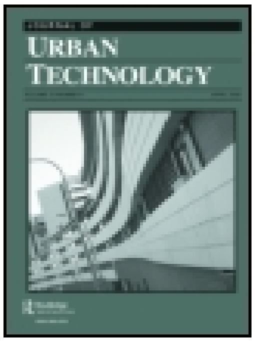

Journal of Urban Technology

ISSN: 1063-0732 (Print) 1466-1853 (Online) Journal homepage: http://www.tandfonline.com/loi/cjut20

Automatic Trip Detection with the Dutch Mobile Mobility Panel: Towards Reliable Multiple-Week Trip Registration for Large Samples

Tom Thomas, Karst T. Geurs, Johan Koolwaaij & Marcel Bijlsma

JOURNAL OF URBAN TECHNOLOGY
https://doi.org/10.1080/10630732.2018.1471874

Routledge
Taylor & Francis Group

OPEN ACCESS   Check for updates

Automatic Trip Detection with the Dutch Mobile Mobility Panel: Towards Reliable Multiple-Week Trip Registration for Large Samples

Tom Thomas, Karst T. Geurs, Johan Koolwaaij, and Marcel Bijlsma

ABSTRACT
This paper examines the accuracy of trip and mode choice detection of the last wave of the Dutch Mobile Mobility Panel, a large-scale three-year, smartphone-based travel survey. Departure and arrival times, origins, destinations, modes, and travel purposes were recorded during a four week period in 2015, using the *MoveSmarter* app for a representative sample of 615 respondents, yielding over 60 thousand trips. During the monitoring period, respondents also participated in a web-based prompted recall survey and answered additional questions. This enables a comparison between automatic detected and reported trips. Most trips were detected with no clear biases in trip length or duration, and transport modes were classified correctly for over 80 percent of these trips. There is strong evidence that smartphone-based trip detection helps to reduce underreporting of trips, which is a common phenomenon in travel surveys. In the Dutch Mobile Mobility Panel, trip rates are substantially higher than trip-diary based travel surveys in the Netherlands, in particular for business and leisure trips which are often irregular. The rate of reporting also hardly decreased during the four-week period, which is a promising result for the use of smartphones in long duration travel surveys.

KEYWORDS
Smartphone app; travel survey; mode detection; underreporting

Introduction

Travel diaries have been used since the 1930s to obtain information about people's travel patterns (Rieser-Schüssler and Axhausen, 2014). Traditional self-administered travel diaries have well-known problems such as inaccuracies in reported distances and travel times due to rounding (e.g., Witlox 2007), respondent burden, and underreporting. Given the inherent and unavoidable downward bias of survey-based methods, it comes as no surprise that passive tracking has become prominent in transportation research since the late 1990s (see for overviews Schönfelder and Axhausen, 2010; Rieser-Schüssler and Axhausen, 2014; Shen and Stopher, 2014). More recently, smartphone-based data collection has been tested because it offers the opportunity to combine GPS tracks with data from other smartphone sensors such as accelerometers, Wifi, and GSM. In the past,

CONTACT Tom Thomas t.thomas@utwente.nl Centre of Transport Studies, University of Twente, PO Box 217,7500, AE Enschede, the Netherlands.

inaccurate GPS measurements and heavy energy-consuming GPS measurements have been obstacles for large-scale applications (Rieser-Schüssler and Axhausen, 2014). Recently, with increased accuracy and employment of battery-management strategies, a growing number of studies have reported successful pilot applications (e.g., Reddy et al., 2010; Nitsche et al., 2012; Geurs et al., 2015; Cottrill et al., 2013; Zhao et al., 2015; Prelipcean et al., 2014, 2017; Safi et al., 2016).

This paper describes results from the Dutch Mobile Mobility Panel project, a three-year, large-scale, long-duration panel study. The paper aims to examine the quality of trip detection using smartphones, using the last wave (2015) comprising four-weeks of smartphone data validated by respondents in a web interface. More specifically, we examine if smartphones are able to record (almost) all trips accurately, even after several weeks of monitoring, and if we can estimate the rate of underreporting and/or control for underreporting. It combines a sample of over 600 respondents with continuous trip registration over four weeks, resulting in more than 60,000 recorded trip legs in 2015 alone. Smartphones were distributed to non-smartphone owners to achieve a good distribution of the population by socioeconomic characteristics and age. This powerful combination enables us to study the quality of trip detection in great detail. We will show to what extent our method is successful in monitoring individual travel behavior over a longer period of time.

The paper is structured as follows. We first provide a brief literature review. Then we describe the collection methodology of the Mobile Mobility Panel project. After the trip rates and quality of the mode detection are presented, we explore the issue of under reporting in more detail. Finally, we present our conclusions.

Literature Review

As a response to the inherent and unavoidable downward bias of survey-based methods, passive tracking has become prominent in transportation research since the late 1990s (see for overviews Schönfelder and Axhausen, 2010; Rieser-Schüssler and Axhausen, 2014; Shen and Stopher, 2014). In the literature, various rates of underreporting are documented ranging from 10 to 80 percent, comparing GPS traces with travel diaries (see Schönfelder and Axhausen, 2010; Bricka et al., 2012; Ribeiro et al. 2014; Kelly et al., 2014; Houston et al., 2014). Kelly et al. (2014) for example used wearable cameras to find an underreporting of trips by 15 percent, while Houston et al. (2014) used GPS loggers to detect 0.39 trips per person per day more than reported in a corresponding survey. Rates in underreporting are, however, not always the same, as differences occur between groups of respondents (e.g., Bricka et al., 2012) and underreporting appears to be most severe for short trips (e.g., Ribeiro et al., 2014). At the same time, there are errors and omissions in GPS and Smartphone measurements as well. Not all trips are detected (missing trips) while detections of non-trips (when sensors detect trips that are actually not taking place) also occur. The magnitude of these errors depends, for example, on the setting of the gap between trip legs, with higher gaps leading to more missing trips but fewer non-trip detections. Typically, gap settings are reported between 45 and 900 seconds (for an overview see e.g., Gong et al., 2012; Reinau et al., 2015). For example, Houston et al. (2014) used a gap setting of 60 seconds. However, their rate of underreporting depends on this gap. It is, therefore, important to consider both missing and non-trip detections when evaluating the quality of automatic detection.

Setting the right gaps to separate between trip legs is only one of the issues of automatic trip detection. Another limitation of GPS and smartphone measurements is related to battery drainage (resulting in empty batteries), which put some limits on their use as collectors of trip data. Due to the limitations of automatic detection, it has been suggested to combine GPS and other smartphone sensors with prompted call survey techniques to increase the quality of the trip-data collection (e.g., Raza et al., 2015).

Prompted call survey techniques are also useful to improve and/or validate mode and activity detection. Recently, significant attention has been paid to the quality of mode choice. Algorithms for GPS data imputation have been applied in different contexts, leading to variations in reported prediction accuracy. Most studies report an average accuracy between 70 and 80 percent (Biljecki et al., 2013). Reported success rates for automatic mode detection are beyond 80 percent in several studies (e.g., Gong et al., 2012; Shin et al., 2015; Rasmussen et al., 2015; Nitsche et al., 2014) and in some cases even beyond 90 percent (e.g., Feng and Timmermans, 2013; Ferrer and Ruiz, 2014). The difference in predicted accuracy depends not only on the algorithm, but also on the number of identified transportation modes, type of input variables, urban setting, and data used to validate the algorithms. In the Netherlands, mode choice detection is particularly challenging in urban settings with high walking and bicycle shares and bicyclists also cycling at relatively high speeds. A large-scale GPS-based seven-day travel study in the Netherlands with over 1,100 respondents (e.g., Bohte and Maat, 2009) deduced car use correctly most often (75 percent of all trips), followed by cycling (72 percent) and walking (68 percent), respectively. Feng and Timmermans (2013) achieved very high accuracy levels (between 83 percent for train trips up to 100 percent for running and cycling) combining GPS and accelerometer data (and trace data recorded every second). However, this study, and also other studies in the literature with high mode choice detection accuracies were small-scale experiments with only a few participants.

More recently, smartphone-based data collection has been tested as that offers a great potential to combine GPS tracks with data from other smartphone sensors such as accelerometer, Wifi, GSM, and Bluetooth. One of the major advantages of smartphone-based data collection is that smartphone owners do not have to carry separate devices while travelling. As a result, large samples of data can be easily collected. Smartphone data has in the literature been used, for example, in macroscopic travel modelling (e.g., Toole et al., 2015), mode-choice forecasting (e.g., Semanjski and Gautama, 2015), and travel-time forecasting (e.g., Semanjski, 2015). In addition, many smartphone apps using GPS and GSM have been developed to provide users with feedback on their travel behavior. The use of smartphones as data collection tools to supplement or replace traditional travel surveys has been the subject of several studies, in the United States (Reddy et al., 2010), Austria (Nitsche et al., 2014), Sweden (Prelipcean et al., 2014), Shanghai, China (Xiao et al. 2015), Singapore (Zhao et al., 2015), and New Zealand (Safi et al., 2016). However, except the Future Mobility Sensing pilots in Singapore with almost 800 users participating for at least 14 days, existing smartphone-based studies have been limited in terms of sample sizes and the number of observed trip legs. |Zhao et al. (2015) conclude from pilots in Singapore that smartphone-based data have a higher resolution and better accuracy than the traditional travel surveys in Singapore, and that common problems in traditional surveys—such as underreporting of trips, overestimation of travel times, and inaccuracy of locations and times—can be reduced. The size of the Dutch Mobile Mobility

Panel project is thus quite unique compared to most of the aforementioned studies. It combined a relatively large sample of about 600 respondents annually with continuous trip registration, and resulted in more than 25,000 trip legs in 2013 (two-week duration), 54,000 in 2014 (four-week duration), and 60,000 in 2015 (four-week duration).

Methodology

Overview

The trip registration in the Dutch Mobile Mobility Panel consists of several steps, which are visualized in Figure 1. First, a smartphone application called *MoveSmarter* (developed for iPhone and Android platforms) is used to automatically detect trip and trip characteristics (left side of Figure 1). Detected trip characteristics are uploaded to a database on a back-end server where a series of algorithms is used to process, clean, and enrich the trip data. These trips are then shown in the *MoveSmarter* app. Finally, trips and trip characteristics are also presented in a web portal, which is an Internet-based prompted recall survey. Respondents are presented with all trips and trip characteristics (grouped by day) and they can change, add, or delete trips in the portal, 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 unchanged *MoveSmarter* trips, changed *MoveSmarter* trips, and added trips). The automatically detected trips (after processing on the back-end server) we call *MoveSmarter* trips. Note that feedback from the recall survey is also used to improve the processing of *MoveSmarter* trips. What follows are detailed descriptions of those steps. The next two sub-sections describe the automatic trip detection process and the Internet-based prompted recall survey. Then we summarize the recruitment method and characteristics of the sample. And finally, we describe how the trip data are analyzed.

In the paper, we make several comparisons using different data sources. These sources are summarized in Table 1. The comparisons enable us to interpret the results and put

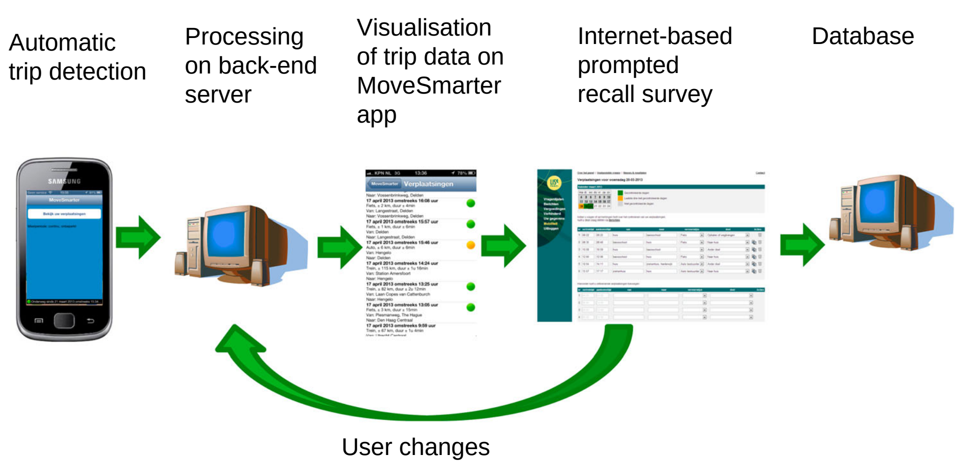

Figure 1. Schematic illustration of the Dutch Mobile Mobility Panel data collection methodology.-Source: Geurs et al., 2015

Table 1. Relevant data sources with descriptions

Name of the data source	Description
Dutch Mobile Mobility Panel	The survey used in this study. The sample is a sub sample of the LISS panel, which is used for all kinds of panel studies
MoveSmarter trips	The automatically detected and processed trips by the *MoveSmarter* app and the back end server
Reported trips	The reported trips in the Internet-prompted recall survey in which respondents check and adjust the *MoveSmarter* trips
One-day trip survey	Self-reported trips on the first day of the experiment during which trips were not detected by *MoveSmarter*. Used as comparison with the reported trips in the recall survey
Mobility Panel for the Netherlands (MPN)	online and place-based three-day trip diaries of about 4,000 individuals (e.g., Hoogendoorn-Lanser et al., 2014)
Dutch National Travel Survey (NTS)	Pen-paper, one-day trip diaries of over 37,000 individuals (Statistics Netherlands, 2014)

them into context. *MoveSmarter* trips are compared with reported trips from the recall survey to check the quality of the automatic trip detection, but also to better understand the rate of underreporting of trips. By comparing observed trip rates (i.e., number of trips per person per day), some estimates of underreporting can be made. First, a direct comparison between the smartphone-based approach used in this study and the traditional travel-survey approach is conducted. We compare a one-day trip survey without *MoveSmarter* detections at the start of each wave (conducted on the first Monday of the monitoring period). Second, we compare trip rates (from reported trips) from the Dutch Mobile Mobility Panel with two other independent Dutch mobility surveys: the Mobility Panel of the Netherlands and the Dutch National Travel Survey. However, this is an indirect comparison given the differences in the sample size and characteristics, duration, and monitoring period.

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 behavior. 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.

Sensing Module and Battery Management. 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 (significant location changes) and trip end. In this way a coherent trace of GPS locations is collected (the "Raw Trip") without any user involvement. The sensing module runs as a background process on the smartphone and restarts itself when the operating system (iOS, Android) boots up. The sensing module is configurable so that different sensing strategies can be deployed depending on situational factors so that needed measurement accuracy is optimized versus battery consumption. 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 24×7 sensing with normal smartphone use throughout the day and charging overnight. Because of this trade-off process, full-sensing accuracy at all times is prevented, and more subtle sensing strategies are required. These sensing strategies use available context information (e.g., if the smartphone "knows" a traveler is on the train, his degrees of freedom is limited and sensing can be less fine-grained) or available historic travel patterns of the user (e.g., stick to the daily commute route unless there is a clear deviation from normal routine). So the system does not use manual start-stop, but autonomously detects trip start and trip end, and uses the historical mobility profile to detect trip ends more quickly to reduce battery consumption, e.g., if someone's mobility profile contains a home location, homebound trips will be detected in one minute, as opposed to 15 minutes for previously unvisited locations. Similarly, the sensing module decides autonomously which quality of input source is required to gather the location information for a new trip. For example, for a daily commute observed for the twenty-first time in a month, Wifi might be sufficient, as opposed to GPS being required for a first-time holiday trip (Geurs et al., 2015: 249).

Additionally, a second stage in the trip detection process is introduced. This back-end process is triggered when Raw Trips are uploaded. Upload starts 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.

Cleaning and Map Matching. After upload, the Raw Trip is processed in the back-end. This is the second stage of the trip detection process. The processing includes filtering, cleaning, enrichment, and analysis of the Raw Data.

Before steps like mode detection can take place, it is essential that the outliers in the raw location trace of a trip are cleaned out, and that sensing gaps in the location trace are filled. Outlier detection is also a complex task, due to the fact that location measurements by smartphones are not always correct. In some cases, off-road locations might actually be accurate, whereas seemingly accurate measurements on the road might in fact be inaccurate. Moreover, the start of a trip is usually detected after a user already has started the trip. That means that we might miss the first few hundred meters of a location trace. These gaps—and also the ones generated by fast moving trains or tunnels—are filled with location estimates based on previously recorded behavior or (if no personal history is available) a logical navigation alternative. This helps significantly to speed up the next step in the process, map matching. We use shortest path algorithms running on a Postgres database extended with PgRouting to determine the best matching route using the available infrastructure (e.g., roads, railroads, canals) given the sensed location trace. *MoveSmarter* provides for multi modal map matching, so also mode transitions from, e.g., car to train via walking from parking lot to platform, are included in the map-matching result.

Mode Detection. Various methods have been used to detect modes. In general, a training set is used to let a mathematical algorithm "learn" to recognize the correct mode, e.g., by Bayesian belief networks (Feng and Timmermans, 2013, Xiao et al., 2015) recurrent neural networks (Ferrer and Ruiz, 2014), and Markov chains (Nitsche et al., 2014). In some cases, geographical context data is added to improve the mode detection algorithm. Rasmussen et al. (2015), for example, combined fuzzy logic with a GIS-based algorithm to detect modes of trip legs in the Greater Copenhagen area.

In the *MoveSmarter* trip analysis process, modes are deduced from the enhanced trip data using probabilistic Bayesian mode deduction models comparable to the Bayesian belief networks used in aforementioned studies. These models are trained on a complete training set (consisting of all types of users and types of trips) of about 3,000 trips, and use a 41-dimensional feature set including:

(1) Speed, altitude, and incline patterns
(2) Location and movement data characteristics (derived from GPS, WiFi, Accelerometer, and cell-ID information, as described above)
(3) The underlying infrastructure network (via road, rails, water, and air, or a combination thereof, acquired from the map-matching step)
(4) Characteristics of the map matched route, including transport hubs along the route, road speeds and surface, access permissions, etc.
(5) Public transport information, i.e., PT lines and time-tables, but most importantly real-time information about PT vehicles, including delayed ones. This information is included in the shortest path algorithm to match the best multi-modal route, as described above
(6) Personal trip history, including daily travel behavior and manual corrections to previous similar trips.

In addition to the probabilistic approach, *MoveSmarter* deploys a quality check and a sanity check per trip. The quality check expresses a combination of the quality of the sensing, the quality of the map matching and the quality of the mode detection, and shows how confident the system is about the timing, route, and mode of the current trip in terms of "Good," "Approximate," "Bad," and "Corrupt." "Good" is basically a trip with accurate sensing and with map matching; "Approximate" has either inaccurate sensing or missing map matching results; "Bad" has both inaccurate sensing and missing map matching result; and "Corrupt" indicates non-trips that are sensed but do not represent an actual travelling activity of the user. The sanity check is to cancel out strange outcomes of the mode detection, which may happen due to underrepresentation of certain situations in the models, or due to low quality sensing or map matching. The main reason for a sanity check is that, although a model is trained on a large training set, not all infrequent real-world situations are accounted for in the model, and hence a model can have an output that is evidently unlikely to any human at first sight. Examples are cycling on motorways or railway tracks and plane trips shorter than 5 km.

Updating and Post-Processing. Frequently used routes and frequently visited places were clustered to deduce the most likely motive to trips (e.g., "daily commute" for a recurrent trip to an office building). By adding spatial context data on activities and land use, trips to regularly visited locations can be detected, even when activity times are very short. Logical transit points along the route were also considered. For example, a temporary stop near a bus stop or trains station points to a transfer between public transport and another mode. This approach compares with for example Liu et al. (2013), who used points of interests, including transfer points, in a post processing procedure to improve the mode detection for (different) trip legs.

Internet-Based Prompted Recall Survey

Respondents of the mobility panel are part of the LISS panel (|Scherpenzeel and Das, 2010). The LISS panel consists of more than 5,000 households and in total over 8,000 respondents. It provides a representative sample of the Dutch population. Representativeness is guaranteed by providing Internet devices and connections to people without them. The respondents get a monetary reward when they fully participate during a research experiment. CentERdata, the owner and manager of the LISS panel, has developed a website where respondents can view their daily trips. Per day, trips are shown using the following fields: departure and arrival time, origin and destination location, mode, and trip purpose. The respondents are asked to check the *MoveSmarter* trips every three days, and make changes where necessary. Departure and arrival times can be changed by changing hour (hh) and minute (mm). The location fields can be freely changed. As the location of an automatic registration is not always recognizable (e.g., the address of a supermarket may be unknown to the respondent), the respondent can change this field and relate it to the activity at the location. 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 (See Figure 1). The mode and trip purpose can be selected from drop-down menus, which values (names) conform to those used in the Dutch National Travel Survey (NTS) and the Mobility Panel for the Netherlands (MPN). In this way, results for different trip purposes can be easily compared among surveys. There is, however, one important difference with the aforementioned surveys. In the Mobile Mobility Panel, each record represents a trip leg. To facilitate the exchange between one transport mode to another mode, the purpose "Change Modes" was added in the drop down menus. The number of trips (from origin to final destination) was simply extracted by excluding the records for which the trip purpose field is equal to "Change Modes." This is done when we compare trip rates and study the underreporting of trips.

The respondent can also delete complete trip legs (that were falsely detected by *Move-Smarter)*, add trips legs (that were not detected by *MoveSmarter*), combine trip legs (that were falsely detected as separate trip legs by *MoveSmarter*) or split trip legs (that were falsely detected as one-trip legs by *MoveSmarter*). In this regard, it is important to mention, that in about 95 percent of the cases, these trip legs correspond with actual trips, i.e., in these cases the respondent only used one travel mode between origin and final destination (note that short walks, for example between parking place and final destination are not considered as separate trip legs). When the trips of a day are submitted, the respondent needs to formally approve the reported trips for that day.

Recruitment and Sample

For this research, about 800 LISS respondents (who were 18 years or older in 2015) showed an interest and gave consent to store their data. From these 800 respondents, 655 respondents actually started the experiment in 2013, and slightly more than 550 respondents participated during the whole experiment (smartphone use and online trip validation). Within the LISS panel, about 36 percent owned a smartphone in 2013.[1] To avoid biased results, respondents without smartphones supported by iOS or Android were provided with a loan smartphone (Samsung Galaxy Gio). To reduce the costs of the data collection, the smartphones were distributed in two batches and could be used twice. The share of loan-phone users in our sample dropped from 59 percent in 2013 to 47 percent in 2015; the share of Android users increased from 24 to 30 percent; and the share of iPhone users went from 17 to 24 percent.

The distribution of personal characteristics (age, education level, income, position in the household) and urbanization level of recruited participants corresponds well with that from the whole LISS panel (16 years and older) (for more details we refer to Geurs et al., 2015). A resulting unique feature of our sample is the share of older adults (39 percent of the sample is between 55 years and older), which corresponds well with the Dutch population (36 percent is 55 years or older). Smartphone-based travel survey studies in the literature typically have fairly skewed samples as older adults are difficult to reach given their low smartphone-ownership levels.

In 2015, trips were recorded from Tuesday, April 7 to Sunday, May 3 (first batch) and from Tuesday, June 9 to Sunday, July 5 (second batch). The sample only includes trips that were approved by respondents. The 2015 sample consists of 615 respondents. This is an average of 308 per batch, of which on average 287 respondents per day approved their trips. Respondents who dropped out due to time limitations or problems with the app have been excluded (71 in total). Another 37 respondents who were immobile respondents (e.g., due to illness) or who were on holiday abroad were not included as they really did not make trips during (a part of) the measurement period. In a small percentage of the cases, trip legs were recorded twice, and duplications were removed.

In total 61,996 trip legs were reported in the recall survey sample. *MoveSmarter* detected 57,839 trip legs, which is 7 percent below the reported number. The rate of missing *MoveSmarter* detections could be higher if we consider that about 10 percent of the *MoveSmarter* trip legs were not accepted by the respondent, and were deleted in the recall survey. However, some of these trip legs were first removed and later added again.

As mentioned earlier, we will exclude trip legs with the purpose equal to "Change Modes" to study trip rates rather than trip leg rates. Trip legs with the purpose "Change Modes" constitute 3 percent of all trip legs. As expected, PT is the dominant mode for this type of trip leg. However, based on the number of PT trips, many access and egress trips appear to be missing in the sample. We checked whether some access or egress trips could have trip purpose "Unknown." However, trip legs with this trip purpose have a different modal split and also show a (much) longer median activity time than the median exchange time in multi-modal trips. We therefore conclude that the access and egress legs in multi-modal trips are underreported in our sample. Some individual tests have shown that *MoveSmarter* has difficulties to separate multi-modal trip legs, especially when access/egress distances and exchange times are short (typically for access/egress by foot and bicycle). Probably, such mistakes are less important for respondents or are not being noticed (as long as the main trip mode is registered) and are, therefore, not corrected in the recall survey.

Data Analysis

In this subsection, we describe the analysis of the trip data. We use success rates of correct detections to describe the quality of the mode detection. Success rates are commonly used, enabling us to compare our results with the literature. The main topic of this study though is trip rate detection. We define the trip rate as the number of trips per person per day. Trip rates are presented as sample means in which we take the average over all respondents in the sample, including those who did not report any trips (0 trips). The standard error is the standard deviation in the sample mean.

We use categorization, i.e., statistical inference based on binning of data in subsamples, to search for patterns in the data. In addition, it enables us to control for differences in demographic composition when we compare different smartphone groups. We distinguish between weekdays, trip characteristics (such as travel time, departure time, mode), smartphone groups (non-smartphone owners, Android smartphone owners, and iPhone owners), and demographics (such as age, gender, income level, education level). We use student t-test statistics with a significance level of 0.01 to indicate statistically significant (i.e., p-values ≤ 0.01) differences in mean trip rates.

To quantify underreporting, Kelly et al. (2014) and Houston et al. (2014) compared trip rates from traditional surveys with automatic trip detection. Similarly, we will compare the results from our one-day trip survey at the beginning of each wave with the recall survey. This will enable us to detect potential underreporting in traditional trip registration. However, automatic trip detection rates may also include biases. The advantage of this study is that we can additionally look for such biases by comparing the automatic trip detection with the trips from the recall survey. Such a comparison is much rarer in the literature. An additional strength is that each comparison is done with the same sample of respondents.

General Results: *MoveSmarter* versus Reported Trips

Detection of Trip Legs

Figure 2 shows the average number of trip legs per person per day (pppd). The standard errors, i.e., the standard deviations in the sample means, are similar for all weekdays and the standard error is on average 0.15 trips pppd for the reported trips, and 0.20 for the *MoveSmarter* trips. The variation is thus slightly lower for self-reported trips. This is not unexpected as respondents will correct some detection errors in the *MoveSmarter* when they report their trips. The figure shows some interesting features. The day-to-day variation is more or less similar for the reported and *MoveSmarter* trips. Both batches show some clear dips in the weekends. Especially on Sundays, rates are relatively low. These results are in line with those from NTS data. Note that in the first batch, the Monday of the last data collection week was a national holiday (Kings Day; April 27). This day clearly shows fewer trips than an average Monday, and this holiday is comparable to a typical Sunday.

The figure indicates the number of registrations is quite stable over the four-week period. If there is respondents' fatigue, this effect is limited, and only may result in a weak trend of declining registrations. This conclusion is further supported by Figure 3, which shows the number of respondents per day in the final sample, i.e., those respondents

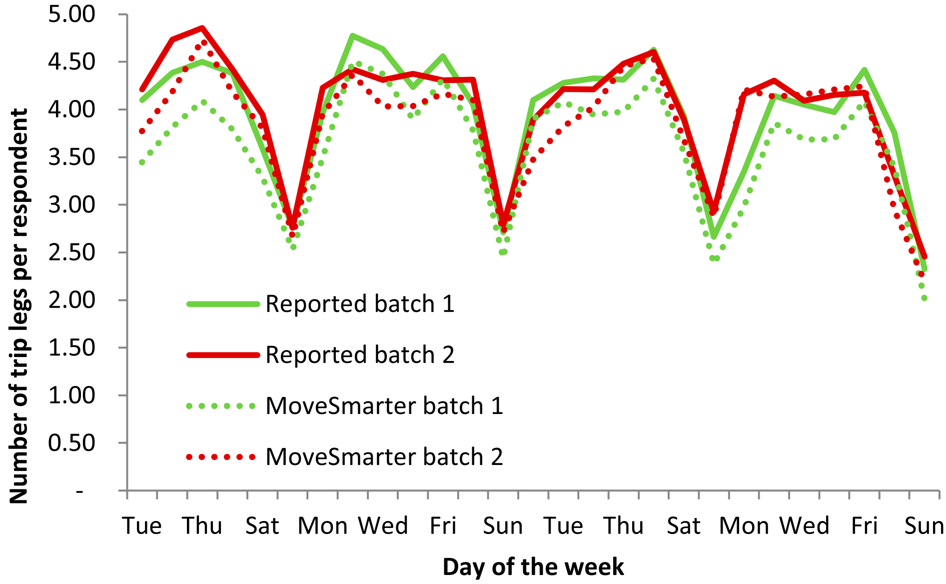

Figure 2. Average number of trip legs per person per day, 2015

who approved their trip diary for the corresponding day. According to the figure, the response is somewhat lower in the beginning (probably due to start-up problems) and at the end (probably due to the fact that some respondents did not approve the last few days after the experiment had formally ended). Although the decline of about 10 percent in the last weekend is not negligibly small, the rate of response is almost constant

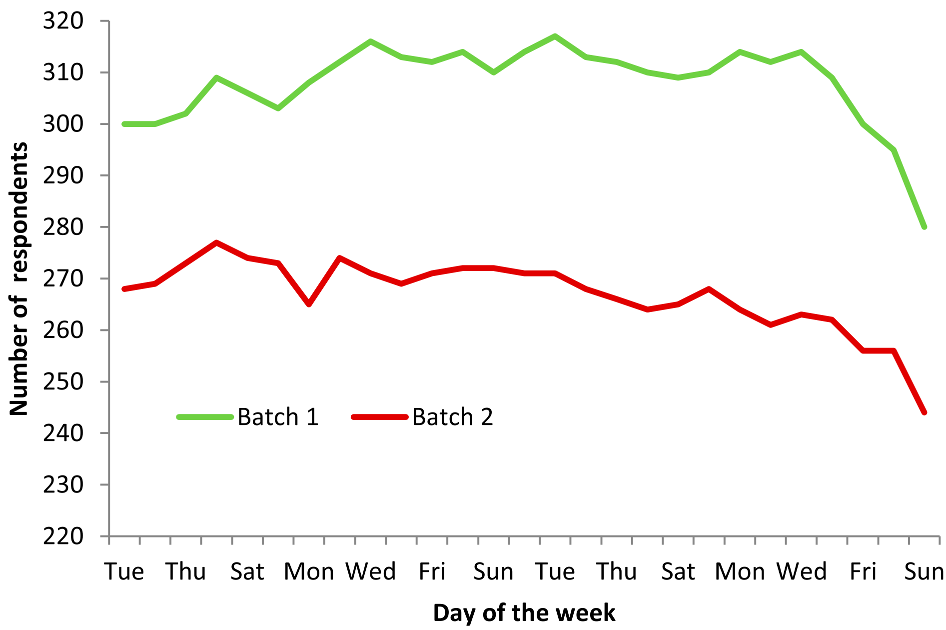

Figure 3. Number of respondents per day in the final 2015 sample

(with only a few percent of variation) during the rest of the period. These results are promising, showing that the monitoring of smartphone-based travel behavior is possible over longer periods with a reasonable respondent burden.

Mode Choice Detection

We measure the quality of the mode detection by the success rate. The success rate is equal to the number of correctly classified trips (of a given reported mode) divided by the total number of trips (of a given reported mode). Correctly classified means that the automatic detection identifies the same mode as reported in the recall survey. Here, we assume that the reported mode in the recall survey is the correct mode. The overall success rate in mode choice detection is 82 percent. Success rates are higher (84 percent) for smartphone owners, depending on trip distance. They range between 78 percent for distances below 2 km to 95 percent for distances beyond 20 km (See Table 2). These detection rates are well within the range of mode choice detection rates found in large-scale applications of GPS- and smartphone-based travel surveys in the literature.

The reason for the relatively small success rate at short distances is that it is quite hard to distinguish between car and bike when trips are short. The results thus suggest that mode detection quality depends on the types of trips being made, which can vary between study areas and respondent groups. Therefore, one should be careful when comparing mode detection quality of different studies. In addition, the quality of mode detection may also depend on the trip detection rate (e.g., Rasmussen et al., 2015). For example, the detection of mode choice is relatively easy for well-defined trip legs, i.e., for high-resolution GPS traces with a clear start and end point. However, when only well-defined trip legs are considered, and less well-defined trip legs are discarded, the actual trip rate may be underestimated (with few non-trips, but [too] many missing trips).

Underreporting

Comparison with Trip Diaries

For the same respondents, Table 3 compares the reported trip rates from the recall survey with the trip rates from the one-day trip survey (without *MoveSmarter* detections) conducted at the start (on the first Monday) of each monitoring period. For a fair comparison, we selected reported trips from the recall survey that were made on the second and third Mondays. Note that the first Monday of the first batch was Easter Monday. We, therefore, used the one-day survey from 2014 for the first batch. Results were not significantly different between the two batches (first batch in 2014 and second batch in 2015). The reported

Table 2. Percentage of successful mode detections for the 2015 wave, distinguishing between trips with different trip lengths (second column) and travel times (third column)

	Success rates per trip length bin	Success rates per travel time bin
0–2 km / 0–7 min	77%	78%
2–7 km / 7–15 min	79%	81%
7–20 km / 15–30 min	87%	85%
20–50 km / 30–60 min	94%	87%
>= 50 km / >= 60 min	95%	83%

Table 3. Trip generation to activities (without transfers). Comparison between self-reporting with and without *MoveSmarter* for a sample of 519 persons

Activity (at destination)	*MoveSmarter* and self-reporting (N trips = 3955*)	One-day trip diary (N trips = 1726)
Work	0.87	0.68
Business	0.26	0.18
Shopping / personal care	0.77	0.68
Education	0.07	0.08
Recreational	0.89	0.62
Touring / walking	0.16	0.23
Bringing / picking up	0.49	0.43
Other	0.27	0.29
Unknown	0.13	0.14
All	**3.92 ± 0.09**	**3,33 ± 0.11**

*Based on reports from the two Mondays after the first Monday (used for the one-day trip diary). These comprise 1,008 person-days (slightly less than two days per person, because in some cases a respondent had not approved his or her trips for that day).

trip rate from the recall survey (excluding trip legs with purpose "Change Modes") is about 4.1 trips per person per working day and 3.2 per weekend day. The table shows significantly ($p < 0.01$ according to a paired t-test) lower trip rates for the one-day survey. The average number of reported trips per person per day is also significantly ($p < 0.01$) higher than in the Dutch National Travel Survey (OViN) (Statistics Netherlands, 2014) and the Mobility Panel for the Netherlands (MPN) (Hoogendoorn-Lanser et al., 2014). The average trip rate from the one-day trip diary is very similar to the MPN (3.3 trips pppd). Differences between the one-day trip diary and *MoveSmarter* trip rates are relatively large for irregular trips such as business trips (47 percent larger in the recall survey compared to the one-day trip survey) and recreational trips (43 percent larger in the recall survey). However, even for commuting, more trips are reported in the recall survey (28 percent more compared to the one-day trip survey). We should note though that commuting is nowadays not always a regular trip as an increasing number of workers have the option to work at home. Moreover, some business trips may also have been registered as commute trips as the difference between the two is not always clear for respondents.

Note that the standard error in the second column of Table 3 (provided for all trips) is lower than the one noted earlier because sample sizes are different. In the comparison with the one-day trip diary, samples of both batches are used, while previously, standard errors per batch were given.

Estimating the Rate of Underreporting

The results from the previous subsection suggest that automatic trip detection (by *MoveSmarter*) may result in the registration of trips that would otherwise not be reported. However, this does not mean that all trips are detected by *MoveSmarter*. We found that some trips were added in the recall survey. If we assume that these trips are real, this implies that *MoveSmarter* is not detecting 100 percent of the trips made. Does this mean that trip detection can be improved by better algorithms, or is the quality of trip detection mainly restricted by the users and the way they use their smartphones?

Although we cannot answer this question directly, we can improve our understanding by exploring the performance of *MoveSmarter* in more detail. This was done by comparing the percentage of newly added trips in the recall survey for different trip distances, travel times, activity times (before and after the trip), modes, times of the day, and smartphone types. The results can be summarized as follows. The percentage of added trips in the recall survey is higher for very long trips (both in distance and travel time), for the last trip of the day, and for users with borrowed Android phones.

The relatively large number of added trips longer than 90 minutes or 100 km may be related to battery constraints. As very long trips have a large effect on battery consumption, they may not all be (completely) recorded by *MoveSmarter*. The performance of *MoveSmarter* also appears to decline towards the end of the day, as relatively many of those trips were added in the recall survey. An explanation may be that batteries draw empty during the day. Surprisingly, however, we do not find any relation between the battery level and the rate of added trips in the recall survey. This result is quite unexpected; it contradicts the notion that trips are not detected because the battery is empty.

Interestingly, we find that most added trips were made (according to the respondents) when battery levels did not decline (much) at all. This implies that power consumption of the smartphone was low during moments when the smartphone should actually be busy monitoring trip movements. This in turn suggests sensors were not working (properly) during these moments. There are several possible explanations for this, ranging from accidental sensor malfunctioning to respondents intentionally shutting off sensors to save power.

Finally, we looked at the differences between smartphone and non-smartphone owners. *MoveSmarter* trip rates (per person per day) are significantly ($p < 0.01$) lower for non-smartphone owners than for smartphone owners. The relatively high percentage of added trips by non-smartphone users does not compensate for this difference, as self-reported trip rates are still significantly ($p < 0.01$) lower for non-smartphone users. We cannot attribute this to the fact that there is a clear difference in demographic composition between both groups (relatively more elderly, and respondents with a low income and/or low education level have no smartphone). For each demographic group (based on gender, age, urbanization level of the residence, income level, and education level), we find a systematically (all significant; $p < 0.01$) lower trip rate for non-smartphone owners. There is a possibility that non-smartphone owners in our sample accidently made fewer trips than the smartphone owners, but it is statistically very unlikely that this would be the case for all demographic groups. We therefore suggest that differences in smartphone use (either the phone itself or the way respondents use it) is the main cause for observing different trip rates between non-smartphone and smartphone owners.

In Figure 4, we show the average *MoveSmarter* and self-reported trip rates for the different smartphone users. In addition, we show the average trip rates for different years, because trip rates were significantly ($p < 0.01$) lower in 2013 and 2014. This is also shown in Figure 4. Differences between years cannot be explained by differences in the sample compositions. Even if we consider the same respondents in 2014 and 2015, trip rates were significantly ($p < 0.01$) lower in 2014. This suggests that the registration process has been improved from 2014 to 2015. In fact, many improvements have been implemented since the start of the project, both with regard to trip registration and mode recognition. Note that the standard errors are of the same order of magnitude as

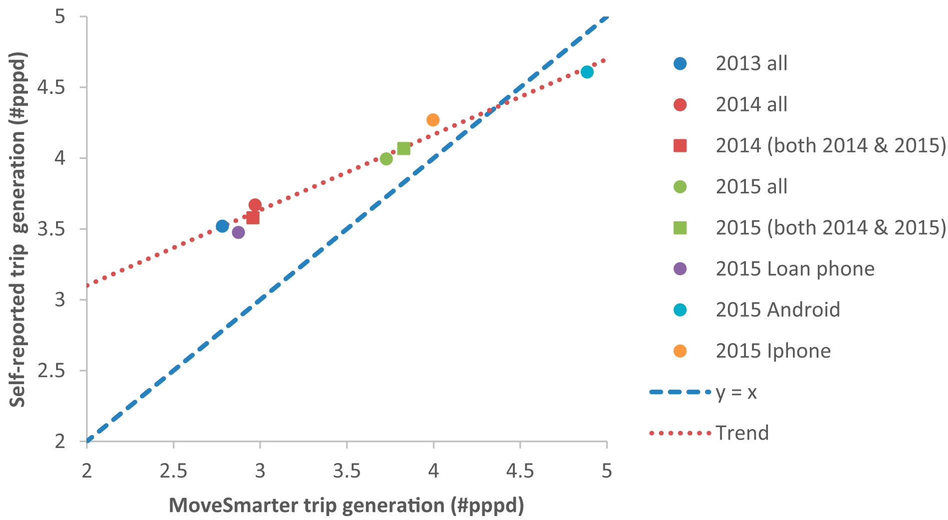

Figure 4. Self-reported trip rates versus trip rates registered by *MoveSmarter*

the symbol size in Figure 4, i.e., slightly less than 0.1 pppd for the complete samples and slightly larger than 0.1 for the smartphone group subsamples.

Figure 4 shows that there is strong positive correlation between the average *MoveSmarter* trip rates and self-reported trip rates, which is as expected *a priori*. However, the relative difference between both trip rates is not constant. This is not self-evident. For a given quality of the *MoveSmarter* registration, one would expect a certain percentage of missing detections that are added in the recall survey. In other words, if the quality of the *MoveSmarter* registration would be the same for all groups and/or years, one would expect the relative difference to be constant between self-reported and *MoveSmarter* trip rates.

In contrast, we observe that when fewer trips are registered by MoveSmarter, relatively *more* trips are added in the recall survey. This suggests that the registration rate of real trips is not the same for different groups and/or years. Unfortunately, it is not trivial to estimate the quality of the trip detection (and from that the real trip rate) directly, because real trip rates can also vary between respondent groups. However, it is quite reasonable to assume that the average real trip rate hardly changes between successive years, especially when the samples contain the same respondents and they are large enough to even out individual changes in behavior. If we assume that real trip rates are similar for the different years, we can interpret Figure 4 as follows.

MoveSmarter misses relatively many trips for groups and/or years on the left side of Figure 4. Many of these trips are reported in the recall survey. However, we expect that respondents can probably not recall all missing trips, implying self-reported trips remain underreported. As the detection rate improves, net rates of missing trips decline, and therefore fewer trips need to be added in the recall survey. This is shown by the dotted line in Figure 4. Where this line intersects the line y = x, the net rate of missing *MoveSmarter* detections would be zero. This would be a good indication of the real trip rate, which is slightly less than 4.5 trips pppd according to Figure 4. Note that the net rate of missing trips is the difference between the rate of missing trips (actual trips that were not detected) and non-trips (detections that were actually not real trips).

If the net rate is zero, the rates of missing trips and non-trips are in balance, and there is no net increase in the number of reported trips in the recall survey. If the rate of non-trips would be actually higher than the rate of missing trips, respondents need to remove more trips than they add. This could in theory lead to overreporting, where respondents forget to remove trips that they had actually never made.

We conclude that we can estimate the real trip rate if we assume that real trip rates are the same for the samples in Figure 4. We think that this assumption is reasonable when we compare the same respondents in different years. Interestingly, Figure 4 shows some scatter with respect to the dotted line if we consider the different smartphone groups. This may suggest some variation in real trip rates between those groups. For example, non-smartphone and iPhone owners lie respectively below and above the dotted line, suggesting that real trip rates are lower for the former group and higher for the latter group. However, the scatter is too small to draw conclusions that are statistically significant.

Conclusions and Discussion

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 (more than 600 respondents), representativeness of the sample for the Dutch population (by age, smartphone ownerships, etc.), and the web-based prompted recall survey. An in-depth comparison between automatic detections and reported trips was conducted for a sample of over 60 thousand trips.

The quality of the mode choice classification is comparable to that in other smartphone-based studies but, in contrast to most of these studies, it is based on a large-scale application which makes it more difficult to achieve high success rates. In this study success rates range between 78 percent for distances less than 2 km to 95 percent for distances more than 20 km. The relatively small success rate for short trips can be attributed to the fact that it is quite hard to distinguish between car and bike when trips are (very) short.

We assume that the quality of mode detection also depends on trip detection rates, because the quality of the data is not the same for all detections. For example, Rasmussen et al. (2015) used different limits for their trip detection algorithm. When they applied relaxed limits, they detected more trips, but the overall quality of the trip detection decreased as a result. Importantly, while the number of missing trips was reduced, the number of non-trip detections increased. In other words, there is a trade-off between the number of missing trips and non-trips. However, it is quite hard to estimate those numbers directly based on reports from respondents, as there are also biases in reported trip rates in the recall survey. Underreporting is a well-known problem, and we have shown that respondents probably do not recall all missing trips. Based on this observation, we have argued that the correct trip rate (where the numbers of missing and non-trips are equal) most likely is observed when *MoveSmarter* and self-reporting trip rates are more or less equal. This is the case for a trip rate of about 4.5 trips per person per day.

The difference between 4.5 trips per person per day and trip rates from Dutch Travel Surveys is quite high, and the difference is higher than reported in the literature. Kelly et al. (2014) and Houston et al. (2014) for example found an underreporting of trips by 15 percent and 0.39 trips per person per day, respectively. However, rates in underreporting are not always the same, and we have confirmed that differences occur between groups of respondents (e.g., Bricka et al., 2012). Comparisons between studies should, therefore, be treated carefully. An important difference with aforementioned studies is also that in those studies comparisons are made between automatic detection and an independent survey. Interestingly, the difference between *MoveSmarter* and our independent survey (on the first Monday of each batch) is more or less 0.4 trips (3.7 versus 3.3) or about 15 percent. However, we have argued that the automatic trip detection itself can also show a bias, which can be corrected for by a recall survey. Earlier smartphone studies that require user validation involve filling in missing information and amending incorrectly inferred data about, for example, trip purpose and modes of travel used (e.g., Safi et al., 2016; Zhao et al., 2015), but to our knowledge, respondents could not add undetected trips. Note from Safi et al. (2016) we can derive an observed trip rate of 4.4 trips per person per day (1,873 recorded trips for 424 travel days according to their Table 3) for their sample in New Zealand, which is quite comparable with our result for The Netherlands.

This paper confirms that underreporting of trips is a substantial problem in self-completion trip-based diaries. Underreporting occurs for all trip motives, although the effect is less severe for commute trips as compared to recreational trips. We also found that underreporting is not only a problem for short trips. Although many short trips are missing in NTS data, long trips are also underreported in traditional surveys. One of the main conclusions is that underreporting mostly occurs for non-regular trips that people most likely forget to report in traditional surveys.

We conclude that automatic detection with the *MoveSmarter* app improves the quality of travel surveys and allows measurements of multi-week trip data with a relatively low respondent burden. However, a prompted recall procedure remains necessary to achieve accurate travel behavior data. It is also useful as comparison to improve the automated trip detection algorithms. The Dutch Mobile Mobility Panel project started in 2013 and since then many improvements in the quality of the trip detection have been made, which (partially) can be attributed to the set-up of this study.

An interesting finding from our study is that the quality of automated trip detection varies significantly among groups of respondents. There are indications that the relatively high rate of underreporting among respondents who were given a loan-smartphone (e.g., older adults) can be attributed to misuse of the smartphone. Given that smartphone ownership is rapidly increasing in the Netherlands and many other countries, smartphone-based data collection will likely become more prominent in travel behavior research.

Note

1. Smartphone ownership in the LISS panel in 2013 is lower compared to surveys based on volunteer access panels in the Netherlands conducted to measure smartphone ownership. The LISS panel is based on a probability sample of the target population and representativeness is guaranteed by providing Internet devices and connections to people without Internet (in particular elderly people). Volunteer access panels, in which respondents self-select themselves into the sample, are attractive due to their low recruitment and maintenance costs but their quality in terms of representativeness and measurement error, however, remains questionable (Blom et al., 2015). They are likely biased towards high shares of smartphone use. Smartphone ownership among elderly has, however, increased rapidly since 2013.

Acknowledgments

This project is financed by the Netherlands Organization 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 Salima Douhou and Maarten Streefkerk from CentERdata and Martin Wibbels from Mobidot for their contributions to this project. In addition, we would like to thank CentERdata for the use of the LISS panel for this research.

Notes on Contributors

Tom Thomas is an assistant professor at the Center for Transport Studies at the University of Twente.

Karst T. Geurs is a professor at the Center for Transport Studies at the University of Twente.

Johan Koolwaaij is a co-founder and software developer of *Mobit*.

Marcel Bijlsma is a co-founder and director of *Mobit*.

Bibliography

F. Biljecki, H. Ledoux, and P. van Oosterom, "Transportation Mode-Based Segmentation and Classification of Movement Trajectories," *International Journal of Geographical Information Science* 27: 2 (2013) 385–407.

A. Blom, M. Bosnjak, A. Cornilleau, A.-S. Cousteaux, M. Das, S. Douhou, and U. Krieger, "A Comparison of Four Probability-Based Online and Mixed-Mode Panels in Europe," *Social Science Computer Review* 34: 1 (2015) 8–25.

W. Bohte and K. Maat, "Deriving and Validating Trip Purposes and Travel Modes for Multi-Day GPS-Based Travel Surveys: A Large-Scale Application in the Netherlands," *Transportation Research Part C* 17: 3 (2009) 285–297.

S. Bricka, S. Sen, R. Paleti, and C. Bhat, "An Analysis of the Factors Influencing Differences in Survey-Reported and GPS-Recorded Trips," *Transportation Research Part C* 21 (2012) 67–88.

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

T. Feng and H. Timmermans, "Transportation Mode Recognition Using GPS and Accelerometer Data," *Transportation Research Part C* 37 (2013) 118–130.

S. Ferrer and T. Ruiz, "Travel Behavior Characterization Using Raw Accelerometer Data Collected From Smartphones," *Procedia: Social and Behavioral Sciences* 160 (2014) 140–149.

H. Gong, C. Chen, E. Bialostozky, and C. Lawson, "A GPS/GIS Method for Travel Mode Detection in New York City," *Computers, Environment and Urban Systems* 36 (2012) 131–139.

K. Geurs, T. Thomas, M. Bijlsma, and S. Douhou, "Automatic Trip and Mode Detection with MoveSmarter: First Results from the Dutch Mobile Mobility Panel," *Transportation Research Procedia* 11 (2015) 247–262.

S. Hoogendoorn-Lanser, N. Schaap, and M.-J. Olde Kalter, "The Netherlands Mobility Panel: An Innovative Design Approach for Web-Based Longitudinal Travel Data Collection," paper presented at the 10th International Conference on Transport Survey Methods (Leura Australia, 16–21 November 2014).

D. Houston, T. Luong, and M. Boarnet, "Tracking Daily Travel; Assessing Discrepancies Between GPS-Derived and Self-Reported Travel Patterns," *Transportation Research Part C* 48 (2014) 97–108.

P. Kelly, A. Doherty, A. Mizdrak, S. Marshall, J. Kerr, and A. Legge, "High Group Level Validity But High Random Error of A Self-Report Travel Diary, As Assessed by Wearable Cameras," *Journal of Transport & Health* 1: 3 (2014) 190–201.

J. Liu, H. Zhenga, T. Feng, S. Yuan, and H. Lu, "Post-Processing Procedures for Passive GPS Based Travel Survey," paper presented at the 13th COTA International Conference of Transportation Professionals (CICTP 2013).

P. Nitsche, P. Widhalm, S. Breuss, and P. Maurer, "A Strategy on How to Utilize Smartphones for Automatically Reconstructing Trips in Travel Surveys," *Procedia: Social and Behavioral Sciences* 48 (2012) 1033–1046.

P. Nitsche, P. Widhalm, S. Breuss, N. Brändle, P. Maurer, "Supporting Large-Scale Travel Surveys with Smartphones: A Practical Approach," *Transportation Research Part C* 43 (2014) 212–221.

A. Prelipcean, G. Gidófalvi, and Y. Susilo, "Mobility Collector," *Journal of Location Based Services* 8: 4 (2014) 229–255.

A. Prelipcean, G. Gidófalvi, and Y. Susilo, "Transportation Mode Detection: An In-Depth Review of Applicability and Reliability," *Transport Reviews* 37: 4 (2017) 442–464.

T. Rasmussen, J. Ingvardson, K. Halldórsdóttir, and O. Nielsen, "Improved Methods to Deduct Trip Legs and Mode From Travel Surveys Using Wearable GPS Devices: A Case Study From the Greater Copenhagen Area," *Computers, Environment and Urban Systems* 54 (2015) 301–313.

A. Raza, L. Knapen, K. Declercq, T. Bellemans, D. Janssens, and G. Wets, "Diary Survey Quality Assessment Using GPS Traces," *Procedia Computer Science* 52 (2015) 600–605.

S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Using Mobile Phones to Determine Transportation Modes," *ACM Transactions on Sensor Networks* 6: 2 (2010) Article 13.

K. Reinau, H. Harder, and M. Weber, "The SMS–GPS-Trip Method: A New Method for Collecting Trip Information in Travel Behavior Research," *Telecommunications Policy* 39 (2015) 363–373.

N. Rieser-Schüssler and K. Axhausen, "Self-Tracing and Reporting: State of the Art in the Capture of Revealed Behaviour," in S. Hess and A. Daly, eds, *Handbook of Choice Modelling* (Cheltenham: Edward Elgar, 2014).

M. Ribeiro, A. Larrañaga, J. Arellana, and H. Cybis, "Influence of GPS and Self-Reported Data in Travel Demand Models," *Procedia - Social and Behavioral Sciences* 162 (2014) 467–476.

H. Safi, B. Assemi, M. Mesbah, and L. Ferreira, "Trip Detection with Smartphone-Assisted Collection of Travel Data," *Transportation Research Record* 2594 (2016) 18–26.

A. Scherpenzeel and M. Das, "True Longitudinal and Probability-Based Internet Panels: Evidence from the Netherlands," in M. Ester and L. Kaczmirek, eds, *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies* (Boca Raton: Taylor & Francis, 2010).

S. Schönfelder and K. Axhausen, *Urban Rhythms and Travel Behavior: Spatial and Temporal Phenomena of Daily Travel* (London: Routledge, 2010)

I. Semanjski, "Potential of Big Data in Forecasting Travel Times," *Traffic & Transportation* 27: 6 (2015) 515–528.

I. Semanjski and S. Gautama, "Smart City Mobility Application: Gradient Boosting Trees for Mobility Prediction and Analysis Based on Crowdsourced Data," *Sensors* 15: 7 (2015) 15974–15987.

L. Shen and P. Stopher, "Review of GPS Travel Survey and GPS Data-Processing Methods," *Transport Reviews* 34: 3 (2014) 316–334.

D. Shin, D. Aliaga, B. Tunçer, S. Arisona, S. Kim, D. Zünd, and G. Schmitt, "Urban Sensing: Using Smartphones for Transportation Mode Classification," *Computers, Environment and Urban Systems* 53 (2015) 76–86.

Statistics Netherlands, *Onderzoek Verplaatsingen in Nederland*, 2014 (2014 National Travel Survey).

J. Toole, S. Colak, B. Sturt, L. Alexander, A. Evsukoff, and M. González, "The Path Most Traveled: Travel Demand Estimation Using Big Data Resources," *Transportation Research Part C* 58 (2015) 162–177.

F. Witlox, "Evaluating the Reliability of Reported Distance Data in Urban Travel Behaviour Analysis," *Journal of Transport Geography* 15: 3 (2007) 172–183.

G. Xiao, Z. Juan, and C. Zhang, "Travel Mode Detection Based on GPS Track Data and Bayesian Networks," *Computers, Environment and Urban Systems* 54 (2015) 14–22.

F. Zhao, F. Câmara Pereira, R. Ball, Y. Kim, Y. Han, C. Zegras, and M. Ben-Akiva, "Exploratory Analysis of a Smartphone-Based Travel Survey in Singapore," *Transportation Research Record* 2494 (2015) 45–56.