2020

Evaluation of roadway spatial-temporal travel speed estimation using mapped low-frequency AVL probe data

Liqun Peng

Zhixiong Li
University of Wollongong, lizhixio@uow.edu.au

Chenhao Wang

Thompson Sarkodie-Gyan

Evaluation of roadway spatial-temporal travel speed estimation using mapped low-frequency AVL probe data

Abstract

© 2020 Elsevier Ltd The rapid increase in the number of vehicles equipped with GPS devices has resulted in using automatic vehicle location (AVL) data as probes to identify traffic flow status as well as route travel speed on a very fine spatial-temporal scale. However, these traffic monitoring approaches heavily rely on the widely distributed probe vehicles in the network and the high frequency of these probe samples, which are rarely implemented in the real world. This study aims to analyze the applicability of providing accurate traffic flow information from four types of low-frequency AVL data. Each data source is applied for speed estimation to develop guidelines on GPS data requirements for travel speed estimation. First, the probe sample size of each data source on each target corridor is studied to reveal the road segments that have the potential for speed estimation, along with the GPS sampling frequency of each data source. Second, the impact of probe vehicle types, sample sizes, and GPS sampling frequency is analyzed. This study offers guidance in using GPS data to conduct speed estimation in different scenarios, which can be further implemented in a prototype software tool for estimating the real-time travel speed. This study has shown the applicability for speed estimation from four types of GPS data, where the transit bus GPS data provides the best mean speed estimation. The speed estimation results are compared with loop detector data on a test road segment to evaluate its accuracy. The comparison results show that given the current GPS data sample size and updating frequency, the transit bus GPS data can provide a reasonably accurate estimation of the traffic flow speed with a mean absolute speed difference of 6.96 km/h.

Disciplines

Engineering | Science and Technology Studies

Publication Details

Peng, L., Li, Z., Wang, C. & Sarkodie-Gyan, T. (2020). Evaluation of roadway spatial-temporal travel speed estimation using mapped low-frequency AVL probe data. Measurement: Journal of the International Measurement Confederation, 165

Evaluation of Roadway Spatial-Temporal Travel Speed Estimation Using Mapped Low Frequency AVL Data as Probes

Liqun Peng[1,3], Zhixiong Li[2*], Chenhao Wang[3], Thompson Sarkodie-Gyan[4]

1 School of Transportation and Logistics, East China Jiaotong University, 808 Shuanggang East Road, Nanchang, Jiangxi, China 330013

2 School of Mechanical, Materials, Mechatronic and Biomedical Engineering, University of Wollongong, NSW 2522, Australia

3 Department of Civil and Environmental Engineering, University of Alberta, ICE 6-327, Edmonton, Alberta, Canada, T6G 1H9

4 Laboratory for Industrial Metrology and Automation, College of Engineering, University of Texas, El Paso, TX, USA

*Corresponding Author: zhixiong_li@uow.edu.au

Abstract: With the road networks covered by rapidly increasing vehicles equipped with GPS devices, it has become common practice to use AVL (Automatic Vehicle Location) data as probes to perform identification of traffic flow status as well as route travel speed at a very fine spatial-temporal scale. However, these traffic monitoring approaches heavily reply upon the wide distributed probe vehicles in the network and high frequency of these probe samples, which are rarely implemented in real world. This study is to analyze the applicability of providing accurate traffic flow information from four types of low frequency AVL data, each data source is applied for speed estimation respectively, so as to develop guidelines on GPS data requirement for travel speed estimation. Firstly, the probe sample size of each data source on each target corridor is studied to reveal the road segments that have the potential for speed estimation, and the GPS sampling frequency of each data source is introduced as well. Secondly, the impact of probe vehicle types, sample sizes and GPS sampling frequency is analyzed. In conclusion, the study developed a guidance for using GPS data to conduct speed estimation in different scenarios, which can be further implemented to a prototype software tool for estimating the real-time travel speed. Our study has shown the applicability for speed estimation from four types of GPS data, where the transit bus GPS data provides the best mean speed estimation. The speed estimation results are compared with loop detector data on a test road segment to evaluate its accuracy. The comparison results have shown that given the current GPS data sample size and updating frequency, the transit bus GPS data can provide a reasonably accurate estimation of the traffic flow speed with the mean absolute speed difference of 6.96 km/h.

Keywords: Travel speed estimation, AVL probes, Low frequency GPS data, Sample requirement

1. Introduction

A major component of ITS (Intelligent Transportation Systems) application is Advanced Traveler Information System (ATIS), with one of its key components being the provision of accurate vehicle travel speed information. Vehicle travel speed is a valuable tool both for network managers and road travellers. It is a key indicator of developing problems on a network and one of the few congestion measures public users really understand. Travel speed is important information for various actors of a transport system, ranging from city planning, to day to day traffic management, to individual travellers. They all make decisions based on average travel speed or variability of travel speed among other factors.

Typically, there are three ways to measure travel speed on an urban and highway network: 1) loop detectors, 2) GPS data, and 3) off-call cellular phone data. Measurement using loop detectors in the highway/urban network, requires detectors and is limited due to speed being measured only at certain points on the network. Additionally, loop detectors are not very effective in arterials with numerous traffic lights due to intermittent traffic movement. The high installation and maintenance cost of loop detector makes them less desirable. Location measurements using GPS data or off-call cellular phone data, can be used to calculate average vehicle speed, and depending on the technology used can also provide point speed along a particular road or highway.

The use of GPS probes in traffic management is growing rapidly as the required data collection infrastructure is increasingly in place, with a significant number of mobile sensors moving around covering expansive areas of the road network. Many travellers carry with them at least one device with a built-in GPS receiver. Furthermore, vehicles are becoming more and more location aware. Vehicles in commercial fleets are now routinely equipped with GPS. This research specifically investigates the estimation of traffic speed from various GPS data sources. The estimation relies on GPS data from City of Edmonton transit buses, Alberta Transportation snow plows, City of Edmonton SmartTravel app, and Shaw GPS Fleet Vehicle Tracking. The main objective of this paper is to evaluate the accuracy and reliability of travel speed estimation using sparse low frequency GPS data sources available in network. Through analysis of the historical GPS data from above datasets, we examine the effects of GPS sampling frequency, GPS sensor accuracy, travelled route, and probe vehicles type on accuracy and reliability of estimating average travel speed.

The remainder of this paper is organized as follows: Section 2 summarizes the related works on travel speed estimation. Then, a description is presented for identifying a set of activities that are necessary to perform the travel time estimation in Section 3. This is followed by a description of experimental data, results and discussions and a conclusion.

2. Related Work

Recent progress in advanced technologies for intelligent transportation systems has enabled the extraction of traffic information from many different sources and in multiple formats. Traffic data sources can be classified in several ways, while in this case the classification presented by Lim and Lee has been chosen [1]. They classify traffic detection systems into two main categories: point detectors and interval detectors. The interval detectors can be further divided into automatic vehicle identification (AVI) techniques and probe vehicle technologies.

Point detectors have been the main source of traffic information in the past decades. This type of detector is set in fixed points of the road and captures traffic variables in these specific points. The most conventional point detectors are the inductive loops that can be further categorized into single and double loop detectors. Single-loop detectors consist of a single induction loop that generates a magnetic field and is able to detect the passing of large metallic objects, in this case vehicles. These detectors output variables such as flow (number of passing vehicles per hour) and occupancy (proportion of the time that the detector is occupied). Substantial studies have focused on this indirect estimation of roadway traffic speeds from single-loop detectors [2, 3], and additional research efforts have also been made in improving the accuracy of single-loop based speed estimation [4, 5]. Double-loop detectors consist of a pair of single-loop detectors set very close to each other. This pair of sensors is capable of obtaining flow and occupancy but they can also collect point speed and vehicle lengths by using the travel time of the vehicles between the two sensors. In practice, it is very common that this velocity information is provided in an aggregated form, where the measurements of several vehicles are combined in different forms [6][7]. Besides the loop detectors, there are other types of point detectors that are available for estimating traffic speeds, including electronic toll collection data [8], infrared and radar technology, video image detection technology [9-11] and so on. These conventional sensors can provide high quality data and are not affected by external factors. They are usually widely deployed along a roadway. However, their installation and maintenance are expensive and complicated, which limit their spatial coverage, especially for arterials.

Another promising approach for measuring traffic speeds is to use probe vehicle technologies, which are capable of tracking probe vehicles by recording position information at regular time interval or space interval. The use of probe vehicles can provide the information of vehicles' trajectories, and travel times between two points can be easily derived. The emerging technologies include smart phones, global positioning systems (GPS), and automatic vehicle location (AVL) systems. Those probe vehicles act as mobile traffic sensors equipped with tracking devices (e.g., GPS or mobile phones), and send location, direction and speed information every few seconds or minutes. They are being used to collect network-wide traffic information such as instantaneous speeds and travel times at any network location without the need of roadside equipment. There are a number of sophisticated methods can be utilized to process probe data for different analyses or applications [12, 13].

However, the most challenge of these services is upon the low sampling (also called reporting, or polling) rate (less than one per minute), which creates difficulties in inferring the travel time of the vehicle traversed between two designated link positions [14]. Furthermore, a few probe vehicles may cause unpredicted data error or missing in the field operation environment due to the GPS device failure and signal interference, in consequence, the fraction of the reported travel time is not accurate. Several non-parameter models are proposed to estimate the probabilities of traffic congestion [15, 16], vehicle stops and travel delays between geo-locations by considering the speed limit, intersections and traffic environment, and allocate the travel time between two consequent observed position into each individual segment [17-21]. These methods mitigate the impact of the sparse and low-frequency vehicle probes on the accuracy of travel time estimation to a certain extent.

On the other hand, the required probe sample size is crucial to represent realistic traffic conditions accurately and reliably. Zou, Xu and Zhu propose a method for arterial speed estimation by utilizing taxi GPS data from 100 vehicles in Guangzhou, China [22]. Their study shows that the number of probe vehicles accounting for 3% of total traffic result in significantly lower errors for travel speed estimation. Lorkowski, Mieth and Schäfer discuss the potential applications of probe vehicle data, such as dynamic routing and automatic congestion detection, using GPS data from 700 taxis in Stuttgart, Germany [23]. Their results indicate that probe vehicles accounting for about 1% of total traffic are required to estimate traffic conditions. Although the previous work clearly demonstrates the feasibility to extract useful information from low-frequency probe vehicle data, it is valuable to leverage the minimum requirement of low frequency GPS data sources available in network with the accuracy and reliability of travel speed estimation. By investing in in-house tools for travel time estimation, we can ensure the integrity of travel speed results and maximize its data assets.

3. Travel Speed Estimation

In this study, a probe-based link travel speed estimation method is proposed. This method calculates estimated link travel speed based on link travel time detected from GPS probes, e.g. transit buses, snow plow trucks, GPS-enabled mobile phones and commercial vehicles. Suggested by Hellinga et al. [17], inferring traffic conditions from positioning data requires five steps: map-matching, path identification, probe filtering, travel time allocation and travel time aggregation. A practical method to estimate link travel speed using different GPS data sources is introduced: (1) the road network model built for map-matching and path identification process, and (2) the proposed link travel time estimation method. It should be noted that three out of four GPS data sources applied in this study are directly collected from GPS-equipped vehicles, and the other one, which is collected from a mobile application, has done the probe filtering before uploading data to the server, so the probe filtering process is not a concern here.

3.1 Road Network

The road network consists of two parts: nodes and links. A node represents a simplified geographical feature of the road network, such as intersections, dead ends of road segments, locations of a change in the road attribute, and on-ramps or off-ramps, as shown in Fig. 1. The set of nodes included the road network is denoted by $\boldsymbol{N} = \{n_i | i \epsilon I\}$, where I is the total number of nodes. A link represents a real road segment connecting two nodes. The set of links in the digital map is denoted by $\boldsymbol{L} = \{l_j | j \in J\}$, where *J* is the total number of links. Note that the two-dimension coordinates of each node and the length of each link can be obtained using spatial analysis application such as ArcGIS.

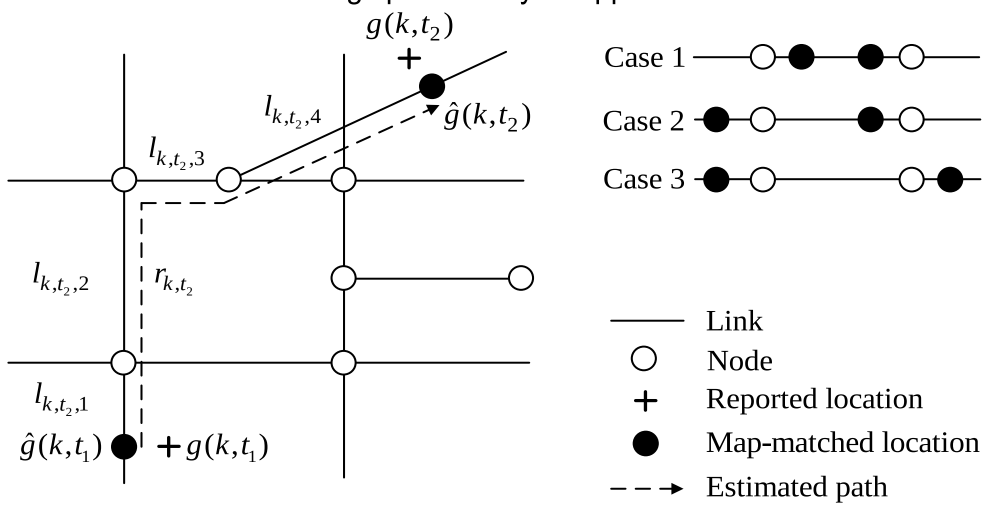


Fig. 1. Examples of Probe Distribution in Network

Considering a GPS location report collected from vehicle *k* at time *t*, denoted by $g(k,t)$, it can be represented by a set of two-dimension coordinates, $g(k,t) = (x_{k,t}, y_{k,t})$. After map-matching process, the reported location is projected to the road link which is inferred to be the actual road segment the vehicle is on. The map-matched location is defined by $\hat{g}(k,t) = \left(l_{k,t}, d\left(l_{k,t}\right)\right)$, where $l_{k,t}$ is the matched road link, and $d\left(l_{k,t}\right)$ is the distance from the map-matched position to the beginning of the link. Thus, the position information is transferred into one-dimension coordinate. After path identification, the path traveled by the GPS probe between two consecutive map-matched locations $\hat{g}(k,t_1)$ and $\hat{g}(k,t_2)$ $(t_2 > t_1)$, namely a GPS report pair $p_{k,t_2} = \left(\hat{g}(k,t_1), \hat{g}(k,t_2)\right)$, is defined as r_{k,t_2}, and the links contained in the travel path can be defined as follows:

$$r_{k,t_2} = \{l_{k,t_2,1}, l_{k,t_2,2} \dots l_{k,t_2,\mathrm{m}} \dots \mathrm{l}_{\mathrm{k},\mathrm{t}_2,\mathrm{M}}\} \quad (1)$$

where *M* is the total number of traversed links. Figure 3.2 gives an example of a probe vehicle *k* with two consecutive location reports. The path travel time and travel distance of two consecutive location report from a probe vehicle can be simply calculated by

$$TT_{k,t_2} = t_2 - t_1 \quad (2)$$

$$\mathrm{D}_{k,t_2} = \left(leng\left(l_{k,t_2,1}\right) - d\left(l_{k,t_1}\right)\right) + \sum_{m=2}^{M-1} leng\left(l_{k,t_2,m}\right) + d\left(l_{k,t_2}\right) \quad (3)$$

where $leng\left(l_{k,t_2,m}\right)$ denotes the length of link $l_{k,t_2,m}$. Note that when M =1, $\mathrm{D}_{\mathrm{k,t_2}} = \left(\mathrm{leng}\left(\mathrm{l}_{\mathrm{k,t_2,1}}\right) - d\left(\mathrm{l}_{k,t_1}\right)\right)$; when M =2, $\mathrm{D}_{k,t_2} = \left(leng\left(l_{k,t_2,1}\right) - d\left(l_{k,t_1}\right)\right) + d\left(l_{k,t_2}\right)$.

3.2 Link Travel Speed Estimation

The time interval between two consecutive GPS reports of one probe vehicle, namely the GPS sampling interval, is one of the key factors that determine the accuracy of link travel time estimation. For example, with a high resolution of GPS sampling interval, e.g. 1 second or even less, we can know the exact time when the vehicle enters and leaves a road link disregarding positioning error caused by GPS device itself; meanwhile, in practice, the GPS sampling interval is much longer, e.g. over 30 seconds. Therefore, to estimate link travel speed, the path travel time should be first allocated to each included link, then by simply dividing the length of the link by the link travel time, the average link travel speed of the probe vehicle can be obtained. This refers to the aforementioned process of travel time allocation. As shown in Fig.1, a GPS pair can be categorized into three cases: Case 1: there is no node on the travel path, that is the two GPS location reports are on the same link; Case 2: there is one node on the travel path, that is the two GPS location reports are on adjacent links; Case 3: there are at least two nodes on the travel path, that is at least one complete link exists between the two GPS location reports.

A benchmark travel time allocation method is proposed to allocate path travel time into each link based. This method is to first estimate the timestamp when a probe vehicle *k* passes a node, then the link travel time can be simply represented by the time difference between the estimated timestamp of its upstream and downstream nodes. The path travel time can be further decomposed into three constituents: (1) Free-flow travel time, which is the time interval that a vehicle traveling through the road segment with free-flow speed; (2) Congestion delay, which is the additional travel time caused by vehicle driving at a speed lower than free-flow speed due to traffic congestions; (3) Stopping delay, which occurs when the vehicles are forced to stop due to traffic control system, e.g. traffic signals. It includes the acceleration and deceleration time caused by stopping.

Considering a GPS pair of vehicle *k*, $p_{k,t_x} = \left(\hat{g}(k, t_{x-1}), \hat{g}(k, t_x)\right)$, which includes at least one node on its travel path (e.g. case 2 and 3), by assuming that the average travel speed on each link is not substantially different,

$$t\left(n_{k,t_x,m}\right) = t_1 + \frac{L^{up}(n_{k,t_x,m})}{L^{up}(n_{k,t_x,m}) + L^{down}(n_{k,t_x,m})} TT_{k,t_2} \quad (4)$$

where $n_{k,t_x,m}$ denotes the downstream node of link $l_{k,t_x,m}$, $t\left(n_{k,t_x,m}\right)$ denotes the estimated timestamp when vehicle *k* passes node $n_{k,t_x,m}$, $L^{up}(n_{k,t_x,m})$ and $L^{down}(n_{k,t_x,m})$ denote the travel distance from node $n_{k,t_x,m}$ to the upstream and downstream map-matched GPS location respectively. Then the estimated travel time of vehicle *k* on link *m* is calculated by

$$tt\left(l_{k,t_x,m}\right) = t\left(n_{k,t_x,m}\right) - t\left(n_{k,t_x,m-1}\right) \quad (5)$$

The assumption of Equation (4) is reasonable for uninterrupted traffic flow, such as freeways, since there is no traffic control device that enforces vehicles to stop on the roadway and the stopping delay can be considered as zero. However, for interrupted traffic flow, e.g. urban arterials, the existence of signalized or un-signalized intersections makes stopping delay no ignorable. In fact, the stopping delay is the main component of traffic delay on arterial. For case 2 and 3, the stopping delay usually happens on the upstream of nodes, and the benchmark method could allocate part of the stopping delay of the upstream link to the downstream link.

To improve the benchmark method, the instantaneous GPS speed is taken into consideration to infer and allocate stopping delay. Given a map-matched trajectory of probe vehicle k, which includes X GPS pairs, $p_{k,t_1}, p_{k,t_2}, \ldots, p_{k,t_x}, \ldots, p_{k,t_X}$, to estimate the timestamp when the vehicle passes node $\mathrm{n}_{k,t_x,m}$, the procedure is described as follows:

Step 1: obtain the adjacent GPS pairs $p_{k,t_{x-1}}, p_{k,t_{x+1}}$;

Step 2: calculate the number of nodes passed in p_{k,t_x}, denoted by $c(p_{k,t_x})$; compared $c(p_{k,t_x})$ with a predefined value $c_{thres}(x)$, if $c(p_{k,t_x}) > c_{thres}(x)$, go to *Step 9*, otherwise go to *Step 3*;

Step 3: calculate average travel path speed $v(p_{k,t_x})$ using Equation (6)

$$v(p_{k,t_x}) = \frac{D_{k,t_x}}{t_x - t_{x-1}} \tag{6}$$

and then compare it with the predefined speed threshold pv_{thres}; if $v(p_{k,t_x}) > pv_{thres}$, go to *Step 9*, otherwise go to *Step 4*;

Step 4: compare instantaneous GPS speed $v_{k,t_{x-1}}, v_{k,t_x}$ of $\hat{g}(k, t_{x-1}), \hat{g}(k, t_x)$ with the predefined point speed threshold v_thres_p; if both $v_{k,t_{x-1}}$ and v_{k,t_x} are higher or not less than v_{thres}, go to *Step x*, else if $v_{k,t_{x-1}} \leq v_{thres}$ and $v_{k,t_{x+1}} > v_{thres}$, go to *Step 5*, otherwise go to *Step 7*;

Step 5: calculate $c(p_{k,t_{x-1}})$; if $c(p_{k,t_{x-1}}) > c_{thres}(x-1)$, go to *Step 6*, otherwise go to *Step 9*;

Step 6: calculate $v(p_{k,t_{x-1}})$ using Equation (6); if $v(p_{k,t_{x-1}}) > pv_{thres}$, calculate $t(n_{k,t_x,m})$ using Equation (7)

$$t(n_{k,t_x,m}) = t_1 + \frac{L^{up}(n_{k,t_x,m})}{v(p_{k,t_{x-1}})} \tag{7}$$

otherwise go to *Step 9* ;

Step 7: calculate $c(p_{k,t_{x+1}})$; if $c(p_{k,t_{x+1}}) > c_{thres}(x+1)$, go to *Step 8*, otherwise go to *Step 9*;

Step 8: calculate $v(p_{k,t_{x+1}})$ using Equation (6); if $v(p_{k,t_{x+1}}) > pv_{thres}$, calculate $t(n_{k,t_x,m})$ using Equation (8)

$$t(n_{k,t_x,m}) = t_2 - \frac{L^{down}(n_{k,t_x,m})}{v(p_{k,t_{x+1}})} \tag{8}$$

otherwise go to *Step 9* ;

Step 9: calculate $t(n_{k,t_x,m})$ using Equation (4).

The threshold values applied in the procedure are explained here. $c_{thres}(x)$ is to clarify which case the GPS pair belongs to, for example when $c_{thres}(x) = 1$, $c(p_{k,t_x}) > c_{thres}(x)$ means p_{k,t_x} is case 3. pv_{thres} is a predefined value of average path travel speed, which is used to determine if significant delay occurs in the path. The v_{thres} is a predefined threshold for instantaneous GPS speed, which is used to determine if the delay occurs on the upstream of the node or on the downstream of the node. For example, if $v_{k,t_{x-1}} \leq v_{thres}$ and $v_{k,t_{x+1}} > v_{thres}$, then it is assumed that the delay occurs on the upstream of the node.

4. Data Collection and Analysis

The research scope of this study is limited to four geographic areas, as shown in Fig.2, including the Anthony Henday Drive, Whitemud Drive, Yellowhead Trail (from 170 Street NW to 75 Street NW) and Queen Elizabeth II Highway (from the Edmonton International Airport to the Anthony Henday Drive) in the City of Edmonton, Alberta, Canada. An attempt will be made to apply developed algorithms to non-urban areas and assess their performance where GPS data availability is more limited.

Fig. 2. Research Scope

4.1 Introduction of GPS Data Sources

There are four GPS data sources involved in this study: ETS Smart Bus GPS Data, Snow Plow GPS Tracking, SmartTravel App GPS Data, and Shaw Commercial Vehicle GPS Tracking. This section gives a brief introduction to these data sources, especially their GPS sampling frequency and sample size. Among the four data sources, SmartTravel data is the only GPS data source collected from general vehicles. Its GPS sampling frequency is 1 second, which makes it much easier to obtain travel path information and accurate link travel speed. If the mobile device has other sensors such as gyroscope or accelerometer, the sensor data can also be collected. Table 1 summarizes the available information from the four data sources. Vehicle ID, timestamp and position are basic information for speed estimation, and speed, direction and route schedule are supplementary information, which is helpful for improving the accuracy and reliability of map-matching and speed estimation. Besides, the real-time capability of ETS Smart Bus data and Snow Plow data gives potential usage for real-time application.

Table 1. Summary of GPS Data Features

Data Source	Basic info			Supplementary info			Real-time availability	
	Vehicle ID	Timestamp	Position	Speed	Direction	Route Schedule	Real-Time	Historical
ETS smart bus	√	√	√			√	√	√
Snow plow	√	√	√	√	√		√	√
SmartTravel	√	√	√	√	√			√
Shaw GPS Tracking	√	√	√					√

4.2 GPS Sampling Frequency

GPS sampling frequency measures how many GPS records can be obtained from one probe vehicle in a certain time period. Higher GPS sampling frequency provides better accuracy of map-matching and speed estimation, since the uncertainty of probe vehicle trajectory is decreased.

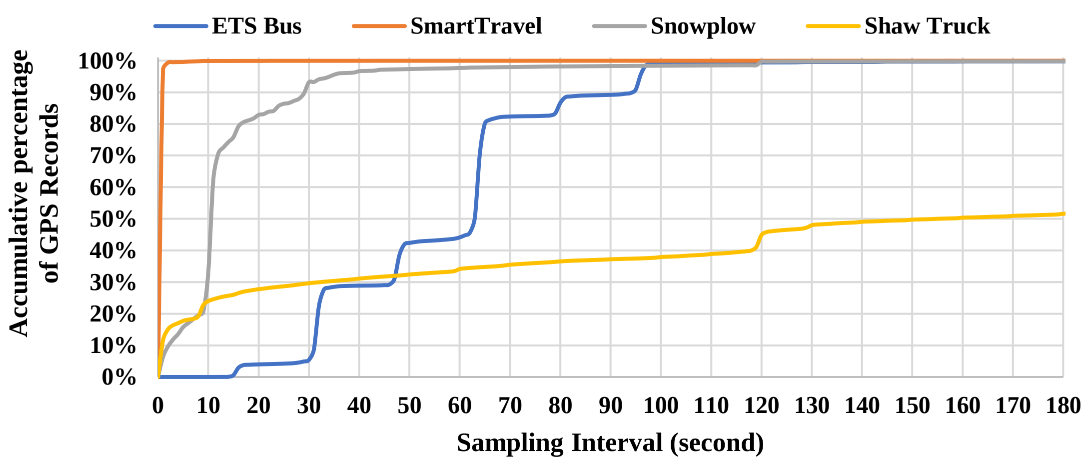


Fig. 3. GPS Sampling Interval of Different Data Sources

GPS sampling interval is the reciprocal of GPS sampling frequency. It refers to the time gap between two consecutive GPS records of one probe vehicle. Fig. 3 shows the accumulative proportion of GPS records with different GPS sampling interval for each data source, and Table 2 shows the proportion of records with different GPS sampling intervals. SmartTravel data has the best sampling frequency, with 97.1% GPS records having 1-second sampling interval. As for Snow Plow data, the sampling interval of about 50% data is between 10 to 12 seconds, and the sampling interval of over 90% data is under 30 second. The third best is ETS Bus data, with 80% GPS records having less than 65-second sampling interval. Regarding the Shaw GPS tracking, 12% data has 1-second sampling interval, while 48.4% data has sampling interval over 180 seconds.

Table 2. GPS Sampling Interval

Sampling Interval (second)	% of GPS records			
	ETS Bus	Snow Plow	Smart Travel	Shaw
(0, 10]	0%	33%	100%	24%
(10, 30]	5%	60%	0%	6%
(30, 60]	39%	5%	0%	4%
(60, 120]	55%	2%	0%	11%
(120, 180]	1%	0%	0%	7%
180+	0%	0%	0%	48%

4.3 Sample Size Distribution

In this study, a sample is defined as an estimate on one road link from one probe vehicle. For travel time estimation, it is important to first analyze the spatial and temporal distribution of the sample size generated from each data source, thus knowing if the GPS data can support speed estimation for a certain road link. For example, the ETS Bus data can cover most urban arterials, while it may not provide a satisfying sample size for freeways; the Snow Plow data is collected from snowplows, which do not work on arterial.

Fig. 4 to 7 illustrates the daily average sample size on each corridor for every 15-minute time slot. The X axis denotes the time of day and the Y axis denotes landmarks (e.g. intersections) along the corridor. For ETS Bus data, it provides relatively large sample size on the road segment between 170 Street and 122 Street on Whitemud Drive, especially during AM and PM peak hours. This is because of that, first, there are more bus routes passing through the road segment; second, more buses of these routes are operated during peak hours, as the bus departure interval reduces from 30 minutes to 15 minutes. Similarly, the samples generated from ETS Bus is large on Yellowhead Trail, especially during peak hours, and less on Anthony Henday Drive, which is consistent with the bus route schedule. Specifically, the bus route 747, known as the Sky Shuttle, provides all samples on Highway 2, and it can cover the whole road segment from Anthony Henday Drive to the airport.

For the Snow Plow data, the samples are mainly on Highway 2, as it is one of the main corridors in the schedule of snowplows. Another factor that limits the sample size of Snow Plow is the limited number of monitored snowplows. In the provided real-time Snow Plow data feed, there are in total 22 snow plow trucks, which, compared with other data sources, is relatively small. It should be noticed that the snow conditions on road also could impact the sample size of Snow Plow. For SmartTravel data, we can find the samples on all four corridors, yet compared with ETS Bus data, the daily average sample size on each corridor is not constant as the users' travel activity is relatively random and cannot be controlled. However, the mobile application can be improved and the number of users will continually increase, and the sample size, especially during peak hours, will hopefully increase and be stable. For Shaw data, the samples are almost on Yellowhead Trail, which is reasonable because the data is collected from commercial vehicles and

Yellowhead Trail is the main access to many commercial and industrial zones.

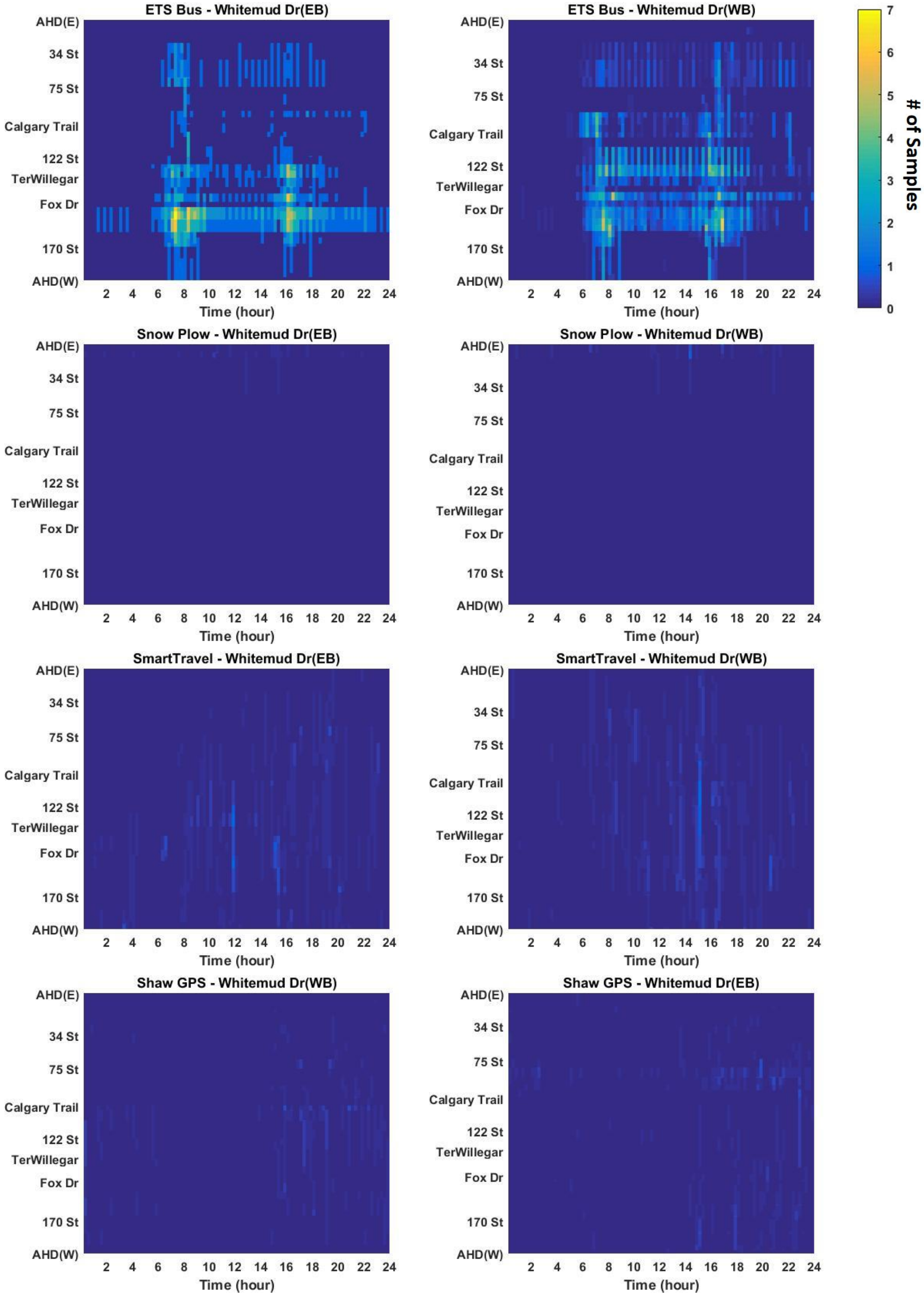


Fig. 4. Distribution of Sample Sizes of GPS Data Sources on Whitemud Drive

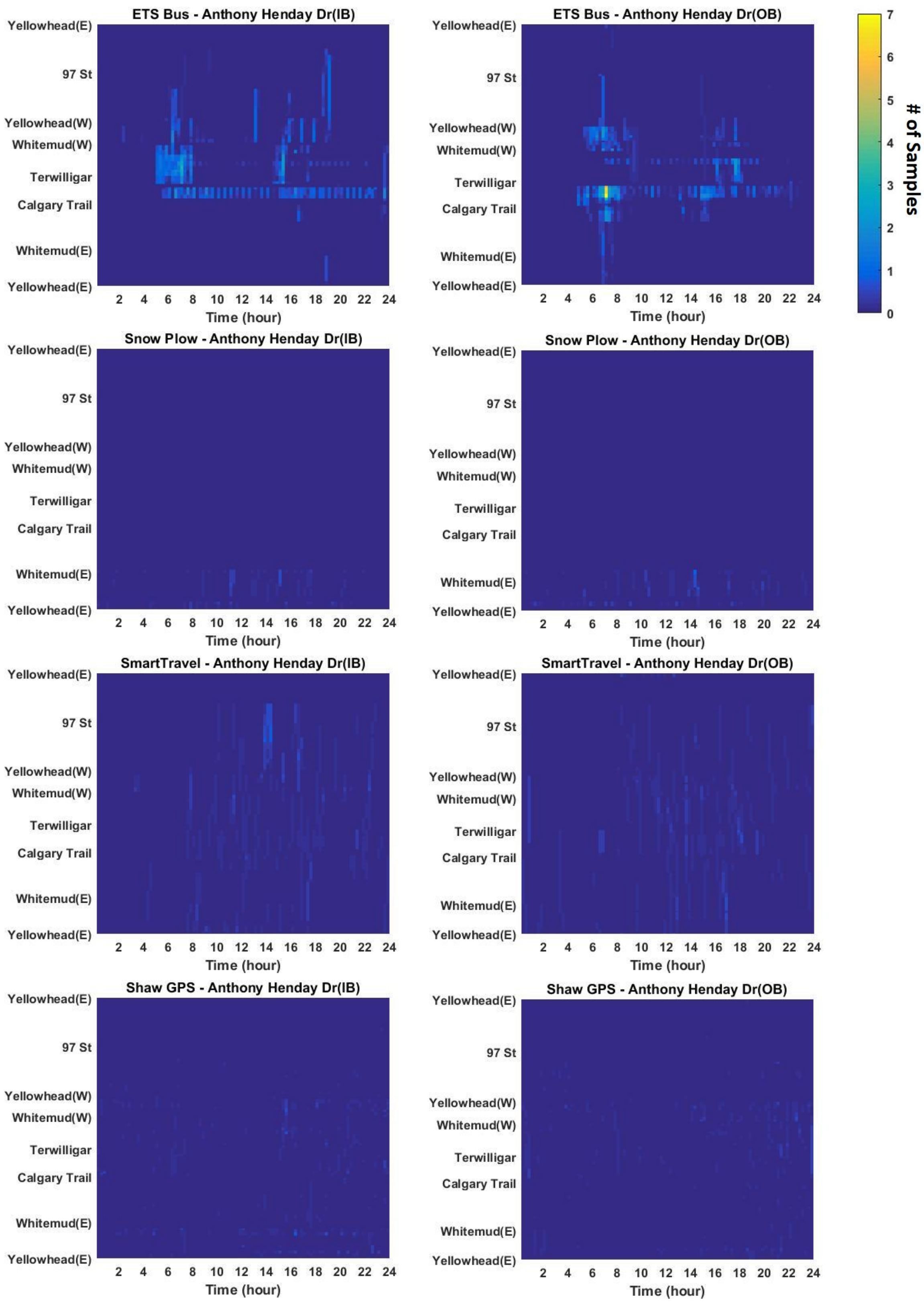


Fig. 5. Distribution of Sample Sizes of GPS Data Sources on Anthony Henday Drive

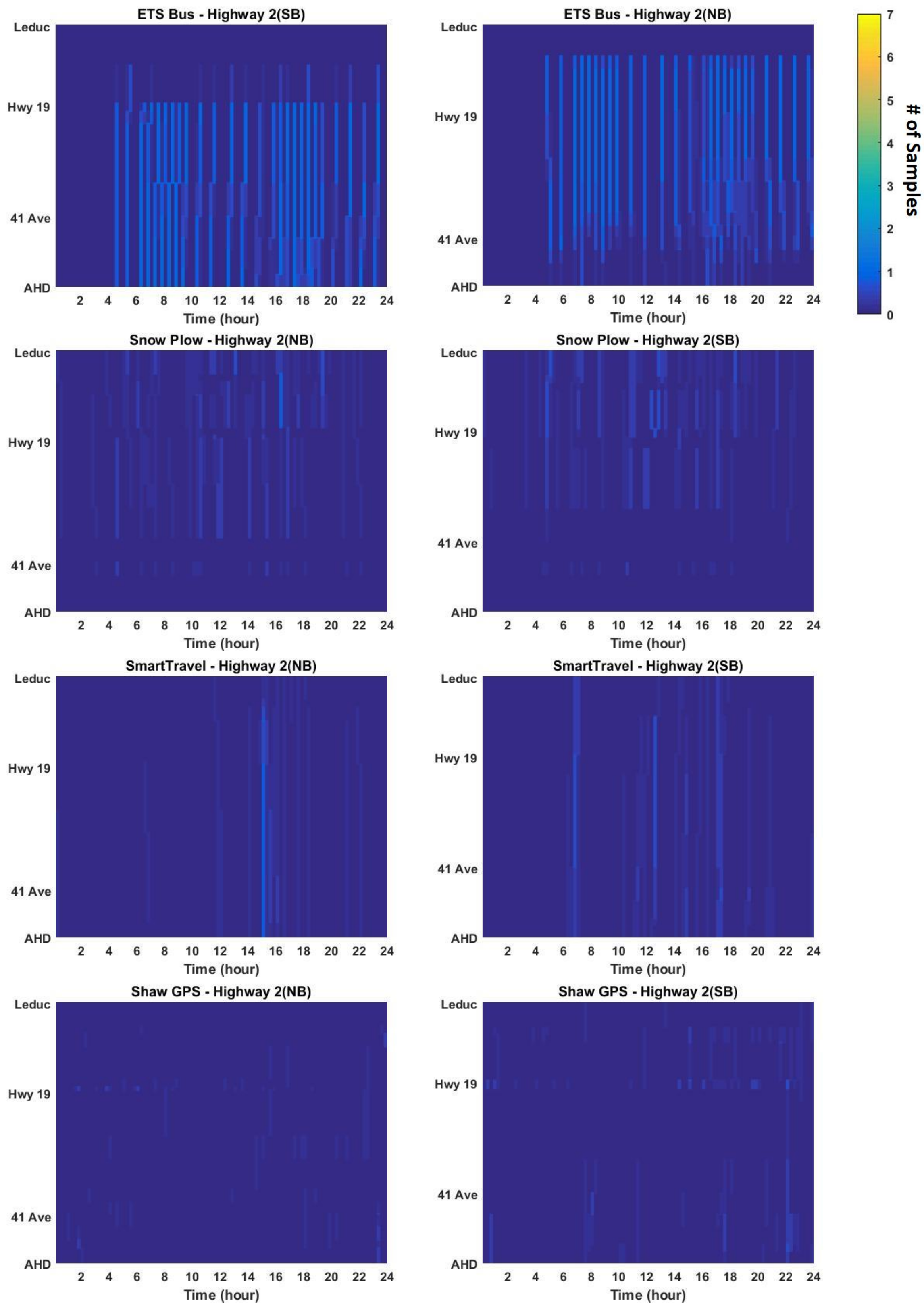


Fig. 6. Distribution of Sample Sizes of GPS Data Sources on Highway 2

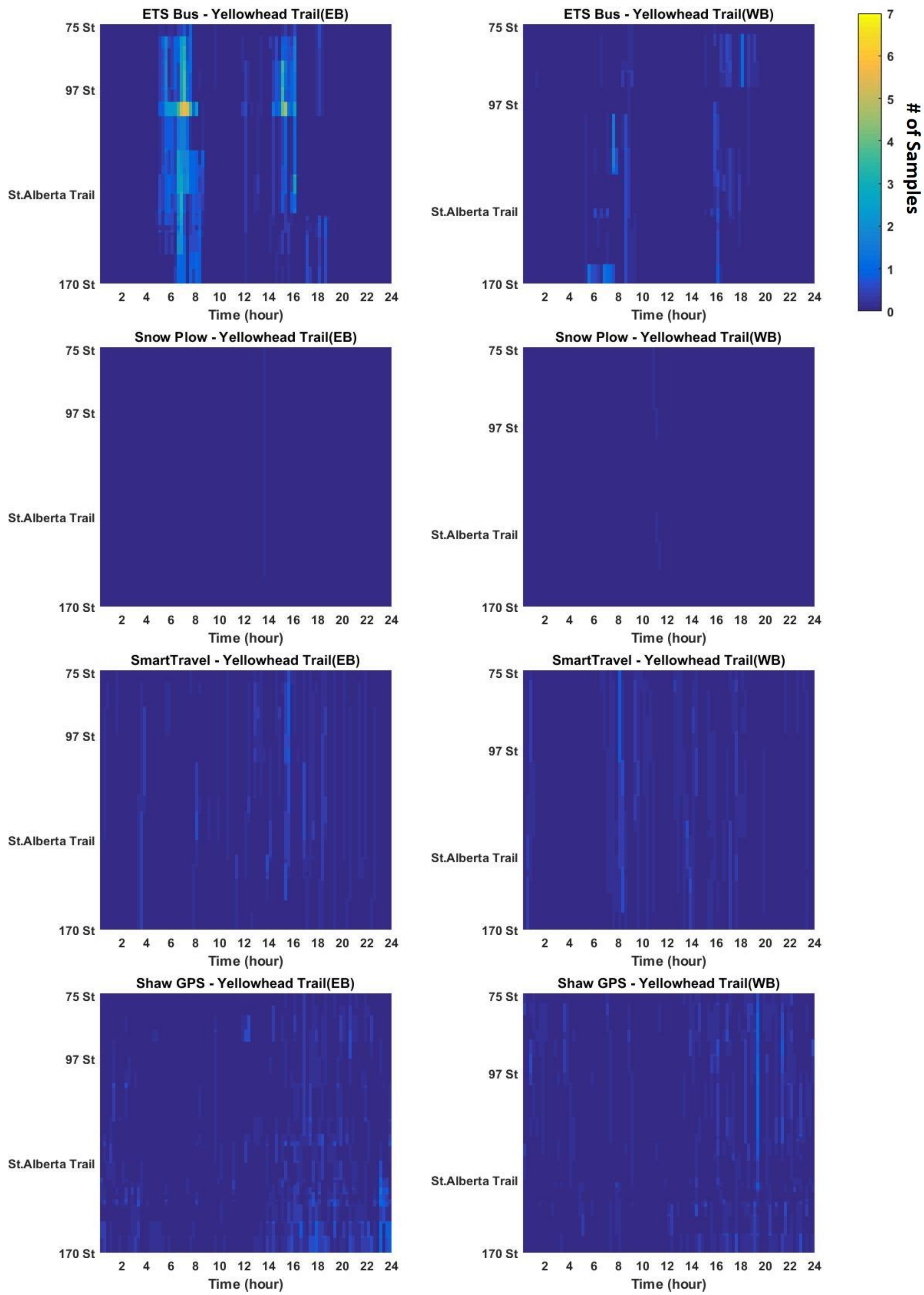


Fig. 7. Distribution of Sample Sizes of GPS Data Sources on Yellowhead Trail

5. Results and Evaluation

In this chapter, we analyze how the characteristics of GPS data and road network influence the accuracy of traffic speed estimation. Section 5.1 analyzes the impact of sample size and probe vehicle type upon link traffic speed estimation. Section 5.2 introduces two data features that may impact the accuracy of sample speed: GPS sampling frequency and road type (e.g. freeway and arterial).

5.1 Impact of Probe Vehicle Type

In this study, Loop Detector Speed Data on the Whitemud Drive are selected as the benchmark to evaluate the speed estimation accuracy of GPS data. The ETS bus data out of four proposed data sources is chosen for this accuracy comparison because it can consistently provide GPS data points due to the pre-determined transit schedule. Furthermore, the ETS bus data also has the greatest sampling sizes as well as the spatial coverage on the Whitemud Drive. However, due to the special type of probe vehicles, the estimated speeds could in general different from the whole traffic flow as buses under certain circumstances travel slower than the general vehicles. The Loop Detectors cover the Whitemud Drive from 122st intersection to 178st intersection as shown in Fig. 8.

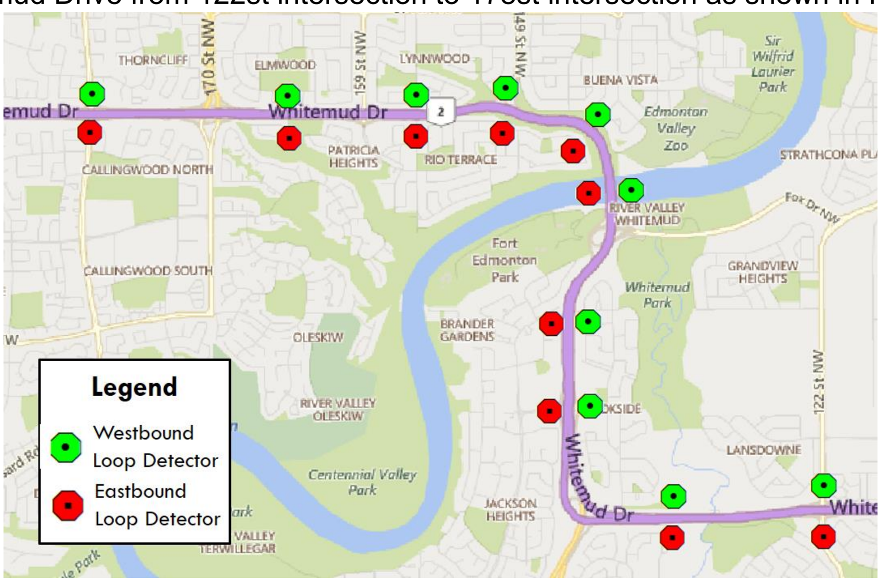


Fig. 8. Loop Detectors (VDS) Locations on Whitemud Drive

In the original Loop Detector Data, loop detectors are constantly recording the speed of passing vehicles in a 20-second interval. If more than one vehicle is passing through the VDS, their average speeds will be noted as the estimated speed within this interval. In this study, the speed data has been aggregated into a 1-minute interval, which means that the average speed of all vehicles passing through one VDS within one minute is noted as the estimated speed in order to smooth the speed profile curve. Furthermore, a 5-minute moving average window is implemented to further smooth out the curve, or otherwise, the speed profile will be very difficult to identify the speed drop. In this case, the VDS speed estimations are used as the reference speed to evaluate the accuracy of speed estimation from the ETS Bus GPS data.

ETS Bus speeds are estimated from the GPS records as mentioned before, and the data interval is 15 minutes. In every fifteen minutes, the average speed of all the sampling records is noted as the estimated speed. Furthermore, the moving average window procedure is also used to smooth out the speed estimation that can better represent the speed drop trend. The loop detector 1034 and 1035 at road segment 25544 and 42896 on the Whitemud Drive is chosen as an example to show the accuracy of the speed estimation. The data covers 10 weekdays from December 5th to 16th, and from 6 A.M. to 10 P.M. on each day.

5.1.1 Evaluation Criteria

In evaluating the speed estimation, the Mean Absolute Speed Difference and Percentage Absolute Speed Difference are used to determine the accuracy of speed estimation. The reason to use the mean absolute speed difference is that the positive and negative speed difference can offset the absolute difference between the ETS and GPS speed estimations. Therefore, applying the absolute speed difference value can represent the deviation of the estimated speed from the reference speed. The percentage absolute speed difference, on the other hand, can help to demonstrate the percentage of the deviation given the reference speeds.

$$MASD = \frac{1}{N}\sum_{k=1}^{N}|v_{GPS} - v_{VDS}| \quad (9)$$

Where MASD = mean absolute speed difference

N = number of samples

v_{GPS} = estimated speed from GPS data

v_{VDS} = estimated speed from VDS data

$$PASD = \frac{1}{N}\sum_{k=1}^{N}\left|\frac{v_{GPS}-v_{VDS}}{v_{VDS}}\right| \quad (10)$$

Where PASD = percentage absolute speed difference

5.1.2 Probe Vehicle Types

The ETS Bus GPS data on the road segments 25544 and 42896 contains in total 2407 samples within 10 weekdays. These samples have covered 1046 15-minute time intervals that at least one sample is presented in these time intervals. The speed differences between the VDS speeds and the estimated speeds from GPS data are considered as the estimation error that is due to the variation of the traffic flow. In the Fig. 9, the box plots have shown the speed difference distribution for different sample size within one time slot.

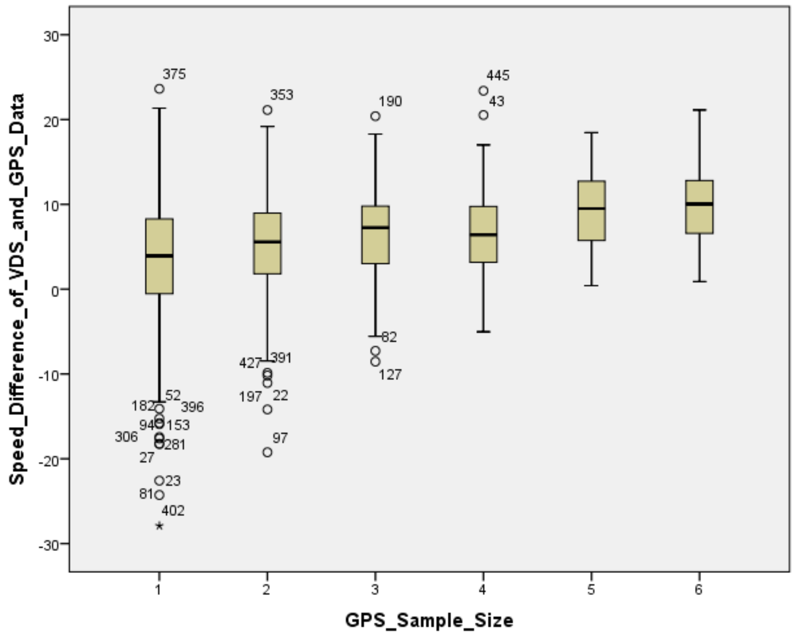


Fig. 9. Box plot of the speed differences between the VDS and GPS speed estimations

From the box plot above, we can see that the distribution for each sample size follows a relative good normal distribution. All the groups have shown that the average VDS travel speed is higher than that of the GPS estimated speed, and the mean is slightly increasing along with the increase of the sample sizes. On the contrary, the standard deviation for each group decreases as the sample size increases. Because larger sample size can provide higher accuracy of the speed estimation, the increasing trend of the mean speed difference could be due to the specificity of probe vehicle type. Because the probe vehicle here is the ETS bus, which could instinctively travel slower than the general traffic flow.

When we take further analysis about the difference between GPS and VDS speed estimations, as we can see from the Fig. 10, the absolute differences between two data sources are shown in a standard normal distribution where the mean difference value is 6.96 km/h. This could due to the bus traffic speed, in general, is slower than the overall traffic speed. And the general distribution is not skewed to the right or left, which means that the speed difference is mostly due to some human random factors. By conducting the one-sample Kolmogorov-Smirnov Test, the result has shown that the speed difference distribution follows the normal distribution with the asymmetric significance less than 0.000. This can further prove that the speed difference between the VDS and GPS speed estimations are due to the random factors with no significant bias.

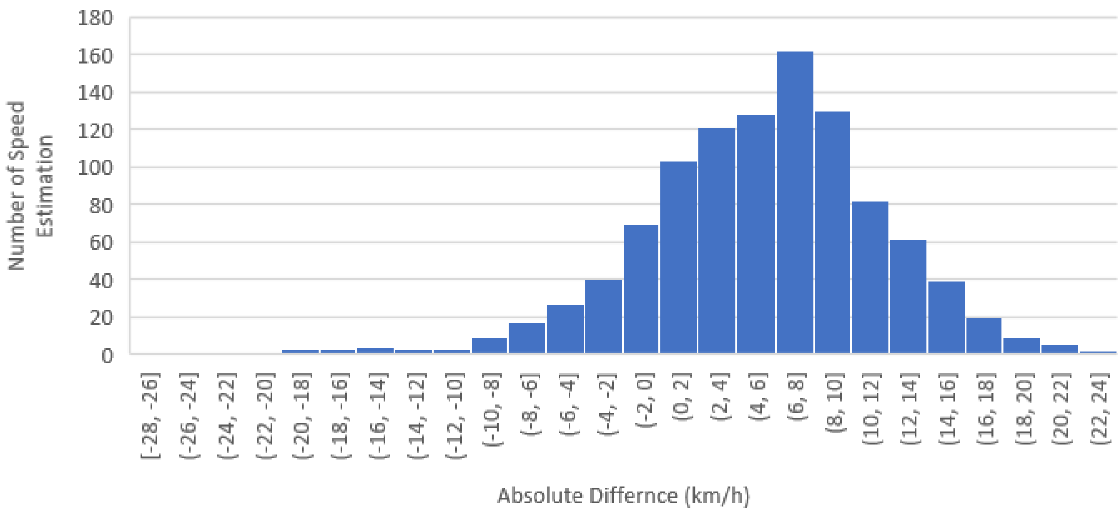


Fig. 10. The distribution of absolute difference of speed estimation between GPS and VDS data

5.2 Sample Size Requirement

In collecting traffic information using probe vehicles, the sample size is one of the critical issues that needs to be considered. It is important to determine the number of probe vehicles because the traffic state can be affected by many variables. Sometimes the variance of traffic speeds within one time interval is very large, and in this case, the sufficient sample size would be important to accurately estimate the current traffic state. Also, higher sample size means that the confidence level of estimation speed will be higher once the standard deviation of speed values and allowable errors are determined. According to the macroscopic traffic flow theory, the required sample size can be calculated using the equation below:

$$n = (\frac{ts}{\varepsilon})^2 \tag{11}$$

where

n = required sample size

t = value for the selected confidence level using t distribution

s = standard deviation of estimated travel speed

ε = allowable error

and T-distribution is used because the sample size is less than 30.

In the equation above, the s represents the standard deviation of travel speed values. In this case, two preliminary assumptions have to be made, where 1) the speed estimation error follows the normal distribution or T-distribution if the sample size is less than 30, and 2) the speed estimation error within one individual time slot follows the same normal or T-distribution of the all the speed estimations. The relationship between the ground truth and the estimated speed can be shown below:

$$\hat{v}_{i,t} = v_t + \varepsilon_{i,t} \tag{12}$$

where

v_t = the ground truth speed at time t

$\hat{v}_{i,t}$ = the ith estimated speed at time t

$\varepsilon_{i,t}$ = the ith estimation error at time t, and the error follows N $((\mu, \sigma^2)$

and the average speed estimations of total N samples at time t equals to:

$$Avg(\hat{v}_{i,t}) = v_t + \frac{1}{N}\sum_{i=1}^{N} \varepsilon_{i,t} \tag{13}$$

Therefore, the standard deviation of the estimated travel speed can be written as the following equation:

$$SD(\hat{v}_t) = [\frac{1}{N}\sum_{i=1}^{N}(\hat{v}_{i,t} - Avg(\hat{v}_{i,t}))^2]^{\frac{1}{2}} \tag{14}$$

By combining the Equation(12) and (13) into Equation (14), $SD(\hat{v}_t)$ can be rewritten as:

$$SD(\hat{v}_t) = [\frac{1}{N}\sum_{i=1}^{N}(\varepsilon_{i,t} - Avg(\varepsilon_{i,t}))^2]^{\frac{1}{2}} \tag{15}$$

The standard deviation of the estimated speeds at time t:$SD(\hat{v}_t)$ can be represented by the standard deviation of the estimation error. And by following the second assumption, the standard deviation of the estimation error is independent from the ground truth speed, and the distribution of estimation error within one time slot equals to the overall distribution of the estimation error for all the time. Thus, the standard deviation of the estimated travel speeds can be calculated by using the Equation (15).

In this study, the required sample size for different combinations of confidence level and errors on road link 25544 and 42896 on Whitemud Drive is shown below in Table 3. There are in total 2407 samples and 1046 time slots, and the standard deviation $SD(\hat{v}_t)$ is 6.411 km/h.

Table 3. Sample size requirements for different combinations of confidence levels and allowable errors.

Confidence level	**ε<3 km/h**	**ε<5km/h**	**ε<10 km/h**	**ε<15 km/h**
70%	5	2	1	1
75%	7	3	1	1
80%	8	3	1	1
85%	10	4	1	1
90%	13	5	2	1
95%	18	7	2	1

As the above table shows the required sample size for different scenarios, the estimation assumes that the GPS points are perfectly accurate. In the real world, due to the location errors of GPS points, the required sample size should be greater than what is displayed in the table, and it depends on the accuracy of the GPS system.

5.3 Impact of GPS Sampling Frequency and Road Type

The GPS sampling frequency can influence the accuracy of estimated sample speed. With lower GPS sampling frequency, probe vehicles may have traversed significant distances between records, which creates difficulties in inferring the true path of the vehicle between two position records. Furthermore, the fraction of the reported travel time that is spent on each individual road link is not observed, which creates challenges for travel speed estimation.

In this section, the historical SmartTravel data of five weekdays is used to evaluate the impact of GPS sampling frequency and the type of travelled routes upon the accuracy of samples. In this study, a sample is defined as a speed estimate of one road link from one probe vehicle, including link ID, vehicle ID, timestamps when the probe vehicle enters and leaves the road link, link travel time and link travel speed.

The SmartTravel mobile application collects vehicle position information every one second, which can offer very accurate vehicle trajectories for sample generation. Due to environmental impacts or network connectivity, sometimes the sampling interval of SmartTravel probes could be more than one second. To build up a reference dataset and provide samples as errorless as possible, an original vehicle trajectory may be split into several sub-trajectories with reassigned vehicle ID. Fig. 11 gives an example demonstrating this process: for a complete vehicle trajectory, when there are two records with sampling interval higher than the predefined threshold (in this study, the threshold is set as 10 seconds), this trajectory is split into two shorter trajectories. Each trajectory is assigned with a new virtual vehicle ID, hence in the processed data the sampling intervals of all virtual vehicle trajectories are lower than the threshold value. Besides, the total travel time of virtual vehicle trajectory should be longer than 5 minutes, and those with total travel time less than 5 minutes are removed. After such data processing, the processed SmartTravel data is considered as the reference dataset, and samples generated the reference dataset are considered as the ground truth for comparison, namely the reference samples.

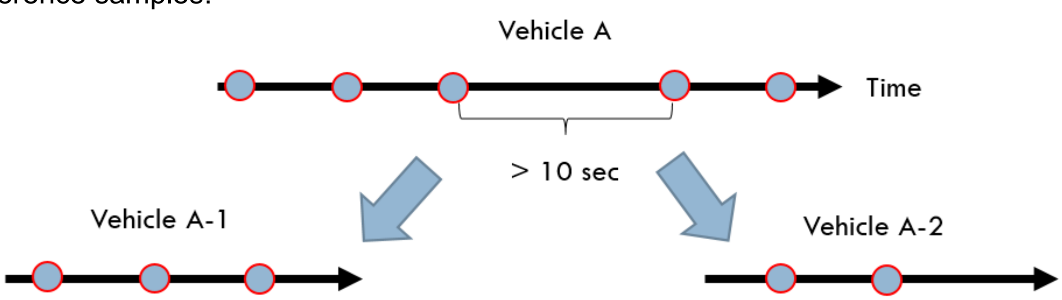


Fig. 11 Example of the process to obtain reference dataset

To evaluate the impact of GPS sampling frequency on the accuracy of samples, by removing partial GPS records of each probe vehicle from the reference dataset, we can obtain several GPS datasets with different sampling intervals, e.g. 10 seconds or 60 seconds. These datasets are considered as the test groups and the samples generated from each test dataset, namely test samples, will finally be compared with the reference samples.

To evaluate the impact of travelled route type on the accuracy of samples, samples on the target four corridors are divided into two groups based on the type of traffic flow. One group is uninterrupted flow, includes Anthony Henday Drive, Whitemud Drive and Highway 2; another group is interrupted flow, includes Yellowhead Trail, on which there are five signalized intersections.

The performance of the test samples is evaluated by the speed difference:

$$\Delta v_k^i = v_k^i - v_{ref}^i \tag{16}$$

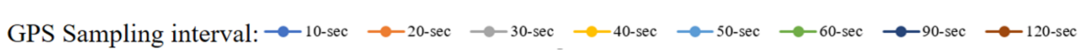
where v_k^i is the speed of sample i generated from test dataset k, and v_{ref}^i is the speed of sample i generated from the reference dataset. Further, the mean value of absolute speed differences is calculated as

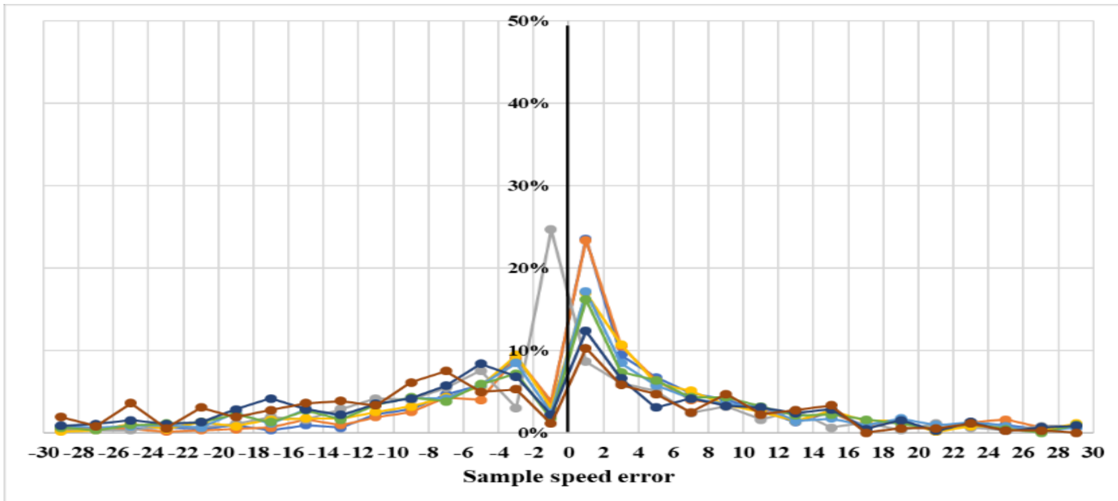
$$MAD_k = \frac{\sum_i^I |\Delta v_k^i|}{I} \tag{17}$$

where I is the total number of samples.

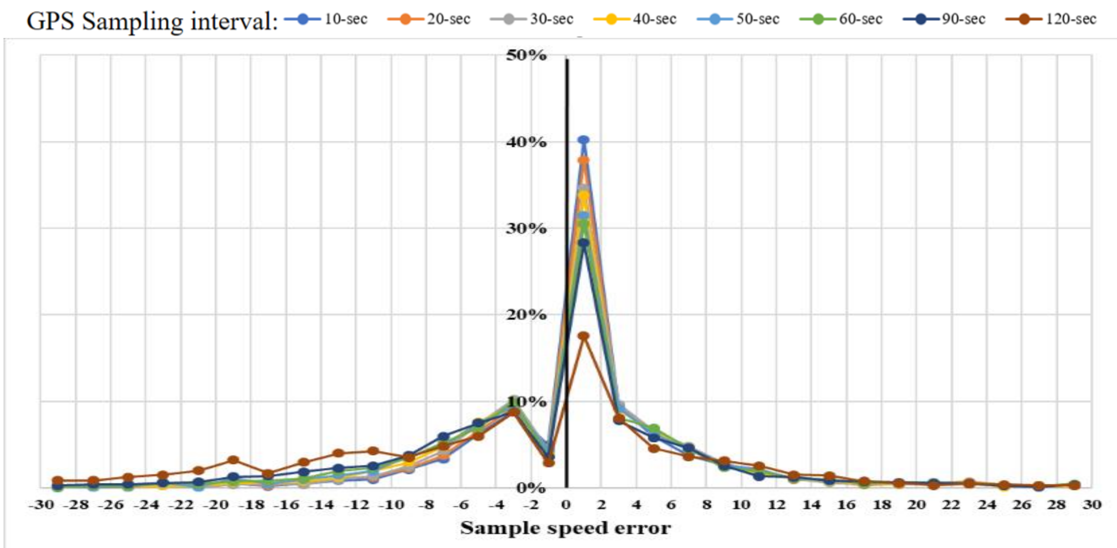
Fig. 12 shows the percentage of samples with different speed difference for each test datasets, and Fig. 13. shows the standard deviation of speed differences for each test datasets. For both freeway and arterial, the speed differences are most within -2~2 km/h, and as the sampling frequency decreases, the standard deviation of speed difference tends to increase, meaning the variance of the sample speed error increases as the sampling frequency decreases. We can also see from Fig. 13. that, with the same sampling frequency, the standard deviation of sample speed error on freeway is lower than that on arterial. This is reasonable since the control delays caused by traffic signals make it more difficult to allocate travel times to road links.

Fig. 14. illustrates the MAD_k for each test dataset k. As the sampling interval increases from 10 seconds to 60 seconds, the mean absolute difference slightly increases from 3.9 km/h to 6.1 km/h for freeway and from 5.2 km/h to 8.0 km/h for arterial; when compare the MAD of 60-second and 90-second dataset, even though the sampling interval increases by 30 seconds, the MAD does not increase significantly for both freeway and arterial. However, as the sampling interval rises to 120 seconds, the MAD of both freeway and arterial datasets significantly increases. The result suggests that the proposed travel speed estimation method performs well for those GPS data with sampling interval shorter than 90 seconds. Besides, generally the proposed method performs better for probe samples on freeway than that on arterial, which, considering the difficulty of allocating stopping delay caused by signalized or un-signalized intersections, is reasonable.





(a) Arterial



(b) Freeway

Fig. 12. Distribution of Sample Speed Error

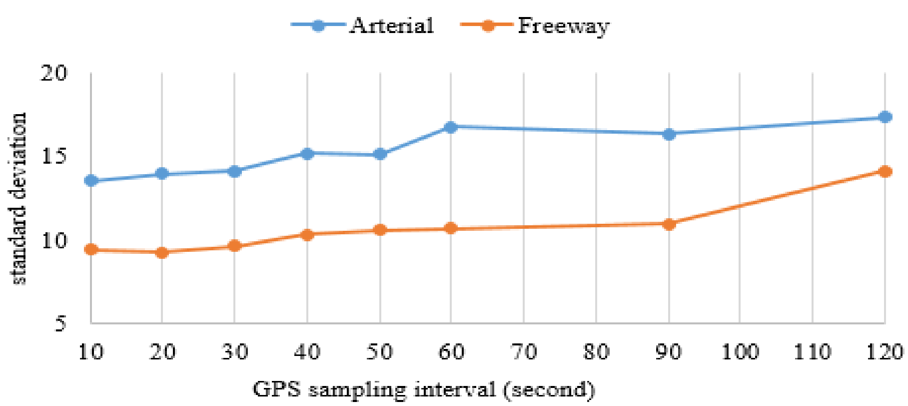


Fig. 13. Standard Deviation of Sample Speed Error with Different GPS Sampling Interval and Road Type

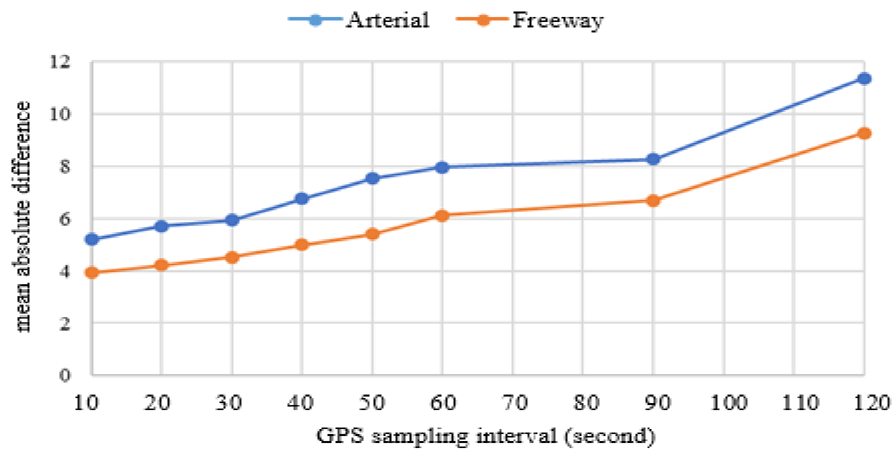


Fig. 14. Mean Absolute Speed Difference with Different GPS Sampling Interval

5.4 Comparison Results

In the following Fig. 15. to 5.10, the speed estimation comparisons between VDS and GPS data on Dec 5th, 6th, and 16th at road link 42896 on Whitemud Drive have been chosen as the example to show the estimation results. In this case, we have chosen the estimated speed between 6 am and 10 pm to conduct the comparison. As shown in the figures, blue solid lines represent the VDS speed estimation, red dot lines show the GPS speed estimation before conducting moving average procedure, and yellow solid lines show the GPS speed estimation after conducting moving average procedure.

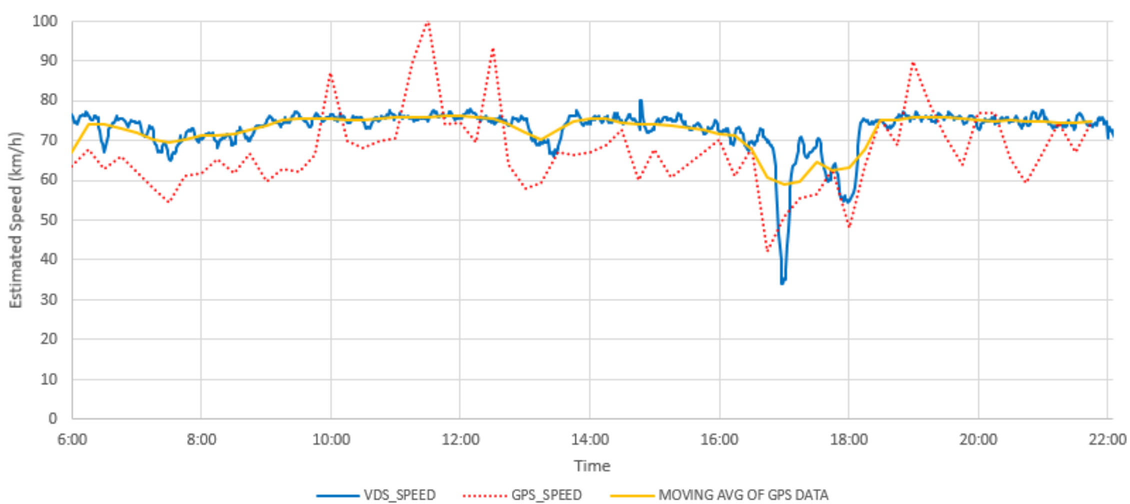


Fig. 15. Speed estimation comparison between VDS (1035) at road segment 42896 and GPS data on Dec. 5th

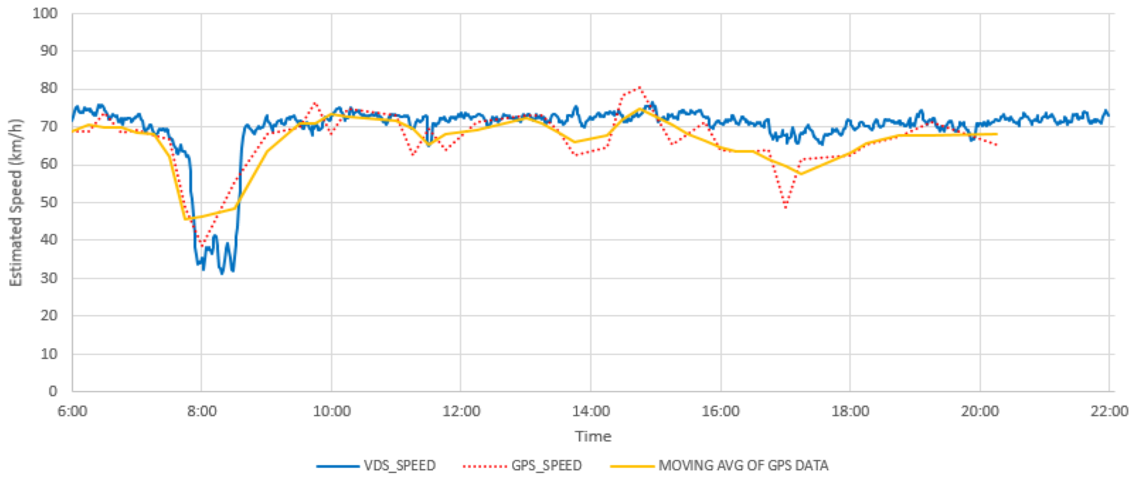


Fig. 16. Speed estimation comparison between VDS (1035) at road segment 42896 and GPS data on Dec. 6th

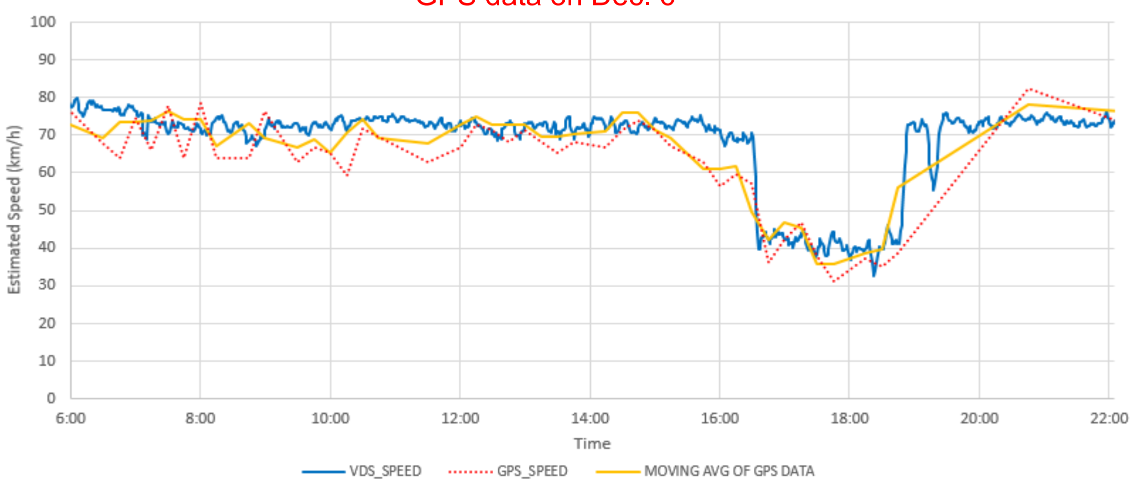


Fig. 17. Speed estimation comparison between VDS (1035) at road segment 42896 and GPS data on Dec. 16th

As the figures showed above, the GPS speed estimation can represent the general trend of speed profile comparing with the VDS data with some variations. For instance, during the morning peak hour on Dec 6th, the GPS speed estimation results did capture the dramatic speed drop caused by a temporary traffic flow disruption between 8 - 9 am. On the Dec 5th and 16th during the afternoon peak hour, the GPS speed can also capture the congestions, which is very common traffic patterns and main concerns for traffic operators as the afternoon peak congestion on this segment is the main cause of traffic delay on roadway, apart from incidents such as traffic accidents and road constructions. But still, there will be some discrepancies in speed estimation at a specific time due to some human factors and the specialty of the transit bus. But in general, this ETS bus can provide a valid estimation of the current traffic state.

Table 4. Speed estimation results for all the samples.

	Sample Size	**Mean Absolute Speed Difference (km/h)**	**Percentage Mean Absolute Speed Difference (%)**
Speed estimation difference between GPS and VDS data	2407	6.96	9.18%

Table 5. Speed estimation comparison between AM peak, PM peak, and off-peak hours.

	AM peak 7 – 10 am (km/h)	**PM peak 4 – 7 pm (km/h)**	**On Average (km/h)**
Speed estimation difference between GPS and VDS data	7.25	7.53	6.96

The Mean Absolute Speed Difference and Percentage Mean Absolute Speed Difference are shown in Table 4. In total, the sample size in these 10 selected dates are 2407, and on average there are 2.3 probe vehicles per speed estimation. The overall mean absolute speed difference is 6.96 km/h, and this difference will be higher in the AM and PM peaks, which are 7.25 km/h and 7.53 km/h respectively (Table 5). These results comply with the sample size requirement calculation above, and the error is within the acceptable range. The percentage mean absolute speed difference showed that the difference between GPS and VDS speed estimations is 9.18% on average according to equation (10). The results also illustrate that the speed difference between the VDS and GPS speed estimations are due to the random factors with no significant bias. It should be emphasized that in the real world, due to the location errors of GPS points, the required sample size in specific scenarios may be greater than what is applied in this study, and it depends on the accuracy of the GPS measurement. In collecting traffic information using probe vehicles, it is important to determine the number of probe vehicles because the traffic state can be affected by many variables. Sometimes the variance of traffic speeds within special time intervals is very large, and in this case, the sufficient sample size would be important to accurately estimate the current traffic state. Also, higher sample size means that the confidence level of estimation speed will be higher once the standard deviation of speed values and allowable errors are determined.

6. Conclusions

In conclusion, the study has proposed one method of conducting the mean speed estimation from the GPS traffic data. In the future, if more GPS data source and/or samples are available, the accuracy of estimation can be further improved. The key piece that this study tries to deliver is the provision of travel speed and congestion heat maps that can provide color coded average speed in each section of the road. This study is aiming to provide the algorithms for estimating travel speed from GPS data and to evaluate the accuracy of various GPS data sources. Here are some key findings from the paper:

- With respect to GPS sampling frequency, more accurate and reliable samples can generate the accurate speed estimation at a higher confidence level. Also, during the morning and afternoon peak hours, the speed estimation accuracy is lower than the speed estimation on the uninterrupted flows.
- In the current four data sources, the transit bus GPS data can provide the best mean speed estimation compared with the reference speed, while other three types of data neither provide enough sample size nor have enough temporal coverage to conduct the analysis.
- From the sample size requirement analysis, the ETS bus data can provide sufficient sampling frequencies for the allowable error of speed estimations less than 10 km/h. However, higher sample size is required for more accurate estimations.
- According to the mean absolute speed difference analysis, the mean speed estimation difference between ETS GPS data and VDS data is 6.96 km/h, and the speed estimation is significantly greater in the morning and afternoon peak hours.
- In the evaluation section, the Transit bus GPS successfully captures the congestion condition given limited sampling size, and the speed profile results fit well with the loop detector data on the Whitemud Drive.

ACKNOWLEDGEMENT

The authors would like to thank Alberta Transportation, Edmonton Transit and Office of Transportation Safety who provided data and information. This work is also partially supported by National Natural Science Foundation of China (Grant No. 61703236) and Jiangxi Provincial Major Science and Technology Project - 5G Research Project (Grant No. 20193ABC03A005)

REFERENCE

1. Lim, S., Lee, C. Data fusion algorithm improves travel time predictions. *IET Intelligent Transport Systems*, 2011, Vol.5, No.4, pp.302-309.
2. Coifman, B. Improved velocity estimation using single loop detectors. *Transportation Research Part A: Policy and Practice*, 2001, Vol.35, No.10, pp.863-880.
3. Hellinga, B.R. Improving freeway speed estimates from single-loop detectors. *Journal of Transportation Engineering*, 2002, Vol.128, No.01, pp.58-67.
4. Catani, L., Tremblay, J.M., Bassani, M., Cirillo, C. Methodology to backcalculate individual speed data originally aggregated by road detectors. *Transportation Research Record*, 2017, No. 2659, pp.1-14.
5. Zefreh, M.M., Torok, A. Single loop detector data validation and imputation of missing data. *Measurement,* 2018, Vol.116, pp.193–198.
6. Li, B., Yu, Z., Huang, L., Guo, B. Vehicle departure pattern and queue length prediction at an isolated intersection with automatic vehicle identity detection. *IET Intelligent Transport Systems,* 2019, Vol.13, No.12, pp.1804-1813.
7. Yuan, Y., Van Lint, J. W. C., Wilson, R. E., Van Wageningen-Kessels, F., Hoogendoorn, S. P. Real-time Lagrangian traffic state estimator for freeways. *IEEE Transactions on*

Intelligent Transportation Systems, 2012, Vol.13, No.01, pp.59-70.
8. Yao, E., Zheng, K., Zhang, J., Zhang, Y. Traffic flow estimation on the expressway network using toll ticket data. *IET Intelligent Transport Systems*, 2019, Vol.13, No.05, pp.886-895.
9. Schoepflin, T.N., Dailey, D.J. Dynamic camera calibration of roadside traffic management cameras for vehicle speed estimation. *IEEE Transactions on Intelligent Transportation Systems*, 2003, Vol.04, No.02, pp.90-98.
10. Abuella, H., Miramirkhani, F., Ekin, S., Uysal, M., Ahmed, S. ViLDAR—visible light sensing-based speed estimation using vehicle headlamps. *IEEE Transactions on Vehicular Technology*, 2019, Vol.68, No.11, pp.10406-10417.
11. Dahl, M., Javadi, S. Analytical modeling for a video-based vehicle speed measurement framework. *Sensors*, 2019, Vol.20, No.01, pp.160-170.
12. Jenelius, E., Koutsopoulos, H.N. Probe vehicle data sampled by time or space: consistent travel time allocation and estimation. *Transportation Research Part B: Methodological*, 2015, Vol.71, pp.120-137.
13. Liu, K., Yamamoto, T., Morikawa, T. Feasibility of using taxi dispatch system as probes for collecting traffic information. *Journal of Intelligent Transportation Systems*, 2009, Vol.13, No.01, pp.16-27.
14. Wang, X., Yin, D., Qiu, T.Z. Applicability Analysis of an Extended METANET Model in Traffic-State Prediction for Congested Freeway Corridors. *Journal of Transportation Engineering, Part A: Systems*, 2018, Vol. 144, No. 9, pp. 04018046.
15. Zhao, J., Gao, Y., Yang, Z., Li, J., Feng, Y., Qin, Z., Bai, Z. Truck traffic speed prediction under non-recurrent congestion: Based on optimized deep learning algorithms and GPS data. *IEEE Access*, 2019, Vol. 7, pp. 9116-9127.
16. Zhao, J., Gao, Y., Bai, Z., Wang, H., Lu, S. Traffic speed prediction under non-recurrent congestion: Based on LSTM method and BeiDou navigation satellite system data. *IEEE Intelligent Transportation Systems Magazine*, 2019, Vol. 11, No. 2, pp. 70-81.
17. Puangprakhon, P., Narupiti, S. Allocating travel times recorded from sparse GPS probe vehicles into individual road segments. *Transportation Research Procedia*, 2017, Vol.25, pp.2208-2221.
18. Zheng, F., Van Zuylen, H. Urban link travel time estimation based on sparse probe vehicle data. *Transportation Research Part C: Emerging Technologies*, 2013, Vol.31, pp.145-157.
19. Jenelius, E., Koutsopoulos, H.N. Travel time estimation for urban road networks using low frequency probe vehicle data. *Transportation Research Part B: Methodological*, 2013, Vol.53, pp.64-81.
20. Hellinga, B., Izadpanah, P., Takada, H., Fu, L. Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments. *Transportation Research Part C: Emerging Technologies*, 2008, Vol.16, No.06, pp.768-782.
21. Wang, X., Ge, Y., Niu, L., He, Y., Qiu, T.Z. Method for imputing missing data using online calibration for urban freeway control. *Transportation research record*, 2018, Vol. 2672, No. 43, pp. 44-54.
22. Zou, L., Xu, J.M. Zhu, L.X. September. Arterial speed studies with taxi equipped with global positioning receivers as probe vehicle. *In: 2005 International Conference on Wireless Communications, Networking and Mobile Computing*, 2005, Vol.02, pp.1343-1347.
23. Lorkowski, S., Mieth, P., Schäfer, R.P. New ITS applications for metropolitan areas based on floating car data. *In: ECTRI Young Researcher Seminar*, 2005, pp.1-10.