Final Reports & Technical Briefs from Mid-America Transportation Center

Mid-America Transportation Center

2021

IDENTIFYING ORIGIN-DESTINATION TRIPS FROM GPS DATA – APPLICATION IN TRAVEL TIME RELIABILITY OF DEDICATED TRUCKS

Li ZHAO

Ying LI

Li ZHAO, Ph.D.[1]
E-mail: lizhao@unl.edu
Ying LI, Ph.D.[2]
E-mail: yingli@chd.edu.cn
[1] Department of Civil Engineering
  University of Nebraska-Lincoln
  262K Prem S. Paul Research Center at Whittier School,
  Lincoln, Nebraska, 68583, USA
[2] School of Information Engineering
  Chang'an University
  Middle-section of Nan'er Huan Road, Xi'an, ShaanXi
  Province, 710064, China

IDENTIFYING ORIGIN-DESTINATION TRIPS FROM GPS DATA – APPLICATION IN TRAVEL TIME RELIABILITY OF DEDICATED TRUCKS

ABSTRACT

The advancement of data collection technologies has brought an upsurge in GPS applications. For example, travel behaviour research has benefited from the integration of multiple sources of Global Positioning System (GPS) data. However, the effective use of such data is still impeded by the challenge in data processing. For instance, GPS data, despite providing detailed spatial movement information, do not label the starting and finishing points of a trip, especially for commercial trucks. Hence, there is a critical need to develop a trip identification method to effectively use the trajectory data provided by GPS without additional information. This paper focused on identifying trips from the raw GPS data. Specifically, a systematic method is proposed to extract trips on the basis of origin-destination (OD) pairs by using a 5-step procedure. An application was provided on estimating the performance of travel time reliability using three metrics based on the OD trips for each dedicated truck. The application showed that, in general, trucks on long-distance routes have less reliable travel times compared to trucks on short-distance routes. This paper provides an example of using GPS data, without further information, to study travel time for freight performance and similar needs of punctuality in logistics.

1. INTRODUCTION

Freight trucks transport goods typically from and to manufacturing plants, retail stores, and distribution centres through localities, which play an important role in serving the needs of today's supply chain and contribute to urban safety, congestion, and pollution [1, 2]. Freight trucks can operate in many modes, such as dedicated, regional, and over-the-road. A dedicated truck is operated in a way that is designated to pick up a load at one regular location and then deliver to a dedicated customer (i.e., another regular location). Unlike many route options for trucks running in the urban area, a dedicated truck, especially running long-distance between cities, is committed to only a few routes and operates on a predictable schedule. It is conceivable that this type is the simplest and the most identifiable mode of truck freight operations.

In recent years, there have been increasing concerns regarding the timeliness of delivery from shippers, contract carriers, and trucking business owners. Therefore, travel time reliability becomes one of the key measures of freight truck performance along intercity or interregional highway networks [3]. However, knowing the truck travel time reliability requires fairly extensive data collection. In fact, most traditional data collection methods, such as established surveys and travel diaries, cannot provide adequate information about truck movements and travel time. An emerging method, using the Global Positioning System (GPS) to monitor and manage the performance of the truck fleet, has gained increasingly more attention because it can obtain large amounts of long-term data while retaining great details at a little cost [4].

GPS data have reasonable accuracy in improving the traditional manual data collection for travel time and trip data [5, 6]. A key issue of GPS data

is that it gives no explicit indication of when a trip begins or ends and is often mixed with noisy and invalid sensor data. Thus, to quantify the level of reliability or the extent of variability in travel time, data purification and trip acquisition techniques are required to turn the continuous GPS data into end-to-end trip data.

This study established a systematic procedure to purify and extract truck trip information from a massive GPS dataset, including target route determination, the trip ends (stops) identification, and trips chains split after GPS data cleaning. The truck trips are applicable to various traffic studies and fleet management. The methodology proposed in this paper contributes to both the existing body of literature and the practices of the traffic management on performance measures of long-haul dedicated freight trucks and other transport modes in need of punctuality.

The paper is organised as follows. Section 2 reviews the relevant studies in the literature. Section 3 defines the scope of the study, which is the trip for a dedicated truck. Section 4 and section 5 are the key parts that present the data and methodology in this study. As an application, section 6 showcases the truck performance measure in travel time reliability. Finally, section 7 concludes the study with contributions, limitations, and future work.

2. LITERATURE REVIEW

Great effort has been made to identify the trip ends automatically from tabular GPS data. In previous studies, the most common way to break the continuous GPS data into trip-based data segments is to search stop dwells [7]. Depending on different GPS settings, a valid stop dwell can be defined by 1) speeds that are close to zero (e.g. less than 5 km/h) over a predefined time in the ongoing GPS recording; or 2) a gap between two consecutive timestamps greater than a predefined time due to signal loss or device outage [8]. Either way, it is critical to determine the dwell time threshold as a gap between any two trips embedded in the GPS data.

The selection of the threshold value for the dwell time that works in all cases is not easy [9]. Literature can be divided into those thresholds for intermediate stops and those thresholds for destination stops. To date, there is no study on how to choose the dwell time threshold to determine the end of a trip. In truck GPS studies, various thresholds for intermediate stops have been reported such as 3 minutes [10, 11],

4 minutes [12], and 5 minutes [13], as well as for destination stops such as 15 minutes [13], 30 minutes [14], and 120 minutes [8]. These studies are focused on distinguishing the intermediate stops from the destination stops [8, 9].

There are also studies that have attempted to combine additional techniques to improve the above dwell time threshold algorithm. For example, it can be upgraded to more complicated multi-stage procedures, supplemented by thresholds corresponding to multiple factors such as spot speed [13], a truck heading angle [14], and trip length [9]. The use of other techniques such as Geographic Information System (GIS) [15] or map matching [16] also found effective improvement. For example, a typical way is to combine GPS data with GIS techniques to provide locations of freight facilities [17, 18]. Recent practices have extended to the use of machine learning methods such as support vector machine [19] and entropy estimation [20]. Regardless, stop dwell time is the predominant determinant in trip extraction from the GPS data [4].

It should be noted that the more important goal of most GPS data studies is to apply the after-processing data to various traffic studies. A significant application of using GPS trajectory is on travel time and travel time reliability studies [10, 21–24]. For example, based on GPS trip data, McCormack and Hallenbeck [10, 21] explored more robust travel time benchmarks and performance statistics related to the reliability of truck trips. Liao [22] derived corridor travel time from GPS data for truck and passenger car travel respectively and found that trucks on the U.S. and interstate highways have about an average of 10% longer travel time than cars. With trip data, these travel time reliability analyses can be carried out at different levels, such as segment or link [17], corridor [22], route [23], and network [24].

As it can be found, the studies based on truck GPS data in the literature are trucks running on a given facility (i.e., segment-based) in a given time frame (i.e., 15-minute aggregates). Unlike the rest of the literature, the uniqueness of this study is that the truck performance on travel time reliability is based on individual trucks over multiple days. In other words, no studies have been found to study a dedicated truck running on a route repeatedly over a period of time.

To achieve this goal, first, the GPS data are filtered and split into trips on the basis of origin-destination (OD) pairs. This part is critical because it provides a detailed and accurate data foundation for

any performance studies. Then, the central tendency (i.e., mean), dispersion (i.e., standard deviation), and position (i.e., 50 and 95 percentiles) of travel times are measured for each dedicated truck. From these measures, the travel time reliability is calculated using different performance metrics such as buffer index, percent variation, and probabilistic measures. To assist the performance comparison of travel time reliability among different trucks on different lengths of routes, a voting score system is adopted so that it integrates different reliability metrics.

With the use of multiday GPS data, this study also provides opportunities for investigating the regularity and variability for long-haul freight truck trips. Other potential applications for splitting truck GPS trips using the method proposed in this paper are, for example, truck travel demand forecasting [25, 26], freight fleeting performances [27], fuel consumption and emission estimations [28], roadway bottlenecks identification [22], and so forth. For the logistics and trucking industry, it would be interesting to know the punctuality of a truck on a route to ensure timely delivery and performance stability [29], which will be showcased later in this paper.

3. DEFINITION OF TRIPS FOR A DEDICATED TRUCK

In this study, a trip is defined as the total movements from a truck's origin to its destination (i.e., OD trip). In GPS data, a trip is simply a set of temporally ordered, timestamped data points that carry geospatial, operational, and other information. The origin can be regarded as the place where a truck stops most often as well as takes the longest rest. The destination is the place where a truck serves its customer (e.g., delivers goods). Thus, each trip either starts from the origin and heads to its destination (O-to-D) or starts from the destination and heads to its origin (D-to-O).

In practice, the origin and destination are like two local zones, as the exact starting and ending locations of each trip may be slightly different. This is because when a truck arrives at the local area, (1) several stops may be needed for distributing goods to more than one place; (2) multiple links may be available for the truck driver from or to the final place; and (3) some activities may be conducted before heading back, which may result in different parking locations. All these reasons may cause the exact trip ends to be unfixed, and the time spent in the local zones is hard to predict. Because of that, the longest common segment is trimmed to represent the OD route when studying the trips from a dedicated truck. As illustrated in *Figure 1a*, the longest common segment for a dedicated truck can be derived when all trip trajectories are lying on the target route.

For a long-haul truck, in addition to the origin and destination stops, a trip may be involved with several intermediate stops for rest, gas, or other purposes. As two trips illustrated in *Figure 1b*, trip 1 passes through O (origin)$\rightarrow S_A \rightarrow$D (destination), where S_A is the intermediate stop during trip 1, and it breaks trip 1 into two trajectories: Tr_1 and Tr_2. Similarly, trip 2 passes through D (destination) $\rightarrow S_B \rightarrow S_C \rightarrow$O (origin), where S_B and S_C are the intermediate stops during trip 2, and they break trip 2 into three trajectories: Tr_3, Tr_4, and Tr_5. These intermediate stops may vary from several minutes to several hours in time.

In summary, for a dedicated truck running on a route where the origin and destination are designated, the OD trips can be characterised as (1) a trip chain which may compose N stops, including origin and destination, and N-1 chain links; (2) a reciprocating trip which moves back-and-forth
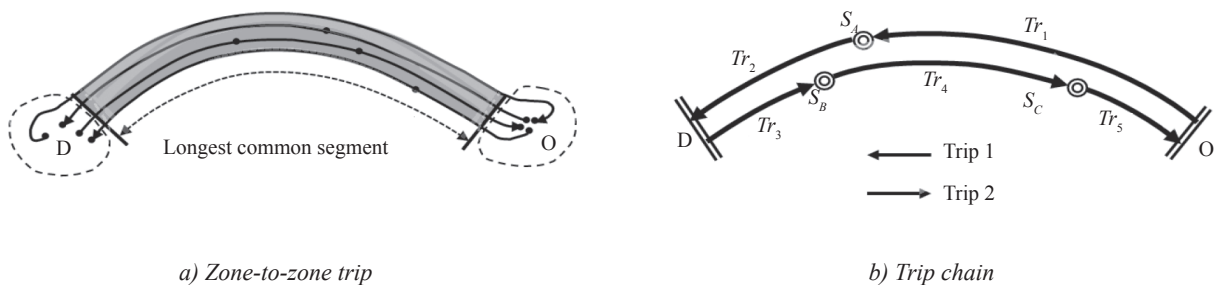

a) Zone-to-zone trip

b) Trip chain

Figure 1 – Illustration of trips between origin and destination on one route (drawing not to scale)

repeatedly. It may be hypothesised that the expected on-road travel time would be very close among trips of each truck running on the same route. Since variations would still exist given that traffic conditions may be different for each trip, this study calculates reliability in travel time to indicate the performance of a dedicated truck in this paper.

4. DATA PREPARATION

The GPS truck data is from the Chinese road freight monitoring and service platform. The platform was established in 2014 and is the only national-level monitoring platform for commercial trucks (e.g. heavy trucks and semi-trailer tractors over 12 tons). All trucks registered on the platform are equipped with a GPS before entering service. The GPS is out of reach of the driver and records data on a basis of a regular time interval regardless of whether it is moving, idle, or parked. The truck transmits a GPS signal with a cellular connection on a platform request basis.

4.1 Overview of the GPS dataset

The GPS dataset used in this paper includes 100 trucks collected on from 1 April to 30 April 2016 that followed their own routes (different trucks run different routes) across China. Each row of the GPS log, representing a data point, includes a uniquely identified number (has been scrambled for anonymity), spatial (i.e., latitude and longitude), temporal (i.e., date and time), instantaneous speed, mileage, and head angle information. It should also be noted that the platform does not contain other useful information (e.g. speed limit, driver information, and traffic conditions), and this information could not be obtained from other sources.

Although most of the trucks take 20 seconds and 30 seconds as the main intervals, the data recording interval varies among trucks. Within one truck, the data interval may also change, possibly due to errors or instabilities of GPS signal reception. For example, truck number 90 has a 20-second data interval taking 89.3 percent of its total data (the highest frequency) and a 120-second interval taking 7.4 percent (the second-highest frequency). For the truck number 54, the most frequent data interval is 30 seconds, which takes 86.1 percent of its total data, and the second most frequent data interval is 180 seconds, which takes 9.3 percent. In the dataset, the majority of the relatively long intervals, e.g. 120 seconds and 180 seconds, respectively, for these two trucks occur during stops (i.e., when speed equals zero). It should be noted that, although different data intervals exist in the dataset, they barely impact the study when compared with the length of truck trips in hours.

4.2 Determining the target route

In this paper, we focused on the target routes, on which the truck transports goods back and forth between two locations on a regular basis. The specific geospatial information of the target routes is estimated from the dataset using a trajectory similarity measure. Interested readers can refer to the previous study [29]. As an example, *Figures 2a and 2b* plot a truck route on a geo-coordinate plane. Colours represent the speed. In total, 15 out of the 100 trucks whose target routes are significant are selected from the dataset.
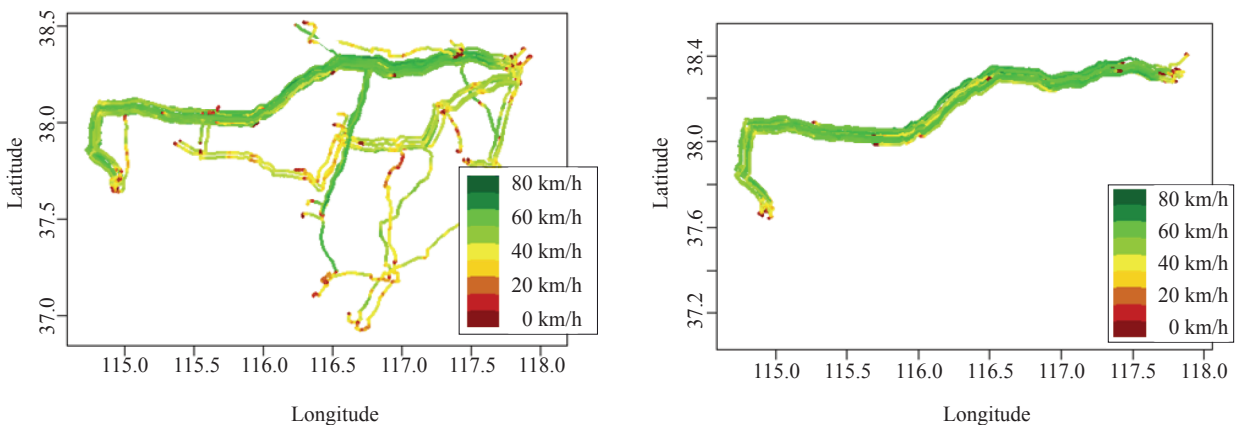

Figure 2 – Target route determination. A slight deviation is added to the coordinates to display multiple trips.

5. METHODOLOGY

The core of this study is to establish a systematic approach to extract purified trips of a dedicated truck from the raw GPS dataset. The methodology includes different types of stops identified in a trip and a generic step-by-step procedure to determine an OD trip.

5.1 Identify the trip stops

A stop can be considered a trip end (i.e., arriving at the destination) or a trip interruption for certain purposes such as gas-filling or dining. In those cases, turning off the engine is necessary to define the truck stop. Therefore, a temporary pause caused by traffic control (e.g. traffic signal) is not considered as a stop; rather, it is defined as a part of the trip. In the GPS dataset of this study, a stop can be identified by a speed less than 5 km/h in any length of dwell time. Though engine shutdown cannot be flagged in the dataset as the GPS device records data continuously, it can be reflected by the purpose and the duration of a stop. According to the functionality, a stop can be divided into either an OD stop (stop between two trips) or an intermediate stop (stop within a trip). This will be described below.

OD stops

As a truck runs in the local area of the origin or destination, its route profile is characterised by relatively low speed, frequent speed changes, frequent turning movements, frequent and longer stop times, and so forth. As an example, in *Figure 3a*, round-trip data is mapped between a local area in Nanyang and a local area in Binzhou, China. For the purpose of illustration, the truck speed is roughly indicated by three colours: green is the speed greater than 60 km/h and red is the speed less than 20 km/h. Intuitively, one can see that the OD stops, as defined in this paper, occur at the two ends of the route.

Technically, the amount of time for a stop (i.e., scale) and the number of stops (i.e., frequency) at a particular place are measured to determine the OD stops. As an example, *Figure 3b* shows all the stops
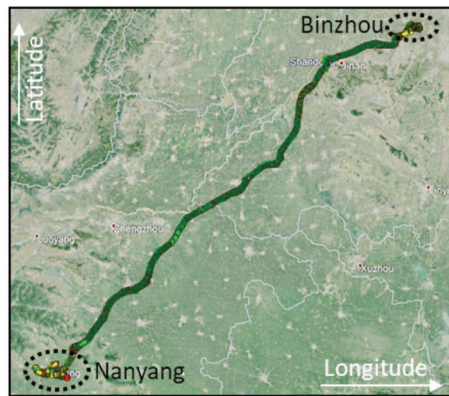
a) A route between Nanyang city and Binzhou city
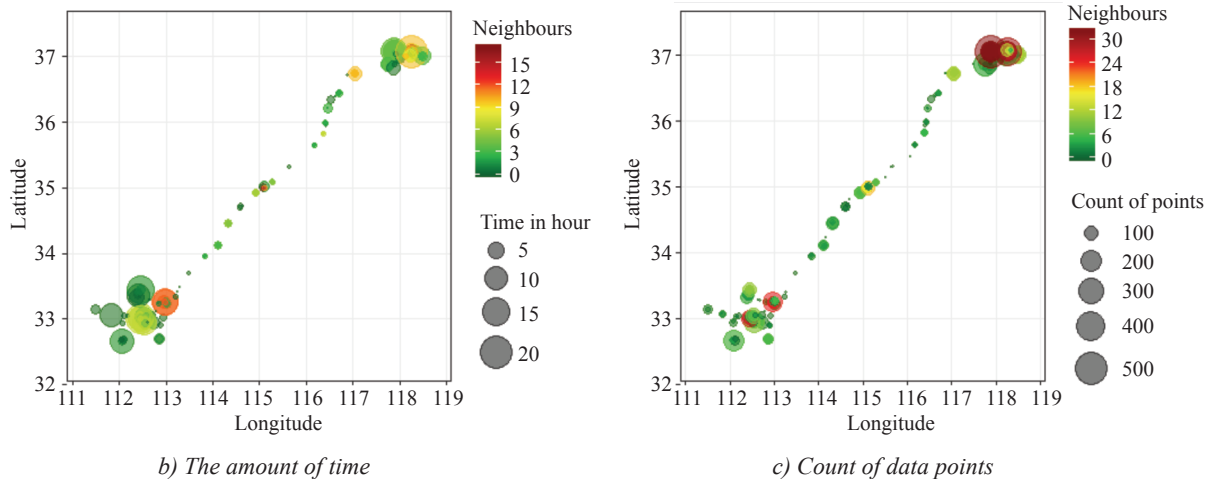
b) The amount of time

c) Count of data points

Figure 3 – Identification of the trip ends of the origin and destination

that occurred in the trips over the data collection month. A circle is plotted on this figure to indicate the latitude and longitude of a stop. The size of the circle indicates the scale of the time length, and the colour of the circle indicates the frequency of a stop. Trucks with the same destination do not necessarily stop at exactly the same location. However, they may park in close proximity [30]. Therefore, the hierarchical cluster analysis, recommended by [11], was adopted with a cluster of stop frequency in the range of 1 km. As may be seen in *Figure 3b*, the larger and redder circles are OD stop candidates and are classified into two groups. The largest and reddest circle in each group is identified as the OD stop.

In addition to the amount of time, this study also examined the density of data points (i.e., GPS pins) considering varying data intervals. As the results show in *Figure 3c*, the count of data points in each stop is located on a longitude-latitude plane. Results of this particular example show that the OD stops obtained using data points are geospatially the same as using the amount of time; however, it is more prominent to determine the OD stops from all those stop candidates. The two methods cluster the density of the stops at a location using different perspectives of data. In this study, both methods are used to inspect the OD stop because there are varying data intervals and missing data in the dataset, albeit in small quantities. It is worth noting that the determination of the OD stops is the core of the trip split method proposed, which does not rely on fixed duration thresholds but an algorithm of the density of the GPS pins or stop durations.

Intermediate stops

There may be a lot of reasons for a driver to take an intermediate stop during a trip and these may be grouped into an intended stop and an unintended stop. Intended stops include dining, resting, or overnight sleeping, which can be planned ahead of time. Unintended stops, such as filling gas and waiting at the toll gate, are those which the driver has little to no control over.

Figure 4 shows two route sub-sections from the route in *Figure 3*. As examples, the stops are identified and mapped to the function of the venue manually. Each arrow matches a data point and shows the travel direction. Stops, where speeds are continuously lower than 5 km/h for more than 1 minute, are circled in *Figure 4* and their time dwells are measured. In *Figure 4a*, the stop at a service area in the trip of the northeast direction took 15.2 minutes. In *Figure 4b*, the stops at a toll station in the southwest and northeast directions took 2.9 minutes and 3.6 minutes, respectively.

As mentioned before, a truck's temporary pause such as waiting for the red light at signalised intersections or being stuck in a traffic jam is treated as a part of the trip and is thus not included in the intermediate stops. This is because the temporary pause (1) represents the running traffic conditions along the route and will be interpreted together in the travel time; (2) is uncontrollable, hard to measure accurately given limited information; and (3) rarely occurs on high-speed corridors of the study segments (e.g. intercity expressway or regional highway). This outlines differences in the intermediate stop definition as compared to the literature [19, 20].
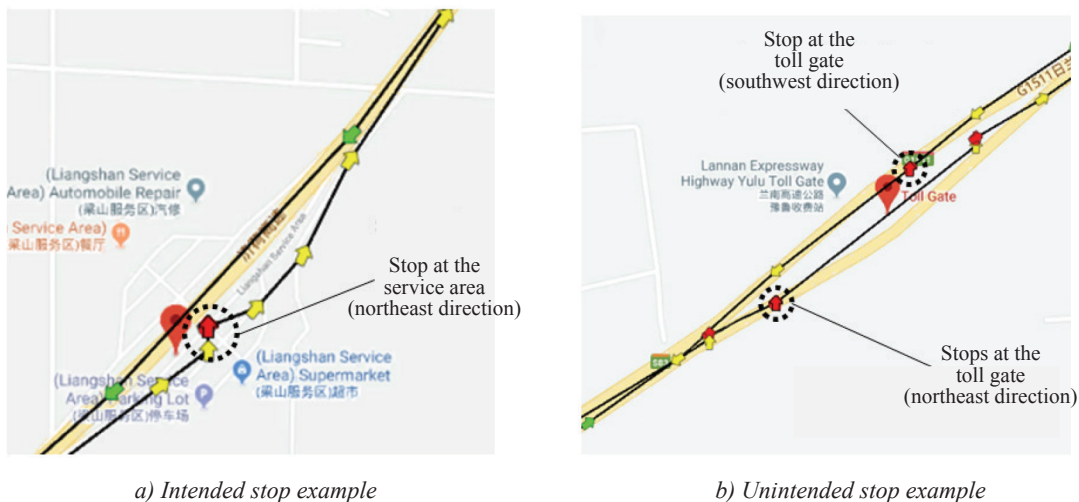

a) Intended stop example

b) Unintended stop example

Figure 4 – Identifying intermediate stops on Google Maps

A manual check of a larger sample data was conducted to locate the latitudes and longitudes of the stops on Google Maps to explore the purpose of a stop. It was found that the average stop time is 20.9 minutes (sample size=122) in the service area and 3.3 minutes (sample size=186) at the toll stations. Both are significantly longer (p<0.001) than a temporary pause at signalised intersections, which takes an average of 0.6 minutes (sample size = 683). Although the intermediate stops are separated as intended and unintended stops, this study does not distinguish them in travel time estimates in the subsequent study. It should be noted that errors may still exist as the method cannot completely separate long pauses from intermediate stops.

5.2 OD trip determination

Intermediate stops break a complete OD trip into a trip chain. The following 5-step procedure is used to purify and convert the raw GPS data into OD segments and eventually OD trips. Results of each step from example truck data are shown in *Figure 5*.

Step 1. Preliminary error filtering

GPS data may be contaminated by errors involving malfunctioning or poor performance (e.g. signal blockage or signal loss) of the sensor device. The main errors with this dataset include: (1) Duplications. In the raw data, two or more consecutive records may have the same timestamp. It can be fixed by keeping the first record of the same data while removing the rest of the duplicate data. (2) Outliers. In the raw data, incorrect values may be evident. For example, it is unreasonable that speed is greater than 150 km/h and that the distance between two consecutive data points is greater than 1.25 km, given a 30-second data interval. In this case, the outlier is replaced with the average before and after the outlier once it is recognised. (3) Missing data. This refers to the timestamp interval between two consecutive data points longer than 180 seconds, which is found as the longest interval in the dataset. Manual remediation is required only when missing data occur at the beginning or end of a "chain" in the trip, which may consider mileage, time, speed, and angle of ten data points before and after the missing gap.

a) Preliminary error filtering

b) LMS smoothing

c) Trip discretization. Note that every small column is a chain in a trip. The data stream is disconnected between any two chains.

d) OD trip assembling. Note that each window represents a "band-pass filter" that delimit the OD trip.

e) Study segment determination. Note that each window represents a "band-pass filter" that delimits the longest common segment.
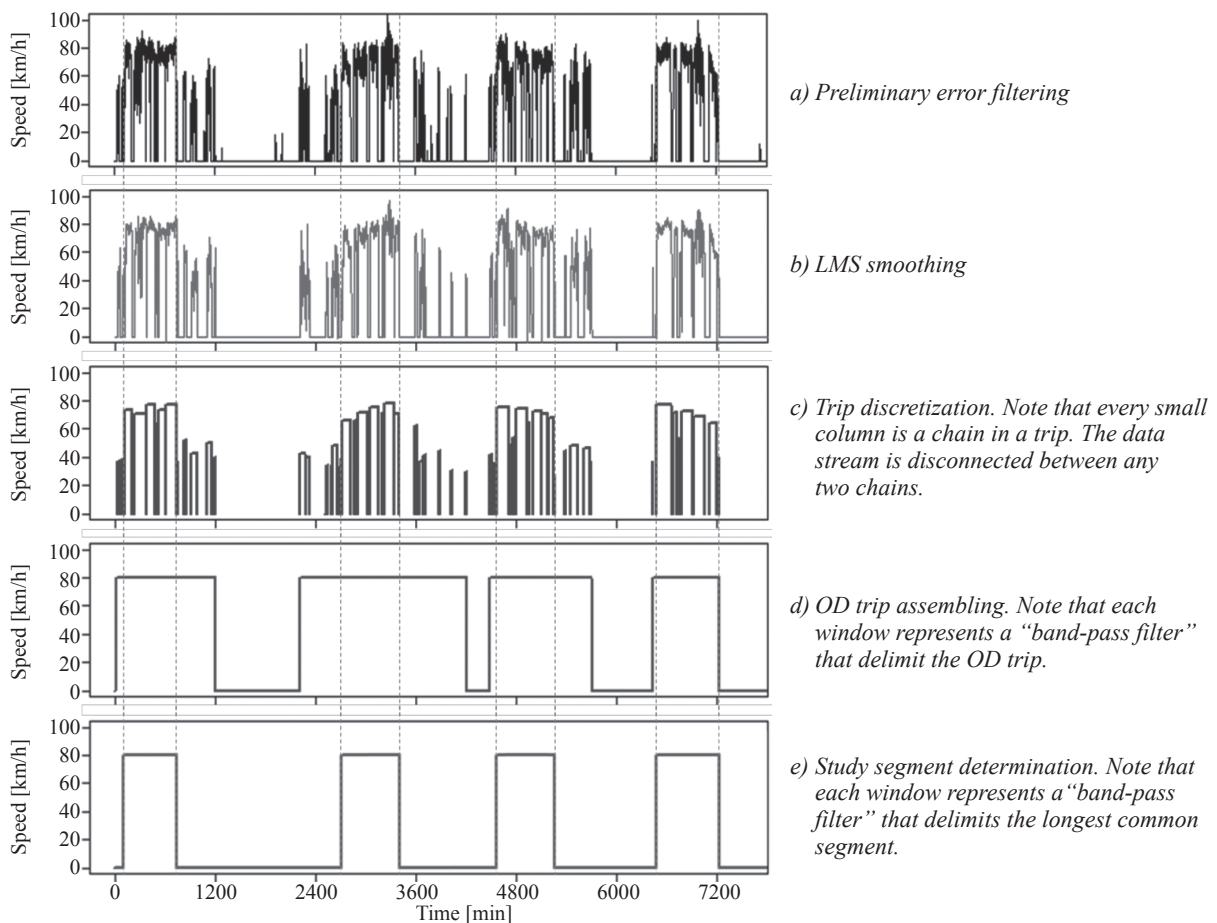
Figure 5 – An example of trip purification and study segment determination results

Step 2. Least Median Square (LMS) smoothing

In some cases, although some data points are suspicious, they might still be correct, leaving a margin of uncertainty. Therefore, this step further filters and smooths the data, as a complement to Step 1. For example, a sequence of consecutive speeds is (0, 0, 0, 0, 0, 0, 100, 0, 0, 0, 0) km/h, in a perspective of data smoothing, the 100 km/h speed is likely an outlier and thus will be replaced. LMS is found suitable for fine filtering and smoothing in this step. It can extract the real trajectories from time series with possible random variations and data shifts [31]. Specifically, in this study, a robust LMS filter package in R statistic software is applied. The width of the moving window filter is set to 180 seconds for the same reason as above.

Step 3. Trip discretisation

In this step, the data is further modified using signal concepts specifically from electrical engineering. At this point, the time-sequenced GPS data can be considered as an analogue signal. The continuous analogue signals are converted into discontinuous digital signals. The benefit of doing so is to separate short stop-and-go movements and amplify signal (i.e., speed) strength at the beginning and end of each chain of the trip. This step produces a trip chain in *Figure 5c*. The dataset is disconnected among the chains, and the amplitude reflects the median of each chain.

Step 4. OD trip assembling

The idea of this step is to flag the OD stops in the dataset using latitude and longitude as discussed previously. Correspondingly, the time frames of the OD stops are flagged, and the chains between the end of one stop and the beginning of the next stop consist of an OD trip. In *Figure 5d*, this step constructs "band-pass" filters to delimit each OD trip from the dataset.

Step 5. Study segment determination

The local OD zones are excluded from the OD trips in this step to determine the study segment, which represents the longest common segment where all trips overlap. The cut-off points of the study segment are decided: (1) after the truck exits a high-speed roadway, (2) before entering the local roads, and (3) as close to the OD zones as possible. The geospatial trajectories of all trips and the speed-time profile for each trip are used to determine the cut-off points from the OD trip. Trip numbers are marked in the dataset only for those data points within the cut-off points, indicated by the "band-pass" filter of the study segment in *Figure 5e*.

The 5-step procedure is programmed in R statistic software so that trips from all the 15 trucks with the 15 target routes are run. The trips determined using the above procedure will serve as the database for the truck performance study next.

6. TRUCK TRAVEL TIME RELIABILITY

The uniqueness of this study is that travel time is based on the performance of individual trucks over multiple days. This may be in contrast to the segment-based and 15-minute aggregates travel time in the Highway Capacity Manual or other reports in the literature [32]. Therefore, travel time reliability in this paper is defined as the consistency or dependability in travel times, as measured from a vehicle's repeatedly running on the same route over a period of time. The results of the travel time for each truck on each route direction are shown in *Table 1*. Note that the OD trip time refers to the time truck used to complete the OD trip, which includes time for the intermediate stops. Travel time is calculated by subtracting the total stop time from the OD trip time. As expected in the data, the OD trip time increases as the travel time increases.

It should be noted that the expected running time is used as a time benchmark. This metric is obtained by plugging the latitudes and longitudes of a route into Google Maps. Basically, the running time is estimated under free-flow conditions using the speed limits along the route. This amount of running time is almost the same for both directions and does not change over time when checking China's routes offline.

It was found in *Table 1* that some trucks, with a p-value<0.05, have significant differences in the average of the total travel time between the two trip directions. Thus, the travel times of the two directions were was examined separately.

In comparison to the estimated running time obtained from Google Maps on each route (see *Table 1*), the travel time for some trucks calculated in this study is dramatically longer. *Figure 6* plots the travel time of each truck from (1) Google Maps estimates, shown in white rectangles, (2) the calculated O-to-D direction, shown in pink rectangles, and (3) the calculated D-to-O direction, shown in blue

Table 1 – Summary of travel time statistics for each dedicated truc

Truck ID	City A (O/D)	City B (D/O)	Total OD length [km]	Study segment length [km]	Segment to OD ratio	Expected running time [min]	Trip direction*	No of trips	OD trip time [min] avg.	OD trip time [min] std.	Travel time [min] avg.	Travel time [min] std.	50th quantile [min]	95th quantile [min]	P-value of the directional travel time**
1	Ru Zhou	Zheng Zhou	121	121	100%	152	1	38	168.1	52.8	140.7	17.4	135.7	171.2	0.0625
							2	38	164.2	25.2	149.2	17.2	148.2	177.9	
11	De Zhou	Cang Zhou	336	336	100%	202	1	26	296.2	38.6	221.9	28.3	269.3	249.2	0.3107
							2	27	283.3	31.1	208.6	22.1	211.7	222.6	
15	Tang Shan	Tang Shan	79	74	94%	93	1	66	149.2	61.5	106.3	12.7	103.9	132.2	0.4798
							2	67	145.4	50.7	108.6	18.2	104.9	152.5	
48	Bao Ding	Cang Zhou	401	401	100%	282	1	27	649.9	140.0	473.0	47.2	470.0	529.0	0.0321
							2	28	729.5	272.9	446.7	41.0	449.5	519.1	
50	Han Dan	Tang Shan	745	708	95%	502	1	9	1219.2	785.3	600.1	10.5	602.0	627.0	0.6675
							2	10	1589.5	672.4	598.8	17.0	598.0	624.6	
52	Pu Yang	Ji Ning	82	64	78%	108	1	55	86.5	72.5	75.6	6.5	74.7	88.1	0.0034
							2	55	53.0	26.4	71.6	5.0	71.1	80.3	
54	Nan Yang	Bing Zhou	819	798	97%	577	1	9	712.7	78.3	570.4	52.0	581.5	614.3	0.1538
							2	10	733.9	114.5	617.8	83.5	639.7	699.8	
57	Su Qian	Bao Ding	877	763	87%	604	1	10	1050.0	422.3	731.2	21.3	734.5	755.5	0.3451
							2	10	1662.5	508.1	716.5	42.2	707.8	778.9	
67	Tian Jin	Han Dan	608	541	89%	483	1	12	1536.9	358.3	771.1	61.9	760.5	866.8	0.0007
							2	12	1727.1	558.6	952.0	135.4	916.6	1176.0	
70	Zheng Zhou	He Bi	173	173	100%	150	1	14	507.3	145.8	237.1	16.7	233.5	263.9	0.0001
							2	14	1097.9	327.4	314.2	50.7	311.8	386.7	
87	Tang Shan	Tang Shan	79	74	94%	92	1	65	167.8	71.5	101.2	8.8	101.0	116.9	0.4042
							2	65	147.9	64.6	99.6	9.1	99.0	115.9	
88	Xi An	Yan An	320	319	100%	237	1	30	271.9	41.7	229.9	39.7	237.8	274.4	0.2461
							2	31	241.1	49.0	216.8	47.4	224.5	258.3	
90	Beng Bu	Wu Xi	402	402	100%	266	1	29	335.0	17.9	333.9	17.4	341.1	347.6	0.0107
							2	29	363.1	38.2	345.6	16.3	342.1	382.0	
91	Jin Cheng	Yu Lin	734	668	91%	626	1	9	1313.2	214.1	877.1	36.5	869.1	926.8	0.0000
							2	9	1383.8	701.6	652.3	29.4	640.7	697.4	
92	Dong Guan	Shan Wei	233	230	99%	186	1	30	182.6	30.1	180.6	30.5	193.0	217.1	0.5165
							2	30	212.1	62.0	187.0	43.6	198.8	243.2	

* 1= O-to-D ; 2 = D-to-O ; For a truck, the longer and more frequent stay of the two OD stops is determined as the origin, and the other is the destination

** H0: no difference in the mean travel time exists between two trip directions at a significant level of 95% using t-test
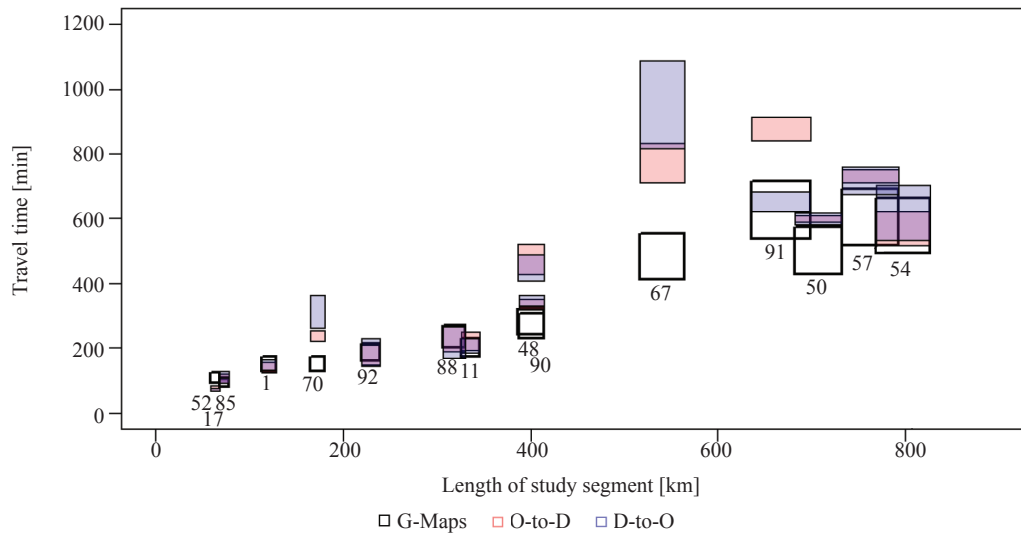
Figure 6 – The travel times obtained from the Google Maps (benchmark) and two travel directions between the origin and the destination, labelled with the truck ID

rectangles. The size of each rectangle is determined by the standard deviation of the average travel time that is located at the centre of the rectangle.

As can be seen in *Figure 6*, generally the longer the study segment length, the greater the difference (i.e., the pink or blue rectangles are far apart from black squares). This may indicate that in the actual truck operation a lot of factors increase the travel time, and the expected time provided by offline Google Maps is not feasible, especially for long-distance trucks.

The truck travel time on the same route may vary from trip to trip due to several factors, such as traffic environment fluctuations and truck operation inconsistencies. For example, temporary traffic events (e.g. accidents, harsh weather, work zone) that cause congestions may dramatically elongate the expected travel time. It should be acknowledged that the study may underestimate the amount of lost time from the intermediate stops. This is because the lost time is measured when speed is below 5 km/h. It did not consider the transition delays when the truck is preparing to stop, i.e., deceleration delay or returning from the stop, i.e., acceleration delay.

6.1 Reliability performance metrics

To measure the truck travel time reliability performance, this paper adopts three commonly used metrics as follows.

Buffer Index (BI). BI represents the extra time (or time cushion) that travellers must add to their average travel time when planning trips to ensure on-time arrival [33]. This study used the median in-

stead of the average travel time in the calculation as it was found to be beneficial [34]. Therefore, BI is calculated by the 95th travel time minus the median travel time and divided by the median travel time. The smaller the BI value, the more reliable the travel time.

Percent Variation (PV). PV is essentially the normalised standard deviation [3, 35]. It is useful when comparing the degree of variation among different routes. The PV is calculated as the ratio of the standard deviation to the mean of travel time and is expressed as a percentage. PV has the same characteristics as the coefficient of variation, thus mathematically a good proxy for a number of other common reliability measures [34]. A smaller PV value indicates a relatively less dispersed distribution and a more reliable travel time.

Probabilistic Measures (PM). PM calculates the probability that the observed travel time is greater than a specified time threshold. This metric pays more attention to the longer travel times (i.e., the right end of the distribution). The probability of travel time is 20 percent greater than the median travel time (i.e., $1.2 \cdot 50^{th}$ percentile) is recommended [36]. The smaller the PM value (close to zero), the more reliable the travel time.

In short, the above three measures represent three categories of indexes to quantify the travel time reliability in literature [37]: the buffer time index (i.e., BI), the statistical index (i.e., PV), and the probability index (i.e., PM). Specifically, BI accounts for unexpected delays and uses the 95th percentile to represent a near-worst case of travel

time for a specific truck route. PV focuses on the distribution of travel time and compares the variation among different routes. PM focuses on the likelihood of a travel time is occurred. The three travel time reliability metrics have their emphases, which are used to characterise the travel time reliability from different perspectives.

6.2 Truck performance results

Using the three metrics selected in this study, the travel time reliability is calculated on the route of each dedicated truck. The values calculated in each metric are ranked among the 15 trucks. The lowest value in each metric is ranked 1 as the best and the

highest value is ranked 15 as the worst. Discrepancies among different reliability measures may lead to different rankings given the same underlying data. Therefore, this study adopted a voting score system to integrate the three metrics.

As can be seen in *Table 2*, the travel time reliability rankings using different metrics do not seem consistent. This may not be unexpected as each metric attempts to measure different aspects of the reliability [38, 39]. The voting score in the last column of *Table 2*, which was calculated by adding the rankings of the three metrics (i.e., columns 6 to 8) in both directions for each truck. The overall scores can assist in finding out which truck performs better than

Table 2 – Truck reliability ranking with different index measures

Truck ID	Trip direction	Reliability measure			Reliability ranking (best = 1)			Voting score
		BI	PV	PM	BI	PV	PM	
1	1	0.272	12.3%	0.133	13	11	15	67
	2	0.201	11.5%	0.067	10	8	10	
11	1	0.104	9.5%	0.034	6	9	9	42
	2	0.146	9.1%	0.000	6	5	7	
15	1	0.262	8.0%	0.014	14	6	1	58
	2	0.240	16.7%	0.152	13	13	14	
48	1	0.126	10.0%	0.037	8	10	10	51
	2	0.155	9.2%	0.036	8	7	8	
50	1	0.008	1.1%	0.000	1	1	1	6 (best)
	2	0.016	1.2%	0.000	1	1	1	
52	1	0.298	14.2%	0.077	15	13	13	73
	2	0.223	12.0%	0.077	11	9	12	
54	1	0.056	9.1%	0.000	4	8	1	27
	2	0.094	13.5%	0.000	3	10	1	
57	1	0.029	2.9%	0.000	3	2	1	15
	2	0.100	5.9%	0.000	4	4	1	
67	1	0.140	11.9%	0.083	10	12	14	76 (worst)
	2	0.283	14.2%	0.167	14	11	15	
70	1	0.130	7.0%	0.067	9	5	12	66
	2	0.454	16.1%	0.143	15	12	13	
87	1	0.158	8.7%	0.043	12	7	11	54
	2	0.171	9.1%	0.043	9	6	9	
88	1	0.154	17.3%	0.033	11	15	8	56
	2	0.151	21.9%	0.000	7	14	1	
90	1	0.019	5.2%	0.000	2	4	1	16
	2	0.117	4.7%	0.000	5	3	1	
91	1	0.125	16.9%	0.000	7	14	1	59
	2	0.223	23.3%	0.000	11	15	1	
92	1	0.066	4.2%	0.000	5	3	1	14
	2	0.089	4.5%	0.067	2	2	11	

others regarding travel time reliability, given there might be a need for comparing different trucks on different routes.

In *Table 2*, truck number 50 has the most reliable travel time performance overall. On the contrary, truck number 67 performs comparatively the worst. It may be found that, in general, trucks on a long-distance route (e.g. truck numbers 52, 15, 87, and 1) have a relatively smaller travel time reliability (i.e., wider dispersion to its mean) compared to the short-distance route (e.g. truck number 54, 57, 91, and 67). This may be because the actual operations of long-haul trucks are more than complicated, which may include unpredictable delays from various aspects.

7. CONCLUDING REMARKS

Effective use of GPS data offers tremendous opportunities for developing new and relevant measures related to truck performance on highways and overall traffic mobility. Given the situation that massive GPS data are available; however, without further information, it is hard to identify when the trip begins and ends. Instead of using a fixed threshold of the gap time between trips, this paper provided a method including a set of steps to disaggregate the valid trips out of the GPS dataset. This is done by determining the OD stops using the density of GPS data pins or stop durations, error data filtering, curve smoothing, trip discretisation, and assembling. The result of the trip split method was used to the travel time reliability to show the potentials of future application. The contributions are:

1) A systematic procedure with a generic step-by-step approach to purify and extract trip information from large amounts of raw GPS dataset. Compared to the most used methods of split GPS data with fixed dwell time threshold, the density of stop scale and stop frequency are used, which reduces errors of identifying the breaks from repeated and long stops. Meanwhile, this study documents a process to determine the target route, identify the trip ends, and split trip chains for each truck.

2) A unique addition to the existing body of literature on performance measures of dedicated trucks with different route lengths. Trip information purification and extraction can be readily extended to other data sources such as the American Transportation Research Institutes (ATRI) truck GPS database.

3) An overall performance measure in travel time reliability is applicable to other similar application scenarios in need of punctuality such as tourist couches, intercity shuttles, school buses, and long-distance working commuter cars. In fact, the logistics and trucking industry and dispatch centres would probably find this paper useful.

Admittedly, there are limitations to this paper. Significant factors that affect travel time reliability would be roadway conditions and the traffic environment, which are unknown in the dataset. In over one month of the study period, each trip may experience different situations (e.g. congestions, special events, etc.). Combining all the trip data without knowing that information for each trip may lead to unnecessary extra errors in travel time reliability measures.

A significant limitation of non-intrusive GPS data is the lack of verification of freight facilities along the route. That is also why there is an increasing trend of combining GPS data with GIS techniques to divide the route into segments on which the analysis is performed [17].

Repeated trips on the same route may incur biases in reliability as drivers may adjust their route choice to avoid traffic events. For example, if a major incident creates very significant congestion, trucks may change their routes to avoid the target route. In that case, the trips (out of the target routes) are removed from the data. Consequently, the travel time calculated from the "clean" OD trips may have understated the worst condition on the right tail of the distribution.

It should be noted that the dataset did not track if the trucks had one or more primary drivers. Also, if more than one driver is assigned to operate a truck, it is unknown which driver takes a specific trip in the dataset. Therefore, different truck drivers may have an impact on the results of the truck travel time reliability.

The next step of the study would continue to investigate the applications of the rich and accurate trip-based data source on truck performance measures and other potential studies. For example, the truck OD trip pairs can be used as an alternative to the conventional 4-step process to advance the accuracy of truck travel demand forecasting models. The truck trips can serve as floating data to identify the time and location of recurring congestions. In addition, the trip trajectory would help

training and optimising the behaviour of autonomous trucking in microsimulation, which will also be the future work of the research team.

赵莉，博士[1]；李颖，博士[2]
[1] 内布拉斯加州林肯大学土木工程学院，地址：美国内布拉斯加州林肯市Prem S. Paul 研究中心，邮编：68583，电子邮件: lizhao@unl.edu
[2] 长安大学信息工程学院，地址：中国陕西省西安市南二环中段，邮编：710064，电子邮件：yingli@chd.edu.cn

从GPS数据中识别起始点的行程：在专用卡车行程时间可靠性中的应用

摘要

数据收集技术的进步带来了GPS应用的热潮。例如，旅行行为的研究就得益于集成了多种来源的全球定位系统（GPS）数据。然而，数据处理的落后仍然阻碍了此类数据的有效使用。尽管GPS数据提供了详细的空间运动信息，但是对于商用卡车GPS数据而言，其起止点并未标识出来。因此，需要开发一种行程识别方法以利用有限信息的轨迹数据。本文着眼于从原始GPS数据中识别出行。具体来说，本文提出了一种系统的方法，通过使用5个步骤提取起始点（OD）的行程。卡车的起始点行程数据被应用在行程时间可靠性的估算中。通过三个常用指标表明，与短途行驶卡车相比，长途行驶卡车具有较差的可靠行驶时间。本文由此提供了一个通过使用GPS数据来研究货运时间可靠性的范例。

关键词：GPS数据；行程提纯和识别；卡车行程时间可靠性；货运性能

REFERENCES

[1] Ambrosini C, Routhier JL. Objectives, methods, and results of surveys carried out in the field of urban freight transport: An international comparison. *Transport Reviews*. 2004;24: 57-77.

[2] De Jong G, Gunn H, Walker W. National and international freight transport models: An overview and ideas for future development. *Transport Reviews*. 2004;24: 103-124. doi: 10.1080/0144164032000122343.

[3] Lomax T, Schrank D, Turner S, Margiotta R. *Selecting travel time reliability measures*. Federal Highway Administration, Washington DC; 2003. https://static.tti.tamu.edu/tti.tamu.edu/documents/TTI-2003-3.pdf [Accessed 11th July 2020].

[4] Hadavi S, et al. Monitoring urban-freight transport based on GPS trajectories of heavy-goods vehicles. *IEEE Transactions on Intelligent Transportation Systems*. 2018;20(10): 3747-3758. doi: 10.1109/TITS.2018.2880949.

[5] Quiroga CA, Bullock D. Travel time studies with global positioning system and geographic information systems: An integrated methodology. *Transportation Research Part C*. 1998;6(1-2): 101-127. doi: 10.1016/S0968-090X(98)00010-2.

[6] Berzina L, Faghri A, Shourijeh MT, Li M. Evaluation of travel time data collection techniques: A statistical analysis. *International Journal of Traffic and Transportation Engineering*. 2013;2(6): 149-158. doi: 10.5923/j.ijtte.20130206.03.

[7] Du JH, Aultman-Hall L. Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues. *Transportation Research Part A*. 2007;41: 220-232. doi: 10.1016/j.tra.2006.05.001.

[8] Huang J, et al. Mining freight truck's trip patterns from GPS data. *17th International IEEE Conference on Intelligent Transportation Systems: 1988-1994*; 2014.

[9] Thakur A, et al. Development of algorithms to convert large streams of truck GPS data into truck trips. *TRR: Journal of the Transportation Research Board*. 2015;2529: 66-73. doi: 10.3141/2529-07.

[10] McCormack E, Hallenbeck M. ITS devices used to collect truck data for performance benchmarks. *TRR: Journal of the Transportation Research Board*. 2006;1957: 43-50.

[11] Sharman BW, Roorda MJ. Analysis of freight global positioning system data: Clustering approach for identifying trip destinations. *TRR: Journal of the Transportation Research Board*. 2011;2246: 83-91. doi:10.3141/1957-07.

[12] Greaves SP, Figliozzi M. Collecting commercial vehicle tour data with passive global positioning system and technology. *TRR: Journal of the Transportation Research Board*. 2008;2049: 158-166. doi: 10.3141/2049-19.

[13] Zanjani AB, et al. Estimation of statewide origin-destination truck flows from large streams of GPS data: Application for Florida statewide model. *TRR: Journal of the Transportation Research Board*. 2015;2494: 87-96. doi: 10.3141/2494-10.

[14] Flaskou M, et al. Analysis of freight corridors using GPS data on trucks. *TRR: Journal of the Transportation Research Board*. 2015;2478: 113-122. doi: 10.3141/2478-13.

[15] Liao CF. *Using archived truck GPS data for freight performance analysis on I-94/I-90 from the Twin Cities to Chicago*. University of Minnesota. Report No. CTS 09-27, 2009.

[16] Kamali M, Ermagun A, Viswanathan K, Pinjari AR. Deriving truck route choice from large GPS data streams. *TRR: Journal of the Transportation Research Board*. 2016;2563: 62-70. doi: 10.3141/2563-10.

[17] Zhao WJ, McCormack E, Dailey DJ, Scharnhorst E. Using truck probe GPS data to identify and rank roadway bottlenecks. *Journal of Transportation Engineering*. 2012;139(1): 1-7. doi: 10.1061/(ASCE)TE.1943-5436.0000444.

[18] Dulebenets MA, Pujats K, Deligiannis N. Development of tools for processing truck GPS data and analysis of freight transportation facilities. *Transportation Research Board 96th Annual Meeting, 8-12 January 2017*.

Washington DC; 2017. Paper No. 17-00717.

[19] Yang X, Sun Z, Ban XJ, Holguín-Veras J. Urban freight delivery stop identification with GPS data. *TRR: Journal of the Transportation Research Board.* 2014;2411: 55-61. doi: 10.3141/2411-07.

[20] Gingerich K, Maoh H, Anderson W. Classifying the purpose of stopped truck events: An application of entropy to GPS data. *Transportation Research Part C.* 2016;64: 17-27. doi: 10.1016/j.trc.2016.01.002.

[21] McCormack E. *Developing transportation metrics from commercial GPS truck data.* University of Washington. Report No. TNW2011-12, 2011.

[22] Liao CF. *Measure of truck delay and reliability at the corridor level.* Minnesota Department of Transportation. Report No. MnDOT 2018-15, 2018.

[23] Yang S, An CC, Wu YJ, Xia J. Origin-destination-based travel time reliability. *TRR: Journal of the Transportation Research Board.* 2017;2643: 139-159. doi: 10.3141/2643-16.

[24] Clark SD, Watling DP. Modelling network travel time reliability under stochastic demand. *Transportation Research Part B.* 2005;39(2): 119-140. doi: 10.1016/j.trb.2003.10.006.

[25] Doustmohammadi E, Sisiopiku VP, Sullivan A. Modeling freight truck trips in Birmingham using tour-based approach. *Journal of Transportation Technologies.* 2016;6: 436-448. doi: 10.4236/jtts.2016.65035.

[26] Bassok A, McCormack ED, Outwater ML, Chilan T. Use of truck GPS data for freight forecasting. *Transportation Research Board 90th Annual Meeting, Washington DC*; 2011. Paper No. 11-3033.

[27] Ma XL, McCormack E, Wang YH. Processing commercial global positioning system data to develop a web-based truck performance measures program. *TRR: Journal of the Transportation Research Board.* 2011;2246: 92-100. doi: 10.3141/2246-12.

[28] Jackson E, Aultman-Hall L, Holmén BA, Du J. Evaluating the ability of global positioning system receivers to measure a real-world operating mode for emissions research. *TRR: Journal of the Transportation Research Board.* 2005;1941: 43-50. doi: 10.1177/0361198105194100106.

[29] Li Y, Zhao L, Rilett LR. Driving performances assessment based on speed variation using dedicated route truck GPS data. *IEEE Access.* 2019;7: 51002-51013. doi: 10.1109/ACCESS.2019.2909572.

[30] Nur Arifin Z. *Route choice modeling based on GPS tracking data: The case of Jakarta.* PhD thesis. ETH Zurich; 2012.

[31] Rousseeuw P. Least median of squares regression. *Journal of the American Statistical Association.* 1984;79(388): 871-880. doi:10.1080/01621459.1984.10477105.

[32] Zegeer J, et al. *Incorporating travel time reliability into the Highway Capacity Manual.* National Research Council. Report No. S2-L08-RW-1, 2014.

[33] Texas Transportation Institute and Cambridge Systems Inc. *Travel time reliability: making it there on time, all the time.* US DOT. http://ops.fhwa.dot.gov/publications/tt_reliability/ [Accessed 26th June 2019].

[34] Pu WJ. Analytic relationships between travel time reliability measures. *TRR: Journal of the Transportation Research Board.* 2011;2254: 122-130. doi: 10.3141/2254-13.

[35] Cambridge Systematics Inc. *Cost-effective performance measures for travel time delay, variation, and reliability.* Transportation Research Board, Washington. NCHRP Report 618. DC, 2008.

[36] Van Lint JW, Van Zuylen HJ, Tu H. Travel time unreliability on freeways: Why measures based on variance tell only half the story. *Transportation Research Part A.* 2008;42(1): 258-277. doi:10.1016/j.tra.2007.08.008.

[37] Chen P, Tong R, Lu G, Wang Y. Exploring travel time distribution and variability patterns using probe vehicle data: Case study in Beijing. *Journal of Advanced Transportation.* 2018; ID 3747632. doi: 10.1155/2018/3747632.

[38] Pu WJ, Meese AJ. Exploring travel time reliability under different circumstances: A case study. *Transportation Research Board 89th Annual Meeting, Washington DC*; 2010. Paper No. 10-0161.

[39] Wang Z, Goodchild A, McCormack E. Measuring truck travel time reliability using truck probe GPS data. *Journal of Intelligent Transportation Systems.* 2016;20(2): 103-112. doi: 10.1080/15472450.2014.1000455.