Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 34 (2018) 131–138

Transportation Research Procedia

www.elsevier.com/locate/procedia

International Symposium of Transport Simulation (ISTS'18) and the International Workshop on Traffic Data Collection and its Standardization (IWTDCS'18)

Traffic Data Characterisation: Review and Challenges

Sara Respati[a*], Ashish Bhaskar[b], Edward Chung[c]

[a]PhD Candidate, School of Civil Engineering & Build Environment, Queensland University of Technology (QUT), Brisbane, Australia
[b]Senior Lecturer, School of Civil Engineering & Build Environment, Queensland University of Technology (QUT), Brisbane, Australia
[c]Professor in Intelligent Transport System, Electrical Engineering, Hong Kong Polytechnic University, Hong Kong, China

Abstract

The growth of technology provides a great amount of traffic data that have distinct characteristics. The absence of the comprehensive understanding of the characteristics and associated challenges leads to resources extravagance. In this paper, we develop data characterisation by disaggregating important traffic data features and present the associated data challenges to provide better insights of traffic data and expand traffic data usage. The paper outlines the opportunity to maximize the data utilisation.

Keywords: Traffic data; Traffic data characteristics; Traffic data challenges

1. Introduction

The technology evolution of road and in-vehicle telematics presented new ways of collecting various traffic data. For instance, Bluetooth Mac Scanners (BMS) and crowdsourced data have been gaining popularity besides the traditional loop detectors in sampling important traffic parameters such as traffic speed and travel time. Lack of the understanding of the detail data characteristics results in its limited applications. A data characteristics catalogue should support comprehensive understanding and suitable utilisation of the data for various traffic and transport

* Corresponding author. *E-mail address:* sara.respati@hdr.qut.edu.au

10.1016/j.trpro.2018.11.024

applications. Recognising the technical strengths and limitations of the sensors data acquisition is also essential. In literature and practice, the comprehensive data characterisation and the associated challenges is generally overlooked.

Therefore, this paper aims to develop a traffic data characterisation that gives a comprehensive view of data, present an overview of the selected traffic sensors data characteristics including loop detectors, BMS, GPS-enabled vehicles, cellular phone and crowdsourced data, and outline the challenges of traffic sensors in offering specific traffic data. This paper also presents the opportunity that can be exploited from traffic data to maximise the data utilisation.

The reminder of this article is organised as follow: first, in section 2, the existing study of data characteristics is discussed. Section 3 presents the proposed traffic data characterisation and summary of selected traffic data characteristics. Based on the identified data characteristics, the outline of challenges in traffic data is provided in section 4. Discussion and conclusion are presented in section 5 that provides suggestions of how to exploit the use of traffic data. The detail of data characterisation and challenges are provided in Appendix.

2. Background

Several lines of research have identified various traffic data characteristics. The term "characteristic" is yet to be specified to exploit the data utilisation. Lin et al. (2012) measured traffic data reliability based on three criteria: *fundamental consistency*, *network consistency* and *historical consistency*. The study argues that a system that incorporates all these three assessments can provide a better understanding of data reliability. They proposed fuzzy-logic based classifier for calculating the reliability based on those three criteria.

Despite identifying data features and features based on the set characteristics, many studies have presented overview of various traffic sensors data individually. For instance, Leduc (2008) presented roadside detectors and the floating car data characteristics. He characterises those traffic sensors based on the variables, or the offered data type, and the strength and weakness of the sensors. The variables presented are either the variables that can be directly generated from the sensors or that are needed to be processed using algorithms. For the emerging traffic sensors, Bhaskar and Chung (2013) have presented the fundamental understanding of BMS traffic data. Review of other traffic data sources include mobile phone (Rose 2006; Steenbruggen et al. 2013; Zhang et al. 2010) and GPS-based traffic sensors (Herrera et al. 2010; Patire et al. 2015).

There is a limited effort to develop data characteristics comprehensively including the comparison between sensors. In fact, it can be used for data requirement identification and selection for various traffic application. The comprehensive data characterisation would give more understanding of the unique data features and help traffic operators to select the most suitable dataset for the designated application.

3. Proposed Traffic Data Characteristics

Traffic data characteristics should present data features or character to give a good understanding of the data. In an effort to improve the understanding and utilisation of the data, this paper proposes the following data characteristics:

- *Data acquisition process*: the process of collecting the traffic data from the network varies over the sensor type.
- *Data type*: data type is categorized into roadside, point to point and floating car sensor data. A roadside sensor collects traffic data at a certain point in the network. Traffic data from point-to-point sensors are collected from a pair of roadside scanners. Floating car data (FCD) is data from which information of devices located in the car are continually recorded.
- *Availability*: to some extent, data availability varies over space and time and it affects the confidence of the data.
- *Quality*: the degree of error in the dataset in measuring or representing the real traffic condition.

According to the proposed data characterisation, Table A. 1 presents the data characteristics of the selected traffic sensors that includes Bluetooth, loop detectors, GPS-enabled taxis, mobile phone and crowdsourcing data. Each data has unique characteristics. They offer different traffic variable, even though the type of sensor is similar. For instance, Bluetooth and mobile phone data are collected from point to point sensors, however they provide distinct traffic variables. Bluetooth data offer timestamp of vehicle detection and duration of vehicles inside the detection zone; whereas mobile phone data includes cell ID at where the vehicle is detected and the timestamp. Point-to-point sensors require vehicles or devices identification to collect the data. The temporal coverage is affected by the availability of the devices scanned by the sensors while the spatial coverage depends on the sensor technology. BMS is generally installed at intersections in arterial and every regular distance in freeway. Meanwhile, cell towers employed to obtain mobile phone data are widely available, but the density may be lower in outside urban areas. GPS-enabled vehicles, which is a FCD, have higher spatial coverage because the detection location is not fixed. The small sample size is the main issue of the GPS-enabled vehicles that produces a low temporal coverage. Data from roadside sensors generally have large temporal coverage such that all the time the vehicle is detected. The sensors are often available only on major roads and installed at specific configuration. The specific placement of detectors and the limited availability (except for major roads and freeway) lead to a low spatial coverage.

The sensors have data quality issue due to their technical limitations. The limitation includes, for example, vehicle counting and speed detection error in loops detection, vehicle positioning error of GPS-enabled vehicles and mobile phone data, and spatial and temporal error in BMS detection. Data acquisition process and type are generally consistent for the same sensors. However, data availability and quality are typically dynamic over space and time. Therefore, overall data characteristic in one place and time is not necessarily similar to a different site and period. Periodic updates for the characteristics is essential to ensure its representativeness.

4. Challenges in Traffic Data

In this section, we present summary of challenges of traffic sensors in deriving certain traffic data including historical 15-min speed, historical 15-min volume, near real-time traffic speed and volume, and historical origin-destination (OD) matrix. The detail of challenges is shown in Table A. 2.

The mentioned traffic data are broadly applied for various traffic application. For example, 15-min historical speed and volume data support network reporting and planning. Network planning can also take benefit of historical OD matrix data to know the traffic movement over the network. Near real-time speed and near real-time volume can be utilised for incident detection, advanced traveller information system (ATIS) and traffic signal operation that require the most up to date data that would help the applications to take the best action.

The challenges of traffic sensors in providing specific traffic data indicators can be categorized into two types: internal and external challenges. In the application of traffic data, we should consider these internal and external challenges, as such, traffic data can accurately represent the real traffic condition.

Internal challenge is related to the accuracy of the devices in collecting the data-the so-called technical limitation, for instance, scanner error that provide inaccurate detected speed (refer to section 3). External challenges pertain to the limitation of the data in reflecting the true traffic condition. For example, in speed data acquisition, loops detect instantaneous speed at one point on a link which is not necessarily reflecting the whole link speed. Bluetooth speed data from a pair of BMSs cannot capture the speed variability between the scanners. This condition is also applied to mobile speed data estimated from two *handover* points or location area. If the speed variability is high over a link stretch, the estimated speed from point-point sensors could mislead the observation. This should be taken into consideration, especially in arterial road at where the speed has high variability mainly due to traffic signal. GPS-enabled vehicles provide data generally every regular time causing a whole link might not be observed. There is a chance of a missed observation for a short link generating no sample available. These two cases require extrapolation to estimate the speed for the link. In missed observation, path estimation is also required if there are more than one possible route that the vehicle has taken.

5. Discussion and Conclusion

The lack of understanding of complexity in the traffic data leads to a failure in inferring traffic condition accurately. This paper presents traffic data characteristics and the challenges of deriving traffic variables from the data with the final intention of providing a comprehensive view of the typical data features. With the development of technology, the richness of traffic data is offered, for instance, the availability of the data discussed in this paper including loop detectors, BMS, GPS-enabled dedicated vehicles, mobile phone and crowdsourced traffic data.

Generally, these traffic sensors data can be differentiated into three groups. Firstly, sensors that are installed on the road and detect the physical presence of vehicles anonymously (roadside sensor data). Secondly, sensors that detect devices inside vehicles with known ID (point-to-point sensor data). Finally, traffic data that are obtained from devices carried by vehicles sending the information at certain time frequency (GPS-enabled data). The challenge of traffic sensors can be categorised into internal and external challenges. Internal challenge is related to the accuracy of the sensors due to technical limitation, while external challenge pertains to the ability of sensors to infer the real traffic condition. These challenges should carefully be considered to obtain a good dataset that can infer the real traffic state.

In the era of mobile internet, crowdsourced data have been extensively growing- it employs passive participation from travellers using GPS-enabled smartphones and GPS-enabled vehicles. The market has increased rapidly- INRIX, Google, TomTom, HERE and Intelematics are the example of crowdsourced traffic data service providers. The sample of crowdsourced data reflects the traffic better than dedicated GPS-based probe vehicles, such as buses and taxis, because the penetration rate is much higher. Moreover, the development of mobile internet services and technology increase these traffic monitoring penetrations and may reach an extensive spatial and temporal coverage in the future.

The mentioned traffic data can provide historical speed data with their limitation such as low sample size and low location precision (refer to Table A. 2). All traffic sensors have a time lag in providing real-time speed data; as such, speed data can only be calculated when vehicles have completed their journey on a certain link. In addition to the speed data, traffic volume is also widely used for different applications. Traffic volume can be measured by a sensor that detects vehicles continuously without any dependency of devices availability in vehicles. Loop detectors have this ability even though literature have reported errors in the loops vehicle counting. However, unlike other discussed sensors, loop detectors cannot measure OD matrix because of no vehicles/devices ID detected. Traditionally it is estimated using loop a bi-level optimization. BMS and mobile phone have a potential to support OD matrix estimation. While each data have their characteristics, strength and weakness distinctively, the future may lie in fusing those data to obtain high temporal and spatial coverage, and reliability. In that way, traffic applications that exploit the data on the basis of fusing traffic data variable may achieve high application performance (e.g. accuracy).

Despite the existence of big traffic data wave, generally, stakeholders do not treat data as an asset. Needless to say, data is an asset that should be carefully managed and refined into information which would allow the optimal use of the data. For example, during data acquisition and storage, stakeholders should broadly think of the applications where the data is useful. This may include:

- Loops data is generally aggregated. However, disaggregated data such as individual vehicle headways is important for safety studies. With aggregated data we lose its potential to study safety.
- Bluetooth data generally has encrypted MAC-IDs. The encryption algorithm can be different between different stakeholders. For instance, state government managing arterial can have different encryption than the federal government managing motorways. This can unnecessarily limit the applicability of the data for large area studies.
- Third party data do not provide information for the data source and confidence-this can cause 'blind' acceptance and provoke challenges in fusing the data with other sources.

Moreover, development of a database with appropriate geospatial and temporal referencing is vital for the multipurpose use and its integration with other sources.

Appendix

Table A. 1. Traffic Data Characterisation

Characteristics		Data				
		Bluetooth	Loops	GPS-enabled Taxis	Mobile Phone	Crowdsourcing
Data acquisition process		BMSs detect Bluetooth devices (identified as MAC ID) passing the scanners zone. The scanners record the time when vehicles enter the scanners zone and duration of vehicles inside the zone.	Loops detect the presence of vehicle passing them. In double loops, the detectors give the vehicle spot speed information	Traffic data obtained from taxi-probe is based on in-vehicle GPS of dedicated commercial vehicle. GPS sends vehicle position at a certain time interval. There is no information of the true path taken by vehicles.	Data from the handovers was used to measure speed. If the call that produced the handover is long enough to cross the new cell entirely, a second handover will be executed. This will provide the time of the 1st and 2nd handover with known handover location.	Data is collected from users using GPS-enabled smartphones and GPS-enabled vehicles. Information of timestamp, location and speed is sent at regular time.
Data Type	Nature of data	Point-to-point sensor	Roadside sensor	FCD	Point-to-point sensor	FCD
	Provided variable	Timestamp, Bluetooth ID (MAC ID), duration in the detection zone.	Vehicle count, spot speed (for dual loops).	Location, timestamp, spot speed, taxi ID.	Time of the call, caller ID, call duration, and cell ID to which the phone is connected while the call is active, LA.	Location, timestamp, spot speed, ID.
Availability	Spatial coverage	In an arterial network, BMSs are generally located at the intersection; and in motorway, BMSs are typically installed at on-ramp and off-ramp and every certain distance interval.	Loops are installed based on requirement.	Taxis have high spatial coverage in busy urban areas and areas that have high passengers demand.	The mobile phone data has high spatial coverage because generally the cell towers are widely available. In rural areas, the distance between towers may be higher contributing less accurate in devices positioning.	Application in smart-phone provides high penetration rate as the developing of mobile internet. For example, INRIX mentioned that it has traffic community around 70+ million devices providing billions of data points *per month* (http://inrix.com).
	Temporal coverage	BMSs detect BT devices actively within their *inquiry mode.* Penetration rate of BT is reported vary from 5% to 30% (Laharotte et al. 2015)	Loops detect the physical presence of vehicle actively.	The interval varies over place. For instance, every 30 seconds for Brisbane data; and for Shanghai, the information sent at different intervals, the most common being 16 seconds when vacant and 61 seconds when occupied. (Liu et al. 2012)	Activities trigger the data collection: in an on-call mode, the network always knows the base station (cell) which means of handover connects the phone. In idle status, the network knows the Location Area (LA) of the phone when the cell phone is just turned on, or when there is mobility to different LA (Caceres, Wideberg and Benitez 2008).	The frequency of crowdsourced traffic data based on smartphones application and in-vehicle GPS is 30-sec or 60-sec in general.

Characteristics		Data				
		Bluetooth	Loops	GPS-enabled Taxis	Mobile Phone	Crowdsourcing
Quality	Reported accuracy	Error in speed data is due to spatial and temporal error in devices detection. The accuracy is related to the distance between scanners and the average speed. T (Bhaskar, Ashish and Edward Chung. 2013). However, long distance between scanners may have a small sample size.	Generally, loops have 5% error of vehicle counting and speed estimation	Spatial resolution is reported between 0.0001 and 0.0002 degrees latitude and longitude, or about 10-20 meters, and bearing is reported at a resolution of 45 degrees. There are noises because of inconsistencies in observed travel time versus the observed distances. (Deng et al. 2015).	The location is an approximation of the geographical area where a phone is located. There is a margin of error that depends on the cell radius. Speed from mobile phones may have an error greater than 20% due to position estimates error (Steenbruggen et al. 2013).	In the study case reported in Tahmasseby (2015, TomTom has a penetration rate of 6% in the city of Calgary, Canada. The estimated average speed from TomTom only has 1.5% difference with BluFAX (Bluetooth data) that has 3.3% penetration rate.

Table A. 2. Traffic Data Characterisation Traffic variables provided by traffic sensors and the challenges

Data	Traffic Indicators	Historical 15-min speed	Historical 15-min volume	Near Real-time 5-minute speed	Near Real-time volume	Historical OD
Bluetooth	Capability	Yes (Bhaskar and Chung 2013)	Possible with high penetration rate	Yes (Bhaskar and Chung 2013)	No	Possible in principle
	Comments	The distance between scanners should be considered; the shorter the distance, the higher the error of travel time estimate. The long distance between scanners may have the issue of small sample size. Speed between two scanners cannot capture speed variability between scanners.	The sensors have limited applicability to detect all passing vehicles as not all vehicles are equipped with Bluetooth devices. The expansion factor is needed to estimate traffic volume from BMS unless the penetration rate is very high.	There is a time lag in the speed/travel time estimation from BMS. Bluetooth travel time match from a vehicle is available only when the vehicle has travelled the corridor. The lag in Bluetooth data is at least equal to the travel time of the corridor.	The direction of vehicles travel is unknown if there is only one BMS used in volume measurement. If two BMSs are used to obtain the matched MAC ID and to know the vehicle direction, there is an issue of time lag. Scaling factor is needed due to sample size that may be small.	The sample size for OD estimation from BMS may become an issue. Especially, for the MAC ID matching for long corridors. The OD estimate should have proper justification, scale factor and confidence level.
Loop Detector	Capability	Yes (with a model)	Yes (Wang and Nihan 2000)	Yes (Dailey 1999)	Yes (disaggregated data)	No
	Comments	The estimated speed may not reflect the speed of the whole link. The location of the loops affects the speed estimate (e.g. loops located in stop line may underestimate the speed). Speed	The location of the loops should be considered. For the loops that are located at stop-line, the vehicle count should be considered per cycle rather than	Speed estimate in freeway is more accurate compared to arterial that has traffic signal control.	Generally, vehicle count is reported/aggregated every certain interval (e.g., 20 secs/30 secs). In fact, if the disaggregated data can provide the real time volume.	Loops do not detect any ID of vehicles passing them, so they cannot record the origin and destination of vehicles trip.

Data	Traffic Indicators	Historical 15-min speed	Historical 15-min volume	Near Real-time 5-minute speed	Near Real-time volume	Historical OD
		may also estimated from single loops as proposed in Dailey (1999).	continuous counting.			
GPS-based Taxis	Capability	Yes (Herrera et al. 2010)	Possible (Zhan et al. 2017)	Possible with limitation	Possible	Yes (Tang et al. 2015)
	Comments	The information is sent every certain time interval rather than at certain location, so the extrapolation method is needed to measure speed on a certain link.	The estimation is possible given a sufficient rich data.	The speed is calculated when taxis send their location at downstream link generating a time lag.	The estimation is possible given a sufficient rich data.	GPS-equipped taxi provides the start and end point of a trip. However, the OD provided is at GPS-level OD that may not represent the OD for the whole network.
Mobile Phone	Capability	Yes (Steenbruggen et al. 2013)	Yes (Caceres, Wideberg and Benitez 2008)	Yes	Possible in principle	Yes (Calabrese et al. 2011)
	Comments	Data from cellular phones have an issue of location estimates accuracy and it should not be ignored. Reported in Steenbruggen et al. (2013), speeds from a cellular phone may have an error greater than 20% due to position estimates error.	The HO in call data and LA data can be utilised to estimate traffic volume. In low calling intensity trend in which the sample size is low. A model is needed to estimate traffic volume, for example by using Cobb-Douglas model (Caceres, Wideberg and Benitez 2008).	The sample size for a near real-time data may become an issue as the estimation depends on the calling data (for the more accurate position).	Using a model, real-time volume can be estimated using the virtual traffic counter located on the cell borders. The position estimates, road geometry and sample size should be monitored as these factors greatly affect the estimation accuracy (Caceres, Wideberg and Benitez 2008).	OD matrix can be obtained from calling data as presented in Iqbal et al. (2014; and for larger sample size, OD can be estimated from LA where a set of LAs or cells are assigned as centroids (Caceres, Wideberg and Benitez 2008).
Crowd-sourced	Capability	Yes	Yes (Zhan et al. 2017)	Yes	No	Possible in principle
	Comments	The data might be based on GPS-enabled mobile phone or in-vehicle GPS that send speed and location information every certain time interval. Extrapolation is needed to estimate the whole link speed. The penetration rate of 2%-3% is enough to provide accurate speed estimation. (Herrera et al. 2010)	A model is needed to estimate volume from crowdsourced-GPS data.	The crowdsourced data are generally collected from GPS that record vehicle location every certain time interval, so the speed cannot be calculated before vehicles passing the link-therefore, there is a time lag for real-time data.	The major issue is sample size that may not enough to represent traffic volume.	The trajectory data from GPS-enabled mobile phone gives origin and destination information of a trip. The use of this data can build an OD matrix.

REFERENCES

Bhaskar, Ashish and Edward Chung. 2013. "Fundamental understanding on the use of Bluetooth scanner as a complementary transport data." *Transportation Research Part C: Emerging Technologies* 37: 42-72. http://www.sciencedirect.com/science/article/pii/S0968090X13002027. doi: http://dx.doi.org/10.1016/j.trc.2013.09.013.

Caceres, N, JP Wideberg and F Garcıa Benitez. 2008. "Review of traffic data estimations extracted from cellular networks." *IET Intelligent Transport Systems* 2 (3): 179-192.

Calabrese, Francesco, Giusy Di Lorenzo, Liang Liu and Carlo Ratti. 2011. "Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area." *IEEE Pervasive Computing* 10 (4): 36-44.

Dailey, Daniel J. 1999. "A statistical algorithm for estimating speed from single loop volume and occupancy measurements." *Transportation Research Part B: Methodological* 33 (5): 313-322.

Deng, Bin, Steve Denman, Vassilis Zachariadis and Ying Jin. 2015. "Estimating traffic delays and network speeds from low-frequency GPS taxis traces for urban transport modelling." *European Journal of Transport & Infrastructure Research* 15 (4).

Herrera, Juan C., Daniel B. Work, Ryan Herring, Xuegang Ban, Quinn Jacobson and Alexandre M. Bayen. 2010. "Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment." *Transportation Research Part C: Emerging Technologies* 18 (4): 568-583. http://www.sciencedirect.com/science/article/pii/S0968090X09001430. doi: http://dx.doi.org/10.1016/j.trc.2009.10.006.

Iqbal, Md Shahadat, Charisma F Choudhury, Pu Wang and Marta C González. 2014. "Development of origin–destination matrices using mobile phone call data." *Transportation Research Part C: Emerging Technologies* 40: 63-74.

Laharotte, P. A., R. Billot, E. Come, L. Oukhellou, A. Nantes and N. E. El Faouzi. 2015. "Spatiotemporal Analysis of Bluetooth Data: Application to a Large Urban Network." *IEEE Transactions on Intelligent Transportation Systems* 16 (3): 1439-1448. doi: 10.1109/TITS.2014.2367165.

Leduc, Guillaume. 2008. "Road traffic data: Collection methods and applications." *Working Papers on Energy, Transport and Climate Change* 1 (55).

Lin, Dung-Ying, Stephen Boyles, Varunraj Valsaraj and S Travis Waller. 2012. "Reliability assessment for traffic data." *Journal of the Chinese Institute of Engineers* 35 (3): 285-297.

Liu, Xuemei, James Biagioni, Jakob Eriksson, Yin Wang, George Forman and Yanmin Zhu. 2012. "Mining large-scale, sparse GPS traces for map inference: comparison of approaches." In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, edited, 669-677: ACM.

Patire, Anthony D, Matthew Wright, Boris Prodhomme and Alexandre M Bayen. 2015. "How much GPS data do we need?" *Transportation Research Part C: Emerging Technologies* 58: 325-342.

Rose, Geoff. 2006. "Mobile Phones as Traffic Probes: Practices, Prospects and Issues." *Transport Reviews* 26 (3): 275-291. https://doi.org/10.1080/01441640500361108. doi: 10.1080/01441640500361108.

Steenbruggen, John, Maria Teresa Borzacchiello, Peter Nijkamp and Henk Scholten. 2013. "Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities." *GeoJournal* 78 (2): 223-243. https://doi.org/10.1007/s10708-011-9413-y. doi: 10.1007/s10708-011-9413-y.

Tahmasseby, Shahram. 2015. "Traffic data: Bluetooth sensors vs. crowdsourcing—A comparative study to calculate travel time reliability in calgary, alberta, canada." *Journal of Traffic and Transportation Engineering* 3: 63-79.

Tang, Jinjun, Fang Liu, Yinhai Wang and Hua Wang. 2015. "Uncovering urban human mobility from large scale taxi GPS data." *Physica A: Statistical Mechanics and its Applications* 438: 140-153. http://www.sciencedirect.com/science/article/pii/S0378437115005853. doi: https://doi.org/10.1016/j.physa.2015.06.032.

Wang, Yinhai and Nancy Nihan. 2000. "Freeway traffic speed estimation with single-loop outputs." *Transportation Research Record: Journal of the Transportation Research Board* (1727): 120-126.

Zhan, X., Y. Zheng, X. Yi and S. V. Ukkusuri. 2017. "Citywide Traffic Volume Estimation Using Trajectory Data." *IEEE Transactions on Knowledge and Data Engineering* 29 (2): 272-285. doi: 10.1109/TKDE.2016.2621104.

Zhang, Yi, Xiao Qin, Shen Dong and Bin Ran. 2010. "Daily OD matrix estimation using cellular probe data." In *89th Annual Meeting Transportation Research Board*, edited.