Estimation of travel time distributions for urban roads using GPS trajectories of vehicles: a case of Athens, Greece

Peer-reviewed author version

ADNAN, Muhammad; Gazder, Uneb; YASAR, Ansar; BELLEMANS, Tom & Kureshi, Ibad (2021) Estimation of travel time distributions for urban roads using GPS trajectories of vehicles: a case of Athens, Greece. In: PERSONAL AND UBIQUITOUS COMPUTING, 25(1), p. 237-246.

DOI: 10.1007/s00779-020-01369-4
Handle: http://hdl.handle.net/1942/30669

Estimation of Travel Time Distributions for Urban Roads Using GPS Trajectories of Vehicles- A case of Athens, Greece

Muhammad Adnan[a,*], Uneb Gazder[b], Ansar-ul-Haque Yasar[a], Tom Bellemans[a], Ibad Kureshi[c]

[a]*UHasselt – Universiteit Hasselt, Transportation Research Institute (IMOB), Agoralaan, 3590 Diepenbeek, Belgium*
[b]*Department of Civil Engineering, University of Bahrain, Bahrain*
[c]*Inlecom Systems, Square de Meeus 38/40,1000 Brussels, Belgium*

Abstract

Investigating travel time distribution and associated variability is important for a variety of transport planning, traffic management and control projects. Studies that investigated travel time distribution tend to be limited to explore changes in characteristics of distribution with respect to space and time-of-day. Given the availability of big data set that contains seven different types of vehicle trajectories in the city of Athens for around 56,000 trips which are traversing on more than 1.8 million road links, this study presents the detailed investigation of travel time distribution in different spatio-temporal settings. The study considered four different types of urban roads and six time intervals along with consideration of weekdays and weekends. The empirical investigation employed Kruskal-Wallis, Chi-square and Kolmogorov-Smirnov tests to fit travel time data into seven uni-modal statistical distributions that are found in the literature to describe travel time distribution. It is found that lognormal distribution outperformed other distribution, and all of the considered categories of travel time data are well-fitted to this distribution. Additionally, parameters of log-normal distribution for different categories of travel time data are not significantly different from each other, which led to the conclusion that travel time distribution is roughly independent of space and time, which is in agreement with a few earlier studies that are limited in their scope especially in relation with availability of data. With this important finding, this study estimate values of travel time variability for different classes of individuals employing a standard approach that requires time-of-day independent standardized distribution of travel time. It is estimated that for Athens population value of travel time variability is approximately half of the value of travel time. This is useful to carry out cost-benefit analyses for mobility-related projects in Athens, Greece.

Keywords: Travel time distribution, GPS-based big data, Value of travel time variability, Statistical distribution, Athens city

[*] Corresponding Author, Tel: +32 (0)11 269111
Fax: +32 (0)11 269199, E-mail: muhammad.adnan@uhasselt.be

1. Introduction

Travel time is an important measure for travelers which affects their choice of mode, route, departure time and other decisions related to trip planning (Beaud et al., 2016, Adnan et al., 2009, Rui et al. 2015, Baqueri et al. 2019, Petrik et al. 2018). To address urban mobility, travel times on the road network and its reliability are used as key indicators. This information is being disseminated to the users regularly to make informed trip decisions (Ma et al., 2017, Khun et al 2013). It becomes essentially important for traffic engineers and planners to have in-depth knowledge of network-wide travel times as it provides inputs for transportation planning and management at many levels. Some of the important applications of travel time include traffic assignment models (Chen and Zhou, 2010), microscopic simulation models (Pel et al., 2012) and economic analysis of highway facilities (Peer et al., 2012).

Studies have also shown that travelers value travel time variability (VTTV) significantly, which is the results of inherent randomness in demand, supply and network performance (Jenelius 2012). Therefore, many recent travel decision models (mode, route and departure time choice) incorporate TTV. Recently, it is emphasized and recommended to include the cost of TTV into the cost-benefit analysis (CBA) for transportation projects (Eliasson 2019, Fosgerau 2017, NZTA 2016, OECD 2016). Fosgerau (2017) developed a method for estimating the ratio of the value of travel time (VTT) and the value of travel time variability (VTTV), which is termed as travel time reliability ratio (TTRR). Zang et al. (2018) stated that TTRR is a dimensionless quantity and can be used in CBA application as recommended by the Organization of Economic Cooperation and Development (OECD, 2016). The calculation of TTRR requires knowledge of standardized travel time distribution function. Therefore, a well-fitted theoretical probability distribution can provide a reasonable estimate of TTRR. Exploration and understanding of travel time distribution characteristics are therefore crucial.

A significant amount of studies exists that analyses and fitted various available statistical distributions onto travel time datasets. Single-mode distribution (one kind of standard distribution) are found more commonly in the literature, such as *Lognormal* (Emam and Al-Deek, 2006, Kaparias et al, 2008, Rakha et al, 2010, Arezoumandi 2011, Lie et al 2014, Chen et al 2018), *Gamma* (Polus, 1979, Nie et al 2012, Lie et al 2014, Chen et al 2018), *Weibull* (Emam and Al-Deek 2006, Lie et al 2014, Chen et al 2018), *Burr* (Susilawati et al 2013, Lie et al 2014, Kieu et al 2015) and *Stable* (Fosgerau and Fukuda, 2012) distributions. Multi-mode distributions are also fitted in some studies as this approach provides much improved model fitting in comparison to single-mode models by considering travel times dataset comprising from two (or more) populations. Within this approach; *normal mixture model* (Guo et al 2010, Yang and Wu 2016), *Lognormal mixture model* (Kazagali and Koustopoulos 2012, Yang and Wu 2016), *Gamma mixture model* (Yang and Wu 2016) and a *finite mixture of the regression model* (Chen et al 2014) are investigated. Maximum likelihood estimation and expectation-maximization algorithms were used to fit multi-mode distribution, which require more computational resources in comparison with single-mode distribution fitting.

Chen et al (2018) mentioned that studies that explore travel time and their distributions used datasets that are collected mainly from one road type. A few studies have used datasets that are collected from different roads, however, in their analysis, they have treated different roads equally. Additionally, travel times can rapidly change due to changes in various circumstances, such as road types, their geometry, type of traffic control employed, presence of incidents and variation in demands. Travel time variability can also differ due to these circumstances, and to appropriately capture this, detailed spatial and temporal exploration of travel time distribution is required. On the other hand, empirical studies that are limited to one-single road and few links found that travel time distribution is roughly independent of time-of-day (Fosgerau and Fukuda 2012). To this end, we found only one single study (Chen et al 2018) that investigated four single-mode distributions such as normal, lognormal, gamma and Weibull, for different road types (Urban expressways, Auxiliary roads of urban expressways, Major roads and secondary roads) and at different time periods(peak hour and off-peak hour during weekdays and weekends) using vehicle GPS probe data of Beijing, China. They found out that lognormal distribution is superior over the other three distributions, however, differences exist in relation to time periods (especially in relation to peak periods) and road types. However, the study does not provide estimated parameters of the distributions, so that the actual magnitude of differences can be measured nor it is helpful to be applied to other similar cities like Beijing. Additionally, the dataset used was came from taxis, and therefore, cannot be generalized with higher confidence. Therefore, it is required to have further investigation to confirm the findings of earlier studies about the dependency of travel time distribution with respect to space and time.

This study further extends the work carried out by Chen et al (2018) by using the similar methodological steps. It attempts to overcome a few shortcomings of the previous study, such as consideration of GPS probe data not just from taxis but from all vehicles in the traffic stream, investigating more distributions (in total seven) instead of only four, four road types with six-time periods along with consideration of weekdays and weekends and using a significantly large dataset available for Athens, Greece. Additionally, this study also provides the estimates of TTRR, and based on that VTTV is also estimated for various population classes and trip purposes utilizing earlier reported values of VTT for Athens, Greece, so that it can be applied in CBA. The remainder of the paper is structured as follows: Section 2 presents details of the methods and dataset employed in the study, section 4 presents results of fitted distributions, estimation of TTRR and VTTV and also discuss the results and compared the findings from previous literature followed by concluding section.

2. Data and Methods

2.1 Data Characteristics and Preparation

GPS trajectory data used in this study was provided by Vodafone Innovus S.A for the vehicles traveling in Athens, Greece. The data is collected at an average rate of 0.96 points per minute, and accommodates rich information such as GPS positioning, time, vehicle speeds, headings, engine status, driving events (e.g. brakes), and fuel levels. According to Vodafone Innovus S.A (a technology provider for fleet management) a device is installed in the vehicle, that include a SIM card and connect via GSM network to transmit the data to the server. While the vehicle in motion, the data was generated more frequently (i.e. for every 300m to 800m distance covered by the vehicle) and when certain event occur (i.e. lane changing etc.) it also generate a data. However, when the vehicle is in stopped condition it generates data with lesser frequency. This study only utilizes GPS positioning, vehicle speed and engine status, to find out trips of the vehicle and then estimated time consumed to travel on the specific type of roads in a trip. The dataset consists of data of 1585 vehicles of 7 different types (such as Passenger car, bus, minibus, taxi, minivans, vans, minitrucks) during 3 months of operation between September to November 2018. The total size of the data is 66.4GB. Usual cleaning methods were employed on the dataset, such as removal of a few outliers (GPS points) that correspond to extremely high instantaneous vehicle speed (i.e. speed more than 170km/hr) after examined its distribution as in this study we examine roads in Athens apart from freeways/motorways/expressways, where speeds limits are in the range of 20-110 km/hr. In addition to this, duplicate points with identical point IDs and timestamp are also removed from the dataset. Furthermore, the distance between the two consecutive points is also examined, which is found in reasonable bounds (i.e. from 0 to 1960m, with a typical sampling rate of approximately 1 minute this corresponds to the speed of 0 and 117 km/hr respectively).

After cleaning the dataset, the next step is to extract vehicle trips trajectories (i.e. from origin to destination), which then later provide travel time. There are a variety of approaches for separating trips: by positional attributes (e.g. taximeter being switched on or off), by temporal cycles (e.g. daily trips), by substantial displacement (e.g. if the next point is at least 5km away), and by temporal gaps between points (stop points) (e.g. no movement for at least 15min). For this study, the temporal gap between stop point (at least 15 minutes) is used for extracting trips from vehicle trajectory data. In addition to this, the trip extraction algorithm also ensures that when positions remained within a small area (instead of only at a stopping point) during a time interval of 15 minutes, that small area is treated as a stopping point. This is adopted for toleration of position measurement errors. From the dataset, in total, we extracted 56,000 trips considering a rectangular area that include an urban area of Athens, Greece.

The next step is to map match extracted trips so that links travel time can be determined. This is the travel time taken by the vehicle to go from start of the link till the end. We used an offline algorithm proposed by Newson and Krumm (2009), which is implemented in the GraphHopper map-matching library (https://github.com/graphhopper/map-matching) that uses the road network extracted from the OpenStreetMap for matching trip trajectories. The algorithm employs a hidden Markov model (HMM) to find the edges in the road network that are more likely to characterize the route, given the locations of the individual GPS coordinates. It is used in many research studies and has shown reliable results (Maalleson et al 2018). Overall, 98% of extracted trips are successfully map-matched to the OpenStreetMap network. For a few random trips, we examine the paths/route obtained after map-matching by visualizing them in QGIS software, and the results seem appropriate. For each link/edge of the network for which we have trips map-matched, we estimated the average travel time. After which, the extreme values were omitted from the datasets to account for any errors in map-matching.

2.2 Methodological Details

Links of the considered road network were categorized into four categories, namely; primary, secondary, tertiary and residential roads. Each of these categories will have different geometric characteristics hence their travel time distribution can differ. The travel time was classified based on the type of link and time of the day. Time of the day variable was divided into six intervals, which are as; 12 midnight to 6:30 a.m., 6:30 to 10:00 a.m., 10:00 a.m. to 1:30 p.m., 1:30 to 5:00 p.m., 5:00 to 8:30 p.m., and 8:30 p.m. to midnight. These classifications are taken from the previous study of Athens by Stathopoulos and Karlaftis (2001), which was the most relevant found for the study area. The temporal variations in traffic for an area can be considered constant over the years as it has been observed in other studies recently (Batterman et al., 2015). Moreover, travel time on weekdays and weekends were also recorded separately. It should be noted at this stage that we did not find any study in which such detailed classification of travel time has been considered. Each link will have a different length which would affect its travel time, therefore, the travel time was converted to a scale of time required to travel a distance of 100m on that link similar to the study conducted by Chen et al (2018).

Once the travel time for considered road categories is spatially and temporally classified, each category was compared with others to investigate whether certain categories can be merged. It was done using a Kruskal Wallis test, which has been recommended as a non-parametric test for comparing 3 or more distributions (McKight and Najab, 2010). The categories, after going through the merging wherever required, were compared with standard distributions using chi-square and Kolmogorov-Smirnov (KS) tests. The following distributions were compared in this study for goodness-of-fit with the travel time data; Normal, Exponential, Gamma, Log-normal, Chi-square, Weibull and Rayleigh. All these distributions include the range of data from zero to infinity and are commonly used statistical distributions. Among the tests for goodness of fit, the Chi-square test is a non-parametric test normally used for estimating the dependence of a variable to specific distributions (McHugh, 2013; Sharpe, 2015). KS test is also a non-parametric test which is advocated for examining asymmetric distributions (Razali and Wah, 2011). The reason for employing two tests for goodness-of-fit is the fact that each of them has its own limitations. Chi-square test is influenced largely by the number and size of bins (intervals) of frequency distribution while the KS test is affected by the sample size. Hence, it seemed appropriate to use both tests to get concrete results.

In order to estimate the value of travel time variability, with the available dataset, we followed the procedure described in Fosgerau (2017) and Zang et al (2019). This requires the distribution of standard travel time to estimate TTRR using the following equations.

$$TTRR = x \int_{y}^{1} F^{-1}(TT)dp \qquad (1)$$

$$x = \frac{\beta+\gamma}{\alpha} \qquad (2)$$

$$y = \frac{\gamma}{\beta+\gamma} \qquad (3)$$

Wherein F^{-1}(TT) is the inverse of the cumulative distribution function for standardized travel time, α is the value of travel time, while β and γ are the delay parameters. The values for these parameters have been set as $\alpha = 2$, $\beta = 1$, and $\gamma = 4$, in this study which has been the observed in a number of previous studies such as Fosgerau and Karlström (2010), Taylor (2017) and Zang et al. (2018). These values yield x to be 2.5 and y to be 0.80. Based on the obtained value of TTRR and using the already available value of travel time for Athens, Greece, we estimate the value of travel time variability. Results are described in the next section.

3. Results and Discussion

The first step was to classify the travel time data on the basis of time of day and type of link. Table 1 presents the descriptive statistics for each category of travel time data. These statistics show that average travel time is higher on residential highways which could be attributed to the slower design speeds on these links. Average travel times were also found to be relatively higher on weekdays and in the evening peak hours (5:00 to 8:30 pm). This could be attributed to higher traffic demand on these days and times.

Table 1. Descriptive Statistics for Travel Time Categories

Categories		Number of links	Minimum travel time /100m (sec)	Maximum travel time /100 m (sec)	Average travel time /100m (sec)	Standard deviation (sec)
Road Types	Primary	259851	2.53	107.55	22.40	24.41
	Secondary	540074	2.29	157.75	16.69	18.28
	Tertiary	472005	5.12	74.24	16.88	12.64

	Residential	556472	7.24	181.30	26.36	29.96
Time-of-day	12:00am-06:30am	208167	2.52	52.94	15.93	11.96
	06:30am-10:00am	287964	2.53	68.97	15.78	10.16
	10:00am-01:30pm	65923	6.25	34.13	17.30	6.43
	01:30pm-05:00pm	166534	5.73	78.43	18.46	17.90
	05:00pm-08:30pm	735523	2.29	181.30	21.28	23.04
	08:30pm-12:00am	364292	2.29	166.04	19.84	23.21
Days of Week	Weekdays	1320090	2.29	181.30	21.05	21.36
	Weekends	508311	2.53	166.04	18.20	18.35

3.1 Kruskal Wallis Test

Each classification of travel time (mentioned in Table 1) was tested through Kruskal Wallis to check whether some categories could be merged if they have the same distribution. The results are shown in Table 2. Firstly, the Kruskal Wallis test was performed on different types of roads. The test statistic (H) was found to be significantly higher which meant that all types of roads do not have the same travel time distribution. After which different combinations were tried and it was found that primary and secondary roads have the same distribution at a significance level of 5% for (H) while others have different distributions. The test statistics for secondary and primary roads are shown in Table 2 while combinations, in which the test statistic was significantly higher, are not shown for conciseness. Histograms of travel time distribution categories, on the basis of remaining classes of roads, are shown in figure 1. This figure shows some visual evidence of change in distribution between the different types of highways.

Table 2. Kruskal Wallis Test

Category of Travel Time	**H**	**D**	**Adjusted H**	**Df***	**P value**
All types of roads	50.529	0.999	50.529	3	6.16E-11
Primary and secondary	3.356	0.999	3.356	1	0.067
All times of day	12.105	0.999	12.106	5	0.033
Times of day excluding 05:00pm-08:30pm	7.272	0.999	2.272	4	0.122
Weekdays and Weekend	1.898	0.999	1.898	1	0.168

* Df = degrees of freedom

Similarly, the test was also performed to check if all times of the day follow the same travel time distribution. Table 2 shows that all time intervals except 5:00 pm to 08:30 pm follow the same distribution. Figure 2 shows the histograms of travel time distributions in these time periods and the difference in distribution can be observed from it.

Kruskal Wallis test performed for weekdays and weekend show that there is no difference in the distribution of travel time for these two categories. Therefore, on the basis of the above tests, travel time data was segregated into the following six categories:

1) Pri+Sec_Eve_peak ,
2) Pri+Sec_Other_time

3) Tertiary_Eve_peak,
4) Tertiary_Other_time,
5) Residential_Eve_peak, and
6) Residential_Other_time

It should be noted that Eve_peak time in the above categories entails the time interval of evening peak i.e. from 05:00 pm-08:30 pm and Other_time include all time intervals except the evening peak time. Further, the first two categories also combine Primary and Secondary roads. The estimation of travel time distributions is applied for the above six categories.

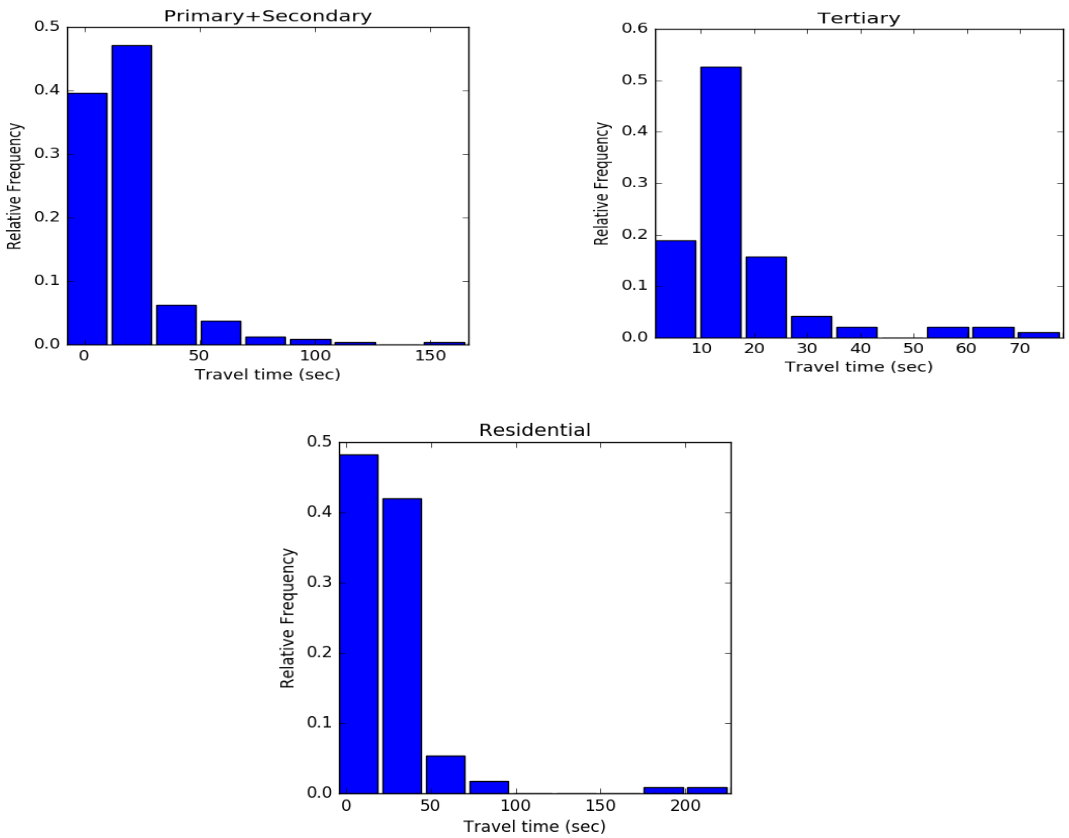

Figure 1. Travel time distribution of different types of roads

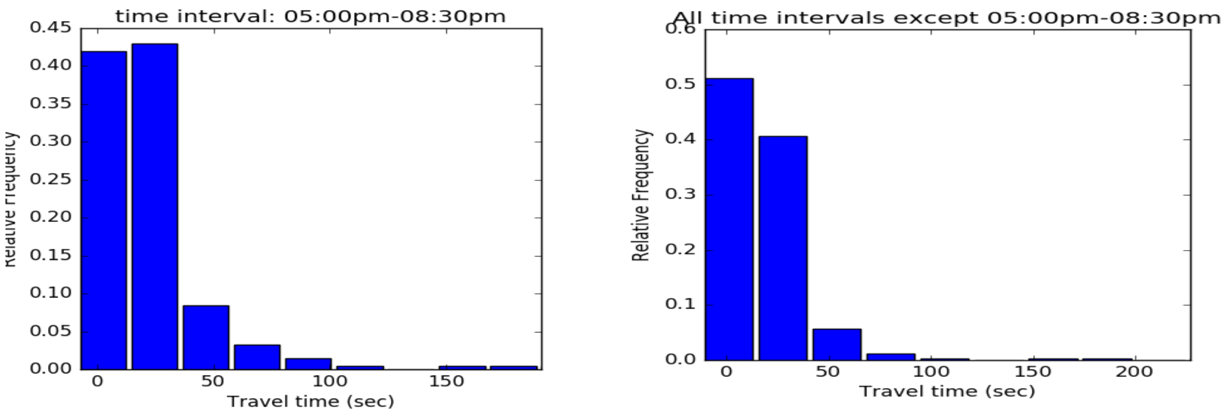

Figure 2. Travel time distributions of different times of day

3.2 Chi-Square Test

All the categories of travel time were tested using chi-square with the distributions mentioned in section 2, the results of the tests can be observed in table 2. Since the Chi-square test is greatly affected by degrees of freedom i.e. number of frequency intervals in the data, hence they were set through trial and error. p-values coming out from this test should be as higher as possible to conclude that travel time data for a category is following a particular statistical distribution. It can be seen from table 2 that normal distribution is significantly different from all categories of travel time. This was expected since the data in figures 1 and 2 shows skewness towards the left side. Exponential distribution was found to fit most of the categories at a significance level of 2.5% (i.e. p-value above 0.025), except for the case of *Residential_Other_time*. Gamma distribution was also found to fit all categories at a significance level of 2.5% except *Pri+Sec_Other_time*. Log-normal distribution was the only distribution which was found to fit all the categories without any exception at the same significance level. Weibull distribution was found to fit the travel time categories in the evening peak hours with all types of road categories. Chi-square and Rayleigh distributions were not found applicable to any categories of travel time.

3.3 KS Test

The results of the KS test for all categories of travel times are also shown in table 2. Similar to the Chi-square test, the p-value should be as higher as possible to suggest whether data is following a particular statistical distribution. Log-normal distribution was found to fit most of the data at 2.5% significance level, except for *Pri+Sec_Other_time*, for which the p-value is 0.01. Gamma distribution was also found to be applicable to tertiary roads in all times of the day. Interestingly, the KS test shows a similarity between tertiary peak travel times and normal distribution at 1% level of significance. However, the data shown in figure 1 shows a clearly visible skewness. Therefore, it could be said that results at 1% significance level or below should not be used to suggest for a particular statistical distribution.

Results from chi-square and KS tests have been combined in table 3. The distributions which were found to be significantly resembling the travel time data at a significance level of 2.5% are marked with 'Yes' otherwise 'No'.

The following observation can be drawn from table 3. Chi-square test is found to provide more conclusive results, in terms of providing goodness-of-fit as compared to KS tests. This is said because it is providing positive results for more distributions with higher chances of similarity, as shown by higher p-values in table 2, than those for KS tests. However, it should be noted that the Chi-square test requires some effort to find the optimum number of frequency intervals for the data. Moreover, KS test is not suggesting any significant resemblance in the data with most of the distributions, this could be that for some categories data may follow multi-modal distributions which are not examined in this study. Moreover, Normal, Rayleigh and Chi-square distribution did not fit the data for any category. Log-normal distribution stands out to be the single most applicable distribution for all cases in chi-square tests. It also shows significant resemblance in KS tests for most cases, for all times of the day. The only exclusion to this is the KS test for Pri+Sec_Other_time. This observation is consistent with another study done for urban roads in Beijing, China (Chen et al., 2018). It must be noted that our study provided a justification for categorization of travel time which is not done in the previous study. Furthermore, the present study covers more ground in the determination of travel time distribution by exploring more distributions and employing two different statistical tests. Other studies as reported in the introduction section also reached to the similar conclusions that travel time usually follows a lognormal distribution. It is evident from the investigation in this study that different categories of travel time follow the same distribution and parameters do not have significantly large differences in values. This confirmed the findings of Fosgerau and Fukuda (2012), where they have analyzed 5 months data of travel time on one single road in Copenhagen, Denmark. The parameters for log-normal distribution for all categories of travel time are given in Table 4.

Table 2. Chi-Square and Kolmogrov-Smirnov Test results for statistical distributions

Categories Considered	**Distribution**	**Chi-square test results**	**Kolmogorov-Smirnov test results**
Pri+Sec_Eve_peak	Normal	Chi = 39.81; df = 5; p = 0.000	D=0.233; p=0.000
	Exponential	Chi = 8.81; df = 5; p = 0.117	D=0.224; p= 0.000
	Gamma	Chi = 8.27; df = 3; p = 0.041	D=0.161; p= 0.009
	Log-normal	Chi = 3.57; df = 3; p = 0.17	D= 0.130 ; p= 0.040
	Chi-square	Chi = 39.94; df = 1; p = 0.000	D= 0.179 ; p= 0.001
	Weibull	Chi = 2.71; df = 1; p = 0.100	D= 0.178 ; p= 0.002
	Rayleigh	Chi = 19.51; df = 1; p = 0.000	D= 0.421 ; p= 0.000
Pri+Sec_Other_time	Normal	Chi = 35.91; df = 1; p = 0.000	D= 0.264 ; p=0.000
	Exponential	Chi = 8.10; df = 3; p = 0.044	D= 0.270 ; p=0.000
	Gamma	Chi = 15.81; df = 2; p = 0.000	D = 0.182 ; p=0.000
	Log-normal	Chi = 7.34; df = 2; p = 0.026	D = 0.128 ; p=0.010
	Chi-square	Chi = 308.76; df = 1; p = 0.000	D= 0.153 ; p=0.000

	Weibull	Chi = 13.93; df = 2; p = 0.001	D = 0.197; p=0.000
	Rayleigh	Chi = 19.51; df = 1; p = 0.000	D= 0.409 ; p=0.000
Tertiary_Eve_peak	Normal	Chi = 7.92; df = 1; p = 0.005	D= 0.251; p=0.010
	Exponential	Chi = 3.46; df = 1; p = 0.063	D= 0.307 ; p=0.001
	Gamma	Chi = 7.84; df = 3; p = 0.049	D = 0.197; p =0.060
	Log-normal	Chi = 2.41; df = 1; p = 0.120	D = 0.149 ; p =0.250
	Chi-square	Chi = 75.96; df = 1; p = 0.000	D = 0.293 ; p =0.001
	Weibull	Chi = 5.03; df = 1; p = 0.025	D = 0.197 ; p =0.060
	Rayleigh	Chi = 19.51; df = 1; p = 0.000	D = 0.358 ; p =0.000
Tertiary_Other_time	Normal	Chi = 11.77; df = 1; p = 0.001	D= 0.248 ; p =0.000
	Exponential	Chi = 4.64; df = 2; p = 0.098	D= 0.274 ; p =0.000
	Gamma	Chi = 12.34; df = 6; p = 0.054	D= 0.178 ; p = 0.040
	Log-normal	Chi = 2.54; df = 1; p = 0.111	D= 0.129 ; p=0.200
	Chi-square	Chi = 154.22; df = 1; p = 0.000	D= 0.238 ; p=0.001
	Weibull	Chi = 5.61; df = 1; p = 0.018	D= 0.179 ; p=0.003
	Rayleigh	Chi = 19.51; df = 1; p = 0.000	D = 0.296 ; p=0.000
Residential_Eve_peak	Normal	Chi = 12.36; df = 2; p = 0.002	D= 0.277 ; p=0.000
	Exponential	Chi = 1.86; df = 1; p = 0.172	D = 0.297 ; p=0.000
	Gamma	Chi = 3.78; df = 2; p = 0.151	D = 0.219 ; p=0.009
	Log-normal	Chi = 3.50; df = 1; p = 0.061	D = 0.171 ; p=0.075
	Chi-square	Chi = 59.84; df = 1; p = 0.000	D = 0.257 ; p=0.001
	Weibull	Chi = 2.04; df = 1; p = 0.154	D = 0.204 ; p=0.020
	Rayleigh	Chi = 12.63; df = 1; p = 0.000	D = 0.411; p=0.000
Residential_Other_time	Normal	Chi = 20.67; df = 2; p = 0.000	D = 0.298 ; p=0.000
	Exponential	Chi = 5.83; df = 1; p = 0.016	D = 0.267 ; p=0.000
	Gamma	Chi = 6.78; df = 2; p = 0.034	D =0.229 ; p=0.000
	Log-normal	Chi = 0.92; df = 1; p = 0.338	D= 0.167 ; p=0.025
	Chi-square	Chi = 78.08; df = 1; p = 0.000	D = 0.223 ; p=0.001
	Weibull	Chi = 6.74; df = 1; p = 0.009	D = 0.212 ; p=0.009
	Rayleigh	Chi = 12.63; df = 1; p = 0.000	D = 0.458 ; p=0.000

Table 3. Summary of Statistical Tests

Categories	**Test**	**Normal**	**Exponential**	**Gamma**	**Log-normal**	**Chi-square**	**Weibull**	**Rayleigh**
Pri+Sec_Eve_Peak	Chi-square	No	Yes	Yes	Yes	No	Yes	No
	KS	No	No	No	Yes	No	No	No
Pri+Sec_Other_time Off	Chi-square	No	Yes	No	Yes	No	No	No
	KS	No	No	No	No	No	No	No
Tertiary_Eve_peak	Chi-square	No	Yes	Yes	Yes	No	Yes	No
	KS	No	No	Yes	Yes	No	Yes	No
Tertiary_Other_time	Chi-square	No	Yes	No	Yes	No	No	No
	KS	No	No	Yes	Yes	No	No	No
Residential_Eve_peak	Chi-square	No	Yes	Yes	Yes	No	Yes	No
	KS	No	No	No	Yes	No	No	No

Residential_Other_time	Chi-square	No	No	Yes	Yes	No	No	No
	KS	No	No	No	Yes	No	No	No

Table 4. Log-normal Distribution Parameters (location parameter = 0)

Categories	**Scale**	**Shape**
Pri+Sec_Eve_peak	2.69	0.76
Pri+Sec_Other_time	2.47	0.68
Tertiary_Eve_peak	2.68	0.59
Tertiary_Other_time	2.60	0.56
Residential_Eve_peak	3.02	0.59
Residential_Other_time	3.02	0.56

Other distributions which were found to be significant in few cases include Exponential, Gamma and Weibull. Travel times on tertiary roads were found to fit more distributions in chi-square as well as KS tests. On the other hand, travel times on primary and secondary roads in the other time only showed significant results for very few distributions and that was also at a lower confidence interval, this provides an indication that travel times in this category may follow the multi-modal distribution.

3.4 Estimation of TTRR and value of travel time variability

The analysis from the previous section established the fact that log-normal distribution fits the travel time distribution in all categories and the close examination of values mentioned in table 5 indicates that for all categories values are more or less similar. In order to estimate TTRR, it is required to fulfil a few conditions such as 1) The travel time distribution needs to be formed based on standardized travel time (i.e. (travel time – mean travel time) / standard deviation of travel time), 2) the distribution needs to be independent of time of day as required by theory, i.e. the formulation shown in equation 1 is only applicable when this condition is roughly satisfied (Fosgerau and Fukuda 2012). Findings of our investigation suggest that all considered categories of travel time that also include different time-of-day followed lognormal distribution and for all categories, the parameters are roughly similar. So it is believed that standardized travel time will also follow a similar trend. Therefore, travel time data of all categories are pooled together and converted standardized travel time. The same 7 statistical distributions are then fitted to this data. Based on our examination, lognormal distribution again provides appropriate results of KS and chi-square tests. The parameters of the distribution are as follows*: Scale parameter = 0.68, Shape parameter = 0.70 and Location parameter = -0.89*. This distribution will be used for further calculations of TTRR as per equation (1) that yields value of *TTRR as 0.495*. We used a numerical approximation to obtain the integral of inverse cumulative distribution. Using a previous study (Wardman et al 2016) that provided a meta-analysis of values of travel time in Europe, the value of travel time (VTT) was taken for different groups of travelers for Greece. We then obtained VTTV by simply dividing the estimated TTRR with VTT. The values of VTTV are shown in Table 5.

Table 5. VTT and estimated VTTV for Athens, Greece in €/hour
(based on 2010 incomes and prices)

	Car Commute		Car Other		Car Business Travel		Public Transport Commute
	Urban Free Flow	Urban Congested	Urban Free Flow	Urban Congested	Urban Free Flow	Urban Congested	Urban
VTT (Wardman et al, 2016)	5.60	7.96	4.93	7.01	14.72	20.93	5.68
VTTV	2.77	3.94	2.44	3.47	7.29	10.36	2.81

It can be observed from table 5 that travelers value travel time differently in free flow and congested situation and for business travel, these values are quite high compared to usual commuting and travelling for other purposes. Based on our investigation, VTTV is approximately half of the VTT, this also confirms results of other studies where TTRR values are estimated to be around this range such as 0.5 to 0.8 (Taylor 2017, Fosgerau 2016, Zang et al 2018). Table 5 is useful to carry out a cost-benefit analysis of mobility projects. An example illustration of this as follows: Suppose that introduction of project A (a new road/bridges etc) in the urban area of Athens improve mean travel time of commuters (i.e. from 60 minutes to 45 minutes) and also improve travel time variability (i.e. standard deviation of travel time is improved from 15 minutes to 10 minutes). Therefore, by simply following the procedure indicated in OECD (2016) and values mentioned in table 5, travel time savings for a single car-based business traveler during free flow would be around (14.72*(60-45)/60 = 3.7€) and variability savings would be around (7.29*(15-10)/60 = 0.6 €) after introduction of project A.

4. Conclusions

This study was aimed at investigating frequency distribution functions of travel times on urban highways. GPS trace data for vehicles traveling in Greece, Athens was used for estimating travel times. Travel times were categorized on the basis, time and day of the trip as well as types of urban roads under the hypothesis that travel time distribution would be different in different space and time settings. Kruskal Wallis test was used to justify the categories for which travel time distribution are estimated. Seven different types of uni-modal statistical distributions were investigated to fit the travel time data of identified categories using non-parametric chi-square and KS tests. It was found that lognormal distribution fitted well with all categories of data, however, there are only slight variations in distribution parameters. This finding provided the evidence to conclude that travel time distribution of links is roughly similar in space and time, that contradicts the findings of the recent study in Beijing, however, in agreement with other previous studies. This conclusion helps in applying the method to estimate TTRR, that in theory requires time-of-day independent travel time distribution. Based on the application of this method, TTRR was estimated and using the available values of travel time, values of travel time variability are estimated for various classes of individuals. These values are helpful to carry out a cost-benefit analysis of transport-related infrastructure projects specifically in Athens. Further research could be carried out to investigate multi-modal distributions in fitting travel time data to further confirm the independency of travel time distribution with time and space. Furthermore, the travel time prediction problem can be researched given the availability of rich data of vehicle trajectories.

Acknowledgement

This work is partially funded by the EU H2020 program under Grant Agreement No. 780754, "Track & Know". Authors also acknowledge the Special Research Fund (BOF) of Hasselt University, to support the research visit of Dr. Uneb Gazder at Hasselt University, who significantly contributed in this study. Authors also wanted to pay special thanks to Vodafone Innovus, S.A. for providing GPS dataset of vehicles for this study.

References

Adnan, M., Watling D.P., & Fowkes, A.S., (2009). Model for integrating home-work tour scheduling with time-varying network congestion and marginal utility profiles for home and work activities, *Transportation Research Record-Journal of Transportation Research Board*, No. 2134, pp:21-30

Arezoumandi, M., (2011). Estimation of travel time reliability for freeways using mean and standard deviation of travel time, *Journal of Transportation Systems Engineering and Information. Technology*, 11 (6), pp: 74–84.

Batterman, S., Cook, R., & Justin, T. (2015). Temporal variation of traffic on highways and the development of accurate temporal allocation factors for air pollution analyses. *Atmospheric environment*, 107, pp: 351-363.

Baqueri, S. F. A., Adnan, M., Kochan, B. & Bellemans, T. 2019. Activity-based model for medium-sized cities considering external activity–travel: Enhancing FEATHERS framework. *Future Generation Computer Systems*, 96, 51-63.

Beaud, M., Blayac, T., & Stéphan, M. (2016). The impact of travel time variability and travelers' risk attitudes on the values of time and reliability. *Transportation Research Part B: Methodological*, 93, pp: 207-224.

Chen, A., & Zhou, Z. (2010). The α-reliable mean-excess traffic equilibrium model with stochastic travel times. Transportation Research Part B: Methodological, 44(4), 493-513.

Chen, M., Yu, G., Chen, P., & Wang, Y. (2017). A copula-based approach for estimating the travel time reliability of urban arterial. *Transportation Research Part C: Emerging Technologies*, *82*, 1-23.

Chen, P., Tong, R., Lu, G., & Wang, Y., (2018). Exploring travel time distribution and variability patterns using probe vehicle data: Case study in Biejing, *Journal of Advanced Transportation*, https://doi.org/10.1155/2018/3747632

Chen, P., Yin, K., & Sun, J., (2014) Application of finite mixture of regression model with varying mixing probabilities to estimation of urban arterial travel times, *Transportation Research Record*, No. 2442, pp. 96–105

Eliasson J., (2019) Modeling reliability benefits, *Transport Findings*, doi:10.32866/7542

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

Fosgerau, M. (2017), "The valuation of travel-time variability", in Quantifying the Socio-economic Benefits of Transport, OECD Publishing, Paris, https://doi.org/10.1787/9789282108093-3-en.

Fosgerau, M., & Fukuda, D. (2012). Valuing travel time variability: Characteristics of the travel time distribution on an urban road. *Transportation Research Part C: Emerging Technologies*, 24, 83-101.

Fosgerau, M., & Karlström, A. (2010). The value of reliability. *Transportation Research Part B: Methodological*, *44*(1), 38-49.

Guo, F., Rakha, H., Park, S., (2010). Multistate model for travel time reliability. *Transportation Research Record, Journal of Transportation Research Board*, No. 2188, pp: 46–54.

Jenelius, E., (2012). The value of travel time variability with trip chains, flexible scheduling and correlated travel times. *Transportation Research. Part B: Methodological*. 46, pp:762–780.

Kaparias, I., Bell, M.G.H., Belzner, H., (2008). A new measure of travel time reliability for in-vehicle navigation systems, *Journal of Intelligent Transportation System*, 12 (4), pp: 202–211.

Kieu, L.-M., Bhaskar, A., Chung, E., (2015). Public transport travel-time variability definitions and monitoring, *Journal of Transportation Engineering*, 141, 1, https://doi.org/10.1061/(ASCE)TE.1943-5436.0000724

Kuhn, B., Higgins, L., Nelson, A., Finley, M., Ullman, G., Chrysler, S., Wunderlich, K., Shah, V., Dudek, C., (2013). Effectiveness of Different Approaches to Disseminating Traveler Information on Travel Time Reliability. No. SHRP 2 Reliability Project L14.

Lei, F., Wang, Y., Lu, G., & Sun, J., (2014). A travel time reliability model of urban expressways with varying levels of service, *Transportation Research Part C: Emerging Technologies*, 48, pp. 453–467.

Malleson. N., Vanky, A., Hashemian, B., Santi, P., Verma, S. K., Courtney, T. K., & Ratti, C., (2018). The characteristics of asymmetric pedestrian behavior: A preliminary study using passive smartphone location data, *Transaction in GIS*, 22, 2, pp: 616-634

Ma, Z., Koutsopoulos, H. N., Ferreira, L. & Mesbah, M. (2017). Estimation of trip travel time distribution using a generalized Markov chain approach. *Transportation Research Part C: Emerging Technologies*, *74*, pp: 1-21.

Marra, A. D., Becker, H., Axhausen, K. W., & Corman, F. (2019). Developing a passive GPS tracking system to study long-term travel behavior. *Transportation Research Part C: Emerging Technologies*, *104*, 348-368.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia Medica*, *23*(2), 143-149.

McKight, P. E., & Najab, J. (2010). Kruskal‐wallis test. *The corsini encyclopedia of psychology*, 1-1.

Newson, P., & Krumm, J. (2009). Hidden Markov map matching through noise and sparseness. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 336-343). ACM.

Nie, Y., Wu, X., Dillenburg, J. F., & Nelson, P.C., (2012). Reliable route guidance: A case study from Chicago, *Transportation Research Part A: Policy and Practice*, 46 (2), pp: 403–419

NZTA, (2016). Economic Evaluation Manual. New Zealand Transport Agency, Wellington, New Zealand. www.nzta.govt.nz.

OECD, (2016) Quantifying the socio-economic benefits of transport roundtable. International Transport Forum. Organization for Economic Cooperation and Development, Paris. www.itf-oecd.org/quantifying-socio-economic-benefits-transport-roundtable

Pel, A. J., Bliemer, M. C., & Hoogendoorn, S. P. (2012). A review on travel behaviour modelling in dynamic traffic simulation models for evacuations. *Transportation*, *39*(1), 97-123.

Peer, S., Koopmans, C. C., & Verhoef, E. T. (2012). Prediction of travel time variability for cost-benefit analysis. *Transportation Research Part A: Policy and Practice*, *46*(1), 79-90.

Petrik, O., Adnan, M., Basak, K. & Ben-Akiva, M. (2018). Uncertainty analysis of an activity-based microsimulation model for Singapore. *Future Generation Computer Systems.* https://doi.org/10.1016/j.future.2018.04.078

Polus, A., (1979). A study of travel time and reliability on arterial routes. *Transportation*, 8 (2), pp: 141–151.

Rakha, H., El-Shawarby, I., Arafeh, M., (2010), Trip travel-time reliability: issues and proposed solutions, *Journal of Intelligent Transportation System*, 14, pp: 232–250.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, *2*(1), 21-33.

Rui, T., Adnan, M., Lee, D.H., Ben-Akiva, M., (2015). New path size formulation in path size logit for route choice modelling in public transport networks, *Transportation Research Record, Journal of Transportation Research Board*, No. 2538, pp: 11–18.

Stathopoulos, A., & Karlaftis, M. (2001). Temporal and spatial variations of real-time traffic data in urban areas. *Transportation Research Record*, *1768*(1), 135-140.

Sharpe, D. (2015). Your chi-square test is statistically significant: now what?. *Practical Assessment, Research & Evaluation*, *20*.

Susilawati, S., Taylor, M.A.P., Somenahalli, S.V.C., (2013). Distributions of travel time variability on urban roads, *Journal of Advanced Transportation*. 47 (8), 72

Taylor, M. A. (2017). Fosgerau's travel time reliability ratio and the Burr distribution. *Transportation Research Part B: Methodological*, *97*, 50-63.

Wardman, M., Chintakayala, V. P. K., and De Jong, G. C. (2016) Values of travel time in Europe: Review and meta-analysis. *Transportation Research Part A: Policy and Practice*, 94. pp.93-111.

Yang, S., Wu, Y., (2016). Mixture models for fitting freeway travel time distributions and measuring travel time reliability, *Transportation Research Record, Journal of Transportation Research Board*, No. 2594, pp: 95–106.

Zang, Z., Xu, X., Yang, C., & Chen, A. (2018). A distribution-fitting-free approach to calculating travel time reliability ratio. *Transportation Research Part C: Emerging Technologies*, *89*, 83-95.