Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 11 (2015) 247 – 262

Transportation Research Procedia

www.elsevier.com/locate/procedia

10th International Conference on Transport Survey Methods

Automatic trip and mode detection with MoveSmarter: first results from the Dutch Mobile Mobility Panel

Karst T. Geurs[a]*, Tom Thomas[a], Marcel Bijlsma[b], Salima Douhou[c]

[a] *Centre for Transport studies, University of Twente, PO Box 217, 7500AE Enschede, the Netherlands*
[b] *Mobidot, Hengelosestraat 511, 7521AG Enschede, the Netherlands*
[c] *Centre for Comparative Social Surveys, City University London, Northampton Square, London EC1V 0HB, UK*

Abstract

This paper describes the performance of a smartphone app called MoveSmarter to automatically detect departure and arrival times, trip origins and destinations, transport modes, and travel purposes. The app is used in a three-year smartphone-based prompted-recall panel survey in which about 600 smartphone and non-smartphone owners participated and over 18,000 validated trips were collected during two weeks. MoveSmarter is concluded to be a promising alternative or addition to traditional trip diaries, reducing respondent burden and increasing accuracy of measurement, but there is room to improve trip and mode detection rates and the efficiency of battery consumption.

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

Keywords: mobile mobility panel; smartphone data collection; trip detection accuracy; mode choice detection

1. Introduction

In most countries, including the Netherlands, the understanding of people's travel behaviour is based on cross-sectional travel surveys where only one day is surveyed for each respondent in 'representative' periods when traffic flows are maximal (Ortuzar et al., 2010). This is not enough to gain a proper understanding of the dynamics in travel behaviour. More specific, cross-section data do not give any information to ascertain how choices will vary over time if the system changes. Moreover, multi-day travel behaviour data (collection using GPS-devices) show a

* Corresponding author. Tel.: +31 534891056; *E-mail address:* k.t.geurs@utwente.nl

Peer-review under responsibility of International Steering Committee for Transport Survey Conferences ISCTSC
doi:10.1016/j.trpro.2015.12.022

strong variation in travel behaviour (Stopher and Zhang, 2011). It has been shown that people are to visit new places even after several months of monitoring (Schönfelder and Axhausen, 2010).

Dynamic information on travel behaviour can be obtained by asking respondents to register their trips during more than one day. Although this might be possible for small samples, it is not a realistic option for large national travel surveys as this implies a high respondent burden. In the literature, it is suggested that automatic detection of trips might be the solution for this problem, especially when GPS and GSM technologies in smartphones are becoming standard (Stopher, 2009, Nitsche et al., 2012). Kracht (2004), for example, showed the great potential of GPS and GMS in monitoring travel behavior. An additional advantage of automatic detection is the possibility of the registration of trips that respondents forget to report in traditional surveys. Survey-based methods have an inherent and downward bias in trip detection which can be significantly reduced using automatic detection. In the literature, underreporting of 10% up to 80% of car trips is documented, when GPS traces are compared with travel diaries (Schönfelder and Axhausen, 2010). At the same time, there are also errors and omissions in GPS and smartphone measurements. In addition, prompted recall surveys are also not necessarily error-free (Feng and Timmermans, 2014).

In the literature, a number of studies are described in which smartphone apps are used to detect trips and trip characteristics such as transport modes (e.g., Reddy et al., 2010; Nitsche et al., 2013; Shin et al., 2015). A number of other studies describe apps aiming to detect trips and provide users with feedback on their travel behaviour (e.g., Fan et al., 2012; Li et al., 2011; Bie et al., 2012; Prelipcean et al., 2014). These studies describe apps in the prototype stage and which have not (yet) tested the accuracy of the tool in large-scale field trials. One exception is the smartphone-based prompted-recall survey developed and deployed in Singapore as part of a subset of a national household travel survey (Cottrill et al., 2013). Furthermore, to the authors' knowledge, smartphones have to date not been used in longitudinal studies in travel behavior. This paper describes the first results from the Dutch Mobile Mobility Panel project. The project tries to answer two main research questions: (1) Are smartphones an effective and efficient tool for trip registration in order to monitor individual travel behavior during a long period of time? (2) What is the variation in travel behavior over time, and which external factors (such as weather) do influence this variation? To answer these questions, a field study is carried out with respondents recruited from the Dutch LISS (Longitudinal Internet Studies for the Social Sciences) panel. This is a non-commercial panel which can be only used by academic researcher and policy makers. For about 600 respondents from the panel, smartphones are used to detect trips automatically during several weeks in the period April – June in 2013, 2014 and 2015. The size of our study is quite unique compared to the studies in the literature.

This paper describes the data collection methodology of the Mobile Mobility Panel project (Section 2) and analyses the accuracy of trip rate detection (Section 3) and mode detection (Section 4), using the data from the 2013 wave. Furthermore, we describe the results of a user evaluation conducted after the fieldwork period (Section 5) Participants were asked to express their opinion about different aspects of the experiment, e.g. the app, accuracy of the trips and mode detections, smartphone use, accuracy of trip corrections, battery use and the prompted recall website. We use the results to explain biases in trip detection (Section 6). Finally, Section 7 presents the conclusions.

2. Methodology

2.1. Overview

The trip registration consists of three parts: (1) automatic trip registration with the smartphone application MoveSmarter (for iPhone and Android), (2) update of the detected trip characteristics on a back-end service, and (3) an internet-based prompted recall survey. In the prompted recall, respondents are asked to check, revise if needed, and approve the MoveSmarter trip detections. For each day of the fieldwork, they can adjust, add or delete trips on a webpage, after which they need to approve the trips for that particular day. In the remaining part of the paper, we call these *reported* trips (consisting of unadjusted MoveSmarter trips, adjusted MoveSmarter trips, and added trips). The automatically detected trips (after processing on the back-end server) are called *MoveSmarter* trips. In addition, after all trips of a fieldwork day have been approved by the respondents, they are asked if there were specific personal or transport related circumstances which affected their travel during that day, e.g. sickness or major disruptions in public transport.

Figure 1 visualizes this approach, and the next sections explain the approach in detail. Section 2.2 describes the automatic trip detection process. Section 2.3 the internet-based prompted recall survey, and section 2.4 the recruitment method and characteristics of the sample.

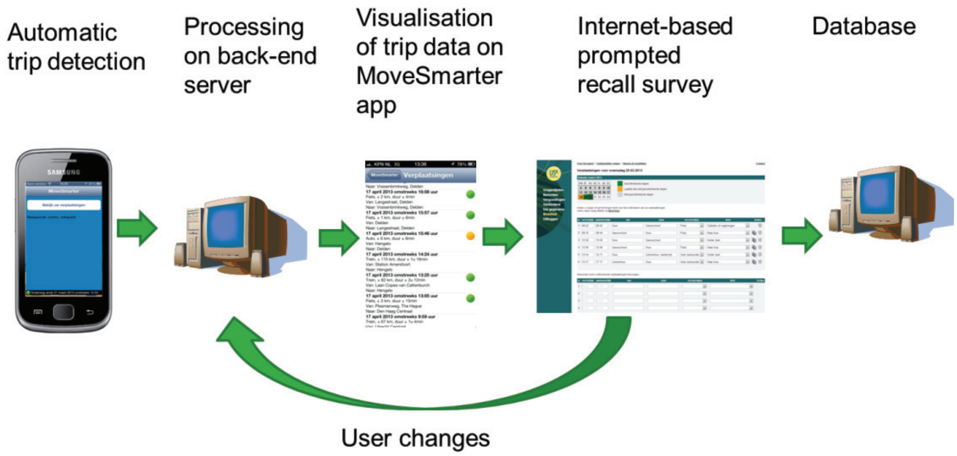

Fig. 1: Schematic illustration of the Dutch Mobile Mobility Panel data collection methodology

2.2. Automatic trip detection with MoveSmarter

The automatic trip detection process consists of two stages. The first stage takes place on the smartphone of a user via the installation of a dedicated app. This app allows end-users to have a visual overview (‘logbook’) of their travel behaviour. The measurement capabilities are created by a sensing module integrated in the app. This module also manages authentication and communication with the MoveSmarter back-end to perform measured data analysis in the second stage of the process.

The sensing module is an intelligent software component that uses an array of available sensors in the Smartphone (GPS, WiFi, Accelerometer and cell-ID information) to automatically sense trip start, movements and trip end. In this way for each trip a coherent trace of GPS locations is collected (a ‘Raw Trip’) without any user involvement. The sensing module runs as a background process on the Smartphone and restarts itself at operating system (iOS, Android) booting. With the current generation of smartphones there is a clear trade-off between measurement accuracy and battery consumption. In some GPS-based studies, GPS measurements are done every second (Feng and Timmermans, 2013). This level of accuracy is not possible using smartphones as it would drain the battery too much during the course of the day. Here, we aim for 24x7 sensing with normal Smartphone use throughout the day. A number of battery saving strategies are deployed to achieve this objective. First of all, the concept of ‘static’ versus ‘en route’ is used to prevent excess battery use by the GPS sensor. Only when the sensing module detects significant location changes, the GPS sensor is triggered and the App becomes ‘en route’. GPS sensing is automatically stopped when the sensing module detects that the Smartphone no longer moves (App becomes ‘static’). During the ‘en route’ mode GPS positions are collected with a frequency of approximately once per 2 seconds leading to typical accuracies between 3 and 30 m depending on speed of the Smartphone and GPS accuracy. Secondly back-end communication is kept to a minimum. Upload is only performed when a trip has been marked as ended by the sensing module and communication with the MoveSmarter back-end is possible. Otherwise the Raw Trip is cached in the Smartphone.

After upload, the Raw Trip is processed in this back-end. The processing includes filtering, cleaning, map-matching (using the open source OpenStreetMap network) and data enrichment. Also via interpolation and averaging routines, the quality of the data is further improved. An illustration of this process is shown in Figure 2.

a b

Fig. 2: (a) An example of the first stage of data processing (raw GPS data) of a measured round trip (trip length ~900 m), (b): Result of the same round trip after the second stage of processing with cleaned and map matched data.

Transport mode deduction takes place using Bayesian probability statistics taking into account: i) speed patterns, ii) sensor data characteristics, iii) infrastructure network (i.e. location of road, rail, water and air infrastructure), iii) public transport information (i.e. location of public transport stops), and iv) personal trip history. Especially for shorter trips, where less measurement data is available, or more densely populated areas, where more mode options are available, this mode deduction is consequently more prone to errors. Furthermore, in combination with advanced trip splitting algorithms and reprocessing unimodal trips are generated with properties such as start location, end location, start time, end time, trip duration and modality. An example of a typical automated trip splitting decision is shown in Figure 3, illustrating a respondent standing still close to a bus stop.

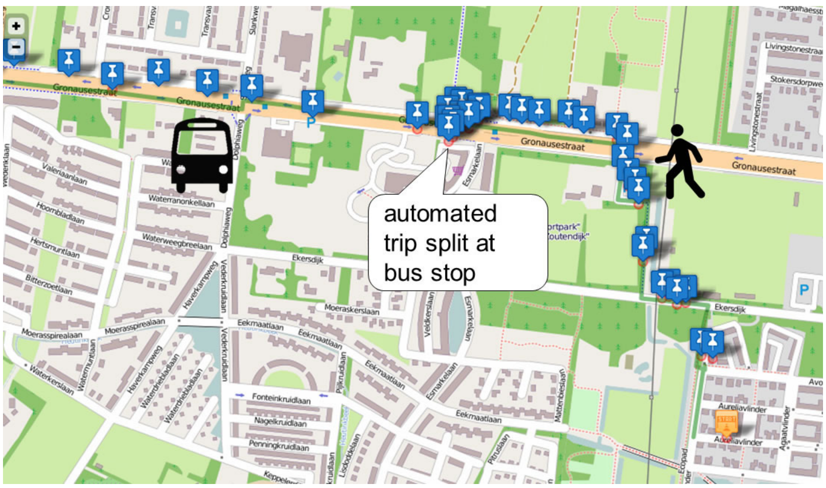

Fig. 3: An example of unimodal trip detection

Given the trace characteristics before arriving at the bus stop and after leaving the bus stop and the duration of inactivity of the respondent, the door-to-door trip shown in Fig. 3 is divided in two single mode trips for this particular case. For convenience, we refer to the unimodal trips as trips in this paper, unless mentioned otherwise.

Finally, learning algorithms deduce frequent routes and places and assign trip motives (e.g. a recurrent home-work trip).

2.3 Internet-based prompted recall survey

CentERdata, the owner and manager of the LISS panel, has developed a webpage where respondents can (re)view their daily trips. Per day, trips are shown using the following fields: departure and arrival time, origin and destination (by street and city) mode, and goal of trip. The respondents are encouraged to check the MoveSmarter trips regularly, preferably every day, and make changes where necessary. Departure and arrival time can be adjusted, both at the level of hours (hh) and minutes (mm). The fields that include location can be adjusted freely. As the location of an automatic trip registration is not always recognized (e.g. the address of a supermarket may be unknown to the respondent), the respondent can adjust this field and relate it to the activity at the location (e.g. doing groceries at a supermarket). Examples are “home”, “work”, and “supermarket”. The string value provided by the respondent is automatically used by the back-end server of MoveSmarter when the respondent visits that location the next time. The mode and goal of the trip can be selected from dropdown menus. These have been harmonized with the terminology for mode and goal of trips used in the Dutch National Travel Survey (NTS) (Statistics Netherlands, 2014).

The respondent can also delete trips completely (that were falsely detected by MoveSmarter), add trips (that were not detected by MoveSmarter), combine multiple trips (that were falsely detected as separate trips by MoveSmarter) or split a trip (that was falsely detected as one trip by MoveSmarter).

When the trips of a particular day are adjusted (where necessary) and confirmed on the webpage, the respondent needs to complete a short questionnaire. In this questionnaire, the respondent can indicate if special circumstances have occurred during that day (e.g., extreme weather, road works that caused delays, delays in public transport, and personal circumstances like illness or travelling with other persons), and if it influenced their mobility pattern of that day. Answers to these questions will provide possible insight in the variability in mobility, which is topic of future research.

Finally, there are two other online questionnaires, each respondent is asked to fill in at different points in time during fieldwork. The first questionnaire is a travel diary on the first day of the study. The results from this diary can be compared with the reported trips based on MoveSmarter registrations as the app is activated on this first day already but respondents are only asked to check trips on the webpage after this first day. In particular, the rate of reporting and expected improvement herein due to automatic detection will be further described in an extended study on trip generation. The second questionnaire is the respondents’ evaluation at the end of the study. Results from this evaluation will be reported in section 7 of this paper.

2.4 Recruitment and sample

Respondents of the Mobile Mobility Panel are part of the LISS panel (Scherpenzeel and Das, 2011). The LISS panel consists of more than 5,000 households and in total over 8,000 respondents. The LISS panel provides a true representative sample of the Dutch population and provides households without internet access cost-free equipment and internet connection. Leenheer and Scherpenzeel (2013) show that including non-internet households in the LISS panel increases the representativeness of the panel. In addition, Scherpenzeel and Bethlehem (2011) show that the LISS panel, in terms of representativeness, is close to a traditional face-to-face study. The reader is also referred to De Vos (2010) for additional information regarding the representativeness of the LISS panel. LISS panel respondents receive a monetary reward for each completed questionnaire. For the purpose of this research, at least 500 participating respondents were needed. From a random subsample of LISS panel respondents about 800 LISS respondents expressed an interest to participate in the study and consented to have their data temporarily stored at a third party in order to be able to run the trip detection analyses for the MoveSmarter app. Of these 800 respondents, 655 respondents actually started the experiment and more than 50 respondents stopped immediately after the start of the study or after some time due to time constraints or problems related to the (installation of the) MoveSmarter app. These respondents were not further considered in the analyses. Another 45 respondents stopped because they were immobile due to injury or disease, or because they were on a holiday. These respondents are counted in the statistics as they (contrary to the other group) did not make trips (in the Netherlands) during (a part of) the measurement period and therefore represent the immobile part of the population.

Within the LISS panel, about 36% has a smartphone (measurement 2013). Respondents without a smartphone or a smartphone that is not supported by iOS or Android were provided with an Android loan smartphone (Samsung Galaxy Gio). Of the slightly more than 550 respondents that participated during the whole experiment, about 59% used the loan smartphone, 24% owned a smartphone supported by Android, and 17% owned a smartphone supported by iOS. To reduce the costs of the data collection, the smartphones were distributed in two batches. In 2013, the first batch of the fieldwork was between Monday April 29 and Sunday May 12, and the second batch was between Monday June 17 and Sunday June 30. As the first day is used for the online travel questionnaire, the actual MoveSmarter measurement period started on Tuesday April 30 and Tuesday June 18 respectively.

Table 1 shows the distribution of personal characteristics of the respondents. The table shows that the distribution from the panel by gender, age, education and income corresponds well with that from the whole LISS panel and the Dutch population. This is also true for other characteristics such as position in the household, main activity, and urbanization level of the residence (available upon request). From this we conclude that we used a representative panel for this research. Differences in car ownership can unfortunately not be examined, as this only available for about 45% and 40% of the respondents and the LISS panel, respectively.

Table 1. Personal characteristics from respondents and all members of the LISS panel.

	respondents	LISS panel	Dutch population (2013)
N	655	~6,400	
Gender			
Male	53%	46%	50%
Female	47%	54%	50%
Age			
15-24	10%	13%	15%
25-34	14%	12%	15%
35-44	19%	16%	17%
45-54	18%	18%	18%
55-64	22%	19%	16%
65 years and older	17%	22%	20%
Education			
Low	26%	34%	33%
Middle	34%	35%	40%
High	40%	31%	27%
Net income (in €)			
0-10,000	23%	31%	20%
10,000-20,000	34%	33%	23%
20,000-30,000	30%	26%	19%
30,000-40,000	10%	7%	15%
40,000-50,000	2%	2%	10%
More than 50,000	1%	1%	14%

*The sample size differs per background variable due to missing values.

3. Trip detection rates

In total 25,605 trips were reported in the dedicated webpage, and 22,898 trips were detected by MoveSmarter. About 10% of the trips were, however, not included. Some of these trips were made by persons that dropped out during the study. Other trips were not formally approved with the follow-up questionnaire (about specific circumstances) and hence were not included. Excluding these cases, Figure 4 shows the number of trips per person per day for both the reported and MoveSmarter trips. The figure confirms the clear underestimation of MoveSmarter detections which is on average slightly higher than 20%. This means that a relatively small but significant number of trips were added in the webpage. The rate of missing MoveSmarter detections could be even close to 25% if we consider that 5% of the MoveSmarter trips were deleted in the webpage by respondents. However, some of these trips were first removed and later added again, perhaps because they considered this easier than adjusting the information already presented. Therefore, we conclude that the rate of missing MoveSmarter detections (compared to the reporting) lies somewhere between 20 and 25%.

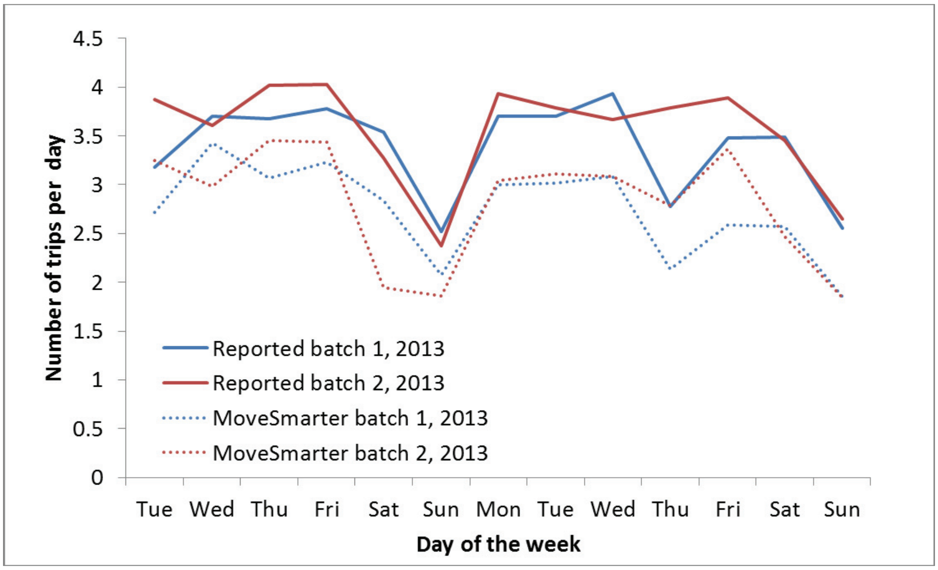

Fig. 4: Average number of trips per respondent per day

Figure 4 shows some interesting features. The day-to-day variation is more or less similar for the reported and MoveSmarter trips which is an encouraging sign. Both batches show some clear dips in the weekends. Especially on Sundays, trip rates are relatively low. These results are in line with those from the Dutch National Travel Survey data. Trip rates are also relatively low on specific days of the measurement period. Tuesday April 30 was "Koningsdag" which is a national holiday in the Netherlands, and the first week of the first batch coincided with a school holiday. This could explain the somewhat lower trip rate in the first week of the first batch, especially on Tuesday. Thursday May 9 was Ascension Day, which is for many people followed by a day off. This could explain the extra dip in the second week of the first batch. If we would consider normal business days the trip rate is above the 3.8 trip per person per day. Taking into account that about 5% of these trip stages are egress or access stages for public transport, the corresponding door-to-door trip rate is about 3.4 trips per person per day (3.6 for working and 3.0 for weekend day). Although a direct comparison can not be done (given differences in sample size, measurement period etc.), trip rates (per person) appear to be clearly higher than those from Dutch national travel surveys. The average number of reported trips per person per day is significantly higher than in the 2013 Dutch National Travel Survey (Statistics Netherlands, 2014) with 2.6 door-to-door trips per day and also higher than the 2013 wave of the Mobility Panel for the Netherlands (MPN) (Hoogendoorn et al., 2014) with 3.1 door-to-door trips per person per day (based on over 3,850 respondents of 15 year and older). The results indicate that underreporting may be significant in traditional travel surveys relying on a single or multiple-day measurement. Schönfelder and Axhausen (2010) describe the variation in travel behaviour using several multi-week data sets (GPS-based or travel diary). Using data from a 6-week travel survey in Switzerland, respondents visited on average 0.3 new (never previously visited) locations per days during a 6-week monitoring period. More than half of the previously never or seldom-visited locations were leisure destinations. Schönfelder and Axhausen (2010) conclude that it takes about five to ten weeks of monitoring to gain relative certainty about individual destination-choice preferences. The bias due to underreporting is probably the largest for infrequently made trips (e.g., trip to the dentist) and smallest for regular trips (e.g. commuting).

Underreporting is also a problem in our smartphone-based measurements. There is some tentative evidence for this when trip rates are compared based on results from the users' evaluation (see sections 6 and 7). More direct evidence for underreporting follows from mismatches between destination and origin location of successive trips. These should be the same while in about 7% of the cases the arrival location of the previous trip appears to be different from the departure location of the next trip. This would indicate that unreported trips were made between these locations. If we would impute 7% of extra trips to account for these mismatches, trip rates would be close to 4 trips per person per day.

Aforementioned results suggest that automatic trip detection may result in the registration of trips that would otherwise not be reported. To tackle the problem of underreporting, it is crucial that almost all trips can be automatically detected. Therefore, it is important to look at possible explanations for the missing MoveSmarter detections. Can trip detection be improved by better algorithms, or is the quality of trip detection mainly restricted by the users and the way they use the smartphone, or is a combination of all of these factors?

In Figure 5, we show the cumulative distribution of the number of days respondents make trips. This figure only includes respondents that reported at least one trip during the measurement period.

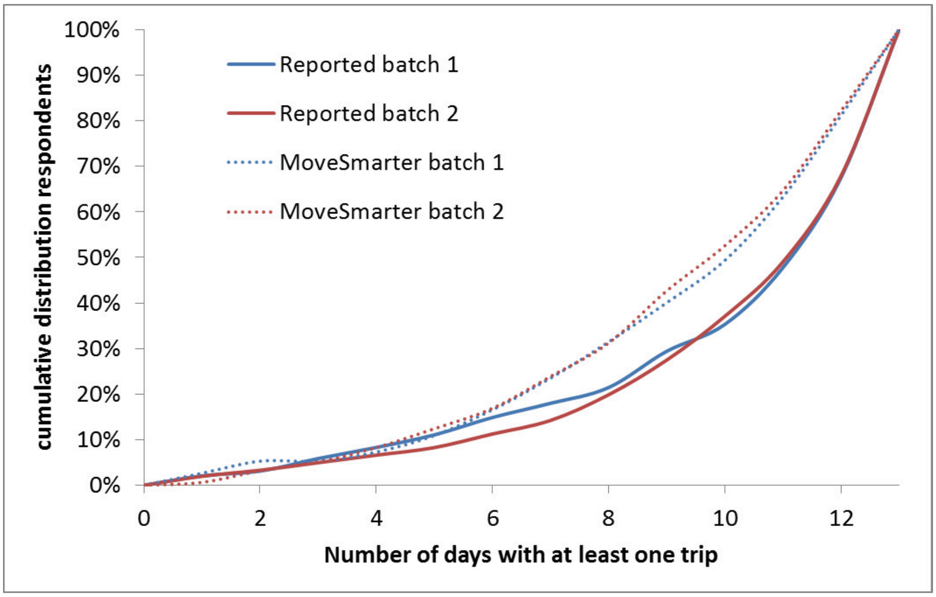

Fig. 5: Cumulative distribution of respondents by number of days with at least one trip

The figure shows that quite some respondents are not mobile during many days. On average, 13% and 21% of the respondents did not make trips during working and weekend days, respectively. Although we cannot be sure which share of the no-trip days are due to underreporting or really due to respondents being immobile, underreporting does not seems to be problematic. According to the 2013 NTS and MPN surveys, about 15% and 25% of the Dutch population is not mobile during working and weekend days, respectively. The respondents in our sample are thus more mobile during the measurement period. This could indicate that multi-week smartphone detections (combined with a prompted recall webpage) can help in getting more accurate trip rates compared to the one or three day trip diaries used in the NTS and MPN, respectively. Also, the results do indicate that the issue of missing detections can probably not be attributed to problems related to limited battery capacity. Otherwise one would only expect missing detections on days that many trips are being made. The percentage of missing detections is also not larger for respondents that make many trips (see also section 7).

As battery constraints do not appear to be the only cause for missing MoveSmarter trips, we extend the analysis by comparing the rates of missing detections for different distances, travel times, activity times (before and after the trip), modes, and smartphone types. We also checked whether the rate of missing detections is higher for the first or last trip of the day. The results can be summarized as follows. Rates of missing detections are significantly higher for very short activity times, both before and after the reported trip, for very long trips (both in distance and travel time), for public transport trips, for the last trip of the day, and for users with an Android loan smartphone. Even if we would not consider these specific circumstances, missing MoveSmarter trip rates are still up or beyond the 10%. It is therefore questionable whether an automatic trip detection app would be able to detect all trips at this point in time.

The high rate of missing trips when period of activity at the destination is very short (about 5% of all trips) can be contributed to the fact that some of these trips may not be recognized as separate trips. That trips are short is in itself not determinative. It appears that the sensing of smartphones is sensitive enough to detect short distance trips, although mode detection for short distance trips is an issue (see Section 4). Significant though is the underdetection of trips longer than 90 minutes or 100 km. These missing detections might be related to battery constraints as very long trips have a large effect on battery consumption, and the rate of missing detections is significantly higher for the last trip of the day.

Other remarkable deviations are related to mode and smartphone type. Public transport trips are less well detected by MoveSmarter. For train trips, the rate of missing detections is limited, and can probably be attributed to the fact that train trips mainly constitute long trips. Although the overall share of bus, tram and metro (BTM) trips is small, a substantial number of BTM trips are not detected by MoveSmarter. As we will explore in section 4, BTM trips are also often classified as bike or car trips. Finally, missing detection rates are higher for users with an Android loan smartphone than for smartphone owners, especially iPhone owners. These differences suggest that smartphone use may also play a role in the detection rates of MoveSmarter. This will be further explored in sections 6 and 7.

We conclude that improved sensing techniques and battery consumption efficiency could improve the detection rate of MoveSmarter trips in specific circumstances. However, even if these improvements are effective, we estimate that still more than half of the missing detections cannot be accounted for. In other words, for all types of trips, MoveSmarter detection rates might always be suboptimal due to inappropriate usage (e.g. leaving the phone at home or forgetting to charge the smartphone, see section 6) or to the limited capabilities of sensors to gather reliable and accurate measurements at all times.

4. Mode detection

As described in the previous section, the number of trips is reduced significantly when we only consider MoveSmarter trips that were validated on the webpage. In total there were 18,189 of such trips for which a one-on-one comparison is possible between MoveSmarter and self-reporting. For this sample, we compare distributions over the modes to establish the quality of the mode recognition in MoveSmarter. In Table 2, we show the modal split for the MoveSmarter and the reported trips on the webpage (i.e. after correction and verification by users), and the average modal split for the Dutch population from the NTS.

Table 2. Modal split; reported trips, MoveSmarter and NTS.

Transport mode	Reported trips N = 18,189	MoveSmarter N = 18,189	NTS 2013
Walking	18.4%	20.9%	18.7%
Bicycle	23.6%	30.5%	26.1%
Car	52.1%	40.2%	47.8%
Train	1.9%	1.3%	2.2%
Bus / Tram / Metro (BTM)	1.4%	0.2%	2.6%
Others	2.4%	3.0%	2.6%
Unknown	0.3%	3.9%	0.0%

Table 2 shows that distribution of trips by mode is quite comparable with those from NTS statistics. There are, however, two peculiarities. First, the fraction of reported BTM trips (1.4%) is relatively low compared to NTS data. This might be related to the fact that MoveSmarter poorly detects BTM trips (only 0.2% are BTM trips). If some respondents are not meticulously checking MoveSmarter registrations, this might lead to an underreporting of BTM trips. From the relatively small fraction of reported unknown modes (0.3% versus 3.9% MoveSmarter detections with an unknown mode) we can conclude that this bias should be quite small. Secondly, the fraction of car trips is quite high compared with NTS data (which is about 45% in the Netherlands). This cannot be attributed to the aforementioned bias. On the contrary, the number of MoveSmarter car trips is clearly too low, while the number of bike trip detections is too high. If we assume mode choice does not only depend on trip characteristics, but also on personal circumstances, a limited sample size and minor differences in demographics may explain this difference with NTS data.

Differences between MoveSmarter and reported mode shares can be interpreted by comparing their modes directly (Table 3). The table shows for each reported mode (rows) how the trips are distributed over the MoveSmarter modes (columns). The number of correct classifications, i.e. the number of trips with the same MoveSmarter and reported mode, are shown bold-face in the diagonal of the table.

Table 3 shows that most trips have the same mode, i.e. in 75% of the cases. There are however some significant differences. If we do not consider public transport, fast modes are underdetected by MoveSmarter. The largest difference is between car and bicycle as 1530 reported car trips are classified as bicycle trips by MoveSmarter. This is 16% of all reported car trips. In addition, of the reported car trips, still 371 (4%) are classified as walk trips by MoveSmarter and 496 (5%) have unknown modes. Misclassifications are much more rare the other way around. Only 48 reported walk trips and 152 reported bicycle trips are classified as car trip by MoveSmarter. A similar trend is visible when comparing bicycle and walk classifications. For the faster mode, i.e. cycling, more reported trips (479) are classified as a slower mode, i.e. walking, by MoveSmarter than the other way around (281 trips). However, in this case differences are relatively small, and the fraction of misclassifications is limited to about 10%.

If we consider public transport, MoveSmarter classifications are clearly not optimal. Train trips are under registered by MoveSmarter. Only 58% of reported train trips are classified as such, while many of them are classified as car trip (about 25%). However, if MoveSmarter classifies a trip as a train trip, this is in 85% of the cases correct. This result suggests that train trips can be correctly detected, but that there is a bias towards other modes (mainly the car). This is mainly caused by inaccuracies in the map-matching process. The performance is much worse for BTM trips. Not only are BTM trips underreported by MoveSmarter, the reliability of the BTM classification is also poor. For the relatively few MoveSmarter BTM trips, only a third of those were indeed reported as BTM trip. To improve mode detection accuracy, the mode choice detection algorithms in the 2014 wave of the Mobile Mobility Panel use, in addition to the locations of public transport stops and routes, real-time public transport time information and public transport time table data available at the national level. The effect of these improvements on the accuracy of mode choice detection is topic of further research.

Table 3. Mode classification reported trips versus MoveSmarter detections.

	MoveSmarter						
	Walking	Bicycle	Car	Train	BTM	Others	Unknown
Walk reported	**2834**	281	48	7	2	149	26
Bicycle reported	479	**3542**	152	4	9	47	58
Car reported	371	1530	**6864**	18	11	183	496
Train reported	11	22	84	**206**	1	13	16
BTM reported	24	87	73	6	**11**	6	41
Others reported	78	89	90	2	0	**150**	19
Unknown reported	0	0	0	0	0	0	**49**

To find an explanation for the MoveSmarter performance, we first look at the overall percentages of correct classifications for various classes of trips. If we make a distinction between the various smartphone types, we can see that iPhone performs better than the Samsung Galaxy Gio loan smartphone and other Android phones, with 78%, 76% and 70% successful classifications, respectively.

In Table 4 we distinguish between different distance and travel time classes. As can be seen from the table, the percentage of correct classifications is relatively low for the smallest and largest travel time trips, while percentages also decline towards more short distance trips. These low percentages, however, do not necessarily have the same cause. As travel modes have different (average) speeds, the distance and travel time classes in Table 5 do not represent the same trips and should therefore be interpreted independently.

Table 4. Successful classifications as percentage of total per distance and travel time class.

	Distance	Travel time
0 - 2 km / 0 - 7 min	70%	69%
2 - 7 km / 7 - 15 min	73%	77%
7 - 20 km / 15 - 30 min	84%	78%
20 - 50 km / 30 - 60 min	87%	79%
>= 50 km / >= 60 min	84%	69%

When we look at the classification of travel time in Table 4, it is quite remarkable how little mode shares vary with travel time. Trends in the mode classification are also more or less independent of travel time. There is one notable exception though. For the longest travel times, 23% of reported bicycle trips are classified as walk trips by MoveSmarter. This fraction is clearly larger than for other travel time classes. As these specific trips mainly constitute recreational trips, one could argue about the relevance of the result. For small travel times, percentages of correct classifications are also relatively low. This cannot be attributed to a peculiar bias in the classification, but is assumed to be merely the result of inaccuracies due to ‘warming-up’ of GPS tracking and the fact that for short trips differences in travel speeds are less distinct for the various modes, which makes mode detection more difficult.

When we look at distance, a completely different picture emerges. Mode composition clearly varies with distance. In particular, the share of fast modes increases towards longer distances. As a result, trends in the mode classification are dependent on distance as is shown in Table 5. Per combined distance class and reported mode, the table lists how the trips are distributed over the MoveSmarter modes. Bold-face percentages are of classifications where MoveSmarter matches the reported mode. Only classes with more than 50 trips are considered. This means that the table only presents modes that are frequently used for corresponding distances. These are slow modes for short distances, BTM for intermediate distances and train for long distance trips. The car is always used frequently and is therefore represented in each distance class.

Table 5: For combined distance class and reported mode, the distribution in percentages of MoveSmarter mode classifications (row total is 100%).

		Walk	Bicycle	Car	Train	BTM	Other	Unknown
0 - 2 km	Walking	**87%**	8%	1%	0%	0%	4%	0%
0 - 2 km	Bicycle	16%	**78%**	3%	0%	0%	1%	1%
0 - 2 km	Car	12%	34%	**44%**	0%	0%	2%	8%
2 - 7 km	Walking	**82%**	11%	4%	0%	0%	0%	2%
2 - 7 km	Bicycle	6%	**89%**	3%	0%	0%	0%	1%
2 - 7 km	Auto	3%	21%	**67%**	0%	0%	1%	8%
2 - 7 km	BTM	11%	46%	20%	0%	**6%**	1%	16%
7 - 20 km	Bicycle	1%	**95%**	4%	0%	0%	0%	0%
7 - 20 km	Car	1%	8%	**87%**	0%	0%	1%	4%
7 - 20 km	Train	0%	3%	30%	**65%**	0%	0%	3%
7 - 20 km	BTM	3%	37%	32%	1%	**5%**	0%	22%
20 - 50 km	Car	0%	2%	**94%**	0%	0%	1%	2%
20 - 50 km	Train	0%	4%	25%	**64%**	1%	1%	5%
>= 50 km	Car	1%	3%	**91%**	0%	0%	5%	1%
>= 50 km	Train	0%	2%	25%	**70%**	0%	3%	1%

From Table 5, we can draw conclusions for each separate mode. For the car trips, misclassifications mainly occur at short distances. At short distances car trips are clearly underdetected in favor of the bicycle trips. Similarly, bicycle trips are often classified as walk trips. The bias towards slower modes mainly occurs at the very short distances. An explanation could be found in the speed profiles, as the profiles from the fast modes start to resemble those from the slow modes. For distances between 2 and 7 km, this bias is already much weaker, and the bias disappears for distances beyond 7 km. Note that for large distances, biases towards faster modes are not present or very weak. We find a similar kind of trend, albeit much weaker, for train trips. The bias (of too few train trips) is strongest at shorter distances. Whereas the overall rate of successful classifications of train trips was 58%, this rate increases to 70% for distances beyond 50 km. It is not unexpected that characteristics used for detecting train trips are more distinct for these long distance trips. Finally, the rate of successful classifications remains very poor for BTM trips and appears to be distance independent.

5. User evaluation

After the fieldwork period, a user evaluation has been carried out. Participants were asked to express their opinion about different aspects of the experiment, e.g. the app, accuracy of the trips and mode detections, smartphone use, accuracy of trip corrections, battery use and the webpage. Table 6 shows the users’ assessment of the MoveSmarter app on a five point Likert scale. The users are most satisfied with the automatic detection of departure and arrival location (> 50% satisfied, and 21% dissatisfied) and least satisfied with the mode detection (22% satisfied, 49% dissatisfied). Note that departure and arrival locations and times were hardly changed in the webpage.

Table 6. User evaluation MoveSmarter (N=549)

	Certainly not true 1	2	3	4	Certainly true 5	Total
MoveSmarter detected my trips correctly	14%	23%	29%	23%	12%	100%
MoveSmarter detected the correct mode	19%	30%	28%	18%	4%	100%
MoveSmarter detected the correct trip purpose	13%	16%	33%	27%	10%	100%
MoveSmarter detected the correct departure location	8%	13%	22%	36%	21%	100%
MoveSmarter detected the correct arrival location	7%	14%	26%	36%	17%	100%

Table 7 shows that about half of the respondents were dissatisfied with the battery consumption during the fieldwork period, while 35% did not have a negative experience. Contrary to what was expected in advance, non-smartphone owners were most negative. As smartphone owners will lose battery capacity due to location tracking, it was expected they might feel this as an extra burden. Possibly, smartphone users are already familiar with apps that use up some battery power and are, therefore, better able to estimate the battery consumption of MoveSmarter compared to non-experienced smartphone users familiar with the low battery consumption of traditional mobile phones.

Table 7. How do you assess battery usage of the smartphone during the experiment?

	N	Very negative 1	2	3	4	Very positive 5	Total
Loan (Android)	326	28%	23%	15%	18%	15%	100%
Android	133	32%	18%	17%	14%	19%	100%
iPhone	91	24%	19%	23%	16%	18%	100%
Total	550	28%	22%	18%	18%	17%	100%

Table 8 shows to what extend respondents forget to take their smartphone with them during trips. As expected, respondents with loan phones forgot to bring their phone more often (17%) than smartphone owners (13% for Android users, and 4% for iPhone users). Interestingly, iPhone users rarely 'forget' their smartphones which is consistent with the fact that iPhone users had the fewest missing detections (section 4).

Table 8. I sometimes forgot to bring my phone with me

	N	Certainly not true 1	2	3	4	Certainly true 5	Total
Loan (Android)	326	51%	21%	12%	11%	6%	100%
Android	133	67%	15%	6%	8%	5%	100%
iPhone	91	73%	10%	13%	3%	1%	100%
Total	550	58%	18%	11%	9%	5%	100%

Table 9 shows the users' assessment of the webpage. A large majority did not have any troubles to correct MoveSmarter trips in the web page. A small but significant fraction (27%) indicated that they hardly had to correct MoveSmarter trips. An interesting result is that more than 10% of respondents indicate they cannot remember all registered trips. This shows that automatic trip detection can reduce the danger of underreporting. MoveSmarter has the potential to take on this task, and results from this evaluation and from previous sections are promising. However, results also show that some improvements are still needed.

Table 9: Evaluation reporting via the web page (N=549)

	Certainly not true 1	2	3	4	Certainly true 5	Total
I found it difficult to correct trip data in the corresponding webpage	44%	25%	13%	12%	7%	100%
I could well remember all trips	4%	7%	13%	32%	43%	100%
I hardly needed to correct my trip data in the webpage	20%	23%	30%	19%	8%	100%
I meticulously checked and corrected my trip data in the webpage	5%	3%	10%	26%	56%	100%
Checking trip data got easier after one week	8%	9%	32%	26%	24%	100%

6. User evaluation and missing detections

In this section we look for relationships between trip rates from section 4 and the outcomes of the user evaluation from the previous section. The idea is that the user evaluation may provide explanations for missing trip detections and underreporting. The user evaluation was conducted after the fieldwork period and about 87% participated in the user evaluation. In this section, we consider only respondents that participated in the study and completed the user evaluation (N=479).

The largest difference is detected between non-smartphone owners (N = 285) and smartphone owners (N = 194). Smartphone owners detected 0.7 more trips per day with MoveSmarter and reported 0.4 more trips per day. These differences are detected for both batches and for both business days and weekends. The likelihood is also very small that these differences are coincidental (due to limited sample sizes). The observed difference in MoveSmarter detections is consistent with the fact that non-smartphone owners are more likely to use their phone inappropriately or forget their phone more often (Table 9). The rate of missing MoveSmarter detections is with 21% also significantly higher for non-smartphone owners than the 13% for smartphone owners. It is possible that more MoveSmarter detections result in fewer forgotten trips and therefore more reported trips. This could be a quite logical explanation for the higher rate of reported trips for smartphone owners. However, it cannot completely be ruled out that smartphone owners are simply more mobile than people without a smartphone.

We also looked at respondents who positively assessed the experiment. Respondents that indicated to always bring their phone along (N = 373), that meticulously corrected their trips on the webpage (N = 330), or that did not have trouble to correct their trips (N = 341) all showed on average about 0.3 more MoveSmarter trips per day. Interestingly, if we consider respondents that gave good assessments on all these points (N = 236), we find that the gap between non-smartphone owners and smartphone owners narrows significantly. Non-smartphone owners with good assessments show significantly higher MoveSmarter trip rates than those with bad assessments, while for smartphone owners there is little variation in MoveSmarter trip rates. For both smartphone groups, reported trips rates are clearly higher when the assessments are good. The latter result suggests that reported trip rates are higher for more motivated respondents, and might indicate underreporting for the remaining group. The former result is consistent with the expectation that missing MoveSmarter detections mainly occur among respondents that hardly have any experience with smartphones and at the same time are less motivated and / or report their trips less meticulously. One of the challenges is to find the right balance between improving trip detection rates and keeping the sample representative.

Finally, we looked at battery usage. For the respondents with a negative assessment on battery usage (N = 238) on average 0.5 more MoveSmarter trips per day were reported than for the respondents with a positive assessment (N = 165). The difference is more or less similar for non-smartphone owners and smartphone owners. At first sight, this result is counterintuitive. Negative experiences with battery consumption should lead to a decline in MoveSmarter use. However, the explanation for this result is probably more straightforward. Respondents with many MoveSmarter detections consume more battery power and hence might complain more often about the battery. Although these respondents also report more trips, the difference in reported trip rates is smaller than the difference between MoveSmarter trip rates. There is therefore little evidence that these respondents detect more trips because they are more mobile Moreover, as we observe these results for all respondents (both for non-smartphone owners and smartphone owners), we conclude that the MoveSmarter trip detection has just performed better for respondents with a negative assessment on battery consumption. This conclusion is further supported by the low rate of missing MoveSmarter detections (8%) for this group.

7. Conclusions

In this paper, we examined the quality of automatic trip detection using a dedicated smartphone app (MoveSmarter). The Mobile Mobility Panel is unique compared to other international studies of automatic trip detection by smartphones, because of the sample size (about 600 respondents), representativeness of the sample for the Dutch population (by age, smartphone ownership, etc.), and the web-based prompted recall survey. An in-depth comparison between automatic detections and reported trips was conducted for a sample of about 20 thousand trips.

The first results are promising. MoveSmarter detects most trips correctly without strong biases in trip length or travel time distributions. Respondents are in general also satisfied with the accuracy of the measurements. However, 20 – 25% of the trips were not detected by MoveSmarter. Trip detection is inaccurate when activity times at the trip destination are small. This bias may be caused by the difficulty to distinguish two successive trips that follow each other rapidly. Relatively many public transport trips were also not classified. Yet, we estimate that most missing MoveSmarter trips are the result of inappropriate use (e.g. forget to take smartphone along) or empty batteries. The latter would also be consistent with results from the users' evaluation in which respondents indicated that they were not very satisfied with battery usage.

There is also direct and indirect evidence for underreporting. Sometimes a reported trip does not start at the same location where the previous trip ended, while less meticulous respondents report fewer trips. Moreover, the rate of reported trips decreases when relatively many MoveSmarter trips are missing. The latter result implies that respondents tend to 'forget' to report more often when the fraction of missing detections increases. Evidence for underreporting is not a drawback that arises from this study. On the contrary, in traditional National Travel Surveys underreporting is more difficult to detect, but no less of an issue. Although a direct comparison can not be done, trip rates (per person) appear to be clearly higher than those from other Dutch national travel surveys. This suggests that the rate of underreporting is reduced when automatic trip detection is included.

The automatic mode detection is also quite accurate with successful mode classifications for 75% of the trips. However, there is clearly room for improvement: short distance car trips are too often classified as bicycle trips, train trips are underdetected, and BTM (bus, tram or metro) trips have very poor detection rates. From the users' evaluation it also became clear that quite some respondents were not satisfied with the mode detection. In the future, we propose three steps to improve the mode detection. First, sensors will be updated (e.g. by including accelerometers) and sensing will be improved (as a result of newer versions of Android and iOS). This will in particular result in improved classifications between slow and fast modes. Secondly, more context information will be considered in the back-end algorithms, such as locations of train and BTM stations & lines, and BTM stops. This should lead to improved classifications for public transport trips. Finally, learning algorithms will be extended, for example by using automatic pattern recognition techniques. This will enable us to improve mode recognition based on previous classifications. Of course, this method cannot easily be transferred to other countries, because it requires large samples of trips for which modes have also been reported.

The Mobile Mobility Panel combines the advantages of automatic detection with self-reporting: a low burden for respondents with high accuracy rates. MoveSmarter can be concluded to be a promising alternative or addition to traditional trip diaries but there is room to improve trip and mode detection rates and the efficiency of battery consumption.

Acknowledgements

This project is financed by the Netherlands Organisation for Scientific Research (NWO) and would not have been possible without co-funding from the KiM Netherlands Institute for Transport Policy Analysis and CentERdata. The authors want to thank Josette Janssen, Maarten Streefkerk, Iggy van der Wielen (CentERdata) and Martin Wibbels (Mobidot) for their contributions to this project, and CentERdata for the use of the LISS panel for this research. We would also like to thank two anonymous referees for their comments on an earlier draft, which helped us to improve this paper.

References

Bie, J., Bijlsma, M., Broll, G., Cao, H., Hjalmarsson, A., Hodgson, F., Holleis, P., Van Houten, Y., Jacobs, K. & Koolwaaij, J. 2012. Move Better with tripzoom. International Journal On Advances in Life Sciences, 4, 125-135.

Cottrill, C., F. Pereira, F. Zhao, I. Dias, H. Lim, M. Ben-Akiva, P. Zegras (2013), Future Mobility Survey. Experience in Developing a Smartphone-Based Travel Survey in Singapore. Transportation Research Record, 2354, 59-67.

De Vos, K. 2010. The representativeness of the LISS panel 2008, 2009, 2010, Retrieved from www.lissdata.nl/assets/uploaded/Representativeness%20of%20the%20LISS%20panel%202008,%202009,%202010.pdf

Fan, Y., Chen, Q., Douma, F. & Liao, C.-F. 2012. Smartphone-Based Travel Experience Sampling and Behavior Intervention among Young Adults. Intelligent Transportation Systems Institute, Center for Transportation Studies, University of Minnesota.

Feng, T, Timmermans, H. J. P., 2013. Transportation mode recognition using GPS and accelerometer data. Transportation Research Part C: Emerging Technologies 37, 118-130.

Hoogendoorn-Lanser, S., Schaap, N., Kalter, M-J. Olde Kalter, 2014. The Netherlands Mobility Panel: An innovative design approach for web-based longitudinal travel data collection, 10th International Conference on Transport Survey Methods, Leura, Australia, 16-21 November, 2014.

Kracht, M. 2004. Tracking and interviewing individuals with GPS and GSM technology on mobile electronic devices. In: Seventh International Conference on Travel Survey Methods, Costa Rica, 2004.

Leenheer, J., Scherpenzeel, A.C. 2013. Does it pay off to include non-internet households in an internet panel? International Journal of Internet Science, 8(1), 17-29.

Li, M., Dai, J., Sahu, S. & Naphade, M. 2011. Trip analyzer through smartphone apps. In: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2011. ACM, 537-540.

Nitsche, P., Widhalm, P., Breuss, S. & Maurer, P. 2012. A Strategy on How to Utilize Smartphones for Automatically Reconstructing Trips in Travel Surveys. Procedia - Social and Behavioral Sciences, 48, 1033-1046.

Nitsche, P., Widhalm, P., Breuss, S., Brändle, N., Maurer, P., 2013. Supporting large-scale travel surveys with smartphones – A practical approach. Transportation Research Part C: Emerging Technologies (article in press)

Prelipcean, A. C., G. Gidófalvi, Y. O. Susilo (2014), Mobility Collector. Journal of Location Based Services, 8(4), 229-255.

Ortúzar, J. D. D., Armoogum, J., Madre, J.-L. & Potier, F. 2010. Continuous Mobility Surveys: The State of Practice. Transport Reviews, 31, 293-312.

Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M. & Srivastava, M., 2010. Using mobile phones to determine transportation modes. ACM Transactions on Sensor Networks 6.

Statistics Netherlands (2014) Onderzoek Verplaatsingen in Nederland, 2013. [2013 National Travel Survey]

Scherpenzeel, A.C., "True" Longitudinal and Probability-Based Internet Panels: Evidence From the Netherlands. Bethlehem, J.G., 2011. How representative are online panels? Problems of coverage and selection and possible solutions. In: Das, M., P. Ester, and L. Kaczmirek (Eds.), Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies. Boca Raton: Taylor & Francis.

Scherpenzeel, A.C., & Das, M., 2011. In: Das, M., P. Ester, and L. Kaczmirek (Eds.), Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies. Boca Raton: Taylor & Francis.

Schönfelder, S. & Axhausen, K. W. 2010. Urban Rhythms and Travel Behaviour. Spatial and Temporal Phenomena of Daily Travel, Farnhem/Burlington, Ashgate Publishing.

Stopher, P.R., Zhang, Y., 2011. The repetitiveness of daily travel, Transportation Research Board Annual Meeting, Washington, January 2011.

Stopher, P. R. 2009. The travel survey toolkit: where to from here? Transport survey methods, keeping up with a changing world, 15-46.

Thomas, T., Jaarsma, C.F., Tutert, S.I.A. 2013. Exploring temporal fluctuations of daily cycling demand on Dutch cycle paths: the influence of weather on cycling, Transportation, 40 (1), 1 – 22.

Shin, D., D. Aliaga, B. Tunçer, S.M. Arisona, S. Kim, D. Zünd, G. Schmitt (2015), Urban sensing: Using smartphones for transportation mode classification. Computers, Environment and Urban Systems (in press).