Travel Time Variability Analysis in Stockholm Vehicular Traffic Network

Zhiwu Dong, Matej Cebecauer and Erik Jenelius

November 5, 2019

Travel time variability analysis in Stockholm vehicular traffic network

Zhiwu Dong[1], Matej Cebecauer[1], Erik Jenelius[1]*

[1]Division of Transport Planning, KTH Royal Institute of Technology, Sweden
[*]Corresponding author: zhiwu@kth.se

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

1. Introduction

Characterizing travel time distribution is often the prerequisite when analysis the model of travel time reliability. Especially the mean deviation and the standard deviation are two key statistics commonly used. Even though the mean of travel time data is easy to obtain, other statistics such as central tendency and tendency of dispersion are hard to obtain due to insufficient travel data. Mahmassani et al. (2013) analysed the characteristics of travel time reliability and showed that the linear relationship exists between mean travel time per kilometre and corresponding standard deviation at different network scales and various levels of analysis (network-wide level, Origin-Destination (OD) level, path level and link level). Van Lint et al. (2018) highlight that the distribution of travel time is often skewed and wide, especially during the onset, peak and dissolve period of congestion. However, conventional methods based on the mean or variance of travel time may result in a biased estimate of reliability when the distribution is skewed. They suggested that skewness also should be one of character. Bogers et al. (2004) highlight the importance and significance of distribution skewness in estimating travel times by regression models and illustrate its role in route choice for drivers. Especially during congested periods is the distribution of travel time often wide and skewed and can influence the reliability of travel time estimates. Additionally, Zheng et al. (2018) confirm that travel time (un)reliability can be characterized by standard deviation and skewness, by using simulation data and ANPR data from Changsha, Huna. However, in Zheng et al. (2018) research, they only use data collected from 3 routes in three days, so they only did analysis on route level and set time interval as 15min. We aim to validate this model on a multiscale (network and route) and multilevel (day-to-day and within-day).

Furthermore, based on the literature introduced above, the skewness attribute in regression models is investigated. In contrast to previous studies, we use Bluetooth data which represents a rich source of individual car travel time measurements on about 19 sensor pairs. This allows us to study network-wide, as well as route level in our analysis. We aim to propose and validate a simple and robust method to estimate the standard deviation of travel time per kilometer, which can be an interesting contribution to the travel time estimation reliability. Importantly thanks to rich per-user data, we discuss the effects of Spatio-temporal aggregation when computing mean and standard deviation for linear regression models. We want to further explore what characteristics of the road lead to different regression models coefficient for each road when estimating standard deviation. These results will be presented at the conference.

2. Data and case study

We use data collected by Bluetooth sensors in Stockholm, Sweden. The sensors mainly cover highways and main arterials around the inner city (see Figure 1 with 19 routes covered by pairs of Bluetooth sensors). Bluetooth sensors are a new technology to measure travel times by detecting Bluetooth or Wi-Fi devices passing at least two of them. The MAC address of these devices will be recorded by the Bluetooth sensor on the roadside. Thus, the raw data set includes the hashed MAC address with timestamps when devices crossed the pair of sensors. This enables to calculate the speed and travel time for each device passing the pair of sensors. We will refer to the pair of sensors as the route from this point. The mean travel times and standard deviations can be extracted on different levels of Spatio-temporal aggregations. The raw observations allow us to calculate the mean and standard deviation for any period of time. Also, the routes can be clustered based on different attributes such as directions, peak performance, spatial coordinates and etc. This study uses data from the whole of February 2019 and it represents 3,114,460 collected observations for computational experiments.

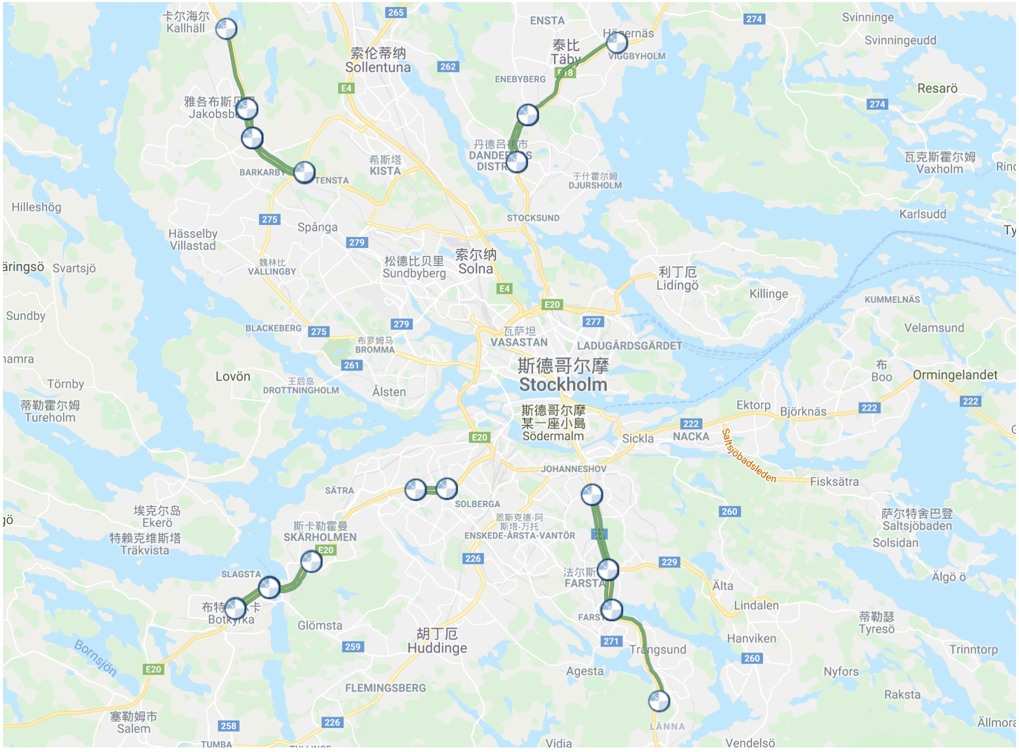

Figure 1: Illustration of the Bluetooth sensors and routes locations.

3. Methodology

Travel time is distance-dependent, and the travel time per kilometer is used in further analysis to guarantee the consistency of collected data across multiple routes with different lengths. At first, we calculate the mean travel time per kilometer *T* and the corresponding standard deviation *TTSD*. The mean travel time *T* is used as an explanatory or independent variable in a linear regression model for *TTSD* according to the Mahmassani et al. (2013):

$$TTSD = a * T + b \tag{1}$$

van Lint et al. (2018) extended Mahmassani et al. (2013) model by skewness *S*. The Formulation of a linear model with skewness is following:

$$TTSD = m * T + n * S + k \tag{2}$$

We investigate both models on different levels of aggregation. We consider network-wide and route levels in combination with 1 day and 5 minutes time intervals. The level of aggregation is affecting the number of data points but it is also expected that large aggregation such for network-wide will be smoothing the variance and standard deviation of travel times.

4. Results

We consider in this paper network-day, route-day, network-5minute and route-5minute time intervals aggregations. For each of such subsets, the mean and standard deviation of travel time is calculated and used in computational experiments to estimate regression models introduced above. Figure 2 illustrates the mean and standard deviation variation of travel time per kilometer. Tables 1 and 2 presents the regression models for network and route levels when considering 5 minutes time intervals.

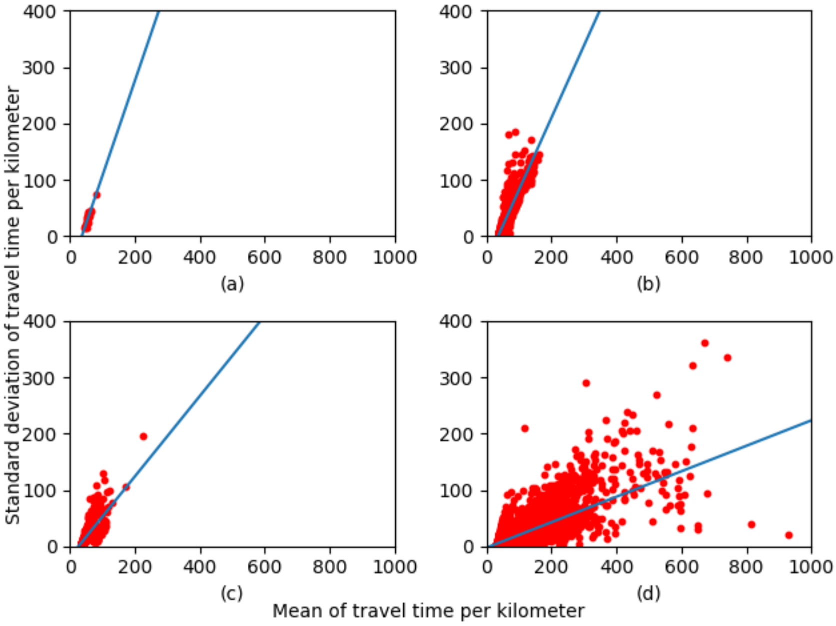

Figure 2: Mean and standard deviation data points and corresponding linear regression models for travel time standard deviation based on mean (a) network-day (b) network-5min (c) route-day and (d) route_5min travel time

Table 1: Linear regression model for network-5min travel time standard deviation based on mean network-5min travel time

	Estimate	Confidence Interval	T statistic	P - value	Coefficient of Determination
a	-46.366	[-47.211, -45.520]	-107.523	<0.001	0.832544
b	1.282	[1.266, 1.297]	161.298	<0.001	
m	-51.234	[-52.016, -50.451]	-128.352	<0.001	0.870484
n	1.265	[1.251, 1.279]	180.575	<0.001	
k	2.875	[2.731, 3.019]	39.134	<0.001	

Table 2: Linear regression model for route-5min travel time standard deviation based on mean route-5min travel time

	Estimate	Confidence Interval	T statistic	P - value	Coefficient of Determination
a	-2.567	[-2.674, -2.461]	-47.314	<0.001	0.493502
b	0.226	[0.225, 0.228]	265.849	<0.001	
m	-5.721	[-5.832, -5.610]	-101.037	<0.001	0.573982
n	0.245	[0.243, .246]	307.037	<0.001	
k	2.102	[2.067, 2.138]	116.978	<0.001	

According to the preliminary results, we found that there exists an improvement of the coefficient of determination (R^2) when skewness is used as an explanatory variable in the model. And the P-value of every coefficient is greater than 0.001, so we can believe that the regressions are statistically significant. Comparing the two corresponding models under the same constitution, the coefficient of the constant term (a, m) and mean travel time per km (b, n) does not change much, but the coefficient of skewness (k) is much larger than other coefficients, which also indicates the importance of skewness in the model. At the same time, the mean coefficient of the regression equation at the route level is much smaller than that under other conditions, especially when the time interval is 5 min. The regression performance for route-5min is much worse than the other three with a significantly higher variance of values. The analysis of regression models for routes individually revealed to us that the network-wide regression model, when data from all routes and 5 minutes time intervals are considered together to build regression model as in Figure 2 (d), is not well−fitting the individual routes and time intervals. The average coefficient of determination across the linear regression models of every single route (there are 19 routes in the case study) and 5 minutes time aggregation is 0.675.

The regression models of network-day or network-5min can represent a robust model for estimating average network standard deviation but are not suitable for estimating the standard deviation for the particular routes.

6. Conclusions

This paper further validates the assumptions introduced by Mahmassani et al. (2013) and such that the linear model is a simple yet robust model that can be commonly used for different scales of networks (Network level and route level). We examined the importance and effect of skewness by using it as a variable in the linear regression model of standard deviation. The coefficient of determination has been improved across different aggregation levels, which reveals the important role of skewness in predicting standard deviation. Results also revealed that time-of-day in 5-minute time intervals aggregation is affecting the performance and standard deviation variance as in case of whole day aggregation peak and off-peak is smoothed out.

For possible future work to be presented at the conference, we would like to analyse why the regression performance falls deeply on the route-5min level and investigate possible effects of additional route attributes such as time-of-day or number of observations as estimates of the flow. The independent dataset will be used for the evaluation of regression models and how good they can predict the standard deviation based on the mean travel times.

References

Mahmassani H S, Hou T, Dong J. (2012). Characterizing travel time variability in vehicular traffic networks: deriving a robust relation for reliability analysis. *Transportation Research Record*, 2315(1): 141-152.

Van Lint J W C, Van Zuylen H J, Tu H. (2008). Travel time unreliability on freeways: Why measures based on variance tell only half the story. *Transportation Research Part A: Policy and Practice*, 42(1), 258-277.

Bogers E A I, Van Zuylen H J. (2004). The importance of reliability in route choices in freight transport for various actors on various levels. *Proceedings of the European Transport Conference*. 149-161.

Zheng F, Li J, Van Zuylen H, et al. (2018) Urban travel time reliability at different traffic conditions. *Journal of Intelligent Transportation Systems,* 22(2), 106-120.