POLYPUBLIE

Workshop synthesis: system based passive data streams systems; smart cards, phone data, GPS

Martin Trépanier, Toshiyuki Yamamoto

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

Date	2015
Type	Article de revue / Article
Version	Version officielle de l'éditeur / Published version
URL de PolyPublie PolyPublie URL	https://publications.polymtl.ca/35502/
Évaluation par les pairs Peer reviewed	Révisé par les pairs / Refereed
Conditions d'utilisation Terms of use	Creative Commons Attribution-Utilisation non commerciale-Pas d'oeuvre dérivée 4.0 International / Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND)
Référence Citation	Trépanier, M., & Yamamoto, T. (2015). Workshop synthesis: system based passive data streams systems; smart cards, phone data, GPS. Transportation Research Procedia, 11, 340-349. Présentée à 10th International Conference on Transport Survey Methods, Leura, Australia. https://doi.org/10.1016/j.trpro.2015.12.029

Publication officielle
Official publication

Titre de la revue Journal title	Transportation Research Procedia (vol. 11)
Maison d'édition Publisher	Elsevier
URL officielle Official URL	https://doi.org/10.1016/j.trpro.2015.12.029
Mention légale Legal notice	©2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal
https://publications.polymtl.ca

Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 11 (2015) 340 – 349

Transportation Research Procedia

www.elsevier.com/locate/procedia

10th International Conference on Transport Survey Methods

Workshop synthesis: System based passive data streams systems; smart cards, phone data, GPS

Martin Trépanier [a,*], Toshiyuki Yamamoto [b]

[a] *Polytechnique Montréal, CIRRELT (Interuniversitary Research Centre on Entreprise Networks, Logistics and Transportation) P.O. box 6079, station Centre-ville, Montréal, Québec H3C 3A7, Canada*

[b] *Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan*

Abstract

Each day, huge quantities of passive data streams are collected by smart card, GPS, Bluetooth and mobile phone systems all over the world. These data happen to be very useful to transport planners, because of the valuable spatial and temporal information they contain. This paper is a synthesis of a workshop on passive data streams held in Australia in November 2014. Many issues are discussed here: definitions, data collection and processing, privacy, how to use these data for transport planning, how to integrate these data with traditional and more "active" data sources, such as household travel surveys that respondents are explicitly asked to participate in, and how to cope with the absence of socio-demographic attributes in passive data streams.

© 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of International Steering Committee for Transport Survey Conferences ISCTSC

Keywords: Passive data stream; smart card; smart phone; GPS; Bluetooth; privacy

1. Introduction

The use of passive data streams has become increasingly popular among transport planners. Streams are said to be "passive" because there is no acknowledgement of the user that "his/her" data are used for planning. Hence, there is no additional input that could be used to qualify, for example, the purpose of the trips or the socio-demographic attributes of the users. The huge quantities of data collected continuously by passive stream technologies, such as

* Corresponding author. Tel.: +1-514-340-4711 ext. 4911; fax: +1-514-340-4173.
E-mail address: martin.trepanier@polymtl.ca

2352-1465 © 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of International Steering Committee for Transport Survey Conferences ISCTSC
doi:10.1016/j.trpro.2015.12.029

mobile phones, GPS devices, Bluetooth sensors and smart card fare collection systems in public transport, can surely be used to analyze the behavior of transport users or vehicles.

This document synthesizes the discussion of the W8 workshop for the International Steering Committee on Travel Survey Conference that was held in Leura, Australia, in November 2014. The workshop can be regarded as the extension of the workshop at the previous conference of the same series in 2011. The previous workshop focused on passive public transportation data streams (Morency, 2013), and several other types of passive data, such as GPS, Bluetooth and mobile phone systems, are added to the focus of the current workshop. It presents the topics as they were discussed during the 6-hour workshop:

- privacy issues related to the use of passive data streams, may it be primary or secondary data as defined by the participants;
- challenges related to the use of passive data (GPS, mobile phone, Bluetooth, smart card);
- the possibility to add socio-demographic information to collected data;
- the role of passive data streams in transport planning, for a wider perspective.

For greater convenience, the different types of data streams are discussed in separate sections because each type has its own issues and challenges. The paper finishes with a concluding section that discusses the open questions that emerged during the workshop. Further research directions are also suggested.

2. Context

This section presents contextual elements about the definition, privacy and control of passive data streams.

2.1. Definition of passive data streams

There are two types of passive data streams. A traveler, through the use of a smart phone application or a portable GPS device, produces a primary data stream voluntarily. However, in all cases there are no interactions between the device and the user during data collection, but post-validation can occur in specific research studies. A secondary data stream is the most common source: data are collected without any knowledge from the user, by a system that sometimes is not intended to collect mobility data. Smart card fare collection system data, mobile phone location collection by phone operators and roadside Bluetooth detection are examples of secondary data.

2.2. Privacy

Passive data streams can be used to trace the mobility behavior of a single transport user. This raises privacy questions because the confidentiality of data must be assured by people collecting and analyzing data. For primary data streams, the issue is the same as for active surveys: a consent agreement can be established between the respondents and the survey organization. A secondary data stream is more complicated because there is usually no consent agreement. In this case, the data stream must be kept anonymous according to the legislations in place. To avoid reverse identification, an irreversible encryption algorithm can be used to encrypt the identification number of the card, phone or Bluetooth device. In this way, the mobility pattern associated to each device can be analyzed, but the link to the transport user cannot be performed. Privacy questions about the usage of data can be addressed through ethical reviews completed by institutional review boards. During the workshop, the privacy issue was presented by Yamamoto (2014). In his study, a driver's willingness to provide vehicle travel information using a GPS device is investigated, and the results suggest that frequent vehicle users are more willing to provide the most information, including travel routes, origins and destinations, but less willing to provide data on speed.

2.3. Data control of secondary data

Although researchers or the transport authority usually collect primary data, this is not the case for secondary data, which causes some issues. First, there can be a cost of data acquisition, and agreements must be signed

between the provider and the users. Commercial licensing of data can raise questions about data quality. In many cases, the researcher may not know how the original data of the provider have been processed before being transmitted. Preprocessing could alter or change the data in a sense that mobility analyses become useless. For example, mobile phone data could be aggregated to a zoning system that is incompatible with transport analysis zones (TAZ). In addition, data can be sampled, and one may not know the universe, sampling rate and scheme required to expand the data. This is often the case for passive data streams: we do not know who owns the mobile phones, smart card and Bluetooth.

3. GPS data

GPS stands for "global positioning system", a system first developed by the United States government. GPS data consist of a date, time, and latitude/longitude, so the location of the vehicle or person with the GPS receiver can be identified globally. The longitudinal GPS data can be used to trace the trajectory of the vehicle or person in time and space.

3.1. How does it work?

GPS has been used for travel surveys since the mid-1990s, just after GPS was made available to the public (Zito et al. 1995; Murakami and Wagner 1999). GPS satellites are constantly emitting signals from the universe. The signals are received and processed by GPS receivers to obtain the latitude, longitude and height coordinates of the receiver. The location of the GPS receiver is identified by triangulation. Thus, more accurate locations can be obtained by clearer signals from each satellite and signals from more satellites. Before 2000, the accuracy of GPS was intentionally degraded for public use by the US Government because of security reasons, which was called selective availability (SA). SA was turned off in 2000, and the accuracy improved. The accuracy of GPS data decreases at locations surrounded by high-rise buildings, where the area of sky visible to the GPS receiver is reduced, and the signal is lost in tunnels, underground and inside buildings. Thus, a given person's trip survey using GPS data may be more incomplete than a vehicle trip survey.

3.2. What can we do with the data?

GPS data can be used to measure vehicle speed on road segments, traffic congestion, waiting time at intersections and stations, route choice, origin and destination of trips, travel mode, etc. However, the observed information is the space-time trajectory only, so the data must be processed before being used as meaningful information. During the workshop, Escobar Garcia et al. (2014) and Lovelace (2014) presented the applications of GPS data. Escobar Garcia et al. (2014) evaluated the road network based on observations of detailed travel routes and operational speeds using GPS data. Data collection occurred during 45 consecutive days, so the reliable day-to-day vehicle flows within the city could be obtained. Potential application for monitoring energy use of travel behavior was explored by Lovelace (2014). It is suggested that GPS data and smart phone systems can be used to improve the estimation of energy use resulting from travel behavior.

3.3. Data processing

The origin and destination of a trip can be observed directly in case the GPS receiver is set in a vehicle and the switch is turned on and off according to the engine start and stop or the GPS receiver is held by a person who turns on and off at the trip start and end. In other cases, the trip ends should be identified from the trajectory. Travel route should be identified by map-matching of the trajectory to the GIS database. Travel mode should be identified by comparing the trajectory and transport network of the GIS database, time table of public transit, and the travel speed calculated from the GPS trajectory. If trip end locations can be identified accurately, the trip purpose is also inferred by the land use of the location, duration of stay, time of day, frequency of visits (in the case of multi-day surveys), etc.

Considerable research has been conducted to develop data processing algorithms, but the validation of the original GPS data is difficult without the ground truth information, and the accuracy of the processing algorithm heavily depends on the original GPS data accuracy and frequency. Also, the algorithms have been developed in parallel by many researchers, and the program codes are not generally available for others. Thus, the comparison of the algorithms is difficult, and the standard methods are not yet established. This prevents the smooth improvement of the algorithms.

Another problem with the algorithm is its computational speed. Since complex data processing is required for GPS data to be useful, the scalability of a GPS survey should be well considered. More in-depth discussion on post-processing of GPS data was held at a different workshop in the previous conference and reported by Stopher and Sneade (2013).

4. Mobile phone data

Mobile phones are basically devices to make and receive telephone calls not limited to locations where landline phones are available. People take their mobile phones with them as they leave the home, and carry it all day, so the location of the user with the mobile phone can be identified by identifying the location of the mobile phone, which is done by the mobile phone network provider to maintain the connection of mobile phone to the telephone network. In addition, technologies related with mobile phones have been extensively developed and they are now equipped with many sensors; these mobile phones are called smart phones. The sensors include GPS, accelerometer, gyroscope, compass, gravity, brightness, proximity, atmospheric pressure, temperature, etc. Such sensors can also be used to identify the location and other aspects of travel behavior, including trip ends, travel speed and mode.

4.1. How does it work?

Whenever a user of a mobile phone makes or receives a call or a text message or connects to the internet, the location of the tower routing the communication is recorded. Thus, the location of the user can be identified as the location of the tower. The spatial accuracy of the observation depends on the density of the towers, which is not as accurate as GPS data. Also, if the user turns off the phone or puts it in airplane mode, the location of the mobile phone is not recorded, so the full trajectory of the user is not observed.

In some applications, the latitude and longitude are observed by GPS receivers installed in the mobile phone, and the information is sent to the data center periodically even when the user does not make any communications intentionally, which enables the mobile phone network provider to monitor the location of the user continuously. In this case, the information based on the tower location is not used, but the GPS data can be used to identify the location of the mobile phone at higher accuracy, and the discussions in Section 3 apply. However, it is known that the GPS consumes considerable battery energy, so the battery life of smart phones shortens, which prevents continuous GPS observations of smart phones.

In addition to GPS data, other information observed by sensors installed in the mobile phone can be sent to the data center of the mobile phone network provider and also to other third party companies, which depends on the applications and agreements by the user. In this case, the accuracy of the estimation of the travel mode discussed in Section 3 can be improved.

4.2. What can we do with the data?

The location information obtained by the tower-based observation provides the trajectory of the mobile phone user at a lower accuracy of the location than the GPS data, so it can be used to measure the same aspects at lower accuracy. However, the sample size of mobile phone data can be much greater than GPS surveys, so the aggregated information, such as origin-destination matrices by time of day, can be estimated at a higher accuracy with the mobile phone data than with smaller sample size GPS data. Gonzalez et al. (2008) among others investigated mobility patterns using mobile phone data with a sample of over 100,000 individuals.

When GPS data are used to identify the location of mobile phones, as mentioned in Section 3, travel speed, traffic congestion, route choice, trip origin and destination, travel mode, etc., can be estimated. When other information

observed by several sensors in addition to GPS are used, the accuracy of the estimation mentioned above can be improved, and other aspects of the travel behavior can be observed. During the workshop, three applications of mobile phone data were presented and discussed. An origin-destination matrix is estimated in two studies, but in different contexts. Bonnel et al. (2014) could access location information of tower-based individual phone data. The results of the comparison of the estimated origin-destination matrix with that from household travel surveys were not satisfactory, and the possible improvements were discussed. Ge and Fukuda (2014) used aggregated phone data and developed an updating algorithm to update the origin-destination matrix with travel survey data. The algorithm was validated using synthetic data, and then applied to the Tokyo area. Another application of mobile phone data was presented by Hara (2014). Geo-tagged tweet data obtained through smartphones were used to investigate travel behavior.

4.3. Representativeness

Although the sample size of the mobile phone data can be very large, it is usual that the data from only one mobile phone network provider are available when there are several mobile phone network providers in the survey area. This makes it difficult to expand the sample to the population. If the data are aggregated by the mobile phone network provider when the data are provided to the researchers because of privacy issues or other reasons, it is more difficult to for researchers to check the data quality. The penetration rate of mobile phones is not 100%, and the sampling scheme is not known. Some people carry multiple mobile phones, such as business and private phones, so the number of mobile phones cannot be simply converted to the number of persons. All of these factors affect the representativeness of the mobile phone data.

5. Bluetooth

Bluetooth is a short-range telecommunication protocol used by small handheld devices (smart phones, tablets, computers, headsets). As a matter of fact, these devices can be used as probes to collect transportation data if they are carried by people or within vehicles. It is mainly used to collect speed and travel time data on road segments, but can also be used to collect origin-destination data.

5.1. How does it work?

Bluetooth is a wireless technology invented in 1994. An ultra-high frequency signal between 2.400 and 2.485 GHz is used to exchange data between short distances. Each Bluetooth device has a unique identification code that is emitted regularly to airwaves. Any Bluetooth device in a close range of another can read its identification code. This principle can be used to capture devices that travel, for example, on a road segment. By looking at different waypoints through the road network, we can "follow" individual devices. In practice, in dense areas, there can be hundreds of devices in front of the capturing device during a short time period.

5.2. What can we do with the data?

Bluetooth data can be used to measure travel time in a closed road segment from point A to B. Due to the huge quantity of data, measurement can be very precise and ventilated through periods of time, e.g., days, weeks, etc. Bluetooth data can also be used in a wider road network to determine origin-destination matrices because of the ability to spot-check the presence of devices at any sensor location.

5.3. Data cleaning

As stated before, many challenges are related to Bluetooth use. The sensors may not be able to capture all the devices because of its capacity. This becomes more severe when the number of devices increases during peak periods because matching IDs becomes computationally harder with the number of devices. There can be double counting (that must be addressed) of the presence of weak signals or parasite signals (i.e., permanent installation on

the side of the road). In consequence, the capturing system may not be able to follow all devices at once. This "probability" of capturing a device is multiplied by the number of sensors, and can reduce the ability to capture travel time and paths. There is also the challenge of imputing travel times without knowing whether 'the device' stopped and engaged in activity between the two successive matches. However, a significant number of devices can be traced, and it may be sufficient to capture the measurement that is needed (travel time, origin and destination).

5.4. Sampling

Like in other passive data streams, sample representativeness can be weak with Bluetooth devices. The possession rate of such devices is not well known. In addition, many devices can be found in a car (car GPS, car telephony system, smartphones, tablets, computers, etc.). The ability to expand data collected to evaluate flows is not trivial and must be further studied.

6. Smart card

Smart card automated fare collection systems are now widely adopted as a mode of payment in most public transport authorities across the world. These systems collect data on every transaction made over the public transport network, day after day. Apart from revenue collection (the true role of these systems), smart card data can be very useful for transport planners because they track the behavior of travelers on a continuous basis. This applies also to other modes of automated payment, including mobile phone payment, that will soon or later complement if not replace smart cards. However, there are still many challenges to solve before using this kind of data in a transparent manner. This section addresses some of these issues.

6.1. How does it work?

Smart card fare collection systems are based on three components: the card, the readers and the central server. At first, transit users acquire fares by different means (ticket booth, vendor, web-based membership, etc.). Fares are then registered on the card (or pre-registered). In some cases, the smart card is used as out-of-pocket money, but the principle remains the same. The second components, the readers, are put aboard vehicles or at station entries and exits. When the user flags his card, the readers exchange information with the card (identification, fare debit) and transfer the transaction data to the third component, the central server that collects all the data on transactions through the system. Transfer is usually asynchronous, for example, buses will download the transactions at the end of the day.

6.2. What can we do with the data?

Raw smart card data can be used to measure the ridership on specific elements of the transit network: buses, subway and train stations, and routes. Of course, the data are used to manage the revenue collection in the system. But more can be done with data, as demonstrated in recent works, including calculating network key performance indicators, analyzing the patterns of transit users, evaluating the influence of weather on ridership, estimating origin-destination matrices for the transit network, etc. (Pelletier et al. 2011). During the workshop, four applications of smart card data were presented and discussed. Bouterfif et al. (2014) estimated the microelasticity of service changes using individual smart card data. It was stated that there is the challenge related to the development of indicators and the modeling of individual behavioral changes with smart card data. Chu (2014) analyzed the longitudinal patterns using a long-term dataset. It was found that systematic card replacement and withdrawal have a significant impact on sample size and introduce spatial and temporal biases. Spurr et al. (2014) evaluated the accuracy of large household surveys with smart card data transformed in "diary data". An algorithm for matching respondents of household travel surveys to smart card data was developed. The matching process considers multiple aspects of travel behavior including departure time, detailed transit itinerary and the regularity of individual trip patterns over multiple days, and half of the respondents of household surveys could be matched to smart card data in the experimental analysis. Munizaga (2014) validated smart card data from external sources. The results suggest that

the accuracy of determining alighting locations is approximately 70%, and the accuracy for determining origins and destinations is 80% and 60% with the threshold of 1 km, respectively.

6.3. Missing or incomplete data

Even though smart card data collection is quasi-universal in a sense, there are still some elements to be considered to ensure complete information:

- many systems do not provide the "tap-out" (exiting) information, so there is a need to apply a destination estimation algorithm;
- some systems do not record the location of the transactions, only the date, time and vehicle information, so data must be combined to automated vehicle location (AVL) data or other sources;
- in some networks, all users may not have smart cards, so data will not represent the universe, but a subset that has to be measured;
- as in any information system, smart card data can contain errors such as missing values, inaccurate device identification or GPS positions, etc.

6.4. User identification

For structural or privacy reasons, no information of users is available in smart card system datasets. The lack of socio-demographic attributes limits the ability to use smart card data in econometric models or to analyze trip purposes, demography of membership, etc. However, like for other indirect and passive data streams, methods can be proposed to derive the home location of users, or other attributes, such as age or working/studying status (looking at the type of fare). For longitudinal datasets, card replacement or loss can cause disruptions in the temporal sequence of transactions, when looking at cards individually.

6.5. Data management

Typically, smart card systems register millions of transactions per day in larger public transit networks. The huge quantity of data can be a challenge to manage. Some operators may be tempted to aggregate or erase the data after a certain amount of time. In addition, the processing time of such data quantities can be counted in hours, especially for more advanced data mining techniques. Thus, the technical challenges have to be examined before using smart card data in the day-to-day planning process. Finally, external datasets of transit network geometry and schedule must be synchronized to smart card data to be able to exploit their full potential.

7. Adding socio-demographic information

Most of the planning models and methods require the use of attributes on trips, such as trip purpose, trip mode and socio-demographic attributes on the traveler. Obviously, this information is not available in passive data streams, so further methods are needed to be able to derive additional attributes on trips. In most cases, origin-destination data could be used to derive the home location of the traveler. For example, with smart card data, we can suppose that the first trip of the day starts from a stop or a station near the home location, given some hypothesis on time of day and the location of residential neighborhoods (people may work night shifts or have break schedules). Then, the average attributes extracted from censuses can be linked to these users (average income, average age, etc.).

In some studies involving smart cards, fare type has been used to distinguish age, because student and elderly people may have different fares. This exercise is much more difficult for other data sources (mobile phone, Bluetooth, GPS), because we usually cannot have any information on the owner of the device. For GPS, mobile phone and Bluetooth traces, the trip mode can be derived using a series of logic related to travel speed, location and readings from accelerometers. The last "hope" to add socio-demographic information may reside in the ability of matching passive streams with home or passenger survey data. Directly requesting the number of the device or the

card whose data can be retrieved will help in the matching process. However, in most jurisdictions, legal restrictions will forbid this action.

8. Added value of using passive data streams in transport planning and modeling

As previously mentioned, passive data streams can be very helpful to obtain continuous and rich datasets of observations on travel behavior of transit passengers, car drivers, and travelers using other modes, such as bikers and walkers. Even though socio-demographic data are not easily available, passive data streams can:

- enhance the understanding of travel patterns (demand) and service level (supply);
- reduce time and cost for analysis (avoiding long and costly surveys);
- see the dynamics of behavior, its distribution, its trend (compared to cross-section surveys);
- be used to manage or analyze real-time operation in the case of special events, incidents, etc.

In addition, data providers can commercially benefit from the use of their data, or can justify their collection. Finally, this data could be used by police and justice officers to resolve crimes by analyzing the spatial and temporal location of individuals.

9. Discussion and conclusions

The discussions and presentations of this workshop have shown the huge potential of using passive data streams for transport planning. The apparent weaknesses of these data (lack of socio-demographic data) are counterbalanced by the richness of the continuous and numerous data records that can be obtained. However, some important issues remain to be addressed properly and are cited here as topics for further research endeavors.

- *Collection*. Although passive data streams have proven useful for transport planning, they are still difficult to obtain. Privacy and legal questions often block access to data streams, or data processing rules impose data aggregation or synthesis that makes data unusable.
- *Data mining*. Even with today's computer standards, processing very large quantities of data is still a challenge. Data mining techniques are not adapted for transport realities (space and time analyses, pattern detection). Further innovations are needed in this field.
- *Preprocessing and enhancing*. Before even thinking of using data for transport planning, there is sometimes a need to preprocess data. For example, smart card data can be enhanced with a destination estimation algorithm, in the case of "tap-in" only systems. The same applies to GPS data, where there is a need to derive trip modes. These methods still require improvement.
- *External validation.* Due to the lack of socio-demographic data, it is sometimes difficult to validate the models, analyses and matrices obtained from passive data because there is no easy way to compare the results with household surveys or other means of direct observation. Further innovations are needed in data fusion and comparison.
- *Real-time availability.* At this time, passive data streams in transport are not obtained and processed on a real-time basis. Data are collected asynchronously and often analyzed for months before any models and results can be obtained. There is a need to propose tools aimed to process huge quantities of data in "short" time, providing useful results for transport planners.
- *Reinventing transport planning.* This last challenge is probably the greatest. Transport planning models are based on the availability of partial datasets, small-sampled household surveys, with coarse spatial definitions (some methods are still working with TAZ). Now that increasingly more passive data streams are being made available, how will this change the way we conduct transport planning? There is a greater need to perform microsimulations and develop agent-based tools capable of integrating these data and literally "replaying" the urban mobility all at once. Developing prediction models at the individual level and evaluating the microelasticity of the population remain as significant yet necessary challenges.

Acknowledgements

The workshop participants were Patrick Bonnel (FR); Matthew Carrigan (AU); Premysl Cech, (CZ); Chongsuk Cho (KR); Ka Kee Alfred Chu (CA); Adrian Ellison (AU); Richard Ellison (AU); Houshang Farabi (AU); Sergio Daher Garcia (AU); Qian Ge (JP); Jane Gould (US); Roniga Hang (AU); Yusuke Hara (JP); Jeffrey LaMondia (US); Wen Liu (AU); Robin Lovelace (UK); Frank Milthorpe (AU); Marcela Munizaga (CL); Hannelie Nel (ZA); Minchoul Park (KR); Virpi Pastinen (FI); Louis Potgieter (ZA); Timothy Spurr (CA); Martin Trépanier (CA); Chester Wilmot (US) and Toshiyuki Yamamoto (JP).

Appendix A. Papers presented during the workshop

How reliable is the information obtained from passive data? Lessons learnt from validation and application experiments. Author: Marcela Munizaga.

Two-year worth of smart card fare validation data – extracting longitudinal observations for the understanding of travel behaviour. Author: Ka Kee Alfred Chu.

Potential of « passive » mobile phone dataset to construct origin-destination matrix. Authors: Patrick Bonnel, Zbigniew Smoreda, Etienne Hombourger and Ana-Maria Olteanu-Raimond.

Behavior Analysis using Tweet Data and Geo-tag Data under Natural Disaster. Author: Yusuke Hara.

Appendix B. Posters associated with the workshop

Harnessing smartphone sensors for tracking location to support travel data collection. Authors: Richard Ellison, Adrian Ellison, Stephen Greaves and Chris Standen.

The impact of GPS based outdoor activity detection on smartphone battery life. Authors: Kalan Nawarathne, Fang Zhao and Francisco Pereira

Battery Patterns and Forecasting in a Large-scale Smartphone-based Travel Survey. Authors: Rudi Ball, Kalan Nawarathne, Rui Baltazar, Francisco Câmara Pereira, Chris Zegras and Moshe Ben-Akiva.

Automating Image Processing for Life-logging Cameras to Support Travel Data Collection. Authors: Li Shen and Peter Stopher

How to improve GPS post-processing with additional information? Authors: Thi-Huong-Thao Pham, Philippe Marchal, Jean-Loup Madre and Jimmy Armoogum.

Introducing a clustering algorithm to derive activity and trip diaries using Global Positioning Systems data: A pilot project in South Africa. Authors: Stephan Krygsman, Hannelie Nel and Klaas van Zyl.

References

Bonnel, P., Smoreda, Z., Hombourger, E., Olteanu-Raimond, A.M., 2014. Passive mobile phone dataset to construct origin-destination matrix: potentials and limitations, 10th International Conference on Transport Survey Methods. Leura, Australia.

Bouterfif, S.E., Trépanier, M., Morency, C., Chu, A.K.K., 2014. Can we estimate the microelasticity of public transport use from smart card data?, 10th International Conference on Transport Survey Methods. Leura, Australia.

Chu, A.K.K., 2014. Two-year worth of smart card transaction data – extracting longitudinal observations for the understanding of travel behaviour, 10th International Conference on Transport Survey Methods. Leura, Australia.

Escobar Garcia, D., Cadenagaitan, C., Restrepo Cadavid, J.M., 2014. ICTS-based accessibility analysis as platform for sustainable mobility projects, 10th International Conference on Transport Survey Methods. Leura, Australia.

Ge, Q., Fukuda, D., 2014. Work/school-related travel demand estimation from aggregate datasets, 10th International Conference on Transport Survey Methods. Leura, Australia.

Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L., 2008. Understanding individual human mobility patterns. Nature 453, 779-782.

Hara, Y., 2014. Behavior Analysis using tweet data and geo-tag data under natural disaster, 10th International Conference on Transport Survey Methods. Leura, Australia.

Lovelace, R. 2014. Estimating energy use in transport from travel surveys: current techniques and future prospects, 10th International Conference on Transport Survey Methods. Leura, Australia.

Morency, C., 2013. Chapter 39. Workshop synthesis: exploiting and merging passive public transportation data streams, in *"Transport Survey Methods: Best Practice for Decision Making"*. In: Zmud, J., Lee-Gosselin, M., Munizaga, M. and Carrasco, J.A. (Eds.). Emerald Group Pub Ltd, UK, pp. 716-720.

Munizaga, M., 2014. How reliable is the information obtained from passive data? Lessons learnt from validation and application experiments, 10th International Conference on Transport Survey Methods. Leura, Australia.

Murakami, E., Wagner, D.P., 1999. Can using global positioning system (GPS) improve trip reporting? Transportation Research Part C 7, 149-165.

Pelletier, M.P., Trépanier, M., Morency, C., 2011. Smart card data use in public transit: a literature review. Transportation Research Part C: Emerging Technologies 19(4), 557-568.

Stopher, P., Sneade, A., 2013. Chapter 43. Workshop synthesis: post-processing of spatio-temporal data, in *"Transport Survey Methods: Best Practice for Decision Making"*. In: Zmud, J., Lee-Gosselin, M., Munizaga, M. and Carrasco, J.A. (Eds.). Emerald Group Pub Ltd, UK, pp. 769-779.

Spurr, T., Chu, K.K.A., Chapleau, R., Piché, D., 2014. A smart card transaction "travel diary" to assess the accuracy of the Montréal household travel survey, 10th International Conference on Transport Survey Methods. Leura, Australia.

Yamamoto, T., 2014. An analysis on willingness to provide personal vehicle travel information using GPS device, 10th International Conference on Transport Survey Methods. Leura, Australia.

Zito, R., D'Este, G., Taylor, M.A.P., 1995. Global positioning systems in the time domain: how useful a tool for intelligent vehicle-highway systems? Transportation Research Part C 3(4), 193-209.