of bus probes to estimate urban street segment travel time. However, the bus probes data are incorporated with stopping time and dwelling time issues. This issue may lead a travel time which may not be representative to non-transit vehicle traffic conditions. Therefore, this is study can overcome this type of issue, as it uses a very high precision data recorded by multimodal trips in an urban street.

Previously, most travel time reliability research focused on the properties of the travel time distribution before developing measures of travel time reliability. Several studies applied simulation methods to construct the travel time distribution (*5,6*). Most of the relevant studies (*7,8)* fitted statistical distribution models by using real-life traffic data. One travel time reliability measure, travel time window, was derived exploiting a Normal distribution (*9*). Skewed statistical distributions are also most common in previous research. In (*10*), the author claimed that the Gamma distribution could best be fitted by the travel time data collected from arterial roads. Moreover, in (*7*), authors conducted travel time reliability research in a transportation network environment and found the Weibull distribution could be representative of travel time distribution, rather than the exponential distribution. Other studies concluded that the Lognormal distribution outperformed other skewed distributions in different traffic flow conditions (*7,11*). The Lognormal distribution was therefore adopted in measures (*8*), or to develop a new measure (*12*). These researchers used a single-model to represent the travel time distribution in a given time. Furthermore, these researches were conducted based on freeway or arterial data. Consequently, this study adopts both Lognormal distribution and Gamma distribution based on their outperformance and applies them to get the reliability measures for urban street.

Earliest researchers used travel time reliability measures as a practical performance measure (*13*). They recommended the use of the percent variation, misery index, and buffer time index and consequently the U.S. Department of Transportation (DOT) recommended 90th or 95th percentile travel time, buffer index, planning time index, and frequency of congestion for travel time reliability measures in a guidance document (*14*). Other researchers also mentioned about the standard deviation and coefficient of variation (COV) to evaluate the reliability of the network (*15, 16*). This study also endeavors to measure these matrices and assess the level of service of an urban street.

STUDY CORRIDOR AND DATA DESCRIPTION

Selection of the Corridor

The study corridor was selected from the city Chengdu, China. Urban street segment of a mile was chosen for this study purpose (FIGURE 1). The following subsection gives an overview on choosing such a small segment based on the challenges with the datasets used.

FIGURE 1: Study corridor (Source: OpenstreetMap).

Data Preparation

Didi Chuxing has served more than 450 million users with a full range of mobility services, including Taxi, Express, Premier, Luxe, Hitch, Bus, Minibus, Designated Driving, Car Rental, Enterprise Solutions and Bike-Sharing. As of December 2017, DiDi completes more than 25 million orders every day. DiDi Datasets report a certain trip with points of trajectory having a precision of 3 seconds. It includes trip ID and order ID as well. The trajectories are presented as latitude and longitude in the file. It provides 30 days data which is from November 1st to November 30th. Individual day contains more than 40 million records which makes the data processing much challenging.

Firstly, a very small subset of data is processed to see how the trajectory looks like in the chosen study corridor (FIGURE 2(a)). Then the data extraction follows a method that is presented in FIGURE 3. The method is divided into three major steps. These include: Map Matching, Export Data and Travel Time Calculation.

Map Matching

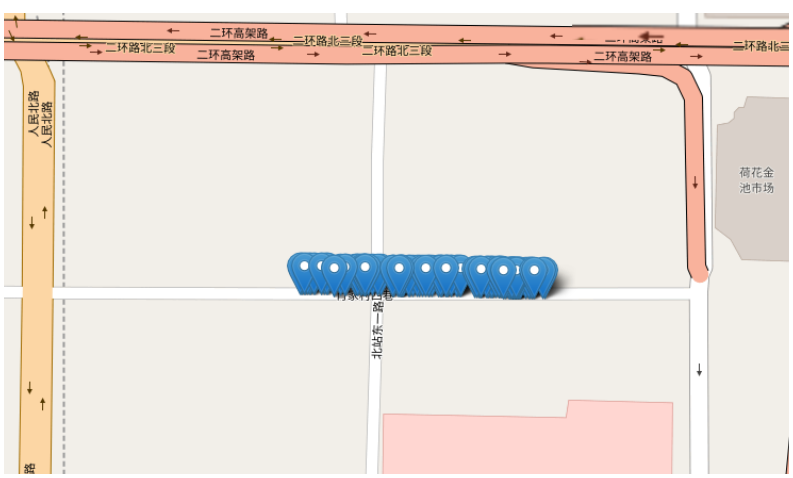
Shapefiles for the roads of China is collected from online website (*18*). ArcMap is utilized to read the shapefile. The DiDi datasets is also imported to the GIS environment. The selected corridor is separated into different layer through ArcMap. Afterwards, 'Join' tool in ArcMap facilitates merge the data to the corridor based on the latitude and longitude attributes. The final trajectory data traversing through that corridor is presented in FIGURE 2(b).

Export Data

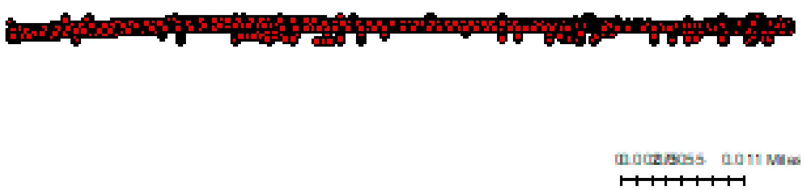
The final output given by the Map Matching process is exported as a .csv file. This trajectory file is then imported in python which helps to process big data and, therefore, critical analysis.

Travel Time Calculation

This step calculates distance from the latitude/longitude coordinates. Due to some errors in the reported GPS data, negative distances are also found, and filter process is done for that. Then, instantaneous speeds are calculated using the distances, which are further aggregated every five minutes in order to avoid short duration travel time fluctuation (*19*). Finally, segment travel time is computed using those aggregated speeds and segment length. To make sure the process went well, a distribution of travel time over the 16 hours (6 am. to 10 pm.) is plotted and the plot seems reasonable showing all the peaks (FIGURE 4). The study corridor usually suffers from recurring traffic congestion during the afternoon peak hours (4 pm - 7 pm) every weekday; and uncongested flow the rest of the day.



(a)



(b)

FIGURE 2: (a) GPS sensor spatial layout. (b) Trajectory data after spatial join in the selected corridor.

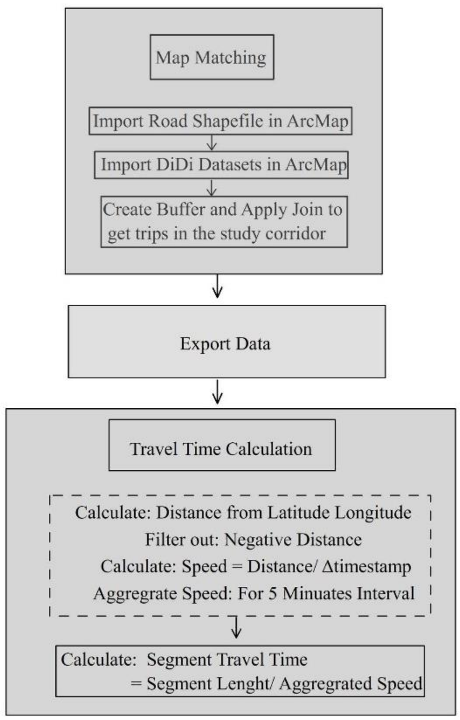


FIGURE 3: Methodological flow chart for data preparation.

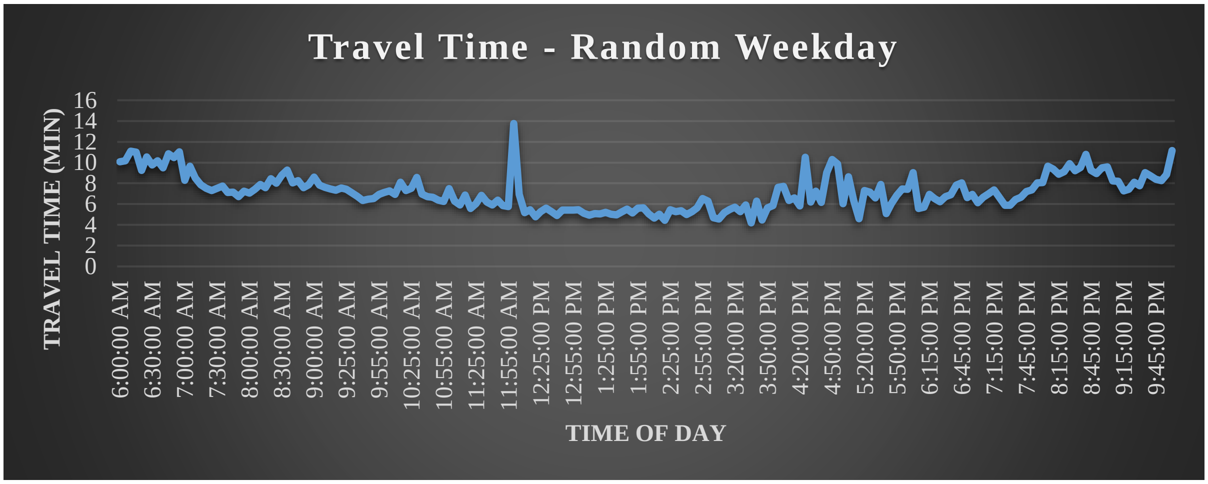


FIGURE 4: Travel time variation for a typical day.

METHODOLOGY

Hypothesis Testing for the Distribution

The accurate measurement of travel time reliability depends on the selection of continuous probability distribution (*16*). This study examines the conformity of the observed travel time distributions to theoretical distributions to allow for a statistical analysis. According to existing research discussed previously, Lognormal distribution and Gamma distribution outperforms among the probability distribution models. Thus, the observed distributions of the travel time datasets in this study appear to follow lognormal distributions and Gamma distribution. The lognormal distributions and Gamma distribution equation is given by Equation 1 and Equation 2.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x}\exp[-(\log x - \mu)^2 / 2\sigma^2]$$

$$(x > 0),\ 0,\ \ (x \le 0) \qquad (1)$$

Here, μ and σ represent the mean and standard deviation of $\log x$, and x represents the travel time.

$$f(x;k,\theta) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}$$

$$x > 0 \ and\ k,\theta > 0 \qquad (2)$$

Here, x represents travel time, shape and scale is defined by k and θ.

To calculate travel time reliability using the distribution, one of the distribution is chosen based on the significant performance for the observed data. For this purpose, one-sample Kolmogorov–Smirnov (K-S) Test is performed for both observed and theoretical travel time distributions. The null hypothesis and the alternative hypothesis in the K-S test can be defined as

$H_{0:}$ The travel times statistically follow the probability distribution models.

H_1: The travel times do not follow the distributions generated from the probability models.

Significance levels of $\alpha = 0.05$ is selected to be the thresholds for indicating whether the travel times follow those generated probability distributions. The resulting value of the one-sample K-S test is the maximum distance between an empirical cumulative probability function (ECDF) generated by the travel times and a cumulative probability function (CDF) produced by one of the probability distribution models.

Travel Time Reliability Measures

Once the distribution is confirmed from the two models been considered, the travel time reliability measures can be calculated based on it. Following sub-sections gives an idea to get those measures relying on the distribution model.

Moment-based Travel Time Reliability Measures

This study attempts to assess the LOS of the urban street from the viewpoint of travel time reliability. As such, it is necessary to use the travel time distribution of the segment to evaluate the LOS for different time periods. Among the various factors that affect the LOS of a road segment, this study focuses on both the efficiency and stability of the corridor. Thus, two momentum-based indices are adopted for the LOS. The average travel time for 1 mile is used to evaluate the efficiency of the segment. The coefficient of variation (COV) of the travel time is used to evaluate the reliability of the segment. Accordingly, the indices using the average and variance of travel time are regarded to be valid for analyzing LOS of the road segment at least. These indices can be calculated from the travel time distribution for the selected segment, and cumulative distributions are used to evaluate the LOS for different period of the day as well as days of the week (weekdays in this case).

Percentile-based Travel Time Reliability Measures

Another approach to report travel time reliability is percentile travel time. Nth percentile travel time is the value that splits the dataset into two parts. The lower part includes the N percent travel time and the upper part includes the rest of the data. The Percentile method is not sensitive to the presence of outliers or rare events as it ranks observations from smallest to largest. Hence it is robust to the presence of any abnormality in the data. This study uses 85^{th} percentile travel time to make a comparison among the LOS of weekdays.

Travel time index (TTI) is a measure of the average travel time that would be experienced by the system users during the analysis period. While the Buffer Index (BI) is a measure of the amount of additional time (above the average) needed to include 95% of the travel data points. In this study, the Buffer Index can be used to compare the LOS for different time of the day. Furthermore, analysis is used to imply that Buffer Index could be able to provide more information to the traveler in terms of travel time reliability and hence the level of service of a facility than the Travel Time Index.

ANALYSIS AND RESULTS

The selected model i.e. Lognormal distribution and Gamma distribution are used for theoretical distribution of the datasets. The distributions are generated for four different periods of the day: Morning (6 am – 9 am), Mid-day (9 am – 3 pm), Evening (3 pm – 6 pm) and Night (6 pm to 10 pm). In FIGURE 5 and FIGURE 6 the bar graph characterizes the observed distribution, which appears to follow a Lognormal distribution because of the longer right-hand tail and also follow Gamma distribution. The line shown in FIGURE 5 and FIGURE 6 gives the theoretical distribution with the same mean and variance as the observed travel time distribution. The shape of this line resembles that of the observed travel time distribution. K-S test is applied for both observed and theoretical travel time distributions.

TABLE 1 shows the results from the K-S test. The null hypothesis is accepted for maximum of the scenarios while fitting Lognormal distribution. Alternatively, two cases already got rejected for the observed travel time fitted in Gamma distribution. The results indicate that Lognormal distribution is more representative for the observed trend of the experimental data. However, the Midday scenario for each distribution is rejected. The reason can be the interruptions between the GPS receivers and satellites during this period which can be from the adjacent tallest building. Although more detailed examination should be carried out for the conformity of the travel time distribution, this study assumes that travel time distribution conforms to a lognormal distribution for the further analysis.

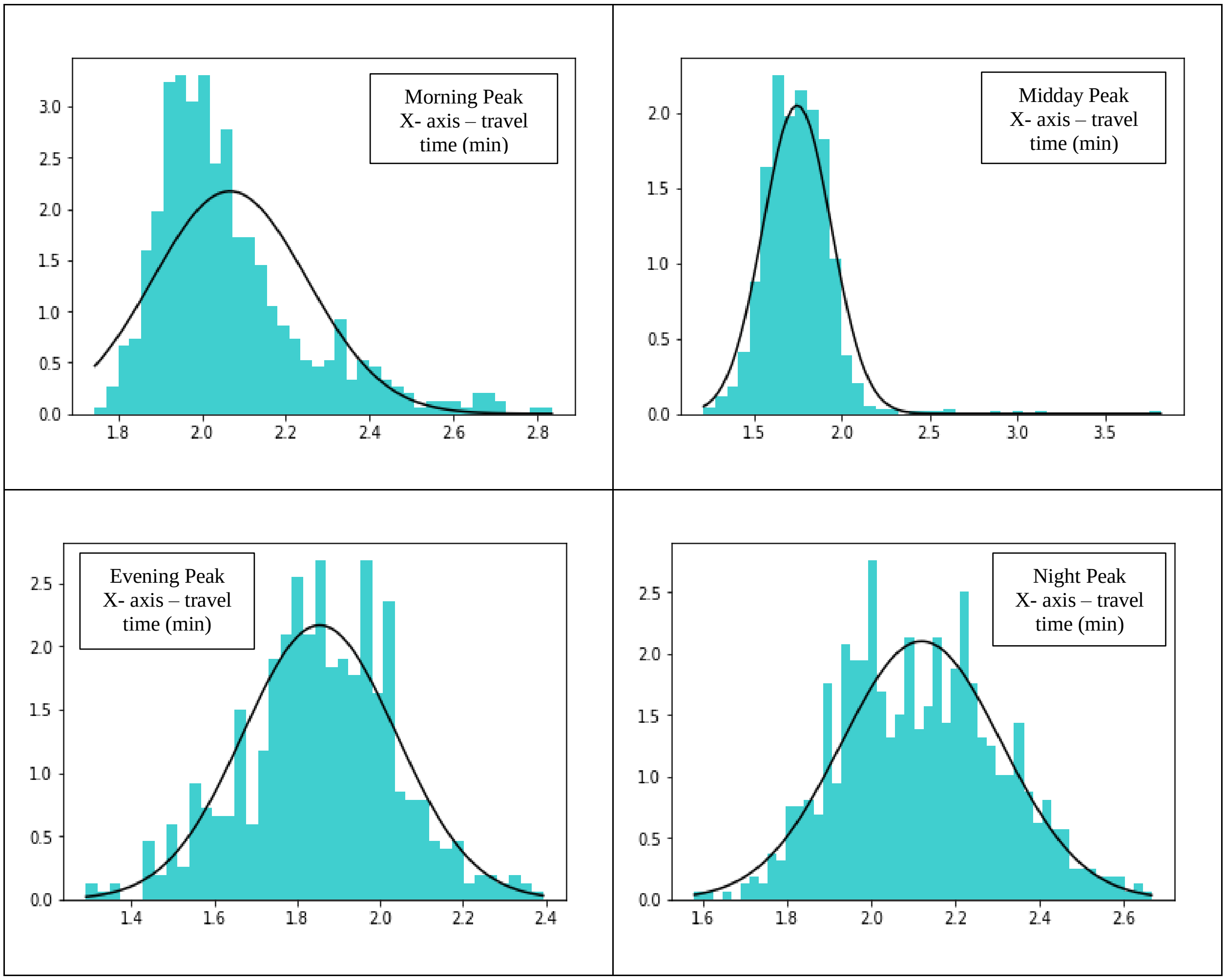


FIGURE 5: Lognormal vs Histogram of Travel Time.

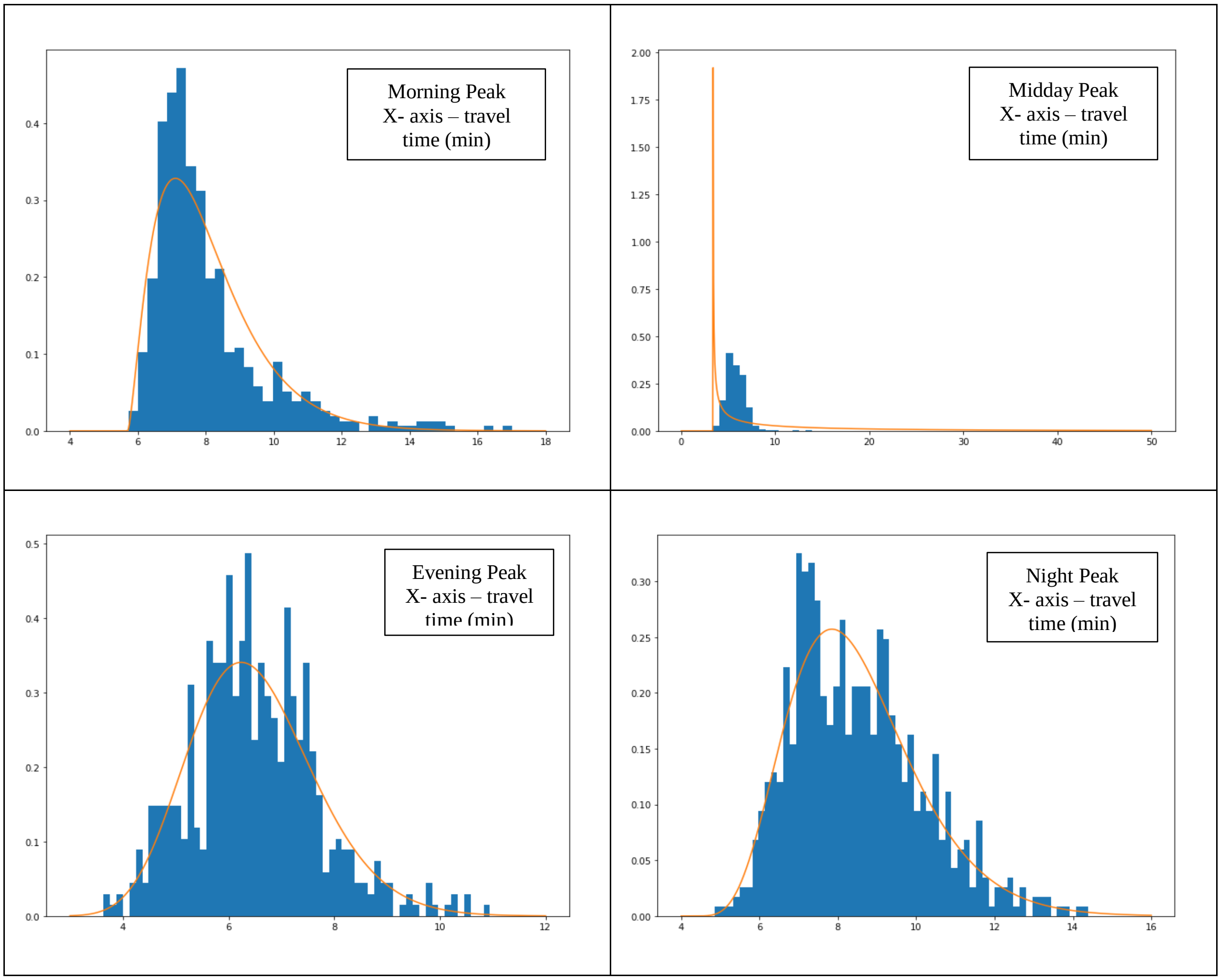


FIGURE 6: Gamma vs Histogram of Travel Time.

TABLE 1: K-S Test for Observed and Theoretical Travel Time Distribution

Test for Log Normal Distribution			
Time of the Day	**Statistic**	**p-Value**	**Result**
Morning Peak (6 am to 9 am)	0.05185	0.09764	Not Rejected
Mid-day Peak (9 am to 3 pm)	0.05013	0.00855	Rejected
Evening Peak (3 pm to 6 pm)	0.03557	0.47942	Not Rejected
Night-time Peak (6 pm to 10 pm)	0.04289	0.13002	Not Rejected

Test for Gamma Distribution			
Time of the Day	**Statistic**	**p-Value**	**Result**
Morning Peak (6 am to 9 am)	0.086879	0.00043	Rejected
Mid-day Peak (9 am to 3 pm)	0.38379	0.0	Rejected
Evening Peak (3 pm to 6 pm)	0.03658	0.44194	Not Rejected
Night-time Peak (6 pm to 10 pm)	0.04177	0.14953	Not Rejected

Following the Lognormal distribution of the observed trend, FIGURE 7 shows the cumulative distributions of both the average travel time for 1 mile and the COV of the travel time. The lines shown for the weekdays (Monday, Tuesday, Wednesday, Thursday and Friday) correspond to the cumulative distributions of travel time for individual days. In both figures, the Tuesday line appears to be the left most, which indicates that the LOS of the road segment in Tuesday is normally better than the other days of the week (i.e. weekdays). Moreover, percentile speeds calculated from the distribution also confirms the LOS for Tuesday is comparatively better. TABLE 2 shows a higher 50th percentile speed and 85th percentile speed for that day with respect to other days.

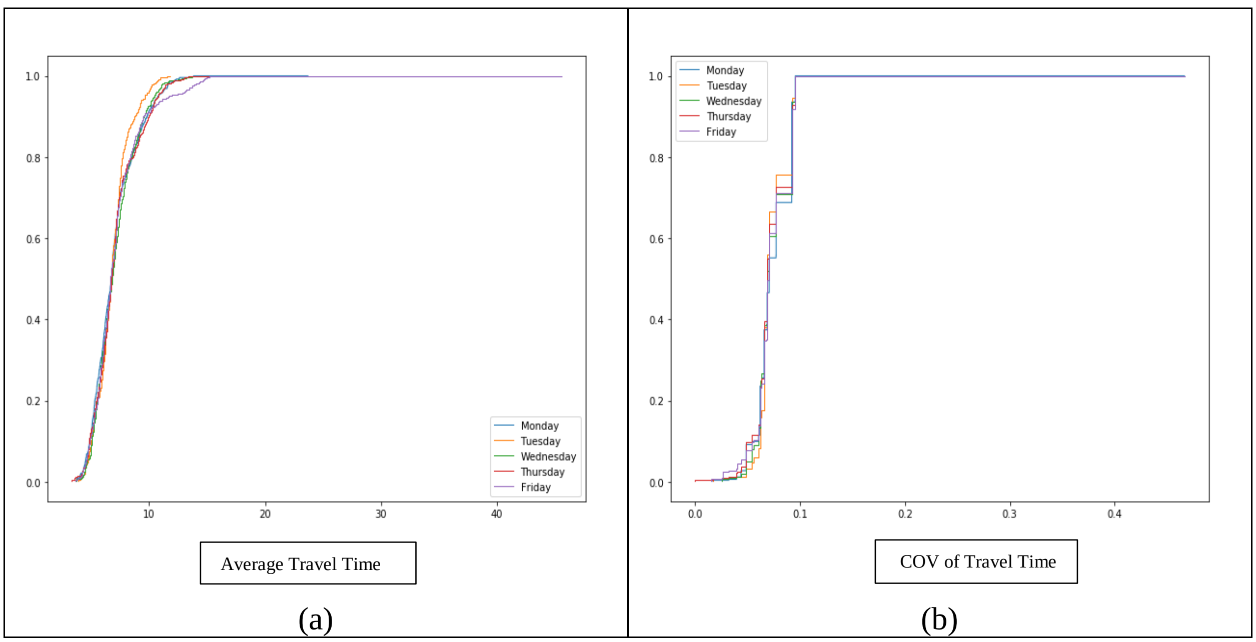


FIGURE 7: Evaluation of LOS of the Corridor using Momentum based Reliability. (a) Cumulative distributions of the average travel time. (b) Cumulative distributions COV of the travel time.

TABLE 2: Percentile Speeds from the distribution

Weekdays	*50th Percentile Speed*	*85th Percentile Speed*
Monday	18.9870143764	13.9944766387
Tuesday	19.5368058049	15.7939929967
Wednesday	18.7879173218	14.1917239226
Thursday	18.9701198746	13.9104659132
Friday	19.2376592721	14.4700534145

To compare the LOS for different periods of a day, it is recommended to compare the Buffer Index (BI). As the road segment shows same LOS in term of speed, it is wise compare BI to have a conclusion. Because, BI is inconclusive in providing an accurate measure for defining LOS. From the existing researches, higher BI values may suggest to road users that the journey time has a higher variability, whereas, lower values of BI may suggest a reliable travel time is available but at two levels (*20*):

- a reliable short travel time in free flow conditions.
- a reliable longer journey time but in congested conditions.

In TABLE 3, the higher value for BI shows for AM peak and Evening suggesting a higher variability of travel time during that period. However, lower values are found for the remaining two periods. The lowest one implies reliable congested period during night peak. This finding also supported by the average speed and TTI for that period. Average speed is lowest for night among the four scenarios. Moreover, it shows higher TTI which necessarily means longer travel time indicating congestion. The second lowest BI implies reliable short travel time for Mid-day peak. Considering average speed, it is highest among the four scenarios. Furthermore, lowest TTI indicates more reliability during the Mid-day. Thus, the selected corridor better serves during Mid-day in a much reliable manner.

TABLE 3: Percentile based Travel Time Reliability

Time of the day	*LOS*	*Average Speed*	*Travel Speed*	*TTI*	*BI*
AM peak	F	16.5	7.5 – 22.5 mph	3.76	42.26%
Mid-day peak	F	22.7	2.8 – 38.1 mph	2.75	26.6%
Evening peak	F	20.5	11.7 -35.3	3.03	32.5%
Night peak	F	15.7	8.9 – 24.5 mph	3.97	16%

**LOS estimated based on HCM 2010*

CONCLUDING REMARKS

This study discusses a method to report travel time reliability including Travel Time Index and Buffer Index. This method is implemented using Global Positioning System (GPS) based trajectory data in Chengdu, China. during four time periods. Evaluating of an urban street segment from the viewpoint of travel time stability and reliability is done using highly precise probe data. Based on the selected travel time distributions of the urban corridor, level of service (LOS) is then assessed following the concept of travel time reliability.

Finally, some conclusive remarks can be summarized. As BI gives more indication of travel time variability, BI and TTI are not showing any specific correlation for that corridor with LOS F. After the assessment of LOS, the night time (between 6 pm to 10 pm) scenario shows reliable longer journey in a congested situation based on TTI, BI and average speeds. Moreover, the midday (between 9 am to 3 pm) scenario shows a reliable short travel time based on the same indices. Lastly, during Thursday, the LOS is comparatively better than other days depicted by cumulative distribution and percentiles speed.

FUTURE RESEARCH

This research performs analysis based on 30 days data. As a part of further research, data for an entire 1-year period is needed to more accurately perform test. Analysis should be carried on several segments instead of one to compare the LOS among the routes. Besides, assessment of LOS can also be done for two directional major roadways to make it more robust.

ACKNOWLEDGEMENTS
The authors would like to express thanks to "Data source: Didi Chuxing" for providing large amount of trajectory data for individual trips.

REFERENCES

1. http://www.baltometro.org/phocadownload/Publications/GPS2003.pdf
2. Second Strategic Highway Research Program (US), Kimley-Horn and Associates and Parsons Brinckerhoff, 2011. *Guide to Integrating Business Processes to Improve Travel Time Reliability*. Transportation Research Board.
3. Gaffney, J., 2006, August. Understanding network performance information provided to users. In *International Seminar on Intelligent Transport Systems (ITS) in Road Network Operations, 2006, Kuala Lumpur, Malaysia*.
4. Pu, W., Lin, J. and Long, L., 2009. Real-time estimation of urban street segment travel time using buses as speed probes. *Transportation Research Record: Journal of the Transportation Research Board*, (2129), pp.81-89.
5. Guo, F., Rakha, H. and Park, S., 2010. Multistate model for travel time reliability. *Transportation Research Record: Journal of the Transportation Research Board*, (2188), pp.46-54.
6. Hollander, Y. and Liu, R., 2008. Estimation of the distribution of travel times by repeated simulation. *Transportation Research Part C: Emerging Technologies*, *16*(2), pp.212-231.
7. Al-Deek, H. and Emam, E.B., 2006. New methodology for estimating reliability in transportation networks with degraded link capacities. *Journal of intelligent transportation systems*, *10*(3), pp.117-129.
8. Pu, W., 2011. Analytic relationships between travel time reliability measures. *Transportation Research Record: Journal of the Transportation Research Board*, (2254), pp.122-130.
9. Lomax, T. and Margiotta, R., 2003. *Selecting travel reliability measures*. The Institute.
10. Polus, A., 1979. A study of travel time and reliability on arterial routes. *Transportation*, *8*(2), pp.141-151.
11. Rakha, H., El-Shawarby, I. and Arafeh, M., 2010. Trip travel-time reliability: issues and proposed solutions. *Journal of Intelligent Transportation Systems*, *14*(4), pp.232-250.

12. Kaparias, I., Bell, M.G. and Belzner, H., 2008. A new measure of travel time reliability for in-vehicle navigation systems. *Journal of Intelligent Transportation Systems*, *12*(4), pp.202-211.
13. Lomax, T. and Margiotta, R., 2003. *Selecting travel reliability measures*. The Institute.
14. Texas Transportation Institute, & Cambridge Systems, I., 2006. Travel Time Reliability: Making It There On Time, All the Time. *Federal Highway Administration Office of Operations*. Retrieved from http://www.ops.fhwa.dot.gov/publications/tt_reliability/
15. Rashidi, S., Pant, R., Hooper, A. and Baker, C., Travel Time Reliability Measure Using Global Positioning System Data.
16. Uno, N., Kurauchi, F., Tamura, H. and Iida, Y., 2009. Using bus probe data for analysis of travel time variability. *Journal of Intelligent Transportation Systems*, *13*(1), pp.2-15.
17. https://gaia.didichuxing.com
18. http://www.mapcruzin.com/free-china-country-city-place-gis-shapefiles.htm
19. Yang, S., Malik, A. and Wu, Y.J., 2014. Travel time reliability using the Hasofer-Lind-Rackwitz-Fiessler algorithm and kernel density estimation. *Transportation Research Record: Journal of the Transportation Research Board*, (2442), pp.85-95.
20. Russell, M.B., 2014. Travel Time Reliability and Level of Service.