Spring 2017

CEE 6603: TRAFFIC ENGINEERING

Term Project:

Goodness of fit of sparse GPS data into Instantaneous Speed dataset

04.14.2017

Han Gyol Kim and Somdut Roy

School of Civil and Environmental Engineering

College of Engineering

Georgia Institute of Technology

INTRODUCTION

With rapid development of technologies in transportation, data collection, and big data analysis, urban road traffic network systems have the potential to produce more efficient, real-time, intelligent, and safe transportation infrastructure. Wireless sensor network technologies, mobile phones, and crowd-sourced traffic data are being used to increase the quality of transportation network services.

Global Positioning Systems (GPS) devices provide a real-time spatial and temporal measurement of a location which most transportation services require. The traffic data attained from GPS is valuable to predict traffic congestion in the near future. Companies, such as Waze, are attempting to crowd-source traffic data, so the real-time wireless automotive communications are expected in the near future.

At present times dynamic real-time traffic data are collected by fixed infrastructure sensors or mobile sensors. The real traffic data are crucial for understanding and analyzing transportation systems, and furthermore better analyzing transportation network needs for predicting traffic pattern more accurately. But there is no tool to handle this large amount of data.

To create efficient and intelligent routing it is essential that historic trajectory data of a large number of vehicles are at disposal. Due to lack of real data, synthetic data from VISSIM 9.0 run had to be used to serve the purpose. In the following sections, we have a brief literature review, description of the methodology of our project and discussion of some of our results before concluding.

BACKGROUND RESEARCH

In the past, most conventional methods to measure speed were based on the availability of equipment in the measured road sections. Because of the relatively high cost of the equipment and its installation, speed measurements were typically conducted at a low frequency in road locations that are selected for different purposes and cover only a small part of the road network. In recent years, however, with the development of cellular technologies and vast penetration of global positioning system (GPS) devices, several attempts were made to develop sampling methods on the basis of the speed assessment of vehicles equipped with cellular phones or GPS to characterize speeds in the road network.

The advantage of these methods stems from the availability of the data without a need to install or deploy equipment of any sort. Therefore, having defined the methods required for data processing, it is possible to receive assessments of the speed distribution on every road section in the road network as well as to evaluate trends of the changes in speed throughout time or because of various intervention activities [1].

Before going into details, two concepts –average and instantaneous speed need to be revisited. Average speed (S) can be defined as the total distance covered (d) over the time taken (t) or just simply S = d/t. Instantaneous speed formula is used to find the rate of change of displacement for any instant of time. It is just a first derivative with respect to time and is given as Instantaneous speed(S) is equal to displacement (dx) over a very small time (dt) or just simply S = dx/dt. Instantaneous speed formula is used to find the rate of change of displacement for any given instant of time. Itis expressed in meter per second. [2]

By using two GPS points (locations) we can calculate the distance covered. We can use the clock inside the GPS receiver (a very accurate clock that synchronizes regularly with the atomic clocks aboard the GPS satellites) to measure how long it took the vehicle to travel between those two points. In the example below the GPS receiver within the truck records its location (latitude and longitude, or lat/lon) at Point A. It takes note of the time as well. A short while later, say two minutes, it records its location again Point B. The GPS receiver can then perform a calculation using these numbers and determine the speed of the vehicle. [3]

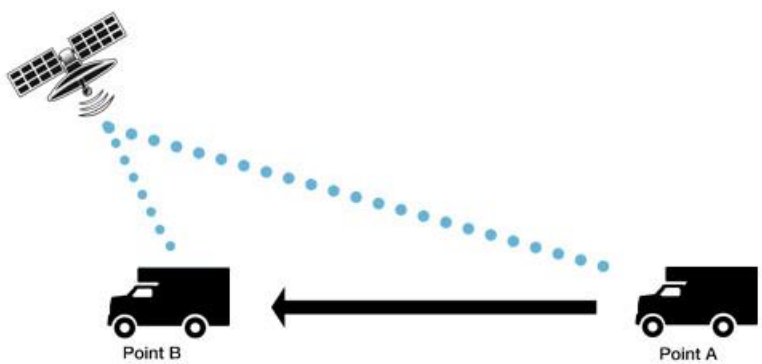

Figure 1: GPS Speed Recording (Telogis, 2014)

In the following study we use similar concepts in analysis of speed data and comparison with sparse GPS Data.

METHODOLOGY

VISSIM 9.0 is a microscopic traffic flow simulation based software which is capable of generating second by second data (or even better) of trajectory for vehicles. Our aim is to clip out minute by minute location data from the big chunk of data, calculate the speed based on that and compare it with the instantaneous data generated by VISSIM. To analyze different traffic scenarios (queuing and signal timing), two different cases had to be studied –arterial and freeway.

A large section of Metro Atlanta was part the VISSIM file that was run for 3 simulation hours.

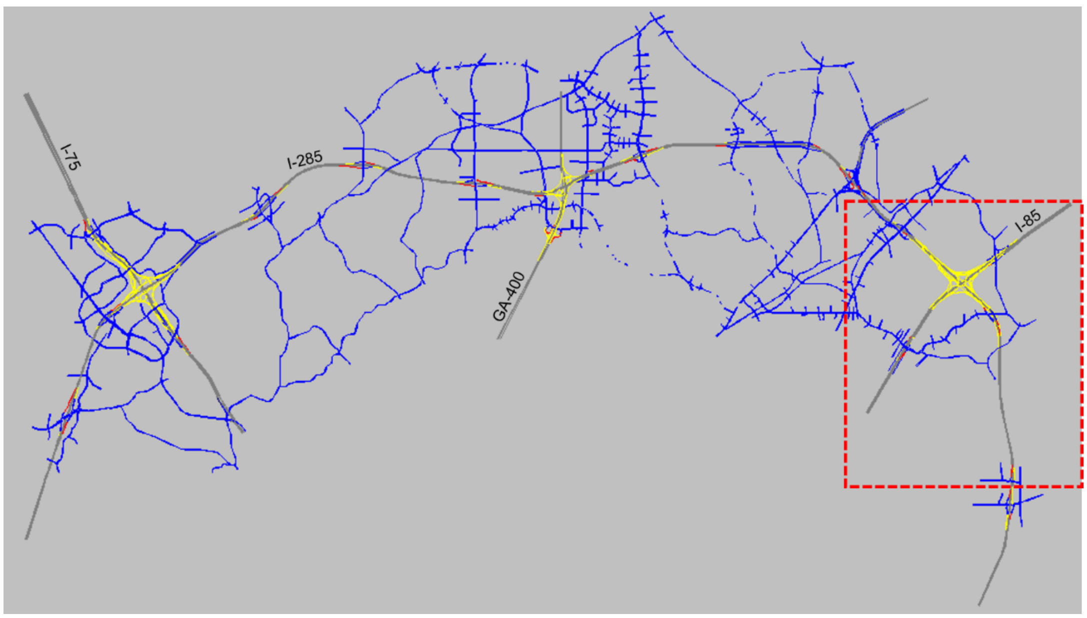

Figure 2: VISSIM file screenshot and the focused area of study (covered in red dash lines)

The VISSIM output file which has a .fzp extension is a readable text file which had to be processed with python to get a section that would cover one freeway and one arterial section. Using the VISSIM link and connector IDs, we got selective data of the sections as depicted in the figure below.

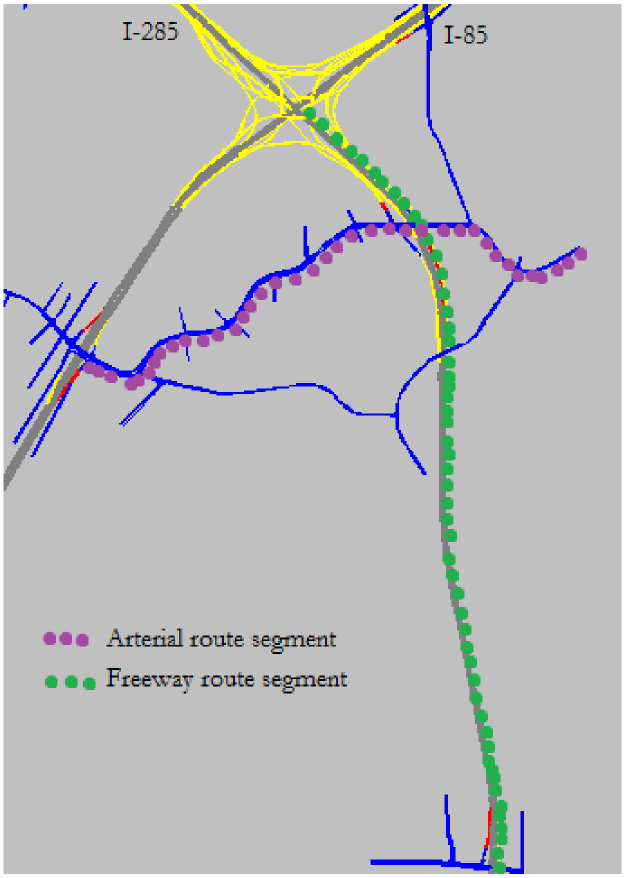

Figure 3: Freeway and Arterial study focus area

We clip out min-by-min vehicle location data as well as sec-by-sec data of the two road sections as indicated above and then process then in MS Excel to get the min interval speed sets and the sec interval speed sets. It is mention-worthy here that VISSIM does provide us with instantaneous speed data which in turn can be compared with the min-interval speed data sample as well as the second-interval data sample to get an idea about the goodness of fit of GPS derived speed data into actual instantaneous speed distribution.

In order to compare the sets of data, it is advisable to use *t-test*. In reference to this, it must be mentioned that a t-test is any statistical hypothesis test in which the test statistic follows a Student's *t*-distribution under the null hypothesis. It can be used to determine if two sets of data are significantly different from each other. In this case the null hypothesis is that the means of the speed data are same and we test that claim at a 95% confidence.

Since the sec-interval data is comparable to the second by second speed data, we compare the two using correlation coefficient. For the minute-interval data, however, we have to use t-test as the two sets of data in question (viz. Sec-by-sec instantaneous data and the min-interval data) have different resolution of data and do not show similar values for a given car at a given point in its trajectory. All the statistical analyses were done in MS Excel.

RESULTS

After processing in python, separate CSV files were generated for the arterial and freeway sections in discussion in two different frequencies (/min and /sec). Assuming that there is a straight line trajectory between two consecutive data-points for a given vehicle, average speed is calculated at that instance. The plots below show us a comparative study of min-interval data with the instantaneous speed data in both arterial and freeway. As we can see, it is all random and it does not suggest any proper trend.

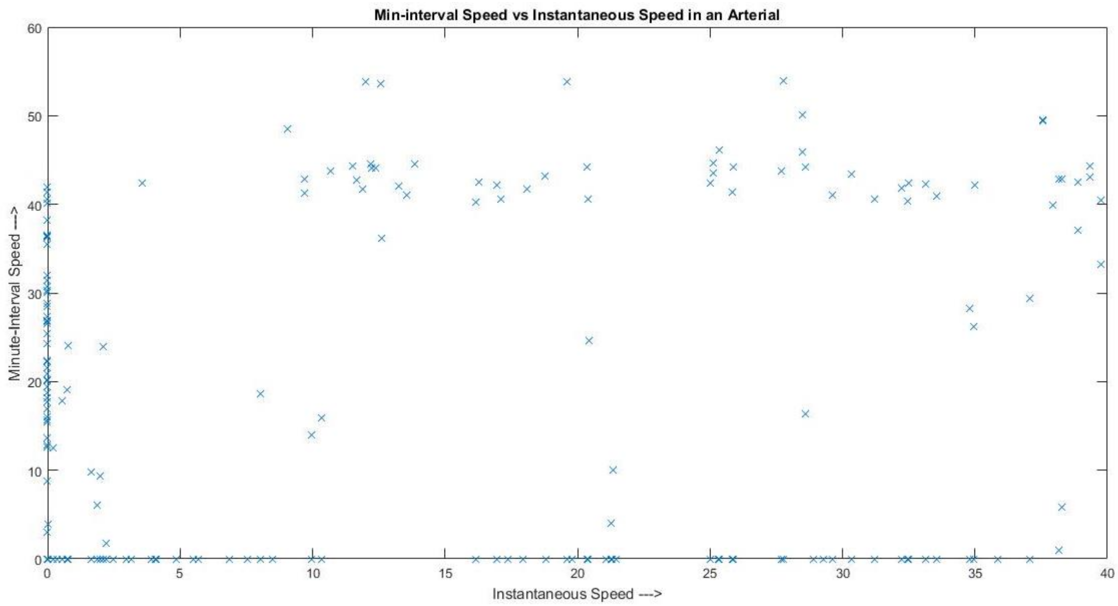

Figure 4: A scatterplot of min-based data with instantaneous speed data in arterial

The summary of Pearson correlation and t-test results of all scenarios are explained in the table below. (Detailed tables are in appendix section.)

Table 1: T-statistics analysis and Pearson correlation coefficient (PCC) results of min data

	T-stat	**Critical t-value**	**Accept null hypothesis**	**PCC**
Arterial	3.43	1.96	No	0.603
Freeway	-6.57	1.96	No	0.895

Table 2: Pearson correlation coefficient for second-interval data

	PCC
Arterial	0.990
Freeway	0.999

DISCUSSION AND CONCLUSION

It was noticed that the t-test for comparison of speed distribution fell apart in all cases probably due to the fact that the data had way too many zero-speed data to make a proper comparison. However if the Pearson correlation coefficient gave us some idea about the adherence of the interval speeds to the instantaneous speed dataset for both arterial and freeway. It was noticed that minute by minute data were poorly correlated with instantaneous speed data. The correlation coefficient, however, improved with second-interval speed data (understandably so). For freeway, however, we have more encouraging trends of adherence of min-based speed data to the original instantaneous speed data having a correlation coefficient of 0.895 (which is comparatively better). For the second by second data, the correlation coefficients were almost touching unity. This means second by second GPS data is quite accurate for meeting practical purposes.

Before concluding, we must mention some of the difficulties and limitations of this study. Firstly, we did not have real time data to analyze. Real-life data will have problems related to accuracy of records which we did not have to solve for synthetic VISSIM data. Due to lack of time we could not device an efficient way to extract the focused data and process them from the large VISSIM output file. Also, further attempts could have been made to automate the whole process and extend this case study to more freeways and arterials under different signal timings and volume conditions to study the variations. To conclude it must, however, be mentioned this study though limited could be extended to study GPS data in a larger scale.

REFERENCES

1. Bekhor et al. (2103), Free-Flow Travel Speed Analysis and Monitoring at the National Level Using Global Positioning System Measurements, ASCE Journal of Transportation Engineering/Volume 139 Issue 12 - December 2013
2. Angkiko, F. EXPERIMENT #5: INSTANTANEOUS SPEED & AVERAGE SPEED. http://www.academia.edu/8315752/EXPERIMENT_5_INSTANTANEOUS_SPEED_and_AVERAGE_SPEED. Accessed Apr. 4, 2017.
3. Ago, J. D. C. A. E. A. C. A. How does GPS tracking report on speeding? https://www.telogis.com/blog/gps-tracking-report-speeding. Accessed Apr. 12, 2017.

APPENDIX

Table 1: Arterial Min Speed Data

t-Test: Paired Two Sample for Means		
	Instantaneous speed	Min based speed
Mean	0.428656	0.315425
Variance	15.85287	8.363011
Observations	9477	9477
Pearson Correlation	0.603693	
Hypothesized Mean Difference	0	
df	9476	
t Stat	3.432358	
P(T<=t) one-tail	0.0003	
t Critical one-tail	1.645014	
P(T<=t) two-tail	0.000601	
t Critical two-tail	1.960214	

Table 2: Freeway Min Speed Data

t-Test: Paired Two Sample for Means		
	Instantaneous speed	Min based speed
Mean	2.795517	2.911684
Variance	175.0421	168.7001
Observations	115545	115545
Pearson Correlation	0.894999	
Hypothesized Mean Difference	0	
df	115544	
t Stat	-6.56797	
P(T<=t) one-tail	2.56E-11	
t Critical one-tail	1.644867	
P(T<=t) two-tail	5.12E-11	
t Critical two-tail	1.959985	

Table 3: Arterial Sec Speed Data

t-Test: Paired Two Sample for Means		
	Instantaneous speed	Sec based speed
Mean	0.450744	1.135814
Variance	16.15185	104.9981
Observations	567469	567469
Pearson Correlation	0.989935	
Hypothesized Mean Difference	0	
df	567468	
t Stat	-81.9918	
P(T<=t) one-tail	0	
t Critical one-tail	1.644856	
P(T<=t) two-tail	0	
t Critical two-tail	1.959968	

Table 4: Freeway Sec Speed Data

t-Test: Paired Two Sample for Means		
	Instantaneous speed	Sec based speed
Mean	24.03609	24.05809
Variance	1032.851	1033.062
Observations	699999	699999
Pearson Correlation	0.999736	
Hypothesized Mean Difference	0	
df	699998	
t Stat	-24.9004	
P(T<=t) one-tail	4.23E-137	
t Critical one-tail	1.644856	
P(T<=t) two-tail	8.46E-137	
t Critical two-tail	1.959967	